MACQUARIE UNIVERSITY

MACQUARIE BUSINESS SCHOOL

Thesis

ON PRICING OF COMMODITY FUTURES USING TWO-FACTOR STATE-SPACE MODEL

by

PEILUN HE

Submitted in partial fulfillment of the

requirements for the degree of

Master by Research

2019

Supervisors

Principal: Prof Pavel Shevchenko

Associate: Dr Nino Kordzakhia

Acknowledgments

First and foremost I would like to thank my supervisors, Prof Pavel Shevchenko and Dr Nino Kordzakhia. They have provided me with limitless guidance in my research. I am also grateful for the support from Macquarie Business School and Faculty of Science and Engineering, especially, for access to the Amazon Web Services (AWS) and space they have generously provided me with. The AWS was very helpful for my computational needs. Finally, I am thankful for the financial support from Macquarie University during the last semester of my enrolment into MRes program.

Peilun He

Department of Actuarial Studies and Business Analytics

Macquarie Business School

Macquarie University

ON PRICING OF COMMODITY FUTURES USING TWO-FACTOR STATE-SPACE MODEL

PEILUN HE

ABSTRACT

We study a bivariate latent factor model for the pricing of commodity futures prices. The two unobservable state variables representing the short and long term factors are modelled as Ornstein-Uhlenbeck (OU) processes and are used for risk-neutral pricing of futures contracts. The Kalman Filter (KF) method is being implemented to estimate the short and long term factors jointly with unknown model parameters. The model parameters are estimated in a form of the Maximum Likelihood Estimators (MLEs). The parameter identification problem arising within the likelihood function in the KF has been addressed by introducing an additional constraint. In the two-dimensional OU model, the consistency and asymptotic variances of conditional MLEs of model parameters are derived. The methodology has been tested on simulated data and also applied to WTI Crude Oil NYMEX futures real market data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Commodity Market

Commodities are usually traded in both spot and futures markets. In the old days, buyers and sellers met on the market where the trade of the commodity and its delivery took place without delay. In the 18th and 19th centuries, some agricultural products were traded in the form of forward agreements. The farmers sold agricultural crops at the time of their planting, in order to finance their production process. In 1848, the Chicago Board of Trade (CBOT) was formed, which is the first commodity trading exchange in the west. After a few years of growth, CBOT decided to standardize the contracts. These contracts are believed to be the first futures-type contracts, also known as 'to-arrive' contracts. Nowadays, the transaction of commodities may be physical or financial, corresponding to trading in the spot or futures market. However, these two types of transactions are highly correlated.

Both a forward and a futures contract is an agreement to purchase an asset at a specified contract maturity date for a fixed price. However, these two types of contracts are quite different. A forward contract is traded in the over-the-counter market. The forward contract usually is traded between two financial institutions, without intermediaries. In contrast, a futures contract is standardized and traded in an exchange, which plays the role of intermediaries. There are other differences between a forward contract and a futures contract. For example, a forward contract settles at the end of the agreement, while the account for a futures contract is maintained daily until its expiry.

Forward contracts mainly are used for hedging. In the spot market, four major risks are identified as follows, Geman (2009):

- Price risk, which corresponds to the change in the product price.
- Transportation risk, which represents the change in the shipping cost.
- Delivery risk.
- Credit risk, which is associated with the default of buyers or sellers.

A forward contract could provide a hedge against price risk. For example, if a corporation knows they will have to pay to their clients in 6 months some amount in a foreign currency, they will buy a 6-month forward contract to hedge against the exchange rate risk. For a forward contract, the credit risk still exists, as one of the participants may not honour their position. However, the credit risk is reduced to almost zero for a futures contract. The clearing house, which guarantees the performance of the participants, and the existence of margin account take away any credit risk. To enter into a futures contract, all participants need to make an initial margin payment, and the participants are required to add daily margin calls to keep a contract ongoing, if the market value of a commodity decreases from the previous day. Moreover, the delivery risk is also reduced in a futures contract, since the exchange standardizes the quality, quantity, and variety of products. Considering the different risks a forward contract and a futures contract take, the prices of these two types of contracts are different, albeit close.

Commodity futures contracts are written on tradable commodities such as metals, energy, livestock/meat and agricultural. In 1999, the Chicago Mercantile Exchange (CME) started trading weather futures. Nowadays futures contracts include Renewable Energy Certificate Futures, Carbon Allowance Futures, and Greenhouse Gas Initiative Futures. Figure 1·1 presents the three fundamental groups of commodities: agricultrual, metals and energy. Because of the physical constraints, the commodities markets show different characteristics comparing to the financial market. In financial

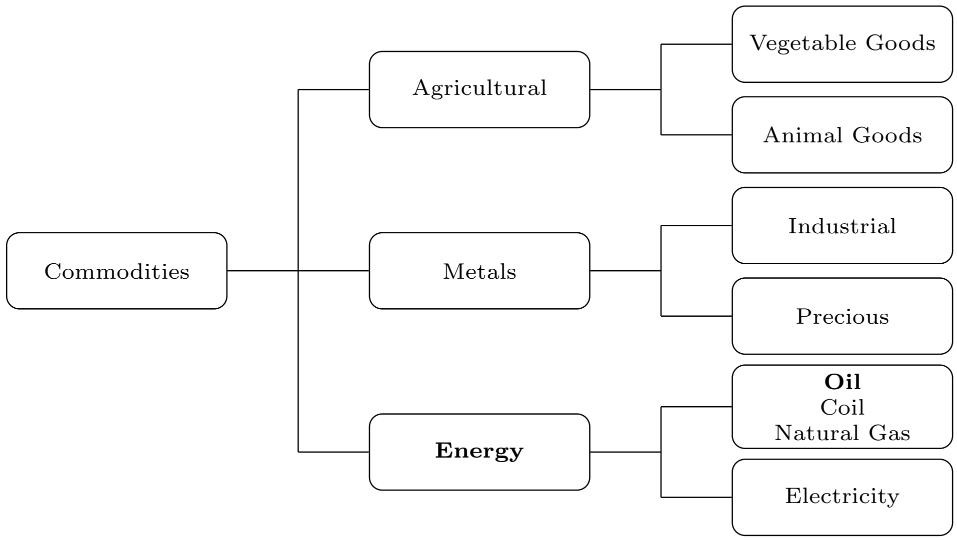

Figure 1·1: Classification of Commodity Markets

markets, cash flow is transferred from one party to another at maturity and without an exchange of underlying goods, while in commodity markets, the commodities are actually delivered. Consequently, commodity prices usually have seasonal and mean-reverting behaviours, and that is why some of the financial theories are not directly applicable to commodity markets.

Energy products are most actively traded commodities in both over-the-counter market and on exchanges. The energy products include crude oil, electricity and natural gas. In the last chapter of this thesis, we use a crude oil historical prices data. There are many grades of crude oil, distinguished by density and sulphur content. The two popular grades are Brent Crude Oil and West Texas Intermediate (WTI) Crude Oil, which is extracted from the North Sea and West Texas, respectively. Figure 1·2 gives a cross-sectional data of WTI Crude Oil Futures settle prices with different maturities on 8th November 2019. The original data was obtained from the CME Group. The major futures exchanges include the New York Mercantile Exchange

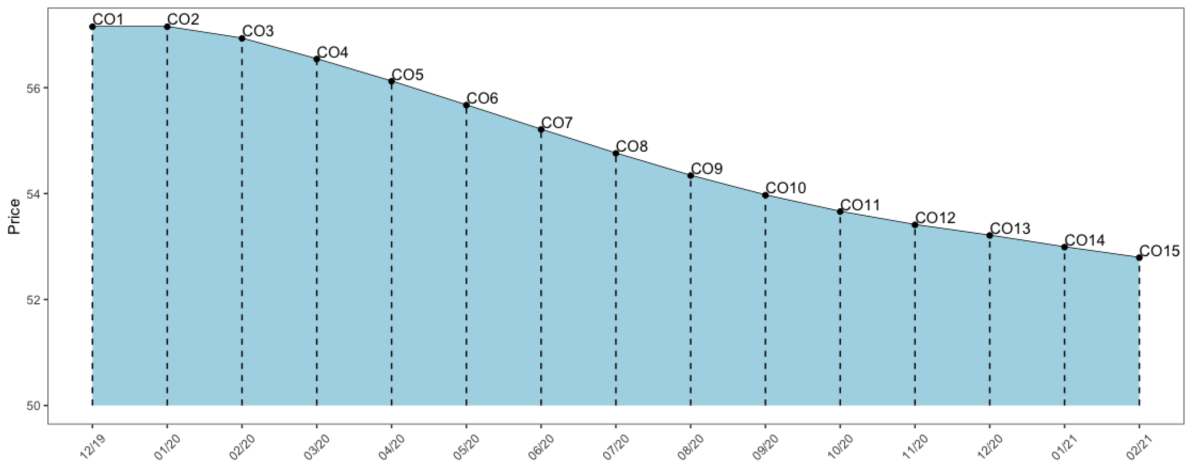

Figure 1·2: WTI Crude Oil Futures Curve

(NYMEX), New York Board of Trade (NYBOT) and London International Financial Futures Exchange (LIFFE).

1.2 Structure and Aim of the Thesis

From 2001 to 2008, the price of oil increased eight times. However, during the global financial crisis (GFC) in 2008, the oil futures prices dropped dramatically. The decline had taken commodity more than halfway back to the 2001 level. One reason for the high-risk of commodity futures is leverage, which enables investors to purchase investments with only a small proportion of their real value. For a futures contract, this proportion could be only 10%. The uncertainty of futures prices makes commodity consumers and market players subjected to elevated long-term risks.

This work uses the Ornstein-Uhlenbeck (OU) two-factor model for modelling of short and long equilibrium commodity spot price levels. A commodity spot price S_t is modelled as the sum of two unobservable factors χ_t and ξ_t. Both processes χ_t and ξ_t represent the mean-reverting processes. In the mean-reverting model, when the commodity price is higher than the equilibrium price level, some new suppliers will enter the market and create downward pressure on the prices. Conversely, when the price is lower than the equilibrium price level, some high-cost suppliers will exit the market and put upward pressure on the prices. In the short term, due to these movements, the price fluctuates temporarily, and it will eventually converge to its equilibrium level over the long term.

Our motivation is driven by the fact that the parameter estimation problem in the linear system using the Kalman Filter cannot be overlooked whilst the estimation of the state variables remains the priority. In the different setup, the parameter estimation problem for the bivariate OU process using Kalman Filter (KF) has been studied by Favetto and Samson (2010) and Kutoyants (2019).

In Chapter 2, we derived the linear partially observable system specific for commodity futures prices developed in the two-factor model, which represents an extension of Schwartz and Smith's (2000) model, in the risk-neutral setting. In Chapter 3, the details of the implementation of the KF algorithm designed for estimation of the parameters of the two-dimensional partially observable linear system jointly with the estimation of unobserved state variables χ_t and ξ_t were presented. Moreover, the algorithm for obtaining the asymptotic variance of the estimates of parameters was introduced. The results of the simulation study are presented in Chapter 4. The parameter identification problem (PIP) arising within the likelihood function in the KF is proved mathematically, which has been resolved numerically by introducing an additional constraint. In Chapter 5, the modelling method has been applied to WTI Cude Oil futures data.

The results were presented at the Research School on Statistics and Data Science which was held from 24th to 26th of July in Melbourne. The paper on parameter estimation in Schwartz and Smith's model (2000) has been accepted for publication, Binkowski et al. (2019).

1.3 Literature Review

Over more than four decades the stochastic processes have been used in modelling of commodity futures prices. In early studies, the researchers assumed that the commodity prices followed a geometric Brownian motion, Black (1976). However, Gibson and Schwartz (1990) have introduced the mean-reverting processes for statistical modelling in finance, also known as Ornstein-Uhlenbeck or Vasicek processes, Ornstein and Uhlenbeck (1930), Vasicek (1977).

The previous studies included one and two-factor models. Sorensen (2002) used the one-factor model for seasonality adjustment in agricultural commodity futures. Carmona and Coulon (2014) applied the one-factor model to the electricity market. However, the disadvantage of the one-factor model is that the futures returns are correlated, which opposes the empirical evidence of Cortazar and Schwartz (2003). Ames et al (2020) introduced the two-factor model with observable factors incorporated into model drift parameters to allow for analysis of the impact of macroeconomic factors on the futures prices. In Cortazar et al. (2019), the authors improved the performance of the Kalman Filter by deriving the commodity spot prices from futures prices which have had incorporated an analyst's forecasts of spot prices. In Cheng et al. (2018) the Kalman Filter is used to study the effect of stochastic volatility and interest rates on commodity spot prices using the market prices of long-dated futures and options. Peters et al. (2013) applied the Kalman technique to calibration, jointly with filtering, of partially unobservable processes using particle Markov chain Monte Carlo approach. Guo (2017) proposed a multi-factor model for risk measurement and modelling using the Monte Carlo method. Ewald et al. (2018) developed the extended Kalman Filter for estimation of the state variables in the two-factor Shwartz (1997) model for the commodity spot price and its yield.

1.4 Computing Environment

The codes for this work were developed in MATLAB R2019b. In the Amazon Web Services (AWS) Australian Sydney computing center, the computations for simulation study and applications to Crude Oil futures data were run in a c5.18xlarge instance (72 vCPUs) and c5.x24xlarge instance (96 vCPUs), respectively. The results presented in Chapters 4 and 5 have taken the total computational time of approximately 1000 hours (42 days) on both AWS platforms.

Chapter 2

Two-Factor Model

We propose the two-factor model of pricing of commodity futures which represents an extension of Schwartz and Smith (2000), where the spot price S_t is modelled as the sum of two unobservable factors χ_t and ξ_t. In Schwartz and Smith (2000), only one factor had a mean-reverting property. In this work, both χ_t and ξ_t are modelled as the mean-reverting processes. We develop Kalman Filter for estimation of unobservable factors χ_t and ξ_t as well as the model parameters.

In this chapter, we provide a brief description of our model. Firstly, we introduce Ornstein-Uhlenbeck processes, which are used for modelling of the short and long term state variables. Then we discuss the two-factor model, followed by a special case, the risk-neutral setup is used for futures pricing. Further, we assume the interest rate $r = 0$ throughout.

2.1 Ornstein-Uhlenbeck Processes

In this section, we introduce the Ornstein-Uhlenbeck process as the building blocks of our modelling approach. These processes are often used in modelling of asset prices and interest rates.

Definition 2.1 $(X_t)_{t\geq 0}$ is said to be an Ornstein-Uhlenbeck (OU) process if

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \tag{2.1}$$

where the parameters $\theta > 0$, μ and $\sigma > 0$ are the mean reversion rate, the mean and the volatility, respectively. And the process $(W_t)_{t\geq 0}$ is a standard Brownian motion.

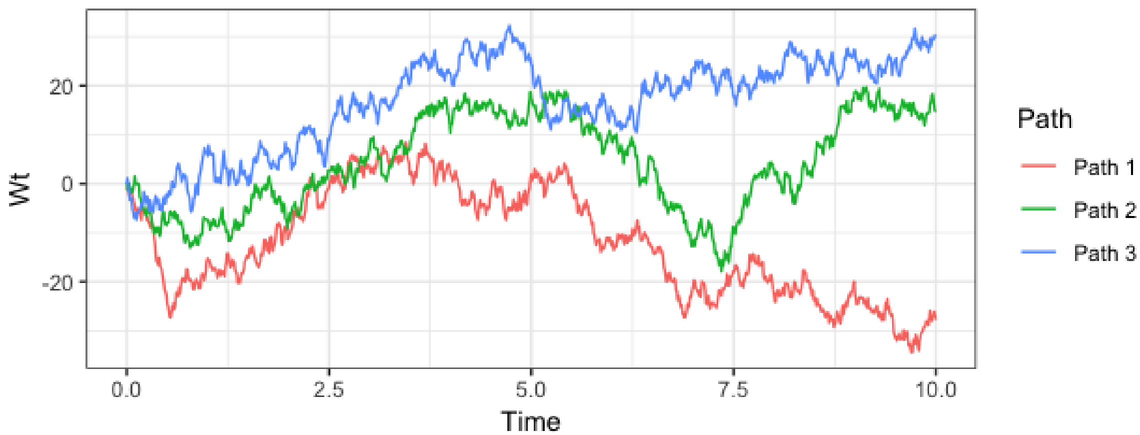

Figure 2·1: Paths of Brownian Motion, $W_0 = 0$

Remark 2.1 The solution of OU equation (2.1) is

$$X_t = X_0 e^{-\theta t} + \mu\left(1 - e^{-\theta t}\right) + \sigma \int_0^t e^{-\theta(t-s)} dW_s. \tag{2.2}$$

Indeed (2.2) can be proved by using Ito's lemma for the function $f(x,t) = xe^{\theta t}$. The mean-reverting property can be explained as follows, if we ignore the stochastic movement in the Brownian motion part, then X_t converge to μ. In Figures 2.1 and 2.2 three different paths of Brownian Motion and Ornstein-Uhlenbeck processes are displayed.

2.2 Spot Price Modelling

Here we provide the description of the two-factor model using the short and long term state variables, Schwartz and Smith (2000). We model the logarithm of the spot price, the price of an asset, using the additive model. Let the spot price of a commodity at time t be S_t, then

$$\log(S_t) = \chi_t + \xi_t,$$

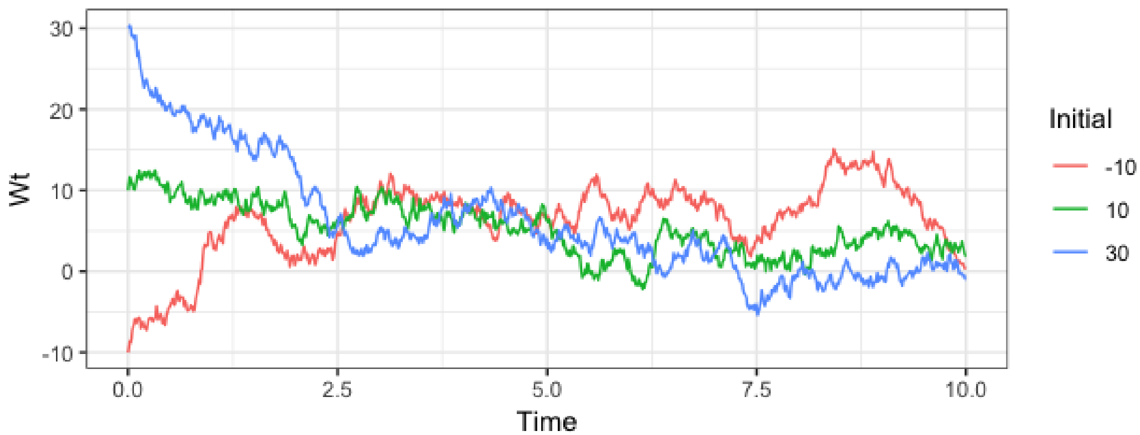

Figure 2·2: Paths of Ornstein-Uhlenbeck process, $X_0 = -10, 10, 30$ with θ=1, μ=5, σ=0.5

where χ_t is the short-term fluctuation in prices and ξ_t is the long-term equilibrium price level. We assume that changes in χ_t are temporary, following an Ornstein-Uhlenbeck process

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dZ_t^\chi, \kappa > 0. \tag{2.3}$$

The changes in the equilibrium level of ξ_t are expected to persist and the process itself is assumed to be mean-reverting

$$d\xi_t = \gamma\left(\frac{\mu_\xi}{\gamma} - \xi_t\right)dt + \sigma_\xi dZ_t^\xi = (\mu_\xi - \gamma\xi_t)dt + \sigma_\xi dZ_t^\xi, \gamma > 0. \tag{2.4}$$

The processes $(Z_t^\chi)_{t\geq 0}$ and $(Z_t^\xi)_{t\geq 0}$ are correlated standard Brownian motions processes with $E\left(dZ_t^\chi dZ_t^\xi\right) = \rho_{\chi\xi}dt$. In (2.3) and (2.4), χ_t and ξ_t converge to 0 and $\frac{\mu_\xi}{\gamma}$ respectively. In discrete time, given the initial values χ_0 and ξ_0, χ_t and ξ_t are jointly normally distributed with mean

$$E[(\chi_t, \xi_t)] = \left(e^{-\kappa t}\chi_0, \frac{\mu_\xi}{\gamma}\left(1 - e^{-\gamma t}\right) + e^{-\gamma t}\xi_0\right), t \geq 0 \tag{2.5}$$

and covariance matrix

$$Cov[(\chi_t, \xi_t)] = \begin{bmatrix} \frac{1-e^{-2\kappa t}}{2\kappa}\sigma_\chi^2 & \frac{1-e^{-(\kappa+\gamma)t}}{\kappa+\gamma}\sigma_\chi\sigma_\xi\rho_{\chi\xi} \\ \frac{1-e^{-(\kappa+\gamma)t}}{\kappa+\gamma}\sigma_\chi\sigma_\xi\rho_{\chi\xi} & \frac{1-e^{-2\gamma t}}{2\gamma}\sigma_\xi^2 \end{bmatrix}. \tag{2.6}$$

Derivation of (2.5) and (2.6) are given in Appendix A. Therefore the logarithm of spot price is normally distributed with mean

$$E[\log(S_t)] = e^{-\kappa t}\chi_0 + \frac{\mu_\xi}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t}\xi_0$$

and variance

$$Var[\log(S_t)] = \frac{1 - e^{-2\kappa t}}{2\kappa}\sigma_\chi^2 + \frac{1 - e^{-2\gamma t}}{2\gamma}\sigma_\xi^2 + 2\frac{1 - e^{-(\kappa+\gamma)t}}{\kappa + \gamma}\sigma_\chi\sigma_\xi\rho_{\chi\xi}.$$

Hence S_t, the spot price, is log-normally distributed and

$$E(S_t) = \exp\left(E[\log(S_t)] + \frac{1}{2}Var[\log(S_t)]\right),$$

or

$$\begin{aligned} \log[E(S_t)] =& e^{-\kappa t}\chi_0 + \frac{\mu_\xi}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t}\xi_0 \\ &+ \frac{1}{2}\left(\frac{1 - e^{-2\kappa t}}{2\kappa}\sigma_\chi^2 + \frac{1 - e^{-2\gamma t}}{2\gamma}\sigma_\xi^2 + 2\frac{1 - e^{-(\kappa+\gamma)t}}{\kappa + \gamma}\sigma_\chi\sigma_\xi\rho_{\chi\xi}\right), \end{aligned} \tag{2.7}$$

where μ_ξ is the mean, σ_χ and σ_ξ are the volatilities, γ and κ are the speed of mean-reversion parameters of χ and ξ processes, respectively.

2.3 Risk-Neutral Approach to Spot Price Modelling

In this section, we introduce the two additional parameters in χ and ξ processes, one parameter in each process, for adjustment to market risk prices. The approach stems from the risk-neutral futures pricing theory developed by Black (1976). In the risk-neutral framework, the commodity futures prices are supposed to be martingales.

However, the spot price process utilised for pricing of the futures contracts is not a tradable asset and, therefore, may not be a martingale. In this framework, the commodity spot price is unobservable and the market is incomplete. Hence, in the extended Schwartz-Smith's model, we do not require the commodity spot price process to be a martingale. These types of models have been used in Gibson and Schwartz (1990), Schwartz and Smith (2000), Cortazar et al. (2019), Ames et al. (2020) and Farkas et al. (2017).

We assume that risk premium adjustments are constants, i.e.

$$d\chi_t = (-\kappa\chi_t - \lambda_\chi)dt + \sigma_\chi dZ_t^{\chi^*},$$

$$d\xi_t = (\mu_\xi - \gamma\xi_t - \lambda_\xi)dt + \sigma_\xi dZ_t^{\xi^*},$$

where λ_χ, λ_ξ are risk-neutral mean corrections and $Z_t^{\chi^*}$ and $Z_t^{\xi^*}$ are correlated standard Brownian motions with $E\left(dZ_t^{\chi^*}dZ_t^{\xi^*}\right) = \rho_{\chi\xi}dt$, the correlation coefficient is $\rho_{\chi\xi}$ the same as that of Z_t^χ and Z_t^ξ. Under the risk-neutral process, χ_t and ξ_t are also jointly normally distributed with mean

$$E^*[(\chi_t, \xi_t)] = \left(e^{-\kappa t}\chi_0 - \frac{\lambda_\chi}{\kappa}(1 - e^{-\kappa t}), \frac{\mu_\xi - \lambda_\xi}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t}\xi_0\right)$$

and covariance matrix

$$Cov^*[(\chi_t, \xi_t)] = Cov[(\chi_t, \xi_t)],$$

where E^* and Cov^* are the expectation and covariance with respect to the risk-neutral measure. The logarithm of spot price is normally distributed with mean

$$E^*[\log(S_t)] = e^{-\kappa t}\chi_0 - \frac{\lambda_\chi}{\kappa}(1 - e^{-\kappa t}) + \frac{\mu_\xi - \lambda_\xi}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t}\xi_0$$

and variance

$$Var^*[\log(S_t)] = Var[\log(S_t)].$$

The spot price is log-normally distributed with

$$\log[E^*(S_t)] = E^*[\log(S_t)] + \frac{1}{2}Var^*[\log(S_t)] = e^{-\kappa t}\chi_0 + e^{-\gamma t}\xi_0 + A(t), \tag{2.8}$$

where

$$\begin{aligned} A(t) = & -\frac{\lambda_\chi}{\kappa}(1 - e^{-\kappa t}) + \frac{\mu_\xi - \lambda_\xi}{\gamma}(1 - e^{-\gamma t}) \\ & + \frac{1}{2}\left(\frac{1 - e^{-2\kappa t}}{2\kappa}\sigma_\chi^2 + \frac{1 - e^{-2\gamma t}}{2\gamma}\sigma_\xi^2 + 2\frac{1 - e^{-(\kappa+\gamma)t}}{\kappa + \gamma}\sigma_\chi\sigma_\xi\rho_{\chi\xi}\right). \end{aligned} \tag{2.9}$$

In (2.8) the parameters λ_χ and λ_ξ appear according to the adjustment made in (2.7).

2.4 Risk-Neutral Approach to Pricing of Futures

For a futures contract, we are interested to know what is the price of such contract at present. Let $F_{0,T}$ be the current market price of the futures contract with maturity T. For elimination of arbitrage, colloquially known as a free-lunch situation, the futures prices must be equal to the expected spot prices at the asset delivery time T. Hence, under the risk-neutral approach from Section 2.2, we obtain (we assume the interest rate is not stochastic)

$$\log(F_{0,T}) = \log[E^*(S_T)] = e^{-\kappa T}\chi_0 + e^{-\gamma T}\xi_0 + A(T).$$

Then for modelling in discrete time, we have the following AR(1) dynamics for bivariate state variable x_t

$$x_t = c + Gx_{t-1} + w_t, \tag{2.10}$$

where

$$x_t = \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix}, c = \begin{bmatrix} -\frac{\lambda_\chi}{\kappa}(1 - e^{-\kappa\Delta t}) \\ \frac{\mu_\xi - \lambda_\xi}{\gamma}(1 - e^{-\gamma\Delta t}) \end{bmatrix}, G = \begin{bmatrix} e^{-\kappa\Delta t} & 0 \\ 0 & e^{-\gamma\Delta t} \end{bmatrix},$$

and w_t is a column vector of uncorrelated normally distributed random variables with $E(w_t) = 0$ and $Cov(w_t) = W = Cov[(\chi_{\Delta t}, \xi_{\Delta t})]$. Δt is the time step between $t - 1$

and t. The relationship between the state variables and the observed futures prices is given by

$$y_t = d_t + F_t' x_t + v_t, \tag{2.11}$$

where

$$y_t' = (\log(F_{T_1}), \log(F_{T_2}), \cdots, \log(F_{T_n})),$$

$$d_t' = (A(T_1), A(T_2), \cdots, A(T_n)),$$

$$F_t = \begin{bmatrix} e^{-\kappa T_1}, e^{-\kappa T_2}, \ldots, e^{-\kappa T_n} \\ e^{-\gamma T_1}, e^{-\gamma T_2}, \ldots, e^{-\gamma T_n} \end{bmatrix},$$

and v_t is a $n \times 1$ vector of uncorrelated normally distributed random variables with $E(v_t) = 0$ and $Cov(v_t) = V$. The variables $T_i, i = 1, ..., n$ are the times to maturity, which are the differences between the futures expiry times and current time t. We assume that V is a diagonal matrix with non-zero diagonal entries $s = (s_1^2, s_2^2, \ldots, s_n^2)$. Here n is the number of the futures contracts. Let $\mathcal{F}_t$ be the history of information generated by the futures contract up to time t. The log-likelihood function of $y = (y_1, y_2, \ldots, y_{n_T})$ is

$$l(\theta; y) = \sum_{t=1}^{n_T} p(y_t | \mathcal{F}_{t-1}),$$

with the set of unknown parameters $\theta = (\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}, \lambda_\chi, \lambda_\xi, s)$, where $p(y_t|\mathcal{F}_{t-1})$ is the probability density of y_t conditioned on the information available up to $t-1$ and n_T is the number of time instances. We assume that the prediction errors $e_t = y_t - E(y_t|\mathcal{F}_{t-1})$ are multivariate normally distributed, then the log-likelihood function is

$$l(\theta; y) = -\frac{n n_T \log 2\pi}{2} - \frac{1}{2} \sum_{t=1}^{n_T} \left[\log \left[\det(L_{t|t-1}) \right] + e_t' L_{t|t-1}^{-1} e_t \right], \tag{2.12}$$

where $L_{t|t-1} = Cov(e_t|\mathcal{F}_{t-1})$. Given y_t, the maximum likelihood estimates (MLE) of unknown parameters θ can be estimated by maximising this log-likelihood function (2.12). Because of the complexity of the log-likelihood function, grid search is necessary in estimating of parameters.

Chapter 3

Filtering and Parameter Estimation

In this chapter, we provide numerical methods for Kalman's algorithms, including Kalman Filter and Kalman Smoother. Given the data $y_1, y_2, \cdots, y_{t_1}$, our main target is to estimate the unobservable state vector $x_{t_2} = (\chi_{t_2}, \xi_{t_2})'$. When $t_1 < t_2$, $t_1 = t_2$ or $t_1 > t_2$, the problem is called the forecasting, filtering or smoothing problem, respectively. Here we do not consider the forecasting problem scenario. The estimation of state vector can be used for predicting the futures prices and their risk management. Then, we introduce the method to obtain the asymptotic variances of the estimates of the parameters, which is based on the product of the score vector. In the final section, we introduce the generalised Cholesky decomposition for dealing with improper empirical estimates of covariance matrices.

3.1 Kalman Filter

The Kalman filter is a recursive process to estimate state vector at time t based on the information available up to t, Harvey (1990). In this section, we are using Kalman Filter to estimate the unobservable vector of state variables $x_t = (\chi_t, \xi_t)'$ jointly with unknown parameters $\theta = (\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}, \lambda_\chi, \lambda_\xi, s)$ by observing y_t. We recall the equations (2.10) and (2.11) for x_t and y_t, respectively

$$x_t = c + Gx_{t-1} + w_t,$$

$$y_t = d_t + F_t' x_t + v_t.$$

We need an initial value to start the Kalman Filter. Binkowski, Shevchenko and Kordzakhia (2009) used the initial expectation and covariance matrix

$$E(x_0) = a_0 = \begin{bmatrix} 0 \\ \frac{\mu_\xi}{\gamma} \end{bmatrix} \tag{3.1}$$

and

$$Cov(x_0) = P_0 = \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa} & \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa+\gamma} \\ \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa+\gamma} & \frac{\sigma_\xi^2}{2\gamma} \end{bmatrix}. \tag{3.2}$$

Starting with a_0 and P_0, the recursive process is constructed. The algorithm of the Kalman Filter is given below.

1. Start with a_0 and P_0 given in (3.1) and (3.2).

2. Given all information until time $t-1$, we predict the expectation of state vector $a_{t|t-1}$ and covariance matrix $P_{t|t-1}$ at time t by

$$a_{t|t-1} = Ga_{t-1} + c$$

and

$$P_{t|t-1} = GP_{t-1}G' + W.$$

3. When a new observation y_t is available, the prediction error e_t and covariance matrix $L_{t|t-1}$ are calculated by

$$e_t = y_t - d_t - F_t' a_{t|t-1}$$

and

$$L_{t|t-1} = F_t' P_{t|t-1} F_t + V.$$

4. a_t and P_t are updated by

$$a_t = a_{t|t-1} + K_t e_t$$

and

$$P_t = (I - K_t F_T')P_{t|t-1},$$

where

$$K_t = P_{t|t-1}F_t(L_{t|t-1})^{-1}$$

is the Kalman gain matrix.

5. The log-likelihood function at time t is calculated by

$$l_t = -\log\left(\det(L_{t|t-1})\right) - e_t' L_{t|t-1}^{-1} e_t.$$

6. Repeat step 2 - 5 for next time point.

By completing the recursive process from $t = 1$ to n_T, we can sum all l_t up to get the log-likelihood l, i.e. $l = \sum_{t=1}^{n_T} l_t$. Then we maximise l for obtaining the estimates θ. The flowchart of Kalman Filter is given in Figure 3·1. In the extended Schwartz-Smith model, the parameter estimation theory is yet to be established. Based on our results from the simulation and empirical studies we demonstrated in our recently published paper the consistency of the MLE estimators obtained through Kalman technique, Binkowski et al. (2019).

3.2 Kalman Smoother

While the Kalman FIlter uses all past observations for estimation, the Kalman Smoother uses the full (past and future) set of observations. That is, we are interested in the estimations of state vector x_t given all information up to time n $(t = 1, 2, ..., n)$.

There are many types of smoothing, such as fixed-interval smoothing, fixed-point smoothing and fixed-lag smoothing. In the fixed-interval smoothing, the full set of observations n is fixed, and the smoothed point t varies. The fixed-point smoothing supposes t remains fixed and n changes. With fixed-lag smoothing, both n and t vary but the lag $n - t$ remains fixed. De Jong (1989) provided different methods for

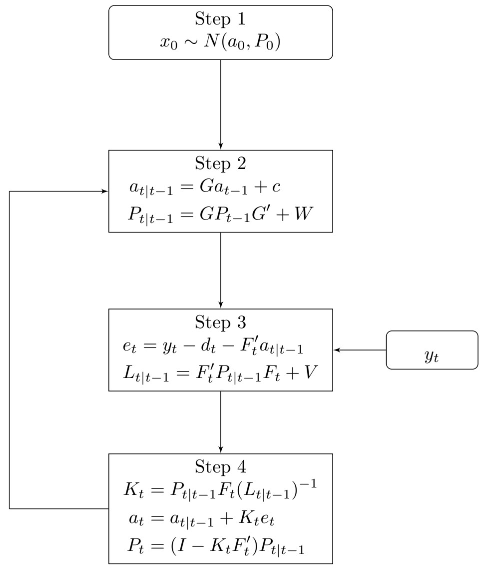

Figure 3·1: Flowchart of Kalman Filter

different types of smoothing. However, we only focus on fixed-interval smoothing in this thesis.

The classic fixed-interval smoothing, also known as Rauch-Tung-Striebel (RTS) smoother, was developed by Rauch, Tung and Striebel (1965). After applying the Kalman Filter, the backwards smoother is given by

$$a_{t-1|n} = a_{t-1|t-1} + C_{t-1}(x_{t|n} - x_{t|t-1}), \tag{3.3}$$

and

$$P_{t-1|n} = P_{t-1|t-1} + C_{t-1}(P_{t|n} - P_{t|t-1})C_{t-1}', \tag{3.4}$$

where $t = n, n-1, ..., 1$ and

$$C_{t-1} = P_{t-1|t-1} G' P_{t|t-1}^{-1}. \tag{3.5}$$

The initial conditions $a_{n|n}$ and $P_{n|n}$ are obtained via Kalman Filter.

However, in RTS smoother, inverse matrix $P_{t|t-1}^{-1}$ is required. When the dimension of the state variable x_t is large, the matrix inversion causes a computational difficulty.

Therefore, Bierman (1973) and De Jong (1989) provided another method, which is called Modified Bryson-Frazier (MBF) smoother. The smoother is given by

$$a_{t|n} = a_{t|t-1} + P_{t|t-1} r_{t-1}, \tag{3.6}$$

and

$$P_{t|n} = P_{t|t-1} - P_{t|t-1} R_{t-1} P_{t|t-1}, \tag{3.7}$$

where

$$r_{t-1} = F_t L_{t|t-1}^{-1} e_t + (G - GK_t F_t)' r_t, \tag{3.8}$$

and

$$R_{t-1} = F_t L_{t|t-1}^{-1} F_t' + (G - GK_t F_t)' R_t (G - GK_t F_t). \tag{3.9}$$

The initial conditions are $r_n = 0$ and $R_n = 0$. The RTS smoother will be used for calculating the score vector in the next section.

3.3 Score Vector

In this section, the score vector is introduced. Let a column vector $\mathcal{G}$ denote the first-order derivative of log-likelihood with respect to the set of unknown parameters θ. The column vector $\mathcal{G}$ is also called gradient or score. Koopman and Shephard (1992) and Durbin and Koopman (2002) introduced a method to calculate the score for Gaussian state-space models by Kalman Smoother. The log-likelihood function is given by

$$\log L(y|\theta) = \log p(y|\theta) = \log p(x, y|\theta) - \log p(x|y, \theta),$$

where $p(\cdot)$ denote density function. By taking expectation with respect to the density $p(x|y,\theta)$ and then differentiating with respect to θ, we have

$$\begin{aligned}\mathcal{G}(\theta) = \frac{\partial \log L(y|\theta)}{\partial \theta} = -\frac{1}{2}\frac{\partial}{\partial \theta}\sum_{t=1}^{n} & \{\log|W| + \log|V| \\ & + tr[\{\omega_{t|n}\omega_{t|n}' + Var(\omega_{t|n})\}W^{-1}] \\ & + tr[\{\nu_{t|n}\nu_{t|n}' + Var(\nu_{t|n})\}V^{-1}]\Big\}, \end{aligned} \quad (3.10)$$

where $\omega_{t|n}$ and $\nu_{t|n}$ are smoothed estimates of w_t and v_t. $tr(\cdot)$ represents the trace of a matrix and $|\cdot|$ represents the determinant of a matrix. Durbin and Koopman (2002) provided formulas for $\omega_{t|n}$, $\nu_{t|n}$, $Var(\omega_{t|n})$ and $var(\nu_{t|n})$ as

$$\omega_{t|n} = Wr_t, \quad (3.11)$$

$$\nu_{t|n} = VL_{t|t-1}^{-1}e_t - VK'G'r_t, \quad (3.12)$$

$$Var(\omega_{t|n}) = W - WR_tW, \quad (3.13)$$

and

$$Var(\nu_{t|n}) = V - V(L_{t|t-1}^{-1} + K'G'R_tGK)V, \quad (3.14)$$

where r_t and R_t are given from (3.8) and (3.9). The prove of (3.10) is given in Appendix B.

3.4 Asymptotic Variance

In this section, the asymptotic variances of MLE estimators are established using the score vector. In Section 2.4, the estimates of unknown parameters are obtained by maximising the log-likelihood function

$$l(\theta; y) = -\frac{nn_T \log 2\pi}{2} - \frac{1}{2}\sum_{t=1}^{n_T}\left[\log\left(\det(L_{t|t-1})\right) + e_t'L_{t|t-1}^{-1}e_t\right],$$

where θ is a vector of the true parameters. Let $\hat{\theta}_n$ be the MLE estimate of θ. Ljung and Caines (1979), Stoffer and Wall (1991) and Davis (2013) stated that under appropriate conditions, given y, as $n \to \infty$, $(\hat{\theta}_n - \theta) \to N(\mathbf{0}, \mathcal{C})$ in distribution. The asymptotic normal distribution has a zero mean vector $\mathbf{0}$ and variance-covariance matrix $\mathcal{C}$. The matrix $\mathcal{C}$ is commonly obtained by taking the inverse of Fisher information matrix (FIM). The Cramer-Rao theorem states that the diagonal of the inverse of FIM gives the lower bound of the variances of the unbiased parameter estimators. For MLEs, the diagonal entries of the FIM's inverse are equal to their variances. Let $\mathcal{I}$ be the FIM, $\mathcal{C}$ is given by

$$\mathcal{C} = \mathcal{I}^{-1}.$$

$\mathcal{I}$ is defined as the variance of the score vector:

$$\mathcal{I}(\theta) = Var(\mathcal{G}) = Var\left(\frac{\partial \log L(y|\theta)}{\partial \theta}\right). \tag{3.15}$$

The general asymptotic properties of MLE can be found in Devore and Berk (2012). Since $L(y|\theta)$ is a density function, we have

$$\int L(y|\theta)dy = 1. \tag{3.16}$$

Then we differentiate both sides of (3.16) with respect to θ. Using the fact

$$\frac{\partial}{\partial \theta} \log L(y|\theta) = \frac{1}{L(y|\theta)} \frac{\partial}{\partial \theta} L(y|\theta),$$

we have

$$\begin{aligned} \frac{\partial}{\partial \theta} \int L(y|\theta)dy &= \int \frac{\partial}{\partial \theta} L(y|\theta)dy \\ &= \int L(y|\theta) \frac{\partial}{\partial \theta} \log L(y|\theta)dy \\ &= E\left(\frac{\partial}{\partial \theta} \log L(y|\theta)\right) = 0. \end{aligned} \tag{3.17}$$

Then from (3.16) and (3.17) we have

$$\mathcal{I}(\theta) = E(\mathcal{G}\mathcal{G}') = E\left(\frac{\partial \log L(y|\theta)}{\partial\theta}\frac{\partial \log L(y|\theta)}{\partial\theta'}\right). \tag{3.18}$$

However, if we take the second-order derivative with respect to θ_i and θ_j in (3.16), we have

$$\begin{aligned}\int \frac{\partial^2}{\partial\theta_i\theta_j} L(y|\theta)dy &= \int L(y|\theta)\frac{\partial^2}{\partial\theta_i\theta_j}\log L(y|\theta)dy + \int \frac{\partial}{\partial\theta_j}L(y|\theta)\frac{\partial}{\partial\theta_i}\log L(y|\theta)dy \\ &= \int L(y|\theta)\frac{\partial^2}{\partial\theta_i\theta_j}\log L(y|\theta)dy + \int L(y|\theta)\frac{\partial}{\partial\theta_j}\log L(y|\theta)\frac{\partial}{\partial\theta_i}\log L(y|\theta)dy \\ &= E\left(\frac{\partial^2}{\partial\theta_i\theta_j}\log L(y|\theta)\right) + E\left(\frac{\partial}{\partial\theta_i}\log L(y|\theta)\frac{\partial}{\partial\theta_j}\log L(y|\theta)\right) \\ &= 0. \end{aligned} \tag{3.19}$$

Then we get

$$E\left(\frac{\partial \log L(y|\theta)}{\partial\theta}\frac{\partial \log L(y|\theta)}{\partial\theta'}\right) = -E\left(\frac{\partial^2 \log L(y|\theta)}{\partial\theta\partial\theta'}\right), \tag{3.20}$$

which implies an alternative expression for $\mathcal{I}$:

$$\mathcal{I}(\theta) = -E(\mathcal{H}) = -E\left(\frac{\partial^2 \log L(y|\theta)}{\partial\theta\partial\theta'}\right), \tag{3.21}$$

where $\mathcal{H}$ is the second-order derivative of the log-likelihood function with respect to vector θ. $\mathcal{H}$ is also called the Hessian matrix. Both (3.18) and (3.21) can be used to calculate FIM. However, since the numerical calculations of the second-order derivatives are more problematic, we report (3.18) throughout even though in some applications we have had both formulae used and achieved a reasonable consistency.

The algorithm for obtaining the variances of the estimates of parameters is outlined below.

1. For a given data set y, get MLE estimates $\hat{\theta}$ for unknown parameters θ.

2. Let $\theta = \hat{\theta}$. Generate a new data set $\tilde{y}$ from (2.10) and (2.11).

3. For the data $\tilde{y}$ and parameters $\hat{\theta}$, obtain the score vector by (3.10) numerically. That is, given an increment h, the score is $\frac{\mathcal{G}(\theta+h)-\mathcal{G}(\theta)}{h}$.

4. Let $\mathcal{F}_i = \mathcal{G}\mathcal{G}'$ be the product of the score vector and its transpose.

5. Repeat step 2 - 4 M times, we get $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_M$. Get the expectation of these $\mathcal{F}_i$'s. to obtain the FIM $\mathcal{I}$.

6. Take the inverse of $\mathcal{I}$. This would be the variance-covariance matrix. Let's denote it $\mathcal{C}$.

At step 2, we generate random values for w_t and v_t in (2.10) and (2.11), so that we have different data $\tilde{y}$ in each iteration, and so $\mathcal{F}_i$ are different. This algorithm will be used for obtaining the asymptotic variances of the estimates of parameters in Chapter 4 and 5.

3.5 Generalised Cholesky Decomposition

In Section 3.4, we provided the algorithm for obtaining the asymptotic varaince-covaraince matrix $\mathcal{C}$. However, to make $\mathcal{C}$ a valid variance-covariance matrix, two conditions must be satisfied:

1. The Fisher information matrix $\mathcal{I}$ is invertible.

2. The variance-covariance matrix $\mathcal{C}$ is a positive definite matrix.

The first condition ensures that $\mathcal{C}$ exists, while the second condition ensures all the variances are positive. If the first condition is not satisfied, the inverse of FIM can be obtained through Moore-Penrose pseudoinverse. In this section, we introduce the generalised Cholesky decomposition (GCD) to deal with the problem of the fulfilment of the second condition.

The GCD was firstly developed by Gill and Murray (1974), which is used to address the indefiniteness of variance-covariance matrix. Given a symmetric and not necessarily positive definite matrix A, we calculate a Cholesky decomposition L and D of $A+\Delta$, where

$$P(A+\Delta)P' = L'DL$$

with P a permutation matrix and Δ which make $A+\Delta$ a positive definite matrix. If A is positive definite, Δ should be zero; if A is indefinite, the norm of Δ should be small.

Schnabel and Eskow (1990) provided an updated algorithm, which gives a considerably smaller norm of Δ than Gill and Murray's algorithm. Cheng and Higham (1998) suggested a new algorithm of GCD. Their algorithm is based on a symmetric indefinite factorization obtained by a so-called bounded Bunch-Kaufman (BBK) pivoting strategy, Ashcraft, Grimes and Lewis (1998). Cheng and Higham (1998) showed that this algorithm is effective and competitive with Gill and Murray's algorithm and Schnabel and Eskow's algorithm. We use Cheng and Higham's algorithm for computing of the asymptotic variances.

Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, Cheng and Higham (1998) defined the distance from A to its generalised Cholesky decomposition $A+\Delta$ as

$$\mu(A,\delta) = \min\{||\Delta|| : \lambda_{\min}(A+\Delta) \geq \delta\},$$

where $\lambda_{\min}$ is the minimum eigenvalue and $\delta \geq 0$. The distances in the 2-norm and Frobenius norms are

$$\mu_2(A,\delta) = \max\{0, \delta - \lambda_{\min}(A)\}$$

and

$$\mu_F(A,\delta) = \left(\sum_{\lambda_i<\delta}(\delta-\lambda_i)^2\right)^{1/2},$$

where λ_i are the eigenvalues of A.

In Cheng and Higham (1998), the tolerance is $\delta = \sqrt{u}||A||_\infty$, where u is the value **eps** in Matlab representing the floating-point relative accuracy. The perturbations Δ

are measured using the ratio

$$r_2 = \frac{||\Delta||_2}{|\lambda_{\min}(A)|}$$

and

$$r_F = \frac{||\Delta||_F}{\mu_F(A, \delta)}.$$

Cheng and Higham's algorithm reinforce r_2 and r_F to be close to 1.

Example: WTI Crude Oil data

Let $\mathcal{A}$ be a 10×10 matrix,

$$\mathcal{A} = \begin{bmatrix} 10056 & -663858 & -22348 & 3415 & -6870 & -255 & 50 & 22348 & 389 & 42957 \\ -663858 & 1.99e{+}09 & 6.82e{+}07 & -3.79e{+}06 & 1.90e{+}07 & 6724 & -152883 & -6.82e{+}07 & -65242 & 1025727 \\ -22348 & 6.82e{+}07 & 2.34e{+}06 & -130354 & 653589 & 170 & -5258 & -2.34e{+}06 & -2219 & 39155 \\ 3415 & -3.79e{+}06 & -130354 & 37878 & -37217 & -207 & 293 & 130354 & 1247 & 420546 \\ -6870 & 19008886 & 653589 & -37217 & 206522 & -222 & -1467 & -653589 & -489 & 95574 \\ -255 & 6724 & 170 & -207 & -222 & 873 & -0.3817 & -170 & -151 & -67503 \\ 50 & -152883 & -5258 & 293 & -1467 & -0.3817 & 12 & 5258 & 5 & -88 \\ 22348 & -68174896 & -2342830 & 130354 & -653589 & -170 & 5258 & 2342830 & 2219 & -39155 \\ 389 & -65242 & -2219 & 1247 & -489 & -151 & 5 & 2219 & 6145697 & 379195 \\ 42957 & 1025727 & 39155 & 420546 & 95574 & -67503 & -88 & -39155 & 379195 & 5.74e{+}09 \end{bmatrix}.$$

$\mathcal{A}$ is the FIM calculated using WTI Crude Oil futures prices from 2001 to 2005, the corresponding parameter estimates have been reported in Section 5.1. The determinant of $\mathcal{A}$ is 0, hence, the matrix is not positive definite.

Table 3.1: Measure of Δ for FIM

	GCD	Method 1	Method 2
r_2	1	1	2
r_F	1	1	2

The ratios r_2 and r_F from GCD are calculated and compared with the values from two other methods, which represent the replacement of negative eigenvalues with a small positive value (e.g. 10^{-3}), Method 1, and, according to Method 2, the negative eigenvalue is replaced by its absolute value. The ratios r_2 and r_F are given in Table 3.1. Evidently, GCD is competitive with Method 1 which is commonly used in practice for "repairing" of indefinite covariance matrices.

Chapter 4

Simulation Study

In this chapter, the simulation study is performed for testing our KF algorithm for parameter estimation in the two-factor model. The convergence of the estimates of the model parameters is also studied. The parameter identification problem arising within the likelihood function has been explored.

4.1 Results

In this section, we present the results of the simulation study conducted for validating the use of Kalman Filter for estimation of the state vector x_t jointly with the model parameters θ. The constraint $\kappa \geq \gamma$ has been introduced to address the parameter identification problem arising in estimation of the x vector using ML method. The simulation study has been programmed as follows.

1. Set $\theta_0 = (\kappa, \gamma, \mu, \sigma_\chi, \sigma_\xi, \rho, s)$ as the vector of true values.

2. Simulate x_t and y_t using the true values of parameters set in θ_0.

3. Set the intervals for searching for unknown parameters. Locate the grid over the Cartesian product of these intervals.

4. Do grid-search for finding the best initial vector θ_{init}.

5. Maximise the log-likelihood function $l(\theta; y)$ using the best initial vector θ_{init}.

6. Obtain the estimates of the model parameters $\hat{\theta}$ and their standard errors.

The grid search for the best initial set of the parameters' values allowed overcoming the problem of sensitivity to the initial values. Further, for simplicity we assume $\lambda_\chi = \lambda_\xi = 0$ and $s_1^2 = s_2^2 = ... = s_n^2 = s_v^2$. The vector y_t is 13-dimensional, i.e. there are 13 simulated futures contracts. The model parameter estimates and the corresponding standard errors obtained by using the above procedure are presented in Table 4.1. The true values of parameters are given in the last row of Table 4.1.

Table 4.1: $\hat{\theta}$ for different sample size with SE; NLL stands for $-l(\theta; y)$

n	κ	γ	μ	σ_χ	σ_ξ	ρ	s_v	NLL
500	1.2568	0.0120	-0.0506	0.7830	0.2568	-0.9381	0.0299	-13129
	(0.0199)	(0.0342)	(0.0425)	(0.0480)	(0.0196)	(0.0164)	(0.0003)	
1000	1.2407	0.0100	-0.0121	0.6243	0.1446	-0.8895	0.0300	-26283
	(0.0127)	(0.0282)	(0.0330)	(0.0295)	(0.0097)	(0.0239)	(0.0002)	
2000	1.5627	1.0068	-2.0290	0.4587	0.3762	-0.4127	0.0297	-52830
	(0.0468)	(0.0260)	(0.0512)	(0.0260)	(0.0188)	(0.0540)	(0.0001)	
4000	1.4555	1.0092	-2.0229	0.5448	0.4224	-0.6837	0.0300	-105576
	(0.0308)	(0.0193)	(0.0381)	(0.0205)	(0.0104)	(0.0292)	(0.0001)	
6000	1.4252	0.9773	-1.9609	0.5276	0.3738	-0.6719	0.0300	-158484
	(0.0263)	(0.0166)	(0.0330)	(0.0181)	(0.0073)	(0.0295)	(0.0001)	
8000	1.4741	1.0004	-2.0074	0.5029	0.3633	-0.6503	0.0300	-211259
	(0.0231)	(0.0150)	(0.0299)	(0.0142)	(0.0068)	(0.0252)	(7.3e-05)	
10000	1.4630	0.9721	-1.9500	0.5080	0.3529	-0.6718	0.0300	-264100
	(0.0202)	(0.0127)	(0.0253)	(0.0128)	(0.0052)	(0.0225)	(6.5e-05)	
16000	1.4917	0.9791	-1.9631	0.4974	0.3248	-0.6573	0.0301	-422708
	(0.0162)	(0.0101)	(0.0201)	(0.0101)	(0.0039)	(0.0202)	(5.0e-05)	
θ_0	1.5	1	-2	0.5	0.3	-0.7	0.03	

As the sample size increses from 500 to 16000, all estimates of parameters converge to their true values. When the sample size is large, all estimates were significant, except σ_ξ. However, as the sample size increases, the difference between the true value of σ_ξ and the width of the 95% confidence interval of the estimates of σ_ξ decreases. It is reasonable to conclude that if the sample size increases continuously, the estimates of σ_ξ will finally be significant.

For some sample sizes n from the range $(500, 16000)$, the best initial values found according to Step 5, are given in Table 4.2. These initial values were used for obtaining the corresponding optimal model parameter estimates in Table 4.1.

Table 4.2: Best initial values for $\hat{\theta}$

n	κ	γ	μ	σ_χ	σ_ξ	ρ	s
500	2.2525	0.7575	-0.5000	1.5025	1.0050	0.5000	0.5000
1000	1.5050	0.7575	-0.5000	0.50750	1.0050	0.5000	0.5000
2000	1.5050	0.7575	-2.7500	0.5075	1.5025	0	0.500
4000	1.5050	0.7575	1.7500	1.5025	1.5025	0	0.5000
6000	2.2525	1.5050	-2.7500	1.5025	1.5025	0.5000	0.7500
8000	2.2525	0.7575	-0.5000	1.0050	1.0050	-0.5000	0.7500
10000	1.5050	1.5050	1.7500	1.5025	1.5025	0.5000	0.7500
16000	0.7575	2.2525	-2.7500	0.5075	0.5075	0	0.2500

The convergence of the parameter estimates can be seen in Figure 4·1, where the estimation errors $\hat{\theta}_i - \theta_{0,i}, i = 1, 2, ..., 7$ are plotted versus the sample size n, with θ_0 be the vector of true parameter values.

The paths of the estimated state variables $\hat{\chi}_t$ and $\hat{\xi}_t$ obtained through Kalman Filter along with the simulated trajectories χ_t and ξ_t are presented in Figure 4·2 and 4·3.

The plots of the paths of the estimated $\log \hat{S}_t = \hat{\chi}_t + \hat{\xi}_t$ and simulated spot prices $\log S_t = \chi_t + \xi_t$ are presented in Figure 4·4.

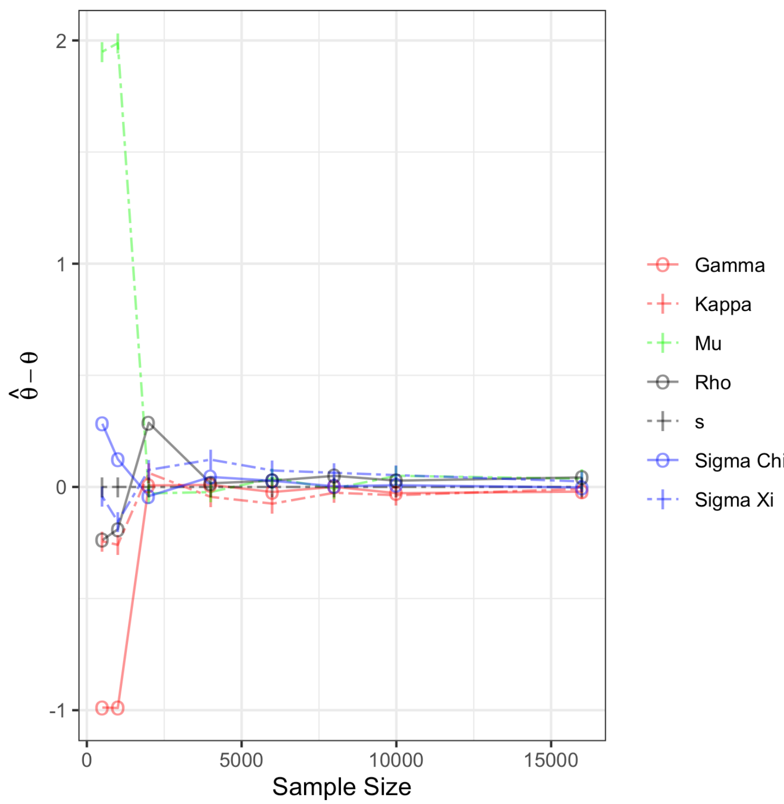

Figure 4·1: θ_0 estimation error plots componentwise versus sample size n

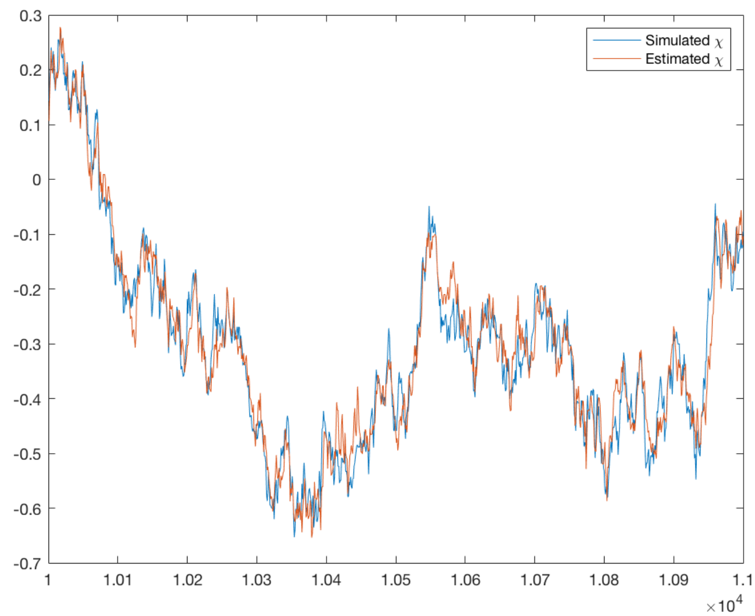

Figure 4·2: Estimated ($n = 16000$) $\hat{\chi}_t$ and simulated χ_t (data points: $10001 - 11000$)

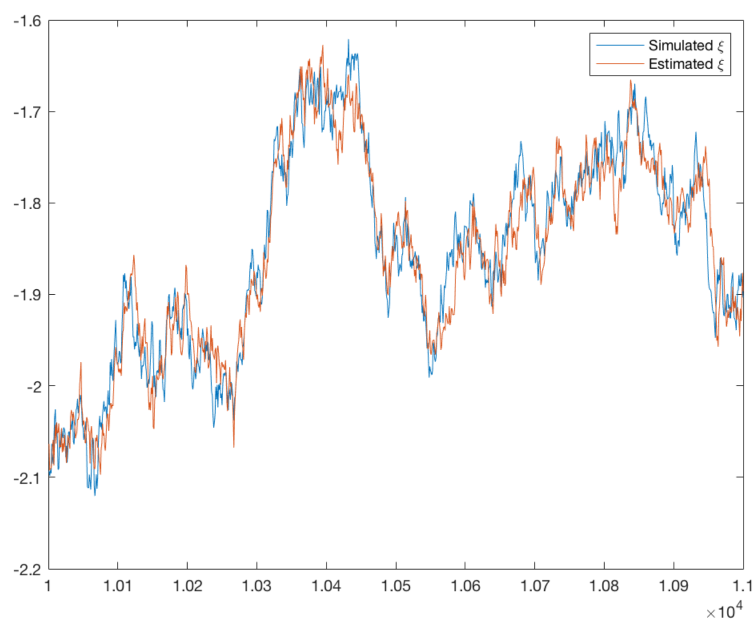

Figure 4·3: Estimated ($n = 16000$) $\hat{\xi}_t$ and simulated ξ_t (data points: $10001 - 11000$)

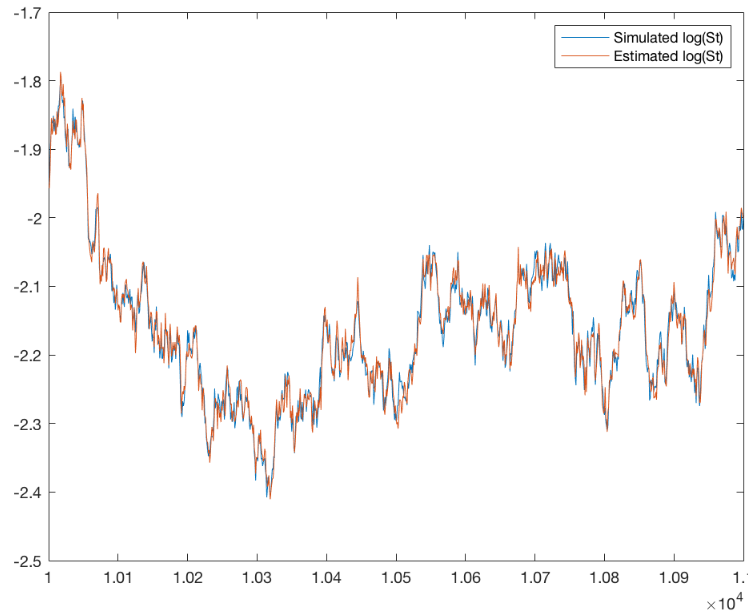

Figure 4·4: Estimated ($n = 16000$) $\log \hat{S}_t$ and simulated $\log S_t$ (data points: $10001 - 11000$)

4.2 Parameter Identification Problem

In the process of obtaining the MLE estimates, the parameters may be identified incorrectly due to what is known as the Parameter Identification Problem (PIP). In this section, we will discuss the presence of this problem in our model and then we suggest a reasonable numerical resolution.

To obtain the MLE estimates of parameters, the log-likelihood function

$$l(\theta; y) = -\frac{nn_T \log 2\pi}{2} - \frac{1}{2}\sum_{t=1}^{n_T}\left[\log\left(\det(L_{t|t-1})\right) + e_t' L_{t|t-1}^{-1} e_t\right]$$

is maximised. However, it is possible that the log-likelihood function has the same value at two different vectors of estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, which gives

$$l(\hat{\theta}_1; y) = l(\hat{\theta}_2; y). \tag{4.1}$$

In Section 2.2, both the short-term factor χ and the long-term factor ξ are modelled as Ornstein-Uhlenbeck processes with different means,

$$d\chi_t = -\kappa \chi_t dt + \sigma_\chi dZ_t^\chi$$

and

$$d\xi_t = (\mu_\xi - \gamma \xi_t)dt + \sigma_\xi dZ_t^\xi.$$

Let $\hat{\theta}_1 = \{\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho, s\}$ and $\hat{\theta}_2 = \{\gamma, \kappa, \tilde{\mu}_\xi, \sigma_\xi, \sigma_\chi, \rho, s\}$. That is, we swap the parameters for χ and ξ. Since ρ represents the correlation coefficient of dZ_t^χ and dZ_t^ξ, it is invariant to swapping. s is also invariant to swapping. μ_ξ is changed to another value $\tilde{\mu}_\xi$. It can be shown numerically and analytically that if we swap parameters this way, the MLE estimates $\tilde{\mu}_\xi$ would be

$$\tilde{\mu}_\xi = \frac{\kappa}{\gamma} \mu_\xi, \tag{4.2}$$

and then (4.1) satisfied.

For numerical evidence, the algorithm is given below.

1. Generate random values for $\hat{\theta}_1 = \{\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho, s\}$. In this step, any reasonable distributions can be considered. However, due to the continuity property and, the fact, most parameters are bounded (e.g. the absolute value of correlation coefficient ρ must less than 1), the uniform distribution is assumed for all parameters.

2. Simulate data with respect to $\hat{\theta}$ with sample size n.

3. Swap parameters as discussed above. We get $\hat{\theta}_2 = \{\gamma, \kappa, \frac{\kappa}{\gamma}\mu_\xi, \sigma_\xi, \sigma_\chi, \rho, s\}$.

4. Calculate log-likelihood for $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively.

5. Repeat step $1 - 3$ m times.

6. Calculate the sum of squares of the difference of $l(\hat{\theta}_1)$ and $l(\hat{\theta}_2)$, where l is the log-likelihood function.

We used the sample size $n = 2000$ and the number of iterations $m = 1000$, which gives the sum of squares 6.8×10^{-15}. Table 4.3 displays the results of the first 10 iterations. It is clear that the log-likelihoods with respect to $\hat{\theta}_1$ and $\hat{\theta}_2$ are identical.

Table 4.3: First 10 iterations of numerical illustration of PIP

κ	γ	μ_ξ	σ_χ	σ_ξ	ρ	s	$l(\hat{\theta}_1)$	$l(\hat{\theta}_2)$
2.79	0.95	-3.34	0.41	1.14	0.19	0.96	90471.60	90471.60
1.69	0.64	-2.23	1.24	1.87	-0.34	0.86	83064.45	83064.45
1.35	2.62	-2.72	1.13	1.40	-0.59	0.65	64519.77	64519.77
1.62	0.25	1.39	1.99	0.62	-0.64	0.51	50036.07	50036.07
0.59	2.21	2.83	1.19	0.77	0.54	0.47	44343.66	44343.66
2.22	2.49	-3.43	0.04	1.11	0.17	0.10	-52740.50	-52740.50
1.08	2.68	-3.59	0.83	0.30	-0.04	0.69	68948.71	68948.71
2.83	2.49	2.46	0.77	0.55	-0.50	0.86	82940.70	82940.70
2.42	0.67	1.26	1.53	0.68	-0.54	0.13	-38563.71	-38563.71
0.88	1.62	-3.28	1.58	0.52	0.87	0.21	-5601.71	-5601.71

The analytical proof is provided below. The log-likelihood (2.12) is a function of prediction error e_t and covariance matrix $L_{t|t-1}$, so if e_t and $L_{t|t-1}$ under $\hat{\theta}_1$ and $\hat{\theta}_2$ are identical, the log-likelihood $l(\hat{\theta}_1)$ and $l(\hat{\theta}_2)$ would be identical.

By Mathematical induction, it can be proved that when we swap κ and γ, and σ_χ and σ_ξ, $\tilde{\mu}_\xi$ is equal to $\frac{\kappa}{\gamma}\mu_\xi$. And then, it can be proved that e_t and $L_{t|t-1}$ would be identical under $\hat{\theta}_1$ and $\hat{\theta}_2$. The detailed proof is available in appendix C. To deal with this problem, a constraint $\kappa \geq \gamma$ is introduced.

Chapter 5

Application: Crude Oil Futures Data

In this section, the Kalman Filter and Smoother are implemented using real data. The in-sample and out-of-sample forecasting performances are also discussed.

We use historical data of prices of WTI Crude Oil NYMEX futures from 02/01/1996 to 30/09/2019. It includes 20 contracts with maturities 1 month to 20 months respectively. The parameters are estimated based on the first 13 contracts and the last 7 contracts are used to study the model out-of-sample forecasting performance. The data set was provided by DataScope.

In Section 5.1, the parameter estimations of Crude Oil futures data are reported. Considering the effect of the global financial crisis (GFC), a moving window with 4 years period and 1 year shift is used. This moving window will provide a view of the changes of parameters in different periods. The in-sample and out-of-sample performances are analysed using the root mean square error (RMSE) in Section 5.2. In Section 5.3, the model performance on raw data is compared with the model performance on interpolated data. The interpolation is a commonly used technique in time series analysis.

5.1 Parameter Estimation

In this section, a moving window over 4 years period with 1-year shift period is used in an observational study of the dynamics of the parameter estimations. This design is efficient for assessment of the effect on the model parameters of some global events, such as the GFC.

Table 5.1: Estimated parameters in different time periods

Period	κ	γ	μ_ξ	σ_χ	σ_ξ	ρ	λ_χ	λ_ξ
2015-2019	1.0784	0.0091	-0.0665	0.2475	0.2584	0.3994	-0.5826	-0.0061
	(0.0074)	(0.0056)	(0.0250)	(0.0062)	(0.0063)	(0.0056)	(0.0017)	(4.72E-05)
2014-2018	1.1293	0.0046	-3.415	0.2441	0.2389	0.4530	-3.5956	-3.3445
	(0.0081)	(0.0140)	(0.0222)	(0.0065)	(0.0059)	(0.0140)	(5.70E-05)	(0.0018)
2013-2017	1.2393	0.0459	0.0099	0.2314	0.2466	0.4010	-0.0636	-0.1080
	(0.0090)	(0.0019)	(0.0062)	(0.0244)	(0.0027)	(0.0060)	(0.0027)	(3.12E-04)
2012-2016	1.2069	0.1289	0.4784	0.1854	0.2413	0.1814	-4.9790	0.5175
	(0.0093)	(0.0020)	(0.0060)	(0.0341)	(0.0046)	(0.0021)	(2.25E-04)	(0.0021)
2011-2015	0.5753	0.5282	-1.3102	2.8602	2.9994	-0.9970	-4.9979	1.0311
	(0.0750)	(0.0032)	(0.0054)	(0.0812)	(0.0034)	(2.88E-04)	(0.0050)	(0.0054)
2010-2014	0.7223	0.4180	-0.1068	0.4309	0.5881	-0.9080	-3.2860	0.0239
	(0.0073)	(0.0041)	(0.0175)	(0.0114)	(0.0100)	(0.0085)	(0.0049)	(0.0085)
2009-2013	2.9491	0.1966	0.8685	0.1598	0.3283	0.0440	2.3081	-0.1323
	(0.0191)	(0.0023)	(0.0074)	(0.0438)	(7.95E-04)	(0.0020)	(1.56E-04)	(0.0023)
2008-2012	2.0965	0.1326	0.1679	0.1952	0.3728	0.1643	-4.9486	-0.0622
	(0.0139)	(0.0010)	(0.0085)	(0.0382)	(0.0020)	(0.0029)	(1.30E-04)	(0.0020)
2007-2011	1.6106	0.0555	0.0100	0.2241	0.3309	0.2935	-4.9896	-0.0135
	(0.0099)	(0.0011)	(0.0077)	(0.0310)	(0.0014)	(0.0046)	(5.30E-04)	(0.0014)
2006-2010	1.5025	0.0242	0.0049	0.2358	0.3118	0.3691	-0.6028	-0.0325
	(0.0098)	(0.0013)	(0.0074)	(0.0268)	(0.0015)	(0.0054)	(0.0015)	(5.32E-04)
2005-2009	1.2082	0.0027	-0.9515	0.2088	0.2811	0.3062	0.6292	-0.8723
	(0.0103)	(0.0236)	(0.0236)	(0.0299)	(0.0071)	(0.005)	(4.10E-04)	(0.0016)
2004-2008	1.2659	0.0047	-0.4566	0.1839	0.2203	0.3875	-4.9447	-0.3659
	(0.0111)	(0.0186)	(0.0057)	(0.0281)	(0.005)	(0.0019)	(0.0186)	(6.90E-05)
2003-2007	1.1373	0.0889	0.0039	0.2491	0.2222	0.1569	0.6550	-0.1924
	(0.0086)	(0.0026)	(0.0057)	(0.0264)	(0.0029)	(0.006)	(2.30E-04)	(0.0029)
2002-2006	1.4316	0.0552	-0.0205	0.2722	0.2208	0.2997	-3.3817	0.0075
	(0.0086)	(0.0021)	(0.0056)	(0.025)	(9.40E-04)	(0.0066)	(9.40E-04)	(4.30E-04)
2001-2005	1.5117	0.0558	-0.0502	0.3036	0.2201	0.4222	-4.0223	0.0093
	(0.0097)	(0.0023)	(0.0057)	(0.0205)	(0.0014)	(0.0079)	(4.70E-04)	(0.0014)
2000-2004	1.5368	0.2187	-0.004	0.2985	0.2696	0.0758	-3.713	-0.0588
	(0.0105)	(0.0026)	(0.0067)	(0.0289)	(0.002)	(0.0067)	(2.84E-04)	(0.002)
1999-2003	1.0755	0.3321	-0.2969	0.4055	0.3648	-0.4456	-3.411	-0.1596
	(0.0075)	(0.0235)	(0.0085)	(0.0076)	(0.0037)	(0.0033)	(0.0011)	(0.0037)

Table 5.2: Estimated parameters in different time periods

Period	s_1	s_2	NLL
2015-2019	0.0132	0.0027	-53082
	(3.15E-04)	(1.38E-05)	
2014-2018	0.0133	0.0029	-52450
	(3.20E-04)	(1.50E-05)	
2013-2017	0.0138	0.0030	-52119
	(9.82E-05)	(1.61E-05)	
2012-2016	0.0105	0.0024	-54807
	(2.36E-04)	(1.11E-05)	
2011-2015	0.0091	0.0023	-55552
	(2.03E-04)	(1.73E-05)	
2010-2014	0.0100	0.0023	-55428
	(2.24E-04)	(1.11E-05)	
2009-2013	0.0179	0.0028	-52505
	(4.06E-04)	(1.49E-05)	
2008-2012	0.0226	0.0030	-51018
	(5.20E-04)	(1.72E-05)	
2007-2011	0.0236	0.0030	-51034
	(4.87E-05)	(1.67E-05)	
2006-2010	0.0238	0.0032	-50170
	(2.39E-05)	(1.92E-05)	
2005-2009	0.0181	0.0032	-50717
	(5.30E-05)	(1.80E-05)	
2004-2008	0.0149	0.0028	-51944
	(3.40E-04)	(1.60E-05)	
2003-2007	0.0167	0.0029	-51578
	(3.80E-04)	(1.60E-05)	
2002-2006	0.0189	0.0032	-50662
	(3.50E-05)	(1.90E-05)	
2001-2005	0.0209	0.0037	-48562
	(5.20E-05)	(2.20E-05)	
2000-2004	0.0206	0.0037	-48518
	(4.63E-04)	(2.24E-05)	
1999-2003	0.0206	0.0033	-49585
	(4.58E-04)	(1.93E-05)	

The parameter estimates and the corresponding standard errors are given in Table 5.1 and 5.2, along with the negative log-likelihood function values. For Crude Oil Futures data, we assume that the covariance matrix V of error term in y_t is diagonal, and has the form

$$V = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \ddots & \\ & & & s_2^2 \end{bmatrix}.$$

That is, the variance of the error term for the first contract is s_1^2 and s_2^2 for other contracts.

The estimates of κ are stable on the interval $[1, 1.6]$, except the periods from 2008-2012 to 2011-2015. In these periods, $\hat{\kappa}$ changes from about 0.6 to 2.9. Comparing to $\hat{\kappa}$, the range of $\hat{\gamma}$ is larger, from only 0.003 to about 0.5. This illustrates that the speed of mean-reverting for short factor is more stable than the speed of mean-reverting for long factor. The estimates of σ_χ and σ_ξ are very close and fluctuate at about 0.3, except the period 2011-2015, in which the estimates are close to 3. In 1999-2003, 2010-2014 and 2011-2015, the estimates of ρ are negative whilst otherwise positive, which illustrates the short and long term factors are usually positively correlated. The estimates of s_1 and s_2 are stable over all periods. However, the estimates of μ_ξ, λ_χ and λ_ξ are sensitive to varying time periods. This observational study conformed with our expectations. The model was designed in the risk-neutral framework

$$d\chi_t = (-\kappa\chi_t - \lambda_\chi)dt + \sigma_\chi dZ_t^{\chi^*},$$

$$d\xi_t = (\mu_\xi - \gamma\xi_t - \lambda_\xi)dt + \sigma_\xi dZ_t^{\xi^*},$$

where the risk premiums λ_χ and λ_ξ were the adjustment scalars for the mean levels of short-term and long-term factors. Hence, it was expected to experience difficulties in the estimation of λ's separately from μ.

5.2 In-Sample and Out-Of-Sample Performances

In Section 5.1, the parameters are estimated by maximising the log-likelihood function. In this section, the in-sample and out-of-sample performances are analysed using the RMSE criterion.

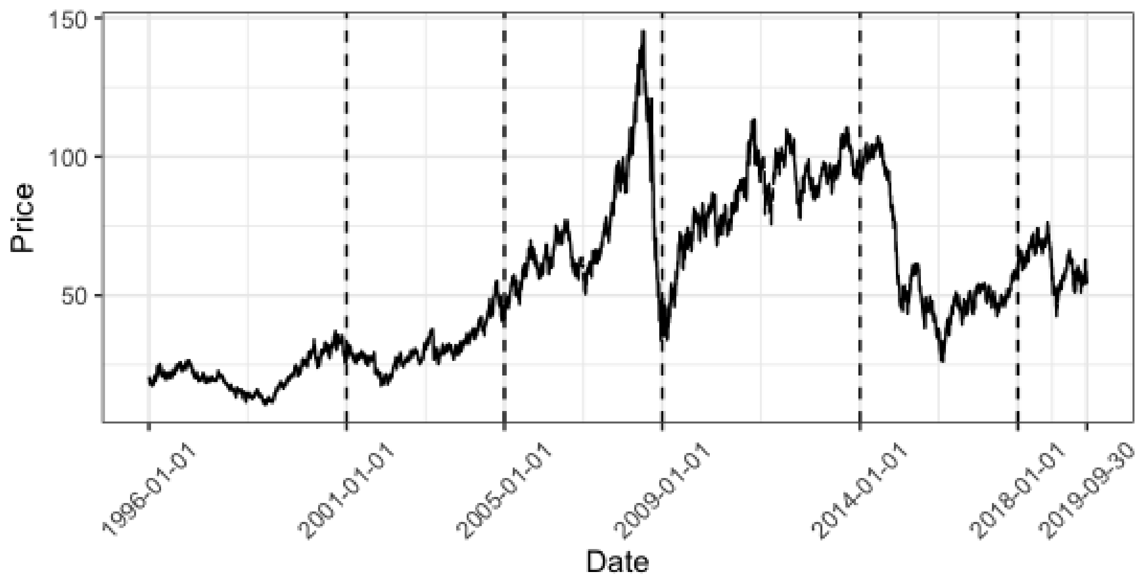

Figure 5·1: WTI Crude Oil futures prices of the first available contract

Figure 5·1 shows the WTI Crude Oil futures prices from 1996 to 2019. It is obvious that the prices dropped dramatically during the Global Financial Crisis in 2008, so the stationarity of the time series may not be satisfied during this period. To study the in-sample and the out-of-sample forecasting performance, the three separate periods are selected, 01/01/2001 - 01/01/2005, 01/01/2005 - 01/01/2009, and 01/01/2014 - 01/01/2018 in Figure 5·1.

The root mean square error (RMSE) would be a criteria for studying the performance. The algorithm for obtaining RMSE are given below:

1. Given the data set in some specified time period, the parameters are estimated.

2. Obtain the estimates of state variable x_t trough Klaman FIlter or Kalman Smoother.

3. Obtain the estimates of the logarithms of futures prices y_t by

$$\hat{y}_t = d_t + F_t' \hat{x}_t.$$

d_t and F_t are given in Section 2.4. $\hat{x}_t$ is the estimate of x_t by Kalman Filter or Kalman Smoother.

4. Calculate RMSE of $y_t - \hat{y}_t$.

Table 5.3: RMSE with data in three different time periods and different estimation methods

Period		2001-2005		2005-2009		2014-2018	
Estimation		Filter	Smoother	Filter	Smoother	Filter	Smoother
In-Sample	C6	0.003786	0.003790	0.003361	0.003362	0.003044	0.003042
	C12	0.002753	0.002768	0.002578	0.002597	0.002146	0.002144
Out-of-Sample	C14	0.005960	0.005953	0.005667	0.005663	0.005231	0.005211
	C15	0.007894	0.007885	0.007462	0.007449	0.007208	0.007182
	C16	0.009770	0.009761	0.009418	0.009398	0.009451	0.009420
	C17	0.011798	0.011785	0.011508	0.011483	0.011905	0.011870
	C18	0.013883	0.013870	0.013658	0.013628	0.014498	0.014462
	C19	0.016108	0.016096	0.015913	0.015877	0.017270	0.017233

Table 5.3 gives the RMSE of data in different time periods. The in-sample forecasting performance is evaluated on the 6th (C6) and 12th (C12) contracts, while the out-of-sample performance is evaluated on the 14th (C14), 15th (C15), 16th (C16), 17th (C17), 18th (C18) and 19th (C19) contracts. The RMSE's are consistent and reasonable in three periods, even for the period from 2005 to 2009, where the futures prices decreased dramatically due to the GFC. In summary, the forecasting performances of different estimation methods have been studied. In each specified time period, the RMSE calculated through Kalman Filter is smaller for short maturity contracts, which provides evidence that the Kalman Filter performs better in estimations of futures prices for short maturity contracts, while Kalman Smoother is better for longer maturity contracts.

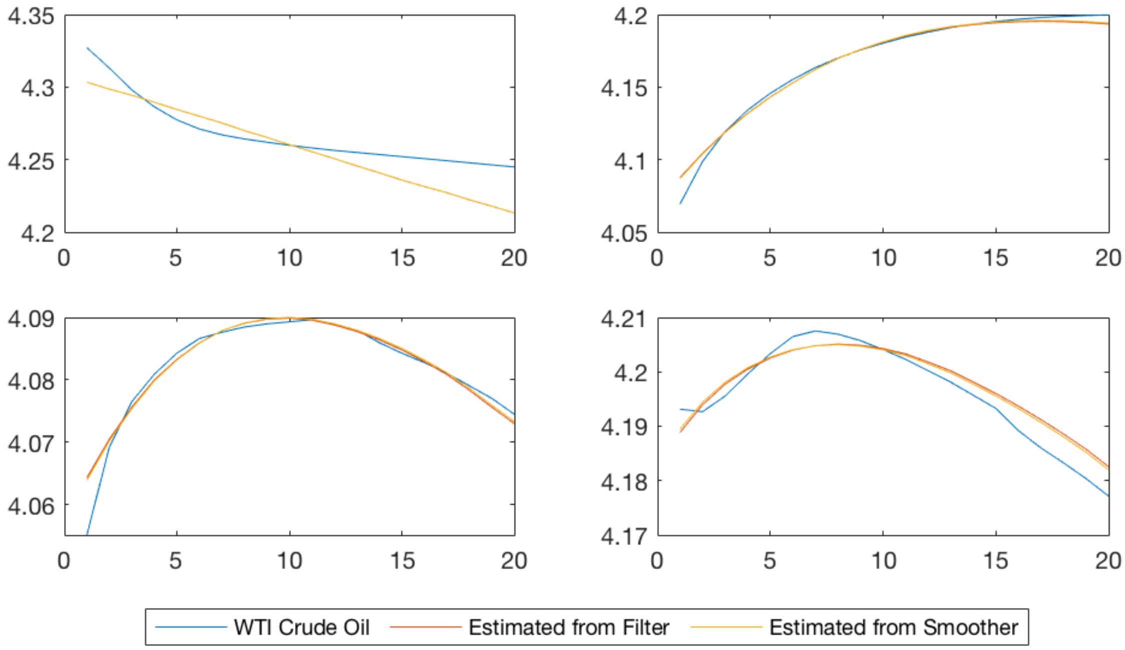

Figure 5·2: Cross-sectional data of logarithm of futures prices and estimates on 4 different days

Figure 5·2 gives the cross-sectional data about the logarithm of futures prices and estimated the logarithm of prices from Kalman Filter and Kalman Smoother on 4 different days with distinct patterns of futures curves, 10/10/2006, 05/09/2007, 14/11/2005 and 20/09/2005. The horizontal axis represents the number of contracts from 1 to 20 and the vertical axis is the logarithm of prices. On these 4 days, the tendencies are different. However, both Kalman Filter and Kalman Smoother gave reasonable estimates.

Figure 5·3 shows the logarithm of the futures prices of the first available contract and the estimated logarithm of prices from Kalman Filter and Smoother from 01/01//2005 to 01/01/2009. Both Kalman Filter and Kalman Smoother gave good estimates for this period.

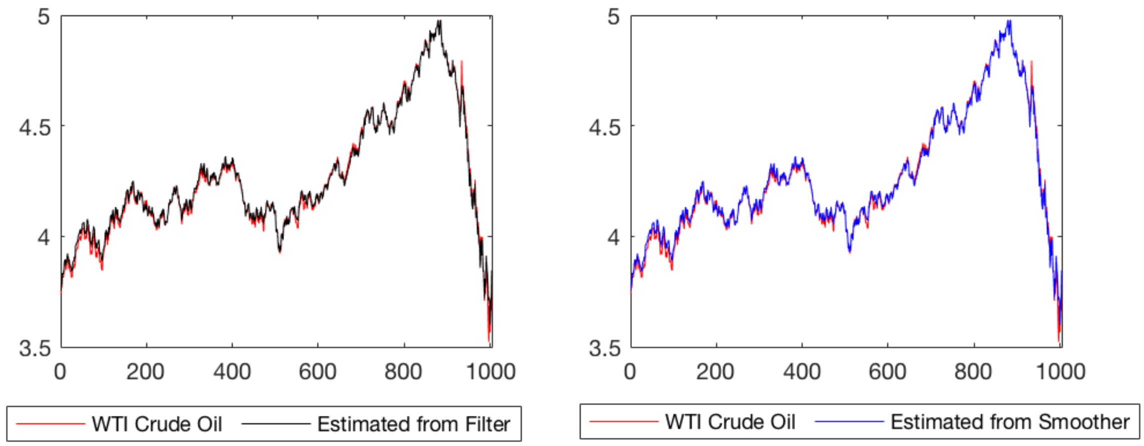

Figure 5·3: Logarithm of futures prices and estimates of the first available contract from 2005 to 2009

5.3 Data Interpolation

Interpolation is commonly used in time series analysis to create a new data point within the discrete known data points. Given the futures prices y_1 and y_2 with two distinct maturities t_1 and t_2 $(t_1 \leq t_2)$ on a specified date, the linear interpolation $\tilde{y}$ with maturity $\tilde{t}$ gives

$$\tilde{y} = y_1 + \frac{\tilde{t} - t_1}{t_2 - t_1}(y_2 - y_1), \tag{5.1}$$

where $t_1 \leq \tilde{t} \leq t_2$. In this section, the raw data is interpolated so that the maturities of each contract over the whole period are fixed, i.e. the maturity of the first contract on each trading day is "one month" (i.e. 21 days), the maturity of the second contract on each trading day is "two months" (i.e. 42 days), etc.

The RMSE's of interpolated data are given in Table 5.4. For interpolated data, the conclusions are similar to the conclusions made in reference of raw data. The RMSE's are consistent across the three periods. The Kalman Filter gives the best RMSE for short maturity contracts, and Kalman Smoother is better for long-maturity contracts. Moreover, the RMSE's in Table 5.4 are larger than the RMSE's in Table

Table 5.4: RMSE with interpolated data in three different time periods and different estimation methods

Period		2001-2005		2005-2009		2014-2018	
Estimation		Filter	Smoother	Filter	Smoother	Filter	Smoother
In-Sample	C6	0.003037	0.003040	0.002604	0.002604	0.002351	0.002349
	C12	0.003575	0.003587	0.003196	0.003201	0.002889	0.002877
Out-of-Sample	C14	0.007065	0.007068	0.006243	0.006232	0.006322	0.006299
	C15	0.008869	0.008871	0.008031	0.008014	0.008453	0.008426
	C16	0.010700	0.010700	0.009975	0.009952	0.010860	0.010828
	C17	0.012634	0.012635	0.012015	0.011988	0.013440	0.013407
	C18	0.014722	0.014724	0.014162	0.014130	0.016192	0.016158
	C19	0.016955	0.016958	0.016424	0.016388	0.019157	0.019122

5.3, except for the 6th contract (C6). Hence, the interpolation of the data can be recommended for modelling of futures prices with shorter maturities.

Chapter 6

Conclusion

We have developed the two-factor model which can be used for the pricing of energy commodity futures. The Kalman Filter has been implemented to estimate the hidden factors jointly with unknown model parameters. The asymptotic variances of the estimates of the model parameters were calculated using the score vector.

The simulation study has been carried out to test the sensitivity to the initial values and consistency of the estimation procedure. Through the simulation study, we illustrated the robustness of the grid-search and consistency of the estimates of the parameters and the estimates of the state variables χ_t and ξ_t. The parameter identification problem established in this model has been resolved numerically by introducing an additional constraint.

The model has been applied to WTI Crude Oil futures historical prices from 1999 to 2019. A moving window was used for monitoring the change of parameters over time. Most of the parameters were stable over time except μ_ξ, λ_χ and λ_ξ. These three parameters appeared difficult to estimate. The asymptotic variances of the estimates were obtained using an empirical analog of the Fisher Information matrix and when it returned indefinite, the matrix was "repaired" using the generalised Cholesky decomposition. The model in-sample and out-of-sample forecasting performances were evaluated using the RMSE criterion. Moreover, Kalman Filter gives a better estimate of state vector x_t for shorter maturity contracts, while Kalman Smoother performs better for contracts with longer maturities.

The topics for further research are as follows:

1. The two-factor model can be extended to a multi-factor model, which would include seasonal effects and stochastic interest rate.
2. Further, we aim to compare the findings of this work with the results which will be obtained through implementation of the Particle Filter for the two-factor model.

References

Ames, M., Bagnarosa, G., Matsui, T., Peters, G. W., and Shevchenko, P. V. (2020). Which risk factors drive oil futures price curves? *Energy Economics*, page 104676.

Ashcraft, C., Grimes, R. G., and Lewis, J. G. (1998). Accurate symmetric indefinite linear equation solvers. *SIAM Journal on Matrix Analysis and Applications*, 20(2):513–561.

Bierman, G. J. (1973). Fixed interval smoothing with discrete measurements. *International Journal of Control*, 18(1):65–75.

Binkowski, K., He, P., Kordzakhia, N., and Shevchenko, P. (2019). On the parameter estimation in the Schwartz-Smith's two-factor model. In Nguyen, H., editor, *Communications in Computer and Information Science*, volume 1150. RSSDS 2019, Springer.

Binkowski, K., Shevchenko, P., and Kordzakhia, N. (2009). Modelling commodity prices. CSIRO technical report CMIS 09/43.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179.

Carmona, R. and Coulon, M. (2014). A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer.

Cheng, B., Nikitopoulos, C. S., and Schlögl, E. (2018). Pricing of long-dated commodity derivatives: Do stochastic interest rates matter? *Journal of Banking and Finance*, 95:148–166.

Cheng, S. H. and Higham, N. J. (1998). A modified Cholesky algorithm based on a symmetric indefinite factorization. *SIAM Journal on Matrix Analysis and Applications*, 19(4):1097–1110.

Cortazar, G., Millard, C., Ortega, H., and Schwartz, E. S. (2019). Commodity price forecasts, futures prices, and pricing models. *Management Science*, 65(9):4141–4155.

Cortazar, G. and Schwartz, E. S. (2003). Implementing a stochastic model for oil futures prices. *Energy Economics*, 25(3):215–238.

Davis, M. (2013). *Stochastic modelling and control*. Springer Science & Business Media.

De Jong, P. (1989). Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association*, 84(408):1085–1088.

Devore, J. L. and Berk, K. N. (2012). *Modern mathematical statistics with applications*. Springer.

Durbin, J. and Koopman, S. J. (2002). *Time series analysis by state space methods*. Oxford University Press.

Ewald, C.-O., Zhang, A., and Zong, Z. (2019). On the calibration of the Schwartz two-factor model to WTI crude oil options and the extended Kalman filter. *Annals of Operations Research*, 282(1-2):119–130.

Farkas, W., Gourier, E., Huitema, R., and Necula, C. (2017). A two-factor cointegrated commodity price model with an application to spread option pricing. *Journal of Banking & Finance*, 77:249–268.

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, 37(2):200–220.

Geman, H. (2009). *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy*. John Wiley & Sons.

Gibson, R. and Schwartz, E. S. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3):959–976.

Gilbert, P. (2005). Brief user's guide: Dynamic systems estimation (dse). *Available in the file doc/dse-guide, downloaded from http://cran.r-project.org.*

Gill, P. E. and Murray, W. (1974). Newton-type methods for unconstrained and linearly constrained optimization. *Mathematical Programming*, 7(1):311–350.

Guo, Z.-Y. et al. (2017). *Models with short-term variations and long-term dynamics in risk management of commodity derivatives*. ZBW.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Helske, J. (2016). Kfas: Exponential family state space models in R. *Preprint arXiv:1612.01907*.

Hull, J. C. (2003). *Options futures and other derivatives*. Pearson Education India.

Koopman, S. J. and Shephard, N. (1992). Exact score for time series models in state space form. *Biometrika Cambridge*, 79:823–823.

Kummer, S. and Pauletto, C. (2012). The history of derivatives: A few milestones. In *EFTA Seminar on Regulation of Derivatives Markets*, pages 431–466.

Kutoyants, Y. A. (2019). On parameter estimation of the hidden Ornstein-Uhlenbeck process. *Journal of Multivariate Analysis*, 169:248–263.

Lennart, L. and Caines, P. E. (1979). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics*, 3(1-4):29–46.

Peters, G. W., Briers, M., Shevchenko, P., and Doucet, A. (2013). Calibration and filtering for multi factor commodity models with seasonality: incorporating panel data from futures contracts. *Methodology and Computing in Applied Probability*, 15(4):841–874.

Rauch, H. E., Striebel, C., and Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450.

Schnabel, R. B. and Eskow, E. (1990). A new modified Cholesky factorization. *SIAM Journal on Scientific and Statistical Computing*, 11(6):1136–1158.

Schwartz, E. and Smith, J. E. (2000). Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911.

Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, 52(3):923–973.

Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 22(5):393–426.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, 86(416):1024–1033.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian Motion. *Physical Review*, 36(5):823.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Appendix A

Derivation of Characteristics of Bivariate OU Process

In Section 2.1, we define a bivariate Ornstein-Uhlenbeck process as

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dZ_t^\chi \tag{A.1}$$

and

$$d\xi_t = (\mu_\xi - \gamma\xi_t)dt + \sigma_\xi dZ_t^\xi, \tag{A.2}$$

where Z_t^χ, Z_t^ξ are correlated standard Brownian motions with $E(dZ_t^\chi dZ_t^\xi) = \rho dt$. Here we provide a somewhat more detailed derivation of (2.5) and (2.6). We will show that (2.5) and (2.6) can be obtained based on the limits of the discretised variants of (A.1) and (A.2) instead of directly using their exact solutions.

Firstly, from (A.1),

$$\Delta\chi_t = -\kappa\chi_t\Delta t + \sigma_\chi\sqrt{\Delta t}\epsilon_\chi.$$

Therefore,

$$\chi_{t+1} = (1 - \kappa\Delta t)\chi_t + \sigma_\chi\sqrt{\Delta t}\epsilon_\chi. \tag{A.3}$$

Similarly, from (A.2), we get

$$\xi_{t+1} = (1 - \gamma\Delta t)\xi_t + \mu_\xi\Delta t + \sigma_\xi\sqrt{\Delta t}\epsilon_\xi, \tag{A.4}$$

where $\epsilon_\chi, \epsilon_\xi \sim N(0,1)$. Let $Corr(\epsilon_\chi, \epsilon_\xi) = \rho$ and $w = \begin{bmatrix} \sigma_\chi \sqrt{\Delta t} \epsilon_\chi \\ \sigma_\xi \sqrt{\Delta t} \epsilon_\xi \end{bmatrix}$, then

$$Var(w) = \begin{bmatrix} \sigma_\chi^2 \Delta t & \rho \sigma_\chi \sigma_\xi \Delta t \\ \rho \sigma_\chi \sigma_\xi \Delta t & \sigma_\xi^2 \Delta t \end{bmatrix} = W.$$

Let $X_t = \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix}$, $c = \begin{bmatrix} 0 \\ \mu_\xi \Delta t \end{bmatrix}$ and $G = \begin{bmatrix} 1 - \kappa \Delta t & 0 \\ 0 & 1 - \gamma \Delta t \end{bmatrix}$. Then from (A.3) and (A.4) we get

$$X_{t+1} = c + G X_t + w_{t+1}.$$

Let $\phi = 1 - \kappa \Delta t$, $\psi = 1 - \gamma \Delta t$. Then

$$\begin{aligned} E(X_t) &= \begin{bmatrix} (1-\kappa\Delta t)\chi_{t-1} \\ (1-\gamma\Delta t)\xi_{t-1} + \mu_\xi \Delta t \end{bmatrix} \\ &= \begin{bmatrix} (1-\kappa\Delta t)^n \chi_0 \\ (1-\gamma\Delta t)^n \xi_0 + (1-\gamma\Delta t)^{n-1}\mu_\xi \Delta t + \cdots + (1-\gamma\Delta t)^0 \mu_\xi \Delta t \end{bmatrix} \\ &= \begin{bmatrix} \phi^n \chi_0 \\ \psi^n \xi_0 + \mu_\xi \Delta t \frac{1-(1-\gamma\Delta t)^n}{\gamma \Delta t} \end{bmatrix} \\ &= \begin{bmatrix} \phi^n \chi_0 \\ \psi^n \xi_0 + \frac{\mu_\xi}{\gamma}(1-\psi^n) \end{bmatrix}, \end{aligned} \tag{A.5}$$

and

$$Var(X_t) = G Var(X_{t-1}) G' + W = G^n Var(X_0)(G')^n + G^{n-1} W (G')^{n-1} + \cdots + G^0 W (G')^0.$$

If we assume $Var(X_0) = 0$, we can get

$$\begin{aligned} Var(X_t) &= G^{n-1} W (G')^{n-1} + \cdots + G^0 W (G')^0 \\ &= \begin{bmatrix} \sigma_\chi^2 \Delta t \sum_{i=0}^{n-1} \phi^{2i} & \rho\sigma_\chi\sigma_\xi \Delta t \sum_{i=0}^{n-1} (\phi\psi)^i \\ \rho\sigma_\chi\sigma_\xi \Delta t \sum_{i=0}^{n-1} (\phi\psi)^i & \sigma_\xi^2 \Delta t \sum_{i=0}^{n-1} \psi^{2i} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_\chi^2 \Delta t \frac{1-\phi^{2n}}{1-\phi^2} & \rho\sigma_\chi\sigma_\xi \Delta t \frac{1-(\phi\psi)^n}{1-\phi\psi} \\ \rho\sigma_\chi\sigma_\xi \Delta t \frac{1-(\phi\psi)^n}{1-\phi\psi} & \sigma_\xi^2 \Delta t \frac{1-\psi^{2n}}{1-\psi^2} \end{bmatrix}. \end{aligned} \tag{A.6}$$

When $n \to \infty$, $\Delta t = t/n \to 0$, $\Delta t^2 = 0$, then

$$\phi^n = (1 - \frac{\kappa t}{n})^n \to e^{-\kappa t},$$

$$\psi^n = (1 - \frac{\gamma t}{n})^n \to e^{-\gamma t},$$

$$(\phi\psi)^n = (1 - (\kappa + \gamma)t/n)^n \to e^{-(\kappa+\gamma)t},$$

$$1 - \phi^2 = 2\kappa\Delta t, 1 - \psi^2 = 2\gamma\Delta t, 1 - \phi\psi = (\kappa + \gamma)\Delta t.$$

From (A.5) and (A.6), we have

$$E(X_t) = \begin{bmatrix} e^{-\kappa t}\chi_0 \\ e^{-\gamma t}\xi_0 + \frac{\mu_\xi}{\gamma}(1 - e^{-\gamma t}) \end{bmatrix}$$

and

$$Var(X_t) = \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa}(1 - e^{-2\kappa t}) & \frac{\rho\sigma_\chi\sigma_\xi}{\kappa+\gamma}(1 - e^{-(\kappa+\gamma)t}) \\ \frac{\rho\sigma_\chi\sigma_\xi}{\kappa+\gamma}(1 - e^{-(\kappa+\gamma)t}) & \frac{\sigma_\xi^2}{2\gamma}(1 - e^{-2\gamma t}) \end{bmatrix}.$$

Appendix B

Gradient of Log Likelihood

The state space model is given by the equations

$$x_t = Gx_{t-1} + c + w_t, \tag{B.1}$$

and

$$y_t = F_t' x_t + d_t + v_t. \tag{B.2}$$

Let $p(x, y|\theta)$ be the joint density of x and y, $p(x|y, \theta)$ be the conditional density of x given y, and $p(y|\theta)$ be the marginal density of y, where θ is the vector of all unknown parameters. We have

$$p(y|\theta) = \frac{p(x, y|\theta)}{p(x|y, \theta)}.$$

Then we have

$$\log p(y|\theta) = \log p(x, y|\theta) - \log p(x|y, \theta). \tag{B.3}$$

Let $\tilde{E}$ denote the expectation with respect to the density $p(x|y, \theta)$. Since the left hand side of equation (B.3) is independent on x, we have

$$\log p(y|\theta) = \tilde{E}\left[\log p(x, y|\theta)\right] - \tilde{E}\left[\log p(x|y, \theta)\right].$$

We differentiate both sides with respect to θ. Since

$$\tilde{E}\left[\frac{\partial \log p(x|y, \theta)}{\partial \theta}\right] = \int \frac{1}{p(x|y, \theta)} \frac{\partial p(x|y, \theta)}{\partial \theta} p(x|y, \theta) dx = \frac{\partial}{\partial \theta} \int p(x|y, \theta) dx = 0,$$

we have

$$\frac{\partial \log p(y|\theta)}{\partial \theta} = \tilde{E}\left[\frac{\partial \log p(x,y|\theta)}{\partial \theta}\right]. \tag{B.4}$$

Since $x_t \sim N(\omega_t, W)$ and $y_t|x_t \sim N(\nu_t, V)$, where ω_t and ν_t are estimates of w_t and v_t, and $p(x,y|\theta) = p(x|\theta)p(y|x,\theta)$, we have

$$\begin{aligned}\log p(x,y|\theta) &= \log p(x|\theta) + \log p(y|x,\theta)\\ &= constant - \frac{1}{2}\sum_{t=1}^{n}\left(\log|W| + \log|V| + \omega_t' W^{-1}\omega_t + \nu_t' V^{-1}\nu_t\right).\end{aligned} \tag{B.5}$$

W and V are independent on x, so

$$\tilde{E}\left[\frac{\partial \log p(x,y|\theta)}{\partial \theta}\right] = -\frac{1}{2}\frac{\partial}{\partial \theta}\sum_{t=1}^{n}\left[\log|W| + \log|V| + \tilde{E}(\omega_t' W^{-1}\omega_t) + \tilde{E}(\nu_t' V^{-1}\nu_t)\right]. \tag{B.6}$$

Let $W_{ij}^{(-1)}$ and $V_{ij}^{(-1)}$ denote the elements at ith row and jth column of W^{-1} and V^{-1}. Assuming ω_t and ν_t are n and m dimensional vectors respectively. Then

$$\tilde{E}(\omega_t' W^{-1}\omega_t) = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{E}(\omega_t^{(i)} W_{ij}^{(-1)}\omega_t^{(j)}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{E}(\omega_t^{(i)}\omega_t^{(j)})W_{ij}^{(-1)}, \tag{B.7}$$

and

$$\tilde{E}(\nu_t' V^{-1}\nu_t) = \sum_{i=1}^{m}\sum_{j=1}^{m}\tilde{E}(\nu_t^{(i)} V_{ij}^{(-1)}\nu_t^{(j)}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\tilde{E}(\nu_t^{(i)}\nu_t^{(j)})V_{ij}^{(-1)}, \tag{B.8}$$

where $\omega_t^{(i)}$ and $\nu_t^{(i)}$ are ith elements of ω_t and ν_t.

1. Calculation of $\tilde{E}(\omega_t' W^{-1}\omega_t)$

Let c_i be the ith element of vector c, and G_i be the ith row of G. At time t, we have

$$\begin{aligned}&\tilde{E}(\omega_t^{(i)}\omega_t^{(j)}) = \tilde{E}\left[(x_t^{(i)} - c_i - G_i x_{t-1})(x_t^{(j)} - c_j - G_j x_{t-1})\right]\\ &= \tilde{E}\left[(x_t^{(i)} - c_i)(x_t^{(j)} - c_j) - G_i x_{t-1}(x_t^{(j)} - c_j) - G_j x_{t-1}(x_t^{(i)} - c_i) + (G_i x_{t-1})(G_j x_{t-1})\right]\end{aligned}$$

$$= \int (x_t^{(i)} - c_i)(x_t^{(j)} - c_j)p(x|y,\theta) - G_i x_{t-1}(x_t^{(j)} - c_j)p(x|y,\theta)$$

$$-G_j x_{t-1}(x_t^{(i)} - c_i)p(x|y,\theta) + (G_i x_{t-1})(G_j x_{t-1})p(x|y,\theta)dx$$

$$= E\left[(x_t^{(i)} - c_i)(x_t^{(j)} - c_j)|y\right] - E\left[G_i x_{t-1}(x_t^{(j)} - c_j)|y\right]$$

$$-E\left[G_j x_{t-1}(x_t^{(i)} - c_i)|y\right] + E\left[(G_i x_{t-1})(G_j x_{t-1})|y\right]$$

$$= (x_{t|n}^{(i)} - c_i)(x_{t|n}^{(j)} - c_j) + Cov\left(x_{t|n}^{(i)} - c_i, x_{t|n}^{(j)} - c_j\right)$$

$$-G_i x_{t-1|n}\left(x_{t|n}^{(j)} - c_j\right) - Cov\left(G_i x_{t-1|n}, x_{t|n}^{(j)} - c_j\right)$$

$$-G_j x_{t-1|n}\left(x_{t|n}^{(i)} - c_i\right) - Cov\left(G_j x_{t-1|n}, x_{t|n}^{(i)} - c_i\right)$$

$$+(G_i x_{t-1|n})(G_j x_{t-1|n}) + Cov\left(G_i x_{t-1|n}, G_j x_{t-1|n}\right).$$

The sum of all covariances is equal to $Cov\left(x_{t|n}^{(i)} - G_i x_{t-1|n}, x_{t|n}^{(j)} - G_j x_{t-1|n}\right)$, which is $Cov\left(\omega_{t|n}^{(i)}, \omega_{t|n}^{(j)}\right)$, and the sum of all other terms is equal to $\left(x_{t|n}^{(i)} - G_i x_{t-|n}\right)\left(x_{t|n}^{(j)} - G_j x_{t-1|n}\right)$, which is $\omega_{t|n}^{(i)}\omega_{t|n}^{(j)}$. So we have

$$\tilde{E}\left(\omega_t^{(i)}\omega_t^{(j)}\right) = \omega_{t|n}^{(i)}\omega_{t|n}^{(j)} + Cov\left(\omega_{t|n}^{(i)}, \omega_{t|n}^{(j)}\right). \qquad \text{(B.9)}$$

By substituting (B.9) in (B.7), we get

$$\begin{aligned}\tilde{E}\left(\omega_t' W^{-1}\omega_t\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\omega_{t|n}^{(i)}\omega_{t|n}^{(j)} + Cov\left(\omega_{t|n}^{(i)}, \omega_{t|n}^{(j)}\right)\right] W_{ij}^{(-1)} \\ &= tr\left[\{\omega_{t|n}\omega_{t|n}' + Var(\omega_{t|n})\}W^{-1}\right]. \qquad \text{(B.10)}\end{aligned}$$

2. Calculation of $\tilde{E}(\nu_t' V^{-1}\nu_t)$

At time t, let $d_t^{(i)}$ be the ith element of vector d, and $F_t^{(i)}$ be the ith column of F_t. Since y_t is independent on x_t, we have

$$\tilde{E}\left(\nu_t^{(i)}\nu_t^{(j)}\right) = \tilde{E}\left[(y_t^{(i)} - d_t^{(i)} - F_t^{(i)'}x_t)(y_t^{(j)} - d_t^{(j)} - F_t^{(j)'}x_t)\right]$$

$$= \tilde{E}\left[(y_t^{(i)} - d_t^{(i)})(y_t^{(j)} - d_t^{(j)}) - F_t^{(i)'}x_t(y_t^{(j)} - d_t^{(j)}) - F_t^{(j)'}x_t(y_t^{(i)} - d_t^{(i)}) + (F_t^{(i)'}x_t)(F_t^{(j)'}x_t)\right]$$

$$= \int (y_t^{(i)} - d_t^{(i)})(y_t^{(j)} - d_t^{(j)})p(x|y,\theta) - F_t^{(i)'}x_t(y_t^{(j)} - d_t^{(j)})p(x|y,\theta)$$

$$-F_t^{(j)'}x_t(y_t^{(i)} - d_t^{(i)})p(x|y,\theta) + (F_t^{(i)'}x_t)(F_t^{(j)'}x_t)p(x|y,\theta)dx$$

$$= \left(y_t^{(i)} - d_t^{(i)}\right)\left(y_t^{(j)} - d_t^{(j)}\right) - F_t^{(i)'}E[x_t|y]\left(y_t^{(j)} - d_t^{(j)}\right)$$

$$-F_t^{(j)'}E[x_t|y]\left(x_t^{(i)} - d_t^{(i)}\right) + E\left[(F_t^{(i)'}x_t)(F_t^{(j)'}x_t)|y\right]$$

$$= (y_t^{(i)} - d_t^{(i)})(y_t^{(j)} - d_t^{(j)}) - F_t^{(i)'}x_{t|n}(y_t^{(j)} - d_t^{(j)}) - F_t^{(j)'}x_{t|n}(x_t^{(i)} - d_t^{(i)})$$

$$+ \left(F_t^{(i)'}x_{t|n}\right)\left(F_t^{(j)'}x_{t|n}\right) + Cov\left(F_t^{(i)'}x_{t|n}, F_t^{(j)'}x_{t|n}\right)$$

$$= \nu_{t|n}^{(i)}\nu_{t|n}^{(j)} + Cov\left(y_t^{(i)} - d_t^{(i)} - \nu_{t|n}^{(i)}, y_t^{(j)} - d_t^{(j)} - \nu_{t|n}^{(j)}\right)$$

Since $y_t^{(i)} - d_t^{(i)} - \nu_{t|n}^{(i)}$ is constant, we have

$$\tilde{E}\left(\nu_t^{(i)}\nu_t^{(j)}\right) = \nu_{t|n}^{(i)}\nu_{t|n}^{(j)} + Cov\left(\nu_{t|n}^{(i)}, \nu_{t|n}^{(j)}\right). \tag{B.11}$$

By substituting (B.11) in (B.8), we get

$$\begin{aligned}\tilde{E}\left(\nu_t' V^{-1}\nu_t\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\nu_{t|n}^{(i)}\nu_{t|n}^{(j)} + Cov(\nu_{t|n}^{(i)}, \nu_{t|n}^{(j)})\right]W_{ij}^{(-1)} \\ &= tr\left[\{\nu_{t|n}\nu_{t|n}' + Var(\nu_{t|n})\}V^{-1}\right]. \end{aligned} \tag{B.12}$$

From (B.4), (B.6), (B.10) and (B.12), we finally get

$$\begin{aligned}\frac{\partial \log L(y|\theta)}{\partial\theta} = -\frac{1}{2}\frac{\partial}{\partial\theta}\sum_{t=1}^{n}\{&\log|W| + \log|V| \\ &+ tr[\{\omega_{t|n}\omega_{t|n}' + Var(\omega_{t|n})\}W^{-1}] \\ &+ tr[\{\nu_{t|n}\nu_{t|n}' + Var(\nu_{t|n})\}V^{-1}]\}. \end{aligned} \tag{B.13}$$

Appendix C

Proof of Parameter Identification Problem

In this appendix, we will show that the log-likelihood does not change when we swap the parameters $(\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho, s)$ to $(\gamma, \kappa, \tilde{\mu}_\xi, \sigma_\xi, \sigma_\chi, \rho, s)$, where $\tilde{\mu}_\xi = \frac{\kappa}{\gamma}\mu_\xi$. For simplification, we assume one-dimensional observable variable y_t.

From equation (2.12), log-likelihood l is a function of prediction error e_t and covariance matrix $L_{t|t-1}$. If e_t and $L_{t|t-1}$ do not change after swapping the parameters, then l does not change.

Mathematical induction will be used to prove the identification problem. Firstly, we will show that the estimates of state vector $a_{t|t-1}$ and the corresponding covariance matrix $P_{t|t-1}$ have the structure

$$a_{t|t-1} = \begin{bmatrix} \alpha_1 \\ \frac{\mu_\xi}{\gamma} + \alpha_2 \end{bmatrix} \tag{C.1}$$

and

$$P_{t|t-1} = \begin{bmatrix} W_1 - Q_{11} & W_2 - Q_{12} \\ W_2 - Q_{21} & W_3 - Q_{22} \end{bmatrix}, \tag{C.2}$$

where $W_1 = \frac{\sigma_\chi^2}{2\kappa}$, $W_2 = \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa+\gamma}$ and $W_3 = \frac{\sigma_\xi^2}{2\gamma}$. Let's denote $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$. The following properties are satisfied:

P1 α_1 and α_2 are linear combinations of $e_1, e_2, \cdots, e_{t-1}$, and all the coefficients are independent to μ_ξ.

P2 $F_t'\alpha = e^{-\kappa T}\alpha_1 + e^{-\gamma T}\alpha_2$ does not change after swapping the parameters, where T is the maturity time.

P3 $Q_{12} = Q_{21}$ so that $P_{t|t-1}$ is a symmetric matrix.

P4 After swapping the parameters, the diagonal entries of $P_{t|t-1}$ swap and the off-diagonal entries do not change. That is, $\tilde{Q}_{11} = Q_{22}$, $\tilde{Q}_{22} = Q_{11}$ and $\tilde{Q}_{12} = \tilde{Q}_{21} = Q_{12}$, where $\tilde{Q}_{11}$, $\tilde{Q}_{22}$ and $\tilde{Q}_{12}$ are values after swapping the parameters.

The prediction error e_t and covariance matrix $L_{t|t-1}$ are calculated by

$$e_t = y_t - d_t - F_t' a_{t|t-1}$$

and

$$L_{t|t-1} = F_t' P_{t|t-1} F_t + V.$$

Under properties P1 - P4, we can prove that e_t has a structure

$$e_t = y_t - \frac{\mu_\xi}{\gamma} - B(T_t) - E_t, \tag{C.3}$$

where

$$B(T_t) = \frac{1}{2}\left((1 - e^{-2\kappa T_t})\frac{\sigma_\chi^2}{2\kappa} + (1 - e^{-2\gamma T_t})\frac{\sigma_\xi^2}{2\gamma} + 2(1 - e^{-(\kappa+\gamma)T_t})\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa+\gamma}\right). \tag{C.4}$$

E_t is a linear combination of $e_1, e_2, \cdots, e_{t-1}$ satisfies:

P5 All coefficients of E_t are independent to μ_ξ.

P6 E_t does not change after swapping the parameters.

1. t=1

Firstly, we start from the simplest model with one data point y_1. Let the vector of unknown parameters be $\theta = (\kappa, \gamma, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}, s)$ and the state vector be $x_t = (\chi_t, \xi_t)'$, where

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dZ_t^\chi$$

and

$$d\xi_t = (\mu_\xi - \gamma\xi_t)dt + \sigma_\xi dZ_t^\xi.$$

The initial expectation and covariance matrix are

$$E(x_0) = a_0 = \begin{bmatrix} 0 \\ \frac{\mu_\xi}{\gamma} \end{bmatrix}$$

and

$$Cov(x_0) = P_0 = \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa} & \frac{\sigma_\chi\sigma_\xi\rho_{\chi\xi}}{\kappa+\gamma} \\ \frac{\sigma_\chi\sigma_\xi\rho_{\chi\xi}}{\kappa+\gamma} & \frac{\sigma_\xi^2}{2\gamma} \end{bmatrix}.$$

Given a_0 and P_0, we predict the expectation of the state vector x_t as

$$a_{1|0} = c + G'a_0 = \begin{bmatrix} 0 \\ \frac{\mu_\xi}{\gamma} \end{bmatrix}$$

and covariance matrix

$$P_{1|0} = GP_0G' + W = \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa} & \frac{\sigma_\chi\sigma_\xi\rho_{\chi\xi}}{\kappa+\gamma} \\ \frac{\sigma_\chi\sigma_\xi\rho_{\chi\xi}}{\kappa+\gamma} & \frac{\sigma_\xi^2}{2\gamma} \end{bmatrix} = \begin{bmatrix} W_1 & W_2 \\ W_2 & W_3 \end{bmatrix}.$$

Obviously, $a_{1|0}$ and $P_{1|0}$ have a structure (C.1) and (C.2), where $\alpha_1 = \alpha_2 = 0$ and $Q_{11} = Q_{12} = Q_{21} = Q_{22} = 0$, and properties P1 - P4 are satisfied.

When a new data point y_1 becomes available, we calculate the prediction error

$$e_1 = y_1 - d_1 - F_1'a_{1|0} = y_1 - \frac{\mu_\xi}{\gamma} - B(T_1)$$

and covariance matrix

$$L_{1|0} = F_1'P_{1|0}F_1 + V = e^{-2\kappa T_1}\frac{\sigma_\chi^2}{2\kappa} + e^{-2\gamma T_1}\frac{\sigma_\xi^2}{2\gamma} + 2e^{-(\kappa+\gamma)T_1}\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa+\gamma} + s,$$

where $B(T_1)$ is given by (C.4). Then $E_1 = 0$ satisties properties P5 and P6.

To obtain the Maximum Likelihood Estimates of μ_ξ, we derive the log-likelihood function (2.12) with respect to μ_ξ

$$\frac{\partial l_1}{\partial \mu_\xi} = -L_{1|0}^{-1} e_1 \frac{\partial e_1}{\partial \mu_\xi} = \frac{L_{1|0}^{-1}}{\gamma}\left(y_1 - \frac{\mu_\xi}{\gamma} - B(T_1)\right) = 0.$$

Then we have

$$\frac{\mu_\xi}{\gamma} = y_t - B(T_1) \tag{C.5}$$

where $B(T_1)$ is given in (C.4).

Now we swap the parameters κ and γ, and σ_χ and σ_ξ. Let's denote a new vector $\tilde{\theta} = (\tilde{\kappa}, \tilde{\gamma}, \tilde{\mu}_\xi, \tilde{\sigma}_\chi, \tilde{\sigma}_\xi, \rho_{\chi\xi}, s)$. Then we have $\tilde{\kappa} = \gamma$, $\tilde{\gamma} = \kappa$, $\tilde{\sigma}_\chi = \sigma_\xi$ and $\tilde{\sigma}_\xi = \sigma_\chi$. From (C.5), we have

$$\frac{\tilde{\mu}_\xi}{\tilde{\gamma}} = y_t - B(T_1),$$

because y_t and $B(T_1)$ do not change after swapping of the parameters. Since $\tilde{\gamma} = \kappa$, we have

$$\frac{\tilde{\mu}_\xi}{\tilde{\gamma}} = \frac{\tilde{\mu}_\xi}{\kappa}.$$

Then from (C.5) we have

$$\tilde{\mu}_\xi = \frac{\kappa}{\gamma}\mu_\xi \tag{C.6}$$

After swapping of the parameters, the prediction error and covariance matrix would be

$$\tilde{e}_1 = y_t - \frac{\tilde{\mu}_\xi}{\tilde{\gamma}} - B(T_1) = y_t - \frac{\mu_\xi}{\gamma} - B(T_1) = e_1$$

and

$$\tilde{L}_{1|0} = L_{1|0},$$

hence the log-likelihood did not change after swapping of the parameters.

2. t = 2

Then we move on to the model with two data points. Given $a_{1|0}$ and $P_{1|0}$, we calculate the Kalman gain matrix

$$K_1 = P_{1|0}F_1L_{1|0}^{-1} = \begin{bmatrix} (W_1e^{-\kappa T_1} + W_2e^{-\gamma T_1})L_{1|0}^{-1} \\ (W_2e^{-\kappa T_1} + W_3e^{-\gamma T_1})L_{1|0}^{-1} \end{bmatrix} = \begin{bmatrix} K_1^{(1)} \\ K_2^{(1)} \end{bmatrix}.$$

Let's denote $K_1^{(1)}$ and $K_2^{(1)}$ as the first and second elements of vector K_1. Obviously $K_1^{(1)}$ and $K_2^{(1)}$ are independent with μ_ξ, and $\tilde{K}_1^{(1)} = K_2^{(1)}$ and $\tilde{K}_2^{(1)} = K_1^{(1)}$. The updating equations are

$$a_1 = a_{1|0} + K_1e_1 = \begin{bmatrix} K_1^{(1)}e_1 \\ \frac{\mu_\xi}{\gamma} + K_2^{(1)}e_1 \end{bmatrix}$$

and

$$P_1 = (I - K_1F_1')P_{1|0}$$
$$= \begin{bmatrix} W_1 - W_1K_1^{(1)}e^{-\kappa T_1} - W_2K_1^{(1)}e^{-\gamma T1} & W_2 - W_2K_1^{(1)}e^{-\kappa T_1} - W_3K_1^{(1)}e^{-\gamma T1} \\ W_2 - W_1K_2^{(1)}e^{-\kappa T_1} - W_2K_2^{(1)}e^{-\gamma T1} & W_3 - W_2K_2^{(1)}e^{-\kappa T_1} - W_3K_2^{(1)}e^{-\gamma T1} \end{bmatrix}$$
$$= \begin{bmatrix} W_1 - P_{11}^{(1)} & W_2 - P_{12}^{(1)} \\ W_2 - P_{21}^{(1)} & W_3 - P_{22}^{(1)} \end{bmatrix},$$

where $P_{12}^{(1)} = P_{21}^{(1)}$. Then we predict the expectation of state vector at time $t = 2$ as

$$a_{2|1} = c + Ga_1 = \begin{bmatrix} K_1^{(1)}e^{-\kappa\Delta t}e_1 \\ \frac{\mu_\xi}{\gamma} + K_2^{(1)}e^{-\gamma\Delta t}e_1 \end{bmatrix}$$

and the covariance matrix

$$P_{2|1} = GP_1G' + W = \begin{bmatrix} W_1 - P_{11}^{(1)}e^{-2\kappa\Delta t} & W_2 - P_{12}^{(1)}e^{-(\kappa+\gamma)\Delta t} \\ W_2 - P_{21}^{(1)}e^{-(\kappa+\gamma)\Delta t} & W_3 - P_{22}^{(1)}e^{-2\gamma\Delta t} \end{bmatrix}.$$

So at time $t = 2$, we have $\alpha_1 = K_1^{(1)}e^{-\kappa\Delta t}e_1$, $\alpha_2 = K_2^{(1)}e^{-\gamma\Delta t}e_1$, $Q_{11} = P_{11}^{(1)}e^{-2\kappa\Delta t}$, $Q_{12} = P_{12}^{(1)}e^{-(\kappa+\gamma)\Delta t}$, $Q_{21} = P_{21}^{(1)}e^{-(\kappa+\gamma)\Delta t}$ and $Q_{22} = P_{22}^{(1)}e^{-2\gamma\Delta t}$. We can show that properties P1 - P4 are satisfied:

P1 Obviously α_1 and α_2 are linear combinations of e_1. $K_1^{(1)}$ and $K_2^{(1)}$ are independent with μ_ξ, so the coefficients are independent with μ_ξ.

P2 In Appendix C.1, we have shown that e_1 does not change after swapping the parameters. Moreover, $\tilde{K}_1^{(1)} = K_2^{(1)}$ and $\tilde{K}_2^{(1)} = K_1^{(1)}$, so $F_2'\alpha = K_1^{(1)} e^{-\kappa(\Delta t + T_2)} e_1 + K_2^{(1)} e^{-\gamma(\Delta t + T_2)} e_1$ does not change.

P3 $P_{12}^{(1)} = P_{21}^{(1)}$, so $Q_{12} = Q_{21}$.

P4 $P_{12}^{(1)}$ and $P_{21}^{(1)}$ do not change, so Q_{12} and Q_{21} do not change. $\tilde{P}_{11}^{(1)} = P_{22}^{(1)}$ and $\tilde{P}_{22}^{(1)} = P_{11}^{(1)}$, so $\tilde{Q}_{11} = Q_{22}$ and $\tilde{Q}_{22} = Q_{11}$.

We calculate the prediction error and covariance matrix at time $t = 2$ as

$$e_2 = y_2 - d_2 - F_2' a_{2|1} = y_2 - \frac{\mu_\xi}{\gamma} - B(T_2) - (e^{-\kappa T_2}\alpha_1 + e^{-\gamma T_2}\alpha_2)$$

and

$$L_{2|1} = F_2' P_{2|1} F_2 + V.$$

It is clearly that e_2 has a sturcture (C.3) where $E_2 = \left(e^{-\kappa T_2}\alpha_1 + e^{-\gamma T_2}\alpha_2\right)$ is a linear combination of e_1. Since $e^{-\kappa T_2}\alpha_1 + e^{-\gamma T_2}\alpha_2 = F_2'\alpha$ does not change after swapping the parameters, E_2 does not change. Properties P5 and P6 are satisfied.

Now we derive the log-likelihood function (2.12) with respect to μ_ξ

$$\frac{\partial l_2}{\partial \mu_\xi} = \frac{L_{1|0}^{-1}}{\gamma}\left(y_1 - \frac{\mu_\xi}{\gamma} - B(T_1)\right) + \frac{L_{2|1}^{-1}(1 - E_2/e_1)}{\gamma}\left[y_2 - \frac{\mu_\xi}{\gamma} - B(T_2) - E_2\right] = 0$$

Then we get

$$\frac{\mu_\xi}{\gamma} = \frac{L_{1|0}^{-1}(y_1 - B(T_1)) + L_{2|1}^{-1}(1 - E_2/e_1)\left[y_2 - B(T_2) - (E_2/e_1)(y_1 - B(T_1))\right]}{L_{1|0}^{-1} + L_{2|1}^{-1}(1 - E_2/e_1)^2} \tag{C.7}$$

The right-hand side of (C.7) does not change after swapping the parameters, so we get

$$\tilde{\mu}_\xi = \frac{\kappa}{\gamma}\mu_\xi. \tag{C.8}$$

Given (C.8) and properties P2 and P6, e_2 does not change after swapping the parameters. Given properties P3 and P4, the diagonal elements of $P_{2|1}$ is swapped and vector F_2 is also swapped, so $L_{2|1}$ does not change. The log-likelihood at time $t = 2$ does not change.

3. t=n+1

We assume

$$a_{n|n-1} = \begin{bmatrix} \alpha_1^{(n)} \\ \frac{\mu_\xi}{\gamma} + \alpha_2^{(n)} \end{bmatrix} \tag{C.9}$$

and

$$P_{n|n-1} = \begin{bmatrix} W_1 - Q_{11}^{(n)} & W_2 - Q_{12}^{(n)} \\ W_2 - Q_{21}^{(n)} & W_3 - Q_{22}^{(n)} \end{bmatrix} \tag{C.10}$$

have structure (C.1) and (C.2) respectively, where properties P1 - P4 are satisfied. The prediction error

$$e_n = y_n - \frac{\mu_\xi}{\gamma} - B(T_n) - E_n \tag{C.11}$$

has structure (C.3) where properties P5 and P6 are satisfied. e_n and $L_{n|n-1}^{-1}$ do not change after swapping the parameters. We derive the log-likelihood function (2.12) with respect to μ_ξ, then we get

$$\frac{\mu_\xi}{\gamma} = \beta_n, \tag{C.12}$$

where β_n is a function independent from μ_ξ and remains invariant after switching the parameters.

We calculate the Kalman gain matrix as

$$
\begin{aligned}
K_n &= P_{n|n-1}F_nL_{n|n-1}^{-1} \\
&= \begin{bmatrix} (W_1e^{-\kappa T_n} - Q_{11}^{(n)}e^{-\kappa T_n} + W_2e^{-\gamma T_n} - Q_{12}^{(n)}e^{-\gamma T_n})L_{n|n-1}^{-1} \\ (W_2e^{-\kappa T_n} - Q_{21}^{(n)}e^{-\kappa T_n} + W_3e^{-\gamma T_n} - Q_{22}^{(n)}e^{-\gamma T_n})L_{n|n-1}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} K_1^{(n)} \\ K_2^{(n)} \end{bmatrix}.
\end{aligned}
$$

The updating equations are

$$
a_n = a_{n|n-1} + K_ne_n = \begin{bmatrix} \alpha_1^{(n)} + K_1^{(n)}e_n \\ \frac{\mu_\xi}{\gamma} + \alpha_2^{(n)} + K_2^{(n)}e_n \end{bmatrix}
$$

and

$$
P_n = (I - K_nF_n')P_{n|n-1} = \begin{bmatrix} W_1 - P_{11}^{(n)} & W_2 - P_{12}^{(n)} \\ W_2 - P_{21}^{(n)} & W_3 - P_{22}^{(n)} \end{bmatrix},
$$

where

$$
P_{11}^{(n)} = Q_{11}^{(n)} + W_1K_1^{(n)}e^{-\kappa T_n} - Q_{11}^{(n)}K_1^{(n)}e^{-\kappa T_n} + W_2K_1^{(n)}e^{-\gamma T_n} - Q_{21}^{(n)}K_1^{(n)}e^{-\gamma T_n},
$$

$$
P_{12}^{(n)} = Q_{12}^{(n)} + W_2K_1^{(n)}e^{-\kappa T_n} - Q_{12}^{(n)}K_1^{(n)}e^{-\kappa T_n} + W_3K_1^{(n)}e^{-\gamma T_n} - Q_{22}^{(n)}K_1^{(n)}e^{-\gamma T_n},
$$

$$
P_{21}^{(n)} = Q_{21}^{(n)} + W_2K_2^{(n)}e^{-\gamma T_n} - Q_{21}^{(n)}K_2^{(n)}e^{-\gamma T_n} + W_1K_2^{(n)}e^{-\kappa T_n} - Q_{11}^{(n)}K_2^{(n)}e^{-\kappa T_n},
$$

$$
P_{22}^{(n)} = Q_{22}^{(n)} + W_3K_2^{(n)}e^{-\gamma T_n} - Q_{22}^{(n)}K_2^{(n)}e^{-\gamma T_n} + W_2K_2^{(n)}e^{-\kappa T_n} - Q_{12}^{(n)}K_2^{(n)}e^{-\kappa T_n}
$$

and $P_{12}^{(n)} = P_{21}^{(n)}$. We predict the expactation of state vector at time $t = n+1$ as

$$
a_{n+1|n} = c + Ga_n = \begin{bmatrix} \alpha_1^{(n)}e^{-\kappa\Delta t} + K_1^{(n)}e^{-\kappa\Delta t}e_n \\ \frac{\mu_\xi}{\gamma} + \alpha_2^{(n)}e^{-\gamma\Delta t} + K_2^{(n)}e^{-\gamma\Delta t}e_n \end{bmatrix}
$$

and covariance matrix

$$
P_{n+1|n} = GP_nG' + W = \begin{bmatrix} W_1 - P_{11}^{(n)}e^{-2\kappa\Delta t} & W_2 - P_{12}^{(n)}e^{-(\kappa+\gamma)\Delta t} \\ W_2 - P_{21}^{(n)}e^{-(\kappa+\gamma)\Delta t} & W_3 - P_{22}^{(n)}e^{-2\gamma\Delta t} \end{bmatrix}.
$$

So at time $t = n+1$, we have $\alpha_1 = \alpha_1^{(n)}e^{-\kappa\Delta t} + K_1^{(n)}e^{-\kappa\Delta t}e_n$, $\alpha_2 = \alpha_2^{(n)}e^{-\gamma\Delta t} +$

$K_2^{(n)}e^{-\gamma\Delta t}e_n$, $Q_{11} = P_{11}^{(n)}e^{-2\kappa\Delta t}$, $Q_{12} = P_{12}^{(n)}e^{-(\kappa+\gamma)\Delta t}$, $Q_{21} = P_{21}^{(n)}e^{-(\kappa+\gamma)\Delta t}$ and $Q_{22} = P_{22}^{(n)}e^{-2\gamma\Delta t}$, where properties P1 - P4 are satisfied:

P1 $\alpha_1^{(n)}$ and $\alpha_2^{(n)}$ are linear combinations of $e_1, e_2, \cdots, e_{n-1}$ and all coefficients are independent to μ_ξ, so α_1 and α_2 are linear combinations of $e_1, e_2, \cdots, e_n$ and all coefficients are independent to μ_ξ.

P2 $F_n'\alpha^{(n)}$ and e_n do not change after swapping the parameters, and $\tilde{K}_1^{(n)} = K_2^{(n)}$ and $\tilde{K}_2^{(n)} = K_1^{(n)}$, so $F_{n+1}'\alpha = \alpha_1^{(n)}e^{-\kappa(\Delta t+T_{n+1})}+\alpha_2^{(n)}e^{-\gamma(\Delta t+T_{n+1})}+K_1^{(n)}e^{-\kappa(\Delta t+T_{n+1})}e_n+K_2^{(n)}e^{-\gamma(\Delta t+T_{n+1})}e_n$ does not change.

P3 $P_{12}^{(n)} = P_{21}^{(n)}$, so $Q_{12} = Q_{21}$.

P4 $P_{12}^{(n)}$ and $P_{21}^{(n)}$ do not change, so Q_{12} and Q_{21} do not change. $\tilde{P}_{11}^{(n)} = P_{22}^{(n)}$ and $\tilde{P}_{22}^{(n)} = P_{11}^{(n)}$, so $\tilde{Q}_{11} = Q_{22}$ and $\tilde{Q}_{22} = Q_{11}$.

Then we calculate the prediction error and covariance matrix as

$$e_{n+1} = y_{n+1} - d_{n+1} - F_{n+1}'a_{n+1|n} = y_{n+1} - \frac{\mu_\xi}{\gamma} - B(T_{n+1}) - F_{n+1}'\alpha$$

and

$$L_{n+1|n} = F_{n+1}'P_{n+1|n}F_{n+1|n} + V.$$

So e_{t+1} has a structure (C.3) where $E_{n+1} = F_{n+1}'\alpha$. Then E_{n+1} is a linear combination of $e_1, e_2, \cdots, e_n$ where properties P5 and P6 are satisfied.

We derive the log-likelihood function (2.12) with respect to μ_ξ

$$\frac{\partial l_{n+1}}{\partial \mu_\xi} = \frac{\partial l_n}{\partial \mu_\xi} + L_{n+1|n}^{-1}e_{n+1}\frac{\partial e_{n+1}}{\partial \mu_\xi} = 0. \tag{C.13}$$

From (C.12) we have

$$\frac{\mu_\xi}{\gamma} + L_{n+1|n}^{-1}e_{n+1}\frac{\partial e_{n+1}}{\partial \mu_\xi} = \beta_n.$$

E_{n+1} is a linear combination of $e_1, e_2, \cdots, e_n$ and all of these e_i's are linear with respect to μ_ξ. Hence $\frac{\partial e_{n+1}}{\partial \mu_\xi}$ in (C.13) is independent from μ_ξ. Then e_{n+1} will be a linear function of μ_ξ. We obtain

$$\frac{\mu_\xi}{\gamma} = \beta_{n+1}, \tag{C.14}$$

where β_{n+1} is the expression similar to (C.7). It must be noted that β_{n+1} does not depend on μ_ξ and remains invariant to switching of parameters. So we have

$$\tilde{\mu}_\xi = \frac{\kappa}{\gamma}\mu_\xi. \tag{C.15}$$

Given (C.15) and all properties P1 - P6, we have e_{n+1} and $L_{n+1|n}^{-1}$ does not change after swapping the parameters, so that the log-likelihood does not change at time $t = n + 1$.