

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/93986>

Copyright and reuse:

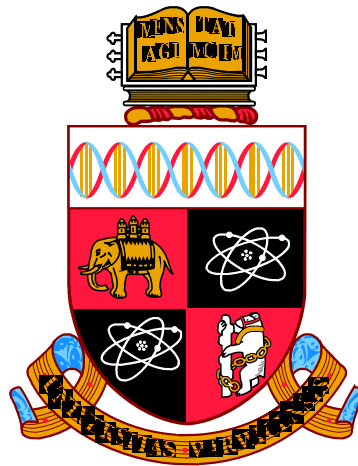
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



**Essays on Multi-Asset Jump Diffusion
Models: Estimation, Asset Allocation and
American Option Pricing.**

by

Cheng-Yu Yang

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Business School

September 2016



Contents

1	General Introduction	1
2	Efficient estimation of lower and upper bounds for pricing higher-dimensional American arithmetic average options by approximating their payoff functions	4
2.1	Introduction	4
2.2	The model and valuation	9
2.2.1	The model	9
2.2.2	Pricing American options by dynamic programming algorithm	10
2.3	Lower and upper bounds	11
2.3.1	Arithmetic mean and geometric mean	11
2.3.2	Constructing lower bounds	15
2.3.3	Constructing upper bounds	17
2.4	Numerical Experiments	18
2.5	Conclusion	37
3	Dynamic Optimal Portfolio Choice Problem under Financial Contagion	38
3.1	Introduction	38
3.2	Contagion Literatures	40
3.2.1	What is financial contagion?	41
3.3	A Multi-Dimensional Jump-Diffusion Model with Stochastic Volatilities	46
3.3.1	A Stochastic Variance-Covariance Process with Jumps	48
3.3.2	A Hawkes-style jump	49
3.3.3	Interpretation of J	50
3.3.4	Parameter setting in this paper	50
3.4	Solve the optimal portfolio choice problem	52
3.4.1	The Solution of the Investment Problem	52

3.4.2	Decomposition	54
3.4.3	Decomposition of Hedging Demands	56
3.5	Simple Examples	58
3.5.1	Two-dimensional case: $n=2$	58
3.5.2	One-dimensional case: $n = 1$	59
3.5.3	No jump cases	59
3.5.4	Intuition given by Sensitivity Analysis	60
3.6	Numerical Analysis: Financial Implications	61
3.6.1	Hedging Demands	62
3.6.2	Capturing financial contagion	64
3.6.3	Effects of Model Misspecification	67
3.7	Conclusion	68
4	Estimation for Multivariate Stochastic Volatility Models	84
4.1	Introduction	84
4.2	Literatures Reviews	86
4.3	The Model	89
4.4	MCMC Estimation Implementation	90
4.4.1	Model Discretization	90
4.5	Posterior distribution derived for our model	93
4.5.1	Posterior for parameters	94
4.5.2	Posterior for latent variables	98
4.5.3	MCMC Procedure Specification	101
4.5.4	Brief Introduction of Slice Sampling methods	102
4.6	Numerical Results	105
4.7	Conclusion	116
4.8	Preliminary Results for empirical applications	118
4.9	Complementary: Derived Moment Generating Functions for moment- matching techniques	122

List of Figures

2.1	Correlation between arithmetic and geometric averages	15
2.2	Comparison of convergence speeds of MLSM and OLSM for pricing American arithmetic average option on 6 assets	26
2.3	Effects of volatility, time-to-maturity and strike price on pricing error. All stocks have the same setting as Example 2.3, i.e., Common Initial Price $S_{i0} = 100$, Strike Price $K=100$, Interest rate $r=3\%$, no Dividend, Maturity $T=1$ Year, Common Volatility $\sigma_{i1} = 20\%$, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 10$, if not otherwise mentioned. In particular, the pricing error is calculated with SPSA estimators as benchmark and $M=10000$. .	29
3.1	Total Hedging Demand against Time Horizon against various models	65
3.2	Hedging Demand against $C_2(1)$ (DJ)	66
3.3	Hedging Demand against $C_1(1)$ (DJ)	66
3.4	Hedging Demand of Asset 1 against $J_{21} J_{12}$ (DJ)	67
3.5	Hedging Demand of Asset 2 against $J_{21} J_{12}$ (DJ)	68
4.1	Illustration of Single-variate Slice Sampling Method	103
4.2	Illustration of Metropolis-Hasting and slice sampling scheme . .	104
4.3	Fitted Variance-Covariance states - Σ_{t11}	115
4.4	Fitted Variance-Covariance states - Σ_{t21}	116
4.5	Fitted Variance-Covariance states - Σ_{t22}	116
4.6	Asymmetric Correlation - Market and Consumer portfolios . . .	119
4.7	Asymmetric Correlation - Market and Manufacturing portfolios	120
4.8	Asymmetric Correlation - Market and High-Technology portfolios	120
4.9	Asymmetric Correlation - Market and Health portfolios	120
4.10	Asymmetric Correlation - Market and Other portfolios	121

List of Tables

2.1	American Arithmetic Average Options on 1 to 6 underlying stocks.	30
2.2	American Arithmetic Average Option on 10, 30 and 50 underlying stocks following MJD process.	31
2.3	American Arithmetic Average Options on 1 to 6 underlying stocks. Verifying the accuracy of MLSM upper bound with SPSA approach.	32
2.4	American Arithmetic Average Option on 30 assets with 20 long positions and 10 short positions.	33
2.5	American Arithmetic Average Option on 20 and 30 underlying stocks with different volatilities.	34
2.6	American Arithmetic Average Option on 10 and 20 assets with Heston's model of stochastic volatility.	35
2.7	American Arithmetic Average Options on 10 underlying stocks.	36
3.1	Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled	70
3.2	Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled	71
3.3	Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled	72
3.4	Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled	73
3.5	Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps	74
3.6	Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps	75
3.7	Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps	76
3.8	Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps	77

3.9	Optimal Variance Hedging Demands for 2 Risky Assets with Jumps in Returns Only	78
3.10	Optimal Covariance Hedging Demands for 2 Risky Assets with Jumps in Returns Only	79
3.11	Optimal Asset Jump Hedging Demands for 2 Risky Assets with Jumps in Returns Only	80
3.12	Optimal Variance Hedging Demands for 2 Risky Assets with Jumps in Covariance Matrix Only	81
3.13	Optimal Covariance Hedging Demands for 2 Risky Assets with Jumps in Covariance Matrix Only	82
3.14	Parameters adopted for numerical experiments.	83
4.1	Simulation tests for entangled MJD model, $T=1000$	109
4.2	Simulation tests for entangled MJD model, $T=3000$	110
4.3	Simulation tests for 1 dimensional stochastic volatility model.	111
4.4	Simulation tests for WJD-DJ model with 500 simulated time series	112
4.5	Simulation tests for WJD-DJ model with 1000 simulated time series	113
4.6	Simulation tests for WJD-DJ model with 2000 simulated time series	114
4.7	Hyper parameters of prior distributions	115

To my family

Acknowledgement

There has been so many people who directly or indirectly help during my Ph.D studies. Firstly, I have to deliver my sincerest gratefulness to my supervisor, Dr. Xing Jin, who not only guides my Ph.D studies, but also shares his philosophy of life. During this long lasting Ph.D life, I have learned a lot from him and without his guidance, I would have wasted more times struggling for finding the right way to make myself prepared for academic research. Moreover, when I felt depressed and stressed, his encouragement, enthusiasm and support always revive me from those negative sentiments. I would also like to thank my second supervisor, Dr. Sarah Qian Wang also shares her experience in preparing for job markets and ways of doing research. Although we have no collaboration work yet, I can still feel her patience and kindness for guiding me.

Apart from my supervisors, I would like to express my heartfelt appreciation to the entire Finance group at Wariwck Business School, where kindly and insightful comments and frequent interactions often provide me new ideas and broaden my mind. In particular, I would especially thank Andrea Gamma and Michael Moore for organizing special events.

Besides, I also appreciate my close friends, including Yao Chen, Linquan Chen and Chunling Xia, Jeff Hung, Harold Contreras Munoz, and many others, made during these years. They are not only colleagues whom I spent those sleepless night with but also friends to share lives with. Without their company, I would not experience a Ph.D life as wonderful as it has been.

The last but not the least, I would like to deliver the deepest appreciation to my family, Mao-Hsiu Yang, Mei-Ying Chen, Chen-Hao Yang and Hsiu-Ying Yang Liu. Without their support, encouragement and heart-touching consideration, I could not recover soon when I feel frustrated or depressed. It is the unconditional love of my family that keeps me encouraged and supported to overcome the difficulties of studying abroad.

Declaration

I declare that the three essays included in this thesis has not been submitted for a degree to any other university. Moreover, Chapter 2 and Chapter 4 are collaboration works with my supervisor, Dr. Xing Jin, while Chapter 3 is a collaboration work with Dr. Xing Jin and Dr. Xudong Zeng (School of Finance, Shanghai University of Finance and Economics).

I further declare that the paper drawn from Chapter 2, which is titled as “Efficient estimation of lower and upper bounds for pricing higher-dimensional American arithmetic average options by approximating their payoff functions” has been published on *International Review of Financial Analysis* 44 (2016): 65-77.

Abstract

In the first essay(Chapter 2), we develop an efficient payoff function approximation approach to estimating lower and upper bounds for pricing American arithmetic average options with a large number of underlying assets. This method is particularly efficient for asset prices modeled by jump-diffusion processes with deterministic volatilities because the geometric mean is always a one-dimensional Markov process regardless of the number of underlying assets and thus is free from the curse of dimensionality. Another appealing feature of our method is that it provides an extremely efficient way to obtain tight upper bounds with no nested simulation involved as opposed to some existing duality approaches. Various numerical examples with up to 50 underlying stocks suggest that our algorithm is able to produce computationally efficient results.

Chapter 3 solves portfolio choice problem in multi-dimensional jump-diffusion models designed to capture empirical features of stock prices and financial contagion effect. To obtain closed-form solution, we develop a novel general decomposition technique with which we reduce the problem into two relative simple ones: Portfolio choice in a pure-diffusion market and in a jump-diffusion market with less dimension. The latter can be reduced further to be a bunch of portfolio choice problems in one-dimensional jump-diffusion markets. By virtue of the decomposition, we obtain a semi-closed form solution for the primary optimal portfolio choice problem. Our solution provides new insights into the structure of an optimal portfolio when jumps are present in asset prices and/or their variance-covariance.

In Chapter 4, we develop a estimation procedure based on Markov Chain Monte Carlo methods and aim to provide systematic ways to estimating general multivariate stochastic volatility models. In particular, this estimation technique is proved to be efficient for multivariate jump-diffusion process such as the model developed in Chapter 3 with various simulation studies. As a result, it contributes to the asset pricing literature by providing an efficient estimation technique for asset pricing models.

Abbreviation

AA: Arithmetic average
GA: Geometric average
OLSM: Ordinary least square Monte Carlo
MLSM: Modified lease square Monte Carlo
MCMC: Markov Chain Monte Carlo
MH: Metropolis-Hasting
WJD: Wishart-Jump-Diffusion (model) DJ: Double jump (model)
iDJ: Independent double jump (model)
AJ: asset jump (model)
CJ: covariance jump (model)
MJD: Merton-Jump-Diffusion (model)
SV: Stochastic volatility (model)
SVMJ: Stochastic volatility with Merton Jump (model)
SVC MJ: Stochastic volatility with common Merton Jump (model)
SDE, ODE: stochastic/ordinary differential equations
HJB: Hamilton-Jacobi-Bellman (equation) ML: Maximum Likelihood (method)
QML: Quasi-Maximum Likelihood (method)
SML: Simulated Maximum Likelihood (method)
GMM: Generalized Method of Moment (method)
EMM: Efficient Method of Moment (method)
WinBUGS: Bayesian inference using Gibbs sampling (in Windows system)

Chapter 1

General Introduction

In this thesis, I aim to provide insightful researches that help investors to optimally manage their wealth, which may be one of the original motivation of finance studies. With the growing innovation in technology and increasingly entangled financial markets, every investor is exposed to all sources of economic risks domestically, or internationally. The first idea that comes into my mind is finding a good hedging vehicle for a portfolio.

In Chapter 2, we develop an efficient payoff function approximation approach to estimating lower and upper bounds for pricing American arithmetic average options with a large number of underlying assets. The crucial step in the approach is to find a geometric mean which is more tractable than and highly correlated with a given arithmetic mean. Then the optimal exercise strategy for the resultant American geometric average option is used to obtain a low-biased estimator for the corresponding American arithmetic average (AA) option. This method is particularly efficient for asset prices modeled by jump-diffusion processes with deterministic volatilities because the geometric mean is always a one-dimensional Markov process regardless of the number of underlying assets and thus is free from the curse of dimensionality. Another appealing feature of our method is that it provides an extremely efficient way to obtain tight upper bounds with no nested simulation involved as opposed to some existing duality approaches. Various numerical examples with up to 50 underlying stocks suggest that our algorithm is able to produce computationally efficient results. With such efficient American AA pricing tool, it is feasible to hedge a large portfolio with the corresponding American AA option.

Apart from hedging, understanding how stock prices move and how markets

react during financial crisis would be of great research value since it helps for preventing investors from portfolio loss. Recent empirical studies find stock prices tend to have big move together and a big jump may be followed by more frequent jumps, which is especially evident during financial crisis.

In Chapter 3, We study the portfolio choice problem in multi-dimensional jump-diffusion models designed to capture these empirical features and capture the financial contagion effects. To obtain closed-form solution, we develop a novel general decomposition technique with which we reduce the problem into two relative simple ones: Portfolio choice in a pure-diffusion market and in a jump-diffusion market with less dimension. The latter can be reduced further to be a bunch of portfolio choice problems in one-dimensional jump-diffusion markets. By virtue of the decomposition, we obtain a semi-closed form solution for the primary optimal portfolio choice problem. Our solution provides new insights into the structure of an optimal portfolio when jumps are present in asset prices and/or their variance-covariance. Our results show that the jumps in the variance-covariance have important effects on the asset allocations, especially when there are jumps in the asset prices simultaneously. Meanwhile, the hedging demands for jumps are much more significant compared to variance or covariance hedging demands for diffusion risks and ignoring jump risk in the variance-covariance may cause large wealth equivalent loss in the presence of jumps in the asset prices. In addition, two novel components integrated to capture empirical features are verified to cause significant effects in the resulting optimal portfolio weights. As a result, the proposed multivariate model provides an potential ideal model to study financial contagion. Moreover, with optimal portfolio problem solved with semi-closed form solution, financial contagion may be studied in the context of asset allocation quantitatively. To some extent, this paper sheds new lights on the financial contagion and portfolio choice literatures.

In Chapter 4, we develop a systematic estimation procedure for multivariate stochastic volatility models based on Markov Chain Monte Carlo (MCMC) and slice sampling techniques. In particular, this estimation technique makes inference problem feasible for a wide class of models, termed as the Wishart-Jump-Diffusion (WJD) models, where the variance-covariance process are allowed to follow a Wishart process and jumps in returns and covariance may be included. By varying the parameters and number of assets, WJD model reduces to several important models, such as the Wishart-Diffusion model by Buraschi et al. (2010) and the widely studied 1-dimension stochastic volatility models,

e.g. stochastic volatility model with common jump in Eraker et al. (2003), among others.

The detailed establishment of this estimation including, derivation of posterior distributions for parameters and state variables and novel transformations incorporated are provided and ready for future researches. Various numerical studies are conducted to verify the accuracy and efficiency of the proposed estimation procedure. In particular, the estimates obtained are all extremely close to the assigned true values. The true values are adopted from existing literatures and hence benchmarks are available for comparison. For example, the efficiency is verified to be improved in terms of number of sampling required for achieving comparable accuracy. In summary, the developed estimation procedure is shown to be efficient and accurate and provides an potential systematic inference tool for researches about multivariate/univariate stochastic volatility models. In the end of this chapter, an augmented Merton-Jump-Diffusion (MJD) model are estimated to fit the 5-Industry portfolios returns in the Fama-French Data Library to study the asymmetric correlation feature as an empirical application of the proposed estimation procedure. The resulting estimates replicate the asymmetric correlation well.

Chapter 2

Efficient estimation of lower and upper bounds for pricing higher-dimensional American arithmetic average options by approximating their payoff functions

1

2.1 Introduction

The importance of American-style options has been growing increasingly and pricing of American options especially high-dimensional cases remains one of the challenging problems both theoretically and practically in the option pricing theory. In particular, high-dimensional American options would be valuable research topics. For example, Shiu et al (2013) document that basket warrants, essentially basket options with multiple underlying assets become more popular over the past decade.² In this paper, we focus on pricing American arithmetic average options. The appealing advantage of an American arithmetic average option lies in the fact that it exactly replicates the evolution of the portfolio

¹This paper has been published on *International Review of Financial Analysis* 44 (2016): 65-77

²Although they focus on dealing with European basket warrants, essentially basket options, the American-style ones would be more important in practice.

formed by the underlying assets. For example, the cost of hedging a portfolio with an American arithmetic average option is much lower than a portfolio of individual options on the same underlying assets since the former takes the correlations among the underlying assets into account and only one option is involved in hedging. Besides, it would be simple for investors to replicate the payoff of any portfolio without actually holding the portfolio if there is such an American arithmetic average option available on the market. Given these significant applications, efficient pricing methods for American arithmetic average options written on the average of multiple underlying assets are of great value from various points of view such as hedging and risk management especially after the recent financial crisis that re-emphasized the importance of risk management. The purpose of this paper is to develop an efficient approach to obtaining lower and upper bounds for American arithmetic average option prices on a large number of underlying assets.

The traditional valuation methods, such as lattice and tree-based techniques, for pricing high dimensional American option pricing problems are typically plagued by the curse of dimensionality and thus, simulation-based numerical methods are inevitably required. Earlier literature about simulation-based approaches can be traced back to Boyle (1977) in which European style claim is priced with Monte Carlo (MC) simulation. American style option pricing techniques with MC simulation include Bundling Methods in Tilley (1993), Stratified State Aggregation (SSA) in Barraquand and Martineau (1995), Stochastic Mesh Method (SMM) in Broadie and Glasserman (2004), regression-based approach in Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001), among others.

The existing simulation-based methods can be categorized into: (1) Primal approach, which aims to obtain a lower bound for an American option by estimating a suboptimal exercise strategy, e.g., regression-based approaches as in Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001); (2) Duality approach, which estimates an upper bound for an American option by using a dual martingale, e.g. Rogers (2002), Haugh and Kogan (2004) and Anderson and Broadie (2004).

Among existing primal approaches, the most important simulation-based method is the regression-based approach, where computational costs are approximately linear in exercise opportunities and the number of simulated paths. The theory

has been well established in Carriere (1996), Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001), etc. Related convergence analysis and simulation issues can be found in Tsitsiklis and Van Roy (2001), Clément et al. (2002), Glasserman and Yu (2004a,b) and Stentoft (2004).

In particular, the least squares method (LSM) developed by Longstaff and Schwartz (2001) is the most widely used method due to its simplicity and generality. A lower bound of an American option can be obtained from a suboptimal optimal exercise strategy derived from linear regression procedure. However, this method and other primal approaches are becoming computationally expensive with the increasing dimension of pricing problem and hence the trade-off between computational costs and efficiency of approximation would be a critical issue.

A variety of methods have been proposed to improve the performance of regression-based approaches. For instance, to address arbitrary style of continuation values, Kohler et al (2010) use least square neural network regression estimates and estimate continuation values from artificial MC simulated paths. Their approach is more general than LSM since the regression is nonparametric. But, compared to LSM, the nonparametric in Kohler et al (2010) would be even worse to implement for pricing high-dimensional American options.³ More recently, Jain and Oosterlee (2012) proposed a stochastic grid method (SGM) which could be regarded as a hybrid of Barraquand and Martineau (1995): stratified sampling along pay-off method, Longstaff and Schwartz (2001): Least square Monte Carlo method and Broadie and Glasserman (1997b): stochastic mesh method. The proposed SGM algorithm is more suitable for pricing some high-dimensional American options than existing methods. However, SGM would be computationally costly when sub-simulations are embedded and more early exercise times are allowed.

To circumvent the curse of the dimensionality problem associated with pricing of multi-dimensional American options, several dimension reduction methods have been proposed. For example, Barraquand and Martineau (1995) introduce a partitioning algorithm. Their method differs from Tilley's bundling algorithm in that they partition the payoff space instead of the state space. Hence, only a one-dimensional space is partitioned at each time step, regardless the dimension of the problem. More recently, Jin et al (2013) further integrate this idea

³We thank an anonymous referee for pointing out this to us.

into state-space partitioning algorithm (SSPM) developed by Jin, Tan and Sun (2007) and improve the computational efficiency significantly with computational accuracy preserved. Those papers, however, do not provide an algorithm for upper bounds.

In the present paper, we follow the dimension reduction approach to pricing high-dimensional American arithmetic average options. The key idea is to find a highly correlated geometric average for a given arithmetic average. As will become clear later, the former is more tractable than the latter in the sense that the geometric average has a lower dimension⁴ than the corresponding arithmetic average, and thus the optimal exercise strategy for the American geometric average option is far easier to obtain than for the American arithmetic average option. In particular, when the asset prices are modeled by jump-diffusion processes with deterministic volatilities, the geometric mean is always a one-dimensional Markov process regardless of the number of underlying assets, and thus is free from the curse of dimensionality. Then the optimal exercise strategy for the American geometric average option is used to obtain a lower bound for the corresponding American arithmetic average option. In addition, by using an inequality similar to (4) in Haugh and Kogan (2004), we provide an extremely fast way to obtain the corresponding upper bound without nested MC simulations. To be more specific, in the inequality (4) in Haugh and Kogan (2004), we approximate the payoff function of given American arithmetic average option by the one of a highly correlated American geometric average option. Unlike Haugh and Kogan (2004), we do not need to find the optimal supermartingale and thus we do not need nested MC simulations.

An important limitation of the lower bound is that it is not easy to evaluate the accuracy of its approximation to the true option price. Upper bounds in combination with the corresponding lower bounds allow us to measure the accuracy of price estimators for American average options. In earlier literature, Broadie and Glasserman (1997, 2004) propose stochastic mesh methods which generate not only lower but also upper bounds and both bounds converge

⁴For assets following GBM, the geometric average is always one-dimensional Markov process. However, the dimension will increase if other state variables are involved such as stochastic volatility. Consider a case where there are ten stocks and the price of each stock follows Heston stochastic volatility model. Then, an arithmetic average depends on twenty state variables, namely, ten stock price processes and ten volatility processes. By contrast, the corresponding geometric average depends on eleven state variables, that is, the geometric average process itself and the ten volatility processes.

asymptotically to the true value. Despite the advantage of obtaining the upper bound, the stochastic mesh methods are quite computationally demanding. Boyle et al (2003) further generalizes Broadie and Glasserman (1997, 2004) with a low-discrepancy sequence for efficiency.

Independently developed by Rogers (2002), Anderson and Broadie (2004) and Hough and Kogan (2004), duality approach is the most general technique among those upper bound related approach. The idea is to introduce a dual martingale in the pricing problem and rewrite the primal problem into a dual minimization problem. For example, Anderson and Broadie (2004) use nested MC simulation to approximate the optimal exercise strategy. On the other hand, Hough and Kogan (2004) apply an intensive neural network algorithm and low discrepancy sequences to estimate the option prices. However, their estimation techniques to estimate dual martingale do not preserve the martingale property in general and the computational cost is generally high.

To improve this, Glasserman and Yu (2004b) proposed a special regression algorithm to preserve the martingale property. Nonetheless, the martingale property Condition (C3) on the basis functions may not be straightforward to verify in practice. In terms of efficiency, Kolodko and Schoenmakers (2004) try to overcome the computational inefficiency of nested simulation by choosing a different estimator to reduce the number of inner path simulations. However, the upper bound is not guaranteed by their estimator as the number of inner path is too few.

Instead of estimating a dual martingale directly, Belomestny et al (2009) estimate the coefficient of the corresponding martingale representation of the dual martingale. By martingale representation theorem, the martingale property of the estimated dual martingale is preserved. The resultant bound is then the true upper bound. More recently, Zhu, Ye and Zhou (2013) extend the method in Belomestny et al (2009) to a jump-diffusion model. Their theoretical analysis shows that the martingale property of the estimated optimal dual martingale is preserved and no nested simulation is used in their algorithm. These methods, however, may become impractical for pricing high-dimensional American options as a regression-based method similar to LSM is employed to estimate dual martingales. By contrast, our upper bound algorithm requires neither nested simulation nor high-dimensional regression especially when the asset prices are modeled by exponential jump-diffusion processes with deterministic volatilities.

In summary, we have made two contributions to the literature of pricing high-dimensional American arithmetic average options. First, we have developed a computationally efficient dimension reduction method to estimate lower bound. Second, we provide an easy-to-implement approach to evaluate the upper bound which involves no nested simulation and is based on a simple linear regression procedure. We are not aware of any research in the current literature that estimates lower and/or upper bounds for pricing high-dimensional American arithmetic average options via geometric mean approximation. As mentioned above, the essence of our algorithm is to approximate an arithmetic average by a lower-dimensional and more tractable geometric average which is highly correlated with the arithmetic average. In contrast, extant literature usually approximates the continuation values of American options.

The remaining of this paper is organized as follows. In Section 2, we introduce basic dynamic programming framework for pricing American-style options. In Section 3, we provide some theoretical considerations and empirical tests that justify using a highly correlated geometric mean to approximate a given arithmetic mean. Then we present procedures for estimating lower and upper bounds for pricing American arithmetic average options. In Section 4, various numerical experiments are provided to illustrate the performance of our algorithms. Section 5 concludes the paper.

2.2 The model and valuation

In this section, we first introduce American arithmetic average options and then formulate the American option pricing framework by using dynamic programming approach.

2.2.1 The model

Following the literature, we consider a Bermudan-style arithmetic average option with n underlying stocks and strike price, K . The option is exercisable at any date in the set $\Gamma = \{t_0 = 0, t_1, \dots, t_N = T\}$ where T is the pre-specified maturity. When N increases to infinity, we may view the discrete exercise dates as an approximation to the continuous exercise dates to value an American-style arithmetic average option. Considering a put option, the exercise value at time

$t \in \Gamma$ of an American arithmetic average option is defined as

$$h_t^A(S_t) = \left(K - \sum_{i=1}^n a_i S_{it} \right)^+, \quad (2.1)$$

where $S_t = (S_{1t}, \dots, S_{nt})$, S_{it} denotes the price of the i th underlying asset at time t , $i = 1, \dots, n$ and a_i represents the weight of the i th stock satisfying conditions: $a_i > 0, i = 1, \dots, n$ and $\sum_{i=1}^n a_i = 1$. It is straightforward to define a call option.

It is worth mentioning that the two restrictions above can be readily relaxed. For example, a short position can be allowed in the i th stock, i.e., $a_i < 0$. For simplicity, assume that $a_i > 0, i = 1, \dots, n_0$ and $a_i < 0, i = n_0 + 1, \dots, n$ with $n_0 \leq n$. In this case, the sum in (2.1) can be expressed as

$$\sum_{i=1}^n a_i S_{it} = \sum_{i=1}^{n_0} a_i S_{it} - \sum_{i=n_0+1}^n (-a_i) S_{it}$$

As shown later, the sum $\sum_{i=1}^n a_i S_{it}$ can be approximated by the difference of two geometric means and thus, the dimension of the pricing problem is reduced.

2.2.2 Pricing American options by dynamic programming algorithm

In this section, we formally illustrate the dynamic programming formulation for pricing American options. To this end, consider an economy described by the probability space $(\Omega, \mathbb{F}, \mathbb{P})$ where Ω is the sample space, \mathbb{F} is the σ -algebra and \mathbb{P} is a risk-neutral probability measure. Following the literature, we formulate a general class of American option pricing problem through an R^d -valued Markov process $X = \{X_t, 0 \leq t \leq T\}$ (with X_0 fixed) defined on the probability space, where the American option can be exercised at any time τ on or before the pre-specified maturity T . The process X represent the prices of underlying assets, volatilities, interest rates and other state variables.⁵ At each time, $t_i, i = 0, \dots, N$, the option buyer makes the exercising decision based on the dynamic programming framework. More specifically, given a nonnegative adapted payoff function, h_{t_i} , the buyer chooses to exercise the option and gains h_{t_i} if the payoff is greater than the continuation value at time t_i . Let Γ_i denote

⁵For notational simplicity, S_t or S is adopted hereafter since we focus on the stock price as the state variable. For cases with additional state variables such as stochastic volatility, X would be adopted as discussed.

the set of stopping times (with respect to the history of S) taking values in $\{t_i, \dots, t_N = T\}$. The option value at time t_i can be defined as:

$$V_{t_i}(x) = \sup_{\tau \in \Gamma_i} E_t \left[e^{-\int_{t_i}^{\tau} r(s) ds} h_{\tau} \middle| S_{t_i} = x \right], x \in \mathbb{R}^d$$

for $i = 0, \dots, N$, where $\{r(t), 0 \leq t \leq T\}$ is the instantaneous short rate process.

As a result, the option value at time 0 is determined by the dynamic programming algorithm:

$$\begin{aligned} V_{t_N}(x) &= h_{t_N}(x) \\ V_{t_i}(x) &= \max \left\{ h_{t_i}(x), E_t \left[e^{-\int_{t_i}^{t_{i+1}} r(s) ds} V_{t_{i+1}}(S_{t_{i+1}}) \middle| S_{t_i} = x \right] \right\}, \quad i = 0, \dots, N-1. \end{aligned}$$

Conventionally, we define the continuation value as

$$C_{t_i}(x) = E_t \left[e^{-\int_{t_i}^{t_{i+1}} r(s) ds} V_{t_{i+1}}(S_{t_{i+1}}) \middle| S_{t_i} = x \right], x \in \mathbb{R}^n, \quad i = 0, \dots, N-1.$$

Then the option value at t_i satisfies

$$V_{t_i}(x) = \max \{h_{t_i}(x), C_{t_i}(x)\}, \quad i = 0, \dots, N-1.$$

2.3 Lower and upper bounds

In this section, we propose simulation-based approaches to estimating lower and upper bounds for pricing American arithmetic average options. The key step, as aforementioned, is to construct a highly correlated geometric average for a given arithmetic average. In the next subsection, we first illustrate the construction of the correlated geometric average and then investigate both theoretically and empirically the correlation between the two variables to justify our approaches.

2.3.1 Arithmetic mean and geometric mean

For illustrative purposes, we assume that the price of each of these n stocks follows the widely used Merton-Jump-Diffusion (MJD) process. In particular,

the price dynamics of the i th stock is as follows:

$$\frac{dS_{it}}{S_{it}} = (r - q_i - \sigma_{i2}\lambda k)dt + \sigma_{i1}dW_{it} + \sigma_{i2}(e^J - 1)dN_t, \quad i = 1, \dots, n,$$

where r is a constant risk free rate, q_i is the dividend yield, σ_{i1} is the volatility, σ_{i2} is the coefficient of (or sensitivity to) the jump, λ is the jump intensity, $e^J - 1$ is the jump size, $k = E(e^J - 1)$ is the expected jump size, W_{it} is a standard Brownian motion and $\langle W_{it}, W_{jt} \rangle = \rho_{ij}t$ for $i \neq j = 1, \dots, n$. $W_t = (W_{1t}, \dots, W_{nt})$, N_t and J are mutually independent. The jump is assumed to be a common jump representing the systemic shock arising from the market. By letting $\sigma_{i2} = 0$, the above MJD process reduces to a geometric Brownian motion (GBM). Further, by applying the Ito's lemma, we obtain

$$S_{it} = S_{i0} \exp \left\{ \left(r - q_i - \frac{1}{2}\sigma_{i1}^2 - \sigma_{i2}\lambda k \right) t + \sigma_{i1}W_{it} + J_i N_t \right\}, \quad (2.2)$$

where $J_i = \ln [\sigma_{i2}(e^J - 1) + 1]$.

To find a geometric mean to approximate the arithmetic mean in (2.1), we use (2.2) to rewrite the arithmetic mean as

$$\begin{aligned} AA_t &= \sum_{i=1}^n a_i S_{it} \\ &= \sum_{i=1}^n a_i S_{i0} \exp \left\{ \left(r - q_i - \frac{1}{2}\sigma_{i1}^2 - \sigma_{i2}\lambda k \right) t + \sigma_{i1}W_{it} + J_i N_t \right\} \\ &= (AA_0) \sum_{i=1}^n \tilde{a}_i \exp \left\{ \left(r - q_i - \frac{1}{2}\sigma_{i1}^2 - \sigma_{i2}\lambda k \right) t + \sigma_{i1}W_{it} + J_i N_t \right\}, \end{aligned} \quad (2.3)$$

where $AA_0 = \sum_{i=1}^n a_i S_{i0}$ and $\tilde{a}_i = \frac{a_i S_{i0}}{AA_0}$ satisfying $\sum_{i=1}^n \tilde{a}_i = 1$.

Next, we define

$$\tilde{S}_{it} = \exp \left\{ \left(r - q_i - \frac{1}{2}\sigma_{i1}^2 - \sigma_{i2}\lambda k \right) t + \sigma_{i1}W_{it} + J_i N_t \right\},$$

and

$$\widetilde{AA}_t = \sum_{i=1}^n \tilde{a}_i \tilde{S}_{it}.$$

Then we present the following approximation:

$$\widetilde{AA}_t \approx \alpha_t^0 + \beta_t^0 GA_t, \quad (2.4)$$

where the coefficients α_t^0 and β_t^0 are deterministic, GA_t is a geometric mean defined by

$$GA_t = \prod_{i=1}^n \widetilde{S}_{it}^{a_i} = \exp \left\{ \sum_{i=1}^n \widetilde{a}_i \left[\left(r - q_i - \frac{1}{2} \sigma_{i1}^2 - \sigma_{i2} \lambda k \right) t + \sigma_{i1} W_{it} + J_i N_t \right] \right\}.$$

Consequently, by setting $\alpha_t = AA_0 \alpha_t^0$ and $\beta_t = AA_0 \beta_t^0$, (2.3) and (2.4) imply

$$AA_t = AA_0 \cdot \widetilde{AA}_t \approx AA_0 (\alpha_t^0 + \beta_t^0 GA_t) = \alpha_t + \beta_t GA_t, \quad (2.5)$$

where the coefficients α_t and β_t are estimated by the ordinary least squares method detailed later. The result (2.5) will play a crucial role in approximating the exercise value of an American arithmetic average option by a highly correlated American geometric average option, leading to lower and upper bounds for the American arithmetic average option price. To better understand the relation between \widetilde{AA}_t and GA_t , we present the following result.

Proposition Let x_1, \dots, x_n be positive numbers and a_1, \dots, a_n positive weights satisfying $\sum_{i=1}^n a_i = 1$. Then we have

$$\lim_{s \rightarrow 0} \left(\sum_{i=1}^n a_i x_i^s \right)^{1/s} = \prod_{i=1}^n x_i^{a_i}. \quad (2.6)$$

In particular, for s sufficiently small,

$$\sum_{i=1}^n a_i x_i^s \approx \left(\prod_{i=1}^n x_i^{a_i} \right)^s = \prod_{i=1}^n x_i^{sa_i}. \quad (2.7)$$

proof. (2.6) can be easily proved by applying l'Hospital's rule. \square

To make the intuition behind the relation between \widetilde{AA}_t and GA_t as clear as possible, we concentrate on a simple example where the all stock prices follow geometric Brownian motions with $\sigma_{i2} = 0, i = 1, \dots, n$, and $\sigma_{11} = \sigma_{21} = \dots = \sigma_{n1} = \sigma$ in (2.2). To apply (2.7) to the model, we let

$$x_i = \exp \left\{ W_{it} / \sqrt{t} \right\},$$

where W_{it}/\sqrt{t} is a standard normal random variable. Then, by (2.7), for small $s = \sigma\sqrt{t}$,

$$\sum_{i=1}^n \tilde{a}_i \exp \{ \sigma_{i1} W_{it} \} = \sum_{i=1}^n \tilde{a}_i \exp \left\{ s(W_{it}/\sqrt{t}) \right\} = \sum_{i=1}^n \tilde{a}_i x_i^s \approx \prod_{i=1}^n x_i^{s\tilde{a}_i}.$$

Furthermore, assuming that the interest rate r , the dividend yields $q_i, i = 1, \dots, n$, the volatility σ and the maturity T are small, we obtain

$$\begin{aligned} \widetilde{AA}_t &= \sum_{i=1}^n \tilde{a}_i \exp \left\{ \left(r - q_i - \frac{1}{2}\sigma^2 \right) t + \sigma W_{it} \right\} \\ &\approx \sum_{i=1}^n \tilde{a}_i \exp \{ \sigma W_{it} \} \\ &\approx \prod_{i=1}^n \exp \{ \tilde{a}_i \sigma W_{it} \} \\ &\approx \prod_{i=1}^n \exp \left\{ \tilde{a}_i \left[\left(r - q_i - \frac{1}{2}\sigma^2 \right) t + \sigma W_{it} \right] \right\} = GA_t, \end{aligned}$$

for $t \leq T$. The above analysis suggests that the smaller the volatility σ of the underlying stock prices and the maturity T of an option, the smaller $s = \sigma\sqrt{t}$ and the more accurate approximation of (2.5). In the following, we test the effects of these parameters on the approximation accuracy of (2.5).

As the high correlation between AA_t and GA_t implies accurate approximation of (2.5), we now empirically evaluate the correlation coefficient between the two variables. We consider two examples. In the first example, there are six stocks with prices following geometric Brownian motions (GBM) given by (2.2) where $\sigma_{i2} = 0, i = 1, \dots, 6$. In this example, we assume that the initial price $S_{i0} = 100$, the volatility $\sigma_{i1} = 50\%, i = 1, \dots, 6$, the interest rate $r = 3\%$, the strick price $K = 100$, the maturity $T = 3$ year, and correlation coefficient $\rho_{ij} = 0.5, i \neq j = 1, \dots, 6$. Here we intentionally choose a high volatility and a long maturity to underscore high correlation between AA_t and GA_t because these high values may adversely affect the correlation. The second example also consists of six stocks where the price of each stock evolves according to a jump-diffusion model given by (2.2) with parameters: $\sigma_{i2} = 1, i = 1, \dots, 6$, The common jump size $J \sim N(-0.1, 0.1)$ and the common jump intensity $\lambda = 5$. The other parameters remain the same as in the example 1. For each model, we simulate 10000 paths to evaluate the correlation coefficients between AA_{t_i}

and GA_{t_i} , where $t_i = iT/50, i = 1, \dots, 50$.

The left and right panels in Figure 2.1 present correlation coefficients for the first example and the second example, respectively, with red segments being the 95% confidence intervals. We can see from both panels that the correlation coefficients decrease with the horizon t_i . This result is consistent with the above theoretical analysis. It is noticeable that the correlation coefficients are close to 1 although the horizon is as long as three years and the volatilities are as high as 50%. These results support our approaches developed below.⁶

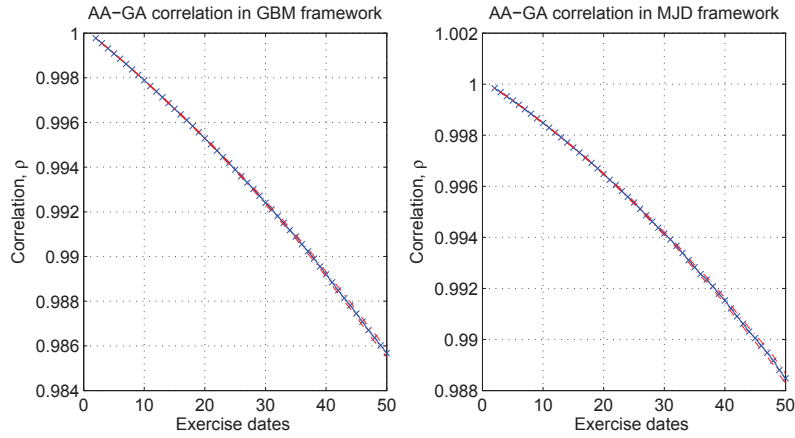


Figure 2.1: Correlation between arithmetic and geometric averages

2.3.2 Constructing lower bounds

Equipped with the results in the last section, we are now able to establish lower bounds for American arithmetic average option prices. More specifically, for an American arithmetic average option with payoff given by (2.1), we first evaluate the optimal exercise strategy for a highly correlated American geometric average option with the time- t exercise value given by $(K - \alpha_t - \beta_t GA_t)^+$, which approximates the time- t exercise value of the American arithmetic average option. Then, this optimal exercise strategy is employed to derive a lower bound for the American arithmetic average option price. The coefficients α_t and β_t are estimated by the least squares method. Without loss of generality, the discretization epochs are assumed to be the same as the set of exercise dates $\Gamma = \{t_0, t_1, \dots, t_N\}$. The procedure is as follows:

⁶We test various examples and find that the correlation coefficients are consistently close to one. To save space, we do not report these results here. The presented simulation results are the averages of 100 simulations with red segments as the 95% confidence interval.

Step 1. Estimating the coefficients α_t and β_t :

1. Simulate M_0 sample paths for the n underlying assets processes, $S_{t_i}^l = (S_{1t_i}^l, \dots, S_{nt_i}^l)$, $l = 1, \dots, M_0, i = 1, \dots, N$.
2. Calculate the arithmetic and geometric averages of $S_{t_i}^l$ and denote them as $AA_{t_i}^l$ and $GA_{t_i}^l$ respectively, $i = 1, \dots, N$.
3. Given $t_i \in \Gamma$, regress $AA_{t_i}^l$ on $GA_{t_i}^l$ based on the equation $AA_{t_i}^l = \alpha_{t_i} + \beta_{t_i} GA_{t_i}^l + \epsilon_i^l$, $l = 1, \dots, N_1$. Store the regression coefficients, $\hat{\alpha}_{t_i}$ and $\hat{\beta}_{t_i}$.

Given $\hat{\alpha}_{t_i}$ and $\hat{\beta}_{t_i}$ estimated in Step 1, we next follow LSM to estimate the optimal exercise strategy for the American geometric average option with time- t_i exercise value defined by

$$h_t^G(S_{t_i}) = \left[K - (\hat{\alpha}_{t_i} + \hat{\beta}_{t_i} GA_{t_i}) \right]^+. \quad (2.8)$$

For expository convenience, we assume that the price process of each stock evolves according to (2.2) and then the geometric average GA_t is a one-dimensional process. Furthermore, we adopt simple basis functions: $1, X$ and X^2 , suggesting that the continuation value at time t_i is represented by $\hat{a}_{t_i} + \hat{b}_{t_i} GA_{t_i} + \hat{c}_{t_i} GA_{t_i}^2$, where $\hat{a}_{t_i}, \hat{b}_{t_i}$ and \hat{c}_{t_i} are constants estimated by following the linear regression method developed by Longstaff and Schwartz (2001)⁷. To save space, we omit this step.

With the regression coefficients $\hat{\alpha}_{t_i}, \hat{\beta}_{t_i}, \hat{a}_{t_i}, \hat{b}_{t_i}$ and $\hat{c}_{t_i}, i = 1, \dots, N$, we are now able to estimate a lower bound for the American geometric average option price.

Step 2. Pricing the American geometric average option and the American arithmetic average option:

1. Simulate a new set of M sample paths for the n underlying assets processes, $S_{t_i}^l = (S_{1t_i}^l, \dots, S_{nt_i}^l)$, $i = 1, \dots, N, l = 1, \dots, M$. And calculate arithmetic average process $AA_{t_i}^l$ and geometric average process $GA_{t_i}^l$, $i = 1, \dots, N, l = 1, \dots, M$.

⁷For simplicity, the number, M , of simulated paths in this step is the same as the one used in Step 2 below.

2. For the American geometric average option, determine the earliest exercise time τ^l as $\tau^l = \min\{t_i \in \Gamma | h_{t_i}^G(S_{t_i}^l) \geq \hat{C}_i^G(S_{t_i}^l), i = 1, \dots, N\}$, if the set $\{t_i \in \Gamma | h_{t_i}^G(S_{t_i}^l) \geq \hat{C}_i^G(S_{t_i}^l), i = 1, \dots, N\}$ is empty, we let $\tau^l = T + 1$, $l = 1, \dots, M$.
3. For τ^l , compute the value function of the American arithmetic average option as $\hat{V}_{\tau^l}^A(S_{\tau^l}^l) = h_{\tau^l}^A(S_{\tau^l}^l)$ if $\tau^l \leq T$; $\hat{V}_{\tau^l}^A(S_{\tau^l}^l) = 0$ if $\tau^l = T + 1$, $l = 1, \dots, M$.
4. For τ^l , compute the value function of the American geometric average option as $\hat{V}_{\tau^l}^G(S_{\tau^l}^l) = h_{\tau^l}^G(S_{\tau^l}^l)$ if $\tau^l \leq T$; $\hat{V}_{\tau^l}^G(S_{\tau^l}^l) = 0$ if $\tau^l = T + 1$, $l = 1, \dots, M$.
5. A lower bound of the American arithmetic average option is estimated by

$$V_0^A = \frac{1}{M} \sum_{l=1}^M e^{-r\tau^l} \hat{V}_{\tau^l}^A(S_{\tau^l}^l). \quad (2.9)$$

6. A lower bound of the American geometric average option is estimated by

$$V_0^G = \frac{1}{M} \sum_{l=1}^M e^{-r\tau^l} \hat{V}_{\tau^l}^G(S_{\tau^l}^l). \quad (2.10)$$

This lower bound (2.10) of the American geometric average option will play an essential role in obtaining upper bound in Section 4.

2.3.3 Constructing upper bounds

We now turn to establishing upper bounds for American arithmetic average option prices by using the approximation results in Section 3.1. Our idea is motivated by the duality approach in Rogers (2002), Anderson and Broadie (2004) and Hough and Kogan (2004) but differs from theirs in that we avoid estimating dual martingales which is in general computationally expensive for

high-dimensional cases. In particular, our idea hinges on the following result:

$$\begin{aligned}
V_0 &= \sup_{\tau \in \Gamma} E[e^{-r\tau} h_\tau^A(S_\tau)] \\
&= \sup_{\tau \in \Gamma} E[e^{-r\tau} h_\tau^A(S_\tau) - e^{-r\tau} h_\tau^G(S_\tau) + e^{-r\tau} h_\tau^G(S_\tau)] \\
&\leq \sup_{\tau \in \Gamma} E[e^{-r\tau} h_\tau^A(S_\tau) - e^{-r\tau} h_\tau^G(S_\tau)] + \sup_{\tau \in \Gamma} E[e^{-r\tau} h_\tau^G(S_\tau)] \\
&\leq E\{\sup_{t \in \Gamma} [e^{-rt} h_t^A(S_t) - e^{-rt} h_t^G(S_t)]\} + \sup_{\tau \in \Gamma} E[e^{-r\tau} h_\tau^G(S_\tau)] \quad (2.11)
\end{aligned}$$

In the last inequality, the first term represents the mean of the largest difference between the two discounted payoffs along a given path⁸ and the second term is the price of the American geometric average option. Apparently, the tightness of the resulting upper bound essentially depends on how well the payoff of American arithmetic average option is approximated by the payoff of the American geometric average option. More specifically, if the arithmetic mean is precisely approximated by the linear function of the geometric mean, then, the first term will be negligible and the price of the American geometric average option given by the second term is close to the price of the American arithmetic average option, implying that the upper bound (2.11) is close to the price of the American arithmetic average option. The resulting upper bounds would be straightforwardly calculated with equation (2.11). Simulation results in Section 4 demonstrate the accuracy of the upper bounds⁹.

2.4 Numerical Experiments

Various numerical experiments of American arithmetic average options are provided to examine the efficiency of our methods proposed in the previous sections, each example containing up to 50 underlying stocks. The first example is taken from Kovalov et al (2007) where American arithmetic average options with up to six underlying assets are valued via their Finite-Element-Method (FEM). Their FEM values are used as our benchmarks in our first example while the other examples do not have price benchmarks. In the mean time, we also compare the

⁸This term can be estimated by two steps. First, given a path, we maximize the function $e^{-rt} h_t^A(S_t) - e^{-rt} h_t^G(S_t)$ across time steps; second, we take the average of the optimal objective function obtained in the first step across paths.

⁹In Section 4, the price of each underlying stock MJD with constant volatility and interest rate and thus the corresponding geometric average GA_t is a one-dimensional process. As documented by the literature, a one-dimensional American put option can be accurately priced by lower bound based on LSM. For this reason, we will use LSM described in Section 3.2 to price the American geometric average option in (2.11) with the estimate given by (2.10).

performance of lower bounds estimated by the least squares regression method (called OLSM) in Longstaff and Schwartz (2001) and our methods (termed MLSM). In all examples except Example 6, we assume that the asset prices follow either geometric Brownian motion or more general Merton-Jump-Diffusion (MJD) processes (2.2) in Section 3 under certain risk-neutral probability measure.

For each example below, we applied 100 runs for both approaches, and the reported statistics were collected from the repeated simulations. The CPU time (in seconds) shows the time of a single run by averaging over the total runs. The simulations were implemented on the Intel(R) Xeon(R) E5-2690, 2.9GHz machine with MATLAB software. The regression is performed with the standard built-in function, REGRESS in MATLAB. As suggested by Longstaff and Schwartz (2001), we only include the in-the-money paths in the cross sectional regression for efficiency. To ensure the accuracy of the estimation of continuation values, no variance reduction technique is employed to reduce the standard errors of the estimations.¹⁰ In Examples 1, 2 and 5, for MLSM, the basis functions are : $1, X$ and X^2 ; for OLSM, the basis functions are : $1, X_i$ and X_i^2 , $i = 1, \dots, n$ for a n -dimensional American arithmetic average option¹¹. The basis functions will be specified for other examples.

Example 1: American arithmetic average options on 1 to 6 assets.

This example is taken from Kovalov et al (2007) where the price of each underlying stock follows a GBM with model parameters: the initial price $S_{i0} = 100, i = 1, \dots, 6$, the strike price $K=100$, the interest rate $r=3\%$, no dividend, the maturity $T=0.25$ year, the volatility $\sigma_{i1} = 20\%, i = 1, \dots, 6$, the correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 6$, the other simulation parameters are $N = 50$, $M_0 = 1000$ and $M = 10000$. The FEM values obtained by their Finite-Element-Method are quoted as benchmark values. The LB and UB represent the lower and upper bounds respectively. The CPU time is the time elapsed for each approximation except for those under MSLM where the CPU time is the total time consumed for constructing both lower and upper bounds. For simplicity and

¹⁰We thank an anonymous referee for his or her comments.

¹¹Here we do not include the cross product terms: $X_i X_j, i \neq j = 1, \dots, n$ for OLSM. We have two reasons for this. First we find in our numerical examples that the estimates of lower bounds by including the cross product terms are much lower with much larger standard deviations than those obtained without the cross product terms. Second and more importantly, incorporating the cross product terms into the set of basis functions suffers from the curse of dimensionality. In addition, we increase the order of basis function to four and find that the results are very similar. For demonstration purpose, we stick to this basis function throughout all examples.

consistency, We keep this set of parameters as the base parameters throughout all examples if not otherwise mentioned.

The simulation results of OLSM and MLSM are summarized in Table 1. Apparently, the lower bounds obtained by OLSM approaches are lower than the true values as expected since estimated exercise strategies are suboptimal. However, in four cases, the lower bounds obtained by MLSM are higher than those by OLSM, which means MLSM provides better lower bounds to OLSM. For example, in the 6-asset case, the estimated values of MLSM and OLSM are 2.71663 and 2.71073 respectively and the former is closer to the benchmark value, 2.71838.

Apart from lower bounds, MLSM also generates fast and good upper bounds¹². We use Mid price defined by $(\text{upper bound} + \text{lower bound})/2$ as the estimator for price of an American arithmetic average option. To gauge the accuracy of MLSM, the pricing error is calculated as $\frac{|\text{Midprice} - \text{Benchmark}|}{\text{Benchmark}} \times 100$ and the results of the pricing errors are reported in the P. E. column. The P. E. column illustrates that the estimated prices obtained by MSLM are comparable to the benchmarks with pricing error up to 0.869% for 6-asset case.

In the examples below, we test the pricing accuracy of our methods by considering American arithmetic average option with various settings. As in Example 1, we use Mid price defined by $(\text{upper bound} + \text{lower bound})/2$ as the estimator for price of an American arithmetic average option. Unlike Example 1, we use the quantity $[|\text{Midprice} - \text{lowerbound}| / \text{lowerbound}] \times 100\%$ to measure the performance of the price estimator because there is no benchmark available. Further, it is worth mentioning that this quantity is a conservative estimator or an upper bound¹³ of the true pricing error given by $\frac{|\text{Midprice} - \text{trueprice}|}{\text{trueprice}} \times 100\%$ because $\text{lowerbound} \leq \text{trueprice}$,

$$|\text{Midprice} - \text{trueprice}| \leq \frac{|\text{upperbound} - \text{lowerbound}|}{2},$$

¹²It is computationally demanding to construct upper bounds for OLSM based on the duality method and thus the upper bounds for OLSM are not constructed here. For general upper bound construction techniques, see Hough and Kogan (2004) and Anderson and Broadie (2004) and among others.

¹³We thank an anonymous referee for suggesting us to clarify this property.

and thus

$$\begin{aligned}
& \frac{|Midprice - trueprice|}{trueprice} \times 100\% \\
& \leq \frac{|upperbound - lowerbound|}{2 \times lowerbound} \times 100\% \\
& = \frac{[Midprice - lowerbound]}{lowerbound} \times 100\%, \tag{2.12}
\end{aligned}$$

the last equality following from the definition of Mid price.

As a result, the MLSM algorithm performs better than what are indicated by the reported upper bounds of pricing errors.

Example 2. American arithmetic average options on 10, 30 and 50 underlying assets following MJD processes.

To further demonstrate the performance of MLSM, we consider American arithmetic average options with underlying assets following the Merton-Jump-Diffusion process, where the jumps are assumed to be co-jump as conventionally adopted in the literature. In particular, the co-jump component parameters are set to be $\lambda = 5$, $J \sim N(-0.1, 0.1)$ and $\sigma_{i2} = 1$ for all $i = 1, \dots, n$.¹⁴ Moreover, we increase the number of underlying assets to 10, 30 and 50 with parameters of diffusion components remaining the same as in Example 1. If not otherwise mentioned, the we stick to the same parameters as in Example 1 for consistency. The results are presented in Table 2.2, where Panel 1 and 2 summarize results simulated with $M = 10000$ and $M = 1000$, respectively¹⁵.

As Panel 1 illustrates, MLSM generates good approximations for the true values with upper bounds of pricing errors (UB of P. E.) up to 0.278%. By contrast, the lower bounds obtained by OLSM are consistently lower than those by MLSM. In particular, it is approximately 5% lower for the 50-asset case. Moreover, MLSM provides both lower and upper bounds with the similar computational time. Besides, Panel 2 indicates the efficiency of MSLM, in terms of the MC paths required. More specifically, with 1/10 MC paths, MLSM generates quantitatively similar results with around 3 times larger standard deviation in comparison with those in Panel 1. OLSM, however, provides even 50% smaller lower bound for 50-asset case relative to the one obtained by MLSM. The reason for this is that the geometric average approximate the arithmetic average

¹⁴For illustrative purpose, we set $\sigma_{i2} = 1$ for all $i=1, \dots, n$. We also tried the cases where the underlying assets are allowed to have different sensitivities to the co-jump, i.e. σ_{i2} varies across different underlying assets. Consistently, MLSM yields good results.

¹⁵We thank an anonymous referee for suggesting us to use smaller numbers of simulated paths to test the efficiency of our method.

accurately and, compared to OLSM, MLSM has only three parameters to be estimated for determining the exercise value the American geometric average option at each step.

Example 3. Simultaneous perturbation stochastic approximation (SPSA) benchmarking

The upper bounds implied by MLSM hinges on the key inequality, (2.11), where the latter term is essentially an American geometric average option and estimated by MLSM in the numerical experiments. Essentially, this American geometric average option would be estimated accurately by MLSM or OLSM since it is a 1-dimensional problem if stock prices are the only state variables. However, to further demonstrate the applicability and accuracy of the estimation for this American geometric average option, we apply the Simultaneous perturbation stochastic approximation (SPSA)¹⁶ for Example 1 as benchmark.¹⁷ The parameters remains the same as in Example 1 except N which is set to be 4 for illustrative purpose.¹⁸ The set of SPSA parameters are: $a_k = \frac{100}{(10+k)^{0.751}}$ and $c_k = \frac{10}{k^{0.25}}$, where k denotes the kth iteration of total 500 iterations. (For details about the SPSA algorithm and the parameters, see Fu et al (2000), Spall (1998) and among others.)

Table 2.3 indicates the accuracy of the MSLM algorithm. The $SPSA_{GA}$ value column summarizes the American geometric average option values obtained by SPSA algorithm, while American GA column shows the geometric average option values obtained by MLSM. As both columns illustrate, the values differ only in the third decimal points in all cases, which indeed imply the accuracy and validity of applying MSLM for estimating the American geometric average option for constructing the MLSM upper bound. Besides, $SPSA_{AA}$ value column reports the values of American arithmetic average options obtained by SPSA algorithm. As the results indicate, SPSA algorithm provides reliable benchmark for American arithmetic average options with low number of exercise opportunities.

¹⁶Please refer to Spall (2012) for details about SPSA algorithm

¹⁷We thank an anonymous referee for the comments about applying the SPSA algorithm.

¹⁸Theoretically, we can choose N=50 as in other examples. However, the SPSA algorithm would deal with an optimization problems with 50 parameters, which may be not easy to get accurate estimates for the 50 parameters. As a result, we choose N=4 for demonstration. Moreover, SPSA could be applied as benchmark for the American arithmetic average options especially for less-exercise-opportunity cases. More specifically, SPSA is essentially an optimization problem and the number of parameters depends on exercise opportunities when it is applied for option pricing.

Example 4. American arithmetic average options on 30 underlying assets with 20 long and 10 short positions.

As mentioned in Section 2.1, the MLSM algorithm also applies to American arithmetic average options with both positive and negative weights. We provide a concrete example here for completeness.¹⁹

More specifically, the exercise value of this option at time $t \in \Gamma$ can be expressed as

$$\sum_{i=1}^{30} a_i S_{it} = \sum_{i=1}^{20} a_i S_{it} - \sum_{i=21}^{30} (-a_i) S_{it}$$

The two arithmetic means are then approximated by their geometric means, denoted as GA_t^+ and GA_t^- respectively, as illustrated in Step 1. Following this, the dimension of the pricing problem is reduced to two for MLSM algorithm. Accordingly, the cross sectional regression for MLSM is now applied with basis functions $1, GA_t^+, GA_t^-, GA_t^+ \times GA_t^-, (GA_t^+)^2, (GA_t^-)^2$ here, while OLSM is implemented as in other examples. With respect to the parameters, they are kept the same as in Example 1 except that we increase the number of underlying assets to 30, where 20 assets are allocated with positive weights and the others are with negative weights and vary the strike price, K , from 34 to 37.²⁰ The results are summarized in Table 2.4.

As Panel 1 shows, MLSM generates higher lower bounds in comparison with OLSM and the accuracy of MLSM are reflected by the upper bounds of pricing errors, which are below 1% except for $K=34$. Moreover, the simulation is repeated with $M = 1000$ and the results are reported in Panel 2. Consistent with previous examples, the efficiency of MSLM in terms of MC paths required is preserved for the case with both positive and negative weights. In contrast, OLSM generates significantly smaller lower bounds (around 11.5% lower as $K=34$).

Example 5. American arithmetic average options on 20 and 30 underlying assets with different volatilities.

In this example, to show that MLSM is applicable in general, we consider two models of 20 and 30 underlying assets with prices following GBM with variety in volatilities. The closeness between arithmetic and geometric averages de-

¹⁹We thank an anonymous referee for pointing out this to improve the completeness of the paper.

²⁰We choose 34 for illustration since it is the smallest integer strike price where the option is in the money

depends on volatility. To examine the effect of volatility on pricing accuracy, we change the volatilities of the underlying assets with other parameters remaining the same as in previous examples. The results obtained with $N=50$ and $M_0 = 1000$, $M = 10000$ are summarized in Table 2.5. In Panel 1, the volatilities are set to be $\sigma_{i1} = 0.15, i = 1, \dots, 10; \sigma_{i1} = 0.2, i = 11, \dots, 20$ for 20-asset case and $\sigma_{i1} = 0.1, i = 1, \dots, 10; \sigma_{i1} = 0.15, i = 11, \dots, 20, \sigma_{i1} = 0.2, i = 21, \dots, 30$ for 30-asset case.

As this table indicates, OLSM provides lower bounds around 1% lower than MLSM. The accuracy of MLSM is reflected by the upper bounds of pricing errors around 0.6% for both cases. Moreover, when the volatility structures among the underlying assets are changed to $\sigma_{i1} = 0.15, i = 1, \dots, 10; \sigma_{i1} = 0.3, i = 11, \dots, 20$ for 20-asset case and $\sigma_{i1} = 0.15, i = 1, \dots, 10; \sigma_{i1} = 0.2, i = 11, \dots, 20, \sigma_{i1} = 0.3, i = 21, \dots, 30$ for 30-asset case with other parameters unchanged to allow more variations in the price processes, Panel 2 shows that MLSM consistently generate higher lower bounds with small upper bounds of pricing errors around 1.067% and 0.821% for both cases respectively.

Similarly, the simulations are repeated with $M = 1000$ and summarized in Panel 3 and 4 to show the convergency speed of MLSM in terms of MC paths required. Both Panel 3 and 4 show that the results obtained by MLSM are quantitatively the same; however, those obtained by OLSM are significantly underestimated. In the worst case (the last row), the lower bound (2.282) is approximately 21% lower than the one obtained by MLSM (2.765).

Example 6. American arithmetic average options on 10 and 20 underlying assets following Heston's model.

In this example, we consider Heston's model (1993) of stochastic volatility to demonstrate the performance of MLSM for pricing American arithmetic average options. For the stock i , its price S_{it} and variance σ_{it}^2 are given by:

$$\begin{aligned} dS_{it} &= rS_{it}dt + \sigma_{it}dW_{it}^s, \\ d\sigma_{it}^2 &= \alpha(\beta - \sigma_{it}^2)dt + \gamma\sigma_{it}dW_{it}^v, i = 1, \dots, n, \end{aligned}$$

where $W_{it}^v, i = 1, \dots, n$ are independent Brownian motions, that is, volatility processes are independent; for the stock i , the Brownian motions W_{it}^s and W_{it}^v are correlated with coefficients, $\rho_i, i = 1, \dots, n$, capturing leverage effect; for $i \neq j$, W_{it}^s and W_{jt}^s are correlated with a coefficient ρ_{ij} , namely, the stock returns are correlated.

In this case, a geometric mean of the stock prices depends on the stochastic

volatility processes as state variables, which should be included in the cross sectional regression. Considering the case of 10 stocks, a geometric mean, denoted as GA_t , depends on 11 state variables, that is, the geometric mean itself and the 10 stochastic volatility processes. In contrast, the original arithmetic mean depends on 20 state variables, that is, the 10 stock price processes and the 10 stochastic volatility processes. More specifically, the basis functions adopted in this example for MLSM are: $1, GA_t, GA_t^2, \sigma_{it}, \sigma_{it}^2, i = 1, \dots, 10/20$; for OLSM: $1, S_{it}, S_{it}^2, \sigma_{it}, \sigma_{it}^2, i = 1, \dots, 10/20$.

In this exercise, we only present lower bound estimates since the true values of American geometric average options are not easy to obtained for upper bounds due to higher dimension. The simulation results are summarized in Table 2.6. Compared with OLSM, MLSM still provides an effective way to estimate lower bounds. As Panel 1 shows, MLSM generate higher lower bounds for both cases compared to those obtained from OLSM. Moreover, Panel 2 further manifests the performance of MLSM by reducing M to 2500.²¹ More concretely, the results obtained by MLSM are quantitatively similar to those in Panel 1. In contrast, OLSM generates significantly lower estimates compared with the results as $M = 10000$.

To show the results are not subject to the specific second-order basis function adopted, we repeat the simulation by increasing the order of basis functions to four and summarize the results in Panel 3.²² As Panel 3 demonstrates, MLSM generates consistent results as in Panel 1; however, the estimates by OLSM deteriorate around 1 to 2 %.

Example 7. Convergence comparisons of MLSM and OLSM.

In this exercise, we repeat the 6-asset case in Example 1 to examine the rate of convergence for OLSM and MLSM by varying the number of simulation paths from 500 to 10000. The results are displayed in Figure 2.2. The lower bounds estimated by MLSM are consistently higher than those by OLSM. Moreover, the estimates obtained by the former start to approach the benchmark value (2.71838) essentially bounded below after the number of simulation paths exceeds 1000. By contrast, the lower bound estimates from OLSM converge much slower than the MLSM estimates. In particular, 10000 simulation paths are

²¹Since the number of state variables increases as volatilities are allowed to be stochastic, the pricing problem for MSLM is not one dimensional for this example anymore. As a result, it requires more MC paths (2500) in comparison with 1000 paths as analyzed in the previous examples.

²²The results obtained with order of basis function to 4 are reported in this example only for demonstration. With respect to other experiments, we can get very similar results.

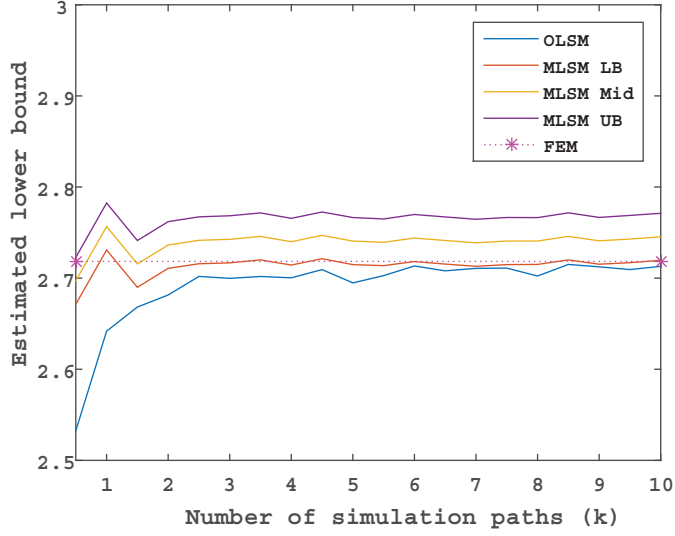


Figure 2.2: Comparison of convergence speeds of MLSM and OLSM for pricing American arithmetic average option on 6 assets

sufficient for a good lower bound estimate. It is also noteworthy that the upper bound estimate from MLSM converges very fast too, in the sense that it converges to the true upper bound implied by MLSM as indicated by Figure 2.2. The desirable efficiency of MLSM arises from two key facts. Firstly, the arithmetic mean can be accurately approximated with the corresponding geometric mean. Secondly, the pricing problem for MLSM is reduced to lower dimension and or even 1 dimension for the case where the underlying assets follow the deterministic GBM process in particular.

Example 8. Robustness test for MLSM and OLSM with SPSA as benchmark²³

As demonstrated in (2.7), the MLSM is especially efficient when the volatility and the time-to-maturity are not very high. In this example, we test performance of our MLSM in cases where volatility and/or time-to-maturity are relatively high.²⁴ More specifically, American arithmetic average options on 10 assets with the time-to-maturity and the volatility ranging from 0.5 years to 5 years and 0.15 to 0.5, respectively, are investigated with other parameters fixed as in Example 3.²⁵ SPSA is adopted as a reasonable benchmark, i.e., true value,

²³FEM method is a PDE-based algorithm and is hard to get accurate prices in high dimensional cases. As a result, we apply SPSA as the main benchmark here as discussed in Example 3.

²⁴We are very grateful to an anonymous referee for suggesting this test.

²⁵The ranges of the time-to-maturity and the volatility are common in EQ and FX markets. We thank an anonymous referee for pointing out this to us.

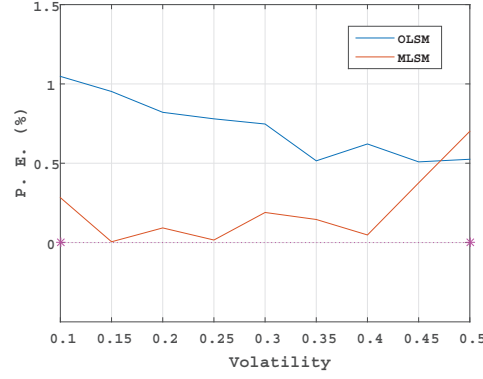
here as justified in Example 3.

First, we investigate the effect of high-volatility case by fixing the level of volatility at 0.5 and vary the time-to-maturity from 0.5 years to 5 years. Then, we fix the time-to-maturity at 5 years and vary the volatility from 0.15 to 0.5 to illustrate the effect of long time-to-maturity. The results of both cases are summarized in the upper and bottom panels of Table 2.7, respectively. As summarized in the table, MLSM provides lower lower bounds relative to the OLSM estimators in high-volatility and long time-to-maturity cases and higher upper bounds of pricing errors relative to those in Examples 1 to 7. In particular, the upper bounds of pricing errors are as high 4.2856% when the volatility is 0.5 and the time-to-maturity is equal to 5 years. However, as the column under “P. E.” indicates that MLSM provides good estimators given by the mid prices for American arithmetic average options in terms of pricing errors with SPSA estimators as benchmark. This finding is consistent with our result (2.12) that the upper bounds of pricing errors are conservative estimators of true pricing errors.

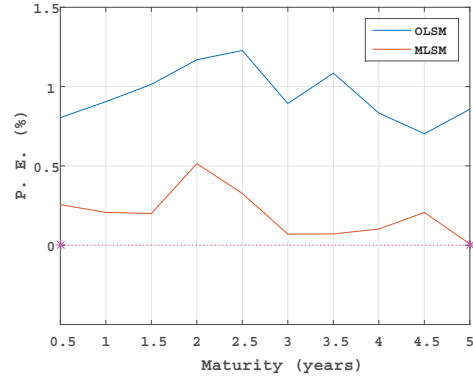
Furthermore, we test the performance of MLSM by varying volatility, time-to-maturity and strike price, respectively.²⁶ In the following tests plotted in Figure 3, we graphically illustrate the pricing errors (P.E.) with SPSA estimators as benchmark and number of simulation path, $M=10000$. Specifically, in Panel (a), we vary volatility with time-to-maturity fixed at 1 year and other parameters fixed as in Example 3; in Panel (b), we vary time-to-maturity with volatility fixed at 0.2 and other parameters fixed as in Example 3; in Panel (c), with time-to-maturity fixed at 1 year, volatility fixed at 0.5 and other parameters fixed as in Example 3, we test performance of MLSM against different strike prices by scaling the strike price by $\exp(rT + a\sqrt{T})$ with a ranging from $-1/4$ to $1/4$. As Figure 2.3 demonstrates, reflected by “P. E.”, mid prices obtained by MLSM are good estimators for American arithmetic average options among a reasonable range of time-to-maturities, volatilities, and strike prices.

²⁶We are very grateful to an anonymous referee for suggesting this test.

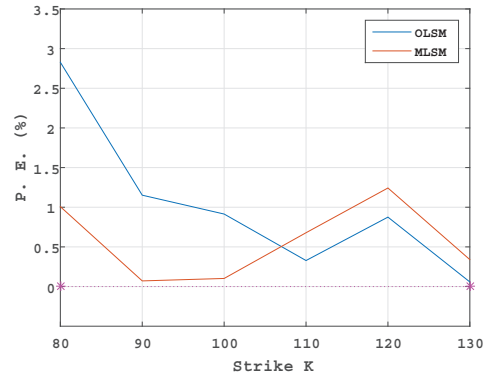
In short, the above numerical experiments illustrate the robustness and limitation of the MLSM algorithm for pricing high dimensional American arithmetic average options. On the one hand, MLSM can generate good lower bound estimates with faster convergence as opposed to OLSM. On the other hand, the performance of MLSM in terms of the upper bound of pricing error may deteriorate in the cases where volatility is very high and/or time-to-maturity is very long.



(a)



(b)



(c)

Figure 2.3: Effects of volatility, time-to-maturity and strike price on pricing error. All stocks have the same setting as Example 2.3, i.e., Common Initial Price $S_{i0} = 100$, Strike Price $K=100$, Interest rate $r=3\%$, no Dividend, Maturity $T=1$ Year, Common Volatility $\sigma_{i1} = 20\%$, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 10$, if not otherwise mentioned. In particular, the pricing error is calculated with SPSA estimators as benchmark and $M=10000$.

Number of Stocks	FEM True Value	CPU		OLSM		P. E.(%)		CPU		MLSM		P. E.(%)		CPU	
		Time(sec)		LB		LB		Time(sec)		LB		UB		Mid	Time(sec)
1	3.6678	0.04		3.6560 (0.0432)		3.6533 (0.0466)		0.3215	0.4330	3.6533 (0.0466)		3.6533 (0.0466)		3.6533	0.7518
2	3.1396	55.03		3.1168 (0.0350)		3.1264 (0.0409)		0.7262	0.6540	3.1264 (0.0409)		3.1802 (0.0412)		3.1533	1.0317
3	2.9445	20.86		2.9187 (0.0374)		2.9312 (0.0334)		0.8767	0.8652	2.9312 (0.0334)		2.9890 (0.0336)		2.9601	1.2755
4	2.8402	7932.63		2.8201 (0.0358)		2.8197 (0.0366)		0.7076	1.1299	2.8197 (0.0366)		2.8758 (0.0368)		2.8478	1.5178
5	2.7719	559.79		2.7559 (0.0337)		2.7584 (0.0378)		0.5772	1.3916	2.7584 (0.0378)		2.8117 (0.0381)		2.7850	1.7714
6	2.7184	26299.28		2.7107 (0.0368)		2.7166 (0.0389)		0.2811	1.7472	2.7166 (0.0389)		2.7674 (0.0391)		2.7420	2.0825

The FEM values are given in the Kovalov et al (2007) solved by PDE method. All stocks have Common Initial Price $S_{i0} = 100$, Strike Price $K=100$, Interest rate $r=3\%$, no Dividend, Maturity $T=0.25$ Year, Common Volatility $\sigma_{i1} = 20\%$, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 6$. Approximate option values are obtained by applying OLSM and MLSM approaches 100 times with 50 equally spaced Exercisable Times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM respectively. The reported values are averages over the 100 runs. LB and UB represent lower and upper bounds respectively. The upper bounds are calculated as described in the upper bound section and the Mid price is defined as the average of the estimated lower and upper bounds. The P. E. is the pricing error. In particular, the reported P. E. for OLSM is calculated by $\frac{|OLSM - Benchmark|}{Benchmark} \times 100$ since the upper bound of OLSM is not available and not the main issue of this paper. The standard errors of the approximate values are shown in the parentheses.

Table 2.1: American Arithmetic Average Options on 1 to 6 underlying stocks.

Panel 1							
Number of Stocks	OLSM LB	CPU Time(sec)	LB	UB	MLSM Mid	UB of P. E.(%)	CPU Time(sec)
10	6.20357 (0.08735)	51.44967	6.26185 (0.09309)	6.29676 (0.09329)	6.27930	0.27873	52.58328
30	6.05943 (0.08261)	163.19413	6.22653 (0.09527)	6.24815 (0.09533)	6.23734	0.17354	163.52299
50	5.90727 (0.08915)	288.28503	6.23717 (0.08713)	6.25419 (0.08721)	6.24568	0.13643	287.32009
Panel 2							
Number of Stocks	OLSM LB	CPU Time(sec)	MLSM LB	UB	Mid	UB of P. E.(%)	CPU Time(sec)
10	5.65246 (0.27316)	5.75452	6.20944 (0.25329)	6.24411 (0.25372)	6.22678	0.27919	7.33546
30	4.54559 (0.27593)	15.51786	6.16732 (0.26138)	6.18890 (0.26158)	6.17811	0.17491	22.06787
50	4.00669 (0.23942)	26.48666	6.15054 (0.25016)	6.16748 (0.25037)	6.15901	0.13771	37.43104

The underlying stocks are modeled to follow MJD process with co-jump component parameters: $\lambda = 5$, $J \sim N(-0.1, 0.1)$. The other parameters in the models are : Common Initial prices $S_{i0} = 100$, Strike Price $K=100$, Interest Rate $r=3\%$, no Dividend, Common volatilities $\sigma_{i1} = 20\%$, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 10/30/50$ and maturity, T is set to be 0.25 Year respectively. The jump sensitivities, σ_{i2} to the jump are set to be 1 for all underlying assets.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times for each case with 50 equally spaced Exercisable times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM, respectively in Panel 1. In contrast, the simulation is repeated with $M = 1000$ to examine the efficiency against the number of the Monte Carlo paths. The results are reported in the Panel 2. The reported values are averages over the 100 runs. LB and UB represent lower and upper bounds, respectively. The upper bounds are calculated as described in the upper bound section and the Mid is defined as the average of the estimated lower and upper bounds. The UB of P. E. is the upper bound of pricing error. The standard errors of the approximate values are shown in the parentheses.

Table 2.2: American Arithmetic Average Option on 10, 30 and 50 underlying stocks following MJD process.

n	<i>SPSA_{AA}</i>		CPU		OLSM		P. E.(%)		CPU		MLSM		<i>SPSA_{GA}</i>		American		P. E.(%)		CPU	
	Value	(sec)	LB	(0.0489)	LB	(0.0398)	0.1747	(0.0132)	3.6543	(0.0517)	LB	(0.0380)	UB	(0.0439)	GA	(0.0517)	0.0063	(0.0063)	(sec)	(sec)
1	3.6536 (0.0479)	1.2058	3.6481 (0.0489)	3.1056 (0.0398)	3.6543 (0.0517)	3.1158 (0.0380)	0.1747 (0.0132)	0.0132 (0.0517)	3.6543 (0.0517)	0.0189 (0.0439)	3.6543 (0.0517)	0.0189 (0.0439)	3.6543 (0.0517)	0.0189 (0.0439)	3.6543 (0.0517)	0.0189 (0.0439)	0.0063 (0.0063)	0.0063 (0.0063)	0.0968 (0.0968)	0.0968 (0.0968)
2	3.1187 (0.0449)	1.2188	3.1056 (0.0398)	3.1056 (0.0398)	3.1158 (0.0380)	3.1158 (0.0380)	0.3148 (0.0323)	0.0323 (0.0383)	3.1548 (0.0383)	0.5321 (0.0383)	3.1154 (0.0419)	0.5321 (0.0384)	3.1154 (0.0419)	0.5321 (0.0384)	3.1145 (0.0384)	0.5321 (0.0384)	0.0274 (0.0274)	0.0274 (0.0274)	0.2180 (0.2180)	0.2180 (0.2180)
3	2.9274 (0.0370)	1.3481	2.9125 (0.0362)	2.9125 (0.0362)	2.9205 (0.0396)	2.9205 (0.0396)	0.2978 (0.0600)	0.0600 (0.0397)	2.9610 (0.0397)	0.4576 (0.0398)	2.9212 (0.0500)	0.4576 (0.0398)	2.9212 (0.0500)	0.4576 (0.0398)	2.9197 (0.0398)	0.4576 (0.0398)	0.0497 (0.0497)	0.0497 (0.0497)	0.2748 (0.2748)	0.2748 (0.2748)
4	2.8184 (0.0373)	1.2834	2.8072 (0.0345)	2.8072 (0.0345)	2.8192 (0.0351)	2.8192 (0.0351)	0.4226 (0.1036)	0.0782 (0.1036)	2.8579 (0.0353)	0.7134 (0.0353)	2.8191 (0.0348)	0.7134 (0.0353)	2.8191 (0.0348)	0.7134 (0.0353)	2.8187 (0.0350)	0.7134 (0.0350)	0.0157 (0.0157)	0.0157 (0.0157)	0.2901 (0.2901)	0.2901 (0.2901)
5	2.7552 (0.0316)	1.3334	2.7458 (0.0364)	2.7458 (0.0364)	2.7543 (0.0433)	2.7543 (0.0433)	0.3926 (0.1060)	0.1036 (0.1060)	2.7906 (0.0435)	0.6246 (0.0435)	2.7566 (0.0372)	0.6246 (0.0435)	2.7566 (0.0372)	0.6246 (0.0435)	2.7533 (0.0434)	0.6246 (0.0434)	0.1187 (0.1187)	0.1187 (0.1187)	0.3178 (0.3178)	0.3178 (0.3178)
6	2.7103 (0.0462)	1.3814	2.7059 (0.0391)	2.7059 (0.0391)	2.7113 (0.0341)	2.7113 (0.0341)	0.1959 (0.0341)	0.1060 (0.0341)	2.7455 (0.0340)	0.6666 (0.0340)	2.7112 (0.0346)	0.6666 (0.0340)	2.7112 (0.0346)	0.6666 (0.0340)	2.7103 (0.0392)	0.6666 (0.0392)	0.0331 (0.0331)	0.0331 (0.0331)	0.2867 (0.2867)	0.2867 (0.2867)

The *SPSA* value is calculated by a parametrized iteration method, simultaneous perturbation stochastic approximation(*SPSA*) and works as a benchmark for completeness. All stocks have the same setting as Example 2.1, i.e. Common Initial Price $S_0 = 100$, Strike Price $K=100$, Interest rate $r=3\%$, no Dividend, Maturity $T=0.25$ Year, Common Volatility $\sigma_{i1} = 20\%$, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 6$, except that $N = 4$ for illustrative purpose.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times with 50 equally spaced Exercisable Times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM respectively. The reported values are averages over the 100 runs. *SPSA_{AA}* and *SPSA_{GA}* represent the values of American arithmetic and geometric average options obtained by *SPSA* algorithm, respectively. LB and UB represent lower and upper bounds respectively. The upper bounds are calculated as described in the upper bound section and the Mid price is defined as the average of the estimated lower and upper bounds. The P. E. is the pricing error. In particular, the second "P. E.(%)" column is the pricing error for our MLSM technique, while the third "P. E.(%)" column is the pricing error for the American geometric average option. The standard errors of the approximate values are shown in the parentheses.

Table 2.3: American Arithmetic Average Options on 1 to 6 underlying stocks. Verifying the accuracy of MLSM upper bound with *SPSA* approach.

Panel 1					
Strike	LSM		MLSM		UB of
Price	LB	LB	UB	Mid	P. E.(%)
34	1.30585 (0.01544)	1.31802 (0.01356)	1.34946 (0.01364)	1.33374	1.19289
35	1.96168 (0.01525)	1.97742 (0.01411)	2.01409 (0.01414)	1.99575	0.92709
36	2.75018 (0.01431)	2.77325 (0.01685)	2.81316 (0.01699)	2.79320	0.71944
37	3.65444 (0.01288)	3.67401 (0.01365)	3.71595 (0.01376)	3.69498	0.57079
Panel 2					
34	1.16107 (0.04149)	1.29531 (0.04743)	1.32679 0.04763	1.31105	1.21501
35	1.83902 (0.04853)	1.97582 (0.04697)	2.01240 0.04721	1.99411	0.92558
36	2.68046 (0.04248)	2.76171 (0.04859)	2.80173 0.04864	2.78172	0.72459
37	3.61886 (0.04049)	3.65690 (0.04741)	3.69850 0.04747	3.67770	0.56879

In this table, the applicability of MLSM for portfolios composed of both long and short positions, i.e. positive and negative weights is examined. In particular, $a_i = 1/30$, for $i = 1, \dots, 20$ and $a_i = -1/30$, for $i = 21, \dots, 30$. The parameters in the model are : Common Initial prices $S_{i0} = 100$, $i = 1, \dots, 30$, Interest Rate $r=3\%$, no Dividend, Maturity $T=0.25$ Year, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 30$. The results are summarized as $K=34, \dots, 37$ for illustration. The computational times for LSM and MLSM are around 20 and 18, respectively and not reported in the table for illustrative brevity.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times for each case with 50 equally spaced Exercisable times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM respectively in Panel 1. In contrast, the simulation is repeated with $M = 1000$ to examine the efficiency against the number of the Monte Carlo path. The results are reported in the Panel 2. The reported values are averages over the 100 runs. LB and UB represent lower and upper bounds respectively. The upper bounds are calculated as described in the upper bound section and the Mid is defined as the average of the estimated lower and upper bounds. The UB of P. E. is the upper bound of pricing error. The standard errors of the approximate values are shown in the parentheses.

Table 2.4: American Arithmetic Average Option on 30 assets with 20 long positions and 10 short positions.

Panel 1							
Number of Stocks	OLSM LB	CPU Time(sec)	LB	MLSM UB	Mid	UB of P. E.(%)	CPU Time(sec)
20	2.19438 (0.02820)	9.21937	2.20738 (0.02664)	2.23399 (0.02672)	2.22069	0.60256	8.86362
30	1.81391 (0.02481)	17.86448	1.83744 (0.02260)	1.85792 (0.02269)	1.84768	0.55748	17.20838
Panel 2							
Number of Stocks	OLSM LB	CPU Time(sec)	LB	MLSM UB	Mid	UB of P. E.(%)	CPU Time(sec)
20	2.90650 (0.03414)	8.38199	2.91865 (0.04189)	2.98095 (0.04207)	2.94980	1.06723	8.05845
30	2.74234 (0.03740)	17.71292	2.77007 (0.03568)	2.81554 (0.03587)	2.79280	0.82060	17.05103
Panel 3							
Number of Stocks	OLSM LB	CPU Time(sec)	LB	MLSM UB	Mid	UB of P. E.(%)	CPU Time(sec)
20	1.93455 (0.07995)	0.69619	2.20587 (0.08961)	2.23243 (0.08985)	2.21915	0.60212	0.89514
30	1.48652 (0.07379)	1.22382	1.82517 (0.07785)	1.84576 (0.07825)	1.83547	0.56410	1.59656
Panel 4							
Number of Stocks	OLSM LB	CPU Time(sec)	LB	MLSM UB	Mid	UB of P. E.(%)	CPU Time(sec)
20	2.55810 (0.13629)	0.69891	2.91531 (0.13169)	2.97758 (0.13242)	2.94645	1.06807	0.88955
30	2.28217 (0.10487)	1.19687	2.76481 (0.09261)	2.81017 (0.09287)	2.78749	0.82029	1.62135

The parameters in the models are : Common Initial prices $S_{i0} = 100$, $i = 1, \dots, 20/30$, Strike Price $K=100$, Interest Rate $r=3\%$, no Dividend, Maturity $T=0.25$ Year, Common Correlation $\rho_{ij} = 0.5$, $i \neq j = 1, \dots, 20/30$.

The underlying stocks have different volatilities as $\sigma_{i1} = 0.15$, $i = 1, \dots, 10$, $\sigma_{i1} = 0.2$, $i = 11, \dots, 20$, for 20-stock case and $\sigma_{i1} = 0.10$, $i = 1, \dots, 10$, $\sigma_{i1} = 0.15$, $i = 11, \dots, 20$, $\sigma_{i1} = 0.2$, $i = 21, \dots, 30$, for 30-stock case in the upper panel and $\sigma_{i1} = 0.15$, $i = 1, \dots, 10$, $\sigma_{i1} = 0.3$, $i = 11, \dots, 20$, for 20-stock case and $\sigma_{i1} = 0.15$, $i = 1, \dots, 10$, $\sigma_{i1} = 0.2$, $i = 11, \dots, 20$, $\sigma_{i1} = 0.3$, $i = 21, \dots, 30$, for 30-stock case in the bottom panel.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times for each case with 50 equally spaced Exercisable times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM in Panel 1 and 2 respectively. In contrast, the simulations for both cases are repeated with $M = 1000$ to examine the convergency against the number of the Monte Carlo path. The results are reported in Panel 3 and 4. The reported values are averages over the 100 runs. LB and UB represent lower and upper bounds respectively. The upper bounds are calculated as described in the upper bound section and the Mid is defined as the average of the estimated lower and upper bounds. The UB of P. E. is the upper bound of pricing error. The standard errors of the approximate values are shown in the parentheses.

Table 2.5: American Arithmetic Average Option on 20 and 30 underlying stocks with different volatilities.

Panel 1				
Number of Stocks	OLSM LB	CPU Time(sec)	MLSM LB	CPU Time(sec)
10	4.61323 (0.05576)	6.45618	4.62770 (0.06643)	6.62625
20	4.44706 (0.05726)	18.25560	4.49729 (0.06253)	18.05300
Panel 2				
Number of Stocks	OLSM LB	CPU Time(sec)	MLSM LB	CPU Time(sec)
10	4.46118 (0.11368)	1.29819	4.58711 (0.12850)	1.49783
20	4.11185 (0.11160)	3.21959	4.37469 (0.11633)	3.62714
Panel 3				
Number of Stocks	OLSM LB	CPU Time(sec)	MLSM LB	CPU Time(sec)
10	4.560722 (0.05581)	9.858611	4.601852 (0.05498)	8.666284
20	4.334726 (0.05590)	25.26982	4.439366 (0.05868)	21.75666

In this table, the applicability of MLSM for stochastic volatility models is examined and for illustration, Heston model is adopted. The parameters in the model are : Common Initial prices $S_{i0} = 100$, $i = 1, \dots, 10/20$, Interest Rate $r=3\%$, no Dividend, Maturity $T=0.25$ Year, Common Correlation $\rho_{ij} = 0.5, i \neq j = 1, \dots, 10/20$. In particular, the parameters of the Heston model are: $\sigma_{i0} = 0.1, \alpha = 0.5, \beta = 0.16, \gamma = 0.9, \rho_i = -0.1, i = 1, \dots, 10/20$.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times for each case with 50 equally spaced Exercisable times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM respectively in Panel 1. In contrast, the simulation is repeated with $M = 2500$ to examine the convergency against the number of the Monte Carlo path. Moreover, consistency is examined by extending cross sectional regression basis, i.e., state variables to the fourth power. The results are reported in the Panel 2 and Panel 3, respectively. The reported values are averages over the 100 runs. LB represents lower bound. The standard errors of the approximate values are shown in the parentheses.

Table 2.6: American Arithmetic Average Option on 10 and 20 assets with Heston's model of stochastic volatility.

Panel 1		Vol	0.5											
T-M (Years)	SPSA Value	CPU(sec)	OLSM		P.E.(%)	CPU(sec)	LB	MLSM		Mid	UB of		P.E.(%)	CPU(sec)
0.5	9.7336 (0.1240)	1.4086	9.6753 (0.1135)	0.5989	0.2158	9.6336 (0.1234)	9.9655 (0.1256)	9.7996	1.7227	0.6778	0.4361			
3	20.9719 (0.1877)	1.4087	20.8878 (0.2067)	0.4008	0.1842	20.2533 (0.2131)	21.7331 (0.2557)	20.9932	3.6531	0.1017	0.4543			
5	25.2607 (0.2234)	1.0088	25.2101 (0.2518)	0.2005	0.1756	24.1432 (0.2688)	26.2125 (0.3675)	25.1779	4.2856	0.3281	0.4334			
Panel 2		T-M	5 years											
Vol	SPSA Value	CPU(sec)	OLSM		P.E.(%)	CPU(sec)	LB	MLSM		Mid	UB of		P.E.(%)	CPU(sec)
0.15	5.0976 (0.0684)	1.3734	5.0478 (0.0679)	0.9775	0.1648	5.0549 (0.0767)	5.2411 (0.0788)	5.1480	1.8420	0.9885	0.3863			
0.3	13.6265 (0.1578)	1.4169	13.5246 (0.1734)	0.7473	0.1595	13.0849 (0.1448)	13.9181 (0.1502)	13.5015	3.1839	0.9170	0.4081			
0.5	25.2607 (0.2234)	1.0088	25.2101 (0.2518)	0.2005	0.1756	24.1432 (0.2688)	26.2125 (0.3675)	25.1779	4.2856	0.3281	0.4334			

The SPSA value is calculated by a parametrized iteration method, simultaneous perturbation stochastic approximation (SPSA) and works as a benchmark for completeness. All stocks have the same setting as Example 2.3 except that the maturities are 0.5, 3, 5 years and the volatilities are 0.15, 0.3, 0.5.

Approximate option values are obtained by applying OLSM and MLSM approaches 100 times with 50 equally spaced Exercisable Times, $M_0 = 1000$, $M = 10000$ Monte Carlo paths and labeled with OLSM and MLSM respectively. The reported values are averages over the 100 runs. LB and UB represent lower and upper bounds respectively. The upper bounds are calculated as described in the upper bound section and the Mid price is defined as the average of the estimated lower and upper bounds. T-M, Vol and CPU are time-to-maturity, volatility and CPU time, respectively. The (UB of) P.E. is the (upper bound of) pricing error. The standard errors of the approximate values are shown in the parentheses. In Panel 1, we fix the level of volatility at 0.5 and vary the time-to-maturity from 0.5 years to 5 years. In Panel 2, we fix the time-to-maturity at 5 years and vary the volatility from 0.15 to 0.5.

Table 2.7: American Arithmetic Average Options on 10 underlying stocks.

2.5 Conclusion

This paper develops an efficient payoff approximation approach to estimating lower and upper bounds for pricing high-dimensional American arithmetic average options. The crucial step of the approach hinges on finding a highly correlated geometric average with a given arithmetic average. Then, the LSM approach proposed in Longstaff and Schwartz (2001) is applied to obtain the optimal exercise strategy for a lower-dimensional and more tractable American geometric average option. This optimal exercise strategy is used to estimate a lower bound for the corresponding American arithmetic average option price. In addition, an upper bound can be simply calculated without computationally expensive nested simulation. Our method is especially efficient when the underlying asset prices are modeled by jump-diffusion processes with deterministic volatilities because the geometric mean is always a one-dimensional Markov process regardless of the number of underlying assets and thus is free from the curse of dimensionality. Numerical examples show that the proposed methods perform well with dimensions up to 50 when volatility is not very high and/or time-to-maturity is not very long, providing encouraging evidence that our methods have the potential of being widely adopted by academics and practitioners for pricing high-dimensional American arithmetic average options and using high-dimensional American arithmetic average options in financial activities such as hedging, diversifying, risk management particularly.

Chapter 3

Dynamic Optimal Portfolio Choice Problem under Financial Contagion

3.1 Introduction

Recent empirical studies have documented the following features during turbulent market conditions: (1) stock prices tend to crash together; (2) a big jump in stock prices is likely to be followed by more frequent jumps; (3) a big jump in stock prices is likely to be associated with a big jump in volatility.

In particular, Ang and Chen (2002) demonstrate that asset returns are more correlated in bear markets than in bull markets as measured by the exceedance correlation, i.e., the stock prices tend to crash together but not boom together. This finding implies a reduction in portfolio diversification during market downturns. In Aït-Sahalia, et al. (2015), they document episodes of highly clustered jumps across world markets, i.e., a big jump in stock prices is likely to be followed by more frequent jumps, which they term *contagion*. In the option pricing literature, Eraker, Johannes and Polson (2003), Eraker (2004), among many others, find strong evidence for co-jumps in volatility and stock returns, i.e., a big jump in stock prices is likely to be associated with a big jump in volatility. The objective of the present paper is to develop tractable multi-asset models to capture all these stylized facts and propose a new approach to solve the optimal portfolio choice problem in order to investigate impacts of time-varying jumps in stock returns and variance-covariance.

While jump-diffusion models have been introduced into the optimal portfolio

choice problem for long time (e.g. Merton (1976)), there are few results which provide closed-form solutions to portfolio choice problems in multi-dimensional jump-diffusion models. In general, one has to solve numerically an n -dimensional nonlinear equation for an n -dimensional jump-diffusion model. For example, Liu et al.(2003) study the effects of rare events affecting price and volatility. But, in their model, there is only one risky asset available for trading. In contrast, we investigate impacts of rare events affecting prices and variance-covariance, and thus, we are able to study the impact of contagion on optimal portfolio strategies. Considering a multi-asset model, Das and Uppal (2004) investigate the effects of systemic risk on portfolio diversification. Unlike our model in the present paper, the volatility and jump intensity in their model are all constants. In a multi-asset model closely related to ours, Aït-Sahalia, et al. (2009) use self- and mutually exciting processes to capture contagion in dynamics of asset returns. In particular, they develop a simple orthogonal decomposition method and possibly remove the restriction on the number of assets involved. Our model in the present paper differs from theirs. First, we use a Wishart process with jumps to model the variance-covariance while they do not incorporate jumps in volatility. Second, their decomposition method requires really special structures for the jump vector and the variance-covariance matrix. In this paper, we present a general decomposition technique which can be applied to a wide class of multi-dimensional jump-diffusion models. As a result, we obtain a semi-closed form solution for the optimal portfolio choice problem.

This paper also extends another closely related work, Buraschi et al. (2010), by developing a general decomposition technique for jump-diffusion models and allowing jumps in both the stock returns and the variance-covariance processes. As a result, these contributions present a venue to study the effects of an abrupt increase in dynamics of correlation on portfolio diversification. Meanwhile, this paper also extends the result on portfolio choice of Leippold and Trojani (2010) where they do not study the problem when asset prices consist of jumps.

Jin and Zhang (2012) also adopts a decomposition method to deal with the models where there are jumps in asset returns only. But it is not straightforward to employ their method to solve the optimal portfolio choice problem in a model with co-jump in asset returns and volatility. In contrary, our decomposition method in this paper can be applied to more general models where the variance-covariance process contains jumps. In principle, our model can be viewed as an multi-dimensional extension of the one-dimensional case studied in Liu et al.(2003).

In order to obtain a semi-closed solution in the stochastic volatility model, we let the jump intensities be an affine function of the variance-covariance. Since the variance-covariance consists of jumps, the jump intensities are mean-reverting with jumps. As a result, the corresponding jump processes have a feature of “mutually exciting“, which is nicely used to model contagion by Aït-Sahalia and Hurd (2015) and Aït-Sahalia, et al. (2015). Apart from the ”mutual-“ and/or ”self-“ exciting feature, the specification of jump intensity also captures one important observed feature, i.e., volatility clustering of returns. More specifically, the jump intensity is defined as the scaled variance-covariance matrix, where the values are essentially linear combinations of variance and covariance between the assets. In other words, it is the current variance-covariance that drives the potential ”clustering“ of future variance-covariance. Moreover, by looking at the values of the scaling matrix, the ”vulnerability“ of each asset to its own risk (reflected by the variance) or to the counterpart’s risk (reflected by the covariance) is measurable.

The rest of this paper is organized as follows. For the sake of completeness, we review the financial contagion literature in Section 2 in order to provide general pictures about how financial contagion influences the economy. In Section 3, we first introduce the jump-diffusion model where the variance-covariance is a continuous-time Wishart process with jumps. Then we develop the decomposition method and solve the optimization problem in Section 4. Numerical illustrations and financial implications are presented in Section 5. Concluding remarks are summarized in Section 6. Tedious proofs are collected in Appendix for clarity.

3.2 Contagion Literatures

There had been several branches of literatures about financial contagion including the transmission of contagion arising from the linkages of the market, correlated information and liquidity shocks and even responses from investors’ *behaviour* such as attitude or psychology. For instance, Kang and Viswanathan (2010) provided an indirect evidence of contagion in terms of liquidity, where illiquidity spills over from an industry to another. In particular, after a large negative shock, the financial intermediaries are forced to liquidate. As the markets go down considerably, most of the assets are also forced to liquidate. Hence,

both level and commonality of liquidity are affected due to the large negative shock. As a result, liquidity short can be regarded as a signal of contagion effects.

Jorion and Zhang (2009) examined the clustering of default i.e. *credit contagion* and documented that counter party risk may be another channel of credit contagion. While, Leitner (2005) developed a model for interdependence of financial markets where linkages not only spread contagion but further induce bailout coming from the threat of contagion. Lonstaff and Rajan (2008) explored the information embedded in CDOs to investigate the contagion effects in the sense that corporate defaults tend to cluster. They found there is significant clustering of default in the market and hence implied that there exists a significant part of corporate credit risk which is not diversifiable.

Some features such as flight-to-quality and home bias may be observed during crisis or market downturn. For example, Nieuwerburgh and Veldkamp (2009) proposed an alternative way to study home bias puzzle. More specifically, they emphasized that the information immobility, the costs of learning and initial advantages of certain information would further induce the home bias since these obstacles tend to stop investors from deviating their initial positions. Giannetti and Laeven (2012) used international syndicated loan market as a channel to investigate the transmission mechanism of adverse shocks to banking. Analogous to the well-known puzzle, home bias, they found that investments in domestic loans increase 20% during financial crises i.e., manifestation of flight home effect. (The authors argued that diversification benefit is reduced during financial crises, probability of bailout associated with domestic investment maybe higher and the level of risk aversion increases, see also Dungey et al (2009))

Following the related literature, it is easy to understand that financial contagion is an integrated phenomenon with several unsolved puzzles entangled. As a result, we need to understand thoroughly what is financial contagion.

3.2.1 What is financial contagion?

There is no consensus definition for contagion and it has been widely investigated in different aspects, e.g. identifying the contagion channels, detecting directly the contagion effect, etc. However, the financial contagion has been studied under some common definitions in the literature.

We review the definitions in Pericoli and Sbracia (2003) to shed some light on

how financial contagion had been studied in the literature.¹

Definition 1. Contagion occurs if there is a significant increase in the probability of future crises, conditional on a crisis occurring in another counterpart in the economy.

Conventionally, this definition is widely adopted in the literature about currency crisis. By construction, it helps to identify the vulnerability of a specific country to external shocks. However, it could be intuitively hard to be incorporated for solving optimal portfolio problems.

Definition 2. Spillovers of the volatility of asset prices from the crisis country to the other counterparts imply the occurrence of contagion.

In general, spillover effects demonstrate explicitly how the contagion is spread across different countries. The common techniques applied for assessing the occurrence and direction of volatility spillover are the multivariate GARCH models. However, the tractability of the model is typically not guaranteed for solving optimal portfolio problems because of the nature of multivariate GARCH model incorporated.

Definition 3. Contagion arises if there exists excessive cross-country movements which could not be explained by fundamentals.

Definition 4. Contagion occurs when there are excessive co-movements of prices and quantities among the markets, conditional on a crisis occurring in one or more counterparts in the economy.

Definition 5. (Shift-) contagion occurs if the transmission channel intensifies or significantly changes after a shock in one country

zz The Definition 3~5 are quite similar. More specifically, they all hinge on the "significant change" resulting from a external shock caused by the country itself or by other counterparts. The difference lies in the "object" investigated, for example, direct assessment of prices, co-movements, etc are commonly analyzed for Definition 3 and 4; while the transmission channels are considered in Definition 5. Analogous to contagion in epidemiology, the direct assessments of the prices, co-movements, etc are analogous to the syndrome and the transmission channels are analogous to the routes of the infection.

¹Pericoli and Sbracia (2003) provides a thorough survey for extant literatures. In particular, a review of the related literature among various methodology are listed.

zz To sum up, based on Definition 3~5, it is more appropriate to test contagion by determining if there is any "structural break" in the data-generating process. The detected "structural break" would imply the deviation resulting from the source of infection, say the crisis-originating country. In addition, the data-generating process, or more generally, the presumed asset pricing models adopted in the existing literatures related to contagion makes the contagion effect explicitly detectable. On the contrary, the probit models and volatility spillover related models adopted in the literatures related to Definition 1 and 2, to some extent, only suggest the outcomes of crises and hence provide only some potential interpretations about financial contagion.

Apart from these common definitions, the immediate way to study financial contagion is to identify and capture the observed features of market returns. Stylized features of contagion could be summarized as: 1. time-series clustering, 2. cross-sectional clustering, 3. directional clustering. For instance, the recent financial crisis demonstrated the first two features conspicuously and brought about devastating recession among the world economics for years. Moreover, the severity resulting from crisis of an economy may differ from others and mainly depends on its own stability. This characteristic is reflected by the third feature, which describes the power of influence to others. A crisis occurs in one place and spread across the world through the increasing connection between different markets like a disease. Finally, this chain reaction deteriorates the financial economics based on the fragility of each market. Indeed, this is exactly how disease contagiously spread among different objects and it is termed "financial contagion". Financial contagion has gained more attention in the literature after the recent crisis. In particular, Choi (2014) proposed a stylized model of self-fulfilling panic and showed heterogeneous fragilities to strategic risks (financial spillovers) of agents are critical components in financial contagion. In fact, the heterogeneous fragilities are reflecting the financial health of each agents and Choi (2014) provided a theoretical model for directional clustering. An interesting result is that agents with stronger financial health should be bolstered instead of the weak ones, which is quite counter intuitive since contagion often starts from weaker parts of the economy.

Based on the above discussion, a comprehensive model of financial contagion should, to great extent capture these three features. A direct vehicle for modeling time-series and cross-sectional clustering had been studied in Aït-Sahalia et

al. (2015), where the Hawkes jump is incorporated for the dynamic of jump intensity because of its inherited properties: self- and mutual- excitations. They developed the estimation technique with GMM method and found self- and mutual- excitation feature in the data under the assumption of constant volatility. To some extent, capture part of financial contagion effect and in fact, U.S has more influence on other markets. Although many efforts had been made in the literature of optimal portfolio choice problem, including stochastic volatility and stochastic volatility with jumps, etc, little is known about the optimal portfolio choice under financial contagion with stochastic volatility and jumps in both return and covariance matrix.

Our main contributions of this paper can be summarized as follows. Compared with Aït-Sahalia et al (2015), our model outperforms in two aspects: 1. Wishart process is incorporated to address the effects from stochastic variance-covariance matrix, 2. we solve the optimal portfolio problem. Complemented with the Hawkes-style jump component, including two novel parameters which would be analyzed in detail later, our model also could capture more empirical features than the one in Buraschi et al (2010). In particular, their model is valid for weekly and monthly data but rejected for daily ones. They suggested that the jumps in daily context result in the rejection of Hansen’s test, which manifests the importance of jumps in return. Furthermore, Corradi et al. (2012) also documented that including jump in volatility is not sufficient to capture contagion and jump in asset price are needed. By incorporating the jump components, our model aim to capture the daily variation of prices and returns much more well and hence the contagion effect could be analyzed in more detail without losing much information as using weekly and monthly data.

The advantages of our model could be summarized as the following:

- Parsimony: Our model consists of two jumps in asset price and covariance matrix process modeled by Wishart process with novel parameter $J, C1$ and $C2$ to capture financial contagion. Hence, we avoid assuming the multivariate factor models commonly adopted in the empirical researches. Moreover, the presumption of the interdependence structure is avoided. As a result, we circumvent the omitted variable problems and some empirical issues commonly faced by existing literatures.
- Tractability: Optimal portfolio choice problem is solved with closed-form

solution.

- Capturing observed clustering: Future jumps are affected/induced by current and past jumps e.g. the market sentiments would be affected by large market moves and the effects may propagate within the region and across countries.
- Asymmetry: The flexibility of the intensity function allows us to set different influencing powers to each element in the economy. As a results, it is feasible to capture the asymmetric influencing power(directional clustering) among different markets.
- Realism: As shown in Aït-Sahalia et al (2015) that the application of Hawkes jump would capture the observed features e.g. time-series clustering and cross-sectional clustering. Besides, the Wishart process is proven to well capture the stochastic covariance, which plays an important role in both optimal portfolio problem. Combined with these two key components, our model are expected to fit the data better given both models are nested in our model. As illustrated in the Numerical Analysis, we find our model captures several observed features of contagion and provides economical interpretations in the context of optimal portfolio choice problem.

3.3 A Multi-Dimensional Jump-Diffusion Model with Stochastic Volatilities

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space with filtration $\mathbf{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ where T denotes a finite time horizon. An investor with constant relative risk aversion (CRRA) utility trades n risky assets and a riskless asset with instantaneous riskless return r over the finite time horizon $[0, T]$. Let $S_t = (S_{1,t}, \dots, S_{n,t})', t \in [0, T]$ denote the price vector of the risky assets at time t , where $'$ denotes the transpose of a vector (or matrix).

Assume we have two stocks and two jumps, i.e., $n = m = 2$.² The dynamics of the price vector $S_t = (S_{1,t}, \dots, S_{n,t}), t \in [0, T]$ of the risky assets is governed by the equation:

$$diag(S_t^{-1})dS_t = (\mu_t + r1_n)dt + \Sigma_t^{1/2}dW_t + J \cdot (Y_t dN_t), \quad (3.1)$$

where $r > 0, 1_n = (1, \dots, 1)' \in \mathcal{R}^{n \times 1}, \mu \in \mathcal{R}^{n \times 1}, J \in \mathcal{R}^{n \times m}$ are constant or deterministic vectors or matrices; $\Sigma^{1/2} = (\sigma_{ij})_{n \times n}$ is the (volatility) coefficient matrix of the diffusion terms; W_t is a $n \times 1$ standard Brownian motion. The operator “*diag*” transposes a vector into a diagonal matrix.

We set the variable

$$\mu_t = \Sigma_t \eta - J \cdot \mathbf{E}[Y] \lambda_t = \Sigma_t \eta_t - J \cdot \mathbf{E}[Y] \cdot C_1 \Sigma_t C_2,$$

where μ_t is an affine structure of Σ_t as usual settings in literature of stochastic volatility models. The affine structures λ_t and μ_t make it possible to obtain a closed-form solution.

An m dimensional jump counting process $N_t = (N_{1,t}, \dots, N_{m,t})'$ is independent of the Brownian motions, and $N_{i,t}$ is independent of $N_{j,t}$ for $1 \leq i \neq j \leq m$. The random jump size $Y = diag(Y_1, \dots, Y_m)$ is a diagonal matrix where Y_i is independent of Y_j for $1 \leq i \neq j \leq m$. J is a constant matrix which is used

²Without loss of generality, I added the assumption $n = m$ in this chapter. For the case $n < m$, it is not solvable since in the model setting, the jumps in the return are set to be scaled by the $n \times n$ matrix, J . In particular, this setting is innovative and specially designed to capture financial contagion effects by allowing jumps to propagate through all components in the economy with jump sizes scaled by the matrix, J . As a result, the case $n > m$ is not included in the current model. However, with respect to the case $n > m$, it is solvable since it is a nested case of $n = m$. More explicitly, we can simply set the jump sizes to zeros for remained $n - m$ dimensions. Hence, it is clear that the case $n > m$ could be solvable as in the case $n = m$ while $n - m$ components of jumps are with zero jump sizes.

to scale the jump sizes. Hence $J \cdot (Y_t dN_t)$ denotes the jump components in the dynamics of the risky assets. If not otherwise mentioned, we use the convention “ \cdot ” to denote the multiplication between vectors or matrices hereafter. Although the matrix multiplication does not need a “dot” operator, adding it would make terms with several matrix multiplications more intuitive. If the element-by-element product is applied, further notification would be noted.

The jump counting process $N_{i,t}$, for $i \in \{1, 2, 3, \dots, m\}$ has an intensity $\lambda_{i,t}$ such that

$$\begin{aligned}\mathbf{P}[N_{i,t+\Delta t} - N_{i,t} = 1 | \mathcal{F}_t] &= \lambda_{i,t} \Delta t + o(\Delta t), \\ \mathbf{P}[N_{i,t+\Delta t} - N_{i,t} > 1 | \mathcal{F}_t] &= o(\Delta t).\end{aligned}$$

In particular, N_t is an m -dimensional Poisson process if $\lambda_t = (\lambda_{1,t}, \dots, \lambda_{m,t})'$ is constant. In summary, the equation (3.1) is modeling a jump-diffusion market with m types of jumps. The jump intensities and the jump sizes are described by λ_t and JY_t respectively.

Let $\Sigma_t = (\Sigma_{ij,t})_{n \times n} = \Sigma_t^{1/2} \cdot (\Sigma_t^{1/2})'$ denote the covariance process. In the following, we consider the case: $n = m = 2$. We take the specific setting of $\lambda_t = C_1 \Sigma_t C_2$, where constants vector $C_1 \in \mathcal{R}^{m \times n}$, $C_2 \in \mathcal{R}^{n \times 1}$. That is, λ_t is affine of Σ_t . Let

$$J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, C_1 = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, C_2 = \begin{pmatrix} C_2(1) \\ C_2(2) \end{pmatrix}. \quad (3.2)$$

Then

$$\lambda_t = \begin{pmatrix} \lambda_{1t} \\ \lambda_{2t} \end{pmatrix} = C_1 \Sigma C_2 = \begin{pmatrix} C_{11}C_2(1)\Sigma_{11} + (C_{11}C_2(2) + C_{12}C_2(1))\Sigma_{12} + C_{12}C_2(2)\Sigma_{22} \\ C_{21}C_2(1)\Sigma_{11} + (C_{21}C_2(2) + C_{22}C_2(1))\Sigma_{12} + C_{22}C_2(2)\Sigma_{22} \end{pmatrix}. \quad (3.3)$$

To keep λ_t non-negative, we may set further that

$$C_1 = \text{diag}(c_1, c_2, \dots, c_m) \cdot \begin{pmatrix} C'_2 \\ \dots \\ C'_2 \end{pmatrix} \quad (3.4)$$

where c_1, \dots, c_m are non-negative constants. This model nests the one dimensional model of Liu et al. (2003) which corresponds to the case of $n = m = 1$ and $\lambda_t = c\sigma_t^2$ for some $c > 0$.

3.3.1 A Stochastic Variance-Covariance Process with Jumps

To model stochastic variance-covariance matrices, Buraschi et al. (2010) adopt a continuous-time Wishart process which is basically a matrix-valued version of the univariate square-root process. In this paper, we extend their model by incorporating jumps. In particular, the variance-covariance matrix process Σ_t is assumed to follow a stochastic (matrix) process with jumps:

The variance-covariance matrix process Σ_t is assumed to follow a Wishart process with jumps:

$$d\Sigma_t = [\Omega\Omega' + M\Sigma_t + \Sigma_t M']dt + \Sigma_t^{1/2}dZ(t)Q + Q'dZ'\Sigma_t'^{1/2} + \sum_{i=1}^m K^{(i)}dN_i, \quad (3.5)$$

where $\Omega, M, Q \in \mathcal{R}^{n \times n}$ are constant coefficient matrices; M is negative definite. Z_t is a $n \times n$ standard Brownian motion, correlated with the process W_t through the way

$$W_t = \sqrt{1 - \rho'\rho}Z_t^\perp + Z_t\rho, \quad (3.6)$$

where Z_t^\perp is an n dimensional standard Brownian motion independent of Z_t , and $\rho = (\rho_1, \dots, \rho_n)' \in \mathcal{R}^{n \times 1}$ is with the properties $\rho_i \in [-1, 1], i = 1, 2, 3, \dots, n$ and $\rho'\rho \leq 1$. Thus,

$$W_{1t} = \sqrt{1 - \rho_1^2 - \rho_2^2}Z_{1t}^\perp + \rho_1 Z_{11t} + \rho_2 Z_{12t} \quad (3.7)$$

$$W_{2t} = \sqrt{1 - \rho_1^2 - \rho_2^2}Z_{2t}^\perp + \rho_1 Z_{21t} + \rho_2 Z_{22t} \quad (3.8)$$

The jump coefficient matrix $K_i \in \mathcal{R}^{n \times n}, i = 1, 2, \dots, m$ are positive semi-definite (symmetric) matrices. When $K_i \equiv 0, i = 1, 2, \dots, m$, the diffusion process (3.5) is introduced by Bru (1991). It is known that if $\Omega'\Omega \gg Q'Q$, then Σ_t is a well-defined covariance matrix process. Under the same condition, one can see that (3.5) also defines a well-defined covariance matrix process when $K_i = (K_{i,jl})_{1 \leq j,l \leq n}, i = 1, 2, \dots, m$ are positive semi-definite. The equation for Σ_t suggests that Σ_t will increase by K_i if the jump $N_{i,t}$ occurs at time t .

The equation (3.1) and (3.5) together with the setting $\lambda_t = C_1 \Sigma_t C_2$ can cap-

ture several stylized phenomenon during turbulent market conditions. In our model, once a jump occurs, the variance-covariance increases simultaneously. The jump intensities increases as well. As a consequence, jumps will occur more frequently. Through the same mechanics, when one type jump occurs, another type of jumps shall occur more frequently. This is a feature of contagion in the same sense as Aït-Sahalia et al. (2015).

In this following paragraphs, we explain the interpretation of model parameters and how economic issues related to contagion may be captured.

3.3.2 A Hawkes-style jump

The affine structure of λ_t not only allow for closed-form solution of optional portfolio but also accounts for the volatility clustering feature i.e., high volatility tends to follow high volatility. Moreover, the resulting high jump intensity in volatile market captures the fact that crisis often occurs during market turmoil. On the other hand, given that M is negative definite, Σ is a mean-reverting process. As a result, the jump intensity $\lambda_t = C_1 \Sigma_t C_2$ is a mean-reverting process as well. By incorporating jump in the covariance structure, it is clear that the counting jump process N_t is a Hawkes-style process. N_t has a feature called mutually exciting in sense that once a type of jump occurs, other types of jumps tend to occur more frequently. This feature is ideal for modeling financial contagions which usually are used to describe the phenomena that a worldwide sequence of market crashes (jumps) follows a crash in one country. For example, Aït-Sahalia and Hurd (2015) and Aït-Sahalia, et al. (2015) model financial contagions by mutually exciting processes. In contrast to theirs, the structure of λ_t in our model exhibits more comprehensive interpretation. In particular, λ_t is associated with the covariance and scaled by $C1$ and $C2$ parameters. As a result, variance and covariance of the assets are the vehicles which drive the magnitude of jump intensity in absence of jumps. The sudden jump incurred by unexpected systemic risks then can be regarded as a natural response based on the private sensitivities to variance and covariance which is measured by the absolute value of $C1$ and $C2$. Moreover, the resulting increase in covariance then changes dynamics of the covariance structure. Actually, this implication illustrates an idea similar to the "financial health" discussed in Choi (2014).

3.3.3 Interpretation of J

As discussed in previous section, the interaction among markets is commonly investigated in the financial contagion literature to determine and show the contagion effects. However, little is known about addressing this issue in portfolio choice problems. To explicitly describe the interaction between markets or assets in our model, we add a novel component to the jump component in the asset process driven by SDE(3.1). The role of J is intuitively clear i.e., it scales the influence of jumps among the assets where all possible jumps are affective to each asset in the economy. More specifically, one jump occurring in one market may be transmitted to other markets even if there is no jump there. This feature allows us to investigate both the power of influence and vulnerability(fragility) for each markets by absolute value of each off-diagonal element in J .

Incorporated with these novel and economical meaningful parameters, our model exhibits good properties for modeling financial contagion. Apart from these properties, the affine structure makes the optimal portfolio choice problem solvable with closed-form solution.

3.3.4 Parameter setting in this paper

Our model exhibits flexible properties based on the novel components incorporated. In this section, we specify the setting of parameters and jump distributions of return and covariance matrix adopted. The jump distribution of return, $Y_{i,t} = e^{Z_{i,t}} - 1$ represents the relative size change in asset prices with $Z_{i,t} \sim \mathcal{N}(\mu_J^{(i)}, (\sigma_J^{(i)})^2)$ for $i = 1, \dots, n$. To preserve the positive semidefinite property of covariance and to make influence of jump intuitive, we construct $K_t^{(i)}$ if $n = 2$ as:

$$K_t^{(i)} = \begin{pmatrix} K_t^{11} & c_i \sqrt{K_t^{11} K_t^{22}} \\ c_i \sqrt{K_t^{11} K_t^{22}} & K_t^{22} \end{pmatrix} \quad (3.9)$$

where $K_t^{ii} \sim \text{Exp}(\mu_v^{(i)})$, $-1 \leq c_i \leq 1$, $i = 1, 2$. Given $K_t^{(i)}$ constructed as above, it's clear that positive semidefinite property of Σ_t is preserved. Furthermore, the jump increments in variance and covariance are correlated with a constant coefficient c_i , and hence exhibit intuitive jump change in covariance structure.

Apart from jump distributions, we construct parameters $J, C1, C2$ and $\Omega\Omega'$ in the following specific structure which is consistent with intuitions. For illus-

trative purpose, as $n = 2$, they are constructed as below:

$$J = \begin{pmatrix} 1 & J_{21} \\ J_{12} & 1 \end{pmatrix}, C_1 = \begin{pmatrix} 1 \cdot C_2(1) & 1 \cdot C_2(2) \\ C_1(1) \cdot C_2(1) & C_1(1) \cdot C_2(2) \end{pmatrix}, C_2 = \begin{pmatrix} C_2(1) \\ C_2(2) \end{pmatrix}. \quad (3.10)$$

These refinements are intuitively straightforward but remain intuitive for economical interpretations. In particular, the diagonal elements of J are set to be 1, which imply the jump resulting from risks of asset 1 are scaled by 1, while the influence of risk reflected by jump in asset 2 are transmitted to asset 1 and scaled by J_{21} . As a result, J_{21} and J_{12} capture how contagion propagates through different assets or markets with the magnitude measured by the absolute value of them. To investigate financial contagion, we can further limit J_{21} and J_{12} to be positive to capture how negative jumps deteriorate returns during market turmoil. With respect to refinement of C_1 and C_2 , positive jump intensities are preserved and $C_1(1)$ captures the ratio of vulnerability between asset 1 and asset 2. More specifically, more vulnerable one would have higher probability to be affected by jumps/risks and hence capture how assets react during crisis.

Without loss of generality, we let $\Omega\Omega' = a + a'$ in the drift term of the Wishart process (3.5), to simplify the estimation procedure. Since $\Omega\Omega' \gg Q'Q$ preserves the positive semidefinite property of covariance matrix process, it is commonly assumed $\Omega\Omega' = kQ'Q, k \geq 10$ as in Buraschi et al. (2010). The replacement of $\Omega\Omega'$ with $a + a'$ relaxes the linear relationship between $\Omega\Omega'$ and $Q'Q$. The positive semidefinite property of Σ_t can be preserved in the MCMC procedure naturally given $a + a'$ constrained to be positive semidefinite.³

With respect to the flexibility of our model, we denote WJD model as Wishart-Jump-Diffusion model, where the suffix $DJ(iDJ), AJ, CJ$ to be the double jump (independent double jump) in return and covariance matrix, asset jump in return only and covariance jump in covariance matrix process only of our model, respectively. On the other hand, *Myopic* and WD (Wishart-Diffusion) are denoted as the Merton-Jump-diffusion (MJD) model and the model in Buraschi et al. (2010) respectively. For example, the most general WJD-DJ model represents the Wishart-Jump-Diffusion model with jumps in both return and variance-covariance. If not otherwise mentioned, we follow the abbreviation

³Note that detailed estimation procedure is developed in Chapter 4.

defined in this section.

3.4 Solve the optimal portfolio choice problem

In the section, we investigate the effect of contagion by solving the portfolio choice problem. The optimization problem of the investor can be described as the following.

Let $\pi_t = (\pi_{1,t}, \dots, \pi_{n,t})'$ denote the vector of the proportion of the wealth X_t invested into the n risky assets, respectively. Then the wealth process X_t follows:

$$dX_t = X_t(r + \pi_t' \mu)dt + X_t \pi_t' \Sigma_t^{1/2} dW_t + X_t \pi_t' J(Y_t dN_t). \quad (3.11)$$

Given the initial wealth $X_t = x_0$, the investor selects the portfolio process π_t maximizing CRRA the utility of terminal wealth, that is, the optimization problem of the investor is

$$\max_{\pi_t} \mathbf{E} \left[\frac{X_T^{1-\gamma} - 1}{1 - \gamma} \right], \quad (3.12)$$

(or $\inf_v \max_{\pi} E[u(X(T))]$ for incomplete markets.) where the risk aversion coefficient γ is greater than 1. The results in this paper shall also hold for the logarithm utility function if we let $\gamma \rightarrow 1^4$.

3.4.1 The Solution of the Investment Problem

Let

$$V(x, \Sigma, t) = \max_{\pi} \mathbf{E} \left[\frac{X_T^{1-\gamma} - 1}{1 - \gamma} \middle| X_t = x, \Sigma_t = \Sigma \right].$$

The HJB equation associated with the optimization problem (3.12) is given by

$$\begin{aligned} 0 = & V_t + \max_{\pi} V_x x(r + \pi' \mu) + \frac{x^2}{2} V_{xx} \pi' \Sigma \pi + 2x \pi' \Sigma \cdot \mathbf{F} V \cdot Q' \rho + \mathcal{A}V \\ & + \sum_{l=1}^m \lambda_l \mathbf{E} [V(x(1 + (\pi' J)_l Y_l), \Sigma + K^{(l)}, t) - V(x, \Sigma, t)], \end{aligned} \quad (3.13)$$

⁴The value of γ is not critical and the results for the case $0 < \gamma < 1$ could be similarly obtained. Following the literature, we focus on the case $\gamma > 1$ since it is conventionally adopted.

where $\mathbf{F} = (\frac{\partial}{\partial x} \frac{\partial}{\partial \Sigma_{ij}})_{n \times n}$, and the infinitesimal differential operator associated to the covariance process is

$$\mathcal{A} = Tr[(\Omega\Omega' + M\Sigma + \Sigma M')D] + Tr[2\Sigma\mathbf{D}Q'Q\mathbf{D}], \quad (3.14)$$

where $\mathbf{D} = (\frac{\partial}{\partial \Sigma_{ij}})_{n \times n}$.

To solve the HJB equation (3.13), we conjecture that

$$V(x, \Sigma, t) = \frac{x^{1-\gamma}}{1-\gamma} e^{Tr(A(t)\Sigma) + B(t)},$$

where $A(t)$ is a symmetric matrix function of t satisfying $A(T) = 0$, and $B(t)$ is a function of t satisfying $B(T) = 0$. Then plugging the conjecture into the HJB equation, we can obtain two equation systems of $A(t)$ and $B(t)$:

$$0 = \frac{1}{1-\gamma} \frac{d}{dt} B(t) + r + \frac{1}{1-\gamma} Tr[\Omega\Omega' A'], \quad (3.15)$$

$$\begin{aligned} 0 = & \frac{1}{1-\gamma} Tr[\frac{d}{dt} A(t)\Sigma] + \max_{\pi} \left\{ \pi' \mu - \frac{\gamma}{2} \pi' \Sigma \pi + 2\pi' \Sigma A' Q' \rho \right. \\ & + \frac{1}{1-\gamma} Tr[(M\Sigma + \Sigma M') A'] + \frac{1}{1-\gamma} Tr[2\Sigma A' Q' Q A'] \\ & \left. + \frac{1}{1-\gamma} \sum_{l=1}^m \lambda_l E[(1 + (\pi' J)_l Y_l)^{1-\gamma} e^{Tr[A(t)K^{(l)}]} - 1] \right\}. \end{aligned} \quad (3.16)$$

We solve the maximization problem in (3.15) by investigating the first order condition for π , which gives

$$\mu - \gamma \Sigma \pi + 2\Sigma A'(t) Q' \rho + \sum_{l=1}^m \lambda_l E[(1 + ((\pi') J)_l Y_l)^{-\gamma} J_l e^{Tr[A(t)K^{(l)}]}] = 0. \quad (3.17)$$

As a result, the ODE (3.15) for $A(t)$ can be transferred into the following form:

$$\begin{aligned} 0 = & \frac{dA(t)}{dt} + \frac{1-\gamma}{2} \{ \eta(\pi^*)' + \pi^* \eta' - C_2(\pi^*)' J \mathbf{E}[Y] C_1 \\ & - C_1' \mathbf{E}[Y] J' \pi^* C_2' - \gamma \pi^* (\pi^*)' + \pi^* g' + g(\pi^*)' \} \\ & + (A(t)' M + M' A(t)) + 2A(t)' Q' Q A(t) \end{aligned} \quad (3.18)$$

$$+ \frac{1}{2} \sum_{l=1}^m (C_2 e_l' C_1 + C_1' e_l C_2') E[(1 + ((\pi^*)' J)_l Y_l)^{1-\gamma} e^{Tr[A(t)K^{(l)}]} - 1], \quad (3.19)$$

where $e_1 = (1, 0, \dots, 0)'$, $g = 2A'(t) Q' \rho$ and π^* is calculated by (3.17).

Then a numerical iteration method can be employed to solve the above non-

linear equation system for the $n \times 1$ vector π for each Σ . Solving n nonlinear equations with n variables is computationally expensive for a large n in general. Below we develop a decomposition approach which reduces the above nonlinear equation system into m 1-dimensional nonlinear equations and $(n - m)$ linear equations. Those 1-dimensional nonlinear equations can be solved by the numerical iteration method while those linear equations can be solved explicitly, leading to significant reduction of the computational burden for large values of n .

3.4.2 Decomposition

Recall $J \in \mathcal{R}^{n \times m}$ and $m \leq n$. Without loss of generality, we assume $\text{rank}(J) = m$. The columns of J generate a subspace of $\mathcal{R}^{n \times 1}$, denoted by $\{J\}$. We decompose $\mathcal{R}^{n \times 1}$ into two orthogonal subspaces $\{J\}$ and $\{J\}^\perp$ such that $\mathcal{R}^{n \times 1} = \{J\} \oplus \{J\}^\perp$. Let $J_1^\perp, \dots, J_{n-m}^\perp$ be an unit orthogonal basis of $\{J\}^\perp$ and let $J^\perp = (J_1^\perp, \dots, J_{n-m}^\perp)$, i.e., the matrix with columns $J_1^\perp, \dots, J_{n-m}^\perp$.

We now lay out the notations used below. If x is a row or column vector with the i^{th} entry x_i for $i = 1, 2, \dots, m$, we define

$$D_Y(x) = \text{diag}(E[(1 + x_1 Y_1)^{-\gamma} Y_1] e^{Tr(A(t)K^{(1)})}, \dots, E[(1 + x_m Y_m)^{-\gamma} Y_m] e^{Tr(A(t)K^{(m)})}) \quad (3.20)$$

Also, we denote

$$g_t = 2A'_t Q' \rho, \quad t \in [0, T].$$

Starting from the first order condition (3.17) and exploiting the orthogonal decomposition, we obtain the following result. Recall $\lambda_t = C_1 \Sigma_t C_2$, and $\mu_t = \Sigma_t \eta - J \mathbf{E}[Y] C_1 \Sigma_t C_2$.

Proposition 1. The optimal portfolio choice π_t^* is given by

$$\pi_t^* = J(J'J)^{-1} \bar{\pi}_t^* + J^\perp \pi_{\perp, t}^* \quad (3.21)$$

where

$$\pi_{\perp, t}^* = \frac{1}{\gamma} (J^\perp)' (\eta + g_t), \quad (3.22)$$

and $\bar{\pi}_t^* = (\bar{\pi}_{1, t}^*, \dots, \bar{\pi}_{m, t}^*)'$ solving the following m -dimensional nonlinear equa-

tions:

$$J'_k J_k [J'_k (\eta + g_t) - \gamma \bar{\pi}_{k,t}^*] + (J'_k C_2)(J'_k J) \cdot (D_Y(\bar{\pi}_t^*; A_t) - \mathbf{E}[Y_t]) \cdot (C_1 J_k) = 0, \quad (3.23)$$

for $k = 1, 2, \dots, m$.

We leave the proof in the appendix.

Clearly, in case J 's columns are mutually orthogonal, the above proposition can be further simplified. In fact, the equation (3.23) turns to be a 1-dimensional nonlinear equation consisting of $\bar{\pi}_{k,t}^*$ only. As a result, a numerical procedure can be further simplified for this case.

In general, if J is not a matrix with pairwise orthogonal columns, we find an invertible matrix $P \in \mathcal{R}^{n \times n}$ such that $\hat{J} = PJ$ has pairwise orthogonal columns. Then following similar derivation as above, we reach the following proposition.

Proposition 2. Let $P \in \mathcal{R}^{n \times n}$ be an invertible square matrix such that $\hat{J} := PJ$ satisfies $\hat{J}'\hat{J}$ is a diagonal matrix. Let $\hat{J}^\perp \in \mathcal{R}^{n \times (n-m)}$ with columns being an unit orthogonal basis of the subspace orthogonal to the space $\{\hat{J}\}$. Let

$$\hat{\eta} = P'^{-1}\eta, \quad \hat{g} = P'^{-1}g, \quad \hat{C}_1 = C_1 P^{-1}, \quad \hat{C}_2 = P'^{-1}C_2. \quad (3.24)$$

Then the optimal portfolio choice π^* is given by

$$\pi^* = P'(\hat{J}(\hat{J}'\hat{J})^{-1}\bar{\pi}^* + \hat{J}^\perp \pi_\perp^*)$$

where

$$\pi_\perp^* = \frac{1}{\gamma}(\hat{J}^\perp)'(\hat{\eta} + \hat{g}),$$

$\bar{\pi}^* = (\bar{\pi}_1^*, \dots, \bar{\pi}_m^*)'$ and for each $k = 1, 2, \dots, m$, $\bar{\pi}_k^*$ solves 1-dimensional nonlinear equation

$$\hat{J}'_k(\hat{\eta} + \hat{g}) - (\hat{J}'_k \hat{C}_2)(\hat{J}'_k \hat{J})L\mathbf{E}[Y_k] \cdot (\hat{C}_1 \hat{J}_k) - \gamma \bar{\pi}_k^* + (\hat{J}'_k \hat{C}_2)(\hat{J}'_k \hat{J}) \cdot D_Y(\bar{\pi}^*) \cdot (\hat{C}_1 \hat{J}_k) = 0 \quad (B.25)$$

The proof of this proposition is left in the appendix.

The advantage of (3.25) is that it is a 1-dimensional nonlinear equation of $\bar{\pi}_{k,t}^*$ because the only nonzero entry of $\hat{J}'_k \hat{J}$ is at the k^{th} position. By virtue of the above proposition, the optimal portfolio can be decomposed into two parts corresponding to a diffusion economy and a diffusion-jump economy. The latter can be further decomposed into a system of 1-dimensional jump-diffusion economies, which can be easily solved numerically.

The above result gives a decomposition expression for the optimal portfolio π_t^* . However, both $\bar{\pi}_t^*$ and $\pi_{\perp,t}^*$ depend on the function A_t which is a solution of ODE (3.15). Next, we show how to solve A_t and hence obtain $\bar{\pi}_t^*$, $\pi_{\perp,t}^*$ and the optimal portfolio π_t^* eventually.

For the case that $\text{rank}(J) = d < m$, we can find d columns of J such that they are a basis of the subspace $\{J\}$. Then following the same procedures as before, we shall obtain similar results as Proposition 1 or Proposition 2. The difference is that we shall obtain a d -dimensional nonlinear equation (corresponding to Prop. 1), or a system of 1-dimensional equations and several equations with dimension no more than $(m - d) + 1$ (corresponding to Prop. 2).

Given Prop. 1 and Prop. 2, the resulting optimal portfolio weight may be obtained by solving the above equation systems. In general, we need to solve m m-dimensional nonlinear equations for $\bar{\pi}_t^*$ and for $A(t)$. In summary, the scheme can be described as: Starting from boundary condition $A(T) = 0$, we solve $\bar{\pi}_t^*$ by (3.23) or (3.25) and $\pi_{\perp,t}^*$ and the ODE (3.18) for $A(t)$ recursively back to time zero by standard finite difference techniques. In particular, I solve the system of ODEs by standard finite difference techniques with second order Runge-Kutta scheme, while the trust-region approach by Coleman and Li (1996) is applied to solve the nonlinear equations. In addition, several numerical examples had been tested to verify the accuracy of the trust-region approach. Other numerical methods also generated similar results for solving the nonlinear equations. For clarity, the tested results are omitted here.

3.4.3 Decomposition of Hedging Demands

To better understand the optimal portfolio rules in the preceding propositions, we decompose the optimal portfolio given by (3.21) into several parts below.

For each $k = 1, 2, \dots, n$, we have

$$\pi_{k,t}^* = J_k(J'J)^{-1}\bar{\pi}_t + J_k^\perp \pi_t^\perp = \dots \quad (3.26)$$

$$= \frac{1}{\gamma} \eta_k + \frac{1}{\gamma} \underbrace{\left\{ \sum_{i=1}^m (J(J'J)^{-1})_{k,i} (J'_k J_k)^{-1} (J'_i C_2) (J'_i J) \cdot (D_Y((\bar{\pi}_t^*); 0) - D_Y(0; 0)) \cdot (C_1 J_i) \right\}}_{(3.27)}$$

$$+ \frac{1}{\gamma} g_{k,t} \quad \text{Asset Jump Hedging Demand} \quad (3.28)$$

$$+ \frac{1}{\gamma} \underbrace{\left\{ \sum_{i=1}^m (J(J'J)^{-1})_{k,i} (J'_k J_k)^{-1} (J'_i C_2) (J'_i J) \cdot (D_Y((\bar{\pi}_t^*); A_t) - D_Y((\bar{\pi}_t^*); 0)) \cdot (C_1 J_i) \right\}}_{(3.29)}$$

$$\text{Variance Jump Hedging Demand} \quad (3.30)$$

where $(J(J'J)^{-1})_{k,i}$ denotes the (k, i) entry of the matrix.

This decomposition analysis resembles those documented in Buraschi et al. (2010) but effects resulting from additional jumps in returns and covariance structure provide more robust economic implications. Following the terms defined in Buraschi et al. (2010), we may call $\frac{1}{\gamma} \eta_k$ the myopic demand, and $\frac{1}{\gamma} g_{k,t}$ the covariance/volatility hedging demand via diffusion since $g_t = 2A_t Q' \rho = 0$ if $\rho = 0$, while the rest is termed as the jump hedging demand.

The covariance/volatility hedging demand can be further separated as:

$$\begin{aligned} g_t &= 2A_t Q' \rho \\ &= 2A_{11,t} \begin{pmatrix} q_{11}\rho_1 + q_{21}\rho_2 \\ 0 \end{pmatrix} + 2A_{22,t} \begin{pmatrix} 0 \\ q_{12}\rho_1 + q_{22}\rho_2 \end{pmatrix} + 2A_{12,t} \begin{pmatrix} q_{12}\rho_1 + q_{22}\rho_2 \\ q_{11}\rho_1 + q_{21}\rho_2 \end{pmatrix}. \end{aligned}$$

According to Bueaschi et al. (2010), the first two terms, proportional to $A_{11,t}$ and $A_{22,t}$ are intuitively defined as volatility hedging portfolios since $A_{11,t}$ and $A_{22,t}$ are hedging demands against $\Sigma_{t,11}$ and $\Sigma_{t,22}$, respectively. Similarly, the third term proportional to $A_{12,t}$ are covariance hedging portfolios. In contrast to pure diffusion case in Buraschi et al (2010), it is worth mentioning that the covariance/volatility hedging demand, $\frac{1}{\gamma} g_t$, is also affected by the jumps. Since g_t is a function of A_t which follows the ODE (3.16) with the jump parameters involved.

On the other hand, The jump hedging demand can also be separated into two

parts by distinguishing the influence from jumps in assets and covariance structure. As equation (3.26) indicates, the first part of the jump hedging demand is for the jumps in the returns since it does not include the jumps in the variance-covariance while the second part is due to the jumps in the variance-covariance. The second part will be zero if there is no jump in the variance-covariance, i.e., $K_t = 0$. In addition, this separation between asset and covariance jumps allows for a direct implication that the jump risks in the variance-covariance can not be hedged through the stocks. Since the hedging demand resulting from jumps in covariance structure is zero if there is no jump in stock returns, i.e., $Y_t = 0$. This is analogous to the case in diffusion hedging variance-covariance via correlation ρ .

3.5 Simple Examples

In this section, to demonstrate the model flexibility, we present several simple examples which are nested by our model. Some of them have already been studied in the literature.

3.5.1 Two-dimensional case: $n=2$

In this case there are only two risky assets in the market. For various values of m , i.e. the number of jump types, we have three simplified models.

Case 1: $m = 0$. There is no jump in the model. Our model is reduced to the one employed in Buraschi et al. (2010). In this case, $J \equiv 0$, P is the identity matrix. By Proposition 2, $\bar{\pi}_t^* \equiv 0$, and $\pi_t^* \equiv \pi_{\perp}^* = \frac{1}{\gamma}(\mu_t + g_t)$. $A(t)$ solves an equation simplified from (3.16) by letting $J = K_t = 0$ for all $l = 1, 2, \dots, m$.

Case 2: $m = 1$. We take $P \equiv I_{2 \times 2}$, the unit square matrix. Then $\bar{\pi}_t^*$ and $\pi_{\perp,t}^*$ are two real numbers solving (3.23) and (3.22). $A(t)$ is a two-by-two matrix solving (3.16). This case will be used in the section of financial implications later.

Case 3: $m = 2$. J is a 2×2 invertible matrix. Let P be an invertible matrix s.t. $\hat{J} = PJ$ has mutually orthogonal columns. $J^{\perp} = 0$ hence $\hat{\pi}_{\perp,t}^* \equiv 0$. We have

the dynamics of the risky assets and the covariance matrix process as follow.

$$\begin{aligned}
(diag(S_t))^{-1}dS_t &= (\Sigma_t\eta - J\mathbf{E}[Y_t]\lambda_t + r)dt + \Sigma_t^{1/2}dW_t + JY_t dN \\
&= \begin{pmatrix} r + \Sigma_{11,t}\eta_1 + \Sigma_{12,t}\eta_2 \\ r + \Sigma_{21,t}\eta_1 + \Sigma_{22,t}\eta_2 \end{pmatrix} - J\mathbf{E}[Y_t]\lambda_t dt + \begin{pmatrix} \sigma_{11,t}dw_{1,t} + \sigma_{12,t}dw_{2,t} \\ \sigma_{21,t}dw_{1,t} + \sigma_{22,t}dw_{2,t} \end{pmatrix} \\
&\quad + \begin{pmatrix} J_{11}Y_{1,t}dN_{1,t} + J_{12}Y_{2,t}dN_{2,t} \\ J_{21}Y_{1,t}dN_{1,t} + J_{22}Y_{2,t}dN_{2,t} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
d\hat{\Sigma}_t &= P(d\Sigma_t)P' \\
&= (\hat{\Omega}\hat{\Omega}' + \hat{M}\hat{\Sigma}_t + \hat{\Sigma}_t\hat{M}')dt + \hat{\Sigma}_t^{1/2}dZ_t\hat{Q} + \hat{Q}'dZ_t\hat{\Sigma}_t^{1/2} + \sum_{l=1}^2 \hat{K}^l dN_{l,t}.
\end{aligned}$$

3.5.2 One-dimensional case: $n = 1$

Consider a simple 1-dimensional model as an example. This is the same model studied in Liu et al.(2003).

$$\begin{aligned}
dS_t/S_t &= (V_t\eta - J\mathbf{E}[Y_t]\lambda_t + r)dt + \sqrt{V_t}dW_t + JY_t dN_t \\
dV_t &= k(\theta - V_t)dt - K\lambda_t dt + \delta\sqrt{V_t}dB_t + KdN_t.
\end{aligned}$$

Assume $\lambda_t = \bar{\lambda}V_t$ for some constant $\bar{\lambda}$, then

$$d\lambda_t \equiv \bar{\lambda}dV_t = k(\bar{\lambda}\theta - \lambda_t)dt - \bar{\lambda}K\lambda_t dt + \sqrt{\bar{\lambda}}\delta\sqrt{\lambda_t}dB + \bar{\lambda}KdN_t$$

where Y_t is a random variable describing jump size. $dB_t \cdot dW_t = \rho dt$.

The above dynamics of the market has a clear feature of self-exciting. As one jump occurs, the volatility jumps up and the jump intensity jumps up as well. As a result, more frequent jumps shall follow. The mean-reverting (decaying) feature prevents the system from explosion, however.

By substituting $m = n = 1$ in the proposition 1 or 2, we can obtain the same solution as that in Liu et al.(2003). Roughly, the model in this paper can be regarded as a multi-dimensional extension of theirs.

3.5.3 No jump cases

When there is no jump either in the asset price dynamics or in the variance-covariance process, the solution to the optimization problem can be simplified

greatly. If there is no jump in the both of them, we obtain the same model as discussed in Buraschi et al. (2010) (given $n = 2$), and the optimal portfolio is consistent with theirs as well. If there is no jump in the asset prices but there is jump in the variance-covariance process, we can obtain simplified results by letting $J \equiv 0$ in the preceding results. A close model and related portfolio choice problem are discussed in Leippold and Trojani (2010). If there is no jump in the variance-covariance process, that's, $K_l \equiv 0$ for all $l = 1, 2, \dots, m$, then $\bar{\pi}_t^*$ is independent of A_t . The optimal portfolio can be determined by the equation (3.23) and the ODE (3.15).

3.5.4 Intuition given by Sensitivity Analysis

To provide an intuitive illustration of our model, we conduct the following sensitivity analysis. If Σ_t is constant i.e., letting $Q = M = \Omega = K \equiv 0$, and the jump size is a constant as well, we can obtain a proposition as follows.

Proposition 3. Suppose $m = 1$ and $\mu_t = \Sigma_t \eta - J \lambda_t u$, where u is the constant jump size. Given $J' \eta > 0$, we have $J' \pi_t^* > 0$ and

$$\begin{aligned} \frac{\partial J' \pi_t^*}{\partial u} &< 0, \text{ if } u \geq 0; \\ \frac{\partial J' \pi_t^*}{\partial u} &> 0, \text{ if } u \leq 0. \end{aligned}$$

The above proposition is consistent with Liu et al.(2003) when we take $n = 1$ and $J = 1$. If $J = [1, 1, \dots, 1]'$, that is, all risky assets response to the jump by the same level, $J' \pi_t^* = \sum_{i=1}^n \pi_i^*$ which is the regular exposure to the risky assets. However, if the risky assets response to the jump by different levels, the proposition suggests that the total exposure shall be measured by $J' \pi^*$ instead of the regular sum of all risky investments. This is one of the important features of our model to capture the contagion effects, where each component in the economy has specific vulnerability against different economic shocks. And this vulnerability is reflected by J . In fact, since $J' \pi^* = \bar{\pi}^*$, we see that the parallel part of the optimal portfolio $\bar{\pi}^*$ is exactly the measure of exposure to the risky assets.

The proposition also implies that the total exposure will get decreased if $u \geq 0$. But it is possible that investment on some asset may increase. It really depends on J , the response matrix to the jumps. We also study the sensitivity of the optimal portfolio to the jump intensity. In order to study the effect, we take the setting (3.4) and assume $c_1 = c_2 = \dots = c_m$. It is straightforward to show

that

$$J'_l \frac{\partial J\bar{\pi}}{\partial c_l} = \frac{E[(1 + \pi'_t J_l Y_l)^{-\gamma} Y_l]}{\gamma + \gamma c_l E[(1 + \pi'_t J_l Y_l)^{-\gamma-1} Y_l^2]}.$$

This quantity is understood as the total exposure to the l^{th} jump risk. As a consequence, if $Y_l \leq 0$ or $\mathbf{E}[(1 + \pi'_t J_l Y_l)^{-\gamma} Y_l] \leq 0$, we have

$$J'_l \frac{\partial J\bar{\pi}_t}{\partial c_l} \leq 0,$$

that is, the exposure to jump risk is a decreasing function of the jump intensity. The above sensitivity analysis are under the condition that Σ_t is a constant. The case of stochastic Σ_t or the case of constant Σ_t but more than two types of jumps in the asset prices is complicated to analyze the sensitivities with explicit formula. The effects of the parameters are mixed with impacts by other factors (e.g. signs of J 's components).

3.6 Numerical Analysis: Financial Implications

To demonstrate how our model capture financial contagion in the context of optimal portfolio choice problem, we investigate the resulting optimal portfolio implied by our model. In particular, we examine the hedging demands of variance, covariance, asset jumps and covariance jumps with various ranges of novel parameters, which are designed to capture the contagion effects as suggested in previous sections. The parameters adopted in this analysis are the monthly estimates in Buraschi et al. (2010)⁵ supplemented with additional jumps in returns and covariance. The jump distributions follow the same setting as discussed in Section 3.3.4 with reasonable jump parameters summarized in footnotes of each Table.⁶

Following this setup, the analysis conducted in this section provides an clear manifestation about about the model and the resulting optimal portfolio problem with stochastic variance-covariance and observed features considered. In particular, it also demonstrates how the additional novel jump components influence the optimal portfolio and capture financial contagion.

⁵The values of parameters adopted from Buraschi et al. (2010) are listed in Table 3.14.

⁶The parameters applied are defined in the footnotes in each table. The mean values of jumps are assumed to be negative in returns and positive in covariances, which is consistent with the feature that crisis often comes after large negative shocks and during market turmoil. similar manner is also adopted in Leippold and Trojani (2010).

3.6.1 Hedging Demands

In this section, we demonstrate how the resulting hedging demands would change given addition jumps components.

Firstly, the optimal hedging demands for 2 risky assets with two jumps scaled by J , i.e., WJD-DJ model, are summarized in Table 3.1 ~ Table 3.4 as the proportion of the myopic portfolio which is in the last column of each table. In general, the variance and covariance hedging demands increase with relative risk aversion (RRA) and investment time horizon and reach a steady level at a short time horizon consistent with the findings in Buraschi et al. (2010) and Kim and Omberg (1996). On the other hand, the demands of asset jumps and covariance jumps overwhelmingly change the resulting demands with large (negative) demands in asset jump and positive covariance jump demands. For example, when RRA equals 2 and T is 5 year, the net demand required by jumps is around -94.68% of the myopic portfolio, while for short horizon, say T is 3 month, the net demand for jumps is -78.18% . With respect to higher RRA, the net demand is still as high as -41.95% as RRA equals 11 and T is 3 month. These large (negative) demands suggest that under potential risks of crisis (reflected by negative jumps and positive jumps in returns and covariance matrix, respectively), the positions in optimal portfolio shrink drastically in order to reduce the exposures to financial contagion.

Secondly, in contrast with jumps scaled by our novel parameter J , we also compare the optimal portfolio with jumps disentangled (WJD-iDJ model⁷).⁸ With all else identical, the optimal hedging demands are summarized in Table 3.5 ~ Table 3.8. The hedging demands among all components are qualitatively similar with those obtained in the previous case, i.e., the hedging demands increase with RRA and investment horizon while demands are slightly larger in variance and covariance demands compared to previous case. However, the magnitude of demands of jumps shrink greatly if the jumps are disentangled.

⁷Recall that WJD-iDJ model denotes model with $J = I_n$, an n -by- n identity matrix

⁸It is worth mentioning that jumps in all existing literatures about stochastic volatility or stochastic volatility with jump are always independent jumps or simultaneous jumps with correlation. This is essentially different from the idea of ours. The empirical feature of contagion often occurs when a big jump, say a critical bad event hits one market with several negative jumps (bad events) following in the same place(self-exciting) or elsewhere(mutual exciting) even if there was no jump there.

For instance, the net hedging demands of jumps decreases to -21.6% of the myopic portfolio as RRA equals 2 and T is 5 year. Moreover, for $RRA = 11$ and $T = 3$ month, the net demand decreases to -9.31% of the myopic portfolio. This result is quite intuitive since assets would be only exposed to its own potential jump risks in returns in this case. As opposed to disentangled jump cases, the entangled jumps allows asset-specific risks in returns to propagate through different assets and result in larger negative jump hedging demands.

Accordingly, we find jumps in return and covariance matrix are essential component for modeling financial contagion. In particular, the parameter J captures the sensitivities to jumps (risks) and can be identified as a measure of sensitivity and/or vulnerability. In other words, by estimating the values of the off-diagonal elements of J , it is possible for people to identify how influential and sensitive a market could be compared with the other markets in the economy with our model.

Furthermore, given the importance of jump components as analyzed above, we examine which jump dominates in the context of portfolio choice problem by comparing the hedging demands where one has only jump in returns (WJD-AJ) and one has only jumps (WJD-CJ) in covariance matrix. Table 3.9 ~ Table 3.12 summarize the corresponding hedging demands of all components. As we may expect, the hedging demands of all components shrink in absolute value for WJD-AJ since there is one less resource of risk compared with DJ model. On the other hand, only variance and covariance hedging are available for WJD-CJ and the values are larger than those of WJD-AJ since the jump in covariance would affect the variance and covariance hedging demands as discussed in Section 3.4.3.

By comparing the Table II of Buraschi et al. (2010), where the model is nested in our model with jump components omitted, we can compare the difference between impacts of jumps in return and covariance. Firstly, as discussed in Section 3.4.3, there is no corresponding jump hedging demands and the hedging demand in variance and covariance hedging increase through g_t . On the other hand, jumps in returns not only lead to negative asset jump hedging demands but also deteriorate the magnitudes of variance and covariance demands. As a result, we can conclude that impact of jumps in covariance are smaller compared with the impact of jumps in returns, while in the presence of jump in covariance, the impact of jumps in return increases drastically.

3.6.2 Capturing financial contagion

In this section, we analyze how the augmented novel parameters capture the financial contagion by comparing the resulting total hedging demands among all various models, including WJD-DJ, WJD-iDJ, WJD-AJ and WJD-CJ models.

For illustration purpose, we plot the total hedging demands implied by various models in Figure 3.1. In general, the plotted total demands are consistent with the above analysis. It is worth mentioning that optimal portfolios implied by jump models shrink substantially even with low relative risk aversion ($\gamma = 2$). Moreover, the *CJ* model suggests slightly higher hedging demands compared with *Wishart* model, while the hedging demands implied by the *AJ* model decrease drastically and suggest least hedging demands due to the lack of positive demands of covariance jumps and large negative hedging demands required by asset jumps. As analyzed in the above paragraph, *iDJ* model suggests higher hedging demands since the potential risks of contagion can not be captured by disentangled jumps. However, the optimal portfolio implied by *DJ* model generally shrink but sustains higher level compared with *AJ* due to the positive demands for covariance jump structure. This significant difference between the total demands of *DJ* and *iDJ* strongly manifests the capability of capturing financial contagion with our novel model parameter J .

Next, we investigate how C_1, C_2 and J capture financial contagion in terms of resulting hedging demands by varying the critical components with different values. Firstly, with else remain identical and relative risk aversion $\gamma = 6$ and $T = 5$ year, we vary $C_2(1)$ from $0.5 \sim 1.5$ and plots all 4 types hedging demands in Figure 3.2. As discussed in Section 3.3.4, $C_2(1)$ controls the sensitivity and it is supported by the increasing (absolute) demands in jump components summarized in panel 3 and 4. With respect to variance and covariance hedging, increases in $C_2(1)$ also imply higher probability of jumps in covariance matrix and hence lead to increasing hedging demand with less magnitude. Similarly, $C_1(1)$ directly captures the vulnerability relation between asset 1 and asset 2 in the sense of jump intensity. As Figure 3.3 depicts, hedging demands of jump components increases drastically in absolute value for asset 2, while variance and covariance hedging demands increase slightly. The last but not the least, we

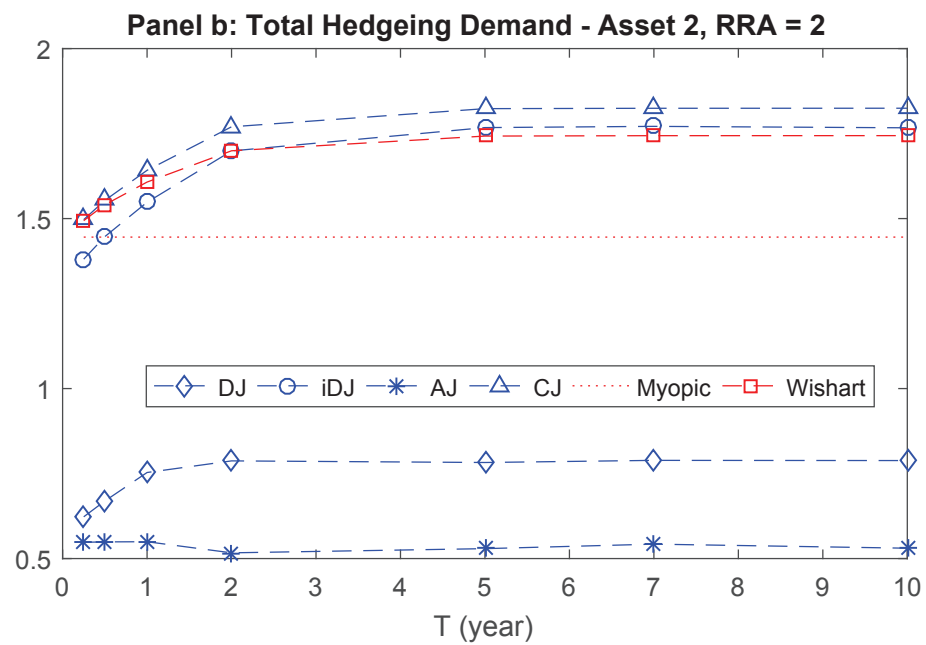
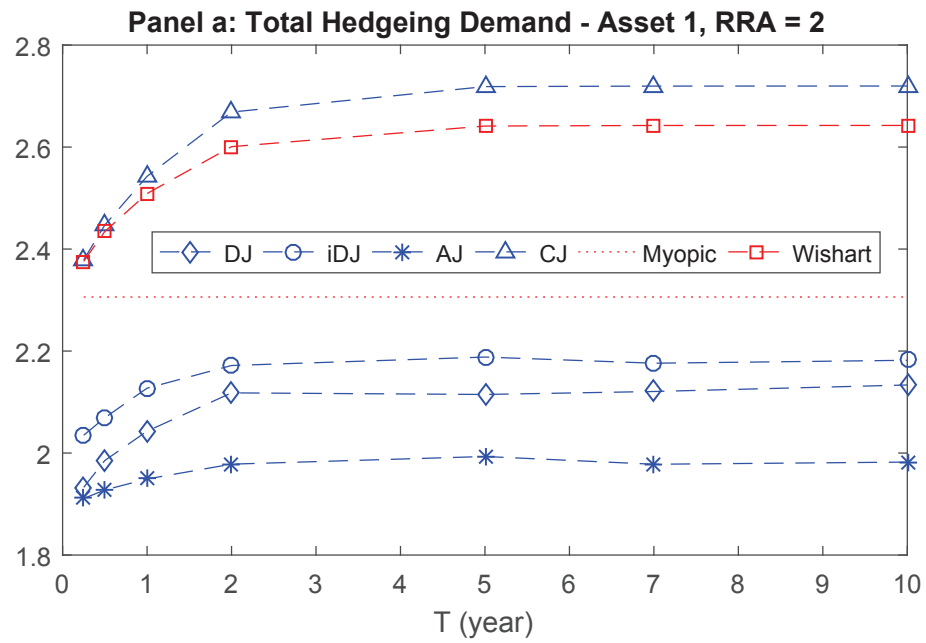


Figure 3.1: Total Hedging Demand against Time Horizon against various models

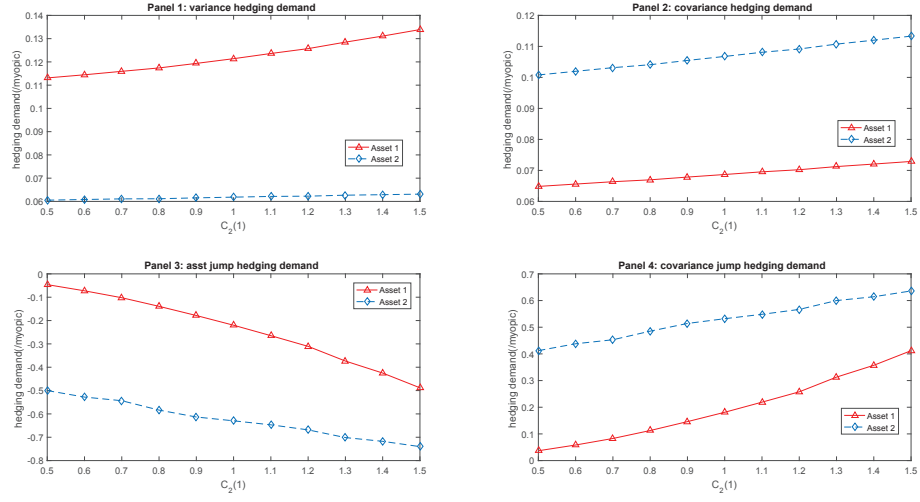


Figure 3.2: Hedging Demand against $C_2(1)$ (DJ)

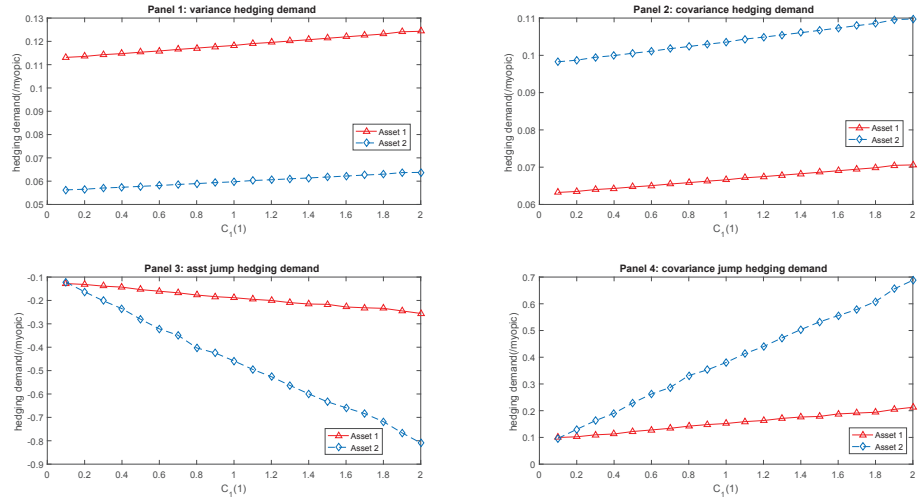


Figure 3.3: Hedging Demand against $C_1(1)$ (DJ)

provide an illustrative comparison for power of influence by varying parameter J_{21}, J_{12} and the resulting hedging demands for asset 1 and asset 2 are summarized in Figure 3.4 and 3.5 respectively. In general, variance and covariance hedging demands are not significantly affected as shown in the first two panels. On the other hand, both asset jump and covariance jump demands are substantially influenced as J_{21} and J_{12} increase. In particular, as J_{12} is low and J_{21} increases, sum of absolute jump hedging demands increases substantially for asset 1, while it remains certain level for asset 2, and vice versa. The increase in the sum of absolute jump hedging demands may be regarded as the vulnerability against mutual-excited risks with the total hedging demands representing the true optimal portfolio implied by our model.

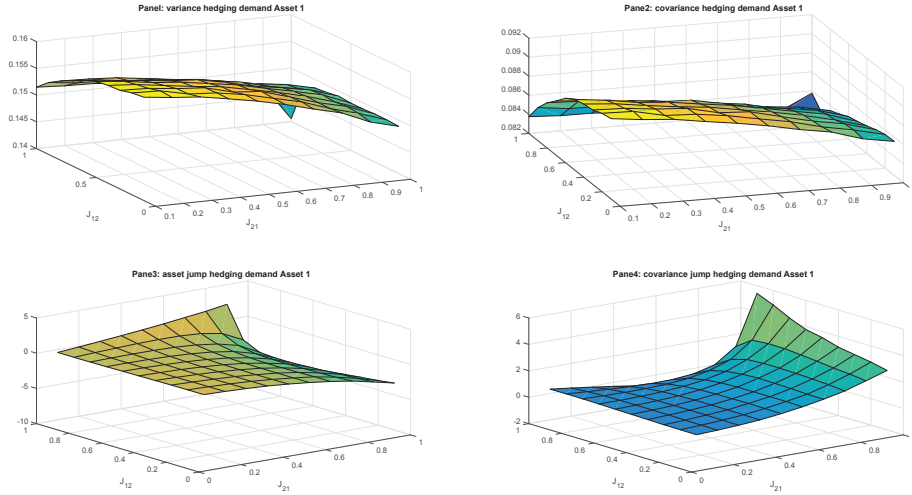


Figure 3.4: Hedging Demand of Asset 1 against J_{21} J_{12} (DJ)

To sum up, integrated with all comparative statistics of our model, the net effects of all features implied by our model provide an parsimonious and realistic model for optimal portfolio choice problem under financial contagion. Moreover, we may be the first to propose a model with closed-form solution to solve optimal portfolio which is especially designed to capture several features of empirical data and financial contagion.

3.6.3 Effects of Model Misspecification

The model misspecification considered here is caused by the investor who mistakenly assume a model to obtain a portfolio strategy. For example, Suppose

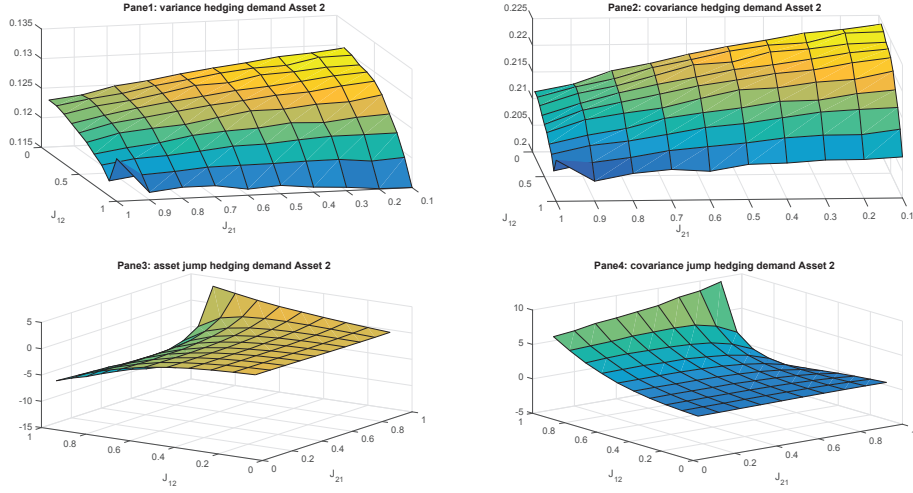


Figure 3.5: Hedging Demand of Asset 2 against J_{21} J_{12} (DJ)

the true model is the double jump model, but the investor ignores the jump risk and adopts a suboptimal portfolio (which is obtained from the model without jumps). We evaluate the economic loss caused by a suboptimal portfolio strategy. Given initial wealth $x_0 = 1$, we define the wealth equivalent loss to be x such that

$$V(1, \Sigma, 0; \pi^*) = V(1 + x, \Sigma, 0; \pi_{00}^*),$$

where π_{00}^* denotes the suboptimal portfolio. That is, the wealth equivalent loss of the investor is that the extra wealth for the suboptimal portfolio to generate the same utility value as the optimal portfolio. Using the form of the indirect utility function, we obtain the expression for x :

$$x = \exp\left(\frac{1}{1 - \gamma}(\text{Tr}(A_{11}\Sigma - A_{00}\Sigma)) + B_{11} - B_{00}\right),$$

where $A_{11}, B_{11}, A_{00}, B_{00}$ are the functions corresponding to the optimal portfolio and the suboptimal portfolio respectively. Plugging appropriate values for the parameters, we obtain the wealth equivalent losses for various RRAs.

3.7 Conclusion

In the present paper, we develop tractable multi-variable jump-diffusion models to capture empirical facts documented recently in financial literature, e.g., stock prices tend to have big move together and a big jump may be followed by more frequent jumps and features of financial contagion. To obtain closed-form solution, we develop a novel general decomposition technique with which we reduce

the problem into two relative simple ones: Portfolio choice in a pure-diffusion market and in a jump-diffusion market with less dimension. The latter can be reduced further to be a bunch of portfolio choice problems in one-dimensional jump-diffusion markets. More importantly, our method is easy to apply and free of the curse of dimensionality.

Our numerical exercises provide new insights into the structure of an optimal portfolio when jumps are present in asset prices and/or their variance-covariance. More specifically, our results show that the jumps in the variance-covariance have important effects on the asset allocations when there are jumps in the asset prices. Meanwhile, the hedging demands for jumps are much more significant compared to volatility or covariance hedging demands for diffusion risks and ignoring jump risk in the variance-covariance may cause large wealth equivalent loss.

By examining the novel component of our model, we also demonstrate the capability of our model to capture features of financial contagion. In particular, the novel parameters have intuitive economic interpretations, which help provide potential measure for financial contagion. To some extent, this paper shed new lights on the financial contagion studies in the optimal portfolio literature. In particular, financial contagion issues among different markets may be one of the top interests for future researches. Moreover, equipped with our model, some existing puzzles such as flight bias and home bias may be interesting topics for future researches.

	RRA	T (year)									Myopic Demand
			0.25	0.5	1	2	5	7	10		
Variance hedging (/myopic)	2		0.0184	0.0346	0.0574	0.0903	0.1057	0.1064	0.1066	2.3060	
			0.0067	0.0166	0.0357	0.0700	0.0882	0.0891	0.0891	1.4455	
	6		0.0321	0.0610	0.1001	0.1450	0.1581	0.1579	0.1579	0.7687	
			0.0131	0.0317	0.0654	0.1129	0.1285	0.1283	0.1284	0.4818	
	8		0.0338	0.0646	0.1055	0.1514	0.1639	0.1644	0.1641	0.5765	
			0.0140	0.0341	0.0691	0.1177	0.1327	0.1332	0.1329	0.3614	
	11		0.0352	0.0674	0.1099	0.1569	0.1691	0.1690	0.1690	0.4193	
			0.0146	0.0357	0.0722	0.1219	0.1366	0.1365	0.1364	0.2628	
	16		0.0364	0.0695	0.1134	0.1613	0.1731	0.1733	0.1733	0.2883	
			0.0152	0.0369	0.0745	0.1252	0.1394	0.1396	0.1396	0.1807	
	21		0.0370	0.0708	0.1154	0.1638	0.1751	0.1754	0.1754	0.2196	
			0.0155	0.0377	0.0759	0.1272	0.1408	0.1412	0.1411	0.1377	
	41		0.0381	0.0727	0.1185	0.1671	0.1788	0.1788	0.1788	0.1125	
			0.0161	0.0388	0.0781	0.1294	0.1435	0.1435	0.1436	0.0705	

This table dictates the variance hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_J^{(i)} = -0.1, \sigma_J^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1, c_2 = 0.5, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.1: Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled

RRA	T (year)	Myopic Demand							
		0.25	0.5	1	2	5	7	10	Demand
2		0.0074	0.0155	0.0287	0.0501	0.0607	0.0612	0.0613	2.3060
		0.0184	0.0384	0.0713	0.1242	0.1506	0.1518	0.1520	1.4455
6		0.0134	0.0281	0.0511	0.0804	0.0895	0.0893	0.0894	0.7687
		0.0332	0.0698	0.1268	0.1994	0.2219	0.2215	0.2216	0.4818
8		0.0142	0.0300	0.0539	0.0839	0.0926	0.0929	0.0927	0.5765
		0.0351	0.0744	0.1338	0.2080	0.2295	0.2303	0.2299	0.3614
11		0.0148	0.0313	0.0562	0.0869	0.0954	0.0953	0.0953	0.4193
		0.0366	0.0776	0.1394	0.2154	0.2365	0.2363	0.2363	0.2628
16		0.0153	0.0323	0.0580	0.0893	0.0974	0.0976	0.0976	0.2883
		0.0379	0.0801	0.1439	0.2214	0.2417	0.2420	0.2420	0.1807
21		0.0156	0.0329	0.0591	0.0907	0.0985	0.0987	0.0987	0.2196
		0.0386	0.0817	0.1465	0.2248	0.2443	0.2449	0.2448	0.1377
41		0.0161	0.0339	0.0607	0.0923	0.1005	0.1005	0.1005	0.1125
		0.0398	0.0840	0.1505	0.2290	0.2493	0.2492	0.2493	0.0705

This table dictates the covariance hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_J^{(i)} = -0.1, \sigma_J^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1, c_2 = 0.5, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.2: Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled

	RRA	T (year)									Myopic	
			0.25	0.5	1	2	5	7	10		Demand	
	2		-0.2251	-0.2696	-0.3647	-0.5621	-0.7213	-0.7310	-0.7183		2.3060	
			-0.7340	-0.8898	-1.1617	-1.8548	-2.2920	-2.3260	-2.3456		1.4455	
	6		-0.1375	-0.1649	-0.2031	-0.2638	-0.2870	-0.2853	-0.2869		0.7687	
			-0.5533	-0.6878	-0.8719	-1.2074	-1.3150	-1.3538	-1.3007		0.4818	
Asset	8		-0.1305	-0.1519	-0.1936	-0.2409	-0.2647	-0.2677	-0.2685		0.5765	
jump			-0.5256	-0.6309	-0.8373	-1.0847	-1.2228	-1.2410	-1.2366		0.3614	
hedging	11		-0.1291	-0.1481	-0.1851	-0.2336	-0.2394	-0.2456	-0.2479		0.4193	
(/myopic)			-0.5211	-0.6196	-0.8128	-1.0783	-1.1014	-1.1124	-1.1316		0.2628	
	16		-0.1245	-0.1458	-0.1810	-0.2240	-0.2392	-0.2389	-0.2396		0.2883	
			-0.5150	-0.6159	-0.7964	-1.0151	-1.1033	-1.0998	-1.0736		0.1807	
	21		-0.1241	-0.1438	-0.1760	-0.2210	-0.2331	-0.2314	-0.2302		0.2196	
			-0.5054	-0.6058	-0.7697	-0.9954	-1.0658	-1.0543	-1.0374		0.1377	
	41		-0.1186	-0.1406	-0.1731	-0.2113	-0.2252	-0.2239	-0.2213		0.1125	
			-0.4885	-0.5877	-0.7536	-0.9612	-1.0351	-1.0130	-1.0059		0.0705	

This table dictates the asset jump hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_J^{(i)} = -0.1, \sigma_J^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1, c_2 = 0.5, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.3: Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled

	RRA	T (year)		0.5	1	2	5	7	10	Myopic Demand	
		0.25	0.5								
Covariance jump hedging (/myopic)	2	0.0375	0.0805	0.1645	0.3402	0.4720	0.4830	0.4755	2.3060		
	6	0.1398	0.2978	0.5757	1.2052	1.5945	1.6309	1.6497	1.4455		
		0.0411	0.0823	0.1376	0.2105	0.2370	0.2349	0.2356	0.7687		
		0.1812	0.3685	0.6182	0.9939	1.1137	1.1376	1.0952	0.4818		
	8	0.0413	0.0797	0.1359	0.1961	0.2221	0.2264	0.2254	0.5765		
		0.1807	0.3540	0.6127	0.9065	1.0475	1.0688	1.0615	0.3614		
	11	0.0427	0.0799	0.1330	0.1950	0.2049	0.2090	0.2112	0.4193		
		0.1880	0.3557	0.6096	0.9180	0.9611	0.9676	0.9843	0.2628		
	16	0.0426	0.0806	0.1327	0.1894	0.2060	0.2065	0.2069	0.2883		
		0.1910	0.3598	0.6045	0.8783	0.9731	0.9691	0.9477	0.1807		
	21	0.0432	0.0807	0.1302	0.1887	0.2021	0.2015	0.1999	0.2196		
		0.1899	0.3596	0.5881	0.8703	0.9411	0.9340	0.9202	0.1377		
	41	0.0424	0.0801	0.1304	0.1816	0.1984	0.1963	0.1943	0.1125		
		0.1886	0.3548	0.5869	0.8444	0.9272	0.9079	0.9012	0.0705		

This table dictates the covariance jump hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_J^{(i)} = -0.1, \sigma_J^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1, c_2 = 0.5, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.4: Optimal Hedging Demands for 2 Risky Assets with Two Jumps Entangled

RRA	T (year)	Myopic Demand									
		0.25	0.5	1	2	5	7	10			
2		0.0195	0.0369	0.0613	0.0953	0.1102	0.1105	0.1105	2.3060		
		0.0091	0.0210	0.0423	0.0787	0.0965	0.0970	0.0969	1.4455		
6		0.0338	0.0644	0.1046	0.1484	0.1598	0.1600	0.1598	0.7687		
		0.0158	0.0367	0.0714	0.1173	0.1309	0.1311	0.1309	0.4818		
8		0.0357	0.0678	0.1099	0.1545	0.1655	0.1655	0.1656	0.5765		
Variance hedging (/myopic)		0.0168	0.0387	0.0749	0.1217	0.1348	0.1348	0.1349	0.3614		
11		0.0371	0.0707	0.1144	0.1597	0.1703	0.1702	0.1703	0.4193		
		0.0175	0.0403	0.0780	0.1255	0.1381	0.1380	0.1382	0.2628		
16		0.0384	0.0732	0.1180	0.1635	0.1745	0.1742	0.1744	0.2883		
		0.0181	0.0418	0.0804	0.1281	0.1411	0.1408	0.1411	0.1807		
21		0.0390	0.0743	0.1197	0.1661	0.1764	0.1764	0.1764	0.2196		
		0.0184	0.0425	0.0815	0.1300	0.1423	0.1424	0.1424	0.1377		
41		0.0400	0.0764	0.1227	0.1694	0.1795	0.1796	0.1796	0.1125		
		0.0189	0.0437	0.0835	0.1324	0.1445	0.1446	0.1446	0.0705		

This table dictates the variance hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = J_{12} = 0$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_j^{(i)} = -0.1$, $\sigma_j^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1$, $c_2 = 0.5$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.5: Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps

	RRA	T (year)		0.5	1	2	5	7	10	Myopic Demand
		0.25	0.5							
	2	0.0085	0.0176	0.0321	0.0546	0.0650	0.0653	0.0653	0.0653	2.3060
		0.0211	0.0437	0.0796	0.1353	0.1612	0.1612	0.1619	0.1618	1.4455
	6	0.0148	0.0307	0.0544	0.0828	0.0907	0.0907	0.0908	0.0907	0.7687
		0.0366	0.0762	0.1349	0.2054	0.2250	0.2250	0.2253	0.2250	0.4818
	8	0.0156	0.0324	0.0571	0.0861	0.0937	0.0937	0.0937	0.0938	0.5765
		0.0386	0.0803	0.1417	0.2135	0.2323	0.2323	0.2324	0.2326	0.3614
	11	0.0162	0.0338	0.0595	0.0889	0.0962	0.0962	0.0962	0.0962	0.4193
		0.0403	0.0838	0.1475	0.2204	0.2385	0.2385	0.2385	0.2386	0.2628
	16	0.0168	0.0350	0.0613	0.0908	0.0984	0.0984	0.0983	0.0984	0.2883
		0.0417	0.0867	0.1520	0.2253	0.2441	0.2441	0.2437	0.2440	0.1807
	21	0.0170	0.0355	0.0622	0.0922	0.0994	0.0994	0.0994	0.0994	0.2196
		0.0423	0.0881	0.1542	0.2288	0.2465	0.2465	0.2466	0.2466	0.1377
	41	0.0175	0.0365	0.0637	0.0940	0.1010	0.1010	0.1011	0.1011	0.1125
		0.0434	0.0906	0.1580	0.2332	0.2506	0.2506	0.2507	0.2507	0.0705

This table dictates the covariance hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = J_{12} = 0$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_j^{(i)} = -0.1$, $\sigma_j^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1$, $c_2 = 0.5$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.6: Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps

RRA	T (year)	Myopic Demand									
		0.25	0.5	1	2	5	7	10			
Asset jump hedging (/myopic)	2	-0.1756	-0.2252	-0.3081	-0.5195	-0.6306	-0.6525	-0.6518	2.3060		
		-0.1016	-0.1161	-0.1401	-0.1822	-0.2048	-0.2066	-0.2111	1.4455		
	6	-0.0816	-0.0918	-0.1079	-0.1316	-0.1373	-0.1345	-0.1409	0.7687		
		-0.0844	-0.0959	-0.1103	-0.1332	-0.1425	-0.1427	-0.1424	0.4818		
	8	-0.0755	-0.0887	-0.1037	-0.1211	-0.1271	-0.1281	-0.1250	0.5765		
		-0.0829	-0.0932	-0.1081	-0.1292	-0.1351	-0.1363	-0.1381	0.3614		
	11	-0.0738	-0.0822	-0.0962	-0.1173	-0.1226	-0.1184	-0.1161	0.4193		
		-0.0818	-0.0961	-0.1080	-0.1273	-0.1313	-0.1340	-0.1329	0.2628		
	16	-0.0694	-0.0807	-0.0918	-0.1086	-0.1131	-0.1149	-0.1111	0.2883		
		-0.0812	-0.0916	-0.1081	-0.1246	-0.1295	-0.1318	-0.1326	0.1807		
41	21	-0.0705	-0.0804	-0.0933	-0.1047	-0.1085	-0.1078	-0.1095	0.2196		
		-0.0817	-0.0922	-0.1048	-0.1288	-0.1274	-0.1328	-0.1289	0.1377		
		-0.0692	-0.0749	-0.0894	-0.1033	-0.1060	-0.1055	-0.1057	0.1125		
		-0.0794	-0.0898	-0.1038	-0.1238	-0.1315	-0.1285	-0.1293	0.0705		

This table dictates the asset jump hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = J_{12} = 0$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_j^{(i)} = -0.1$, $\sigma_j^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1$, $c_2 = 0.5$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.7: Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps

RRA	T (year)	Myopic Demand							
		0.25	0.5	1	2	5	7	10	Demand
2		0.0301	0.0680	0.1372	0.3114	0.4044	0.4204	0.4222	2.3060
		0.0267	0.0524	0.0891	0.1435	0.1700	0.1731	0.1748	1.4455
6		0.0241	0.0451	0.0717	0.1035	0.1114	0.1084	0.1136	0.7687
		0.0365	0.0653	0.0968	0.1310	0.1421	0.1437	0.1418	0.4818
8		0.0236	0.0456	0.0712	0.0974	0.1048	0.1053	0.1030	0.5765
Covariance		0.0376	0.0660	0.0977	0.1291	0.1366	0.1375	0.1398	0.3614
jump		0.0240	0.0432	0.0680	0.0967	0.1031	0.0986	0.0970	0.4193
hedging		0.0385	0.0703	0.0996	0.1294	0.1342	0.1369	0.1368	0.2628
(/myopic)		0.0234	0.0441	0.0660	0.0892	0.0968	0.0973	0.0946	0.2883
16		0.0395	0.0693	0.1014	0.1277	0.1351	0.1362	0.1381	0.1807
21		0.0240	0.0443	0.0679	0.0883	0.0928	0.0921	0.0938	0.2196
		0.0400	0.0697	0.0982	0.1325	0.1338	0.1392	0.1351	0.1377
41		0.0244	0.0423	0.0661	0.0877	0.0920	0.0916	0.0911	0.1125
		0.0399	0.0693	0.0978	0.1296	0.1374	0.1348	0.1367	0.0705

This table dictates the covariance jump hedging demand as a proportion to myopic demand for the model with two jumps scaled by $J_{21} = J_{12} = 0$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are as follows: $\mu_J^{(i)} = -0.1, \sigma_J^{(i)} = 0.2$ for $Z_t^{(i)}$ and $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1, c_2 = 0.5, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.8: Optimal Hedging Demands for 2 Risky Assets with Two Independent Jumps

RRA	T (year)	Myopic Demand							
		0.25	0.5	1	2	5	7	10	Demand
2		0.0176	0.0311	0.0471	0.0661	0.0759	0.0760	0.0760	2.3060
		0.0058	0.0127	0.0240	0.0419	0.0533	0.0538	0.0537	1.4455
6		0.0305	0.0543	0.0818	0.1107	0.1198	0.1202	0.1201	0.7687
		0.0113	0.0241	0.0436	0.0714	0.0817	0.0824	0.0824	0.4818
8		0.0321	0.0573	0.0864	0.1161	0.1253	0.1255	0.1253	0.5765
		0.0120	0.0257	0.0465	0.0748	0.0857	0.0859	0.0857	0.3614
11		0.0335	0.0598	0.0898	0.1202	0.1297	0.1299	0.1298	0.4193
		0.0126	0.0269	0.0484	0.0774	0.0885	0.0888	0.0886	0.2628
16		0.0346	0.0617	0.0929	0.1240	0.1334	0.1333	0.1336	0.2883
		0.0131	0.0278	0.0503	0.0799	0.0910	0.0908	0.0914	0.1807
21		0.0352	0.0630	0.0948	0.1261	0.1352	0.1352	0.1350	0.2196
		0.0133	0.0286	0.0518	0.0814	0.0923	0.0922	0.0917	0.1377
41		0.0362	0.0644	0.0970	0.1289	0.1378	0.1382	0.1382	0.1125
		0.0138	0.0291	0.0528	0.0833	0.0937	0.0940	0.0942	0.0705

This table dictates the variance hedging demand as a proportion to myopic demand for the model with only jumps in returns scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are: $\mu_j^{(i)} = -0.1, \sigma_j^{(i)} = 0.2$ for $Z_t^{(i)}, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2, 1.5 * C_2)$. Please refer to model details for the explicit formulation of parameters.

Table 3.9: Optimal Variance Hedging Demands for 2 Risky Assets with Jumps in Returns Only

RRA	T (year)	Myopic Demand							
		0.25	0.5	1	2	5	7	10	
Covariance hedging (/myopic)	2	0.0069	0.0132	0.0218	0.0336	0.0403	0.0404	0.0404	2.3060
		0.0171	0.0327	0.0541	0.0833	0.0998	0.1003	0.1002	1.4455
	6	0.0123	0.0237	0.0385	0.0566	0.0628	0.0631	0.0631	0.7687
		0.0306	0.0587	0.0955	0.1403	0.1557	0.1564	0.1564	0.4818
	8	0.0130	0.0250	0.0408	0.0593	0.0657	0.0658	0.0657	0.5765
		0.0323	0.0621	0.1013	0.1471	0.1628	0.1632	0.1629	0.3614
	11	0.0136	0.0262	0.0425	0.0614	0.0679	0.0681	0.0680	0.4193
		0.0337	0.0649	0.1053	0.1522	0.1684	0.1688	0.1686	0.2628
	16	0.0141	0.0270	0.0440	0.0633	0.0698	0.0697	0.0700	0.2883
		0.0349	0.0669	0.1091	0.1571	0.1732	0.1729	0.1736	0.1807
	21	0.0143	0.0276	0.0450	0.0644	0.0707	0.0707	0.0705	0.2196
	41	0.0355	0.0685	0.1116	0.1598	0.1754	0.1754	0.1749	0.1377
		0.0148	0.0282	0.0460	0.0659	0.0720	0.0722	0.0723	0.1125
		0.0366	0.0699	0.1141	0.1634	0.1785	0.1792	0.1793	0.0705

This table dictates the covariance hedging demand as a proportion to myopic demand for the model with only jumps in returns scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are: $\mu_j^{(i)} = -0.1$, $\sigma_j^{(i)} = 0.2$ for $Z_t^{(i)}$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2, 1.5 * C_2)$. Please refer to model details for the explicit formulation of parameters.

Table 3.10: Optimal Covariance Hedging Demands for 2 Risky Assets with Jumps in Returns Only

RRA	T (year)	Myopic Demand									
		0.25	0.5	1	2	5	7	10			
2		-0.1958	-0.2084	-0.2237	-0.2419	-0.2518	-0.2587	-0.2568	2.3060		
		-0.6439	-0.6655	-0.6980	-0.7677	-0.7870	-0.7789	-0.7869	1.4455		
6		-0.1230	-0.1319	-0.1424	-0.1523	-0.1583	-0.1562	-0.1578	0.7687		
		-0.4775	-0.4980	-0.5521	-0.5928	-0.6209	-0.6125	-0.6089	0.4818		
8		-0.1188	-0.1249	-0.1344	-0.1470	-0.1515	-0.1513	-0.1530	0.5765		
		-0.4594	-0.4856	-0.5291	-0.5792	-0.5917	-0.5909	-0.5952	0.3614		
11		-0.1151	-0.1219	-0.1325	-0.1452	-0.1465	-0.1455	-0.1466	0.4193		
		-0.4465	-0.4732	-0.5232	-0.5707	-0.5783	-0.5770	-0.5807	0.2628		
16		-0.1112	-0.1206	-0.1287	-0.1395	-0.1414	-0.1439	-0.1421	0.2883		
		-0.4419	-0.4713	-0.5107	-0.5609	-0.5670	-0.5723	-0.5592	0.1807		
21		-0.1117	-0.1157	-0.1255	-0.1372	-0.1418	-0.1424	-0.1431	0.2196		
		-0.4376	-0.4580	-0.4921	-0.5502	-0.5539	-0.5612	-0.5714	0.1377		
41		-0.1072	-0.1170	-0.1249	-0.1367	-0.1407	-0.1374	-0.1378	0.1125		
		-0.4272	-0.4635	-0.4977	-0.5401	-0.5557	-0.5574	-0.5524	0.0705		

This table dictates the asset jump hedging demand as a proportion to myopic demand for the model with only jumps in returns scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are: $\mu_f^{(i)} = -0.1, \sigma_f^{(i)} = 0.2$ for $Z_t^{(i)}, i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2, 1.5 * C_2)$. Please refer to model details for the explicit formulation of parameters.

Table 3.11: Optimal Asset Jump Hedging Demands for 2 Risky Assets with Jumps in Returns Only

RRA	T (year)	Myopic Demand								
		0.25	0.5	1	2	5	7	10	10	Demand
2		0.0218	0.0411	0.0670	0.1001	0.1129	0.1132	0.1132	0.1132	2.3060
		0.0111	0.0248	0.0477	0.0827	0.0978	0.0981	0.0981	0.0981	1.4455
6		0.0365	0.0685	0.1084	0.1487	0.1584	0.1584	0.1584	0.1584	0.7687
		0.0186	0.0410	0.0755	0.1177	0.1293	0.1292	0.1293	0.1293	0.4818
8		0.0384	0.0720	0.1136	0.1545	0.1638	0.1639	0.1639	0.1639	0.5765
		0.0195	0.0430	0.0790	0.1217	0.1329	0.1330	0.1330	0.1330	0.3614
11		0.0399	0.0748	0.1177	0.1592	0.1683	0.1683	0.1684	0.1684	0.4193
		0.0203	0.0446	0.0817	0.1250	0.1359	0.1359	0.1360	0.1360	0.2628
16		0.0412	0.0772	0.1212	0.1631	0.1720	0.1721	0.1720	0.1720	0.2883
		0.0210	0.0460	0.0840	0.1277	0.1383	0.1385	0.1384	0.1384	0.1807
21		0.0418	0.0784	0.1230	0.1651	0.1739	0.1740	0.1740	0.1740	0.2196
		0.0213	0.0467	0.0851	0.1290	0.1396	0.1397	0.1397	0.1397	0.1377
41		0.0429	0.0803	0.1259	0.1683	0.1769	0.1770	0.1770	0.1770	0.1125
		0.0218	0.0479	0.0870	0.1312	0.1416	0.1417	0.1417	0.1417	0.0705

This table dictates the variance hedging demand as a proportion to myopic demand for the model with only jumps in covariance matrix scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are: $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1$, $c_2 = 0.5$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.12: Optimal Variance Hedging Demands for 2 Risky Assets with Jumps in Covariance Matrix Only

RRA	T (year)	Myopic Demand								
		0.25	0.5	1	2	5	7	10		
Covariance hedging (/myopic)	2	0.0098	0.0201	0.0356	0.0571	0.0660	0.0661	0.0662	2.3060	
		0.0244	0.0499	0.0882	0.1417	0.1636	0.1640	0.1641	1.4455	
	6	0.0165	0.0334	0.0569	0.0831	0.0898	0.0897	0.0898	0.7687	
		0.0409	0.0828	0.1411	0.2060	0.2226	0.2226	0.2226	0.4818	
	8	0.0173	0.0351	0.0596	0.0861	0.0926	0.0926	0.0926	0.5765	
		0.0430	0.0869	0.1477	0.2134	0.2295	0.2297	0.2296	0.3614	
	11	0.0180	0.0364	0.0617	0.0886	0.0949	0.0949	0.0949	0.4193	
		0.0446	0.0903	0.1529	0.2197	0.2352	0.2353	0.2354	0.2628	
	16	0.0186	0.0376	0.0635	0.0906	0.0967	0.0968	0.0968	0.2883	
		0.0461	0.0931	0.1574	0.2246	0.2399	0.2401	0.2400	0.1807	
	21	0.0189	0.0382	0.0643	0.0916	0.0978	0.0978	0.0978	0.2196	
		0.0468	0.0946	0.1595	0.2272	0.2424	0.2425	0.2425	0.1377	
	41	0.0193	0.0391	0.0658	0.0933	0.0993	0.0994	0.0993	0.1125	
		0.0480	0.0969	0.1632	0.2313	0.2462	0.2464	0.2464	0.0705	

This table dictates the covariance hedging demand as a proportion to myopic demand for the model with only jumps in covariance matrix scaled by $J_{21} = 0.5$ and $J_{12} = 0.05$ with various relative risk aversion level and investment time horizon. The last two columns report total demand and myopic demand, respectively. The hedging demands are computed with the monthly parameter obtained in Buraschi et al. (2010) supplemented with reasonable jump sizes in returns and covariance. More specifically, the parameters of the jump distributions applied are: $\mu_v^{(i)} = 0.1$ for K_t^{ii} with $c_1 = 0.1$, $c_2 = 0.5$, $i = 1, 2$. While, the jump intensity parameters, C_1, C_2 are set as: $C_2 = (1, 1)'$ and $C_1 = (C_2', 1.5 * C_2')$. Please refer to model details for the explicit formulation of parameters.

Table 3.13: Optimal Covariance Hedging Demands for 2 Risky Assets with Jumps in Covariance Matrix Only

$vec(M)'$	-1.122	0.884	0.747	-0.888
$vec(Q)'$	0.16	-0.021	0.083	0.009
ρ'	-0.279	-0.247		
η'	4.612	2.891		

This table summarizes the monthly estimates adopted from Buraschi et al. (2010). This set of parameters is augmented with reasonable jump parameters and applied for the numerical experiments. With respect to the jump parameters, they are reported in the footnotes below each table for clarity.

Table 3.14: Parameters adopted for numerical experiments.

Chapter 4

Estimation for Multivariate Stochastic Volatility Models

4.1 Introduction

Coming up with a model which composes good properties and capable of capturing desired empirical features may be hard, nevertheless, most of the time, developing an estimation procedure for the inference problem is even harder. In this paper, we developed the detailed estimation technique and procedures based on the Bayesian Markov Chain Monte Carlo (MCMC) method for the general multivariate jump diffusion model with stochastic covariance matrix modeled as Wishart process, denoted as WJD (Wishart-Jump-Diffusion) model¹ hereafter. This novel model is introduced in Jin et al. (2016), where they solved the optimal portfolio choice problem under contagion with closed-form solution. Besides, several features and implications of financial contagion were demonstrated by various comparative statistics and numerical analysis. Moreover, several existing models are embedded in this more sophisticated model. As a result, an easy and sufficiently efficient estimation technique for inferring this kind of complex multivariate models would be of great research value. Given inference tools at hand, theoretical models can be verified by fitting them with empirical data and hence provide potential explanations to existing puzzles in the literature such as under diversification, home bias, financial contagion, etc.

To achieve this goal, we rely on the desirable properties of MCMC, especially

¹For instance, WJD-DJ model denotes the double jump model with stochastic covariance matrix modeled as Wishart process, while, WD (Wishart-Diffusion) model denotes the model without jump components as in Buraschi et al. (2010), if not mentioned otherwise.

its efficiency. Andersen et al. (1999) documented that MCMC may be one of the most efficient method for inference, where various finite sample comparisons were conducted by Monte Carlo studies. The other reason why we decided to apply MCMC method for our estimation procedure other than existing popular methods is that the growing dimension of asset pricing models. It is reasonable that adding parameters brings more desirable properties for a model, however, the resulting complexity makes the estimation even more difficult. For example, WJD model consists of 4 state variables, including covariance structure, jump times, jump sizes of return and covariance structure and 31-dimensional parameter space are incorporated.² As a result, if moment-matching techniques, e.g., generalized method of moments (GMM) or Efficient method of moments (EMM) methods are applied, at least 32 moments are required for over relaxation condition. And it immediately brings about three main drawbacks of moment-matching techniques. Firstly, the computing complexity increases as the number of moments required. In particular, in the context of optimal portfolio choice problem, moments are often obtainable from solving the ODEs of the corresponding moment generating functions. Secondly, how to choose the proper set of moment conditions remains tricky especially for high dimensional cases. In some extreme cases, different sets of moment conditions even yield different results. The last but not the least, the minimization problem involved in moment-matching methods also needs to be examined carefully. Since the key idea of moment-matching method is to find a set of parameter that minimize a certain weighted quadratic form of deviation from the target moment conditions, finding the optimal set of parameters which has the global minimum of the deviation is hence the most important part of the estimation question. In general, there is no universal way of optimization and trial and error is often applied. In some worse cases, similar values of some plausible global minimums may be obtained but each composes greatly different parameters. Hence it remains tricky to find the ideal/true parameters interested. For more detailed discussions about simulation-based estimation techniques, we refer to Yu et al (2010) where several methods such as ML (Maximum Likelihood) method, GMM (Generalized Method of Moment) method, EMM (Efficient Method of Moment) method, etc are well surveyed.

The remaining of the paper is organized as follows: Section 4.2 briefly reviews the importance of multivariate stochastic volatility models and existing estima-

²The exact number of parameter space is subject to the specification of the model parameters. The 31 is calculated based on the specification defined in the Section 4.3.

tion techniques for comparison with our technique, meanwhile, the drawbacks of other methods manifest why the proposed technique is of great research value. Section 4.3 describes the details of the model, including the specification of jump components and parameter settings. Section 4.4 illustrates the building blocks of proposed MCMC estimation procedure with detailed derivations for the posterior distributions. Next, Section 4.6 demonstrates the accuracy of our estimation procedure with various models, including standard Merton-Jump-Diffusion model with entangled jumps, common stochastic volatility models, Wishart model and the most complicated one, WJD-DJ model. Finally, Section 4.7 summarizes the concluding remarks and potential applications for future researches.

4.2 Literatures Reviews

The multivariate stochastic volatility models has gained more and more attention because stochastic covariance structure and correlations are found in the literature. For example, Asai and McAleer (2009) documented that dynamic correlation is persistent even in long time horizon, but constant in traditional DCC model. Ang and Chen (2002) investigated the asymmetric correlation by calculating the exceedance correlation and found correlation are market-status dependent. This kind of asymmetric feature of correlation may be linked the famous home puzzle, first introduced in French and Poterba (1991). Moreover, ignoring the stochastic property of correlation and covariance structure in fact leads to great economic loss (see Buraschi et al. (2010) and Jin et al. (2016)).

The diffusion Wishart process, introduced in Bru (1991), has become popularized since Gouriéroux (2006) where detailed properties, theories and financial application about Wishart process are thoroughly explained and discussed. In fact, Wishart process, or Wishart-Diffusion model is an ideal candidate for modeling the dynamic of variance-covariance matrix because it is positive-semidefinite with some moderate conditions satisfied. In addition, it's affine structure makes the optimal portfolio choice problem solvable in closed-form solution, e.g. Buraschi et al. (2010). More recently, a more general model, Wishart-Jump-Diffusion (WJD) model has been studied in Jin et al. (2016) with novel parameters integrated in the jump components to capture observed features of log returns. Moreover, the optimal portfolio problem is also solved with closed-form solution. By alternating the novel parameters, the resulting

optimal portfolio weights reflect the impacts resulting from the observed features. Hence, the WJD model can be regarded as a new alternative to study financial contagion

Regarding the estimation problem, Buraschi et al. (2010) provided an estimation procedure for the Wishart-Diffusion model based on GMM (Generalized Method of Moments) method. Nevertheless, inference problem for the more general WJD model where jumps are allowed to be incorporated in returns and variance-covariance is not well explored and established. In summary, the estimation techniques for multivariate stochastic volatility models remain challenging and not well established because of several difficulties, e.g. higher dimension of parameter and state space, required positive definitive property, complicated likelihood functions.

The main contribution of this paper is to develop a general estimation procedure for the WJD model, which is nested with various multivariate stochastic volatility models. In addition, by reducing the dimension to one, common univariate stochastic volatility models, e.g. Eraker et al (2003), among others, are also embedded. To circumvent the difficulties encountered with other estimation techniques such as GMM and SMM (Simulated Method of Moments) methods as discussed in the Introduction section, the developed estimation procedure is based on Markov Chain Monte Carlo (MCMC) (see Johannes and Polson (2003), among others) and slice sampling techniques (Neal (2003), etc).

There are quite a few MCMC techniques and are often categorized into two types: the Gibbs sampler and the Metropolis-Hasting (MH) algorithm, which includes random walk and independent Metropolis schemes. The difference between Gibbs sampling and MH algorithm is whether the posterior distribution to be sampled is a standard distribution or not. Given the posterior distribution standard and direct random sampling is feasible, Gibbs sampling is what to be applied. While for those with non-standard posterior distributions, MH algorithm is typically the first attempt for sampling.

In particular, Gibbs sampling had gained more and more popularity since Gelfand and Smith (1992). Introductions for MCMC can be found in Casella and George (1992), Gelman et al. (2014), Tierney (1994), etc. Since MH algorithm often requires careful tuning procedure to ensure efficient sampling results, there have been some advanced algorithms developed to improve the ef-

efficiency of MH algorithm. For example, Adaptive rejection metropolis (ARMS) in Gilk et al. (1995), Adaptive Metropolis (AM): Haario et al. (2001), Delayed rejection adaptive Metropolis (DRAM): Haario et al (2006), are shown have better performance compared to pure MH algorithm. Among these improvements of MH scheme, the key idea is the essentially same, i.e., improve the efficiency of the acceptance-rejection step of MH algorithm (see Gilk et al. (1995)). Yu and Meyer (2006) documented various comparisons between stochastic volatility models based on Bayesian estimation. In particular, the estimation is done with WinBuG, a program designed for Bayesian estimation using Gibbs sampler.

These methods work good in single dimensional problems and "well-shaped" posterior distributions, e.g., log concave distributions, nevertheless, dealing with higher dimensional models remains difficult and needs tricky tuning procedure. Advantage of Bayesian MCMC methods is three-fold. Firstly, with augmented state variables, no asymptotic argument required as other methods. Secondly, smoothed estimates of latent variables are by-products of the MCMC algorithm, which may provide general pictures about the observed states. The developed model may be further extended with existing techniques about identifying state variables augmented. For instance, if high-frequency data are available, the volatility state may be approximated as realized volatility (Andersen et al. (2003)), while the jump events could be also identified with jump detection techniques (Lee and Hannig (2010), Lee and Mykland (2008), among others). Given these approximated state variables, the estimation procedure proceeds as these states fixed and yields the resulting parameters. In particular, approximating state variables in advance may not be necessary but indeed provides better understanding about the data since analysis of high frequency data is involved. Moreover, the time consumption would be greatly reduced because the volatility (or covariance) needs to be sampled sequentially and parallel computing is not feasible. Hence, it is always the most time-consuming component.

Apart from augmenting existing techniques, the estimation procedure developed in this paper is based on single-move algorithm. Extending it to a multi-move algorithm such as Kim et al. (1998), Liesenfel and Richard (2006), Omori et al. (2007), among others may be of great value and potential interest for future researches.

4.3 The Model

For illustrative purpose, the 2 dimensional WJD model is reviewed here.³ Assume we have two stocks and two jumps, i.e., $n = m = 2$. The dynamics of the price vector $S_t = (S_{1,t}, \dots, S_{n,t})$, $t \in [0, T]$ of the risky assets and covariance matrix $\Sigma_t = (\Sigma_{ij,t})_{n \times n} = \Sigma_t^{1/2} \cdot (\Sigma_t^{1/2})'$ are governed by the stochastic differential equations (SDE):

$$\begin{aligned} \text{diag}(S_t^{-1})dS_t &= (\mu_t + r1_n)dt + \Sigma_t^{1/2}dW_t + J \cdot (Y_t dN_t) \\ d\Sigma_t &= [\Omega\Omega' + M\Sigma_t + \Sigma_t M']dt + \Sigma_t^{1/2}dZ(t)Q + Q'dZ'\Sigma_t^{1/2} + \sum_{i=1}^m K^{(i)}dN_i \end{aligned} \quad (4.1)$$

where $r > 0$, $1_n = (1, \dots, 1)' \in \mathcal{R}^{n \times 1}$, $\mu_t \in \mathcal{R}^{n \times 1}$, $J \in \mathcal{R}^{n \times m}$ are constant or deterministic vectors or matrices; $\Omega, M, Q \in \mathcal{R}^{n \times n}$ are constant matrix with M is negative definite. $\Sigma^{1/2} = (\sigma_{ij})_{n \times n}$ is the (volatility) coefficient matrix of the diffusion terms. W_t and Z_t are $n \times 1$ and $n \times n$ standard Brownian motions. In particular, they are correlated through the way

$$W_t = \sqrt{1 - \rho'\rho}Z_t^\perp + Z_t\rho, \quad (4.3)$$

where Z_t^\perp is an n dimensional standard Brownian motion independent of Z_t , and $\rho = (\rho_1, \dots, \rho_n)' \in \mathcal{R}^{n \times 1}$ is with the properties $\rho_i \in [-1, 1]$, $i = 1, 2, 3, \dots, n$ and $\rho'\rho \leq 1$.

Moreover, the variable μ_t is set as

$$\mu_t = \Sigma_t \eta_t - J \cdot \mathbf{E}[Y] \lambda_t = \Sigma_t \eta_t - J \cdot \mathbf{E}[Y] \cdot C_1 \Sigma_t C_2,$$

, which is an affine structure of Σ_t as conventionally adopted in the literature of stochastic volatility models. While the m dimensional jump counting process $N_t = (N_{1,t}, \dots, N_{m,t})'$ is independent of the Brownian motions, and $N_{i,t}$ is independent of $N_{j,t}$ for $1 \leq i \neq j \leq m$. In particular, the jump intensity is specified by $\lambda_t = C_1 \Sigma_t C_2$ with constants matrix $C_1 \in \mathcal{R}^{m \times n}$, $C_2 \in \mathcal{R}^{n \times 1}$. Moreover, the Y is a diagonal matrix composed of independent random jump sizes, (Y_1, \dots, Y_m) i.e., $Y = \text{diag}(Y_1, \dots, Y_m)$ where Y_i is independent of Y_j for $1 \leq i \neq j \leq m$. In particular, $Y \sim \exp(Z_t) - 1$. J is a constant scaling matrix for jump sizes in return. It is worth mentioning that the novel features of parameter J, C_1, C_2 and the jump intensity specification $\lambda_t = C_1 \Sigma_t C_2$ were analyzed in Jin et al.

³The nested models could be easily obtained by varying the corresponding parameters and ignored here.

(2016) in the context of optimal portfolio choice problem, where details and interpretations about WJD model had been explored well.

For illustrative simplicity, we follow the specification defined in Jin et al. (2016), i.e.,

$$K_t^{(i)} = \begin{pmatrix} K_t^{11} & c_i \sqrt{K_t^{11} K_t^{22}} \\ c_i \sqrt{K_t^{11} K_t^{22}} & K_t^{22} \end{pmatrix} \quad (4.4)$$

$$C_1 = \begin{pmatrix} 1 \cdot C_2(1) & 1 \cdot C_2(2) \\ C_1(1) \cdot C_2(1) & C_1(1) \cdot C_2(2) \end{pmatrix}, C_2 = \begin{pmatrix} C_2(1) \\ C_2(2) \end{pmatrix}. \quad (4.5)$$

where $K_t^{ii} \sim \text{Exp}(\mu_v^{(i)}), -1 \leq c_i \leq 1, i = 1, 2$

4.4 MCMC Estimation Implementation

The estimation technique developed in this paper hinges on the Markov Chain Monte Carlo (MCMC) method, a Bayesian style inference technique and slice sampling technique is integrated for parameters/state variables without standard posterior distribution. Other than matching the moments or other characteristics of target models, MCMC method tackles inference problems by drawing samples from the corresponding posterior distributions implied by the target model and empirical data. Besides, unlike maximum likelihood method, all target posterior distributions are only required to be known up to certain constant, i.e., calculation of complicated normalization constants of the distributions are circumvented. This important advantage of MCMC method makes complicated inference problem feasible since sometimes the normalization constants required in maximum likelihood methods incorporate several complicated integrals and hence hard and impossible to be calculated in practice. In addition, the MCMC technique also avoids optimization procedure, which is the key component in moment-matching methods and the global optimum is often not guaranteed.

4.4.1 Model Discretization

The building block of an MCMC estimation algorithm relies on the time-discretization of the model. Given time t and predefined time step, $dt = \Delta$, the

discretized model could be expressed as:

$$\begin{cases} y_{t+\Delta} = \mu \Delta + \Sigma_t^{1/2} dW_{t+\Delta} + \ln(1 + J \cdot (Y_{t+\Delta} dN_{t+\Delta})), \\ \Sigma_{t+\Delta} = \Sigma_t + [\Omega\Omega' + M\Sigma_t + \Sigma_t M'] \Delta + \Sigma_t^{1/2} dZ_{t+\Delta} Q + Q' dZ'_{t+\Delta} \Sigma_t^{1/2} + \sum_{i=1}^m K_{t+\Delta}^{(i)} dN_{t+\Delta} \end{cases}$$

where

$$\mu_t = \Sigma_t \eta - J \cdot \mathbf{E}[Y] \cdot C_1 \Sigma_t C_2$$

is the stationary mean of the log return, $y_t = d(\ln S_t)$ and $D(x)$ denotes the diagonal element of square matrix x .

Note that η would be obtained based on the estimated μ, Σ_t, C_1, C_2 . The reason why we estimate μ instead of η is that all other terms involved in μ_t possess their characteristics in specific parts of the model. Besides, μ is the drift term of the process, which determines the characteristic mean of observable log returns and setting it as a random variable like μ_t only increases the complexity of estimation procedure. In fact, similar procedures are adopted in the MCMC literature, see Eraker et al. (2003). Hence, without loss of generality, we estimate μ and extract η given other estimates obtained in the estimation procedure. $[t, t + dt]$,

Although the above discretized expression makes deriving posterior distributions of the state variables and parameters feasible, the symmetric expression of Σ_t would lead to degenerate multivariate normal distributions in the likelihood posterior distribution and make the sampling procedure complicated. We make the following decomposition of Σ_t :

$$\begin{cases} d\Sigma_t = dX_t + dX_t', \\ dX_t = [a + M\Sigma_t] \Delta + \Sigma_t^{1/2} dZ_{t+\Delta} Q + \frac{1}{2} \sum_{i=1}^m K_{t+\Delta}^{(i)} dN_{t+\Delta}. \end{cases}$$

where $a + a' = \Omega\Omega'$ is a positive semidefinite matrix to ensure Σ_t is a well defined variance covariance matrix.⁴ For notational simplicity, let $\xi_t^y = \ln(1 + J \cdot (Y_t dN_t))$ and $\xi_t^X = \text{vec}(\frac{1}{2} \sum_{i=1}^m K_t^{(i)} dN_t)$ ⁵, then the resulting time discretized model are converted into:

$$\begin{cases} y_{t+\Delta} = \mu \Delta + \Sigma_t^{1/2} dW_{t+\Delta} + \xi_{t+\Delta}^y, \\ X_{t+\Delta} = X_t + (a + M\Sigma_t) \Delta + \Sigma_t^{1/2} dZ_{t+\Delta} Q + \text{vec}^{-1}(\xi_{t+\Delta}^X). \end{cases} \quad (4.6)$$

⁴Please refer to Chapter 4 for further discussion

⁵ vec and vec^{-1} denote the vectorization and inverse vectorization of a matrix.

Equation (4.6) is the building block of the derivation of all likelihood function and posterior distribution. More specifically, the likelihood function of our model, described by y_t and X_t , could be expressed as the following:

$$\varepsilon_{t+\Delta} = \begin{pmatrix} \varepsilon_{t+\Delta}^y \\ \varepsilon_{t+\Delta}^X \end{pmatrix} = \begin{pmatrix} \Sigma_t^{1/2} dW_t \\ \text{vec}(\Sigma_t^{1/2} dZ_t Q) \end{pmatrix} \sim \mathcal{N}(0, V_t) \quad (4.7)$$

where

$$\varepsilon_{t+\Delta}^y = y_{t+\Delta} - \mu \Delta - \xi_{t+\Delta}^y \quad (4.8)$$

$$\varepsilon_{t+\Delta}^X = \text{vec}(X_{t+\Delta} - X_t + (a + M\Sigma_t)\Delta) - \xi_{t+\Delta}^X \quad (4.9)$$

$$V_{t+\Delta} = \begin{pmatrix} \Sigma_t & \rho' Q \otimes \Sigma_t \\ Q' \rho \otimes \Sigma_t & Q' Q \otimes \Sigma_t \end{pmatrix} \cdot \Delta \quad (4.10)$$

is the covariance matrix of $\varepsilon_t = (\varepsilon_t^y, \varepsilon_t^X)$. By blockwise inverse, we can obtain the inverse of V_t as:

$$V_t^{-1} = \frac{1}{\Delta} \begin{pmatrix} \frac{1}{1-\rho'\rho} \cdot \Sigma_t^{-1} & -\rho'(I_n - \rho\rho')^{-1}(Q')^{-1} \otimes \Sigma_t^{-1} \\ -Q'(I_n - \rho\rho')^{-1}\rho \otimes \Sigma_t^{-1} & Q^{-1}(I_n - \rho\rho')^{-1}(Q')^{-1} \otimes \Sigma_t^{-1} \end{pmatrix} \quad (4.11)$$

Given this expression of V_t^{-1} , the posterior distributions of Q and ρ are non-standard distribution and slice sampling (MH scheme for usual MCMC methods) are required. However, we exploit the following transformation to circumvent this. This transformation in deed substantially improved the performance of our estimation procedure since Q and ρ involves 6 dimensions of non-standard sampling even if n is only 2.

Let $\Omega = Q'(I_n - \rho\rho')Q$, and $\Psi = Q'\rho$, we transform (4.10) and (4.11) into:

$$V_t = \begin{pmatrix} \Sigma_t & \Psi' \otimes \Sigma_t \\ \Psi \otimes \Sigma_t & (\Omega + \Psi\Psi') \otimes \Sigma_t \end{pmatrix} \cdot \Delta \quad (4.12)$$

$$V_t^{-1} = \frac{1}{\Delta} \begin{pmatrix} (1 + \Psi'\Omega^{-1}\Psi) \cdot \Sigma_t^{-1} & -\Psi'\Omega^{-1} \otimes \Sigma_t^{-1} \\ -\Omega^{-1}\Psi \otimes \Sigma_t^{-1} & \Omega^{-1} \otimes \Sigma_t^{-1} \end{pmatrix} \quad (4.13)$$

In addition , the blockwise determinant of V_t is:

$$|V_t| = |\Sigma_t| \cdot |\Omega \otimes \Sigma_t| = |\Omega|^n \cdot |\Sigma_t|^{n+1} \quad (4.14)$$

From the above derivation, the likelihood function of the random variable, ε_t , which fully describes our model (4.6) could be calculated. In particular, the following derivation of posterior distributions among all state variables and parameters are based on terms calculated above.

The transformation of $(Q, \rho) \rightarrow (\Omega, \Psi)$ yields standard Normal-Inverse-Wishart (NIW) conjugate prior of the posterior distribution and hence improves the efficiency of our proposed MCMC estimation procedure. Similar transformation was applied in the literature, e.g., Jacquier et al. 1994, Li et al. (2008). Our transformation could be regarded as a high-dimensional extension and it is firstly derived for MCMC procedure in this paper. With this transformation, the leverage effects could be captured by the negative correlations between Brownian motion components in price and stochastic volatility.

To extract (Q, ρ) from estimates of (Ω, Ψ) , we calculate $Q'Q = \Omega + \Psi\Psi' = A$ and obtain Q by eigendecomposition of symmetric A . More specifically, simple linear algebra implies:

$$A = VD V' = \underbrace{V D^{\frac{1}{2}} V'}_{Q'} \cdot \underbrace{V D^{\frac{1}{2}} V'}_Q \quad (4.15)$$

$$\rho = (Q')^{-1} \Psi \quad (4.16)$$

where V and D are the eigen vector and diagonal eigen value matrices of A . Moreover, the identified Q and ρ are uniquely determined by the property of the symmetric matrix.

4.5 Posterior distribution derived for our model

In this section, we derive all posterior distributions applied in this paper and aim to provide a general guideline for related future researches. The general case $m = n$ is applied here and simplified cases as $m \leq n$ can easily be obtained. For notational clarity, we denote $Y, \Gamma(\Gamma_-), \Theta(\Theta_-), \theta$ to be the sets of observable time series, i.e., log return here, latent variables, model parameters and hyper parameters of priors, respectively, while the "–" denotes the corresponding complement set of a state variable or parameter.⁶ More specifically, Γ, Θ, θ rep-

⁶For example, the " Θ_- " in the likelihood function, $\mathcal{P}(a|Y, \Gamma, \Theta_-, \theta)$ denotes the set of all model parameters except for the parameter a

resent specific sets for each derivation if not noted otherwise. Besides, constant terms with respect to each posterior derivation are omitted for clarity and it is indeed an advantage of applying MCMC technique.

4.5.1 Posterior for parameters

- Posterior for μ : The prior of μ is assumed to follow a normal distribution, i.e., $\mu \sim \mathcal{N}(m_0, M_0)$ and the corresponding posterior is then derived as follows:

$$\begin{aligned}
\mathcal{P}(\mu|Y, \Gamma, \Theta_-) &\propto \mathcal{P}(Y, X|\mu, \Theta_-) \cdot \mathcal{P}(\mu|\theta) \\
&\propto \prod_{t=0}^{T-1} \mathcal{P}(y_{t+\Delta}, X_{t+\Delta}|\mu, \Theta_-, \Gamma_{t+\Delta}) \times e^{\frac{-1}{2}(\mu-m_0)'M_0^{-1}(\mu-m_0)} \\
&\propto \exp\left(\frac{-1}{2}\left\{\mu'\left(\sum_{t=0}^{T-1}((1 + \Psi'\Omega^{-1}\Psi) \cdot \Sigma_t^{-1}\Delta) + M_0^{-1})\mu\right.\right.\right. \\
&\quad \left.\left.\left.- \mu'\left[\sum_{t=0}^{T-1}((1 + \Psi'\Omega^{-1}\Psi) \cdot \Sigma_t^{-1}B_{t+\Delta} - \Psi'\Omega^{-1} \otimes \Sigma_t^{-1} \cdot \varepsilon_t^X) + M_0^{-1}m_0\right]\right.\right. \\
&\quad \left.\left.- \left[\sum_{t=0}^{T-1}((1 + \Psi'\Omega^{-1}\Psi) \cdot B_{t+\Delta}'\Sigma_t^{-1} - (\varepsilon_t^X)' \cdot \Omega^{-1}\Psi \otimes \Sigma_t^{-1}) + m_0'M_0^{-1}\right]\mu\right\}\right) \\
&\sim \mathcal{N}(\mathcal{W}^{-1}\mathcal{S}, \mathcal{W}^{-1})
\end{aligned}$$

which follows a normal distribution, $\mathcal{N}(\mathcal{W}^{-1}\mathcal{S}, \mathcal{W}^{-1})$ with $B_{t+\Delta}$, W and S expressed as:

$$\begin{aligned}
B_{t+\Delta} &= y_{t+\Delta} - \xi_{t+\Delta}^y \\
\varepsilon_{t+\Delta}^X &= \text{vec}(X_{t+\Delta} - X_t - a \Delta - \xi_{t+\Delta}^X) - \Sigma_t \Delta \otimes I_n \cdot \text{vec}(M) \\
W &= \Delta \sum_{t=0}^{T-1} ((1 + \Psi'\Omega^{-1}\Psi) \cdot \Sigma_t^{-1}) + M_0^{-1} \\
S &= \sum_{t=0}^{T-1} ((1 + \Psi'\Omega^{-1}\Psi) \cdot \Sigma_t^{-1}B_{t+\Delta} - \Psi'\Omega^{-1} \otimes \Sigma_t^{-1} \cdot \varepsilon_{t+\Delta}^X) + M_0^{-1}m_0
\end{aligned}$$

- Posterior for M^* : The prior of $M^* = \text{vec}(M)$ is assumed to be a normal distribution, i.e., $M^* \sim \mathcal{N}(m_0, M_0)$ and the posterior distribution can be similarly obtained as:

$$\begin{aligned}
\mathcal{P}(M^*|Y, \Gamma, \Theta_-) &\propto \mathcal{P}(Y, X|M^*, \Theta_-) \cdot \mathcal{P}(M^*|\theta) \\
&\propto \prod_{t=0}^{T-1} \mathcal{P}(y_{t+\Delta}, X_{t+\Delta}|M^*, \Theta_-, \Gamma_{t+\Delta}) \times e^{\frac{-1}{2}(M^*-m_0)'M_0^{-1}(M^*-m_0)} \\
&\propto \exp\left(\frac{-1}{2}\left\{(M^*)'(\Delta \sum_{t=0}^{T-1} (\Sigma_t \Omega^{-1} \Sigma_t \otimes \Sigma_t^{-1}) + M_0^{-1})M^* \right. \right. \\
&\quad - (M^*)' \left[\sum_{t=0}^{T-1} (-\Sigma_t \Omega^{-1} \Psi \otimes \Sigma_t^{-1} \varepsilon_{t+1}^y + \Sigma_t \Omega^{-1} \otimes \Sigma_t^{-1} \cdot B_{t+\Delta}) + M_0^{-1} m_0 \right] \\
&\quad \left. \left. - \left[\sum_{t=0}^{T-1} (-\Psi \Sigma_t \Omega^{-1} \otimes (\varepsilon_{t+1}^y)' \Sigma_t^{-1} + B_{t+\Delta}' \cdot \Omega^{-1} \Sigma_t \otimes \Sigma_t^{-1}) + m_0' M_0^{-1} \right] M^* \right\} \right) \\
&\sim \mathcal{N}(\mathcal{W}^{-1} \mathcal{S}, \mathcal{W}^{-1})
\end{aligned}$$

which also follows a normal distribution $\mathcal{N}(\mathcal{W}^{-1} \mathcal{S}, \mathcal{W}^{-1})$, where

$$\begin{aligned}
B_{t+\Delta} &= \text{vec}(X_{t+\Delta} - X_t - a \Delta - \xi_{t+\Delta}^X) \\
\varepsilon_{t+\Delta}^y &= y_{t+\Delta} - \mu \Delta - \xi_{t+\Delta}^y \\
W &= \Delta \sum_{t=0}^{T-1} (\Sigma_t \Omega^{-1} \Sigma_t \otimes \Sigma_t^{-1}) + M_0^{-1} \\
S &= \sum_{t=0}^{T-1} (-\Sigma_t \Omega^{-1} \Psi \otimes \Sigma_t^{-1} \varepsilon_{t+\Delta}^y + \Sigma_t \Omega^{-1} \otimes \Sigma_t^{-1} \cdot B_{t+\Delta}) + M_0^{-1} m_0
\end{aligned}$$

- Posterior for a^* : Similar with M^* : $a^* = \text{vec}(a)$ also has a normal distribution conjugate prior, $a^* \sim \mathcal{N}(m_0, M_0)$ and the posterior distribution is expressed in a similar form:

$$\begin{aligned}
\mathcal{P}(a^*|Y, \Gamma, \Theta_-) &\propto \mathcal{P}(Y, X|a^*, \Theta_-) \cdot \mathcal{P}(a^*|\theta) \\
&\propto \exp\left(\frac{-1}{2}\{(a^*)' W a^* - (a^*)' S - S' a^*\}\right) \\
&\sim \mathcal{N}(\mathcal{W}^{-1} \mathcal{S}, \mathcal{W}^{-1})
\end{aligned}$$

where

$$\begin{aligned}
B_{t+\Delta} &= \text{vec}(X_{t+\Delta} - X_t - a \Delta - \xi_{t+\Delta}^X) \\
\varepsilon_{t+\Delta}^y &= y_{t+\Delta} - \mu \Delta - \xi_{t+\Delta}^y \\
W &= \Delta \sum_{t=0}^{T-1} (\Omega^{-1} \otimes \Sigma_t^{-1}) + M_0^{-1} \\
S &= \sum_{t=0}^{T-1} (\Omega^{-1} \otimes \Sigma_t^{-1} \cdot B_{t+\Delta} - \Omega^{-1} \Psi \otimes \Sigma_t^{-1} \varepsilon_{t+\Delta}^y) + M_0^{-1} m_0
\end{aligned}$$

- Posterior for (Ω, Ψ) : As we would derive below, (Ω, Ψ) has a conjugate prior $\mathcal{N}\mathcal{IW}(\mu_0, k_0, P_0, v_0)$, and the corresponding posterior distribution follows $\mathcal{N}\mathcal{IW}(\mu^*, k^*, P^*, v^*)$. Note that (Ω, Ψ) only controls covariance matrix, V_t of the likelihood and hence ε_t is independent with it. Besides, we need to keep the $|V_t|$ term in the likelihood. More specifically,

$$\begin{aligned}
\mathcal{P}(\Omega, \Psi | Y, \Gamma, \Theta_-) &\propto \mathcal{P}(Y, X | \Omega, \Psi, \Theta_-) \cdot \mathcal{P}(\Omega, \Psi | \theta) \\
&\propto \prod_{t=0}^{T-1} \left(|V_t|^{-\frac{1}{2}} e^{-\frac{1}{2} \varepsilon'_{t+\Delta} V_t^{-1} \varepsilon_{t+\Delta}} \cdot \underbrace{|\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2} \text{tr}((\frac{\Omega}{k_0})^{-1} (\Psi - \mu_0)(\Psi - \mu_0)')} \cdot |\Omega|^{-\frac{v_0+n+1}{2}} e^{-\frac{1}{2} \text{tr}(P_0 \Omega^{-1})}}_{\mathcal{N}\mathcal{IW}} \right) \\
&\propto |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2} \text{tr}((\frac{\Omega}{k^*})^{-1} (\Psi - \mu^*)(\Psi - \mu^*))} \cdot |\Omega|^{-\frac{v^*+n+1}{2}} e^{-\frac{1}{2} \text{tr}(P^* \Omega^{-1})}
\end{aligned}$$

where the second proportionality comes from simplifying the likelihood term with (4.11), (4.14) and completing the square. In particular,

$$\begin{aligned}
A &= \frac{1}{\Delta} \sum_{t=0}^{T-1} (\varepsilon_{t+\Delta}^y)' \Sigma_t^{-1} \varepsilon_{t+\Delta}^y \\
B &= \frac{1}{\Delta} \sum_{t=0}^{T-1} \text{vec}^{-1}(\varepsilon_{t+\Delta}^X)' \Sigma_t^{-1} \varepsilon_{t+\Delta}^y \\
C &= \frac{1}{\Delta} \sum_{t=0}^{T-1} \text{vec}^{-1}(\varepsilon_{t+\Delta}^X)' \Sigma_t^{-1} \text{vec}^{-1}(\varepsilon_{t+\Delta}^X) \\
\mu^* &= \frac{B + k_0 \mu_0}{A + k_0} \\
k^* &= A + k_0, v^* = v_0 + nT \\
P^* &= C + P_0 + k_0 \mu_0 \mu_0' - k^* \mu^* (\mu^*)'
\end{aligned}$$

From the above derivation, it is shown that the resulting posterior dis-

tribution follows a $\mathcal{NIW}(\mu^*, k^*, P^*, v^*)$, which improves the efficiency greatly compared with directly sampling Q and ρ . Without our transformation, Metropolis-Hasting algorithm could be applied as in Eracker et al (2003), but the efficiency of sampling would be very low compared with ours.

- Posterior for $\delta \in \{J, C_1, C_2, c_i\}$: There is no standard conjugate prior existing for these terms, which imply essential economic interpretations. Other than Metropolis-Hasting style techniques, slice sampling is applied in this paper to draw samples from their posterior distributions in the MCMC procedure. For the sake of completeness and clarity, the posterior distributions are provided here in a semi-closed form by the expression of ε_t where J, C_1, C_2 and c_i are incorporated in the corresponding components. Since J, C_1, C_2 and c_i are constants, uninformative uniform priors could be applied, which are also constants. As a result, the posterior distributions could be obtained by calculating their likelihoods as follows:

$$\mathcal{P}(\delta|Y, \Gamma, \Theta_-) \propto \mathcal{P}(Y, X|\delta, \Theta_-) \cdot \mathcal{P}(\delta|\theta) \quad (4.17)$$

$$\propto \prod_{t=0}^{T-1} e^{\frac{-1}{2} \varepsilon_{t+\Delta} V_t^{-1} \varepsilon_{t+\Delta}} \quad (4.18)$$

(4.18) is a quite complicated function for each δ but we can draw sam

- Posterior for $(\mu_J^{(i)}, (\sigma_J^{(i)})^2)$, hyper parameters of Z_t : The jump sizes of log returns are assumed to follow independent normal distributions, i.e., $Z_t^{(i)} \sim \mathcal{N}(\mu_J^{(i)}, (\sigma_J^{(i)})^2), i = 1, \dots, n$. Each conjugate prior follows $\mathcal{NIG}(\mu_0, v_0, \alpha_0, \beta_0)$ and the posterior distribution follows $\mathcal{NIG}(\mu^*, v^*, \alpha^*, \beta^*)$ where

$$\mu^* = \frac{v_0 \mu_0 + T \mathbf{E}(Z_t^{(i)})}{v_0 + T} \quad (4.19)$$

$$v^* = v_0 + T, \alpha^* = \alpha_0 + \frac{T}{2} \quad (4.20)$$

$$\beta^* = \beta_0 + \frac{1}{2} \sum_{t=0}^{T-1} (Z_t^{(i)} - \mathbf{E}(Z_t^{(i)}))^2 + \frac{T v_0}{v_0 + T} \frac{(\mathbf{E}(Z_t^{(i)})^2 - \mu_0)^2}{2} \quad (4.21)$$

- Posterior for $(\mu_v^{(i)})$, hyper parameters of $K_t^{(i)}$: The jump sizes of variance-covariance matrix are assumed to follow exponential distributions with mean, $\mu_v^{(i)}$ in variance components and specially designed in covariance components, e.g. $c_i \sqrt{K_t^{11} K_t^{22}}$ in two-dimensional case. For clarity, i is

omitted here. The conjugate prior is standard, a Γ distribution with hyperparameter of α_0, β_0 . And the posterior follows $\Gamma(\alpha^*, \beta^*)$, where

$$\alpha^* = \alpha_0 + T \quad (4.22)$$

$$\beta^* = \beta + \sum_{t=0}^{T-1} K_t \quad (4.23)$$

4.5.2 Posterior for latent variables

- Posterior for $Z_{t+\Delta}$: For notation clarity, denote $\xi_{t+\Delta}^y = \ln(1 + J(Y_{t+\Delta}) \cdot dN_{t+\Delta})$ and $B_{t+\Delta} = y_{t+\Delta} - \mu\Delta$, then the resulting posterior distribution of $Z_{t+\Delta}$ equals the product of its likelihood and prior, which is assumed to follow $\mathcal{N}(\mu_{J,i}, \sigma_{J,i}^2), i = 1, \dots, n$. The explicit posterior distribution is calculated as:

$$\mathcal{P}(Z_{t+\Delta}^{(i)} | Y_t, \Gamma, \Theta) \propto \mathcal{P}(Y_t, X_t | Z_{t+\Delta}^{(i)}, \Theta) \cdot \mathcal{P}(Z_{t+\Delta}^{(i)}, \Psi | \theta) \quad (4.24)$$

$$\propto e^{-\frac{1}{2}\varepsilon'_{t+\Delta} V_t^{-1} \varepsilon_{t+\Delta}} \cdot \prod_{l=1}^n e^{-\frac{(Z_{t+\Delta}^i - \mu_{J,i})^2}{2\sigma_{J,i}^2}} \quad (4.25)$$

$$\propto e^{-\frac{1}{2\Delta}((\xi_{t+\Delta}^y)' A_1 \xi_{t+\Delta}^y - (\xi_{t+\Delta}^y)' A_2 - A_2' \xi_{t+\Delta}^y)} \cdot \prod_{l=1}^n e^{-\frac{(Z_{t+\Delta}^i - \mu_{J,i})^2}{2\sigma_{J,i}^2}} \quad (4.26)$$

where

$$A_1 = (1 + \Psi' \Omega^{-1} \Psi) \Sigma_t^{-1}$$

$$A_2 = (1 + \Psi' \Omega^{-1} \Psi) \Sigma_t^{-1} B_{t+\Delta} - \Psi' \Omega^{-1} \otimes \Sigma_t^{-1} \cdot \varepsilon_{t+\Delta}^X$$

for $t = 1, \dots, T$ and $i = 1, \dots, n$.

- Posterior for $N_{t+\Delta}$: The likelihood of $N_{t+\Delta}$ is straightforward, while the prior depends on $\lambda_t = C1\Sigma_t C2\Delta$. In fact, the posterior distribution which describes 2^n combinations of jump events follows a categorical distribution, where the exact distribution could be defined if probability density

for each event are obtained. More specifically, we calculate:

$$\mathcal{P}(N_{t+\Delta} = N_i | Y_t, \Gamma, \Theta) \propto \mathcal{P}(Y_t, X_t | N_{t+\Delta} = N_i, \Theta) \cdot \mathcal{P}(N_{t+\Delta} = N_i, \Psi | \theta) = p_i \quad (4.27)$$

$$\mathcal{P}(N_{t+\Delta} = N_i | Y, \Gamma, \Theta_-) = \frac{p_i}{\sum_{j=1}^{n^2} p_j} \quad (4.28)$$

where N_i is the i^{th} event in the sample space composed of all jump events and p_i is the corresponding unnormalized probability density. As a results, $N_{t+\Delta}$ can be sampled from a categorical distribution for $t = 1, \dots, T$.

- Posterior for $K_{t+\Delta}$: Similarly, by denoting $\xi_{t+\Delta}^X = \frac{1}{2} \sum_{i=1}^n K_{t+\Delta}^{(i)} dN_{t+\Delta}^{(i)}$, the posterior distribution of $K_{t+\Delta}$ can be calculated as the product of its likelihood and prior, which follows $Exp(\mu_v^{(i)})$. The explicit formula is expressed as follows:

$$\mathcal{P}(K_{t+\Delta}^{(i)} | Y_t, \Gamma, \Theta) \propto \mathcal{P}(Y_t, X_t | K_{t+\Delta}^{(i)}, \Theta) \cdot \mathcal{P}(K_{t+\Delta}^{(i)}, \Psi | \theta) \quad (4.29)$$

$$\propto e^{-\frac{1}{2} \varepsilon'_{t+\Delta} V_t^{-1} \varepsilon_{t+\Delta}} \cdot \prod_{l=1}^{n^2} \frac{1}{\mu_v^{(l)}} e^{\frac{K_{t+\Delta}^{(l)}}{\mu_v^{(l)}}} \quad (4.30)$$

$$\propto e^{\frac{-1}{2\Delta} ((\xi_{t+\Delta}^X)' A_1 \xi_{t+\Delta}^X - (\xi_{t+\Delta}^X)' A_2 - A_2' \xi_{t+\Delta}^X + A_2' A_1^{-1} A_2)} \cdot \prod_{l=1}^{n^2} \frac{1}{\mu_v^{(l)}} e^{\frac{K_{t+\Delta}^{(l)}}{\mu_v^{(l)}}} \quad (4.31)$$

where

$$\begin{aligned} A_1 &= \Omega^{-1} \otimes \Sigma_t^{-1} \\ A_2 &= \Omega^{-1} \otimes \Sigma_t^{-1} \cdot B_{t+\Delta} - \Omega^{-1} \Psi \otimes (\Sigma_t^{-1} \varepsilon_{t+\Delta}^y) \\ B_{t+\Delta} &= vec(X_{t+\Delta} - X_t - (a + M \Sigma_t) \Delta - \xi_{t+\Delta}^X) \end{aligned}$$

for $t = 1, \dots, T$ and $i = 1, \dots, n$. The $K_t^{(l)}$ denotes all jump distributions belonging to K_t^i .

- Posterior for $X_{t+\Delta}$: Based on the time discretization, $X_{t+\Delta}$ depends on both the neighbor of the state variables, $X_{t+2\Delta}$ and X_t and hence the posterior distribution of $X_{t+\Delta}$ is quite complicated. Nevertheless, we can

express the posterior distribution with $\varepsilon_{t+\Delta} = (\varepsilon_{t+\Delta}^y, \varepsilon_{t+\Delta}^X)$ as:

$$\mathcal{P}(X_{t+\Delta}|Y, \Gamma, \Theta_-) \propto \mathcal{P}(Y_t, X_{t+\Delta}|X_t, X_{t+2\Delta}, \Theta_-) \quad (4.32)$$

$$\propto \mathcal{P}(X_t|Y_{t+\Delta}, X_{t+\Delta}, \Gamma, \Theta) \cdot \mathcal{P}(X_{t+2\Delta}|Y_{t+2\Delta}, X_{t+\Delta}, \Gamma, \Theta) \quad (4.33)$$

$$\propto (e^{-\frac{1}{2}\varepsilon_{t+\Delta} V_t^{-1} \varepsilon_{t+\Delta}}) \cdot \left(\frac{1}{|V_{t+\Delta}|^{\frac{1}{2}}} e^{-\frac{1}{2}\varepsilon_{t+2\Delta} V_{t+\Delta}^{-1} \varepsilon_{t+2\Delta}} \right) \quad (4.34)$$

where $X_{t+\Delta}$ is incorporated in $\varepsilon_{t+\Delta}$, $\varepsilon_{t+2\Delta}$ and $V_{t+\Delta}$. Furthermore, to improve the computational cost, we calculate the explicit formula for both parts in the product:

$$\mathcal{P}(X_{t+2\Delta}|\dots) \propto |\Omega|^{-\frac{n}{2}} \cdot |\Sigma_{t+\Delta}|^{-\frac{n+1}{2}} \exp\left(-\frac{1}{2\Delta}(a_{t+\Delta} - 2c_{t+\Delta} + b_{t+\Delta})\right) \quad (4.35)$$

$$\mathcal{P}(X_t|\dots) \propto \exp\left(-\frac{1}{2\Delta}(a_t - 2c_t + b_t)\right) \quad (4.36)$$

$$a_{t+\Delta} = (\varepsilon_{t+2\Delta}^y)'(1 + \Psi'\Omega^{-1}\Psi)\Sigma_{t+\Delta}^{-1}\varepsilon_{t+2\Delta}^y \quad (4.37)$$

$$b_{t+\Delta} = \text{tr}(\text{vec}^{-1}(\varepsilon_{t+2\Delta}^X)'\Sigma_{t+\Delta}^{-1}\text{vec}^{-1}(\varepsilon_{t+2\Delta}^X)\Omega^{-1}) \quad (4.38)$$

$$c_{t+\Delta} = \varepsilon_{t+2\Delta}^y \Sigma_{t+\Delta}^{-1} \text{vec}^{-1}(\varepsilon_{t+2\Delta}^X)\Omega^{-1}\Psi \quad (4.39)$$

In particular, the posterior of $X_{t+\Delta}$ depends on the first exponential part as $t + \Delta = T$ because X_T depends only on $X_{T-\Delta}$, while X_0 depends on both V_0 and the second exponential part.

As equation (4.37) and (4.36) indicate, the complexity of the posterior distribution leads to difficulty of sampling. Even for the popular square-root stochastic volatility model, drawing samples from the posterior distribution of volatility state, say V_t is often tricky if techniques such as Metropolis-Hasting algorithm are applied. In particular, the poor mixing deteriorate the efficiency in the MCMC procedure.

On the contrary, the slice sampling technique adopted in this paper greatly circumvented the tuning procedure. Nevertheless the posterior distribution of X_t is matrix-variate, our estimation procedure exhibits high efficiency compared with existing methods such as Metropolis-Hasting style algorithms or adaptive Metropolis-Hasting algorithms.

4.5.3 MCMC Procedure Specification

With all posterior distributions derived in the above subsection, we complete the MCMC estimation algorithm with specification of the entire procedure. With respect to those nonstandard distributions, slice sampling technique in Neal (2003) are applied for sampling. Johannes and Polson(2003), (2006) provides details about MCMC backgrounds and related theory(see also Casella and George (1992) and Gelman et al. (2014) for further details).

In general, the MCMC algorithm works for inference problem by drawing samples iteratively from the corresponding posterior distributions of parameters and state variables. More specifically, given initial values of $(\Gamma^{(0)}, \Theta^{(0)})$, the MCMC algorithm initiates by drawing $\Theta^{(1)} \sim \mathcal{P}(\Theta^{(0)}|Y, \Gamma, \Theta^{(0)})$ and then $\Gamma^{(1)} \sim \mathcal{P}(\Gamma^{(1)}|Y, \Gamma, \Theta^{(1)})$, where the notations follow those defined in Section 4.5. By iteratively drawing in this manner, the MCMC algorithm generates a sequence of posterior random samples, $\{\Theta^{(j)}, \Gamma^{(j)}\}_{j=1}^N$. where each sample path constitutes the posterior distribution of the corresponding parameter(or state variable). Given enough iterations, the drawn samples then approximates the desired posterior distribution of the parameter(or state variables).

For the j_{th} iteration, the procedure can be explicitly summarized as follows:

1. Updating model parameters:

$$\mathcal{P}(\Theta_i^j) \propto \mathcal{P}(\Theta_-^{(j-1)}, Z_t^{(j-1)}, N_t^{(j-1)}, K_t^{(j-1)}, X_t^{(j-1)}, Y), i = 1, \dots, k \quad (4.40)$$

where k is the total number of model parameters.

2. Updating state variables of jump components:

$$\mathcal{P}(Z_t^{(j)}) \propto \mathcal{P}(\Theta^{(j)}, N_t^{(j-1)}, K_t^{(j-1)}, X_t^{(j-1)}, Y), t = 1, \dots, T \quad (4.41)$$

$$\mathcal{P}(N_t^{(j)}) \propto \mathcal{P}(\Theta^{(j)}, Z_t^{(j)}, K_t^{(j-1)}, X_t^{(j-1)}, Y), t = 1, \dots, T \quad (4.42)$$

$$\mathcal{P}(K_t^{(j)}) \propto \mathcal{P}(\Theta^{(j)}, Z_t^{(j)}, N_t^{(j)}, X_t^{(j-1)}, Y), t = 1, \dots, T \quad (4.43)$$

$$(4.44)$$

3. Updating covariance structure described by X_t :

$$\mathcal{P}(X_t^{(j)}) \propto \mathcal{P}(\Theta_-^{(j)}, Z_t^{(j)}, N_t^{(j)}, K_t^{(j)}, Y), t = 1, \dots, T \quad (4.45)$$

$$(4.46)$$

Following this updating scheme, the resulting MCMC estimates for parameters and state variables can be obtained by taking expectation for all drawn samples with standard error obtained by the corresponding standard deviation of the drawn samples.

With the MCMC specification, we complete the estimation procedure developed in this paper.

4.5.4 Brief Introduction of Slice Sampling methods

For the sake of completeness, a brief introduction of single-variable slice sampling technique is provided in this subsection. Without loss of generality, single-variable slice sampling method could be extended for multivariate distributions easily by applying it for each variable of the target multivariate distribution repeatedly in turn. Furthermore, updating all variables may be also feasible to improve the efficiency of sampling procedure. For details and theories about slice sampling, we refer to Neal (2003), among others.

Given the target distribution f with univariable x and initial point x_0 , the scheme of slice sampling technique could be divided into 3 steps:

Step 1. Uniformly draw a value "y" from $[0, f(x_0)]$. The horizontal "slice" is defined by $S = \{x : y \leq f(x)\}$.

Step 2. Find an interval $I = [L, R]$ in the neighbourhood of x_0 with predefined width w , which includes all or most of the slice S defined in Step 1. The L and R represent the left and right end points, respectively.

Step 3. The next stage of x , i.e., x_1 is updated by drawing a point uniformly from the intervals within the slice S . In particular, stepping-out and shrinkage procedures are applied for sampling.

The entire slice sampling is completed by repeating Step 1 to Step 3 iteratively and the resulting updated points of (x_1, x_2, \dots) would be the desired samples drawn from the target distribution. In particular, the stepping-out and shrinkage processes as discussed in Neal (2003) are applied in the MCMC procedure

developed in this chapter.

Figure (4.1) illustrates the Step 1 to Step 3. More explicitly, by drawing

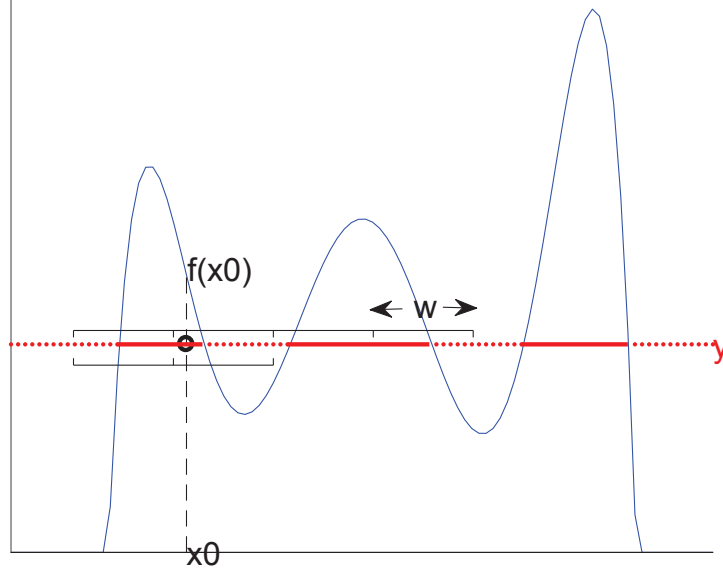


Figure 4.1: Illustration of Single-variate Slice Sampling Method

The blue solid curve represents the target distribution i.e., $f(x)$ to be sampled and x_0 is the initial state of sampling procedure. The black dashed line may be regarded as the vertical "slice" and y is uniformly drawn from $[0, f(x_0)]$. Accordingly, the solid red line is the horizontal "slice", S . The interval composed by line segments with width w is the interval for each iterative sampling.

$y \in [0, f(x_0)]$, the slice S is defined and labelled in bold red line segments. Subsequently, the interval I is constructed by expanding the neighbourhood around x_0 with predefined width w and the stepping-out/shrinkage procedures. The resulting interval I would include all or most of the slice. In particular, the stepping-out procedure is implemented by randomly positing w around x_0 and expanding the neighbourhood of x_0 by w until both ends of the interval are outside the slice. Noticeably, this interval would contain points or subintervals which are outside the slice. Hence, shrinkage is applied if the updated point (uniformly drawn from I) is not in the slice. These points outside the slice would be used as new end points to shrink the interval. The next point would then be drawn from the shrunk interval.

For the sake of completeness, we also provide the general picture of slice sampling and Metropolis-Hasting style MCMC techniques here.

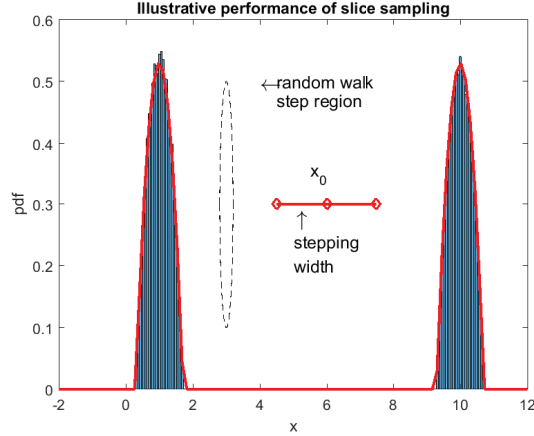


Figure 4.2: Illustration of Metropolis-Hasting and slice sampling scheme

For illustrative purpose, Figure (4.2) provides one exaggerated image of drawing samples from a two-peak distribution, where "random walk step region" and "stepping width" represent the key steps of Metropolis-Hasting style MCMC techniques and slice sampling technique, respectively. The red curve is the target probability density function and the shaded area is the resulting probability density function calculated with the obtained samples by slice sampling.

The key difference between these two methods lies in the feature of exploring steps. The details about the theory and practical issues are well documented in Johannes and Polson (2006) and Neal (2003), among others. By the shrinkage and stepping-out procedure of slice sampling, it is clear that the samples drawn by slice sampling indeed follow the desired distribution even though the distribution is unusual. On the contrary, as illustrated as the dotted circle, the random walk step region of Metropolis-Hasting style techniques only allows one side of the distribution to be effectively sampled, while the remained part may even remain unexplored. Even though tuning the random walk step size partly improved the exploration of the entire distribution, the resulting acceptance ratio may often be too low for efficient sampling.

As a results, we developed an MCMC estimation procedure with slice sampling technique augmented, which is capable of estimating both existing stochastic volatility models and more advanced multivariate stochastic volatility models where the stochastic covariance structure follows a Wishart process. Moreover, the developed estimation procedure also makes inference feasible even when jumps in return, covariance and some novel parameters are integrated. As we

would demonstrate in the following section, various simulation studies and the intuitive demonstration of slice sampling technique introduced in this section would verify the accuracy and advantage of the developed estimate method. Hence, we complete the establishment of the proposed estimation technique.

4.6 Numerical Results

In this section, we examine the performance of our MCMC procedure through simulation studies. Various models are embedded in the WJD model, including Merton-Jump-Diffusion model (Merton (1976)) with entangled jumps⁷, stochastic volatility models with common Merton jump in return and volatility, SVCMJ model. For illustration purpose, we provide the simulation studies for these models with the proposed estimation procedure. The results are compared against those obtained in the existing literatures, where the parameters are adopted here as true values. Lastly, WJD model, the target multivariate stochastic covariance matrix model with jumps is estimated with reasonable parameters assigned. For the sake of completeness and clarity, only hyperparameters used for WJD-DJ model is listed in Table 4.7.

Table 4.1 and 4.2 depict that the MJD model with entangled jumps are well approximated, except for Σ_J , which is often slightly biased. As we can expect, as number of observation increases, the accuracy is improved because with more observations, the time series is more informative about the true dynamics.

Next, we estimate the SVCMJ model which is more complicated because of the additional stochastic volatility component. Although SVCMJ model is nested in WJD-DJ, we recall the conventional expression for comparison with the literature. The SDE of SVCMJ is expressed as:

$$dS_t/S_t = (\mu + r)dt + \sqrt{(v_t)}dW_{t+1}^y + Y_{t+1}dN_{t+1} \quad (4.47)$$

$$dv_t = k(\theta - v_t)dt + \sigma_v\sqrt{(v_t)}dW_{t+1}^v + K_{t+1}dN_{t+1} \quad (4.48)$$

⁷The jumps are scaled by the constant matrix, J and hence they are essentially "entangled" with each other. As a result, it is denoted as "entangled jump" to separate them from conventional independent jumps.

where dW_t^y and dW_t^v are Brownian motions with $\text{corr}(dW_t^y, dW_t^v) = \rho$. The jump component is determined by jump intensity, $P(N_{t+1} = 1) = \lambda_t$ and jump sizes, $Y_t = \exp(Z_t) - 1$, $Z_t \sim N(\mu_y + \rho_J K_t, \sigma_y^2)$ and $K_t \sim \text{Exp}(\mu_v)$ in return and volatility, respectively. For illustrative purpose, the length of simulated time series is set to be 2000 in daily scale. The parameters are adopted directly from Eraker et al. (2003) for an clear comparison. Li et al. (2008) also applied the same setting to support their estimation procedure. We choose uninformative priors close to those applied in Li et al. (2008) and Table 4.3 summarizes the estimated results.

In general, the estimates obtained by our technique are close to the true values and superior to Eraker et al. (2003) and comparable with Li et al. (2008). It is worth mentioning that, much smaller number of sampling is required by our methods to sustain similar accuracy. Commonly, the number of posterior samples required for us is 5000, while it is over 50000 or even more to assure similar accuracy for other methods. In fact, this advantage comes from the augmented slice sampling technique, which is applied for drawing samples from nonstandard posterior distribution. Compared with normal MCMC techniques, where Metropolis-Hasting style methods are applied, the acceptance ratio is often set to be 20% to 40% for efficiency and the resulting number of sampling increases for accurate estimates. Besides, careful tuning is tricky for other sampling techniques, especially for high dimensional problem discussed in the following simulation tests. In some worse case, such as multi-modal posterior distributions, the target posterior distribution may be only partly explored and the estimation may yield possible misleading results. By incorporating slice sampling in our estimation procedure, efficiency and accuracy are greatly improved as demonstrated in the Table 4.3.

Finally, we conduct the simulation studies for WJD model with both jumps in return and variance-covariance, i.e., WJD-DJ model, the primary target for our estimation technique. Since the WJD-DJ model is at least n dimensional with one $n \times n$ stochastic variance-covariance structure and two jump components in return and variance-covariance, the simulation bias would be augmented in the simulated paths and the resulting stable state may deviate from the true parameter set for simulation studies.

As a result, it is not feasible to conduct a simulation test as done for the single dimensional cases. However, given every simulated path, all state variables are

known. It is reasonable to divide the inference problem into two parts: the variance-covariance state and all the remained parts, including all parameters and other state variables. More specifically, the estimation procedure is verified by examining the accuracy of estimated parameters given variance-covariance states fixed as the simulated state and the smoothed variance-covariance state given all other parameters and state variables fixed.

For demonstration, we focus on 2 dimensional case. Firstly, we assign reasonable parameters for each model and the values are presented in real value and daily scale. In particular, these parameters are randomly chosen with the corresponding requirements satisfied and yield similar patterns of log returns observed in real stock indices. The J , c_i and $C1C2$ reported represent (J_{21}, J_{12}) , (c_1, c_2) and $(C_2(1), C_2(2), C_1(1))$ respectively. To further demonstrate the accuracy of our MCMC procedure, we compare short($T=500$ observations) and long($T=2000$ observations) time scales, while the number of sampling is set to be 5000 with the first 3000 as burn-in period. Noticeably, the number of observation is set to be low (500~2000) for demonstrating the performance in the short time scale. More specifically, financial contagion occurs unexpectedly and gradually shades into the resulting next stage, where possible future events start to dominate. The time scale for specific financial contagion would be therefore limited and only moderate period prior and post to it would be of specific value. Hence, accuracy for estimation in short time horizon is essential, especially for financial contagion studies.

As Table 4.4 to Table 4.6 illustrate, the estimated results verified the performance of our estimation technique for complicated WJD-DJ models with different time horizons. In general, as numbers of observations increase, the accuracy is improved. As summarized in the tables, the parameters are reasonably well estimated and accuracy improved as time scale increases. This is natural since the more observation, the more informative the time series is. For shorter time scale, only few jumps would occur and hence lead to less accuracy for parameters related to jumps.

There are parameters worth emphasizing, the $J, C_1(1), C_2(1)$ and $C_2(2)$. These parameters control the sensitivity of each asset in terms of jump magnitude and intensity. More specifically, $J_{21}(J_{12})$ control how a jump in asset 2 (1) affects asset 1 (2). On the other hand, $C_1(1), C_2(1)$ and $C_2(2)$ specify the jump intensity by scaling the variance-covariance matrix, where $C_2(1), C_2(2)$ reflect

sensitivities to variance and covariance for both assets. The larger the $C_2(1)$, the more sensitive to the variance of asset 1 for jump intensity. Moreover, $C_1(1)$ determines the ratio of jump intensity between two assets. For example, $C_1(1) = 1.5$ means the probability of a jump event is 50% higher in asset 2 compared with asset 1. Even in the shortest time scale case, the estimates, $C_1(1) = 3.45, C_2(1) = 4.82, C_1(1) = 1.73$ are reasonably accurate compared with true values 3.5, 5.5, 1.5.

Regarding the second part where variance-covariance state are estimated with the remained parameters and state variables fix, we demonstrate the smoothed variance-covariance states as $T=1000$ observation for clarity. The fitted results are displayed in Figure 4.3 to Figure 4.5 with simulated path(treated as true value) in blue-solid line and fitted variance-covariance states (estimated by our technique) in red-dot line. Clearly, the estimated (smoothed) variance covariance state variables coincide with the true values well. To a great extend, the fitted state indicates the dynamic of this latent, unobserved variance-covariance state.⁸ As a result, as a by-product of MCMC estimation, the smoothed state variable provides an image of unobserved states given the only observed log returns. Integrated with these two parts of simulation studies, we can therefore verify the performance of the developed estimation procedure.

In summary, through various simulation studies, our estimation technique is verified by estimating various types of models and in different time scales. Hence, the developed estimation procedure may be ready for future researches where estimation of model parameters are important, such as optimal portfolio choice problem and option pricing literatures, etc.

Our technique also can be extended by incorporating other techniques to estimate state variables. For example, the realized volatility are commonly applied to estimated volatilities (Andersen et al. (2003)) or jump techniques (Lee and Hannig (2010) and Lee and Mykland (2008)). Given these states identified, the estimation procedure would be more efficient for estimating real empirical data. In particular, only parameters needs to be estimated and the other state variables are approximated with the augmented techniques and treated as known in our procedure. In fact, this augmentation of our estimation procedure and others may be regarded as a analogy of moment matching, where the states are "matched" by the augmented techniques while the parameters are estimated by

⁸Similarly, fitted jump times, and jump sizes would be obtained during the estimation procedure but ignored here for clarity. Since they are normally well approximated.

our estimation procedure.

Entangled MJD						
	μ			Σ		
True	0.0198	0.0278	0.0397	0.0198	0.0198	0.0397
Mean	0.0197	0.0274	0.0395	0.0196	0.0196	0.0408
RMSE	0.0012	0.0015	0.0021	0.0016	0.0016	0.0018
	μ_J			Σ_J		
True	-3	-3		3	3	
Mean	-2.7737	-2.6633		3.6273	3.2133	
RMSE	0.5888	0.3628		0.6126	0.5836	
	J			C1C2		
True	0.5	0.5		0.5	0.5	1
Mean	0.4965	0.4983		0.4885	0.4927	0.8861
RMSE	0.0120	0.0095		0.0710	0.0468	0.1118

The simulation test is conducted by simulating MJD model for 1000 daily data with 50 simulation paths, and parameters are set as displayed in the row labeled as True. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

Table 4.1: Simulation tests for entangled MJD model, T=1000

Entangled MJD						
	μ			Σ		
True	0.0198	0.0278	0.0397	0.0198	0.0198	0.0397
Mean	0.0199	0.0277	0.0400	0.0198	0.0198	0.0400
RMSE	0.0009	0.0008	0.0011	0.0008	0.0008	0.0012
	μ_J			Σ_J		
True	-3	-3		3	3	
Mean	-2.8348	-2.8648		3.4054	3.1267	
RMSE	0.2103	0.2448		0.6033	0.4853	
	J			C1C2		
True	0.5	0.5		0.5	0.5	1
Mean	0.5000	0.4993		0.4988	0.4579	0.9581
RMSE	0.0061	0.0054		0.0833	0.0691	0.0926

The simulation test is conducted by simulating MJD model for 3000 daily data with 50 simulation paths, and parameters are set as displayed in the row labeled as True. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

Table 4.2: Simulation tests for entangled MJD model, T=3000

Panel A.			SVMJ			Model		
μ	k	θ	σ_v	ρ	λ	μ_y	σ_y	
TRUE	0.00050	0.01500	0.000090	0.00100	-0.40	0.01984	-0.0300	0.0350
Mean	0.00048	0.01490	0.000094	0.00098	-0.42	0.02005	-0.0297	0.0349
RMSE	0.00005	0.00153	0.000003	0.00007	0.07	0.00352	0.0065	0.0030
Panel B.			SVCMJ			Model		
μ	k	θ	σ_v	ρ	λ	μ_y	σ_y	μ_v
TRUE	0.00050	0.01500	0.000090	0.00100	-0.40	0.01984	-0.0300	0.000050
Mean	0.00048	0.01492	0.000090	0.00093	-0.42	0.01988	-0.0310	0.000050
RMSE	0.00009	0.00190	0.000015	0.00010	0.10	0.00291	0.0152	0.000011
Panel C.			SVCMJ			independent		
μ	k	θ	σ_v	ρ	λ	μ_y	σ_y	μ_v
TRUE	0.00050	0.01500	0.000090	0.00100	-0.40	0.01984	-0.0300	0.000050
Mean	0.00050	0.01500	0.000088	0.00091	-0.40	0.01919	-0.0306	0.000052
RMSE	0.00009	0.00181	0.000016	0.00011	0.10	0.00295	0.0053	0.000012

The simulation test is conducted by simulating SVCMJ model for 2000 daily data with 100 simulation paths, and parameters are mainly adopted from Li et al (2008), except for ρ_J and μ_v , which is increased to test more extreme case. The values is set as displayed in the row labeled as True and expressed in real number instead of % to avoid confusion. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

SVCMJ independent in Panel C. represent s the jumps in the return and volatility are independent.

Table 4.3: Simulation tests for 1 dimensional stochastic volatility model.

WJD-DJ									
	μ_1	μ_1	J_{21}	J_{12}	c_1	c_2	Ψ_1	Ψ_2	
TRUE	0.0005	0.0005	0.2	0.5	0.1	0.5	-0.0002375	-0.0002	
Mean	0.000558	0.000506	0.23	0.47	0.10	0.50	-0.00023	-0.00020	
RMSE	0.000111	0.000059	0.09	0.09	0.01	0.01	0.00001	0.00000	

	μ'_J	σ'_J	$C_1(1)$	$C_2(1)$	$C_2(2)$	
TRUE	-0.03	-0.03	0.035	0.035	1.5	
Mean	-0.0302	-0.0298	0.0351	0.0347	1.73	
RMSE	0.0013	0.0017	0.0011	0.0013	0.08	

	$\text{vec}(\mathbf{a})'$					$\text{vec}(\mathbf{M})'$				
TRUE	1.50E-05	1.50E-05	7.50E-06	1.50E-05	-0.025	-0.01667	-0.01667	-0.04167	-0.04167	
Mean	1.51E-05	1.52E-05	7.59E-06	1.52E-05	-0.02513	-0.01688	-0.01692	-0.04210	-0.04210	
RMSE	9.57E-07	8.91E-07	8.59E-07	7.26E-07	0.00133	0.00124	0.00211	0.00169	0.00169	

	$\text{vec}(\Omega)'$					μ_v^i				
TRUE	1.34E-07	1.03E-07	1.03E-07	8.25E-08	5.00E-05	5.00E-05	5.00E-05	5.00E-05	5.00E-05	
Mean	1.52E-07	1.10E-07	1.10E-07	9.05E-08	5.06E-05	5.02E-05	5.02E-05	5.02E-05	4.97E-05	
RMSE	6.80E-09	4.32E-09	4.32E-09	3.69E-09	2.18E-06	2.42E-06	2.25E-06	2.25E-06	2.40E-06	

The simulation test is conducted by simulating WJD-DJ model for 500 daily data with 50 simulation paths, and parameters are set as displayed in the row labeled as True. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. To investigate the accuracy of parameter estimation, all state variables are fixed in this test. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

Table 4.4: Simulation tests for WJD-DJ model with 500 simulated time series

WJD-DJ									
	μ_1	μ_1	J_{21}	J_{12}	c_1	c_2	Ψ_1	Ψ_2	
TRUE	0.0005	0.0005	0.2	0.5	0.1	0.5	-0.0002375	-0.0002	
Mean	0.00050	0.00047	0.23	0.53	0.09	0.50	-0.00023	-0.00020	
RMSE	0.00009	0.00013	0.06	0.09	0.09	0.02	0.00001	0.00001	
	μ'_I	σ'_I	$C_1(1)$	$C_2(1)$	$C_2(2)$				
TRUE	-0.03	-0.03	0.035	0.035	3.5	5.5	1.5		
Mean	-0.0256	-0.0296	0.0334	0.0361	3.45	5.08	1.53		
RMSE	0.0070	0.0043	0.0051	0.0055	0.35	0.32	0.27		
	$\text{vec(a)'}'$			$\text{vec(M)'}'$					
TRUE	1.50E-05	1.50E-05	7.50E-06	1.50E-05	-0.025	-0.01667	-0.01667	-0.04167	
Mean	1.50E-05	1.51E-05	7.50E-06	1.51E-05	-0.02496	-0.01676	-0.01666	-0.04183	
RMSE	6.23E-07	4.57E-07	5.09E-07	4.33E-07	0.00082	0.00057	0.00113	0.00104	
	$\text{vec}(\Omega)'$			μ_{η}^i					
TRUE	1.34E-07	1.03E-07	1.03E-07	8.25E-08	5.00E-05	5.00E-05	5.00E-05	5.00E-05	
Mean	1.51E-07	1.10E-07	1.10E-07	8.85E-08	4.71E-05	5.17E-05	4.82E-05	4.89E-05	
RMSE	3.09E-09	2.47E-09	2.47E-09	1.97E-09	8.91E-06	9.24E-06	1.12E-05	9.57E-06	

Table 4.5: Simulation tests for WJD-DJ model with 1000 simulated time series

The simulation test is conducted by simulating WJD-DJ model for 1000 daily data with 50 simulation paths, and parameters are set as displayed in the row labeled as True. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. To investigate the accuracy of parameter estimation, all state variables are fixed in this test. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

WJD-DJ									
	μ_1	μ_1	J_{21}	J_{12}	c_1	c_2	Ψ_1	Ψ_2	
TRUE	0.0005	0.0005	0.20	0.50	0.10	0.50	-0.0002375	-0.0002	
Mean	0.000489	0.000507	0.21	0.53	0.09	0.50	-0.0002344	-0.000197	
RMSE	0.000075	0.000055	0.07	0.05	0.03	0.01	5.03E-06	3.44E-06	

	μ'_I	σ'_I	$C_1(1)$	$C_2(1)$	$C_2(2)$	
TRUE	-0.03	-0.03	0.035	0.035	3.5	5.5
Mean	-0.0266	-0.0313	0.0327	0.0341	3.29	5.31
RMSE	0.0047	0.0074	0.0044	0.0033	0.73	0.95

	$\text{vec}(\mathbf{a})'$	$\text{vec}(\mathbf{M})'$	
TRUE	1.50E-05	1.50E-05	1.50E-06
Mean	1.53E-05	1.50E-05	1.50E-06
RMSE	4.77E-07	2.88E-07	2.73E-07

	$\text{vec}(\Omega)'$	μ_v^i	
TRUE	1.34E-07	1.03E-07	1.03E-07
Mean	1.47E-07	1.08E-07	1.08E-07
RMSE	1.93E-09	1.55E-09	1.55E-09

	μ_v^i		
TRUE	1.34E-07	1.03E-07	1.03E-07
Mean	1.47E-07	1.08E-07	1.08E-07
RMSE	1.93E-09	1.55E-09	1.55E-09

The simulation test is conducted by simulating WJD-DJ model for 2000 daily data with 50 simulation paths, and parameters are set as displayed in the row labeled as True. The Mean and RMSE represent mean values and root-mean-square error respectively. The specification of parameters follows the description in Numerical Results section. To investigate the accuracy of parameter estimation, all state variables are fixed in this test. For each simulation path, 5000 samples are drawn from their posterior distributions. In particular, the first 3000 samples are discarded as burn-in period and each parameter are estimated as the mean of the next 2000 samples.

Table 4.6: Simulation tests for WJD-DJ model with 2000 simulated time series

$N(m_0, M_0)$	m_0	M_0		
μ	0_2	$10I_2$		
$vec(M)$	0_4	$10I_4$		
$vec(a)$	0_4	$10I_4$		
$NIW(\mu_0, k_0, P_0, v_0)$	μ_0	k_0	P_0	v_0
(Ω, Ψ)	0_2	10	5×10^{-5}	2
$NIG(\mu_0, v_0, \alpha_0, \beta_0)$	μ_0	v_0	α_0	β_0
$(\mu_J^{(i)}, (\sigma_J^{(i)})^2)$	0	2	2	0.001
$\Gamma(\alpha_0, \beta_0)$	α_0	β_0		
$\mu_v^{(i)}$	2	2×10^{-5}		
$U(a, b)$	a	b		
J	0	1		
C_1, C_2	0	10		
c_i	-1	1		

Detailed values for all hyperparameters are listed here. 0_n and I_n represents $n \times 1$ zero vector and $n \times n$ identity matrix, respectively. All priors are extremely uninformative for the parameters compared with the obtained estimates. Hence, the estimates are not driven by the prior distributions.

Table 4.7: Hyper parameters of prior distributions

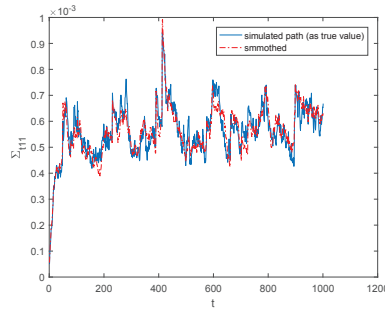


Figure 4.3: Fitted Variance-Covariance states - Σ_{t11}

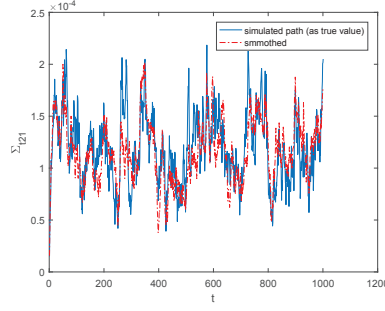


Figure 4.4: Fitted Variance-Covariance states - Σ_{t21}

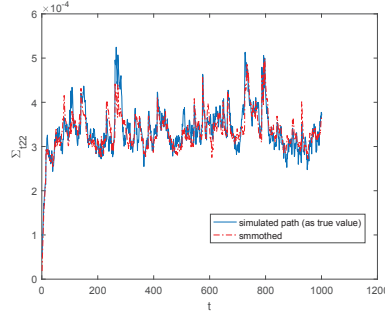


Figure 4.5: Fitted Variance-Covariance states - Σ_{t22}

4.7 Conclusion

A Bayesian based Markov Chain Monte Carlo estimation technique is developed for estimating parameters of a wide class of asset pricing models. In particular, the proposed estimation technique works well for WJD model, which solves the optimal portfolio choice problem under contagion and captures observed features during market turmoil. Detailed derivation of the posterior distributions are provided and can be ready for future researches related to stochastic volatility and financial contagion. The performance is verified with various simulation tests, where common and widely applied models are estimated with simulated data.

To sum up, the major contribution of our paper is two-fold. Firstly, we developed an estimation procedure for the general multivariate jump diffusion model with stochastic covariance matrix modeled as the Wishart process, which is novel in the literature. The estimation procedure is based on the Bayesian Markov Chain Monte Carlo (MCMC) technique. The augmented slice sampling technique developed by Neal (2003) substantially improves the efficiency of our MCMC technique. To the best of our literature review, this may be the first attempt to estimate the WJD models with MCMC based methods. In addition, our paper sheds new lights in the literature about estimating asset pricing mod-

els and the developed estimation procedure may be ready for future researches where inference problems are going to be solved.

4.8 Preliminary Results for empirical applications

In this complementary section, we roughly demonstrate the application of our estimation procedure. In particular, we are interested whether the novel component J of WJD model could help captured stylized features of empirical data. As a first attempt, we study the asymmetric correlation as investigated in Ang and Chen (2002).

Following similar procedures as adopted in Ang and Chen (2002), we investigate asymmetric correlation thoroughly with the 5 industry classifications, which is categorized as "Consumer", "Manufacturing", "High-Technology", "Health" and "Others". The data is collected from the Data Library on the website of Kenneth R. French, where the data are mainly collected from COMPUSTA and details about how these portfolios are constructed could be found. To have a clear picture of asymmetric correlation, we calculate the exceedance correlation following Longin and Solnik (2001) in different time scales and periods. By doing so, it could be shown that asymmetric correlation is persistent among different contexts and remains significant even with the innovation of computer technology and the resultantly increasing entanglement of markets. The data ranges from 1st January, 1980 to 31st August, 2016. The weekly data is constructed by aggregating the daily data from the end of last Wednesday to the end of current Wednesday as widely adopted in the literature.

In this preliminary result, we group each industrial categorization with market portfolio and estimate each group with a 2-dimensional MJD model with entangled jump as done in Section 4.6. Then we simulate 500 paths with the estimates for each group and calculate the corresponding exceedance correlations for each path. Consequently, the resulting fitted asymmetric correlations for each group are calculated as the mean of 500 trials within each group. In addition, given these simulated paths, we plot 95% credible interval for each fitted asymmetric correlation for benchmark. In other words, we determine whether asymmetric correlation observed in the real market is well captured by the MJD with entangled jump model.

Figure 4.6 to Figure 4.10 summarizes the fitted asymmetric correlation for all 5 industry classification. As depicted in the graphs, the asymmetric correlations are well replicated by the model fitted results, except for "Other". However,

most of the graph lie within the 95% credible interval in this case. It is worth mentioning that J is the dominating component that successfully capture this observed asymmetry. By contrast, for ordinary MJD models, only symmetric, i.e., tent-shaped graphs may be obtained due to the symmetric Poisson jump incorporated. Compared with those models studied in Ang and Chen (2002), such as GARCH and regime-switching models, MJD model with entangled J is incredibly simple and has more intuitive economical meanings. More specifically, J allows jump risks to propagate through different portfolios (or markets in international context) even if there is only one jump event. In fact, this feature assembles financial contagion or the initiate of market downturns, where there is often one starting point that triggers the subsequent disasters and spread to the counterparts.

To sum up, this example further justifies the validity of augmenting J in the multivariate models such as WJD models. As a result, these novel components introduced in WJD models may shed new light on the researches where multivariate stochastic models may be applied and WJD models may be of great potential for future researches.

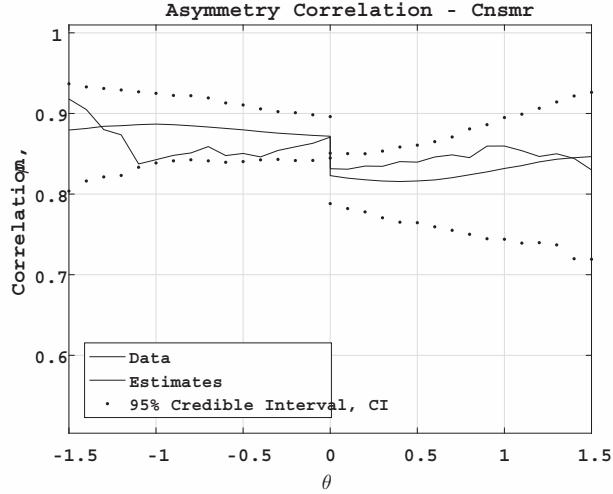


Figure 4.6: Asymmetric Correlation - Market and Consumer portfolios

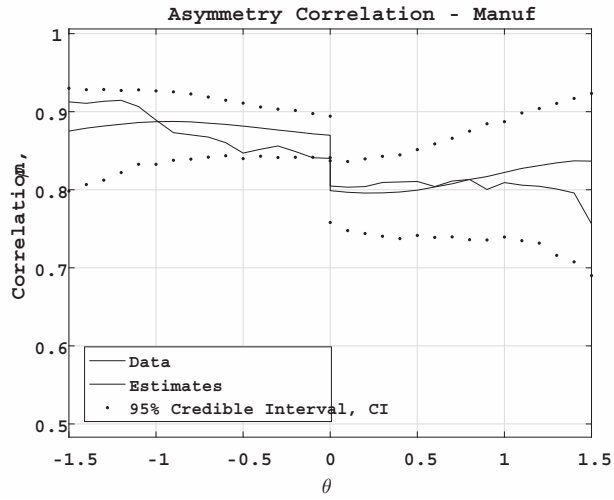


Figure 4.7: Asymmetric Correlation - Market and Manufacturing portfolios

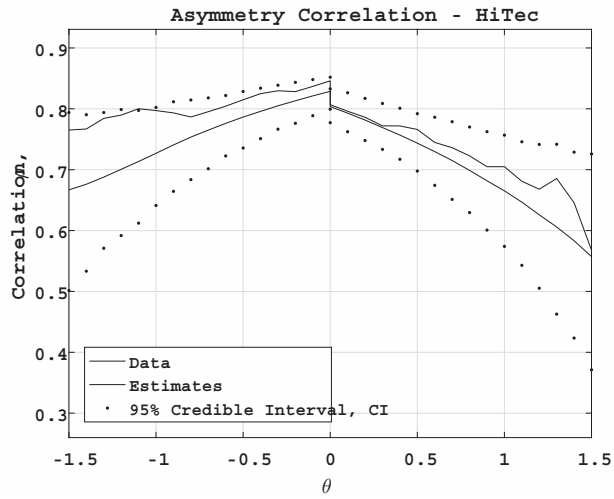


Figure 4.8: Asymmetric Correlation - Market and High-Technology portfolios

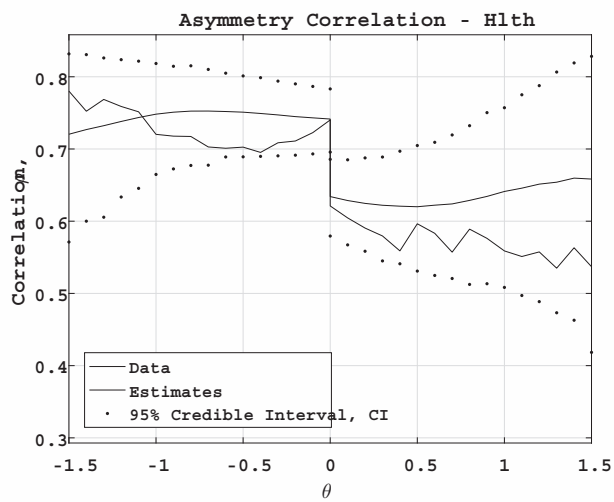


Figure 4.9: Asymmetric Correlation - Market and Health portfolios

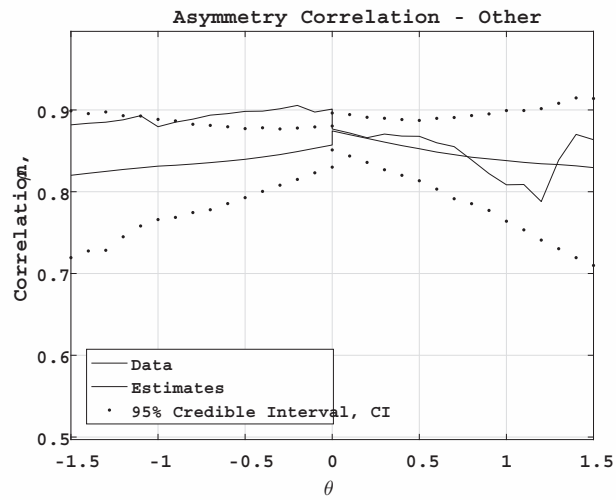


Figure 4.10: Asymmetric Correlation - Market and Other portfolios

4.9 Complementary: Derived Moment Generating Functions for moment-matching techniques

In this complementary section, we provide some critical closed-form formula obtained by deriving the corresponding moment generating function, which can be ready for moment-matching methods, e.g. GMM for future researches. This kind of estimation technique may be preferable sometimes, if high frequency data are available such that matching moments help capturing stylized features observed from data. For example, realized volatility is often applied as a proxy for volatility.

Following the notations defined in previous sections, the unconditional risk premia of stock i is thus

$$M_{1i} = [E[\mu_{it}] - \frac{1}{2}e_i' E[\Sigma(t)]e_i + E[\lambda_{1t}]E(Z_{i1}) + E[\lambda_{2t}]E(Z_{i2})]\tau,$$

where

$$\begin{aligned} E[\lambda_{1t}] &= C_{11}C_2(1)E[\Sigma_{11}(t)] + (C_{11}C_2(2) + C_{12}C_2(1))E[\Sigma_{12}(t)] + C_{12}C_2(2)E[\Sigma_{22}(t)] \\ E[\lambda_{2t}] &= C_{21}C_2(1)E[\Sigma_{11}(t)] + (C_{21}C_2(2) + C_{22}C_2(1))E[\Sigma_{12}(t)] + C_{22}C_2(2)E[\Sigma_{22}(t)] \\ E[\mu_{it}] &= e_i' E[\Sigma(t)]\eta - (J_{i1}E(Y_1)E[\lambda_{1t}] + J_{i2}E(Y_2)E[\lambda_{2t}]) \end{aligned}$$

2) Quadratic variation matrix $QV(t, t+\tau) = (QV_{ij}(t, t+\tau))_{1 \leq i, j \leq 2}$ of log returns:

$$\begin{aligned} QV_{11}(t, t+\tau) &= \int_t^{t+\tau} \Sigma_{11}(s)ds + \int_t^{t+\tau} Z_{11}^2 dN_{1s} + \int_t^{t+\tau} Z_{12}^2 dN_{2s}, \\ QV_{22}(t, t+\tau) &= \int_t^{t+\tau} \Sigma_{22}(s)ds + \int_t^{t+\tau} Z_{21}^2 dN_{1s} + \int_t^{t+\tau} Z_{22}^2 dN_{2s}, \\ QV_{12}(t, t+\tau) &= QV_{12}(t, t+\tau) = \int_t^{t+\tau} \Sigma_{12}(s)ds + \int_t^{t+\tau} Z_{11}Z_{21}dN_{1s} + \int_t^{t+\tau} Z_{12}Z_{22}dN_{2s}. \end{aligned}$$

Hence $QV(t, t+\tau)$ can be written as

$$QV(t, t+\tau) = \int_t^{t+\tau} \Sigma(s)ds + \int_t^{t+\tau} Z_1 dN_{1s} + \int_t^{t+\tau} Z_2 dN_{2s},$$

where

$$Z_1 = \begin{pmatrix} Z_{11}^2 & Z_{11}Z_{21} \\ Z_{11}Z_{21} & Z_{21}^2 \end{pmatrix}, Z_2 = \begin{pmatrix} Z_{12}^2 & Z_{12}Z_{22} \\ Z_{12}Z_{22} & Z_{22}^2 \end{pmatrix}, \quad (4.49)$$

Unconditional mean of the realized Quadratic variation matrix of log returns is:

$$E[QV(t, t + \tau)] = (E[\Sigma(t)] + E(Z_1)E[\lambda_{1t}] + E(Z_2)E[\lambda_{2t}])\tau,$$

3) Generating function of variance-covariance matrix $\Sigma(t)$ of is given by

$$\Psi^\Sigma(\Gamma, \Sigma(t), t, T) = E_t \left[\exp \left(- \int_t^T r_s ds + \text{tr}(\Gamma \Sigma(T)) \right) \right].$$

We have the following result:

$$\Psi^\Sigma(\Gamma, \Sigma(t), t, T) = \exp(B(T - t) + \text{tr}(A(T - t)\Sigma(t))),$$

where the functions A and B solve the system of matrix differential equations:

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= M'A(\tau) + A(\tau)M' + 2A(\tau)Q'QA(\tau) + \frac{1}{2}(C_2e'_1C_1 + C'_1e_1C'_2)[E[\exp(\text{tr}(A(\tau)K^{(1)}))] - 1] \\ &\quad + \frac{1}{2}(C_2e'_2C_1 + C'_1e_2C'_2)[E[\exp(\text{tr}(A(\tau)K^{(2)}))] - 1], \\ \frac{dB(\tau)}{d\tau} &= -r + \text{tr}(A(\tau)\Omega\Omega'), \end{aligned}$$

with boundary conditions: $B(0) = 0$ and $A(0) = \Gamma$. We now calculate $E[\Sigma(t)]$.

Let $\Gamma = (\Gamma_{ij})_{1 \leq i, j \leq 2}$ with $\Gamma_{ij} = \Gamma_{ji}$.

$$\begin{aligned} E[\Sigma_{ii}(t)] &= \frac{\partial \Psi^\Sigma(\Gamma, \Sigma(0), 0, t)}{\partial \Gamma_{ii}} \Big|_{\Gamma=0} \\ &= \left[\left(\frac{\partial B(t)}{\partial \Gamma_{ii}} + \frac{\partial(\text{tr}(A(t)\Sigma(0)))}{\partial \Gamma_{ii}} \right) \exp(B(t) + \text{tr}(A(t)\Sigma(0))) \right] \Big|_{\Gamma=0}, \end{aligned}$$

$$\begin{aligned} E[\Sigma_{ij}(t)] &= \frac{\partial \Psi^\Sigma(\Gamma, \Sigma(0), 0, t)}{\partial \Gamma_{ij}} \Big|_{\Gamma=0} \\ &= \frac{1}{2} \left[\left(\frac{\partial B(t)}{\partial \Gamma_{ij}} + \frac{\partial(\text{tr}(A(t)\Sigma(0)))}{\partial \Gamma_{ij}} \right) \exp(B(t) + \text{tr}(A(t)\Sigma(0))) \right] \Big|_{\Gamma=0}. \end{aligned}$$

More results: let $F = (F_{ij})_{1 \leq i, j \leq 2}$ with $F_{ij} = F_{ji}$.

$$\begin{aligned} E[\Sigma_{ii}(t) \exp(\text{tr}(F\Sigma(t)))] &= \frac{\partial \Psi^\Sigma(\Gamma, \Sigma(0), 0, t)}{\partial \Gamma_{ii}} \Big|_{\Gamma=F} \\ &= \left[\left(\frac{\partial B(t)}{\partial \Gamma_{ii}} + \frac{\partial(\text{tr}(A(t)\Sigma(0)))}{\partial \Gamma_{ii}} \right) \exp(B(t) + \text{tr}(A(t)\Sigma(0))) \right] \Big|_{\Gamma=F}, \end{aligned}$$

$$\begin{aligned} E[\Sigma_{ij}(t) \exp(\text{tr}(F\Sigma(t)))] &= \frac{\partial \Psi^\Sigma(\Gamma, \Sigma(0), 0, t)}{\partial \Gamma_{ij}} \Big|_{\Gamma=F} \\ &= \frac{1}{2} \left[\left(\frac{\partial B(t)}{\partial \Gamma_{ij}} + \frac{\partial(\text{tr}(A(t)\Sigma(0)))}{\partial \Gamma_{ij}} \right) \exp(B(t) + \text{tr}(A(t)\Sigma(0))) \right] \Big|_{\Gamma=F}. \end{aligned}$$

4) Generating function of quadratic variation matrix $QV(t, t + \tau) = (QV_{ij}(t, t + \tau))_{1 \leq i, j \leq 2}$ of log returns is given by

$$\Psi^{QV}(\Gamma, \Sigma(t), t, t + \tau) = E[\exp(\text{tr}(\Gamma QV(t, t + \tau)))].$$

We first define the conditional generating function of quadratic variation matrix as

$$\Psi^{QV}(\Gamma, \Sigma(t), 0, T, t) = E_t[\exp(\text{tr}(\Gamma QV(0, T)))].$$

We have the following result:

$$\Psi^{QV}(\Gamma, \Sigma(t), 0, T, t) = \exp(B(T - t) + \text{tr}(\Gamma QV(0, t)) + \text{tr}(A(T - t)\Sigma(t))),$$

where the functions A and B solve the system of matrix differential equations:

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= \Gamma + M'A(\tau) + A(\tau)M' + 2A(\tau)Q'QA(\tau) + \frac{1}{2}(C_2e'_1C_1 + C'_1e_1C'_2)[\Theta_1^\Sigma(A(\tau)) - 1] \\ &\quad + \frac{1}{2}(C_2e'_2C_1 + C'_1e_2C'_2)[\Theta_2^\Sigma(A(\tau)) - 1], \\ \frac{dB(\tau)}{d\tau} &= \text{tr}(A(\tau)\Omega\Omega'), \end{aligned}$$

where $\Theta_1^\Sigma(A(\tau)) = E[\exp(\text{tr}(A(\tau)K^{(1)} + \text{tr}(\Gamma Z_1)))]$ and $\Theta_2^\Sigma(A(\tau)) = E[\exp(\text{tr}(A(\tau)K^{(2)} + \text{tr}(\Gamma Z_2)))]$. The boundary conditions are: $B(0) = 0$ and $A(0) = 0$.

Thus,

$$\Psi^{QV}(\Gamma, \Sigma(t), t, t + \tau) = E[\exp(\text{tr}(\Gamma QV(t, t + \tau)))] = E[\exp(B(\tau) + \text{tr}(A(\tau)\Sigma(t)))].$$

We now calculate unconditional second moment M_3 of the realized quadratic

variation matrix $QV(t, t+\tau) = (QV_{ij}(t, t+\tau))_{1 \leq i, j \leq 2}$ of log-returns. Let $f(\tau, t) = \exp(B(\tau) + \text{tr}(A(\tau)\Sigma(t)))$.

$$\begin{aligned} E[QV_{ii}^2(t, t+\tau)] &= \frac{\partial^2 \Psi^{QV}(\Gamma, \Sigma(t), t, t+\tau)}{\partial \Gamma_{ii}^2} \Big|_{\Gamma=0} \\ &= E \left[\left[\left(\frac{\partial^2 B(\tau)}{\partial \Gamma_{ii}^2} + \frac{\partial^2 (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ii}^2} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right] \\ &+ E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{ii}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ii}} \right)^2 f(\tau, t) \right] \Big|_{\Gamma=0} \right], \end{aligned}$$

$$\begin{aligned} E[QV_{ij}^2(t, t+\tau)] &= \frac{\partial^2 \Psi^{QV}(\Gamma, \Sigma(t), t, t+\tau)}{\partial \Gamma_{ij}^2} \Big|_{\Gamma=0} \\ &= \frac{1}{4} E \left[\left[\left(\frac{\partial^2 B(\tau)}{\partial \Gamma_{ij}^2} + \frac{\partial^2 (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ij}^2} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right] \\ &+ \frac{1}{4} E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{ij}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ij}} \right)^2 f(\tau, t) \right] \Big|_{\Gamma=0} \right], \end{aligned}$$

$$\begin{aligned} E[QV_{ii}(t, t+\tau)QV_{ij}(t, t+\tau)] &= \frac{\partial^2 \Psi^{QV}(\Gamma, \Sigma(t), t, t+\tau)}{\partial \Gamma_{ij} \partial \Gamma_{ii}} \Big|_{\Gamma=0} \\ &= \frac{1}{2} E \left[\left[\left(\frac{\partial^2 B(\tau)}{\partial \Gamma_{ij} \partial \Gamma_{ii}} + \frac{\partial^2 (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ij} \partial \Gamma_{ii}} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right] \\ &+ \frac{1}{2} E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{ij}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ij}} \right) \left(\frac{\partial B(\tau)}{\partial \Gamma_{ii}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ii}} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right]. \end{aligned}$$

$$\begin{aligned} E[QV_{ii}(t, t+\tau)QV_{jj}(t, t+\tau)] &= \frac{\partial^2 \Psi^{QV}(\Gamma, \Sigma(t), t, t+\tau)}{\partial \Gamma_{jj} \partial \Gamma_{ii}} \Big|_{\Gamma=0} \\ &= E \left[\left[\left(\frac{\partial^2 B(\tau)}{\partial \Gamma_{jj} \partial \Gamma_{ii}} + \frac{\partial^2 (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{jj} \partial \Gamma_{ii}} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right] \\ &+ E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{jj}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{jj}} \right) \left(\frac{\partial B(\tau)}{\partial \Gamma_{ii}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ii}} \right) f(\tau, t) \right] \Big|_{\Gamma=0} \right]. \end{aligned}$$

5) Generating function of quadratic variation matrix $QV(t, t+\tau) = (QV_{ij}(t, t+\tau))_{1 \leq i, j \leq 2}$ of log returns and log return $Y_{t, t+\tau} = (\ln(S_{1, t+\tau}/S_{1, t}), \ln(S_{2, t+\tau}/S_{2, t}))'$ is given by

$$\Psi_1^{QV}(\Gamma, \Sigma(t), t, t+\tau) = E [\exp(u' \cdot Y_{t, t+\tau} + \text{tr}(\Gamma QV(t, t+\tau)))] ,$$

where $u = (u_1, u_2)'$. As in 5), we first define the conditional generating function

of quadratic variation matrix and log prices as

$$\Psi_2^{QV}(\Gamma, \Sigma(t), 0, T, t) = E_t [\exp(u' \cdot X_T + \text{tr}(\Gamma QV(0, T)))] ,$$

where $X_t = (\ln(S_{1,t}), \ln(S_{2,t}))'$. We have the following result:

$$\Psi_2^{QV}(\Gamma, \Sigma(t), 0, T, t) = \exp(u' \cdot X_t + B(T - t) + \text{tr}(\Gamma QV(0, t)) + \text{tr}(A(T - t)\Sigma(t))),$$

where the functions A and B solve the system of matrix differential equations:

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= \Gamma + \frac{1}{2}(\eta u' + u\eta') - \frac{1}{2}(C_2 u' J \cdot \mathbf{E}[Y] \cdot C_1 + C_1' \cdot \mathbf{E}[Y] J' u C_2') \\ &+ \frac{1}{2}(u u' - u_1 e_1 e_1' - u_2 e_2 e_2') + A(\tau)(M + Q' \rho u') \\ &+ (M' + u \rho' Q)A(\tau) + 2A(\tau)Q' Q A(\tau) + \frac{1}{2}(C_2 e_1' C_1 + C_1' e_1 C_2')[\Theta_1^\Sigma(u, A(\tau)) - 1] \\ &+ \frac{1}{2}(C_2 e_2' C_1 + C_1' e_2 C_2')[\Theta_2^\Sigma(u, A(\tau)) - 1], \\ \frac{dB(\tau)}{d\tau} &= \text{tr}(A(\tau)\Omega\Omega'), \end{aligned}$$

where

$$\begin{aligned} \Theta_1^\Sigma(A(\tau)) &= E[\exp(\text{tr}(A(\tau)K^{(1)}) + \text{tr}(\Gamma Z_1) + u_1 Z_{11} + u_2 Z_{21})], \\ \Theta_2^\Sigma(A(\tau)) &= E[\exp(\text{tr}(A(\tau)K^{(2)}) + \text{tr}(\Gamma Z_2) + u_1 Z_{12} + u_2 Z_{22})]. \end{aligned}$$

The boundary conditions are: $B(0) = 0$ and $A(0) = 0$.

Thus,

$$\Psi_1^{QV}(\Gamma, \Sigma(t), t, t + \tau) = E[\exp(u' Y_{t,t+\tau} + \text{tr}(\Gamma QV(t, t + \tau)))] = E[\exp(B(\tau) + \text{tr}(A(\tau)\Sigma(t)))] .$$

Unconditional covariance between assets' simple excess returns and the quadratic variation matrix of log returns. For asset 1, let $u = (1, 0)'$, then

$$\begin{aligned} E[\exp(u' \cdot Y_{t,t+\tau}) QV_{ii}(t, t + \tau)] &= \frac{\partial \Psi_1^{QV}(\Gamma, \Sigma(t), t, t + \tau)}{\partial \Gamma_{ii}} \Big|_{\Gamma=0} \\ &= E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{ii}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ii}} \right) \exp(B(\tau) + \text{tr}(A(\tau)\Sigma(t))) \right] \Big|_{\Gamma=0} \right], \end{aligned}$$

$$\begin{aligned}
E [\exp(u' \cdot Y_{t,t+\tau})] QV_{ij}(t, t + \tau) &= \frac{\partial \Psi_1^{QV}(\Gamma, \Sigma(t), t, t + \tau)}{\partial \Gamma_{ij}} \Big|_{\Gamma=0} \\
&= \frac{1}{2} E \left[\left[\left(\frac{\partial B(\tau)}{\partial \Gamma_{ij}} + \frac{\partial (\text{tr}(A(\tau)\Sigma(t)))}{\partial \Gamma_{ij}} \right) \exp(B(\tau) + \text{tr}(A(\tau)\Sigma(t))) \right] \Big|_{\Gamma=0} \right],
\end{aligned}$$

6) Processes:

$$d[\log(S_{1t})] = [r + \mu_{1t} - \frac{1}{2}(\sigma_{11}^2 + \sigma_{12}^2)]dt + \sigma_{11}dW_{1t} + \sigma_{12}dW_{2t} + Z_{11}dN_{1t} + Z_{12}dN_{2t},$$

$$\begin{aligned}
W_{1t} &= \sqrt{1 - \rho_1^2 - \rho_2^2} Z_{1t}^\perp + \rho_1 Z_{11t} + \rho_2 Z_{12t} \\
W_{2t} &= \sqrt{1 - \rho_1^2 - \rho_2^2} Z_{2t}^\perp + \rho_1 Z_{21t} + \rho_2 Z_{22t}
\end{aligned}$$

$$\begin{aligned}
d\Sigma_{11}(t) &= [\bar{\Omega}_{11} + 2(M_{11}\Sigma_{11}(t) + M_{12}\Sigma_{12}(t))]dt \\
&+ 2Q_{11}(\sigma_{11}dZ_{11t} + \sigma_{12}dZ_{21t}) \\
&+ 2Q_{21}(\sigma_{11}dZ_{12t} + \sigma_{12}dZ_{22t}) \\
&+ K_{11}^{(1)}dN_{1t} + K_{11}^{(2)}dN_{2t},
\end{aligned}$$

$$\begin{aligned}
d\Sigma_{12}(t) &= [\bar{\Omega}_{12} + M_{21}\Sigma_{11}(t) + (M_{11} + M_{22})\Sigma_{12}(t) + M_{12}\Sigma_{22}(t)]dt \\
&+ (Q_{12}\sigma_{11} + Q_{11}\sigma_{21})dZ_{11t} \\
&+ (Q_{22}\sigma_{11} + Q_{21}\sigma_{21})dZ_{12t} \\
&+ (Q_{12}\sigma_{12} + Q_{11}\sigma_{22})dZ_{21t} \\
&+ (Q_{22}\sigma_{12} + Q_{21}\sigma_{22})dZ_{22t} \\
&+ K_{12}^{(1)}dN_{1t} + K_{12}^{(2)}dN_{2t},
\end{aligned}$$

$$\begin{aligned}
d\Sigma_{22}(t) &= [\bar{\Omega}_{22} + 2(M_{21}\Sigma_{12}(t) + M_{22}\Sigma_{22}(t))]dt \\
&+ 2Q_{12}(\sigma_{21}dZ_{11t} + \sigma_{22}dZ_{21t}) \\
&+ 2Q_{22}(\sigma_{21}dZ_{12t} + \sigma_{22}dZ_{22t}) \\
&+ K_{22}^{(1)}dN_{1t} + K_{22}^{(2)}dN_{2t},
\end{aligned}$$

Generating function of variance-covariance matrix $\int_t^{t+\tau} \Sigma_{ij}(s)ds$ of log returns

and log return $Y_1(t, t + \tau) = \ln(S_{1,t+\tau}/S_{1,t})$ is given by

$$\begin{aligned} & \Psi_1^\Sigma(\Gamma, \Sigma(t), t, t + \tau) \\ = & E \left[\exp \left(u_1 \cdot Y_1(t, t + \tau) + \Gamma_{11} \int_t^{t+\tau} \Sigma_{11}(s) ds + \Gamma_{12} \int_t^{t+\tau} \Sigma_{12}(s) ds + \Gamma_{22} \int_t^{t+\tau} \Sigma_{22}(s) ds \right) \right], \end{aligned}$$

As in 5), we first define the conditional generating function of variance-covariance matrix and log prices as

$$\begin{aligned} & \Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t) \\ = & E_t \left[\exp \left(u_1 \cdot Y_1(T) + \Gamma_{11} \int_0^T \Sigma_{11}(s) ds + \Gamma_{12} \int_0^T \Sigma_{12}(s) ds + \Gamma_{22} \int_0^T \Sigma_{22}(s) ds \right) \right], \end{aligned}$$

where $Y_1(t) = \ln(S_{1,t})$. Clearly, $\Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t)$ is a martingale. We have the following result:

$$\Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t) = \exp[G],$$

where

$$\begin{aligned} G &= u_1 \cdot Y_1(t) + \Gamma_{11} \int_0^t \Sigma_{11}(s) ds + \Gamma_{12} \int_0^t \Sigma_{12}(s) ds + \Gamma_{22} \int_0^t \Sigma_{22}(s) ds \\ &+ B(\tau) + A_{11}(\tau) \Sigma_{11}(t) + A_{12}(\tau) \Sigma_{12}(t) + A_{22}(\tau) \Sigma_{22}(t) \end{aligned}$$

From Ito's lemma, we have

$$\begin{aligned} dG &= u_1 \cdot dY_1(t) + (\Gamma_{11} \Sigma_{11}(t) + \Gamma_{12} \Sigma_{12}(t) + \Gamma_{22} \Sigma_{22}(t)) dt \\ &- B'(\tau) dt - A'_{11}(\tau) \Sigma_{11}(t) dt + A_{11}(\tau) d\Sigma_{11}(t) \\ &- A'_{12}(\tau) \Sigma_{12}(t) dt + A_{12}(\tau) d\Sigma_{12}(t) - A'_{22}(\tau) \Sigma_{22}(t) dt + A_{22}(\tau) d\Sigma_{22}(t) \\ &= \mu_G dt + \sigma_{11}^G dZ_{11} + \sigma_{12}^G dZ_{12} + \sigma_{21}^G dZ_{21} + \sigma_{22}^G dZ_{22} + \sigma_1^G dZ_1^\perp + \sigma_2^G dZ_2^\perp \\ &+ \sigma_3^G dN_{1t} + \sigma_4^G dN_{2t}, \end{aligned}$$

where

$$\begin{aligned}
\mu_G &= u_1(r + \mu_{1t} - \frac{1}{2}\Sigma_{11}(t)) \\
&+ \Gamma_{11}\Sigma_{11}(t) + \Gamma_{12}\Sigma_{12}(t) + \Gamma_{22}\Sigma_{22}(t) - B'(\tau) \\
&- A'_{11}(\tau)\Sigma_{11}(t) - A'_{12}(\tau)\Sigma_{12}(t) - A'_{22}(\tau)\Sigma_{22}(t) \\
&+ A_{11}(\tau)[\bar{\Omega}_{11} + 2(M_{11}\Sigma_{11}(t) + M_{12}\Sigma_{12}(t))] \\
&+ A_{12}(\tau)[\bar{\Omega}_{12} + M_{21}\Sigma_{11}(t) + (M_{11} + M_{22})\Sigma_{12}(t) + M_{12}\Sigma_{22}(t)] \\
&+ A_{22}(\tau)[\bar{\Omega}_{22} + 2(M_{21}\Sigma_{12}(t) + M_{22}\Sigma_{22}(t))]
\end{aligned}$$

$$\bar{\Omega} = (\bar{\Omega}_{ij}) = \Omega\Omega',$$

$$\begin{aligned}
\sigma_{11}^G &= u_1\sigma_{11}\rho_1 + 2A_{11}(\tau)Q_{11}\sigma_{11} + A_{12}(\tau)(Q_{12}\sigma_{11} + Q_{11}\sigma_{21}) \\
&+ 2A_{22}(\tau)Q_{12}\sigma_{21}
\end{aligned}$$

$$\begin{aligned}
\sigma_{12}^G &= u_1\sigma_{11}\rho_2 + 2A_{11}(\tau)Q_{21}\sigma_{11} + A_{12}(\tau)(Q_{22}\sigma_{11} + Q_{21}\sigma_{21}) \\
&+ 2A_{22}(\tau)Q_{22}\sigma_{21}
\end{aligned}$$

$$\begin{aligned}
\sigma_{21}^G &= u_1\sigma_{12}\rho_1 + 2A_{11}(\tau)Q_{11}\sigma_{12} + A_{12}(\tau)(Q_{12}\sigma_{12} + Q_{11}\sigma_{22}) \\
&+ 2A_{22}(\tau)Q_{12}\sigma_{22}
\end{aligned}$$

$$\begin{aligned}
\sigma_{22}^G &= u_1\sigma_{12}\rho_2 + 2A_{11}(\tau)Q_{21}\sigma_{12} + A_{12}(\tau)(Q_{22}\sigma_{12} + Q_{21}\sigma_{22}) \\
&+ 2A_{22}(\tau)Q_{22}\sigma_{22}
\end{aligned}$$

$$\sigma_1^G = u_1\sigma_{11}\sqrt{1 - \rho_1^2 - \rho_2^2}$$

$$\sigma_2^G = u_1\sigma_{12}\sqrt{1 - \rho_1^2 - \rho_2^2}$$

$$\sigma_3^G = u_1Z_{11} + A_{11}(\tau)K_{11}^{(1)} + A_{12}(\tau)K_{12}^{(1)} + A_{22}(\tau)K_{22}^{(1)}$$

$$\sigma_4^G = u_1Z_{12} + A_{11}(\tau)K_{11}^{(2)} + A_{12}(\tau)K_{12}^{(2)} + A_{22}(\tau)K_{22}^{(2)}$$

Applying Ito's lemma to $H = \exp(G)$, we have

$$\begin{aligned} dH &= H\mu_H dt + H(\sigma_{11}^G dZ_{11} + \sigma_{12}^G dZ_{12} + \sigma_{21}^G dZ_{21} + \sigma_{22}^G dZ_{22} + \sigma_1^G dZ_1^\perp + \sigma_2^G dZ_2^\perp \\ &+ H(\exp(\sigma_3^G) - 1)dN_{1t} + H(\exp(\sigma_4^G) - 1)dN_{2t}, \end{aligned}$$

where

$$\mu_H = \mu_G + \frac{1}{2}[(\sigma_{11}^G)^2 + (\sigma_{12}^G)^2 + (\sigma_{21}^G)^2 + (\sigma_{22}^G)^2 + (\sigma_1^G)^2 + (\sigma_2^G)^2].$$

Since H is a martingale, we have

$$\mu_H + \lambda_{1t}E(\exp(\sigma_3^G) - 1) + \lambda_{2t}E(\exp(\sigma_4^G) - 1) = 0.$$

And thus the functions A and B solve the system of matrix differential equations:

$$\begin{aligned} \frac{dA_{11}(\tau)}{d\tau} &= u_1[\eta_1 - J_{11}E(Y_1)C_{11}C_2(1) - J_{12}E(Y_2)C_{21}C_2(1) - \frac{1}{2}] \\ &+ \Gamma_{11} + 2M_{11}A_{11}(\tau) + M_{21}A_{12}(\tau) \\ &+ \frac{1}{2}(u_1\rho_1 + 2A_{11}(\tau)Q_{11} + A_{12}(\tau)Q_{12})^2 \\ &+ \frac{1}{2}(u_1\rho_2 + 2A_{11}(\tau)Q_{21} + A_{12}(\tau)Q_{22})^2 \\ &+ \frac{1}{2}u_1^2(1 - \rho_1^2 - \rho_2^2) \\ &+ C_{11}C_2(1)E(\exp(\sigma_3^G) - 1) + C_{21}C_2(1)E(\exp(\sigma_4^G) - 1) \end{aligned}$$

$$\begin{aligned} \frac{dA_{12}(\tau)}{d\tau} &= u_1[\eta_2 - J_{11}E(Y_1)(C_{11}C_2(2) + C_{12}C_2(1)) - J_{12}E(Y_2)(C_{21}C_2(2) + C_{22}C_2(1))] \\ &+ \Gamma_{12} + 2M_{12}A_{11}(\tau) + (M_{11} + M_{22})A_{12}(\tau) + 2A_{22}(\tau)M_{21} \\ &+ (u_1\rho_1 + 2A_{11}(\tau)Q_{11} + A_{12}(\tau)Q_{12})(A_{12}(\tau)Q_{11} + 2A_{22}(\tau)Q_{12}) \\ &+ (u_1\rho_2 + 2A_{11}(\tau)Q_{21} + A_{12}(\tau)Q_{22})(A_{12}(\tau)Q_{21} + 2A_{22}(\tau)Q_{22}) \\ &+ (C_{11}C_2(2) + C_{12}C_2(1))E(\exp(\sigma_3^G) - 1) + (C_{21}C_2(2) + C_{22}C_2(1))E(\exp(\sigma_4^G) - 1) \end{aligned}$$

$$\begin{aligned}
\frac{dA_{22}(\tau)}{d\tau} &= u_1[-J_{11}E(Y_1)C_{12}C_2(2) - J_{12}E(Y_2)C_{22}C_2(2)] \\
&+ \Gamma_{22} + 2M_{22}A_{22}(\tau) + M_{12}A_{12}(\tau) \\
&+ \frac{1}{2}(A_{12}(\tau)Q_{11} + 2A_{22}(\tau)Q_{12})^2 \\
&+ \frac{1}{2}(A_{12}(\tau)Q_{21} + 2A_{22}(\tau)Q_{22})^2 \\
&+ C_{12}C_2(2)E(\exp(\sigma_3^G) - 1) + C_{22}C_2(2)E(\exp(\sigma_4^G) - 1)
\end{aligned}$$

$$\frac{dB(\tau)}{d\tau} = u_1r + A_{11}(\tau)\bar{\Omega}_{11} + A_{12}(\tau)\bar{\Omega}_{12} + A_{22}(\tau)\bar{\Omega}_{22}.$$

7) Generating function of variance-covariance matrix $\int_t^{t+\tau} \Sigma_{ij}(s)ds$ of log returns and log return $Y_2(t, t+\tau) = \ln(S_{2,t+\tau}/S_{2,t})$ is given by

$$\begin{aligned}
&\Psi_1^\Sigma(\Gamma, \Sigma(t), t, t+\tau) \\
&= E \left[\exp \left(u_2 \cdot Y_2(t, t+\tau) + \Gamma_{11} \int_t^{t+\tau} \Sigma_{11}(s)ds + \Gamma_{12} \int_t^{t+\tau} \Sigma_{12}(s)ds + \Gamma_{22} \int_t^{t+\tau} \Sigma_{22}(s)ds \right) \right],
\end{aligned}$$

As in 5), we first define the conditional generating function of variance-covariance matrix and log prices as

$$\begin{aligned}
&\Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t) \\
&= E_t \left[\exp \left(u_2 \cdot Y_2(T) + \Gamma_{11} \int_0^T \Sigma_{11}(s)ds + \Gamma_{12} \int_0^T \Sigma_{12}(s)ds + \Gamma_{22} \int_0^T \Sigma_{22}(s)ds \right) \right],
\end{aligned}$$

where $Y_2(t) = \ln(S_{2,t})$. Clearly, $\Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t)$ is a martingale. We have the following result:

$$\Psi_2^\Sigma(\Gamma, \Sigma(t), 0, T, t) = \exp[G],$$

where

$$\begin{aligned}
G &= u_2 \cdot Y_2(t) + \Gamma_{11} \int_0^t \Sigma_{11}(s)ds + \Gamma_{12} \int_0^t \Sigma_{12}(s)ds + \Gamma_{22} \int_0^t \Sigma_{22}(s)ds \\
&+ B(\tau) + A_{11}(\tau)\Sigma_{11}(t) + A_{12}(\tau)\Sigma_{12}(t) + A_{22}(\tau)\Sigma_{22}(t)
\end{aligned}$$

From Ito's lemma, we have

$$\begin{aligned}
dG &= u_2 \cdot dY_2(t) + (\Gamma_{11}\Sigma_{11}(t) + \Gamma_{12}\Sigma_{12}(t) + \Gamma_{22}\Sigma_{22}(t))dt \\
&- B'(\tau)dt - A'_{11}(\tau)\Sigma_{11}(t)dt + A_{11}(\tau)d\Sigma_{11}(t) \\
&- A'_{12}(\tau)\Sigma_{12}(t)dt + A_{12}(\tau)d\Sigma_{12}(t) - A'_{22}(\tau)\Sigma_{22}(t)dt + A_{22}(\tau)d\Sigma_{22}(t) \\
&= \mu_G dt + \sigma_{11}^G dZ_{11} + \sigma_{12}^G dZ_{12} + \sigma_{21}^G dZ_{21} + \sigma_{22}^G dZ_{22} + \sigma_1^G dZ_1^\perp + \sigma_2^G dZ_2^\perp \\
&+ \sigma_3^G dN_{1t} + \sigma_4^G dN_{2t},
\end{aligned}$$

where

$$\begin{aligned}
\mu_G &= u_2(r + \mu_{2t} - \frac{1}{2}\Sigma_{22}(t)) \\
&+ \Gamma_{11}\Sigma_{11}(t) + \Gamma_{12}\Sigma_{12}(t) + \Gamma_{22}\Sigma_{22}(t) - B'(\tau) \\
&- A'_{11}(\tau)\Sigma_{11}(t) - A'_{12}(\tau)\Sigma_{12}(t) - A'_{22}(\tau)\Sigma_{22}(t) \\
&+ A_{11}(\tau)[\bar{\Omega}_{11} + 2(M_{11}\Sigma_{11}(t) + M_{12}\Sigma_{12}(t))] \\
&+ A_{12}(\tau)[\bar{\Omega}_{12} + M_{21}\Sigma_{11}(t) + (M_{11} + M_{22})\Sigma_{12}(t) + M_{12}\Sigma_{22}(t)] \\
&+ A_{22}(\tau)[\bar{\Omega}_{22} + 2(M_{21}\Sigma_{12}(t) + M_{22}\Sigma_{22}(t))]
\end{aligned}$$

$$\bar{\Omega} = (\bar{\Omega}_{ij}) = \Omega\Omega',$$

$$\begin{aligned}
\sigma_{11}^G &= u_2\sigma_{21}\rho_1 + 2A_{11}(\tau)Q_{11}\sigma_{11} + A_{12}(\tau)(Q_{12}\sigma_{11} + Q_{11}\sigma_{21}) \\
&+ 2A_{22}(\tau)Q_{12}\sigma_{21}
\end{aligned}$$

$$\begin{aligned}
\sigma_{12}^G &= u_2\sigma_{21}\rho_2 + 2A_{11}(\tau)Q_{21}\sigma_{11} + A_{12}(\tau)(Q_{22}\sigma_{11} + Q_{21}\sigma_{21}) \\
&+ 2A_{22}(\tau)Q_{22}\sigma_{21}
\end{aligned}$$

$$\begin{aligned}
\sigma_{21}^G &= u_2\sigma_{22}\rho_1 + 2A_{11}(\tau)Q_{11}\sigma_{12} + A_{12}(\tau)(Q_{12}\sigma_{12} + Q_{11}\sigma_{22}) \\
&+ 2A_{22}(\tau)Q_{12}\sigma_{22}
\end{aligned}$$

$$\begin{aligned}
\sigma_{22}^G &= u_2\sigma_{22}\rho_2 + 2A_{11}(\tau)Q_{21}\sigma_{12} + A_{12}(\tau)(Q_{22}\sigma_{12} + Q_{21}\sigma_{22}) \\
&+ 2A_{22}(\tau)Q_{22}\sigma_{22}
\end{aligned}$$

$$\sigma_1^G = u_2\sigma_{21}\sqrt{1 - \rho_1^2 - \rho_2^2}$$

$$\sigma_2^G = u_2 \sigma_{22} \sqrt{1 - \rho_1^2 - \rho_2^2}$$

$$\sigma_3^G = u_2 Z_{21} + A_{11}(\tau) K_{11}^{(1)} + A_{12}(\tau) K_{12}^{(1)} + A_{22}(\tau) K_{22}^{(1)}$$

$$\sigma_4^G = u_2 Z_{22} + A_{11}(\tau) K_{11}^{(2)} + A_{12}(\tau) K_{12}^{(2)} + A_{22}(\tau) K_{22}^{(2)}$$

Applying Ito's lemma to $H = \exp(G)$, we have

$$\begin{aligned} dH &= H \mu_H dt + H(\sigma_{11}^G dZ_{11} + \sigma_{12}^G dZ_{12} + \sigma_{21}^G dZ_{21} + \sigma_{22}^G dZ_{22} + \sigma_1^G dZ_1^\perp + \sigma_2^G dZ_2^\perp \\ &+ H(\exp(\sigma_3^G) - 1) dN_{1t} + H(\exp(\sigma_4^G) - 1) dN_{2t}, \end{aligned}$$

where

$$\mu_H = \mu_G + \frac{1}{2}[(\sigma_{11}^G)^2 + (\sigma_{12}^G)^2 + (\sigma_{21}^G)^2 + (\sigma_{22}^G)^2 + (\sigma_1^G)^2 + (\sigma_2^G)^2].$$

Since H is a martingale, we have

$$\mu_H + \lambda_{1t} E(\exp(\sigma_3^G) - 1) + \lambda_{2t} E(\exp(\sigma_4^G) - 1) = 0.$$

And thus the functions A and B solve the system of matrix differential equations:

$$\begin{aligned} \frac{dA_{11}(\tau)}{d\tau} &= u_2[-J_{21}E(Y_1)C_{11}C_2(1) - J_{22}E(Y_2)C_{21}C_2(1)] \\ &+ \Gamma_{11} + 2M_{11}A_{11}(\tau) + M_{21}A_{12}(\tau) \\ &+ \frac{1}{2}(2A_{11}(\tau)Q_{11} + A_{12}(\tau)Q_{12})^2 \\ &+ \frac{1}{2}(2A_{11}(\tau)Q_{21} + A_{12}(\tau)Q_{22})^2 \\ &+ C_{11}C_2(1)E(\exp(\sigma_3^G) - 1) + C_{21}C_2(1)E(\exp(\sigma_4^G) - 1) \end{aligned}$$

$$\begin{aligned} \frac{dA_{12}(\tau)}{d\tau} &= u_1[\eta_1 - J_{21}E(Y_1)(C_{11}C_2(2) + C_{12}C_2(1)) - J_{22}E(Y_2)(C_{21}C_2(2) + C_{22}C_2(1))] \\ &+ \Gamma_{12} + 2M_{12}A_{11}(\tau) + (M_{11} + M_{22})A_{12}(\tau) + 2A_{22}(\tau)M_{21} \\ &+ (2A_{11}(\tau)Q_{11} + A_{12}(\tau)Q_{12})(u_2\rho_1 + A_{12}(\tau)Q_{11} + 2A_{22}(\tau)Q_{12}) \\ &+ (2A_{11}(\tau)Q_{21} + A_{12}(\tau)Q_{22})(u_2\rho_2 + A_{12}(\tau)Q_{21} + 2A_{22}(\tau)Q_{22}) \\ &+ (C_{11}C_2(2) + C_{12}C_2(1))E(\exp(\sigma_3^G) - 1) + (C_{21}C_2(2) + C_{22}C_2(1))E(\exp(\sigma_4^G) - 1) \end{aligned}$$

$$\begin{aligned}
\frac{dA_{22}(\tau)}{d\tau} &= u_1[\eta_2 - J_{21}E(Y_1)C_{12}C_2(2) - J_{22}E(Y_2)C_{22}C_2(2) - \frac{1}{2}] \\
&+ \Gamma_{22} + 2M_{22}A_{22}(\tau) + M_{12}A_{12}(\tau) \\
&+ \frac{1}{2}(u_2\rho_1 + A_{12}(\tau)Q_{11} + 2A_{22}(\tau)Q_{12})^2 \\
&+ \frac{1}{2}(u_2\rho_2 + A_{12}(\tau)Q_{21} + 2A_{22}(\tau)Q_{22})^2 \\
&+ \frac{1}{2}u_2^2(1 - \rho_1^2 - \rho_2^2) \\
&+ C_{12}C_2(2)E(\exp(\sigma_3^G) - 1) + C_{22}C_2(2)E(\exp(\sigma_4^G) - 1)
\end{aligned}$$

$$\frac{dB(\tau)}{d\tau} = u_2r + A_{11}(\tau)\bar{\Omega}_{11} + A_{12}(\tau)\bar{\Omega}_{12} + A_{22}(\tau)\bar{\Omega}_{22}.$$

where

$$\begin{aligned}
\Theta_1^\Sigma(A(\tau)) &= E[\exp(\text{tr}(A(\tau)K^{(1)}) + \text{tr}(\Gamma Z_1) + u_1Z_{11} + u_2Z_{21})], \\
\Theta_2^\Sigma(A(\tau)) &= E[\exp(\text{tr}(A(\tau)K^{(2)}) + \text{tr}(\Gamma Z_2) + u_1Z_{12} + u_2Z_{22})].
\end{aligned}$$

The boundary conditions are: $B(0) = 0$ and $A(0) = 0$.

Bibliography

Aït-Sahalia, Y., et al. (2009). "Portfolio choice with jumps: A closed-form solution." *The Annals of Applied Probability* 19(2): 556-584.

Aït-Sahalia, Y., et al. (2015). "Modeling financial contagion using mutually exciting jump processes." *Journal of Financial Economics* 117(3): 585-606.

Aït-Sahalia, Y. and T. R. Hurd (2015). "Portfolio choice in markets with contagion." *Journal of Financial Econometrics* 14(1): 1-28.

Andersen, L. and M. Broadie (2004). "Primal-dual simulation algorithm for pricing multidimensional American options." *Management Science* 50(9): 1222-1234.

Andersen, T. G., et al. (2003). "Modeling and forecasting realized volatility." *Econometrica* 71(2): 579-625.

Ang, A. and J. Chen (2002). "Asymmetric correlations of equity portfolios." *Journal of Financial Economics* 63(3): 443-494.

Asai, M. and M. McAleer (2009). "The structure of dynamic correlations in multivariate stochastic volatility models." *Journal of Econometrics* 150(2): 182-192.

Barraquand, J. and D. Martineau (1995). "Numerical valuation of high dimensional multivariate American securities." *Journal of financial and quantitative analysis* 30(03): 383-405.

Belomestny, D., et al. (2009). "TRUE UPPER BOUNDS FOR BERMUDAN PRODUCTS VIA NON-NESTED MONTE CARLO." *Mathematical Finance* 19(1): 53-71.

Boyle, P. P. (1977). "Options: A monte carlo approach." *Journal of Financial Economics* 4(3): 323-338.

Boyle, P. P. and Y. Tse (1990). "An algorithm for computing values of options on the maximum or minimum of several assets." *Journal of financial and quantitative analysis* 25(02): 215-227.

Broadie, M. and P. Glasserman (1997). "Pricing American-style securities using simulation." *Journal of Economic Dynamics and Control* 21(8): 1323-1352.

Broadie, M. and P. Glasserman (1997). A Stochastic Mesh Method for Pricing High-Dimensional American Options.

Broadie, M. and P. Glasserman (2004). "A stochastic mesh method for pricing high-dimensional American options." *Journal of Computational Finance* 7: 35-72.

Broadie, M., et al. (1997). "Enhanced Monte Carlo estimates for American option prices." *The Journal of Derivatives* 5(1): 25-44.

Broadie, M. N. and P. Glasserman (1998). "A stochastic mesh method for pricing high-dimensional American options."

Bru, M.-F. (1991). "Wishart processes." *Journal of Theoretical Probability* 4(4): 725-751.

Buraschi, A., et al. (2010). "Correlation risk and optimal portfolio choice." *The Journal of Finance* 65(1): 393-420.

Carriere, J. F. (1996). "Valuation of the early-exercise price for options using simulations and nonparametric regression." *Insurance: mathematics and Economics* 19(1): 19-30.

Casella, G. and E. I. George (1992). "Explaining the Gibbs sampler." *The American Statistician* 46(3): 167-174.

Choi, D. B. (2014). "Heterogeneity and stability: Bolster the strong, not the

weak.“ *Review of Financial Studies* 27(6): 1830-1867.

Clark, C. E. (1961). "The greatest of a finite set of random variables.“ *Operations Research* 9(2): 145-162.

Clément, E., et al. (2002). "An analysis of a least squares regression method for American option pricing.“ *Finance and Stochastics* 6(4): 449-471.

Coleman, T. F. and Y. Li (1996). "An interior trust region approach for non-linear minimization subject to bounds.“ *SIAM Journal on optimization* 6(2): 418-445.

Corradi, V., et al. (2012). "International market links and volatility transmission.“ *Journal of Econometrics* 170(1): 117-141.

Das, S. R. and R. Uppal (2004). "Systemic risk and international portfolio choice.“ *The Journal of Finance* 59(6): 2809-2834.

Dungey, M., et al. (2009). "Flight-to-quality and asymmetric volatility responses in US Treasuries.“ *Global Finance Journal* 19(3): 252-267.

Eraker, B. (2004). "Do stock prices and volatility jump? Reconciling evidence from spot and option prices.“ *The Journal of Finance* 59(3): 1367-1403.

Eraker, B., et al. (2003). "The impact of jumps in volatility and returns.“ *The Journal of Finance* 58(3): 1269-1300.

Fan, Z., et al. (2015). "Asymmetric excitation and the US bias in Portfolio Choice.“

Forbes, K. J. and R. Rigobon (2002). "No contagion, only interdependence: measuring stock market comovements.“ *The Journal of Finance* 57(5): 2223-2261.

French, K. R. and J. M. Poterba (1991). *Investor diversification and international equity markets*, National Bureau of Economic Research.

Fu, M. C., et al. (2001). "Pricing American options: A comparison of Monte

Carlo simulation approaches.“ *Journal of Computational Finance* 4(3): 39-88.

Gelfand, A. E. and A. F. Smith (1990). "Sampling-based approaches to calculating marginal densities.“ *Journal of the American statistical association* 85(410): 398-409.

Gelman, A., et al. (2014). *Bayesian data analysis*, Chapman and Hall/CRC Boca Raton, FL, USA.

Giannetti, M. and L. Laeven (2012). "The flight home effect: Evidence from the syndicated loan market during financial crises.“ *Journal of Financial Economics* 104(1): 23-43.

Gilks, W. R., et al. (1995). "Adaptive rejection Metropolis sampling within Gibbs sampling.“ *Applied Statistics*: 455-472.

Glasserman, P. and B. Yu (2004). "Number of paths versus number of basis functions in American option pricing.“ *The Annals of Applied Probability* 14(4): 2090-2119.

Glasserman, P. and B. Yu (2004). *Simulation for American options: regression now or regression later?* Monte Carlo and Quasi-Monte Carlo Methods 2002, Springer: 213-226.

Gouriéroux, C. (2006). "Continuous time Wishart process for stochastic risk.“ *Econometric Reviews* 25(2-3): 177-217.

Haario, H., et al. (2006). "DRAM: efficient adaptive MCMC.“ *Statistics and Computing* 16(4): 339-354.

Haario, H., et al. (2001). "An adaptive Metropolis algorithm.“ *Bernoulli*: 223-242.

Hameed, A., et al. (2010). "Stock market declines and liquidity.“ *The Journal of Finance* 65(1): 257-293.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications.“ *Biometrika* 57(1): 97-109.

Haugh, M. B. and L. Kogan (2004). "Pricing American options: a duality approach." *Operations Research* 52(2): 258-270.

Heppenger, P. (2013). "Pricing high-dimensional Bermudan options using variance-reduced Monte Carlo methods." *Journal of Computational Finance* 16(3): 99-126.

Jacquier, E., et al. (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics* 12(4).

Jacquier, E., et al. (2004). "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors." *Journal of Econometrics* 122(1): 185-212.

Jain, S. and C. W. Oosterlee (2012). "Pricing high-dimensional Bermudan options using the stochastic grid method." *International Journal of Computer Mathematics* 89(9): 1186-1211.

Jin, X., et al. (2013). "A computationally efficient state-space partitioning approach to pricing high-dimensional American options via dimension reduction." *European Journal of Operational Research* 231(2): 362-370.

Jin, X., et al. (2007). "A STATE-SPACE PARTITIONING METHOD FOR PRICING HIGH-DIMENSIONAL AMERICAN-STYLE OPTIONS." *Mathematical Finance* 17(3): 399-426.

Jin, X. and A. X. Zhang (2012). "Decomposition of optimal portfolio weight in a jump-diffusion model and its applications." *Review of Financial Studies* 25(9): 2877-2919.

Jin, Xing, Cheng-Yu Yang and Xudong Zeng, Dynamic optimal portfolio choice problem under contagion, working paper, 2016.

Johannes, M. and N. Polson (2006). "MCMC Methods for Continuous-Time Financial Econometrics."

Johannes, M. S. and N. Polson (2003). "MCMC methods for continuous-time financial econometrics." Available at SSRN 480461.

Jorion, P. and G. Zhang (2009). "Credit contagion from counterparty risk." *The Journal of Finance* 64(5): 2053-2087.

Kim, S., et al. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models." *The Review of Economic Studies* 65(3): 361-393.

Kim, T. S. and E. Omberg (1996). "Dynamic nonmyopic portfolio behavior." *Review of Financial Studies* 9(1): 141-161.

Kohler, M., et al. (2010). "PRICING OF HIGH-DIMENSIONAL AMERICAN OPTIONS BY NEURAL NETWORKS." *Mathematical Finance* 20(3): 383-410.

Kolodko, A. and J. Schoenmakers (2004). "Upper bounds for Bermudan style derivatives." *Monte Carlo Methods and Applications mcma* 10(3-4): 331-343.

Kovalov, P., et al. (2007). "Pricing multi-asset American options: A finite element method-of-lines with smooth penalty." *Journal of Scientific Computing* 33(3): 209-237.

Lee, S. S. and J. Hannig (2010). "Detecting jumps from Lévy jump diffusion processes." *Journal of Financial Economics* 96(2): 271-290.

Lee, S. S. and P. A. Mykland (2008). "Jumps in financial markets: A new nonparametric test and jump dynamics." *Review of Financial Studies* 21(6): 2535-2563.

Leippold, M. and F. Trojani (2010). "Asset Pricing with Matrix Jump Diffusions." *SSRN Working Paper Series*.

Leitner, Y. (2005). "Financial networks: Contagion, commitment, and private sector bailouts." *The Journal of Finance* 60(6): 2925-2953.

Li, X. and Z. Wu (2006). "A semi-analytic method for valuing high-dimensional options on the maximum and minimum of multiple assets." *Annals of Finance* 2(2): 179-205.

- Liesenfeld, R. and J.-F. Richard (2006). "Classical and Bayesian analysis of univariate and multivariate stochastic volatility models." *Econometric Reviews* 25(2-3): 335-360.
- Liu, J., et al. (2003). "Dynamic asset allocation with event risk." *The Journal of Finance* 58(1): 231-259.
- Longstaff, F. A. and A. Rajan (2008). "An empirical analysis of the pricing of collateralized debt obligations." *The Journal of Finance* 63(2): 529-563.
- Longstaff, F. A. and E. S. Schwartz (2001). "Valuing American options by simulation: A simple least-squares approach." *Review of Financial Studies* 14(1): 113-147.
- Longin, F. and B. Solnik (2001). "Extreme correlation of international equity markets." *The journal of finance* 56(2): 649-676.
- Merton, R. C. (1976). "Option pricing when underlying stock returns are discontinuous." *Journal of Financial Economics* 3(1-2): 125-144.
- Metropolis, N., et al. (1953). "Equation of state calculations by fast computing machines." *The journal of chemical physics* 21(6): 1087-1092.
- Neal, R. M. (2003). "Slice sampling." *Annals of statistics*: 705-741.
- Omori, Y., et al. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference." *Journal of Econometrics* 140(2): 425-449.
- Pericoli, M. and M. Sbracia (2003). "A primer on financial contagion." *Journal of Economic Surveys* 17(4): 571-608.
- Rogers, L. C. (2002). "Monte Carlo valuation of American options." *Mathematical Finance* 12(3): 271-286.
- Shiu, Y.-M., et al. (2013). "A closed-form approximation for valuing European basket warrants under credit risk and interest rate risk." *Quantitative Finance* 13(8): 1211-1223.

Spall, J. C. (1998). "Implementation of the simultaneous perturbation algorithm for stochastic optimization." *Aerospace and Electronic Systems, IEEE Transactions on* 34(3): 817-823.

Spall, J. C. (2012). *Stochastic optimization. Handbook of computational statistics*, Springer: 173-201.

Stentoft, L. (2004). "Assessing the least squares Monte-Carlo approach to American option valuation." *Review of Derivatives research* 7(2): 129-168.

Tibbits, M. M., et al. (2011). "Parallel multivariate slice sampling." *Statistics and Computing* 21(3): 415-430.

Tierney, L. (1994). "Markov chains for exploring posterior distributions." *the Annals of Statistics*: 1701-1728.

Tilley, J. A. (1993). "Valuing American options in a path simulation model." *Transactions of the Society of Actuaries* 45(83): 104.

Tsitsiklis, J. N. and B. Van Roy (1999). "Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives." *Automatic Control, IEEE Transactions on* 44(10): 1840-1851.

Tsitsiklis, J. N. and B. Van Roy (2001). "Regression methods for pricing complex American-style options." *Neural Networks, IEEE Transactions on* 12(4): 694-703.

Van Nieuwerburgh, S. and L. Veldkamp (2009). "Information immobility and the home bias puzzle." *The Journal of Finance* 64(3): 1187-1215.

Yu, J., et al. (2010). "Simulation-based estimation methods for financial time series models."

Yu, J. and R. Meyer (2006). "Multivariate stochastic volatility models: Bayesian estimation and model comparison." *Econometric Reviews* 25(2-3): 361-384.

Zhu, H., et al. (2013). "Fast Estimation of True Bounds on Bermudan Op-

tion Prices under Jump-diffusion Processes.” arXiv preprint arXiv:1305.4321.

Appendix: Proofs

For illustrative simplicity, we consider constant jump case in the proofs below. Following similar steps, the general case where both jumps are random variables could be easily obtained.

Proof of Proposition 1

Proof. Recall the first order condition (3.17) is

$$\Sigma(\eta + g_t) - J\mathbf{E}[Y]C_1\Sigma C_2 - \gamma\Sigma\pi + \sum_{l=1}^m \lambda_l \mathbf{E}[(1 + \pi' J_l Y_l)^{-\gamma} J_l Y_l] e^{tr(A_t K_l)} = 0. \quad (50)$$

It can be re-written as

$$0 = \Sigma(\eta + g) - J\mathbf{E}[Y]C_1\Sigma C_2 - \gamma\Sigma\pi + JD_Y(\pi' J; A_t)\lambda \quad (51)$$

$$= \Sigma(\eta + g) - J\mathbf{E}[Y]C_1\Sigma C_2 - \gamma\Sigma\pi + JD_Y(\pi' J; A_t)C_1\Sigma C_2. \quad (52)$$

Hence for any $k = 1, 2, \dots, m$, we have

$$0 = J'_k \Sigma(\eta + g_t) - \gamma J'_k \Sigma \pi + J'_k J \cdot (D_Y(\pi' J; A_t) - \mathbf{E}[Y]) \cdot C_1 \Sigma C_2. \quad (53)$$

Since Σ is a variable symmetric matrix, the above equality implies

$$g_t J'_k + J_k g'_t + \eta J'_k + J_k \eta' - \gamma(\pi J'_k + J_k \pi) \quad (54)$$

$$+ C_2 J'_k J \cdot (D_Y(\pi' J; A_t) - \mathbf{E}[Y]) \cdot C_1 + C'_1 \cdot (D_Y(\pi' J; A_t) - \mathbf{E}[Y]) \cdot J' J_k C'_2 = 0 \quad (55)$$

Multiplying J'_k from the left side and J_k from the right side on the both sides of (54), we obtain

$$J'_k(g_t + \eta) - \gamma J'_k \pi + (J'_k C_2)(J'_k J) \cdot (D_Y(\pi' J; A_t) - \mathbf{E}[Y]) \cdot (C_1 J_k) = 0. \quad (56)$$

Note that we simply the above equation by exploiting the equalities, e.g. $J'_k g_t = g'_t J_k$.

In virtue of the orthogonal decomposition of $\mathcal{R}^{n \times 1}$, we have the orthogonal decomposition for π :

$$\pi = J(J'J)^{-1}\bar{\pi} + J^\perp \pi_\perp \quad (57)$$

for some $\bar{\pi} \in \mathcal{R}^{m \times 1}, \pi_{\perp} \in \mathcal{R}^{n-m \times 1}$.

Plugging (57) into (56), we obtain the nonlinear equations for $\bar{\pi}_k, k = 1, 2, \dots, m$:

$$J'_k(g_t + \eta) - \gamma \bar{\pi}_k + (J'_k C_2)(J'_k J) \cdot (D_Y(\bar{\pi}'; A_t) - \mathbf{E}[Y]) \cdot (C_1 J_k) = 0. \quad (58)$$

On the other hand, multiplying $(J_k^{\perp})'$ from the left side on the both sides of (51), and similarly we can obtain

$$(J_k^{\perp})'(g_t + \eta) - \gamma (J_k^{\perp})' \pi = 0. \quad (59)$$

Hence $\pi_{\perp} = \frac{1}{\gamma} (J^{\perp})'(g_t + \eta)$. □

Proof of Proposition 2

Proof. Since $\text{rank}(J) = m$, such a square matrix P always exists. (In fact, J can be transferred to a matrix with pairwise orthogonal columns by *row transformations* only if $\text{rank}(J) = m$). Denote

$$\hat{\Sigma} = P \Sigma P'.$$

Then under the transformation of P , the first order condition (3.17) becomes

$$\hat{\Sigma}(\hat{\mu} + \hat{g}_t) - \gamma \hat{\Sigma} \hat{\pi} + \hat{J} \cdot (D_Y(\hat{\pi}; A_t) - \mathbf{E}[Y]) \cdot \hat{C}_1 \hat{\Sigma} \hat{C}_2 = 0.$$

Similar to the proof of Prop. 1, exploiting the symmetricalness of $\hat{\Sigma}$ and the prosperities of matrix trace, we obtain

$$\hat{J}'_k(\hat{g}_t + \hat{\eta}) - \gamma \bar{\pi}_k + (\hat{J}'_k \hat{C}_2)(\hat{J}'_k \hat{J}) \cdot (D_Y(\hat{\pi}; A_t) - \mathbf{E}[Y]) \cdot (\hat{C}_1 \hat{J}_k) = 0.$$

□

Proof of Proposition 3

Proof. Under the conditions given, the first order condition (3.17) becomes

$$\Sigma \eta - \gamma \Sigma \pi + J(D_Y(\pi' J; 0) - \mathbf{E}[Y]) \lambda_t = 0. \quad (60)$$

Multiplying Σ^{-1} and J' from the left hand side on the both side of the above equality, we obtain

$$J' \eta - \gamma J' \pi + (J' \Sigma^{-1} J) \lambda \cdot ((1 + \pi' J u)^{-\gamma} - 1) u = 0, \quad (61)$$

where u is the constant jump size. Then since Σ is positive semi-definite, λ_t is non-negative, and $J'\eta$ is positive as given, we obtain the conclusion straightforwardly. \square