The Pennsylvania State University
The Graduate School
Eberly College of Science

THREE ESSAYS ON QUANTITATIVE FINANCE

A Dissertation in
Mathematics
by
Jun Ni

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2018

The dissertation of Jun Ni was reviewed and approved* by the following:

Jingzhi Huang
David H. McKinley Professor of Business, Professor of Finance and Professor of Mathematics
Dissertation Advisor, Chair of Committee

Xiantao Li
Professor of Mathematics

Anna Mazzucato
Professor of Mathematics
Chair of Graduate Program

Runze Li
Verne M. Willaman Professor of Statistics

*Signatures are on file in the Graduate School.

Abstract

This dissertation contains three essays.

The first part studies the continuous-time dynamics of VIX with stochastic volatility and jumps in VIX and volatility. Built on the general parametric affine model with stochastic volatility and jumps in the logarithm of VIX, we derive a linear relationship between the stochastic volatility factor and the VVIX index. We detect the existence of a co-jump of VIX and VVIX and put forward a double-jump stochastic volatility model for VIX through its joint property with VVIX. Using the VVIX index as a proxy for stochastic volatility, we use the MCMC method to estimate the dynamics of VIX. Comparing nested models of VIX, we show that the jump in VIX and the volatility factor are statistically significant. The jump intensity is also stochastic. We analyze the impact of the jump factor on VIX dynamics.

The second part establishes a forecast framework for the bond excess return based on macroeconomics fundamentals. Empirical evidence has suggested that excess bond returns are forecastable with macroeconomics fundamentals. In our study, we build new links to tie the forecastable variation in excess bond returns to underlying macroeconomic series. Based on two types of models, the linear model and additive model, and utilizing different combinations of screening methods, nonlinearization techniques and regularization techniques, we extract different factor combinations from 131 macroeconomic series, including employment, housing, financial, and inflation factors. This approach results in stronger forecast power for the excess bond returns compared with existing macro-based return predictors. The nonlinear effect of the macroeconomic predictors on the excess bond returns is recovered if we incorporate nonlinearized macro data in the analysis. A horse race comparing different variable selection approaches allows us to propose a robust model that generates highly accurate predictions of bond risk premia. Finally, we perform a comprehensive analysis of risk premia with an ETF dataset.

The third part of this dissertation is a summary of traditional asset allocation methods performance on Chinese market. Since traditional asset allocation methods

are well analyzed in US capital market, similarly, we want to conduct a comprehensive analysis of asset allocation techniques on Chinese market. Based on a horserace comparison among the trading performance by different asset allocation approaches with investment universe of Chinese capital market indices and the associated ETFs, we achieve a clear understanding on the relative ranking of different methods, finding the link between trading performance with different parameter estimation time windows and different investment universe as well. To explain the difference in the trading performance of several methods, we perform a simulation study and attribute bad performance as the inaccuracy of return estimation.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to express my sincere gratitude to my thesis advisor Professor Jingzhi Huang for the continuous support throughout my Ph.D. life and research, for his patient guidance, motivation, and immense knowledge, but also for his valuable advice on my career development. I started to work with Professor Huang from my third year here, it is Professor Huang who gave me a future, without his incredible patience and timely help, I am not the one today. Besides my advisor, I am also indebted to my committee members. I would like to thank Professor Anna Mazzucato, Professor Xiantao Li and Professor Runze Li, for their kindness to serve on my dissertation committee, as well as for their helpful comments and suggestions on my dissertation. Without their motivation and valuable advice, the thesis work would not have been successful. I would also thank Professor Qiang Du who brought me into the world of research, showed me the beauty of math and gave me lots of patient guidance during my early years study. Same appreciation would be given to Professor Yuxi Zheng, who gave me tons of support and courage on my Ph.D. life here. In addition, I want to thank all the people who helped me throughout my academic exploration. Last but not the least, I would like to thank my parents for their unconditional love and support during my whole life.

Dedication

To my parents, Jie Zhao and Yanhui Ni.

Chapter 1 Introduction

The main discussion of finance can be summarized as the research on the relation between risk and return. Some concern on how to model the risks, and others discuss on the predictions of returns or excess returns In this dissertation we will try to delve into the two financial aspects.

1.1 VIX Volatility Modelling

Investors on stock make decisions based on their opinions of whether the stock price goes up or down, bond investors make trading decisions based on their opinions on the interest rate. Similarly, if investors have some sense on the prediction of the volatility of some asset class, they can also profit from some trading activity. A kind of classical trading strategy is to trade some asset as well as the option on this underlying with the method of delta hedging. Unfortunately, there are two drawbacks: firstly, the risk not only comes from the direction of the volatility changing, but also from the price change of the underlying; secondly, delta hedging is based on the assumption of Black-Scholes model, where it assumes that there is no friction on the market, absolute liquidity, continuous trading and constant volatility, which can not be satisfied in the practical market.

Although the strategy above is not perfect, fortunately, the market provides us some more financial products to trade volatility, that is the volatility swap. Let's introduce the concepts of realized variance and realized volatility. For a probability space which satisfies the usual conditions, $\{\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P}\}$, where there is a semimartingale $\mathcal{S}$ defined to describe the process of asset prices. Assume within the interval $[t, t+\tau]$, there are $n+1$ trading days, $t = t_0 < t_1 < \cdots, < t_n = t+\tau$,

then the realized annualized variance defined within the interval $[t, t+\tau]$ is

$$RV_{t,t+\tau} = \frac{252}{n}\sum_{i=1}^{n}(\log\frac{S_{t_i}}{S_{t_{i-1}}})^2 \tag{1.1}$$

realized annualized volatility is just the square root of the realized annualized variance. The volatility swap on a stock is just a forward contract with the underlying of its realized annualized volatility, the payment when it is matured is

$$(\sigma_R - K_{vol}) * N$$

where σ_R is the realized annualized volatility calculated with equation 1.1, K_{vol} is the strike price determined in advance, N is the nominal price associated with 1 point of the annualized volatility. When the contract matures, the investor who purchases the contract will receive N dollars for every point where the realized annualized volatility σ_R exceeds the strike price K_{vol}. Namely, the investor actually is trying to exchange the future realized volatility σ_R with a fixed volatility level K_{vol}. In reality, much earlier than volatility swap, a usual contract is variance swap. Similar to the payment of volatility swap, variance swap has the payment

$$(\sigma_R^2 - K_{var}) * N$$

similarly, σ_R^2 is the realized annualized variance within the contract life, K_{var} is the strike price for the annualized variance. Let v_t and J_t be the instantaneous volatility and jump size of the asset return, if let $n \to \infty$ in (1.1), for a general semimartingale $\mathcal{S}$, that will converge to the quadratic variance of log prices in probability, that is

$$\frac{252}{n}\sum_{i=1}^{n}(\log\frac{S_{t_i}}{S_{t_{i-1}}})^2 \to \frac{1}{\tau}\int_t^{t+\tau} v_s ds + \frac{1}{\tau}\sum_{t\le s\le t+\tau} J_s^2 = QV_{t,t+\tau} \tag{1.2}$$

Under risk neutral measure Q, let $VS_{t,t+\tau}$ be the price of the variance swap within the time interval $[t, t+\tau]$, then

$$VS_{t,t+\tau} = E_t^Q[QV_{t,t+\tau}]$$

if we fix t and make τ change, we will get a term structure of the variance swap at

time t. After 2005, more derivatives of volatility are provided, including the option on realized variance, corridor variance swap, gamma swap, timer option and others.

The prosperity of volatility market made Chicago Board Options Exchange (CBOE) provide the first volatility index in 1993. which is well known as VIX index. As a measure of the market expectations for the 30-day implied volatility of the S&P500 index, VIX provides rich information for the prediction of future market trends. VIX can be seen as a compression of the information involved in S&P500 options. Usually, VIX and the S&P500 index are negatively correlated, the larger the VIX index, the volatile the market is and the more fearful the investors are, meaning that the VIX index is often referred to as the fear index or the fear gauge. When VIX comes to market at first, VIX calculation is the weighted average of some short call and put options on S&P100 index, in 2003, to describe the expect on the future volatility of the market accurately, CBOE changed its calculation on VIX index with the incorporation of options on S&P500. Under the new definition, VIX index at time t is defined as

$$VIX^2(t,T) = \frac{2}{T-t}\sum_i \frac{\Delta K_i}{k_i^2}\exp r_t(T-t)O_i(K_i,T) - \frac{1}{T-t}[\frac{F_t}{K_0}-1]^2$$

where T is the common maturity of all the options used, F_t is the forward price of S&P500 index, K_i is the strike price of the ith out of money option, $O_i(k_i,T)$ is the mid price of ith option, K_0 is the first strike price below F_t, r_t is the risk free rate with the time interval from t to T, Δ is the difference between to close strike prices, define $\Delta K_i = (K_{i+1} - K_i)/2$. Since then, the futures and options with the underlying of VIX come to the market, and the trading is more and more active. Similar to the calculation of VIX index, CBOE also calculate another index VVIX with the options on VIX, to describe the implied volatility of VIX, which is called "volatility of volatility" in academy. The volatility derivatives enrich the market, and also stimulate research work in academy.

There is a wide research on the volatility derivatives, this section will mainly summarize the research on variance swap, VIX index and derivatives on VIX. [Gatheral 2011] is an important paper on volatility models, [Carr and Lee 2009] conducts an extensive effort to describe the development and evolvement history of volatility derivatives market, by summarizing part of the literatures before 2009, the author clarifies some basic mathematical theory which serves as the basis of

this research area.

[Demeterfi et al. 1999] describes the basic mathematical properties of variance swap and volatility swap, and explains how to replicate and hedge the variance swap theoretically. [Carr and D. Madan 1998] shows three methods on realized volatility (replication and hedging).

There is also an extensive research result on the statistical properties on realized variance and realized volatility, including [McAleer and Medeiros 2008; Andersen, Bollerslev, et al. 2003; Barndorff-Nielsen and Shephard 2002b; Barndorff-Nielsen and Shephard 2002a; Bollerslev, Gibson, and Zhou 2011; L. Zhang, Mykland, and Aït-Sahalia 2011; Christoffersen, Jacobs, and Mimouni 2010] and [Andersen and Teräsvirta 2009].

For the pricing of variance swap term structure, some literatures give the results based on some specific assumptions. [Egloff, Leippold, Wu, et al. 2010] and [Ait-Sahalia, Karaman, and Mancini 2014] assume v_t follows an affine process in [1.1]. [Filipović, Gourier, and Mancini 2016] assumes $VS_{t,t+\tau} = \frac{1}{\tau}G(\tau, X_t)$, where X_t satisfies a second order linear multi-factor model, G is a second order functional. [Amengual and Xiu 2014] assumes v_t follows a nonaffine model. [Swishchuk 2004] provides some new analytical methods to price variance swap and volatility swap.

Many other papers are trying to talk about the problem of pricing and hedging for variance swap and derivatives with the underlying of realized volatility and realized variance, of course, under different model assumptions. [Broadie and Jain 2008a] assumes the volatility of asset prices follows Heston stochastic volatility model. [Broadie and Jain 2008b] shows the impact of asset price jumps on the pricing of variance and volatility swaps, and gives a comprehensive comparison on different nested models. [Elliott, Kuen Siu, and Chan 2007] talks about the pricing and hedging under the model of Heston stochastic volatility model and regime switching model. [Bernard and Cui 2014] considers the discretely hedging and its asymptotic property under the set up of general time homogeneous model. [Jordan and Tier 2009] considers the variance swap pricing problem under the assumption that asset prices follow a constant elasticity of variance model. [S.-P. Zhu and Lian 2011] exploits PDE techniques to price the discretely sampled variance swap under Henson model set up. [Carr, Lee, and Wu 2012] assumes the asset prices follow a Levy process with a time transformation, and comes to the conclusion that the variance swap price can be determined by the product of log contract and a

multiplier, the multiplier is only determined by the property of the Levy process, and having nothing to do with the time transformation. [Hobson and Klimmek 2012] considers a super-replicate strategy with a general asset prices model. [Jarrow et al. 2013] gives some theoretical requirement for the hold of equation (1.1).

Some other papers talked about the pricing problem on the options with underlying of realized variance and realized volatility, including [Carr, Hélyette Geman, et al. 2005; Carr and Lee 2007; G. G. Drimus 2012; Carr, Helyette Geman, et al. 2011; Itkin and Carr 2010; Crosby and Davis 2012; Di Graziano and Torricelli 2012; Kallsen, Muhle-Karbe, and Voß 2011] and [Sepp 2008a]. [Carr and Lee 2008] utilizes the technique of integral transform, generalizes the pricing and hedging on variance swap to a product with the payment of a function of realized variance. [Friz and Gatheral 2005] gives some more detailed analysis for the contents in [Carr and Lee 2008], and shows two call option examples on volatility swap and realized variance. [Hörfelt et al. 2012] talks about covariance swap and [Schoutens 2005] discusses on moment swap.

Besides the white paper for VIX index by CBOE, [Carr and Wu 2005] gives an extensive description of the history and basic properties of VIX index. Based on the white paper and [Carr and Wu 2005], [Whaley 2008] clarifies the role of VIX in the capital market. [Jiang and Tian 2007] talks about how to construct VIX index more accurately. [Gonzalez-Perez 2015] shows the impact of VIX index in financial literature, summarizes the benefits and drawbacks of VIX in some important fields. In building a parameterized stochastic model for VIX, there are two typical starting points. The first starting point is to model a multi-factor stochastic volatility process for the S&P500 index and then to derive a calculation formula for the VIX index under these circumstances. For more on multi-factor model setup, see, for example, [Duffie, Pan, and K. Singleton 2000; Gatheral 2008a; Egloff, Leippold, Wu, et al. 2010; Cont and Kokholm 2013] and [Papanicolaou and Sircar 2014]. Based on this model's assumption, the final VIX may be a combination of one or more factors (see e.g. [Ait-Sahalia, Karaman, and Mancini 2014; Song and Xiu 2012; Lin and Chang 2009] and [Luo and J. E. Zhang 2012]. The second starting point is to directly model the VIX index, the models are different, but often have a mean-reverting property. For example, [Mencía and Sentana 2013; Kaeck and Alexander 2013] and [Goard and Mazur 2013]. Others talk about the significance and statistical properties on the jump of volatility, like [Todorov and Tauchen 2011;

Khagleeva 2012] and [Du and Kapadia 2012].

Similar to the term structure of variance swap, [Luo and J. E. Zhang 2012] and [Johnson 2017] talk about the term structure of VIX index. As for VIX option and future modeling, under different model assumption, there are a lot of works. [Lin and Chang 2009] gives the option pricing model firstly, [Dupoyet, Robert T Daigler, and Chen 2011; J. E. Zhang and Y. Zhu 2006; S.-P. Zhu and Lian 2012; Lin 2007; Y. Zhu and J. E. Zhang 2007] consider the pricing model for VIX futures. [Huskaj and Nossman 2013] and [Z. Lu and Y. Zhu 2010] try to work on VIX term structures. [Shu and J. E. Zhang 2012] shows the price transform in VIX futures. [Asensio 2013] shows some puzzles in practical VIX future market. [Marabel Romo 2017] clarifies a double stochastic volatility model. [Z. Wang and Robert T Daigler 2011] examines the performance of several available option pricing models. [Psychoyios, Dotsis, and Markellos 2010] raises a jump diffusion model to price VIX option and futures. [Sepp 2008a] and [Sepp 2008b] give a pricing approach by PDE and generalized Fourier Transform. [Lian and S.-P. Zhu 2013] generates a analytic solution for option pricing by integral transform and asymptotic method. [Cheng et al. 2012] shows the bias of pricing formula in [Lin and Chang 2009] and [Lin and Chang 2010]. [Lin 2013] simplifies the VIX option pricing with the information of VIX term structure and VIX futures. [Völkert 2015] extracts the VIX risk neutral distribution from VIX option prices and analyzes its impact on the whole economy. [Lo et al. 2013] tries to incorporate jumps and other factors to improve the model for volatility.

Since the fact that VIX index is calculated from the weighted average of a series of S&P500 option prices, between VIX index associated with its derivatives, and S&P500 index associated with its derivatives, i.e., between VIX market and S&P500 market, there must be some intrinsic relationship. How to model the relation is a main consern in academy as well. [Branger and Voelkert 2013; Gatheral 2011; Gatheral 2008b] and [Gatheral 2007] try to model VIX and S&P500 adjointly from the basis of multi-factor model. [Song and Xiu 2012] considers the characteristics of two markets from the point of state price density. [Papanicolaou and Sircar 2014] conducts analysis adjoint model in the framework of regime-switch. [D. B. Madan and Yor 2011] assumes log of S&P500 follows a Sato process, its changing velocity is proportional to VIX index. [Carr and D. B. Madan 2014] utilizes double-side gamma model to describe VIX, and assumes log of stock prices follows a

variance gamma process. [Lin and Chang 2010] suggests to model stock price and its volatility with double-jump models. [Amengual and Xiu 2012] examines and compares many different models performance on S&P500 indix and its volatility data. [Bardgett, Gourier, and Leippold 2013] and [Chung et al. 2011] compares some common information and individual information owned by the two markets.

Similar to VIX, which is extracted from a weighted average of S&P500 index options, and describes the implied volatility of S&P500 market, if we utilize the same method to VIX options, and extract a series of values which represent the implied volatility of VIX index, that is, the volatility of volatility, the well known VVIX index. [D. Huang and Shaliastovich 2014] tries to describe the volatility level by the realized volatiltiy of VIX, shows it has a significant neagtive risk premia. [Barndorff-Nielsen and Veraart 2013] suggests a probability model of the volatility of volatility, gives a variance risk premia as well. Empirical analysis on the volatility of volatility can also refer [Park 2013] and [Z. Wang and Robert T. Daigler 2012].

As for modeling of volatility of volatility, [Mencía and Sentana 2013] and [Kaeck and Alexander 2013] are two main literature. They both model VIX directly, and try to suggest a reasonable model from the point of empirical analysis. To better fit the real data, directly modelling the logarithm of VIX is better substantiated empirically than modelling on VIX level directly. They both suggest the logarithm of VIX satisfy a stochastic volatility model, and try to prove the benefit on description of data characteristics by incorporating the stochastic volatility to the models, where [Kaeck and Alexander 2013] tries to model the volatility of VIX with a square root diffusion model, [Mencía and Sentana 2013] uses a pure jump Ornstein-Uhlenbeck(OU) process to model volatility. [Curato et al. 2012] gives a nonparametric estimator of covariance of asset prices and volatility series. [R. Wang, Kirby, and S. P. Clark 2013] discusses about the relationship among the risk premia of stocks, variance and volatility of volatility. [G. Drimus and Farkas 2013] conducts the modelling of VIX with a mean-revision volatility of volatility. [Song 2012] analyzes volatility jumps as well as the risk premia.

In our study of Chapter 2, we will make an extensive empirical analysis of the dynamics of VIX, concentrating particularly on modeling its stochastic volatility under the physical measure via additional information provided by the VVIX index. Our contribution consists of three aspects. First, we find evidence of the co-jumps between the VIX and VVIX index through statistical test of the historical data

of both indices. Second, we show that the VVIX index and the volatility of VIX satisfy the criteria of a linear relationship under the general affine assumption on the dynamics of the logarithm of the VIX index and its stochastic volatility. Thus the modeling of VIX and its volatility could be transformed into the joint modeling of the VIX and VVIX index. Empirically, both of VIX and VVIX are mean-reverting and have co-jumps. Based on these facts we propose a double-jump stochastic volatility model for the VIX and its volatility. Third, we provide a Markov-Chain-Monte-Carlo (MCMC) method to estimate the double-jump model and its nested models using historical data regarding VIX and VVIX. We obtain both a unified set of model parameters and a series of outcomes of latent variables such as stochastic volatility, jump intensity and jump sizes. The results can be exploited to understand the economics of the market further. We compare the model performance through several criterion such as residual analysis, p-value and deviance information criterion (DIC) method. We show that the jumps in volatility of VIX is statistically significant and the jump intensity is not deterministic which may imply a more complex structure of it.

1.2 Bond Risk Premia Prediction

Evidence from recent empirical research in financial economics has supported and verified the forecastability in the excess returns of the U.S. Treasury bonds based on some financial and macroeconomic variables. Previous literature mainly focus on constructing forecast factors from combination of forward rates, spreads, yield curve principal components and macro-series principal components, or just use cointegration theory to identity cycle factors. For instance, [Fama and Bliss 1987] state the forecastability of excess bond returns by the spread between the n-year forward rate and one-year yield. Campbell and Shiller (1991) Campbell and Shiller 1991 recover that Treasury yield spreads can be utilized to predict excess bond returns. [Cochrane and Piazzesi 2005] construct a tent-shaped linear combination of five forward spreads, which explains 30 to 35 percent of the variation of one-year excess returns on Treasury bonds maturing from two to five years. Following [Cochrane and Piazzesi 2005], [Ludvigson and Ng 2009a] shed new links between excess bond returns and macroeconomic variables, through dynamic factor analysis, they construct a macro factor by extracting variables form a monthly panel of 132

macroeconomic series, the macro factor explains 21-26 percent of the one year ahead variation in their excess returns. More recently, by incorporating a new model selection method- the supervised adaptive group lasso procedure, [J.-z. Huang and Shi 2011] extract a new macro factor from a monthly panel of 131 macroeconomic variables, which results in a stronger predictive power for future excess bond returns, with in sample R^2 up to 43 percent, findings in [J.-z. Huang and Shi 2011] provide new evidence on potential links between bond risk premia and macroeconomic fundamentals and provide further support for the notion of dynamic term structure models with hidden/unspanned factors as well. All the studies mentioned above spawn an extensive literature on the determinants of bond risk premia and serve as the pioneers of our study. Nonetheless, despite the theoretical insights, the fast development of statistical methodology extends more and more potential to uncover the deeper link between bond risk premia and macroeconomic fundamentals. For example, as stated in [Ludvigson and Ng 2009a], models before them are imperfect descriptions of reality and only restrict attention to a small set of variables and fail to span the information sets of financial market participants. To overcome the difficulty, [Ludvigson and Ng 2009a] apply the methodology of dynamic factor analysis to large macroeconomic datasets, and eliminate the arbitrary reliance on a small number of imperfectly measured indicators, by summarizing information from a large number of macroeconomic series into a few estimated factors(principal components), they make it feasible to incorporate all the information from a vast set of economic variables. However, as criticized in [J.-z. Huang and Shi 2011], factor analysis approach neglects the association between dependent variable and independent variables, standard principal components(linear combination of all the variables considered) contain all the information with respect to the data matrix of independent variables, some of which may even not be correlated with the dependent variable to be forecasted. Moreover, since the incorporating of all the information in principal components, not any specific important variables from the large dataset are selected, it arises difficulty to achieve a better interpretation. With all these considerations, [J.-z. Huang and Shi 2011] apply the SAGLasso procedure for variable selection and achieve a better forecast performance for excess bond returns and realize a better understanding on the underlying economic determinants of bond risk premia. Under the prosperity of related disciplines, like statistics, a vast of well performed variable selection techniques have been developed, which

enable us to facilitate the research on excess bond returns forecast. With more tools in hand today, we can confront more challenges. For instance, for a stronger predictive power(in-sample and out-of-sample) on excess bond returns and for a better understanding of the determinants of bond risk premia, can we achieve a better performance on variable selection from high dimensional dataset, like macroeconomic series with lags used in [J.-z. Huang and Shi 2011]? Besides the linear structure between excess bond returns and macroeconomic fundamentals, can we recover some nonlinear relationship which may benefit the prediction purpose? Can we propose a robust procedure of macro variables selection for predicting bond risk premia? In our study of Chapter 3, we will give a comprehensive discussion on these questions.

1.3 Markov Chain Monte Carlo(MCMC)

Empirical analysis of dynamic asset pricing models always try to tackle the problem of extracting information about latent state variables, structural parameters and market prices from the observed prices, and the Bayesian inference problem is trying to figure out the distribution of the parameters, Θ, and state variables, X, conditional on observed prices, Y. The posterior distribution, $p(\theta, X|Y)$, combines the information of the model as wll as the observed prices. The main target of Markov Chain Monte Carlo(MCMC) is trying to sample from these high-dimensional, complex distributions by generating a Markov Chain over (Θ, X), $\{\Theta^(g), X^(g)\}_{g=1}^{G}$, whose equilibrium distribution is $p(\Theta, X|Y)$. After which, the Monte Carlo method will use these samples for numerical integration for parameter estimation, state estimation or model comparison. As is known to all, that characterizing $p(\Theta, X|Y)$ in continuous-time asset pricing models is a tough job, because prices are observed discretely while the theoretical models specify that prices and state variables evolve continuously in time, second, in many cases, the state variables are the so called latent variables from our perspective, third, due to the high dimension of $p(\Theta, X|Y)$, we can not use standard sampling methods, fourth, the continuous-time models of interest always generate transition distributions for prices and state variables which are usually non-normal and non-standard, complicating standard estimation methods such as MLE or GMM, finally, parameters enter nonlinearly or even in a non-analytic form as the implicit solution to ordinary or partial differential

equations in term structure and option pricing models. MCMC methods are proved to be particularly well-suited for continuous-time finance applications and will tackle all of these issues for the following reasons. 1. Continuous-time asset models specify that prices and state variables solve parameterized stochastic differential equations (SDEs) which are built from Brownian motions, Poisson processes and other i.i.d. shocks whose distributions are easy to characterize. When discretized at any finite time-interval, the models take the form of familiar time series models with normal, discrete mixtures of normals or scale mixtures of normals error distributions. This implies that the standard tools of Bayesian inference directly apply to these models. 2. MCMC is a unified estimation procedure, simultaneously estimating both parameters and latent variables. MCMC directly computes the distribution of the latent variables and parameters given the observed data. This is a stark alternative the usual approach in the literature of applying approximate filters or noisy latent variable proxies. This allows the researcher, for example, to separate out the effects of jumps and stochastic volatility in models of interest rates or equity prices using discretely observed data 3. MCMC methods allow the researcher to quantify estimation and model risk. Estimation risk is the inherent uncertainty present in estimating parameters or state variables, while model risk is the uncertainty over model specification. Increasingly in practical problems, estimation risk is a serious issue whose impact must be quantified. In the case of option pricing and optimal portfolio problems, [Merton 1980] argues that the âĂIJmost important direction is to develop accurate variance estimation models which take into account of the errors in variance estimatesâĂİ (p. 355). 4. MCMC is based on conditional simulation, therefore avoiding any optimization or unconditional simulation. From a practical perspective, MCMC estimation is typically extremely fast in terms of computing time. This has many advantages, one of which is that it allows the researcher to perform simulation studies to study the algorithms accuracy for estimating parameters or state variables, a feature not shared by many other methods.

Next we want to describe the foundations and mechanics of MCMC algorithms, including the Clifford-Hammersley theorem, the Gibbs sampler, the Metropolis-Hastings algorithm, and theoretical convergence properties of MCMC algorithms as well.

1.3.1 Overview of MCMC and Bayesian Inference

In asset pricing models, we are trying to use MCMC to generate random samples from the distribution of parameters and state variables given the observed prices, $p(\Theta, X|Y)$. One way to motivate the construction of MCMC algorithms is based on a result called Clifford-Hammersley theorem, which states that a joint distribution can be described by its complete conditional distributions. Namely, the theorem states that $p(X|\Theta, Y)$ and $p(\Theta|X, Y)$ as the complete conditional distributions, will completely characterize the joint distribution of $p(\Theta, X|Y)$. MCMC will provide the recipe for combining the information in these distributions to generate samples from $p(\Theta, X|Y)$. For example, suppose we are given two initial values, $\Theta(0)$ and $X(0)$, draw $X(1) \sim p(X|\Theta^{(0)}, Y)$and then $\Theta(1) \sim p(\Theta|X^{(0)}, Y)$, Repeat this process again and again, we will finally generate a series of random variables,$\{X^{(g)}, \Theta^{(g)}\}_{g=1}^{G}$. This sequence is not i.i.d., but instead it will form a Markov Chain, under a number of metrics and mild conditions, the distribution of the Markov Chain will converge to $p(\Theta, X|Y)$, the target distribution. The key benefit of MCMC is that it is typically easier to sample from the complete conditional distributions, $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$, than to analyze the higher-dimensional joint distribution, $p(\Theta, X|Y)$ directly.

MCMC algorithms generically consist of two types of sampling. Case 1: if the complete conditional distribution is known in closed form or can be sampled directly, the step in the MCMC algorithm is called a âĂIJGibbsâĂİ step. Case 2: if one or more of the conditionals don't have a closed form or too complicated to be sampled directly, a âĂIJMetropolis-Hastings algorithmsâĂİ will apply, unlike Gibbs type, these algorithms sample a candidate from a proposal density first and then decide to accept or reject the candidate based on some acceptance criterion. We will discuss this later in detail. No matter for which case, the algorithms will always generate random samples with the appropriate equilibrium distribution. Generally, an algorithm can include only Gibbs steps, only Metropolis-Hastings steps or any combination of the two. This latter case, usually encountered in practice, generates a hybrid MCMC algorithm. The samples $\{X^{(g)}, \Theta^{(g)}\}_{g=1}^{G}$ from the joint posterior can be used for parameter and state variable estimation using the Monte Carlo method.

As for Bayesian inference, there are three important building concepts,

- The posterior distribution

  In asset pricing models, the posterior distribution summarizes the information embedded in prices regarding latent state variables and parameters. Bayes rule factors the posterior distribution into is constituent components:

$$p(\Theta, X|Y) \propto p(Y|X, \Theta)p(X|\Theta)p(\Theta) \tag{1.3}$$

  where $X = X_{t=1}^{T}$ are the unobserved state variables, $Y = Y_{t=1}^{T}$ are the observed prices, Θ are the parameters, $p(Y|X, \Theta)$ is the likelihood function, $p(X|\Theta)$ is the distribution of the state variables, and $p(\Theta)$ is the distribution of the parameters, commonly called the prior. The parametric asset pricing model generates $p(Y|X, \Theta)$ and $p(X|\Theta)$ and $p(\Theta)$ summarizes any non-sample information about the parameters.

- The likelihood

  Usually there are two types of likelihood functions of interest. The distribution $p(Y|X, \Theta)$ is the full-information (or data-augmented) likelihood and conditions on the state variables and parameters. This is related to marginal likelihood function, $p(Y|\Theta)$, which integrates the latent variables from the augmented likelihood:

$$p(Y|\Theta) = \int p(Y, X|\Theta)dX = p\int (Y|X, \Theta)p(X|\Theta)dX.$$

  In most continuous-time asset pricing models, $p(Y|\Theta)$ is not available in closed form and simulation methods are required to perform likelihood-based inference. On the other hand, the full-information likelihood is usually known in closed form which is a key to MCMC estimation.

- The prior distribution

  The prior distribution, as an implication of Bayes rule, enters in the posterior distribution in (1.3). Because $p(\Theta)$ cannot be ignored: its presence in the posterior, like the presence of the likelihood, is merely an implication of the laws of probability. Additionally, this distribution provides important economic and statistical roles. The prior allows the researcher to incorporate

nonsample information in a consistent manner. For example, the prior provides a consistent mechanism to impose important economic information such as positivity of certain parameters or beliefs over the degree of mispricing in a model. Statistically, the prior can impose stationarity, rule out near unit-root behavior, or separate mixture components, to name a few applications.

With these concepts in hand, we will give the mechanics of MCMC algorithms in detail, their theoretical underpinnings and convergence properties.

1.3.2 MCMC: Methods and Theory

In this section, we describe the mechanics of MCMC algorithms and their convergence properties.

1.3.2.1 Clifford-Hammersley Theorem

As described in previous section, in many continuous-time asset pricing models, $p(\Theta, X|Y)$ is usually with a complicated, high-dimensional distribution and it is quite difficult to directly generate samples from this distribution. On the other hand, we can break the joint distribution into its complete set of conditionals, which are of lower dimension and are easier to sample. It is in this manner that MCMC algorithms attacks the curse of dimensionality that plagues other methods.

The theoretical justification for breaking $p(\Theta, X|Y)$ into its complete conditional distributions is a remarkable theorem by Clifford and Hammersley. The general version of the Clifford-Hammersley theorem [Hammersley and Clifford 1971] and [Besag 1974] provides conditions for when a set of conditional distributions characterizes a unique joint distribution. In our setting, the theorem indicates that $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$ uniquely determine p $(\Theta, X|Y)$. If it is still not easy to sample from $p(\Theta|X, Y)$ and $p(X|\Theta, Y)$ directly. We can apply the Clifford-Hammersley theorem again and break the complete distributions further. For example, consider $p(\Theta|X, Y)$ and assume that the KâĹŠdimensional vector Θ can be partitioned into kâĽdK components $\Theta = (\Theta_1, \cdots, \Theta_k)$ where each component could be uni- or multidimensional. Given the partition, the Clifford- Hammersley theorem implies that the following set of conditional distributions

$$\Theta_1 \quad |\Theta_2, \Theta_3, \cdots, \Theta_k, X, Y$$

$$\begin{aligned}
\Theta_2 \quad & |\Theta_1, \Theta_3, \cdots, \Theta_k, X, Y \\
\cdots & \\
\Theta_k \quad & |\Theta_1, \Theta_2, \cdots, \Theta_{k-1}, X, Y
\end{aligned}$$

uniquely determines $p(\Theta|X,Y)$. For state vector, the joint distribution $p(X|\Theta,Y)$ can be described by its own complete set of conditionals: $p(X_t|\Theta, X_{(-t)}, Y)$ for $t = 1, \cdots, T$ where $X_{(\hat{a}\acute{L}\check{S}t)}$ denotes the elements of X excluding X_t. In the extreme case, the Clifford-Hammersley theorem implies that instead of drawing from a $T + K$ dimensional posterior, the same information is contained in $T + K$ one dimensional distributions. A proof of the Clifford-Hammersley theorem based on the Besag formula [Besag 1974] uses the insight that for any pair (Θ^0, X^0) of points, the joint density $p(\Theta, X|Y)$ is determined as

$$\frac{p(\Theta, X|Y)}{p(\Theta^0, X^0|Y)} = \frac{p(\Theta|X^0, Y)p(X|\Theta, Y)}{p(\Theta^0|X^0, Y)P(X^0|\Theta, Y)}$$

as long as a positivity condition is satisfied. Thus, knowledge of $p(\Theta|X,Y)$ and $p(X|\Theta,Y)$, up to a constant of proportionality, is equivalent to knowledge of the joint distribution. The positivity condition in our case requires that for each point in the sample space, $p(\Theta, X|Y)$ and the marginal distributions have positive mass. Under very mild regularity conditions the positivity condition is always satisfied.

1.3.2.2 Gibbs Sampling

The simplest MCMC algorithm is called the Gibbs sampler. If it is possible and easy to directly sample from all of the complete conditionals, the Gibbs sampler applies. For example, the following defines a Gibbs sampler: given $\Theta^{(0)}, X^{(0)}$

1. Draw $\Theta^{(1)} \sim p(\Theta|X^{(0)}, Y)$
2. Draw $X^{(1)} \sim p(X|\Theta^{(1)}, Y)$

Repeat this process again and again, the Gibbs sampler will generate a sequence of random variables, $\{\Theta^{(g)}, X^{(g)}\}_{g=1}^{G}$, which, as we discuss earlier, will converge to $p(\Theta, X|Y)$. The Gibbs sampler requires that one can conveniently draw from the complete set of conditional distributions. In many cases, implementing the Gibbs

sampler requires drawing random variables from standard continuous distributions such as Normal, t, Beta or Gamma or discrete distributions such as Binomial, Multinomial or Dirichlet. The reference books by [Devroye 1986] or [Ripley 2009] provide algorithms for generating random variables from a wide class of recognizable distributions.

1.3.2.3 Metropolis-Hastings

In some cases, one or more of the complete conditional distribution cannot be conveniently sampled, then a very general approach called the Metropolis-Hastings algorithms will often apply.

Consider the case where one of the parameter posterior conditionals, generically, $\pi(\Theta_i) = p(\Theta_i|\Theta_{(\hat{a}\acute{I}\check{S}i)}, X, Y)$, can be evaluated (as a function of Θ_i), (for simplicity, let's denote it by $\pi(\Theta)$), but it is not possible to generate a sample from the distribution directly. Then we can first specify a recognizable proposal or candidate density $q(\Theta^{(g+1)}|\Theta^{(g)})$, then the Metropolis-Hastings algorithm will first draws a candidate point that will be accepted or rejected based on the acceptance probability. The Metropolis-Hastings algorithm replaces a Gibbs sampler step with the following two stage procedure:

1. Draw $\Theta^{(g+1)}$ from the proposal density $q(\Theta^{(g+1)}|\Theta^{(g)})$
2. Accept $\Theta^{(g+1)}$ with probability $\alpha(\Theta^{(g)}, \Theta^{(g+1)})$

where

$$\alpha(\Theta^{(g)}, \Theta^{(g+1)}) = \min(\frac{\pi(\Theta^{(g+1)})/q(\Theta^{(g+1)}|\Theta^{(g)})}{\pi(\Theta^{(g)})/q(\Theta^{(g)}|\Theta^{(g+1)})}, 1)$$

The acceptance criterion can insure that the algorithm has the correct equilibrium distribution. Continuing in this manner, the algorithm generates samples $\{\Theta^{(g)}\}_{g=1}^{G}$ whose limiting distribution is $\pi(\Theta)$. It needs to point out, Gibbs sampling is also a special case of Metropolis-Hastings, if we choose $q(\Theta^{(g+1)}|\Theta^{(g)}) \propto \pi(\Theta^{(g+1)})$.

There are also another two important special cases of the general Metropolis-Hastings algorithm which deserve special attention, Independence Metropolis-Hastings and Random-Walk Metropolis.

If we draw the candidate $\Theta^{(g+1)}$ from a distribution independent of the previous state, $q(\Theta^{(g+1)}|\Theta^{(g)}) = q(\Theta^{(g+1)})$. This is known as an independence Metropolis-

Hastings, and the corresponding acceptance criterion is reduced to

$$\alpha(\Theta^{(g)}, \Theta^{(g+1)}) = \min(\frac{\pi(\Theta^{(g+1)})q(\Theta^{(g)})}{\pi(\Theta^{(g)})q(\Theta^{(g+1)})}, 1)$$

When using independence Metropolis, it is common to pick the proposal density to closely match certain properties of the target distribution.

If we draw candidate $\Theta^{(g+1)}$ from a proposal density where $q(\Theta^{(g+1)}|\Theta^{(g)}) = q(\Theta^{(g)}|\Theta^{(g+1)})$, we come to the Random-Walk Metropolis. The corresponding acceptance criterion is

$$\alpha(\Theta^{(g)}, \Theta^{(g+1)}) = \min(\frac{\pi(\Theta^{(g+1)})}{\pi(\Theta^{(g)})}, 1)$$

In random walk Metropolis-Hastings algorithms, the researcher controls the variance of the error term and the algorithm must be tuned, by adjusting the variance of the error term, to obtain an acceptable level of accepted draws, generally in the range of 20-40%.

1.3.2.4 Convergence Theory

MCMC algorithm generates sequence of draws for parameters, $\Theta^{(g)}$, and state variables, $X^{(g)}$. By construction, this sequence is Markov and the chain is characterized by its starting value, $\Theta^{(0)}$ and its conditional distribution or transition kernel $P(\Theta^{(g+1)}, \Theta^{(g)})$, without any loss of generality, we abstract from the latent variables. One of the main advantages of MCMC is the attractive convergence properties that this sequence of random variables inherits from the general theory of Markov Chains.

We are interested in verifying that the chain produced by the MCMC algorithm converges and then identifying the unique equilibrium distribution of the chain as the correct joint distribution, the posterior. We now briefly review the basic theory of the convergence of Markov Chains. Where, we have the conclusion that if an irreducible and aperiodic chain has a proper invariant distribution π, then π is unique and is also the equilibrium distribution of the chain.

The easiest way to verify and find an invariant distribution is to check time-reversibility. Recall that for a Metropolis-Hastings algorithm, that the target distribution, π, is given and is proper being the posterior distribution. The easiest

way of checking that π is an invariant distribution of the chain is to verify the detailed balance (time-reversibility) condition: a transition function P satisfies the detailed balance condition if there exists a function π such that

$$P(x,y)\pi(x) = P(y,x)\pi(y)$$

for any points x and y in the state space, for the general Metropolis-Hasting algorithm, the transition function is

$$P(x,y) = \alpha(x,y)q(x,y)$$

if $x \neq y$, then

$$\begin{aligned}
\alpha(x,y)q(x,y)\pi(x) &= \min(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)},1)q(x,y)\pi(x) \\
&= \min(\pi(y)q(y,x), q(x,y)\pi(x)) \\
&= \min(\frac{\pi(x)q(x,y)}{\pi(y)q(y,x)},1)q(y,x)\pi(y) \\
&= \alpha(y,x)q(y,x)\pi(y)
\end{aligned}$$

for the case where $x = y$, the proof of time-reversibility is trivial. Then we can conclude that Metropolis-Hastings algorithms generate Markov Chains that are time-reversible and have the target distribution as an invariant distribution.

It is straightforward to verify π-irreducibility, see [Roberts and N. G. Polson 1994] for the Gibbs samplers and [Roberts and Smith 1994] and [Christian and Casella 1999] for Metropolis-Hastings algorithms. One sufficient condition is that $\pi(y) > 0$ implies that $q(x,y) > 0$ (see, e.g., [Mengersen, Tweedie, et al. 1996]). To verify aperiodicity, one can appeal to a theorem in [Tierney 1994] which states that all πâĹŠirreducible Metropolis algorithms are Harris recurrent. Hence, there exists a unique stationary distribution to which the Markov chain generated by Metropolis-Hastings algorithms converges and hence the chain is ergodic.

In summary, we have the following two theorems, specifying results on sample averages of functionals along the chain.

Theorem 1.3.1. *(Ergodic Averaging) Suppose $\Theta^{(g)}$ is an ergodic chain with stationary distribution π and suppose f is a real-valued function with $\int |f| d\pi < \infty$.*

Then for all $\Theta^{(g)}$ for any initial starting value $\Theta^{(g)}$,

$$\lim_{G\to\infty} \frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)}) = \int f(\Theta)\pi(\Theta)d\Theta$$

almost surely.

We can even go further with an ergodic central limit theorem,

Theorem 1.3.2. *(Central Limit Theorem) Suppose $\Theta^{(g)}$ is an ergodic chain with stationary distribution π and suppose f is a real-valued function with $\int |f|d\pi < \infty$. Then there exists a real number $\sigma(f)$, such that*

$$\sqrt{G}(\frac{1}{G}\sum_{g=1}^{G} f(\Theta^{(g)}) - \int f(\Theta)\pi(\Theta)d\Theta)$$

converges in distribution to a mean zero normal distribution with variance $\sigma^2(f)$ for any starting value.

1.4 Stochastic Calculus Preliminery

In this section, we review some preliminery of stochastic differential equations, including Markov process, diffusion process, backward and forward Kolmogorov equations, as well as Ito formulas.

1.4.1 Diffusion Process and Kolmogorov Equations

Beginning from Markov process, which is defined as follows,

Definition 1.4.1. *Let X_t be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in $\mathbb{R}^d$, and let $\mathcal{F}_t^X$ be the filtration generated by $\{X_t, t \in R_+\}$, then $\{X_t, t \in R_+\}$ is a Markov process if*

$$\mathbb{P}(X_t \in \Gamma|\mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma|X_s)$$

for all $t, s \in [0, T]$ with $t \geq s$, and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, $\mathcal{B}$ is a $\sigma-$algebra.

If denote $P(\Gamma, t|x, s) := \mathbb{P}(X_t \in \Gamma|\mathcal{F}_s^X)$, then we have the Chapman-Kolmogorov equation:

$$P(\Gamma, t|x, s) = \int_{\mathbb{R}^d} P(\Gamma, t|y, u)P(dy, u|x, s)$$

which is true for all $x \in \mathbb{R}^d$, $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, and $s, u, t \in \mathbb{R}_+$ with $s \leq u \leq t$.

In the set up for the following content, we assume for our continuous-time markov process, the conditional probability density exists, i.e. the transition probability has the following expression,

$$\mathbb{P}(X_{t+h} \in \Gamma|X_t = x) = \int_\Gamma p(y, t + h|x, t)dy$$

For Markov process, it comes naturally that the transition function and the initial distribution of X_t are sufficient to determine this process uniquely, if we consider Markov processes whose transition function has a density with respect to the Lebesgue measure, the Chapman-Kolmogorov equation becomes

$$\int_\Gamma p(y, t|x, s)dy = \int_{\mathbb{R}^d} \int_\Gamma p(y, t|z, u)p(z, u|x, s)dzdy$$

since $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ is arbitrary, we have the Chapman-Kolmogorov equation for the transition probability density:

$$p(y, t|x, s) = \int_{\mathbb{R}^d} p(y, t|z, u)p(z, u|x, s)dz$$

Next, we talk about the generator of a Markov process, suppose X_t denotes a time-homogeneous Markov process, let $p(t, x, y) := p(y, t|x, 0)$ be the transition function, $f \in \mathcal{C}_b(\mathbb{R}^d)$, the space of continuous bounded functions on $\mathbb{R}^d$, define the operator

$$(\mathcal{P}_t f)(x) := \mathbb{E}(f(X_t)|X_0 = x) = \int_{\mathbb{R}^d} f(y)p(t, x, dy)$$

it is easy to prove the operator has the properties of $(\mathcal{P} \circ f)(x) = f(x)$, $(\mathcal{P}_{t+s}f)(x) = (\mathcal{P}_t \circ (\mathcal{P}_s f))(x)$, i.e. $\mathcal{P}_0 = I$, $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s$, for all $t, s \geq 0$, the operator forms a semigroup.

If we assume $\mathcal{P}_t f$ is also a $\mathcal{C}_b(\mathbb{R}^d)$ function, and define $\mathcal{D}(\mathcal{L})$ the set of all $f \in \mathcal{C}_b(\mathbb{R}^d)$ such that the strong limit

$$\mathcal{L}f := \lim_{t \to 0} \frac{\mathcal{P}_t f - f}{t}$$

exists, the operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \to \mathcal{C}_b(\mathbb{R}^d)$ is called the infinitesimal generator of the operator semigroup $\mathcal{P}_t$, we will also refer to $\mathcal{L}$ as the generator of the Markov process X_t.

Following the semigroup property and the definition of the generator, we have the following equation formally,

$$\mathcal{P}_t = e^{t\mathcal{L}}$$

If we consider $u(x,t) = (\mathcal{P}_t f)(x) = \mathbb{E}(f(X_t)|X_0 = x)$ and calculate its time derivative, u satisfies

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{d}{dt}(\mathcal{P}_t f) = \frac{d}{dt}(e^{t\mathcal{L}} f) = \mathcal{L}(e^{t\mathcal{L}} f) = \mathcal{L}u \\ u(x,0) = f(x) \end{cases} \tag{1.4}$$

this is called the backward Kolmogorov equation, which governs the evolution of the expection of an observation $f \in \mathcal{C}_b(\mathbb{R}^d)$, when the Markov process is the solution of a stochastic differential equation, the generator is a second-order elliptic differential operator, and the backward Kolmogorov equation will become an initial value problem for a parabolic partial differential equation, this will be explained in detail later.

The semigroup $\mathcal{P}_t$ is a operator acts on bounded continuous functions, we can also define teh corresponding adjoint semigroup $\mathcal{P}_t^*$, which acts on the probability measure,

$$(\mathcal{P}_t^* \mu)(\Gamma) = \int_{\mathbb{R}^d} \mathbb{P}(X_t \in \Gamma | X_0 = x) d\mu(x) = \int_{\mathbb{R}^d} p(x, t.\Gamma) d\mu(x)$$

the image of a probability measure μ under $\mathcal{P}_t^*$ is again a probability measure. The operators $\mathcal{P}_t^*$ and $\mathcal{P}_t$ are adjoint in the L^2-sense.

$$\int_{\mathbb{R}^d} (\mathcal{P}_t f)(x) d\mu(x) = \int_{\mathbb{R}^d} f(x) d(\mathcal{P}_t^* \mu)(x)$$

we can write $\mathcal{P}_t^* = e^{t\mathcal{L}^*}$, $\mathcal{L}^*$ is the L^2-adjoint of the generator of the process.

Suppose for Markov process X_t. $X_0 \sim \mu$, define $\mu_t := \mathcal{P}_t^* \mu$, then we have the initial problem holds,

$$\frac{\partial \mu_t}{\partial t} = \mathcal{L}^* \mu_t, \quad \mu(0) = \mu$$

which governs the dynamic of the distribution of X_t.

Next, let's review the diffusion process and its forward and backward Kolmogorov equations. As is known to us, a Markov process consists of three parts, a drift, a random part and a jump process. A diffusion process is a Markov process that has continuous sample paths, i.i. a Markov process with no jump.

Definition 1.4.2. *A Markov process X_t in $\mathbb{R}$ with transition function $P(\Gamma, t|x, s)$ is called a diffusion process if the following conditions are satisfied:*

- *Continuity: For every x and every $\epsilon > 0$,*

$$\int_{|x-y|>\epsilon} P(dy, x|x, s) = o(t-s) \tag{1.5}$$

  uniformly over $s < t$.

- *Definition of drift coefficient: There exists a function $b(x, s)$ such that for every x and every $\epsilon > 0$,*

$$\int_{|x-y|\leq\epsilon} (y-x)P(dy, x|x, s) = b(x, s)(t-s) + o(t-s) \tag{1.6}$$

  uniformly over $s < t$.

- *Definition of diffusion coefficient: There exists a function $\Sigma(x, s)$ such that for every x and every $\epsilon > 0$,*

$$\int_{|x-y|\leq\epsilon} (y-x)^2 P(dy, x|x, s) = \Sigma(x, s)(t-s) + o(t-s) \tag{1.7}$$

  uniformly over $s < t$.

If we assume the first two moments exist, for the drift and diffusion coefficients, we will have

$$\lim_{t\to s} \mathbb{E}\left(\frac{X_t - X_s}{t-s}|X_s = x\right) = b(x, s)$$
$$\lim_{t\to s} \mathbb{E}\left(\frac{|X_t - X_s|^2}{t-s}|X_s = x\right) = \Sigma(x, s)$$

For the diffusion process, we can now try to obtain an explicit formula for the generator of such a process and to derive a partial differential equation for the

conditional expectation $u(x,s) = \mathbb{E}(f(X_t)|X_s = x)$, as well as for the transition probability density $p(y,t|x,s)$. These are the so-called backward and forward Kolmogorov equations.

Theorem 1.4.3. *(Kolmogorov) Let $f(x) \in \mathcal{C}_b(\mathbb{R})$, and let*

$$u(x,s) = \mathbb{E}(f(X_t)|X_s = x) = \int f(y)P(dy,t|x,s)$$

with t fixed. Assume furthermore, that the functions $b(x,s)$, $\Sigma(x,s)$ are smooth in both x and s. Then $u(x,s)$ solves the final value problem,

$$\begin{aligned} -\frac{\partial u}{\partial s} &= b(x,s)\frac{\partial u}{\partial x} + \frac{1}{2}\Sigma(x,s)\frac{\partial^2 u}{\partial x^2} \\ u(t,x) &= f(x), \ \textit{for } s \in [0,t] \end{aligned}$$

This is a final value problem for a partial differential equation of parabolic type, for the time-homogeneous diffusion processes, where the drift and the diffusion coefficients are independent of time, $b = b(x)$ and $\Sigma = \Sigma(x)$, if let $T = t - s$, $U(x,T) = u(x,t-s)$, we can write the backward Kolmogorov equation as follows:

$$\begin{aligned} \frac{\partial U}{\partial t} &= b(x)\frac{\partial U}{\partial x} + \frac{1}{2}\Sigma(x)\frac{\partial^2 U}{\partial x^2} \\ U(x,0) &= f(x) \end{aligned}$$

If we reset the initial time $s = 0$, let $u(x,t) = \mathbb{E}(f(X_t)|X_0 = x$, then we have the following backward Kolmogorov equation for time-homogeneous equation:

$$\begin{aligned} \frac{\partial u}{\partial t} &= b(x)\frac{\partial u}{\partial x} + \frac{1}{2}\Sigma(x)\frac{\partial^2 u}{\partial x^2} \\ u(x,0) &= f(x) \end{aligned}$$

The differential operator on the right-hand side is the generator of the diffusion process of X_t, i.e. $\mathcal{L} = b(x)\frac{\partial}{\partial x} + \frac{1}{2}\Sigma(x)\frac{\partial^2}{\partial x^2}$.

Assume the transition function has a density with respect to the Lebesgue measure that is a smooth function of its arguments, $P(dy,t|x,s) = p(y,t|x,s)dy$, we will have the following forward Kolmogorov equation or Fokker-Planck equation.

Theorem 1.4.4. *(Kolmogorov) Assume that conditions (1.5), (1.6), (1.7) are satisfied, and that $p(y,t|\cdot,\cdot)$, $b(y,t)$ and $\Sigma(y,t)$ are smooth functions of y,t. Then*

the transition probability density is the solution to the initial value problem

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}(b(t,y)p) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\Sigma(t,y)p)$$
$$p(y,s|x,s) = \delta(x-y)$$

If we assume the initial distribution of $X_0 \sim \rho_0(x)$, then define $p(y,t) := \int p(y,t|x,0)\rho_0(x)dx$, we multilply the forward Kolmogorov equation by $\rho_0(x)$ and integrate with respect to x, we have the following equation,

$$\frac{\partial p(y,t)}{\partial t} = -\frac{\partial}{\partial y}(a(y,t)p(y,t)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y,t)p(y,t))$$
$$p(y,0) = \rho_0(y)$$

which is the Fokker-Planck equation, which provides the probability of the diffusion process X_t.

For the case of multi-dimensional diffusion process in $\mathbb{R}^d$, define

$$b(x,s) = \lim_{t\to s}\int_{|y-x|<\epsilon}(y-x)P(dy,t|x,s)$$
$$\Sigma(x,s) = \lim_{t\to s}\frac{1}{t-s}\int_{|y-x|<\epsilon}(y-x)\otimes(y-x)P(dy,t|x,s)$$

where the drift coefficient $b(x,s)$ is a d-dimensional vector field, and the diffusion coefficient $\Sigma(x,s)$ is a d-by-d symmetric nonnegative matrix. The generator of a d-dimensional diffusion process is given as:

$$\begin{aligned}\mathcal{L} &= b(x,s)\cdot\nabla + \frac{1}{2}\Sigma(x,s) : \nabla\nabla \\ &= \sum_{j=1}^{d} b_j(x,s)\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{i,j=1}^{d}\Sigma_{ij}(x,s)\frac{\partial^2}{\partial x_i \partial x_j}\end{aligned}$$

If we assume the first and second moments of the multi-dimensional diffusion exist, we can write the formulas for the drift vector and diffusion matrix as

$$\lim_{t\to s}\mathbb{E}\left(\frac{X_t - X_s}{t-s}|X_s = x\right) = b(x,s)$$
$$\lim_{t\to s}\mathbb{E}\left(\frac{(X_t - X_s)\otimes(X_t - X_s)}{t-s}|X_s = x\right) = \Sigma(x,s)$$

the backward and forward Kolmogorov equations are

$$-\frac{\partial u}{\partial s} = b(x,s)\cdot\nabla_x u + \frac{1}{2}\Sigma(x,s):\nabla_x\nabla_x u, \quad u(t,x) = f(x)$$
$$\frac{\partial p}{\partial t} = \nabla_y\cdot(-b(t,y)p + \frac{1}{2}\nabla_y\cdot(\Sigma(t,y)p)), \quad p(y,s|x,s) = \delta(x-y)$$

1.4.2 Stochastic Differential Equation

In this section, let's consider the SDE:

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \quad X_0 = x$$

where $b(\cdot,\cdot):[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, and $b\sigma(\cdot,\cdot):[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times m}$ are measurable vector-valued or matrix-valued functions, W_t denotes standard Brownian motion in $\mathbb{R}^m$. Assume initial condition is a random variable which is independent of W_t, we define a strong solution as

Definition 1.4.5. *A process X_t with continuous paths defined on the probability space $(\Omega,\mathcal{F},P)$ is called a strong solution if*

- *X_t is almost surely continuous and adapted to the filtration $\mathcal{F}$.*
- *$b(\cdot,X)\in L^1((0,T);\mathbb{R}^d)$ and $\sigma(\cdot,X)\in L^2((0,T);\mathbb{R}^{d\times m})$ almost surely.*
- *For every $t\geq 0$, teh stochastic integral equation*

$$X_t = x + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s, X_0 = x$$

  holds almost surely.

If we have the condition that there exists a positive constant C such that for all $x\in\mathbb{R}^d$ and $t\in[0,T]$,

$$|b(t,x)| + |\sigma(t,x)|_F \leq C(1+|x|) \tag{1.8}$$

and for all $x,y\in\mathbb{R}^d$ and $t\in[0,T]$

$$|b(t,x)-b(t,y)| + |\sigma(t,x)-\sigma(t,y)|_F \leq C(|x-y|) \tag{1.9}$$

a theorem about existence and uniqueness of strong solution follows,

Theorem 1.4.6. *Let $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ satisfy assumption (1.8) and (1.9), further more, the initial condition x is a random variable independent of the Brownian motion W_t, with $\mathbb{E}(|x|^2) < \infty$, then the SDE has a unique strong solution, X_t, with*

$$\mathbb{E}[\int_0^t |X_s|^2 ds] < \infty \text{ for all } t > 0$$

The solution satisfies the markov property and has continuous paths, it is a diffusion process.

As for the application in Chapter 2, we use a SDE with a diffusion term of square function, which doesn't satisfy the Lipschitz condition, then following theorem from [Karatzas and Shreve 2012] can relax the Lipschitz condition on the diffusion term.

Theorem 1.4.7. *Suppose the coefficients of the one-dimensional equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

satisfy the conditions

$$|b(t,x) - b(t,y)| \leq K|x-y|$$
$$|\sigma(t,x) - \sigma(t,y)| \leq h(|x-y|)$$

for every $0 \leq t < \infty$, and $x \in \mathbb{R}$, $y \in \mathbb{R}$, K is a positive constant and $h : [0,\infty) \to [0,\infty)$ is a strictly increasing function with $h(0) = 0$ and

$$\int_{(0,\epsilon)} h^{-2}(u)du = \infty, \text{ for any } \epsilon$$

then the strong uniqueness holds.

If we choose $h(u) = \sqrt{u}$, that would be the equation used in our case. As for the strong existence and uniqueness of high dimensional square root diffusion equations, please refer [Duffie and Kan 1996].

Except the discussion on X_t itself, we can also consider a functional on X_t, and talk about the rate of change in time , which is the famous Ito's formula.

Theorem 1.4.8. *(Ito's formula) Assume that the conditions of Theorem(1.4.6) hold, let X_t be the solution and let $V \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d)$, tehn the process $V(X_t 0$ satisfies:*

$$V(t, X_t) = V(0, X_0) + \int_0^t \frac{\partial V}{\partial s}(s, X_s)ds + \int_0^t \mathcal{L}V(s, X_s)ds + \int_0^t < \nabla V(s, X_s), \sigma(X_s)dW_s >$$

i.e.

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + < \nabla V(t, X_t), \dot{X}_t > + \frac{1}{2} < \dot{X}_t, D^2V(t, X_t)\dot{X}_t >$$

More general, if we have the assumption $f \in \mathcal{C}_0^2(\mathbb{R}^d)$ and $V \in \mathcal{C}(\mathbb{R}^d)$ bounded from below, then the function

$$u(x,t) = \mathbb{E}\left[e^{-\int_0^t V(X_s)ds} f(X_t)\right]$$

is the solution to the initial value problem

$$\frac{\partial u}{\partial t} = \mathcal{L}u - Vu, u(0,x) = f(x)$$

this is the well known Feynman-Kac formula which establishes a link between parabolic partial differential equations and stochastic processes, which is very import in option pricing models.

1.5 Jump-Diffusion Models

To illustrate jump-diffusion models, we need start from the Levy process.

Definition 1.5.1. *A stochastic process X_t is a Levy process if*

- $X_0 = 0$ *almost surely.*
- *Independent increments: for any choice of $n \geq 1$, $0 \leq t_0 < \cdots < t_n$, the random variables of $X_{t_0}, X_{t_1} - X_{t_0}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent.*
- *Stationary increments: distribution of $X_{t+s} - X_t$ does not depend on s.*
- *Stochastic continuity: for any $t \geq 0$, $\epsilon > 0$, $\lim_{t \to s} \mathbb{P}(|X_t - X_s| > \epsilon) = 0$*

- *Cadlag: there exists a space* $\Omega_0, \mathbb{P}(\Omega_0) = 1$*, for each* $\omega \in \Omega_0$*,* $X_t(\omega)$ *is right-continuous in* t *for* $t \geq 0$ *and has left limits in* t *for* $t > 0$*.*

Some typical examples of Levy processes include Poisson process, Compound Poisson process, Jump-diffusion process, Gamma process, Variance Gamma process, Inverse Gaussian process, and Normal inverse Gaussian process.

- Poisson process: let $\{t_i\}_{i\geq 1}$ be a series of independent exponential random variables with parameter λ, $T_n = \sum_{i=1}^{n} t_i$, the Poisson process is defined by

$$N_t = \sum_{n\geq 1} I_{t\geq T_n}$$

  N_t follows Poisson distribution of parameter λt,

$$\mathbb{P}(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

  the corresponding characteristic function of a Poisson process X_t with parameter λ is

$$\Phi_X(u) = \mathbb{E}(e^{iuX_t}) = \exp\{\lambda t(e^{iu} - 1)\}$$

- Compound Poisson process: if we set the jump size free in Poisson process, more precisely, let $\{Y_i\}_{i\geq 1}$ be a sequence of independent random variable with law f, define the Compound Poisson process as

$$X_t = \sum_{i=0}^{N_t} Y_i$$

  the corresponding characteristic function is

$$\mathbb{E}(e^{iuX_t}) = \exp\{\lambda t \int_{\mathbb{R}} (e^{iux} - 1) f(dx)\}$$

- Jump-diffusion process: if we combine a Brownian motion drift and a Compound Poisson process as a jump, we get a simplest jump-diffusion process,

$$X_t = \mu t + \sigma W_t + \sum_{i=0}^{N_t} Y_i$$

the corresponding characteristic function is given by

$$\mathbb{E}(e^{iuX_t}) = \exp\{t(i\mu u - \frac{\sigma^2 u^2}{2} + \lambda \int_{\mathbb{R}} (e^{iux} - 1) f(dx))\}$$

- Gamma process: a Gamma process is a random process with independent gamma distributed increments, whose marginal density with parameters λ and ct is given by

$$p_t(x) = \frac{\lambda^{ct}}{\Gamma(ct)} x^{ct-1} e^{-\lambda t}$$

where $\Gamma(x) = \int_0^{+\infty} e^{-t} t^{x-1} dt$, the corresponding characteristic function is

$$\mathbb{E}(e^{iuX_t}) = \frac{1}{(1 - iu/\lambda)^{-ct}}$$

- Variance Gamma process: if we take a Gamma process as drift and change the time scale with a Gamma process:

$$Y_t = \mu X_t + \sigma W_{X_t}$$

where X_t is a standard Gamma process with parameters λ and c, the corresponding characteristic function is

$$\mathbb{E}(e^{iuY_t}) = \frac{1}{(1 + \frac{\sigma^2 u^2}{2} - i\mu c u)^{ct}}$$

- Inverse Gaussian process: Inverse Gaussian process with parameters λ and c is a Levy process X_t, who follows an inverse Gaussian distribution of parameters λ and ct;

$$p_t(x) = \frac{ct}{x^{1.5}} e^{-\frac{(\sqrt{\lambda}x - \sqrt{\pi}ct)^2}{x}}$$

whose characteristic function is

$$\mathbb{E}(e^{iuX_t}) = \exp\left(-2ct\sqrt{\pi}(\sqrt{\lambda - iu} - \sqrt{\lambda})\right)$$

- Normal inverse Gaussian process: if we take a Brownian motion as drift and

change the time scale with Inverse Gaussian process:

$$Y_t = \theta X_t + \sigma W_{X_t}$$

the corresponding characteristic function is

$$\mathbb{E}(e^{iuY_t}) = \exp\left(t\frac{1-\sqrt{1+u^2\sigma^2\kappa - 2i\theta u\kappa}}{\kappa}\right)$$

As for the model of stock prices or some financial asset dynamics, we always require the positivity and the independence and stationarity of log-returns, this implies another large class: exponentials of Levy process, $S_t = S_0 e^{X_t}$, where X_t is a Levy process. The popular jump models include jump diffusion model, Variance Gamma model, CGMY model, Normal Inverse Gaussian model, Bates volatility model and general Affine model.

- Jump diffusion model: general jump diffusion model has the following form

  $$dS_t = \mu S_{t-}dt + \sigma S_{t-}dW_t + S_{t-}dJ_t$$

  where W_t is Brownian motion and J_t is a Compound Poisson process with jump intensity λ, $J_t = \sum_{i=1}^{N_t} Y_i$, μ is a parameter chosen to make $e^{-rt}S_t$ a martingale, for different distribution of Y_i, we have different model categories:

  - Poisson jump model if $Y_i = d_+$ with probability p and $Y_i = d_-$ with probability $1-p$.
  - Merton's jump model if $Y_i \sim N(a, b)$
  - Kou's model is $\log(Y_i)$ follows an asymmetric double exponential distribution with density $f(y) = p\eta_1 e^{-\eta_1 y} I_{y\geq 0} + (1-p)\eta_2 e^{\eta_2 y} I_{y<0}$, where $0 \leq p \leq 1$, and $\eta_1 > 1$, $\eta_2 > 0$

- Variance Gamma model: Let X_t be an independent Gamma process with mean rate unity. Y_t is an inver gamma process based on X_t, then the variance gamma model is defined by:

  $$S_t = S_0 \exp(\mu t + Y_t)$$

- CGMY model, based on Variance Gamma model, if

$$S_t = S_0 \exp\left(\mu t + \sigma W_t + X_t\right)$$

where X_t is a Levy process without drift and diffusion, the corresponding jump measure is given by $K_{CGMY}(x) = \frac{C}{|x|^{1+Y}}(\exp\left(-G|x|\right) * I_{x<0} + \exp\left(-M|x|\right) * I_{x>0})$.

- Normal Inverse Gaussian model: suppose X_t is a normal inverse Gaussian process with parameters $\nu, \alpha, \beta, \delta$, the Normal Inverse Gaussian model is defined by

$$S_t = S_0 \exp \mu t + \sigma W_t + X_t$$

- Bates volatility model: this is a stochastic volatility model, with dynamics

$$\begin{aligned} dS_t &= (r-\delta)S_{t-}dt + \sqrt{V_t}S_{t-}dW_t^1 + S_{t-}dJ_t \\ dV_t &= \kappa(\theta - V_t) + \sigma_t\sqrt{V_t}dW_t^2 \end{aligned}$$

where W_t^1, W_t^2 are Brownian motions such that $dW_t^1 dW_t^2 = \rho dt$, J_t is a Compound Poisson process with independent jumps Y, where $\log(1+Y)$ follows a normal distribution.

- General Affine model: this is a large class of models which has been deeply analized by Duffee. It follows the form

$$\begin{aligned} S_t &= S_0 e^{X_t} \\ dX_t &= \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t \end{aligned}$$

where dJ_t is a Compound Poisson process with intensity $\lambda(X_t)$, also the following affine assumption is required:

$$\begin{aligned} &\mu(x) = K_0 + K_1 x, K = (K_0, K_1) \in \mathbb{R}^N \times \mathbb{R}^{N\times N} \\ &(\sigma(x)\sigma(x)^T)_{ij} = (H_0)_{ij} + (H_1)_{ij} \cdot x, H = (H_0, H_1) \in \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N} \\ &\lambda(x) = l_0 + l_1 \cdot x, l = (l_0, l_1) \in \mathbb{R} \times \mathbb{R}^N \end{aligned}$$

1.6 Stochastic Differential Equation with Jumps

Next is a brief review on SDE with jumps. Let there be given a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}), \mathcal{P}$, with $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t\geq 0}$, satisfying the usual conditions, and with a set of marks $\mathcal{E} = \mathbb{R} - \{0\}$. We define on $\mathcal{E} \times [0,T]$ an $\underline{\mathcal{A}}$-adapted Poisson measure $p_\phi(dv, dt)$, with intensity measure $\nu_\phi(dv, dt) = \phi(dv)dt$, $T \in [0, \infty)$. Thus, $p_\phi = \{p_\phi(t) = p_\phi(\mathcal{E}, [0,t]), t \in [0,T]\}$ is a stochastic process that counts the number of jumps occuring in the time interval $[0,T]$. The Poisson random measure $p_\phi(dv, dt)$ generates a sequence of pairs $\{(\tau_i, \xi_i), i \in \{1, 2, 3, \cdots, p_\phi(T)\}\}$, where $\{\tau_i \in [0,T], i \in \{1, 2, 3, \cdots, p_\phi(T)\}\}$ is a sequence of increasing nonnegative random variables representing the jump times of a standard Poisson process with intensity λ, and $\{\xi_i \in \mathcal{E}, i \in \{1, 2, 3, \cdots, p_\phi(T)\}\}$ is a sequence of independent, identically distributed random variables. Here ξ_i is distributed according to $\frac{\phi(dv)}{\lambda} = F(dv)$. $F(\cdot)$ is the distribution function of the marks. We interpret τ_i as the time of the ith event and the mark ξ_i as its amplitude. We can consider the d-dimensional SDE with jumps

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t + \int_{\mathcal{E}} c(t, X_{t-}, v)p_\phi(dv, dt) \tag{1.10}$$

for $t \in [0,T]$, with initial value $X_0 \in \mathbb{R}^d$, an $\underline{\mathcal{A}}$-adapted m-dimensional Wiener process $W = \{W_t = (W_t^1, W_t^2, \cdots, W_t^m), t \in [0,T]\}$ and the previously introduced Poisson random measure p_ϕ. X_{t-} is the almost sure left-hand limit of X at time t. A solution of an SDE of the type (1.10) is called a jump diffusion or an Ito process with jumps.

The coefficients $a(t,x)$ and $c(t,x,v)$ are d-dimensional vectors of Borel measurable real valued functions on $[0,T] \times \mathbb{R}^d$ and on $[0,T] \times \mathbb{R}^d \times \mathcal{E}$, respectively. Additionally, $b(t,x)$ is a $d \times m$-matrix of Borel measurable real valued functions on $[0,T] \times \mathbb{R}^d$. The SDE (1.10) is only a short hand notation for its integral form

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s + \sum_{i=1}^{p_\phi(t)} c(\tau_i, X_{\tau_i -}, \xi_i) \tag{1.11}$$

where the pair $\{(\tau_i, \xi_i), i \in \{1, 2, 3, \cdots, p_\phi(T)\}\}$ is the above described sequence of pairs of jumps times and corresponding marks generated by the Poisson random measure p_ϕ.

Next, we can continue to define the strong solution and uniqueness of strong solution.

Definition 1.6.1. *Assume that we have given a filtered probability space* $(\Omega, \mathcal{A}, \underline{\mathcal{A}}), \mathcal{P}$. *And we call a triplet* $(X, W_{,\phi})$, *consisting of a stochastic process* $X = \{X_t, t \in [0,T]\}$, *an* $\underline{\mathcal{A}}$*-adapted standard Wiener process* W *and an* $\underline{\mathcal{A}}$*-adapted Poisson measure* p_ϕ, *a strong solution of the Ito integral equation (1.11) if* X *is* $\underline{\mathcal{A}}$*-adapted, the integrals on the right hand side are well-defined and the equality in (1.11) holds almost surely.*

As for the uniqueness, it is defined as follows,

Definition 1.6.2. *If any two strong solutions* X *and* $\tilde{X}$ *are indistinguishable on* $[0,T]$, *that is if*

$$X_t = \tilde{X}_t$$

almost surely for all $t \in [0,T]$, *then we say that the solution of (1.11) on* $[0,T]$ *is a unique strong solution.*

Next, we can state a standard theorem on the existence and uniqueness of strong solution of SDEs with jumps.This ensures that the objects we model are well defined. We assume the coefficient functions of SDE (1.11) satisfy the Lipschitz conditions

$$|a(t,x) - a(t,y)| \leq C_1|x-y| \tag{1.12}$$

$$|b(t,x) - b(t,y)| \leq C_2|x-y| \tag{1.13}$$

$$\int_{\mathcal{E}} |c(t,x,v) - c(t,y,v)|^2 \phi(dv) \leq C_3|x-y|^2 \tag{1.14}$$

as well as the linear growth conditions

$$|a(t,x)| \leq K_1(1+|x|) \tag{1.15}$$

$$|b(t,x)| \leq K_2(1+|x|) \tag{1.16}$$

$$\int_{\mathcal{E}} |c(t,x,v)|^2 \phi(dv) \leq K_3(1+|x|^2) \tag{1.17}$$

for every $t \in [0,T]$ and $x, y \in \mathbb{R}^d$, moreover, we assume the initial value X_0 is $\mathcal{A}_0$-measurable with $E(|X_0|^2) < \infty$.

Then we have the following existence and uniqueness theorem for SDE (1.11)

Theorem 1.6.3. *Suppose that the coefficient function $a(\cdot), b(\cdot)$ and $c(\cdot)$ of SDE (1.11) satisfy the Lipschitz conditions (1.14), the linear growth conditions (1.17) and the initial condition. Then the SDE (1.11) admits a unique strong solution. Moreover, the solution X_t of the SDE satisfies the estimate*

$$E\left(\sup_{0\leq s\leq t}|X_s|^2\right)\leq C(1+E(|X_0|^2))$$

with $T<\infty$, where C is a finite positive constant.

Moreover, we have the moment estimates theorem

Theorem 1.6.4. *Suppose that the coefficient function $a(\cdot), b(\cdot)$ and $c(\cdot)$ of SDE (1.11) satisfy the Lipschitz conditions (1.14), the linear growth conditions (1.17) and the initial condition. Then the SDE (1.11) admits a unique strong solution. Additionally, let for $n\in\{1,2,\cdots\}$*

$$E(|X_0|^{2n})<\infty$$

then the solution X_t satisfies

$$E\left(\sup_{0\leq s\leq t}|X_s|^{2n}\right)\leq C(1+E(|X_0|^{2n}))$$

for $t\in[0,T]$, with $T<\infty$, where C is a positive constant depending only on T,n and the linear growth bound.

Chapter 2
Double-Jump Diffusion Model for VIX: Evidence from VVIX

2.1 Introduction

Modeling the VIX index and its derivatives has become increasingly popular among researchers. As a measure of the market's expectations for the 30-day implied volatility of the S&P500 index, VIX provides rich information for the prediction of future market trends. VIX can be seen as a compression of the information involved in S&P500 options. Usually, VIX and the S&P500 index are negatively correlated, meaning that the VIX index is often referred to as the fear index or the fear gauge. For more about VIX, see for example [Carr and Wu 2005].

In considering VIX's importance, researchers have focused significant attention on directly modelling the dynamics of VIX. Earlier work has sought to use the geometric Brownian motion, square-root diffusion, or log-normal Ornstein-Uhlenbeck (OU) diffusion to model VIX. Some researchers have also considered jumps in VIX. Recently, an innovative parameterized stochastic volatility model for VIX was put forward by [Mencía and Sentana 2013] and [Kaeck and Alexander 2013]. These authors specified a new process for modeling the volatility of VIX, which may be correlated with VIX, thus revealing its empirical advantage over other traditional models. Both sets of authors have also pointed out that stochastic volatility reduces the impact of the jump on VIX. However, the specifications of this innovative stochastic volatility model are different depending on the researchers. [Mencía and Sentana 2013] have adopted a pure-jump OU process stemming from analytical

treatability, while [Kaeck and Alexander 2013] have characterized volatility as a square-root diffusion that takes the correlation of the volatility and VIX into account. How to better specify and estimate this volatility factor and further interpret the dynamics of VIX remains an unresolved problem.

In 2012, Chicago Board Options Exchange (CBOE) introduced a new volatility index named VVIX onto the market. VVIX measures the 30-day implied volatility of the VIX index. [D. Huang and Shaliastovich 2014] have constructed the realized volatility of the VIX index (i.e. the realized volatility of volatility), showing that the VIX index itself is not a good predictor of realized volatility. Rather, VVIX serves as a better candidate for this. We can thus infer from the researchers' empirical conclusion that the VVIX index may provide additional information about the volatility of VIX beyond that provided by VIX itself.

To emphasis the role of VVIX on quantifying the stochastic volatility of VIX, first we recalibrate the stochastic volatility out of logVIX model in [Kaeck and Alexander 2013] using the same prior and sampling method specified in [Kaeck and Alexander 2010]. Here we choose the VIX data from January 2007 to September 2014 on account of the VVIX data starting from January 2007. We then plot both the estimated stochastic volatility and VVIX time series in Figure 2.1. As a result, the correlation between the posterior volatility and VVIX index in this experiment is only 0.4193. Although some of peaks of the estimated volatility coincide with VVIX, there is more inconsistency. This indicates that when we use the VIX index as the sole data source for sampling the latent stochastic volatility, limited information about the dynamics of stochastic volatility is obtained. By the observation, we suspect the stochastic volatility was not just spanned by the VIX in some sense. To obtain a more accurate volatility of VIX, we should explore the relationship between it and the VVIX index.

In this paper we make an extensive empirical analysis of the dynamics of VIX, concentrating particularly on modeling its stochastic volatility under the physical measure via additional information provided by the VVIX index. Our contribution consists of three aspects. First, we find evidence of the co-jumps between the VIX and VVIX index through statistical test of the historical data of both indices. Second, we show that the VVIX index and the volatility of VIX satisfy the criteria of a linear relationship under the general affine assumption on the dynamics of the logarithm of the VIX index and its stochastic volatility. Thus the modeling of VIX

and its volatility could be transformed into the joint modeling of the VIX and VVIX index. Empirically, both of VIX and VVIX are mean-reverting and have co-jumps. Based on these facts we propose a double-jump stochastic volatility model for the VIX and its volatility. Third, we provide a Markov-Chain-Monte-Carlo (MCMC) method to estimate the double-jump model and its nested models using historical data regarding VIX and VVIX. We obtain both a unified set of model parameters and a series of outcomes of latent variables such as stochastic volatility, jump intensity and jump sizes. The results can be exploited to understand the economics of the market further. We compare the model performance through several criterion such as residual analysis, p-value and deviance information criterion (DIC) method. We show that the jumps in volatility of VIX is statistically significant and the jump intensity is not deterministic which may imply a more complex structure of it.

Two papers that are mostly relevant to our work are [Mencía and Sentana 2013] and [Kaeck and Alexander 2013]. Both studies conclude that modeling the logarithm of the VIX index is better than modeling VIX directly and modeling VIX with stochastic volatility performs better in the sense of fitting historical data of the VIX market. [Mencía and Sentana 2013] use a pure-jump OU model to describe the volatility of VIX. They apply maximum and pseudo-maximum likelihood for estimating parameters and use extended Kalman filter to generate the outcome of stochastic volatility. The data sources are VIX and its derivatives. [Kaeck and Alexander 2013] model this volatility with a square-root diffusion and estimate the model using MCMC method and nearly 20 years data of VIX. Compared to their work, we obtain an expression of the VVIX index in terms of the affine model of the stochastic volatility and make more reasonable assumptions of the joint dynamics of both indices, especially the co-jumps. For empirical analysis, we calibrate our model using both information of VIX and VVIX. We also justify our model specification with several statistical tests and make a detailed analysis of the latent variables.

The structure of this paper is as follows: Section 2.2 shows the linear relationship between the VVIX index and stochastic volatility of the VIX index. Section 2.3 analyzes model specifications and sets up our model. Section 2.4 gives our empirical method. Section 2.5 describes the VIX and VVIX data that we use in this paper. Section 2.6 summarizes the estimation results and provides our empirical analysis. Section 2.7 outlines the conclusions and implications of our study.

2.2 Linear Relationship between VVIX and Volatility of VIX

In this part and following of the paper we build a model based on the logarithm of VIX instead of VIX directly. We show that if the logVIX mean-reverts to a constant central tendency with stochastic volatility as well as jumps in logVIX itself and volatility, then there will be a linear relationship between the VVIX index and the stochastic volatility factor of the logVIX. This relationship can serve as a gauge for determining whether a particular stochastic volatility model for VIX is reliable. A similar argument regarding how to find a proxy for some unobservable factor has been made in [Ait-Sahalia and Kimmel 2007; Duan and Yeh 2010] and [Ait-Sahalia, Karaman, and Mancini 2014]. In this paper, the VIX index is the underlying factor; this means its dynamics are observed under the P measure. As VVIX is compiled from VIX options, which are calculated under the pricing measure Q, all of the following derivations involving VVIX are likewise computed under the Q measure.

Let $Y(t) = \log VIX(t)$ and assume that under Q, $Y(t)$ and $\omega(t)$ follow a general affine jump diffusion model

$$\begin{aligned} dY(t) &= \kappa_V(\theta - Y(t))\,dt + \sqrt{\omega(t)}dW_Y^Q(t) + J_Y^Q dN(t) \\ d\omega(t) &= \left(\alpha_\omega - \kappa_\omega^Q \omega(t)\right)dt + \sigma_\omega\sqrt{\omega(t)}dW_\omega^Q(t) + J_\omega^Q dN(t) \end{aligned} \tag{2.1}$$

where we assume $\left\langle dW_Y^Q(t), dW_\omega^Q(t)\right\rangle = \rho dt$. $N(t)$ is a Poisson process with stochastic jump intensity $\lambda(t) = \lambda_0 + \lambda_1\omega(t)$ at time t for analytical tractability. The jump magnitudes for logVIX and its volatility factor are characterized by

$$J_Y^Q \sim N\left(\mu_y^J, \left(\sigma_y^J\right)^2\right)$$
$$J_\omega^Q \tilde{} N\left(\mu_\omega^J, \left(\sigma_\omega^J\right)^2\right)$$

Similar to the idea that regarding VIX square as the expectation of quadratic variation of the logarithm of the S&P500 index under the pricing measure approximately (see, for example, [Ait-Sahalia, Karaman, and Mancini 2014]), we set the

theoretical VVIX value to be

$$VVIX^2_{t,t+\tau} = \frac{1}{\tau}\left[E^Q_t\left(\int_t^{t+\tau}\omega(s)\,ds\right) + E^Q_t\left(\sum_{s\geq 0}\triangle Y^2(s)\right)\right] \tag{2.2}$$

With the simple calculation from (2.1) we obtain

$$\begin{aligned} E^Q_t\left(\int_t^{t+\tau}\omega(s)\,ds\right) &= \frac{1-e^{-\left(\kappa^Q_\omega-\lambda_1\mu^J_\omega\right)\tau}}{\kappa^Q_\omega-\lambda_1\mu^J_\omega}\omega(t) + \left(\tau - \frac{1-e^{-\left(\kappa^Q_\omega-\lambda_1\mu^J_\omega\right)\tau}}{\kappa^Q_\omega-\lambda_1\mu^J_\omega}\right)\frac{\alpha_\omega+\lambda_0\mu^J_\omega}{\kappa^Q_\omega-\lambda_1\mu^J_\omega} \\ &\triangleq \alpha_Q(\tau)\,\omega(t) + \beta_Q(\tau) \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} E^Q_t\left(\sum_{s\geq 0}\triangle Y^2(s)\right) &= \left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)E^Q_t\left(\int_t^{t+\tau}\left(\lambda_0+\lambda_1\omega(s)\right)ds\right) \\ &= \left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\left(\lambda_0\tau + \lambda_1\beta_Q(\tau) + \lambda_1\alpha_Q(\tau)\,\omega(t)\right) \end{aligned} \tag{2.4}$$

where $\alpha_Q(\tau) = \frac{1-e^{-\left(\kappa^Q_\omega-\lambda_1\mu^J_\omega\right)\tau}}{\kappa^Q_\omega-\lambda_1\mu^J_\omega}$ and $\beta_Q(\tau) = \left(\tau - \frac{1-e^{-\left(\kappa^Q_\omega-\lambda_1\mu^J_\omega\right)\tau}}{\kappa^Q_\omega-\lambda_1\mu^J_\omega}\right)\frac{\alpha_\omega+\lambda_0\mu^J_\omega}{\kappa^Q_\omega-\lambda_1\mu^J_\omega}$.

After combining (2.2), (2.3) and (2.4), we have

$$\begin{aligned} VVIX^2_{t,t+\tau} &= \frac{1}{\tau}\left[\alpha_Q(\tau)\,\omega(t) + \beta_Q(\tau) + \left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\left(\lambda_0\tau + \lambda_1\beta_Q(\tau) + \lambda_1\alpha_Q(\tau)\,\omega(t)\right)\right] \\ &= \frac{1}{\tau}\left(\beta_Q(\tau) + \left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\left(\lambda_0\tau + \lambda_1\beta_Q(\tau)\right)\right) \\ &\quad + \frac{1}{\tau}\left(1+\lambda_1\left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\right)\alpha_Q(\tau)\,\omega(t) \qquad (2.5) \\ &\triangleq A(\tau) + B(\tau)\,\omega(t) \qquad (2.6) \end{aligned}$$

where

$$\begin{cases} A(\tau) = \lambda_0\left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right) + \frac{1}{\tau}\left(1+\lambda_1\left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\right)\beta_Q(\tau) \\ B(\tau) = \frac{1}{\tau}\left(1+\lambda_1\left(\left(\mu^J_y\right)^2 + \left(\sigma^J_y\right)^2\right)\right)\alpha_Q(\tau) \end{cases} \tag{2.7}$$

only depend on time to maturity τ and parameters.

Relationship (2.6) can be seen as a benchmark for the estimated volatility

factor. The dynamics of VVIX reflect the property of $\omega(t)$ more directly than the VIX option does. It also provides more intuitive empirical evidence for the model specifications for $\omega(t)$, as will be seen in Section 2.3.

2.3 Model Specification and Setup

2.3.1 Model Specification

The log-normal Ornstein-Uhlenbeck model has been proposed by [Detemple and Osakwe 2000]. Modelling the logarithm of VIX or VIX futures has also been considered in [Psychoyios, Dotsis, and Markellos 2010] and [Huskaj and Nossman 2013]. [Mencía and Sentana 2013] and [Kaeck and Alexander 2013] have compared and examined different model specifications for VIX dynamics. Both sets of authors have concluded that the setup for modelling logVIX as an affine jump process is superior to the setup for modelling VIX directly. This conclusion is consistent among all of the model specifications.. In our model, we also study the affine property of logVIX.

Since there is a linear relationship between VVIX and $\omega(t)$, the VVIX index can be considered a proxy for this unobservable variable. Jointly modelling the VIX index and its stochastic volatility is thus equivalent to jointly modelling the VIX index and the VVIX index. In this sense, the model should reflect some of their joint properties.

Both VIX and VVIX have the mean-reverting property. VIX may mean-revert to a constant or stochastic central tendency. [Mencía and Sentana 2013] make both assumptions, examining the model's performance in this case. As their data is comprised of VIX, VIX futures, and VIX options, the stochastic central tendency of VIX model performs better. In fact, the specifications of the central tendency of VIX are mainly characterized by the information from VIX futures, while the VIX options play a relatively less influential role. However, based on the derivation in Section 2.2, we know that the expression of the VVIX index is irrelevant to the drift part of VIX. In fact, this is consistent with its stochastic volatility role. As our paper mainly concerns the impact of VVIX data on the estimation, we make a simple assumption that the VIX mean-reverts to a constant central tendency. Since VVIX has a similar empirical property, we assume the stochastic volatility of VIX

also has a constant central tendency.

Based on the historical data regarding VIX and VVIX collected daily from January 2007 to November 2014, we can observe that there is a co-jump between the two indices, no matter whether a positive or negative jump occurs. To conduct a formal test to verify this phenomenon, we adapt the method in [Bollerslev, Law, and Tauchen 2008] to our lower sampling frequency (see also [Gilder 2009] for an illustration of using this method with daily data). The testing procedure is divided into two parts: first, we show that there are jumps in both VIX and VVIX; secondly, we demonstrate that VIX and VVIX have common jumps.

As for the test on whether there are significant jumps, the general procedure is: we assume that a process X (to be VIX or VVIX) is observed daily $[0, T]$ at times $t = 0, 1, \ldots, T$ and denote the time series by $X_t, t = 1, 2, \ldots, T$. The return process $r_t = X_t - X_{t-1}, t = 1, 2, \ldots, T$ is also defined. Define the n-day rolling sample estimates of realized volatility,

$$RV_t = \sum_{k=0}^{n} r_{t-k}^2 \tag{2.8}$$

and bipower variation

$$BV_t = \frac{\pi}{2} \sum_{k=0}^{n-1} |r_{t-k}| \, |r_{t-k-1}| \tag{2.9}$$

The relative contribution measure

$$RJ_t = \frac{RV_t - BV_t}{RV_t} \tag{2.10}$$

immediately follows from (2.8) and (2.9). The tripower quarticity for daily changes is defined by

$$TP_t = \mu_{4/3}^{-3} \frac{n^2}{n-2} \sum_{k=0}^{n-1} |r_{t-k}|^{4/3} \, |r_{t-k-1}|^{4/3} \, |r_{t-k-2}|^{4/3} \tag{2.11}$$

where $\mu_{4/3} = 2^{2/3} \Gamma\left(\frac{7}{6}\right) \Gamma\left(\frac{1}{2}\right)$. Finally, the statistic

$$z_t = \frac{RJ_t}{\sqrt{\left[(\pi/2)^2 + \pi - 5\right] \frac{1}{n} \max\left(1, \frac{TP_t}{BV_t^2}\right)}} \tag{2.12}$$

is determined using (2.9), (2.10), and (2.11) and tests whether a jump occurs at day

t. We reject the null hypothesis of no jumps at $\alpha\%$ confidence level if $|z_t| > \Phi^{-1}_{1-\alpha/2}$ where Φ is the cumulative normal distribution for a given day t.

To implement the second part of testing co-jump, we denote VIX and VVIX by X^1 and X^2. Assume VIX and VVIX are observed daily $[0, T]$ at times $t = 1, 2, \ldots, T$; the time series is thus $X^i_t, t = 1, 2, \ldots, T, i = 1, 2$, respectively. Given the return processes $r^i_t = X^i_t - X^i_{t-1}, t = 1, 2, \ldots, T, i = 1, 2$, we can calculate the contemporaneous correlation

$$cp_t = \sum_{k=0}^{n-1} r^1_{t-k} r^2_{t-k}$$

and study the studentized statistic

$$z_{cp,t} = \frac{cp_t - \overline{cp}}{s_{cp}} \tag{2.13}$$

where

$$\overline{cp} = \frac{1}{T-(n-1)} \sum_{t=n}^{T} cp_t$$

and

$$s_{cp} = \left[\frac{1}{T-(n-1)} \sum_{t=n}^{T} (cp_t - \overline{cp})^2 \right]^{1/2}$$

at time t. We reject the null hypothesis of no common jumps at $\alpha\%$ confidence level if $|z_{cp,t}| > \Phi^{-1}_{1-\alpha/2}$. For details about the two tests, please refer (120).

We test the jump behavior of VIX and VVIX from March 3, 2007 to November 26, 2014. Given the 5% significance level, 222 days for VIX and 141 days for VVIX out of 1939 days total indicate a significant jump for the first step. In the second step, 131 days call for a co-jump. Thus the specification for a co-jump in VIX and VVIX is justified. This phenomenon provides an important foundation for our model setup.

2.3.2 Basic Model

Section 2.3.1 demonstrates that in addition to the diffusion part, we should assume jumps in both $Y(t)$ and $\omega(t)$ and moreover, that the jumps are dominated by a single Poisson process. The jump intensity may be constant or state-dependent on the affine factor. In this paper, we assume the jump is affected by $\omega(t)$. The assumption of constant or stochastic jump intensity is examined below. As both

positive and negative jumps occur, we make the normal distribution assumption regarding the jump size. For logVIX, this may be a sensible assumption. For square root diffusion plus a jump for $\omega(t)$, given that a jump is a rare event for the historical path, this assumption is also acceptable. For previous work on jumps in volatility, we refer to [Duffie, Pan, and K. Singleton 2000; Eraker, M. Johannes, and Nicholas Polson 2003; Eraker 2004; Todorov and Tauchen 2011] and [Amengual and Xiu 2014].

We thus assume that under measure Q,

$$\begin{aligned} dY(t) &= \kappa_V(\theta - Y(t))dt + \sqrt{\omega(t)}dW_Y^Q(t) + J_Y^Q dN(t) \\ d\omega(t) &= \left(\alpha_\omega - \kappa_\omega^Q \omega(t)\right)dt + \sigma_\omega\sqrt{\omega(t)}dW_\omega^Q(t) + J_\omega^Q dN(t) \end{aligned} \tag{2.14}$$

where $\left\langle dW_Y^Q(t), dW_\omega^Q(t)\right\rangle = \rho dt$ and $N(t)$ is a Poisson process with stochastic jump intensity $\lambda(t) = \lambda_0 + \lambda_1\omega(t)$ at time t.

$$\begin{aligned} J_Y^Q &\sim N\left(\mu_y^J, \left(\sigma_y^J\right)^2\right) \\ J_\omega^Q &\tilde{} N\left(\mu_\omega^J, \left(\sigma_\omega^J\right)^2\right) \end{aligned}$$

We specify the risks of price between Q and P regarding Brownian motions as

$$\begin{aligned} dW_Y^Q(t) &= dW_Y^P(t) - \varsigma_V\sqrt{\omega(t)}dt \\ dW_\omega^Q(t) &= dW_\omega^P(t) - \varsigma_\omega\sqrt{\omega(t)}dt \end{aligned}$$

then under P,

$$\begin{aligned} dY(t) &= \left[\kappa_V(\theta - Y(t)) - \varsigma_V\omega(t)\right]dt + \sqrt{\omega(t)}dW_Y^P(t) + J_Y^P dN(t) \\ d\omega(t) &= \left(\alpha_\omega - \kappa_\omega^P\omega(t)\right)dt + \sigma_\omega\sqrt{\omega(t)}dW_\omega^P(t) + J_\omega^P N(t) \end{aligned} \tag{2.15}$$

where $\left\langle dW_Y^P(t), dW_\omega^P(t)\right\rangle = \rho dt$ and $N(t)$ is a Poisson process with stochastic jump intensity $\lambda(t) = \lambda_0 + \lambda_1\omega(t)$ at time t. $\kappa_\omega^P = \kappa_\omega^Q + \varsigma_\omega\sigma_\omega$ is the speed of mean reversion under P. The jump sizes are characterized by

$$J_Y^P \sim N\left(\mu_y^{JP}, \left(\sigma_y^J\right)^2\right)$$

$$J_\omega^P \sim N\left(\mu_\omega^{JP}, \left(\sigma_\omega^J\right)^2\right)$$

The parameter set under P is denoted by

$$\Theta_P = \left\{\kappa_V, \varsigma_V, \theta, \kappa_\omega^P, \mu_y^{JP}, \mu_\omega^{JP}, \sigma_\omega^J, \rho, \sigma_\omega\right\} \tag{2.16}$$

while the parameter set under Q is summarized as

$$\Theta_M = \left\{\alpha_\omega, \kappa_\omega^Q, \lambda_0, \lambda_1, \mu_y^J, \mu_\omega^J, \sigma_y^J\right\} \tag{2.17}$$

We also assume that there exists a residual or pricing error between the observed VVIX which is denoted by $VVIX_{t,t+\tau}^{obs}$ and our theoretical value, i.e.

$$\left(VVIX_{t,t+\tau}^{obs}\right)^2 = A(\tau) + B(\tau)\,\omega(t) + \varepsilon_t \tag{2.18}$$

where the expressions of $A(\tau)$ and $B(\tau)$ are given by (2.7). As we concern more about the statistical performance of approximating VVIX with the linear expression of $\omega(t)$, we assume simply here that the pricing error is i.i.d., i.e.

$$\varepsilon_t \sim N\left(0, \sigma_P^2\right) \text{ for all } t$$

The further discussion of the econometric property of the time series of this pricing error is beyond this paper and may be left for further research. For the model (2.18), we need to estimate

$$\Theta_E = \{\sigma_P\} \tag{2.19}$$

In the following of the paper, the general model (2.15) was called SVJJ-S model (stochastic λ). If we let $\lambda_1 = 0$, the model reduces to the SVJJ-C model (constant λ) model. If we further let $J_\omega^P\left(J_\omega^Q\right) = 0$, the model collapses to the SVJ-C model (constant λ) model. Finally, when there are no jumps, i.e., $J_y = J_\omega = 0$, the model becomes the SV model. We want to give empirical analysis of these models using the real-market historical data of VIX and VVIX to show that whether the additional jump parts of VIX and $\omega(t)$ can improve the VIX modeling significantly and whether the jump intensity is constant or stochastic.

2.4 Model Inference with VIX and VVIX

In this section, we use the data of VIX and VVIX indices from January 3, 2007, to November 26, 2014, to estimate the models. For the daily VVIX, $\tau = \frac{1}{12}$ in (2.18), so we denote by $\left(VVIX_t^{obs}\right)^2 := \left(VVIX_{t,t+1/12}^{obs}\right)^2$ for brevity. In total, we draw from 1,991 daily observations each for VIX and the VVIX index. We adopt the MCMC method as our estimation method. Compared with the maximum-likelihood estimation (MLE) method, the generalized method of moments (GMM), and some additional methods, MCMC has two advantages that are particularly suitable for our study. First, not only does MCMC estimate the unknown parameters, it also provides posterior estimated latent variables such as stochastic volatility, jump times, and jump sizes. These variables are fundamental for subsequent empirical analysis and model comparison. Secondly, MCMC is very efficient for implementation. For more details about applications of the MCMC method in finance, we refer to [M. S. Johannes and Nick Polson 2003; Eraker, M. Johannes, and Nicholas Polson 2003] and [Amengual and Xiu 2012].

With the parameters set denoted by $\mathbf{\Theta} = (\Theta_P, \Theta_M, \Theta_E)$, the latent variables by $\mathbf{Z}$, and the observed data by $\mathbf{Y} = (VIX, VVIX)$, for some model M we are interested in the joint posterior of parameters and latent variables given data:

$$p\left(\mathbf{\Theta}, \mathbf{Z} | \mathbf{Y}, M\right) \propto p\left(\mathbf{Y} | \mathbf{\Theta}, \mathbf{Z}, M\right) \cdot p\left(\mathbf{\Theta}, \mathbf{Z} | M\right)$$

We assume market data are observed daily. Let the time interval $\Delta = 1/252$ be one day, with the assumption that we have T observations $Y_{i\Delta}, 1 \leq i \leq T$ for the logarithm of VIX and $VVIX_{i\Delta}^{obs}, 1 \leq i \leq T$ for VVIX respectively. A time discretization of the dynamics (2.15) with time interval Δ gives

$$\begin{aligned} Y_{i\Delta} - Y_{(i-1)\Delta} &= \left(\kappa_V \theta - \kappa_V Y_{(i-1)\Delta} - \varsigma_V \omega_{(i-1)\Delta}\right)\Delta + \sqrt{\omega_{(i-1)\Delta}\Delta}\epsilon_{i\Delta}^y + j_{i\Delta}^y n_{i\Delta} \\ \omega_{i\Delta} - \omega_{(i-1)\Delta} &= \left(\alpha_\omega - \kappa_\omega^P \omega_{(i-1)\Delta}\right)\Delta + \sigma_\omega \sqrt{\omega_{(i-1)\Delta}\Delta}\epsilon_{i\Delta}^\omega + j_{i\Delta}^\omega n_{i\Delta} \end{aligned} \tag{2.20}$$

where $\epsilon_{i\Delta}^y$ and $\epsilon_{i\Delta}^\omega$ are correlated Normal variables with correlation ρ, $j_{i\Delta}^y$ and $j_{i\Delta}^\omega$ are normal with different parameters.

Denote by $\widetilde{Y}_{i\Delta} = Y_{i\Delta} - j_{i\Delta}^y n_{i\Delta}$ for $2 \leq i \leq T$ and $\widetilde{\omega}_{i\Delta} = \omega_{i\Delta} - j_{i\Delta}^\omega n_{i\Delta}$ for

$2 \leq i \leq T$, then move from (2.20) to the jump-adjusted processes

$$\begin{aligned} \widetilde{Y}_{i\Delta} &= a_0 + a_1 Y_{(i-1)\Delta} + a_2 \omega_{(i-1)\Delta} + \sqrt{\omega_{(i-1)\Delta}\Delta}\epsilon^y_{i\Delta} \\ \widetilde{\omega}_{i\Delta} &= c_0 + c_1 \omega_{(i-1)\Delta} + \sigma_\omega \sqrt{\omega_{(i-1)\Delta}\Delta}\epsilon^\omega_{i\Delta} \end{aligned} \tag{2.21}$$

where $a_0 = \kappa_V \theta \Delta$, $a_1 = 1 - \kappa_V \Delta$, $a_2 = -\varsigma_V \Delta$, $c_0 = \alpha_\omega \Delta$, $c_1 = 1 - \kappa^P_\omega \Delta$. A time discretization of (2.18) gives

$$\left(VVIX^{obs}_{i\Delta}\right)^2 = A(\tau) + B(\tau)\,\omega_{i\Delta} + \varepsilon_{i\Delta}$$

where $\varepsilon_{i\Delta} \sim N(0, \sigma^2_P)$. Next, we apply $Y_{i\Delta}$ and $VVIX^{obs}_{i\Delta}, 1 \leq i \leq T$ to estimate latent variables

$$\begin{aligned} \omega_{i\Delta},\ 1 &\leq i \leq T \\ n_{i\Delta}, j^y_{i\Delta} \text{ and } j^\omega_{i\Delta},\ 2 &\leq i \leq T \end{aligned}$$

and parameters $\mathbf{\Theta} = (\Theta_P, \Theta_M, \Theta_E)$.

As the joint posterior distribution $p(\mathbf{\Theta}, \mathbf{Z}|\mathbf{M})$ is not known in closed-form, the MCMC algorithm samples these parameters and latent variables sequentially from the posterior conditional distributions as follows:

$$\begin{aligned} \text{spot volatility} &: p\left(\omega^{(g)}_{i\Delta} | \omega^{(g)}_{<i\Delta}, \omega^{(g-1)}_{>i\Delta}, n^{(g-1)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\ \text{jump time} &: p\left(n^{(g)}_{i\Delta} | \omega^{(g)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\ \text{jump size in VIX} &: p\left(j^{y(g)}_{i\Delta} | n^{(g)}, \omega^{(g)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\ \text{jump size in volatility} &: p\left(j^{\omega(g)}_{i\Delta} | n^{(g)}, \omega^{(g)}, j^{y(g)}, \Theta^{(g-1)}, Y\right) \\ \text{parameters} &: p\left(\Theta^{(g)} | n^{(g-1)}, \omega^{(g)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \end{aligned}$$

where g represents the g-th iteration with total of M times. In this study, we sample the data 5000 times ($M = 5000$) and discard the first 2000 samples.

2.4.1 Estimation Strategy

In this subsection we consider the sampling method for the latent variables and parameters. For the prior of the parameters, we follow the assumption of several

previous literature. The prior of jump time and jump size is obtained through analysis of historical VVIX data by digging out large jumps. The prior of stochastic volatility $\omega(t)$ is set to be the square VVIX. Next we will provide details for the sampling schemes we discuss the corresponding algorithms for stochastic volatility ω_t, jump times, jump sizes, and parameters set Θ_P in (2.16), the Q-parameters Θ_M in (2.17) and the pricing error parameter Θ_E in (2.19).

2.4.1.1 Sampling Volatility

Sampling stochastic volatility ω_t necessitates taking information from both VIX and VVIX into consideration. Utilizing the linear relationship in (2.18), we employ the random-walk Metropolis-Hastings algorithms to sample ω_t. When we have obtained the first $(g-1)$-th samples, for the g-th sample, let $\omega_{(-i)}^{(g-1)} = \left(\omega_{1\Delta}^{(g)}, \cdots, \omega_{(i-1)\Delta}^{(g)}, \omega_{(i+1)\Delta}^{(g-1)}, \cdots, \omega_{T\Delta}^{(g-1)}\right)$, $g = 1, 2, \cdots, M$. We specify the full conditional density as

$$
\begin{aligned}
& p\left(\omega_{i\Delta}^{(g)} | \omega_{(-i)}^{(g-1)}, n^{(g-1)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\
\propto\ & \frac{1}{\omega_{i\Delta}^{(g)}} \exp\left[-\frac{\left(C_{i\Delta}^2 + D_{i\Delta}^2 - 2\rho C_{i\Delta} D_{i\Delta}\right)}{2\left(1-\rho^2\right)}\right] \exp\left[-\frac{\left(C_{(i+1)\Delta}^2 + D_{(i+1)\Delta}^2 - 2\rho C_{(i+1)\Delta} D_{(i+1)\Delta}\right)}{2\left(1-\rho^2\right)}\right] \\
& \cdot \exp\left(-\frac{\left(\left(VVIX_{i\Delta}^{obs}\right)^2 - A(\tau) - B(\tau)\,\omega_{i\Delta}^{(g)}\right)^2}{2\sigma_P^2}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
C_{i\Delta} &= \frac{Y_{i\Delta} - j_{i\Delta}^{y(g-1)} n_{i\Delta}^{(g-1)} - a_0 - a_1 Y_{(i-1)\Delta} - a_2 \omega_{(i-1)\Delta}^{(g)}}{\sqrt{\omega_{(i-1)\Delta}^{(g)} \Delta}} \\
D_{i\Delta} &= \frac{\omega_{i\Delta}^{(g)} - j_{i\Delta}^{\omega(g-1)} n_{i\Delta}^{(g-1)} - c_0 - c_1 \omega_{(i-1)\Delta}^{(g)}}{\sigma_\omega \sqrt{\omega_{(i-1)\Delta}^{(g)} \Delta}}
\end{aligned}
$$

and

$$
C_{(i+1)\Delta} = \frac{Y_{(i+1)\Delta} - j_{(i+1)\Delta}^{y(g-1)} n_{(i+1)\Delta}^{(g-1)} - a_0 - a_1 Y_{i\Delta} - a_2 \omega_{i\Delta}^{(g)}}{\sqrt{\omega_{i\Delta}^{(g)} \Delta}}
$$

$$D_{(i+1)\Delta} = \frac{\omega^{(g-1)}_{(i+1)\Delta} - j^{\omega(g-1)}_{(i+1)\Delta} n^{(g-1)}_{(i+1)\Delta} - c_0 - c_1 \omega^{(g)}_{i\Delta}}{\sigma_\omega \sqrt{\omega^{(g)}_{i\Delta} \Delta}}$$

for $2 \leq i \leq T-1$. The case for $i = 1$ and $i = T$ is similar.

$$\begin{aligned} & p\left(\omega^{(g)}_{1\Delta} | \omega^{(g-1)}_{(-1)}, n^{(g-1)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\ \propto & \frac{1}{\omega^{(g)}_{1\Delta}} \exp\left[-\frac{(C^2_{1\Delta} + D^2_{1\Delta} - 2\rho C_{1\Delta} D_{1\Delta})}{2(1-\rho^2)}\right] \exp\left(-\frac{\left(\left(VVIX^{obs}_{1\Delta}\right)^2 - A(\tau) - B(\tau)\omega^{(g)}_{1\Delta}\right)^2}{2\sigma_P^2}\right) \end{aligned}$$

and

$$\begin{aligned} & p\left(\omega^{(g)}_{T\Delta} | \omega^{(g-1)}_{(-T)}, n^{(g-1)}, j^{y(g-1)}, j^{\omega(g-1)}, \Theta^{(g-1)}, Y\right) \\ \propto & \frac{1}{\omega^{(g)}_{T\Delta}} \exp\left[-\frac{(C^2_{T\Delta} + D^2_{T\Delta} - 2\rho C_{T\Delta} D_{T\Delta})}{2(1-\rho^2)}\right] \exp\left(-\frac{\left(\left(VVIX^{obs}_{T\Delta}\right)^2 - A(\tau) - B(\tau)\omega^{(g)}_{T\Delta}\right)^2}{2\sigma_P^2}\right) \end{aligned}$$

2.4.1.2 Sampling Jump Times

Since the model assumes the co-jump condition on VIX index and its volatility, we just need to sample the co-jump parameter, which will describe the jumps of the two. Given the information the $(g-1)$th sampling, for the case where $i = 2, 3, \cdots, T$, the jump time of day i, $n_{i\Delta}$, the probability where there is a jump on day i is

$$\begin{aligned} & p(n_{i\Delta} = 1 | X, \Theta_P, Y) \\ = & \frac{p(Y_{i\Delta}, \omega_{i\Delta} | Y_{(i-1)\Delta}, \omega_{(i-1)\Delta}, n_{i\Delta} = 1, j^y_{(i-1)\Delta}, j^\omega_{(i-1)\Delta}, \Theta_P) * p(n_{i\Delta} = 1 | \omega_{(i-1)\Delta}, Y_{(i-1)\Delta})}{\sum_{s=0}^{1} p(Y_{i\Delta}, \omega_{i\Delta} | Y_{(i-1)\Delta}, \omega_{(i-1)\Delta}, n_{i\Delta} = s, j^y_{(i-1)\Delta}, j^\omega_{(i-1)\Delta}, \Theta_P) * p(n_{i\Delta} = s | \omega_{(i-1)\Delta}, Y_{(i-1)\Delta})} \end{aligned}$$

where $\{(Y_{i\Delta}, \omega_{i\Delta}) | Y_{(i-1)\Delta}, \omega_{(i-1)\Delta}, n_{i\Delta} = s, j^y_{(i-1)\Delta}, j^\omega_{(i-1)\Delta}, \Theta_P\}$ follows the following bivariate normal distribution, whose mean is

$$\begin{bmatrix} a_0 + a_1 Y_{(i-1)\Delta} + a_2 \omega_{(i-1)\Delta} + s j^y_{i\Delta} \\ c_0 + c_1 \omega_{(i-1)\Delta} + s j^\omega_{i\Delta} \end{bmatrix}$$

covariance matrix is

$$\omega_{(i-1)\Delta}\Delta \begin{bmatrix} 1 & \rho\sigma_\omega \\ \rho\sigma_\omega & \sigma_\omega^2 \end{bmatrix}$$

and $p(n_{i\Delta} = 1|\omega_{(i-1)\Delta}, Y_{(i-1)\Delta}) = (\lambda_0 + \lambda_1\omega_{(i-1)\Delta})\Delta$.

2.4.1.3 Sampling Jump Sizes

If there is a jump on the ith day, we will use a normal distribution $N(\frac{B}{A}, \frac{1}{A})$ to sample $j^{\omega}_{i\Delta}$, where

$$\begin{aligned} A &= \frac{1}{\sigma_\omega^2\omega_{(i-1)\Delta}\Delta} + \frac{1}{(\sigma_\omega^J)^2} \\ B &= \frac{\omega_{i\Delta} - c_0 - c_1\omega_{(i-1)\Delta}}{\sigma_\omega^2\omega_{(i-1)\Delta}\Delta} + \frac{\mu_\omega^{JP}}{(\sigma_\omega^J)^2} \end{aligned}$$

Then, use the sampled $j^{\omega}_{i\Delta}$, we can use the normal distribution $N(\frac{B}{A}, \frac{1}{A})$ to sample $j^{y}_{i\Delta}$

$$\begin{aligned} A &= \frac{1}{\omega_{(i-1)\Delta}(1-\rho^2)\Delta} + \frac{1}{(\sigma_y^J)^2} \\ B &= \frac{Y_{i\Delta} - a_0 - a_1Y_{(i-1)\Delta} - a_2\omega_{(i-1)\Delta} - \frac{\rho}{\sigma_\omega}(\omega_{i\Delta} - c_0 - c_1\omega_{(i-1)\Delta} - j^{\omega}_{i\Delta})}{\omega_{(i-1)\Delta}(1-\rho^2)\Delta} + \frac{\mu_y^{JP}}{(\sigma_y^J)^2} \end{aligned}$$

if there is not a jump, the posterior distribution would be the same as its prior distribution, because the current VIX and volatility provide no further information to update jump size, that is:

$$\begin{aligned} J_Y^P &\sim N(\mu_y^{JP}, (\sigma_y^J)^2) \\ J_\omega^P &\sim N(\mu_\omega^{JP}, (\sigma_\omega^J)^2) \end{aligned}$$

No matter whether there is a jump on day i, we still need to update μ_y^{JP}, σ_y^J, μ_ω^{JP} and σ_ω^J.

2.4.1.4 Sampling Parameters under P Measure

The parameter set under P measure is $\Theta_P = \left\{\kappa_V, \varsigma_V, \theta, \kappa_\omega^P, \mu_y^{JP}, \mu_\omega^{JP}, \sigma_\omega^J, \rho, \sigma_\omega\right\}$, next, we will explain the details of sampling each parameter, here we will ignore the notation g for sampling count.

- Sample θ

  Suppose the prior distribution of θ is $N(\mu_\theta, \sigma_\theta^2)$, we will use posterior distribution $\theta \sim N(B/A, 1/A)$ to sample θ, where

$$\begin{aligned} A &= \frac{1}{\sigma_\theta^2} + \sum_{i=2}^{T} \frac{\kappa_V^2 \Delta}{(1-\rho^2)\omega_{(i-1)\Delta}} \\ B &= \frac{\mu_\theta}{\sigma_\theta^2} + \kappa_V \sum_{i=2}^{T} \frac{\widetilde{Y}_{i\Delta} - Y_{(i-1)\Delta} + \kappa_V Y_{(i-1)\Delta}\Delta + \varsigma_V \omega_{(i-1)\Delta}\Delta - \rho D_{i\Delta}\sqrt{\omega_{(i-1)\Delta}\Delta}}{(1-\rho^2)\omega_{(i-1)\Delta}} \end{aligned}$$

- Sample κ_V

  Suppose the prior distribution of κ_V is $N(\mu_{\kappa_V}, \sigma_{\kappa_V}^2)$, we will use posterior distribution $\kappa_V \sim N(B/A, 1/A)$ to sample κ_V, where

$$\begin{aligned} A &= \frac{1}{\sigma_{\kappa_V}^2} + \sum_{i=2}^{T} \frac{(\theta - Y_{(i-1)\Delta})^2 \Delta}{(1-\rho^2)\omega_{(i-1)\Delta}} \\ B &= \frac{\mu_{\kappa_V}}{\sigma_{\kappa_V}^2} + \sum_{i=2}^{T} \frac{(\widetilde{Y}_{i\Delta} - Y_{(i-1)\Delta} + \varsigma_V \omega_{(i-1)\Delta}\Delta - \rho D_{i\Delta}\sqrt{\omega_{(i-1)\Delta}\Delta})(\theta - Y_{(i-1)\Delta})}{(1-\rho^2)\omega_{(i-1)\Delta}} \end{aligned}$$

- Sample ς_V

  Suppose the prior distribution of ς_V is $N(\mu_{\varsigma_V}, \sigma_{\varsigma_V}^2)$, we will use posterior distribution $\varsigma_V \sim N(B/A, 1/A)$ to sample ς_V, where

$$\begin{aligned} A &= \frac{1}{\sigma_{\varsigma_V}^2} + \sum_{i=2}^{T} \frac{\omega_{(i-1)\Delta}\Delta}{1-\rho^2} \\ B &= \frac{\mu_{\varsigma_V}}{\sigma_{\varsigma_V}^2} - \sum_{i=2}^{T} \frac{\widetilde{Y}_{i\Delta} - Y_{(i-1)\Delta} - \kappa_V(\theta - Y_{(i-1)\Delta})\Delta - \rho D_{i\Delta}\sqrt{\omega_{(i-1)\Delta}\Delta}}{1-\rho^2} \end{aligned}$$

- Sample κ_ω^P

  Suppose the prior distribution of κ_ω^P is $N(\mu_{\kappa_\omega^P}, \sigma_{\kappa_\omega^P}^2)$, we will use posterior distribution $\kappa_\omega^P \sim N(B/A, 1/A)$ to sample κ_ω^P, where

$$\begin{aligned} A &= \frac{1}{\sigma_{\kappa_\omega^P}^2} + \sum_{i=2}^{T} \frac{\omega_{(i-1)\Delta}\Delta}{(1-\rho^2)\sigma_\omega^2} \\ B &= \frac{\mu_{\kappa_\omega^P}}{\sigma_{\kappa_\omega^P}^2} - \sum_{i=2}^{T} \frac{\widetilde{\omega}_{i\Delta} - \omega_{(i-1)\Delta} - \alpha_\omega\Delta - \sigma_\omega \rho C_{i\Delta}\sqrt{\omega_{(i-1)\Delta}\Delta}}{(1-\rho^2)\sigma_\omega^2} \end{aligned}$$

- Sample μ_y^{JP}, μ_ω^{JP}, $(\sigma_\omega^J)^2$

  Suppose the prior distributions of these parameters are $\mu_y^{JP} \sim N\left(\mu_{\mu_y^{JP},\sigma^2_{\mu_y^{JP}}}\right)$, $\mu_\omega^{JP} \sim N\left(\mu_{\mu_\omega^{JP},\sigma^2_{\mu_\omega^{JP}}}\right)$, $(\sigma_\omega^J)^2 \sim$ InvGam $\left(\alpha^*_{(\sigma_\omega^J)^2 1}, \alpha^*_{(\sigma_\omega^J)^2 2}\right)$, we will use the following posterior distributions to sample the three parameters,

$$\begin{aligned}
\mu_\omega^{JP} &\sim N\left(\frac{(\sigma_\omega^J)^2\mu_{\mu_\omega^{JP}} + \sigma^2_{\mu_\omega^{JP}}\sum_{i=2}^T j^\omega_{i\Delta}}{(\sigma_\omega^J)^2 + T\sigma^2_{\mu_\omega^{JP}}}, \left(\frac{T}{(\sigma_\omega^J)^2} + \frac{1}{\sigma^2_{\mu_\omega^{JP}}}\right)\right) \\
\mu_y^{JP} &\sim N\left(\frac{(\sigma_y^J)^2\mu_{\mu_y^{JP}} + \sigma^2_{\mu_y^{JP}}\sum_{i=2}^T j^y_{i\Delta}}{(\sigma_y^J)^2 + T\sigma^2_{\mu_y^{JP}}}, \left(\frac{T}{(\sigma_y^J)^2} + \frac{1}{\sigma^2_{\mu_y^{JP}}}\right)\right) \\
(\sigma_\omega^J)^2 &\sim InvGam\left(\alpha^*_{(\sigma_\omega^J)^2 1} + \frac{T}{2}, \alpha^*_{(\sigma_\omega^J)^2 2} + \frac{1}{2}\sum_{i=2}^T (j^\omega_{i\Delta} - \mu_\omega^{JP})^2\right)
\end{aligned}$$

  Remark: Inverse Gamma distribution random variable $X \sim InvGam(\alpha, \beta)$ has a density function of

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right), x > 0$$

  where shape parameter $\alpha > 0$, and scale parameter $\beta > 0$, if $X \sim InvGam(\alpha, \beta)$, we have $\frac{1}{X} \sim Gamma(\alpha, \beta)$

2.4.1.5 Sampling Parameters under Q Measure

The Q-parameters $\Theta_M = \left\{\alpha_\omega, \kappa_\omega^Q, \lambda_0, \lambda_1, \mu_y^J, \mu_\omega^J, \sigma_y^J\right\}$ in (2.17) are related to the observed VVIX index model of (2.18). We thus use random-walk Metropolis-Hastings algorithms to sample these parameters with the target density given as

$$\frac{1}{\sqrt{2\pi}\sigma_P}\exp\left(-\frac{\sum_{i=1}^T\left(\left(VVIX^{obs}_{i\Delta}\right)^2 - A(\tau) - B(\tau)\,\omega_{i\Delta}\right)^2}{2\sigma_P^2}\right) \tag{2.22}$$

Take α_ω as an example, assume the $(g-1)$th sampling is $\alpha_\omega^{(g-1)}$, make $\alpha_\omega^{(g)} = \alpha_\omega^{(g-1)} + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$, take equation (2.22) as the density function of α_ω, denoted by $\pi(\alpha_\omega)$, then we can accept $\alpha_\omega^{(g)}$ as the gth sampling of α_ω with a probability of $\alpha(\alpha_\omega^{(g-1)}, \alpha_\omega^{(g)})$, if we don't accept, we will keep $\alpha_\omega^{(g-1)}$ as the sampling

of $\alpha_{\omega}^{(g)}$, where

$$\alpha(\alpha_{\omega}^{(g-1)}, \alpha_{\omega}^{(g)}) = \min\left(\frac{\pi(\alpha_{\omega}^{(g)})}{\pi(\alpha_{\omega}^{(g-1)})}, 1\right)$$

we will adjust the variance of ϵ, that is, the magnitude of σ, to make sure the accept probability between 20% and 40%. For the other parameters in Θ_M, the sampling method is similar, the only difference is the magnitude of the variance of ϵ.

2.4.1.6 Sampling Pricing Error Parameter

For the pricing error parameter $\Theta_E = \{\sigma_P\}$, conditional on $\left(VVIX_{i\Delta}^{obs}\right)^2$ and $\omega_{i\Delta}$, $\epsilon_{i\Delta} = \left(VVIX_{i\Delta}^{obs}\right)^2 - A(\tau) - B(\tau)\omega_{i\Delta} \sim N(0, \sigma_P^2)$. Assume the prior for σ_P^2 is $\pi_{\sigma_P^2}(\sigma_P^2) \sim InvGam\left(\alpha_{\sigma_P^2}, \beta_{\sigma_P^2}\right)$, we then sample σ_P^2 using inverse gamma distribution, that is

$$\sigma_P^2 \sim InvGam\left(\alpha_{\sigma_P^2}^*, \beta_{\sigma_P^2}^*\right)$$

with $\alpha_{\sigma_P^2}^* = \alpha_{\sigma_P^2} + \frac{T-1}{2}$ and $\beta_{\sigma_P^2}^* = \beta_{\sigma_P^2} + \frac{\sum_{i=2}^{T}\left(\left(VVIX_{i\Delta}^{obs}\right)^2 - A(\tau) - B(\tau)\omega_{i\Delta}\right)^2}{2}$.

2.4.2 Model Diagnostics and Specification Tests

Given the sampled posterior latent variables (spot volatility, jump times and jump sizes) and parameters, we can construct several statistics to test and assess the ability of the model to fit the historical data. More specifically, we use detailed analysis of the residual, p-value method and DIC method.

2.4.2.1 Residual Analysis

Recall the discretization of $Y(t)$ in (2.21), the representation of $\epsilon_{i\Delta}^y$ follows immediately and is given by

$$\epsilon_{i\Delta}^y = \frac{\widetilde{Y}_{i\Delta} - a_0 - a_1 Y_{(i-1)\Delta} - a_2\omega_{(i-1)\Delta}}{\sqrt{\omega_{(i-1)\Delta}\Delta}}, 2 \le i \le T+1 \tag{2.23}$$

With the estimated variables and parameters at hand, we can calculate these residuals immediately. We can then compare the Q-Q plot of the residuals of different models. If a given model's residuals approximate standard normal distribution, then the model can be said to perform well for fitting the historical VIX index. If

there is a substantial discrepancy between the residuals and the standard normal distribution, the given model has potential for further improvement. With the expression for $\widetilde{\omega}_{i\Delta}$ in (2.21), we can calculate the residual for $\omega(t)$ similarly

$$\epsilon^{\omega}_{i\Delta} = \frac{\widetilde{\omega}_{i\Delta} - c_0 - c_1\omega_{(i-1)\Delta}}{\sigma_\omega\sqrt{\omega_{(i-1)\Delta}\Delta}}, 2 \leq i \leq T \tag{2.24}$$

This residual can be used to compare the various models for $\omega(t)$.

The sample of jump times of $Y(t)$ and $\omega(t)$ can be used to test whether the jump intensity is constant or stochastic. If the posterior sampled jump times are clustered, the constant jump intensity assumption is rejected.

2.4.2.2 p-value Method

We also perform a simulation study using the posterior parameters to test different specifications. We first specify some statistics that reflect the dynamics of VIX and calculate these statistics for the logVIX data. Then for each model, we simulate many trajectories for Y with the same sample size as the VIX data using the estimated parameters from the MCMC results. With the simulated trajectory, we calculate the sample statistics and compare them with those obtained from the original VIX data.

More specifically, we use the following 10 reference statistics: standard deviation, skewness, kurtosis, maximum, minimum, maxjump (the highest positive changes in the index), minjump (the highest negative changes in the index), avgmax10 (the average over the 10 largest positive changes), avgmin10 (the average over the 10 largest negative changes) and various percentiles of daily changes (The percentiles are denoted by $percNUM$ where NUM indicates the percentage level).

Denote these statistics calculated from the logVIX data by $\phi_k, k = 1, 2, \cdots, 10$. Then for each given model, simulate N trajectories for Y using the estimated parameters from the MCMC results. For the n-th simulated trajectory Y, $1 \leq n \leq N$, calculate the statistics above which are denoted by $\phi_k^{(n)}, k = 1, 2, \cdots, 10$. For every k, $1 \leq k \leq 10$, calculate

$$p_k = \frac{\sum_{n=1}^{N} 1_{\left\{\phi_k^{(n)} > \phi_k\right\}}}{N} \tag{2.25}$$

where 1_A is the indicator function. A too high or too low $p_k, 1 \leq k \leq 10$ indicates that the given model may be distorted from the genuine form. For more details about this method, we refer to [Gelman, Meng, and Stern 1996] and [Kaeck and Alexander 2013].

2.4.2.3 DIC Method

Last but not the least we employ the deviance information criterion (DIC) method. This strategy is a good replacement of the Bayes factor which imposes heavy calculation burden under our circumstance. Assume we have observations $y = (y_1, y_2, \cdots, y_n)$, θ is a p-dimensional parameters which need to estimate. (126) suggests the following empirical statistic to test the posterior distribution of θ,

$$D(\theta) = -2 \ln f(y|\theta) + 2 \ln g(y)$$

where $f(y|\theta)$ is the adjoint distribution of the observations if given the unknown parameter θ, in our analysis, we set $g(y) = 1$, that is we don't have the second term. (127) gives the definition of DIC based on (126), as a criterion for model selection. DIC has two components, one is the measure of data fitness, the other is a penalty of the model complexity, DIC is defined by

$$DIC = \overline{D} + p_D \tag{2.26}$$

in equation (2.26), $\overline{D}$ is a Bayes measure, defined as a posterior expectation, in our example, it is reduced as

$$\overline{D} = E_{\theta|y}[D(\theta)] = E_{\theta|y}[-2 \ln f(y|\theta)]$$

the larger the value of likelihood function is, the smaller $\overline{D}$ is, and the better the data fitness is.

The second term in equation (2.26) p_D is a measure of the complexity of the model, it is defined by

$$\begin{aligned} p_D &= \overline{D} - D(\overline{\theta}) = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}[\theta]) \\ &= E_{\theta|y}[-2 \ln f(y|\theta)] + 2 \ln f(y|\overline{\theta}) \end{aligned}$$

the larger p_D is, the more complex the model is. In summary, the smaller of the DIC, the better the model is. In our analysis, we can get the expectation of $\overline{D}$ from the MCMC sampling trajectories of the parameters. From the mean of different parameters, we can get the value of $\ln f(y|\overline{\theta})$, the likelihood function is as follows:

$$f(Y|\Theta) = \frac{1}{(2\pi)^{\frac{T}{2}}} \exp\left[\sum_{i=1}^{T} \frac{1}{2(1-\rho^2)}(\epsilon_{i\Delta}^{y} - \rho\epsilon_{i\Delta}^{\omega})^2\right]$$

where $\epsilon_{i\Delta}^{y} = \frac{\widetilde{Y}_{i\Delta} - a_0 - a_1 Y_{(i-1)\Delta} - a_2\omega_{(i-1)\Delta}}{\sqrt{\omega_{(i-1)\Delta}\Delta}}$, $\epsilon_{i\Delta}^{\omega} = \frac{\widetilde{\omega}_{i\Delta} - c_0 - c_1\omega_{(i-1)\Delta}}{\sigma_\omega\sqrt{\omega_{(i-1)\Delta}\Delta}}$

For more details of this method, we recommend [Berg, Meyer, and Yu 2004] and the reference therein.

2.5 Data

In 1993, CBOE introduced the VIX index, which serves as a benchmark for the volatility of the market. On September 22, 2003, CBOE revised the calculation method of VIX to utilize a wider range of S&P500 options. It also back-calculated the new VIX to 1990. The now well-known generalized formula for calculating VIX is

$$VIX^2(t,T) = \frac{2}{T-t}\sum_i \frac{\triangle K_i}{K_i^2} e^{r_t(T-t)} Q(K_i) - \frac{1}{T-t}\left[\frac{F_t}{K_0} - 1\right]^2$$

This formula utilizes a strip of OTM S&P500 option prices $Q(K_i)$, where F_t is the forward S&P500 index level derived from S&P500 options.

On March 14, 2012, CBOE released a new volatility of volatility index called VVIX. VVIX is a measure of volatility of volatility which represents the expected volatility of the 30-day forward price of the CBOE volatility index. The calculation method of VVIX is similar to that of VIX, VVIX is calculated from the price of a strip of at- and out- of the money VIX options, i.e.

$$VVIX^2(t,T) = \frac{2}{T-t}\sum_i \frac{\triangle K_i}{K_i^2} e^{r_t(T-t)} O(K_i) - \frac{1}{T-t}\left[\frac{F_t}{K_0} - 1\right]^2$$

where $O(K_i)$ is the midpoint of the bid-ask spread of VIX options with strike K_i and F_t is the forward VIX index level derived from VIX option prices. K_0 is the first strike below the forward index level F_t. Using this method, CBOE has also

calculated the VVIX index before the release date up to the start of 2007. We plot the historical time series of VIX and VVIX from January 2007 to November 2014 in Figure 2.2.

From Figure 2.2 we can see that the level of VVIX is significantly higher overall than that of VIX. Yet like VIX, VVIX also mean-reverts to its historical mean value, which is nearly 80. Furthermore, both indices do share some of their peak values, especially at points during the 2008 financial crisis. Compared with VIX, VVIX is more volatile. The range of variation of VVIX becomes particularly broad when VIX is high. The statistics of VIX and VVIX from January 2007 to November 2014 are summarized in Table 2.1.

2.6 Empirical Results

In this section, we discuss the estimation results for VIX dynamics among different models. The parameter estimations for the four models are summarized in Table 2.2, while the simulation results are shown in Table 2.3. For all of the proposed models, the estimates of ρ are positive and around 0.52, which is similar to the findings of [Kaeck and Alexander 2013] ($\rho = 0.659$ for the SVJ model in their paper). As the parameter ς_V enters into the drift of VIX, the estimation for θ remains relatively low compared to the mean value of logVIX market data during the same period. κ_ω^P is significantly larger than κ_V, reflecting the fact that the volatility of VIX or VVIX is more volatile than VIX itself. We point out that to determine the sign of μ_y^J which cannot be detected using (2.6), we specify another expression of VVIX from its definition and perform similar MCMC estimation for test. The result indicates that μ_y^J is significantly negative.

Figure 2.3 gives the estimated volatility processes of VIX for all four models. As we use the VVIX index as a proxy for volatility, the four processes are similar. The correlations between estimated spot volatility and VVIX for all four models are 0.9781, 0.9782, 0.9822, and 0.9807, respectively. The average posterior volatility values for the SV and SVJ models are slightly higher than those of the other two models. This can be explained by the fact that the addition of jumps in volatility reduces the demand on the volatility process.

Figure 2.4 shows the Q-Q plot of the residuals of VIX calculated by (2.23) for all four models, and Figure 2.5 plots the time series form. Based on the upper-left

panel in Figure 2.4, we can see that the SV model is misspecified, as it requires very large shocks to the Brownian motion. This can also be seen based on the upper-left panel in Figure 2.5. Compared to the other three models, the range of the residuals for the SV model is significantly larger, and there are many large innovations.

From the two upper panels in Figure 2.4, we can see that the tail of the residuals becomes slightly thin so SVJ model improves SV model better. Many of the big Brownian shocks can be absorbed into the jump part. The estimated jump size in VIX is reported in the upper-left panel in Figure 2.6.

However, from the simulation results in Table 2.3, we find that for both the SV and SVJ models, there are one or more statistics whose p-values are out of the $[0.05, 0.95]$ range. In contrast, [Kaeck and Alexander 2013] also test the SV and SVJ models (with normal jumps) and demonstrate that all of the p-values are within the $[0.05, 0.95]$ range. This shows that the addition of VVIX as the proxy for the volatility of VIX helps detect an area for improvement in the stochastic volatility of volatility model for VIX. We also show Figure 2.7, created using (2.24) and which compares the residuals of the volatility processes among all four models. The residuals of the SV and SVJ models are larger than those of the SVJJ-C and SVJJ-S models.

Next we come to the SVJJ-C and SVJJ-S models. As mentioned above, the residuals of the volatility processes for these two models perform better than those of the SV and SVJ models. This shows the impact of jump on volatility. Figure 2.8 describes the estimated daily jump probability for the SVJ, SVJJ-C, and SVJJ-S models. With the jump in volatility added, the jump occurs a bit more frequently. We recall that for the SVJ model in which there is no jump in volatility, the jump time is determined mainly by the information from the VIX index. For the SVJJ-C and SVJJ-S models, however, we sample the jump time using either information or signals from VIX and volatility (VVIX). As we assume that the jumps of VIX and its volatility factor are determined by the same Poisson process, a large jump in volatility may increase the jump probability. This means that not only does the volatility jump, but also that it jumps more intensely than that of VIX. Return to Figure 2.4, in which the bottom panels for SVJJ-C and SVJJ-S perform better than the upper ones. This shows that the jump in volatility can also have an impact on the dynamics of VIX. The channel of influence can be through moments of high order or extreme values, as shown in Table 2.3.

Unlike transient Brownian motion shocks, the influence of jump on volatility is more persistent. After a positive or negative jump, volatility enters a new regime. As the diffusion part of VIX, volatility continues to experience these effects for a period of time. A simple empirical method for judging the influence of a jump on volatility on a particular day is to compare the fluctuations within a period of VIX data before and after the day of a jump. For example, on February 27, 2007, VIX jumped from 11.15 to 18.31. Before this day, VIX had been very stable, positioned around 11. However, after this turning point, VIX became more volatile with large ups and downs ranging from 12.19 to 19.63 occurring frequently over the next 20 days. In fact, on February 27, the VVIX index jumped from 70.33 to 110.42. If we calculate the average of the VVIX index for 20 days before and after the day of the jump, the results are 72.54 and 96.55, respectively. This indicates that the volatility shifted to a new and higher state and thus made the VIX index more active. This effect cannot be achieved by a single Brownian shock but is instead caused by a jump. Thus the SVJ model is also misspecified from empirical observation.

Figure 2.9 describes the jump in volatility for the SVJJ-C and SVJJ-S models, respectively. The jump sizes are almost positive, with only a large negative jump in the SVJJ-S models. Because of the mean-reverting property, volatility reduces to its mean level through negative Brownian innovations after a large positive jump. This indicates that the impact of a positive jump can be persistent and significant.

From Figure 2.8, we can observe that the SVJJ-C model's the jump times are clustered. This is extremely unlikely under the constant jump intensity assumption. We can also see from Table 2.2 that the estimation of λ_1 in the SVJJ-S model is significant when it is above zero. These facts indicate that the SVJJ-S model is superior to the SVJJ-C model, as it depicts the dynamics of VIX more accurately. When stochastic volatility $\omega(t)$ enters into a relatively high state, more jumps occur and affect the dynamics of VIX.

Table 2.4 reports the DIC value of the four models. We find the SVJJ-C and SVJJ-S model outperform the SV and SVJ model which justify the role of the jumps in volatility of VIX. However under this criterion SVJJ-C is better than SVJJ-S. As pointed above the jump time is indeed stochastic instead of possessing constant intensity. This suggest that a better specification of jump intensity may be needed to describe the dynamics of VIX better.

2.7 Conclusion

This paper discusses the specifications for stochastic volatility models of VIX using information provided by VVIX. We construct a volatility proxy for VIX using the VVIX index as the benchmark and study its role in improving the model assumptions of VIX from empirical observations. Based on the joint behavior of VIX and VVIX, we propose a double-jump stochastic volatility model for VIX. We use the MCMC method to estimate and compare different nested models using daily data on VIX and VVIX. Based on the results, we point out that the jumps in VIX and volatility are essential and statistically significant, and analyze the impact of the jumps on VIX dynamics. We show that the jump intensity is behaved stochastic instead of being constant. However a better process to characterize the jump times is still needed. The use of VVIX allows for the estimation of some Q-parameters. Compared with richer dataset composed of VIX futures and options, the accuracy of these parameters has potential for further improvement. In particular, we suggest that in future research, the corresponding risk premia should be further specified.

Table 2.1: Summary Statistics of VIX and VVIX

This table provides summary statistics for VIX and VVIX index from January 3, 2007, to November 26, 2014.

	Mean	Volatility	Skewness	Kurtosis	Min	Max
VIX	21.9101	10.3966	2.1241	5.7181	9.89	80.86
VVIX	85.9204	12.8226	0.8289	1.0079	59.74	145.12

Table 2.2: VIX Parameter Estimates

This table shows the parameter estimation results for the four models using VIX and VVIX index data from January 3, 2007 to November 26, 2014. Four each parameter, we give the mean and the standard deviation of the posterior. "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

	SV		SVJ-C		SVJJ-C		SVJJ-S	
	Mean	Stddev	Mean	Stddev	Mean	Stddev	Mean	Stddev
κ_V	1.6800	0.5733	1.5765	0.5600	1.8611	0.5688	2.1093	0.5866
ς_V	-1.1869	0.7718	-0.8046	0.7673	-0.2702	0.8305	-0.1538	0.7820
θ	2.3500	0.4404	2.3090	0.4446	2.2704	0.3954	2.3312	0.3120
κ_ω^P	4.5162	1.0284	4.4308	0.9973	6.1132	1.0650	6.2849	1.0645
κ_ω^Q	7.5104	0.4314	7.6866	0.4676	2.5996	0.2584	2.5674	0.1958
α_ω	3.8549	0.7807	3.7683	0.7540	4.0781	0.7882	3.7938	0.7308
ρ	0.5392	0.0161	0.5596	0.0141	0.5204	0.0169	0.4998	0.0190
σ_ω	0.8560	0.0724	0.8207	0.0117	0.8848	0.0853	0.8461	0.0372
λ_0			0.6550	0.0492	2.4295	0.1787	2.7557	0.1332
λ_1							1.6086	0.1262
μ_y^{JP}			0.1593	0.0220	0.1999	0.0279	0.1551	0.0171
μ_y^J			-0.0520	0.0037	-0.0556	0.0746	-0.0960	0.0306
σ_y^J			0.1075	0.0172	0.1121	0.0132	0.1231	0.0108
μ_ω^{JP}					0.1872	0.0226	0.1430	0.0239
μ_ω^J					-2.0084	0.0882	-1.2046	0.0547
σ_ω^J					0.1307	0.0165	0.1420	0.0161
σ_P	0.0599	0.0077	0.0592	0.0071	0.0563	0.0082	0.0612	0.0076

Table 2.3: Simulation Results on VIX for Different Models

This table reports the p-values calculated by (2.25) for all the statistics of simulation results of VIX for different models. It describes the average comparisons of the statistics of historical data and the simulation paths from every given model. Very high or low p-values indicate the model's inability to capture the VIX dynamics. "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

	Data	SV	SVJ-C	SVJJ-C	SVJJ-S
stadev	0.4493	0.0670	0.3351	0.8593	0.3733
skewness	0.9005	0.1462	0.0359	0.2194	0.7479
kurtosis	0.7806	0.2453	0.0150	0.6250	0.7186
maximum	4.7558	0.0079	0.1383	0.6875	0.4744
minimum	2.3984	0.2418	0.6040	0.0769	0.6896
maxjump	0.2267	0.4897	0.1016	0.2805	0.6358
minjump	-0.2422	0.0953	0.9539	0.7698	0.7509
avgmax10	0.1852	0.5697	0.1508	0.4308	0.2614
avgmin10	-0.1671	0.3262	0.9069	0.7996	0.5826
perc0.01	-0.1345	0.4271	0.8488	0.5005	0.6311
perc0.05	-0.0931	0.4504	0.4957	0.5048	0.6277
perc0.95	0.0928	0.5604	0.6182	0.8245	0.4581
perc0.99	0.1370	0.4478	0.0264	0.9379	0.1622

Table 2.4: DIC Model Comparison

This table reports the DIC value for alternative models. DIC consists of $\overline{D}$ which measures model fit and p_D which penalizes the number of the parameters. The smaller of DIC, the better the model is. "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

	SV	SVJ-C	SVJJ-C	SVJJ-S
$\overline{D}$	6503.797	6273.394	5976.237	6023.068
p_D	435.4473	553.4345	469.4931	512.1815
DIC	6939.244	6826.828	6445.73	6535.25

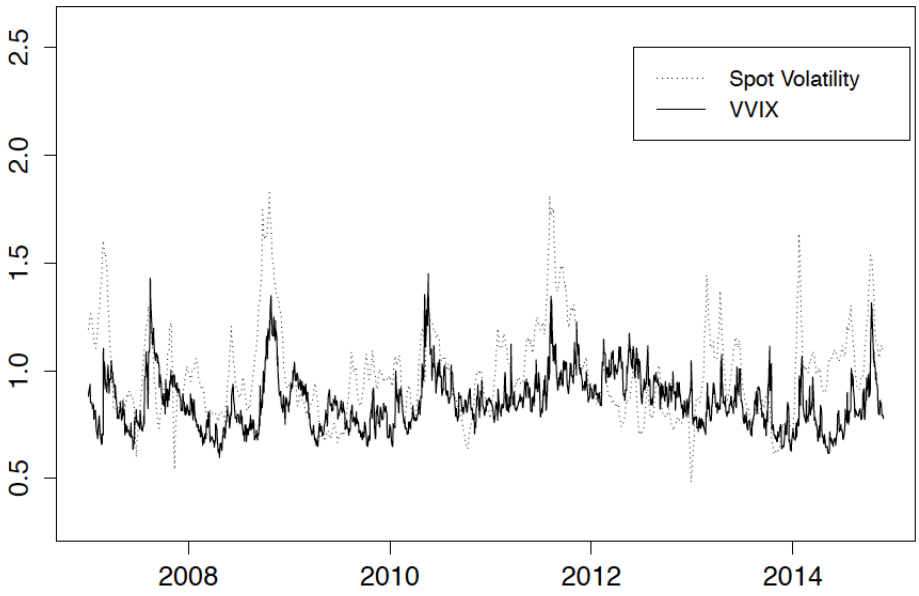


Figure 2.1: Spot Volatility from VIX Estimation vs VVIX

The spot volatility in this figure is the estimated posterior volatility of logVIX in SVJ model only using the VIX index itself. It shows a visual comparison of this volatility and the contemporaneous VVIX index

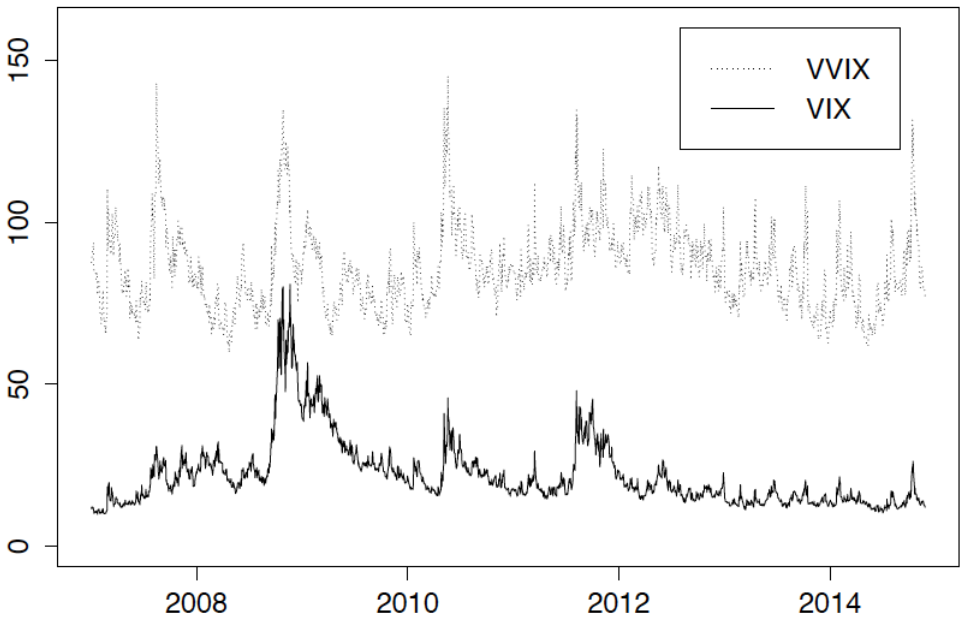


Figure 2.2: VIX and VVIX Indices

This figure shows the time series of VIX and VVIX index from January 3, 2007 to November 26, 2014. Both of them are mean-reverting and VVIX is at a significant higher level than VIX in terms of the range of values.

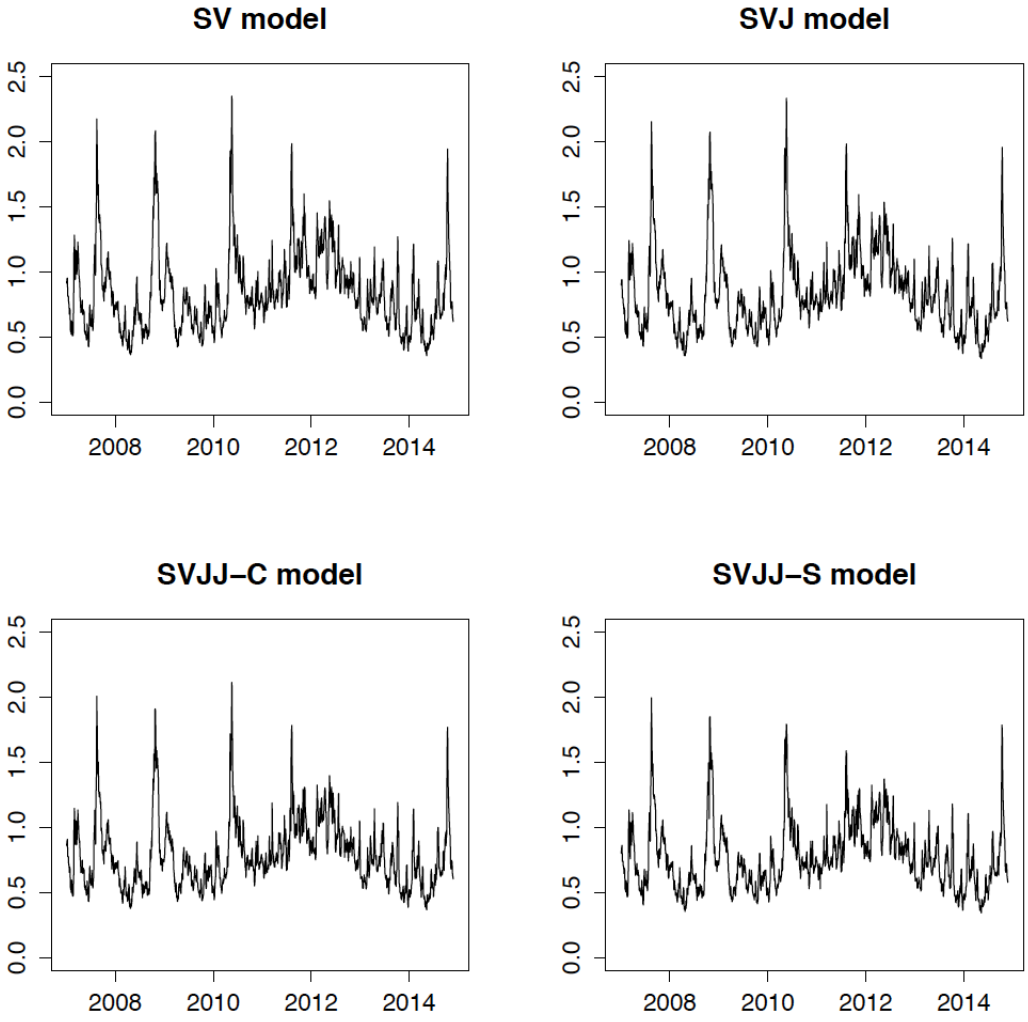


Figure 2.3: Posterior Volatility of VIX for Each Model

The figures show the estimated paths of posterior volatility $\omega(t)$ for four models. All of them are highly correlated with VVIX index. The level of the volatility in SV and SVJ models is slightly higher than that in SVJJ-C and SVJJ-S models. "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

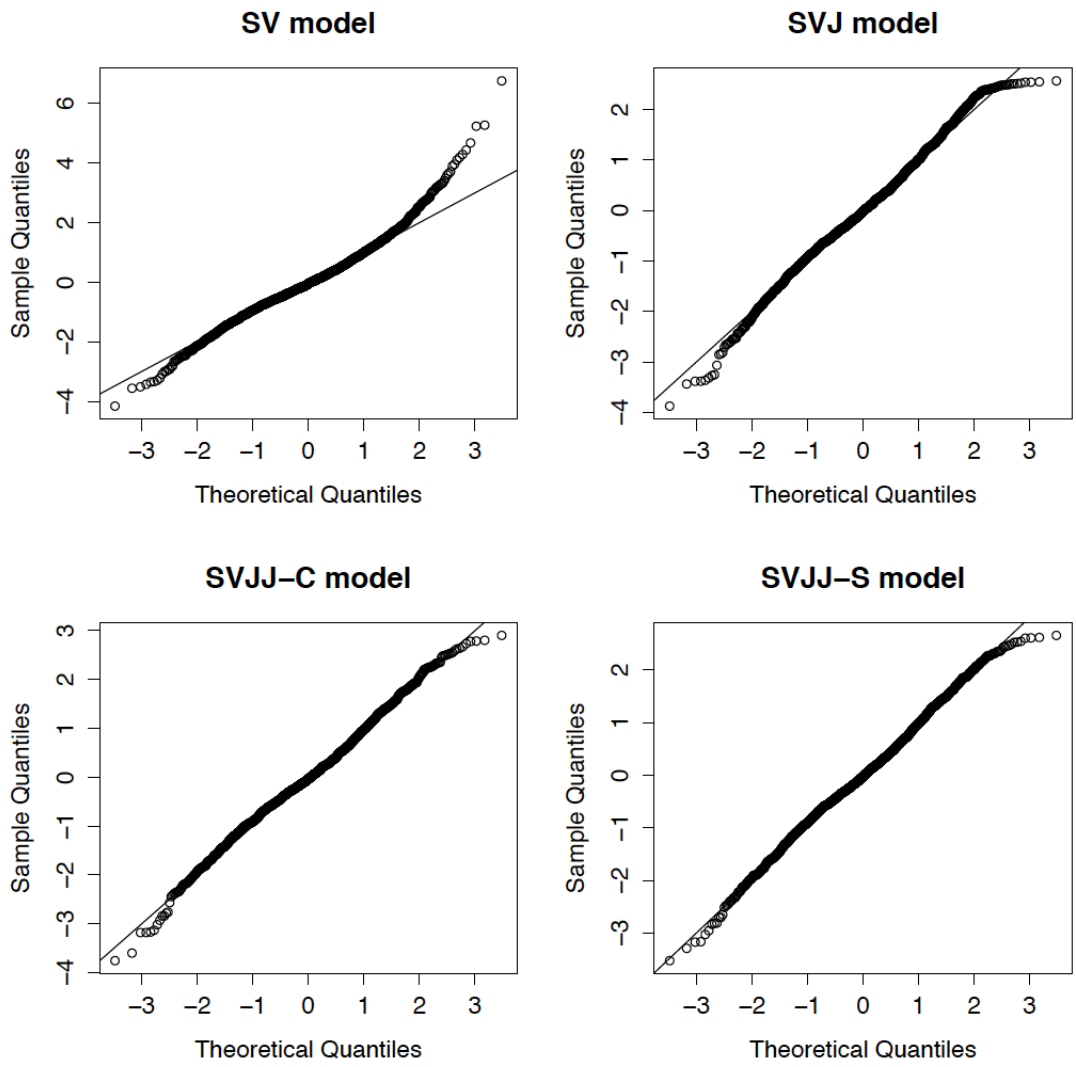


Figure 2.4: Q-Q Plot of the Residuals

The figures show the Q-Q plot of the residuals calculated from each of the models using (2.23) with the estimated parameters as input. SVJJ-C and SVJJ-S models perform relatively better than SV and SVJ models."SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

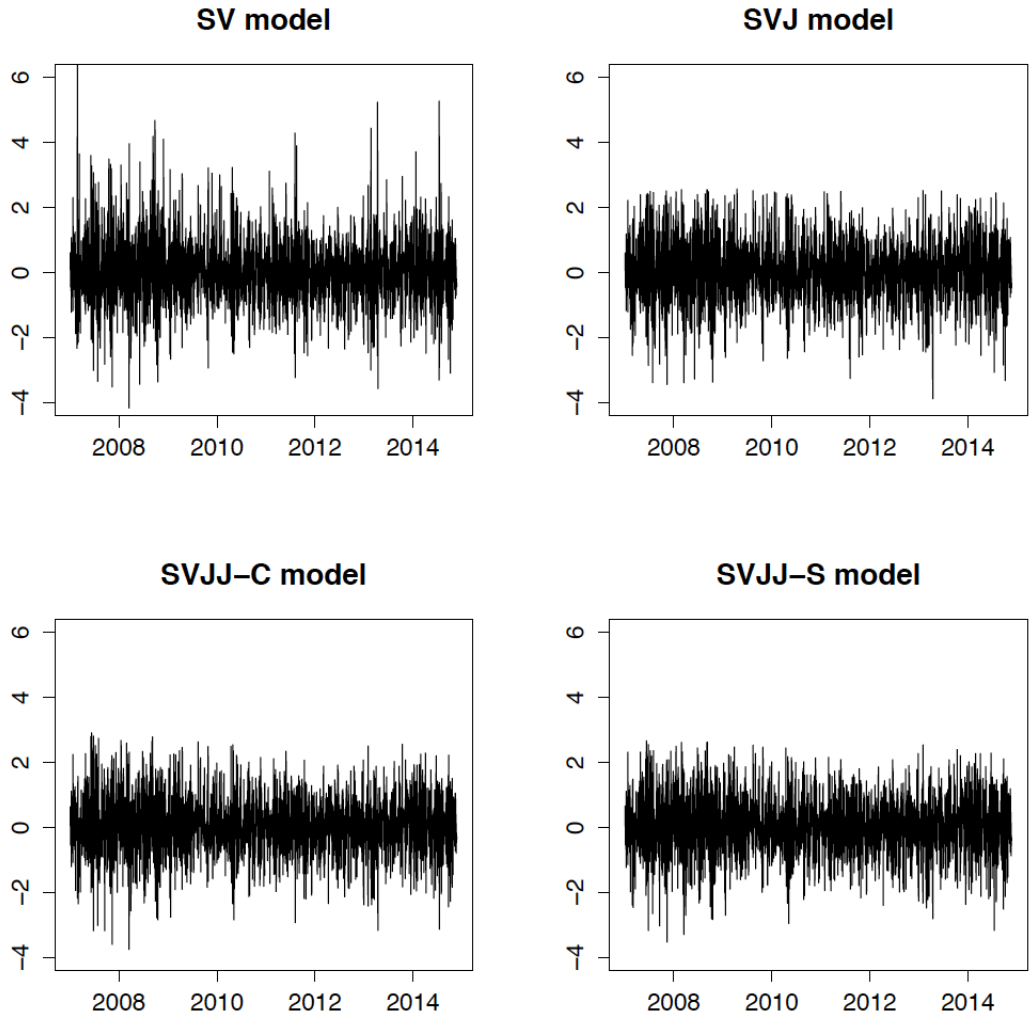


Figure 2.5: VIX Residuals

The figures show the time series of standard innovations or residuals of VIX calculated from the estimated parameters using (2.24). "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

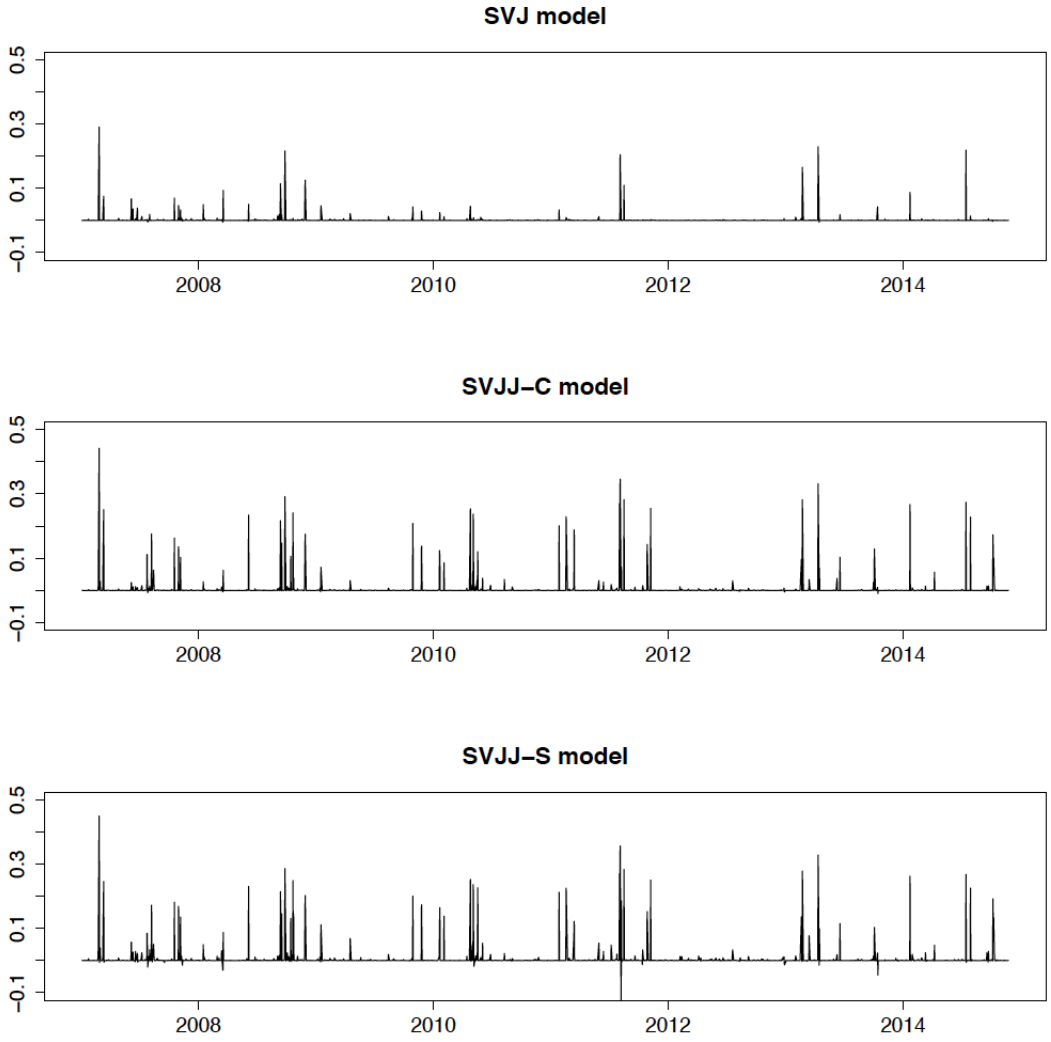


Figure 2.6: Posterior Mean of Jumps in VIX

The figures show the time series of average jump sizes in VIX. "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

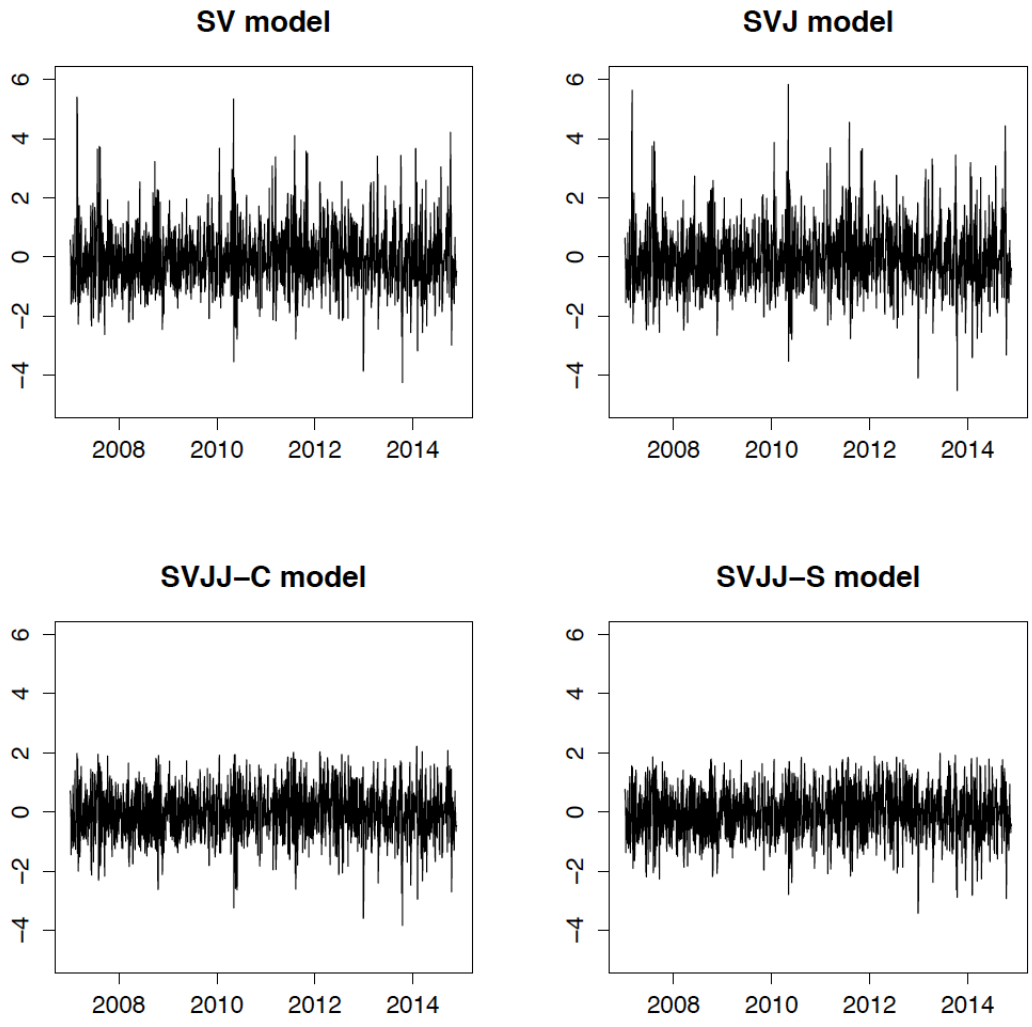


Figure 2.7: Volatility Residuals

The figures show the time series of standard innovations or residuals of volatility of VIX calculated from the estimated parameters using . "SV" denotes diffusion model with no jumps. "SVJ" introduces jumps in VIX in the SV model with constant jump intensity, "SVJJ-C" adds double jumps in VIX and its volatility in the SV model with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

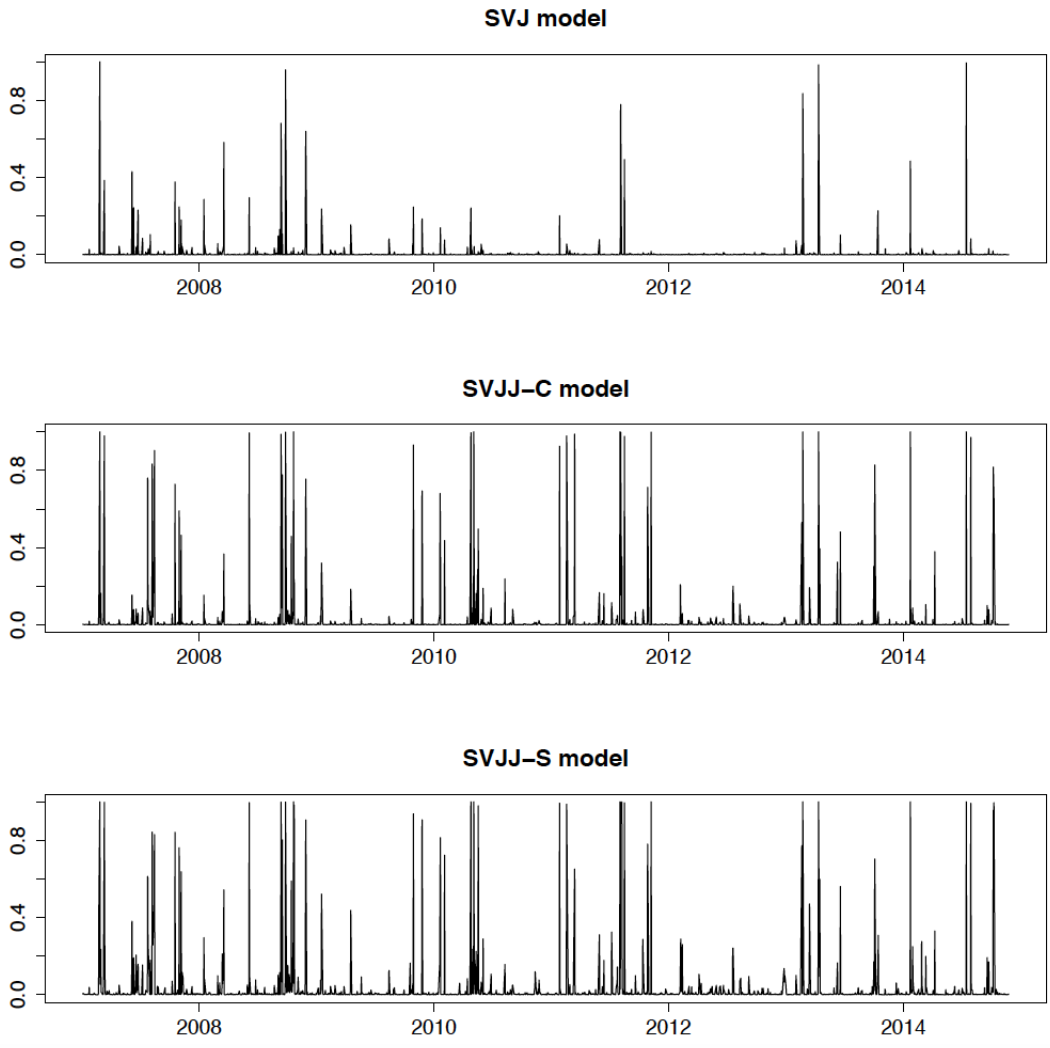


Figure 2.8: Estimated Jump Times for SVJ, SVJJ-C and SVJJ-S Models

The figures show the estimated jump probability of SVJ, SVJJ-C and SVJJ-S models. "SVJ" introduces jumps in VIX with constant jump intensity and models the volatility using square root diffusion model, "SVJJ-C" introduces double jumps in VIX and its volatility with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

Figure 2.9: Posterior Mean of Jumps in Volatility of VIX

The figures show the time series of average jump sizes in the volatility of VIX. "SVJJ-C" introduces double jumps in VIX and its volatility with constant jump intensity. "SVJJ-S" assumes the jump intensity to be stochastic in the SVJJ-C model.

Chapter 3
Forecasting Bond Returns Using High-dimensional Model Selection

3.1 Introduction

Prediction of bond risk premia based on the availability of macroeconomic fundamentals is a popular research topic in recent years, like described in Chapter 1, with the fast development of fundamental statistics, a vast of well performed variable selection techniques have been developed, which enable us to facilitate the research on excess bond returns forecast. With more tools in hand today, we can indeed try to overcome the challenges like, trying to achieve a better performance on variable selection from high dimensional dataset, trying to explore a stronger predictive power(in-sample and out-of-sample) on excess bond returns and for a better understanding of the determinants of bond risk premia, trying to recover some nonlinear relationship which may benefit the prediction purpose besides the linear structure between excess bond returns and macroeconomic fundamentals, as well as trying to propose a robust procedure of macro variables selection for predicting bond risk premia.

In this paper, we will introduce a linear variable selection framework and a nonlinear variable selection framework for ultrahigh-dimensional dataset, aiming to select important variables with significant forecast power on bond risk premia, as well as improving both the prediction accuracy for excess bond returns and model

interpretability in this high dimensional setting. we will try to construct several different variable selection approaches based on the linear and nonlinear variable selection frameworks we proposed. In addition, the same dataset considered in [Ludvigson and Ng 2009a], and [J.-z. Huang and Shi 2011], a panel of 131 macroeconomic variables as well as their lags up to seven will be used in our analysis, it is worth to mention that, the data analysis in our paper is an initiated comprehensive study on ultrahigh-dimensional macroeconomic series. Like the analysis in this paper, high-dimensional data analysis has now become increasingly frequent and necessary in finance, moreover, one even more important challenge facing both researchers and practitioners nowadays is the tractability of ultrahigh-dimensional data, where the number of predictors, p, is usually much larger than the sample size n, ($p \gg n$)-namely, $\log p = O(n^{\alpha})$ for some $\alpha > 0$. The macroeconomic series associated with their lags analyzed in our study falls into the ultrahigh-dimensional setting as introduced. Considering the issues of computational expediency, statistical accuracy and algorithmic stability, we will incorporate two newly developed statistical modeling techniques for dimension reduction and variable selection: Screening and Regularization. In our variable selection approaches, screening procedure is performed first to remove as many irrelevant variables as possible in order to reduce the dimensionality from ultrahigh p to a relatively large-scale d which is less than n, as such, the screening procedure can dramatically reduce the computational complexity. A benefit of the screening methods applied in our analysis is that all of them enjoy a sure screening property, that is all truly important predictors can be selected with probability approaching one as the sample size goes to the infinity. After the reduction of the predictors' dimension to a relative large-scale d, which is less than n, well-established regularization methods for variable selection will be used to simultaneously select significant variables and estimate statistical effects of those selected variables. Regularization techniques can improve both the prediction accuracy and model interpretability compared to dynamic factor analysis used in [Ludvigson and Ng 2009a], and like the sure screening property shared by screening techniques, regularization techniques possess an oracle proper, which states the variable selection procedure in regularization is performed as if the true underlying model is given in advance. Compared to [Ludvigson and Ng 2009a], where dynamic factor analysis is used to achieve dimension reduction, the principal components constructed are all linear combinations of all the underlying factors, maybe some

predictors are irrelative to the bond risk premia, as is criticized in [J.-z. Huang and Shi 2011], our approaches can outperform their work by the improvement of the forecast accuracy as well as the interpretability of the predicting model constructed. While in [J.-z. Huang and Shi 2011], for the purpose of dimension reduction, a specific regularization method, supervised adaptive group lasso, is performed, where they divide the set of 131 macro series into eight subgroups first, then perform lasso regularization separately to the macro factors along with their lags in each individual group, since regularization can not deal with the data when predictor dimension p is greater than sample size n, finally, group level lasso is performed on the reduced subgroups and complete the variable selection procedure. Compared to their work, the advantage of our screening procedure lies in that it can handle the ultrahigh-dimensional macroeconomic series directly, and with the guarantee of sure screening property, our approach outshines the proposed method in [J.-z. Huang and Shi 2011]. On the other hand, as is pointed in many statistics literature, lasso regularization, as a penalized least square with the application of L_1 penalty, the resulted penalized estimator from lasso approach is biased and increases the model bias, especially for the large true coefficients. However, in our analysis, SCAD penalty is utilized, which is short for Smoothly Clipped Absolute Deviation, satisfies all the important conditions of a good penalty function as proposed in [Fan and R. Li 2001], namely, unbiasedness, sparsity and continuity. Based on the appealing properties, this concave penalty can give us an unbiased estimator with oracle property in the end. To sum, because of the advantages of our statistics methodology, we can construct some predicting models that can improve the real-time predictive power on the excess bond returns significantly then the already existing macro-based return factors.

To be brief in our study, we will perform two kinds of screening procedures, Sure Independence Screening (SIS) and Feature Screening via Distance Correlation (DCSIS), the regularization method used here will be the penalized least square with concave penalty Smoothly Clipped Absolute Deviation (SCAD). In our linear variable selection framework, an outstanding approach for prediction purpose is SIS $\rightarrow$ SCAD, for ease of reference, the single macro factor constructed from this approach is referred as G_1. In the nonlinear variable framework, between screening and regularization steps, we perform a nonlinearization on the macroeconomic factors selected from screening procedure, the nonlinearization technique considered

in our study is B-spline expanding. Similarly, like linear framework, we have some well-performed approaches, including DCSIS $\rightarrow$ BSPLINE $\rightarrow$ SCAD, we refer the constructed single macro factors from this approach as G_2, respectively. Compared to the factors proposed in [Cochrane and Piazzesi 2005] (CP hereafter), [Ludvigson and Ng 2009a] (LN hereafter) and [J.-z. Huang and Shi 2011] (SAGLasso hereafter), which achieve the prediction on 2- to 5-year maturity excess bond returns with in-sample R^2 up to 0.357, 0.253, and 0.43 respectively, the in-sample performance(R^2) of our proposed factors G_1, G_2 are up to 0.587, 0.664, respectively. Under a careful analysis when incorporating LN factor, we find all of our proposed factors can subsume LN, and if augmented by CP, the forecasting in-sample R^2 increase to 0.606 and 0.689. In addition to the impressive in-sample performance, both of G_1, G_2 exhibit a surprisingly much stronger out-of-sample predictive power than the earlier macro-based return factors, in the measure of out-of-sample R^2 introduced in [Campbell and Thompson 2008], which range up to 0.403, 0.439, 0.440, 0.444 and 0.482, while out-of-sample R^2 of CP, LN and SAGLasso are up to 0.274, 0.189 and 0.278. In general, we can draw the conclusion that, the proposed approaches in our study can outperform the existing literature in the sense of bond risk premia forecast, regardless in sample and out of sample.

As a summary, in this paper, we reconsider the link between macroeconomic series and bond risk premia, compared to existing literature, the important innovations of our study lie in the following aspects, (1) We introduce a linear variable selection framework and a nonlinear variable selection framework for ultrahigh-dimensional dataset. (2) Under the linear and nonlinear variable selection frameworks, we propose different variable selection approaches which consist of different type of combinations among feature screening, nonlinearizaion and regularization. (3) Corresponding to different approaches, we construct different single predictive macro factors respectively, not limit to achieving a much better in sample performance than ever, all of the macro factors we construct can improve the real-time forecast power on the excess bond returns significantly then the already existing macro-based factors as well. (4) Based on the nonlinearization analysis and appealing forecast performance of our new factors, we realize the nonlinear effect of the macroeconomic predictors on excess bond returns, namely, prediction superiority of the nonlinear structure between excess bond returns and macroeconomic fundamentals compared to simple linear structure. (5) Through a

horse race comparison of different approaches, we propose a robust well-performed macro-based predictive model on bond risk premia. (6) A comprehensive analysis on ETF dataset is performed, where in-sample and out-of-sample performance on ETF returns prediction reemphasizes the robustness of the proposed approach.

The organization of the paper is as follows: The next section is mainly about methodology description, introduces the basic set up of our problem, model specification, describes all the advanced statistical techniques used in our analysis, and gives a detailed discussion on the construction of several variable selection approaches and how to construct different macro factors to predict bond risk premia. Section 3 describes the data used in our analysis. In section 4, based on our linear and nonlinear variable selection frameworks, empirical evidence on in-sample and out-of-sample forecast performance will be presented. In section 5, a comprehensive analysis based on our proposed approaches is performed in ETF dataset. Section 6 concludes.

3.2 Empirical Method

In this section, basic set up for the analysis goes first, and then model specification is shown, where the two kinds of models considered in our analysis, linear model and additive model are described. Following model specification, we introduce our linear variable selection framework and nonlinear framework based on the models we consider. Three important building blocks for our frameworks, screening, nonlinearization and regularization are reviewed for the ease of elaboration . Under the constructed frameworks, different linear and nonlinear variable selection approaches are given as well as discussion on how to construct single predictive macro factors to improve forecast power on excess bond returns.

3.2.1 Basic Setup

Following [Fama and Bliss 1987], continuously compounded annual log returns on an n-year zero-coupon Treasury bond in excess of the annualized yield on a one-year zero-coupon Treasury bond is used. With the same notations used in [J.-z. Huang and Shi 2011], from the period $t = 1, \ldots, T$, the excess return is defined as

$$rx^n_{t\to t+12} = r^n_{t\to t+12} - y^{(1)}_t = ny^{(n)}_t - (n-1)y^{(n-1)}_{t+12} - y^{(1)}_t \tag{3.1}$$

where $r^n_{t\to t+12}$ is the one-year log holding-period return on an n-year bond purchased at month t and sold as a $(n-1)$-year bond at month $t+12$, $y^{(n)}_t$ is the month-t log yield on the n-year bond.

3.2.2 Model Specification

In our analysis, we will consider two kinds of models, linear model and additive model, additive model serves as our nonlinear extension beyond linear case.

3.2.2.1 Linear Model

Following [Ludvigson and Ng 2009a] and [J.-z. Huang and Shi 2011], the linear predictive regression in our analysis is defined as follows:

$$rx^n_{t\to t+12} = \beta_0 + \beta_1 F^1_t + \cdots + \beta_p F^p_t + \gamma_1 Z^1_t + \cdots + \gamma_q Z^q_t + e_{t+12} \tag{3.2}$$

where $F^1, \ldots, F^p$ are macroeconomic factors, maybe predetermined, such as GDP growth or CPI, etc. $Z^1, \ldots, Z^q$ are other factors. But in our analysis, we will never predetermine which macroeconomic factor to be incorporated into our final predictive model, our proposed approaches are applied to achieve variable selection, which is an important contribution of this paper.

3.2.2.2 Additive Model

The nonlinear model considered in our study is additive model, additive model (AM) is a nonparametric regression method suggested by [Friedman and Stuetzle 1981], as a restricted class of nonparametric regression models built from one-dimensional smoother, it is less affected by the curse of dimensionality than e.g. a p-dimensional smoother, in addition, compared to a standard linear model, AM is more flexible, while being more interpretable than a general regression surface at the cost of approximation errors. Definition of additive model gives as follows:

$$rx^n_{t\to t+12} = f_1(F^1_t) + \cdots + f_p(F^p_t) + g_1(Z^1_t) + \cdots + g_q(Z^q_t) + e_{t+12} \tag{3.3}$$

where F and Z are predictors introduced as above, $f_1, \ldots, f_p$ and $g_1, \ldots, g_q$ are unknown smooth functions fit from the data.

3.2.3 Variable Selection Framework

Hereafter, for ease of explanation, we will denote the excess bond returns $rx^n_{t\to t+12}$ as variable Y, and denote macroeconomic factors $F^1, \ldots, F^p$ as a predictor matrix X. Detailed description of our proposed variable selection frameworks are as follows.

3.2.3.1 Linear Variable Selection Framework

Based on the standard linear model, we will divide our linear variable selection framework from ultrahigh-dimensional dataset into two stages. First, we perform the well-established screening techniques to dataset X, in the screening procedure, we remove as many irrelevant variables to Y as possible in order to reduce the dimensionality from ultrahigh p(dimension of predictor matrix X) to a relatively large-scale d which is less than n(sample size), as such, the reduced predictor matrix X^* is tractable by the methodology applied in the second stage. After the reduction of the predictor's dimension in the first stage, the newly developed regularization techniques for high-dimensional variable selection will be utilized to simultaneously select significant variables and estimate statistical effects of those selected variables. It is worth mentioning that regularization technique as a modern statistical contribution can improve both the prediction accuracy and model interpretability, which is guaranteed by the oracle property introduced in [Fan and R. Li 2001]. As a conclusion, the two-stage linear variable selection framework from from ultrahigh-dimensional dataset has the structure: SCREENING $\rightarrow$ REGULARIZATION. In the following subsection of 2.4 to 2.6, we will specify the screening techniques and regularization techniques used in our analysis.

3.2.3.2 Nonlinear Variable Selection Framework

Similar to the linear variable selection framework, we still have the stages of screening and regularization, what's more, under the nonlinear framework, we need incorporate the stage of nonlinearization on the macroeconomic variables between screening and regularization. In detail, after the screening stage, natural spline expanding or B-spline expanding to our selected variable matrix X^* are

performed, which results in an augmented predictor space X^{**}, linear combination of the enlarged predictors in X^{**} is an estimator of the smooth function components introduced in the additive model definition, of course, each factor in the augmented predictor space X^{**} can be seen as a nonlinear trend of the corresponding macroeconomic series in X^{*}. With the augmented predictor space, regular regularization techniques are applied, and finish the construction of the final predicting model. From empirical evidence of existing literature and empirical evidence on our in-sample and out-of-sample analysis, we will only consider natural spline expanding and B-spline expanding up to second order. In summary, our nonlinear variable selection framework from ultrahigh-dimensional dataset possesses the structure: SCREENING $\rightarrow$ NONLINEARIZATION $\rightarrow$ REGULARIZATION. However, for the purpose of checking whether variables selected under linear framework has nonlinear effect on excess bond returns, we have another variation of the above nonlinear framework proposed, SCREENING $\rightarrow$ REGULARIZATION $\rightarrow$ NONLINEARIZATION $\rightarrow$ REGULARIZATION. which can be seen as an nonlinear extension of the linear framework in Section 2.3.1. For ease of understanding, following gives a brief review on the screening, nonlinearization and regularization techniques used in our analysis.

3.2.4 Screening

In our analysis, we consider two types of screening methods, Sure Independence Screening (SIS) and Feature Screening via Distance Correlation Learning (DCSIS), a brief description of the two screening approaches is as followes:

A. Sure Independence Screening (SIS), proposed by [Fan and Lv 2008], is a variable screening procedure under ultrahigh-dimensional setting, via Pearson correlation comparison, SIS reduces the ultrahigh dimension down to a relative large scale. Consider the linear regression model:

$$Y = X * \beta + \varepsilon \tag{3.4}$$

where Y denotes average excess returns in our case, and X are the macroeconomic series along with their lags. Under sparsity assumption, we denote $\mathcal{M}_* = \{1 \leq j \leq p, \beta_j \neq 0\}$ as the true model, and $|\mathcal{M}_*|$ denotes the true

model size. If we denote the standardized columnwise design matrix as X_s, and define $\omega = (\omega_1, \omega_2, \cdots, \omega_p)^T$ as

$$\omega = X_s^T * Y \tag{3.5}$$

then ω_j is the marginal Pearson correlation between predictor X_j and Y, scaled by the deviation of Y, moreover, ω_j can also be seen as the least squares estimated coefficient of β_j, as a result, it is reasonable to characterize the marginal relationship between X_j and Y by the magnitude of $|\omega_j|$. As for the practical implement of SIS, calculation of vector ω is conducted first, and then rank all the predictors according to the magnitude of $|\omega_j|$, i.e. we select all the predictors from the subset:

$$\widehat{\mathcal{M}}_{d_n} = \{1 \leq j \leq p : |\omega_j| \text{ is among the first } d_n \text{ largest of all}\} \tag{3.6}$$

In practice, we choose d_n in the level of $O(\frac{n}{\log n})$, under some regularity conditions, [Fan and Lv 2008] gives

$$P(\mathcal{M}_* \subseteq \widehat{\mathcal{M}}_{d_n}) \to 1, \text{as n} \to \infty \tag{3.7}$$

As a result, it demonstrates that SIS can reduce the ultrahigh dimensionality to a relatively large scale $d_n \ll n$, while the reduced model still contains all the true predictors with an overwhelming probability, to be specific, the truly important predictors can be selected to $\widehat{\mathcal{M}}_{d_n}$ with probability approaching one as sample size tends to infinity, i.e. sure independence property by [Fan and Lv 2008] holds.

B. Feature Screening via Distance Correlation (DCSIS), proposed by [R. Li, Zhong, and L. Zhu 2012] is a new strategy which employs the measure of independence to efficiently detect linearity and nonlinearity between predictors and constructs feature screening procedure for ultrahigh-dimensional data.

   Unlike other correlation coefficients defined for two random variables, distance covariance is defined from two random vectors, which are allowed to have different dimensions. For two random vectors $U \in \mathcal{R}^{q_1}$ and $V \in \mathcal{R}^{q_2}$, the

distance covariance is defined by

$$dcov^2(U,V) = \int_{\mathcal{R}^{q_1+q_2}} ||\phi_{U,V}(t,s) - \phi_U(t)\phi_V(s)||^2 \omega(t,s)dtds \tag{3.8}$$

where $\phi_{U,V}(t,s)$ is the joint characteristic function between U and V, $\phi_U(t)$ and $\phi_V(s)$ are corresponding marginal characteristic functions, $\omega(t,s) = \{c_{q_1}c_{q_2}||t||_{q_1}^{1+q_1}||s||_{q_2}^{1+q_2}\}^{-1}$, with $c_d = \pi^{(1+d)/2}/\Gamma\{(1+d)/2\}$. Thanks to the contribution of [Szekely et al. 2007], $dcov^2(U,V)$ becomes data tractable. Where

$$dcov^2(U,V) = S_1 + S_2 - 2S_3 \tag{3.9}$$

$$S_1 = E(\|U - \tilde{U}\|\|V - \tilde{V}\|) \tag{3.10}$$

$$S_2 = E(\|U - \tilde{U}\|)E(\|V - \tilde{V}\|) \tag{3.11}$$

$$S_3 = E\{E(\|U - \tilde{U}\||U)E(\|V - \tilde{V}\||V)\} \tag{3.12}$$

the sample counterparts are

$$\widehat{dcov}^2 = \widehat{S}_1 + \widehat{S}_2 - 2\widehat{S}_3 \tag{3.13}$$

$$\widehat{S}_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|U_i - U_j\|\|V_i - V_j\| \tag{3.14}$$

$$\widehat{S}_2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|U_i - U_j\|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|V_i - V_j\| \tag{3.15}$$

$$\widehat{S}_3 = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\|U_i - U_l\|\|V_j - V_l\| \tag{3.16}$$

The estimated distance correlation between U and V is

$$\widehat{dcorr}(U,V) = \widehat{dcov}(U,V)/\sqrt{\widehat{dcov}(U,U)\widehat{dcov}(V,V)} \tag{3.17}$$

[Szekely et al. 2007] theoretically proved that $dcorr(U,V) = 0$ if and only if U and V are independent, in addition, $dcorr(U,V)$ is a strictly increasing function of the absolute value of Pearson correlation of U and V. Motivated by the appealing properties of distance correlation, [R. Li, Zhong, and L. Zhu 2012] proposed DCSIS procedure to rank the importance of predictors by the

marginal utility:

$$\widehat{\omega_j} = \widehat{dcorr}^2(X_j, Y) \tag{3.18}$$

Under some regularity conditions, Li also proves the proposed feature screening procedure (DCSIS) enjoys the sure screening property, as introduced in SIS section. In addition, DCSIS is asymptotically equivalent to the SIS when (U, V) follows a bivariate normal distribution. Moreover, since the DCSIS can be used for screening features without specifying a regression model between the response and the predictors, it can be seen as a model-free extension of SIS. To summarize, similar to SIRS, model-free is a very appealing feature in that it may be very difficult to specify an appropriate regression model for the actual model in ultrahigh-dimensional setting, this virtue makes the proposed procedure robust to model mis-specification.

3.2.5 Nonlinearization

In our study, we mainly perform one type of nonlinearization techniques, B-spline expanding.

3.2.5.1 B-Spline Expanding

B-spline, or basis spline, is a spline function that has minimal support with respect to a given degree, smoothness, and domain partition. Any function of given degree can be expressed as a linear combination of B-splines of that degree. In B-spline expanding procedure, the domain is subdivided by knots(some predetermined points in the domain), and each B-spline basis function is non-zero on a few adjacent subintervals and, namely, B-spline basis functions are quite "local".

In our analysis, we choose a specific macroeconomic factor X_j^*(after screening stage, and for the consistency with the notation used above) as an example to clarity the B-spline expanding on X_j^*.

In our approach, denote $t_0 = \min_i X_{ij}^*$, and $t_k = \max_i X_{ij}^*$, we choose equidistance partition in the range of X_j^*, i.e. our partition sequence is given as $T = \{t_0 < t_1 < \cdots < t_k\}$, the value of k is predetermined. The t_i's are called knots, the set T the knot vector, and the half-open interval $[t_i, t_{i+1})$ is the i-th knot span.

To define B-spline basis functions, we need one more parameter, the degree of these basis functions, p, in our analysis, second order B-spline functions are applied,

where $p = 2$. The i-th B-spline basis function of degree p, written as $N_{i,p}(t)$, is defined recursively as follows:

$$N_{i,0}(t) = I_{(t_i \leq t < t_{i+1})} \tag{3.19}$$

$$N_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} N_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} N_{i+1,p-1}(t) \tag{3.20}$$

Above is usually referred to as the Cox-de Boor recursion formula. Our target is the calculation of $N_{i,2}(t)$, after the evaluation of $N_{i,2}(t)$ on each points in series X_j^*, we get $X_j^{**} = N_{i,2}(X_j)$, then X_j^{**} is a B-spline expanding basis of X_j^*, in practice, we can predetermine the number of basis *df* for the B-spline expanding, and in our study, the relation $df = k + 1$ holds. As a summary, for each macro series X_j^*, whenever the knots are predetermined, or already relatively fixed by the quantiles of X_j^*, the basis function $N_{i,p}(t)$ has nothing to do with the specific values in the series X_j^*, by evaluating the *df* basis functions $N_{i,p}(t)$, where $i = 1, 2, \cdots, df$, we can construct *df*'s B-spline basis, and each one can be seen as a nonlinear trend of X_j^*. All of the B-spline basis consist the final augmented nonlinear predictor matrix X^{**}

3.2.6 Regularization

As is introduced in Wikipedia(*https* : *//en.wikipedia.org/wiki/Regularization_(mathematics)*), in mathematics, statistics and particularly in the fields of machine learning and inverse problem, regularization refers to a process of introducing additional information in order to solve an ill-posed problem or to prevent overfitting.

In general, a regularization term $\mathcal{R}(f)$ is introduced to a general loss function:

$$\min_f \sum_{i=1}^{n} \mathcal{V}(f(X_i), Y_i) + \lambda \mathcal{R}(f) \tag{3.21}$$

where $\mathcal{V}$ is a loss function describing the cost of predicting f, such as square loss or hinge loss. λ controls the importance of the regularization term, $\mathcal{R}(f)$ is typically a penalty on the complexity of f, such as restriction for smoothness or bounds on the vector space norm. In our setting, we consider a specific form of regularization: penalized least squares problem, and we focus on the discussion about penalties.

Consider the linear regression model,

$$Y = X\beta + \varepsilon \tag{3.22}$$

where $X = (x_1, \cdots, x_n)^T$, $Y = (y_1, \cdots, y_n)^T$, and ε is an n-dimensional noise vector. The penalized least squares(PLS) problem considered is given as followed,

$$\min_{\beta \in \mathcal{R}^p} \{\frac{1}{2n}\|Y - X\beta\|^2 + \sum_{j=1}^{p} P_\lambda(|\beta_j|)\} \tag{3.23}$$

where $\|\cdot\|$ denotes the L_2-norm, $P_\lambda(|\cdot|)$ is the penalty function, λ is the regularity parameter that controls the size of the penalty. The PLS estimator $\widehat{\beta}^{PLS}$ is obtained from the optimization problem (31), if there is no penalty, the resulting estimator is just equal to the ordinary least squares estimate $\widehat{\beta}^{OLS}$.

In the existing literature, most of the well-known penalty functions fall into the category of L_q penalties, with the form $P_\lambda(|\beta_j|) = |\beta_j|^q$, for example, Best Subset Selection corresponds to L_0 penalty, Ridge Regression proposed by [Hoerl and Kennard 1970] uses L_2 penalty, Bridge Regression proposed by [Frank and Friedman 1993] corresponds to L_q penalty with $0 < p < 1$. The famous penalty: Least Absolute Shrinkage and Selection Operator (LASSO) proposed by [Tibshirani 1996] falls into the category of L_q as well, where L_1 penalty is applied. LASSO performs better than other L_q form penalties, as it can shrink some coefficients to zero exactly and results in a sparse model, for the predicting accuracy as well as interpretability, a significant improvement is shown in LASSO approach, this is the reason that LASSO is chosen in [J.-z. Huang and Shi 2011].

However, there are still some shortcomings for L_q penalty, even for LASSO approach. For example, best subset selection can be used to conduct variable selection, but the resulting set of selected predictors is unstable, and computation is often extensive. Ridge regression doesn't possess the variable selection feature(shrinking the coefficient of irrelevant predictor to 0), even though it can shrink the estimated coefficients and make the model stable. LASSO is better than best subset selection and ridge regression, which possesses variable selection feature and can shrink the estimation, but the resulting estimator is biased even for large true coefficients. Based on the unsatisfactory fact, Fan and Li (2001) gives the three important properties which should be satisfied in a good penalty functions,

1. Unbiasedness: The penalized estimator should be nearly unbiased to reduce model bias, especially for the large true coefficients.

2. Sparsity: The penalized estimator can automatically set small estimated coefficients to zero to achieve variable selection and reduce model complexity.

3. Continuity: The penalized estimator is continuous in the data in the sense that it can avoid instability in model prediction.

To construct a penalty satisfying all the conditions mentioned above simulstaneously, [Fan and R. Li 2001] introduced smoothly clipped (SCAD) penalty, whose derivative is given by

$$p_\lambda^{'}(t) = \lambda\{I(t \leq \lambda) + \frac{(a\lambda - t)_+}{(a-1)\lambda} I(t > \lambda)\} \tag{3.24}$$

where $p_\lambda(0) = 0$, and $a = 3.7$ is used(suggested by a Bayesian argument), SCAD takes off at the origin as the L_1 penalty and then gradually levels off.

There are also some good penalties in the sprit of SCAD, like minimax concave penalty (MCP) proposed by [C.-H. Zhang et al. 2010], which contribute a lot to the research in regularization, In this paper, for the ease of comparison and analysis, SCAD is applied and incorporated to our variable selection framework.

3.2.7 Approaches Considered in Our Study

Under linear variable selection framework from ultrahigh-dimensional dataset, we consider the following two-stage approach:

- SIS $\rightarrow$ SCAD

And under nonlinear variable selection framework, we mainly consider the following approach:

- DCSIS $\rightarrow$ Bspline expanding $\rightarrow$ SCAD

SIS and DCSIS denote the specific method used in the screening stage, Bspline expanding means B spline expanding on the predictors selected from screening stage, SCAD denotes the regularization stage, where we always use SCAD penalty in our study.

3.2.8 Construction of Single Predictive Macro Factors from Different Approaches

This subsection is dedicated to clarify the construction procedure of the single predictive macro factors under different approaches proposed above.

For linear approaches, we take SIS $\rightarrow$ SCAD as an example to give a detailed description. Before the implement of our algorithms, we calculate the average excess return (the bond market return), $arx_{t\rightarrow t+12} = \frac{1}{4}\sum_{n=2}^{5} rx_{t\rightarrow t+12}^{n}$, which serves as the dependent series. For the independent predictors, the panel of 131 macroeconomic series matrix, we incorporate their lags up to seven, the enlarged predictor dataset matrix of n by p is named by X, where n is the sample size, in our data set described in next section, equals 528, and p is the dimension of predictor space, in our case, it equals 1048. Dataset X falls into the category of ultrahigh-dimensional setting. In the first stage, screening stage, we perform SIS (Sure Independence Screening) to the predictor matrix X, and screen out a large portion of exogenous explanatory variables (candidate predictors), such that reduce the predictors' dimension to a relative large-scale d, which is less than n. Usually, according to [Fan and Lv 2010], d should be in the level of $O(\frac{n}{\log n})$, in our analysis, we set $d = c * (\frac{n}{\log n})$, where c takes the values of 1, 1.5, 2, 2.5 and 3. The selected predictors from screening stage consist the predictor matrix X^*, which is n by d, where $d < n$, the reduced predictor matrix X^* is tractable by the well-established regularization techniques, in our study, we will perform SCAD penalty to guarantee the unbiased property. Of course, the application of regularization still takes lots of effort, since a suitable tuning parameter needs to be chosen, in our study, we apply the parameter chosen criteria of Bayesian information criterion(BIC) by [Schwartz, 1978]. After the second stage, we have picked up all the important predictors needed in the final predicting regression model, we denote it as $\widetilde{X}$, if we perform standard linear regression of $rx_{t\rightarrow t+12}^{2}$, $rx_{t\rightarrow t+12}^{3}$, $rx_{t\rightarrow t+12}^{4}$, $rx_{t\rightarrow t+12}^{5}$ as well as $arx_{t\rightarrow t+12}$ on $\widetilde{X}$, we will result in our final predicting model, which correspond to 2- to 5-year maturity excess bond return predicting regression model. Moreover, the linear regression estimator on $arx_{t\rightarrow t+12}$,

$$\widehat{arx}_{t\rightarrow t+12} = \widetilde{X} * \widetilde{\beta} \tag{3.25}$$

serves as the single macro predicting factor extract from SIS $\rightarrow$ SCAD approach. Other approaches under linear framework are more or less the same, except that

we perform different types of screening methods in the first stage.

For nonlinear approaches, we take DCSIS $\rightarrow$ Bspline expanding $\rightarrow$ SCAD as an example, similar to linear approaches, we use the same average excess bond return as dependent variable, and macroeconomic series associated with their lags up to seven as initial predictors. After the fist stage work, screening, we shrink the predictor space into the same scale as in linear approaches, which result in a matrix named by X^*, now, the difference between linear and nonlinear approaches appears, following screening stage, we perform second order Bspline expanding to each predictor in matrix X^*, and form an augmented predictor space X^{**}, each series in X^{**} is a nonlinear trend of the corresponding series in X^*, the linear combination of series in X^{**} serves as an estimator of the additive(nonlinear) model introduced above. Next stage is the same as that in linear case, we perform regularization on matrix X^{**} with SCAD penalty, and complete the final variable selection, each selected factor is a nonlinear version of the macro series or their lags. As for the construction of single predicting macro factor, the same procedure as that in linear approaches applies here.

3.3 Data Description

3.3.1 Macroeconomic Time Series

For the ease of comparison with the existing literature, we use the same macroeconomic variables over the time period January 1964 to December 2007 as that in [J.-z. Huang and Shi 2011; Duffee 2011] and [Ludvigson and Ng 2009a]. The macro data set we use consists 131 monthly macroeconomic time series, which have been transformed to induce stationarity initially. Macro data set represent 15 broad categories: real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories; orders; commercial credit; stock indexes; exchange rates; interest rates and spreads; money and credit quantity aggregates; inflation indexes; average hourly earnings; and miscellaneous.

3.3.2 Zero-coupon Treasury Bond Return

The bond return data are taken from the Fama-Bliss dataset available from the Center for Research in Securities Prices (CRSP), and contain observations on one- through five-year zero coupon U.S. Treasury bond prices spanning the period January 1964 to December 2007. Based on the monthly prices, we construct annual excess returns, where annual returns are constructed by continuously compounding monthly return observations, which has been shown more predictability of the annual excess returns rather than monthly excess returns.

3.3.3 ETF Dataset

In the study of ETF dataset, the iShares Treasury Bond ETF managed by BlackRock is analyzed, in specific, prices of the iShares 1-3 Year Treasury Bond ETF with ticker SHY(1 yr - 3 yr), iShares 7-10 Year Treasury Bond ETF with ticker IEF(7 yr - 10 yr) and iShares 20+ Year Treasury Bond ETF with ticker TLT(20+ yr) are utilized, where SHY seeks to track the investment results of an index composed of U.S. Treasury bonds with remaining maturities between one and three years, IEF tracks the investment results of an index composed of U.S. Treasury bonds with remaining maturities between seven and ten years and TLT tracks the investment results of an index composed of U.S. Treasury bonds with remaining maturities greater than twenty years. After computing the annualized log quarterly returns, excess returns are computed from these returns by subtracting the CRSP Riskfree Rates for three-month tenors, respectively, the average of bid and ask yields provided in the Riskfree Rates file are used. Excess returns and macroeconomic variables are over the time period August 2002 to April 2014.

3.4 Empirical Analysis

In this section, we extract important macro factors from the 131 monthly macroeconomic time series based on the proposed approaches, with the constructed single macro factors, we examine their predictive power on the excess bond returns. Section 4.1 clarifies our procedure of analysis, both the in-sample and out-of-sample

case, where we give a detailed description of the data analysis. Section 4.2 reports the performance of our selected factors from the 15 approaches discussed above, including the case of in-sample and out-of-sample, since we mainly stress on the real time prediction, we will focus on the comparison of the out-of-sample performance and give a ranking on the out-of-sample forecast ability. In section 4.3, after choosing the top 5 approaches with the best out-of-sample forecast ability, we construct the corresponding single macro factors and examine their in-sample performance, with the incorporation of CP's and LN's single factors for the ease of comparison purposes, we investigate whether our new factors can capture any information about bond risk premia that is not contained in CP and LN factors. In Section 4.4 we examine the out-of-sample forecasting performance of our single factors considered in Section 4.3 and see whether our single macro factor has a significant incremental predictive power over the existing macro-based factors, from the out-of-sample results we will explain the nonlinear effects of the macro data on excess bond returns prediction, and finally we propose a robust approach for variable selection and forecast. In Section 4.4, we conduct a comprehensive analysis on ETF dataset, and show the methods proposed in the draft have some forecast power on quarterly excess bond returns of ETF as well.

3.4.1 Procedure of the Analysis

This subsection describes the procedure of our in-sample analysis and out-of-sample analysis.

For in-sample analysis stage, we will apply our full sample set (from 1964:1 to 2007:12) to the proposed variable selection approaches in section 2.7, in detail, corresponding to different approaches, different combinations of screening, nonlinearization(if needed, like in nonlinear approaches) and regularization are performed. Each approach results in a set of macro variables, which serves as the final variable selection output. Of course for the approaches in the nonlinear framework, we also extract some nonlinear-formed macro variables. With the selected variables form a specific approach, we can construct the single macro predicting factor according to the description in Section 2.8 and check its in-sample performance on the full sample level, or we can just consider the multivariate regression of the excess bond returns on the variables selected in this approach. Same in-sample analysis can be

carried out for different approaches. As a summary, we extract important macro variables for different approaches, construct single macro predicting factors and conduct the regression analysis using the full sample of data in our in-sample analysis stage.

For out-of-sample analysis stage, we divide the full sample into training portion and testing portion. In detail, for the full sample which spans from 1964:1 to 2007:12, we treat the period of 18 years from 1964:1 to 1981:12 as the initial estimation period (dependent variables from 1965:1 to 1981:12, independent variables from 1964:1 to 1980:12), and the rest data serves as the testing portion, examination of the real time forecast performance (out-of-sample performance) is conducted on the time period between 1982 and 2007. This procedure involves fully recursive factor estimation and parameter estimation using data only through time t for forecasting at time $t+1$. For example, in the month $t = R$, we would like to forecast the values of $\{rx^{(n)}_{R\to R+12}, n = 2, \cdots, 5\}$, the available monthly observation macro series are $\{X_t,\ t = 1, 2, \cdots, R\}$ and the available excess bond returns are $\{rx^{(n)}_{t\to t+12},\ n = 2, \cdots, 5,\ t = 1, 2, \cdots, R-12\}$, for each approach, with the selected variables from in-sample stage, we can use the important predictors to forecast one-step ahead yearly excess bond returns. For $t = R+1$, our available data set is augmented by incorporating the new observation of month R, and we repeat the same exercise on the month $R+1$ forecast. In our analysis, we also incorporate the out-of-sample performance of CP factor(from [Cochrane and Piazzesi 2005]) and LN factor(from [Ludvigson and Ng 2009a]) as a benchmark.

3.4.2 Evidence of the Group Macro Factors from Different Approaches

In this subsection, we conduct ultrahigh-dimensional variable selection based on the 15 different approaches discussed above, as well as in sample regression analysis and out of sample real time forecast examination. But it is worth mentioning that, in the analysis of this subsection, we don't perform single macro factors construction, in the stage of in sample analysis, we perform multivariate regression of the excess bond returns on the selected variables from each approach, and as for the real time forecast examination, we also use the set of selected variables instead of single predictive factor to achieve recursive forecast, that is the reason

we name this section as evidence of the group macro factors instead of evidence of the single macro predictive factors. For the ease of explanation, as discussed before, we divide all the 15 approaches into two different categories, approaches under linear framework and approaches under nonlinear framework. In our evidence of the group macro factors, not limit to the in-sample and out-of-sample performance are reported, according to the comparison among the out-of-sample performance (out-of-sample R squared) of different approaches, we also present the performance ranking, over all the approaches considered as well as ranking within each group, namely, linear approach group and nonlinear approach group. Detailed results are reported in Table 2.

From Table 2, based on the out-of-sample performance ranking across all the approaches proposed, the 2nd best approach is SIS-SCAD, which comes from linear variable selection framework, the 1st best approach is DCSIS-BSPLINE-SCAD. The top performed approach is from nonlinear variable selection framework. Out-of-sample R-squared of SIS-SCAD factors range form 0.387 of 5-year excess bond return to 0.409 of 2-year excess bond return, regardless of the bond maturity, the performance of the nonlinear approach improves significantly, from 0.421 to 0.441 for DCSIS-BSPLINE-SCAD factors. Compared to the out-of-sample forecast power of the proposed factors described in [Cochrane and Piazzesi 2005], [Ludvigson and Ng 2009a] and even [J.-z. Huang and Shi 2011], we have an extremely significant increment. In addition, it is also worth to mention that, in the sense of out-of-sample forecast, there seems a class of robust approaches under the nonlinear framework, that is the form of SCREENING $\rightarrow$ BSPLINE EXPANDING $\rightarrow$ REGULARIZATION, which shows the out-of-sample R-squared ranging from 0.411 for 4-year excess bond return and 0.473 for 2-year excess bond return.

As for the evidence of in-sample analysis, we also achieve extremely appealing in sample performance than previously constructed predicting factors. For the approaches discussed above, in sample performance is also satisfying, ranging from 0.550 to 0.566 for SIS-SCAD approach, from 0.623 to 0.638 for DCSIS-BSPLINE-SCAD. Again, as is seen from the results of the approaches SCREENING $\rightarrow$ BSPLINE EXPANDING $\rightarrow$ REGULARIZATION, in sample evidence is still robust and outperforms than previous literature.

3.4.3 Evidence of the Single Macro Factors from Different Approach: In-sample Analysis

In that the appealing performance of the 2 approaches discussed in last section, regardless of the in-sample case and out-of-sample case, further analysis is worth to be conducted, especially on the analysis on single macro predicting factors from the 2 proposed approaches. For the convenience of explanation, we denote the single macro factors constructed from SIS-SCAD, DCSIS-BSPLINE-SCAD as $\widehat{G}_1$ and $\widehat{G}_2$.

From evidence in Table 3 to Table 4, extensive evidence has indicated that our proposed factors have unconditional statistically significant and economically important predictive power compared with those existing macro-based factors, like CP, LN, SAGLasso, Output Gap factor, Yield-based factor, Cycle factor, and Realized jump factors. All the comparison evidence among these factors can be found in [J.-z. Huang and Shi 2011].

Table 3 represents the estimation results of univariate predictive regressions on $\widehat{G}_1$. The results show that $\widehat{G}_1$ is significant regardless of the bond maturity, the regression R-squared ranges from 0.567 for the 2-year bond to 0.587 for the 3-year bond, this provides evidence that $\widehat{G}_1$ has substantially higher explanatory power than existing macro-based factors. In addition, if we augment $\widehat{G}_1$ by $\widehat{CP}$, the forecasting power is improved, for example, the regression R-squared increases from 0.585 against $\widehat{G}_1$ alone to 0.606 against $\widehat{G}_1$ and $\widehat{CP}$ together, for the 4-year bond, both $\widehat{G}_1$ and $\widehat{CP}$ are significant regardless of bond maturity, which means our $\widehat{G}_1$ factor doesn't subsume $\widehat{CP}$. If we augment $\widehat{G}_1$ by $\widehat{LN}$, the in-sample R-squared doesn't increase too much, and $\widehat{LN}$ becomes insignificant under either HH or NW t-statistic, which means $\widehat{LN}$ factor can capture the macroeconomic information about term premia contained in $\widehat{LN}$ macro factor, and subsume $\widehat{LN}$ as well. Due to the subsumption of $\widehat{LN}$ and the fact that $\widehat{LN}$ and $\widehat{LN}$ are constructed from the same macroeconomic series dataset, our proposed approach SIS-SCAD seems enjoy an appealing information digging potential.

In Table 4, similar analysis is conducted, we find $\widehat{G}_2$ have higher in-sample R-squared compared with $\widehat{G}_1$, in detail, from 0.634 to 0.664 for factor $\widehat{G}_2$. In addition, the same fact as that in Table 3, when we augment the single macro factors by $\widehat{CP}$ and $\widehat{LN}$, we can only improve the explanatory power a little bit, and $\widehat{CP}$ is always significant, but $\widehat{LN}$ is not, which reemphasize the advantage

of our new approaches compared with those old forecast methods, in the sense of information extraction.

For the in-sample analysis of $\widehat{G}_2$, we also incorporate the factor $\widehat{G}_1$. The estimation results in the bottom right of Table 4 shows that augmenting $\widehat{G}_2$ by $\widehat{G}_1$, can substantially improve the forecasting power. And, $\widehat{G}_1$ is significant when it is incorporated with $\widehat{G}_2$.

3.4.4 Evidence of the Single Macro Factors from Different Approach: Out-of-sample Analysis

In this subsection, we examine the out-of-sample R-squared of our single macro factors $\widetilde{G}_1$ and $\widetilde{G}_2$, as well as $\widetilde{CP}$ and $\widetilde{LN}$.

In table 5, we first examine the out-of-sample R-squared of different factors combination, where historical excess returns serve as the benchmark. The second column reports the R-squared of the macro predictors set extracted directly from the approach of SIS $\rightarrow$ SCAD, which equals 0.409, 0.400, 0.406 and 0.387 for 2- to 5-year bonds, the third column shows the R-squared of single factor $\widetilde{G}_1$ against historical average, the results are very close to that in column 2, with out-of-sample R-squared 0.401, 0.389, 0.403, and 0.398, respectively. Evidence from the comparison between column 2 and column 3 shows that the construction of single factor is robust in the sense of real time forecast. When we incorporate $\widetilde{CP}$, we can see some increment in the forecast power, but when augmented by $\widetilde{LN}$ factor, the out-of-sample performance is not improved that much, however, the combination of three factors $\widetilde{G}_1+\widetilde{CP}+\widetilde{LN}$ gives the most substantial improvement of out-of-sample predictive power, making the R-squared equal 0.447, 0.437, 0.455 and 0.434 respective to 2-, 3-, 4- and 5-bonds.

Next, in table 6 we consider the MSE and encompassing tests for four different pairs of unrestricted and restricted specifications and report the results of model comparisons. Column labeled by $\frac{MSE_u}{MSE_r}$ shows the ratio of the MSE of unrestricted model to that of the restricted one. Column labeled "Test Statistic" reports the ENC-NEW test statistics of [T. E. Clark and McCracken 2001], whose 95% (asymptotic) critical value is 1.584. Results reported in panel A of table 6 indicate that by including our single macro factor $\widetilde{G}_1$, the unrestricted model clearly improves over the restricted model that includes only $\widetilde{CP}$ factor. By augmenting by $\widetilde{G}_1$, the MSE

is reduced by about 22-23 percent. Similarly, when we incorporate $\widetilde{G}_1$ to single factor $\widetilde{LN}$, the MSE is reduced by about 26-30 percent. Results about restricted model $AR(6)$ is also reported in table 9. In conclusion, the predictor $\widetilde{G}_1$ is shown to help significantly improve the forecast power of macro variables, not only in sample but more importantly also out of sample.

In table 7-8, we conduct similar analysis to $\widetilde{G}_2$. In detail, for the results in Table 7, we compare the out-of-sample performance against excess bond return historical average, all of $\widetilde{G}_2$ exhibit a better out-of-sample performance compared to factor $\widetilde{G}_1$, from 0.401 to 0.439 for $\widetilde{G}_2$. When we augment the single factors by $\widetilde{CP}$ and $\widetilde{LN}$, we can also see the forecast power increment to some extent, but what is more worth to mention lies that, if we incorporate $\widetilde{G}_1$ to our new factors $\widetilde{G}_2$, we can realize a really extreme significant increment in real time forecast power, to be specific, $\widetilde{G}_2 + \widetilde{G}_1$ results in the out-of-sample R-squared 0.485, 0.495, 0.510 and 0.503 for different bond maturity. Compared to the macro-based factors introduced in previous literature, we really find a quite stronger out-of-sample evidence that our approaches can improve forecasting power than all the factors existed in history, even for financial factors.

In the panel A of Table 8, we also compare the unrestricted model $\widetilde{G}_2 + \widetilde{G}_1$ and restricted model $\widetilde{G}_1$, results show that we can reduce the MSE significantly, and also the Test Statistic ENC-NEW shows there is a significant improvement in the out-of-sample prediction power when we augment the linear factor $\widetilde{G}_1$ by nonlinear single factor like $\widetilde{G}_2$. Similar results when incorporating $\widetilde{CP}$ and $\widetilde{LN}$ are reported in the tables.

To sum, among all the combinations of macro-based single factors we construct in the analysis, there is a robust approach that can result in an extreme satisfying out-of-sample forecast ability, as well as appealing in sample performance, which is single factor constructed by approach of "SCREENING(DCSIS) $\rightarrow$ BSPLINE EXPANDING $\rightarrow$ REGULARIZATION". Moreover, the approach "SCREENING(DCSIS) $\rightarrow$ BSPLINE EXPANDING $\rightarrow$ REGULARIZATION" combined with "LINEAR SCREENING(SIS) $\rightarrow$ REGULARIZATION" can also exhibit a perfect results, both in sample and out of sample.

3.5 Analysis on ETF Data

In this subsection, we conduct in-sample and out-of-sample analysis on the ETF quarterly excess returns against the same macroeconomic time series we used in previous analysis. The ETF data ranges from August 2002 to April 2014, and following [J.-Z. Huang and Y. Wang 2013], we divide our time series into three parts, pre-crisis, crisis and post-crisis, pre-crisis denotes the time period August 2002 - June 2007, crisis denotes the period July 2007 - March 2010, and post-crisis denotes the period April 2010 - April 2014. We conduct the two approaches "SIS $\rightarrow$ SCAD" and "DCSIS $\rightarrow$ BSPLINE $\rightarrow$ SCAD" to our ETF excess returns, and detailed results are listed in table 18 and table 19.

In the quarterly returns table 19, we achieve out-of-sample R-squared ranging from 0.265 to 0.263 for SHY(1-year - 3-year) in the period crisis&post-crisis, out-of-sample R-squared of 0.194 to 0.468 for IEF(7-year - 10-year), and 0.347 to 0.548 for TLT(20+ year). Our factors always have additional forecast power against historical average except for the SHY during period post-crisis. In addition, the whole time period in-sample R-squared are above 30 percent. If we compare the out-of-sample R-squared between nonlinear approaches "DCSIS-BSPLINE-SCAD" against the linear approach"SIS-SCAD", we will find a substantial increment, which also reemphasizes the nonlinear effect of macroeconomic predictors on ETF excess returns.

3.6 Concluion

In our study, we conduct the application of a recently developed methodology on variable selection for high-dimensional data, i.e. screening and regularization techniques to large financial dataset. Through a horse race, a comprehensive comparison of different approaches consisting different combinations of screening methods, nonlinearization techniques and regularization technique, we get a clear picture about the relationship between excess bond returns and macroeconomic predictors. Via the examination of real-time forecast performance of the various approaches, we construct several new single macro factors which can improve the real-time forecast power on the excess bond returns significantly than the already existing macro-based factors. In addition, if we incorporate the nonlinear versions

(B-spline expanding) of original macro data to our analysis, we can result in factors with even more forecast power increment, which indicates the nonlinear effect of the macroeconomic predictors on excess bond returns forecast. Finally, a robust approach for variable selection in macro dataset and excess bond return forecast is suggested, which gives a satisfying out-of-sample forecast ability, as well as in sample performance. Analysis on ETF dataset reemphasizes our main results in Treasury excess bond returns analysis.

Table 3.1: Summary Statistics for the Zero-coupon Treasury Bond Excess Return

This table reports the main Statistics of the zero-coupon Treasury Bond Excess Returns of different maturities, where the sample spans the period January 1964 **to December** 2007

maturity (yr)	2	3	4	5
Min.	-5.5950	-10.4300	-13.5500	-17.5500
25% Qu.	-0.9073	-1.5830	-2.2660	-3.0430
Median	0.2053	0.4101	0.4364	0.4863
75% Qu.	1.6590	2.9600	3.8820	4.4900
Max.	5.9680	10.2600	14.3800	16.8900
Mean	0.4233	0.7150	0.9124	0.8983
Std. Dev.	1.8662	3.4080	4.7121	5.7684
Kurtosis	3.2376	3.3439	3.3522	3.3762
Skewness	0.1192	0.0654	0.1155	0.1201

Table 3.2: Predictability of Factors from Different Approaches

This table reports results from one-year-ahead in-sample and out-of-sample forecast comparisons of n-period log excess bond returns, $rx_{t+1}^{(n)}$. The sample spans the period January 1964-December 2007, and a burn-in period of 18 years is used. Predictors considered include SIS-SCAD factors, DCSIS-SCAD factors, SIRS-SCAD factors, SIS-SCAD-NSPLINE-SCAD factors, SIS-SCAD-BSPLINE-SCAD factors, SIS-NSPLINE-SCAD factors, SIS-BSPLINE-SCAD factors, DCSIS-SCAD-NSPLINE-SCAD factors, DCSIS-SCAD-BSPLINE-SCAD factors, DCSIS-NSPLINE-SCAD factors, DCSIS-BSPLINE-SCAD factors, SIRS-SCAD-NSPLINE-SCAD factors, SIRS-SCAD-BSPLINE-SCAD factors, SIRS-NSPLINE-SCAD factors, and SIRS-BSPLINE-SCAD factors. Columns labeled "Out-of-sample R-squared" report 1 less the ratio of the mean-squared forecasting error of the unrestricted model to the mean-squared forecasting error of the restricted benchmark model (constant). Columns labeled "Rank-overall" and "Rank-within-group" report the rank of forecast performance overall (linear approaches group and nonlinear approaches group together) and within-group (linear approaches group or nonlinear approaches group) respectively.

Approach	R-squared for different maturities							
	Out-of-sample				In-sample			
	2	3	4	5	2	3	4	5
Group 1: Linear Approaches								
SIS-SCAD	0.409	0.400	0.406	0.387	0.561	0.566	0.560	0.550
SIS-SCAD Single Factor	0.401	0.389	0.403	0.398	0.567	0.587	0.585	0.572
Group 2: Nonlinear Approaches								
DCSIS-BSPLINE-SCAD	0.434	0.424	0.441	0.421	0.632	0.634	0.638	0.623
DCSIS-BSPLINE-SCAD Single Factor	0.401	0.426	0.439	0.432	0.634	0.658	0.664	0.646

Table 3.3: Predictability of SIS-SCAD Single Macro Factor: G_1

The return to an n-year zero-coupon Treasury bond from month t to month $t+12$ less the month-t yield on a one-year Treasury bond is regressed on $\widehat{G_{1t}}$ (the SIS-SCAD single factor), $\widehat{CP_t}$ (the Cochrane-Piazzesi 2005 forward-rate factor), $\widehat{LN_t}$ (the Ludvigson-Ng 2011 macro factor). Rows labeled "HH" report test statistics computed using standard errors with the Hansen-Hedrick GMM correction for overlap. Rows labeled "NW" report test statistics computed using standard errors with 18 Newey-West lags to correct serial correlation. Column labeled "Joint Test" reports Wald tests of the hypothesis that all coefficients equal zero. Asymptotic p-values, based on $\chi^2(1)$ distribution, are in brackets. The sample spans the period January 1964 to December 2007

maturity (yr)	$\widehat{G_{1t}}$	$\widehat{LN_t}$	R^2	Joint Test	P-val	maturity (yr)	$\widehat{G_{1t}}$	$\widehat{CP_t}$	$\widehat{LN_t}$	R^2	Joint Test	P-val
2	0.469		0.567			2	0.417	0.108		0.576		
HH	(15.462)			239.06	[0.00]	HH	(13.011)	(1.830)			286.83	[0.00]
NW	(14.425)			208.07	[0.00]	NW	(12.081)	(2.012)			234.59	[0.00]
3	0.872		0.587			3	0.765	0.222		0.599		
HH	(15.669)			245.51	[0.00]	HH	(12.932)	(1.922)			305.08	[0.00]
NW	(14.607)			213.35	[0.00]	NW	(12.414)	(2.144)			249.87	[0.00]
4	1.203		0.585			4	1.013	0.394		0.606		
HH	(13.340)			177.96	[0.00]	HH	(13.087)	(2.383)			252.33	[0.00]
NW	(13.096)			171.49	[0.00]	NW	(12.272)	(2.670)			217.73	[0.00]
5	1.457		0.572			5	1.260	0.468		0.587		
HH	(12.482)			155.80	[0.00]	HH	(12.334)	(1.905)			219.09	[0.00]
NW	(12.518)			156.69	[0.00]	NW	(12.067)	(2.131)			202.15	[0.00]
2	0.433	0.094	0.572			2	0.385	0.104	0.088	0.581		
HH	(7.176)	(1.114)		450.29	[0.00]	HH	(7.938)	(1.682)	(1.058)		409.21	[0.00]
NW	(7.847)	(1.245)		278.04	[0.00]	NW	(8.262)	(1.854)	(1.174)		277.44	[0.00]
3	0.813	0.154	0.591			3	0.714	0.215	0.142	0.603		
HH	(8.080)	(1.119)		348.46	[0.00]	HH	(8.849)	(1.796)	(1.038)		349.93	[0.00]
NW	(8.698)	(1.240)		251.86	[0.00]	NW	(9.115)	(2.003)	(1.142)		265.35	[0.00]
4	1.122	0.212	0.589			4	0.944	0.386	0.191	0.609		
HH	(7.386)	(1.100)		259.75	[0.00]	HH	(8.566)	(2.236)	(1.008)		286.35	[0.00]
NW	(8.083)	(1.229)		209.74	[0.00]	NW	(8.831)	(2.504)	(1.119)		232.57	[0.00]
5	1.369	0.229	0.575			5	1.185	0.399	0.207	0.590		
HH	(7.268)	(0.993)		224.90	[0.00]	HH	(8.386)	(1.792)	(0.910)		258.84	[0.00]
NW	(7.968)	(1.102)		192.37	[0.00]	NW	(8.803)	(2.005)	(1.006)		220.08	[0.00]

Table 3.4: Predictability of DCSIS-BSPLINE-SCAD Single Macro Factor: G_2

The return to an n-year zero-coupon Treasury bond from month t to month $t+12$ less the month-t yield on a one-year Treasury bond is regressed on $\widehat{G_{2t}}$ (the DCSIS-BSPLINE-SCAD single factor), $\widehat{CP_t}$ (the Cochrane-Piazzesi 2005 forward-rate factor), $\widehat{LN_t}$ (the Ludvigson-Ng 2011 macro factor) and $\widehat{G_{1t}}$ (the SIS-SCAD single factor). Rows labeled "HH" report test statistics computed using standard errors with the Hansen-Hedrick GMM correction for overlap. Rows labeled "NW" report test statistics computed using standard errors with 18 Newey-West lags to correct serial correlation. Column labeled "Joint Test" reports Wald tests of the hypothesis that all coefficients equal zero. Asymptotic p-values, based on $\chi^2(1)$ distribution, are in brackets. The sample spans the period January 1964 to December 2007

maturity (yr)	$\widehat{G_{2t}}$	$\widehat{LN_t}$	R^2	Joint Test	P-val	maturity (yr)	$\widehat{G_{2t}}$	$\widehat{CP_t}$	$\widehat{G_{1t}}$	R^2	Joint Test	P-val
2	0.467		0.634			2	0.418	0.119		0.648		
HH	(22.207)			493.14	[0.00]	HH	(13.232)	(2.843)			605.18	[0.00]
NW	(19.518)			380.93	[0.00]	NW	(12.477)	(2.877)			470.24	[0.00]
3	0.869		0.658			3	0.770	0.239		0.674		
HH	(19.475)			379.28	[0.00]	HH	(13.408)	(3.106)			440.43	[0.00]
NW	(18.720)			350.43	[0.00]	NW	(13.112)	(3.136)			411.49	[0.00]
4	1.207		0.664			4	1.042	0.400		0.689		
HH	(17.548)			307.92	[0.00]	HH	(14.149)	(3.854)			380.16	[0.00]
NW	(17.596)			309.61	[0.00]	NW	(13.537)	(3.862)			392.63	[0.00]
5	1.458		0.646			5	1.281	0.427		0.665		
HH	(15.644)			244.74	[0.00]	HH	(12.954)	(3.094)			303.05	[0.00]
NW	(16.101)			259.23	[0.00]	NW	(12.717)	(3.101)			325.00	[0.00]
2	0.449	0.048	0.635			2	0.312		0.215	0.682		
HH	(12.046)	(0.799)		600.68	[0.00]	HH	(7.720)		(4.962)		800.27	[0.00]
NW	(12.069)	(0.824)		411.56	[0.00]	NW	(7.229)		(4.782)		470.45	[0.00]
3	0.844	0.066	0.658			3	0.583		0.397	0.708		
HH	(11.805)	(0.649)		474.37	[0.00]	HH	(8.103)		(5.778)		607.62	[0.00]
NW	(12.184)	(0.662)		387.95	[0.00]	NW	(7.653)		(5.298)		441.48	[0.00]
4	1.178	0.079	0.664			4	0.824		0.532	0.711		
HH	(10.511)	(0.547)		490.05	[0.00]	HH	(7.947)		(5.575)		387.23	[0.00]
NW	(11.244)	(0.563)		383.34	[0.00]	NW	(7.665)		(5.116)		242.07	[0.00]
5	1.431	0.073	0.646			5	0.988		0.653	0.694		
HH	(9.991)	(0.411)		356.15	[0.00]	HH	(7.124)		(5.076)		282.89	[0.00]
NW	(10.732)	(0.420)		312.69	[0.00]	NW	(6.984)		(4.739)		276.93	[0.00]

Table 3.5: Out-of-Sample Predictive Power of SIS-SCAD Factors: G_1 (v.s. Constant)

This table reports results from one-year-ahead out-of-sample forecast comparisons of n-period log excess bond returns, $rx_{t+1}^{(n)}$. The sample spans the period January 1964-December 2007, and a burn-in period of 18 years is used. Predictors considered include the Cochrane-Piazzesi forward-rate ($\widetilde{CP}$), Ludvigson-Ng macro ($\widetilde{LN}$), our single SIS-SCAD factor ($\widetilde{G_1}$) and also the macro factors chosen denoted by $\{g_i^1\}_{i=1,2,...,30}$. Predictors with a tilde indicate that they are constructed recursively month-by-month using only information available at the time of estimation. Columns labeled "Out-of-sample R-squared" report 1 less the ratio of the mean-squared forecasting error of the unrestricted model to the mean-squared forecasting error of the restricted benchmark model (constant).

	Out-of-sample R-square: v.s. constant				
maturity (yr)	$\{\widetilde{g_i^1}\}_{i=1,2,...,30}$	$\widetilde{G}_t$	$\widetilde{G_{1t}} + \widetilde{CP}_t$	$\widetilde{G_{1t}} + \widetilde{LN}_t$	$\widetilde{G_{1t}} + \widetilde{CP}_t + \widetilde{LN}_t$
2	0.409	0.401	0.436	0.412	0.447
3	0.400	0.389	0.424	0.402	0.437
4	0.406	0.403	0.442	0.417	0.455
5	0.387	0.398	0.424	0.409	0.434

Table 3.6: Out-of-Sample Predictive Power of the SIS-SCAD Single Factor: G_1

This table reports results from one-year-ahead out-of-sample forecast comparisons of n-period log excess bond returns, $rx_{t+1}^{(n)}$. The sample spans the period January 1964-December 2007, and a burn-in period of 18 years is used. Predictors considered include the Cochrane-Piazzesi forward-rate ($\widetilde{CP}$), Ludvigson-Ng macro ($\widetilde{LN}$), and our single SIS-SCAD factor ($\widetilde{G_1}$). Predictors with a tilde indicate that they are constructed recursively month-by-month using only information available at the time of estimation. Column labeled "Test Statistic" reports the ENC-NEW test statistic of Clark and McCracken (2001) for the null hypothesis that the benchmark model encompasses the unrestricted model with addition predictors, whose 95% (asymptotic) critical value is 1.584 for testing one additional predictor, the alternative is that the unrestricted model contains information that could be used to improve the benchmark model's forecast.

maturity (yr)	$\frac{MSE_u}{MSE_r}$	Test Statistic	$\frac{MSE_u}{MSE_r}$	Test Statistic
	Panel A: $\widetilde{G}_{1t} + \widetilde{CP}_t$ v.s. $\widetilde{CP}_t$		Panel B: $\widetilde{G}_{1t} + \widetilde{LN}_t$ v.s. $\widetilde{LN}_t$	
2	0.766	136.408	0.709	177.411
3	0.782	125.780	0.734	171.189
4	0.770	120.444	0.719	169.537
5	0.765	122.933	0.718	167.131
	Panel C: $\widetilde{G}_{1t} + \widetilde{CP}_t + \widetilde{LN}_t$ v.s. $\widetilde{CP}_t + \widetilde{LN}_t$		Panel D: $\widetilde{G}_{1t} + AR(6)$ v.s. $AR(6)$	
2	0.840	99.830	0.620	264.549
3	0.878	88.810	0.616	262.711
4	0.863	84.385	0.593	266.564
5	0.853	87.205	0.599	257.976

Table 3.7: Out-of-Sample Predictive Power of DCSIS-BSPLINE-SCAD Factors: G_2 (v.s. Constant)

This table reports results from one-year-ahead out-of-sample forecast comparisons of n-period log excess bond returns, $rx_{t+1}^{(n)}$. The sample spans the period January 1964-December 2007, and a burn-in period of 18 years is used. Predictors considered include the Cochrane-Piazzesi forward-rate ($\widetilde{CP}$), Ludvigson-Ng macro ($\widetilde{LN}$), single SIS-SCAD factor ($\widetilde{G_1}$), our single DCSIS-BSPLINE-SCAD factor ($\widetilde{G_2}$) and also the macro factors chosen denoted by $\{g_i^2\}_{i=1,2,...,39}$. Predictors with a tilde indicate that they are constructed recursively month-by-month using only information available at the time of estimation. Columns labeled "Out-of-sample R-squared" report 1 less the ratio of the mean-squared forecasting error of the unrestricted model to the mean-squared forecasting error of the restricted benchmark model (constant).

	Out-of-sample R-square: v.s. constant					
maturity (yr)	$\{\widetilde{g_i^2}\}_{i=1,2,...,39}$	$\widetilde{G}_{2t}$	$\widetilde{G}_{2t}+\widetilde{G}_{1t}$	$\widetilde{G}_{2t}+\widetilde{CP}_t$	$\widetilde{G}_{2t}+\widetilde{LN}_t$	$\widetilde{G}_{2t}+\widetilde{CP}_t+\widetilde{LN}_t$
2	0.434	0.401	0.485	0.430	0.401	0.428
3	0.424	0.426	0.495	0.463	0.428	0.463
4	0.441	0.439	0.510	0.484	0.442	0.484
5	0.421	0.432	0.503	0.465	0.434	0.465

Table 3.8: Out-of-Sample Predictive Power of the DCSIS-BSPLINE-SCAD Single Factor: G_2

This table reports results from one-year-ahead out-of-sample forecast comparisons of n-period log excess bond returns, $rx_{t+1}^{(n)}$. The sample spans the period January 1964-December 2007, and a burn-in period of 18 years is used. Predictors considered include the Cochrane-Piazzesi forward-rate ($\widetilde{CP}$), Ludvigson-Ng macro ($\widetilde{LN}$), single SIS-SCAD factor ($\widetilde{G_1}$) and our single DCSIS-BSPLINE-SCAD factor ($\widetilde{G_2}$). Predictors with a tilde indicate that they are constructed recursively month-by-month using only information available at the time of estimation. Column labeled "MSE_u/MSE_r" reports the ratio of the mean-squared forecasting error of the unrestricted model to the mean-squared forecasting error of the restricted benchmark model that excludes additional forecasting variables. Column labeled "Test Statistic" reports the ENC-NEW test statistic of Clark and McCracken (2001) for the null hypothesis that the benchmark model encompasses the unrestricted model with addition predictors, whose 95% (asymptotic) critical value is 1.584 for testing one additional predictor, the alternative is that the unrestricted model contains information that could be used to improve the benchmark model's forecast.

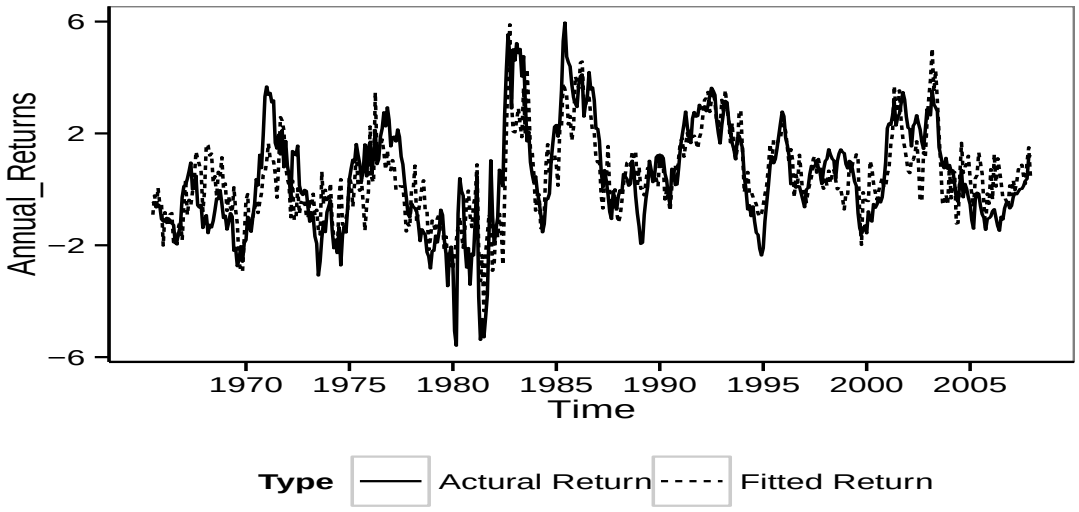
maturity (yr)	$\frac{MSE_u}{MSE_r}$	Test Statistic	$\frac{MSE_u}{MSE_r}$	Test Statistic
	Panel A: $\widetilde{G}_{2t} + \widetilde{G}_{1t}$ v.s. $\widetilde{G}_{1t}$		Panel B: $\widetilde{G}_{2t} + \widetilde{CP}_t$ v.s. $\widetilde{CP}_t$	
2	0.860	70.121	0.774	152.058
3	0.825	78.335	0.730	165.853
4	0.821	81.067	0.713	166.051
5	0.826	79.493	0.712	167.493
	Panel C: $\widetilde{G}_{2t} + \widetilde{LN}_t$ v.s. $\widetilde{LN}_t$		Panel D: $\widetilde{G}_{2t} + \widetilde{CP}_t + \widetilde{LN}_t$ v.s. $\widetilde{CP}_t + \widetilde{LN}_t$	
2	0.723	169.017	0.868	97.881
3	0.703	182.123	0.837	108.615
4	0.689	184.453	0.817	110.353
5	0.688	182.916	0.807	113.896

Table 3.9: Predictability on The iShares Treasury Bond ETF Quarterly Returns

This table reports results from three-month-ahead out-of-sample forecast of log excess returns of the iShares Treasury Bond ETF managed by BlackRock, as well as in sample analysis based on the macro factors selected from approaches of SIS-SCAD, DCSIS-BSPLINE-SCAD, SIRS-BSPLINE-SCAD, SIS-BSPLINE-SCAD. SHY is the ticker of iShares 1-3 Year Treasury Bond ETF, which seeks to track the investment results of an index composed of U.S. Treasury bonds with remaining maturities between one and three years, IEF is the ticker of iShares 7-10 Year Treasury Bond ETF, which tracks the investment results of an index composed of U.S. Treasury bonds with remaining maturities between seven and ten years, and TLT is the ticker of iShares 20+ Year Treasury Bond ETF, which tracks the investment results of an index composed of U.S. Treasury bonds with remaining maturities greater than twenty years. The sample spans the period November 2002-April 2014, and a burn-in period of 56 months is used. Columns labeled "Out-of-sample R-squared" report 1 less the ratio of the mean-squared forecasting error of the unrestricted model to the mean-squared forecasting error of the restricted benchmark model (constant). Pre crisis denotes the time period November 2002 - June 2007, crisis denotes the period July 2007 - March 2010, post crisis denotes the period April 2010 - April 2014, "whole period" denotes the time period November 2002 - April 2014.

Return Type	Approach	Out-of-sample R-squared			In-sample R-squared			
		crisis	post crisis	crisis & post crisis	pre crisis	crisis	post crisis	whole period
SHY(1-3)	SIS-SCAD	0.327	-0.308	0.265	0.230	0.419	0.169	0.426
quarterly	DCSIS-BSPLINE-SCAD	0.258	0.311	0.263	0.570	0.583	0.463	0.601
IEF(7-10)	SIS-SCAD	0.090	0.271	0.194	0.021	0.223	0.434	0.173
quarterly	DCSIS-BSPLINE-SCAD	0.454	0.478	0.468	0.471	0.835	0.690	0.645
TLT(20+))	SIS-SCAD	0.450	0.292	0.347	0.095	0.771	0.353	0.401
quarterly	DCSIS-BSPLINE-SCAD	0.541	0.552	0.548	0.496	0.632	0.829	0.726

(a) In-Sample Fitting of 2-yr Bond Returns



(b) Out-of-Sample Forecasts of 2-yr Bond Returns

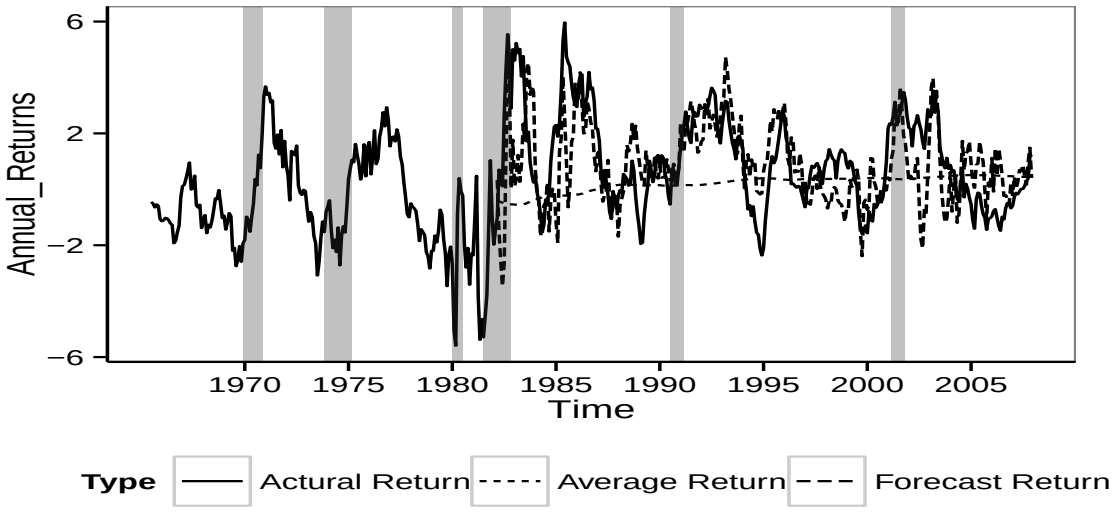
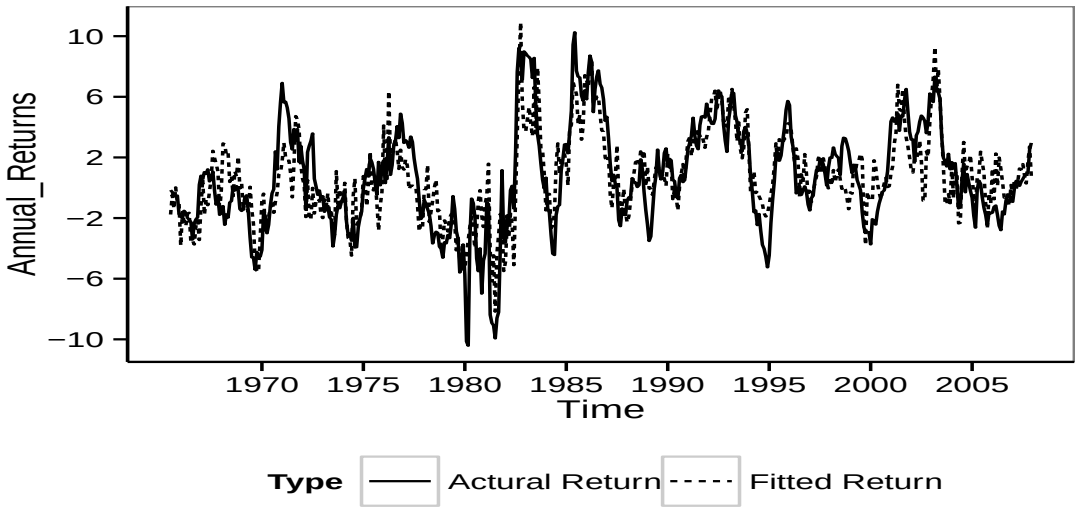


Figure 3.1: Time Variations of 2-yr Bond Returns

The first figure plots the actural 2-year excess bond return as well as the fitted curve on the full sample over January 1964 to December 2007. The second figure plots the actual 2-year excess bond return spanning the time period of January 1964 to December 2007, the forecast 2-year excess bond return form single DCSIS-BSPLINE-SCAD factor as the macro predictor, as well as the historical average of excess return denoted by average return in the graph, both are from January 1982 to December 2007. Shaded bars denote months designated as recessions by the National Bureau of Economic Research.

(a) In-Sample Fitting of 3-yr Bond Returns



(b) Out-of-Sample Forecasts of 3-yr Bond Returns

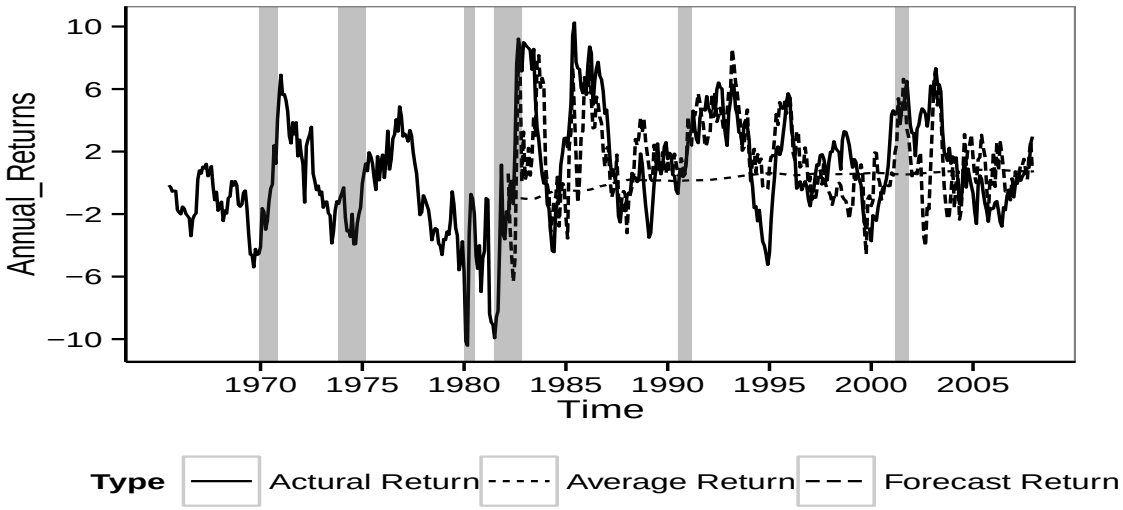
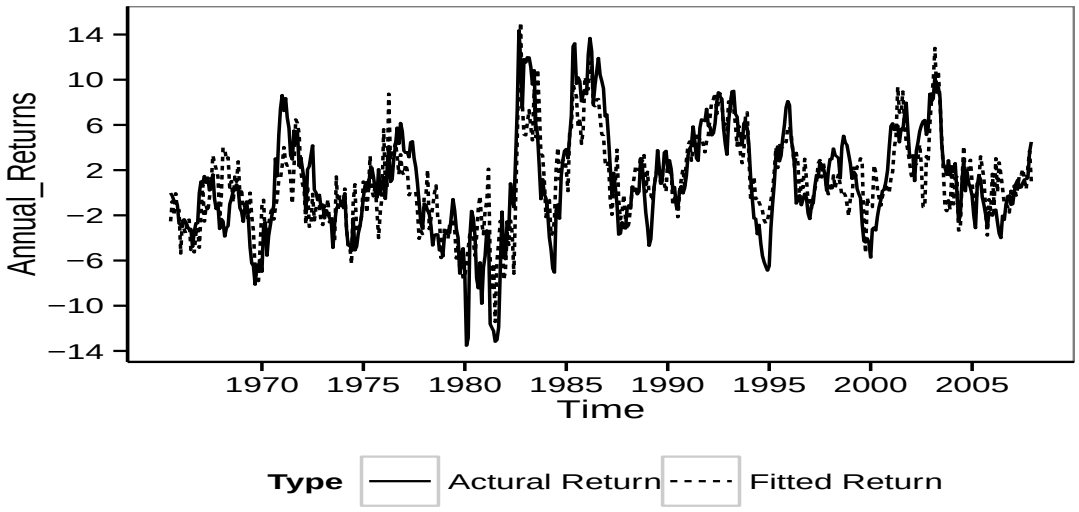


Figure 3.2: Time Variations of 3-yr Bond Returns

The first figure plots the actural 3-year excess bond return as well as the fitted curve on the full sample over January 1964 to December 2007. The second figure plots the actual 3-year excess bond return spanning the time period of January 1964 to December 2007, the forecast 3-year excess bond return form single DCSIS-BSPLINE-SCAD factor as the macro predictor, as well as the historical average of excess return denoted by average return in the graph,, both are from January 1982 to December 2007. Shaded bars denote months designated as recessions by the National Bureau of Economic Research.

(a) In-Sample Fitting of 4-yr Bond Returns



(b) Out-of-Sample Forecasts of 4-yr Bond Returns

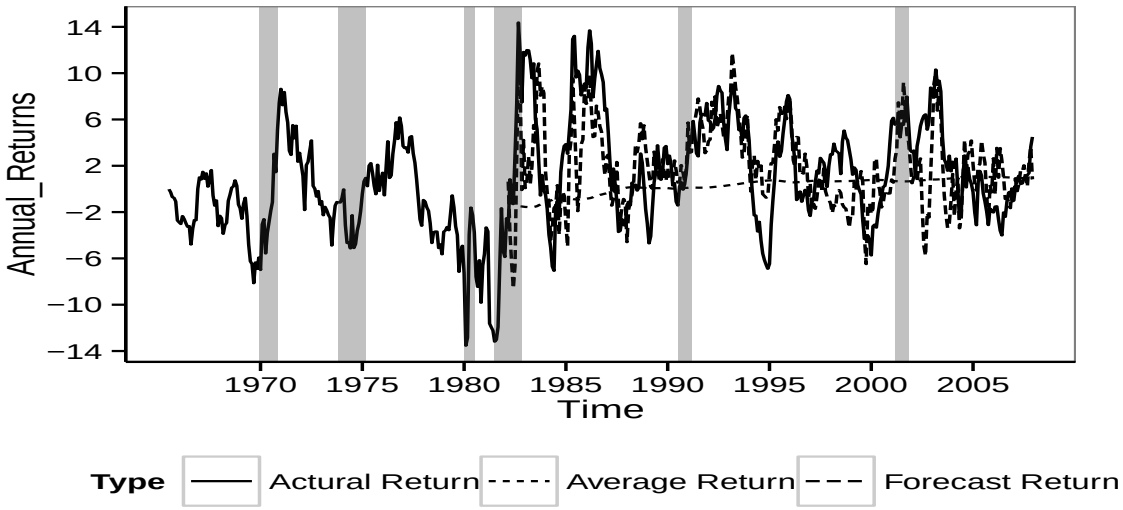
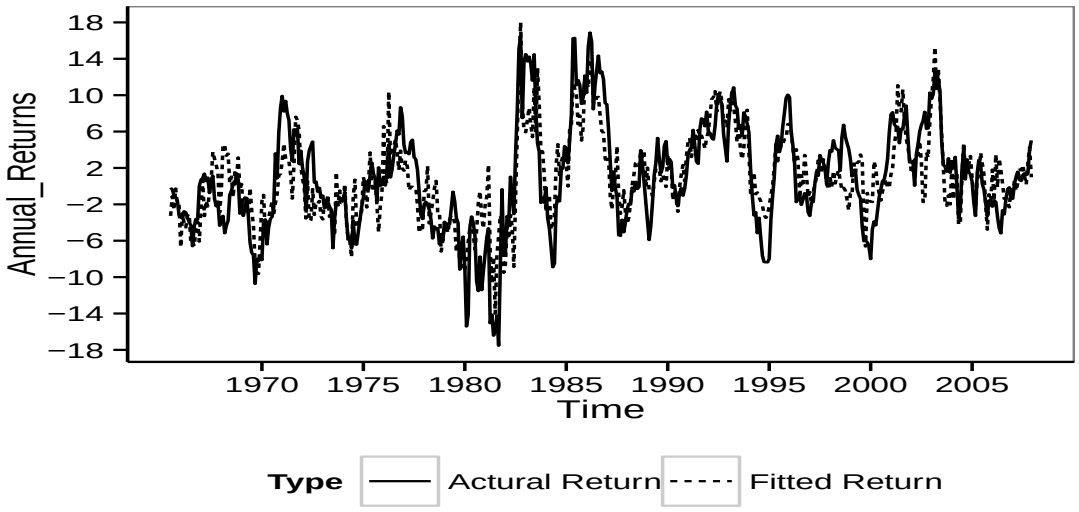


Figure 3.3: Time Variations of 4-yr Bond Returns

The first figure plots the actural 4-year excess bond return as well as the fitted curve on the full sample over January 1964 to December 2007. The second figure plots the actual 4-year excess bond return spanning the time period of January 1964 to December 2007, the forecast 4-year excess bond return form single DCSIS-BSPLINE-SCAD factor as the macro predictor, as well as the historical average of excess return denoted by average return in the graph,, both are from January 1982 to December 2007. Shaded bars denote months designated as recessions by the National Bureau of Economic Research.

(a) In-Sample Fitting of 5-yr Bond Returns



(b) Out-of-Sample Forecasts of 5-yr Bond Returns

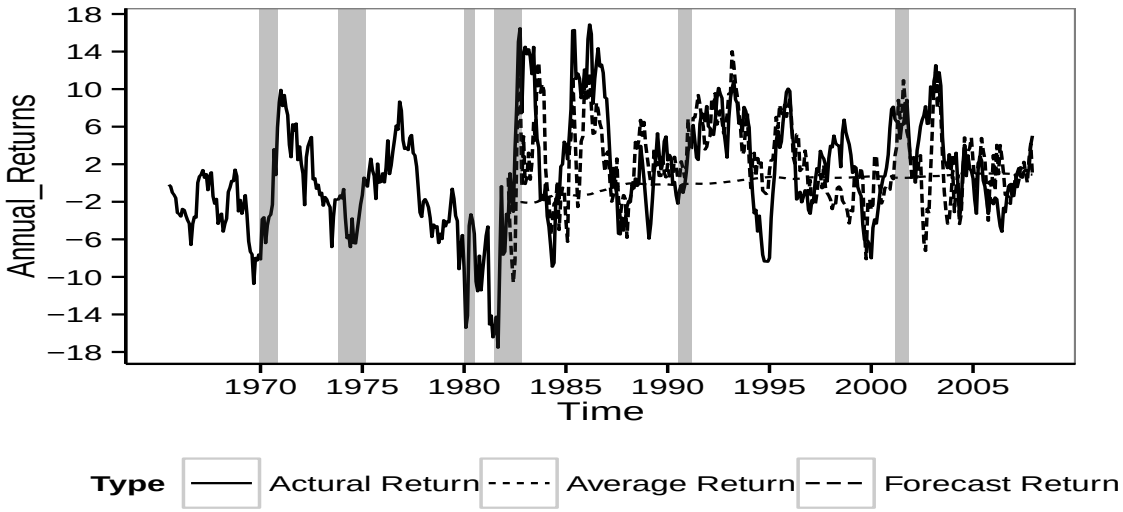
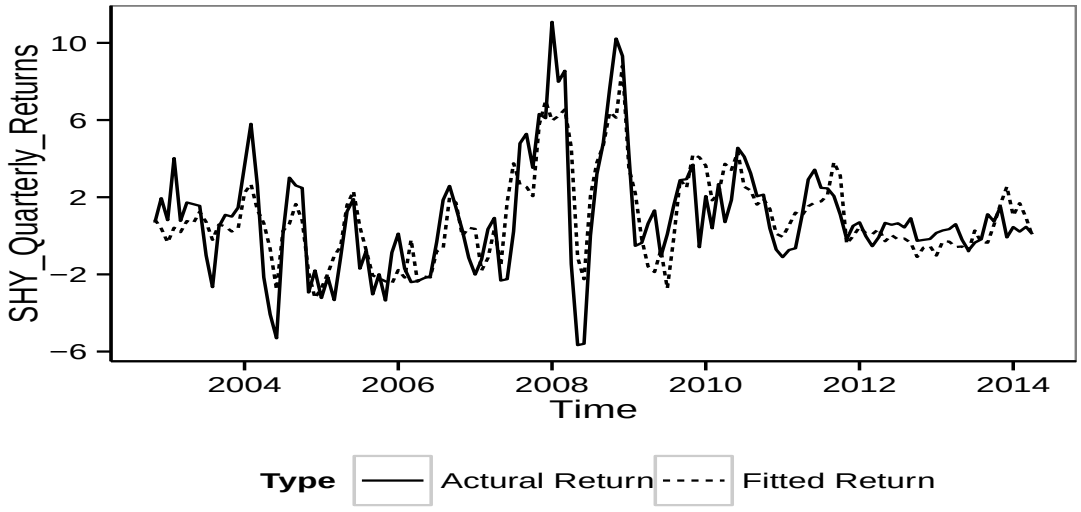


Figure 3.4: Time Variations of 5-yr Bond Returns

The first figure plots the actural 5-year excess bond return as well as the fitted curve on the full sample over January 1964 to December 2007. The second figure plots the actual 5-year excess bond return spanning the time period of January 1964 to December 2007, the forecast 5-year excess bond return form single DCSIS-BSPLINE-SCAD factor as the macro predictor, as well as the historical average of excess return denoted by average return in the graph,, both are from January 1982 to December 2007. Shaded bars denote months designated as recessions by the National Bureau of Economic Research.

(a) In-Sample Fitting of SHY Quarterly Returns



(b) Out-of-Sample Forecasts of SHY Quarterly Returns

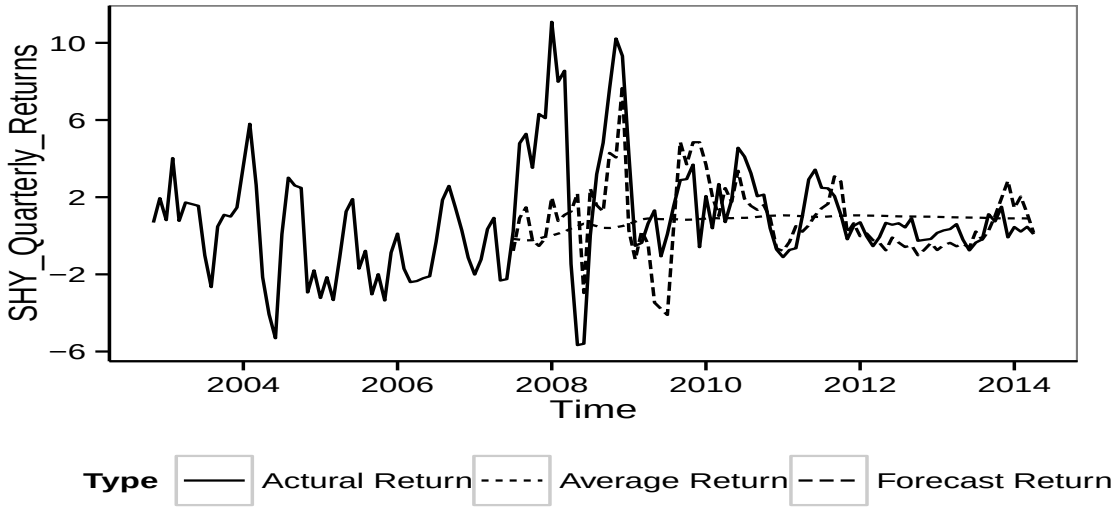
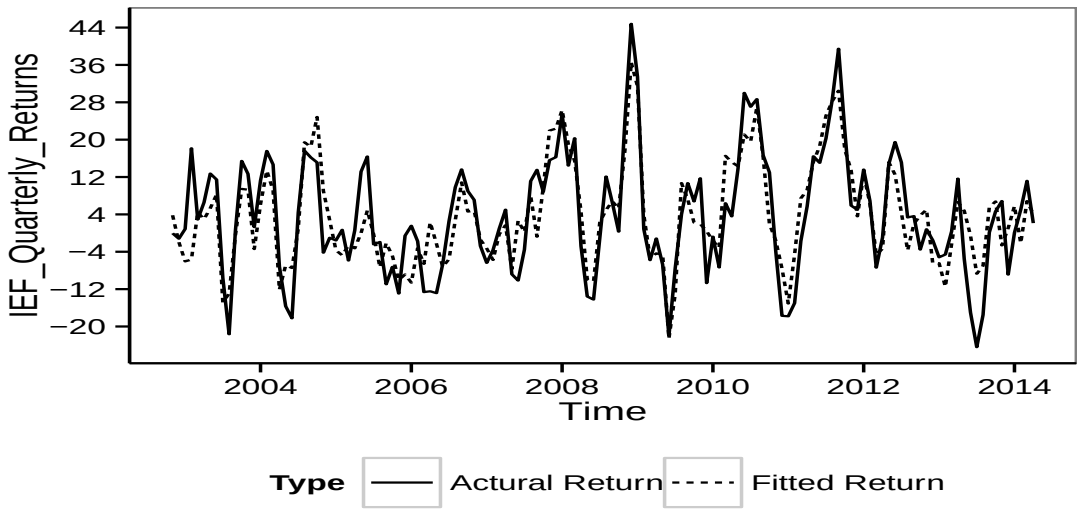


Figure 3.5: Time Variations of Quarterly ETF Returns: SHY(1-yr to 3-yr)

The first figure plots the quarterly annualized log excess returns of the iShares 1-3 Year Treasury Bond ETF(ticker: SHY) managed by BlackRock, as well as the fitted curve on the full sample over November 2002 to April 2014. The second figure plots the actual SHY excess bond return spanning the time period of November 2002 to April 2014, the forecast SHY excess bond return form DCSIS-BSPLINE-SCAD, spanning from July 2007 to March 2010(the 2008 financial crisis period) and from April 2010 to April 2014 (the period after financial crisis), as well as the historical average of excess return denoted by average return in the graph, from July 2007 to April 2014.

(a) In-Sample Fitting of IEF Quarterly Returns



(b) Out-of-Sample Forecasts of IEF Quarterly Returns

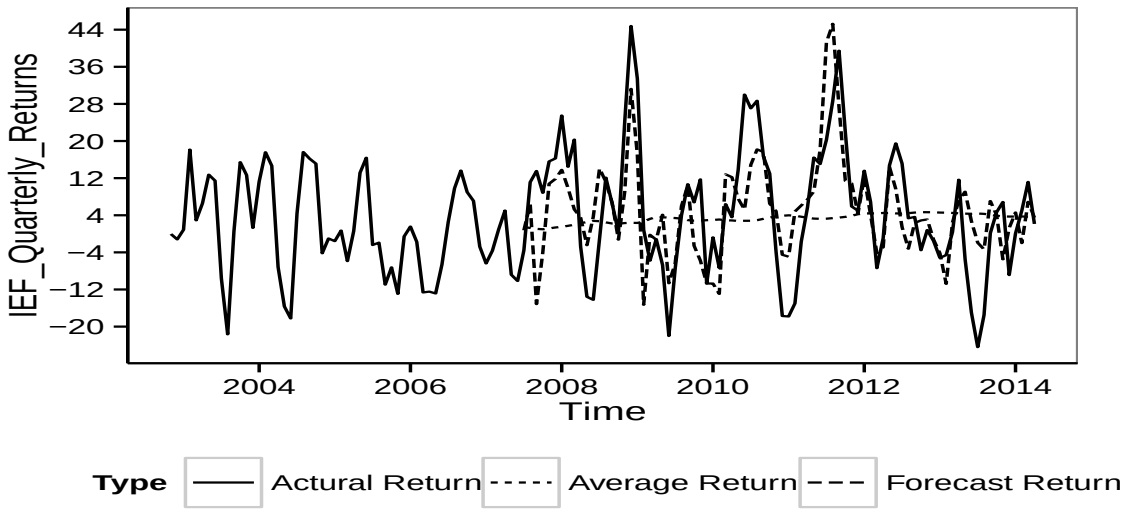
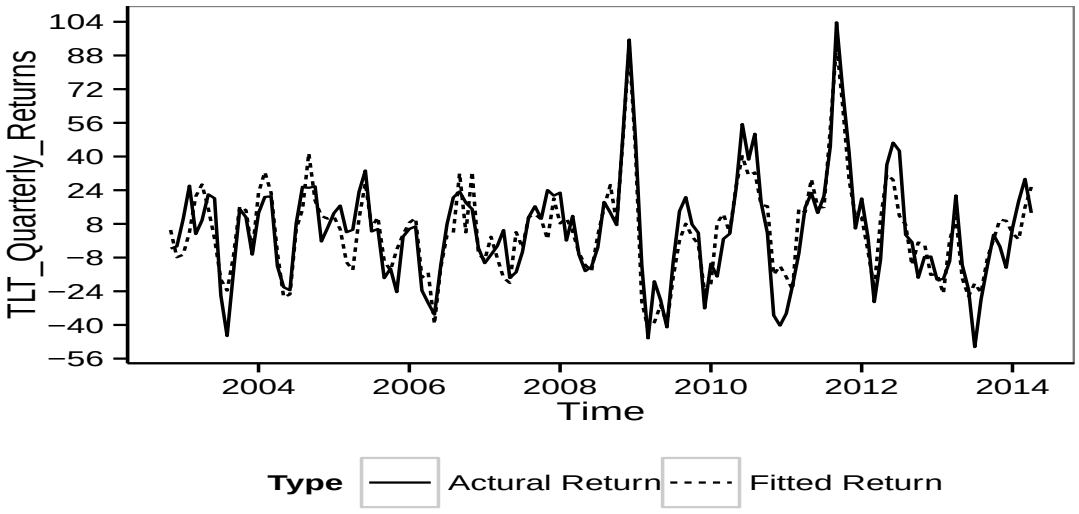


Figure 3.6: Time Variations of Quarterly ETF Returns: IEF(7-yr to 10-yr)

The first figure plots the quarterly annualized log excess returns of the iShares 7-10 Year Treasury Bond ETF(ticker: IEF) managed by BlackRock, as well as the fitted curve on the full sample over November 2002 to April 2014. The second figure plots the actual IEF excess bond return spanning the time period of November 2002 to April 2014, the forecast IEF excess bond return form DCSIS-BSPLINE-SCAD, spanning from July 2007 to March 2010(the 2008 financial crisis period) and from April 2010 to April 2014 (the period after financial crisis), as well as the historical average of excess return denoted by average return in the graph, from July 2007 to April 2014.

(a) In-Sample Fitting of TLT Quarterly Returns



(b) Out-of-Sample Forecasts of TLT Quarterly Returns

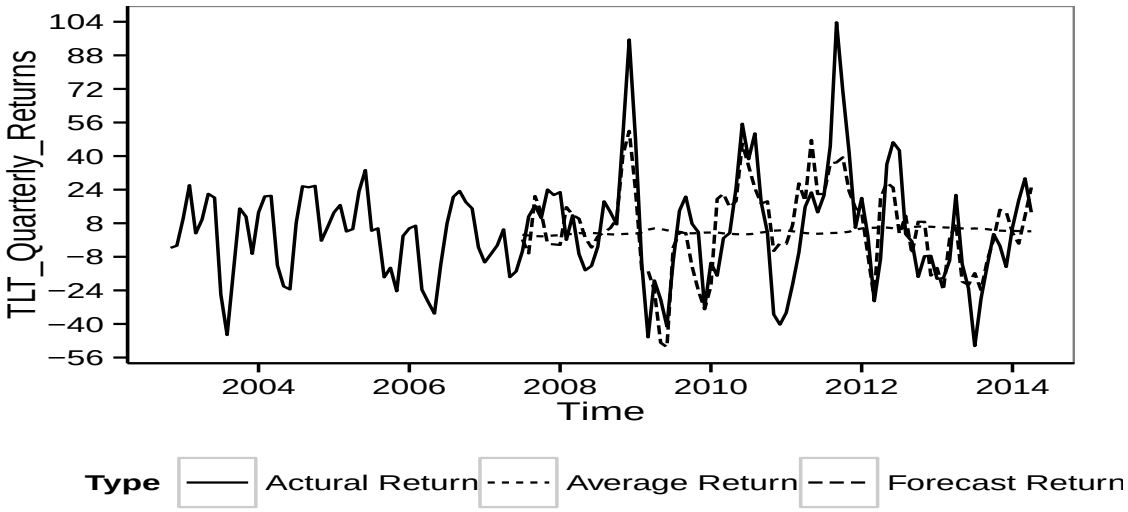


Figure 3.7: Time Variations of Quarterly ETF Returns: TLT(20+ yr)

The first figure plots the quarterly annualized log excess returns of the iShares 20+ Year Treasury Bond ETF(ticker: TLT) managed by BlackRock, as well as the fitted curve on the full sample over November 2002 to April 2014. The second figure plots the actual TLT excess bond return spanning the time period of November 2002 to April 2014, the forecast TLT excess bond return form DCSIS-BSPLINE-SCAD, spanning from July 2007 to March 2010(the 2008 financial crisis period) and from April 2010 to April 2014 (the period after financial crisis), as well as the historical average of excess return denoted by average return in the graph, from July 2007 to April 2014.

Chapter 4
Asset Allocation Methods on Chinese Market

4.1 Introduction

Evidence from empirical research and practical investment have supported and verified the benefit of diversification. Mean-Variance investing is all about diversification. By exploiting the interaction of assets with each other, one asset's gains can make up for another asset's losses. Diversification allows investor's to increase expected returns while reducing risks. In practice, mean-variance portfolios that constrain the mean, volatility and correlation inputs to reduce sampling error have performed much better than unconstrained portfolios.

Mean-variance investing by far is the most common way to choose optimal portfolios. The main takeaway is that diversified portfolios should be selected because investors can reduce risk and increase returns. Many of the popular approaches today, like risk parity, and minimum variance portfolios, are special cases of unconstrained mean-variance portfolios, and the advantage of mean-variance investing is that it allows diversification benefits (and losses) to be measured in a simple way.

Mean-variance frontiers depict the best set of portfolios that an investor can obtain, in the measures of means and volatilities. Given some assets holding. The formal theory behind diversification was developed by Harry Markowitz(1952), who was awarded the Nobel Prize in 1990. The revolutionary capital asset pricing model (CAPM) is laid on the capstone of mean-variance investing. The CAPM pushes the

diversification concept further and derives that an asset's risk premium is related to the (lack of) diversification benefits of that asset. This turns out to be the asset's beta.

Mathematically, diversification benefits are measured by covariance or correlations. Suppose we have two assets, A and B. We use r_A, r_B, σ_A, σ_B to denote their returns and volatilities. We use r_p to denote the portfoio return, then the volatility of portfolio is

$$\sigma_p = \sqrt{\omega_A^2\sigma_A^2 + \omega_B^2\sigma_B^2 + 2\rho_{AB}\omega_A\omega_B\sigma_A\sigma_B} \tag{4.1}$$

where ω_A and ω_B are the portfolio weights held in A and B.

Large diversification benefits correspond to low correlations. Mathematically, the low correlation reduces the portfolio volatility. Economically, the low correlation means that A is more likely to pay off when B does poorly, and the insurance value of A increases. This allows the investor to lower her overall portfolio risk, the more A doesn't look like B, the greater the benefits to add A to a portfolio of B holdings. Mean variance investor love adding investments that act differently from those that they currently hold, the more dissimilar, or the lower the correlation, the better.

The mathematical statement related to a mean-variance frontier is

$$\begin{aligned} \underset{\{\omega_i\}}{\text{minimize}} \quad & Var(r_p) \\ \text{subject to} \quad & E(r_p) = \mu^*, \text{and } \sum_i \omega_i = 1. \end{aligned} \tag{4.2}$$

we want to find the combination of portfolio weights, $\{\omega_i\}$, that minimizes the portfolio variance subject to two constrains, the First is that the expected return on the portfolio is equal to a target return μ^*, the second is that the portfolio must be a valid portfolio, which is the admissibility condition.

From above, we have introduced the unconstrained mean-variance frontiers, but investors oftern face constraints on what types of portfolios that they can hold, one constraint faced by many investors is the inability to short, that is $\omega_i \geq 0$, we can also add this constraint to the optimization problem in equantion (2). And adding short-sale constraints changes the mean-variance frontier, sometimes dramatically. The constrained mean-variance frontier is much smaller than the unconstrained frontier, and it lies inside the unconstrained frontier. The constrained frontier is also not bullet shaped, which is the unconstrained case like. Generally constraints

cause an investor to achieve a worse risk-return trade-off. Nevertheless, even with constraints, the concept of diversification holds: the investor can reduce risk by holding a portfolio of assets rather than a single asset.

In this paper, I mainly take analysis on two data sets, the indices data set, including CSI 300 Total Return Index, CSI 500 Total Return Index, SME Prime Mkt Total Return Index and CSI Aggregate Bond Index; The other is ETF data set, including the ETFs associated with CSI 300, CSI 500, SME Mkt index, as well as Treasury bond index. Through the analysis on these data, I track the performance of various portfolios based on different mean-variance investing. Including Equal Weights Portfolio, Risk Parity Variance Portfolio, Risk Parity Volatility Portfolio, Equal Risk Contribution Portfolio, Minimum Variance Portfolio, Mean-Variance Weights Portfolio, and Kelly-Rule Portfolio. From the horserace among the portfolios, we can compare the out-of-sample performance of each portfolio, as well get a sense on how the parameters affect the performance of each mean-variance method.

As a summary, in this paper, we reconsider the asset allocation methods performance on Chinese market, compared to existing literature, the important innovations of our study lie in the following aspects, (1) We perform the analysis on Chinese equity and bond indices as well as the ETF data set, which gives us some guidance on the real investment. (2) Through the horserace between different mean-variance investment approach, we can get a sense on when and why some specific methods perform better than others. (3) To recheck our explanation on the performance of each asset allocation method, we even perform a simulation and give a proof of our views.

The organization of the paper is as follows: The next section describes the data used in our analysis. Section 3 is mainly about methodology description, introduces the basic set up of our problem, model specification. In section 4, based on our selected mean-variance investing methods, empirical evidence on the out-of-sample trading performance will be presented. In section 5, a comprehensive analysis based on our proposed a pproaches is performed in ETF dataset. In section 6, I perfrom a simulation to further explain why the methods perform well or bad. And section 7 concludes.

4.2 Data Description

4.2.1 Chinese Equity and Bond Indices

Three Chinese equity indices and one Chinese bond index from Wind are incorporated in the analysis. Chinese equity indices are CSI 300 Total Return Index, CSI 500 Total Return Index, and SME Prime Mkt Total Return Index, the Chinese bond index I used is CSI Aggregate Bond Index. The corresponding inception date of each series are 04/03/2016, 01/15/2007, 12/30/2005 and 12/31/2002, the unified end date in our analysis is 03/23/2017. For detailed descriptions, please refer Table 1. In our analysis on the indices data set, the full data set we use is from January 2007 to December 2016, and our testing out-of-sample paper trading period extends from January 2010 through December 2016.

4.2.2 Chinese Equity and Bond ETF Dataset

In the study of ETF dataset, the Chinese equity and bond ETFs with source from WInd are used, in specific, Chinese equity ETFs include Huataiborui Shangzheng Zhongxiaoban ETF (510200.OF), Huataiborui Hushen 300 ETF (510300.OF), and Nanfangzhongzheng 500 ETF (510500.OF), the Chinese bond ETF usd in the analysis is Guotai Shangzheng 5Nianqi Guozhai ETF (511010.OF). Compared to indices data, the inception dates are near, with 01/26/2011, 05/04/2012, 02/06/2013 and 03/05/2013 accordingly, and we use the data before 09/14/2017. Table 12 gives detailed description of ETF data set. The full data used in this study extends from May 2013 to August 2017, and data from April 2014 through August 2017 are used as the testing out-of-sample paper trading period.

4.3 Empirical Method

In this section, basic set up for the analysis goes first, and then model specification is shown, where the eight kinds of models considered in our analysis, Equal Weights, Risk Parity Variance, Risk Parity Volatility, Equal Risk Contribution, Proportional to Sharpe Ratio, Minimum Variance, Mean-Variance Weights, and Kelly-Rule are described.

4.3.1 Basic Setup

Following traditional asset allocation analysis, we fix our reblancing horizon as one month. Namely, in the beginning of each month in the testing trading period, for example, January 2010 of indices data set, and April 2014 of ETFs data. we will estimate the expected return and expected covariance matrix based on the historical return series available upon that time, and then calculate the proper weights invested on each asset based on the proposed asset allocation methods. After one month, we reblance again, we repeat the reblance action till the end of our testing trading period, for indices data, December 2016 and August 2017 for ETFs data. For the step we estimate the expected return and covariance matrix, we have the flexibility to choose the lengh of training data, we name it as time window, it ranges from 3 months to 36 months for indices study, and ranges from 1 month to 12 months for ETF study, due to the shorter history compared to indices data set. As this is a study on Chinese market, we restrict us without short sale and full investment on all asset classes, which means $\{\omega_i \geq 0, \sum_i \omega_i = 0\}$

4.3.2 Model Specification

In our analysis, we will consider eight kinds of models, Equal Weights, Risk Parity Variance, Risk Parity Volatility, Equal Risk Contribution, Proportional to Sharpe Ratio, Minimum Variance, Mean-Variance Weights, and Kelly-Rule. Detailed description are as follows.

- Equal Weights Portfolio
  Equal Weights, or 1/N rule, is a strategy which simply holds 1/N weight in each asset class, if we consider N asset classes in total. Namely, the asset weights are $\{\omega_i = \frac{1}{N},\ \text{for i in } 1, 2, \cdots, N\}$

- Risk Parity Variance Portfolio
  Risk Parity Variance is just a strategy which chooses asset weights proportional to the inverse of return variance. The corresponding asset weights are $\{\omega_i = \frac{1/var(r_i)}{\sum_{j=1}^{N} 1/var(r_j)},\ \text{for i in } 1, 2, \cdots, N\}$

- Risk Parity Volatility Portfolio
  Risk Parity Volatility is similar to Risk Parity Variance, but the weights

are proportional to the inverse of return volatility, which means we will put more weights on the asset class with smaller risk. The asset weights are $\{\omega_i = \frac{1/std(r_i)}{\sum_{j=1}^{N} 1/std(r_j)},\ \text{for i in } 1, 2, \cdots, N\}$

- Equal Risk Contribution Portfolio
  Given the volatility of a portfolio consisting N asset classes,

$$\sigma_p = \sqrt{\sum_{i=1}^{N} \omega_i^2 \sigma_i^2 + \sum_{i \neq j} 2\rho_{ij}\omega_i\omega_j\sigma_i\sigma_j} \tag{4.3}$$

  we can define the risk contribution of asset i as

$$\sigma_{x_i} = \omega_i * \frac{\partial \sigma_p}{\partial \omega_i} \tag{4.4}$$

  then, the asset weights are defined as the solution to the following equations:

$$\begin{aligned} &\underset{\{\omega_i\}}{\text{solve}} && \sigma_{x_1} = \sigma_{x_2} = \cdots = \sigma_{x_N} \\ &\text{subject to} && \omega \geq 0, \text{and } \sum_i \omega_i = 1. \end{aligned} \tag{4.5}$$

- Proportional to Sharpe Ratio Portfolio
  Proportional to Sharpe Ratio is a strategy with weights proportional to historical sharpe ratio, which holds larger positions in assets that have larger realized sharpe ratios. The asset weights are $\{\omega_i = \frac{SR(r_i)}{\sum_{j=1}^{N} SR(r_j)},\ \text{for i in } 1, 2, \cdots, N\}$

- Minimum Variance Portfolio
  Minimum Variance is the portfolio on the left-most tip of the mean-variance frontier, where the asset weights $\{\omega_i\}$ are the solution of the following optimization problem,

$$\begin{aligned} &\underset{\{\omega_i\}}{\text{minimize}} && Var(r_p) \\ &\text{subject to} && \omega \geq 0, \text{and } \sum_i \omega_i = 1. \end{aligned} \tag{4.6}$$

- Mean-Variance Weights Portfolio
  In Mean-Variance Weights, the asset weights are chosen to maxmize the

historical portfolio sharpe ratio. Namely, the asset weights are solutions to the following problem,

$$\begin{aligned} \underset{\{\omega_i\}}{\text{minimize}} \quad & \frac{r_p - r_f}{std(r_p - r_f)} \\ \text{subject to} \quad & \omega \geq 0, \text{and } \sum_i \omega_i = 1. \end{aligned} \tag{4.7}$$

- Kelly-Rule Portfolio
  Kelly-Rule is a portfolio strategy that maximizes the expected log return, in detail, define the expected log return of the portfolio as

$$\begin{aligned} & E[\log(\omega_1(1+r_1) + \omega_2(1+r_2) + \cdots + \omega_N(1+r_N))] \\ =& E[1 + \log(\omega_1 r_1 + \omega_2 r_2 + \cdots + \omega_N r_N)] \\ \approx& E[\sum_{i=1}^{N} \omega_i r_i - \frac{1}{2}(\sum_{i=1}^{N} \omega_i^2 r_i^2 + 2\sum_{i \neq j} \omega_i \omega_j r_i r_j)] \\ =& (\omega_1, \cdots, \omega_N) * E[r_1, \cdots, r_N]^, \\ & - \frac{1}{2}(\omega_1, \cdots, \omega_N) * E[(r_1, \cdots, r_N)^, * (r_1, \cdots, r_N)] * (\omega_1, \cdots, \omega_N)^, \end{aligned} \tag{4.8}$$

  the asset weights are chosen to maximaze the quantity above, of course with the restirctions $\omega_i \geq 0$ and $\sum_{i=1}^{N} \omega_i = 1$.

4.4 Empirical Analysis

In this section, we examine out-of-sample paper trading performance of different asset allocation methods, check the relationship between the methodology behaviors and their training horizons used for parameters estimation, from the observed empirical results, we try to explain why and when each asset allocation method performs better or worse compared to each other. Section 4.1 introduces main procedures of our analysis. Section 4.2 clarifies the details of the indices data set we used, which includes the description of average returns, annualized volatility and sharpe ratio, as well as the correlations between each pair of asset class indices. Section 4.3 reports the performance of each asset allocation methods from the 8 approaches discussed above, including different cases of incorporating different asset classes into the universe, since we mainly stress on the real time trading

performance, we will focus on the comparison of the out-of-sample performance and give a comparison on the out-of-sample trading behaviors. From section 4.4, we conduct a comprehensive analysis on ETF dataset. In section 4.4, similar to section 4.2, we give a extensive description on the statistics of ETFs data. Section 4.5 shows the similar empirical trading analysis on ETFs data. And finally, section 4.6 gives a simulation result to explain unconditional mean-variance weights work better than risk parity if parameters are estimated accurately.

4.4.1 Procedure of the Analysis

This subsection describes the procedure of our paper trading analysis. For both the indices and ETF data sets, we extract an out-of-sample trading sample from the full sample. In detail, indiecs data's full sample period spans from 2007:1 to 2016:12, we treat the first 3 years as the initial estimation period , which will be explained later, and rest as the out-of-sample trading period, from 2010:1 to 2016:12. Similarly to indices data, the corresponding full sample and out-of-sample trading period for ETF data is 2013:5-2017:8 and 2014:4-2017:12.

Suppose in the beginning of month t, we have V_t dollars in our account, we need to decide how much to invest in each asset class, let's take mean-variance weights as an example, before we figure out the weights $\{\omega_{it}, i = 1, 2, \cdots, N\}$ on all the asset classes, we need to estimate the expected returns $\{r_{it}, i = 1, 2, \cdots, N\}$ and covariance matrix of asset classes returns Σ_t, where some historical data is needed, this estimation used sample length(we will name it as time window in our analysis, denoted by n months) is some parameter we can choose, namely, we will use the data from month $t-n$ to month t to as the traing sample to estimate the parameters we need, in our analysis, we take 3, 6, 12, 24 and 36 months for the indices study, and 1, 2, up to 12 months for ETF study. After calculation, we can decide the investment weights $\{\omega_{it}\}$ on each asset class, then we hold the proposed positions for one month, until the beginning of next month, we rebalance our positions, we will repeat what we do in the previous month, namely, in the beginning of each trading month till the end of trading month, we will always rebalance one time and record the total value V_t in our account and the corresponding investment weights $\{\omega_{it}, t = 1, 2, \cdots, T\}$.

Finally, with the series of account value $\{V_t\}$, we can calculate our trading

performance like annualized return, volarility and sharpe ratio, as well as average weights on each asset class. For our analysis, we have the flexibility to change the asset allocation method, parameter estimation time window as well as the asset class universe, from different combinations, we can give a clear picture of the horserace between different asset allocation methodology. We will show the analysis results in the following sections.

4.4.2 Statistics of Indices Data

In this subsection, we give an extensive description on the indices data statistics. In the indices data set, we have four series in total, CSI 300 Total Return Index, CSI 500 Total Return Index, SME Prime Mkt Total Return Index and CSI Aggregate Bond Index. The full sample spans from January 2007 to December 2016, and the out-of-sample trading sample we used spans from January 2010 to December 2016.

Table 2 summarizes the main statistics of sample daily returns. In panel A, for the full sample case, average annualized return of CSI300, CSI500, SME Mkt are 10.63%, 20.49% and 21.20%, annualized return of the Agg Bond is much smaller, which is 4.25%, the annualized volatilities are 30.37%, 34.58%, 32.56% and 1.84%, due to the extremely small volatility of bond, the resulting annualized sharpe ratio of Agg Bond is 1.0808, while the sharpe ratios of CSI300, CSI500 and SME Mkt are not that big, only 0.2913, 0.5407, and 0.5916. If we consider only the out-of-sample trading period sample, the average annualized returns of equities decreased a lot compared to the full sample case, only 4.49%, 9.64% and 11.68%, which implies the equity market didn't perform well from 2007 to 2009. However, The return of Agg Bond didn't change too much, around 4.83%, and the volatilities keep in the same level as the fulls sample, around 25.03%, 28.94%, 28.55% and 1.57% for the equities and bond. The corresponding sharpe ratios are .1022, 0.2661, 0.3414 and 1.8481 accordingly. Panel B summarizes the correlations among different asset classes, if we compare the full sample case with the out-of-sample case, we will find the correlations keep stable, didn't change to much, the equity indices keep a extremely high correlation, from 0.813 to 0.979, however, correlation between bond and equities are around 0 during these years.

Similar statistics for monthly returns are reported in Table 3. We find Agg Bond still achieves the least volatility and largest sharpe ratio compared to equities,

as well, equity indices show high correlations, but with a close 0 correlation with Agg Bond.

4.4.3 Evidence of Indices Study

In this subsection, we conduct the horserace comparison of all the out-of-sample paper trading based on different asset allocation methodologies. Including the comparison of their annualized returns, volatilities, sharpe ratios and average weights on different asset classes, corresponding to different parameter estimation time windows and investment universe.

For the ease of explanation, as discussed before, we divide the time windows into five different categories, 3, 6, 12, 24 and 36 months. As for the investment universe, we consider the first combination of Agg Bond and CSI300 (bond and large value equity index), the second is the combination of Agg Bond, CSI300 and SME Mkt (bond, large cap value equity and mid cap value equity), after that, we take the third combination of Agg Bond, CSI300, SME Mkt and CSI500 (bond, and all cap value equities), finally, we take only the stocks CSI300, SME Mkt and CSI500 as the last combination to check the performance of each asset allocation method on pure equities.

We examine the out-of-sample monthly trading performance of different asset allocation methods on the investment universe of only CSI300 and Aggregate Bond. The parameter estimation time windows we use are 3, 6, 12, 24 and 36 months. From Table 4, we find that Risk Parity Volatility, Risk Parity Variance, Equal Risk Contribution and Minimum Variance are the most best ones in terms of sharpe ratio, and the performance is stable regardless of time windows. Risk Parity Volatility has the sharpe ratios ranging from 0.7078 to 0.764, sharpe ratios of Risk Parity Variance range from 0.786 to 0.7929, sharpe ratios of Equal Risk Contribution are from 0.6141 to 0.8295, and time window of 6 months achieves the largest sharpe ratio. Minimum variance has the sharpe ratios between 0.7399 and 0.7896. These methods perform better compared to others is because that they put a lot of weight on bond. Another reason is that, we all know asset returns are difficult to estimate, while the covariance matrix between asset classes is easy to estimate relatively. These four methods only require the estimation of expected covariance matrix, and get rid of the trouble of estimation error on expected asset

returns. As for Proportional to Sharpe Ratio, Mean Variance Weights and Kelly Rule, the performance is not stable with the time windows chosen. In Table 5, we list the best performance (in terms of sharpe ratio) of each method if we choose the best time window associated with each method, and we also list the corresponding average positions on each asset class for each method. We can find Equal Risk Contribution achieves the best performance with sharpe ratio 0.829471, Equal Weights has the least sharpe ratio of 0.144253. If we look at Panel B, we find the best performed methods really have a lot of weights on bond.

In Table 6 and Table 7, we enlarge our investment universe to CSI300, SME Mkt and Aggregate Bond. If we compare Table 6 with Table 4, we find after we incorporate one more asset class, Equal Weights performs better tremendously than before, Risk Parity Volatility, Risk Parity Variance, Equal Risk Contribution and Minimum Variance don't change too much in terms of sharpe ratio, are still the best four. Besides, we find Mean Variance Weights is better than before, and there is the evidence that Mean Variance Weights will perform better if we consider a longer time window for parameter estimation. Similar to that in Table 5, Table 7 shows the performance of different methods and the corresponding average positions, the best four methods still have a large position in bond.

When we enlarge the investment universe to include CSI500 as well, the detailed results are shown in Table 8 and 9. Finally, we consider the case when we only have the equity indices in our universe, as shown in Table 10, if the volatilities of the asset classes are around in the same level, Risk Parity Volatility, Risk Parity Variance, Equal Risk Contribution and Minimum Variance can not show a dominant benefit than others, on the contrary, Mean Variance Weights and Kelly Rule show us a better performance, and we find the longer time window we use, the higher the sharpe ratios are. For average holding positions of each method on each asset class, please check Table 11.

4.4.4 Statistics of ETF Data

Similar to the statistics for indices data, in Table 13, daily returns statistics of ETF data are reported. For CSI300 ETF, CSI500 ETF, SME Mkt ETF and Treasury Bond ETF, the full sample annualized returns are 17.08%, 23.55%, 19.53% and 2.44%, the corresponding annualized volatilities are 25.12%, 29.60%, 28.66% and

2.14%, the resulted the sharpe ratios are 0.596, 0.7245, 0.608, and 0.1569. If we burn in the first year of the ETF data, only consider the out-of-sample period, the returns increase significantly, around 25.56%, 24.87%, 23.51% and 3.78%, the volatilities keep the same level, around 26.15%, 31.15%, 30.35% and 2.00% accordingly, because the high returns, the sharpe ratios are higher than the full sample case, which are 0.9021, 0.7353, 0.7096 and 0.9045. In panel B of Table 13, equities still possess high positive correlations, and the correlation between equities and bond is close 0. Similar results are shown for the monthly ETF returns case, for detail please check Table 14.

4.4.5 Evidence of ETFs Study

In this subsection, we compare the trading performance of different asset allocation methods on ETF data. In this ETF study, we still consider four cases of different investment universe. Case 1: CSI300, Treasury Bond ETF; Case 2: CSI300, SME Mkt and Treasury Bond ETF; Case 3: CSI300, SME Mkt, CSI500 and Treasury Bond ETF; Case 4: only CSI300, SME Mkt and CSI500.

For case 1, results as shown in Table 15 and Table 16, we find the largest sharpe ratio for the single asset is that of CSI300, 0.8249, however, in the performance shown in Table 15, nearly all the achieved sharpe ratios are larger than that, in this ETF case, Equal Risk Contribution has the best performance with the sharpe ratio greater than 1.33 regardless time windows, Risk Parity Volatility ranks second, and Risk Parity Variance follows with sharpe ratio greater than 0.8. Minimum Variance doesn't perform that well as in the indices case, with sharpe ratios ranging from 0.7 to 0.8. Mean Variance Weights and Kelly Rule have a relatively better performance compared that to indices case. And we can find Mean Variance Weights have the best performance when we choose a long time window, like 10, 11 and 12 months. In Table 17, detailed holding positions are reported for different asset allocation methods.

If we incorporate SME Mkt ETF into the investment universe, as shown in Table 18, we find the average performance of each method improves, this is due to the benefit of diversification. The relative ranking of different methods don't change. Similar results for case three are shown in Table 20 and 21.

Finally, if we just consider the pure equity portfolios, namely, we just consider

the equity ETFs, Table 22 shows that Risk Parities, including Risk Parity Volatility, Risk Parity Variance and Equal Risk Contribution are not the best ones, this time, Minimum Variance ranks first, and Mean Variance Weights and Kelly Rule performs relatively better, which is just like the equity indices case.

4.4.6 Evidence of Simulation Study

From the previous sectionsâĂŹ discussion, we find in average, Mean Variance Weights will perform worse than other asset allocation methods, this is due to the following reasons: Mean Variance Weights as an unrestricted mean-variance portfolio, it contains too many parameters to estimate, which results in the instability of the portfolio performance even for a small error for each of the parameter estimation. Unlike other asset allocation methods, Minimum Variance, Risk Parity, and even Equal Weights, are just a special case of the full unconstrained mean-variance portfolio, for example, Minimum Weights assumes the means of the asset returns are equal, Risk Parity assumes the returns have the equal means, and the correlations among asset classes are equal to 0, Equal Weights requires the means, volatilities and even correlations of asset returns are equal. Just with these restrictions, we conclude that special case of mean-variance perform better than the full mean-variance procedure, because fewer things can go wrong with estimations of the inputs, which results the better performance of these asset allocation methods. As for the parameters estimation, asset expected returns are far more difficult to estimate than the volatilities.

In this section, we try to perform a simulation study to compare the performance of Equal Risk Contribution and Mean Variance Weights, intuitively, we attribute the bad performance of Mean Variance Weights by the inaccurate estimation of asset returns, in this simulation study, we will simulate the relationship between return estimation and trading performance. For the implementation of Equal Risk Contribution, only expected covariance matrix needs to estimate, we will use the historical sample covariance matrix to approximate, time window of 1 to 12 months are included in the study. As for Mean Variance Weights, besides the covariance, we still need to estimate the asset returns, here, we divide our study into three cases, case 1: just usd the exact returns μ as the expected returns in the weights calculation; case 2: generate returns from a normal distribution with mean of the

exact returns μ and standard deviation of $0.1 * \sigma$, where σ is the real volatility; case 3: generate returns from a normal distribution with mean of the exact returns μ and standard deviation of $0.1 * \sigma$. For each case, we perform a 1000 round simulations, in each round, we generate 500 trading months data, including CSI300, SME Mkt, CSI500 and Treasury Bond returns, based on their sample average returns and sample covariance matrix, normal distribution is used in this study, after the data generation, we will try to perform the paper trading with Equal Risk Contribution and Mean Variance Weights, finally record the sharpe ratios. After all the simulation, we take the average sharpe ratio and compare.

From Table 23, if we consider case 1, we will find Equal Risk Contribution behaves worse than Mean Variance Weights, that means, if we can estimate the asset returns accurately, Mean Variance Weights in theory can outperform risk parity, as we extend the time window to estimate the covariance matrix, the sharpe ratios increase, that means, within 1 year horizon, the more data used, the more accurate the covariance matrix estimation is. As we consider case 2 and case 3, we will find, if we add some noise to the return estimation, Equal Risk Contribution will outperform the Mean Variance Weights, this is just the reason why Mean Variance Weights performs bad in practice.

4.5 Concluion

In our study, we conduct an analysis of traditional asset allocation methods applied in Chinese market, through a comprehensive horserace comparison among different asset allocation methods on two data sets, asset indices and asset ETFs, we get a clear picture about the relative performance rank of different methods, as well as how the trading performance changes as we use different parameter estimation time windows and different investment universe. Finally, we perform a simulation to study why Mean Variance Weights perform worse than others in practice, and we attribute the bad performance by the bad estimation of asset returns.

Table 4.1: Index Data Description

This table describes the sources and inception dates of the total return indices used in the paper, the five total return indices are two benchmark asset classes – Chinese Equity and Chinese Bond

Asset class	Description	Source	Inception date	End date
Chinese Equity	CSI 300 Total Return Index	Wind	04/03/2006	03/23/2017
Chinese Equity	CSI 500 Total Return Index	Wind	01/15/2007	03/23/2017
Chinese Equity	SME Prime Mkt Total Return Index	Wind	12/30/2005	03/23/2017
Chinese Bond	CSI Aggregate Bond Index	Wind	12/31/2002	03/23/2017
Chinese Bond	CSI Treasury Bond Index	Wind	12/31/2002	03/23/2017

Table 4.2: Index Data: Summary Statistics on Daily Returns

This table reports the summary statistics of the five daily data series used in this paper. See Table 1 for detailed descriptions of each data series

Panel A: Mean, standard deviation and Sharpe ratio, (annualized percentage)

January 2007 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
Mean	10.63	20.49	21.20	4.52	4.25
StdDev	30.37	34.58	32.56	1.84	2.28
Sharpe	29.13	54.07	59.61	148.08	108.08

January 2010 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
Mean	4.49	9.64	11.68	4.83	4.39
StdDev	25.03	28.94	28.55	1.57	2.07
Sharpe	10.22	26.61	34.14	184.81	118.27

Panel B: Correlation matrix

January 2007 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
CSI300	1.000				
CSI500	0.892	1.000			
SME Mkt	0.849	0.954	1.000		
Agg Bond	-0.013	-0.022	-0.011	1.000	
Treas Bond	-0.025	-0.037	-0.031	0.863	1.000

January 2010 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
CSI300	1.000				
CSI500	0.860	1.000			
SME Mkt	0.813	0.979	1.000		
Agg Bond	0.007	-0.002	0.004	1.000	
Treas Bond	-0.018	-0.032	-0.024	0.842	1.000

Table 4.3: Index Data: Summary Statistics on Monthly Returns

This table reports the summary statistics of the five monthly data series used in this paper. See Table 1 for detailed descriptions of each data series

Panel A: Mean, standard deviation and Sharpe ratio, (annualized percentage)

January 2007 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
Mean	11.04	20.23	20.05	4.41	4.17
StdDev	33.22	37.20	35.21	3.15	3.58
Sharpe	26.11	48.02	50.23	64.99	50.45

January 2010 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
Mean	4.29	10.39	12.71	4.68	4.26
StdDev	26.23	29.43	30.84	2.67	3.02
Sharpe	6.47	26.50	32.80	78.44	55.37

Panel B: Correlation matrix

January 2007 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
CSI300	1.000				
CSI500	0.869	1.000			
SME Mkt	0.795	0.946	1.000		
Agg Bond	-0.221	-0.191	-0.176	1.000	
Treas Bond	-0.215	-0.2	-0.199	0.963	1.000

January 2010 – December 2016

	CSI300	CSI500	SME Mkt	Agg Bond	Treas Bond
CSI300	1.000				
CSI500	0.820	1.000			
SME Mkt	0.704	0.966	1.000		
Agg Bond	-0.034	-0.058	-0.054	1.000	
Treas Bond	-0.032	-0.088	-0.097	0.942	1.000

Table 4.4: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300 and Agg Bond

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Annualized return/volatility/sharpe from different time window(m) training				
Strategies	3	6	12	24	36
Equal Weights	4.48/13.14/14.43	4.48/13.14/14.43	4.48/13.14/14.43	4.48/13.14/14.43	4.48/13.14/14.43
Risk Parity Volatility	4.63/2.88/70.78	4.78/2.87/76.4	4.71/2.88/73.68	4.7/2.94/71.8	4.68/2.92/71.51
Risk Parity Variance	4.67/2.65/78.6	4.69/2.65/79.29	4.68/2.65/79.14	4.69/2.65/79.15	4.68/2.65/78.93
Equal Risk Contribution	4.46/2.74/68.39	4.85/2.73/82.95	4.21/2.64/61.41	4.52/2.74/70.61	4.54/2.65/73.81
Proportional to Sharpe Ratio	5.67/10.54/29.26	3.14/10.12/5.46	0.86/8.45/-20.45	4.46/7.63/24.59	2.38/5.81/-3.6
Minimum Variance	4.56/2.66/73.99	4.65/2.65/77.6	4.68/2.67/78.33	4.64/2.66/77.13	4.68/2.65/78.96
Mean Variance Weights	3.63/10.49/9.96	2.35/10.54/-2.26	1.06/8.54/-17.91	2.83/3.72/6.53	4.41/2.65/68.82
Kelly Rule	6.75/19.96/20.86	3.96/19.04/7.18	2.61/18.82/0.12	10.7/20.39/39.8	1.31/19.37/-6.59

Table 4.5: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300 and Agg Bond Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.044838	0.131400	0.039674	0.144253
Risk Parity Volatility	0.047789	0.028674	0.210310	0.763961
Risk Parity Variance	0.046867	0.026465	0.218228	0.792904
Equal Risk Contribution	0.048536	0.027310	0.228377	0.829471
Proportional to Sharpe Ratio	0.056724	0.105417	0.080558	0.292563
Minimum Variance	0.046833	0.026532	0.217320	0.789612
Mean Variance Weights	0.044098	0.026466	0.189215	0.688215
Kelly Rule	0.107039	0.203922	0.107924	0.397976

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.500000	0.500000	0.000000	0.000000
Risk Parity Volatility	0.940639	0.059361	0.000000	0.000000
Risk Parity Variance	0.995417	0.004583	0.000000	0.000000
Equal Risk Contribution	0.973433	0.026567	0.000000	0.000000
Proportional to Sharpe Ratio	0.661496	0.338504	0.000000	0.000000
Minimum Variance	0.994931	0.005069	0.000000	0.000000
Mean Variance Weights	0.990063	0.009937	0.000000	0.000000
Kelly Rule	0.537250	0.462750	0.000000	0.000000

Table 4.6: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, Agg Bond and SME Mkt

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Annualized return/volatility/sharpe from different time window(m) training				
Strategies	3	6	12	24	36
Equal Weights	7.16/17.55/26.05	7.16/17.55/26.05	7.16/17.55/26.05	7.16/17.55/26.05	7.16/17.55/26.05
Risk Parity Volatility	4.89/3.39/67.95	5.16/3.46/74.17	5.07/3.51/70.64	5.19/3.7/70.24	5.11/3.65/69.14
Risk Parity Variance	4.67/2.63/79.38	4.71/2.63/80.86	4.7/2.63/80.51	4.73/2.64/81.11	4.71/2.64/80.55
Equal Risk Contribution	4.61/2.95/68.64	4.8/3.02/73.3	4.86/3.03/74.91	5.07/3.15/78.73	4.99/3.0/80.02
Proportional to Sharpe Ratio	7.61/14.94/33.62	4.27/13.36/12.58	3.7/12.36/8.97	6.33/12.11/30.85	1.95/10.49/-6.1
Minimum Variance	4.53/2.66/72.72	4.63/2.64/77.3	4.67/2.66/78.24	4.66/2.66/77.78	4.7/2.65/79.99
Mean Variance Weights	6.0/13.2/25.87	3.19/12.03/5.01	3.79/8.81/13.61	3.71/5.14/21.92	4.21/2.66/61.09
Kelly Rule	3.3/24.73/2.86	2.2/23.02/-1.69	2.19/23.01/-1.73	8.93/26.22/24.2	2.86/25.86/1.06

Table 4.7: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, Agg Bond and SME Mkt Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.071610	0.175521	0.071509	0.260525
Risk Parity Volatility	0.051551	0.034607	0.204272	0.741697
Risk Parity Variance	0.047271	0.026368	0.223273	0.811149
Equal Risk Contribution	0.049925	0.030046	0.220353	0.800176
Proportional to Sharpe Ratio	0.076102	0.149387	0.092151	0.336169
Minimum Variance	0.047048	0.026460	0.220164	0.799899
Mean Variance Weights	0.042104	0.026552	0.167757	0.610930
Kelly Rule	0.089334	0.262201	0.066047	0.241996

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.333333	0.333333	0.333333	0.000000
Risk Parity Volatility	0.894336	0.056057	0.049607	0.000000
Risk Parity Variance	0.991760	0.004506	0.003734	0.000000
Equal Risk Contribution	0.927237	0.037885	0.034878	0.000000
Proportional to Sharpe Ratio	0.558288	0.199314	0.242398	0.000000
Minimum Variance	0.994748	0.002995	0.002258	0.000000
Mean Variance Weights	0.981173	0.001202	0.017626	0.000000
Kelly Rule	0.334439	0.030702	0.634859	0.000000

Table 4.8: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, Agg Bond, SME Mkt and CSI500

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Annualized return/volatility/sharpe from different time window(m) training				
Strategies	3	6	12	24	36
Equal Weights	7.96/20.39/26.35	7.96/20.39/26.35	7.96/20.39/26.35	7.96/20.39/26.35	7.96/20.39/26.35
Risk Parity Volatility	5.14/4.26/59.82	5.48/4.38/65.97	5.35/4.47/61.73	5.51/4.74/61.64	5.4/4.66/60.4
Risk Parity Variance	4.68/2.61/80.24	4.74/2.62/82.38	4.73/2.62/81.73	4.76/2.63/82.62	4.74/2.63/81.76
Equal Risk Contribution	4.7/3.3/63.88	5.29/3.32/81.29	5.19/3.33/78.21	5.26/3.56/75.25	5.14/3.5/73.06
Proportional to Sharpe Ratio	8.19/16.43/34.12	4.44/14.96/12.35	4.98/14.27/16.75	7.2/14.53/31.71	1.97/13.23/-4.69
Minimum Variance	4.52/2.67/72.62	4.63/2.64/77.14	4.67/2.66/78.34	4.66/2.66/77.73	4.7/2.65/79.99
Mean Variance Weights	5.93/13.08/25.54	2.65/12.3/0.48	4.64/8.97/22.91	4.44/4.75/39.04	4.27/2.67/63.12
Kelly Rule	3.49/23.86/3.8	-0.75/21.37/-15.61	5.0/23.3/10.34	8.96/24.68/25.8	3.77/24.56/4.82

Table 4.9: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, Agg Bond, SME Mkt and CSI500 Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.079600	0.203890	0.072140	0.263460
Risk Parity Volatility	0.054792	0.043821	0.181684	0.659714
Risk Parity Variance	0.047608	0.026296	0.227425	0.826167
Equal Risk Contribution	0.052864	0.033192	0.223883	0.812890
Proportional to Sharpe Ratio	0.081947	0.164328	0.093348	0.341173
Minimum Variance	0.047044	0.026455	0.220164	0.799899
Mean Variance Weights	0.042715	0.026666	0.173405	0.631223
Kelly Rule	0.089569	0.246835	0.070412	0.258009

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.250000	0.250000	0.250000	0.250000
Risk Parity Volatility	0.853091	0.053148	0.046998	0.046763
Risk Parity Variance	0.988214	0.004488	0.003719	0.003579
Equal Risk Contribution	0.926218	0.027393	0.023494	0.022895
Proportional to Sharpe Ratio	0.515725	0.150014	0.183823	0.150438
Minimum Variance	0.994782	0.002782	0.001781	0.000655
Mean Variance Weights	0.980942	0.000347	0.013399	0.005312
Kelly Rule	0.334439	0.024355	0.437187	0.204018

Table 4.10: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt and CSI500

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Annualized return/volatility/sharpe from different time window(m) training				
Strategies	3	6	12	24	36
Equal Weights	9.07/27.22/23.83	9.07/27.22/23.83	9.07/27.22/23.83	9.07/27.22/23.83	9.07/27.22/23.83
Risk Parity Volatility	9.17/27.11/24.26	9.3/27.2/24.7	9.25/27.18/24.52	9.03/27.11/23.75	9.03/27.12/23.75
Risk Parity Variance	9.24/27.02/24.64	9.53/27.18/25.53	9.44/27.15/25.22	8.98/27.01/23.67	8.98/27.03/23.66
Equal Risk Contribution	9.05/27.04/23.88	9.18/27.11/24.31	9.14/27.1/24.19	8.96/27.03/23.58	8.97/27.05/23.6
Proportional to Sharpe Ratio	9.37/27.93/24.29	9.12/27.79/23.49	8.02/28.03/19.38	9.57/27.58/25.31	9.74/27.89/25.66
Minimum Variance	8.5/26.01/22.71	9.19/26.44/24.97	9.21/26.64/24.86	7.29/26.02/18.08	8.0/26.43/20.46
Mean Variance Weights	10.15/28.93/26.14	8.18/28.46/19.66	8.75/27.43/22.48	9.42/28.42/24.04	10.53/29.21/27.18
Kelly Rule	11.13/29.84/28.63	3.94/27.48/4.91	7.28/27.38/17.12	9.43/28.64/23.89	9.91/29.18/25.08

Table 4.11: Index Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt and CSI500 Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.090736	0.272170	0.065002	0.238283
Risk Parity Volatility	0.093046	0.271964	0.067313	0.246955
Risk Parity Variance	0.095277	0.271799	0.069536	0.255316
Equal Risk Contribution	0.091787	0.271128	0.066284	0.243074
Proportional to Sharpe Ratio	0.097428	0.278851	0.069824	0.256573
Minimum Variance	0.091892	0.264396	0.068078	0.249661
Mean Variance Weights	0.105281	0.292143	0.073749	0.271777
Kelly Rule	0.111314	0.298364	0.077523	0.286330

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from January 2010 through December 2016. See Table 1 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.000000	0.333333	0.333333	0.333333
Risk Parity Volatility	0.000000	0.363952	0.318982	0.317066
Risk Parity Variance	0.000000	0.396090	0.303756	0.300154
Equal Risk Contribution	0.000000	0.374837	0.316506	0.308657
Proportional to Sharpe Ratio	0.000000	0.197334	0.455082	0.347585
Minimum Variance	0.000000	0.809514	0.169003	0.021483
Mean Variance Weights	0.000000	0.014642	0.820908	0.164450
Kelly Rule	0.000000	0.361087	0.419722	0.219191

Table 4.12: ETF Data Description

This table describes the sources and inception dates of the ETF data used in the paper

Asset class	Description	Source	Inception date	End date
Chinese Equity	Huataiborui Shangzheng Zhongxiaopan ETF (510220.OF)	Wind	01/26/2011	09/14/2017
Chinese Equity	Huataiborui Hushen 300 ETF (510300.OF)	Wind	05/04/2012	09/14/2017
Chinese Equity	Nanfangzhongzheng 500 ETF (510500.OF)	Wind	02/06/2013	09/14/2017
Chinese Bond	Guotai Shangzheng 5Nianqi Guozhai ETF (511010.OF)	Wind	03/05/2013	09/14/2017

Table 4.13: ETF Data: Summary statistics on daily returns

This table reports the summary statistics of the four daily ETF data series used in this paper. See Table 17 for detailed descriptions of each data series

Panel A: Mean, standard deviation and Sharpe ratio, (annualized percentage)

May 2013 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
Mean	17.08	23.55	19.53	2.44
StdDev	25.12	29.60	28.66	2.14
Sharpe	59.60	72.45	60.80	15.69

April 2014 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
Mean	25.56	24.87	23.51	3.78
StdDev	26.15	31.15	30.35	2.00
Sharpe	90.21	73.53	70.96	90.45

Panel B: Correlation matrix

May 2013 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
CSI300	1.000			
CSI500	0.835	1.000		
SME Mkt	0.918	0.970	1.000	
Treas Bond	0.039	0.041	0.040	1.000

April 2014 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
CSI300	1.000			
CSI500	0.836	1.000		
SME Mkt	0.919	0.969	1.000	
Treas Bond	0.041	0.036	0.040	1.000

Table 4.14: ETF Data: Summary statistics on monthly returns

This table reports the summary statistics of the four daily ETF data series used in this paper. See Table 17 for detailed descriptions of each data series

Panel A: Mean, standard deviation and Sharpe ratio, (annualized percentage)

May 2013 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
Mean	16.50	22.89	18.60	2.44
StdDev	26.49	30.18	28.15	2.71
Sharpe	54.33	68.87	58.59	12.27

April 2014 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
Mean	24.82	23.96	22.27	3.78
StdDev	27.70	31.26	29.29	2.57
Sharpe	82.49	70.37	69.32	70.50

Panel B: Correlation matrix

May 2013 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
CSI300	1.000			
CSI500	0.785	1.000		
SME Mkt	0.926	0.951	1.000	
Treas Bond	0.022	0.098	0.071	1.000

April 2014 – August 2017

	CSI300	CSI500	SME Mkt	Treas Bond
CSI300	1.000			
CSI500	0.773	1.000		
SME Mkt	0.923	0.948	1.000	
Treas Bond	0.016	0.125	0.109	1.000

Table 4.15: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300 and Treas Bond ETFs

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

Strategies	Annualized return/volatility/sharpe from different time window(m) training					
	1	2	3	4	5	6
	7	8	9	10	11	12
E-W	13.89/13.76/0.86 13.89/13.76/0.86	13.89/13.76/0.86 13.89/13.76/0.86	13.89/13.76/0.86 13.89/13.76/0.86	13.89/13.76/0.86 13.89/13.76/0.86	13.89/13.76/0.86 13.89/13.76/0.86	13.89/13.76/0.86 13.89/13.76/0.86
R-P-Vo	7.56/3.91/1.41 6.7/3.36/1.39	6.85/3.73/1.29 6.52/3.34/1.34	6.41/3.52/1.24 6.3/3.32/1.29	6.41/3.41/1.28 6.28/3.35/1.26	6.64/3.39/1.36 6.26/3.36/1.26	6.68/3.38/1.37 6.21/3.38/1.23
R-P-Va	4.63/2.55/1.01 4.23/2.49/0.88	4.36/2.54/0.91 4.19/2.49/0.86	4.23/2.5/0.87 4.13/2.5/0.83	4.19/2.5/0.86 4.12/2.51/0.83	4.22/2.5/0.87 4.11/2.51/0.82	4.23/2.5/0.87 4.09/2.52/0.81
E-R-C	7.61/3.84/1.45 7.24/3.14/1.65	6.71/3.39/1.38 6.79/3.16/1.5	6.5/3.23/1.38 6.63/3.13/1.46	6.76/2.99/1.58 6.46/3.2/1.38	7.41/3.16/1.7 6.95/3.07/1.6	7.19/3.19/1.61 6.37/3.25/1.33
P-to-SR	15.29/14.53/0.91 12.96/11.93/0.92	16.76/12.96/1.14 12.11/11.79/0.85	14.75/12.79/0.99 12.14/11.65/0.87	13.4/14.51/0.78 10.3/11.34/0.73	14.89/13.26/0.97 10.29/10.45/0.79	14.27/12.33/0.99 11.1/9.95/0.91
M-V	3.92/2.58/0.73 4.0/2.54/0.77	3.94/2.59/0.73 3.96/2.54/0.76	3.84/2.53/0.71 3.98/2.55/0.76	3.93/2.54/0.74 4.04/2.52/0.79	3.96/2.52/0.76 4.0/2.53/0.77	3.96/2.54/0.75 4.0/2.53/0.77
M-V-W	11.4/12.86/0.73 9.3/4.86/1.49	13.39/11.14/1.02 7.73/5.93/0.96	8.6/7.67/0.85 9.37/4.64/1.58	7.53/8.11/0.68 9.66/4.65/1.64	9.93/6.1/1.29 9.05/4.36/1.61	10.46/5.36/1.57 11.29/6.48/1.43
K-R	9.18/7.31/0.98 6.67/3.9/1.19	6.5/5.23/0.85 6.25/3.44/1.22	6.18/4.39/0.94 5.93/3.34/1.16	6.43/3.91/1.12 5.52/3.26/1.07	7.44/4.31/1.25 5.26/3.04/1.06	7.16/4.24/1.21 4.91/2.96/0.97

Table 4.16: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300 and Treas Bond ETFs Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.138933	0.137600	0.231545	0.861205
Risk Parity Volatility	0.075638	0.039053	0.389579	1.413608
Risk Parity Variance	0.046318	0.025507	0.282165	1.014896
Equal Risk Contribution	0.074068	0.031631	0.467597	1.695728
Proportional to Sharpe Ratio	0.167582	0.129563	0.301919	1.135739
Minimum Variance	0.040449	0.025193	0.220998	0.794579
Mean Variance Weights	0.096570	0.046511	0.447561	1.636998
Kelly Rule	0.074399	0.043056	0.345591	1.253429

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from October 2013 through August 2017. See Table 17 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.500000	0.500000	0.0	0.0
Risk Parity Volatility	0.902753	0.097247	0.0	0.0
Risk Parity Variance	0.984582	0.015418	0.0	0.0
Equal Risk Contribution	0.922539	0.077461	0.0	0.0
Proportional to Sharpe Ratio	0.544286	0.455714	0.0	0.0
Minimum Variance	0.993221	0.006779	0.0	0.0
Mean Variance Weights	0.700981	0.299019	0.0	0.0
Kelly Rule	0.921302	0.078698	0.0	0.0

Table 4.17: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt and Treas Bond ETFs

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

Strategies	Annualized return/volatility/sharpe from different time window(m) training					
	1	2	3	4	5	6
	7	8	9	10	11	12
E-W	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79
	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79	16.72/18.48/0.79
R-P-Vo	9.72/5.33/1.44	8.76/5.09/1.32	8.21/4.85/1.27	8.31/4.77/1.31	8.69/4.79/1.39	8.78/4.76/1.42
	8.81/4.74/1.43	8.54/4.77/1.36	8.23/4.76/1.3	8.2/4.82/1.28	8.17/4.84/1.27	8.08/4.84/1.25
R-P-Va	5.21/2.67/1.18	4.79/2.62/1.05	4.61/2.54/1.01	4.58/2.53/1.0	4.63/2.54/1.02	4.65/2.54/1.03
	4.65/2.52/1.03	4.59/2.51/1.02	4.5/2.52/0.98	4.5/2.53/0.97	4.48/2.53/0.96	4.44/2.55/0.94
E-R-C	9.04/4.28/1.63	8.27/4.02/1.55	7.22/3.62/1.43	7.86/3.65/1.59	8.01/3.73/1.6	8.43/3.63/1.76
	8.02/3.7/1.62	7.81/3.66/1.57	7.07/3.75/1.34	7.08/3.75/1.34	7.05/3.71/1.35	6.88/3.74/1.29
P-to-SR	19.01/14.95/1.13	21.41/14.39/1.35	18.12/16.48/0.98	15.96/17.45/0.8	16.86/16.05/0.92	14.73/15.64/0.81
	14.69/15.02/0.84	15.19/15.08/0.87	13.98/15.09/0.79	11.69/15.4/0.63	13.46/15.3/0.75	15.11/14.93/0.88
M-V	4.05/2.58/0.78	3.95/2.6/0.73	3.86/2.54/0.71	3.99/2.56/0.76	3.99/2.54/0.76	4.02/2.56/0.77
	4.04/2.57/0.78	3.98/2.56/0.76	3.97/2.56/0.75	3.98/2.53/0.77	3.94/2.54/0.75	3.95/2.54/0.75
M-V-W	8.85/13.59/0.5	13.83/13.58/0.87	11.63/9.51/1.01	8.55/10.15/0.64	7.37/7.42/0.72	8.6/6.81/0.96
	8.71/5.37/1.24	8.4/5.69/1.12	9.59/5.13/1.47	10.44/5.16/1.63	11.32/5.4/1.72	11.95/6.43/1.54
K-R	10.61/7.78/1.1	8.47/5.95/1.08	7.39/4.54/1.18	7.05/4.17/1.2	7.39/4.03/1.33	7.06/3.91/1.28
	7.05/3.77/1.33	6.93/3.61/1.35	6.57/3.62/1.25	6.29/3.68/1.15	6.05/3.52/1.14	5.72/3.39/1.09

Table 4.18: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt and Treas Bond ETFs Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.167225	0.184767	0.211230	0.794478
Risk Parity Volatility	0.097192	0.053293	0.393698	1.440338
Risk Parity Variance	0.052055	0.026723	0.328631	1.183399
Equal Risk Contribution	0.084283	0.036296	0.483251	1.759182
Proportional to Sharpe Ratio	0.214094	0.143888	0.351381	1.345926
Minimum Variance	0.040354	0.025677	0.215796	0.775879
Mean Variance Weights	0.113224	0.053969	0.467007	1.719353
Kelly Rule	0.069343	0.036127	0.373967	1.353890

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from October 2013 through August 2017. See Table 17 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.333333	0.333333	0.333333	0.0
Risk Parity Volatility	0.832996	0.087934	0.079070	0.0
Risk Parity Variance	0.972689	0.015030	0.012281	0.0
Equal Risk Contribution	0.892693	0.055188	0.052119	0.0
Proportional to Sharpe Ratio	0.473373	0.288004	0.238623	0.0
Minimum Variance	0.992315	0.004334	0.003351	0.0
Mean Variance Weights	0.687618	0.205773	0.106609	0.0
Kelly Rule	0.931942	0.034485	0.033573	0.0

Table 4.19: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt, CSI500 and Treas Bond ETFs

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

Strategies	Annualized return/volatility/sharpe from different time window(m) training					
	1	2	3	4	5	6
	7	8	9	10	11	12
E-W	18.58/20.98/0.79 18.58/20.98/0.79	18.58/20.98/0.79 18.58/20.98/0.79	18.58/20.98/0.79 18.58/20.98/0.79	18.58/20.98/0.79 18.58/20.98/0.79	18.58/20.98/0.79 18.58/20.98/0.79	18.58/20.98/0.79 18.58/20.98/0.79
R-P-Vo	10.93/6.26/1.42 10.37/5.95/1.4	10.07/6.06/1.32 10.06/6.05/1.32	9.52/5.92/1.26 9.7/6.08/1.26	9.72/5.92/1.3 9.67/6.16/1.24	10.19/5.98/1.36 9.65/6.19/1.23	10.34/5.95/1.39 9.55/6.18/1.22
R-P-Va	5.47/2.71/1.26 4.98/2.57/1.15	5.06/2.68/1.12 4.92/2.55/1.13	4.87/2.57/1.1 4.81/2.56/1.08	4.87/2.56/1.11 4.81/2.57/1.08	4.94/2.58/1.12 4.79/2.57/1.07	4.98/2.59/1.14 4.74/2.59/1.04
E-R-C	10.22/4.14/1.97 8.64/4.03/1.64	9.21/4.09/1.75 8.47/4.04/1.59	7.89/4.24/1.38 7.84/4.13/1.4	7.86/4.21/1.38 7.49/4.21/1.29	8.64/4.07/1.62 7.99/4.24/1.4	8.6/4.07/1.61 7.94/4.21/1.4
P-to-SR	20.13/16.42/1.1 15.88/16.48/0.84	21.16/16.12/1.19 18.52/16.7/0.99	18.4/18.7/0.87 17.13/16.87/0.89	15.93/19.11/0.73 14.15/18.07/0.67	16.23/17.31/0.82 16.81/18.5/0.8	13.41/17.56/0.65 17.56/17.93/0.87
M-V	3.95/2.61/0.73 3.99/2.58/0.75	3.97/2.61/0.74 3.94/2.57/0.74	3.86/2.55/0.71 3.89/2.56/0.72	3.99/2.58/0.75 3.88/2.54/0.72	3.97/2.56/0.75 3.88/2.55/0.72	4.02/2.57/0.77 3.93/2.55/0.74
M-V-W	7.23/13.93/0.37 8.14/5.69/1.07	12.17/13.73/0.74 8.48/5.92/1.09	12.19/11.17/0.91 9.78/5.29/1.46	8.3/11.44/0.55 10.68/5.51/1.57	5.61/8.6/0.41 12.41/5.93/1.75	7.26/7.97/0.65 12.98/6.91/1.58
K-R	11.34/7.94/1.17 6.78/3.73/1.27	8.66/6.31/1.05 6.77/3.65/1.29	7.68/5.09/1.11 6.51/3.65/1.22	7.44/4.64/1.16 6.15/3.73/1.1	7.01/4.07/1.22 5.65/3.59/1.0	6.71/3.87/1.2 5.37/3.49/0.95

Table 4.20: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt, CSI500 and Treas Bond ETFs Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.185761	0.209821	0.207980	0.787954
Risk Parity Volatility	0.109303	0.062642	0.385949	1.418726
Risk Parity Variance	0.054708	0.027133	0.350576	1.263289
Equal Risk Contribution	0.102202	0.041439	0.538317	1.973293
Proportional to Sharpe Ratio	0.211621	0.161157	0.310017	1.186355
Minimum Variance	0.040171	0.025699	0.213631	0.768103
Mean Variance Weights	0.124117	0.059278	0.473045	1.749120
Kelly Rule	0.067685	0.036502	0.357781	1.294542

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from October 2013 through August 2017. See Table 17 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.250000	0.250000	0.250000	0.250000
Risk Parity Volatility	0.780580	0.081290	0.072958	0.065172
Risk Parity Variance	0.963751	0.014758	0.012047	0.009443
Equal Risk Contribution	0.865968	0.052542	0.042571	0.038920
Proportional to Sharpe Ratio	0.419010	0.226696	0.176465	0.177829
Minimum Variance	0.991772	0.004166	0.000397	0.003664
Mean Variance Weights	0.675886	0.196124	0.014344	0.113646
Kelly Rule	0.928782	0.034163	0.021857	0.015198

Table 4.21: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt and CSI500 ETFs

This table reports the annualized return, volatility and sharpe ratio of monthly trading by each strategy with different time window for stragegy training. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

Strategies	Annualized return/volatility/sharpe from different time window(m) training					
	1	2	3	4	5	6
	7	8	9	10	11	12
E-W	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79
	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79	23.96/27.89/0.79
R-P-Vo	24.17/27.81/0.8	24.51/27.89/0.81	24.65/27.87/0.81	24.79/27.85/0.82	24.63/27.86/0.81	24.62/27.87/0.81
	24.6/27.88/0.81	24.59/27.89/0.81	24.56/27.9/0.81	24.56/27.9/0.81	24.56/27.89/0.81	24.52/27.88/0.81
R-P-Va	24.4/27.76/0.81	25.06/27.9/0.82	25.32/27.85/0.84	25.6/27.81/0.85	25.27/27.82/0.83	25.27/27.86/0.83
	25.22/27.87/0.83	25.21/27.89/0.83	25.15/27.9/0.83	25.16/27.9/0.83	25.15/27.89/0.83	25.08/27.87/0.83
E-R-C	24.19/27.8/0.8	24.55/27.88/0.81	24.69/27.86/0.81	24.82/27.84/0.82	24.65/27.84/0.81	24.64/27.86/0.81
	24.62/27.86/0.81	24.62/27.87/0.81	24.58/27.88/0.81	24.58/27.88/0.81	24.57/27.87/0.81	24.53/27.86/0.81
P-to-SR	24.02/29.37/0.75	23.02/28.35/0.74	24.08/28.17/0.78	23.32/28.09/0.76	21.91/28.68/0.69	23.8/28.19/0.77
	25.91/28.17/0.85	25.56/28.93/0.81	26.27/29.09/0.83	25.08/28.29/0.81	23.97/28.06/0.78	24.75/28.15/0.81
M-V	26.97/27.83/0.9	30.14/28.03/1.0	32.22/27.42/1.1	32.33/27.59/1.1	30.61/27.36/1.04	31.71/27.71/1.07
	31.09/27.77/1.05	30.89/27.84/1.04	30.96/28.0/1.03	31.18/28.18/1.03	30.86/28.1/1.03	30.16/27.92/1.01
M-V-W	22.87/31.71/0.66	21.61/30.76/0.64	22.31/28.86/0.7	22.39/29.65/0.69	18.59/29.11/0.57	20.51/28.47/0.65
	26.82/28.68/0.86	25.23/29.47/0.79	26.33/29.69/0.82	24.08/29.55/0.75	18.96/28.15/0.6	17.78/28.26/0.56
K-R	27.86/29.29/0.88	29.66/28.18/0.98	30.84/27.32/1.05	30.64/27.4/1.04	29.1/27.2/0.99	28.91/27.08/0.99
	29.16/27.27/0.99	29.92/27.57/1.01	30.48/27.88/1.02	30.09/27.82/1.01	29.04/27.53/0.98	28.18/27.27/0.96

Table 4.22: ETF Data: Monthly Trading Performance of Each Strategy on Asset Classes: CSI300, SME Mkt, and CSI500 ETFs Cont

Panel A: This table reports the summary performance of monthly trading by each strategy if we choose the best time window parameter. Trading period extends from April 2014 through August 2017. See Table 12 for detailed descriptions of each data series

	Raw Return	Volatility	Sharpe(m)	Sharpe(y)
Equal Weights	0.239576	0.278889	0.203158	0.785776
Risk Parity Volatility	0.247907	0.278484	0.210525	0.816835
Risk Parity Variance	0.256036	0.278074	0.217704	0.847275
Equal Risk Contribution	0.248242	0.278358	0.210904	0.818409
Proportional to Sharpe Ratio	0.259122	0.281686	0.217476	0.847367
Minimum Variance	0.322190	0.274167	0.276005	1.100637
Mean Variance Weights	0.268195	0.286779	0.220982	0.863952
Kelly Rule	0.308389	0.273231	0.265608	1.053897

Panel B: This table reports the average weights of each strategy on each asset, when we choose the best time window parameter. Trading period extends from October 2013 through August 2017. See Table 17 for detailed descriptions of each data series

	Agg Bond	CSI300	SME Mkt	CSI500
Equal Weights	0.00000	0.333333	0.333333	0.333333
Risk Parity Volatility	0.00000	0.368494	0.329304	0.302201
Risk Parity Variance	0.00000	0.404926	0.321729	0.273345
Equal Risk Contribution	0.00000	0.376972	0.318943	0.304085
Proportional to Sharpe Ratio	0.00000	0.337467	0.251329	0.411204
Minimum Variance	0.00000	0.813264	0.072885	0.113851
Mean Variance Weights	0.00000	0.363522	0.191358	0.445120
Kelly Rule	0.00000	0.779808	0.122699	0.097493

Table 4.23: Simulation of 500 Monthly Trading Performance based on ETF Asset Classes: CSI300, SME Mkt, CSI500 and Treas Bond

This table reports the simulated annualized sharpe ratios of monthly trading by each strategy with different time window for parameter estimation. Trading period includes 500 trading months.

Strategies	1	2	3	4	5	6
	7	8	9	10	11	12
E-R-C	0.439	0.4582	0.4728	0.4848	0.4928	0.4933
μ	0.4958	0.5063	0.5137	0.5179	0.5253	0.5205
M-V-W	0.4528	0.5232	0.5611	0.5711	0.5715	0.5738
μ	0.5737	0.5745	0.5756	0.5777	0.5782	0.5778
E-R-C	0.4351	0.4569	0.4721	0.4831	0.494	0.4941
$\mu + 0.1 * N(0, \sigma^2)$	0.4952	0.5099	0.5186	0.5194	0.522	0.5208
M-V-W	0.4569	0.445	0.441	0.4661	0.4578	0.4683
$\mu + 0.1 * N(0, \sigma^2)$	0.4485	0.4553	0.4607	0.4653	0.4623	0.4277
E-R-C	0.4341	0.4535	0.4758	0.4824	0.4937	0.4958
$\mu + 0.2 * N(0, \sigma^2)$	0.496	0.5031	0.5132	0.5161	0.5252	0.5217
M-V-W	0.4397	0.4223	0.4246	0.44	0.4525	0.4307
$\mu + 0.2 * N(0, \sigma^2)$	0.4273	0.469	0.4424	0.4482	0.4495	0.457

Chapter 5
Future Research Directions

In Chapter 1, we used Markov Chain Monte Carlo (MCMC) to conduct parameter simulation, and compare the nested models under several criteria. Further more, we have derection which deserve further exploration:

1. From our statistic test, we show that in the 1931 trading days, VIX has significant jumps in 222 days, VVIX has significant jumps in 141 days, and they have significant co-jumps in 131 days, we can further consider a model where VIX and VVIX have different jump intensity, that is, we assume a model where $Y(t) = \log VIX(t)$ has a stronger jumper intensity than $\omega(t)$.

2. Since we believe that the volatility $\omega(t)$ and jump intensity $\lambda(t)$ are positively correlated, in our study, we assume they have a linear relationship, we can further explore some more different specifications, even though we can model the positive correlated property.

3. In our paper, we estimated the parameters and latent variables on the joint dynamics of logarithm of VIX and its volatility, we can also incorporate the estimation on the Greeks of the dynamic.

In Chapter 2, we introduced two high dimensional variable selection frameworks, and successfully constructed new macro-based predictors to forecast bond premia and excess ETF returns, which results to achieve a tremendous improvement in the forecast performance, both in sample and out-of-sample, but we still have some more directions which can be extended.

4. For the nonlinear approach, in the second step of nonliearization, we can further incorporate some other types of splines and construct new kinds of nonlinear trend of the original features.

5. Originally, DCSIS and SIS screening techniques are designed for the variable

selection of the i.i.d. data, we applied those to the area of excess return prediction, which has the autocorrelation, a further research direction is trying to develop some high-dimensional variable screening methods which can directly handle time series data.

Bibliography

[1] Yacine Ait-Sahalia, Mustafa Karaman, and Loriano Mancini. "The term structure of variance swaps and risk premia". In: *Available at SSRN 2136820* (2014).

[2] Yacine Ait-Sahalia and Robert Kimmel. "Maximum likelihood estimation of stochastic volatility models". In: *Journal of Financial Economics* 83.2 (2007), pp. 413–452.

[3] Dante Amengual and Dacheng Xiu. "Delving into risk premia: reconciling evidence from the S&P 500 and VIX derivatives". In: *Work in progress* (2012).

[4] Dante Amengual and Dacheng Xiu. "Resolution of policy uncertainty and sudden declines in volatility". In: *Chicago Booth Research Paper* 13-78 (2014).

[5] Torben G Andersen, Tim Bollerslev, et al. "Modeling and forecasting realized volatility". In: *Econometrica* 71.2 (2003), pp. 579–625.

[6] Torben G Andersen and Timo Teräsvirta. "Realized volatility". In: *Handbook of financial time series*. Springer, 2009, pp. 555–575.

[7] Andrew Ang. *Asset management: A systematic approach to factor investing*. Oxford University Press, 2014.

[8] Andrew Ang and Monika Piazzesi. "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables". In: *Journal of Monetary economics* 50.4 (2003), pp. 745–787.

[9] Ivan O Asensio. "The vix-vix futures puzzle". In: *Unpublished paper, University of Victoria* (2013).

[10] Jan Baldeaux and Alexander Badran. "Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model". In: *Applied Mathematical Finance* 21.4 (2014), pp. 299–312.

[11] Chris Bardgett, Elise Gourier, and Markus Leippold. "Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets". In: *Swiss Finance Institute Research Paper* 13-40 (2013).

[12] Ole E Barndorff-Nielsen and Neil Shephard. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.2 (2002), pp. 253–280.

[13] Ole E Barndorff-Nielsen and Neil Shephard. "Estimating quadratic variation using realized variance". In: *Journal of Applied econometrics* 17.5 (2002), pp. 457–477.

[14] Ole E Barndorff-Nielsen and Almut ED Veraart. "Stochastic volatility of volatility and variance risk premia". In: *Journal of Financial Econometrics* 11.1 (2013), pp. 1–46.

[15] Shlomo Benartzi and Richard H Thaler. "Naive diversification strategies in defined contribution saving plans". In: *American economic review* 91.1 (2001), pp. 79–98.

[16] Andreas Berg, Renate Meyer, and Jun Yu. "Deviance information criterion for comparing stochastic volatility models". In: *Journal of Business & Economic Statistics* 22.1 (2004), pp. 107–120.

[17] Carole Bernard and Zhenyu Cui. "Prices and asymptotics for discrete variance swaps". In: *Applied Mathematical Finance* 21.2 (2014), pp. 140–173.

[18] Julian Besag. "Spatial interaction and the statistical analysis of lattice systems". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1974), pp. 192–236.

[19] Tim Bollerslev, Michael Gibson, and Hao Zhou. "Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities". In: *Journal of econometrics* 160.1 (2011), pp. 235–245.

[20] Tim Bollerslev, Tzuo Hann Law, and George Tauchen. "Risk, jumps, and diversification". In: *Journal of Econometrics* 144.1 (2008), pp. 234–256.

[21] David G Booth and Eugene F Fama. "Diversification returns and asset contributions". In: *Financial Analysts Journal* 48.3 (1992), pp. 26–32.

[22] Nicole Branger and Clemens Voelkert. *The fine structure of variance: Consistent pricing of VIX derivatives*. Tech. rep. Working Paper, Westfaelische Wilhelms Universitaet Muenster, 2013.

[23] Mark Broadie and Ashish Jain. "Pricing and hedging volatility derivatives". In: *Journal of Derivatives* 15.3 (2008), p. 7.

[24] Mark Broadie and Ashish Jain. "The effect of jumps and discrete sampling on volatility and variance swaps". In: *International Journal of Theoretical and Applied Finance* 11.08 (2008), pp. 761–797.

[25] John Y Campbell and Robert J Shiller. "Yield spreads and interest rate movements: A bird's eye view". In: *The Review of Economic Studies* 58.3 (1991), pp. 495–514.

[26] John Y Campbell and Samuel B Thompson. "Predicting excess stock returns out of sample: Can anything beat the historical average?" In: *Review of Financial Studies* 21.4 (2008), pp. 1509–1531.

[27] Peter Carr, Helyette Geman, et al. "Options on realized variance and convex orders". In: *Quantitative Finance* 11.11 (2011), pp. 1685–1694.

[28] Peter Carr, Hélyette Geman, et al. "Pricing options on realized variance". In: *Finance and Stochastics* 9.4 (2005), pp. 453–475.

[29] Peter Carr and Roger Lee. "Realized volatility and variance: Options via swaps". In: *Risk* 20.5 (2007), pp. 76–83.

[30] Peter Carr and Roger Lee. "Robust replication of volatility derivatives". In: *Prmia award for best paper in derivatives, mfa 2008 annual meeting*. Citeseer. 2008.

[31] Peter Carr and Roger Lee. "Volatility derivatives". In: *Annu. Rev. Financ. Econ.* 1.1 (2009), pp. 319–339.

[32] Peter Carr, Roger Lee, and Liuren Wu. "Variance swaps on time-changed Lévy processes". In: *Finance and Stochastics* 16.2 (2012), pp. 335–355.

[33] Peter Carr and Dilip Madan. "Towards a theory of volatility trading". In: *Volatility: New estimation techniques for pricing derivatives* 29 (1998), pp. 417–427.

[34] Peter Carr and Dilip B Madan. "Joint modeling of VIX and SPX options at a single and common maturity with risk management applications". In: *IIE Transactions* 46.11 (2014), pp. 1125–1131.

[35] Peter Carr and Liuren Wu. "A tale of two indices". In: (2005).

[36] Denis Chaves et al. "Risk parity portfolio vs. other asset allocation heuristic portfolios". In: *Journal of Investing* 20.1 (2011), p. 108.

[37] Jun Cheng et al. "A remark on Lin and Chang's paper fffdfffdfffdConsistent modeling of S&P 500 and VIX derivativesfffdfffdfffd". In: *Journal of Economic Dynamics and Control* 36.5 (2012), pp. 708–715.

[38] Mikhail Chernov and Philippe Mueller. "The term structure of inflation expectations". In: *Journal of financial economics* 106.2 (2012), pp. 367–394.

[39] P ROBERT Christian and George Casella. *Monte Carlo statistical methods*. 1999.

[40] Peter Christoffersen, Kris Jacobs, and Karim Mimouni. "Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices". In: *The Review of Financial Studies* 23.8 (2010), pp. 3141–3189.

[41] San-Lin Chung et al. “The information content of the S&P 500 index and VIX options on the dynamics of the S&P 500 index”. In: *Journal of Futures Markets* 31.12 (2011), pp. 1170–1201.

[42] Anna Cieslak and Pavol Povala. “Understanding bond risk premia”. In: *Unpublished working paper. Kellogg School of Management, Evanston, IL* (2011), pp. 677–691.

[43] Todd E Clark and Michael W McCracken. “Tests of equal forecast accuracy and encompassing for nested models”. In: *Journal of econometrics* 105.1 (2001), pp. 85–110.

[44] Roger Clarke, Harindra De Silva, and Steven Thorley. “Minimum-variance portfolios in the US equity market”. In: *Journal of Portfolio Management* 33.1 (2006), p. 10.

[45] John H Cochrane and Monika Piazzesi. “Bond risk premia”. In: *The American economic review* 95.1 (2005), pp. 138–160.

[46] John H Cochrane and Monika Piazzesi. “Decomposing the yield curve”. In: *AFA 2010 Atlanta Meetings Paper*. 2009.

[47] Rama Cont and Thomas Kokholm. “A consistent pricing model for index options and volatility derivatives”. In: *Mathematical Finance* 23.2 (2013), pp. 248–274.

[48] Ilan Cooper and Richard Priestley. “Time-varying risk premiums and the output gap”. In: *Review of Financial Studies* 22.7 (2009), pp. 2801–2833.

[49] John Crosby and Mark Davis. “Variance derivatives: pricing and convergence”. In: (2012).

[50] Imma Valentina Curato et al. “Asymptotics for the Fourier estimators of the volatility of volatility and the leverage”. In: *DYNSTOCH 2013* (2012), p. 36.

[51] Kresimir Demeterfi et al. “More than you ever wanted to know about volatility swaps”. In: *Goldman Sachs quantitative strategies research notes* 41 (1999).

[52] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. “Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?” In: *The review of Financial studies* 22.5 (2007), pp. 1915–1953.

[53] Jérôme Detemple and Carlton Osakwe. “The valuation of volatility options”. In: *European Finance Review* 4.1 (2000), pp. 21–50.

[54] Luc Devroye. “Sample-based non-uniform random variate generation”. In: *Proceedings of the 18th conference on Winter simulation*. ACM. 1986, pp. 260–265.

[55] Giuseppe Di Graziano and Lorenzo Torricelli. "Target volatility option pricing". In: *International Journal of Theoretical and Applied Finance* 15.01 (2012), p. 1250005.

[56] Gabriel G Drimus. "Options on realized variance by transform methods: a non-affine stochastic volatility model". In: *Quantitative Finance* 12.11 (2012), pp. 1679–1694.

[57] Gabriel Drimus and Walter Farkas. "Local volatility of volatility for the VIX market". In: *Review of derivatives research* 16.3 (2013), pp. 267–293.

[58] Jian Du and Nikunj Kapadia. "Tail and volatility indices from option prices". In: *Unpublished working paper. University of Massachusetts, Amhurst* (2012).

[59] Jin-Chuan Duan and Chung-Ying Yeh. "Jump and volatility risk premiums implied by VIX". In: *Journal of Economic Dynamics and Control* 34.11 (2010), pp. 2232–2244.

[60] Gregory R Duffee. "Are variations in term premia related to the macroeconomy". In: *Unpublished working paper. University of California, Berkeley* (2007).

[61] Gregory R Duffee. "Information in (and not in) the term structure". In: *Review of Financial Studies* 24.9 (2011), pp. 2895–2934.

[62] Darrell Duffie and Rui Kan. "A yield-factor model of interest rates". In: *Mathematical finance* 6.4 (1996), pp. 379–406.

[63] Darrell Duffie, Jun Pan, and Kenneth Singleton. "Transform analysis and asset pricing for affine jump-diffusions". In: *Econometrica* 68.6 (2000), pp. 1343–1376.

[64] Brice Dupoyet, Robert T Daigler, and Zhiyao Chen. "A simplified pricing model for volatility futures". In: *Journal of Futures Markets* 31.4 (2011), pp. 307–339.

[65] Daniel Egloff, Markus Leippold, Liuren Wu, et al. "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments". In: *Journal of Financial and Quantitative Analysis* 45.05 (2010), pp. 1279–1310.

[66] Robert J Elliott, Tak Kuen Siu, and Leunglung Chan. "Pricing volatility swaps under Heston's stochastic volatility model with regime switching". In: *Applied Mathematical Finance* 14.1 (2007), pp. 41–62.

[67] Bjørn Eraker. "Do stock prices and volatility jump? Reconciling evidence from spot and option prices". In: *The Journal of Finance* 59.3 (2004), pp. 1367–1404.

[68] Bjørn Eraker, Michael Johannes, and Nicholas Polson. "The impact of jumps in volatility and returns". In: *The Journal of Finance* 58.3 (2003), pp. 1269–1300.

[69] Eugene F Fama and Robert R Bliss. "The information in long-maturity forward rates". In: *The American Economic Review* (1987), pp. 680–692.

[70] Jianqing Fan and Runze Li. "Variable selection via nonconcave penalized likelihood and its oracle properties". In: *Journal of the American statistical Association* 96.456 (2001), pp. 1348–1360.

[71] Jianqing Fan and Jinchi Lv. "A selective overview of variable selection in high dimensional feature space". In: *Statistica Sinica* 20.1 (2010), p. 101.

[72] Jianqing Fan and Jinchi Lv. "Sure independence screening for ultrahigh dimensional feature space". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.5 (2008), pp. 849–911.

[73] Damir Filipović, Elise Gourier, and Loriano Mancini. "Quadratic variance swap models". In: *Journal of Financial Economics* 119.1 (2016), pp. 44–68.

[74] LLdiko E Frank and Jerome H Friedman. "A statistical view of some chemometrics regression tools". In: *Technometrics* 35.2 (1993), pp. 109–135.

[75] Jerome H Friedman and Werner Stuetzle. "Projection pursuit regression". In: *Journal of the American statistical Association* 76.376 (1981), pp. 817–823.

[76] Peter Friz and Jim Gatheral. "Valuation of volatility derivatives as an inverse problem". In: *Quantitative Finance* 5.6 (2005), pp. 531–542.

[77] Jim Gatheral. "Consistent modeling of SPX and VIX options". In: *Bachelier Congress*. 2008.

[78] Jim Gatheral. "Developments in volatility derivatives pricing". In: *ICBI Global Derivatives and Risk Management conference, Paris*. Citeseer. 2007.

[79] Jim Gatheral. "Further developments in volatility derivatives modeling". In: *Global Derivatives Trading & Risk Management, Paris* (2008).

[80] Jim Gatheral. *The volatility surface: a practitioner's guide*. Vol. 357. John Wiley & Sons, 2011.

[81] Andrew Gelman, Xiao-Li Meng, and Hal Stern. "Posterior predictive assessment of model fitness via realized discrepancies". In: *Statistica sinica* 6.4 (1996), pp. 733–760.

[82] Dudley Gilder. "An empirical investigation of intraday jumps and cojumps in US equities". In: *Available at SSRN 1343779* (2009).

[83] Joanna Goard and Mathew Mazur. "Stochastic volatility models and the pricing of VIX options". In: *Mathematical Finance* 23.3 (2013), pp. 439–458.

[84] Maria T Gonzalez-Perez. "Model-free volatility indexes in the financial literature: A review". In: *International Review of Economics & Finance* 40 (2015), pp. 141–159.

[85] John M Hammersley and Peter Clifford. “Markov fields on finite graphs and lattices”. In: (1971).

[86] Lars Peter Hansen and Robert J Hodrick. “Forward exchange rates as optimal predictors of future spot rates: An econometric analysis”. In: *The Journal of Political Economy* (1980), pp. 829–853.

[87] David Hobson and Martin Klimmek. “Model-independent hedging strategies for variance swaps”. In: *Finance and Stochastics* 16.4 (2012), pp. 611–649.

[88] Arthur E Hoerl and Robert W Kennard. “Ridge regression: Biased estimation for nonorthogonal problems”. In: *Technometrics* 12.1 (1970), pp. 55–67.

[89] Per Hörfelt et al. “An easy-to-hedge covariance swap”. In: *Risk* 25.6 (2012), p. 74.

[90] Darien Huang and Ivan Shaliastovich. “Volatility-of-volatility risk”. In: (2014).

[91] Jing-Zhi Huang and Zhan Shi. “Understanding Term Premia on Real Bonds”. In: *Manuscript, Penn State University* (2012).

[92] Jing-zhi Huang and Zhan Shi. *Determinants of bond risk premia*. Tech. rep. Citeseer, 2011.

[93] Jing-Zhi Huang, Zhan Shi, and Wei Zhong. “Model Selection for High-Dimensional Problems”. In: *Handbook of Financial Econometrics and Statistics*. Springer, 2015, pp. 2093–2118.

[94] Jing-Zhi Huang and Ying Wang. “Should investors invest in hedge fund-like mutual funds? Evidence from the 2007 financial crisis”. In: *Journal of Financial Intermediation* 22.3 (2013), pp. 482–512.

[95] Jing-zhi Huang and Liuren Wu. “Specification analysis of option pricing models based on time-changed Lévy processes”. In: *The Journal of Finance* 59.3 (2004), pp. 1405–1439.

[96] Jingzhi Huang, Lei Lu, et al. *Macro factors and volatility of Treasury bond returns*. Tech. rep. Working paper, 2009.

[97] Bujar Huskaj and Marcus Nossman. “A term structure model for VIX futures”. In: *Journal of Futures Markets* 33.5 (2013), pp. 421–442.

[98] Antti Ilmanen. *Expected returns: An investor’s guide to harvesting market rewards*. John Wiley & Sons, 2011.

[99] Andrey Itkin and Peter Carr. “Pricing swaps and options on quadratic variation under stochastic time change modelsfffdfffdfffddiscrete observations case”. In: *Review of Derivatives Research* 13.2 (2010), pp. 141–176.

[100] Jean Jacod and Viktor Todorov. “Testing for common arrivals of jumps for discretely observed multidimensional processes”. In: *The Annals of Statistics* (2009), pp. 1792–1838.

[101] Robert Jarrow et al. “Discretely sampled variance and volatility swaps versus their continuous approximations”. In: *Finance and Stochastics* 17.2 (2013), pp. 305–324.

[102] George J Jiang and Yisong S Tian. “Extracting model-free volatility from option prices: An examination of the VIX index”. In: (2007).

[103] Michael S Johannes and Nick Polson. “MCMC methods for continuous-time financial econometrics”. In: *Available at SSRN 480461* (2003).

[104] Travis L Johnson. “Risk premia and the vix term structure”. In: *Journal of Financial and Quantitative Analysis* 52.6 (2017), pp. 2461–2490.

[105] Richard Jordan and Charles Tier. “The variance swap contract under the CEV process”. In: *International Journal of Theoretical and Applied Finance* 12.05 (2009), pp. 709–743.

[106] Scott Joslin, Marcel Priebsch, and Kenneth J Singleton. “Risk premium accounting in macro-dynamic term structure models”. In: *Manuscript, Standford University* (2009).

[107] Scott Joslin, Marcel Priebsch, and Kenneth J Singleton. “Risk premiums in dynamic term structure models with unspanned macro risks”. In: *The Journal of Finance* 69.3 (2014), pp. 1197–1233.

[108] Andreas Kaeck and Carol Alexander. “Continuous-time VIX dynamics: On the role of stochastic volatility of volatility”. In: *International Review of Financial Analysis* 28 (2013), pp. 46–56.

[109] Andreas Kaeck and Carol Alexander. “VIX dynamics with stochastic volatility of volatility”. In: *ICMA Centre, Henley Business School, University of Reading, UK* (2010).

[110] Jan Kallsen, Johannes Muhle-Karbe, and Moritz Voß. “Pricing options on variance in affine stochastic volatility models”. In: *Mathematical Finance* 21.4 (2011), pp. 627–641.

[111] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. Vol. 113. Springer Science & Business Media, 2012.

[112] Inna Khagleeva. “Understanding jumps in the high-frequency VIX”. In: (2012).

[113] Chenxu Li. “Closed-form expansion, conditional expectation, and option valuation”. In: *Mathematics of Operations Research* 39.2 (2013), pp. 487–516.

[114] Runze Li, Wei Zhong, and Liping Zhu. “Feature screening via distance correlation learning”. In: *Journal of the American Statistical Association* 107.499 (2012), pp. 1129–1139.

[115] Guang-Hua Lian and Song-Ping Zhu. "Pricing VIX options with stochastic volatility and random jumps". In: *Decisions in Economics and Finance* 36.1 (2013), pp. 71–88.

[116] Yueh-Neng Lin. "Pricing VIX futures: Evidence from integrated physical and risk-neutral probability measures". In: *Journal of Futures Markets* 27.12 (2007), pp. 1175–1217.

[117] Yueh-Neng Lin. "VIX option pricing and CBOE VIX Term Structure: A new methodology for volatility derivatives valuation". In: *Journal of Banking & Finance* 37.11 (2013), pp. 4432–4446.

[118] Yueh-Neng Lin and Chien-Hung Chang. "Consistent modeling of S&P 500 and VIX derivatives". In: *Journal of Economic Dynamics and Control* 34.11 (2010), pp. 2302–2319.

[119] Yueh-Neng Lin and Chien-Hung Chang. "VIX option pricing". In: *Journal of Futures Markets* 29.6 (2009), pp. 523–543.

[120] Chien-Ling Lo et al. *Volatility model specification: Evidence from the pricing of VIX derivatives*. Tech. rep. Working paper, National Taiwan University, 2013.

[121] Zhongjin Lu and Yingzi Zhu. "Volatility components: The term structure dynamics of VIX futures". In: *Journal of Futures Markets* 30.3 (2010), pp. 230–256.

[122] Sydney C Ludvigson and Serena Ng. *A factor analysis of bond risk premia*. Tech. rep. National Bureau of Economic Research, 2009.

[123] Sydney C Ludvigson and Serena Ng. "Macro factors in bond risk premia". In: *Review of Financial Studies* 22.12 (2009), pp. 5027–5067.

[124] Xingguo Luo and Jin E Zhang. "The term structure of VIX". In: *Journal of Futures Markets* 32.12 (2012), pp. 1092–1123.

[125] Dilip B Madan and Marc Yor. "The S&P 500 Index as a Sato Process Travelling at the Speed of the VIX". In: *Applied Mathematical Finance* 18.3 (2011), pp. 227–244.

[126] Sébastien Maillard, Thierry Roncalli, and Jérôme Teïletche. "On the properties of equally-weighted risk contributions portfolios". In: (2008).

[127] Jacinto Marabel Romo. "Pricing volatility options under stochastic skew with application to the VIX index". In: *The European Journal of Finance* 23.4 (2017), pp. 353–374.

[128] Michael McAleer and Marcelo C Medeiros. "Realized volatility: A review". In: *Econometric Reviews* 27.1-3 (2008), pp. 10–45.

[129] Javier Mencía and Enrique Sentana. "Valuation of VIX derivatives". In: *Journal of Financial Economics* 108.2 (2013), pp. 367–391.

[130] Kerrie L Mengersen, Richard L Tweedie, et al. "Rates of convergence of the Hastings and Metropolis algorithms". In: *The annals of Statistics* 24.1 (1996), pp. 101–121.

[131] Robert C Merton. "On estimating the expected return on the market: An exploratory investigation". In: *Journal of financial economics* 8.4 (1980), pp. 323–361.

[132] Richard O Michaud. "The Markowitz optimization enigma: Is fffdfffdfff-doptimizedfffdfffdfffdoptimal?" In: *Financial Analysts Journal* 45.1 (1989), pp. 31–42.

[133] Vasily Nekrasov. "Kelly criterion for multivariate portfolios: A model-free approach". In: (2014).

[134] Whitney K Newey and Kenneth D West. *A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix*. 1986.

[135] Andrew Papanicolaou and Ronnie Sircar. "A regime-switching Heston model for VIX and S&P 500 implied volatilities". In: *Quantitative Finance* 14.10 (2014), pp. 1811–1827.

[136] Yang-Ho Park. "Volatility of volatility and tail risk premiums". In: (2013).

[137] Dimitris Psychoyios, George Dotsis, and Raphael N Markellos. "A jump diffusion model for VIX volatility options and futures". In: *Review of Quantitative Finance and Accounting* 35.3 (2010), pp. 245–269.

[138] Edward Qian. "Risk parity portfolios: Efficient portfolios through true diversification". In: *Panagora Asset Management* (2005).

[139] Edward E Qian. *Risk Parity Fundamentals*. CRC Press, 2016.

[140] Brian D Ripley. *Stochastic simulation*. Vol. 316. John Wiley & Sons, 2009.

[141] Gareth O Roberts and Nicholas G Polson. "On the geometric convergence of the Gibbs sampler". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1994), pp. 377–384.

[142] Gareth O Roberts and Adrian FM Smith. "Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms". In: *Stochastic processes and their applications* 49.2 (1994), pp. 207–216.

[143] Wim Schoutens. "Moment swaps". In: *Quantitative Finance* 5.6 (2005), pp. 525–530.

[144] Artur Sepp. "Pricing options on realized variance in the Heston model with jumps in returns and volatility". In: (2008).

[145] Artur Sepp. "VIX option pricing in a jump-diffusion model". In: (2008).

[146] Jinghong Shu and Jin E Zhang. "Causality in the VIX futures market". In: *Journal of Futures Markets* 32.1 (2012), pp. 24–46.

[147] Zhaogang Song. “Expected VIX option returns”. In: (2012).

[148] Zhaogang Song and Dacheng Xiu. “A tale of two option markets: State-price densities implied from S&P 500 and VIX option prices”. In: *Unpublished working paper. Federal Reserve Board and University of Chicago* (2012).

[149] Robert F Stambaugh. “The information in forward rates: Implications for models of the term structure”. In: *Journal of Financial Economics* 21.1 (1988), pp. 41–70.

[150] Anatoliy Swishchuk. “Modeling of variance and volatility swaps for financial markets with stochastic volatilities”. In: *WILMOTT magazine* 2 (2004), pp. 64–72.

[151] Robert Tibshirani. “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288.

[152] Luke Tierney. “Markov chains for exploring posterior distributions”. In: *the Annals of Statistics* (1994), pp. 1701–1728.

[153] Viktor Todorov and George Tauchen. “Volatility jumps”. In: *Journal of Business & Economic Statistics* 29.3 (2011), pp. 356–371.

[154] Ruey S Tsay. *Analysis of financial time series*. Vol. 543. John Wiley & Sons, 2005.

[155] Clemens Völkert. “The distribution of uncertainty: Evidence from the VIX options market”. In: *Journal of Futures Markets* 35.7 (2015), pp. 597–624.

[156] Hansheng Wang, Runze Li, and Chih-Ling Tsai. “Tuning parameter selectors for the smoothly clipped absolute deviation method”. In: *Biometrika* 94.3 (2007), pp. 553–568.

[157] Ruoyang Wang, Chris Kirby, and Steven P Clark. “Volatility of volatility, expected stock return and variance risk premium”. In: (2013).

[158] Zhiguang Wang and Robert T. Daigler. “The option skew index and the volatility of volatility”. In: *Working paper* (2012).

[159] Zhiguang Wang and Robert T Daigler. “The performance of VIX option pricing models: empirical evidence beyond simulation”. In: *Journal of Futures Markets* 31.3 (2011), pp. 251–281.

[160] Robert E Whaley. “Understanding vix”. In: (2008).

[161] Jonathan H Wright and Hao Zhou. “Bond risk premia and realized jump risk”. In: *Journal of Banking & Finance* 33.12 (2009), pp. 2333–2345.

[162] Dacheng Xiu. “Hermite polynomial based expansion of European option prices”. In: *Journal of Econometrics* 179.2 (2014), pp. 158–177.

[163] Cun-Hui Zhang et al. “Nearly unbiased variable selection under minimax concave penalty”. In: *The Annals of statistics* 38.2 (2010), pp. 894–942.

[164] Jin E Zhang and Yingzi Zhu. “VIX futures”. In: *Journal of Futures Markets* 26.6 (2006), pp. 521–531.

[165] Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. “Edgeworth expansions for realized volatility and related estimators”. In: *Journal of Econometrics* 160.1 (2011), pp. 190–203.

[166] Li-Ping Zhu et al. “Model-free feature screening for ultrahigh-dimensional data”. In: *Journal of the American Statistical Association* (2012).

[167] Song-Ping Zhu and Guang-Hua Lian. “A CLOSED-FORM EXACT SOLUTION FOR PRICING VARIANCE SWAPS WITH STOCHASTIC VOLATILITY”. In: *Mathematical Finance* 21.2 (2011), pp. 233–256.

[168] Song-Ping Zhu and Guang-Hua Lian. “An analytical formula for VIX futures and its applications”. In: *Journal of Futures Markets* 32.2 (2012), pp. 166–190.

[169] Yingzi Zhu and Jin E Zhang. “Variance term structure and VIX futures pricing”. In: *International Journal of Theoretical and Applied Finance* 10.01 (2007), pp. 111–127.

[170] Hui Zou. “The adaptive lasso and its oracle properties”. In: *Journal of the American statistical association* 101.476 (2006), pp. 1418–1429.

[171] Hui Zou and Runze Li. “One-step sparse estimates in nonconcave penalized likelihood models”. In: *Annals of statistics* 36.4 (2008), p. 1509.

Vita

Jun Ni

Jun Ni was born in 1989 in Jilin Province, China. In 2008, he was admitted by the Department of Mathematics, University of Science and Technology of China (USTC). He obtained the Bachelor degree in Applied Mathematics from USTC in 2012. Then he was accepted by the PhD program of applied mathematics at Penn State University, where his research was focused on VIX modeling, bond risk premia prediction and asset allocation. He will graduate on April 2018.