

# **Three Essays on Pricing and Hedging in Incomplete Markets**

A thesis presented by

**Dan Chen**

London, October 19, 2011

Submitted for the for the degree of Doctor of Philosophy in the subject of Statistics,  
London School of Economics and Political Science.



# Acknowledgement

I would like to thank Dr. Thorsten Rheinländer for his support and encouragement throughout my master and PhD study. I cannot complete this dissertation without his original ideas and technical advices. I also would like to thank him for the research assistant position funded by the EPSRC grant.

My thanks also goes to Prof. Qiwei Yao, Dr. Kostas Kalogeropoulos, Sujin Park, Hongbiao Zhao, Roy Rosemarin and many other people from the Statistics Department for discussions and suggestions on my research. I also appreciate the improvement advices from my two examiners, Prof. Dr. Rüdiger Kiesel and Dr. Angelos Dassios.

The funding from the department made it is possible for me to live in London independently and thus is greatly appreciated.

The biggest thanks, however, goes to my parents for their support, understanding and encouragement.



# Abstract

The thesis focuses on valuation and hedging problems when the market is incomplete. The first essay considers the quadratic hedging strategy. We propose a generalized quadratic hedging strategy which can balance a short-term risk (additional cost) with a long-term risk (hedging errors). The traditional quadratic hedging strategies, i.e. self-financing strategy and risk-minimization strategy, can be seen as special cases of the generalized quadratic hedging strategy. This is applied to the insurance derivatives market.

The second essay compares parametric and nonparametric measure-changing techniques. The essay discusses three pricing approaches: pricing via Esscher measure, via calibration and via nonparametric risk-neutral density; and empirically compares the performance of the three approaches in the metal futures markets.

The last essay establishes the concept of stochastic volatility of volatility and proposes several estimation methods.



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| 1.1      | Financial Mathematics and Financial Markets . . . . .                | 1         |
| 1.1.1    | Insurance derivatives . . . . .                                      | 1         |
| 1.1.2    | Metal futures . . . . .  | 2         |
| 1.1.3    | Equity and volatility derivatives . . . . .                          | 3         |
| 1.2      | Hedging in an incomplete market . . . . .                            | 3         |
| 1.2.1    | Dynamic quadratic hedging strategies . . . . .                       | 3         |
| 1.2.2    | Other hedging strategies . . . . .                                   | 4         |
| 1.3      | Pricing in an incomplete market . . . . .                            | 5         |
| <b>2</b> | <b>Generalized Quadratic Hedging Strategies</b>                      | <b>9</b>  |
| 2.1      | Introduction . . . . .   | 9         |
| 2.2      | Quadratic hedging strategies for payment streams . . . . .           | 10        |
| 2.3      | Generalised quadratic hedging strategy . . . . .                     | 12        |
| 2.4      | An example of a hedging annuity portfolio . . . . .                  | 16        |
| 2.4.1    | Annuity portfolio . . . . .  | 17        |
| 2.4.2    | Mortality forward . . . . .  | 18        |
| 2.4.3    | Hedging strategies and remaining risks . . . . .                     | 18        |
| 2.4.4    | Numerical analysis . . . . .   | 20        |
| 2.5      | Conclusion . . . . .   | 29        |
| <b>3</b> | <b>Risk-Neutral Densities and Metal Futures</b>                      | <b>33</b> |
| 3.1      | Introduction . . . . .   | 33        |
| 3.2      | Method1: price commodity futures under the Esscher measure . . . . . | 34        |

|          |  |           |
|----------|--|-----------|
| 3.2.1    | Dynamics for the commodity spot prices . . . . .                             | 34        |
| 3.2.2    | Risk-neutral distribution estimated from spot price data . . . . .           | 36        |
| 3.2.3    | Futures price under the Esscher measure . . . . .                            | 37        |
| 3.3      | Method 2: model spot prices under the risk-neutral measure $Q$ . . . . .     | 40        |
| 3.4      | Method 3: non-parametric estimation of risk-neutral densities . . . . .      | 41        |
| 3.5      | Empirical analysis . . . . .   | 42        |
| 3.5.1    | Estimation method and simulation techniques . . . . .                        | 42        |
| 3.5.2    | Gold and gold future market . . . . .  | 44        |
| 3.5.3    | Copper and copper future market . . . . .                                    | 52        |
| 3.5.4    | Aluminum and aluminum future market . . . . .                                | 58        |
| 3.6      | Conclusion . . . . .   | 65        |
| <b>4</b> | <b>Stochastic Volatility of Volatility</b>                                   | <b>69</b> |
| 4.1      | Introduction . . . . .   | 69        |
| 4.2      | The Concept of Volatility of Volatility . . . . .                            | 71        |
| 4.3      | Stochastic Volatility of Volatility Model and Estimation Methods . . . . .   | 73        |
| 4.3.1    | Bayesian MCMC . . . . .  | 74        |
| 4.3.2    | Maximum likelihood estimation via closed-form likelihood expansion . . . . . | 81        |
| 4.3.3    | Calibration . . . . .  | 83        |
| 4.4      | Conclusion . . . . .   | 87        |

# Chapter 1

## Introduction

### 1.1 Financial Mathematics and Financial Markets

Earlier works in financial mathematics (Black and Scholes (1973), Merton (1973), etc.) are based on the strong assumption that the market is complete, i.e. that all contingent claims are replicable by investing in the underlying asset.

However, empirical studies suggest that the market is not complete in practice due to various reasons like illiquidity, stochastic volatility or jumps in price processes [6]. Once we remove the completeness assumption, we have to face a so-called incomplete market. In such markets, the pricing by replication is not working so optimal pricing and hedging depend on the criteria chosen.

Before going into the details of pricing and hedging strategies designed for incomplete markets, we would like to introduce the financial markets that will be studied later on.

#### 1.1.1 Insurance derivatives

Insurance-linked securities have been used as a tool to transfer insurance risks from the insurance industry to the capital market. At the same time, they provide the financial market with a diversification tool, since insurance risks are often uncorrelated with existing financial risks. According to Barrieu and Albertini (2009) [1], there were approximately \$13bn of tradable non-life insurance-linked securities and \$24bn in tradable life insurance-linked securities by the end of 2008.

The most successful non-life insurance-linked security is the catastrophe bond (cat bond). A cat bond can be understood as an 'insurance contract' with the insurance or reinsurance companies as the insured and capital market investors as the insurer. The investors can receive interest payments and get back the principal as long as the natural disaster does not occur before the maturity of the bond. More details about cat bonds and other non-life insurance-linked securities can be found in [1].

Life insurance-linked securities include longevity bonds, survivor swaps, mortality

forwards etc. The typical underlying risks are mortality and longevity risk. Mortality risk indicates the risk that the actual death rate exceeds the expected rate, whereas longevity risk is the risk that the actual mortality becomes lower than expected. Clearly there are plenty of organizations, such as pension funds and life-insurance companies, which have exposures to mortality risk or longevity risk. Therefore they are natural players in the life insurance-linked securities market. Capital market speculators, like hedge funds, have also entered this market to diversify their portfolio or earn speculating returns.

Insurance derivatives markets are in general incomplete due to illiquidity and basis risks. The basis risks arise from the inconsistency between the underlying insurance risk of the derivatives and the insurance risk of the hedger. For instance, an annuity portfolio can have exposure to the longevity risks of a certain group of people in England, but the tradable life-insurance derivatives, which the portfolio manager can choose as hedging instruments, are based on nationwide mortality rates.

### 1.1.2 Metal futures

Metal futures markets have a long history and are in general mature markets. Precious metal futures like gold are traded on Commodity Exchange, Inc. (COMEX) and lots of local exchanges. Industrial metal futures, like copper, aluminium, zinc, etc. are traded on London Metal Exchange (LME), COMEX, Shanghai Future Exchange, etc.

At a certain time point, we have a futures price curve consisting of spot price and prices of futures with different maturities. We can describe the curve to be *normal* if it is upward sloping over different maturities (starting with time 0) and to be *inverted* if it is downward sloping. When we look at the evolvement of the futures price over time, we will find that it will converge to the spot price when closer to the maturity. We call the market is *contango* when the futures price is expected to decrease over time to the futures spot price, and call the market is *normal backwardation* if the futures price is expected to increase to the futures spot price as time goes by. (See [27] for more details)

Unlike financial assets such as stock and bond, commodities can be consumed. Moreover, it is sometimes beneficial to hold the commodity rather than the derivatives, especially when the futures curve is inverted. However, it is not always a good idea to hold a great deal of commodities since the storage and maintenance costs can be huge. A concept called *convenience yield* has been introduced to describe the advantage to hold commodities. Geman (2005) [14] and Carmona and Ludkovshi (2005) [2] define convenience yield as the difference between benefit of direct access and cost of storage. Empirical work implies that convenience yields arise endogenously as a result of the interaction between supply, demand and storage decisions [7]. Many authors propose methods to model convenience yield ([21], [15], [7], etc.).

Another concept, which frequently occurs in commodity future studies, is *cost-of-carry*. Hull (2003) [17] defines the cost-of-carry as the storage cost plus the interest

less the income earned on the asset.

Metal futures can serve as price discovery and risk management tools for metal spot markets. However the metal future cannot be evaluated using complete market techniques. If we assume future price as the conditional expectation of the spot, market will be incomplete mainly because: (i) the convenience yield is unobservable and unhedgeable; (ii) there could be basis risks between the existing position and hedging instruments; (iii) many metal spot markets are not liquid enough.

### 1.1.3 Equity and volatility derivatives

Methods to define and compute volatility can be broadly divided into two groups: the volatility measurements based on underlying asset prices and the ones based on derivatives prices. Please see chapter four for details.

Volatility is an important concept in financial mathematics. It is a key input to the Black-Scholes formula and some other option pricing formulas. The positive derivatives of option prices with respect to volatility increase the value of options when the market becomes more volatile. Before the existence of volatility derivatives, there were several trading strategies to bet on volatility using options on the underlying asset. [26] gives a survey on these strategies: delta-neutral portfolio of stocks and options, straddles and strangles, volatility surface trading, etc.

With the development of volatility indices (check [5] for more details), volatility derivatives begin to appear on the financial markets. After the reversion of the CBOE volatility index (VIX), exchange-traded volatility derivatives have been introduced. Successful examples of volatility derivatives include variance swaps, volatility swaps, volatility index (VIX) futures and options [4]. Today variance and volatility swaps are traded over-the-counter whereas VIX futures and options are actively traded on the Chicago Futures Exchange (CFE), a division of the Chicago Board Options Exchange (CBOE). The VIX options are the CBOE's most liquid option contract after the SPX index options [4].

Volatility derivatives cannot be priced or hedged using complete market methods because the underlying 'asset' is typically a non-tradable index.

## 1.2 Hedging in an incomplete market

### 1.2.1 Dynamic quadratic hedging strategies

One of the reasons that quadratic hedging is popular among scholars and practitioners is that it can be formulated in a mathematically elegant way and solved by relatively easy approaches. Dynamic quadratic hedging in a complete market is always done by perfectly replicating the derivatives with a dynamic self-financing strategy. In an incomplete market, this is not feasible. Either we need to give up the self-financing strategy or give up the perfect replication [22].

Föllmer and Sondermann (1986) [12] formulate the risk-minimization strategy, which is a breakthrough in quadratic hedging in incomplete market. The strategy adopts a mean-self-financing strategy rather than self-financing strategy to minimize the expected future costs. The optimal strategy, consisting of units of risky asset and units of riskless asset, can be uniquely determined. Møller (2001) [16] extends Föllmer and Sondermann (1986) to payment stream cases. Please check the chapter two for more information on Föllmer and Sondermann (1986) and Møller (2001)'s work.

Föllmer and Sondermann (1986)'s model aims to hedge under the risk-neutral measure, i.e., the price processes of hedging instruments are martingales under  $P$ . Schweizer (2001) [22] gives an introduction to quadratic hedging strategies without payment streams in the semimartingale case. The local risk-minimization strategies can be found when the cost process is a martingale which must be orthogonal to the martingale part of the price process. One needs to utilize the Föllmer-Schweizer decomposition, the classical Kunita-Watanabe decomposition computed under the minimal martingale measure, to construct this strategy. Another hedging method in the semimartingale case, the mean-variance optimal strategy, is done by a  $L^2$  projection under the variance-optimal martingale measure.

Cont, Tankov and Voltchkova (2007) [8] study quadratic hedging strategies when the underlying asset process has jumps.

### 1.2.2 Other hedging strategies

There are some other popular hedging strategies, which do not belong to the quadratic hedging family, and which we survey briefly for completeness. The first one we would like to discuss is the *super hedging strategy*. This approach, according to Pham (1999), is to look for an initial capital  $x \geq 0$  and an admissible trading strategy  $\theta$  such that

$$x + \int_0^T \theta_t dS_t \geq H = g(S_T), \text{ a.s.}$$

The weakness of the super hedging strategy is that the cost of hedging could be too high to be acceptable. If a hedger only has limited initial funding, she could adopt a *quantile hedging strategy*. Föllmer and Leukert (1999) [13] describe this approach. We are looking for an admissible strategy  $(V_0, \theta)$  such that

$$P \left[ V_0 + \int_0^T \theta_t dS_t \geq H \right] = \max, \\ \text{when } V_0 \leq \tilde{V}_0$$

where  $\tilde{V}_0$  is the capital constraint.

A drawback of quadratic hedging strategy is that it minimizes the potential losses as well as potential gains. Therefore researchers also proposed several non-quadratic hedging strategies which target on only one side of the risk. *Shortfall risk-minimization* belongs to this category. According to Pham (2002) [19] and Favero

(2004) [11], shortfall risk-minimization is to minimize the criterion

$$\min_{\varphi} E \left[ l \left( (H(S_T) - V_T(\varphi))^+ \right) \right].$$

The problem with strategies targeting one side of risk is the absence of a nice mathematical formulation and explicit optimal solution.

All the hedging strategies mentioned above are dynamic hedging strategies, which is not feasible in reality due to market frictions. Static hedging strategies have been studied by Derman, Ergener and Kani (1995) [10], Carr, Ellis and Gupta (1998) [3], Poulsen(2006) [20], etc.

Dahl, Glar and Møller (2009) [9] proposed a modified quadratic hedging and name it as mixed dynamic and static hedging strategy. Recall that a strategy is call risk-minimizing if it can minimize the criterion

$$R(t, \varphi) = E \left[ (C_T(\varphi) - C_t(\varphi))^2 \middle| \mathcal{F}_t \right].$$

where  $\varphi$  represents the hedging strategy and  $C(\cdot)$  represents the accumulated cost process. A mixed dynamic and static hedging strategy is defined as the strategy  $\varphi^*$  that minimizes

$$R(t_i, \varphi), \text{ for } i = 0, 1, 2, \dots, n-1$$

where the  $(t_i)$  is a fixed time grid.

### 1.3 Pricing in an incomplete market

In a complete market, the derivatives price equals the cost of the self-financing replication portfolio in order to rule out arbitrage opportunities. In an incomplete market, we can construct several different hedging portfolios with different initial costs, and thus the derivative price cannot be uniquely determined. An alternative way to evaluate derivatives is to find out a suitable risk-neutral probability measure and then take the (conditional) expectation under this measure.

[25] describes an 'economic' interpretation of risk-neutral probabilities. Let us imagine there is a kind of security call *Arrow security*, whose payoff is associated with a particular state of the world. If this state occurs, the holder of the Arrow security will be paid £1, and nothing otherwise. The risk-neutral probability at a state is nothing but the price of the Arrow security associated with that state, given the risk-free rate is zero. Obviously, any financial asset can be expressed as a portfolio of Arrow securities. By the non-arbitrage rule, the price of the financial asset should equal the price of the Arrow security portfolio.

Harrison and Kreps (1979) and Harrison and Pliska (1981) established the link between no-arbitrage pricing and martingale theory.

**Theorem 1** (*First fundamental theorem of asset pricing, from [24]*) *If a market model has a risk-neutral probability measure, then it does not admit arbitrage.*

**Theorem 2** (*Second fundamental theorem of asset pricing, from [24]*) *Consider a market model that has a risk-neutral probability measure. The model is complete if and only if the risk-neutral probability measure is unique.*

Put in a different way, in an incomplete market, there exist different risk-neutral measures and hence different no-arbitrage prices. Popular risk-neutral measures include minimal entropy martingale measure (MEMM), Esscher Measure, minimal martingale measure, etc.

[23] gives the definition of MEMM: Fix a time horizon  $T < \infty$ . An equivalent local martingale measure  $Q^E$  for  $S$  on  $[0, T]$  is called *minimal entropy martingale measure (MEMM)* if  $Q^E$  minimises the relative entropy  $H(Q | P)$  over all equivalent local martingale measures  $Q$  for  $S$  on  $[0, T]$ . The MEMM is closely linked to the exponential utility maximization problem, and has connection with the *Esscher measure*. Chapter three gives the definition of the Esscher measure. The minimal martingale measure and the variance-optimal martingale measure mentioned in the previous subsection can also be used as pricing measures.

There are some nonparametric estimation methods to determine the risk-neutral density, hence the probability measure. The biggest advantage of nonparametric methods is that they do not rely on concrete assumptions of underlying asset price processes. Chapter three introduces a nonparametric method and compares it with parametric methods.

# Bibliography

- [1] P. Barrieu, L. Albertini (2009). The Handbook of Insurance-Linked Securities. Wiley, Chichester.
- [2] R. Carmona, M. Ludkovski (2004). Spot Convenience Yield Models for Energy Markets. AMS Mathematics of Finance, Contemporary Mathematics vol. 351, pp. 65–80.
- [3] P. Carr, K. Ellis and V. Gupta (1998). The Journal of Finance. Vol. 53, No. 3, 1165-1190.
- [4] P. Carr and R. Lee (2009). Volatility Derivatives. Annual Review of Financ. Econ. 1:1–21.
- [5] P. Carr and D.B. Madam (1998). Towards a Theory of Volatility Trading. Risk Books. chap. 29, 417–427.
- [6] P. Carr, H. Geman, D.B. Madan (2001). Pricing and Hedging in Incomplete Markets. Journal of Financial Economics 62, 131–167.
- [7] J. Casassus, P. Collin-Dufresne (2001). 'Maximal' Convenience Yield Model Implied by Commodity Futures. EFA 2002 Berlin Meetings Presented Paper.
- [8] R. Cont, P. Tankov, and E. Voltchkova (2007), Hedging with options in models with jumps. Stochastic Analysis and Applications - the Abel Symposium 2005, Springer. Vol 2, 197–218.
- [9] M. Dahl, S. Glar and T. Møller (2010). Mixed Dynamic and Static Risk-minimization with an Application to Survivor Swaps. To appear in European Actuarial Journal.
- [10] E. Derman, D. Ergener and I. Kani (1995). Static options replication. The Journal of Derivatives. Vol 2, No. 4, 78–95.
- [11] G. Favero (2005). Shortfall Risk Minimisation versus Symmetric (quadratic) Hedging. Decisions in Economics and Finance. Vol 28. No. 1, 1-8.
- [12] H. Föllmer, D. Sondermann (1986). Hedging of Non-redundant Contingent Claims. Contributions to Mathematical Economics. North-Holland, 205-223.

- [13] H. Föllmer, P. Leukert (1999). Quantile Hedging. *Finance and Stochastics*. Vol 3, No. 3. 251-274.
- [14] H. Geman (2005). *Commodities and Commodity Derivatives*. Wiley Finance Series.
- [15] R. Gibson and E.S. Schwartz (1990). Stochastic Convenience Yield and the Pricing of Oil Contingent Claims. *Journal of Finance*. 45, 959-976.
- [16] T. Møller (2001). Risk-minimizing Hedging Strategies for Insurance Payment Processes. *Finance and Stochastics*. Vol 5, No. 4. 419-446.
- [17] J.C. Hull (2003). *Options, Futures, and Other Derivatives*. Fifth Edition. Pearson Education.
- [18] H. Pham (1999). Hedging and Optimization Problems in Continuous-Time Financial Models. Lecture presented at ISFMA Symposium on Mathematical Finance, Fudan University.
- [19] H. Pham (2002). Minimizing Shortfall Risk and Applications to Finance and Insurance Problems. *Annals of Applied Probability*. Vol 12, No. 1, 143-172.
- [20] R. Poulsen (2006). Barrier options and their static hedges: simple derivations and extensions. *Quantitative Finance* Vol 6, No. 4, 327-335.
- [21] E.S. Schwartz (1997). The Stochastic Behavior of Commodity Prices: Implication for Valuation and Hedging. *Journal of Finance*. Vol 52, No. 3, 923-973.
- [22] M. Schweizer (2001). A Guided Tour through Quadratic Hedging Approaches. *Option Pricing, Interest Rates and Risk Management*. Cambridge University Press (2001), 538-574.
- [23] M. Schweizer (2010). The Minimal Entropy Martingale Measure. *Encyclopedia of Quantitative Finance*, Wiley, 1195-1200
- [24] S.E. Shreve (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer.
- [25] R. Sundaram (1997). Equivalent Martingale Measures and Risk-Neutral Pricing: An Expository Note. *The Journal of Derivatives*. 85-98.
- [26] Volatility Arbitrage Supplement (2006). [www.incisivemedia.com/hfr/specialreport/VolArbSupp.pdf](http://www.incisivemedia.com/hfr/specialreport/VolArbSupp.pdf)
- [27] D. Harper (2007). Contango Vs. Normal Backwardation. Investopedia. [http://www.investopedia.com/articles/07/contango\\_backwardation.asp#axzz1WiyBfZxG](http://www.investopedia.com/articles/07/contango_backwardation.asp#axzz1WiyBfZxG)

## Chapter 2

# Generalized Quadratic Hedging Strategies

### 2.1 Introduction

The incentive of this study is to find a dynamic hedging strategy in the context of insurance claims which can balance a short-term risk (additional costs) with a long-term risk (hedging errors). Quadratic hedging approaches have been studied very intensively in recent decades. A survey of dynamic hedging strategies based on a quadratic criterion for contingent claims without payment stream is given by Schweizer (2001) [9]. For an incomplete market, to find a perfect hedge (self-financing and without hedging error at maturity) for every claim is by definition impossible. The first remedy is to look for a strategy without hedging error and with a "small" cost. This so-called risk-minimization approach has been firstly formulated by Föllmer and Sondermann (1986) [7]. Another possibility is to find a self-financing strategy with a "small" hedging error. Møller (2001) [8] and Schweizer (2008) [10] extended the dynamic hedging approach to the payment streams case.

The trade-off between cost and hedging error can be seen as balancing a short-term and a long-term goal. While the aim of the risk-minimization strategy is to fulfil the long-term goal, the self-financing strategy puts emphasis on the short-term one. These are two fundamentally different methods. In this paper we are pursuing a multi-criteria optimization which aims for interpolation between these two extremes. It turns out that it is possible to find some kind of strategy enabling a hedger to decide how important one goal is relative to the other.

In the last part, an application is considered to the hedge of an annuity portfolio, which comprises a typical payment stream, by mortality forwards. As a new type of hedging and speculating instrument, insurance derivatives, such as mortality forwards, have recently been received some attention from industry, see Barrieu and Albertini [1] for a detailed overview of the securitization of mortality risk. Major investment banks,

such as JP Morgan, Credit Suisse and Goldman Sachs, have introduced mortality indices, thereby promoting the transactions of insurance derivatives.

The structure of this study can be outlined as follows. The first section summarizes risk minimization and self-financing hedging strategies for payment streams; the second section introduces different criteria to represent risks and considers the optimal hedging strategies under those criteria; the final section provides an example of hedging an annuity portfolio by mortality forwards.

## 2.2 Quadratic hedging strategies for payment streams

Let  $(\Omega, \mathcal{G}, Q)$  be a probability space equipped with a filtration  $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T}$  satisfying the usual conditions. We assume that  $\mathcal{G}_0$  is trivial (apart from containing the  $(P, \mathcal{G})$ -zero sets) and that  $\mathcal{G}_T = \mathcal{G}$ . In the following, equalities between random variables are understood in the almost sure sense.

An agent faces a *payment stream*  $V$  which is modelled as a square-integrable semimartingale, i.e.  $V_t \in L^2(Q)$  for all  $t \in [0, T]$ . To hedge her risks, she invests into a risky asset with value process  $U$  and holds some amount of cash on a savings account (we assume that interest rates are zero). We assume that  $U$  is a square-integrable  $Q$ -martingale.

The space  $L^2(U)$  consists of all  $\mathbb{G}$ -predictable processes  $\theta$  satisfying  $E^Q \left[ \int_0^T \theta_s^2 d[U]_s \right] < \infty$ . An *admissible strategy* is a pair of processes  $\varphi = (\theta, \eta)$ , where  $\theta \in L^2(U)$  and  $\eta$  is  $\mathbb{G}$ -adapted. Intuitively,  $\theta_t$  is the number of shares held in the risky asset  $U$  and  $\eta_t$  represents the value of the savings account, at time  $t \geq 0$ . We define the *value process* of the trading strategy  $\varphi$  by

$$Y_t(\varphi) = \theta_t U_t + \eta_t, \quad t \in [0, T]. \quad (2.1)$$

**Definition 3** *The accumulated cost process  $C$  is defined by*

$$C_t = Y_t(\varphi) - \int_0^t \theta_s dU_s + V_t, \quad t \in [0, T]. \quad (2.2)$$

*A strategy  $\varphi$  is called self-financing if the cost process  $C \equiv C_0$  is constant, which implies that*

$$dY_t(\varphi) = \theta_t dU_t - dV_t.$$

*If there is a self-financing strategy  $\varphi$  such that  $Y_T(\varphi) = 0$ , then the payment stream  $V$  is called attainable.*

According to Møller (2001),  $Y_t$  can also be interpreted as the value of our asset after a payment  $dV_t$  has been made at time  $t$ . If there is a self-financing strategy  $\varphi$  which can achieve  $Y_T = 0$ , which means that the liability  $V$  has been perfectly hedged, then we say that the claim  $V$  is attainable. In general, our market is incomplete which means that not all payment streams are attainable.

**Definition 4** *The risk process of  $\varphi$  is defined by*

$$R_t(\varphi) = E^Q \left[ (C_T(\varphi) - C_t(\varphi))^2 \mid \mathcal{G}_t \right].$$

*A strategy is said to be risk-minimising if it minimises  $R_t(\varphi)$  for all  $t$ . A 0-admissible risk-minimising strategy  $\varphi$  is a risk-minimising strategy which satisfies*

$$Y_T(\varphi) = 0.$$

A risk-minimising strategy can be obtained by the Kunita-Watanabe projection theorem. By projecting the martingale  $\widehat{V}_t := E^Q[V_T \mid \mathcal{G}_t], 0 \leq t \leq T$  on the stable subspace generated by the square-integrable martingale  $U$  under  $Q$ , we can get the decomposition

$$\widehat{V}_t = E^Q[V_T \mid \mathcal{G}_t] = E^Q[V_T] + \int_0^t \theta_s^V dU_s + L_t^V, \quad 0 \leq t \leq T$$

where the  $(\mathcal{G}_t)$ -adapted process  $L^V$  is a square-integrable martingale strongly orthogonal to  $U$  with  $L_0^V = 0$ . Therefore,

$$V_T = \widehat{V}_T = E^Q[V_T \mid \mathcal{G}_T] = E^Q[V_T] + \int_0^T \theta_s^V dU_s + L_T^V. \quad (2.3)$$

Møller (2001) shows that there exists a unique 0-admissible risk-minimizing strategy  $\varphi = (\theta, \eta)$  for  $V$  given by

$$(\theta_t, \eta_t) = \left( \theta_t^V, \widehat{V}_t - V_t - \theta_t^V U_t \right), \quad 0 \leq t \leq T.$$

The associated risk process is given by  $R_t(\varphi) = E^Q \left[ (L_T^V - L_t^V)^2 \mid \mathcal{F}_t \right]$ .

Now we try a different kind of quadratic hedging. Our goal is now to minimise the hedging error using a self-financing strategy

$$\begin{aligned} \min_{\varphi} E^Q[Y_T(\varphi)] &= \min_{\varphi} E^Q \left[ \left( C + \int_0^T \theta_s dU_s - V_T \right)^2 \right] \\ &= \min_{\varphi} E^Q \left[ \left( V_T - C - \int_0^T \theta_s dU_s \right)^2 \right]. \end{aligned}$$

Employing the Kunita-Watanabe decomposition (2.3), the problem can be rewritten-

ten as

$$\begin{aligned}
& \min_{\varphi} E^Q \left[ \left( V_T - C - \int_0^T \theta_s dU_s \right)^2 \right] \\
&= \min_{\varphi} E^Q \left[ \left( E^Q[V_T] + \int_0^T \theta_s^V dU_s + L_T^V - C - \int_0^T \theta_s dU_s \right)^2 \right] \\
&= \min_{\varphi} E^Q \left[ (C - E^Q[V_T])^2 + \left( \int_0^T (\theta_s - \theta_s^V) dU_s \right)^2 + (L_T^V)^2 \right].
\end{aligned}$$

Thus, the optimal strategy is

$$\begin{aligned}
\theta_t &= \theta_t^V, \\
C &= E^Q[V_T], \\
\eta_t &= Y_t - \theta_t^V U_t.
\end{aligned}$$

## 2.3 Generalised quadratic hedging strategy

In this section, we are trying to deal with the trade-off between the terminal hedging error and the additional cost over the whole dynamic hedging process. The problem can be formulated as:

$$\min_{(\theta, \eta)} (J_1(\theta, \eta), J_2(\theta, \eta)) \equiv \min_{(\theta, \eta)} \left( E^Q[Y_T^2], E^Q[(C_T - C_0)^2] \right).$$

Here  $Y_T$  represents the hedging error at terminal time  $T$ , while  $C_T - C_0$  represents the additional cost. In this case, we have a multi-objective problem and thus need to re-define the concept of an optimal solution. The following definition is adapted from Ehr Gott (2005) [6], p.38.

**Definition 5** *An admissible solution  $(\theta^*, \eta^*)$  is called weakly Pareto optimal if there is no admissible pair  $(\theta, \eta)$  such that*

$$J_1(\theta, \eta) < J_1(\theta^*, \eta^*), \quad J_2(\theta, \eta) < J_2(\theta^*, \eta^*).$$

Moreover, it seems to be part of the folklore that this weakly optimal solution of the vector optimization problem can be obtained by solving the following scalar optimisation problem, since the objective functionals in our case are (componentwise) strictly positive and the domain is also convex.

$$\min_{(\theta, \eta)} (\lambda J_1(\theta, \eta) + (1 - \lambda) J_2(\theta, \eta)) \equiv \min_{(\theta, \eta)} \left( \lambda E^Q[Y_T^2] + (1 - \lambda) E^Q[(C_T - C_0)^2] \right) \quad (2.4)$$

Here  $\lambda \in (0, 1)$ . In fact, a proof of this has been provided in Ehrgott (2005), p.78 in the context of a finite dimensional vector space setting. However, despite that this result is mentioned in an infinite dimensional setting in several articles, we were not able to spot a proof of this conjecture. Nevertheless, we will use a dynamic version of the scalar optimization problem (2.4) which is in itself meaningful as a starting point.

**Definition 6** *The dynamic hedging criterion is defined as follows for every  $t \in [0, T]$ :*

$$\text{ess inf } E^Q \left[ \lambda (Y_T)^2 + (1 - \lambda) (C_T - C_t)^2 \middle| \mathcal{G}_t \right],$$

where we minimise over admissible hedging strategies  $\varphi = (\theta, \eta)$  from  $t$  to  $T$ , and under the initial condition

$$C_0 = E^Q[C_T] = E^Q[V_T].$$

**Lemma 7** *The cost process  $C(\varphi)$  is a martingale if the strategy  $\varphi = (\theta, \eta)$  can achieve the essential infimum of the dynamic hedging criterion.*

**Proof.** Let  $s \in [0, T]$  be arbitrary. Define a strategy  $\tilde{\varphi}$  by setting  $\tilde{\theta} = \theta$ , and choosing  $\tilde{\eta}$  such that  $Y_t(\tilde{\varphi}) = Y_t(\varphi)$  for  $t \in [0, s]$ , and, for  $t \in [s, T]$ ,

$$Y_t(\tilde{\varphi}) = E^Q \left[ Y_T(\varphi) - \int_t^T \theta_s dU_s + V_T - V_t \middle| \mathcal{G}_t \right].$$

We have  $Y_T(\tilde{\varphi}) = Y_T(\varphi)$ ,  $C_T(\tilde{\varphi}) = C_T(\varphi)$ . Therefore,

$$C_T(\varphi) - C_s(\varphi) = C_T(\tilde{\varphi}) - C_s(\tilde{\varphi}) + E^Q [C_T(\tilde{\varphi}) | \mathcal{G}_s] - C_s(\varphi).$$

Note that since  $C_s(\tilde{\varphi}) = E^Q [C_T(\tilde{\varphi}) | \mathcal{G}_s]$  by construction, we have

$$E^Q \left[ (C_T(\varphi) - C_s(\varphi))^2 \middle| \mathcal{G}_s \right] = E^Q \left[ (C_T(\tilde{\varphi}) - C_s(\tilde{\varphi}))^2 \middle| \mathcal{G}_s \right] + (E^Q [C_T(\tilde{\varphi}) | \mathcal{G}_s] - C_s(\varphi))^2.$$

Hence,

$$\begin{aligned} J_s(\varphi) &= E^Q \left[ \lambda (Y_T(\varphi))^2 + (1 - \lambda) (C_T(\varphi) - C_s(\varphi))^2 \middle| \mathcal{G}_s \right] \\ &= E^Q \left[ \lambda (Y_T(\tilde{\varphi}))^2 + (1 - \lambda) (C_T(\tilde{\varphi}) - C_s(\tilde{\varphi}))^2 \middle| \mathcal{G}_s \right] \\ &\quad + (1 - \lambda) (E^Q [C_T(\tilde{\varphi}) | \mathcal{G}_s] - C_s(\varphi))^2 \\ &= J_s(\tilde{\varphi}) + (1 - \lambda) (E^Q [C_T(\tilde{\varphi}) | \mathcal{G}_s] - C_s(\varphi))^2. \end{aligned}$$

Since the essential infimum is achieved at  $\varphi = (\theta, \eta)$ , we can conclude that

$$E^Q [C_T(\varphi) | \mathcal{G}_s] = C_s(\varphi).$$

■

Because of the relation between  $\theta, \eta$  and  $C$  and the Kunita-Watanabe decomposition of  $C_T \in L^2(Q)$ ,

$$\begin{aligned} C_T &= E^Q[C_T] + \int_0^T \theta_s^C dU_s + \int_0^T \xi_s^C dU_s^\perp \\ &= C_t + \int_t^T \theta_s^C dU_s + \int_t^T \xi_s^C dU_s^\perp \end{aligned}$$

the dynamic hedging criterion for every  $t \in [0, T]$  can be equivalently defined as follows:

$$\text{ess inf } E^Q \left[ \lambda (Y_T)^2 + (1 - \lambda) (C_T - C_t)^2 \middle| \mathcal{G}_t \right]$$

over the admissible control variables

$$\begin{aligned} \theta &\in L^2(t, T; U), \\ \theta^C &\in L^2(t, T; U), \quad \xi^C \in L^2(t, T; U^\perp), \end{aligned}$$

and under the initial condition

$$C_0 = E^Q[C_T] = E^Q[V_T].$$

**Theorem 8** *The optimal hedging strategy can be uniquely determined as*

$$\begin{aligned} \theta_t &= \theta_t^V, \\ \eta_t &= C_t + \int_0^t \theta_s^V dU_s - \theta_t^V U_t - V_t, \end{aligned}$$

where the accumulated cost process at time  $t$  is

$$C_t = E^Q[V_T] + \lambda L_t^V,$$

for all  $0 \leq t \leq T$ .

The optimal remaining risk process at any time  $t$  is given by

$$\lambda(1 - \lambda)^2 E^Q \left[ (L_T^V)^2 \middle| \mathcal{G}_t \right] + \lambda^2(1 - \lambda) E^Q \left[ (L_T^V - L_t^V)^2 \middle| \mathcal{G}_t \right].$$

**Proof. Existence:** Let us recall first the decompositions

$$\begin{aligned} V_T &= E^Q[V_T | \mathcal{G}_t] + \int_t^T \theta_s^V dU_s + \int_t^T \xi_s^V dU_s^\perp, \\ C_T &= E^Q[C_T] + \int_0^T \theta_s^C dU_s + \int_0^T \xi_s^C dU_s^\perp. \end{aligned}$$

For some  $t \in [0, T]$ , we get by equation (2.2) and by plugging in the decompositions,

$$\begin{aligned}
& E^Q \left[ \lambda (Y_T)^2 + (1 - \lambda) (C_T - C_t)^2 \middle| \mathcal{G}_t \right] \\
= & E^Q \left[ \lambda \left( E^Q [C_T] + \int_0^t \theta_s^C dU_s + \int_0^t \xi_s^C dU_s^\perp + \int_t^T \theta_s^C dU_s + \int_t^T \xi_s^C dU_s^\perp + \int_0^T \theta_s dU_s \right. \right. \\
& \left. \left. - E^Q [V_T | \mathcal{G}_t] - \int_t^T \theta_s^V dU_s - \int_t^T \xi_s^V dU_s^\perp \right)^2 \right. \\
& \left. + (1 - \lambda) \left( E^Q [C_T] + \int_0^t \theta_s^C dU_s + \int_0^t \xi_s^C dU_s^\perp + \int_t^T \theta_s^C dU_s + \int_t^T \xi_s^C dU_s^\perp - C_t \right)^2 \middle| \mathcal{G}_t \right].
\end{aligned}$$

By splitting terms and noting that the corresponding mixed terms vanish this equals

$$\begin{aligned}
& \lambda \left( E^Q [C_T] + \int_0^t \theta_s^C dU_s + \int_0^t \xi_s^C dU_s^\perp + \int_0^t \theta_s dU_s - E^Q [V_T | \mathcal{G}_t] \right)^2 \quad (2.5) \\
& + (1 - \lambda) \left( E^Q [C_T] + \int_0^t \theta_s^C dU_s + \int_0^t \xi_s^C dU_s^\perp - C_t \right)^2 \\
& + E^Q \left[ \lambda \left( \int_t^T \theta_s^C dU_s + \int_t^T \theta_s dU_s - \int_t^T \theta_s^V dU_s \right)^2 + (1 - \lambda) \left( \int_t^T \theta_s^C dU_s \right)^2 \middle| \mathcal{G}_t \right] \\
& + E^Q \left[ \lambda \left( \int_t^T \xi_s^C dU_s^\perp - \int_t^T \xi_s^V dU_s^\perp \right)^2 + (1 - \lambda) \left( \int_t^T \xi_s^C dU_s^\perp \right)^2 \middle| \mathcal{G}_t \right].
\end{aligned}$$

Consider first the initial time point  $t = 0$ , for which the whole term can be made to vanish as follows. The first term is zero iff  $E^Q [C_T] = E^Q [V_T]$ . Setting the second term to zero yields  $C_0 = E^Q [C_T]$ . The third term equals zero if we choose  $\theta = \theta^V$  as well as  $\theta^C = 0$ . As for the fourth term, note that by the Itô-isometry as well as by the definition of the predictable compensator

$$\begin{aligned}
& E^Q \left[ \lambda \left( \int_0^T \xi_s^C dU_s^\perp - \int_0^T \xi_s^V dU_s^\perp \right)^2 + (1 - \lambda) \left( \int_0^T \xi_s^C dU_s^\perp \right)^2 \right] \\
= & E^Q \left[ \int_0^T \left( \xi_s^C \right)^2 - 2\lambda \xi_s^V \xi_s^C + \lambda \left( \xi_s^V \right)^2 d[U^\perp]_s \right] \\
= & E^Q \left[ \int_0^T \left( \xi_s^C \right)^2 - 2\lambda \xi_s^V \xi_s^C + \lambda \left( \xi_s^V \right)^2 d\langle U^\perp \rangle_s \right],
\end{aligned}$$

which equals zero for  $\xi^C = \lambda \xi^V$ .

For arbitrary  $t \in (0, T]$ , note that since  $E^Q [C_T]$  already has been determined in the step for  $t = 0$  that the first term in (2.5) need not be optimised since it depends only on strategies up to time  $t$ . The third and fourth term vanish again for the same choices as for  $t = 0$ , and the second term vanishes if the cost process  $C$  is a martingale.

With these choices, the first term equals

$$\begin{aligned} & \lambda \left( E^Q [V_T | \mathcal{G}_t] - E^Q [C_T] - \int_0^t \theta_s dU_s - \int_0^t \xi_s^C dU_s^\perp - \int_0^t \theta_s^C dU_s \right)^2 \\ &= \lambda \left( \int_0^t \theta_s^V dU_s + \int_0^t \lambda \xi_s^V dU_s^\perp \right)^2. \end{aligned}$$

In summary, the essential infimum of (2.5) is achieved for

$$\theta_s = \theta_s^V, \quad \theta_s^C = 0, \quad \xi_s^C = \lambda \xi_s^V, \quad (2.6)$$

for  $t \leq s \leq T$ , and

$$C_t = E^Q [C_T] + \int_0^t \theta_s^C dU_s + \int_0^t \xi_s^C dU_s^\perp.$$

The optimal amount  $\eta$  in the savings account is now uniquely determined by (2.1) and (2.2).

**Uniqueness:** Clearly once we determine  $\theta$  and  $C$ , the units of riskless asset  $\eta$  will be automatically uniquely determined. The uniquenesses of  $\theta$  and  $C$  are guaranteed by the above proof, so the uniqueness of the optimal strategy holds. ■

**Remark 9** When  $\lambda = 0, 1$ , the optimal strategy cannot be uniquely determined without extra conditions. However, the risk-minimization strategy and self-financing strategy are among the set of strategies obtained when  $\lambda = 0, 1$ .

## 2.4 An example of a hedging annuity portfolio

We will follow Dahl and Møller's (2007) [4] framework, but will consider different hedging instruments and apply modified quadratic hedging strategies in addition to the traditional quadratic hedging strategies. Let  $\mathbb{H}$  be the filtration generated by the counting process of death and  $\mathcal{F}$  be the filtration generated by the mortality intensity. Define  $\mathbb{G} = \mathbb{H} \vee \mathcal{F}$ , and assume that all  $\mathbb{F}$ -martingales remain martingales in the larger filtration  $\mathbb{G}$ . Further assume that the interest rate remains zero.

Consider two portfolios consisting of  $l_j$ ,  $j = 1, 2$  lives, all aged  $x$  years at time 0. Denote the initial mortality intensity as  $\mu_j^o(x)$ , which is a deterministic function of  $x$ . Assume the mortality intensity for the considered cohort at time  $t$  to be

$$\mu_j(x, t) = \mu_j^o(x + t) \xi_j(t).$$

where  $\xi(t)$  follows a CIR process. Put in a different way, we assume the mortality intensity changes in a stochastic way around the initial mortality intensity. Therefore, we will end up with a CIR-type mortality intensity process

$$d\mu_j(x, t) = (\gamma_j(x, t) - \delta_j(x, t)\mu_j(x, t)) dt + \sigma_j(x, t) \sqrt{\mu_j(x, t)} dW_j(t).$$

Define the remaining lifetimes as non-negative random variables  $T_{j,1}, \dots, T_{j,n}$ ,  $j = 1, 2$ . It follows that the survival probability of a single person, given the information on the mortality intensity as contained in  $\mathcal{F}_t$ , is given by

$$P(T_{j,i} > t | \mathcal{F}_t) = E \left[ e^{-\int_0^t \mu_j(x,s) ds} \middle| \mathcal{F}_t \right], \quad j = 1, 2.$$

The number of deaths at time  $t$  in each portfolio is given by

$$N_j(x, t) = \sum_{i=1}^{l_j} \mathbf{1}_{[T_{j,i} \leq t]}, \quad j = 1, 2.$$

The process  $\Lambda = \int \lambda_j(x, \cdot) ds$  is the compensator of  $N$  so that

$$N - \Lambda \quad \text{is a } \mathbb{G} - \text{martingale.}$$

Note that for  $j = 1, 2$ ,

$$\begin{aligned} \lambda_j(x, t) dt &= E[dN_j(x, t) | \mathcal{G}_t], \\ &= (n_j - N_j(x, t_-)) \mu_j(x, t) dt. \end{aligned}$$

### 2.4.1 Annuity portfolio

Consider a simple structured annuity contract: a premium is paid at time 0, and then the insurance company is going to pay a rate of  $a_t$  to the survivors in the portfolio 1 continuously at all  $t \in (0, T]$ . Here the payment at each time  $t$  is  $a_t(l_1 - N_1(x, t))$ .

Denote the intrinsic value process as  $\widehat{V}$ , which amounts to

$$\begin{aligned} \widehat{V}_t &= E^Q \left[ \int_0^T a_s(l_1 - N_1(x, s)) ds \middle| \mathcal{G}_t \right], \\ &= \int_0^t a_s(l_1 - N_1(x, s)) ds + \int_t^T a_s E^Q[l_1 - N_1(x, s) | \mathcal{G}_t] ds, \\ &= \int_0^t a_s(l_1 - N_1(x, s)) ds + (l_1 - N_1(x, t)) \int_t^T a_s Q(\tau_1 > s | \mathcal{G}_t) ds. \end{aligned}$$

$\widehat{V}$  admits a stochastic representation under a risk-neutral measure  $Q$  as follows:

$$\begin{aligned} d\widehat{V}_t &= - \int_t^T a_s Q(\tau_1 > s | \mathcal{G}_t) ds dM_1(x, t), \\ &\quad + (l_1 - N_1(x, t)) \int_t^T a_s \frac{\partial Q}{\partial \mu_1} ds \sigma_1^Q(x, t) \sqrt{\mu_1^Q(x, t)} dW_1(x, t). \end{aligned}$$

### 2.4.2 Mortality forward

Consider a  $T$ -year mortality forward as the hedging instrument. The payoff is linked to the death ratio  $N_2(t, T)/l_2$ . Then the value process  $U$  of the forward contract is

$$\begin{aligned} U_t &= E^Q \left[ k \left( \frac{N_2(x, T)}{l_2} - F \right) \middle| \mathcal{G}_t \right], \\ &= \frac{k}{l_2} E^Q [N_2(x, T) | \mathcal{G}_t] - kF, \\ &= k - \frac{k}{l_2} (l_2 - N_2(x, T)) Q(\tau_2 > T | \mathcal{G}_t) - kF. \end{aligned}$$

Here  $k$  is the nominal and  $F$  is the forward price, which is a constant determined at time 0.

The value process  $U$  also admits a stochastic representation under  $Q$ :

$$\begin{aligned} dU_t &= \frac{k}{l_2} Q(\tau_2 > T | \mathcal{G}_t) dM_2^Q(x, t), \\ &\quad - \frac{k}{l_2} (l_2 - N_2(x, T)) \frac{\partial Q}{\partial \mu} \sigma_2^Q(x, t) \sqrt{\mu_2^Q(x, t)} dW_2(x, t). \end{aligned}$$

### 2.4.3 Hedging strategies and remaining risks

Let

$$W_2^{Q, \perp}(t) = \frac{1}{\sqrt{1 - \rho_t^2}} W_1^Q(t) - \frac{\rho_t}{\sqrt{1 - \rho_t^2}} W_2^Q(t),$$

where  $\rho_t$  represents the correlation coefficient between the two Brownian motions at time  $t$ . Set

$$d\widehat{V}_t^Q = \theta_t dU_t^Q + dL_t,$$

where  $\xi_t$  is the number of the mortality forwards one should hold at each point. Moreover, in our case,

$$\theta_t = - \frac{\rho_t (l_1 - N_1(x, t)) \int_t^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, t) \sqrt{\mu_1^Q(x, t)}}{\frac{k}{l_2} (l_2 - N_2(x, T)) \frac{\partial P_2^Q}{\partial \mu} \sigma_2^Q(x, t) \sqrt{\mu_2^Q(x, t)}}.$$

The error term  $L_t$  is given as

$$\begin{aligned} L_t &= - \int_0^t \theta_u \frac{k}{l_2} P_2^Q(\tau > T | \mathcal{G}_u) dM_2^Q(x, u) - \int_0^t \int_u^T a_s P_1^Q(\tau > s | \mathcal{G}_u) ds dM_1^Q(x, u) \\ &\quad + \int_0^t \sqrt{1 - \rho_u^2} (l_1 - N_1(x, u)) \int_u^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, u) \sqrt{\mu_1^Q(x, u)} dW_1^Q(x, u). \end{aligned}$$

**Self-financing strategy**

The value of the money market account we should prepare is

$$\eta_t^{sf} = \int_0^t \theta_s dU_s^Q - V_t - \theta_t U_t^Q + E^Q[V_T].$$

The additional cost risk is by construction then zero.

The hedging error risk is given as follows:

$$\begin{aligned} & E^Q[L_T^2 | G_t] \\ = & \int_0^T E^Q \left[ \left( \theta_u \frac{k}{l_2} P_2^Q(\tau > T | G_u) \right)^2 (l_2 - N_2^Q(x, u)) \mu_2^Q(x, u) \middle| G_t \right] du, \\ & + \int_0^T E^Q \left[ \left( \int_u^T a_s P_1^Q(\tau > s | G_u) ds \right)^2 (l_1 - N_1^Q(x, u)) \mu_1^Q(x, u) \middle| G_t \right] du, \\ & + \int_0^T E^Q \left[ \left( \sqrt{1 - \rho_u^2} (l_1 - N_1(x, u)) \int_u^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, u) \sqrt{\mu_1^Q(x, u)} \right)^2 \middle| G_t \right] du. \end{aligned}$$

**Risk-minimization strategy**

The value of the money market account we should prepare is

$$\eta_t^{rm} = \widehat{V}_t^Q - V_t - \xi_t U_t^Q.$$

Additional cost risk:

$$\begin{aligned} & E^Q[(L_T - L_t)^2 | G_t] \\ = & \int_t^T E^Q \left[ \left( \theta_u \frac{k}{l_2} P_2^Q(\tau > T | G_u) \right)^2 (l_2 - N_2^Q(x, u)) \mu_2^Q(x, u) \middle| G_t \right] du, \\ & + \int_t^T E^Q \left[ \left( \int_u^T a_s P_1^Q(\tau > s | G_u) ds \right)^2 (l_1 - N_1^Q(x, u)) \mu_1^Q(x, u) \middle| G_t \right] du, \\ & + \int_0^T E^Q \left[ \left( \sqrt{1 - \rho_u^2} (l_1 - N_1(x, u)) \int_u^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, u) \sqrt{\mu_1^Q(x, u)} \right)^2 \middle| G_t \right] du. \end{aligned}$$

By construction, the hedging error risk is zero.

**Generalized quadratic hedging strategies**

The value of the money market account we should prepare is

$$\eta_t^\lambda = \lambda \eta_t^{rm} + (1 - \lambda) \eta_t^{sf}.$$

Additional cost risk:

$$\begin{aligned}
& \lambda^2(1-\lambda)E^Q[(L_T - L_t)^2|G_t] \\
= & \lambda^2(1-\lambda) \int_t^T E^Q \left[ \left( \theta_u \frac{k}{l_2} P_2^Q(\tau > T | G_u) \right)^2 (l_2 - N_2^Q(x, u)) \mu_2^Q(x, u) \middle| G_t \right] du, \\
& + \lambda^2(1-\lambda) \int_t^T E^Q \left[ \left( \int_u^T a_s P_1^Q((\tau > s | G_u) ds \right)^2 (l_1 - N_1^Q(x, u)) \mu_1^Q(x, u) \middle| G_t \right] du, \\
& + \lambda^2(1-\lambda) \int_0^T E^Q \left[ \left( \sqrt{1 - \rho_u^2} (l_1 - N_1(x, u)) \int_u^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, u) \sqrt{\mu_1^Q(x, u)} \right)^2 \middle| G_t \right] du.
\end{aligned}$$

Hedging error risk:

$$\begin{aligned}
& \lambda(1-\lambda)^2 E^Q[L_T^2|G_t] \\
= & \lambda(1-\lambda)^2 \int_t^T E^Q \left[ \left( \theta_u \frac{k}{l_2} P_2^Q(\tau > T | G_u) \right)^2 (l_2 - N_2^Q(x, u)) \mu_2^Q(x, u) \middle| G_t \right] du, \\
& + \lambda(1-\lambda)^2 \int_t^T E^Q \left[ \left( \int_u^T a_s P_1^Q((\tau > s | G_u) ds \right)^2 (l_1 - N_1^Q(x, u)) \mu_1^Q(x, u) \middle| G_t \right] du, \\
& + \lambda(1-\lambda)^2 \int_0^T E^Q \left[ \left( \sqrt{1 - \rho_u^2} (l_1 - N_1(x, u)) \int_u^T a_s \frac{\partial P_1^Q}{\partial \mu_1} ds \sigma_1^Q(x, u) \sqrt{\mu_1^Q(x, u)} \right)^2 \middle| G_t \right] du.
\end{aligned}$$

Finally, the aggregate risk is given as

$$\lambda^2(1-\lambda)E^Q[(L_T - L_t)^2|G_t] + \lambda(1-\lambda)^2 E^Q[L_T^2|G_t].$$

## 2.4.4 Numerical analysis

### Mortality data

According to Coughlan et al. (2007) [3], there are two kinds of mortality rates for the  $x$ -year-old: initial rate of mortality  $q_x$  and central rate of mortality  $m_x$ . The initial rate of mortality represents the probability of deaths within one year, defined as the following ratio

$$q_x = \frac{\# \text{ of death over the year}}{\# \text{ of lives at the start of the year}},$$

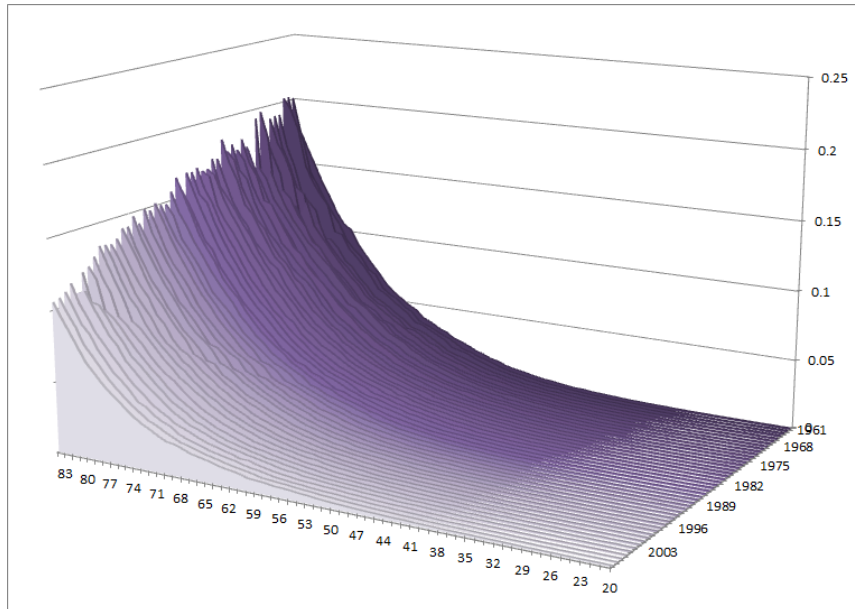
The central rate of mortality reflects deaths per unit of exposure, by replacing the denominator of the above fraction with the number of lives at the middle of the year, a proxy for the exposure-to-risk.

$$m_x = \frac{\# \text{ of death over the year}}{\# \text{ of lives at the middle of the year}},$$

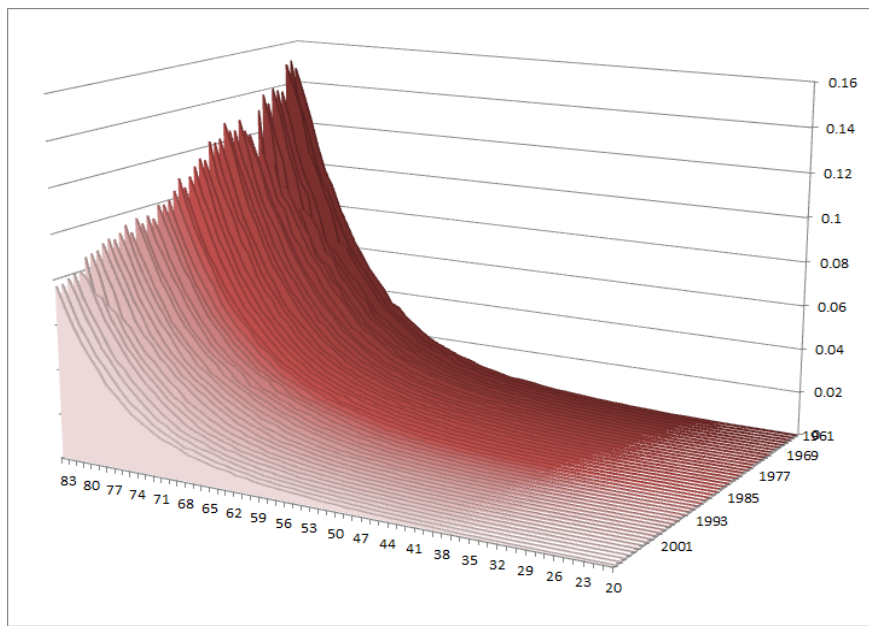
The connection between the two mortality rates is

$$q_x \approx \frac{m_x}{1 + 0.5 * m_x}.$$

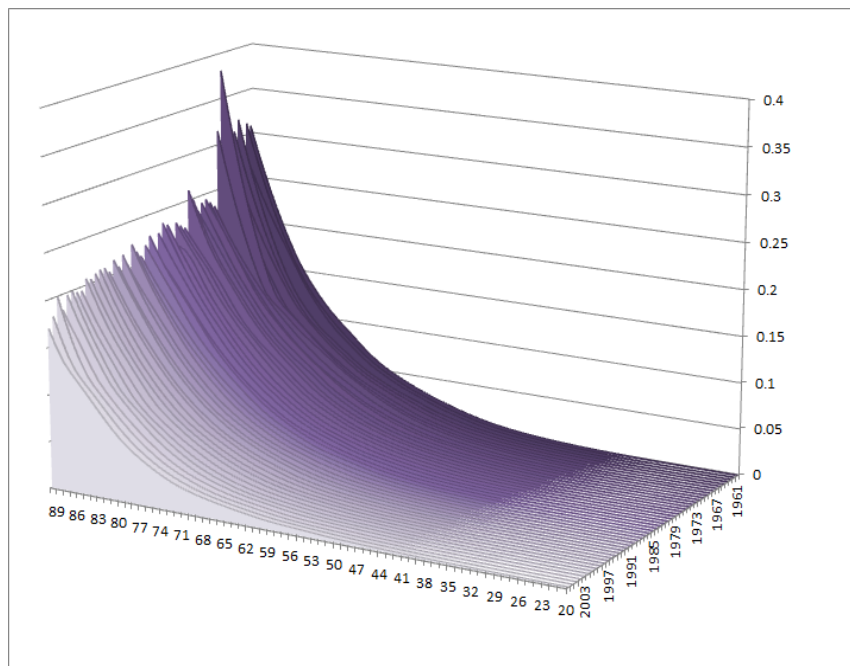
We collect both mortality rates from *www.lifemetrics.com*. The data contains the mortality rates of English and Welsh males and females between age 20 and age 89, from 1961 to 2007. The initial rate of mortality has been 'graduated' in order to eliminate the noise. The smoothing techniques can be found in Coughlan et al. (2007). The following figures summarize the data and show the evolvement of the mortality rate at different ages over years. We can see from the figures that females have a much lower mortality rate than males. And the mortality rates for senior citizens have decreased significantly in recent decades.



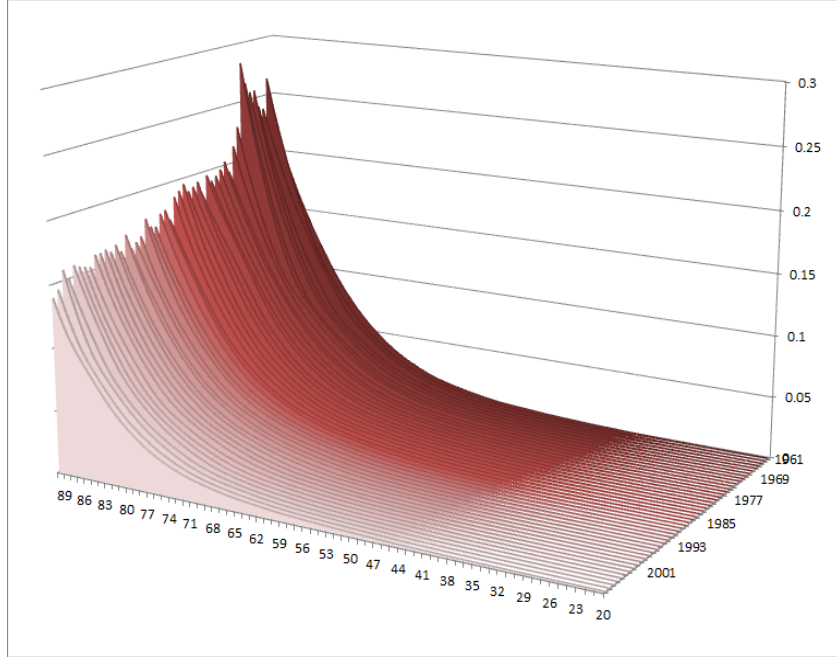
Crude central rate of mortality (England and Welsh male)



Crude central rate of mortality (England and Welsh female)



Graduate initial rate of mortality (England and Welsh, male)



Graduate initial rate of mortality (England and Welsh, female)

### Mortality intensity

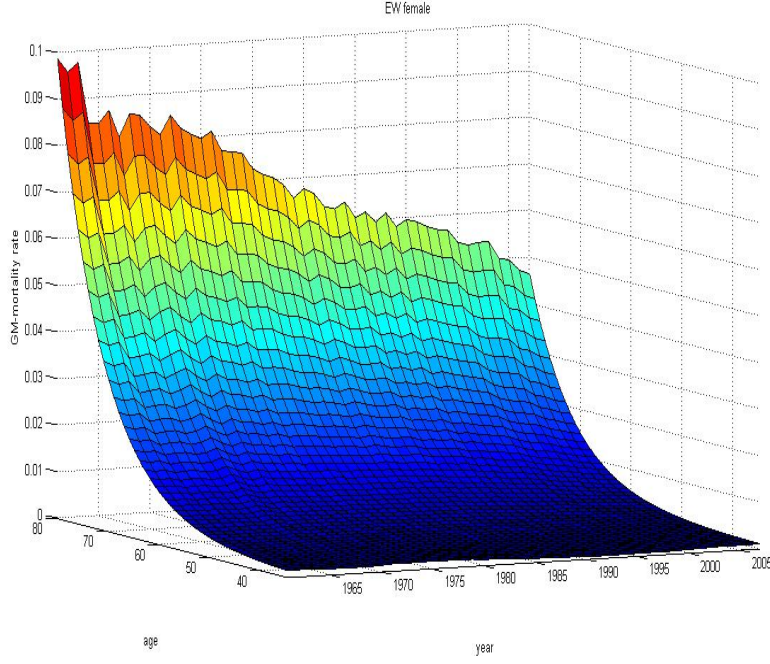
Inspired by Dahl, Melchior and Møller (2008) [5], we assume the initial mortality intensity to take the Gompertz-Makeham form. That is to say, the force of mortality can be expressed as

$$\mu^o(x+t) = a + b c^{x+t},$$

where  $a$ ,  $b$  and  $c$  are positive constants. We focus on this age group between 30 and 80-year old. The next task is to fit the Gompertz-Makeham mortality curve to the data. At this stage, we assume the mortality rate remains constant within a year, which means that the central mortality rate equals the mortality intensity (See Cairns et al. (2007) [2] for details). Therefore, the data considered here are the observed central mortality rates of all the ages at one year. Here we adopt a least square estimation method. Denote the observations as  $y_i$ , where  $i$  represents the age. The parameters of the model,  $a, b, c$ , are estimated by minimizing the objective function

$$\sum_i \left( y_i - \left( a + b(c)^i \right) \right)^2, \quad i = 29.5, 30.5, \dots, 79.5.$$

Here 0.5 has been subtracted from the age since crude mortality data have been used. The following figures show the fitted Gompertz-Makeham mortality curves for English and Welsh females between years 1961 and 2007.



However, a deterministic mortality intensity curve is not enough. Following the idea proposed by Dahl, Melchior and Møller (2008), we multiply the initial deterministic mortality curve by a CIR process to capture the stochastic evolvement of the mortality curve over time. Consider the following CIR noise

$$d\xi_t = (\delta \exp(-\gamma t) - \delta \xi_t) dt + \sigma \sqrt{\xi_t} dW(t).$$

It can be easily shown that the mortality intensity  $\mu(x, t) = \mu^o(x + t)\xi(t)$  is again a CIR process.

The trick here is to leverage the probability to survive from time  $t$  to  $T$ , which resembles the zero-coupon bond price in terms of its mathematical form

$$P(T_{j,i} > T \mid \mathcal{F}_t) = E \left[ e^{-\int_t^T \mu_j(x,s) ds} \mid \mathcal{F}_t \right] = \exp \{ A(T-t) - B(T-t) \mu_j(x, t) \}.$$

where  $A(T-t)$  and  $B(T-t)$  are deterministic functions, containing the parameters  $\gamma, \delta, \sigma$  (See Dahl, Melchior and Møller (2008) for the forms of these two functions). Please notice that the central rate of mortality is an analogue of the yield to maturity of a one-year zero-coupon bond.

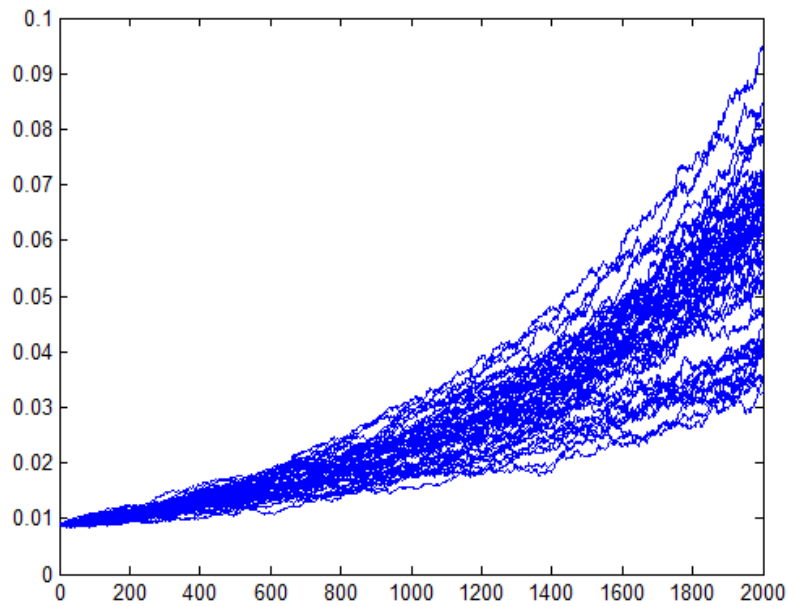
The parameters are estimated by minimizing the difference between the observed central mortality rate  $y$  and the theoretical central mortality rate  $-A(T-t) + B(T-t)(T-t)$

$$\min_{\gamma, \delta, \sigma} \sum (y + A(T-t; \gamma, \delta, \sigma) - B(T-t; \gamma, \delta, \sigma) \mu_j(x, t))^2.$$

**Remark 10** *The mortality intensity and the survival probability obtained in this way are the ones under the physical measure  $P$ . For pricing purposes, one should either change the measure to  $Q$  or estimate directly under the risk-neutral measure  $Q$ . The latter method would involve market price data of some life-insurance contracts or derivatives.*

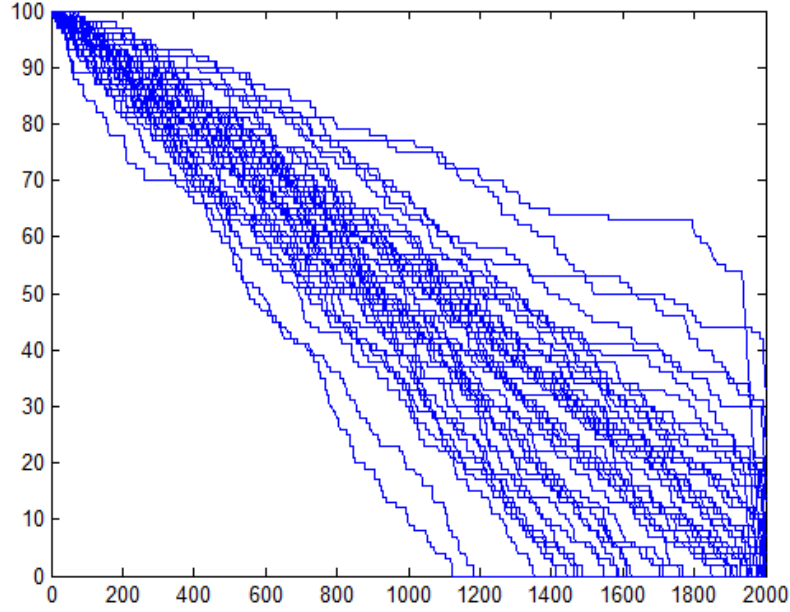
The following graph shows the stochastic intensity process over 20 years for England and Welsh females aged 60 at time 0. The estimated parameters are

| $a$      | $b$      | $c$      | $\gamma$ | $\delta$ | $\sigma$ |
|----------|----------|----------|----------|----------|----------|
| 0.000729 | 0.000014 | 1.112223 | 0.3316   | 0.0138   | 0.0552   |



Stochastic mortality intensity (50 paths, 100 steps per year)

For a pool of 100 60-year old females ( $l = 100$ ), the number of people surviving over the next 20 years based on the proposed mortality intensity are shown by the following figure.



Survivors counting process (50 paths, 100 steps per year)

### Annuity portfolio and hedging

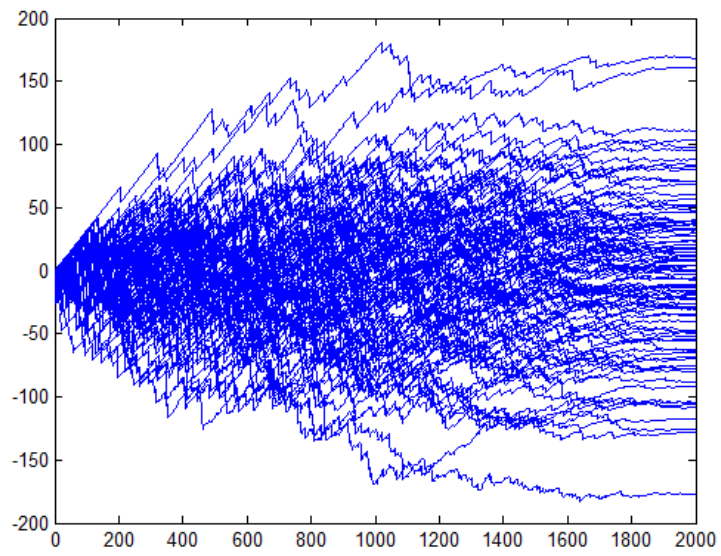
In this section, we show the simulated value of the annuity portfolio and mortality forwards and some examples of hedging. For pricing and hedging purposes, we adopt the following parameters for the mortality intensity process of the annuity portfolio

$$\begin{array}{cccccc} a_1 & b_1 & c_1 & \gamma_1 & \delta_1 & \sigma_1 \\ -0.0013 & 0.0001 & 1.0956 & 0.3316 & 0.0138 & 0.0552 \end{array}$$

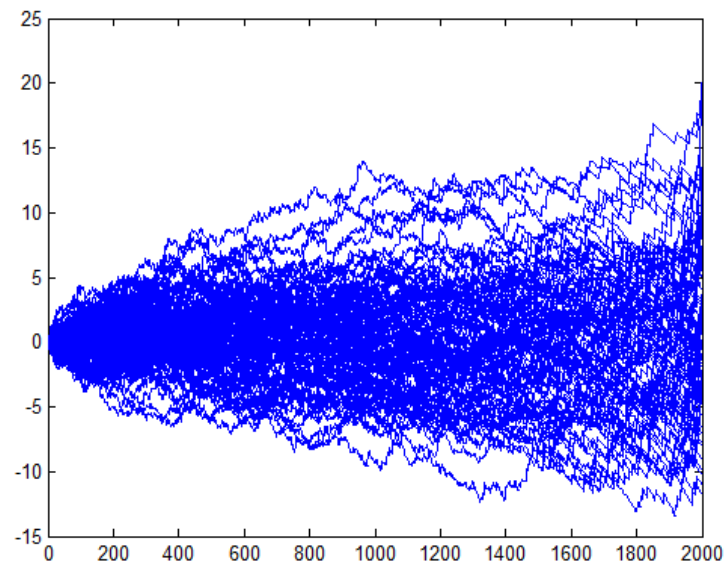
and the following parameters for the mortality forward

$$\begin{array}{cccccc} a_2 & b_2 & c_2 & \gamma_2 & \delta_2 & \sigma_2 \\ -0.0013 & 0.0001 & 1.0956 & 0.1230 & 0.0207 & 0.0354 \end{array}$$

Further assume that the correlation coefficient is  $\rho = 0.85$ . The constant payment to the survivors remains 1. The following figures show the simulated paths of the values of annuity portfolios and the mortality forwards.

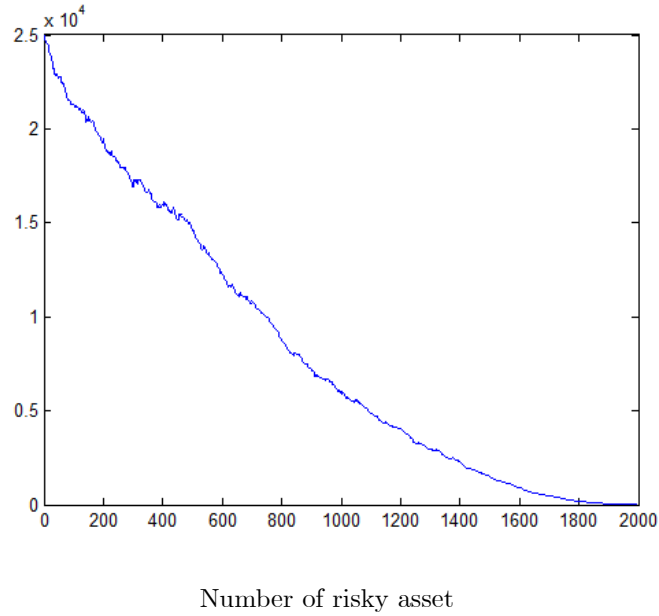


Annuity portfolio (100 paths, 100 steps per year)



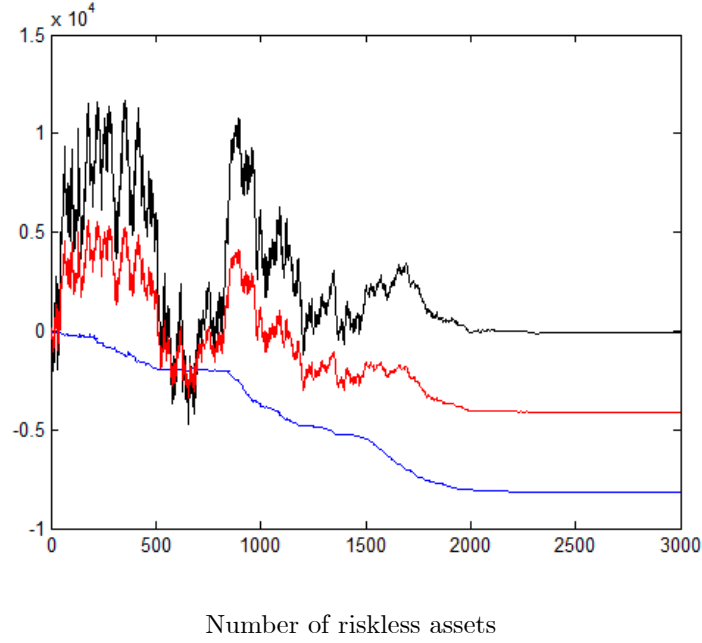
Mortality forward (100 paths, 100 steps per year)

Right now let us look at a hedging example. Assume  $\lambda = 0.5$ . The following graphs give the units of the risky assets and riskless assets which should be used for the hedging portfolio. All the three quadratic hedging strategies, self-financing, risk-minimization and generalized hedging, follow the same strategy for risky asset.



However, the strategy for riskless asset are different. In the following figures, the blue indicates the units of riskless asset used in self-financing strategy, the black one is for risk-minimization strategy and the red one is for the generalized quadratic strategy

with  $\lambda = 0.5$ .



## 2.5 Conclusion

In this study, we design a new set of quadratic hedging strategies, which can balance the short-term risk (additional cost risk) and the long-term risk (hedging error). The unique optimal hedging strategy under the risk-neutral measure can be obtained via a Kunita-Watanabe decomposition when  $\lambda = (0, 1)$ . Though the model is more complicated, the optimal trading strategy for the risky asset is the same with the ones in the risk-minimization strategy and the self-financing strategy, however the trading strategy for the riskless asset is very different.

In the empirical and numerical analysis part, we adopt mortality forwards to hedge the longevity risk in an annuity portfolio. We apply the generalized quadratic hedging strategy with  $\lambda = 0.5$  and compare with the two traditional hedging strategies.



# Bibliography

- [1] P. Barrieu, L. Albertini (2009). The Handbook of Insurance-Linked Securities. Wiley, Chichester.
- [2] Cairns, A.J.G., Blake, D., Dowd, K. (2008) Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal* 2008(2-3): 79-113.
- [3] G. Coughlan, D. Epstein, A. Ong, A. Sinha, J. Hevia-Portocarrero, E. Gingrich, M. Khalaf-Allah, P. Joseph (2007). LifeMetrics: A toolkit for measuring and managing longevity and mortality risks. [www.lifemetrics.com](http://www.lifemetrics.com)
- [4] M. Dahl, T. Møller (2007). Valuation and Hedging of Life Insurance Liabilities with Systematic Mortality Risk, *Insurance. Mathematics and Economics* 39, 193-217.
- [5] M. Dahl, M. Melchior and T. Møller (2008). On systematic mortality risk and risk-minimization with survivor swaps. *Scandinavian Actuarial Journal*. Vol 2008, Issue 2-3.
- [6] M. Ehrgott (2005). Multicriteria optimization. Birkhäuser.
- [7] H. Föllmer, D. Sondermann (1986). Hedging of Non-redundant Contingent Claims. *Contributions to Mathematical Economics*. North-Holland, 205-223.
- [8] T. Møller (2001). Risk-minimizing Hedging Strategies for Insurance Payment Processes. *Finance and Stochastics*. Vol 5, No. 4. 419-446.
- [9] M. Schweizer (2001). A Guided Tour through Quadratic Hedging Approaches. *Option Pricing, Interest Rates and Risk Management*. Cambridge University Press (2001), 538–574.
- [10] M. Schweizer (2008). Local Risk-Minimization for Multidimensional Assets and Payment Streams. *Banach Center Publications* 83, 213-229.



## Chapter 3

# Risk-Neutral Densities and Metal Futures

### 3.1 Introduction

Harrison and Kreps (1979) [10] and Harrison and Pliska (1981) [11] established the link between no-arbitrage pricing and martingale theory. The *first fundamental asset pricing theorem* states that the absence of arbitrage in the market is equivalent to existence of an *equivalent martingale measure*  $Q$  for the price process of the underlying financial asset. The *second fundamental asset pricing theorem* states that the completeness of a market is equivalent to *uniqueness* of the equivalent martingale measure. (see Kiesel (2002) [14] for more details.) In reality, the markets are seldom complete due to undiversifiable and unhedgeable risks. As a result, there could exist multiple pricing measures that can rule out arbitrage opportunities. This comparative study considers three pricing measures obtained by different approaches and applies them to different commodity future markets.

It is important to identify the difference between forward and future prices when the interest rate is stochastic. At time  $t$ , denote the prices of forward and future contracts expiring at time  $T$  to be  $For_S(t, T)$  and  $Fut_S(t, T)$ , respectively. Let  $S(t)$  be the spot price at time  $t$  and  $B(t, T)$  be the price of a zero-coupon bond paying 1 at time  $T$ .

**Definition 11** (Shreve(2000) [19]) *Assume that zero-coupon bonds of all maturities can be traded. Then the forward price is determined by*

$$For_S(t, T) = \frac{S(t)}{B(t, T)}, \quad 0 \leq t \leq T,$$

**Definition 12** (Shreve(2000) [19]) *The futures price of an asset is given by the formula*

$$Fut_S(t, T) = E^Q [S(T) \mid \mathcal{F}_t].$$

*A long position in the futures contract is an agreement to receive as a cash flow the changes in the futures price (which may be negative as well as positive) during the time the position is held. A short position in the futures contract received the opposite cash flow.*

The difference between forward and future prices is the so-called forward-futures spread.

Commodity futures have been intensively studied in the literature. Schwartz (1997) [18] compared three models and found out that the dynamics of futures prices can be captured by three factors: driving spot prices, interest rates and convenience yields. The *convenience yield*, according to Carmona and Ludkovshi (2005) [4], is defined as the difference between benefit of direct access and cost of storage. Later on, a lot of new models for commodity spot prices and future prices have been proposed and empirically studied. (Ross (1997) [17], Casassus and Collin-Dufresne (2002) [5], Carmona and Ludkovski (2005), etc)

The focus of this study is not on the dynamics of commodity spot or future prices, but on the link between spot and future prices. The essay considers three measure-changing approaches. We aim to find out the best measure-changing technique in terms of empirical performance. We also would very much like to quantify the performance of the frequently studied measure-changing techniques like Esscher transform compared to calibration method and the non-parametric method. This is important because techniques like Esscher transform do not require the availability of derivatives prices and thus have often been adopted to study illiquid, not yet mature markets, like the insurance derivatives market discussed in the previous chapter.

The article is structured as follows. The first section describes the three measure-changing techniques. The second section empirically analyzes the performance of the methods on three commodity markets, gold, copper, and aluminum. The last section concludes the results.

## 3.2 Method1: price commodity futures under the Esscher measure

### 3.2.1 Dynamics for the commodity spot prices

We work on a probability space  $(\Omega, \mathcal{F}, P)$ . In this section, we will adopt the following SDE to describe the dynamics of the commodity spot prices.

$$S_t = S_0 \exp(G_t),$$

where

$$\begin{aligned} dG_t &= (\mu^G - \lambda^G G_{t-}) dt + \sigma_t dL_t, \\ d\sigma_t^2 &= (\kappa - \eta \sigma_{t-}^2) dt + \phi \sigma_{t-}^2 d[L, L]_t^d. \end{aligned}$$

The process  $G$  follows a COGARCH process, as proposed by Klüppelberg, Lindner and Maller (2004) [15]. We denote the augmented filtration generated by  $G$  and  $\sigma^2$  with  $\mathbb{F}$ .

$L$  is a Lévy process with differential triplet  $(b, c, F(dx))$ . For simplicity, we assume the  $L$  process to be a  $NIG$  process. As we will see later, this process provides a good fit to the commodity data. With parameters  $\alpha, \beta, \mu, \delta$ , a  $NIG$  random variable has the density function. Here  $K_1(\cdot)$  is the modified Bessel function of the third kind with index 1.

$$\begin{aligned} g_{NIG(\alpha, \beta, \mu, \delta)}(z) &= \frac{\alpha \exp(\varsigma + \beta(z - \mu)) K_1(\alpha \delta q((z - \mu)/\delta))}{\pi q((z - \mu)/\delta)}, \\ \varsigma &= \delta \sqrt{\alpha^2 - \beta^2}, \quad q(x) = \sqrt{1 + x^2}, \\ 0 &< |\beta| < \alpha, \quad -\infty < \mu < \infty, \quad \delta > 0. \end{aligned}$$

For a COGARCH model, we always require that  $L$  has mean zero and unit variance. To this purpose, we reparameterize  $NIG(\alpha, \beta, \mu, \delta)$  to  $NIG(\xi, \rho, \kappa_1, \kappa_2)$ , where  $\kappa_1$  represents the mean value and  $\kappa_2$  represents the variance.

$$\begin{aligned} \xi &= (1 + \varsigma)^{-1/2}, & \rho &= \beta/\alpha, \\ \kappa_1 &= \mu + \delta \rho / \sqrt{1 - \rho^2} = 0, & \kappa_2 &= \delta^2 / (\varsigma(1 - \rho^2)) = 1. \end{aligned}$$

The  $NIG$  process is a pure jump process. Therefore its differential triplet reduces to  $(b, 0, F(dx))$ , where the drift term  $b$  and Lévy measure  $F(dx)$  are given by

$$\begin{aligned} b &= \mu + \frac{2\delta\alpha}{\pi} \int_0^1 \sinh(\beta x) K_1(\alpha x) dx \\ F(dx) &= \frac{\delta\alpha}{\pi |x|} e^{\beta x} K_1(\alpha |x|) dx \end{aligned}$$

**Proposition 13** Assume  $(\sigma_t^2)_{t \geq 0}$  is the stationary version of the process with  $\sigma_0^2 = \sigma_\infty^2$ . If the underlying asset price  $S$  follows a COGARCH model then the future price reduces to

$$Fut_S(t, T) = S(t) \varphi_G^Q(t, T; 1)$$

where  $\varphi_G^Q$  is the Laplace transform of  $G$  with parameter 1 and  $Q$  is a risk-neutral measure.

**Proof.** We have

$$\begin{aligned} Fut_S(t, T) &= E^Q[S(T) \mid \mathcal{F}_t] \\ &= E^Q[S_0 \exp(G_T - G_t + G_t) \mid \mathcal{F}_t] \\ &= S(t) E^Q[\exp(G_T - G_t) \mid \mathcal{F}_t] \\ &= S(t) E^Q[\exp(G_{T-t})] \\ &= S(t) \varphi_G^Q(t, T; 1). \end{aligned}$$

Here the fourth equal sign holds because of the stationary increment property of  $G$ .

■

### 3.2.2 Risk-neutral distribution estimated from spot price data

For two probability measures  $Q$  and  $P$  defined on a measurable space  $(\Omega, \mathcal{F})$ ,  $Q$  is said to be *absolutely continuous* with respect to  $P$  if all  $P$ -zero sets are also  $Q$ -zero sets, denote as  $Q \ll P$ .  $P$  and  $Q$  are considered to be *equivalent* if  $Q \ll P$  and  $P \ll Q$ . If  $Q \ll P$ , there exists a unique density  $Z = dQ/dP$  so that for  $f \in L^1(Q)$  the following equation holds

$$E_Q[f] = E_P[Zf].$$

One can associate a martingale with respect to  $Z$ ,

$$Z_t = E_P[Z | \mathcal{F}_t].$$

This martingale is the *density process* of  $Q$ .

For the following definitions and notation we refer to Jacod and Shiryaev(2003) [12]. Assume that  $X$  is a semimartingale and there exists  $\theta \in L(X)$  such that  $\theta \cdot X$  is exponentially special, where  $\cdot$  denotes stochastic integration. The *Laplace cumulant*  $\tilde{K}^X(\theta)$  is defined as the compensator of the special semimartingale  $\mathcal{L}og(e^{\theta \cdot X})$ . We have

$$\tilde{K}^X(\theta) = \tilde{\kappa}(\theta) \cdot A,$$

where

$$\tilde{\kappa}(\theta)_t = \theta_t b_t + \frac{1}{2} \theta_t^2 c_t + \int (e^{\theta_t x} - 1 - \theta_t h(x)) F_t(dx).$$

Here  $b, c, F$  are the differential characteristics of  $X$ .  $h : \mathbb{R} \rightarrow \mathbb{R}$  is a truncation function, which are bounded and satisfy  $h(x) = x$  in a neighbourhood of 0.

The *modified Laplace cumulant*  $K^X(\theta)$  is defined as

$$K^X(\theta) = \log \mathcal{E}(\tilde{K}^X(\theta)),$$

If  $X$  is quasi-left continuous then  $K^X(\theta) = \tilde{K}^X(\theta)$  and  $K^X(\theta)$  is continuous. A càdlàg process  $X$  is defined as *quasi-left continuous* if  $\Delta X_T = 0$  a.s. on the set  $\{T < \infty\}$  for every predictable time  $T$ . In our case,  $G$  is quasi-left continuous because  $L$  is a NIG process.

In an incomplete market, which is the case when we adopt Lévy-driven asset process, we could face plenty of different risk-neutral measures with different density process. In this study, we focus on the widely used *Esscher measure* [12] with the density process

$$Z_t = \exp \left( \int_0^t \theta_s dX_s - K^X(0, t; \theta) \right).$$

### 3.2.3 Futures price under the Esscher measure

Our next goal is to find the process  $\theta$  making the discounted spot price a martingale under the Esscher measure  $Q$ . Because we are dealing with commodities, we have to define the discounted spot price in a different way.

Commodities, unlike stocks or bonds, can be consumed or stored with some costs. The difference between the consumption value and storage expenses per unit of time is termed the convenience yield  $\delta$ . Intuitively, the convenience yield corresponds to a dividend yield for stocks (Carmona and Ludkovski (1991)). Therefore, the discount rate should become to

$$r_t - \delta_t,$$

and the discounted spot price changes to

$$\tilde{S}_t = \exp \left( - \int_0^t (r_s - \delta_s) ds \right) S_t.$$

It remains to specify a dynamic for  $\delta_t$ . In this study, we adopt a convenience yield proportional to  $G$ , i.e.

$$\delta_t = \lambda^G G_{t-}$$

for some constant  $\lambda^G$ . With these specifications, the Esscher measure, and therefore the risk-neutral spot price, can be determined.

**Theorem 14** *The discounted spot price is a  $Q$ -(local) martingale where  $\theta_t$  satisfies*

$$\theta_t \sigma_{t-}^2 c^2 - r_t + \mu^G + \sigma_{t-} b + \frac{1}{2} \sigma_{t-}^2 c^2 + \int \left( e^{(\theta_t + 1)\sigma_{s-y}} - e^{\theta_t \sigma_{s-y}} - \sigma_{t-} h(y) \right) F(dy) = 0.$$

Here  $(b, c^2, F)$  is the Lévy triplet of  $L$ .

**Proof.** The result can be proved by two different methods.

The first approach rests on Girsanov's theorem. The density process admits the representation

$$dZ_t = Z_t \theta_t \sigma_t c dW_t + Z_t \int (e^{\theta_t \sigma_{s-x}} - 1) (\mu - \nu)(ds, dt).$$

On the other hand, the discounted spot price process under  $P$  admits a stochastic representation as

$$\begin{aligned} & d \left( e^{-\int_0^t (r_u - \lambda^G G_{u-}) du} S_t \right) \\ &= \tilde{S}_t \sigma_t c dW_t + \tilde{S}_t \int (e^{\sigma_t x} - 1) (\mu - \nu)(ds, dt) \\ & \quad + \tilde{S}_t \left( -r_t + \lambda^G G_{t-} + \sigma_t b + \frac{1}{2} \sigma_t^2 c^2 + \int (e^{\sigma_t x} - 1 - \sigma_t h(x)) F(dx) \right) dt. \end{aligned}$$

Therefore the predictable finite variation part in the decomposition of the discounted

spot price process as a special semimartingale is

$$dA_t = \tilde{S}_t \left( -r_t + \lambda^G G_{t-} + \sigma_t b + \frac{1}{2} \sigma_t^2 c^2 + \int (e^{\sigma_t x} - 1 - \sigma_t h(x)) F(dx) \right) dt.$$

Whereas the predictable finite variation part of the discounted spot price process under  $Q$  takes the form

$$\begin{aligned} dA + \frac{1}{Z_-} d\langle Z, \tilde{S} \rangle \\ = \tilde{S}_t \left( \theta_t \sigma_t^2 c^2 + \int \left( e^{(\theta_t+1)\sigma_s x} - e^{\sigma_t x} - e^{\theta_t \sigma_s x} + 1 \right) F(dx) \right. \\ \left. - r_t + \lambda^G G_{t-} + \sigma_t b + \frac{1}{2} \sigma_t^2 c^2 \right). \end{aligned}$$

This part must be sent to zero in order to guarantee that the discounted spot price process is a martingale under  $Q$ . In this way we end up with a Feynman–Kac type equation

$$\theta_t \sigma_t^2 c^2 - r_t + \lambda^G G_{t-} + \sigma_t b + \frac{1}{2} \sigma_t^2 c^2 + \int \left( e^{(\theta_t+1)\sigma_s x} - e^{\theta_t \sigma_s x} - \sigma_t h(x) \right) F(dx) = 0,$$

which proves the result.

In the alternative method, we first determine the differential characteristics of  $\tilde{G} := -r_t + \lambda^G G_{t-} + G_t$ . Recall that in our case, the discounted spot price can be written as

$$S_t = S_0 \exp \left( \tilde{G}_t \right).$$

Let us define a new process

$$\tilde{G}(h) = \tilde{G} - \tilde{G}'(h),$$

where

$$\tilde{G}'(h)_t = \sum_{s \leq t} \left[ \Delta \tilde{G}_s - h(\Delta \tilde{G}_s) \right].$$

$\tilde{G}(h)_t$  admits the representation

$$\begin{aligned} d\tilde{G}(h)_t &= d\tilde{G}_t - \int (\sigma_{t-y} - h(\sigma_{t-y})) \mu(dy, dt) \\ &= -r_t dt + \lambda^G G_{t-} dt + dG_t - \int (\sigma_{t-y} - h(\sigma_{t-y})) \mu(dy, dt). \end{aligned}$$

Therefore the differential characteristics of  $\tilde{G}$  are

$$\begin{aligned} b_t^{\tilde{G}} &= -r_t + \lambda^G G_{t-} + \mu^G - \lambda^G G_{t-} + \sigma_{t-} b + \int (h(\sigma_{t-y}) - \sigma_{t-} h(y)) \mu(dy, dt) \\ c_t^{\tilde{G}} &= \sigma_{t-} c \\ F^{\tilde{G}}(A)_t &= \int 1_A(\sigma_{t-y}) F(dy). \end{aligned}$$

According to Jacod and Shiryaev (2003) (page 224, theorem 7.18), the process  $e^{\tilde{G}}$  is a  $Q$ -local martingale if and only if  $\theta \cdot \tilde{G}$  is exponentially special and if we have

$$K^{\tilde{G}}(\theta + 1) = K^{\tilde{G}}(\theta).$$

Therefore, by the definition of the Laplace cumulant and modified Laplace cumulant, the equation in the theorem holds. ■

The above equation for  $\theta$  can be explicitly solved when the driving Lévy process is a *NIG* process.

**Corollary 15** *If  $L$  follows  $NIG(\alpha, \beta, \mu, \delta)$ , the explicit solutions of the equation for  $\theta$  are*

$$\begin{aligned} & \frac{1}{2} \frac{1}{(R_t^2 + \sigma_{t-}^2 \delta^2) \sigma_{t-} \delta} \left( -\delta^3 \sigma_{t-}^3 - 2\delta^3 \beta \sigma_{t-}^2 - \delta \sigma_{t-} R_t^2 - 2\delta \beta R_t^2 \right. \\ & \quad \left. + \sqrt{-\delta^4 \sigma_{t-}^4 R_t^2 - 2\delta^2 \sigma_{t-}^2 R_t^4 + 4R_t^4 \delta^2 \alpha^2 - R_t^6 + 4\sigma_{t-}^2 \delta^4 R_t^2 \alpha^2} \right), \\ & \text{and} \\ & -\frac{1}{2} \frac{1}{(R_t^2 + \sigma_{t-}^2 \delta^2) \sigma_{t-} \delta} \left( \delta^3 \sigma_{t-}^3 + 2\delta^3 \beta \sigma_{t-}^2 + \delta \sigma_{t-} R_t^2 + 2\delta \beta R_t^2 \right. \\ & \quad \left. + \sqrt{-\delta^4 \sigma_{t-}^4 R_t^2 - 2\delta^2 \sigma_{t-}^2 R_t^4 + 4R_t^4 \delta^2 \alpha^2 - R_t^6 + 4\sigma_{t-}^2 \delta^4 R_t^2 \alpha^2} \right), \end{aligned}$$

subject to an integrability condition. Here

$$R_t = -r_t + \mu^G + \mu \sigma_{t-}.$$

**Proof.** The cumulant function of  $L$  is

$$\mu u + \delta \left( \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2} \right),$$

where  $u \in \mathbb{R}$ . In this case, if we set  $u = \theta_t \sigma_{t-}$ , then

$$K_t^{\tilde{G}}(\tilde{\theta}) = \int_0^t \left( -r_s \theta_s + \mu^G \theta_s + \mu \theta_s \sigma_{s-} + \delta \left( \sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + \theta_s \sigma_{s-})^2} \right) \right) ds.$$

Therefore the equation becomes to

$$\int_0^t \left( \delta \sqrt{\alpha^2 - (\beta + \theta_s \sigma_{s-})^2} - r_s + \mu^G + \mu \sigma_{s-} - \delta \sqrt{\alpha^2 - (\beta + (\theta_s + 1) \sigma_{s-})^2} \right) dt = 0.$$

Since the above integral equation holds for every  $t$ , the result can be solved by equating the integrand to zero. ■

The two solution branches have been tested empirically on the commodity data considered in this study. It seems that, in most cases, only the second solution can prevent the risk-neutral dynamic from exploding.

### 3.3 Method 2: model spot prices under the risk-neutral measure $Q$

A problem with pricing under the Esscher measure is that the density process is solely determined by the spot prices and thus cannot take any information from the futures price data. A remedy is to model  $Q$ -dynamics of the spot prices directly and calibrate the model parameters from the futures prices. This amounts to selecting a risk-neutral measure 'chosen by the market' among a set of structure preserving measures.

The risk-neutral spot price can be described by the following stochastic differential equations:

$$\begin{aligned} dS_t &= S_{t-} (r_t - \lambda G_{t-}) dt + S_{t-} \int (e^{\sigma_{t-x}} - 1) (\tilde{\mu} - \tilde{\nu}) (dx, dt), \\ dG_t &= \left( r_t - \lambda G_{t-} - \int (e^{\sigma_{t-x}} - 1) \tilde{F}(dx) \right) dt + \int (\sigma_{t-x}) \tilde{\mu}(dx, dt). \end{aligned}$$

We assume again that the driving Lévy process is a standard  $NIG$  process with parameters  $(\xi, \rho)$ .

The model parameters can be obtained by minimizing the aggregate difference between theoretical and realized futures prices.

$$\sum_i \varpi_i (F_i(\chi) - F_i^{obs})^2$$

Here  $\chi$  is the vector of the parameters to calibrate. In this case, we have

$$\chi = (\kappa, \eta, \phi; \xi, \rho).$$

$F(\chi)$  denotes the theoretical futures price based on the model and the value of  $\chi$ , while  $F^{obs}$  denotes the observed futures prices gathered from the market.  $\varpi_i \in [0, 1]$  is the weight for the  $i$ th future contract, which reflects the relative importance of the contract.

However, this kind of calibration problem is usually ill-posed (see Cont and Tankov (2004) [7] and Chiarella Carddack and El-Hassan (2007) [?] for details). A remedy for the ill-posed problem is to add a regularization term to the objective functional. See Cont and Tankov (2004), Chiarella, Carddack and El-Hassan (2007) and Galluccio and Le Cam (2005) [8] for details. Therefore, we change the objective functional into the following form:

$$\min_{\chi} \left\{ \sum_i \varpi_i (F_i(\chi) - F_i^{obs})^2 + \alpha l_f(\nu_{\chi}, \nu_{initial}) \right\}$$

with

$$L_f(\nu_{\chi}, \nu_{initial}) := \begin{cases} \int_{(0, \infty)} f\left(\frac{d\nu_{\chi}}{d\nu_{initial}}\right) d\nu_{initial} & \text{if } \nu_{\chi} \ll \nu_{initial} \\ +\infty & \text{else.} \end{cases}$$

The function  $f$  is chosen to be

$$f(x) = x \log(x) - x + 1.$$

The existence of a solution for this regularization calibration problem has been proved by Keller-Ressel (2006) [13].

### 3.4 Method 3: non-parametric estimation of risk-neutral densities

Grith, Härdle and Schienle (2010) [9] proposed a non-parametric approach to estimate risk-neutral density from option prices. In this study, their idea will be applied to futures markets.

There is a link between the  $p$ , the conditional density function of the physical measure, and  $q$ , the conditional density function of the risk-neutral measure, namely

$$q(S_T | \mathcal{F}_t) = m(S_T | \mathcal{F}_t) p(S_T | \mathcal{F}_t).$$

Here  $m$  is the pricing kernel, which summarizes information related to asset pricing. We cannot incorporate all the factors driving the form of the pricing kernel, so we consider the projection of the pricing kernel on the set of available payoff functions and denote it as  $m^*$ . Assume it is close to  $m$  in the sense that

$$\|m - m^*\|^2 = \int |m(x) - m^*(x)|^2 dx < \epsilon.$$

Further assume  $m^*$  has a Fourier series expansion

$$m^*(S_T | \mathcal{F}_t) = \sum_{l=1}^{\infty} \alpha_{l,t} g_l(S_T | \mathcal{F}_t)$$

where  $\alpha_l$  are Fourier coefficients and  $g_l$  is a fixed collection of basis functions. Following Grith, Härdle and Schienle (2010), we adopt the Laguerre polynomials to conduct empirical analysis. In practice, we can only expand  $m^*$  up to a finite number  $L$ , which gives us an approximation

$$\hat{m}(S_T | \mathcal{F}_t) = \sum_{l=1}^L \hat{\alpha}_{l,t} g_l(S_T | \mathcal{F}_t).$$

The remaining task is to estimate  $\hat{\alpha}_{l,t}$  from the derivative prices data. The commodity future prices can be expressed in the following way:

$$\begin{aligned} Y_{i,t} &= e^{-rt} \int_0^{\infty} S_T \sum_{l=1}^L \hat{\alpha}_{l,t} g_l(S_T | \mathcal{F}_t) p(S_T | \mathcal{F}_t) dS_T + \varepsilon_i \\ &= \sum_{l=1}^L \hat{\alpha}_{l,t} \left\{ e^{-rt} \int_0^{\infty} S_T g_l(S_T | \mathcal{F}_t) p(S_T | \mathcal{F}_t) dS_T \right\} + \varepsilon_i. \end{aligned}$$

Set

$$\psi_{il} = e^{-rt} \int_0^\infty S_T g_l(S_T | \mathcal{F}_t) p(S_T | \mathcal{F}_t) dS_T.$$

Finally we can obtain a feasible estimator of  $\alpha$  by least-square estimation

$$\tilde{\alpha} = \left( \hat{\Psi}^T \hat{\Psi} \right)^{-1} \hat{\Psi}^T Y$$

which will give us the pricing kernel

$$\hat{m}(S_T | \mathcal{F}_t) = g(S_T | \mathcal{F}_t)^\top \tilde{\alpha}$$

and hence the risk-neutral density.

## 3.5 Empirical analysis

### 3.5.1 Estimation method and simulation techniques

#### Estimate $P$ -dynamic parameters

The COGARCH model can be estimated by a pseudo-maximum likelihood (PML) method proposed by Maller, Müller and Szimayer (2008) [16]. Suppose we have observations  $G(t_i), 0 = t_0 < t_1 < \dots < t_N = T$ . In Maller, Müller and Szimayer (2008)'s case,  $Y_i$  is defined as the difference between  $G(t_i)$  and  $G(t_{i-1})$ , i.e., the observed returns.

In our case, we define  $Y_i$  in a different way in order to facilitate the estimation procedure, namely

$$Y_i := G(t_i) - G(t_{i-1}) - \left( \mu^G - \lambda^G G(t_{i-1}) \right) \Delta t_i = \int_{t_{i-1}}^{t_i} \sigma(s-) dL(s).$$

Assuming the  $Y_i$  are conditionally normal, we can write a pseudo-log-likelihood function for  $Y_1, \dots, Y_N$  as

$$L_N(\varpi, \eta, \phi) = -\frac{1}{2} \sum_{i=1}^N \frac{Y_i^2}{\rho_i^2} - \frac{1}{2} \sum_{i=1}^N \log(\rho_i^2) - \frac{N}{2} \log(2\pi)$$

where

$$\begin{aligned} \rho_i^2 &: = E[Y_i^2 | \mathcal{F}_{t_{i-1}}] \approx \sigma^2(t_{i-1}) \Delta t_i \\ \sigma^2(t_i) &= \varpi \Delta t_{i-1} + \phi e^{-\eta \Delta t_{i-1}} Y_{i-1}^2 + e^{-\eta \Delta t_{i-1}} \sigma^2(t_{i-1}). \end{aligned}$$

This method can handle both regularly spaced data and irregularly spaced data.

### Simulate $Q$ -dynamics

Another problem with the Esscher measure is that it sometimes changes the model class of the driving process, which happens in this case. After measure-changing the driving process  $L$  is not a  $NIG$  process anymore due to the time-varying  $\sigma$  and  $\theta$ . This increases the difficulty to simulate  $G^Q$ , the  $G$  process under  $Q$ . One way to do it is to approximate  $dG^Q$  by the following process:

$$d\widehat{G}_t^Q = r_t dt + \widetilde{\sigma}_t dW_t + \sum_{i=1}^d c_{i,t} \Delta N_t^{(i)} - c_{i,t} \lambda_{i,t} 1_{|c_{i,t}| \leq 1} dt$$

The logic here is to approximate large jumps by  $d$  nonhomogeneous Poisson processes  $N_t^{(i)}$ ,  $1 \leq i \leq d$ , and small jumps by a Brownian motion with volatility

$$\widetilde{\sigma}_t^2 = \int_{|x| < \epsilon} x^2 F_t^*(dx)$$

where

$$F_t^*(dx) = e^{\theta_t \sigma_t - x} F(dx).$$

Approximating small jumps by a Brownian motion is a commonly used technique and can achieve more accurate results than certain alternative approaches (Asmussen and Glynn (2007) [1]). However, this approximation is not always feasible. A rigorous discussion for Lévy case was provided by Asmussen and Rosiński (2001) [2], suggesting the necessary condition for the approximation,

$$\varepsilon / \sigma_\varepsilon \rightarrow 0,$$

holds if and only if

$$\sigma_{c\sigma_\varepsilon \wedge \varepsilon} \sim \sigma_\varepsilon.$$

For a  $NIG$  process, Asmussen and Rosiński (2001) prove that the approximation is valid. In our case, after the measure-changing,  $F_t^*(dx)$  resembles the Lévy measure of a  $NIG$  process with time-varying parameter  $\beta^* = \beta - \theta_t \sigma_t$ ,

$$F_t^*(dx) = \nu(dx) = \frac{\delta \alpha}{\pi |x|} e^{(\beta - \theta_t \sigma_t)x} K_1(\alpha |x|) dx.$$

Next, let us look at the approximation of the big jumps. We adopt a nonhomogeneous Poisson process with intensity process

$$\lambda_{i,t} = \begin{cases} F_t^*([a_{i-1}, a_i)) & 1 \leq i \leq k \\ F_t^*([a_i, a_{i+1})) & k+1 \leq i \leq d \end{cases}$$

and time-changing jump sizes

$$c_{i,t}^2 \lambda_{i,t} = \begin{cases} \int_{a_{i-1}}^{a_i} x^2 F_t^*(dx) & 1 \leq i \leq k \\ \int_{a_i}^{a_{i+1}} x^2 F_t^*(dx) & k+1 \leq i \leq d. \end{cases}$$

Here,  $\{a_i\}$  are real-numbers satisfying

$$\begin{aligned} a_1 &< \dots < a_k = -\epsilon \\ \epsilon &= a_{k+1} < \dots < a_d \\ \epsilon &\in (0, 1). \end{aligned}$$

We choose  $\epsilon = 0.3$ ,  $a_1 = -5.3$ ,  $a_d = 5.3$ . All the integrals are approximated by the composite Simpson's rule and the nonhomogeneous Poisson processes are generated by the algorithm in [3].

### Numerical approximation of integrals

This study involves quite a lot of integrals such that explicit solutions are very difficult to obtain. In this case, a numerical approximation has been used. An integral between two finite numbers can be approximated by the composite Simpson's rule

$$\int_a^b f(x)dx \approx \frac{h}{3} \left[ f(x_0) + 2 \sum_{j=1}^{n/2-1} f(x_{2j}) + 4 \sum_{j=1}^{n/2} f(x_{2j-1}) + f(x_n) \right].$$

For integrals without finite boundary, a transformation must be adopted first, e.g.

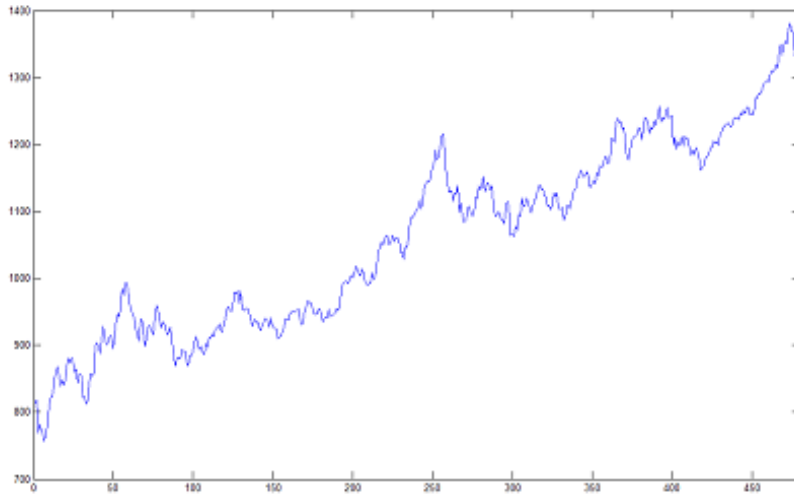
$$\int_0^\infty f(x)dx = \int_0^1 f\left(\frac{t}{1-t}\right) \frac{1}{(1-t)^2} dt.$$

### 3.5.2 Gold and gold future market

Gold has played an important role in the development of civilization. It was firstly used for decoration and then served as a store of wealth and a medium of trade. Nowadays people possess gold to hedge the risks generated from economic or political events. Figure 1 show the gold spot prices and the logarithm of gold spot prices between 26/11/2008 and 26/10/2010. The data is of daily scope and all the non-trading days have been excluded.

Gold futures provide hedging tools for commercial producers and users of gold, opportunities for portfolio diversification and global gold price discovery (<http://www.cmegroup.com>). I considered the 11 gold futures traded on COMEX (New York Commodity Exchange, Inc.) for empirical study: GCG09, GCJ09, GCM09, GCQ09, GCV09, GCZ9, GCG0, GCJ0, GCM0, GCQ0, GCV0. Here GC is the product symbol. The third letter and the followed numbers state the expiry month and year of the contract. Trades of the contract are allowed until three working days before the settlement month. Figure 2 shows the gold future prices surface. We are considering the settlement prices of 11 gold future contracts between 26/11/2008 and 26/10/2010.

All the data are downloaded from the bloomberg terminal.



Gold spot price

#### Fit the model under the $P$ measure

The first task is to fit the model to the gold spot price. If we assume the data points are regularly spaced, i.e., ignore the non-trading days, the estimates of parameters are

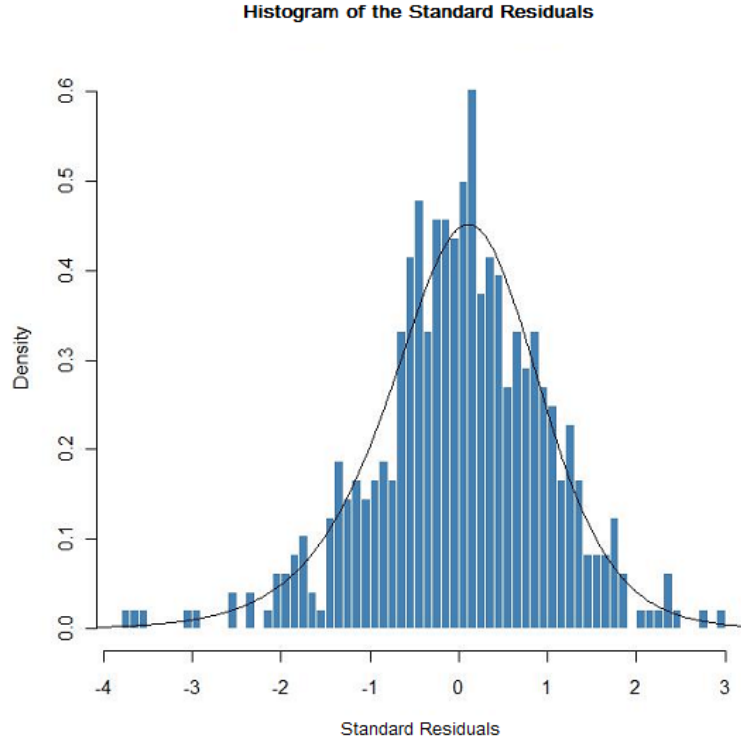
$$\begin{array}{ccccc} \kappa & \eta & \phi & \mu^G & \lambda \\ 2.1662e-06 & 0.0558 & 0.0380 & 0.0016 & 0.0023 \end{array}$$

The estimates were obtained by PMLE. So we need to have a closer look of the standardized residuals in the return function. First of all, we did a Jarque-Bera test to test the normal assumption of the PMLE. It can be seen from the result that the normal distribution assumption can be rejected.

$$\begin{array}{cc} \text{Jarque - Bera} & p\text{-value} \\ 34.6801 & 0.00000 \end{array}$$

Secondly, we try to fit a standardized  $NIG$  distribution. The following table and figure give the maximum likelihood estimates to the four parameters of the  $NIG$  distribution and the comparison between the histogram of the standardized residual and the density function of the  $NIG$  distribution with the four estimated parameters.

$$\begin{array}{cccc} \alpha & \beta & \mu & \delta \\ 1.7182171 & -0.3124525 & 0.3021202 & 1.6336979 \end{array}$$



Finally, we check the fitted *NIG* distribution by the Chi-square test. Unfortunately, the test result suggests the standard residuals do not follow the fitted *NIG* distribution.

| <i>Chi2stat</i> | <i>p - value</i> |
|-----------------|------------------|
| 39.6754         | 0.0000           |

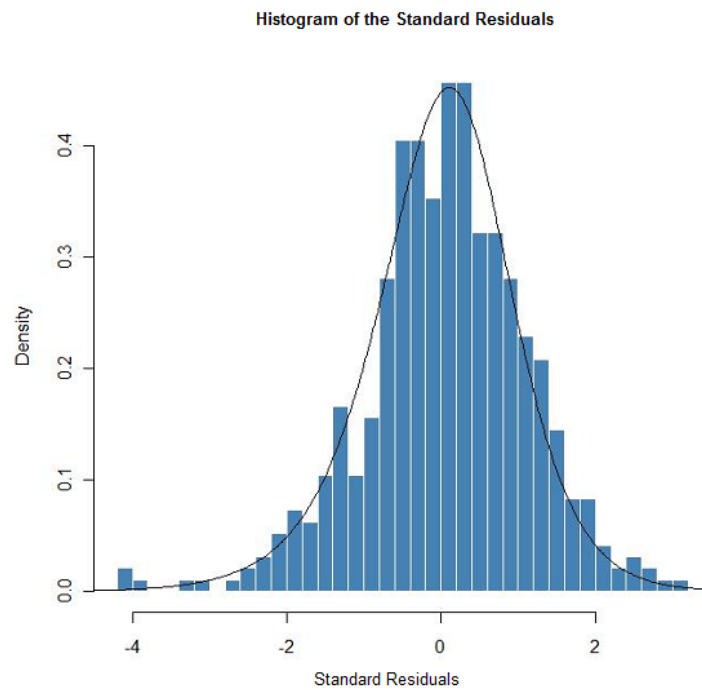
Right now let us look at the case when the non-trading days are considered. Put it in another way, the data points are irregularly spaced. Let us do the PMLE again to fit the model parameters and check the distribution assumption of the standard residuals.

| $\kappa$       | $\eta$ | $\phi$ | $\mu^G$         | $\lambda$ |
|----------------|--------|--------|-----------------|-----------|
| $1.0526e - 06$ | 0.0333 | 0.0275 | $-9.3388e - 05$ | -0.0014   |

| <i>Jarque - Bera</i> | <i>p - value</i> |
|----------------------|------------------|
| 39.048               | 0.00000          |

Again, the normality assumption is rejected. So I tried the standard *NIG* distribution again. The following results are the estimated values of the parameters and the histogram.

| $\alpha$  | $\beta$    | $\mu$     | $\delta$  |
|-----------|------------|-----------|-----------|
| 1.6959514 | -0.3029201 | 0.2932561 | 1.6154437 |



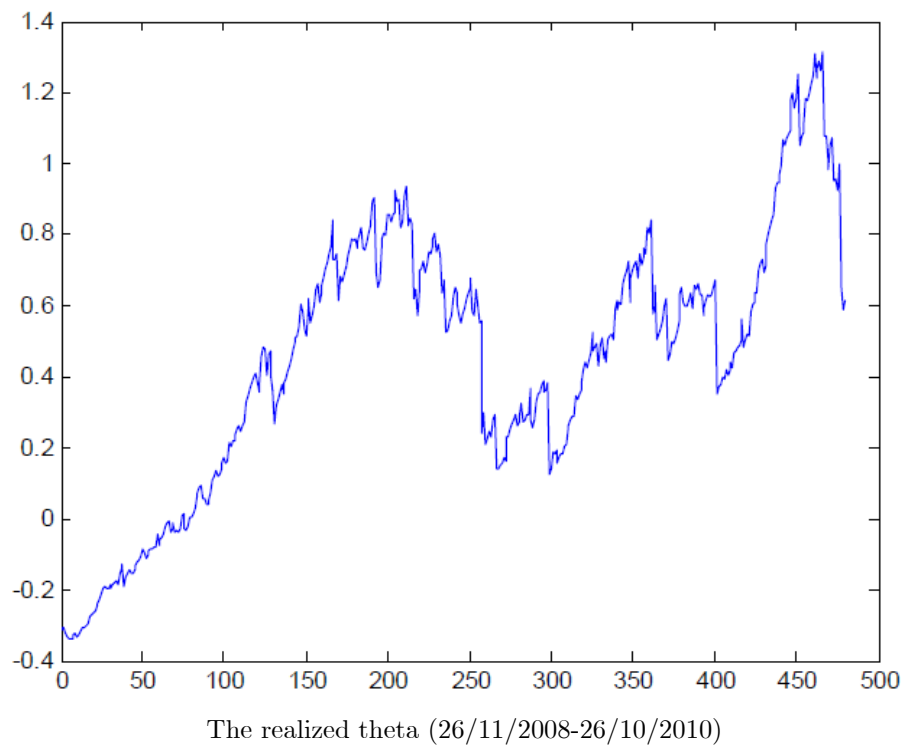
This time the standard *NIG* assumption cannot be rejected by the goodness-of-fit test. Therefore, in the following part, we will only focus on the model estimated under the irregular spaced data assumption.

| <i>Chi2stat</i> | <i>p - value</i> |
|-----------------|------------------|
| 7.35            | 0.3934           |

### The realized $\theta$ process

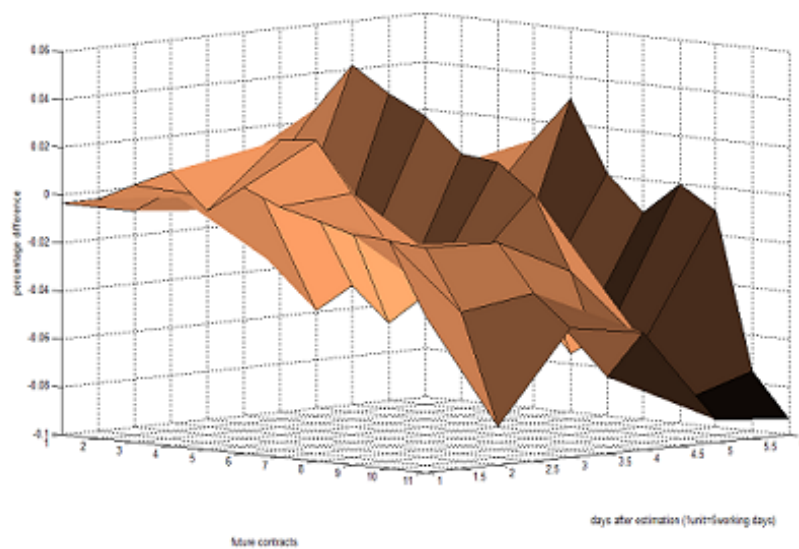
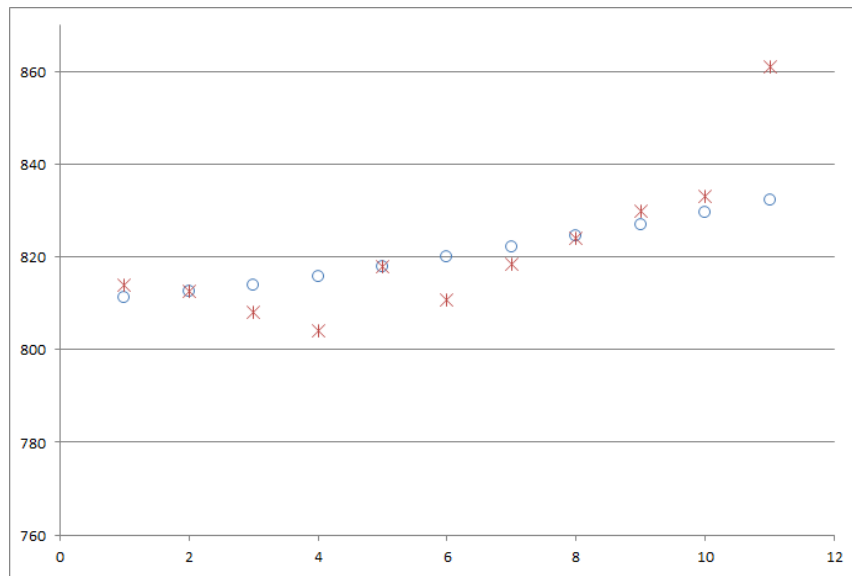
The next task is to find out the realized  $\theta$ , which determines the realized Esscher measure. We can see from the picture that  $\theta$ , the sole factor to determine the risk premium, is indeed a stochastic process in the gold future market. The positive value

of  $\theta$  means that jump sizes tend to be bigger under the pricing measure  $Q$ .

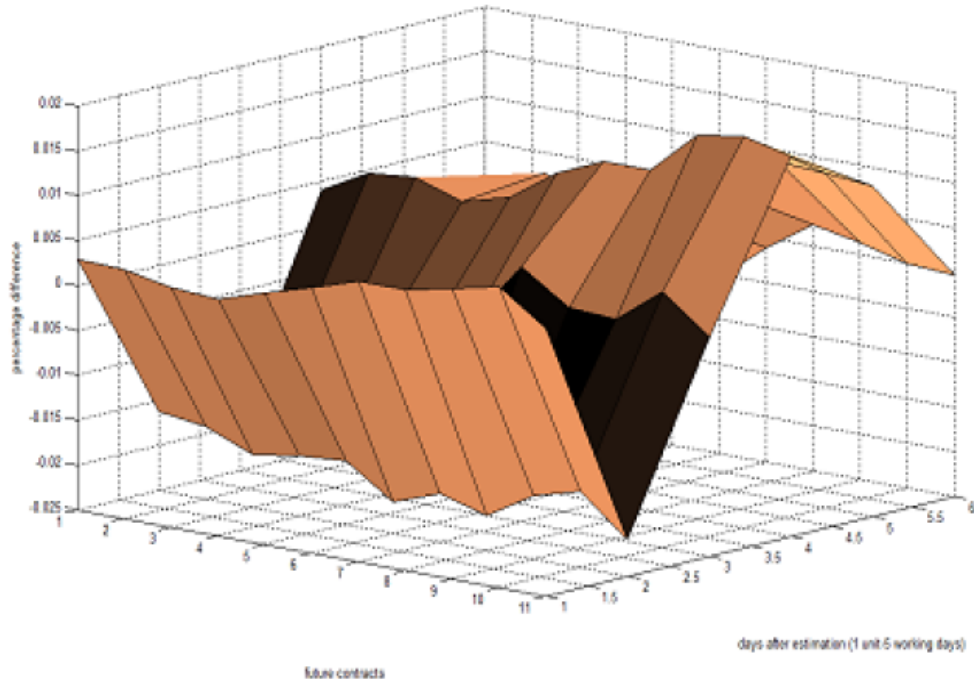
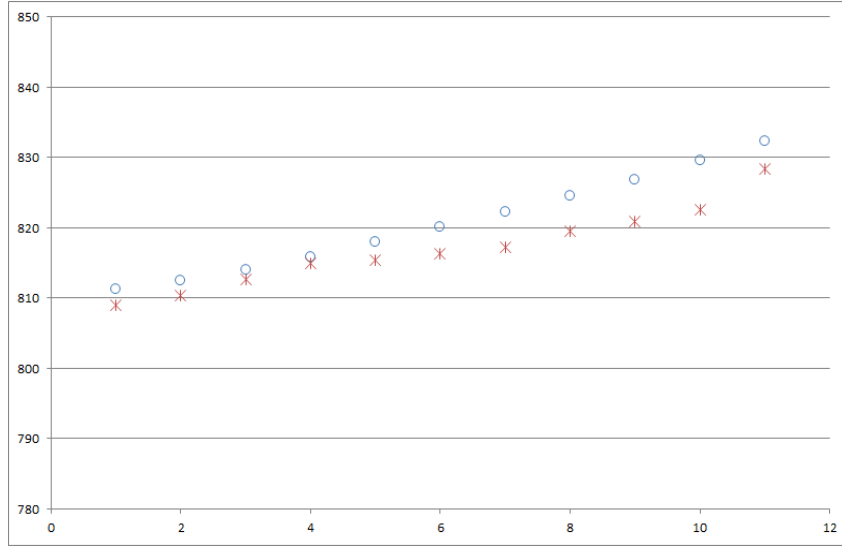


### Theoretical future prices

The last task is to find out the theoretical future prices determined by the three approaches and compare them with the realized gold future prices.

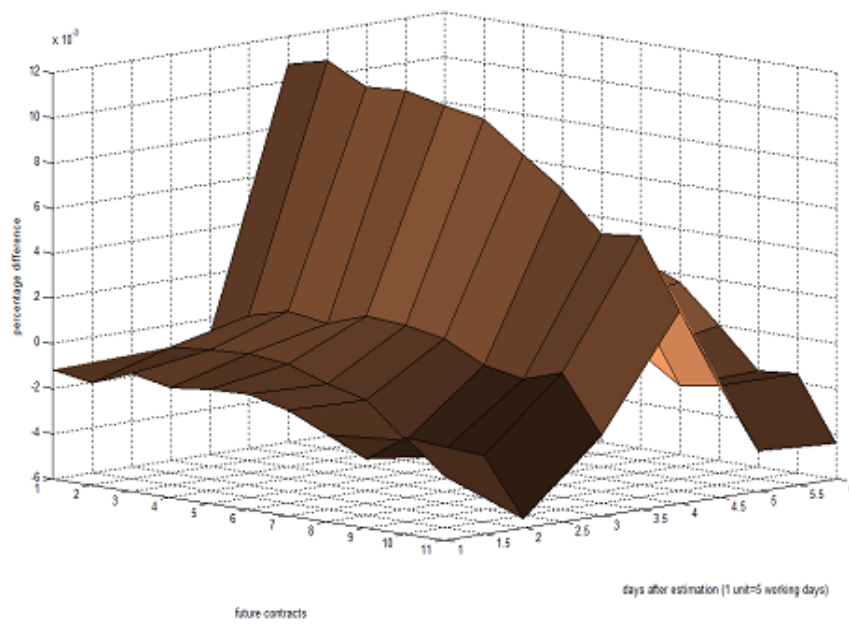
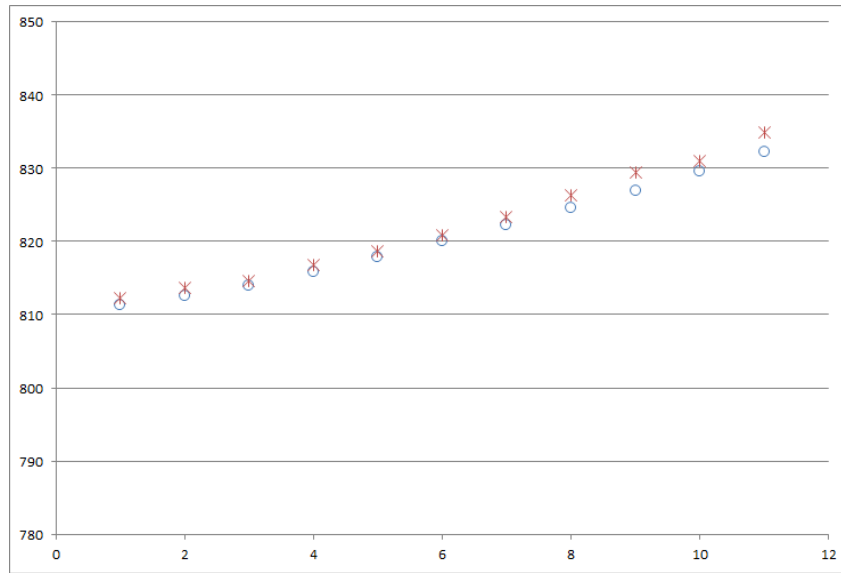


The above show the result of the first method. The first figure shows the fitness of the theoretical future prices to the real future curve on 28 Nov 2008. The second figure tells the percentage difference of the model prices and real prices during the following 30 trading days. The average of the absolute values of percentage difference in this period is 2.706%.



The above is the result of the second method. The model is fitted by the historical future curves on 5 trading days until (including) 28 Nov 2008. The first figures shows the performance of the model on 28 Nov 2008. The second shows the performance of the same model for the following 30 trading days. The average of the absolute values of percentage difference is 0.8509%. The calibrated parameters during this period are

| $\kappa$   | $\eta$ | $\phi$ | $\lambda$ | $\alpha$ | $\beta$ | $\mu$ | $\delta$ |
|------------|--------|--------|-----------|----------|---------|-------|----------|
| $6.34E-07$ | 0.559  | 0.0275 | 0.0024    | 1.3863   | -1.2147 | 0.282 | 0.1551   |

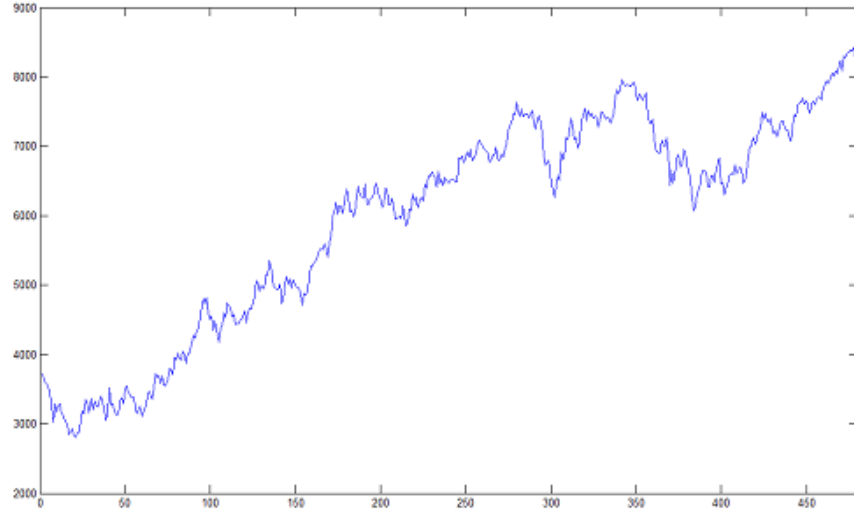


The above is the result of the third method. The model is fitted by the historical future curves on 5 trading days until (including) 28 Nov 2008. The first figures shows the performance of the model on 28 Nov 2008. The second shows the performance of the same model for the following 30 trading days. The average of the absolute values of percentage difference is 0.2876%. Obviously, the third method achieves the best fit to the real market prices.

### 3.5.3 Copper and copper future market

Due to its vast quantity, copper has been for a long time an important material for weapon and household objects in the history. In the modern era, copper has been widely used in electrical and electronic products. The following figure shows the copper spot prices between 26/11/2008 and 26/10/2010.

Copper futures are traded on several exchanges: LME (London Metal Exchange), COMEX, SHME (Shanghai Metal Exchange) etc. This study considers 11 future contracts traded on LME: LPG09, LPJ09, LPM09, LPQ09, LPV09, LPZ9, LPG0, LPJ0, LPM0, LPQ0, LPV0.



Copper spot price

#### Fit the model under the $P$ measure

We firstly fit the  $P$  model to the spot data. Under the regular space assumption, the estimated value of the parameters are

|                |        |        |         |           |
|----------------|--------|--------|---------|-----------|
| $\kappa$       | $\eta$ | $\phi$ | $\mu^G$ | $\lambda$ |
| $7.1434e - 06$ | 0.0951 | 0.0828 | 0.0051  | 0.0055    |

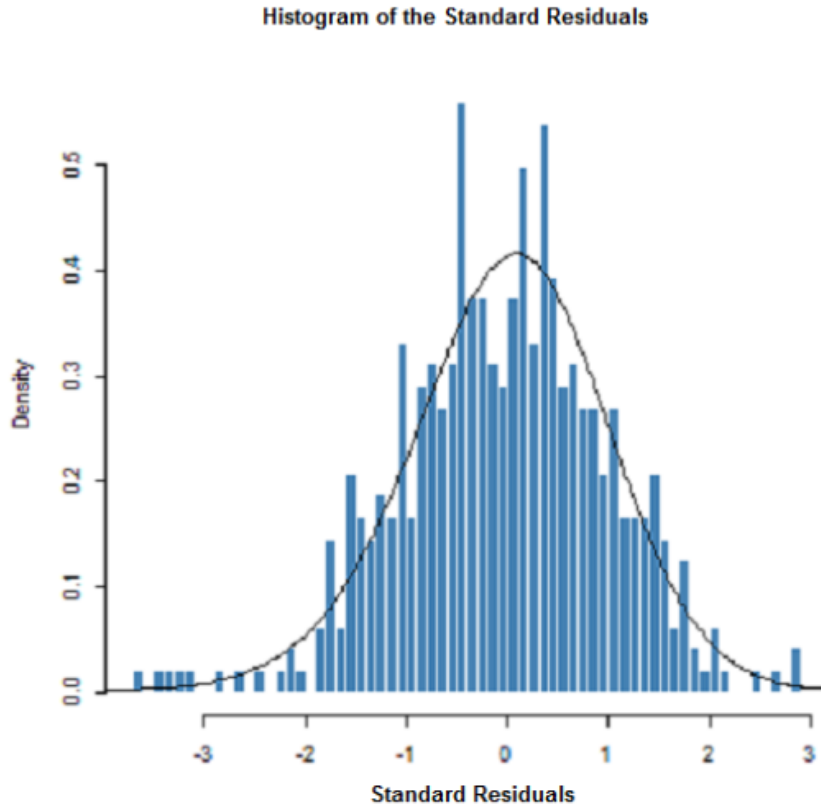
Normality assumption of the standardized residuals can be rejected.

|                      |                  |
|----------------------|------------------|
| <i>Jarque – Bera</i> | <i>p – value</i> |
| 9.3126               | 0.0095           |

Again we choose standard  $NIG$  as a candidate distribution for the standardized

residuals and check this assumption by the goodness-of-fit test. The *SNIG* cannot be rejected.

$$\begin{array}{cccc} \alpha & \beta & \mu & \delta \\ 3.2182479 & -0.6807412 & 0.6502828 & 3.0046911 \end{array}$$



$$\begin{array}{cc} Chi2stat & p - value \\ 4.3212 & 0.8270 \end{array}$$

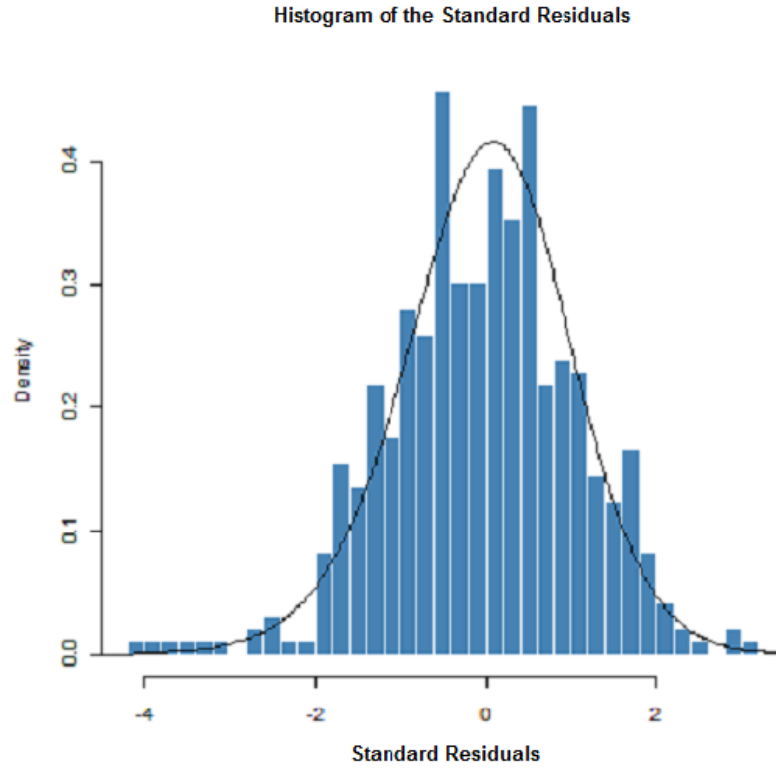
And then we repeated the above procedure to fit the model under the irregular spaced data assumption. The following are the estimated parameters. Jarque-Bera test is again significant.

$$\begin{array}{ccccc} \varpi & \eta & \phi & \mu^G & \lambda \\ 3.2007e-06 & 0.0540 & 0.0568 & 0.0028 & 0.0023 \end{array}$$

$$\begin{array}{cc} Jarque - Bera & p - value \\ 10.9453 & 0.0042 \end{array}$$

Therefore I tried the *SNIG* distribution and check the new assumption by the goodness-of-fit test. The *SNIG* assumption cannot be rejected.

| $\alpha$  | $\beta$    | $\mu$     | $\delta$  |
|-----------|------------|-----------|-----------|
| 3.1315263 | -0.6307958 | 0.6052008 | 2.9428774 |

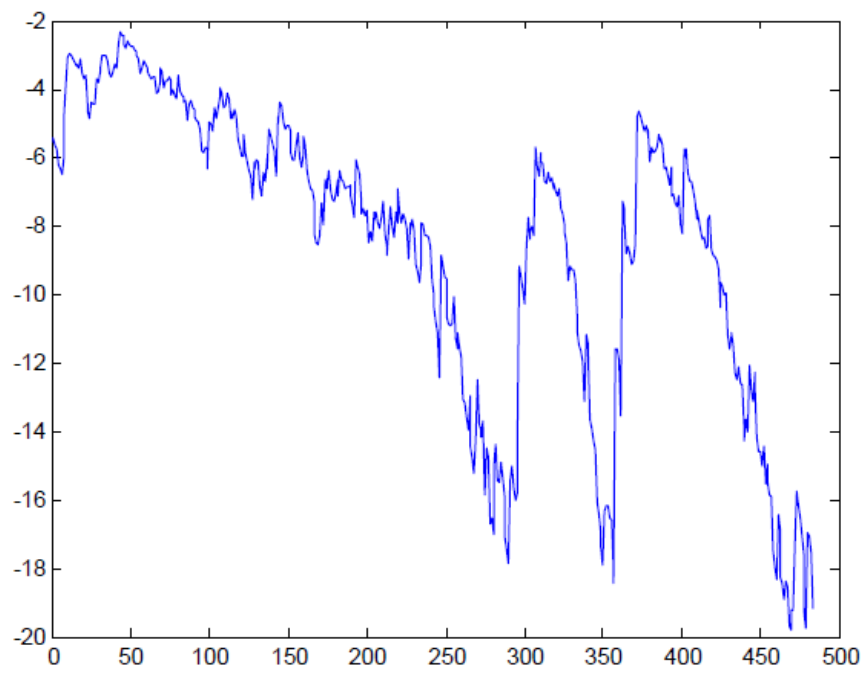


| $Chi2stat$ | $p - value$ |
|------------|-------------|
| 11.3803    | 0.0773      |

Although the model is valid under both regularly spaced data assumption and irregularly spaced data assumption, we will only adopt the model under the irregularly spaced data assumption to estimate the theoretical future price.

### Realized $\theta$ process

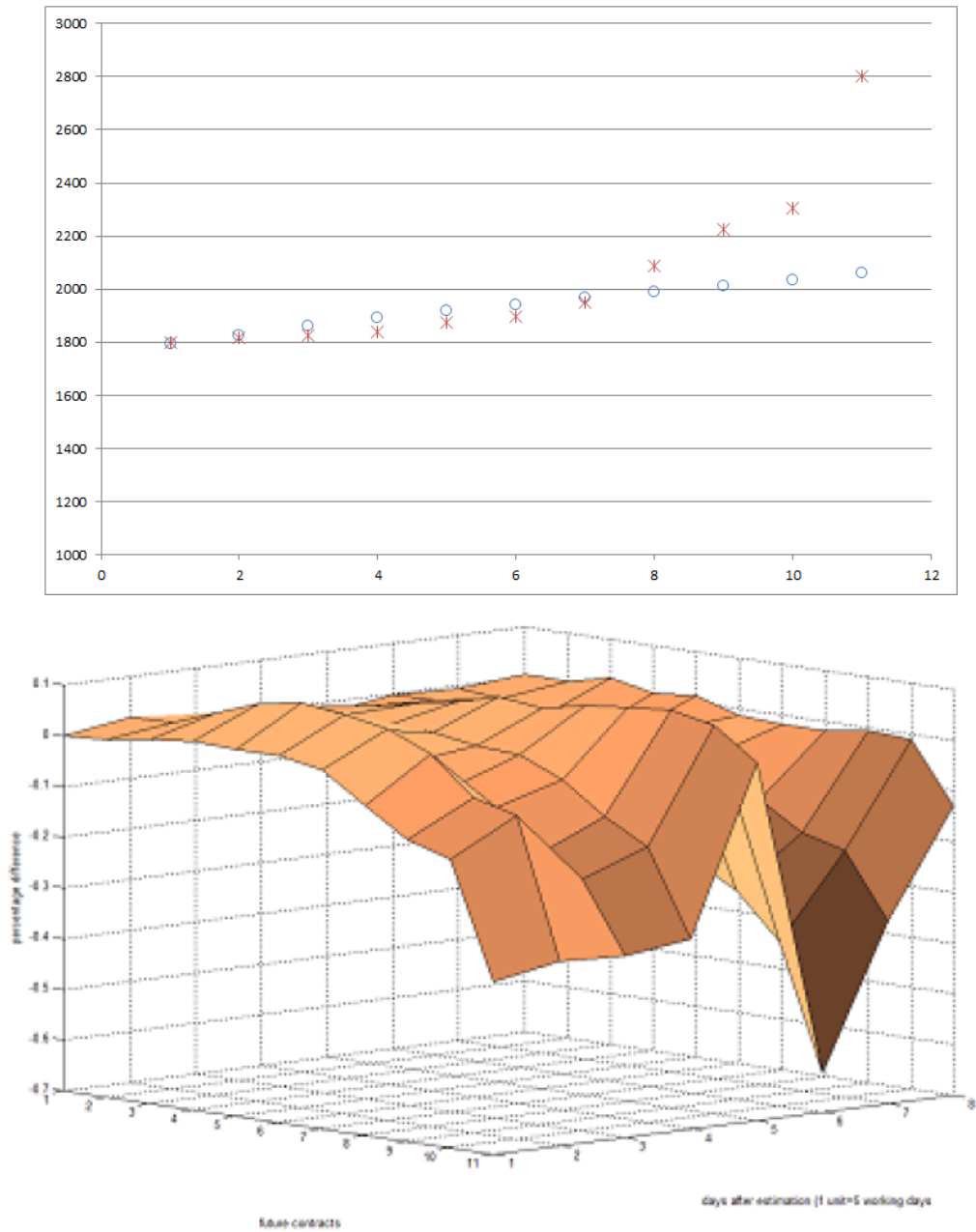
The copper future market also witnesses a stochastic  $\theta$ . However, unlike the gold future market, the realized  $\theta$  process in the copper future market always take negative value, which means the jump size under  $Q$  tend to be smaller than the one under  $P$ .



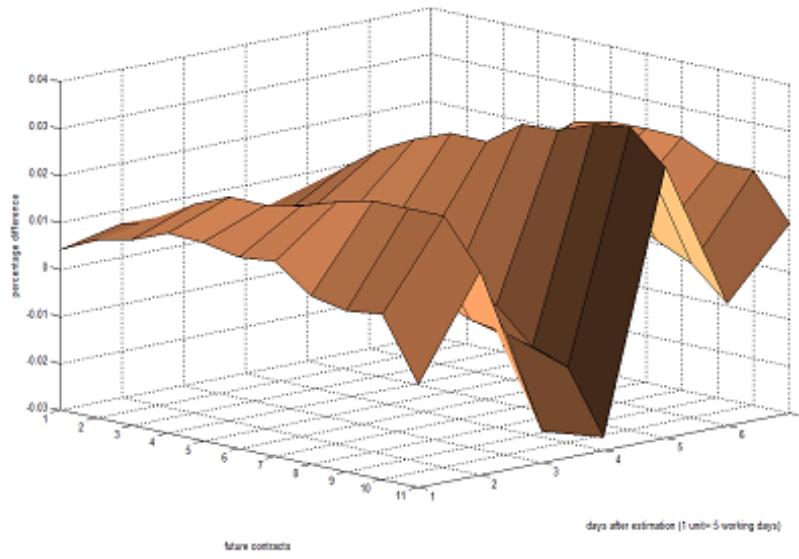
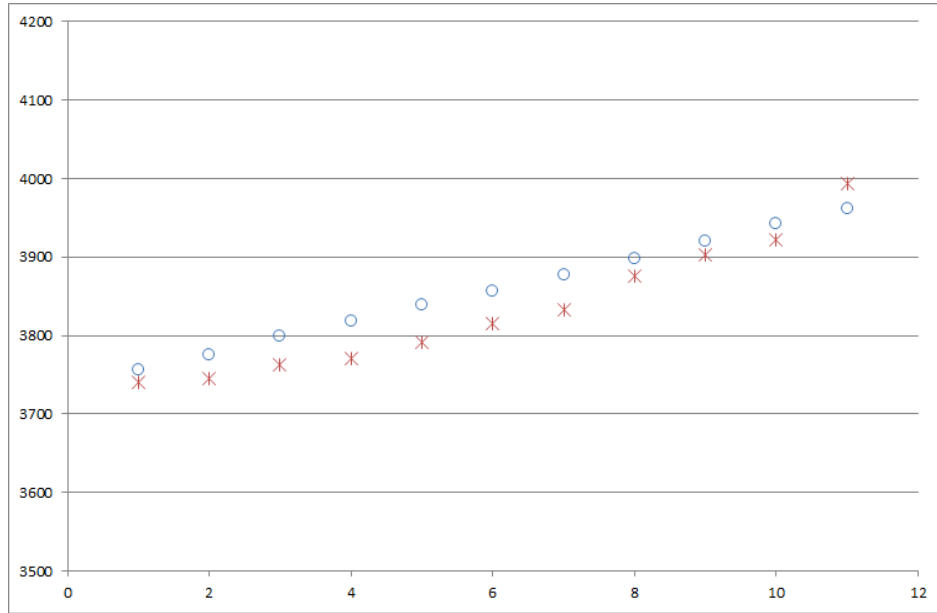
The realized theta (26/11/2008-26/10/2010)

### Theoretical future prices

Finally let us look at the comparison between the theoretical copper future prices and the realized ones.

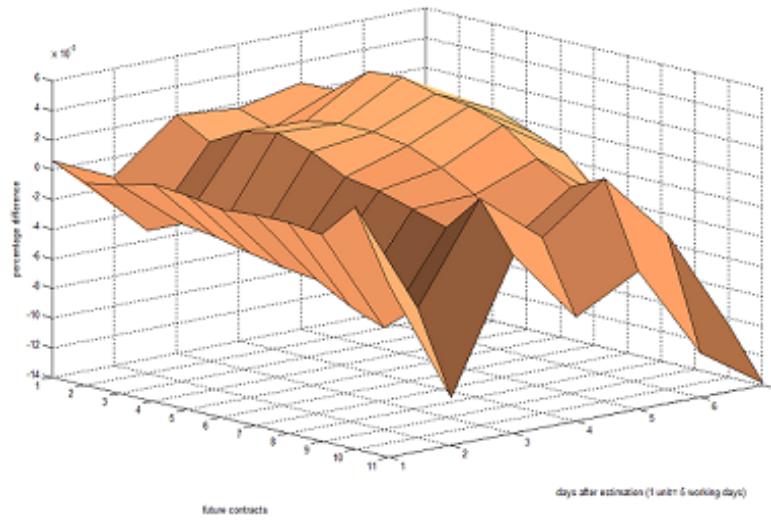
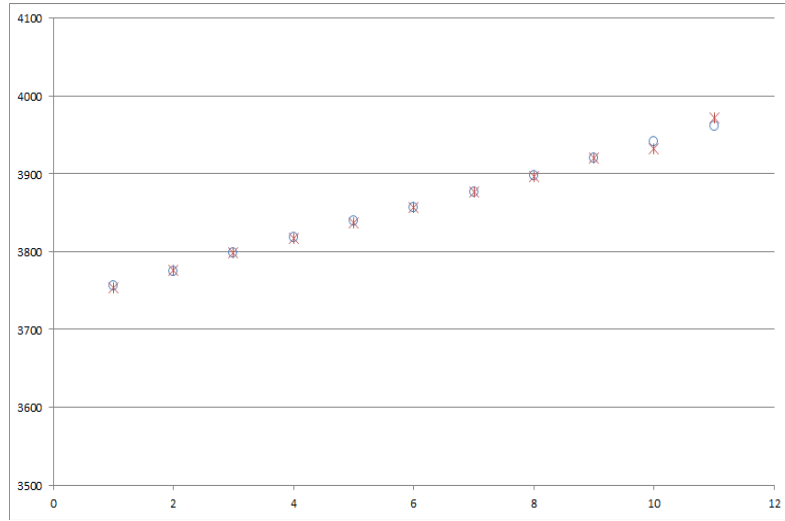


The above are the results of the first method. In-the-sample experiment shows an average absolute values of percentage difference of 4.8347%.



The above show the results of the second method. The average of the absolute values of percentage difference is 1.3073%. The calibrated parameters during this period are as follows

| $\kappa$   | $\eta$ | $\phi$ | $\lambda$ | $\alpha$ | $\beta$ | $\mu$  | $\delta$ |
|------------|--------|--------|-----------|----------|---------|--------|----------|
| $2.97E-06$ | 0.802  | 0.0339 | 0.0028    | 0.2528   | -0.2385 | 0.0263 | 0.0093   |

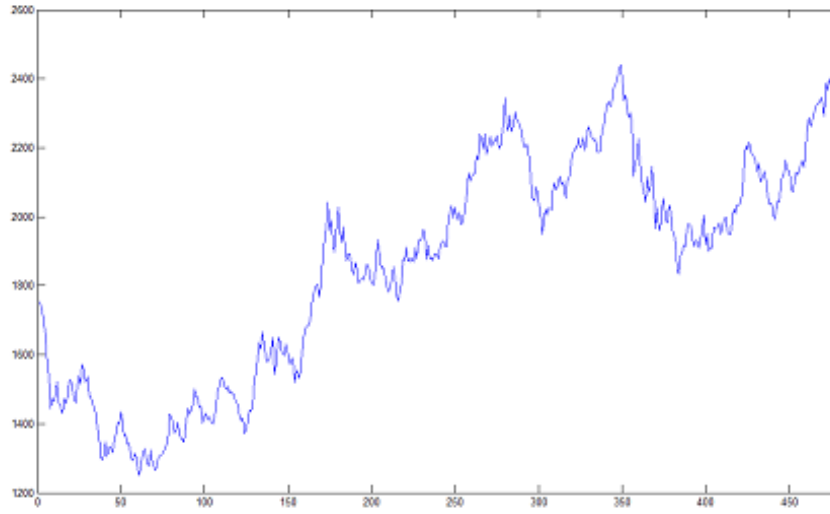


The above are the results of the third method. The average of the absolute values of percentage difference is 0.291%. Again, the third method outperformed the other two.

### 3.5.4 Aluminum and aluminum future market

Aluminium is an important industrial material and has wide application in construction and manufacturing. Aluminum also substitutes for copper in many areas. The following figure show the aluminum spot price between 26/11/2008 and 26/10/2010.

We considered 11 aluminum future contracts traded on LME between 26/11/2008 and 26/10/2010.



Aluminium spot price

**Fit the model under the  $P$  measure**

The empirical analyzes starts again with fitting the  $P$  model to the spot price. The following are the values of the estimated parameters under the regular space assumption.

|              |        |        |         |           |
|--------------|--------|--------|---------|-----------|
| $\kappa$     | $\eta$ | $\phi$ | $\mu^G$ | $\lambda$ |
| $1.3167e-05$ | 0.0706 | 0.0312 | 0.0011  | 0.0037    |

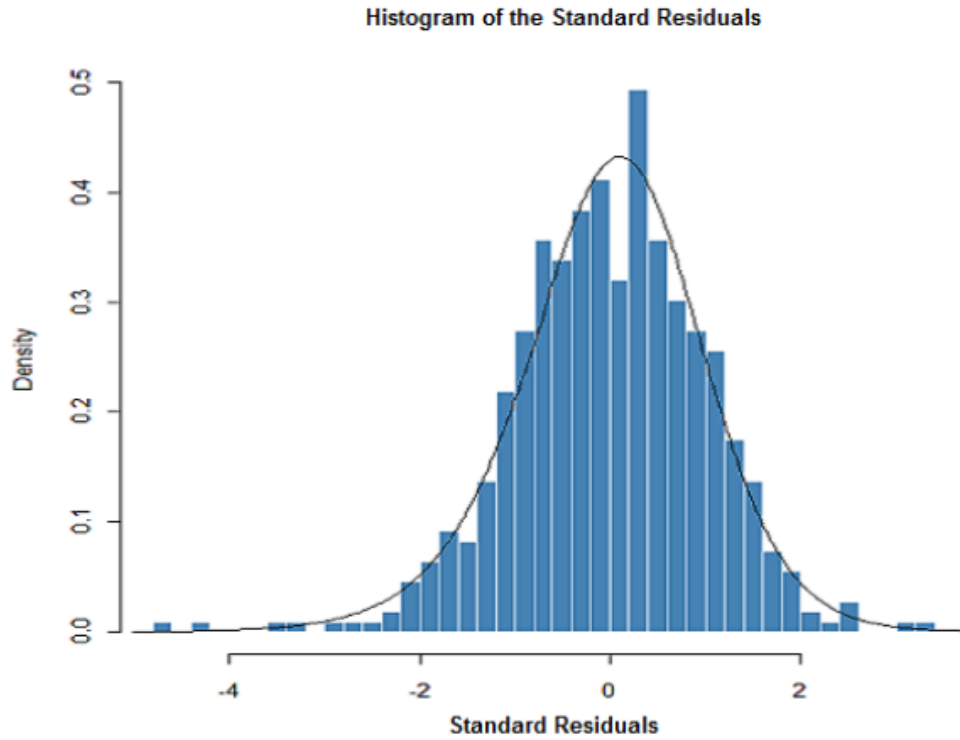
Normality assumption of the standardized residuals can be again rejected.

|                      |                  |
|----------------------|------------------|
| <i>Jarque – Bera</i> | <i>p – value</i> |
| 51.4619              | 0.0000           |

A standard  $NIG$  distribution has been adopted to describe the distribution of the standardized residuals. The following table show the estimates.

|           |            |           |           |
|-----------|------------|-----------|-----------|
| $\alpha$  | $\beta$    | $\mu$     | $\delta$  |
| 2.2156312 | -0.4329320 | 0.4164023 | 2.0899585 |

The shape histogram shows similarity to the shape of the density function of the fitted  $SNIG$  distribution. The distribution assumption cannot be rejected by the test.



|                 |                  |
|-----------------|------------------|
| <i>Chi2stat</i> | <i>p - value</i> |
| 3.4897          | 0.7453           |

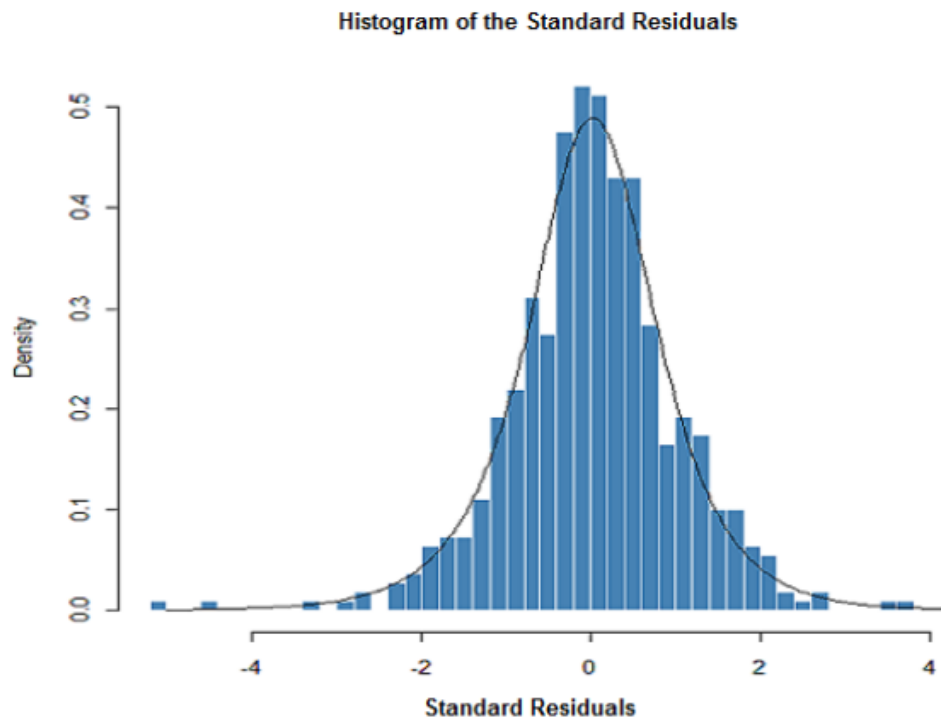
Consider the model again under the irregular spaced data assumption. The following are the estimated parameters and the result of the Jarque-Bera test.

|                |        |        |                |           |
|----------------|--------|--------|----------------|-----------|
| $\kappa$       | $\eta$ | $\phi$ | $\mu^G$        | $\lambda$ |
| $9.6356e - 06$ | 0.0591 | 0.0302 | $6.4010e - 05$ | -0.0023   |

|                      |                  |
|----------------------|------------------|
| <i>Jarque - Bera</i> | <i>p - value</i> |
| 129.4878             | 0.0000           |

Fit a *SNIG* distribution to the data and test it with the goodness-of-fit test. The *SNIG* assumption cannot be rejected.

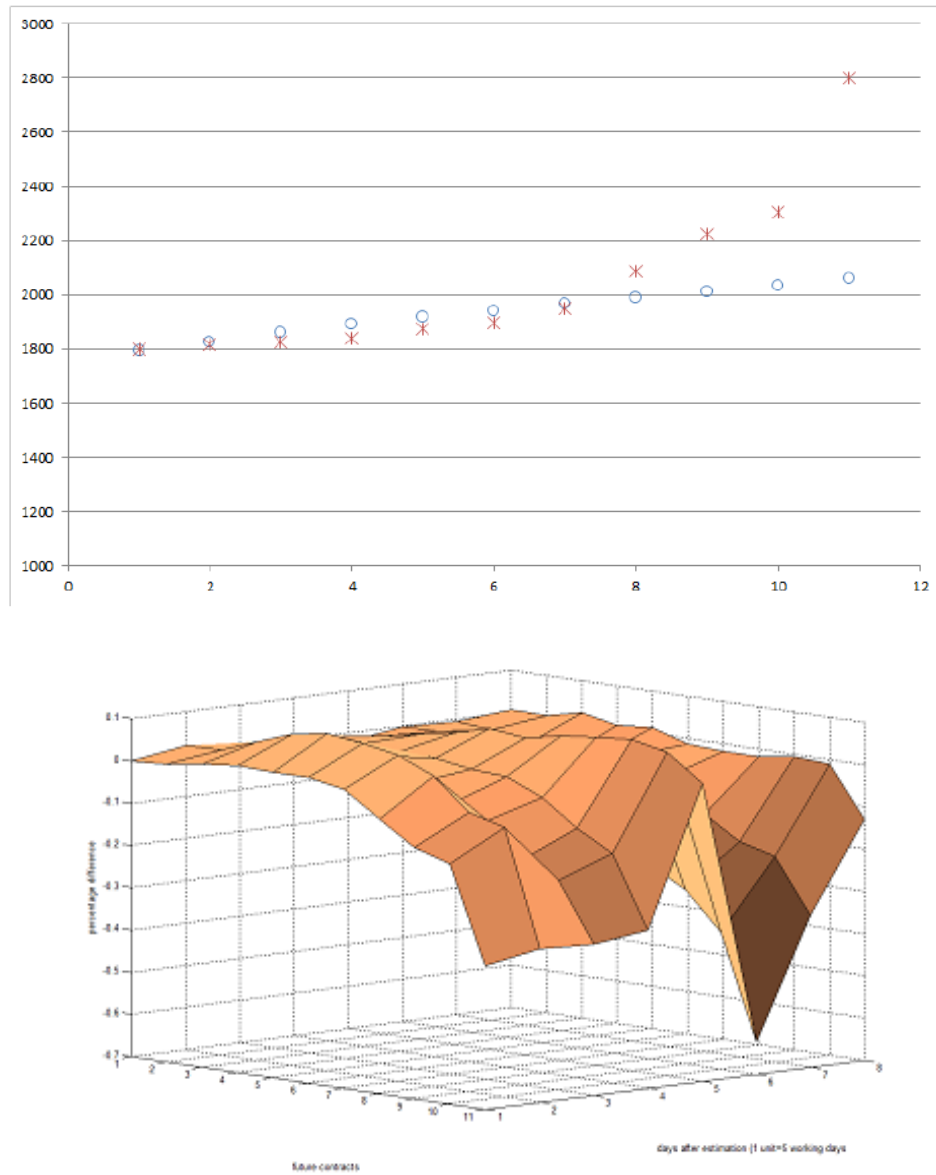
|            |             |            |            |
|------------|-------------|------------|------------|
| $\alpha$   | $\beta$     | $\mu$      | $\delta$   |
| 1.19206241 | -0.05224162 | 0.05214129 | 1.18862986 |



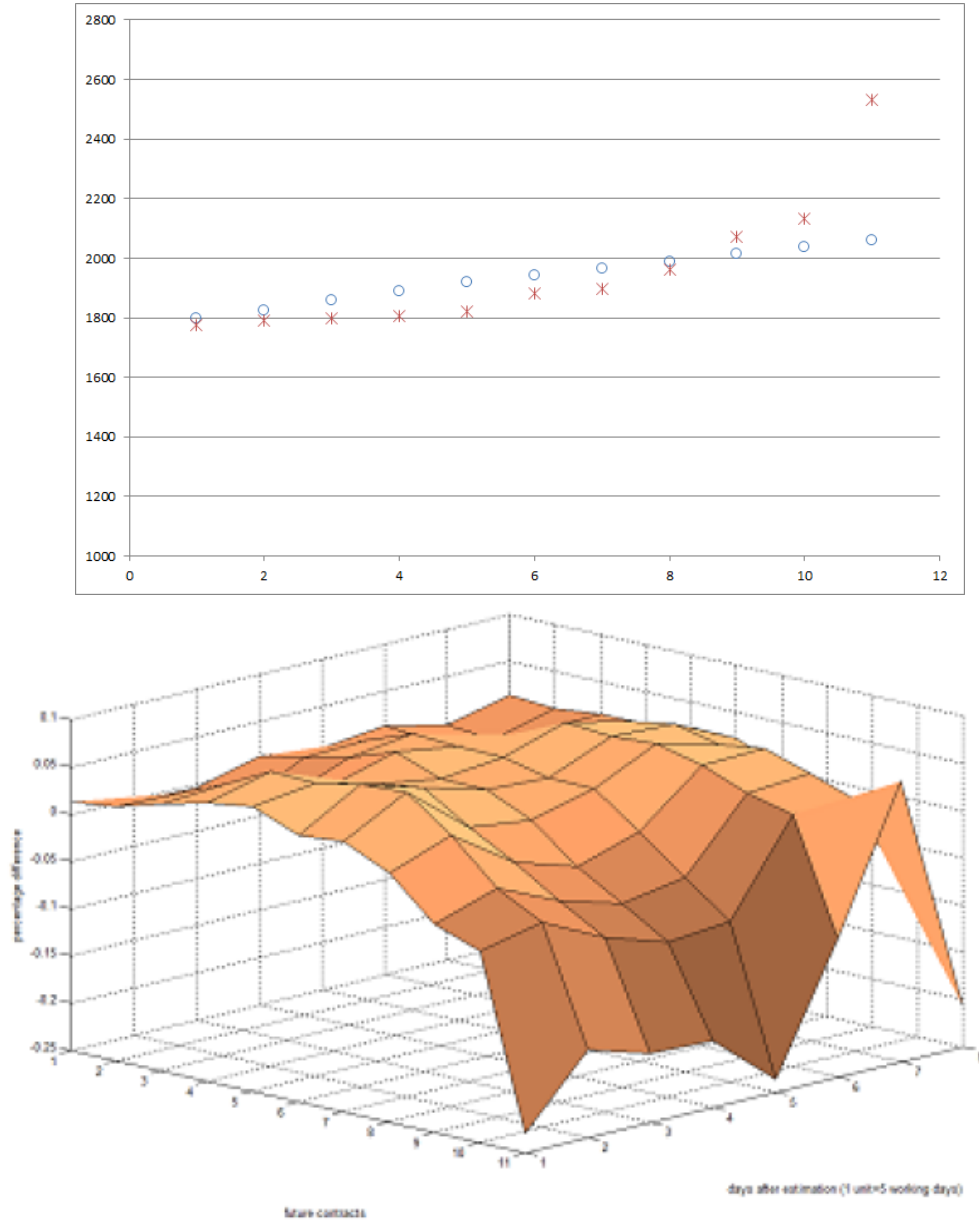
|                 |                  |
|-----------------|------------------|
| <i>Chi2stat</i> | <i>p - value</i> |
| 3.3442          | 0.7646           |

### Theoretical future prices

The followings compare the theoretical and observed future prices.

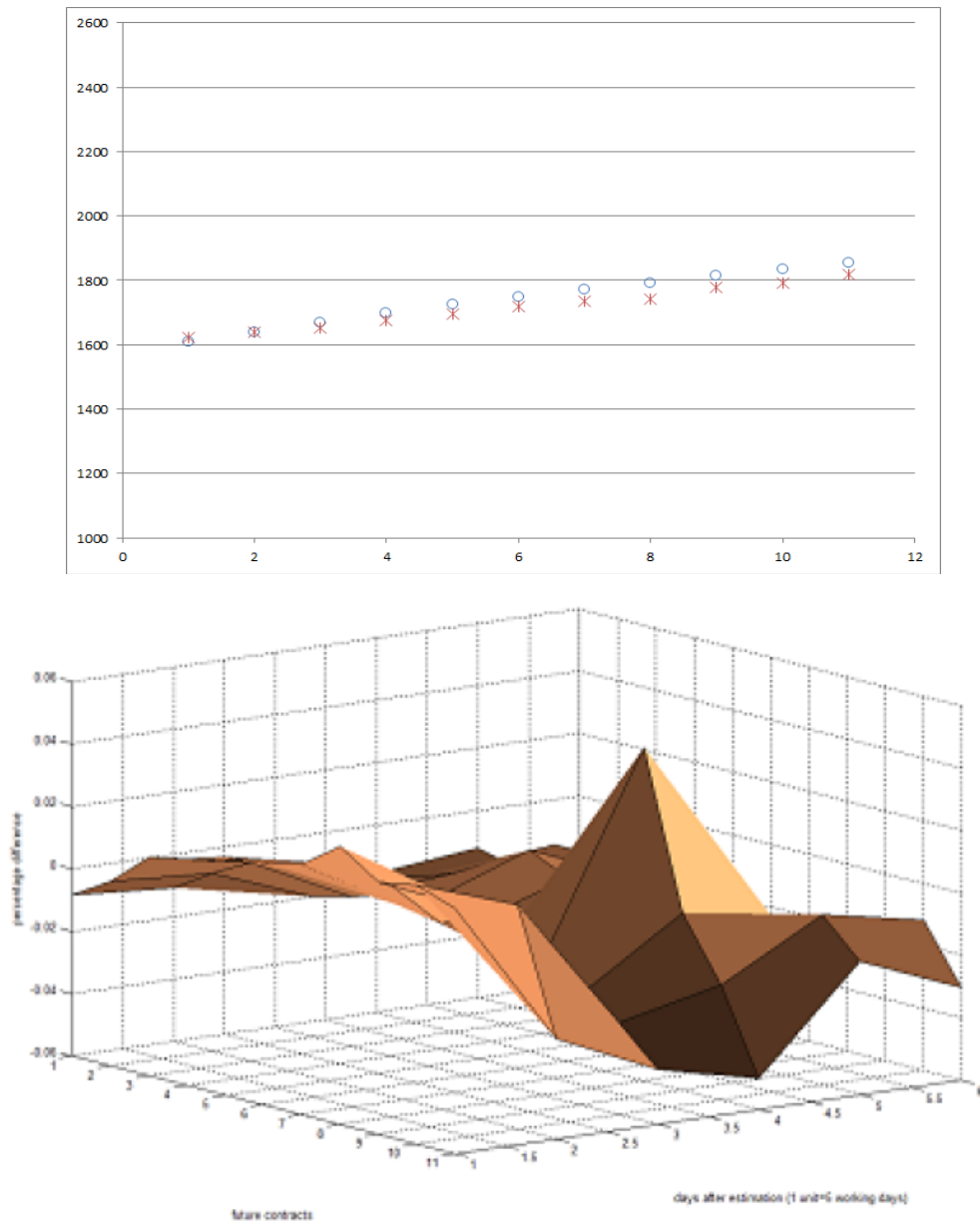


The above are the results of the first method. For the in-the-sample performance, the average of the absolute values of percentage difference is 8.0913% .



The above are the results of the second method. The average of the absolute values of percentage difference is 4.4504%. The calibrated parameters are

| $\kappa$   | $\eta$ | $\phi$ | $\lambda$ | $\alpha$ | $\beta$ | $\mu$   | $\delta$ |
|------------|--------|--------|-----------|----------|---------|---------|----------|
| $9.45e-06$ | 0.0273 | 0.0862 | -0.0862   | 0.6206   | 0.0564  | -0.0560 | 0.6129   |



The above are the results of the third method. The average of the absolute values of percentage difference is 1.6252%.

Not surprisingly, the third method showed for the third time to be the best among all the methods considered.

## 3.6 Conclusion

This research studied three measure-changing approaches and applied them to three commodity future markets: gold, copper and aluminum. The empirical analysis shows that the nonparametric measure-changing method give the best fit to the observed future prices in all the three markets.

In my opinion, the main reason for the relative underperformance of the first approach is that a model fitted from time-series data is probably incapable of capturing cross-section phenomenon. We may change this by modeling the whole futures curve and the stochastic evolution of the curve, just like what we did in the previous chapter to model the mortality intensity.

The reason for the underperformance of the second model is the structure of the spot price dynamics. It would be beneficial to model the convenience yield or the volatility in a different way.

The main purpose of this essay is to quantify the relative performance of the three methods in a market with many liquid futures present, in view that in some of the new financial markets with only very few futures or derivatives available, one has to resort to pricing by optimal martingale measures like the Esscher measure.



# Bibliography

- [1] S. Asmussen, P.W. Glynn (2007). Stochastic simulation: algorithms and analysis. Springer.
- [2] S. Asmussen, J. Rosiński (2001). Approximations of small jumps of Lévy processes with a view towards simulation. Journal of Applied Probability. Volume 38, Number 2 (2001), 482-493.
- [3] K. Burnecki (2009). Simulation of Counting Processes. Presentation slides. [www.im.pwr.wroc.pl/~burnecki](http://www.im.pwr.wroc.pl/~burnecki)
- [4] R. Carmona, M. Ludkovski (2004). Spot Convenience Yield Models for Energy Markets. AMS Mathematics of Finance, Contemporary Mathematics vol. 351, pp. 65–80.
- [5] J. Casassus, P. Collin-Dufresne (2001). 'Maximal' Convenience Yield Model Implied by Commodity Futures. EFA 2002 Berlin Meetings Presented Paper.
- [6] C. Chiarella, M. Craddock, and N. El-Hassan (2000). The calibration of stock option pricing models using inverse problem methodology. QFRQ Research Papers, UTS Sydney.
- [7] R. Cont, P. Tankov (2004). Financial Modelling With Jump Processes. Chapman & Hall/CRC Financial Mathematics Series, Boca Raton.
- [8] S. Galluccio, Y. Le Cam (2005). Implied Calibration of Stochastic Volatility Jump Diffusion Models. Working Paper.
- [9] M. Girth, W.K. Härdle, and M. Schienle (2010). Nonparametric Estimation of Risk-Neutral Densities. SFB 649 Discussion Paper 2010-021.
- [10] J.M. Harrison, D.M. Kreps (1979). Martingale and Arbitrage in Multiperiod Securities Markets. Journal of Economic Theory, 20. 381-408.
- [11] J.M. Harrison, S.R. Pliska (1981). Martingales and Stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11 (3): 215–260.

- [12] J. Jacod and A.N. Shiryaev (2003). Limit Theorems for Stochastic Processes. A Series of Comprehensive Studies in Mathematics. Springer.
- [13] M. Keller-Ressel (2008). Affine processes and applications to stochastic volatility modelling. Presentation slides. Weierstrass Institute, Berlin.
- [14] R. Kiesel (2002). Nonparametric Statistical Methods and the Pricing of Derivative Securities. Journal of Applied Mathematics and Decision Sciences. 6(1), 1-22.
- [15] C. Klüppelberg, A. Lindner, and R. Maller (2004). A Continuous Time GARCH Process Driven by a Lévy process: stationarity and second order behaviour. Journal of Applied Probability. 41(3):601–622.
- [16] R.A. Maller, G. Müller, and A. Szimayer (2008). GARCH modelling in continuous time for irregularly spaced time series data. Bernoulli Vol. 14, No. 2, 519-542.
- [17] S.A. Ross (1997). Hedging Long Run Commitments: Exercises in Incomplete Market Pricing. Economic Notes by Banca Monte 26, 99–132.
- [18] E.S. Schwartz (1997). The Stochastic Behavior of Commodity Prices: Implication for Valuation and Hedging. Journal of Finance. Vol 52, No. 3, 923-973.
- [19] S.E. Shreve (2004). Stochastic Calculus for Finance II: Continuous-Time Models. Springer.

## Chapter 4

# Stochastic Volatility of Volatility

### 4.1 Introduction

Volatility is one of the most important concepts in financial mathematics. Volatility is unobservable so there are different ways to define it. In my point of view, they can be categorized into two groups: volatility measurements determined by the price of the underlying asset and the ones determined by the prices of the derivatives.

Let us firstly look at the volatility defined via the underlying asset price. Here the measurement depends on the parametric model of the asset price. This model can be either discrete or continuous. Popular discrete models include ARCH and GARCH. Famous continuous models include constant instantaneous volatility models (e.g. Black-Scholes model), local volatility (LV) models (e.g. constant elasticity variance (CEV)) and stochastic volatility (SV) models. Swishchuk(2010) [15] summarized some stochastic volatility models in the existing literature:

1. Continuous-time one factor SV models: Ornstein-Uhlenbeck model, Hull and White model, Wiggins model, Scott model, Stein and Stein model, Heston model, etc. Famous discrete-time one factor SV models include the ARV model, stochastic variance model and the ARCH/GARCH family models.
2. Some studies found that one factor is not capable of matching the high conditional kurtosis of returns and the full term structure of implied volatility surface, so a set of generalizations have been proposed. These generalized stochastic volatility models include: 1) Multi-factor SV models; 2) Allow for jumps in the volatility SDE; 3) Discrete and continuous-time long memory SV; 4) Multivariate models: introducing volatility clustering into traditional factor models.

We can use the so-called realized volatility to estimate integrated volatility. Denoting the underlying asset price by  $S$ , according to Hsu and Murray (2007) [10],

$n$ -day realized volatility can be defined as follow

$$RVol_{t,t+n} = 100 \times \sqrt{\frac{365}{n} \sum_{i=1}^n \left[ \ln \left( \frac{S_{t+i}}{S_{t+i-1}} \right) \right]^2}$$

This concept can be extended to the high frequency case. Studies in this area include Andersen and Benzoni (2008) [1], Carr and Lee (2007) [6], Aït-Sahalia et al. (2011) [4], etc. With the ultra high frequency data, we cannot directly use the realized volatility estimator to estimate the integrated volatility due to the microstructure noise. The Two Scales Realized Volatility (TSRV) can solve this problem. More details on TSRV can be found in [4].

The other way to define volatility is via derivatives price. The first attempt to extract volatility directly from derivatives price is the implied volatility calibrated from the option prices based on the Black-Scholes formula. Later on, this concept has been extended to other models.

Dupire (1994) [8] assumed a local volatility model with deterministic volatility and proposes the following formula to calculate the local volatility from option prices:

$$\sigma^2(K, T, S_0) = \frac{\frac{\partial C}{\partial T}}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}}$$

In 1993, the Chicago Board Options Exchange (CBOE) introduced the CBOE Volatility Index (VIX). Ten years later, CBOE together with Goldman Sachs, modified the definition of the VIX to better reflect the expected volatility. The new VIX is based on the S&P 500 Index (SPX) and estimates expected volatility by averaging the weighted prices of SPX puts and calls over a wide range of strike prices ([16]). The general formula to calculate the VIX is ([16]):

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T} \left[ \frac{F}{K_0} - 1 \right]^2$$

where  $\sigma = VIX/100$ ;

$T$  : Time to expiration;

$F$  : Forward index level derived from index option prices;

$K_0$  : First strike below the forward index level,  $F$ ;

$K_i$  : Strike price of  $i$ th out-of-the-money option;

$\Delta K_i$  : Interval between strike prices – half the difference between the strike on either side of  $K_i$ ;

$r$  : Risk-free interest rate to expiration;

$Q(K_i)$  : The midpoint of the bid-ask spread for each option with strike  $K_i$ .

Hsu and Murray (2007) found that a change in the VIX does not predict a change in the 30-day realized volatility of the SPX.

Volatility has been long served as indicator of risk of the underlying asset. Right now there are plenty of financial instruments purely written on volatility. Carr and

Lee (2009) [7] give an overview of the historical development of volatility derivatives. The first liquid volatility derivatives were variance swaps. And then volatility swaps appeared because practitioners prefer to think in terms of volatility rather than variance. Later, more OTC volatility products, like conditional and corridor variance swaps and timer puts and calls, were introduced. In 2003, the CBOE revised the definition of the VIX and then launched VIX futures and options. The VIX options have become the CBOE's most liquid option contract after the SPX index options.

Volatility of volatility is a new area to study the property of volatility. Previous studies include Herath and Kumar (2002) [9], Ingber and Wilson(1999) [11], Kaeck and Alexander (2010) [12]. We consider the two biggest challenges in studying volatility of volatility to be: (i) what is volatility of volatility exactly? (ii) how to estimate it. This essay aims to answer these two questions.

## 4.2 The Concept of Volatility of Volatility

In this study, we define two kinds of volatility of volatility:

1. Volatility of the return's volatility;
2. Volatility of the VIX.

Later on, we will discuss the dynamics of both of them. A very straightforward method to estimate volatility of volatility from the return data or VIX data is to calculate 30-day volatilities using standard deviation and then to calculate the standard deviation of the obtained standard deviation. However, we consider this method not to be feasible on daily data since there are not enough data points available.

Herath and Kumar (2002) [9] proposed a Jackknife estimation method, which includes three steps:

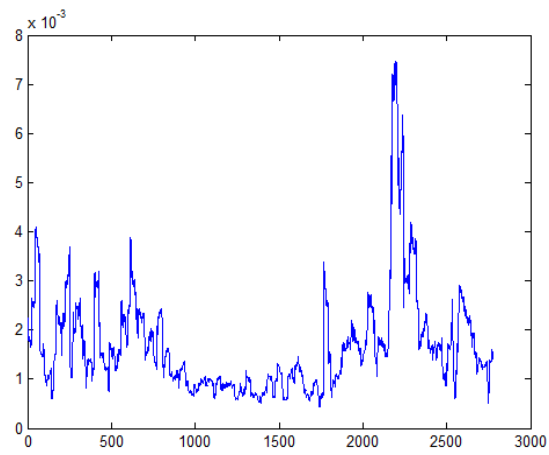
1. Partition the sample  $n$  into  $m$  sub-samples with same size;
2. Calculate a pseudovalue  $\theta_{n-1,j}$ , which is the standard deviation of returns after each observation is omitted;
3. Estimate the volatility of volatility  $\xi$  by

$$\hat{\xi} = \sqrt{\frac{n-1}{n} \sum_{j=1}^n (\theta_{n-1,j} - \bar{\theta}_{n-1})^2}$$

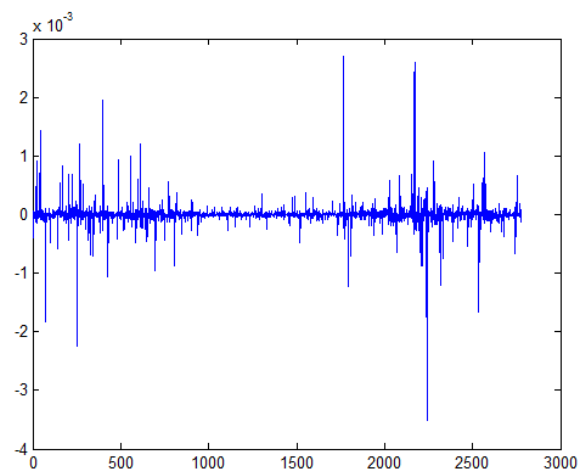
where

$$\bar{\theta}_{n-1} = \sum_{j=1}^n \theta_{n-1,j} / n$$

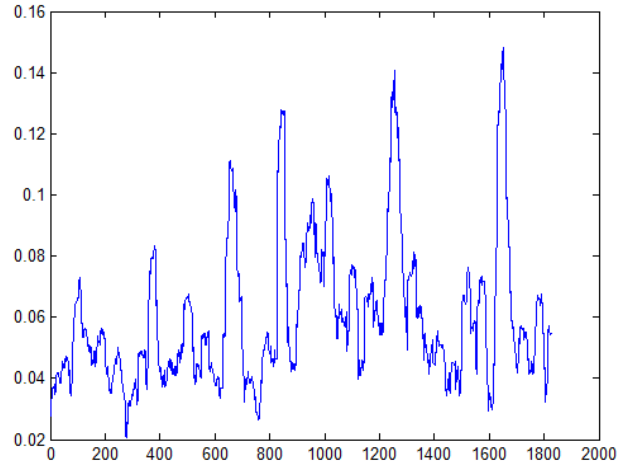
We apply this algorithm to each 30-day period log return (S&P 500) and VIX data and get the following figures. The results support Ingber and Wilson (1999)'s [11] argument that volatility of volatility itself should be stochastic.



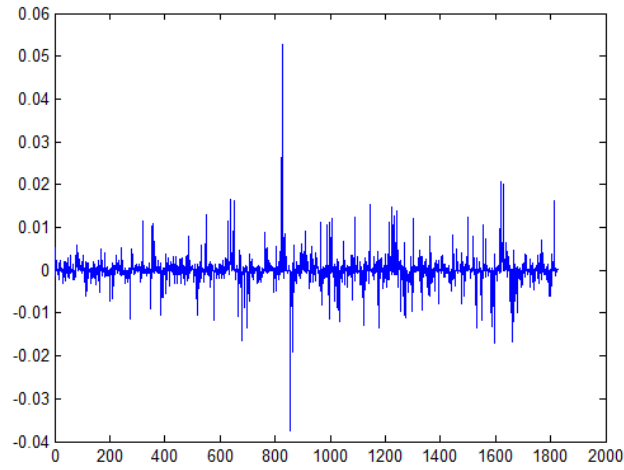
Volatility of volatility of S&amp;P500



The change of volatility of volatility



Volatility of VIX



Change of volatility of VIX

### 4.3 Stochastic Volatility of Volatility Model and Estimation Methods

Similar to volatility, volatility of volatility can be measured in a parametric way. In this section, we assume that the volatility of volatility is stochastic.

For VIX, incorporating stochastic volatility of volatility is nothing but using a stochastic volatility model to describe VIX. SV models are well established and can

be easily fitted to the data. The biggest problem is to incorporate stochastic volatility of volatility to the stock model. The following is an example to incorporate stochastic volatility of volatility into the S&P 500 index model. Assume a 3-SDE model to describe the dynamics of S&P 500 index.  $X, \sqrt{V}, \sqrt{\tilde{V}}$  represent the return process, the volatility process and the volatility of volatility process. Inspired by Kaeck and Alexander (2010), we adopt the process  $J$  to model the jump time and  $Z$  to model the jump size.

$$\begin{aligned} dX_t &= \kappa(\theta - X_{t-})dt + \sqrt{V_t}dW_t^X + Z_t J_t, \\ dV_t &= \kappa_V(\theta_V - V_t)dt + \tilde{V}_t\sqrt{V_t}dW_t^V, \\ d\tilde{V}_t &= \kappa_{\tilde{V}}(\theta_{\tilde{V}} - \tilde{V}_t)dt + \sigma_{\tilde{V}}\sqrt{\tilde{V}_t}dW_t^{\tilde{V}}. \end{aligned}$$

For simplicity, assume the following relations between  $W^X, W^V$  and  $W^{\tilde{V}}$ .

$$\begin{aligned} \langle W^X, W^V \rangle &= 0, \\ \langle W^X, W^{\tilde{V}} \rangle &= 0, \\ \langle W^V, W^{\tilde{V}} \rangle &= 0. \end{aligned}$$

Stochastic volatility of the VIX can be defined by modelling VIX dynamics as a stochastic volatility process. The biggest challenge to apply those models is to estimate the parameters. We consider three methods: Markov Chain Monte Carlo (MCMC), Maximum likelihood estimation via closed-form likelihood expansion and Calibration methods.

### 4.3.1 Bayesian MCMC

The idea of Bayesian statistics is to consider both the observations  $\mathbf{x}$  and the parameter  $\theta$  as random variables. Here, there are two key concepts of the distribution of the parameter. One is the *prior distribution*  $\pi(\theta)$ , which represents the prior information about the distribution of parameter; the other is the *posterior distribution*  $\pi(\theta|\mathbf{x})$ , which represents the distribution updated by the information given by the observations.

The posterior distribution can be calculated by the Bayes' formula

$$\pi(\theta|\mathbf{x}) = \frac{L(\mathbf{x}|\theta)\pi(\theta)}{\int L(\mathbf{x}|\zeta)\pi(\zeta)d\zeta}.$$

And then we can simulate samples from the posterior distribution and obtain the Bayesian estimate from the sample mean. The simulation procedure is typically done by the Markov Chain Monte Carlo (MCMC) approach.

**Definition 16** (Robert and Casella (1999) [14], page 142) *Given a transition kernel*

$K$ , a sequence  $X_0, X_1, \dots, X_n, \dots$  of random variables is a Markov chain, denoted by  $(X_n)$ , if, for any  $t$ , the conditional distribution of  $X_t$  given  $x_{t-1}, x_{t-2}, \dots, x_0$  is the same as the distribution of  $X_{t-1}$  given  $x_{t-1}$ ; that is,

$$\begin{aligned} P(X_{k+1} \in A | x_0, x_1, \dots, x_k) &= P(X_{k+1} \in A | x_k), \\ &= \int_A K(x_k, dx). \end{aligned}$$

**Definition 17** (Robert and Casella (1999) [14], page 231) A Markov Chain Monte Carlo (MCMC) method for the simulation of a distribution  $f$  is any method producing an ergodic Markov chain  $(X^{(t)})$  whose stationary distribution is  $f$ .

Here comes a problem: how to simulate a Markov chain having stationary distribution, given a distribution with density or probability mass function? A commonly used method is the Metropolis-Hastings algorithm.

**Algorithm 18** Metropolis-Hastings algorithm ([14], page 233)

Given  $x^{(t)}$ ,

1. Generate  $Y_t \sim q(y|x^{(t)})$ .

2. Take

$$X^{(t+1)} = \begin{cases} Y_t & \text{with probability } \rho(x^{(t)}, Y_t), \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, Y_t), \end{cases}$$

where

$$\rho(x^{(t)}, Y_t) = \min \left\{ \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right\}.$$

The above is a general Metropolis-Hastings algorithm. In practice, one can use a random walk Metropolis-Hastings algorithm.

**Algorithm 19** Random walk Metropolis-Hastings algorithm ([14], page 245)

Given  $x^{(t)}$ ,

1. Generate  $Y_t \sim g(y - x^{(t)})$ .

2. Take

$$X^{(t+1)} = \begin{cases} Y_t & \text{with probability } \min \left\{ \frac{f(Y_t)}{f(x^{(t)})}, 1 \right\}, \\ x^{(t)} & \text{otherwise.} \end{cases}$$

Another widely used algorithm is the Gibbs Sampler. According to Robert and Casella (1999), the Gibbs sampling method is equivalent to the composition of p Metropolis-Hastings algorithms, with acceptance probabilities uniformly equal to 1.

**Algorithm 20** The Gibbs Sampler ([14], page 285)

Given  $\mathbf{x}^{(t)} = (x_1^{(t)}, \dots, x_p^{(t)})$ , generate

1.  $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \dots, x_p^{(t)})$ ;

$$2. X_2^{(t+1)} \sim f_2 \left( x_2 | x_1^{(t+1)}, \dots, x_p^{(t)} \right);$$

...

$$3. X_p^{(t+1)} \sim f_1 \left( x_p | x_1^{(t+1)}, \dots, x_{p-1}^{(t+1)} \right);$$

### Prior and posterior distributions

In our case, we can adopt the following Gibb's sampler.

$$\begin{aligned} & P \left( Z^{(g)} \mid \tilde{V}^{(g)}, V^{(g)}, J^{(g-1)}, \Theta^{(g-1)}, X \right) \\ & P \left( J^{(g)} \mid \tilde{V}^{(g)}, V^{(g)}, Z^{(g)}, \Theta^{(g-1)}, X \right) \\ & P \left( \tilde{V}^{(g)} \mid V^{(g-1)}, Z^{(g-1)}, J^{(g-1)}, \Theta^{(g-1)}, X \right) \\ & P \left( V^{(g)} \mid \tilde{V}^{(g)}, Z^{(g-1)}, J^{(g-1)}, \Theta^{(g-1)}, X \right) \\ & P \left( \Theta^{(g)} \mid \tilde{V}^{(g)}, V^{(g)}, Z^{(g)}, J^{(g)}, X \right) \end{aligned}$$

This section specifies the prior distribution for parameters and derives the corresponding posterior distribution. It is often more convenient to use distributions from *conjugate families*: the posterior distribution belongs to the same parametric family as the prior (Asmussen and Glynn (2000) [5]).

The prior and posterior distributions of the parameters are (some priors are inspired by [12]):

1.  $\kappa$ : prior distribution  $N(\mu_\kappa, \sigma_\kappa^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned} A &= \frac{1}{\sigma_\kappa^2} + \sum_i \frac{(\theta - X_{i-1})^2 h}{V_{i-1}} \\ B &= \frac{\mu_\kappa}{\sigma_\kappa^2} + \sum_i \frac{(\theta - X_{i-1})(X_i - X_{i-1} - Z_i J_i)}{V_{i-1}} \end{aligned}$$

**Proof.** The posterior distribution is obtained by the following way.

∴

$$\begin{aligned} \kappa &\sim N(\mu_\kappa, \sigma_\kappa^2) \\ X_i - X_{i-1} &\sim N(\kappa(\theta - X_{i-1})h + Z_i J_i, hV_{i-1}) \end{aligned}$$

∴

$$\begin{aligned}
 P(\kappa|X) &\propto \exp \left\{ -\frac{1}{2} \left[ \left( \frac{\kappa - \mu_\kappa}{\sigma_\kappa} \right)^2 + \left( \frac{X_i - X_{i-1} - \kappa(\theta - X_{i-1})h - Z_i J_i}{\sqrt{hV_{i-1}}} \right)^2 \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} [\kappa^2 A + 2\kappa B] \right\}
 \end{aligned}$$

■

2.  $\theta$ : prior distribution  $N(\mu_\theta, \sigma_\theta^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned}
 A &= \frac{1}{\sigma_\theta^2} + \sum_i \frac{\kappa^2 h}{V_{i-1}} \\
 B &= \frac{\mu_\theta}{\sigma_\theta^2} + \sum_i \frac{\kappa(X_i - X_{i-1} - \kappa h X_{i-1} - Z_i J_i)}{V_{i-1}}
 \end{aligned}$$

3.  $\kappa_V$ : prior distribution  $N(\mu_{\kappa_V}, \sigma_{\kappa_V}^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned}
 A &= \frac{1}{\sigma_{\kappa_V}^2} + \sum_i \frac{(\theta_V - V_{i-1})^2 h}{\tilde{V}_{i-1}} \\
 B &= \frac{\mu_{\kappa_V}}{\sigma_{\kappa_V}^2} + \sum_i \frac{(\theta_V - X_{i-1})(V_i - V_{i-1})}{\tilde{V}_{i-1}}
 \end{aligned}$$

4.  $\theta_V$ : prior distribution  $N(\mu_{\theta_V}, \sigma_{\theta_V}^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned}
 A &= \frac{1}{\sigma_{\theta_V}^2} + \sum_i \frac{\kappa_V^2 h}{\tilde{V}_{i-1}} \\
 B &= \frac{\mu_{\theta_V}}{\sigma_{\theta_V}^2} + \sum_i \frac{\kappa_V(V_i - V_{i-1} - \kappa_V h V_{i-1})}{\tilde{V}_{i-1}}
 \end{aligned}$$

5.  $\kappa_{\tilde{V}}$ : prior distribution  $N(\mu_{\kappa_{\tilde{V}}}, \sigma_{\kappa_{\tilde{V}}}^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned}
 A &= \frac{1}{\sigma_{\kappa_{\tilde{V}}}^2} + \sum_i \frac{(\theta_{\tilde{V}} - \tilde{V}_{i-1})^2 h}{\tilde{V}_{i-1} \sigma_{\tilde{V}}^2} \\
 B &= \frac{\mu_{\kappa_{\tilde{V}}}}{\sigma_{\kappa_{\tilde{V}}}^2} + \sum_i \frac{(\theta_{\tilde{V}} - \tilde{V}_{i-1})(\tilde{V}_i - \tilde{V}_{i-1})}{\tilde{V}_{i-1} \sigma_{\tilde{V}}^2}
 \end{aligned}$$

6.  $\theta_{\tilde{V}}$ : prior distribution  $N(\mu_{\theta_{\tilde{V}}}, \sigma_{\theta_{\tilde{V}}}^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$A = \frac{1}{\sigma_{\theta_{\tilde{V}}}^2} + \sum_i \frac{\kappa_V^2 h}{\tilde{V}_{i-1} \sigma_V^2}$$

$$B = \frac{\mu_{\theta_{\tilde{V}}}}{\sigma_{\theta_{\tilde{V}}}^2} + \sum_i \frac{\kappa_{\tilde{V}} (\tilde{V}_i - \tilde{V}_{i-1} - \kappa_{\tilde{V}} h \tilde{V}_{i-1})}{\tilde{V}_{i-1} \sigma_V^2}$$

7.  $\sigma_{\tilde{V}}$  can be simulated by random walk Metropolis algorithm.

8.  $J$ : prior  $Bernoulli(h\lambda_0)$ , posterior  $Bernoulli\left(\frac{A}{A+B}\right)$

$$A = h\lambda_0 \exp \left[ -\frac{1}{2} \left[ \frac{X_i - X_{i-1} - \kappa(\theta - X_{i-1})h - Z_i}{\sqrt{hV_{i-1}}} \right]^2 \right]$$

$$B = (1 - h\lambda_0) \exp \left[ -\frac{1}{2} \left[ \frac{X_i - X_{i-1} - \kappa(\theta - X_{i-1})h}{\sqrt{hV_{i-1}}} \right]^2 \right]$$

**Proof.** The posterior distribution is obtained by the following way:

$\therefore$

$$f(X|J) \propto \exp \left\{ -\frac{1}{2} \left[ \frac{X_i - X_{i-1} - \kappa(\theta - X_{i-1})h - Z_i J_i}{\sqrt{hV_{i-1}}} \right]^2 \right\}$$

$\therefore$

$$g(J|X) = \frac{f(X|J)g(J)}{\sum f(X|J)g(J)}$$

$$= \left( \frac{A}{A+B} \right)^J \left( 1 - \frac{A}{A+B} \right)^{1-J}$$

■

9.  $h\lambda_0$ : prior distribution  $Beta(\alpha_0, \beta_0)$ , posterior distribution  $Beta(A, B)$ .

$$A = \alpha_0 + \sum_i J_i$$

$$B = \beta_0 + n - \sum_i J_i$$

**Proof.** The posterior distribution is obtained by the following way:

$$f(h\lambda_0|J) \propto f(J|h\lambda_0)f(h\lambda_0)$$

$$\propto (h\lambda_0)^{\sum_i J_i} (1 - h\lambda_0)^{n - \sum_i J_i} (h\lambda_0)^{\alpha_0 - 1} (1 - h\lambda_0)^{\beta_0 - 1}$$

$$= (h\lambda_0)^{\alpha_0 + \sum_i J_i - 1} (1 - h\lambda_0)^{\beta_0 + n - \sum_i J_i - 1}$$

■

10.  $Z$ : prior distribution  $N(\mu_J, \sigma_J^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned} A &= \frac{1}{\sigma_J^2} + \sum_i \frac{J_i}{hV_{i-1}} \\ B &= \frac{\mu_J}{\sigma_J^2} + \sum_i \frac{J_i(X_i - X_{i-1} - \kappa h(\theta - X_{i-1}))}{hV_{i-1}} \end{aligned}$$

11.  $\mu_J$ : prior distribution  $N(\mu_{\mu_J}, \sigma_{\mu_J}^2)$ , posterior distribution  $N(B/A, \sqrt{1/A})$ .

$$\begin{aligned} A &= \frac{1}{\sigma_{\mu_J}^2} + \frac{n}{\sigma_J^2} \\ B &= \frac{\mu_{\mu_J}}{\sigma_{\mu_J}^2} + \sum_i \frac{Z_i}{\sigma_J^2} \end{aligned}$$

12.  $\sigma_J^2$ : prior distribution  $InvGamma(\alpha, \beta)$ , posterior distribution  $InvGamma(A, B)$ .

$$\begin{aligned} A &= \alpha + \frac{n}{2} \\ B &= \beta + \sum_i \frac{(Z_i - \mu_J)^2}{2} \end{aligned}$$

**Proof.** The posterior distribution is obtained in the following way:

$$\begin{aligned} f(\sigma_J^2 | Z) &\propto f(Z | \sigma_J^2) f(\sigma_J^2) \\ &\propto \frac{1}{(\sigma_J^2)^{n/2}} \exp\left(-\frac{1}{2} \sum_i \frac{(Z_i - \mu_J)^2}{\sigma_J^2}\right) (\sigma_J^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma_J^2}\right) \\ &= (\sigma_J^2)^{-(\alpha + \frac{n}{2})-1} \exp\left(-\frac{\beta + \sum_i \frac{(Z_i - \mu_J)^2}{2}}{\sigma_J^2}\right) \end{aligned}$$

■

13. Update  $V$  and  $\tilde{V}$ :

$$\begin{aligned} P(\tilde{V}_i | \tilde{V}_{-i}, X, V, Z, J, \Theta) \\ P(V_i | V_{-i}, X, \tilde{V}, Z, J, \Theta) \end{aligned}$$

The appendix shows the result of MCMC on the simulated data for a reduced model.

$$\begin{aligned} dX_t &= \kappa X_t dt + \sqrt{V_t} dW_t^X, \\ dV_t &= \kappa_V (\theta_V - V_t) dt + V_t \tilde{V}_t dW_t^V, \\ d\tilde{V}_t &= \kappa_{\tilde{V}} (\theta_{\tilde{V}} - \tilde{V}_t) dt + \sigma \tilde{V}_t dW_t^{\tilde{V}}. \end{aligned}$$

### Generating random variables in C++

MCMC algorithm needs to simulate random variables from different distributions. Commonly used distributions include normal, truncated normal, beta and inverse gamma distributions. In this study, we adopted the C function on [13]: use 'ran1' to generate uniform random variables, 'gasdev' to generate normal random variables and 'expdev' to generate exponential random variables.

In the following part, I will show the code for the inverse Gamma distribution. Let us firstly recall that the probability density function of inverse Gamma distribution is

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

The easiest way to simulate inverse Gamma variables is to firstly simulate a Gamma random variable and use the following property.

$$\text{If } X \sim \text{Gamma}\left(\alpha, \frac{1}{\beta}\right), \text{ then } \frac{1}{X} \sim \text{InvGamma}(\alpha, \beta)$$

The algorithm to generate a Gamma random variable is recorded in Robert and Casella (1999), (Page 47 and 55). Here 'betarv' is the function for beta random variables.

```
//Inverse Gamma
//Reference: C.P.Robert and G.Casella 'Monte Carlo Statistical Methods'
double invgamma(double aa, double bb, long seed)
{
    double y,z,gamma;
    double b,c,u,v,w;

    //generate Gamma(aa,1)
    if (aa<1)
    {
        y=betarv(aa,1-aa,seed); //betarv generates beta r.v.
        z=expdev(&seed);
        gamma=y*z;
    }//if

    if (aa==1) gamma=expdev(&seed);
    if (aa>1) {
        b=aa-1;
        c=(12*aa-3)/4;

        for(;;){
            u=ran1(&seed);
            v=ran1(&seed);
```

```

w=u*(1-u);
y=sqrt(c/w)*(u-0.5);
gamma=b+y;

if(gamma<=0) continue;
z=64*v*v*w*w*w;

if(z<=1-2*y*y/gamma || 2*(b*log(gamma/b)-y)>=log(z)) break;
} //for
} //else
gamma/=bb;
//make it inverse gamma (aa,bb)
return 1/gamma;
} //function

```

#### 4.3.2 Maximum likelihood estimation via closed-form likelihood expansion

Aït-Sahalia (2008) [2] proposed a method to find out the explicit form of a log-likelihood function, which is the key to the maximum likelihood estimation. Let the log likelihood function be approximated by a power series in the time interval  $\Delta$ :

$$l_X^{(J)}(\Delta, x|x_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(-1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^J C_X^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!}$$

The coefficients  $C_X^{(k)}$  solve the equation

$$f_X^{(k-1)}(x|x_0) = 0$$

where

$$\begin{aligned}
f_X^{(-2)}(x|x_0) &= -2C_X^{(-1)}(x|x_0; \theta) - \sum_{i,j} v_{i,j}(x) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_j} \\
f_X^{(-1)}(x|x_0) &= -G_X^{(0)}(x|x_0; \theta) - \sum_{i,j} v_{i,j}(x) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_j} \\
\text{for } k &\geq 1 \\
f_X^{(k-1)}(x|x_0) &= C_X^{(k)}(x|x_0; \theta) - \frac{1}{k} \sum_{i,j} v_{i,j}(x) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(k)}(x|x_0; \theta)}{\partial x_j} \\
&\quad - G_X^{(k)}(x|x_0; \theta)
\end{aligned}$$

The functions  $G_X^{(k)}(x|x_0; \theta)$  are specified by Aït-Sahalia (2008).

To better determine the coefficients  $C_X^{(k)}$ , one may consider the expansion  $C_X^{(j_k, k)}$

in  $(x - x_0)$  of each coefficient  $C_X^{(k)}$ . Let  $i = (i_1, i_2, \dots, i_m)$  denote a vector of integers and define  $I_k = \{i = (i_1, i_2, \dots, i_m) \in \mathbb{N}^m : 0 \leq \text{tr}[i] \leq j_k\}$  so that the form of  $C_X^{(j_k, k)}$  is

$$C_X^{(j_k, k)}(x|x_0) = \sum_{i \in I_k} \beta_i^{(k)}(x_0) (x_1 - x_{01})^{i_1} \dots (x_m - x_{0m})^{i_m}$$

The coefficients  $\beta_i^{(k)}(x_0)$  are determined by setting the expansion  $f_X^{(j_k, k-1)}$  of  $f_X^{(k-1)}$  to zero. In particular, when  $\text{tr}[i] = 2$ , we obtain the equation

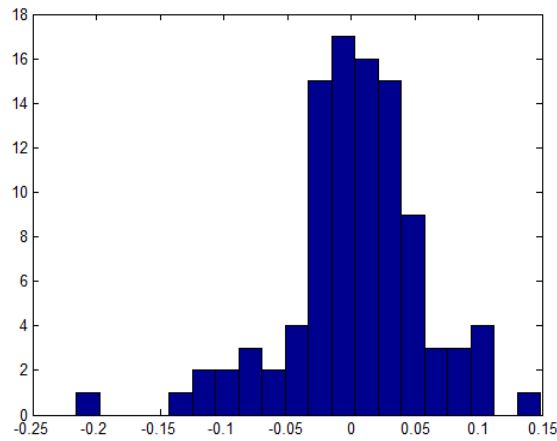
$$\sum_{\text{tr}[i]=2} \beta_i^{(-1)}(x_0) (x_1 - x_{01})^{i_1} \dots (x_m - x_{0m})^{i_m} = -\frac{1}{2} (x - x_0)^T v^{-1}(x_0) (x - x_0)$$

where  $v(x)$  is the infinitesimal variance-covariance matrix of the process.

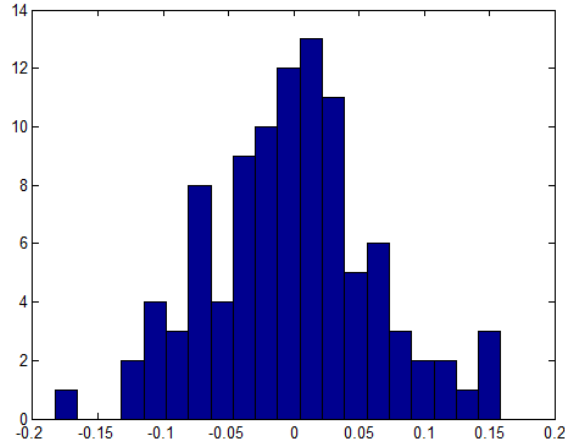
### Fit Heston model to VIX and VIX option data

We applied this method to fit a Heston model to the VIX data between 03 Jan 2011 and 15 Apr 2011 with Black-Scholes implied volatility exacted from in-the-money VIX option as the proxy volatility. The following figures compare the histogram of the realized increments of the log VIX series and the increments determined by the fitted Heston model. It is obvious that they do not share the same properties.

We experienced difficulty to apply this method to the SVV model for S&P500 due to the huge number of equations need to solve.



realized increments of the log VIX series



increments determined by the fitted Heston model

### 4.3.3 Calibration

The third method we have tried is the calibration method. This technique can be implemented to all the models considered.

The object of calibration is to find out the optimal parameters for the underlying asset model which can best match the observed derivative prices. Mathematically the idea can be expressed as:

$$\min_{\Theta} \|C^{theoretical}(\Theta) - C^{real}\|$$

### VIX and VIX option

Let us firstly look at VIX and VIX derivatives. Kaeck and Alexander (2010) has found an improvement of the model for VIX after incorporating the volatility of volatility risk. But there is still a question left: what kind of structure the dynamic of the volatility of VIX should take? We consider the following models for log VIX:

1. Jump volatility of volatility model:

$$\begin{aligned} d\ln(VIX)_t &= (a - b\ln(VIX)_t)dt + \sqrt{\tilde{V}_t}dW_t \\ d\tilde{V}_t &= (\tilde{a} - \tilde{b}\tilde{V}_t)dt + \sqrt{\tilde{V}_t}dJ_t \\ J_t &\sim NIG(\alpha, \beta, \mu, \delta) \end{aligned}$$

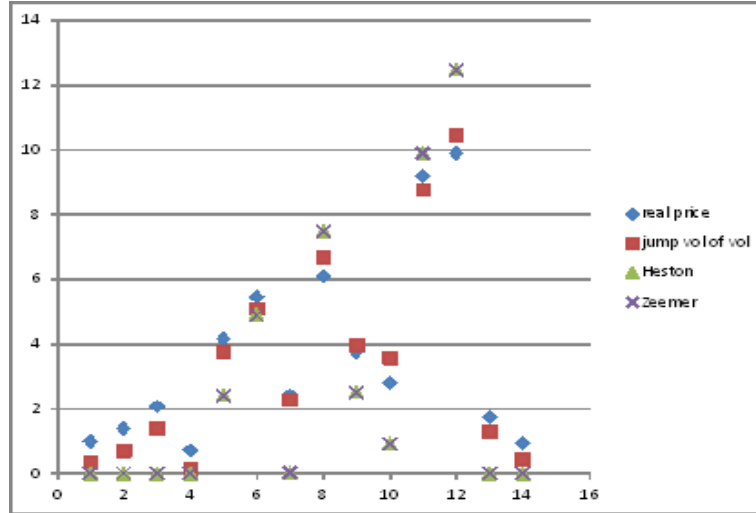
2. Heston model:

$$\begin{aligned} d\ln(VIX)_t &= (a - b\ln(VIX)_t)dt + \sqrt{\tilde{V}_t}dW_t^{VIX} \\ d\tilde{V}_t &= (\tilde{a} - b\tilde{V}_t)dt + \sigma\sqrt{\tilde{V}_t}dW_t^{\tilde{V}} \end{aligned}$$

3. Zeeman's market model:

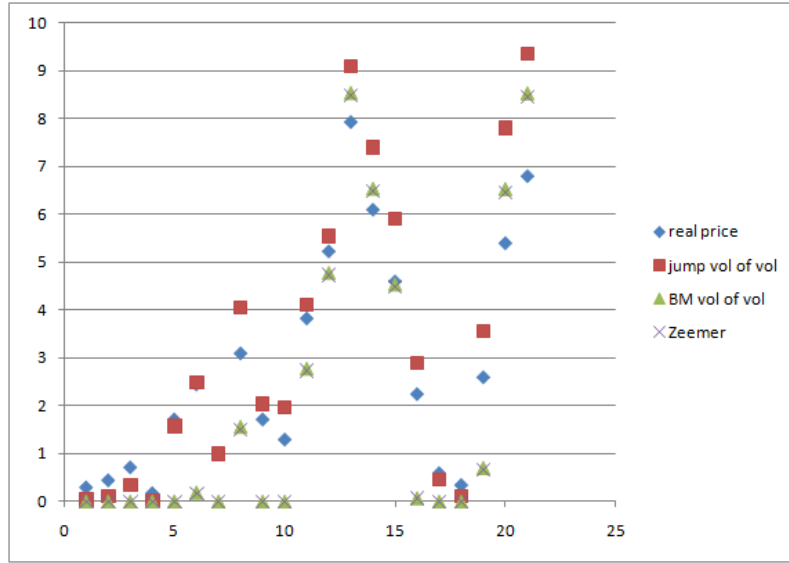
$$\begin{aligned} d\ln(VIX)_t &= (a - b\ln(VIX)_t)dt + \sqrt{\tilde{V}_t}dW_t^{VIX} \\ d\tilde{V}_t &= (\tilde{a} - b\tilde{V}_t + c\tilde{V}_t^3)dt + \sigma\sqrt{\tilde{V}_t}dW_t^{\tilde{V}} \end{aligned}$$

We calibrated all the three models to the VIX options with different strikes and maturities. The following figures show the realized VIX option prices in 03 Jan 2011 and theoretical VIX option prices obtained by calibrated models. A clear improvement can be seen if we model the volatility of the log VIX as a jump process, especially for out-of-the-money options.



realized and theoretical option prices in 04 Jan 2011

Now let us try another date, 04 Feb 2011. The jump volatility of volatility model still outperforms the other two models when price the out-of-the-money options, however underperforms when price highly in-the-money options.



### SPX and SPX option

We consider the following SVV models for S&P500 series.

1. Jump stochastic volatility of volatility model:

$$\begin{aligned}
 d \ln S_t &= (a - b \ln S_t) dt + V_t dW_t^S \\
 dV_t &= (a^V - b^V V_t) dt + \tilde{V}_t dW_t^V \\
 d\tilde{V}_t &= (a^{\tilde{V}} - b^{\tilde{V}} \tilde{V}_t) dt + \sigma \tilde{V}_t dJ_t \\
 J_t &\sim NIG(\alpha, \beta, \mu, \delta), \quad \langle W^X, W^V \rangle = \rho
 \end{aligned}$$

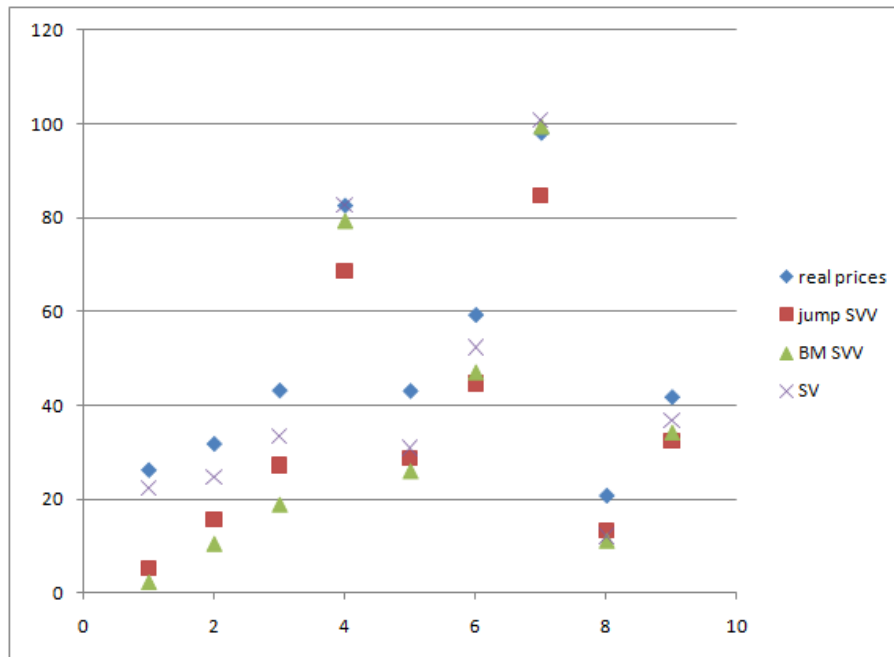
2. BM stochastic volatility of volatility model:

$$\begin{aligned}
 d \ln S_t &= (a - b \ln S_t) dt + V_t dW_t^S \\
 dV_t &= (a^V - b^V V_t) dt + \tilde{V}_t dW_t^V \\
 d\tilde{V}_t &= (a^{\tilde{V}} - b^{\tilde{V}} \tilde{V}_t) dt + \sigma \tilde{V}_t dW_t^{\tilde{V}} \\
 J_t &\sim NIG(\alpha, \beta, \mu, \delta)
 \end{aligned}$$

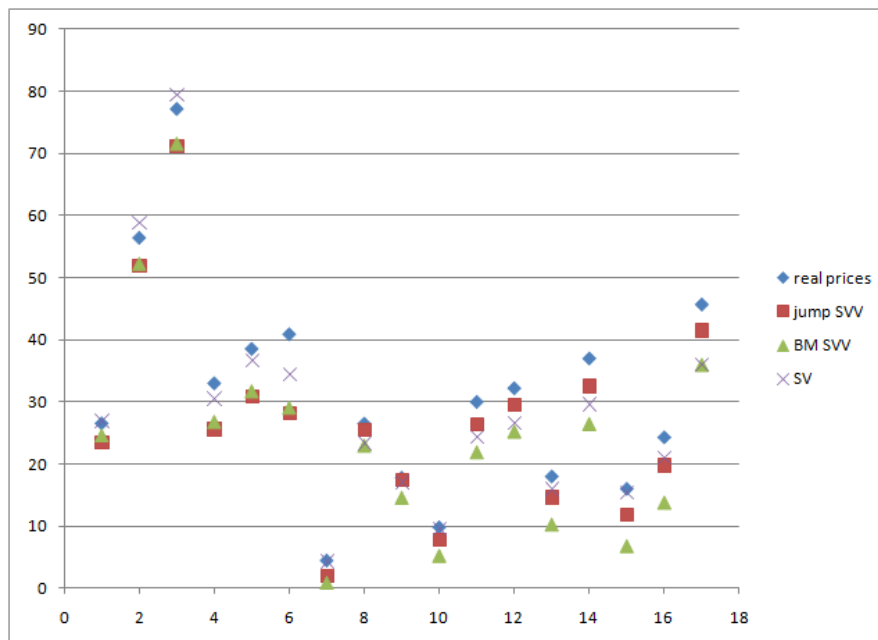
3. Stochastic volatility model:

$$\begin{aligned}
 d \ln S_t &= (a - b \ln S_t) dt + V_t dW_t^S \\
 dV_t &= (a^V - b^V V_t) dt + \sigma V_t dW_t^V
 \end{aligned}$$

Surprisingly, we found out that the adding stochastic volatility of volatility worsens the option pricing.



realized and theoretical option prices in 04 Jan 2011



realized and theoretical option prices in 18 Feb 2011

## 4.4 Conclusion

This essay studied two kinds of volatility of volatility: i) the volatility of the volatility of stock return; ii) the volatility of VIX. The stochastic volatility of volatility model for S&P500 index involves three SDEs whereas the SVV model for VIX is in fact a SV model.

The essay also tried three methods to fit SVV models to real data: i) Markov Chain Monte Carlo; ii) Maximum likelihood estimation via closed-form likelihood expansion; iii) Calibration. The last method was eventually worked out for all the models and reveals the influence of incorporating volatility of volatility risk on the option evaluation. Surprisingly, the SVV model fails to improve the efficiency of option pricing significantly.



# Bibliography

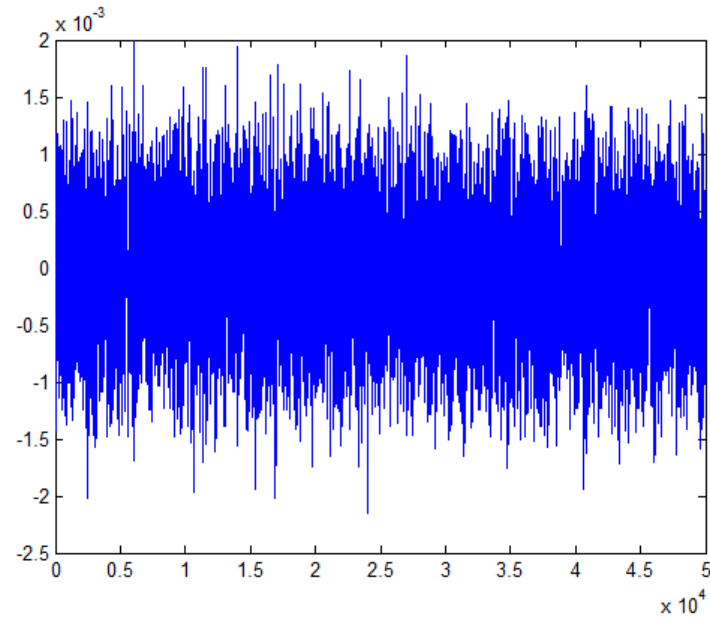
- [1] T.G. Andersen, L. Benzoni (2008). Realized Volatility. Working Paper.
- [2] Y. Aït-Sahalia (2008). Closed-Form Likelihood Expansions for Multivariate Diffusions. *Annals of Statistics*, 36, 906-937.
- [3] Y. Aït-Sahalia, R. Kimmel (2007). Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics*, 83, 413-452.
- [4] Y. Aït-Sahalia, L. Zhang and P. Mykland (2011). Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. *Journal of Econometrics*, 160, 190-203.
- [5] S. Asmussen, P.W. Glynn (2007). *Stochastic simulation: algorithms and analysis*. Springer.
- [6] P. Carr, R. Lee (2007). Realized Volatility and Variance: Options via Swaps. *Risk*. May.
- [7] P. Carr, R. Lee (2009). Volatility Derivatives. *Annual Review of Financ. Econ.* 1:1–21.
- [8] B. Dupire. (1994). Pricing with a Smile. *Risk Magazine*.
- [9] H.S.B. Herath, P. Kumar (2002). The Jackknife Estimator for Estimating the Volatility of Volatility of a Stock. *Corporate Finance Review*. Vol 7, No. 3, 13-21.
- [10] S.D.H. Hsu, B.M. Murray (2007). On the volatility of volatility. *Physica A: Statistical Mechanics and its Applications*, vol. 380. 366-376.
- [11] L. Ingber, J.K. Wilson (1999). Volatility of volatility of financial markets, *Mathl. Computer Modelling* 29, 39-57.
- [12] A. Kaeck, C. Alexander (2010). VIX Dynamics with Stochastic Volatility of Volatility. Working Paper.
- [13] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition. Cambridge University Press.

- [14] C.P. Robert and G. Casella (1999). Monte Carlo Statistical Methods. Springer text in Statistics.
- [15] A. Swishchuk (2010). Stochastic Volatility and Change of Time: Overview. Presentation Slides.
- [16] CBOE (2009). The CBOE Volatility Index - VIX. [www.cboe.com/micro/vix/vixwhite.pdf](http://www.cboe.com/micro/vix/vixwhite.pdf)

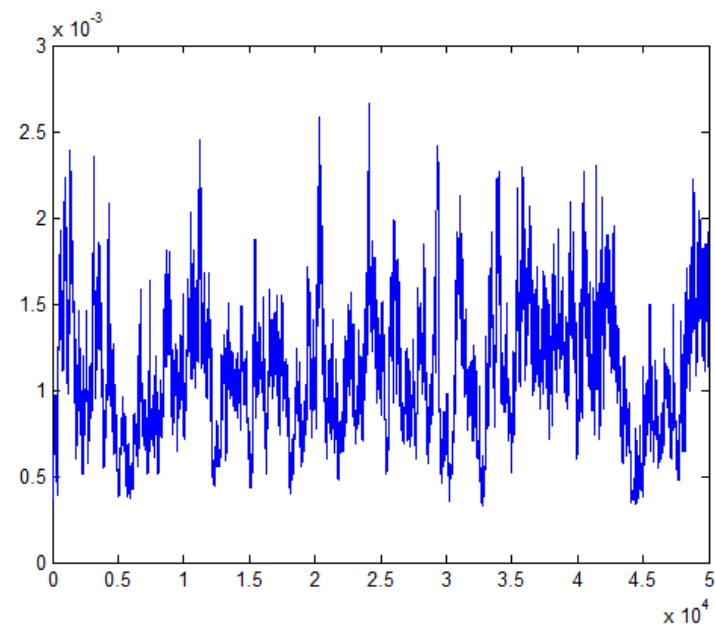
## Appendix

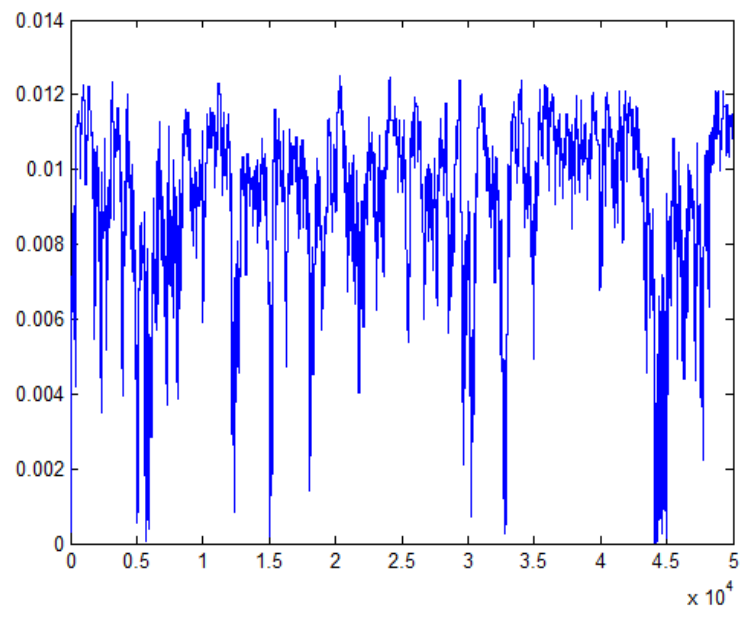
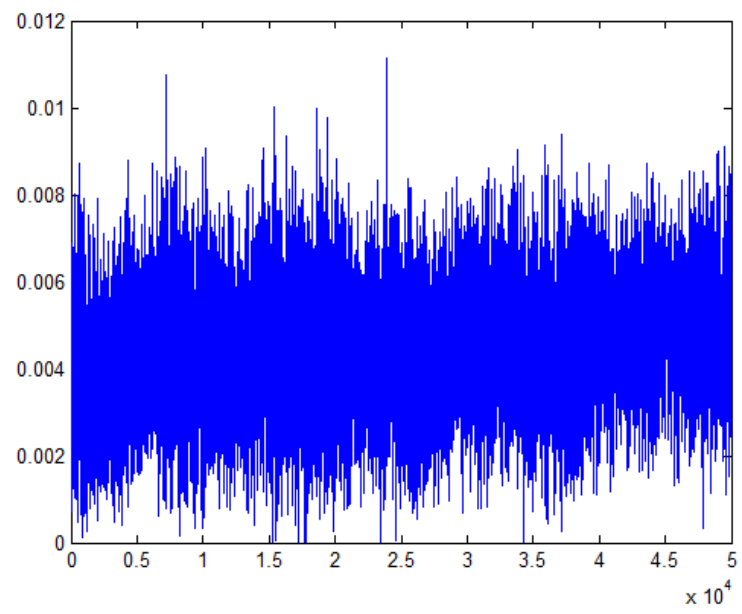
The following shows the result of MCMC on the data simulated by the 3-SDE model. The algorithm starts from the true value and include 50,000 iterations.

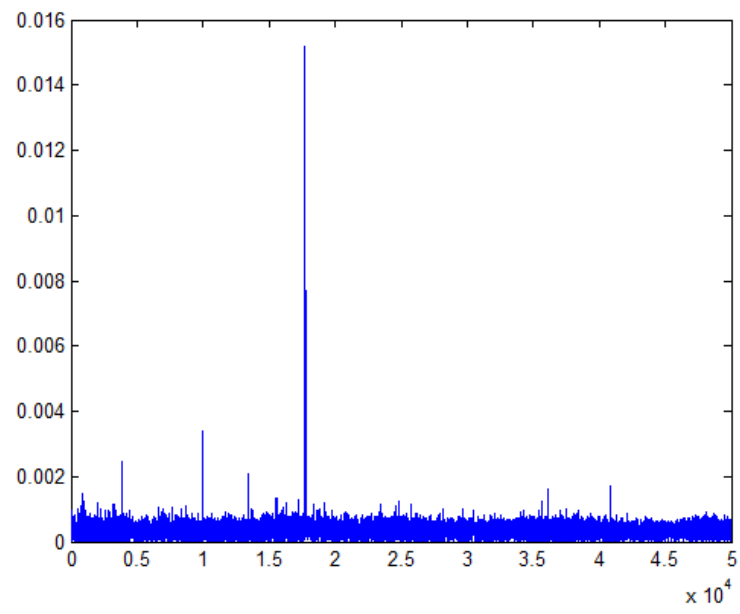
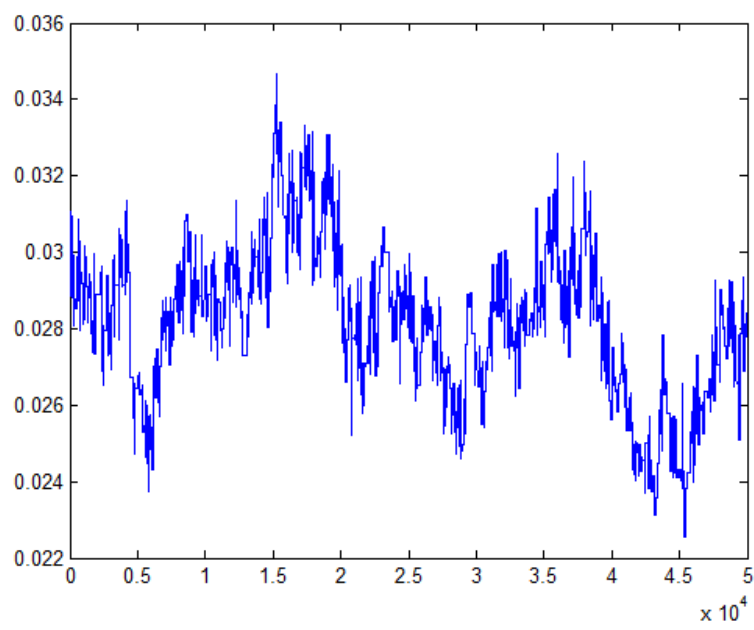
$\kappa$  :



$\kappa_V$  :



$\theta_V :$  $\kappa_{\widetilde{V}} :$ 

$\theta_{\tilde{V}}$  : $\sigma$  :

*Likelihood :*

