Instituto de Matemática Pura e Aplicada

Calibration of the Schwartz-Smith Model for Commodity Prices

Ana Luiza Abrão Roriz Soares de Carvalho

March 2, 2010

Advisor: Jorge Passamani Zubelli
Co-Advisor: Max Oliveira de Souza

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere acknowledgments to:

IMPA for providing a great structure, world-class professors and for maintaining the Finance Program.

All the professors I had during the Master's Program for their commitment and dedication.

My supervisor and professor Jorge Zubelli. His great Mathematical knowledge enriched my understanding of Finance both during the courses and the thesis. I thank him for always helping me to accomplish my goals, both academic and professional. I am also thankfull to all the valuable opportunities he gave me in the last years.

My co-supervisor Max Souza. His commitment towards this thesis was remarkable, and I am very thankful for all the time and effort he dedicated to it. He was the great responsible for making me believe I could indeed write computer programs. I thank Max for his perfectionism, which ended up by always pushing me to do my best.

My mom Leda for the endless prayers during the exam weeks, for the financial support that allowed to focus on my studies, and for the horrible (but efficient) threats she made every time I mentioned quitting IMPA.

Petrobras in the person of Fernando Aiube for the financial support to the research activities at IMPA, specially in the Real Options Project.

Sebastian Jaimungal, from the University of Toronto, who contributed on the early discussions of this work.

Leonardo Müller who was my T.A. during several courses, and was always very nice and helpful both with academic and computational issues.

My classmates and great friends: Pedro and Tereza. The study afternoons were a lot more pleasant with them around.

Last but not least, to the lifetime friends I made at IMPA; Augusto Quadros, Bruno Agrelio, Cassio Alves, Daniel Valesin, Emilio Vital-Brazil, Ives Macedo, Julio Daniel, Marcelo Hilário and Tertuliano Franco.

Chapter 1

Introduction

Commodity markets are unique in financial modeling. Their assets are not like equities, interest rates or currencies, but real assets that are physically produced, transported, stored and consumed. That way, it is natural to expect that they should be treated differently from financial security markets.

Many market participants, such as large industries and governments, are indeed interested in the physical liquidation of the contracts and delivery of the assets. For this reason, a substantial amount of goods must be carried across the globe, implying in considerable delivery costs and delays. As it is virtually impossible to liquidate agricultural, metal or energy contracts with delivery in real time, spot markets for commodities do not exist per se. Contracts are made for future delivery or financial settlement.

The major commodity markets are the Chicago Board of Trade (CBOT) which deals with corn, wheat, soy, silver and American Treasury Bond contracts among others, and the New York Mercantile Exchange (NYMEX). Other exchanges also play an important role in futures and commodities markets, such as the Chicago Mercantile Exchange, the London International Financial Futures Exchange (LIFFE), the Singapore International Monetary Exchange - SIMEX and the São Paulo Mercantile and Futures Exchange (BM&F), see Hull (2006).

The fact that for most commodity contracts, future but not spot prices are available, poses the problem of how to recover those prices. Spot prices are necessary for derivative pricing and investment valuation. For this reason, there is interest from both academics and practitioners in developing methods to recover spot prices from futures data.

A number of models build up state space variables, in order to describe the futures prices. The state space form builds the relationship between a vector of observable time series (future prices for different maturities), and a vector of unobservable ones. The components of this unobserved vector are called

state variables. Spot prices are an example of a state variable, as they are not directly observable. The spot prices are generated by a discrete time version of the stochastic process used, called transition equation. The Kalman Filter is then applied recursively to calculate the optimal estimator for the state vector at time t using the information available at t. We shall treat this in detail in section 2.1.

As mentioned by Schwartz (1997), the main difficulty in the empirical implementation of commodity price models is that frequently the state variables, such as the spot price for example, are not directly observable. The financial literature traditionally estimates the parameters of a model with hidden factors by rewriting it in state space form and using the Kalman Filter.

Barlow, Gusev and Lai (2004) provide a good explanation of the Kalman's Filter application, and use it to calibrate three different models for electricity prices. Schwartz (1997) also exposes the construction of the Kalman Filter in a simple, but complete way.

In this work, we propose a methodology to recover spot prices from futures without the use of Kalman Filter. We use maximum likelihood and nonlinear least squares in a two-factor affine model based on Schwartz and Smith (2000).

The model we use describes commodity spot prices as a sum of short term deviations over a long term pattern. When performing a Principal Component Analysis in the data set of natural gas prices, for instance, we find that the behavior of the two most important eigenvectors are very much like the simulated paths for the short term deviations and long term dynamics considered by Schwartz and Smith (2000). It is important to note also that our calibration method can be applied to other energy commodities other than natural gas.

Our estimation procedure resembles to some extent the work done in Hikspoors and Jaimungal (2007). We use a nonlinear least-squares optimization (NLLS), as in Hikspoors and Jaimungal (2007), to recover the hidden short and long term factors from the futures data. However, we used a maximum likelihood estimation in order to recover the model's parameters, instead of the further NLLS minimization used in Hikspoors and Jaimungal (2007).

The main advantage of our calibration procedure is its robustness. As we show in Chapter 4, the model converges quickly for different initial guesses and even with large perturbations. Besides that, our algorithm is computationally cheaper than the Kalman Filter. It does not need to impose restrictions in order to decrease the number of estimated parameters and it is not dependent of a wise initial-guess selection in order to obtain convergence.

This work shall be divided as follows: Chapter 2 explores the literature concerning commodity modeling and estimation via Kalman Filter. We detail

some of the models and their estimation procedures.

Chapter 3 explains the Schwartz and Smith model, and detail the model's equations both in the "real world" and the risk neutral probability measures. Besides, we also review the theoretical framework of risk neutral pricing.

Chapter 4 details the estimation procedure and the computational methods used. We also validate our method using different numerical tests. Chapter 5 applies the algorithm to real data from one of the most liquid gas markets, namely the Henry Hub. We draw some final remarks and conclusions in Chapter 6.

Chapter 2

Commodity Price Models

In this Chapter, we detail commodity price models and some tools used in their construction. We begin by explaining the estimation via Kalman Filter, which is one of the most used tools for models with latent variables. Next, we introduce commodity price models exploring the concept of convenience yield and the theories of Storage and Expected Risk Premium. Finally, we review some of the most classical commodity prices models, building a historical literature review.

2.1 Theory of Storage, Expected Premium and Convenience Yield

As mentioned by Fama and French (1987), there are two popular views of commodity prices; the theory of storage, and an alternative view that splits a futures price into an expected risk premium and a forecast of a future spot price. We call this second approach simply "risk premium models".

The theory of storage explains the difference between contemporaneous spot and future prices in terms of interest forgone in storing a commodity, warehousing costs and a convenience yield on inventory.

The notion of convenience yield is very important and deserves some attention. As defined in Brennan and Schwartz (1985), the convenience yield is "the flow of services that accrues to an owner of the physical commodity, but not to the owner of a contract for future delivery of the commodity". The owner of the physical commodity can choose where it will be stored and when to liquidate the inventory. Therefore, the convenience yield can be thought of as the value of being able to take advantage of temporary local shortages by profiting from its physical ownership.

Brennan and Schwartz (1985) develop a model for evaluating natural resources investment. The authors value the uncertain cash flow stream gen-

erated by an investment project using a self-financing portfolio, whose cash flows replace those which are to be valued. The construction of this portfolio rests on the assumption that the convenience yield on the output commodity can be written as a function of the output price alone, and that the interest rate is non-stochastic.

Our main interest in Brennan and Schwartz (1985) lies in the relation between spot and futures prices of a commodity and the convenience yield. The authors assume that the convenience yield is a function only of the current spot price, together with the assumption that the risk free interest rate is a constant ρ. Formally:

Let $F(S, \tau)$ represent the futures price at time t for delivery of one unit of the commodity at time T, where $\tau = T - t$. Let S denote the spot price of a single homogeneous commodity at time t, assumed to follow the following stochastic process:

$$\mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sigma_t S_t \mathrm{d}Z_t, \quad t \geq 0$$

where $\mathrm{d}Z_t$ is the increment to a standard Brownian motion, σ_t is the instantaneous standard deviation of the spot price, (assumed to be known) and μ_t is the trend in price, which may be stochastic.

The marginal convenience yield $C(.)$ can be defined as a function of the current spot price. For tractability, it is assumed to be:

$$C(S, t) = cS, \quad c > 0$$

been simply a proportion of the current spot price. The authors then express the instantaneous change in the futures prices as:

$$\mathrm{d}F_s = F_s[S(\mu - \rho) + C]\mathrm{d}t + F_s S \sigma \mathrm{d}Z_s. \tag{2.1}$$

Since the convenience yield is not the same across producers, Fama and French (1987) remark that in equilibrium only individuals with high convenience yields will hold inventories. However, it is important to note that the marginal convenience yield, net of storage costs, is inversely proportional to the amount of the commodity held in inventory and directly proportional to the current spot price.

Turning back to risk premium models, we address an important concept called basis. Brennan and Schwartz (1985) define basis as "the difference between the future prices and the current spot price". Forecast power and premium models state that it can be expressed as the sum of an expected

premium and an expected change in the spot price. Formally:

$$F(t,T) - S(t) = \mathbb{E}_t[P(t,T)] + \mathbb{E}_t[S(T) - S(t)],$$

where $\mathbb{E}_t[P(t,T)] = \mathbb{E}[P(t,T)|\mathcal{F}_t]$ in the historical (physical) measure $\mathbb{P}$.

The expected premium $\mathbb{E}_t[P(t,T)]$ can be defined as the bias of the futures price as a forecast of the future spot price. That is:

$$\mathbb{E}_t[P(t,T)] = F(t,T) - S(t) - \mathbb{E}_t[S(T)].$$

Note, however, that the storage and premium theories are not concurrent. In the opposite, they give different explanations to the same event. Fama and French (1987) clarify this point with an example. Consider the case where the basis is negative[1]. The premium approach implies that a negative basis is explained by the expectation that prices will fall because some event (a harvest, for example) will substantially increase inventories. The storage theory, in its way, would explain the same situation with the argument that a negative basis is caused by low inventories and a convenience yield larger than interest and storage costs.

The situation that occurs when the basis has a positive sign, that is, futures prices are higher than spot prices, is called *contango* or premium markets. Similarly, when spot prices are higher than future prices (negative basis), we have what is called *backwardation*, or discount markets. Further references can be found in Hull (2006).

In this work, we will develop and calibrate a model based on the expected premium theory. Nevertheless, some applications of the storage theory can be found in Fama and French (1987) and Dixit and Pindyck (1994). We review some models dealing with expected premium and convenience yields, focusing on one factor and two-factor models.

Some of the two-factor models that will be studied in this Chapter have been estimated by Kalman Filter. Therefore we include in Appendix C an explanation of the state space representation and the Kalman Filter.

2.2 One-Factor and Two-Factor Models

In this Section, we build a literature review of one and two-factor commodity models. Following a historical ordering of the models, we describe their equations, estimation methods and main conclusions.

[1]Equivalently we can say that the spot price is greater than the futures price at time t for delivery at time T.

We first address the Gibson and Schwartz two-factor model for pricing financial and real assets contingent on the price of oil. The factors used are the spot price of oil and the convenience yield. They consider the spot price of oil as the fundamental, but not the only, determinant of contingent claim prices. That is why the convenience yield is added to the model.

The model assumes that the spot price of oil and the net convenience yield follow a joint diffusion process specified as:

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma_1 \mathrm{d}Z_{1t}, \tag{2.2}$$

$$\mathrm{d}\delta_t = \kappa(\alpha - \delta_t)dt + \sigma_2 \mathrm{d}Z_{2t}. \tag{2.3}$$

where δ is the instantaneous net convenience yield of oil, κ is the speed of mean reversion, α is the long run equilibrium level and $\mathrm{d}Z_1 \mathrm{d}Z_2 = \rho \mathrm{d}t$, where ρ is the correlation coefficient between the two standard Brownian motions.

The settlement price of the closest maturity crude oil futures contract trading on the NYMEX is used as a proxy for the spot price S, and the instantaneous convenience yield was computed using the relationship between the futures and the spot price of a commodity defined in Brennan and Schwartz (1985), namely:

$$F(S,T) = Se^{(r-\delta)(T-t)},$$

where r is the annualized riskless interest rate and δ the annualized convenience yield.

The market price of risk λ is estimated from market prices of all crude oil futures contracts traded on the NYMEX from 1984 to 1986. Then, they are compared to their theoretical prices computed by solving numerically a partial differential equation that describes the behavior of a futures contract. For more details on the estimation of λ, refer to Gibson and Schwartz (1990).

Equations (2.2) and (2.3) are discretized and estimated using unrestricted regressions on weekly data in order to estimate $\kappa, \alpha, \sigma_2, \rho$ and σ_1, while the market price of convenience yield, λ, was estimated separately.

The model's performance depends on the assumptions made on λ. However, their results show that the two-factor model is a reliable instrument for the purpose of valuing short term financial instruments, such as futures contracts.

An important contribution to the factors model in the literature was given by Schwartz (1997). In this paper Schwartz compares three models of stochastic behavior of commodity prices that take into account mean reversion. The first of these three models, a simple one-factor model where the logarithm of

the prices are assumed to follow a mean reverting process, is the most cited one, which we shall address now.

The spot price is assumed to follow the stochastic process:

$$\mathrm{d}S_t = \kappa(\mu - lnS_t)S\mathrm{d}t + \sigma S\mathrm{d}Z_t$$

Let X denote the logarithm of the spot price S. Applying Ito's Lemma to the previous equation, the log price is characterizes by an Ornstein-Uhlenbeck (OU) stochastic process:

$$\mathrm{d}X_t = \kappa(\alpha - X_t)\mathrm{d}t + \sigma\mathrm{d}Z_t; \quad \text{where} \quad \alpha = \mu - \frac{\sigma^2}{2\kappa}. \tag{2.4}$$

For the risk-neutral model, consider $\alpha^* = \alpha - \lambda$, where λ is the market price of risk, and $\mathrm{d}Z^*$ is the increment to the Brownian motion under the equivalent martingale measure.

The OU process can be written under the risk-neutral measure as:

$$\mathrm{d}X_t = \kappa(\alpha^* - X_t)\mathrm{d}t + \sigma\mathrm{d}Z_t^*. \tag{2.5}$$

Calculating the first and second moments of X_t, we can obtain the futures price of a commodity with maturity T under the equivalent martingale measure:

$$F(S,T) = \widetilde{\mathbb{E}}[S(T)] = \exp\left(\mathbb{E}_0[X(T)] + \frac{1}{2}\mathbb{V}\mathrm{ar}_0[X(T)]\right).$$

$$\text{Thus,} \quad F(S,T) = \exp\left[e^{-\kappa T}X + (1 - e^{-\kappa T})\alpha^* + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa T})\right],$$

and

$$\ln F(S,T) = e^{-\kappa T}X + (1 - e^{-\kappa T})\alpha^* + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa T}). \tag{2.6}$$

The coefficient κ determines the speed of the mean reversion to the long run log price α, the coefficient σ characterizes the volatility process, and $\mathrm{d}Z$ is an increment to a standard Brownian motion.

The estimation is carried in the state-space form, using the Kalman Filter. Let $\mathbf{y}_t = [\ln\mathbf{F}(T_i)]$, $i = 1, 2, \ldots, N$, be an $(N \times 1)$ vector of observable variables, $\mathbf{x}_t$ the state vector, and

$$\mathbf{d}_t = \left[(1 - e^{-\kappa T_i})\alpha^* + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa T_i})\right],$$

for $i = 1, 2, \ldots, N$, an $(N \times 1)$ a vector of deterministic variables. Let also $\mathbf{Z}_t = [e^{-\kappa T_i}]$, for $i = 1, 2, \ldots, N$ an $(N \times 1)$ vector and ϵ_t, a $(N \times 1)$ vector

of serially uncorrelated disturbances with $\mathbb{E}[\epsilon_t] = 0$ and $\mathbb{V}\text{ar}(\epsilon_t) = \mathbf{H}$. We can then write the measurement equation as:

$$\mathbf{y}_t = \mathbf{d}_t + \mathbf{Z}_t\mathbf{x} + \epsilon_t \quad t = 1, \ldots, T.$$

For the transition equation let $\mathbf{c}_t = \kappa\alpha\Delta t$, $\mathbf{Q}_t = 1 - \kappa\Delta t$ and η_t serially uncorrelated disturbances with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{V}\text{ar}[\eta_t] = \sigma^2\Delta t$. From (2.5) the transition equation can be written as:

$$\mathbf{x}_t = \mathbf{c}_t + \mathbf{Q}_t\mathbf{x}_{t-1} + \eta_t, \quad t = 1, \ldots, T.$$

Schwartz (1997) uses weekly observations of future prices for oil, copper and gold, and considers five maturity contracts. The interest rates used are 3-Month Treasury Bills.

The results of the one factor model show highly significant speed adjustment coefficients for all four markets, and a market price of risk not significantly different from zero. Logarithms of spot prices for the one-factor model follows closely (but not identically) the logarithm of future prices for the oil market. Encouraging results were also found for the copper data, differently from the gold market, in which the one-factor model could not be fitted to the data.

The second model is a two-factor model where the first factor is the spot price of the commodity, and the second is the instantaneous convenience yield δ. These factors are modeled as:

$$\mathrm{d}S = (\mu - \delta)S\mathrm{d}t + \sigma_1 S dZ_1 \tag{2.7}$$

$$\mathrm{d}\delta = \kappa(\alpha - \delta)\mathrm{d}t + \sigma_2\mathrm{d}Z_2, \tag{2.8}$$

$$\mathrm{d}Z_1\mathrm{d}Z_2 = \rho\mathrm{d}t \tag{2.9}$$

The spot price is a standard process that allows a convenience yield, and follows a OU stochastic process described in (2.8).

The author remarks that assuming a risk free interest rate following a OU process, the model can can be extended to a three-factor model of commodity contingent claims. It models the spot price of the commodity, the instantaneous yield and the instantaneous interest rate.

Schwartz's second model two is estimated by a traditional application of the Kalman Filter, whose transition and measurement equations can be seen at Schwartz (1997), while model three has a simplified estimation. Instead of simultaneously estimating the equations, Schwartz first estimates the interest rate process and then uses model three to estimate the parameters of the spot

price and the convenience yield process.

As pointed by Schwartz, the statistical comparison between the empirical results is not straightforward as the models are not nested. Even though models two and three are nested, they have the same number of estimated parameters due to the simplification hypothesis done to estimate the interest rate process. This complicates the how the likelihood's function value could be interpreted.

However, in relative performance, models two and three outperform model one for all data sets, but alternate better relative performances among them, depending on the data set used.

Lucia and Schwartz (2002) develop both one-factor and two-factor models to recover the dynamics of the spot price. Their models differ with respect to the stochastic process assumed for the spot price, the number of factors considered, and the way the deterministic component (such as a trend) is incorporated into the model.

The one-factor models assume that interest rates are constant, making futures and forward prices equal. This implies that these prices always move in the same direction (which is something contradicted by the data). Although the constant interest rate assumption simplifies the calculations, it makes the results relative to the long maturity contracts less accurate. Their one-factor model for the logarithm of spot prices is:

$$\ln P_t = f(t) + Y_t$$

where $f(t)$ is a deterministic function of the time (totally predictable) that models seasonality and Y_t is an Orstein-Uhlenbeck process with a long term mean equal to zero.

Lucia and Schwartz's two-factor model allows for a long term equilibrium price level and a short term mean reverting component, in the spirit of the Schwartz and Smith model.[2]

$$\begin{aligned} \ln P_t &= f(t) + X_t + \epsilon_t \\ \mathrm{d}X_t &= -\kappa X_t \mathrm{d}t + \sigma_X \mathrm{d}Z_x \\ \mathrm{d}\epsilon_t &= \mu_\epsilon \mathrm{d}t + \sigma_\epsilon \mathrm{d}Z_\epsilon \end{aligned}$$

[2]The model developed by Schwartz and Smith (2000), and which we follow in this work suggests that besides mean reverting, some commodity present uncertainty about the equilibrium price to which prices revert. Their model then captures both mean reverting and uncertainty effects. Further details of the model's structure can be found on Chapter 2, while its estimation is addressed in Chapter 3.

The idea behind mean reverting processes is that when commodity prices are higher than some long-run mean or equilibrium price level, the supply of the commodity will increase because producers will have incentives to produce more, or enter the market. That movement will end by increasing supply and putting downward pressure on prices until they return to the equilibrium level. The exact opposite happens when prices are bellow their long-run equilibrium level. Therefore, even with the presence of short-run disequilibriuns there is a natural correction movement inherent to commodity prices.

Hikspoors and Jaimungal (2007) also propose a two-factor model in order to correct some of the deficiencies of the one-factor models. The authors build mean-reverting spot price processes, with and without a jump process. Hikspoors and Jaimungal classify their model as a perturbation on the standard one-factor mean-reverting approach, improving the fit of forward price curves without loosing tractability.

Instead of modeling the second factor as a geometric Brownian motion in the spirit of Pilipovic (1997), the authors choose the mean-reverting level of the first factor to revert to a second long-run mean. This second factor's perturbation does not change the stationarity behavior presented by the one-factor model. The spot-prices are defined as:

$$
\begin{aligned}
S_t^{(i)} &= exp(g_t^{(i)} + X_t^{(i)}), \quad i = 1, 2 \\
g_t^{(i)} &= A_0^{(i)} + \sum_{k=1}^{n} (A_k^{(i)} \sin(2\pi kt) + B_k^{(i)} \cos(2\pi kt)) \\
\mathrm{d}X_t^{(i)} &= \beta_i (Y_t^{(i)} - X_t^{(i)}) \mathrm{d}t + \sigma_X^{(i)} \mathrm{d}W_t^{(i)} \\
\mathrm{d}Y_t^{(i)} &= \alpha_i (\phi_i - Y_t^{(i)}) \mathrm{d}t + \sigma_Y^{(i)} \mathrm{d}Z_t^{(i)}
\end{aligned}
$$

where $g_t^{(i)}$ models seasonal movements, β_i controls the speed of mean reversion of $X_t^{(i)}$ to the stochastic long-run level $Y_t^{(i)}$, α_i controls the speed of mean-reversion of the long-run level $Y_t^{(i)}$ to the target long-run mean ϕ_i, and $\sigma_X^{(i)}$ and $\sigma_Y^{(i)}$ are the volatility parameters that control the size of fluctuations around the means. Finally,

$$\mathrm{d}W_t^{(1)} \mathrm{d}W_t^{(2)} = \rho_{12} \mathrm{d}t$$

and

$$\mathrm{d}W_t^{(i)} \mathrm{d}Z_t^{(i)} = \rho_i \mathrm{d}t, \quad i = 1, 2.$$

The calibration process is an alternative to the Kalman Filter, and it is performed in two-steps. First the authors calibrate the pure diffusion two-factor

model to market futures prices, resulting in the risk-neutral model parameters. Next, they reference a two-stage nonlinear least squares method to estimate the real-world measures and extract the implied market prices of risk. This method is commonly used in interest rate models calibration, and when applied to NYMEX Light Sweet Crude Oil data produced realistic and stable parameters.

We illustrate the difference between one factor and two factor models in Figures 2.1 and 2.2. We plot a five-year term structure of futures prices, using the futures equation of Schwartz (1997) for the one-factor model and of Schwartz and Smith (2000) for the two-factor model. The expression for the Schwartz (1997) model is:[3]

$$F(S,T) = \exp[e^{-\kappa(T-t)}X + (1-e^{\kappa(T-t)})\alpha^* + \frac{\sigma^2}{4\kappa}(1-e^{-2\kappa(T-t)})],$$

while the expression for Schwartz-Smith's model is:

$$\begin{aligned} F(S,T) =& \exp\left[e^{-\kappa(T-t)}X_0 + Y_0 + \mu^*(T-t) - (1-e^{-\kappa(T-t)})\frac{\lambda_X}{\kappa}\right] \\ =& \exp\left[\frac{1}{2}\left((1-e^{-2\kappa(T-t)})\frac{\sigma^2}{2\kappa} + \eta^2(T-t) + 2(1-e^{-\kappa(T-t)})\frac{\rho\sigma\eta}{\kappa}\right)\right]. \end{aligned}$$

We use the parameters in Table 4.1 to generate the data with which we shall build both term structures. The parameter α^* of the 1997 Schwartz model corresponds to $-\lambda_x/\kappa$ in the Schwartz and Smith 2000 model.

In Figure 2.1, we can see that the one-factor model is not able to model the long end of the curve, resulting in flat prices for the longest maturities. This problem is corrected by the two-factor model, which is able to model the long end movements of the futures curve. The price surface for the two-factor model is shown in Figure 2.2.

Finally, in figure 2.3 we plot the absolute difference between the prices generated by the one and two-factor models. Note that there is almost no difference for the short maturities. This feature changes when we consider longer contracts, making the difference clear.

[3]Note that this futures equation is equivalent to the exponential of the log futures prices shown in equation 2.6

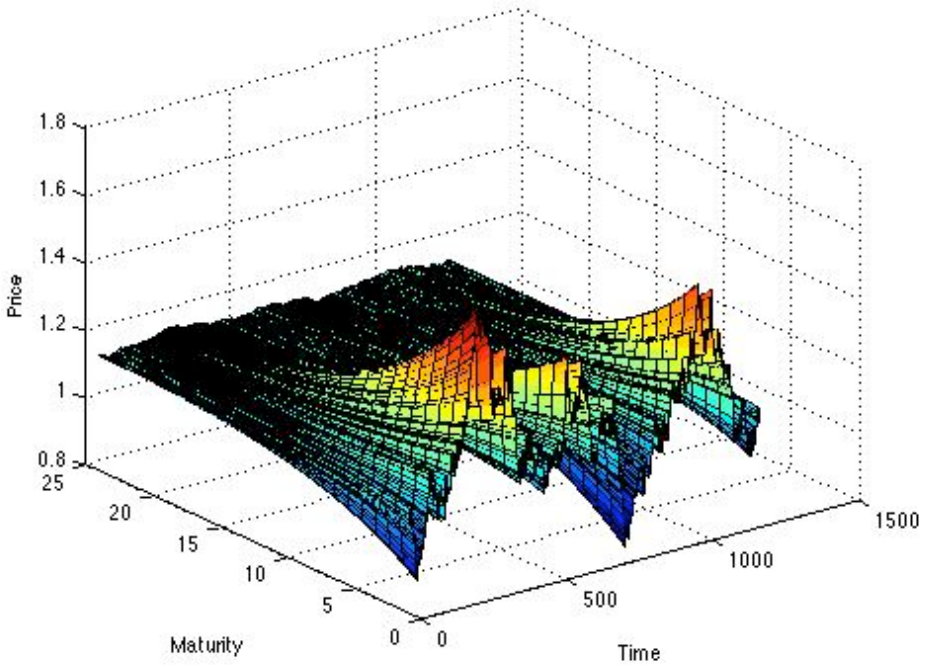


Figure 2.1: Futures prices generated by the Schwartz 1997 one-factor model

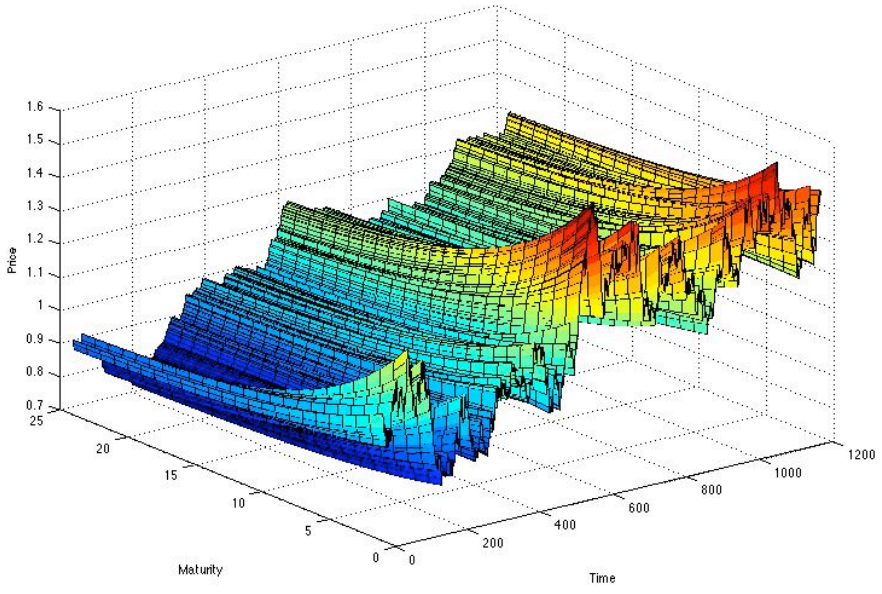


Figure 2.2: Futures prices generated by the Schwartz and Smith 2000 two-factor model

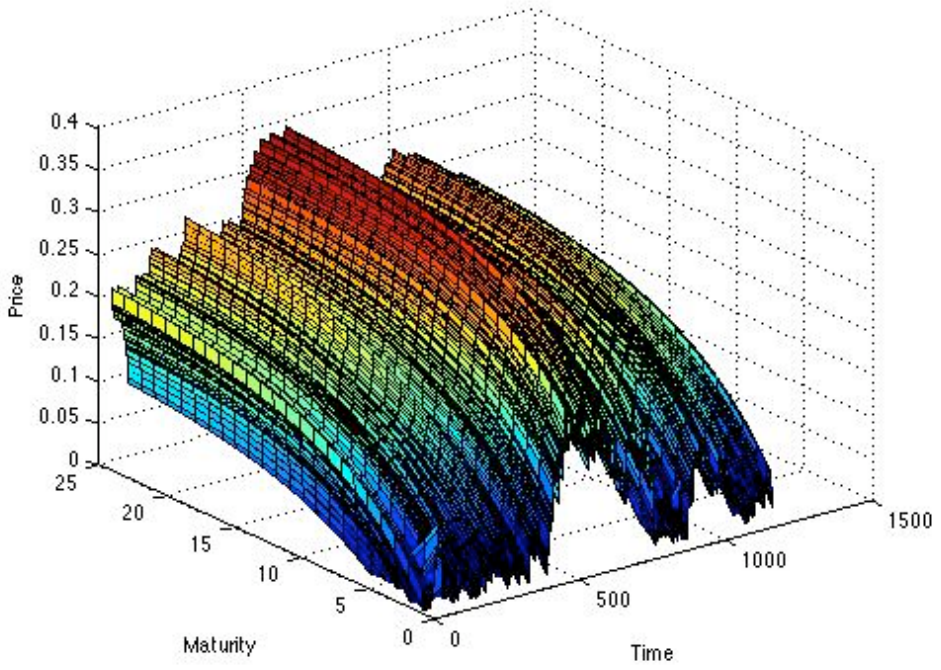


Figure 2.3: Absolute difference between futures prices generated with one and two factor models

Chapter 3

The Schwartz and Smith Model

3.1 Some basic results

In this chapter, we study in detail the derivation of the Schwartz and Smith model, which is our working model.

Schwartz and Smith develop a two-factor model of commodity prices that allows mean-reversion in short-term prices and uncertainty in the equilibrium level to which prices revert.

They model long-run prices as a Brownian motion, reflecting expectations of the exhaustion of existing supply, improving technology for the production and discovery of the commodity, inflation, as well as political and regulatory effects.

The short-term prices reflect the expected vanishing of the difference between spot and equilibrium prices. They are modeled as a Ornstein-Uhlenbeck process. These deviations can be interpreted as short-term changes in demand, resulting from weather changes of intermitent supply disruptions, which are eased by the adjustment of inventory levels made by market participants.

While from an economical point view, as mentioned in Section 2.1, the risk premium and the convenience yield views are quite different, it is interesting to notice that the model in Schwartz and Smith (2000) can be viewed also as a sthocastic convenience model. More precisely, it was shown in their work that the long/short term is equivalent to the stochastic convenience model in Gibson and Schwartz (1990), and a dictionary between the two models was also presented. It was also argued in Schwartz and Smith (2000) that the long/short term model leads to results that are more easily interpreted.

We begin by deriving the model in the physical probability measure $\mathbb{P}$ and then move to the risk-neutral measure $\widetilde{\mathbb{P}}$.

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The set Ω represents all possible outcomes for a financial variable in the future. The set of all possible scenarios has an actual probability measure $\mathbb{P}$, which we sometimes call "real world" measure. $\mathcal{F}$ is a σ-algebra, which is the collection of subsets of Ω whose probability is defined, and $\mathbb{P}$ is a probability measure.

Our model equations are:

$$\mathrm{d}X_t = -\kappa X_t \mathrm{d}t + \sigma \mathrm{d}W_t^X, \tag{3.1}$$

$$\mathrm{d}Y_t = \mu \mathrm{d}t + \eta \mathrm{d}W_t^Y, \tag{3.2}$$

$$S_t = \exp(X_t + Y_t). \tag{3.3}$$

Equation (3.1) describes the short-term deviations in prices which reverts toward zero, following an Orstein-Uhlenbeck (OU) process. The coefficient κ describes the rate at which the short-term deviations are expected to vanish, while σ is the volatility parameter for the short-term process.

In equation (3.2), the process Y_t describes the equilibrium level of prices as a standard Brownian motion with a drift parameter μ and volatility parameter η. The changes in this process denote fundamental changes that are expected to persist.

Finally, S_t models the joint dynamics of the processes. In order to keep the model as simple as possible, and following Schwartz and Smith (2000), we choose not to include seasonal adjustments (although it can be incorporated by including deterministic time-depending constants).

To find the solution for the process X_t, we apply Ito's Lemma using $f(x,t) = e^{kt}X$ and integrate both sides of the resulting equation. That is:

$$\mathrm{d}(e^{kt}X_t) = ke^{kt}X_t \mathrm{d}t + e^{kt}[(-kX_t dt + \sigma \mathrm{d}W_t^X].$$

Integrating both sides from t to T:

$$e^{kT}X_T - e^{kt}X_t = \sigma \int_t^T e^{ks} \mathrm{d}W_s^X$$

$$e^{kT}X_T = e^{kt}X_t + \sigma \int_t^T e^{ks} \mathrm{d}W_s^X.$$

I.e,

$$X_T = e^{-k(T-t)}X_t + \sigma e^{-\kappa T} \int_t^T e^{ks} \mathrm{d}W_s^X. \tag{3.4}$$

The first and second conditional moments of (3.4) can be calculated as follow:

$$\mathbb{E}[X_T|X_t] = \mathbb{E}\left[e^{-k(T-t)}X_t + \sigma e^{-\kappa T}\int_t^T e^{ks}\mathrm{d}W_s^X|X_t\right]$$

which yields,

$$\mathbb{E}[X_T] = e^{-k(T-t)}X_t. \tag{3.5}$$

Note that the only stochastic part in (3.4) is $\int_0^t e^{ks}\mathrm{d}W_s^X$, so X_t will have the integral's distribution. As a Brownian motion is normally distributed, so are $\int_0^t e^{ks}\mathrm{d}W_s^X$ and X_t. (see Theorem 4.4.9 Shreve (2004)).
Besides, also by Theorem 4.4.9 in Shreve (2004), $I(t) = \int_0^t e^{ks}dW_s^X$ has zero mean and variance $\int_0^t e^{2ks}\mathrm{d}s$. Therefore we have that:

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X_T] &= \mathbb{V}\mathrm{ar}\left[\sigma\int_t^T e^{ks}dW^X s\right]\\ &= \sigma^2 e^{-2\kappa T}\int_t^T e^{-2ks}\mathrm{d}s\end{aligned}$$

Hence,

$$\mathbb{V}\mathrm{ar}[X_T] = (1-e^{-2k(T-t)})\frac{\sigma^2}{2k}. \tag{3.6}$$

For the process $\mathrm{d}Y_t$, we simply integrate both sides of equation (3.2) from t to T yielding

$$Y_T = Y_t + \mu(T-t) + \eta\int_t^T \mathrm{d}W_t^Y, \tag{3.7}$$

from which

$$\mathbb{E}[Y_T] = Y_t + \mu(T-t), \tag{3.8}$$

and

$$\mathbb{V}\mathrm{ar}[Y_T] = \eta^2(T-t) \tag{3.9}$$

are directly calculated.

We assume that the processes are correlated and have a correlation coefficient ρ_{XY}. That is, we assume that $\mathrm{d}W_t^X\mathrm{d}W_t^Y = \rho_{XY}\mathrm{d}t$. We can calculate their covariance as

$$\mathbb{C}\mathrm{OV}[X_t, Y_t] = \mathbb{E}[(X_t - \mathbb{E}(X_t))(Y_t - \mathbb{E}(Y_t))].$$

Using equations (3.4) and (3.7), we have that:

$$\begin{aligned}\mathbb{COV}[X_t, Y_t] &= \mathbb{E}\left[\sigma e^{-\kappa T}\int_t^T e^{ks}\mathrm{d}W_s^X \eta \int_t^T \mathrm{d}W_s^Y\right] \\ &= e^{-\kappa T}\sigma\eta\mathbb{E}\left[\int_t^T e^{ks}\mathrm{d}W_s^X \int_t^T \mathrm{d}W_s^Y\right] \\ &= \sigma e^{-\kappa T}\eta\rho\int_t^T e^{ks}\mathrm{d}s \\ &= (1-e^{-\kappa(T-t)})\frac{\sigma\eta\rho}{k}.\end{aligned}$$

Therefore the variance-covariance matrix for the processes is

$$\mathbb{COV}(X_t, Y_t) = \begin{pmatrix} (1-e^{-2k(T-t)})\frac{\sigma^2}{2k} & (1-e^{-\kappa(T-t)})\frac{\sigma\eta\rho}{k} \\ (1-e^{-\kappa(T-t)})\frac{\sigma\eta\rho}{k} & \eta^2(T-t)\end{pmatrix}. \tag{3.10}$$

3.1.1 Brief Review of Risk Neutral Pricing

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. For purposes of derivative pricing, we work with a risk-neutral probability measure $\widetilde{\mathbb{P}}$,[1] but both measures $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent. That is, they agree on what is possible and what is not. Formally:

Definition 3.1.1. *Let Ω be a nonempty set and $\mathcal{F}$ a σ-algebra of subsets of Ω. Suppose that $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are two probability measures on $(\Omega, \mathcal{F})$. $\widetilde{\mathbb{P}}$ is said to be absolute continuous with respect to $\mathbb{P}$ on the σ-algebra $\mathcal{F}$, and we write $\widetilde{\mathbb{P}} << \mathbb{P}$, if for all $A \in \mathcal{F}$,*

$$\mathbb{P}[A] = 0 \Rightarrow \widetilde{\mathbb{P}}[A] = 0$$

If both $\widetilde{\mathbb{P}} << \mathbb{P}$ and $\mathbb{P} << \widetilde{\mathbb{P}}$ hold, we say that $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are equivalent, and we shall write $\widetilde{\mathbb{P}} \approx \mathbb{P}$.

The risk-neutral or risk-free measure imposes no-arbitrage restrictions for the calculated price. We define arbitrage as:

Definition 3.1.2. *An arbitrage is a portfolio value process X_t satisfying $X_0 = 0$ and also satisfying, for some time $T > 0$,*

$$\mathbb{P}[X_t \geq 0] = 1, \quad \mathbb{P}[X_T > 0] > 0.$$

[1] A probability measure $\widetilde{\mathbb{P}}$ is called risk neutral because when used for asset valuation its formula does not take into account any risk aversion, in contrast to valuation in terms of expected utility. For more on utility pricing, see Section 2.3 of Föllmer and Schied (2004).

That is, an arbitrage is a way of trading where one starts with zero capital, and at time T has not lost any money. Furthermore, there is a positive probability of having made money.

A very important theorem, called the First Fundamental Theorem of Asset Pricing, relates risk neutral measures and arbitrage opportunities. Following Shreve (2004), we have:

Theorem 3.1.1. *If a model has a risk-neutral probability measure, then its does not admit arbitrage.*[2]

Proof. Define a discount process, or discount factor as

$$D_T = e^{-\int_0^T r(t)\mathrm{d}t}. \tag{3.11}$$

If a market has a risk-neutral probability measure $\widetilde{\mathbb{P}}$, then every discounted portfolio value process is a martingale under $\widetilde{\mathbb{P}}$. In particular, every portfolio value process satisfies $\widetilde{\mathbb{E}}[D_T X_T] = X_0$. Let X_T be a portfolio value process with $X_0 = 0$. Then we have

$$\widetilde{\mathbb{E}}[D_T X_T] = 0.$$

Suppose that X_T satisfies $P[X_T < 0]$. Since $\widetilde{\mathbb{P}}$ is equivalent to $\mathbb{P}$, we have also $\widetilde{\mathbb{P}}[X_T < 0] = 0$. This, together with $\widetilde{\mathbb{E}}[D_T X_T] = 0$, implies

$$\widetilde{\mathbb{P}}[X_T > 0] = 0,$$

for otherwise we would have

$$\widetilde{\mathbb{P}}[D_T X_T > 0] > 0,$$

which would imply

$$\widetilde{\mathbb{E}}[D_T X_T] > 0.$$

Because $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent, we have also $\mathbb{P}[X_T > 0] = 0$. Hence, X_t is not an arbitrage. In fact, there cannot exist an arbitrage since every portfolio value process X_t satisfying $X_0 = 0$ cannot be an arbitrage. □

This change of measures from the "real world" measure to the risk neutral measure is performed with the help of a random variable Z, and is guaranteed by the Girsanov Theorem. Before we state this theorem, we give some basic definitions.

Definition 3.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For each $w \in \Omega$, suppose there is a continuous function W_t, $t \geq 0$, that satisfies $W_0 = 0$ and*

[2]With additional assumptions, the converse is also true. See Föllmer and Schied (2004).

that depends on w. Then, W_t, $t \geq 0$, is a Brownian motion if for all $0 = t_0 \leq t_1 \leq \cdots \leq t_m$ the increments

$$W_{t_1} = W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}}$$

are independent and each of these increments is normally distributed with

$$\begin{aligned} \mathbb{E}[W_{t_{i+1}} - W_{t_i}] &= 0, \\ \mathbb{V}\text{ar}[W_{t_{i+1}} - W_{t_i}] &= t_{i+1} - t_i. \end{aligned}$$

Definition 3.1.4. *A filtration for the Brownian motion W_t is a collection of σ-algebras of $\mathcal{F}_t$ such that*

1. *$\mathcal{F}_s \subset \mathcal{F}_t$, for $s < t$*
2. *W_t is $\mathcal{F}_t$-measurable*
3. *$W_t - W_s$ is independent of $\mathcal{F}_s$, for $0 \leq s < t$*

Definition 3.1.5. *A stochastic process X_t is said to be "adapted" to $\mathcal{F}_t$, if X_t is $\mathcal{F}_t$-measurable, for every t.*

Now we are ready to state Girsanov's Theorem for a simple Brownian and a new probability measure $\widetilde{\mathbb{P}}$. The proof of this theorem can be found at Shreve (2004).

Theorem 3.1.2. *(Girsanov) Let W_t, $0 \leq t \leq T$, be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}_t$, $0 \leq t \leq T$, be a filtration for W_t. Let θ_t, $0 \leq t \leq T$, be an adapted process. Then define*

$$Z(t) = \exp\left(-\frac{1}{2}\int_0^t \theta^2(u)\mathrm{d}u - \int_0^t \theta(u)\mathrm{d}W_u\right), \tag{3.12}$$

and assume $\mathbb{E}\left[\int_0^t \theta^2(u)Z^2(u)\mathrm{d}u\right] < \infty$.
We define a new probability measure $\widetilde{\mathbb{P}}$ as $\widetilde{\mathbb{P}}(A) = \int_A Z(T)\mathrm{d}\mathbb{P}$. Then:

$$\mathbb{E}[Z(T)] = 1 \quad \textit{and} \tag{3.13}$$

$$\widetilde{W}_t = W_t + \int_0^1 \theta(u)\mathrm{d}u. \tag{3.14}$$

It is important to say that if a model allows multiple risk-neutral measures, they might lead to different prices for a derivative security. Since it is still a matter of debate as how to correctly price in incomplete markets with multiple measures, it is important to have a criteria to guarantee that the risk-neutral measure is unique.

With a unique risk-neutral measure, it is possible to know the present value of the future payments of a derivative security. This duty is performed by a discount function as defined in equation (3.11) In other words, the price or value at time t of any derivative that pays $V(T)$ at time T is

$$V(t) = \frac{1}{D(t)}\widetilde{\mathbb{E}}[D(T)V(T)|\mathcal{F}_t].$$

Where $\widetilde{\mathbb{E}}$ is the expected value taken with respect to the risk neutral measure. Formally we can say that:

Definition 3.1.6. *A derivative is replicable if there is a trading strategy $X(t)$ such that $X(T) = V$, where V is a payoff received at the maturity time T.*

Claim 1. *Any derivative defined by a $\mathcal{F}_t$-measurable random variable V is replicable, and the value at time t of any replicating portfolio $X(t)$ is*

$$X(t) = \widetilde{\mathbb{E}}[e^{-\int_t^T r(u)\mathrm{d}u}V(t)|\mathcal{F}_t], \tag{3.15}$$

where $r(t)$ is an adapted ($\mathcal{F}_t$-measurable) interest rate process and $e^{-\int_0^t r(u)\mathrm{d}u}$ is the discount process. We should also note that if a claim is replicable, then its price is the same in all risk-neutral measures.

The Second Fundamental Theorem of Asset Pricing guarantees the uniqueness of the risk neutral measure, if the market is complete. That is, if all contingent claims are replicable. Formally:

Definition 3.1.7. *An arbitrage free model is called complete if every contingent claim is replicable.*

Theorem 3.1.3. *Consider a market that has a risk-neutral probability measure. The market is complete if and only if the risk-neutral measure is unique.*

Proof. See Shreve (2004). □

We rely on the risk neutral framework that we have built, to formally define futures prices.

Definition 3.1.8. *The futures price of an asset whose value at time T is $S(T)$ is given by*

$$\textsc{Fut}_S(t,T) = \tilde{\mathbb{E}}[S(T)|\mathcal{F}_t], \quad 0 \leq t \leq T.$$

Recall that $S_t = \exp(X_t + Y_t)$. It is immediate that the log of the future spot price is normally distributed with

$$\mathbb{E}[\ln(S_t)] = e^{-\kappa(T-t)}X_t + \mu(T-t). \tag{3.16}$$

and

$$\mathbb{V}\text{ar}[\ln(S_t)] = (1 - e^{-2\kappa(T-t)})\frac{\sigma^2}{2\kappa} + \eta^2(T-t) + (1 - e^{\kappa(T-t)})\frac{\sigma\eta\rho}{\kappa}. \quad (3.17)$$

3.1.2 Risk Neutral Model

We now apply the background of risk neutral pricing we reviewed in the previous section to derive a risk neutral version of the Schwartz and Smith model.

The risk-neutral version of the model can be written as:

$$\mathrm{d}X_t^* = (-\kappa X_t^* - \lambda_X)\mathrm{d}t + \sigma\mathrm{d}\widetilde{W}_t^X, \quad (3.18)$$

$$\mathrm{d}Y_t^* = (\mu - \lambda_Y)\mathrm{d}t + \eta\mathrm{d}\widetilde{W}_t^Y. \quad (3.19)$$

where $\mathrm{d}\widetilde{W}^X t$ are $\mathrm{d}\widetilde{W}^Y t$ are increments of Brownian motions with

$$\mathrm{d}\widetilde{W}_t^X \mathrm{d}\widetilde{W}_t^Y = \rho_{XY}\mathrm{d}t.$$

The market price of risk settings adopted by Schwartz and Smith is fairly standard, assuming constant market prices of risk. The risk premiums λ_x and λ_y take the form of adjustments to the drift, being subtracted from the drift processes X and Y.

Schwartz and Smith explain that, if we assume constant correlations between changes in the state variables and the aggregate wealth of the economy, we would arrive at the same result of a constant drift reduction to the market prices of risk.

A possible generalization of this constant subtraction assumption is also pointed by the authors. They suggest that the short term risk premium to be modeled as a linear function of the short-term deviations. That is:

$$\lambda_X = \beta X_t + \alpha. \quad (3.20)$$

This functional form would allow for the possibility that the short-term risk premium would be higher (or lower) in periods when spot prices are higher than the equilibrium level. Substituting equation (3.20) into (3.18) would yield:

$$\mathrm{d}X_t^* = -[(\kappa + \beta)X_t^* + \alpha]\mathrm{d}t + \sigma\mathrm{d}\widetilde{Z}_t^X,$$

and using that $\tilde{\kappa} = \kappa + \beta$, we would have that:

$$\mathrm{d}X_t^* = (-\tilde{\kappa}X_t^* - \alpha)\mathrm{d}t + \sigma\mathrm{d}\widetilde{Z}_t^X.$$

Returning to the risk-neutral model's settlement, the solutions for its equa-

tions are given by:

$$X_T^* = X_t^* e^{-\kappa(T-t)} - \frac{\lambda_X}{\kappa} e^{\kappa t} + \sigma e^{-\kappa T} \int_t^T e^{\kappa s} \mathrm{d}\widetilde{W}_s^X, \tag{3.21}$$

and

$$Y_T^* = Y_t^* + (\mu - \lambda_Y)(T-t) + \eta \int_t^T d\widetilde{W}_s^Y, \tag{3.22}$$

which were calculated using Ito's Lemma similarly to what was done in Section 3.1.

The model's moments were also calculated in the same way as previously. The first conditional moments are given by:

$$\widetilde{\mathbb{E}}(X_T) = X_t^* e^{-\kappa(T-t)} - \frac{\lambda_X}{\kappa} e^{\kappa t} \tag{3.23}$$

and

$$\widetilde{\mathbb{E}}(Y_T) = Y_t^* + \mu^*(T-t). \tag{3.24}$$

where $\mu^* = (\mu - \lambda_Y)$.

The variance-covariance matrix is exactly the same as in (3.10). So we have that

$$\mathbb{C}\mathrm{OV}(X^*, Y^*) = \mathbb{C}\mathrm{OV}(X, Y).$$

Note that the change from the physical measure $\mathbb{P}$ to the risk-neutral $\widetilde{\mathbb{P}}$ changes the mean rate of return of the stochastic processes, (see equations (3.5) and (3.23) for X_t and (3.8) and (3.24) for Y_t, but not the volatility.

As remarked by Shreve (2004), the volatility tells us which price paths are possible. After the change of measure, we are still considering the same set of stock price paths, but we have shifted their probability.

Finally, we address futures prices. Under the risk neutral process, the log of the future spot price, $\ln(S_t) = X_t + Y_t$ is normally distributed with

$$\widetilde{\mathbb{E}}[\ln(S_t)] = X_t^* e^{-\kappa(T-t)} - \frac{\lambda_X}{\kappa} e^{\kappa t} + Y_t^* + \mu^*(T-t)$$

and

$$\widetilde{\mathbb{V}\mathrm{ar}}(\ln(S_t)) = \mathbb{V}\mathrm{ar}(\ln(S_t)).$$

Now that we have obtained the moments in their risk neutral framework, we are ready to obtain the futures prices. Those prices will be the focus of our attention in the subsequent chapters.

Recall that futures prices are defined as the expected future spot price under the risk neutral measure. The market price at time t for a futures

contract with time T until maturity is given by

$$\ln(F_{t,T}) = \ln(\widetilde{\mathbb{E}}(S_{t,T})).$$

Then, we have that:

$$\begin{aligned}
\ln(F_{t,T}) =& \widetilde{\mathbb{E}}[\ln(S_{t,T})] + \frac{1}{2}\widetilde{\mathbb{V}\text{ar}}[\ln(S_{t,T})] \\
=& e^{-\kappa(T-t)}X_t^* + Y_t^* + \mu^*(T-t) - \frac{\lambda_X}{\kappa}e^{\kappa t} \\
&+ \frac{1}{2}\left[(1-e^{-2\kappa(T-t)})\frac{\sigma^2}{2\kappa} + \eta^2(T-t) + 2(1-e^{-\kappa(T-t)})\frac{\rho\sigma\eta}{\kappa}\right],
\end{aligned}$$

i.e,

$$\ln(F_{t,T}) = e^{-\kappa(T-t)}X_t + Y_t + A(T,t), \tag{3.25}$$

where

$$\begin{aligned}
A(T,t) =& \mu^*(T-t) - (1-e^{-\kappa(T-t)})\frac{\lambda_X}{\kappa} \\
&+ \frac{1}{2}\left[(1-e^{-2\kappa(T-t)})\frac{\sigma^2}{2\kappa} + \eta^2(T-t) + 2(1-e^{-\kappa(T-t)})\frac{\rho\sigma\eta}{\kappa}\right].
\end{aligned}$$

Chapter 4

Calibration

4.1 Discretization and Parametrization of the Stochastic Processes

In spite of its continuous-time representation, the two-factor model has to be estimated with real data, therefore, in discrete time. We must then use some discretization scheme in order to give a discrete version of the continuous-time equations.

As we have seen in Chapter 3, the Schwartz and Smith model consists of a set of stochastic differential equations with two correlated Brownian motions W_t^X and W_t^Y.

For computational purposes, it is useful to consider discretized Brownian motions, where W_t is specified at discrete t values. We must set $\delta t = T/N$ for some positive integer N and let W_j denote W_{t_j} with $t_j = j\delta t$. Using $W_0 = 0$, (recall from the definition of the Brownian motion that $W_0 = 0$) we have that:

$$W_j = W_{j-1} + \mathrm{d}W_j \quad \text{for} \quad j = 1, 2, \ldots, N, \tag{4.1}$$

where each $\mathrm{d}W_j$ is an independent random variable of the form $\sqrt{\delta t}N(0,1)$.

Consider the SDE

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t, \quad X(0) = X_0 \quad , 0 \leq t \leq T. \tag{4.2}$$

Integrating (4.2) from 0 to t we have that:

$$X_t = X_0 + \int_0^t f(X_s)\mathrm{d}s + \int_0^t g(X_s)\mathrm{d}W_s, \quad 0 \leq t \leq T, \tag{4.3}$$

where f and g are scalar functions, and X_0 is the initial condition.

As (4.2) satisfies the conditions of the Variation of Constants Theorem (see Korn and Korn (2000)), the existence and uniqueness of its solution are

guaranteed.

The second integral on the right-hand side is a stochastic integral, and $\mathrm{d}W_s$ is an increment to a Brownian motion. Note that X_t is a random variable for each t.

The stochastic integral can be approximated by:

$$\sum_{j=0}^{N-1} g_{t_j}(W_{t_{j+1}} - W_{t_j}),$$

where we integrate g with respect to the Brownian motion.

The first step to apply a numerical method to (4.4) is to discretize the interval $[0, T]$. Set then $\Delta t = T/L$ for some positive integer L, and $\tau_j = j\Delta t$. Following Higham (2001), we shall denote the numerical approximation X_{τ_j} as X_j.

Considering the step τ_j, we integrate (4.2) from τ_{j-1} to τ_j, such that:

$$X_j = X_{j-1} + \int_{\tau_{j-1}}^{\tau_j} f(X_s)\mathrm{d}s + \int_{\tau_{j-1}}^{\tau_j} g(X_s)\mathrm{d}W_s. \tag{4.4}$$

A discretization scheme is important when a stochastic differential equation (SDE) cannot be solved explicitly. It provides a grid from which we will simulate the trajectory of the SDE. The Euler-Maruyama (EM) method approximates (4.4) as:

$$X_j = X_{j-1} + f(X_{j-1})\Delta t + g(X_{j-1})(W_{\tau_j} - W_{\tau_{j-1}}), \quad j = 1, 2, \ldots, L \tag{4.5}$$

Further issues regarding the order of convergence can be found at Higham (2001).

We apply the EM method to the traditional price SDE:

$$\mathrm{d}X_t = \lambda X_t \mathrm{d}t + \mu X_t \mathrm{d}W_t, \quad X_0 = 0,$$

whose solution is given by:

$$X_t = X_0 \exp\left[\left(\lambda - \frac{1}{2}\mu^2\right)t + \mu W_t\right], \tag{4.6}$$

were λ and μ are real constants.

In figure 4.1 we show the results of the discretization of (4.6) computed using a discretized Brownian path over [0,1], $\delta t = 2^{-8}$, $\lambda = 2$, $\mu = 1$ and 10,000 observations. The Matlab codes for this approximation can be found in the Appendix. We can see that the results obtained with the EM method

are good, and that the Brownian path is well approximated.

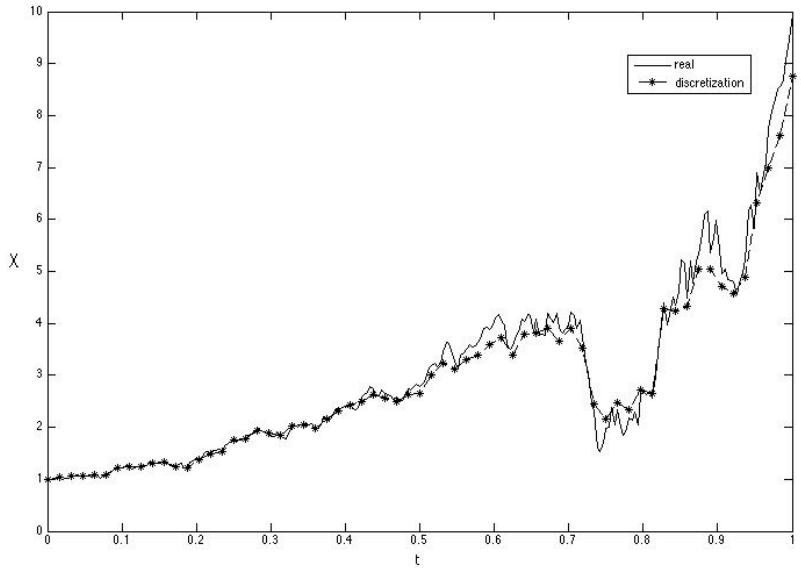


Figure 4.1: Discretization using the Euler-Maruyama method

Although we can find closed-form solutions to the integrals in our model, the discretization is useful for representing the model equations as a series of discrete time integrals and calculating its likelihood functions.

We will measure time in years, and the model is intended to describe prices in a daily basis. Therefore $\Delta t = 1/252$, and every observed series is $S_{n,\Delta t}$, $0 \leq n \leq N$, where N is the number of days in the data sample.

The discrete time versions of equations (3.21), (3.22) and (3.3) can be written as:[1]

$$X_j = \kappa\theta\Delta t + (1 - k\Delta t)X_{j-1} + \sigma\epsilon_j^1, \tag{4.7}$$

$$Y_j = \mu\Delta t + Y_{j-1} + \eta\epsilon_j^2, \tag{4.8}$$

$$\text{and} \quad S_j = \exp(X_j + Y_j), \tag{4.9}$$

where

$$\epsilon_j^1 = (W_{\tau_j}^X - W_{\tau_{j-1}}^X) \quad \text{and} \quad \epsilon_j^2 = (W_{\tau_j}^Y - W_{\tau_{j-1}}^Y).$$

It is easy to see that equation (4.7) is equivalent to a first order autoregressive, AR(1), process and equation (4.8) is a random walk plus drift (RW).

An AR(1) is a stochastic process which satisfies the following difference

[1]Note that using the discretization scheme on equation (4.7) or (3.4) yields the same results for small Δt and $T \leq \mathcal{O}(\Delta t^{-1})$. We chose to work with (4.7) for notational simplicity.

equation:

$$X_{j+1} = \phi_0 + \phi_1 X_j + u_j$$

where u_j is a white noise sequence with zero mean, variance σ^2 and $\mathrm{E}(u_j u_s) = 0$ for $u \neq s$. Hamilton (1994) shows that when $|\phi_1| < 1$ the effects of the u's for X die out over time, so there is a covariance-stationary process for X_t. The stationarity property of the AR(1) is what makes it suitable for representing a mean-reverting process in discrete time.

However, as remarked by Hamilton (1994), when $|\phi_1| \geq 1$, the effects of u's for X accumulate rather than die out over time, and there is no covariance-stationary process for X_t with finite variance that satisfies the AR(1) equation.

When $\phi_1 = 1$, the process is defined as a random walk, and when considering a constant term, as a random walk plus drift. The latter can be written as:

$$Y_{j+1} = \delta + Y_j + v_j.$$

Using the general form of the AR(1) and of the RW, we can parametrize these processes for our model as follows:

For the discrete-time version of the AR(1),

$$X_{j+1} = \theta(1 - e^{\kappa \Delta t}) + e^{-\kappa \Delta t} X_j + \sigma e^{-k\Delta t}(W^*_{j+1} - W^*_j),$$

we have the following parameters

$$\phi_0 = \theta(1 - e^{\kappa \Delta T}),$$

$$\phi_1 = e^{-\kappa \Delta T},$$

$$\mathrm{Var}(u_n) = \sqrt{\frac{\sigma_X^2 2\kappa}{(1 - e^{-2\kappa \Delta t})}},$$

where σ_X^2 is the variance of X_t given by equation (3.6).

From these relations, we have that:

$$\theta = -\frac{\lambda_X}{\kappa} \tag{4.10}$$

Note that κ^* stands for $-\kappa$ in equation (3.18).

We discretize the Random Walk as

$$Y_j = \delta + Y_{j-1} + \epsilon_j^Y, \tag{4.11}$$

where

$$\delta = (\mu - \lambda_Y)\Delta t \tag{4.12}$$

and

$$\epsilon_j^Y = v_j \eta^2 \Delta t. \tag{4.13}$$

The parameters of the RW can be estimated by Maximum Likelihood, and the functional form of its estimators will be derived in the next section. It is important to note that in our methodology we do not need to estimate λ_X and λ_Y. They are obtained by the relations shown in equations 4.12 and 4.10, using the estimated parameters.

4.2 Overall Strategy

In this section, we explain the overall strategy for estimating the model's parameters. We start by noting that we have a data set of futures prices which represent two processes (X_t and Y_t) which are not observable. Our aim is to recover these processes from the observable futures term structure.

One possible approach would be to represent our model in a state-space form and estimate its parameters using the Kalman Filter. That is the technique used by Schwartz (1997), Schwartz and Smith (2000) and Barlow, Gusev and Lai (2004).

Another possible method to recover spot prices from futures data, is the use of nonlinear least squares as suggested by Hikspoors and Jaimungal (2007). In this work we use a two-step approach similar to this one.

Figure 4.2 explains the overall strategy of our estimation algorithm, which we now detail.

4.2.1 Hidden Processes Calibration

As addressed earlier, the main difficulty in implementing commodity models lies in the non-observance of spot prices. Our aim is to obtain an algorithm that recovers these hidden processes in a two-step approach using Non Linear Least Squares and Maximum Likelihood estimation.

Using a given set of parameters, we shall first implement the Non Linear Least Squares method in order to find candidates for the hidden processes. We apply this method to the futures data matrix and solve the following optimization problem for X.

Our aim is to find X_i, where $\mathbf{X} = \text{argmin} \quad g(\mathbf{X}, \mathbf{Y}, \Theta)$, where $\mathbf{X} = (X_i)$, $\mathbf{Y} = (Y_i)$ for $i = 1, 2, \ldots, N$, and Θ is the parameter vector. So we have that:

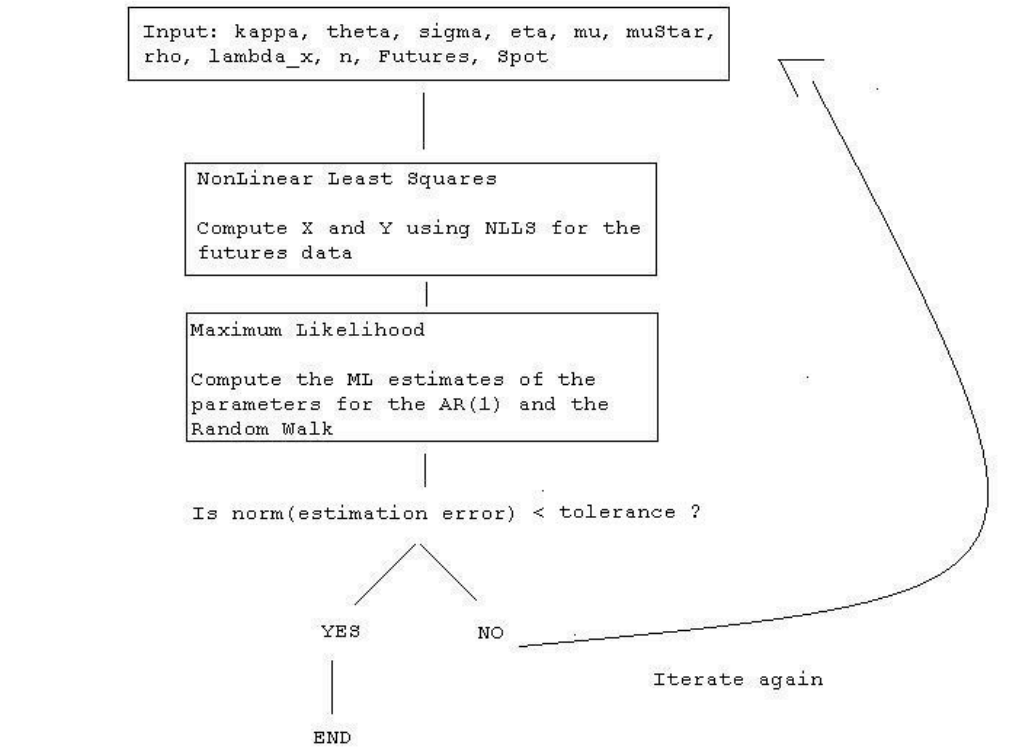


Figure 4.2: Estimation strategy

$$\begin{aligned} \min \quad & g(X_i, Y_i, \Theta) := \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} (Fut_{T_j,t_i} - \bar{F})^2, \\ & \text{subject to} \quad Y_i = \ln(S_i) - X_i, \end{aligned} \tag{4.14}$$

where $\bar{F}$ is the matrix of logarithms of the observed prices and$Fut_{T,t}$ is the logarithm of the futures price. That is:

$$\ln(F_{T,t}) = X_t \mathrm{e}^{-\kappa(T_j - t_i)} + Y_t + A(T_j - t_i), \tag{4.15}$$

and

$$\begin{aligned} A(T_j - t_i) =& \mu(T_j - t_i) \\ & + \frac{1}{2} \left[\frac{\sigma^2}{2k} \left(1 - \mathrm{e}^{-2k(T_j - t_i)} \right) + \eta^2 (T_j - t_i) + \frac{2\sigma\eta\rho}{k} \left(1 - \mathrm{e}^{-k(T_j - t_i)} \right) \right]. \end{aligned}$$

Upon imposing $Y_i = \ln(S_i) - X_i$ into (4.15), and differentiating for X_i, we have that:

$$\frac{\partial g}{\partial X_i} = \sum_{j=2}^{N} \left[X_i \mathrm{e}^{-k(T_j - t_i)} + (S_i - X_i) + A(T_j - t_i) - \bar{F} \right] \left(\mathrm{e}^{-k(T_j - t_i)} - 1 \right).$$

Setting the last expression equal to zero yields

$$\sum_{j=2}^{N} \left(\mathrm{e}^{-k(T_j - t_i)} - 1 \right)^2 X_i = \sum_{j=2}^{N} \left(\bar{F} - A(T_j - t_i) - S_i \right) \left(\mathrm{e}^{-k(T_j - t_i)} - 1 \right).$$

Hence,

$$X_i = \frac{\sum_{j=2}^{N} \left(\bar{F} - A(T_j - t_i) - S_i \right) \left(\mathrm{e}^{-k(T_j - t_i)} - 1 \right)}{\sum_{j=2}^{N} \left(\mathrm{e}^{-k(T_j - t_i)} - 1 \right)^2} \tag{4.16}$$

To recover Y, we only need to use the restriction $Y_i = \ln(S_i) - X_i$, remembering that the shortest maturity futures contract are a proxy for the spot prices.

Next, we estimate the parameters that govern this process X_i using a maximum likelihood approach for the parameter κ. The estimation of κ will be detailed in the next section. The other parameters of X_i are obtained from κ using the change of measure equations shown before.

A similar likelihood approach is done for the parameters' estimations of the random walk process Y_i. We estimate its mean and variance using traditional likelihood estimation methods, which shall be described in the following Section.

4.2.2 Likelihood Functions and Parameter Estimation

We shall address now the model's calibration. We shall start by posing the parameter estimation problem and deriving likelihood functions for this purpose.

Consider a random variable X that is normally distributed and has a vector of parameters $\Theta = (\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance of X. Its probability density function (pdf) is given by:

$$g_\Theta(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2} \right).$$

The likelihood function is defined as

$$L(\Theta; X) = \prod_{i=1}^{N} g_{\Theta}(X_i),$$

where N is the sample size. We can apply a positive monotonic transformation to the likelihood function and write is natural logarithm as

$$\begin{aligned}
\ell(\Theta; X) &= \log L(\Theta, X) \\
&= \sum_{i=1}^{N} \log g_{\Theta}(X_i) \\
&= \sum_{i=1}^{N} \log \left[\frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(X-\mu)^2}{2\sigma^2} \right) \right] \\
&= -\frac{N}{2}\log(2\pi\sigma) - \sum_{i=2}^{N} \left(-\frac{(X-\mu)^2}{2\sigma^2} \right).
\end{aligned}$$

The Maximum Likelihood (ML) method chooses the value $\theta = \hat{\theta}$ that maximizes $\ell(\Theta; X)$.

In our model, X is an Orstein-Uhlenbeck (OU) process with conditional mean $e^{-k\Delta t}X_t$ and conditional variance $(1 - e^{-2k\Delta t})\frac{\sigma^2}{2k}$, where Δt is the time increment. However, in order to allow a more general estimation, we derive the likelihood estimators considering a process that reverts to θ.[2] Plugging these expressions for mean and variance into the likelihood function yields the following likelihood function for the OU process:

$$f(X_{t_i}, \theta, \kappa, \sigma) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta t})}} \exp\left[-\frac{(X_{t_i} - \theta - (X_{t_{i-1}} - \theta)e^{-\kappa\Delta t})^2}{\frac{\sigma^2}{\kappa}(1 - e^{-2\kappa\Delta t})} \right].$$

Its log-likelihood is then given by:

$$\begin{aligned}
\ell(X, \theta, \kappa, \sigma) = &-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\left(\frac{\sigma^2}{2\kappa}\right) \\
&- \frac{1}{2}\sum_{i=2}^{N} \log(1 - e^{-2\kappa\Delta t}) - \frac{\kappa}{\sigma^2}\sum_{i=2}^{N} \frac{(X_{t_i} - \theta - (X_{t_i} - \theta)e^{-\kappa\Delta t})^2}{2\frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta t})}.
\end{aligned}$$

The first step to find the maximum likelihood estimators $\hat{\theta}$,$\hat{\kappa}$ and $\hat{\sigma}$ is to obtain the first order conditions for the maximization problem of the log-likelihood function. Therefore, the estimators must satisfy:

[2] In our estimations we show that θ is in fact very close to zero.

$$\frac{\partial \ell(X,\theta,\kappa,\sigma)}{\partial \theta} = \frac{\partial \ell(X,\theta,\kappa,\sigma)}{\partial \kappa} = \frac{\partial \ell(X,\theta,\kappa,\sigma)}{\partial \sigma} = 0,$$

Deriving the log-likelihood function and setting the gradient equal to zero we have that:

$$\frac{\partial \ell(X,\theta,\kappa,\sigma)}{\partial \theta} = -\frac{2\kappa}{\sigma^2}\sum_{i=2}^{N}\frac{(X_{t_{i-1}} - \theta - (X_{t_i} - \theta)e^{-\kappa\Delta t})(e^{-\kappa\Delta t})}{1 - e^{-2\kappa\Delta t}},$$

which yields

$$\hat{\theta} = \sum_{i=2}^{N}\frac{X_{t_i} - X_{t_i}(e^{-\kappa(T-t)} - 1)}{1 - e^{-2\kappa\Delta t}}(N-1)\frac{1 - e^{-2\kappa\Delta t}}{1 - e^{-\kappa\Delta t}}. \tag{4.17}$$

Also,

$$\frac{\partial \ell(X,\theta,\kappa,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2\kappa}{\sigma^3}\sum_{i=2}^{N}\frac{(X_{t_i} - \theta - (X_{t_{i-1}} - \theta)e^{-\kappa(T-t)})^2}{1 - e^{-2\kappa\Delta t}}.$$

Hence,

$$\hat{\sigma} = \sqrt{\frac{2\kappa}{N}\sum_{i=2}^{N}\frac{(X_{t_i} - \theta - (X_{t_i} - \theta)e^{-\kappa\Delta t})^2}{1 - e^{-2\kappa\Delta t}}}. \tag{4.18}$$

Note that expressions (4.17) and (4.18) show that the estimates for $\hat{\theta}$ and $\hat{\sigma}$ are functions of κ. We define that $\hat{\theta} = f(\hat{\kappa})$ and $\hat{\sigma} = h(\hat{\theta}, \hat{\kappa})$.

There would then be two approaches to obtain maximum likelihood estimates for the parameters. The first implies solving the system given by $f(\hat{\kappa})$, $h(\hat{\theta}, \hat{\kappa})$ and $\dfrac{\partial \ell}{\partial \kappa} = 0$, which would be solved numerically as it does not have a closed form solution. In the second approach we could substitute $\hat{\theta} = f(\hat{\kappa})$ and $\hat{\sigma} = h(\hat{\theta}, \hat{\kappa})$ directly into the likelihood function and maximize it with respect to κ. This would lead to the following problem:

$$\kappa = \text{argmax} V(\kappa),$$

where

$$V(\kappa) = -\frac{n}{2}\log\left[\frac{g(f(\kappa),\kappa)^2}{2\kappa}\right] - \frac{1}{2}\sum_{i=2}^{N}\log\left[1 - e^{-2\kappa(T-t)}\right]. \tag{4.19}$$

Note that $V(\kappa)$ equals minus the log likelihood, so we can minimize $-V(\kappa)$ or

maximize $V(\kappa)$. Note also that $V(\kappa)$ should have a unique maximum.

Solving for $\hat{\kappa}$ we can then obtain $\hat{\theta} = f(\hat{\kappa})$ and $\hat{\sigma} = g(\hat{\theta}, \hat{\kappa})$, using only a one dimensional search and requiring the evaluation of less complex expressions.

Finally, we address the estimation of the RW's parameters. As mentioned before, its estimators for the drift and variance parameters can be obtained by Maximum Likelihood. The general form of the likelihood function can be written as:

$$f_{X_N, X_{N-1}, \ldots, X_1 | X_1}(X_N, X_{N-1}, \ldots, X_2 | X_1; \Theta) = \prod_{i=2}^{T} f_{X_N | X_{N-1}},$$

which for the random walk process takes the form:

$$\log f_{Y_N, Y_{N-1}, \ldots, Y_1 | Y_1} = -\frac{N-1}{2}\log(2\pi) - \frac{N-1}{2}\log(\sigma^2) - \sum_{i=2}^{T}\left[\frac{(Y_{t_i} - \delta - Y_{t_{i-1}})^2}{2\sigma^2}\right].$$

Applying the first order conditions and setting the expressions obtained equal to zero, we have that

$$\frac{\partial \log(L)(\mathbf{y}|\mathbf{y_1}, \theta)}{\partial \delta} = -\sum_{i=2}^{N}\left[\frac{(Y_{t_i} - \delta - Y_{t_{i-1}})}{\sigma^2}\right](-1) = 0$$

$$\frac{1}{\sigma^2}\sum_{i=2}^{N} Y_{t_i} - Y_{t_{i-1}} = \frac{\delta}{\sigma^2}(N-1).$$

Thus,

$$\hat{\delta} = \frac{1}{N-1}\sum_{i=2}^{N}(Y_{t_i} - Y_{t_{i-1}}). \tag{4.20}$$

$$\frac{\partial \log(L)(\mathbf{y}|\mathbf{y_1}, \theta)}{\partial \sigma^2} = -\frac{N-1}{2\sigma^2} - \sum_{i=2}^{N}\left[\frac{(Y_{t_i} - \delta - Y_{t_{i-1}})}{\sigma^2}\right]\left[-\frac{(Y_{t_i} - \delta - Y_{t_{i-1}})}{2\sigma^2}\right] = 0$$

$$(N-1)\sigma^2 = \sum_{i=2}^{N}(Y_{t_i} - \delta - Y_{t_{i-1}})^2.$$

Thus,

$$\hat{\sigma^2} = \frac{\sum_{i=2}^{N}(Y_{t_i} - \delta - Y_{t_{i-1}})^2}{N-1} \tag{4.21}$$

From the estimates in (4.20) and (4.21), we recover the parameters μ and η using equations (4.12) and (4.13).

	κ	σ	θ	λ_X	η	μ	λ_Y	ρ
Initial Guess	1.5	0.28	0.1	0.15	0.14	-0.01	-0.03	0.3

Table 4.1: Initial Guess for model parameters

4.3 Numerical Implementation

In this section we address the numerical procedures used in order to validate our methodology.

First we generate synthetic autoregressive and random walk processes, representing X_t and Y_t. We use Matlab's normal random number `randn` generator for this purpose. Each call to `randn` produces an independent "pseudorandom" number from the $N(0,1)$ distribution.

Afterwards we create a matrix of log-futures prices with 24 maturities and use our nonlinear least squares and maximum likelihood method to recover the parameter vector from this matrix. The efficiency of our method will be measured by its ability to recover the original parameters that generated the log-futures matrix, and is addressed in the next section.

Following the common procedure in the literature, we keep the first maturity contract for a proxy of the spot prices $\ln(S)$.

Using the new futures matrix F we start the parameter estimation procedures. First we generate X using the first order conditions of the nonlinear least squares problem stated in equation (4.16). We then recover Y using the restriction

$$Y_i = \ln(S_i) - X_i.$$

In the second part of the estimation we recover the parameters governing the two processes. Due to the complex nature of the optimization problem involving the mean reverting process, we use different approaches for each of the processes X and Y. For the parameters' estimation of process X we use the optimization problem described by equation (4.19).

We input the initial values from table 4.1 based on Schwartz and Smith (2000), and define the error tolerance to 10^{-6}.[3] If the difference between the estimated parameters in two consecutive round of estimations is smaller than the 10^{-6} tolerance, the estimation finishes and we compute the estimation error, which is the difference between the initial guesses and the estimated values. Otherwise, there is another iteration with an estimation round performed.

During the iteration process, we solve the optimization problem in (4.19)

[3]We discuss why we use this value in the next Section.

using Matlab's function "fminbnd". This command minimizes a single-variable bounded nonlinear function in a given interval. Our inputs are the functional form in (4.19) and an upper and lower bound for X. The bounds we initially considered for X are 0.001 and 5.

Our aim is to estimate the parameters κ, σ and θ for the mean reverting short-term process and η and μ for the long term geometric Brownian motion.

The risk neutral parameters λ_X and λ_Y are recovered from θ and κ by equations

$$\lambda_X = -\theta\kappa \tag{4.22}$$

and

$$\lambda_Y = \mu - \mu^* \tag{4.23}$$

where μ^* is the United State's Treasuries risk free interest rate.

The random walk parameters μ and η are recovered by traditional maximum likelihood estimates. The estimators for the drift and error variance of the random walk are given by equations (4.20) and (4.21). We have then that:

$$\mu = \lambda_Y + \frac{\delta}{\Delta t},$$

and

$$\eta = \sqrt{\frac{\sigma_Y^2}{\Delta t}}.$$

4.4 Estimation Results

As mentioned before, the first results from our algorithm are obtained with the synthetic data. We generate a matrix of futures prices F using the initial guesses in Table 4.1. Then, we apply our algorithm to estimate the parameters of X and Y, and compare the estimate results with the true values that generated the data.

We define the estimation error as the difference between the known parameter value shown in Table 4.1 and the ones obtained from the algorithm. If the estimation error is small and the estimates robust to different initial guesses, it is natural to expect that the algorithm will converge when applied to real data.

We display parameters estimates, estimation and iteration errors for samples of 1,000, 3,000, 10,000, 30,000 and 65,000 observations.

From the results obtained, we can see that the parameters' estimates for κ, σ and η are the most accurate ones. However, the estimates for θ and μ are

$N = 1,000$

	Toler.	κ	θ	σ	η	μ	n. iter
True Values		1.5	0.1	0.28	0.14	-0.01	
Param. Est.	10^{-6}	1.8670	-0.0720	0.2700	0.1415	0.0003	4
Estim. Error		-0.3670	-0.0720	0.0100	-0.0015	-0.0103	

$N = 3,000$

	Toler.	κ	θ	σ	η	μ	n. iter
True Values		1.5	0.1	0.28	0.14	-0.01	
Param. Est.	10^{-6}	1.6188	0.0595	0.2709	0.2130	0.0005	4
Estim. Error		-0.1188	-0.0595	0.0091	-0.0730	-0.0105	

$N = 10,000$

	Toler.	κ	θ	σ	η	μ	n. iter
True Values		1.5	0.1	0.28	0.14	-0.01	
Param. Est.	10^{-6}	1.6695	-0.0791	0.2719	0.1472	-0.0002	4
Estim. Error		-0.1695	0.1791	0.0081	-0.0072	-0.0102	

$N = 30,000$

	Toler.	κ	θ	σ	η	μ	n. iter
True Values		1.5	0.1	0.28	0.14	-0.01	
Param. Est.	10^{-6}	1.5761	-0.0489	0.2779	0.1362	-0.0002	4
Estim. Error		-0.0761	0.1489	0.0021	0.0038	-0.0100	

$N = 65,000$

	Toler.	κ	θ	σ	η	μ	n. iter
True Values		1.5	0.1	0.28	0.14	-0.01	
Param. Est.	10^{-6}	1.4987	-0.0771	0.2822	0.142	0.0000	5
Estim. Error		0.0013	0.1771	-0.0022	-0.0020	-0.01	

Table 4.2: Estimate Results - 1,000, 3,000, 10,000, 30,000 and 65,000 observations

Param. Est. stands for the estimated parameters, n. iter is the number of iterations, Estim. Error is the estimation error, and Toler. is the tolerance.

also very good. The greater gains in error reduction appear when increasing the sample size from 10,000 to 30,000 observations for θ, σ and η, while the best results for κ are obtained for 65,000 observations. It is important to note that even with small sample sizes (1,000 and 3,000) observations, the algorithm was able to recover the original parameters with reasonable estimation errors.

When we increase the sample size, the estimation errors decrease and we recover the parameters more accurately. Figures 4.3, 4.4, 4.5, 4.6 and 4.7 show the behavior of absolute value of the estimation error for κ, θ, σ, η and μ. The plots were obtained creating by simulating the processes X and Y with 65,000 observations, and evaluating the parameter estimates and their errors in estimation steps of size 500. So, we plot 129 different observed errors for each of the parameters.

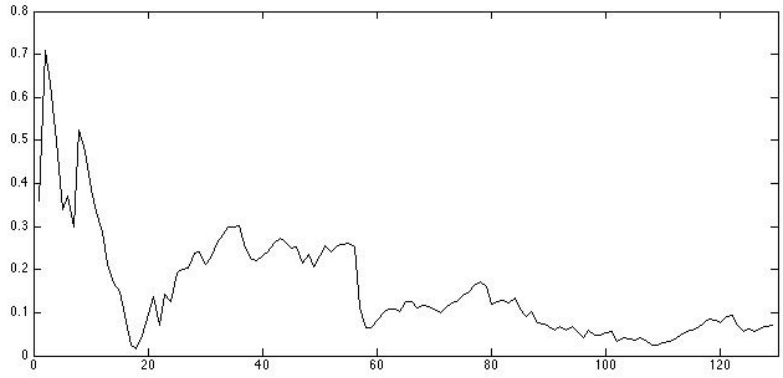


Figure 4.3: Estimation error behavior - κ

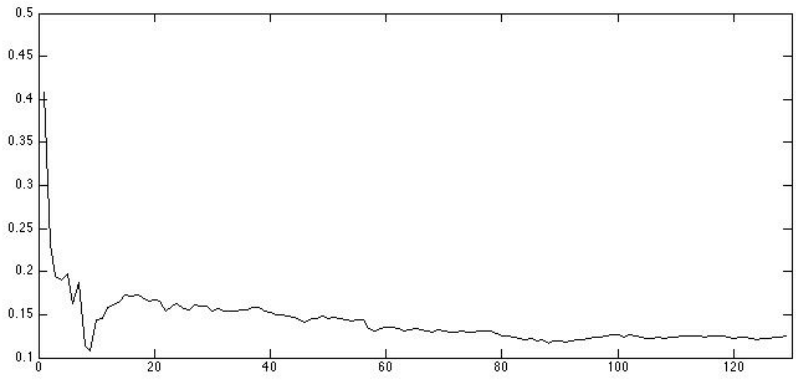


Figure 4.4: Estimation error behavior - θ

Note from figure 4.3 that κ has the most erratic error convergence pattern. There is a sharp reduction in the absolute error around a sample size of 10,500 followed by a region of quite higher errors until a sample size of 60,000, when the absolute error decreases again and remains in a lower level.

The "U" behavior present in κ is also observed for σ at a sample size of 10,500, but with less intensity. For θ, η and μ, it occurs around a sample size of 550.

This phenomenon may be due to the roundoff error in computations. As

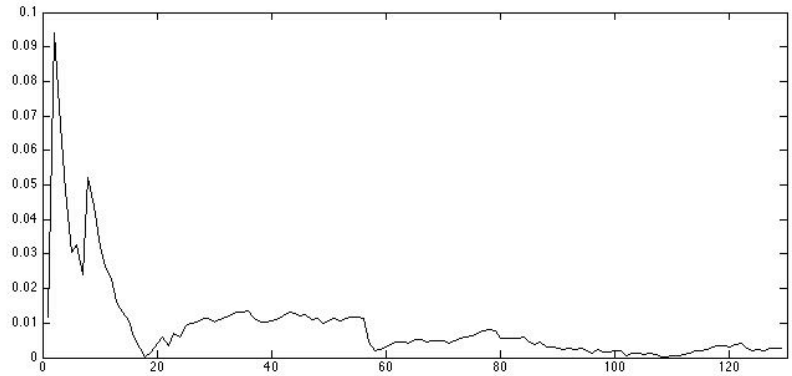


Figure 4.5: Estimation error behavior - σ

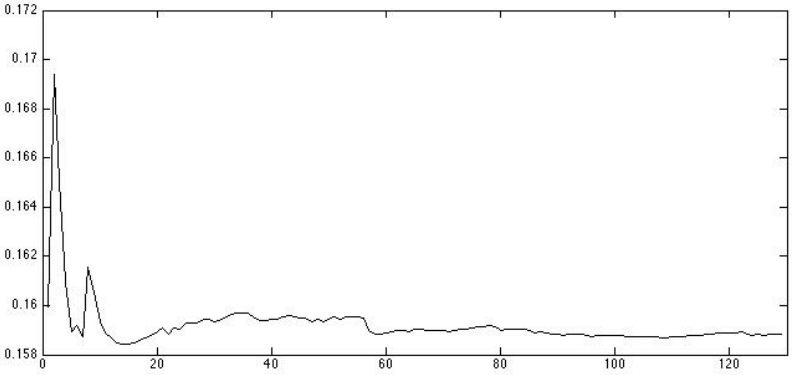


Figure 4.6: Estimation error behavior - η

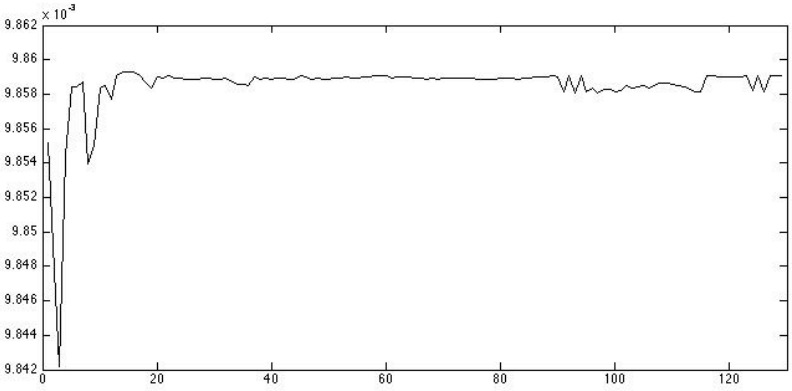


Figure 4.7: Estimation error behavior - μ

mentioned by Press *et al* (1992), computers store numbers not with infinite precision, but rather in some approximation. When data is stored in a floating point (also called "real") format, the arithmetic among numbers is not exact.

Therefore, when performing calculation with the floating numbers, we end up adding a fractional error of at least ϵ_m.[4]

Roundoff errors accumulate with increasing amount of calculation. What we can see in the absolute error plot, is a tradeoff between the error and the number of observations added to the sample. For some parameters, the gain of adding more data to the sample (from 10,500 to 30,500 observations) is offset by the additional roundoff error introduced. The additional data only proves to be efficient to reduce the estimation error instead of increasing it, after a sample size of 30,500.

4.4.1 Choosing the best tolerance

We use a tolerance of 10^{-6} in the estimation process. This value is the optimal stop criteria for the algorithm, based on tests using different initial guesses. When two estimation steps are smaller than this tolerance, we say that the algorithm has converged, and we call the difference between these steps of "Iteration Error".

We have tested several tolerances using different sets of true values for generating the processes. The tolerances ranged from 10^{-2} to 10^{-12} in steps of 10^{-2}. For the sake of parsimony we display the results for one of these tests for only two sets of initial guesses. Set 1: $\kappa = 0.005$, $\theta = 0.002$, $\sigma = 0.005$, $\eta = 0.0004$ and $\mu = 0.0001$ and set 2: $\kappa = 2$, $\theta = 0.5$, $\sigma = 1.2$, $\eta = 1.4$ and $\mu = 0.5$

Table 4.3 illustrates the kind of tests performed. For 10,000 observations our algorithm converges within seven iterations, up to a tolerance of 10^{-8} for the iteration error. While with 30,000 observations we obtain convergence up to a tolerance of 10^{-6}.

It is not worth-while choosing a tolerance lower than 10^{-6} because we can still improve the iteration errors, and when using a tolerance higher than 10^{-6} the convergence (when obtained) requires a heavy computational burden.

4.5 Numerical Tests and Validation Exercises

We created a routine of tests in order to verify if our estimation procedure was robust to other initial guesses apart from the one we used based on Schwartz and Smith (2000).

In order to verify this, we generate a data set with a fixed parameter vector and try to recover this vector with our estimation method, but using starting

[4]The machine accuracy, or simply ϵ_m, is the fractional accuracy to which floating-point numbers are represented, corresponding to a change of one in the least significant bits of the mantissa.

$N = 10,000$

	Toler.	κ	θ	σ	η	μ	Iter.
True Val.		0.005	0.002	0.005	0.0004	0.0001	
Estimates		0.2081	-0.0067	0.0001	0.0050	-0.0001	
Est Error		-0.2031	0.0087	0.0049	-0.0046	0.0002	
It Error	1e-02	-6.10e-11	7.15e-04	3.62e-14	2.89e-12	-4.89e-07	3
It Error	1e-04	-6.12e-11	-2.73e-06	3.49e-14	1.75e-13	2.60e-09	4
It Error	1e-06	-8.23e-13	1.45e-08	3.50e-14	-8.06e-13	-2.55e-09	5
It Error	1e-08	-1.71e-13	-2.44e-10	1.28e-16	2.19e-13	7.25e-10	7
It Error	1e-10	-6.10e-11	6.88e-09	-2.46e-16	-6.35e-13	-2.10e-09	100

$N = 30,000$

	Toler.	κ	θ	σ	η	μ	Iter.
True Val.		2	0.5	1.2	1.4	0.5	
Estimates		1.9241	1.9536	1.2139	2.3386	0.0014	
Est Error		0.0759	-1.4536	-0.0139	-0.9386	0.4986	
It Error	1e-02	-9.4e-03	2.43e-09	-4.19e-12	2.85e-06	8.198e-10	3
It Error	1e-04	-3.69e-09	-3.66e-07	3.72e-10	-3.09e-11	-6.51e-08	4
It Error	1e-06	-3.69e-09	-3.66e-07	3.72e-10	-3.09e-11	-6.51e-08	4
It Error	1e-08	5.53e-09	2.37e-07	3.94e-10	-3.37e-11	-3.34e-07	100

Table 4.3: Tolerance test for 10,000 and 30,000 observations
True Val. states for the true parameter values. Est Errors is the estimation error, and It Error is the iteration error.

values significantly different from the ones used to generate the data.

There are two different tests performed. First, we disturb all parameters in the same direction, creating a single gradient for all of them. Next, we generate random disturbances with a uniform distribution in order to allow each parameter to head to a different direction.

We created processes using 10,000, 30,000 and 65,000 observations that can be seen in Figures 4.8, 4.9 and 4.10.

In figures 4.9 and 4.10 we can see a better picture of the behavior of the price processes X and Y. Note that the short-term process X has a stationary behavior and can be seen as random shocks around zero, as stated in equation (4.7).

The long term process Y behaves as expected for a geometric Brownian motion. It has an erratic behavior that represents the structural shocks, such as improving technology or discovery of a commodity.

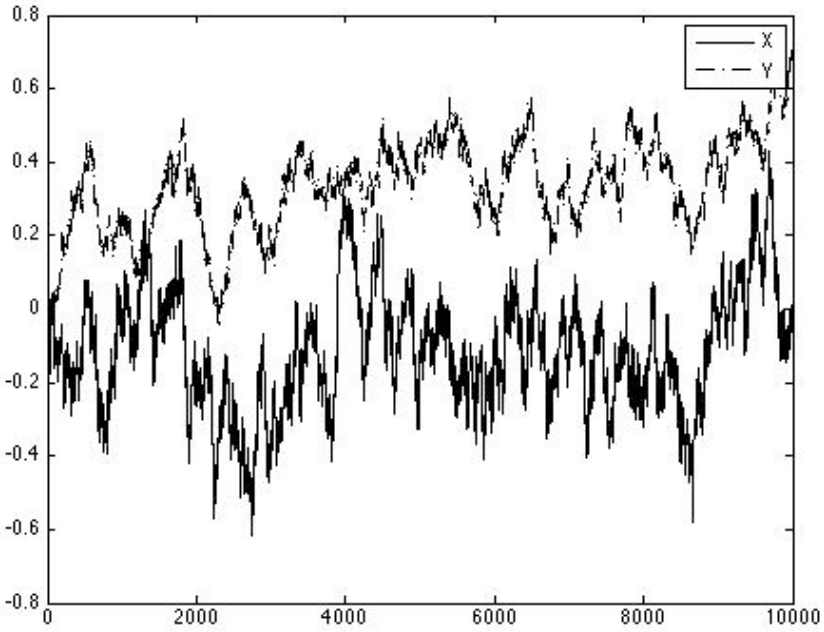


Figure 4.8: X and Y processes - 10,000 observations

For test 1 we consider perturbations of 1, 5, 10, 15, 20, 50, 250, 300, 600 and 1000% of the original guesses. All data was generated with 1,000, 10,000 and 30,000 observations.

It is important to note that the algorithm converges to the **same estimated parameters** for all initial guesses considered.

For test 2 we generate 150 different random shocks uniformly distributed in the interval [0.8, 1.2], allowing each parameter to go to a different direction, and then perturb each of them in the same amount of 1, 5, 10, 15, 20, 50, 250, 300, 600 and 1000% of the original guesses.

The results of this experiment can be seen in table 4.5. We disclose the iteration errors for one random shock for each of the initial disturbances. To see how the error behaves for all of the shocks we plot the iterative error for each of the parameters against κ in table 4.6. The plots for the remaining

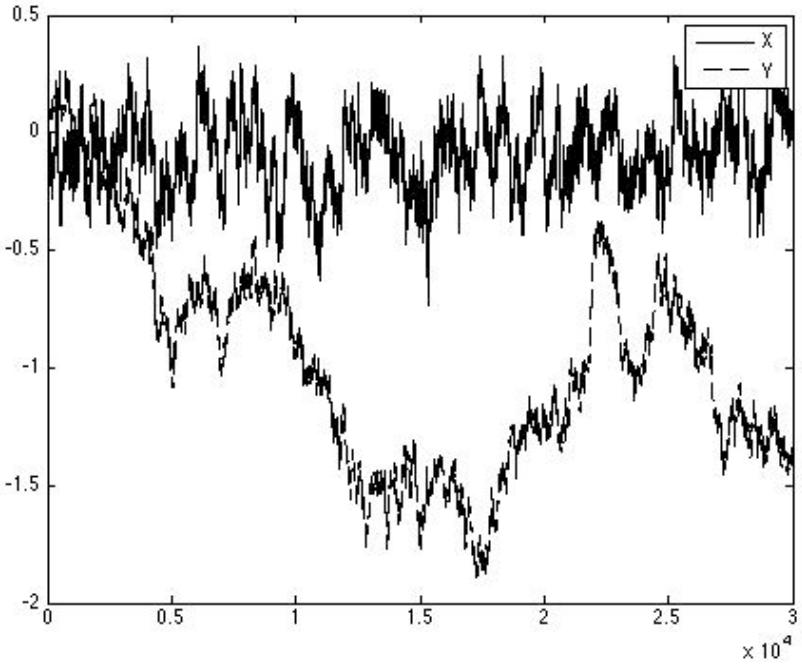


Figure 4.9: X and Y processes - 30,000 observations

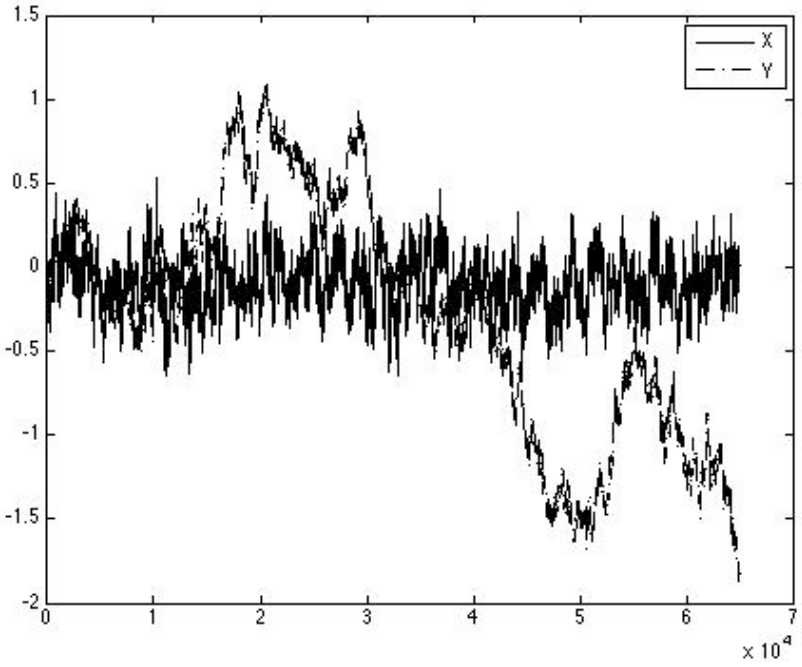


Figure 4.10: X and Y processes - 65,000 observations

parameters are shown in Appendix A.

As we can see from tables 4.4 and 4.5, the iteration error is very small for

Perturbation	κ	θ	σ	η	μ	n. iter
$N = 1.000$						
1%	5.9325e-10	-2.2675e-11	2.8907e-13	4.9444e-13	-1.0726e-10	5
5%	5.9309e-10	-1.1188e-10	6.0988e-12	-9.2545e-13	-2.3135e-10	5
10%	6.9500e-13	-1.9989e-11	2.8533e-14	-2.0134e-13	-5.5791e-12	5
15%	1.0076e-12	-1.7620e-10	3.3112e-14	2.3678e-13	9.3100e-11	5
20%	5.8985e-10	-1.8433e-10	2.3007e-13	-1.0090e-12	2.0484e-10	5
50%	-1.1822e-09	-4.1515e-10	-6.4102e-12	1.6834e-12	-7.3711e-12	4
200%	5.4894e-12	6.0609e-09	-2.9976e-15	-3.8219e-14	-3.0179e-11	4
250%	5.9325e-10	5.7230e-09	3.1433e-12	1.8438e-13	-7.7243e-11	4
300%	-6.0694e-10	5.0850e-09	-3.2226e-12	1.1498e-12	-7.0817e-11	4
600%	-4.7948e-12	-3.8267e-10	2.1094e-15	-1.2157e-12	1.3567e-10	5
10000%	5.6457e-12	-5.5882e-10	-1.8208e-14	-9.0902e-13	1.5562e-10	5
$N = 10,000$						
1%	2.3670e-12	-4.9847e-07	6.7991e-11	-1.2335e-11	1.2601e-09	4
5%	1.0919e-09	-3.9759e-07	2.3135e-11	-4.1050e-12	-7.9858e-10	4
10%	1.8211e-09	-2.9591e-07	-2.1688e-11	4.1137e-12	7.2994e-10	4
15%	-1.0936e-09	-2.1491e-07	2.2051e-11	-4.1084e-12	-2.3953e-10	4
20%	1.8214e-09	-1.4381e-07	3.4997e-11	-6.1812e-12	4.8950e-10	4
50%	1.0927e-09	8.7733e-08	2.3242e-11	-4.1188e-12	-7.2493e-10	4
200%	2.8639e-12	1.9378e-07	-1.1271e-11	2.0312e-12	-2.5109e-09	4
250%	-2.9439e-12	2.1500e-07	-1.1415e-11	2.0632e-12	-1.0328e-09	4
300%	1.4544e-09	2.1790e-07	3.4747e-11	-6.1748e-12	-1.2711e-10	4
600%	-3.6200e-10	2.6043e-07	3.3869e-11	-6.1890e-12	-3.8770e-11	4
10000%	-3.6482e-10	3.3302e-07	-2.3725e-13	5.4678e-15	-1.1194e-10	4
$N = 30,000$						
1%	-2.8000e-09	-4.1206e-10	-2.2479e-10	1.0572e-12	-2.1716e-10	4
5%	3.0479e-09	-8.6024e-10	2.8132e-10	-1.3250e-12	1.0485e-10	4
10%	-2.4639e-10	-1.2308e-09	3.7967e-11	-1.8066e-13	-2.6056e-10	4
15%	1.2602e-09	-1.0083e-09	2.6070e-10	-1.2311e-12	-7.3939e-11	4
20%	2.5459e-09	-1.4718e-09	-3.5716e-11	1.7633e-13	1.4604e-10	4
50%	-2.8001e-09	-1.9617e-09	-2.4352e-10	1.1466e-12	2.5638e-10	4
200%	2.2899e-09	-2.0578e-09	1.8794e-10	-8.8413e-13	-1.5517e-10	4
250%	5.1538e-10	-1.8375e-09	-5.4875e-11	2.6132e-13	-2.0853e-10	4
300%	-3.3174e-09	-2.0758e-09	-1.6993e-10	7.9647e-13	-2.6000e-11	4
600%	31e-09	-2.4154e-09	-2.0371e-10	9.7028e-13	1.6014e-10	4
1000%	1.7844e-09	-2.5120e-09	3.8336e-11	-1.7747e-13	3.3644e-10	4

Table 4.4: Iteration errors - Perturbation on initial parameter guesses

all parameters and perturbations. The smaller error has the order of 10^{-12} and the largest of 10^{-7}, showing that our algorithm is stable.

The best results are obtained for the parameters σ, μ and η for both perturbation tests. However, they do not differ significantly from the results for κ and θ.

The relative superiority of the error order for η and μ over the others is due to the nature of their optimization problem. The random walk estimators have closed form solutions and can be easily obtained, while the estimates for κ were obtained after an optimization routine, as there was no closed form solution to the first order conditions in problem (4.19).

Perturbation	κ	θ	σ	η	μ	n. iter
$N = 1.000$						
1%	1.5588e-13	7.6841e-07	-5.8137e-12	-1.1837e-12	-1.3874e-10	4
5%	-1.1814e-09	-2.8537e-10	-9.3051e-12	1.9404e-12	2.0020e-10	5
10%	1.1448e-11	-8.6391e-11	-2.8537e-12	2.5069e-13	3.6887e-11	5
15%	-5.9224e-10	-7.4801e-11	-3.1463e-12	1.6671e-12	-9.9810e-11	5
20%	6.0129e-10	-6.6498e-11	6.1151e-12	-1.7255e-12	-4.4398e-11	5
50%	-2.0342e-11	-1.0245e-08	-2.9691e-12	1.0871e-12	6.7044e-11	4
200%	-1.3846e-11	6.1162e-09	-1.8846e-14	6.5919e-13	-1.1103e-10	4
250%	5.8590e-10	6.1349e-09	2.6507e-13	5.2497e-13	-1.3818e-10	4
300%	1.7630e-09	5.8865e-09	9.5484e-12	-1.8963e-12	-1.6861e-10	4
600%	1.5588e-13	-6.1339e-10	-2.8658e-12	2.6648e-13	1.4188e-10	5
1000%	1.6573e-12	5.3872e-10	3.2687e-12	-1.6723e-12	2.1287e-10	5
$N = 10,000$						
1%	-3.5824e-09	-5.0543e-08	-1.5187e-10	4.2077e-12	4.5406e-10	4
5%	-1.0745e-08	2.3059e-07	-2.3130e-10	6.3195e-12	-1.0899e-09	4
10%	3.5820e-09	5.2839e-07	4.5236e-10	-1.2645e-11	-8.5655e-10	4
15%	4.7738e-09	7.8015e-07	2.6931e-12	2.7756e-17	-9.7599e-11	4
20%	-4.7730e-09	9.9379e-07	-7.7797e-11	2.0987e-12	-1.9439e-09	4
50%	-1.3131e-08	2.1060e-10	-1.5730e-10	4.1981e-12	-2.0405e-09	4
200%	4.7769e-09	1.2537e-09	-4.4745e-10	1.2631e-11	-6.1636e-10	4
250%	-1.1942e-09	-4.3383e-09	-4.5071e-10	1.2635e-11	6.3791e-10	4
300%	2.3824e-09	-3.8130e-09	-1.4914e-10	4.2277e-12	4.8304e-10	4
600%	-7.1585e-09	2.0419e-10	-5.2886e-10	1.4724e-11	-1.6384e-09	4
1000%	-1.4784e-12	-2.8677e-09	-2.2540e-10	6.3286e-12	4.3748e-10	4
$N = 30,000$						
1%	-1.5112e-09	1.3035e-07	8.6574e-12	7.7210e-13	-3.6861e-10	4
5%	-1.9736e-10	1.4043e-07	-1.6828e-10	-1.3635e-11	5.2711e-10	4
10%	2.5352e-09	1.5255e-07	1.1496e-10	9.1974e-12	-2.2781e-10	4
15%	6.0871e-10	1.6578e-07	1.1424e-10	9.2376e-12	-3.1179e-10	4
20%	-1.1224e-09	1.6928e-07	-4.7910e-11	-3.8302e-12	2.1726e-10	4
50%	8.1720e-10	1.9788e-07	-1.3084e-10	-1.0658e-11	9.2033e-11	5
200%	1.2249e-09	2.1444e-07	4.8476e-11	3.8727e-12	5.3027e-11	5
250%	-5.1551e-10	2.2030e-07	-5.8102e-11	-4.6928e-12	-2.9875e-10	5
300%	2.4318e-09	2.2245e-07	1.4267e-10	1.1454e-11	-2.5337e-10	5
600%	7.1693e-10	2.2161e-07	-1.6890e-10	-1.3737e-11	-1.0010e-10	5
1000%	2.4198e-09	1.9912e-07	6.7814e-11	5.3883e-12	2.3257e-10	5

Table 4.5: Iteration errors - Random perturbation on initial parameter guesses

The gain from increasing the number of observations from 1,000 to 10,000 and to 30,000 is not very significant in both tests, what again confirms the stability of the algorithm. When considering the same perturbation size to all elements of the parameter vector (test 1), we had some gain in parameters θ and η. The error in θ decreased from an average order of 10^{-7} to 10^{-9}, while the one from η decreased from 10^{-12} to 10^{-13} in average.

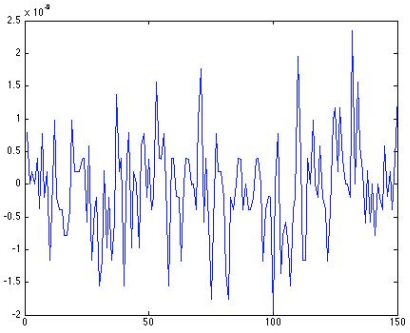
The gain in the iteration error of κ and σ is not very obvious. In some cases there was a decrease in the order of the error, such as in the 1000% perturbation in σ in test 1, from 10^{-15} to 10^{-10}. There was no significant change in the iteration error for μ in either tests.

In test 2, for the random perturbations, the results do not differ much. We have the same blur results for κ when passing from 10,000 to 30,000 observations and when increasing the perturbation size.

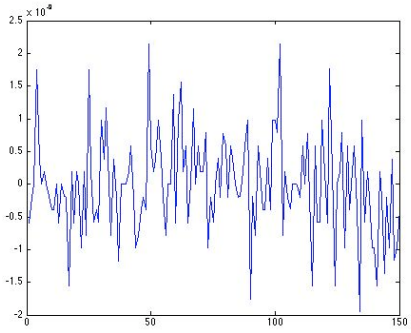
There is also not much change in the order of the error for σ and η when increasing the perturbations or the number of observations. The same happened to μ, κ and η either.

These findings gives strong evidence in favour of our method. The stability of the method, even for very different starting values, indicates that the maximization problem may indeed have a global maximum. A formal mathematical proof of this fact shall be addressed in future research.

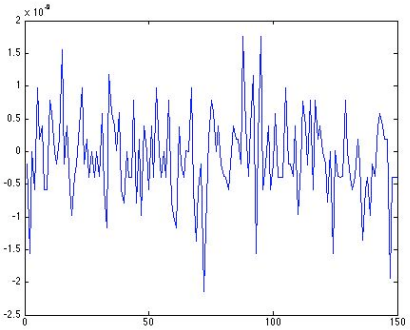
N = 10,000



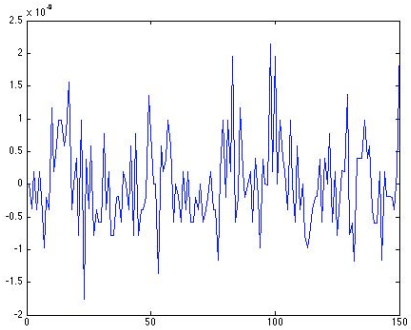
1% perturbations



20% perturbations

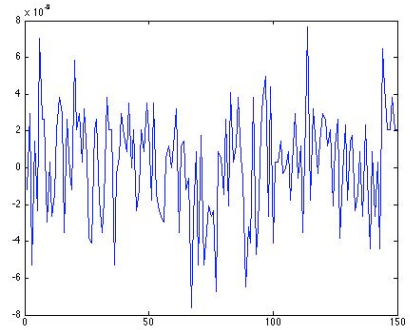


250% perturbations

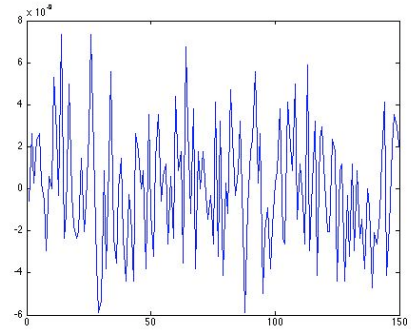


1000% perturbations

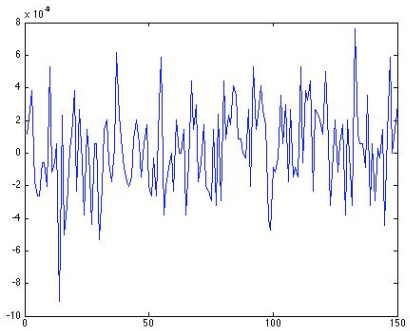
N = 30,000



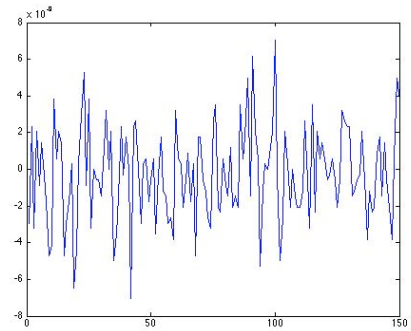
1% perturbations



20% perturbations



250% perturbations



1000% perturbations

Table 4.6: Iteration errors plots - κ

Chapter 5

Energy Market Data

5.1 Data Exploration

In this Section, we apply the theoretical framework developed for calibrating a two-factor model to real data. As noted by Hikspoors and Jaimungal (2007) and Lucia and Schwartz (2002), the energy commodity markets are different from the traditional financial security ones. They lack the high level of liquidity that many financial markets enjoy, the storage costs of most commodities translate into peculiar price behaviors and commodity prices tend to have strong mean-reverting trends.

We use data from the energy market, more specifically, natural gas prices from the Henry Hub. The Henry Hub is a pipeline owned and operated by Sabine Pipeline, LLC. The Sabine Pipeline discloses that the lines start in eastern Texas, runs through and ends in Louisiana. The hub is physically situated at Sabine's Henry Gas Processing Plant.

The Henry Hub interconnects several pipelines in the United States that collectively provide access to markets across the country. The Energy Information Administration (EIA) estimates that approximately 49 percent of the U.S. wellhead production of natural gas occurs near the Henry Hub or passes close to it when moving to downstream consumption markets.

The EIA explains in its website that the Henry Hub prices are reported on a daily basis, differently from its reports, which show an average wellhead price for natural gas in monthly and annual basis. Therefore, market participants have adopted spot and futures prices from the Henry Hub as a substitute measure for the current wellhead price, making it the largest centralized point for natural gas spot and futures trading in the United States.

Our data set consists of daily prices of Henry Hub Natural Gas Futures from May 1996 to April 2008. The contracts' trading unit is 10,000 million British thermal units (mmBtu), and the price quotation is in U.S. dollars

and cents per mmBtu. We work with 24 maturities, ranging from 1 to 24 month future contracts. The data has been adjusted so that each time series correspond to a maturity, avoiding problems with the contracts' expiry.

We can see the futures price surface plot in figure 5.1. We plot the data from May 1996 to April 2008 for the 24 first maturities.

We begin by displaying some descriptive statistics and plots from the Henry Hub data. We disclose the information relative to 1, 12 and 24 months maturity contracts, believing that this is a proxy to the short, medium and long term futures contracts.

ACF	Lag 1	Lag 21	Lag 126	Lag 252
P_t	0.995	0.912	0.628	0.506
$P_t - P_{t-1}$	0.0034	0.018	0.023	-0.019
$\log(P_t)$	0.996	0.936	0.726	0.564
$\log(P_t) - \log(P_{t-1})$	-0.018	0.027	0.015	-0.019

Table 5.1: Autocorrelation Functions for 1 month futures

ACF	Lag 1	Lag 21	Lag 126	Lag 252
P_t	0.998	0.974	0.825	0.731
$P_t - P_{t-1}$	-0.0052	0.047	0.005	0.057
$\log(P_t)$	0.998	0.977	0.851	0.741
$\log(P_t) - \log(P_{t-1})$	-0.013	0.024	-0.011	0.017

Table 5.2: Autocorrelation Functions for 12 month futures

ACF	Lag 1	Lag 21	Lag 126	Lag 252
P_t	0.998	0.977	0.842	0.754
$P_t - P_{t-1}$	-0.011	0.024	-0.0016	0.091
$\log(P_t)$	0.998	0.978	0.864	0.766
$\log(P_t) - \log(P_{t-1})$	0.0 03	-0.001	-0.017	0.040

Table 5.3: Autocorrelation Functions for 24 month futures

The autocorrelation functions (ACF) show that the price series for all maturities considered have high autocorrelation coefficients, suggesting a high degree of persistence in the price data, both in level and logarithm formats. Note that the autocorrelation is very high in a one-day lag, but it decreases while we move to more distant points in time. One month futures have a smaller degree of persistence than 12 or 24 months, suggesting that the short term prices behavior is less driven by their historical behavior than the long term ones. One of the reasons for this is that short-term prices are more

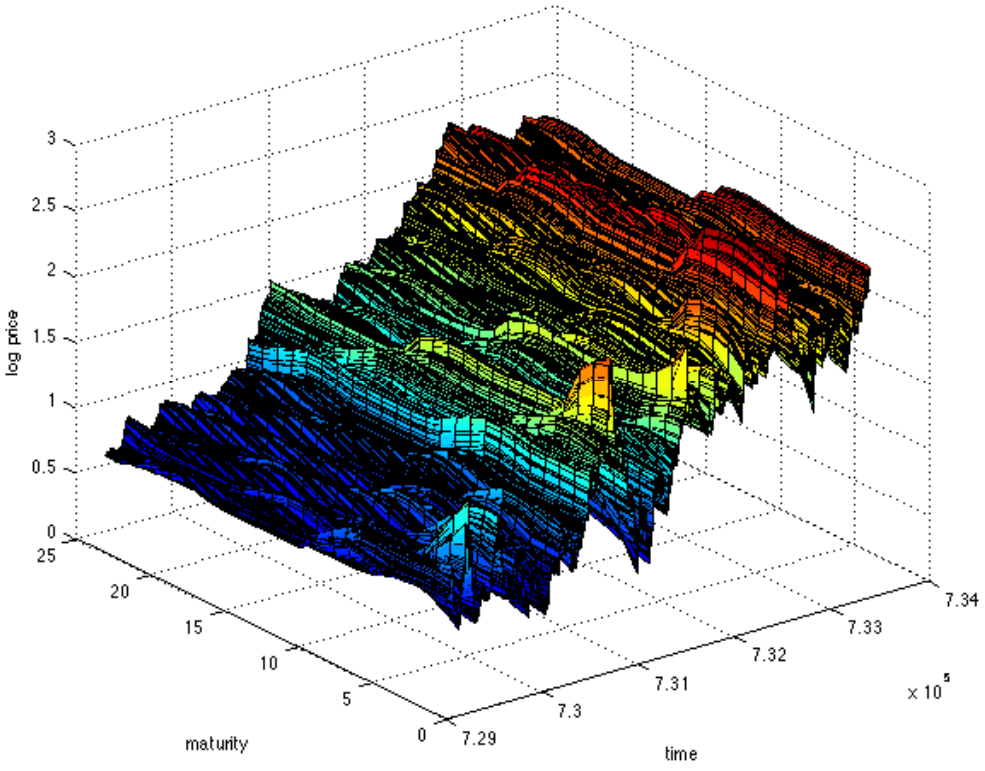


Figure 5.1: Henry Hub futures prices plot

sensible to external shocks than longer ones. As long term contracts usually reflect fundamentals, they respond less to momentary shocks.

Autocorrelation functions for return series and logarithm returns behave

differently from the price ones. They have very low autocorrelation estimates across all lags and maturities.

Tables 5.4, 5.5 and 5.6 show that the data behaves like typical financial time series. They obey some of the stylized facts of financial time series stated in Tsay (2002). That is, there is excess kurtosis for all of the price returns and logarithm returns and their mean is close to zero. The most volatile series considered are the ones for the 1-month maturity contracts, followed by the 12-month contracts and the 24-month ones, showing respectively 58.7%, 27% and 23% of annualized volatility for the logarithm returns. This suggests that short term contracts are more susceptible to external shows and market movements, which is an expected result.

	Mean	Median	Std Deviation	Skewness	Kurtosis
P_t	4.852	4.635	2.567	0.925	3.808
$P_t - P_{t-1}$	0.002	0.002	0.214	0.601	18.095
$\log(P_t)$	1.440	1.5336	0.532	0.037	1.818
$\log(P_t) - \log(P_{t-1})$	0.000	0.000	0.037	0.346	7.568

Table 5.4: Descriptive Statistics for 1 month futures

	Mean	Median	Std Deviation	Skewness	Kurtosis
P_t	5.020	4.368	2.712	0.645	2.179
$P_t - P_{t-1}$	0.002	0.002	0.111	-3.210	51.344
$\log(P_t)$	1.461	1.474	0.546	0.118	1.624
$\log(P_t) - \log(P_{t-1})$	0.000	0.000	0.017	-1.352	16.472

Table 5.5: Descriptive Statistics for 12 month futures

	Mean	Median	Std Deviation	Skewness	Kurtosis
P_t	4.793	3.99	2.53	0.678	2.155
$P_t - P_{t-1}$	0.002	0.002	0.091	-6.42	136.047
$\log(P_t)$	1.429	1.385	0.526	0.168	1.681
$\log(P_t) - \log(P_{t-1})$	0.000	0.000	0.0148	-2.279	33.657

Table 5.6: Descriptive Statistics for 24 month futures

A first look to the prices plot reveals an erratic behavior to the futures price. There is a clear upward trend in all of the maturities considered, beginning in 2002 and going until the end of the sample. The most significant hikes happen in the begining of 2001 and 2003 and at the end of 2005. Note that these dates correspond to the winter season in the northern hemisphere, when the demand for natural gas tends to be higher.

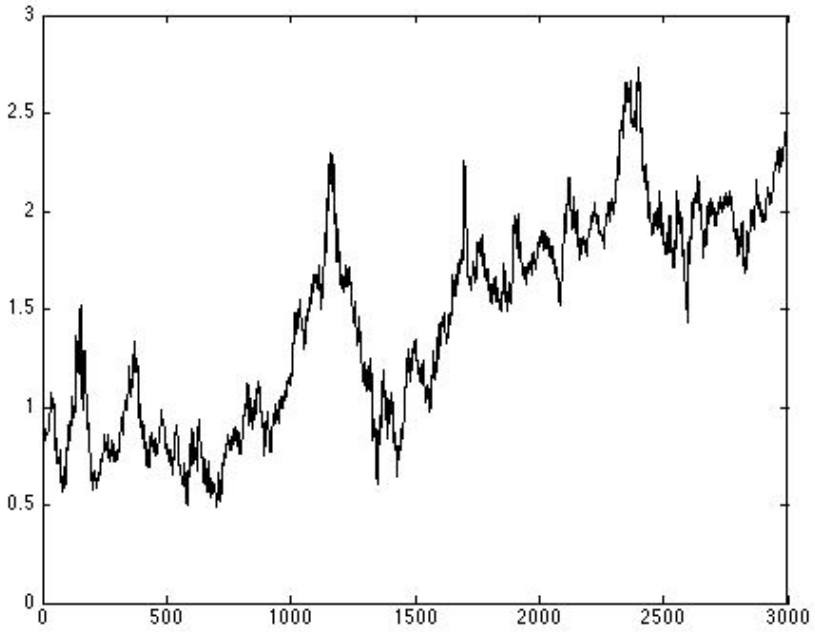


Figure 5.2: Henry Hub data plot - 1 month futures

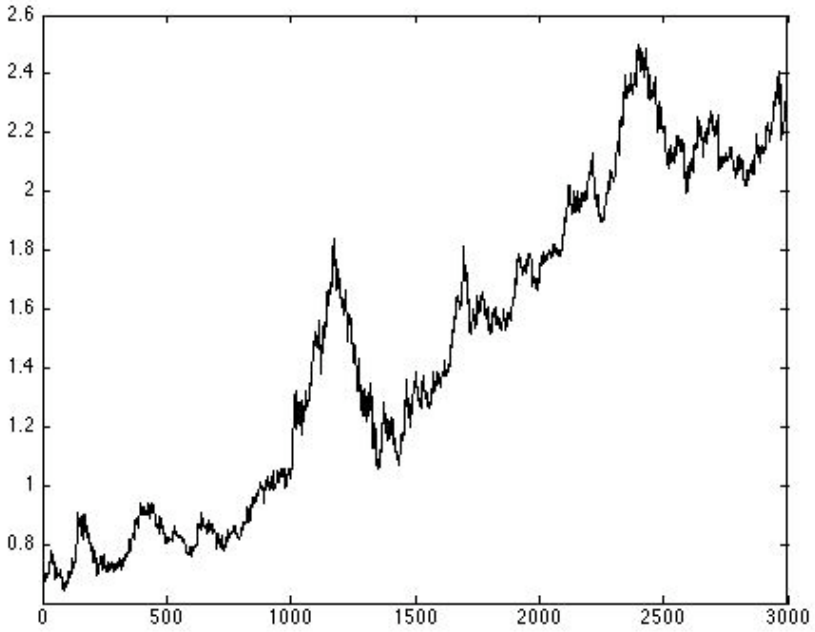


Figure 5.3: Henry Hub data plot - 12 month futures

We can conclude from the analysis of the descriptive statistics of the data that a two-factor model like Schwartz and Smith (2000) is suited to model the

Figure 5.4: Henry Hub data plot - 24 month futures

data. Note that the 24 month futures contract has an annualized standard deviation of 23% for the logarithm returns. This suggests that the long end of the price curve is not flat. Therefore, fitting a one-factor model like Schwartz (1997) to this data set would yield results like the ones shown in figure 2.1, where the long end of the curve is not modeled.

5.2 Estimation Results

In this section, we present the estimation results of our algorithm when applied to the Henry Hub data. We initially analyze the factor decomposition of the data, trying to infer how suitable a two-factor model would be.

As defined in Johnson and Wichern (1998), factor analysis is a method which aims to describe, if possible, the covariance relationship among many variables in terms of a few underlying but unobserved random quantities called factors. A simple, but useful factor model is the Orthogonal Factor Model (OFM).

Consider a random vector X with p components, mean μ and covariance matrix Σ. The OFM postulates that X is linearly dependent upon a few unobserved random variables $F_1, F_2, \ldots, F_m$, called common factors, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$ random shocks, or errors. We can write this model as

$$X - \mu = LF + \epsilon \tag{5.1}$$

where $X - \mu$ is a $(p \times 1)$ matrix of demeaned variables, L (the loading matrix) is a $(p \times m)$ matrix of $l_{i,j}$ coefficients called loadings of the i-th variable on the j-factor, and ϵ is a $(p \times 1)$ matrix of shocks.

We assume that $\mathbb{E}[F] = 0_{(m\times 1)}$, $\mathbb{COV}[F] = \mathbb{E}[FF'] = I_{(m\times m)}$, $\mathbb{E}[\epsilon] = 0_{(p\times 1)}$ and $\mathbb{COV}[\epsilon] = \mathbb{E}[\epsilon\epsilon'] = \Psi_{(p\times p)}$ where Ψ is a diagonal matrix.
Finally, $\mathbb{COV}[\epsilon F] = \mathbb{E}[\epsilon F'] = 0_{(p\times m)}$.

There are many methods available for estimating the factor loadings. One of the most used is Principal Components. The advantage of the method lies on its simplicity. Principal components are easy to implement, can be readily calculated even for large matrices and can be generalized to handle missing values in the data. In spite of not having a direct interpretation, simple linear regressions and correlation exercises can help identify the extracted factors.

The estimation of (5.1) relies on imposing restrictions on the covariance matrix, Σ. As the population covariance matrix in unknown, we carry all estimations on the sample covariance matrix S. Although we keep the notation on Σ for practicity, we shall always be dealing with S.

In the principal components model we factor the matrix Σ in terms of its eigenvalues λ_i, and eigenvectors e_i, and order the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$. Consequently, the principal components depend only on the data's covariance matrix Σ, without any assumption on the data's distribution.

We use Matlab to extract the eigenvalues and eigenvectors of the data matrix. The eigenvectors picture the factors' behavior. As we can see in Figure 5.5, the first and second eigenvectors behave in a similar path to the ones generated with the Schwartz and Smith model, and shown in Figures 4.9 and 4.10.

However, the relative weight of the first factor is considerably higher to the second one. Normalizing the factors to one, we can see that the first factor corresponds to 23% of the data total explained variance, while the second one accounts only for 0.32%. Even with a relative smaller influence of the second factor, it should not be disconsidered. As mentioned in Chapter 1, factor 2 is responsible for providing a better fitting to the futures term structure. Besides, Litterman and Scheinkman (1991) shows that even with 97% of the total explained variance of the yield curve explained by the first factor, the second and third factor play an important role in adding more variability to the term structure of interest rates.

Since the two-factor model seems to be a reasonable choice to calibrate the Henry Hub data, we proceed with the estimation process. In order to show the

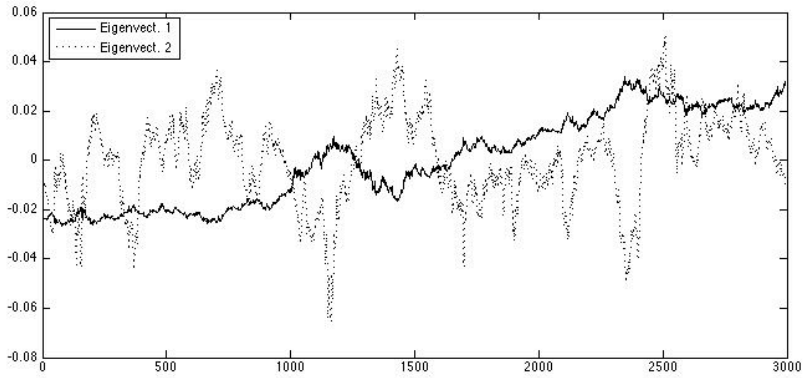


Figure 5.5: First two eigenvectors from the Henry Hub data

results of our estimation procedure, we follow the steps described in Chapter 4 and calculate estimates for κ, θ, σ, η and μ. The parameter ρ is calculated as the correlation between the series X and Y.The results are displayed in Table 5.7.

We can see that our algorithm works well also with real data sets. The error is sufficiently small to give us confidence that our procedure is able to recover the parameters from the stochastic processes X and Y.

As there is an upward shift in natural gas prices beginning in 2002, we also estimate the model's parameters in two sub-samples besides the original data set. We consider estimates from May 1996 to December 2001 and from January 2002 to April 2008. The results are displayed in table 5.7.

For future research it would be interesting to compare the results obtained with our procedure to the ones generated by a concurrent estimation method, such as the Kalman Filter.

	κ	θ	σ	η	μ	ρ
1996-2008						
Estimates	3.281	0.122	0.608	0.369	0.0009	-0.2857
Error	-1.330e-07	-4.159e-07	-7.339e-08	2.443e-10	-4.620e-08	
1996-2001						
Parameter Estimates	4.638	0.168	0.770	0.429	-0.001	0.0021
Error	-1.250e-07	-8.972e-07	-1.443e-07	2.055e-09	-1.024e-07	
2002-2008						
Parameter Estimates	4.781	0.053	0.368	0.302	0.002	-0.3218
Error	-6.010e-08	-5.746e-07	-4.599e-08	-8.385e-10	-5.049e-08	

Table 5.7: Estimation Results for the Henry Hub data

Chapter 6

Conclusion

In this work we have built a robust methodology to calibrate a Schwartz-Smith model for commodity prices. This methodology is a two stage process that is related to a similar method in Hikspoors and Jaimungal (2007), and is an alternative to Kalman Filter estimation.

Given the model parameters as an input, it first obtains the hidden processes fitting the term structure of commodity prices via Non Linear Least Squares (NLLS). In the second stage, we use the processes obtained via NLLS to generate new guesses of the parameters via Maximum Likelihood Estimation. The result is then iterated until convergence.

sting its parameters to the risk-neutral measure. We have also detailed the estimation algorithm and the likelihood functions used.

The methodology was validated using synthetic data. We have first estimated the parameters with known initial guesses, which were then recovered by our algorithm. That way, we could measure accurately the estimation error and therefore, the algorithm's efficiency to recover the parameters. The estimation results for synthetic data sets showed that the parameters estimates are accurate, and that the convergence was achieved in very few iterations. Figures 4.3 - 4.7 show the error patterns for the estimated values, which converge to small values quickly. The only exception is the parameter κ. It converges at a slower pace to a small estimation error, as the result of a highly nonlinear optimization problem on its estimation.

The optimal tolerance level obtained for the iteration steps was 10^{-6}. This number was obtained after tests using starting values ranging from 10^{-2} to 10^{-12}, and was the value we used in our optimization routines.

Next, we attested our methodology's robustness by imposing different disturbances on the initial values. This procedure tested whether the algorithm still converged even if the starting values were not close to the final solution. Not only we disturbed the parameters in one single direction, but we also cre-

ated random shocks that made the initial guesses move in different directions. It is important to mention that in all cases our algorithm converged to the same parameter vector. The method shows here an advantage to the Kalman Filter. While the latter depends on the quality of the initial guess given, our algorithm converged even with very distressed starting values.

Finally, we tested our methodology in real market data. We used futures prices of natural gas from the Henry Hub, from 1996 to 2008, considering the 24 first maturities. After statistically describing the data, we ran our algorithm and obtained very small iteration convergence errors. This was an expected result, as the model proved to be robust to recover the true values used to construct the synthetic data.

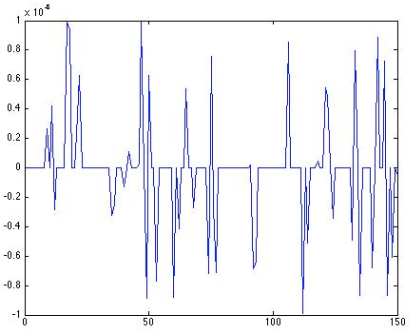
The robust convergence we obtained even with all the perturbations motivates suggests that the optimization problem we solve here has a global maximum. That is why our future research shall focus on the formalization of this problem and a proof of a global and unique solution. Besides, confidence intervals could also be constructed for the parameters estimates in order to allow further statistics investigations.

Another interesting topic for future research would be to compare our estimates with ones generated by a concurrent model, such as the Kalman Filter.

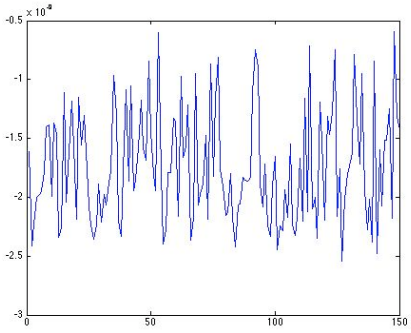
Appendix A

Additional Figures

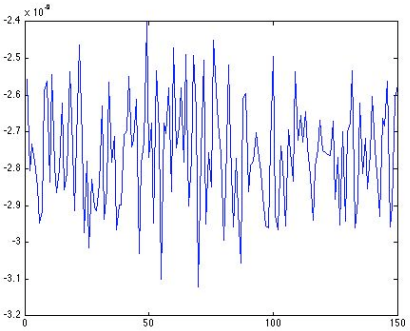
N = 10,000



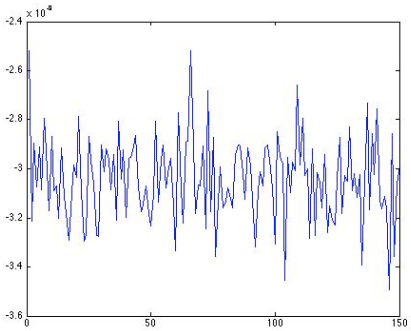
1% perturbations



20% perturbations

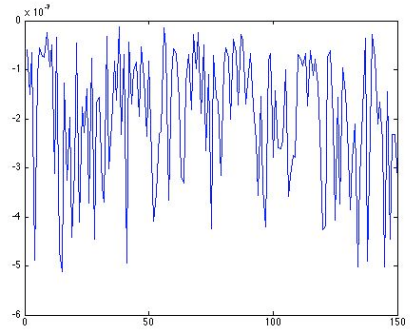


250% perturbations

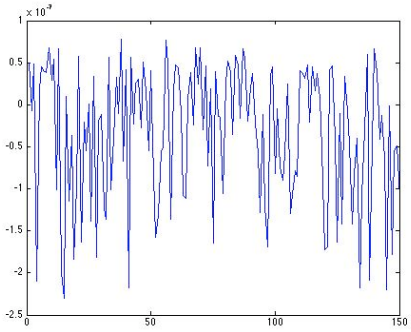


1000% perturbations

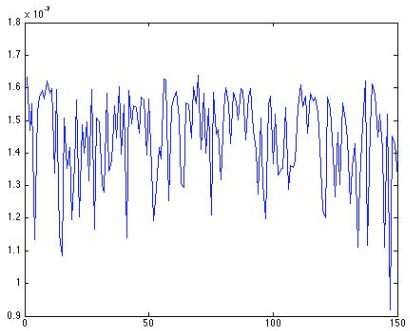
N = 30,000



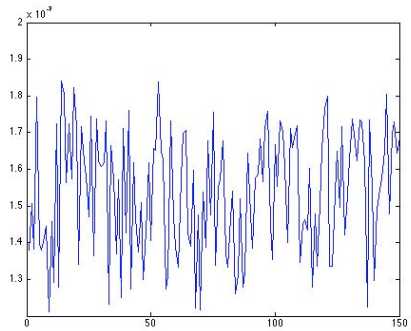
1% perturbations



20% perturbations



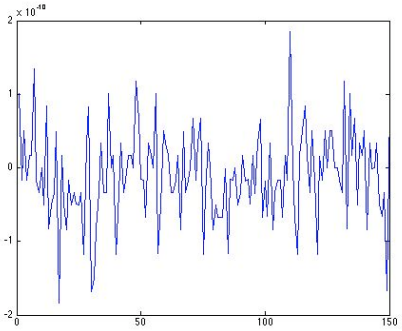
250% perturbations



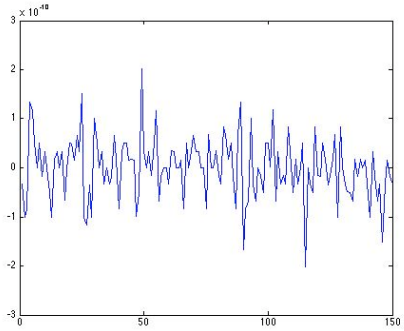
1000% perturbations

Table A.1: Iteration errors plots - θ

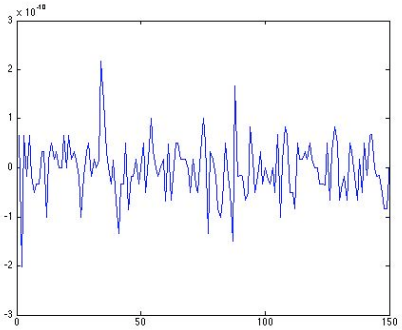
N = 10,000



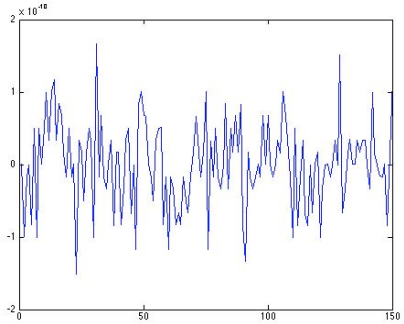
1% perturbations



20% perturbations

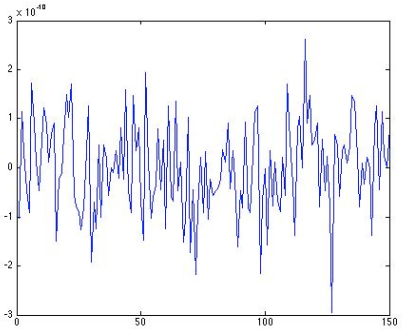


250% perturbations

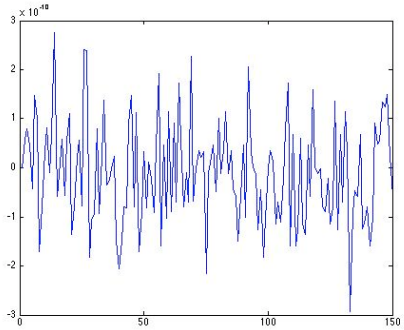


1000% perturbations

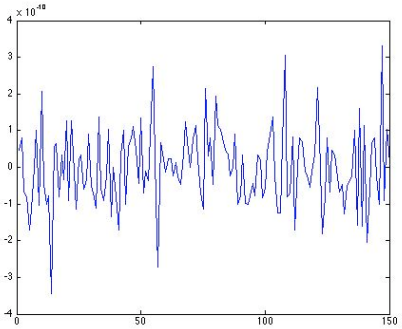
N = 30,000



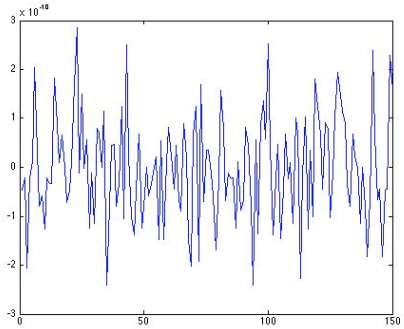
1% perturbations



20% perturbations



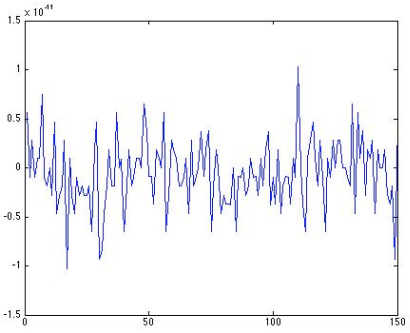
250% perturbations



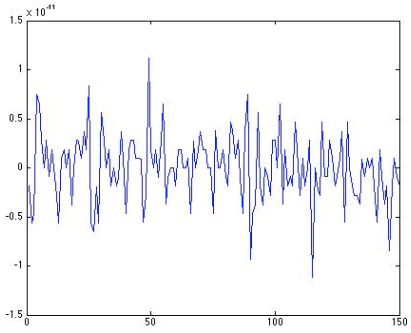
1000% perturbations

Table A.2: Iteration errors plots - σ

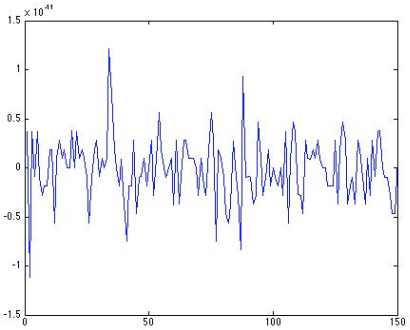
N = 10,000



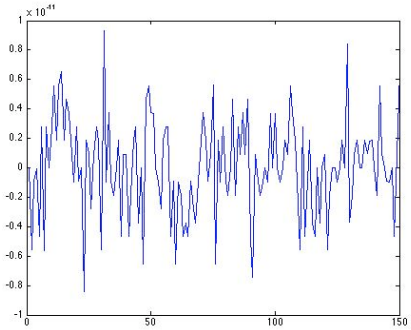
1% perturbations



20% perturbations

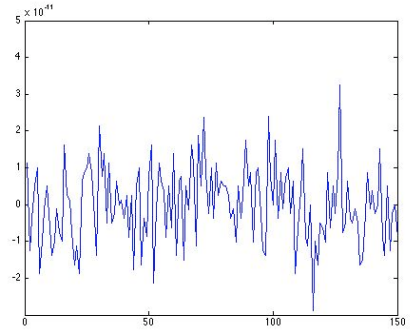


250% perturbations

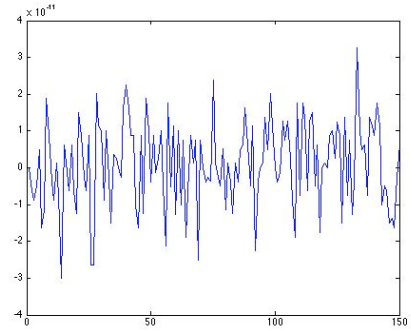


1000% perturbations

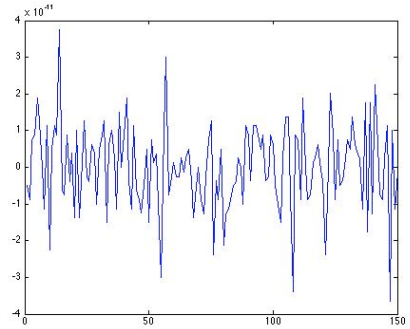
N = 30,000



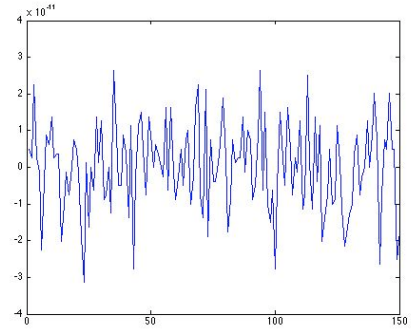
1% perturbations



20% perturbations



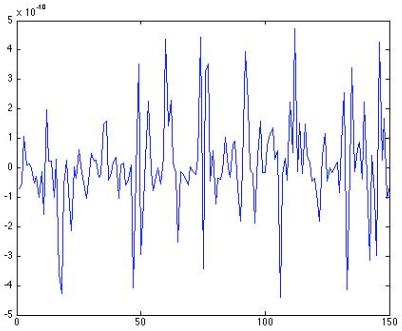
250% perturbations



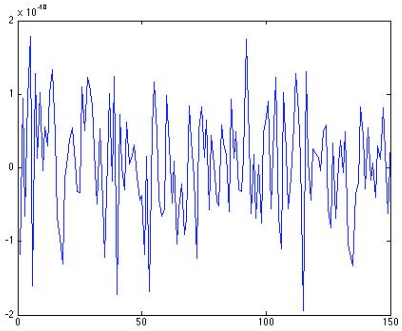
1000% perturbations

Table A.3: Iteration errors plots - η

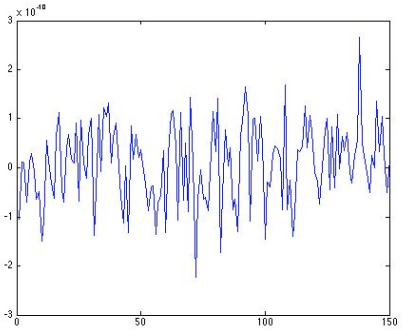
N = 10,000



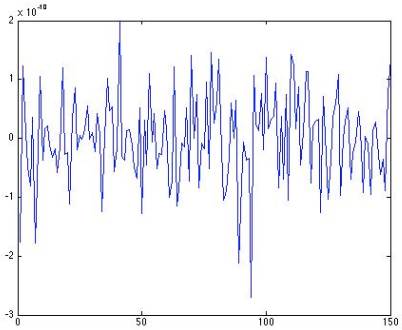
1% perturbations



20% perturbations

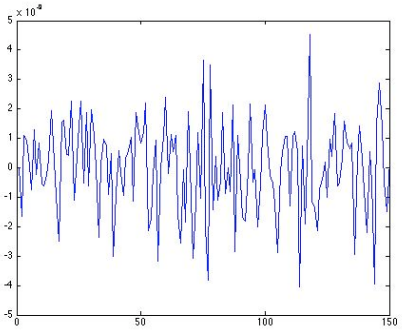


250% perturbations

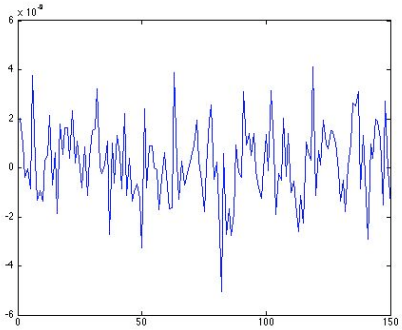


1000% perturbations

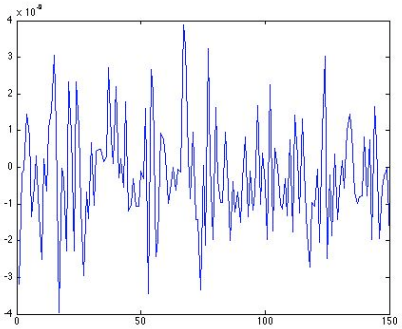
N = 30,000



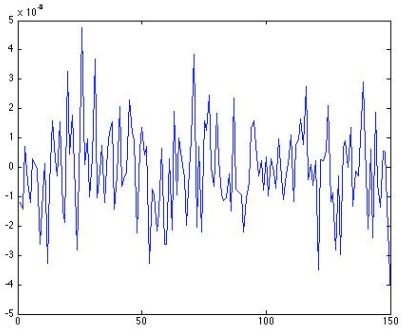
1% perturbations



20% perturbations



250% perturbations



1000% perturbations

Table A.4: Iteration errors plots - μ

Appendix B

Computer Codes

B.1 Euler-Maruyama Discretization

```
%Y/EM Euler-Maruyama method on linear SDE
%. SDE is dX = lambda*X dt + mu*X dW, X(O) = Xzero,
% where lambda = 2, mu = 1 and Xzero = 1.
% Discretized Brownian path over [0,1] has dt = 2"(-8).
% Euler-Maruyama uses timestep R*dt.
randn('state',1000)
lambda = 2; mu = i; Xzero = 1; %
%problem parameters
T = 1; N = 2^8; dt = 1/N;
dW = sqrt(dt)*randn(1,N); % Brownian increments
W = cumsum(dW); %/ discretized Brownian
path
Xtrue = Xzero*exp((lambda-0.5*mu^2)*([dt:dt:T])+mu*W);
plot([0:dt:T],[Xzero,Xtrue],'k-'), hold on
R = 4; Dt = R*dt; L = N/R; % L EM
%steps of size Dt = R*dt
Xem = zeros(1,L); %/
%preallocate for efficiency
Xtemp = Xzero;
for j = 1:L
Winc = sum(dW(R*(j-1)+1:R*j));
Xtemp = Xtemp + Dt*lambda*Xtemp + mu*Xtemp*Winc;
Xem(j) = Xtemp;
end
plot([0:Dt:T],[Xzero,Xem],'k--*'), hold off
xlabel('t','FontSize',12)
```

```
ylabel('X', 'FontSize' ,16, 'Rotation' ,0,'HorizontalAlignment', 'right')
emerr = abs(Xem(end)-Xtrue(end))
```

B.2 Calibration Algorithms

```
function [Xb,Yb,tau,F_barra,brownX,brownY,X,Y] =
 main_code(kappa,theta,sigma,eta,mu,muStar,rho,lambda_x,n)

load tau;
[Ma,M]=size(tau);
tau = tau(end-(n-1):end,:);

N=n;
deltaT=1/252;

varY=eta^2*deltaT;
brownX=randn(n,1);
brownY=randn(n,1);

theta = -lambda_x/kappa;
lambda_y = mu - muStar;

%% generating  AR and RW


X=genAR(n,deltaT,kappa,theta,sigma,brownX);
Y=genRW(n,deltaT,eta,mu,rho,brownY,brownX);

rho_hat=corr(X,Y);

 %%generating futures matrix


[logFut,F_barra,S,A]=genFUT(N,M,X,Y,mu,tau,sigma,kappa,eta,rho);


%% Recovering X and Y from the generated Futures Matrix
%% testing if the generated X and the estimated X are the same.
```

```
[difference,Xa]=testeX(tau,F_barra,kappa,S,A,X);

if difference > 0.001 disp('Attention! X differ too much')
end

[Xb,Yb] = estXY(tau, F_barra,kappa,sigma, eta,rho,mu);

function X=genAR(n,deltaT,kappa,theta,sigma,brownX)


epsilon=sigma*sqrt(deltaT)*brownX;

X=zeros(n,1);
X(1)=theta;

for i = 2:n
X(i)=X(i-1)+kappa*(theta-X(i-1))*deltaT+epsilon(i); % Euler-Maruyama
end



function Y=genRW(n,deltaT,eta,mu,rho,brownY,brownX)

epsilonY=eta*sqrt(deltaT)*((rho^2)*brownX + sqrt(1-rho^2)*brownY);

Y=zeros(n,1);
Y(1)=mu;

for i = 2:n
Y(i)=mu*deltaT+Y(i-1)+epsilonY(i);
end


function [IterationError, estimationError, estimatedParameters,numiter] =

iterations(F_barra,tau,Xb,Yb,kappa,theta,sigma,eta,mu,muStar,rho,lambda_x,n,eps)



disp('iterations...')
```

```
%tolerance
%eps = 1e-6;
numiter =0;
max_numiter =100;
deltaT=1/252;
N=n;

lambda_y=mu-muStar;

initial_kappa=kappa;
initial_theta=theta;
initial_sigma=sigma;
initial_eta = eta;
initial_mu=mu;
initial_lambda_y = lambda_y;
initial_rho=rho;
initial_lambda_x=lambda_x;

current_kappa=kappa;
current_theta=theta;
current_sigma=sigma;
current_etaY = eta;
current_mu=mu;
current_lambda_y = lambda_y;
current_rho=rho;
current_lambda_x=lambda_x;

IterationError = ones(1,8);
estimationError = ones(1,8);

while numiter < max_numiter && norm(IterationError)>eps

    numiter= numiter +1;

    if numiter == max_numiter
        disp('Maximum number of iterations reached without obtaining convergence')
    end


[Xb,Yb] = estXY(tau, F_barra,current_kappa,current_sigma,
```

```
current_etaY,current_rho,current_mu);



[kappa,mLL]=fminbnd(@(x)V(x,n,deltaT,Xb),0.001,5);

theta=f(kappa,n,deltaT,Xb);
sigma=g(theta,kappa,n,deltaT,Xb);
lambda_xE = -kappa*theta;

[aY,bY,varY]=calibrateGARCH(Yb,n);
mu=aY;
lambda_y = mu - muStar;
etaY=sqrt(varY/deltaT);

IterationError=[(current_kappa-kappa),(current_theta-theta),
(current_sigma-sigma), (current_etaY-etaY),(current_mu-mu),
(current_rho-rho),(current_lambda_x-lambda_xE), (current_lambda_y-lambda_y)];

current_kappa = kappa;
current_theta = theta;
current_sigma = sigma;
current_etaY = etaY;
current_mu = mu;
current_rho=rho;
current_lambda_x=lambda_xE;
current_lambda_y=lambda_y;

end

estimatedParameters = [current_kappa,current_theta,current_sigma,
current_etaY,  current_mu,  current_rho, current_lambda_x, current_lambda_y];

estimationError = [(initial_kappa-current_kappa),(initial_theta-current_theta),
(initial_sigma-current_sigma),(initial_eta-current_etaY),(initial_mu-current_mu),
(initial_rho-current_rho), (initial_lambda_x-lambda_x), (initial_lambda_y-lambda_y)];
```

Appendix C

Kalman Filter Estimation

The basic idea of the Kalman Filter is to express a dynamic system in a particular form, called the state-space representation, and use an algorithm for sequentially updating a linear projection of the system.

We start by defining the state-space representation of a dynamic system. This section is based on Hamilton (1994) and Schwartz (1997).

Definition C.0.1. *Let* $\mathbf{y}_t$ *denote a* $(n \times 1)$ *vector of observed variables at time* t*, and* ξ_t *a* $(r \times 1)$*, possibly unobserved, vector called state vector. The state-space representation of the dynamics of* $\mathbf{y}$ *is given by the following system of equations:*

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{v}_{t+1} \tag{C.1}$$

$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\xi_t + \mathbf{w}_t, \tag{C.2}$$

where $\mathbf{F}$, $\mathbf{A}'$ and $\mathbf{H}'$ are matrices of parameters, with dimension $(r \times r)$, $(n \times k)$, and $(n \times r)$, respectively, and $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous or predetermined variables[1].

Equation (C.1) is called the state equation, and generates the unobserved state variables, while equation (C.2) is called the observation equation, and relates observed variables to an unobserved vector of state variables, the state vector.

We also have that $\mathbf{v}_t$ and $\mathbf{w}_t$ are white noise $(r \times 1)$, and $(n \times 1)$ vectors such that:

$$\mathbb{E}(\mathbf{v}_t, \mathbf{v}'_\tau) = \begin{cases} \mathbf{Q} & \text{for } t = \tau, \\ 0 & \text{otherwise} \end{cases} \tag{C.3}$$

[1]The statement that $\mathbf{x}_t$ is exogenous or predetermined means that $\mathbf{x}_t$ provides no information about ξ_{t+s} or $\mathbf{w}_{t+s}$, for $s = 0, 1, 2, \ldots$ beyond that contained in $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_t$.

and

$$\mathbb{E}(\mathbf{w}_t, \mathbf{w}'_\tau) = \begin{cases} \mathbf{R} & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{C.4}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are $(r \times r)$ and $(n \times n)$ matrices and $\mathbf{v}'_\tau$ and $\mathbf{w}'_\tau$ are transposed vectors. Also, we assume that

$$\mathbb{E}(\mathbf{v}_t, \mathbf{w}'_\tau) = 0 \quad \text{for all } t \text{ and } \tau, \tag{C.5}$$

which means that the disturbances are uncorrelated at all lags.

The objective of the calibration procedure is to estimate the values of any unknown parameters in the system based on the observations of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$.

We consider an algorithm using the Kalman Filter for calculating linear least square forecasts of the state vector on the basis of data observed through date t:

$$\hat{\xi}_{t+1|t} \equiv \mathbb{E}(\xi_{t+1}|\mathcal{Y}_t). \tag{C.6}$$

where $\mathcal{Y}_t \equiv (\mathbf{y}'_t, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_1, \mathbf{x}'_t, \mathbf{x}'_{t-1}, \ldots, \mathbf{x}'_1)'$, with associated mean square error (MSE) for each of the forecasts represented by the $(r \times r)$ matrix:

$$\mathbf{P}_{t+1|t} \equiv \mathbb{E}[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})']. \tag{C.7}$$

The Kalman Filter estimations can be started with $\hat{\xi}_{1|0} = 0$ and

$$\mathbf{P}_{1|0} = \mathbb{E}\{[\xi_1 - \mathbb{E}(\xi_1)][\xi_1 - \mathbb{E}(\xi_1)]'\}.$$

The next step is to perform the calculations for $\hat{\xi}_{2|1}$ and $\hat{P}_{2|1}$, and so on. Then we iterate on:

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + F\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\xi}_{t|t-1}) \tag{C.8}$$

We denote $\hat{\xi}_{t+1|t}$ as the best forecast of ξ_{t+1} based on an affine function of $\mathbf{y}$. The coefficient matrix in (C.8) is known as the Kalman Gain. That is:

$$\mathbf{K}_t = F\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}, \tag{C.9}$$

so (C.8) can be re-written as:

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\xi}_{t|t-1}). \tag{C.10}$$

Bibliography

Gusev Yuri Barlow Martin and Manpo Lai. Calibration of multifactor models in electricity prices. *International Journal of Theoretical and Applied Finance*, 7(2):101–120, 2004.

Michael J Brennan and Eduardo S Schwartz. Evaluating natural resource investments. *Journal of Business*, 58(2):135–57, April 1985.

Avinash K. Dixit and Robert S. Pindyck. *Investment under Uncertainty.* Princeton University Press, 1994.

Eugene F Fama and Kenneth R French. Commodity futures prices: Some evidence on forecast power, premiums,and the theory of storage. *Journal of Business*, 60(1):55–73, January 1987.

Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time.* de Gruyter Series in Mathematics, 2004.

Rajna Gibson and Eduardo S Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 45(3):959–76, July 1990.

James D. Hamilton. *Time Series Analysis.* Princeton, 1994.

Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

Samuel Hikspoors and Sebastian Jaimungal. Energy spot price models and spread options pricing. *International Journal of Theoretical & Applied Finance*, pages 1111–1135, 2007.

John C. Hull. *Options, Futures and Other Derivatives.* Prentice Hall, 2006.

Richard Johnson and Dean Wichern. *Multivariate Statistical Analysis.* Prentice Hall, 1998.

Ralf Korn and Elke Korn. *Option Pricing and Portfolio Optimization.* American Mathematical Society, 2000.

Robert Litterman and José Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, pages 54–61, June 1991.

Julio J. Lucia and Eduardo S. Schwartz. Electricity prices and power derivatives. - evidence from the nordic power exchange. *Review of Derivatives Research*, 5:5–50, 2002.

Dragana Pilipovic. *Valuing and Managing Energy Derivatives*. McGraw-Hill, 1997.

Eduardo Schwartz and James E. Smith. Short-term variations and long-term dynamics in commodity prices. *Manage. Sci.*, 46(7):893–911, 2000.

Eduardo S Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–73, July 1997.

Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance, 2004.

Ruey S. Tsay. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics, 2002.

William T. Vetterling William H. Press, Saul A. Teukolsky and Brian P. Flannery. *Numerical Recipes - The Art of Scientific Computing*. Cambridge University Press, 1992.