Norwegian School of Economics
Bergen, December, 2015

Modeling the Steel Price for Valuation of Real Options and Scenario Simulation

Master thesis in Finance

Norwegian School of Economics
Written by: Lasse Berggren
Supervisor: Jørgen Haug

Summary

Steel is widely used in construction. I tried to model the steel price such that valuations and scenario simulations could be done. To achieve a high level of precision this is done with a continuous-time continuous-state model. The model is more precise than a binomial tree, but not more economically interesting. I have treated the nearest futures price as the steel price. If one considers options expiring at the same time as the futures, it will be the same as if the spot were traded. If the maturity is short such that details like this matters, one should treat the futures as a spot providing a convenience yield equal to the interest rate earned on the delayed payment. This will in the model be the risk-free rate.

Then I have considered how the drift can be modelled for real world scenario simulation. It involves discretion, as opposed to finding a convenient AR(1) representation, because the ADF-test could not reject non-stationarity (unit root).

Given that the underlying is traded in a well functioning market such that prices reflect investors attitude towards risk, will the drift of the underlying disappear in the one-factor model applied to value a real-option. The most important parameter for the valuation of options is the volatility. I have estimated relative and absolute volatility. The benefit of the relative volatility is the non-negativity feature.

Then I have estimated a model where the convenience yield is stochastic. This has implications for the risk-adjusted model. I have difficulties arriving at reliable parameter estimates. Here small changes in arguments have large effects on the option value. Therefore should this modelling be carried through only if one feels comfortable that it is done properly.

I finish by illustrating how real-option valuation can be performed. The trick is to translate the real-world setting into a payoff function. Then one can consider Monte Carlo simulation if the payoff function turns out to be complicated or if there are decisions to be made during the life of the project. For projects maturing within the horizon traded at the exchange, the expectation of the spot price under the pricing measure is observable.

To truly compare models, plots of the value of derivative should be created to graphically compare the difference in dependence on parameter values. Alternatively, the derivative of the expressions with respect to the parameters are compared. The value of the option to do something, as opposed to be committed to do something, increases with the volatility of future outcomes. Such known results are used instead in the comparison, because the two reliable models (the one-factor models) are pretty similar. This known result is not contradicted by the present values computed in the real option example.

Modeling the Steel Price for Valuation of Real Options and Scenario Simulation

*Contents

Part I:

Introduction

1 Introduction

Motivation for this thesis Steel is a central input factor in a variety of construction projects. Such projects typically extend far into the future and has associated exposure to price changes. Thus, to understand and identify a model for the dynamics of the steel price is interesting for *scenario simulations*, computing the *value of an insurance* and for valuations of the *option to act* in a particular way during projects, i.e. real options in projects.

In which situations is this useful? Consider a construction company involved with a skyscraper and the nearby metro. They want to get an overview of their exposure to price fluctuations in one of their input factors, steel, and manage this risk. How do they achieve risk management?

Consider a shipyard building ships. A major input factor in the construction process is steel. The supplier of steel adjust the price charged to the shipyard to cover its varying expenses. The shipyard is therefore exposed to steel price fluctuations. Assume that the shipyard wants an insurance against high prices. What is the fair price of the insurance?

Consider a steel supplier with a planned production level two years into the future. What is the value of the option to increase production in the event of a high price?

How will I answer these questions? I will identify and calibrate models for the steel price that incorporate time and uncertainty. I will apply what is known as risk-neutral valuation when the task is to compute present values.

Part II:

Steel

2 Global Steel Markets

China plays a key role. The Economist (2015, 9Dec) [1] reports that China produced 822 million metric tonnes in 2014, about half the worlds annual output. China is predicted (late 2015) to produce over 400 million metric tonnes more than it will consume in 2015.[2] China has of November exported over 100 million metric tonnes. To compare these numbers, Japan produce roughly 110 metric tonnes[3], while US produced about 90 million metric tonnes in 2014.[4] Financial Times (2015)[5] report that as of 2014, the volume on the Shanghai Futures Exchange is larger than the London Metal Exchange and Commodity Exchange, Inc (New York) aggregated.

3 Characteristics of Steel

Steel is cheap and has high tensile strength. It can take on a great variety of forms and is widely used in construction, offshore installations and shipbuilding. The properties depends on the particular alloy and the particular production process. Steel can be recycled without loosing its quality — it is the most recycled material in the world.[6] Scrap metal is back in the market after three months.

Steel is cheap to store and transport. Cheap storage leads to stable production for storage — it is more cost effective to produce and store it than to constantly adjust produced output. Warehouses also works as a buffer for shocks in demand, dampening price volatility. Cheap storage costs reduce the difference in prices between steel delivered far in the future and steel delivered soon.

Steel is durable. Steel in warehouses does not degrade, so situations where consumers are unwilling to delay consumption should be rare. That is, situations in which futures contracts with short maturity are more expensive than longer contracts with longer maturity, should be rare.

[1]The Economist Newspaper Ltd: "China's soaring steel exports may presage a trade war", Dec 9th 2015. "Nervousness of steel", Sep 19th 2015, "It's a steel", Jul 13th 2015

[2]Reuters: "China apparent steel consumption falls 5.7 pct from Jan-Oct -CISA", Nov 13th 2015.

[3]The Japan Iron and Steel Federation, Review 2015, http://www.jisf.or.jp/en/statistics/sij/documents/P02_03.pdf

[4]Profile 2015. American Iron and Steel Institute.

[5]http://www.ft.com/intl/cms/s/0/a2df3018-9feb-11e4-9a74-00144feab7de.html#axzz3ukMbHj00

[6]Profile 2015. American Iron and Steel Institute.

Steel compared to other commodities Steel is easy and cheap to store as opposed to gold, oil and in particular electricity. Steel production is not dependent on weather as coffee and oranges are. It is also easier to regulate steel production than cattle production on short notice. Steel is consumed, contrary to silver and gold. Steel is durable as opposed to seasonal commodities. Steel is recycled as opposed to live cattle.

4 Rebar to Represent Steel in General

***The* steel price** Steel is used all over the world and it is not clear what *the* steel price is. There exists several steel price indexes trying to represent a particular type of steel or a geographical area. There are also steel futures prices on different steel products. A *futures contract* is a contract traded on an exchange and standardized with respect to type, quality, location, and maturity for future delivery of the underlying. The underlying is in this case steel. Rebar is the most traded metal futures contract in the world, with 408 million contracts in 2014.[7] Futures prices are easy to interpret because the steel can in fact be delivered to the quoted price. Also, because the prices are on contracts traded on an exchange as opposed to the indices, the prices can be used as an instrument in cash flow management.

Based on the large role of China, and the fact that the SHFE steel futures contracts are relatively new, do I choose to focus on steel in China and use futures contracts traded on the Shanghai Futures Exchange to represent the steel price. I will in particular focus on rebar futures contracts.

What is rebar? Rebar is short for reinforcement bar, and is used in concrete constructions. Concrete has high compression strength but low tensile strength. Steel is therefore used as reinforcement.

Why Rebar? Rebar is the most traded futures contract and should be the best reflection of the steel price in China and Asia. [8] Companies involved with other steel products like wire rod and hot-rolled coil, find the dynamics of rebar interesting as rebar contracts can be used to *cross-hedge*. Cross-hedging is to hedge your direct exposure to price volatility in X by trading a sufficiently correlated product Y. The correlations of rebar versus different related products, given that they are cointegrated:[9]

- Wire Rod SHFE 98% for F2.[10]

[7] The Financial Times Limited, Commodities Explained: Metals trading in China, April 2nd 2015

[8] Volumes: http://www.csidata.com/factsheets.php?type=commodity&format=html&exchangeid=56. On the arbitrary day 1 December, rebar had 350 times (9,194,734/26,300) higher volume than hot-rolled coil while steel wire rod were not even traded. On Decemeber 12, rebar volume were 3,228,720, hot-rolled coil 14,742 and steel wire rod not traded.

[9] F2 contracts matures between 1 month and 2 months, F4 matures between 3 and 4 months, as time passes. I have chosen the shortest maturity available. E.g. Nickel F4 because F1, F2, F3 had shorter time series. The correlation is similar for all maturities for Wire Rod SHFE and Hot Rolled Coil SHFE.

[10] Sample period: 05 Jan 2010-21 Sept 2015.

- Nickel SHFE 90% for F4.[11]
- Hot Rolled Coil SHFE 98% for F4.[12]
- Hot Rolled Coil NYMEX F1 79%, F2 83%, F3 86%, F4 88%.[13]
- Iron Ore (62% Fe, CFR China) NYMEX 97% for F2.[14]

Cross-hedging is typically done if the trading volume (i.e. liquidity) is low in the product you have direct exposure in. Liquidity is important to be able to take the desired hedging positions at any time.

[11] Sample period: 27 Mar 2015-21 Sept 2015.
[12] Sample period: 21 Mar 2014-21 Sept 2015.
[13] Sample period: 05 Jan 2010-21 Sept 2015. Observe how the correlation rises with maturity, likely due to the geographical differences between the exchanges.
[14] Sample period: 05 Jan 2010-21 Sept 2015.

5 Data Description

Rebar Futures Contracts Traded on Shanghai Futures Exchange (SHFE)

Contract specifications The contracts are traded in yuan, while the data set is in dollar. I have daily observations on ten futures contracts, maturing up to ten months. Rebar contracts are physically settled each month. Delivery must take place within five days after the last trading day.[15] The fee is one yuan per metric ton for physical delivery. Minimum Delivery Size is 300 metric tonnes, each contract is on 10 metric tonnes. Certified warehouses are located in Shanghai, Jiangsu (near Shanghai), Guangzhou (near Hong Kong) and Tianjin (near Beijing). For detailed contract specifications, see the appendix on page 80.

Futures Prices The yuan price of the nearest rebar futures contract and the one with the longest maturity, 10 months, are graphed in Figure 2(a). The negative price trend in many commodities over recent years applies to steel as well. The number of observations per different contract maturity is 211.

[15] Last trading day: 15th, or first trading day following 15th

Figure 1: Rebar Futures Prices on Shanghai Futures Exchange in yuan and dollar in (a) and (b), the exchange rate used to convert is in (c). January 2010 - September 2015.

(a) Yuan denominated rebar futures prices $F_{0,1}$ and $F_{0,10}$.

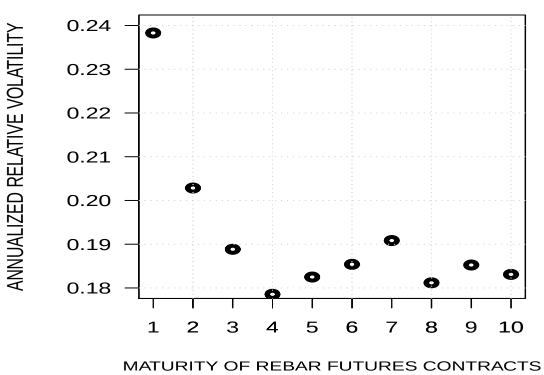
(b) Dollar denominated rebar futures prices $F_{0,1}$ and $F_{0,10}$.

(c) Yuan per US Dollar exchange rate.

Volatility Define relative volatility per week as $\sigma_w = \text{std.dev}[\ln(F_t/F_{t-1})]$ and absolute volatility per week as $\gamma_w = \text{std.dev}[F_t - F_{t-1}]$. To annualize, multiply by $\sqrt{52}$. In Figure 2 is the annualized relative volatility associated with the different maturities plotted and both relative and absolute volatilities are tabulated.

Rebar futures prices with short maturity are more volatile to than those with longer maturities. This is referred to as the Samuelson effect, stating that the arrival of new information has more impact on short maturity contracts than long maturity contracts.

Figure 2: Annualized relative and absolute volatility associated with the different maturities of the rebar contracts. Contract one has from one month to 0 days maturity as time passes, contract two has two to one month maturity and so on.
Relative: $\sigma_{ann} = \text{std.dev}\left[\ln(F_t/F_{t-1})\right] \times \sqrt{52}$.
Absolute: $\gamma_{ann} = \text{std.dev}[F_t - F_{t-1}] \times \sqrt{52}$



Contract	Annualized Volatility	
	Relative	Absolute
1	23.83 %	794 Yuan
2	20.29 %	728 Yuan
3	18.88 %	705 Yuan
4	17.86 %	676 Yuan
5	18.25 %	685 Yuan
6	18.54 %	708 Yuan
7	19.08 %	733 Yuan
8	18.12 %	691 Yuan
9	18.52 %	709 Yuan
10	18.31 %	708 Yuan

Figure 3: Relative volatility through time for $F_{0,1}$ and $F_{0,10}$. The horizontal bands represents one standard deviation on a weekly frequency.

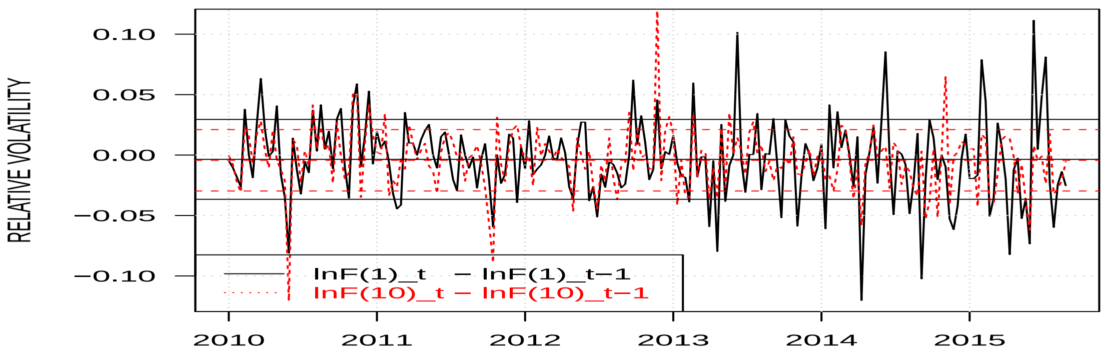
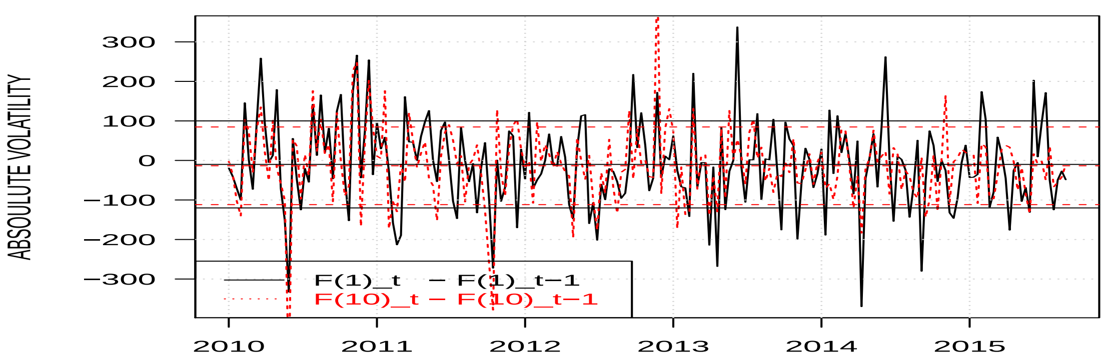
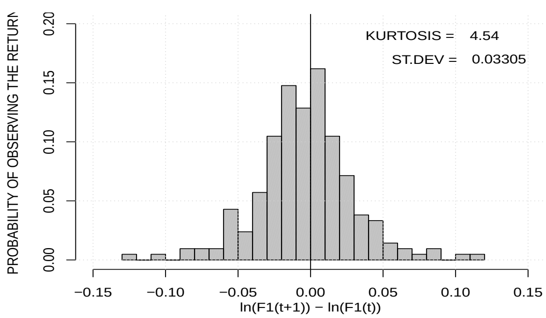


Figure 4: Absolute volatility through time for $F_{0,1}$ and $F_{0,10}$. The horizontal bands represents one standard deviation on a weekly frequency.

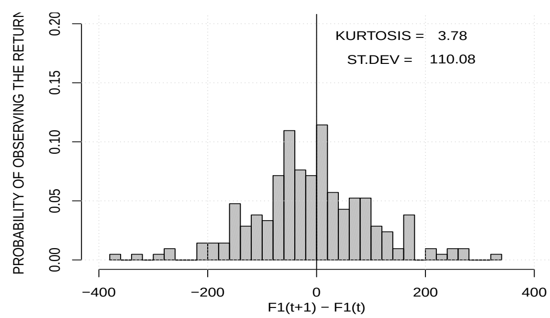


The kurtosis associated with the distribution of returns for commodities is typically[16] above 3, meaning that it has fatter tails than the normal distribution. The normal distribution is nevertheless often, and will also here be, the main tool in modelling the stochastic price process. Often, a separate jump process is added in models for prices that jumps, rather than widening the distribution. The distribution of log returns in Figure 6(a) appears normal and it is not obvious that the absolute changes are log-normally distributed in Figure 6(b). The distribution appear to be wider in Figure 6(b).

Figure 5: The computed kurtosis shows that the distributions has fatter tails than the normal distribution, but is considerably less than what Geman (2005) reports. See footnote 16.



(a) Distribution of the percentage return of the rebar price.



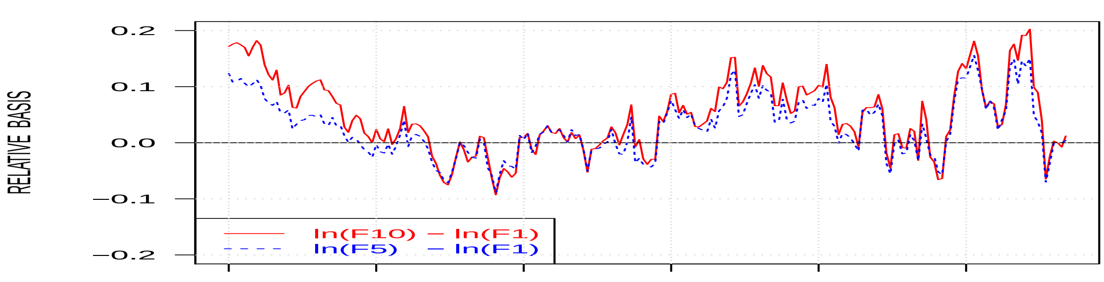
(b) Distribution of the absolute change of the rebar price.

[16] Geman (2005), page 59 calculates kurtosis, from July 1993 to November 2000, here written compact with no decimals: Crude oil 6, Brent 6, Natural gas 30, Heating fuel 11, Unleaded gasoline 4, Corn 51, Soybeans 19, Soymeal 15, Soy oil 5, Wheat 59, Oats 23, Coffee 7, Aluminium 6, Copper 7, Zinc 12, Nickel 5, Tin 6 Lead 6.

Term structure of the futures prices Basis is defined as a longer maturity contract minus a shorter maturity contract (or the spot price), e.g. $F_{t,10} - F_{t,1}$. Define the relative basis as the difference in the log-prices. The relative basis is displayed in Figure 6 for $\ln(F_{t,10}/F_{t,1})$ and $\ln(F_{t,5}/F_{t,1})$. A positive value is known as "contango" and a negative value is known as "backwardation". Define the futures term structure as the curve observed when the prices observed at date t for futures contracts maturing at different dates are plotted with the maturity of the contract on the horizontal axis and the price on the vertical axis. Contango is lingo to describe a situation where the futures term structure is upwards sloping, and backwardation is a name for when the slope is neqative. The term structure of the futures prices can possibly add information on price dynamics.

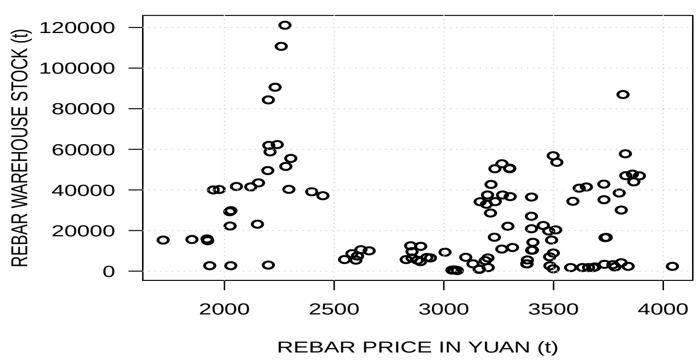
There are cost-of-carry associated with possessing steel physically, which in the absence of other effects make the futures curve slope upwards. Backwardation can be regarded as a situation where agents are unwilling to delay consumption, or a situation where sellers give up a portion of the price to achieve certainty of the future price instead. I have included $\ln(F_{t,5}/F_{t,1})$ in Figure 6 to shed light on the shape of the futures curve. The futures curve can be hump shaped and need not be monotonically increasing or monotonically decreasing as the graph hints towards. The graph hints towards monotonically increasing term structure because $\ln(F_{t,5}/F_{t,1})$ is for the most part below $\ln(F_{t,10}/F_{t,1})$ in the figure. The number of cases in which both the shorter maturity and longer maturity contracts were more expensive than the middle contract is as follows, F2: 27, F3: 28 F4: 34, F5: 45, F6: 41, F7: 40, F8: 35, F9: 34. Per contract series are there 211 observations.

Figure 6: Relative basis. January 2010 - September 2015. Positive: Contango. Negative: Backwardation. Relative basis can also be regarded as a measure of "cost-of-carry".

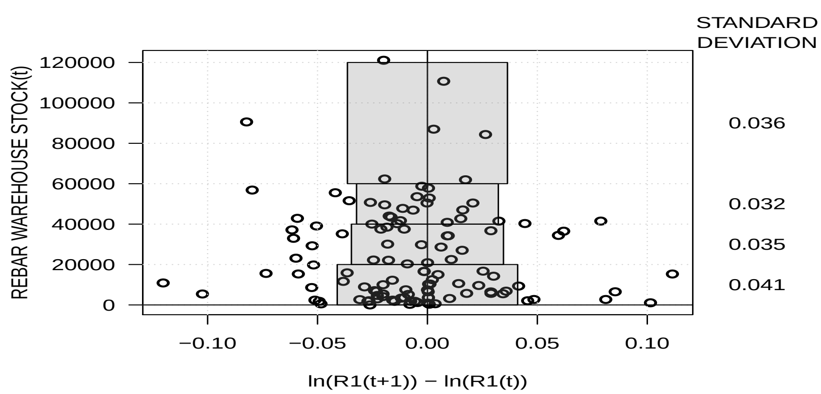


Inventories The theory of storage summarized: Low warehouse stock leads to higher nearby futures prices and higher volatility. The price versus stock at the warehouse is plotted in Figure 7. A linear regression equation to estimate the fit does not improve casual inspection in this case. The data do not confirm a relationship as predicted by the theory of storage. Volatility is somewhat higher for lower inventory levels, seen in Figure 8(b). The obvious shortcoming in the data is that it is for the warehouses associated with the SHFE only. Buyers and sellers can easily stock steel at their sites, in addition to deliver to the exchanges, so producers and consumers easily influence observable inventory level analysed here. The expectation of a relationship vanish as a result of the shortcoming. The lack of relation is demonstrated here because inventory data could easily come to mind as a possible source of information for modelling prices.

Figure 7: Rebar warehouse exchange stock versus price level and volatility.



(a) The relationship in levels is weak with a correlation of - 0.2.



(b) Percentage change in price over a week given warehouse stock in t. One standard deviation indicated by grey boxes. Sample period: July 2012 to September 2015.

Part III:

Modelling I: General Considerations

6 General Modelling Considerations

Computing a present value.

The present value of a certain future cash flow takes the time value of money into account by discounting with a risk-free interest rate. When there are associated uncertainty, the present value is affected. To compute the present value, assumptions must be made, often via an explicit model. A common way is to model the uncertainty going forward with real probabilities and then add a premium — reflecting the risk associated with the expected outcome — to the risk-free rate when discounting the future value back to a present value. An alternative way is to model the uncertainty going forward with artificial probabilities and then use only the risk-free rate to discount back into a present value. This is referred to as “risk-neutral valuation”. Both approaches involves setting up a model reflecting the relevant economy and use parameter values in the computation. I will use risk-neutral valuation. Risk-neutral valuation is explained in Appendix: Deriving the Risk-Adjusted Process on page 72.

Model of uncertainty.

The simplest model of uncertainty and time is one in which a variable takes on a value today and can only take on two different values in the next time period — a binomial model. This specification is reasonable to model a coin flip, but not to model the steel price, because the steel price can take on more than two values and uncertainty is considered over more than one period ahead. To model the steel price, the binomial model can be extended by adding the binomial structure to it self for each of the two possible outcomes, resulting in a “tree”. Doing this n times results in a tree with $n+1$ different possible outcomes if the tree does not keep track the specific path taken through the tree. In this discrete model, a reasonable fine outcome grid for a fixed future time period can be constructed by increasing n. To be completely precise in the modelling, one can construct a mathematical object that moves continuously through time and can take on a continuous set of values. The exchange has trading hours, so the price is not traded continuously. The price quotations are not continuous either, rather they are quoted in fixed intervals, e.g. yuan per ton. The continuous time can be thought of as what the price would have been, if it were traded in the closing hours. The continuous outcome space will ‘fit’ every discrete version, and is easier to scale in time while preserving precision. The benefit of continuous-time, continuous-set of values lies in the precision of the model and importantly, not in the economic principles and intuition.

Stochastic differential equation (SDE) to model uncertainty and time

I will consider SDE's that consists of a drift term and a dispersion term.

$$dX_t = \mu(t, X_t) \times dt + \sigma(t, X_t) \times dW_T$$
$$\text{change in } X_T = (\text{drift function}) \times dt + (\text{dispersion function}) \times \text{randomness}$$

Drift takes care of the deterministic part of the change, while the dispersion term takes care of stochastic part. The stochastic 'motor' is the Wiener process. Both the drift and dispersion can be functions of the stochastic variable and time. Hull (2015) is sufficient to be able to work with SDE's, Øksendal (1998) is more rigorous. When dealing with SDE's normal calculus does not apply, rather Itô calculus must be applied. I have stated Itô's formula in Appendix: Itô's Formula on page 56 because it is used later.

Modelling implications from economic theory

Existing producers have the possibility to regulate production and supply in response to the steel price, thus the quantity in the market should be a function of the price. The price should be driven down by producers when the price is high. When the price is low, supply is expected to decrease and some firms forced out of business if the low price persists — eventually the price moves up again. Demand is also a source of mean-reversion, because a high price makes it more attractive to use substitute materials, decreasing steel demand. Similarly steel is more attractive when it is cheap, increasing demand. This economic reasoning implies that a property of the drift function should be that it makes sure that the process mean-reverts.

The fundamental value of the steel price is likely to change with the improvements in production technology and prices of input factors. The prices of the input factors, demand and supply are likely a function of the economical environment in a broad sense. Hence, it is not obvious how to model the level for which the process seeks to return to. An alternative is to identify the relationship between some key variables, and use that to orient on whether the current price is above or below what is historically observed as the (cointegrated) relationship. If the interest rate is a proxy for the state of the economy, it could be allowed to vary in the model. The interest rate observed at the Peoples Bank of China has been around 5% from 2010 to 2015. I will use a constant interest rate of 5% throughout my work. Trading it self will induce price fluctuations, so the prices will be an imperfect measure of fundamentals. Using a constant interest rate also simplify away the interest effect for margin payments on futures contracts (adjustment of your cash balance in the broker account on a daily basis to avoid inability to meet your obligations).

I argue that when the purpose of modelling is to characterize dynamics for subsequent simulation, the source of the random movements is not critical as long as the model reflects the characteristics of the price movements. The purpose in this thesis is not to predict the direction of the future movements of the steel price, although a prediction on the direction is an implication of a process that mean-reverts.

7 Particular Modelling Considerations

Yuan versus dollar

The contracts are traded in yuan, while the price series are in dollars. I want to model the steel price process in Shanghai from an independent geographical perspective. Therefore, to avoid measurement noise by dollar-yuan fluctuations I convert the observations back into yuan via daily observations of the currency pair accessed at the webpage of the St.Louis Fed.[17] The rebar prices in dollar and the exchange rate used is seen in Figure 2(b) and 2(c).

Sample period

In the time period considered, the prices declines. One must be careful with the interpretation of the data. In any case the history is just a particular realization of possible outcomes. If the prices did not have any tendency to increase nor decline, it would be tempting to assume a constant simple mean as the equilibrium price. It would possibly be the statistical equilibrium price, i.e. the mean over the period, but the true equilibrium is a function of economic activity and the costs of production. Hence, to evaluate the equilibrium level, I argue that there are two options: (i) Industry knowledge about production costs and predicted demand, (ii) Statistical approach is to consider variables that are likely to be co-integrated with rebar. A few candidates are iron ore, coking coal, steel scrap, electricity, labour, purchasing managers index, interest rates, and unemployment.

[17] https://research.stlouisfed.org/fred2/series/DEXCHUS

Frequency of data

Daily observations on trading days are not equidistant with respect to calendar time, but could be with respect to the flow of relevant information.[18] Daily observations increase the sample size, and there is some evidence that increased frequency for a given sample size increase the power when testing for unit root, but the time span is considered to be more important.[19]

I choose to use a weekly frequency motivated by, (i) no need to distinguish between calendar time and abstract information flow time. Annually expressed time step will be 1/52. (ii) Unit root test are not adversely affected. (iii) The fact that the construction sector plan and act with a long time perspective, and are not interested in micro dynamics. If one wants to investigate micro dynamics, a daily frequency is likely to be too infrequent anyway. (iv) A weekly frequency is common in the literature, e.g. Schwartz (1997).

[18] If observations are equidistant with respect to relevant information, it means that relevant information is not three times as large over a weekend (three nights) than for example Wednesday to Thursday (one night). The arrival of relevant information induces price changes, i.e. volatility. Hence, relevant information flow is tested by comparing volatilities over a night versus over three consecutive nights where there were no trading.

[19] E.g. Page 130 in Maddala and Kim (2004)

Part IV:

Modelling II: One Factor Model

8 One Factor Model

Mean-Reverting Model With Absolute Volatility The economics of steel markets call for mean-reversion in the model, although there is reasonable doubt when eye-balling figure 2(a). In Figure 3 and 4 it is seen that the volatility is not necessarily proportional to the price. When price changes are proportional to the price level, the distrubution of the prices will be log-normal and the percentage return will be normal. I choose the main model to be one where the volatility is absolute.

Based on this do I suggest the mean-reverting Orhnstein-Uhlenbeck process for the futures price level:

$$dF_{t,T} = \kappa\left(\mu - F_{t,T}\right)dt + \gamma dW_t \tag{1}$$

Mu μ is the long run mean of $F_{t,T}$. If $F_{t,T} > \mu$, the deterministic drift term is negative. Because the term is negative when $F_{t,T} > \mu$, and positive when $F_{t,T} < \mu$, the process mean-reverts. The deviation from the mean is scaled by kappa κ. Thus the speed of reversion is determined by κ. This deterministic part of the process is disturbed by a random variable W_t, and scaled by a parameter gamma γ. The random variable W_t is the standard Wiener process.[20]

$F_{t,T}$ can become negative, not consistent with the price of a commodity. If the mean is sufficiently far away from zero and the mean-reversion is strong, this is unlikely.

Discrete version The discrete version of (1) with time step Δt is

$$F_{t+\Delta t,T} - F_{t,T} = \kappa\left(\mu - F_{t,T}\right)\Delta t + \gamma\sqrt{\Delta t}\,\eta_{t+\Delta t}\ , \text{ where } \quad \eta_{t+\Delta t} \sim \eta_t \sim N(0,1) \tag{2}$$

This version could be used to estimate parameters if the analytical solution were difficult to find. If this form is used, the size of the time increment Δt is important. Solving (1) analytically is preferable because the discrete version will be exact for all choices for the size of Δt.[21] When estimating parameters in a later section, the analytical solution will be used and the chosen weekly frequency implies $\Delta t = 1/52$.

[20] Properties of the standard Wiener process: $W_0 = 0$, $W_t - W_s \sim N(0, t-s)$ where $s \leq t$, increments of $W_t - W_{t-1}, W_{t-1} - W_{t-2}, \ldots$ are independent of one another, W_t is the sum of its increments and W_t is continuous but not differentiable.

[21] See Appendix: The (Possible) Role of Δt on page 55 for an illustration of the role of Δt.

Analytical solution The analytical solution is[22]

$$F_{T,T} = F_{t,T}e^{-\kappa T} + \mu\left(1 - e^{-\kappa T}\right) + \gamma\int_t^T e^{\kappa(s-T)}dW_s \tag{3}$$

$F_{T,T}$ is a random variable because of the presence of the stochastic integral. The expectation and variance at t for date T are:

$$E\{F_{T,T}\} = F_{t,T}e^{-\kappa T} + \mu\left(1 - e^{-\kappa T}\right) \tag{4}$$

$$Var\left(F_{T,T}\right) = \gamma^2\int_t^T e^{2\kappa(s-T)}ds \tag{5}$$

Mean-Reverting Model With Relative Volatility This model cannot become negative in the continuous case. One has to take care in an Euler scheme here as well, because the system is not updated over Δt. If the distance Δt is large, the system is not updated frequently enough to sufficiently scale down the dispersion term.

$$dF_{t,T} = \kappa\left(\mu - F_{t,T}\right)dt + \sigma F_{t,T}dW_t \tag{6}$$

[22]See Appendix: Solving the Mean-Reverting Ornstein–Uhlenbeck Process on page 57 for a detailed way to the solution.

9 Estimation, One Factor Model

Estimate an AR(1)

I now turn to Box-Jenkins methodology.

Strategy The discrete version of (3) with a time step Δt is

$$F_{t+\Delta t,T} = F_{t,T}e^{-\kappa\Delta t} + \mu\left(1 - e^{-\kappa\Delta t}\right) + \gamma\sqrt{\frac{1 - e^{-2\kappa\Delta t}}{2\kappa}}\eta_{t+\Delta t}\ , \qquad \eta_{t+\Delta t} \sim N(0,1) \tag{7}$$

AR(1) with non-zero mean:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim N\left(0, \sigma^2\right) \tag{8}$$

The estimation strategy is to equate:

$$y_t = F_{t+\Delta t,T} \tag{9}$$
$$a_0 = \mu\left(1 - e^{-\kappa\Delta t}\right) \tag{10}$$
$$a_1 = e^{-\kappa\Delta t} \tag{11}$$
$$y_{t-1} = F_{t,T} \tag{12}$$
$$sd.err(\varepsilon) = \gamma\sqrt{\frac{1 - e^{-2\kappa\Delta t}}{2\kappa}} \tag{13}$$

and solve for μ, κ and γ using the computed numbers for a_0, a_1 and $sd.err(\varepsilon)$, to obtain:

$$\kappa = \frac{-\ln a_1}{\Delta t} \tag{14}$$
$$\mu = \frac{a_0}{1 - a_1} \tag{15}$$
$$\gamma = sd.err(\varepsilon)\sqrt{\frac{-2\ln a_1}{\Delta t(1 - a_1^2)}} \tag{16}$$

Testing the underlying assumptions A condition for estimating the linear model, is that the process is stationary. To test for stationarity, I employ an Augmented Dickey-Fuller test[23] without drift or trend. The idea is to test the true data generating process,

[23] See appendix on Augmented Dickey Fuller test on page 51

and not fit an equation to the observed sample in the testing. I do not want to allow for a trend or drift in the test, because it is not theoretically compatible with the price of a commodity. In fact, a positive trend reflecting inflation could be present, but in the sample the general tendency is a decreasing price. Although the production process is likely to be cost-improved over time, the price cannot decrease in a deterministic nor in an average fashion forever. Hence, I do not allow for a trend or drift. Interestingly, if the prices had an overall tendency to increase, it would be tempting to reason that it is for example inflation or a stable growth in demand, and include drift in the testing equation. If prices follow mean-reversal along an upward sloping trend it would be reasonable to include a trend.

The ADF test statistic for F1 is -1.16, not even significant at the 10% level[24]. Hence, the process is not proven to be stationary. This means that the numbers $a_0, a_1, sd.err(\varepsilon)$ in (10), (11), (13) cannot be reliably computed and plugged into the expressions for κ, μ, γ in (14), (15), (16).

Hence, either the price reverts slowly towards the mean, or it is a true unit root process. Over a short time horizon a true unit root and a near unit root are similar, and they appear more similar the larger the volatility is. ADF test and other unit root tests are known to have low power, so it is not good in differentiating a near unit root from a true unit root.

When failing to reject unit root In general, when one wants to identify a pattern and predict the future, but fail to reject unit root, one must identify whether detrending or differencing is the proper method to make it stationary, and proceed. Differencing results in working with ARIMA(p,1,q), rather than ARIMA(p,0,q). ARIMA(p,1,0) is on the form $\Delta y_t = \sum_{i=1}^{p} \Delta y_{t-i}$. I can report that a_1 is insignificant in ARIMA(1,1,0), so there is no predictive pattern in such a simple model. Using ARIMA(0,1,0) as the model of the mean, I reject conditional variance by testing the squared forecasting errors with the Ljung-Box test and McLeod Li test.[25] This is true for ARIMA(1,1,0) as well.

To take the first difference is not compatible with (1) and (3), because (1) and (3) describe a change from t to T based on the level in t. The aim is not to predict the future, but to find parameters for (1) and (3), with the restriction to keep it in levels. Now, there are two options. (i) Do not take the difference, but assume unit

[24]Critical values, given sample size and test regression specification without drift or trend are -2.58, -1.95, and -1.62, for significance levels 1%, 5%, and 10% respectively. Test statistics computed are F1: -1.16, F10

[25]See subsection in on page 53 for an informal explanation.

root. (ii) Use discretion on the reversal. This implies setting the speed of reversion κ and long term mean μ to values such that the process mean reverts and are consistent with the failure of rejecting unit root. The low power of the unit root test could be used to argue that a higher value of kappa is also consistent with mean-reversal and failure to reject unit root.

10 Assume Unit Root

Cannot reject unit root for AR(1), so assume that the coefficient a_1 is one. Because $a_1 = 1$, I impose $a_0 = 0$ or else the process has a deterministic trend.

$$y_t = a_0 + y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim N\left(0, \sigma^2\right) \tag{17}$$

$$y_t = y_{t-1} + \varepsilon_t \tag{18}$$

The solution to (18) is[26]

$$y_t = y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \tag{19}$$

This is known as a random walk. Each new realization of the innovation ε_t has a permanent effect on y_{t+s}, so the process is not stationary, but finite because of the given initial value y_0 and terminal time t.

Model implication To equate this with the mean-reverting process described by (1) and (3) implies a drift of zero, by setting $\kappa = 0$. The mean-reversal property vanishes, as already pointed out by the fact that the series is not stationary. The continuous model with is thus

$$dF_{t,T} = \gamma dW_t \tag{20}$$

$$F_{T,T} = F_{t,T} + \gamma \int_t^T dW_s \tag{21}$$

The discrete version of (20) is

$$F_{t+\Delta t,T} = F_{t,T} + \gamma\sqrt{\Delta t}\, \eta_{t+\Delta t} \, , \quad \eta_{t+\Delta t} \sim \eta_t \sim N(0,1) \tag{22}$$

[26]See Appendix: Solve AR(1), $a_0 = 0$ and $a_1 = 1$ on page 54 for the derivation.

The discrete version of (21) is

$$F_{T,T} = F_{t,T} + \gamma\sqrt{(T-t)}\,\eta_T\ , \qquad \eta_T \sim \eta_t \sim N(0,1) \tag{23}$$

(22) and (23) are similar because they are now nothing more than an initial condition plus a scaled draw from the standard normal distribution. They are analoge to (19), the difference is that $\varepsilon_t \sim N\left(0,\sigma^2\right)$ and $\eta_t \sim N\left(0,1\right)$, so γ is needed in (22) and (23) to take care of the scaling of the volatility.

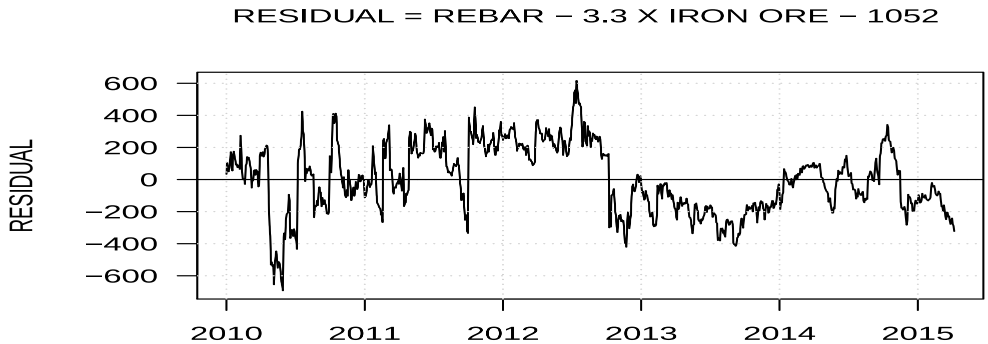
The absoulte annualized volatility γ is computed in Figure 2 to be 794 yuan. The last observation in the sample is 1900 yuan, so the lower bound in a two standard deviation confidence interval equals $1900 - 2 \times 794 = 312$. It is clear that this model can take on negative values. Negative values will not be observed in reality, so this a drawback with this model.

11 Use Discretion On The Drift

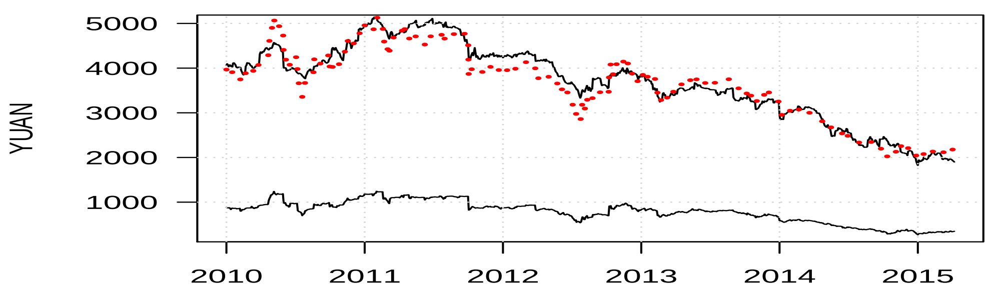
Before even discussing the speed of reversal to the long run mean level, the long run mean level must be determined. It can be constant or time-varying, and the relevant time horizon must be determined. The mean can also be constant or time-varying based on industry and production knowledge.

The average price of $F1$ from January 2010 to September 2015 is 3638 yuan. The average for September 2014 to September 2015 is 2295 yuan. Does it matter what the price were 5 years ago? These are clearly discretionary considerations. Iron ore (also integrated of order one) and rebar are cointegrated. They are cointegrated because the residuals u_t, from a linear regression of the type $\text{Rebar}_t = a + b \times \text{Iron ore}_t + u_t$ are stationary. If the residuals are not stationary, then there are no a and b to transform from one variable to the other that will be correct in expectation. The last observation for iron ore is 347 yuan, and 1900 yuan for rebar. The OLS estimate and these values gives $1900 - 1052 + 3.3 \times 347 = u$, $u = -297$. Because the error is negative, it means that they are closer than they have been historically, and they should therefore diverge.

Figure 8: Iron ore and rebar are cointegrated, because the residuals from a regression is stationary. It implies that OLS is valid, and that this relationship is the average transformation from one to the other.



(a) The residual from a linear regression. When it is negative, it means that they are closer than they have been on average historically, and therefore should diverge in the future.



(b) The dotted line is the prediction of rebar based on iron ore. At the end of the sample, it predict a higher value for rebar than the realized value. The relation in the regression could have been specified the other way around, because there are no explainatory intentions, just a linear relationship. One must remember that there are no information on which variable that is going to move in a particular direction. The information here is that they will diverge, but one can only use this information in a relative sense, not to undertake level predictions.

12 Convenience Yield

Another possible source of information is the term structure of futures prices. A simple relation between two futures prices is[27]

$$F_{t,T} = F_{t,s}e^{r(T-s)} \quad , \quad t < s < T. \tag{24}$$

This relation states that you move through the term structure of futures prices vià the constant interest rate r and the time distance $T - t$. This is violated empirically. It is seen in Figure 2(a) that the distance between the prices do not obey this relation. Introducing a varying *convenience yield* δ_t (expressed as a rate) is the standard way to treat the empirical discrepancy between the futures prices (e.g. Schwartz (1997), Cassusus and Collin-Dufresne (2005)).

Definition of convenience yield, Schwartz (1997) page 927, footnote 9,

> The convenience yield can be interpreted as the flow of services accruing to the holder of the spot commodity but not to the owner of a futures contract.

With deterministic convenience yield, (24) is modified to

$$F_{t,T} = F_{t,s}e^{(r-\delta_{s,T})(T-s)} \quad , \quad t < s < T. \tag{25}$$

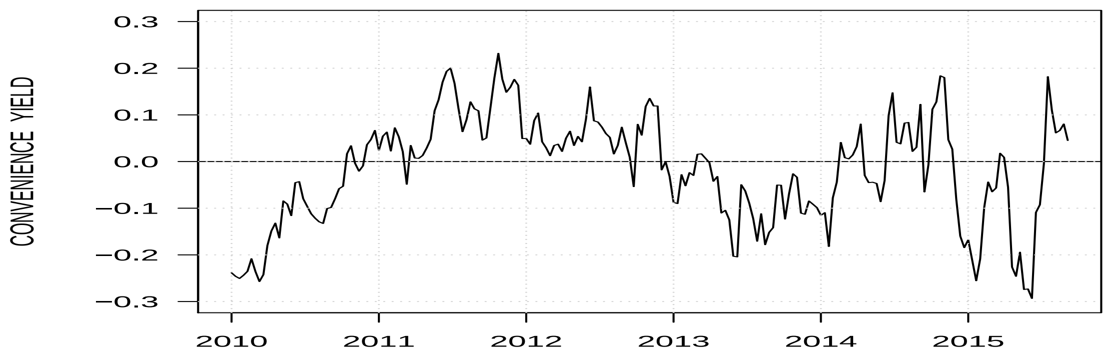
The convenience yield is analogue to a dividend yield (rate) on a stock. If the convenience yield is negative, it is sometimes referred to as *cost-of-carry* — the cost of storing the commodity is paid by the party possessing the commodity and not the holder of the contract. I solve for the convenience yield in (25) to obtain

$$\delta_{t,s,T} = r - \left(\frac{1}{T-s}\right) ln\left(\frac{F_{t,T}}{F_{t,s}}\right) \quad , \quad t < s < T$$

and use the particular contracts $F1$ and $F10$ to plot the empirical convenience yield as a residual in Figure 9. The convenience yield is stationary (ADF-test statistic is -3.95).

[27] See the paragraph preceeding (60) on page 79 for why the relation is as it is.

Figure 9: Convenience yield. Annualized convenience yield as a residual $\delta_{ANN} = \frac{4}{3}\delta_{1,10}$ where $\delta_{1,10} = 0.05 - \frac{4}{3}\, ln\left(\frac{F_{0,10}}{F_{0,1}}\right)$, obtained from the commonly assumed relation between futures prices: $F_{t,T} = F_{t,s}e^{(r-\delta_{s,T})(T-s)} \quad , \quad t < s < T$. January 2010 - September 2015.



13 Summary One Factor Model

Mean-reversal in the drift term is not supported by the employed statistical ADF-test to check for stationarity. The test is known to be weak in disentangling near unit root and true unit root processes. The importance of the distinction decreases with a decreasing horizon. I have shown the implications for the model under Assume Unit Root on page 28. If one really wants to incorporate a drift that is not zero, I have suggested discretionary ways. Possibilities are, setting kappa directly, using cointegration to update the changing mean level and use convenience yield that depends on the spot price.

Part V:

Modelling III: Two Factor Model

14 Motivation for Stochastic Convenience Yield

As the two futures contracts in (25) are stochastic, it is interesting to allow for it in the model. Then at least r or δ must be allowed to vary. Schwartz (1997) and Collin-Duffresne (2005) shows how a mean-reverting convenience yield significantly matters for pricing of contingent claims.

According to the theory of storage, is the convenience yield related to the economy wide inventory levels. When the inventory is low, rebar is scarce so the price and the volatility should rise. The volatility rises because the price sensitivity to changes in demand when the inventory is low, is larger. Furthermore, the importance of low inventory is more important for the immediate future than for a longer horizon. The flow of information should in general be more important for the immediate future. Hence, the nearby futures prices should have a higher volatility than the longer maturity futures.

15 Modelling an Unobservable Variable

I now estimate a model for the unobservable *spot* price, by taking the whole futures curve into consideration. Because the underlying variable is unobservable, it must first be estimated. To do this the log futures prices are put in a state-space form so that the Kalman-filter[28] and Maximum Likelihood Estimation[29] can be used to estimate the time series of the spot price and the associated model parameters.

Schwartz (1997) adds noise to his data to reflect bid-ask spreads, and non-simultaneously observed variables, price limits or errors in the data and then use the Kalman-filter to filter out this noise. The aim is that the serial correlation and cross-correlation is a result of variation in the spot price and convenience yield. I do not add noise to my prices.

[28]See Appendix: Kalman Filter on page 68 for an informal explanation.
[29]See Appendix: Maximum Likelihood Estimation on page 69

16 Theoretical Model

In this part I will impose the two-factor model (26) (27) (28) developed in Schwartz and Gibson (1990) and extended in Schwartz (1997) on the rebar futures data. The rebar spot price follows a mean-reverting geometric Brownian motion. The convenience yield is a variant of the mean-reverting Ornstein-Uhlenbeck process, similar in structure as (1). There are now two sources of randomness driving the two SDE's, with correlation ρ.

Historical measure:

$$dS_t = (\mu - \delta_t) S_t \, dt + \sigma_S S_t \, dW_t^S \tag{26}$$

$$d\delta_t = \kappa \left(\alpha - \delta_t\right) dt + \sigma_\delta \, dW_t^\delta \tag{27}$$

$$dW_t^\delta dW_t^S = \rho dt \tag{28}$$

The model under the *historical measure* is the version used for scenario simulation.[30] There is an modified version of this used to compute present values — the risk-neutral valuation tool.[31]. I present the modified model used for pricing here for completeness. It is then clear where the risk-adjustment λ_δ for the non-traded risk factor convenience yield seen in the estimations belong. The model under the pricing measure is represented by (29) (30) (31).

Pricing measure:

$$dS_t = (r - \delta_t) S_t dt + \sigma_S S_t \, d\tilde{W}_t^S \tag{29}$$

$$d\delta_t = \left[\kappa \left(\alpha - \delta_t\right) - \lambda_\delta \sigma_\delta\right] dt + \sigma_\delta \, d\tilde{W}_t^\delta \tag{30}$$

$$\tilde{W}_t^\delta \tilde{W}_t^S = \rho \, dt \tag{31}$$

17 Estimation Strategy

To fit the Two-Factor Model to the rebar data I have used the function `fit.schwartz2f` in the package `Schwartz97` avaiable at the Comprehensive R Archive Network (CRAN). It employs the Kalman filter technique to estimate the state variable—the unobservable spot price, which then enables Maximum Likelihood Estimation in order to find optimal parameters.

[30] See https://cran.r-project.org/web/packages/schwartz97/vignettes/TechnicalDocument.pdf for details.

[31] The risk-neutral valuation tool with the one-factor model is explained in Appendix: Deriving the Risk-Adjusted Process on page 72.

To the optimization function I provide what I will refer to as “settings”:

- A matrix of weekly observations of one to ten month futures contracts, and specify the time increment to be 1/52.
- A matrix of days to maturity for each contract
- Which parameters to be optimized and which to be held constant
- Initial values of the parameters
- A constant interest rate of 5%
- All maturities receive the same weight in the estimation of the spot price.

. The estimation procedure is unfortunately not as simple as one function call with the “settings” above. The output from the optimization function is sensitive to initial values. Further, there exists local paths where two parameter values can increase pairwise to values not desirable from a modelling perspective (e.g. $\sigma_\delta = \text{sigmaE} = 6 \times 10^{153}$ is not what I want to plug into my model). Therefore, to test the robustness of the estimates do I need to vary the settings in a number of ways. In addition to vary the initial values, it is recommended to vary which parameters to be fixed and which to be optimized. A more detailed description with illustrations and a comparison of different parameter-sets are in Appendix: Two-Factor Model Robustness test on page 59.

18 Estimation Results

In Table 1, I present the results of the estimations. The results presented are based on a corrected mean of the values in the filtrated matrix shown in Figure 12 and 13. The filtrated matrix consists only of vectors of parameter values where all parameters are estimated to be within discretionary filter limits.

Table 1: Estimated parameters for the Two-Factor Model. *Filtering:* Numbers are based on a filtered list of vectors. For each vector in a list of returned vectors from a parameter optimization function, only returned vectors where all parameter values are inside its respective limits are stored in the filtered list of returned vectors. *Correcting:* The mean is computed over the observations where the parameter were in fact free to vary, i.e. the constant parts seen in figure 12 and 13 are not included in the computation of the mean. Vectors with some parameters constant are included in the filtered list of vectors because the other parameters vary, and contribute with observations to compute their means.

Parameters		Mean	Std.dev	Filter Limits Low	Filter Limits High
mu	μ	0.1899	0.1704	0.001	1
sigmaS	σ_S	0.3421	0.134	0.001	5
kappaE	κ	12.4609	4.997	5	30
alpha	α	0.2995	0.1366	0.001	5
sigmaE	σ_δ	2.1907	1.1217	0.001	5
rhoSE	$\rho_{S\delta}$	0.9886	0.0124	0.9	1
lambdaE	λ_δ	3.7291	1.7689	0.001	10

Some estimated parameter values are strange. The correlation of 98% is very high. Also the drift in the real world mu is estimated to be 19%. A relative volatility of 34% is reasonable, but it does not agree with the relative volatility calculated in the one-factor model. By setting up an Euler scheme it is easy to investigate properties of the model. The stationary level to which the convenience yield converts to is -0.35 under the pricing measure. This is means that the net convenience yield is negative. This is not surprising because the term structure of the futures curve is upwards sloping for the most part. The level of kappa implies that if the convenience yield is either 0 or -0.7, it takes two months before it is close to -0.35 (-0.39 and -0.31) and three months to reach -0.36 and -0.34. While the negative sign on the convenience yield long run level is expected, the level is low.

Figure 10: Simulated convenience yield over 26 weeks. The convenience yield has a much higher volatility than observed in the real world data in Figure 9.

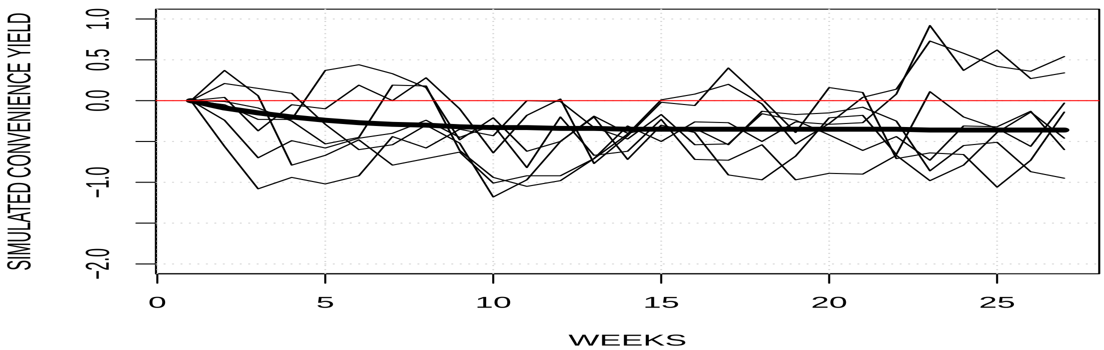


Figure 11: Simulated steel price under the pricing measure. The positive trend in geometric brownian motion is clear.

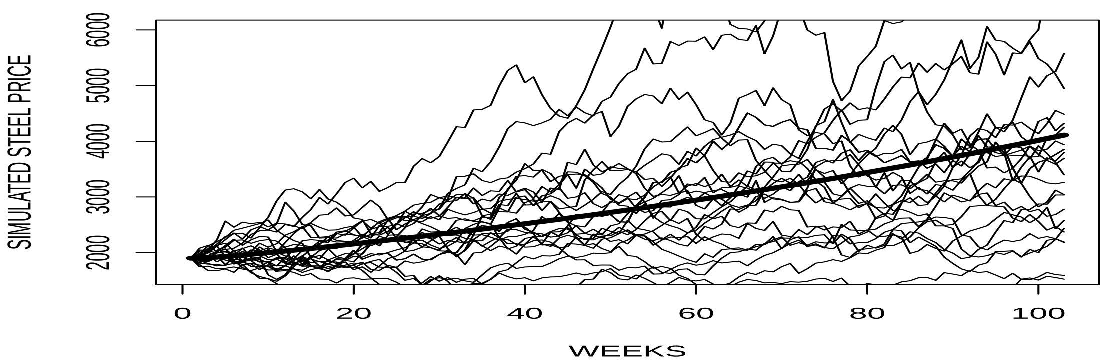
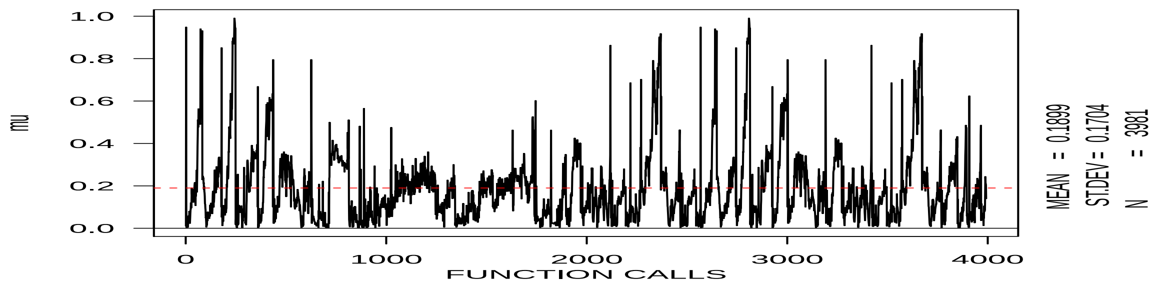
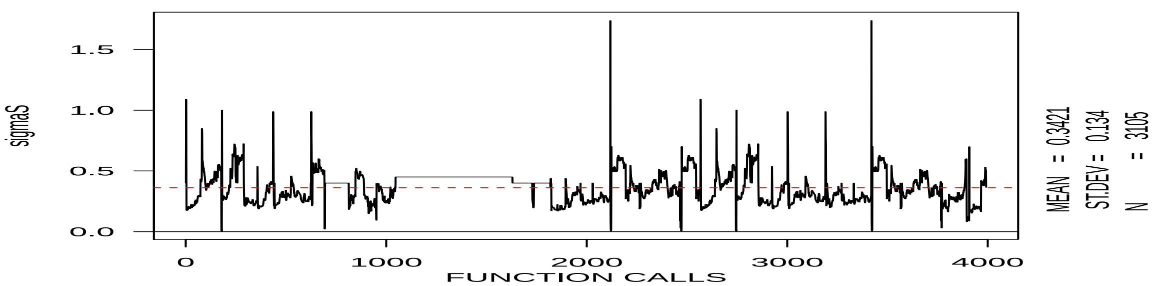


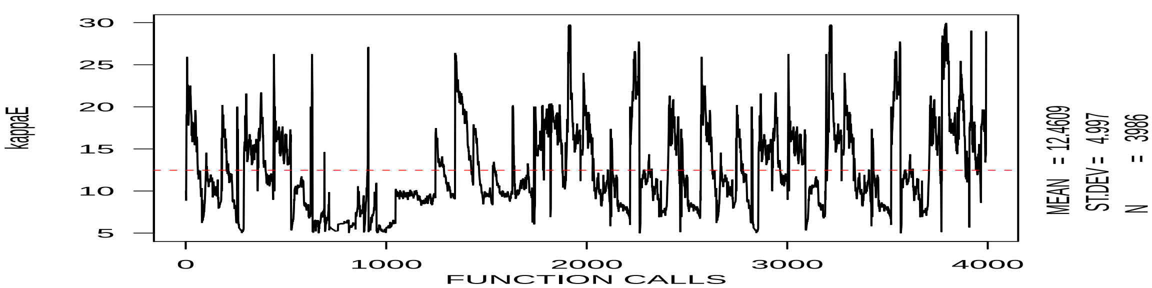
Figure 12: The filtrated matrix of acceptable values of μ, σ_s, κ. The mean is computed over the observations where the parameter were in fact free to vary, i.e. the constant parts are not included in the computation of the mean. N is the sample size for which the mean and standard deviation are computed.



(a) μ

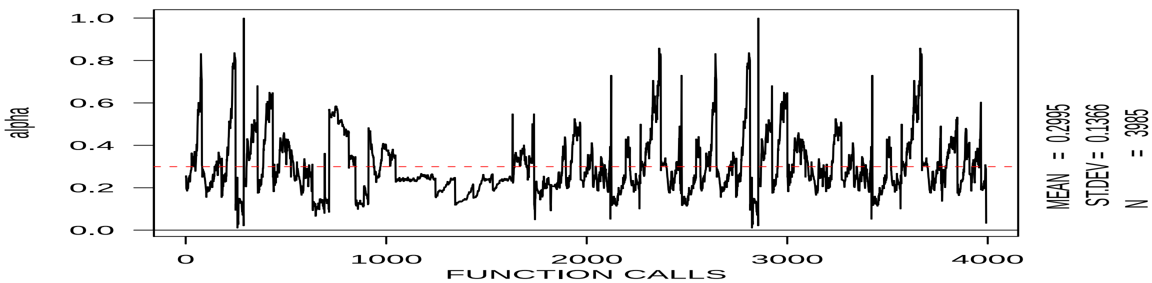


(b) σ_s

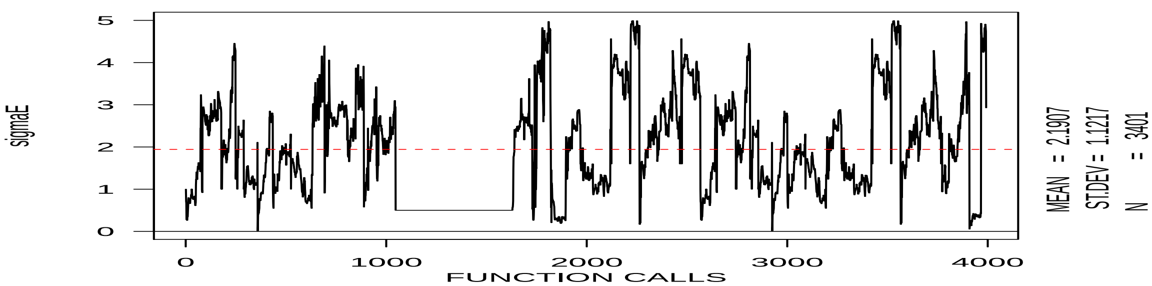


(c) κ

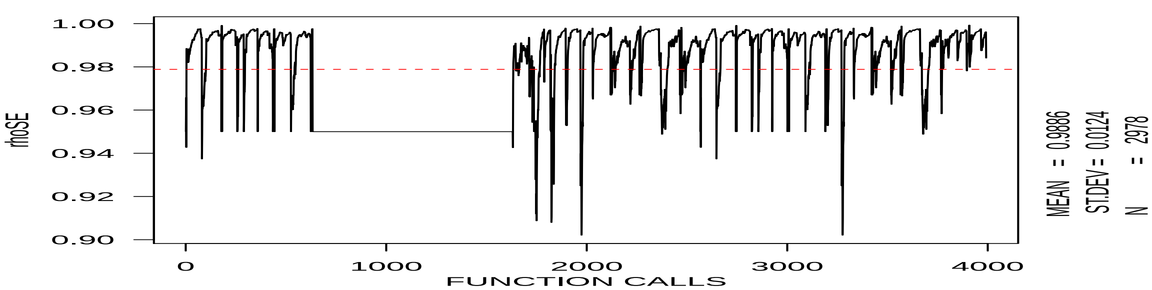
Figure 13: The filtrated matrix of acceptable values of $\alpha, \sigma_\delta, \rho_{SE}, \lambda_\delta$. The mean is computed over the observations where the parameter were in fact free to vary, i.e. the constant parts are not included in the computation of the mean. N is the sample size for which the mean and standaard deviation are computed.



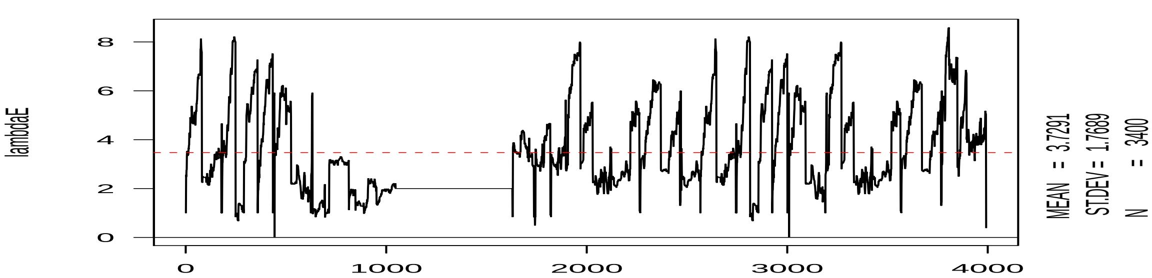
(a) α



(b) σ_δ



(c) ρ_{SE}



(d) λ_δ

Part VI:

Application

19 Application

For the risk-neutral valuation framework, see Appendix: Deriving the Risk-Adjusted Process on page 72.

Real Options At the date a futures contract has delivery, it is effectively a spot price. For an option maturing at this date, it will be the same whether you evaluate an option on the futures or the spot. If they do not expire on the same date, but lets say by a two week difference, one can treat the option on the futures contract as a stock paying a dividend yield and, set the yield to the risk-free rate. This reflects the delayed payment of the futures. Alternatively, if the option expires sufficiently far in the future, e.g. two years, there will be other sources of uncertainty more important than this small detail. E.g. constant interest rate over the two years. The trick with real option valuation is the translation of the real world problem into a derivative interpretation, i.e. a cash flow function must be specified. The option to expand production is a call option on the underlying. If there is a continuous decision considerations in reality, it will be an American option rather than European, complicating the valuation. Also, one must remember that risk-neutral valuation relies on the assumption that the prices reflects risk preferences. For steel, the futures contracts are traded in well functioning markets, the exchange, so I rely on the risk-correction using this technique. In comparing the models, examples are really not necessary, because it is known that a greater volatility increases the option value. Hence, a model which mean-reverts under the pricing measure will yield lower prices than the model with unbounded variance. When the volatility is absolute, the volatility gives the option a higher value the lower the spot price. I will here present an example to illustrate how real option pricing theory can be used, rather than to highlight key differences between the models.

Scenario Simulation For scenario simulation the 'real world dynamics' are used. Complicated cases can be evaluated based on how the single processes ends up. These processes yield a distribution for the many paths simulated, thereby giving a measure on the uncertainty of the future outcomes. Here, caution has to be paid to whether there are correlations between the variables that should be taken into account in the simulation.

Option to expand production

The numbers are inspired by accounting figures from a large Chinese steel maker. Consider a steel manufacturer who have predicted that optimal production level is $Q_1 = 46$ million metric tonnes in two years. $r = 5\%$. The current steel price is 1900 yuan. They have agreed on terms of delivery with its suppliers, fixing the unit input costs to 0.9 of the current steel price plus. In addition 6,000 million yuan covers plants, financial capital and administration expenses. This amounts to $1,900 \times 0.9 \times 46m + 6,000m = 84,660m$ yuan. The present value of this project if there is no convenience yield is like a forward contract. In particular, if the project where inside 10 months, the expected steel price under the pricing measure would be observable at the exchange. For a two year horizon the present value is

$$
\begin{aligned}
&PV(project) = e^{-r(T-t)}(Q_1 E\{S_T\} - F_1) = e^{-0.1}(46m \times 1900e^{0.1} - 84,660m) \\
&PV(project) = 10,796m \quad \text{yuan}
\end{aligned}
$$

Now introduce the option to produce additional 10 million metric tonnes in two years. The suppliers of scrap and iron ore do not demand a premium for 'giving' the steel maker the *right to require* delivery in two years, so the agreed price is 0.9 of the current steel price. Because the production plant has a 60m metric tonnes capacity, the existing plant can absorb this increase in production. $0.90 \times 1900 = 1710$ yuan per metric ton will be the lowest price in which it is worth undertaking the production expansion. This is a call option on the steel price. Use (57) for a call with this particular model

$$
\begin{aligned}
PV(expand) &= e^{-r(T-t)} Q_2 \mathbb{E}_t^Q\{max(S_T - F, 0)\} \\
&= e^{-r(T-t)} Q_2 \left(\mathbb{E}_t^Q \left\{ 1_{\{S_T \geq K\}} S_T \right\} - K \mathbb{E}_t^Q \left\{ 1_{\{S_T \geq K\}} \right\} \right) \\
&= e^{-r(T-t)} Q_2 \left(mN\left(\frac{m-K}{v}\right) + \frac{v}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{K-m}{v}\right)^2} - KN\left(\frac{m-K}{v}\right) \right) \\
&= 526 \text{ yuan} \times Q_2 = 5,260m \text{ yuan}
\end{aligned}
$$

As a reference, Black-Scholes with the same arguments used except for proportional volatility, calculated for the F1 rebar contracts to be 23.83%, gives a present value of the expansion option equal to 445 yuan $\times Q_2$ = 4,450m yuan. The Black-Scholes formula becomes relevant in this case because the risk-adjusted dynamics get the drift of the numéraire. The mean-reverting drift in the relative volatility model differentiated it from geometric brownnian motion. The Black-Scholes formula applies when the process (6) is assumed and developed under the Q measure.

The Two-Factor model is calculated to have a relative volatility of 34.21%, considerably higher than 23.83%, but there are now additional parameters which could work in the opposite direction (reduce value). I have set up a Euler scheme and find that Kappa is a key driver of the option value. Rho is not. Kappa is estimated to be around 12 with a standard deviation of 5. PV when kappa = 2: 832,000. Kappa=12: 224. Kappa=22: 697. The role of the speed of mean-reversion in the convenience yield has an extreme effect on the value. I do not find the estimation results of the two-factor model appealing. The takeaway is that when modelling the drift — and therefore the expectation — under the pricing measure, it matters what you do, so do it properly. Schwartz (1997) shows that convenience yield matters, Casassus and Collin-Dufresne (2005) shows that a varying risk premium also matters. That is, lambda matters. This is now clear, because it is to directly modify the drift of the risk-adjusted underlying.

So what are the options to do something worth compared to being locked in to doing something similar? Value the project without options in it, but as a futures contract

$$PV(LargeProject) = e^{-r(T-t)}(Q_3 E\{S_T\} - F_3) = e^{-0.1}(56m \times 1900e^{0.1} - 101,760m)$$
$$PV(LargeProject) = 14,324m \quad \text{yuan}$$

The conclusion is that the *option* to produce 10,000m metric tonnes in additon to the commitment to produce 46,000m metric tonnes, rather than commit to produce 56,000m metric tonnes right away, has a value of $10,796m+5,260m-14,324m = 1732m$ yuan.

20 Discussion

In short, a model can be precise in it self, but after all, discretion has to be shown somewhere, because the model aims at describing the real world. Proportional volatility or absolute? Mean-reversion under the pricing measure? Stochastic mean-reversion? What is gained by complicating the model, making it harder to get an overview of the implications of various arguments (just test it with an Euler scheme), more difficult to solve, and more parameters to be estimated. The risk of model misspecification is present with both a parsimonious and a large model, but with a large model there are more uncertainty in total because there are more parameters to be estimated. Moreover, there can be interactions between the parameters, such that uncertainty does not simply add up linearly. There are many different specifications on how the volatility relates to the level of the underlying, how the process mean-reverts and processes that jumps. The models considered here are simple ones compared to extensions in the literature. When determining what matters one should be motivated by what is economically important. Also one must judge what is relevant, concerning both what to include as discussed, but also when selecting the sampling period.

Part VII:

Conclusion

21 Conclusion

In identifying and calibrating a model for the steel price I find that there are unit root in the steel price, making scenario simulation a discretionary exercise if one insists on mean-reversion in the dynamics. For present value computations, the specification of the drift is unimportant. The two versions of the one-factor model identified are simple, but well calibrated. They succeed to describe time and uncertainty. I then try to extend the model as recommended by Schwartz (1997), without being completely successful in arriving at a model that can be used to soundly model time and uncertainty. The results from the two-factor model calibration is that the two sources of randomness are highly correlated, rising the question whether they really are two different sources of randomness. The difficulties in parameter estimation are highlighted in the attempt to calibrate the model. The two-factor model is for the unobservable spot price, whereas I have used the nearest futures price directly in the one-factor model as the steel price. I do not reject the two-factor model. I conclude that I would need more time to calibrate the model to incorporate the desirable pricing implications of the mean-reversion in the risk-adjusted process. Mean-reversion bounds the variance and thereby reduce the value of asymmetric derivatives. It also makes the two-factor model more appealing than the simple one-factor model the longer the horizon of the valuation is.

Part VIII:

References

22 References

Casassus, Jaime, and Pierre Collin-Dufresne. "Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates." *The Journal of Finance*, 2005, 2283-331.

Enders, Walter. *Applied Econometric Time Series.* 3rd ed. Hoboken, NJ: Wiley, 2010.

Geman, Hélyette. *Commodities and Commodity Derivatives: Modelling and Pricing for Agriculturals, Metals, and Energy.* West Sussex: John Wiley & Sons, 2005.

Gibson, Rajna, and Eduardo S. Schwartz. "Stochastic Convenience Yield and the Pricing of Oil Contingent Claims." *The Journal of Finance:* 959.

Hamilton, James D. "Chapter 50 State-space Models." *Handbook of Econometrics:* 3039-080.

Hull, John. *Option, Futures, and Other Derivatives.* 9th ed. Boston MA [etc.: Pearson, 2015.

Maddala, Gangadharrao S., and In Kim. *Unit Roots, Cointegration and Structural Change.* 1. Publ. ed. Cambridge: Cambridge Univ. Press, 1998.

Schwartz, Eduardo S. "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging." *The Journal of Finance:* 923.

Øksendal, B. K. *Stochastic Differential Equations: An Introduction with Applications.* 5th ed. Berlin: Springer, 1998.

Part IX:

APPENDICES

23 Appendix: ADF-test

The problem Consider the AR(p) model,

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \varepsilon_t \quad , \quad \varepsilon_t \sim N\left(0, \sigma^2\right). \tag{32}$$

The problem is that if the coefficients in the true data generating process (DGP) sum to one, the series exhibit unbounded variance.[32] Unbounded variance is a violation of one ordinary least squares (OLS) assumption, so using OLS to model the true DGP will not yield reliable results.

Unit root. If the sum of the coefficients sum to one, at least one characteristic root is unity, referred to as *unit root*.

Consider the simpler case of eq. (32) with $p = 1$,

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \tag{33}$$

here, unit root corresponds with $\alpha_1 = 1$ so eq. (33) becomes

$$y_t = y_{t-1} + \varepsilon_t \tag{34}$$

In (34) new realizations of ε_t, called innovations, have a *permanent* effect on all future values of the series. This is why the variance is unbounded. In equation (34) the expected value of the series for an arbitrary future date is the current level of the series. It implies that the history of the sequence is not useful in predicting the future direction of the sequence. The contrary case is if $\alpha_1 \neq 1$. If $|\alpha_1| > 1$ the process will explode. If $|\alpha_1| < 1$ the process will revert to a long term mean. When $|\alpha_1| < 1$, the value of y_t is by it self useful in predicting the future direction of the sequence, because the process will revert to its mean in the absence of shocks. The smaller α_1 is, the faster will the series revert to the mean.

Ordinary critical t-values cannot be used to test whether $\alpha_1 \neq 1$. Dickey-Fuller critical values are needed.

[32] Almost one is almost unbounded variance, as the system will be very slow to revert to the mean.

The Augmented Dickey-Fuller test

The cautionary way to test α_1 is to assume $\alpha_1 = 1$, and try to reject the assumption. The general testing equation with $\gamma = \alpha_1 - 1$, is

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_2 t + \Sigma_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t. \tag{35}$$

But (35) is a misspecification if the data generating process do not have a drift or trend. Drift is incorporated by α_0 and trend is t. Based on the discussion in the text, I employ the testing equation without drift and trend:

$$\Delta y_t = \gamma y_{t-1} + \Sigma_{i=2}^{p} \beta_i \Delta y_{t-i+1} + \varepsilon_t. \tag{36}$$

The null hypothesis is that the coefficient of interest, γ, is zero. The alternative hypothesis is that γ is less than zero. The power of the test is known to be poor. The important point is that the t-values to test the coefficient of interest are not the ordinary ones. Suitable critical values are obtained by Dickey and Fuller in a simulation study. The augmented part of the test is the inclusion of more than one lag in the testing equation. The reason to include more lags is that the test assumes that the errors are independent and have a constant variance. Hence, if the series contains auto regressive components, they must be included. The problem then is to choose the proper lag length in the testing equation. I use the AIC information criterion as my tool to choose the proper lag length.

$$\text{AIC} = -2\ln(\text{maximized value of log-likelihood}) + \frac{(1+p)}{N} \tag{37}$$

Where p is the order of the auto regressive process and N is the sample size.

In short, the criterion appreciates a better fit of the model, but punishes the inclusion of additional parameters to be estimated. You necessarily improve the fit by including parameters, hence the need of a punishment for including more lags. Start with many lags and decrease the number of lags when finding the optimal number of lags. The other way around is found to be biased to selecting to few lags. Enders (2010), page 217.

Alternatives when failing to reject unit root

Univariate analysis To proceed with the analysis in a meaningful way, it is necessary to make the series stationary in the correct way, either by differencing or detrending. It is important which way it is made stationary, as a difference stationary series will not be stationary by de-trending and a trend stationary will not be stationary by differencing. Differencing means that you loose the information on levels, you are just left with the information stemming from changes in the series. Detrending by a deterministic trend preserves level information.

Multivariate analysis

- As in the univariate case — make the series stationary.
- Check for co-integration. If variables are co-integrated they are related to each other relatively, but the system as a whole still contains a (common) stochastic trend. For variables that are co-integrated, OLS can be used to estimate an equilibrium relationship. Hence, it is possible to build a model predicting local co-movements, but predicting the future level of the system as a whole is not possible. On way to test for co-integration is to run a linear regression and test whether the residuals from the regression is a stationary series.

Ljung-Box test and McLeod Li test for conditional volatility

For the exact test specifications, see Enders (2010) page 131-132. Volatility modelling is a model of the certainty in the prediction, measured by a confidence interval around the mean model. Therefore is it necessary to build a mean model first, take the residuals from the predictions by the mean model versus actual observations, square it (the ordinary residuals are for mean modelling) and then see if this sequence is white noise or not. If it is white noise, there is no pattern in the volatility. If there is conditional volatility in the series, a volatility shock enters the model, you predict the mean according to the mean model, but you are now more uncertain about the predictions because there are, lets say, auto-regressive volatility. This means that a confidence interval decrease gradually. What is a volatility shock? There will be errors each period, so there will be volatility shocks each period.

24 Appendix: Solve AR(1), $a_0 = 0$ and $a_1 = 1$

Because $a_1 = 1$, I straight away impose $a_0 = 0$ or else there is a deterministic time trend.

$$\begin{aligned} y_t &= a_0 + a_1 y_{t-1} + \varepsilon_t \\ y_t &= y_{t-1} + \varepsilon \end{aligned} \tag{38}$$

Solve this by iteration

$$\begin{aligned} y_t &= (y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ y_t &= ((y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t \\ y_t &= \sum_{i=0}^{\infty} \varepsilon_{t-i} \end{aligned}$$

This is an improper solution because it is not finite. Impose an initial condition $y_0 = y_0$. I set $t = 3$ to understand how it plays out.

$$\begin{aligned} y_3 &= ((y_0 + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t \\ y_3 &= y_0 + \sum_{i=0}^{2} \varepsilon_{t-i} \end{aligned}$$

Generally:

$$y_t = y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \tag{39}$$

This is finite, but not stationary. This is because each new realization of ε_t has a *permanent* effect on all future values of y_{t+s}.

25 Appendix: The (Possible) Role of Δt

As discussed in Appendix: ADF-test on page 51, in $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$, a necessary and sufficient condition is $|a_1| < 1$ for the process to be stationary. Consider the approximation

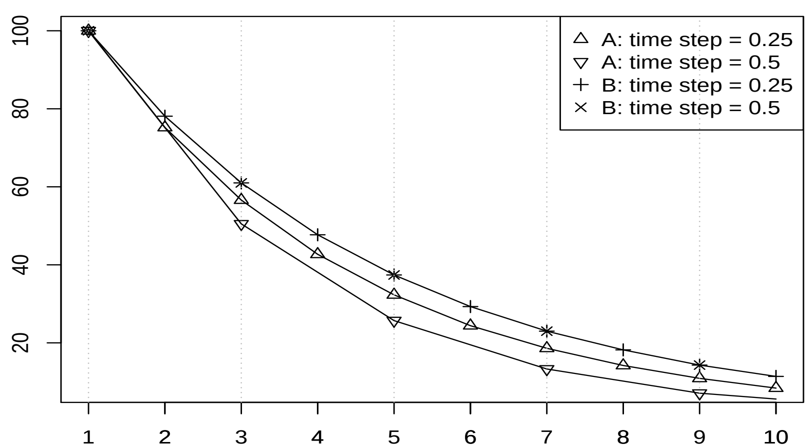
$$F_{t+\Delta t,T} - F_{t,T} = \kappa\left(\mu - F_{t,T}\right)\Delta t + \gamma\sqrt{\Delta t}\eta_{t+\Delta t}\ , \text{ where } \quad \eta_{t+\Delta t} \sim \eta_t \sim N(0,1) \tag{40}$$
$$F_{t+\Delta t,T} = \kappa\mu\Delta t + (1-\kappa\Delta t)F_{t,T} + \gamma\sqrt{\Delta t}\eta_{t+\Delta t} \tag{41}$$

$|1-\kappa\Delta t| < 0$ is necessary in (41). κ is restricted to be positive. Restriction on Δt is:

$$-1 < 1-\kappa\Delta t \quad \text{and} \quad 1-\kappa\Delta t < 1 \quad \Leftrightarrow \quad 0 < \Delta t < 2/\kappa \tag{42}$$

When simulating a differential equation on differential form, or solving a differential equation numerically, the size of the time step Δt matters for the accuracy. This is illustrated in Figure 14. Assume a weekly frequency $\Delta t = 1/52$. Then kappa should be estimated to be below $1/52 < 2/\kappa \Leftrightarrow \kappa < 104$ for the weekly model to be compatible with mean reversal, and not 'overshooting' / amplified oscillations.

Figure 14: Illustration with $\kappa = 1$, $\mu = 1$, initial value of 100 and $\Delta t = 0.25$ & $\Delta t = 0.5$.
Differential form A: $F_{t+\Delta t,T} = F_{t,T} + \kappa\left(\mu - F_{t,T}\right)\Delta t$
Solved form B: $F_{t+\Delta t,T} = F_{t,T}e^{-\kappa\Delta t} + \mu\left(1 - e^{-\kappa\Delta t}\right)$
Equation B yields the same values regardless of the time step Δt. The approximation in A is more accurate the smaller the time step is.



26 Appendix: Itô's Formula

To derive dynamics for a SDE A general diffusion D_t is given by

$$dD_t = \mu(t, D_t)\ dt + \sigma(t, D_t)\ dW_t.$$

Consider two diffusions X_t and Y_t, dependent on the same Wiener process.
Consider a diffusion $Z_t = f(X_t, Y_t)$, the Itô formula is

$$dZ_t = \frac{\partial f}{\partial X} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX_t)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial X \partial Y} dX_t dY_t + \frac{\partial f}{\partial Y} dY_t + \frac{1}{2}\frac{\partial^2 f}{\partial Y^2}(dY_t)^2 \tag{43}$$

Accompanied by the multiplication rules: $dt \times dt = dt \times dW = 0$, and $dW \times dW_t = dt$.
Alternative form of Itô's formula, where $Z_t = f(t, X_t)$.

$$dZ_t = \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X}\mu(t, X_t) + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}\sigma(t, X_t)^2\right] dt + \frac{\partial f}{\partial X}\sigma(t, X_t) dW_t \tag{44}$$

Itô isometry to compute variance

$$\mathbb{E}\left\{\left(\int_0^T D_t dW_t\right)^2\right\} = \mathbb{E}\left\{\int_0^T D_t^2 dt\right\} \tag{45}$$

27 Appendix: Solving the Mean-Reverting Ornstein–Uhlenbeck Process

$$dX_t = \kappa(\mu - X_t)dt + \gamma dW_t \tag{46}$$

Introduce $f(X_t, t) = X_t e^{\kappa t}$

$$\begin{aligned}
df(t, X_t) =& \kappa X_t e^{\kappa t} dt + e^{\kappa t} dX_t \\
=& \kappa X_t e^{\kappa t} dt + e^{\kappa t} \left[\kappa(\mu - X_t)dt + \gamma dW_t\right] \\
=& \cancel{\kappa X_t e^{\kappa t} dt} - \cancel{\kappa X_t e^{\kappa t} dt} + \kappa \mu e^{\kappa t} dt + e^{\kappa t} \gamma dW_t \\
df(t, X_t) =& \kappa \mu e^{\kappa t} dt + e^{\kappa t} \gamma dW_t \\
dX_t e^{\kappa t} =& \kappa \mu e^{\kappa t} dt + e^{\kappa t} \gamma dW_t \\
\int_t^T dX_s e^{\kappa s} =& \int_t^T \left[k \mu e^{\kappa s} ds + e^{\kappa s} \gamma dW_s\right]
\end{aligned}$$

Part one:
$\int dX_s e^{\kappa s}$, $u = X_s e^{\kappa s}$, $\frac{du}{ds} = \kappa X_s e^{\kappa s}$
$\int \kappa X_s e^{\kappa s} ds = X_s e^{\kappa s} + C$
$\int_t^T \kappa X_s e^{\kappa s} ds = [X_s e^{\kappa s}]_t^T = X_T e^{\kappa T} - X_t e^{\kappa t}$

Two:
$\int_t^T \kappa \mu e^{\kappa s} ds = [\mu e^{\kappa s}]_t^T = \mu(e^{\kappa T} - e^{\kappa t})$

Three:
$\int_t^T e^{\kappa s} \gamma dW_s = \gamma \int_t^T e^{\kappa s} dW_s$

Hence:

$$X_T e^{\kappa T} - X_t e^{\kappa t} = \mu(e^{\kappa T} - e^{\kappa t}) + \gamma \int_t^T e^{\kappa s} dW_s$$

$$X_T = X_t e^{\kappa t} e^{-\kappa T} + e^{-\kappa T}\mu(e^{\kappa T} - e^{\kappa t}) + \gamma e^{-\kappa T} \int_t^T e^{\kappa s} dW_s$$

$$X_T = X_t e^{-\kappa(T-t)} + \mu(1 - e^{-\kappa(T-t)}) + \gamma e^{-\kappa T} \int_t^T e^{\kappa s} dW_s \tag{47}$$

$$\mathbb{E}_t\{X_t\} = X_0 e^{-\kappa t} + \mu\left(1 - e^{-\kappa t}\right) \tag{48}$$

$$\mathrm{Var}_t(X_T) = \gamma^2 \int_t^T e^{2\kappa(s-t)} ds \tag{49}$$

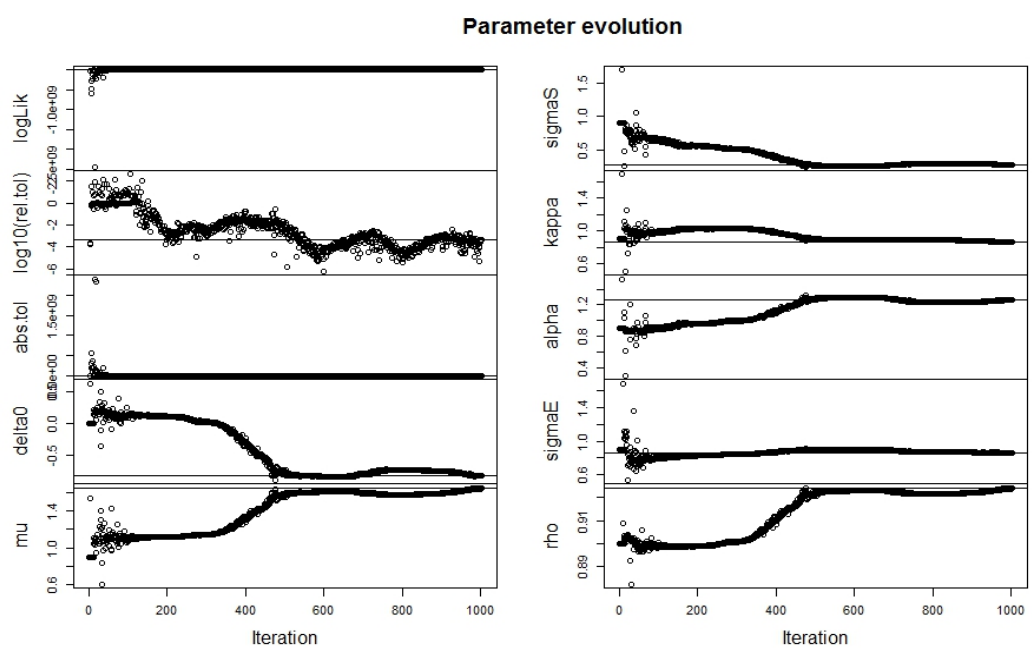
$e^{-\kappa t}$ and $1 - e^{-\kappa t}$ can be interpreted as weights on X_0 and μ in the expectation, where the initial value X_0 has less weight the larger κ and t are. It is intuitive that the initial value has less weight the farther ahead in time we are and also the stronger the reversion towards the mean is.

28 Appendix: Two-Factor Model Robustness test

When observing how sensitive the estimated parameter values from one function call are to initial values and which parameters held constant, it is natural to look more into it. The motivation behind the choices made are robustness. I want to find out how robust the estimates are, and this appendix describes how it is done.

Calling once I call the optimization function once with arbitrary initial values for all parameters, and let all be free to vary. Free to vary means that the optimization function tries to find the optimal value. The optimization function uses an maximum likelihood[33] motivated algorithm to find the optimal parameter values. I allow the optimization function to iterate 1000 times. After it converges or hits the maximum number of iterations, it contains a vector of suggested parameter values. The evolution as the function iterates as a result of *one* single function call is seen in Figure 15.

Figure 15: *One* function call of the optimization function `fit.schwartz2f` in the package `Schwartz97`. The initial value is 0.9 for all parameters.



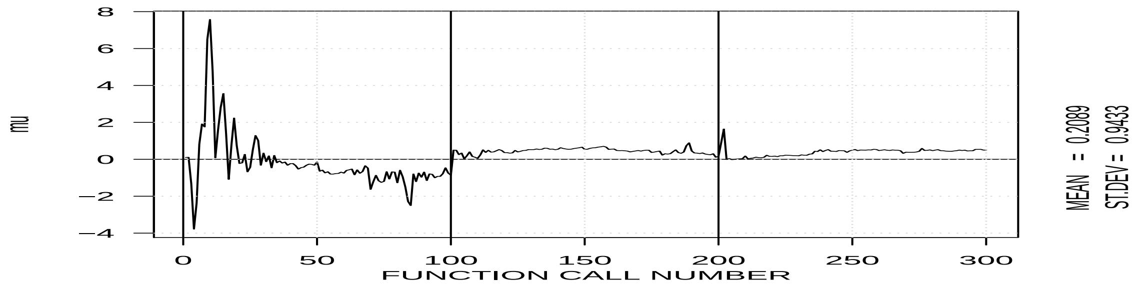
[33]See Appendix: Maximum Likelihood Estimation on page 69 for an introduction to maximum likelihood estimation.

Calling 100 times Then, I use the returned output-vector of results — from one function call — as input-vector for initial values in a new call of the optimization function. Taking the output as input without interaction from me is done 100 times for each initial set of "settings". Let me define 100 function calls in this way as "a round". In a round I effectively gives the optimization function more time to move around. Parameter values through three rounds, with the arbitrary initial values 0.1, 0.5 and 0.9 for all parameters, is graphed in Figure 16 and 17.

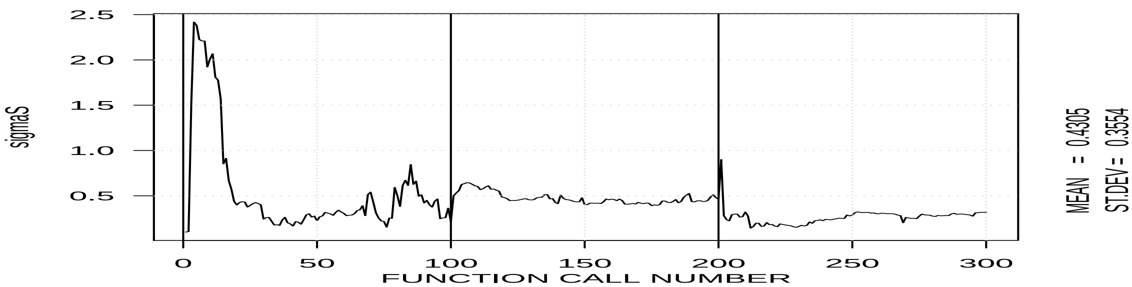
Kappa, alpha, sigmaE and lambda are clearly dependent on their initial values as they do not converge to the same level in the three different rounds. Rho shows strong evidence of convergence to roughly the same high level, 0.95. Mu and sigmaS appears to converge. The initial high sigmaS is likely a result of the initial low rho.

The benefit of taking a round rather than a single function call is clear from the first round — the parameters need time to settle. Seeing a rapid convergence, followed by a stable evolution increases my confidence in the estimates.

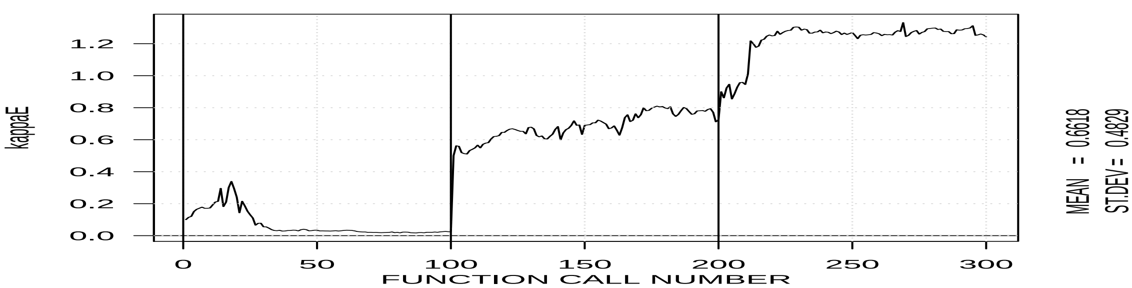
Figure 16: Evolution of μ, σ_s, κ through three rounds, with initial values: [(function call number, initial value) (0, 0.1) (100, 0.5) (200, 0.9)]. Early in the robustness testing.



(a) μ

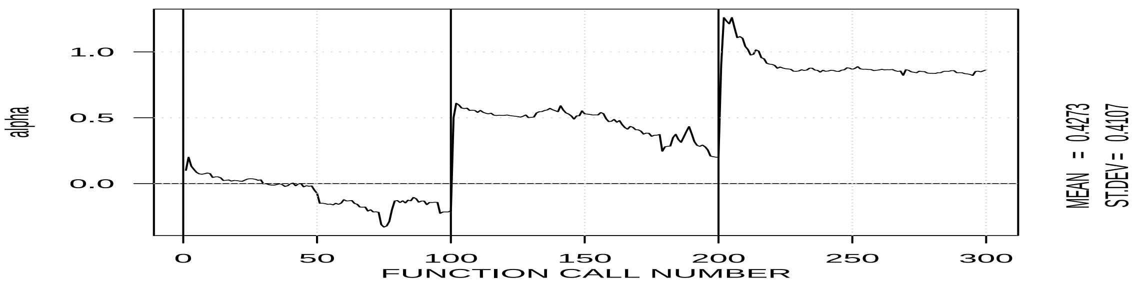


(b) σ_s

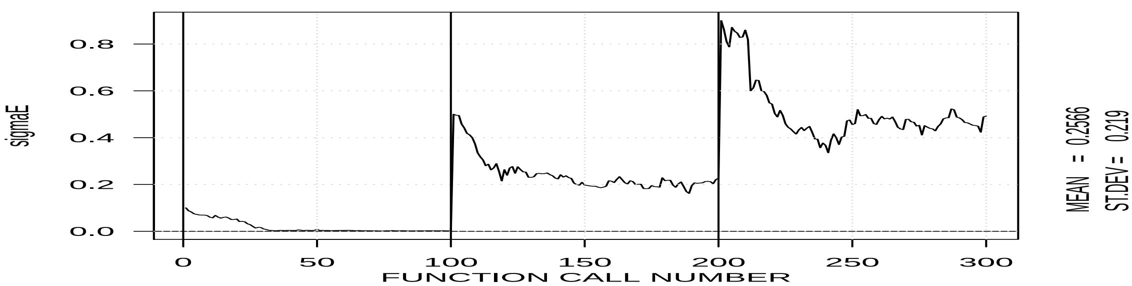


(c) κ

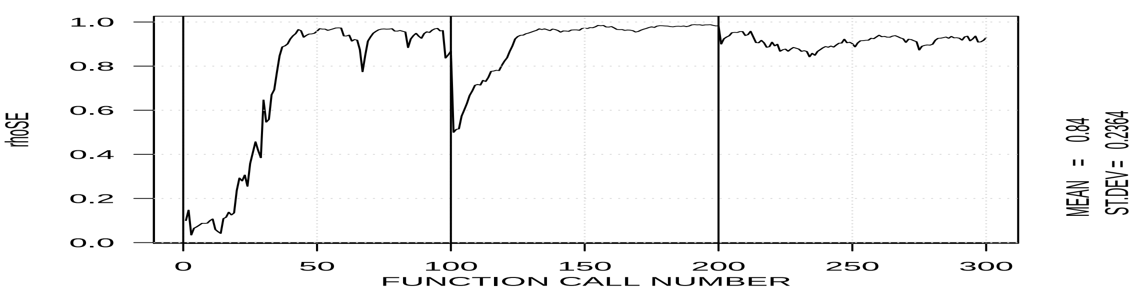
Figure 17: Evolution of $\alpha, \sigma_\delta, \rho_{SE}, \lambda_\delta$ through three rounds, with initial values: [(function call number, initial value) (0, 0.1) (100, 0.5) (200, 0.9)]. Early in the robustness testing.



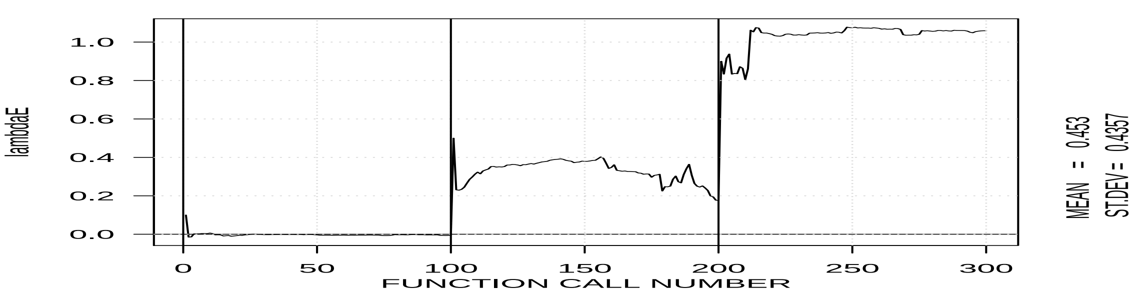
(a) α



(b) σ_δ



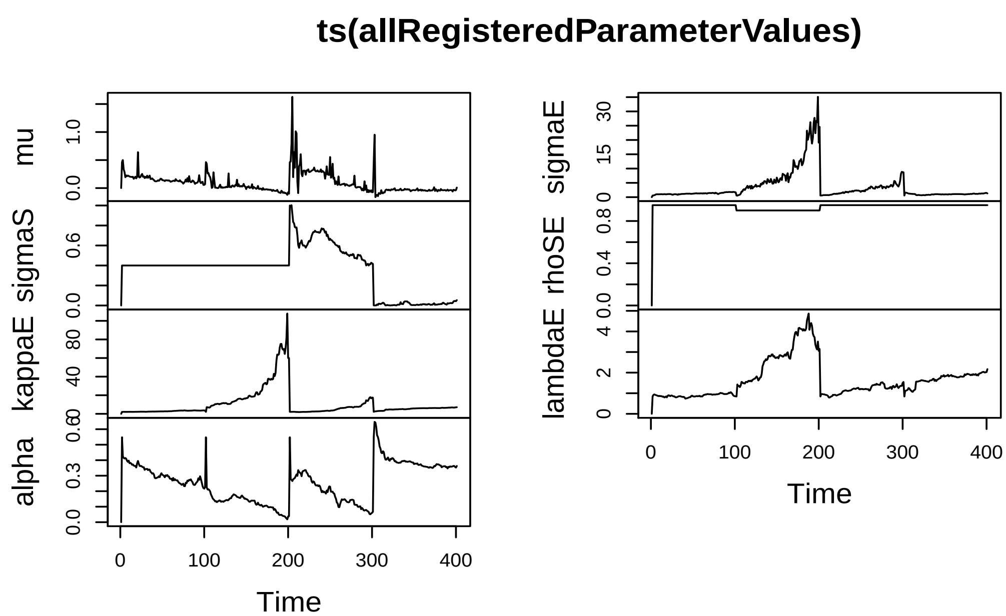
(c) ρ_{SE}



(d) λ_δ

Fixing some parameters and optimize others I continue doing rounds with different initial values and different parameters being fixed. The choices made here are clearly discretionary. As an example, I show in Figure 18 four rounds where I first fix sigmaS at 0.4 and then in the second round change the level rho is fixed at from 0.95 to 0.9. kappaE, sigmaE and lambdaE responds to this change by increasing. In the last two rounds I fix rho at 0.95, and let sigmaS be free to vary and set the initial of sigmasS both below and above the value it appears to converge to. The initial value was 1 for 200 to 300 and 0.001 for 300 to 400. Observe how sigmaS fail to reach the frequently observed level around 0.4 in round four. A sigmaS value of 0.001 is not reasonable to have in the model anyway. Another problem when trying different initial conditions is the interaction between the parameter values. That kappaE and sigmaE interact is clear from round 2 (horizontal axis at 100-200) in Figure 18. For example in another instance (not graphed) kappa and sigmaE rose together, reaching kappa $= 2.2 \times 10^{112}$ and sigmaE $= 1.02 \times 10^{48}$.

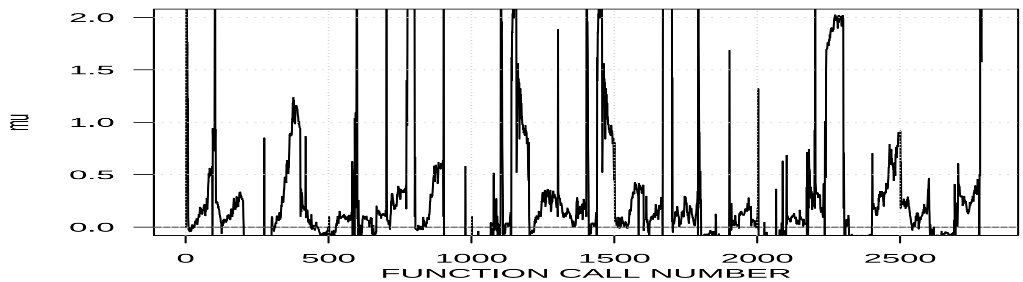
Figure 18: Robustness testing. Here I fix sigmaS (σ_S) and vary rhoSE (ρ_{SE}). (Same information in the plot as in Figure 16 and 17, but a compact version.)



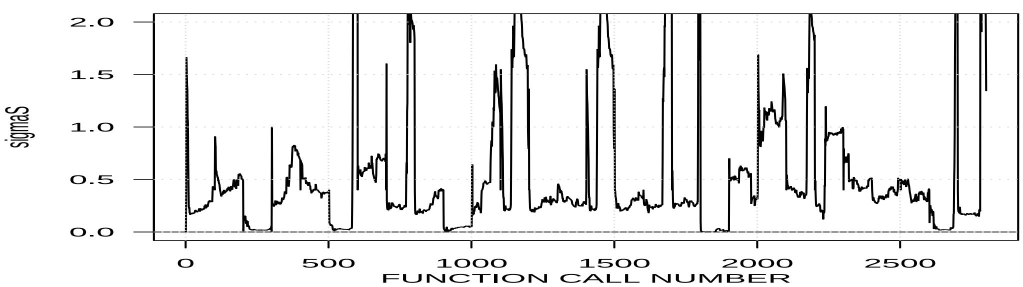
Let them vary with informed initial values After rounds with different parameters being fixed, I again let all vary, and set each parameter with an initial value both below and above the level it appears to converge to. In some rounds are high values of one parameter associated with high values of another parameter, resulting in asymmetric behaviour and extreme values when they take off upwards away from the level they should converge to. Some of the history of such later rounds are shown in Figure 19 and 20. At this point, instead of throwing out the vectors with values outside the limit, I set the values on the axes to reasonable values. Then it is also seen for which sets the estimates are out of bounds.

Curiously, to see how robust the convergence is, I set the initial value of rhoSE to 0.001 and 0.999. I get a kappaE of 8×10^{34}, which is not desirable to put in the model. A value of rhoSE equal to 0.001 is not reasonable, so whether the model converges or not from such values is not important. I argue that it suffices if the model is locally convergent, if all parameters take on reasonable values in the local optimum. I therefore now set initial values equal to the boarder of reasonable values and see whether they converge from there. These later rounds are inclueded in Figure 19 and 20.

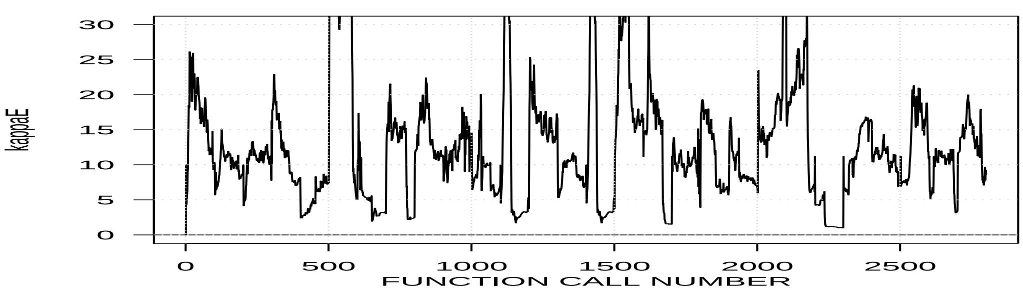
Figure 19: Evolution of μ, σ_s, κ later in the robustness testing.



(a) μ

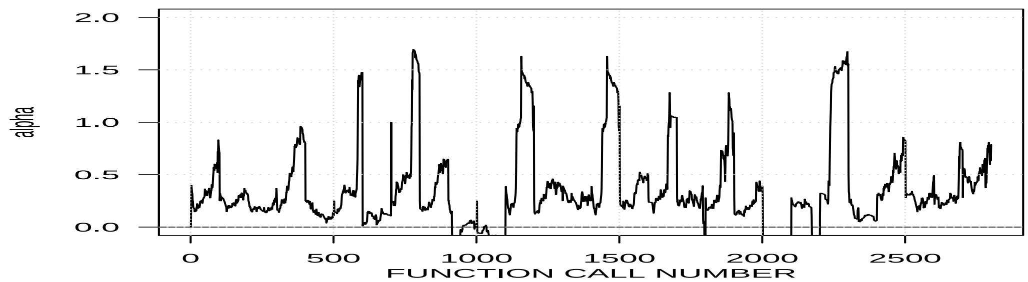


(b) σ_s

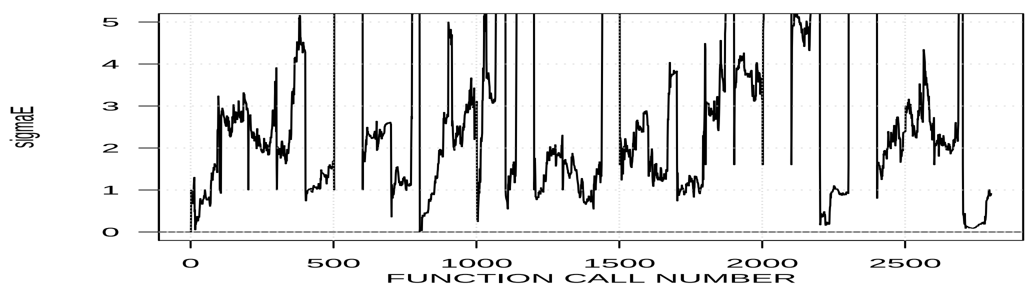


(c) κ

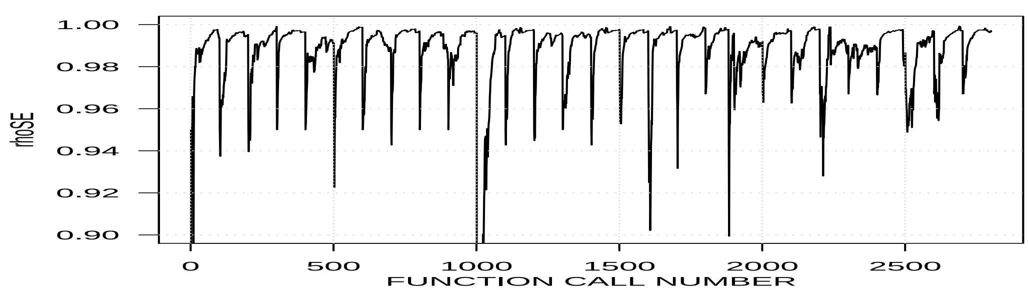
Figure 20: Evolution of $\alpha, \sigma_\delta, \rho_{SE}, \lambda_\delta$ later in the robustness testing.



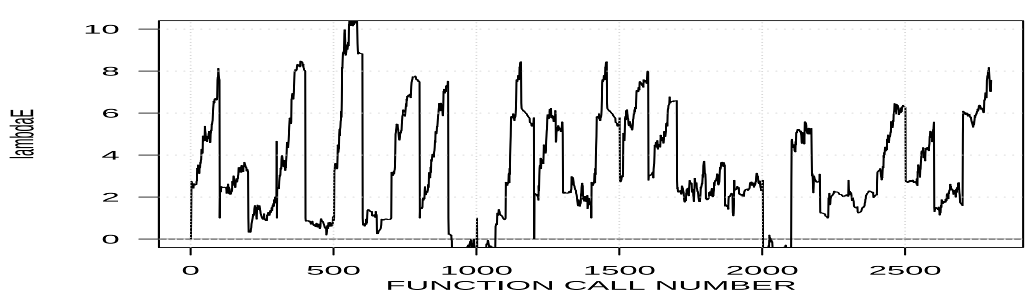
(a) α



(b) σ_δ



(c) ρ_{SE}



(d) λ_δ

Filtering the sample To increase the sample of output vectors from the optimization function, do I initiate many rounds with a variety of settings. I then filter all these stored output vectors according to my discretionary limits. I want to keep track off the output vector, vector by vector, because the parameters are estimated together as a set in the optimization function. Hence, I throw out the whole vector even if only one of the parameters ends up outside its limits. After this, I am left with a filtered list of vectors, where some vectors contain parameters which were not optimized (allowed to vary). These vectors are not thrown out of the filtered matrix because they contribute to variation for the other parameters.

Correcting What I do is to compute the mean and variance of the parameters on a 'corrected subset' of the parameters. The corrected subset consist only of observations where two observations in a row are not equal, to take out the values where they were fixed. I do this to avoid a bias in the computation of the mean and variance. E.g., if I have set σ_S to be 0.4, I do not want 100 observations of 0.4 to be included in the computation of the mean of the optimized value of sigma.

Conclusion regarding robustness The results (numbers found) are semi-robust in the sense that they showed some evidence of convergence.

29 Appendix: Kalman Filter

Hamilton (1994) explains the Kalman filter. The spot price is unobservable, that is the problem. If the futures prices reflect the spot price, but not in a clear mapping from one of the prices into the spot price, the idea is to use several prices to reason on what the current spot price is. The futures prices are therefore noisy reflections of the spot price. To have a system in state space form is a way of describing a dynamic system. You need the current state and description on how you move into the next state. When this is defined the Kalman filter is a technique when there are noise in the data. It updates the prediction on the next state by punishing measurements with high variance and high uncertainty. The Kalman filter assumes that the futures prices are normally distributed, and it works best on linear systems, so the log-prices are preferred in the representation to do the optimization.

30 Appendix: Maximum Likelihood Estimation

Define:

- *Sample* = A particular sequence of observations: $\sum_{i=1}^{N} u_i$.
- *Argument* = A specific parameter value

MLE in short

Find the arguments that maximize the likelihood of actually observing the sample. How:

(i) For each u_i, assume a probability density function $pdf(u_i)$ to represent the probability of observing that particular observation.

(ii) The probability of observing the sample is $L \equiv \prod_{i=1}^{m} pdf(u_i)$ if the observations are independent.

(iii) Take the natural logarithm of the function L to work with sums rather than products.

(iv) Find the arguments of L that maximize L :

   (a) Analytically, take the derivative of $\ln L$ with respect to the parameters of interest.

   (b) Iteratively, try different arguments in the $\ln L$ function and see which argument that yields the highest value of the function.

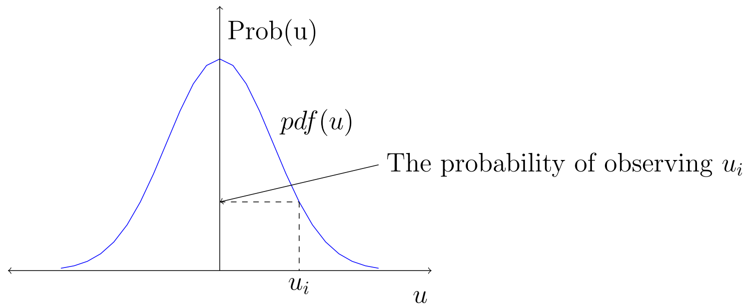
MLE in more detail

- To associate a probability with an observation, assume a probability density function $pdf(u_i)$ for the observation. For example the White-Noise pdf:

$$pdf(u; \mu = 0, v) \equiv \frac{1}{\sqrt{(2\pi v)}} \exp\left[\frac{u_i^2}{2v}\right] = a \exp[b_i]$$

$$a \equiv \frac{1}{\sqrt{2\pi v}}, b_i \equiv \frac{-u_i^2}{2v}$$

- The likelihood of observing u_i is defined as the output of a probability density function.



- The likelihood of observing the sequence $u_1, u_2, \ldots, u_m$ is the product of the probabilities if they are independent, so the likelihood is represented by the product of the pdfs. Define this product as the likelihood L. If the pdfs are equal, the likelihood of observing the particular sequence observed in the data set can be written compactly as

$$L = \prod_{i=1}^{m} a \exp[b_i]$$

- The idea behind MLE is to identify the arguments for the function L such that we maximize the likelihood of observing the sample. So it translates into finding the values of the parameters that maximize L.
- A convenient trick is to take the natural logarithm before computing the derivative as it greatly simplifies the computation of the derivative. This transformation is monotonic, so the argument maximizing L will maximize $\ln L$.
- Deriving the log-likelihood in detail for this particular White-Noise pdf:

$$\begin{aligned}
L = \prod_{i=1}^{m} a e^{b_i} &= a^m \prod_{i=1}^{m} e^{b_i}, \\
\ln\left(a^m \prod_{i=1}^{m} e^{b_i}\right) &= m \ln a + \Sigma_{i=1}^{m} b_i \ln e \\
&= m \ln a + \Sigma_{i=1}^{m} b_i \\
&= m \ln\left(\frac{1}{\sqrt{2\pi v}}\right) + \Sigma_{i=1}^{m} \frac{-u_i^2}{2v} \\
&= m \ln\left(\frac{1}{\sqrt{2\pi v}}\right) - \frac{1}{2v} \Sigma_{i=1}^{m} u_i^2 \\
&= -m \ln\left((2\pi v)^{0.5}\right) - \frac{1}{2v} \Sigma_{i=1}^{m} u_i^2 \\
&= -\frac{m}{2} \ln(2\pi v) - \frac{1}{2v} \Sigma_{i=1}^{m} u_i^2
\end{aligned}$$

- Maximizing the log-likelihood:

- Iteratively, a computer program records the output value of the log-likelihood function given different arguments. In this case, v will be varied until the algorithm finds the argument for v that maximizes the output value of the log-likelihood function. This is an iterative procedure, and it is an alternative to the analytical solution.
- Analytically, take the derivative with respect to the parameter of interest and set it equal to zero. Solve for the parameter to get an expression for the optimal argument. To find the analytical solution for v in this case,

$$\frac{\partial \ln L}{\partial v} = -\frac{m}{2v} + \frac{1}{2v^2}\Sigma_{i=1}^{m} u_i^2$$

Set the derivative equal to zero and solve for v

$$\begin{aligned} -\frac{m}{2v} + \frac{1}{2v^2}\Sigma_{i=1}^{m} u_i^2 &= 0 \\ v^* &= \frac{\Sigma_{i=1}^{m} u_i^2}{m} \end{aligned}$$

This is the optimal argument for the variance, v, in order to describe the sample, given the assumed White-Noise pdf.

- For a simple regression, u_i is the residuals from the regression $u_i = y_i - \beta x_i$.
- For GARCH, v is a non-constant function of u_i. For example, Enders (2010) page 154, the model AR(1) $u_t = y_t - \beta y_{t-1}$ with ARCH(1) $v_t = a_0 + a_1 u_{t-1}^2$ has associated log-likelihood

$$\ln L = -\frac{(m-1)}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=2}^{m}\ln\left(a_0 + a_1 u_{t-1}^2\right) - \frac{1}{2}\sum_{t=2}^{m}\frac{\left(y_t - \beta y_{t-1}\right)^2}{a_0 + a_1 u_{t-1}^2}$$

where you can substitute in $(y_{t-1} - \beta x_{t-1})^2$ for u_{t-1}^2

$$\ln L = -\frac{(m-1)}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=2}^{m}\ln\left(a_0 + a_1(y_{t-1} - \beta x_{t-1})^2\right) - \frac{1}{2}\sum_{t=2}^{m}\frac{\left(y_t - \beta y_{t-1}\right)^2}{a_0 + a_1(y_{t-1} - \beta x_{t-1})^2}$$

and then maximize with respect to a_0, a_1, β.

31 Appendix: Deriving the Risk-Adjusted Process

Compute the present value of a uncertain cash flow.

To compute the present value c_t of a claim c_T in the future that depends the stochastic price F_t and time, a model is constructed. I assume that the market is *complete* such that all states of the uncertain cash fclow can be replicated. In this model *Results* are exploited to make the model internally consistent, i.e. free from arbitrage opportunities. A discussion on martingales and measures can be found in Chapter 28, Hull (2015).

Setting up the model:

Define the economy:

S_t	$=$	Futures price.
D_t	$= \int_0^t S_s \delta_s ds$	Accumulated convenience yield. $\delta_s = \delta(s)$ a function of time only.
		It is the money 'saved' because the futures cash flow is in the future.
A_t	$= e^{rt}$	A risk-free asset earning the constant risk-free rate r continuously.
c_t	$= f(t, c_T(S_T))$	Value of a claim as a function of time and its value at T.
c_T	$= g(T, S_T)$	A cash flow at time T depending on S_T.

Define gains in the economy:

$$
\begin{aligned}
G_t^{(1)} &= S_t + D_t \\
G_t^{(2)} &= A_t \\
G_t^{(3)} &= c_t
\end{aligned}
$$

Discount by a numéraire security. Before presenting result i), discounting by a numéraire security must be introduced. It means to denominate a security in terms of another security, e.g. ${}^{S_t}/_{A_t}$. Here A_t is the numéraire security. Requirements for the numéraire security are,

(i) that it must always have a strictly positive price to avoid division by zero.
(ii) The numéraire must not pay dividends if this is not accounted for.

If dividends are not incorporated, the identification of the drift correction would be wrong, because it would fail to make the gains process a martingale. A_t is suitable to be the numéraire security.

Define the discounted economy with A_t as the numéraire:

$$
\begin{aligned}
S_t^A &\equiv S_t/A_t \\
D_t^A &\equiv \int_0^t (\delta_s S_s)/A_s ds \\
A_t^A &\equiv A_t/A_t \\
c_t^A &\equiv c_t/A_t \\
c_T^A &\equiv c_T/A_T
\end{aligned}
$$

Define discounted gains:

$$
\begin{aligned}
G_t^{A(1)} &= S_t^A + D_t^A \\
G_t^{A(2)} &= A_t^A \\
G_t^{A(3)} &= c_t^A
\end{aligned}
$$

Specify the processes:

$dS_t = \mu(t, D_t, S_t)\ dt + \gamma dW_t$, the drift $\mu(t, D_t, S_t)$ is specified generally.

$$
\begin{aligned}
dD_t &= \delta_t S_t\ dt \\
dA_t &= rA_t\ dt \\
dc_t &= df(t, c_T(S_T)) \\
dG_t^{A(1)} &= d(S_t^A + D_t^A) \\
dG_t^{A(2)} &= dA_t^A \\
dG_t^{A(3)} &= dc_t^A
\end{aligned}
$$

Definition [**1**] The probability measure Q is *equivalent* relative to probability measure P, if all events E assigned zero probability with measure P, also are assigned zero probability with measure Q. I.e.

$$Q(E) = 0 \quad \text{iff} \quad P(E) = 0$$

.

Definition [**2**] A random variable is a *martingale* if the expected value at time t, of its future state in time T, for all $T \geq t$, is equal to the state at time t. I.e.

$$\mathbb{E}_t\{X_s\} = X_t, \quad \forall \quad s \geq t$$

. The drift of a process must be zero for the process to be a martingale.

***Definition* [3]** Using Definition [2] and [3] to define that Q is an *equivalent martingale measure* relative to P if $(D1)$ and $(D2)$ holds,

$$Q(E) = 0 \quad \text{iff} \quad P(E) = 0 \qquad (D1),$$
$$G_t^{*(n)} = \mathbb{E}_t^Q\{G_s^{*(n)}\} \quad , \quad \forall \quad s \geq t \qquad (D2).$$

***Result* [1]** There are no arbitrage opportunities in the defined economy if and only if there exists an equivalent martingale measure such that all discounted gains processes are martingales.

***Result* [2]** When we change from P to Q the standard Wiener process W_t relative to P, will typically not be a standard Wiener process under Q. Girsanov's theorem shows that for arbitrary, but given λ_t the process $dW_t^\lambda = dW_t + \lambda dt$ is a standard Wiener process under some probability measure M. When $M = Q$, M is the EMM. The expected return changes, but the volatility remains the same when we move between F and Q.

Exploiting Result [1] All discounted gains must be martingales to avoid arbitrage:

$$\begin{aligned} G_t^{A(1)} &= S_t^A + D_t^A && = \mathbb{E}_t^Q\{S_T^A + D_T^A\} \\ G_t^{A(2)} &= A_t^A && = \mathbb{E}_t^Q\{A_T^A\} && = \mathbb{E}_t^Q\{1\} \\ G_t^{A(3)} &= c_t^A && = \mathbb{E}_t^Q\{c_T^A\} \end{aligned}$$

Under the probability measure Q, $G_t^{A(2)}$ is a martingale by construction. $G_t^{A(1)}$ is the restriction,, it is where Q is identified. Then, this Q is imposed on $G_t^{A(3)}$ such that $G_t^{A(3)}$ is a martingale by construction as well. Solve for c_t (under Q) to obtain the price of the security c_t. The price c_t is the price in both P and Q worlds, or else it would not be useful. Because of the discounting with a particular numéraire security, the probability measure found is also a particular one — it corresponds to the particular numéraire.

Make all discounted gains process martingales:

$$
\begin{aligned}
dG_t^{A(3)} &= \text{will be by construction} \\
dG_t^{A(2)} &= d(A_t^A) = 0 \\
dG_t^{A(1)} &= d(S_t^A + D_t^A) \\
&= \left(\frac{1}{A_t} dS_t + 0 \frac{1}{2} dS_t^2 - \frac{S_t}{A_t^2} dA_t + \frac{S_t}{A_t^3} \frac{1}{2} \cancelto{0}{dA_t^2} - \frac{1}{A_t} \frac{1}{2} \cancelto{0}{dA_t dS_t} \right) + \left(\frac{\delta(t) S_t}{A_t} dt \right) \\
&= \frac{1}{A_t} dS_t - \frac{S_t}{A_t^2} dA_t + \frac{\delta(t) S_t}{A_t} dt \\
&= \frac{1}{A_t} [\mu(t, D_t, S_t)\ dt + \gamma dW_t] - \frac{S_t}{A_t^2} [r A_t dt] + \frac{\delta(t) S_t}{A_t} dt \\
&= \frac{1}{A_t} \left[\mu(t, D_t, S_t) - S_t \left(r - \delta(t) \right) \right] dt + \gamma dW_t
\end{aligned}
$$

$dG_t^{A(1)}$ is not a martingale because the drift $\frac{1}{A_t} \left[\mu(t, D_t, S_t) - S_t \left(r - \delta(t) \right) \right] dt$ is not zero. Now use Girsanov's theorem to make it a martingale under another probability measure. Insert $dW_t^\lambda - \lambda_t dt$, for dW_t.

$$
\begin{aligned}
dG_t^{A(1)} &= \frac{1}{A_t} \left[\mu(t, D_t, S_t) - S_t \left(r - \delta(t) \right) \right] dt + \gamma \left[dW_t^\lambda - \lambda_t dt \right] \\
&= \frac{1}{A_t} \left[\mu(t, D_t, S_t) - S_t \left(r - \delta(t) \right) - \gamma \lambda_t \right] dt + \gamma dW_t^\lambda
\end{aligned}
$$

Can set λ_t as you want, so choose it so the drift is zero.

$$
\lambda_t = \frac{\mu(t, D_t, S_t) - S_t \left(r - \delta(t) \right)}{\gamma}
$$

For this choice of lambda, is $dG_t^{A(1)}$ a martingale. Apply the measure found to risk adjust dS_t. Denote S_t with λ to emphasise that it is under the risk-neutral measure.

$$\begin{aligned}
dS_t &= \mu(t, D_t, S_t)\; dt + \gamma dW_t \\
dS_t^\lambda &= \mu(t, D_t, S_t)\; dt + \gamma \left[dW_t^\lambda - \lambda_t dt\right] \\
&= \left(\mu(t, D_t, S_t) - \gamma\lambda_t\right) dt + \gamma dW_t^\lambda \\
dS_t^\lambda &= \left(\mu(t, D_t, S_t) - \gamma \left[\frac{\mu(t, D_t, S_t) - S_t^\lambda\,(r - \delta(t))}{\gamma}\right]\right) dt + \gamma dW_t^\lambda
\end{aligned}$$

$$dS_t^\lambda = S_t^\lambda\,(r - \delta(t))\, dt + \gamma dW_t^\lambda \tag{50}$$

The particular drift $\mu(t, D_t, S_t)$ disappears. The volatility is the same in both worlds. Here the convenience yield $\delta(t)$ is allowed to vary through time deterministically. This is useful if one wants to incorporate (possibly through an Euler scheme) an opinion on the convenience yield in the pricing model.

Specify the function $C_T = g(T, S_T)$ as desired. The present value of the derivative in this model is

$$c_t = A_t \mathbb{E}_t^Q \left\{\frac{C_T}{A_T}\right\} = e^{-r(T-t)}\mathbb{E}_t^Q \left\{C_T\right\}. \tag{51}$$

A_T is deterministic, so it goes out of the expectation.
c_t is found by using the Q measure in the expectation to find c_t — that is, to use the dynamics, (50), developed under the Q measure in the computation of the expectation. If it is difficult to obtain an analytical solution is it possible to take the expectation by simulation of the discrete version of the differential equation developed under Q. The time step could have a role, as discussed in Appendix: The (Possible) Role of Δt.

Constant convenience yield For a constant $\delta_t = \delta$ the solution to (50) is

$$S_T^\lambda = S_t^\lambda e^{(r-\delta)(T-t)} + \gamma e^{T(r-\delta)} \int_t^T e^{-s(r-\delta)} dW_s^\lambda \tag{52}$$

$$\mathbb{E}_t^Q\{S_T^\lambda\} = S_t^\lambda e^{(r-\delta)(T-t)} \tag{53}$$

$$\mathrm{Var}(S_T^\lambda) = \gamma^2 e^{2T(r-\delta)} \int_t^T e^{-2s(r-\delta)} ds \quad = \frac{\gamma^2}{2(r-\delta)} \left(e^{2(r-\delta)(T-t)} - 1\right) \tag{54}$$

Convenient to have the following expressions for a call option computation

$$\begin{aligned}
\mathbb{E}_t^Q\{S_T^\lambda\} &\equiv m \\
\mathrm{Var}(S_T^\lambda) &\equiv v^2 \\
S_T^\lambda &\overset{dist.}{=} m + v\,\eta \quad , \quad \eta \sim N(0,1) \\
1_{\{S_T \geq K\}} &= 1 \quad \text{if} \quad S_T \geq K \\
1_{\{S_T \geq K\}} &= 0 \quad \text{if} \quad S_T < K \\
\mathbb{E}\{1_{\{S_T \geq K\}} S_T\} &= \int_{-\infty}^{\infty} 1_{\{m+v\,\eta \geq K\}} (m + v\,\eta) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2} d\eta \\
&= m \int_{\frac{K-m}{v}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2} d\eta + \int_{\frac{K-m}{v}}^{\infty} v\,\eta \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2} d\eta \\
&= mN\left(-\left(\frac{K-m}{v}\right)\right) + \frac{v}{\sqrt{2\pi}} \int_{\frac{K-m}{v}}^{\infty} \eta e^{-\frac{1}{2}\eta^2} d\eta \\
&= mN\left(\frac{m-K}{v}\right) + \frac{v}{\sqrt{2\pi}} \left[-e^{-\frac{1}{2}\eta^2}\right]_{\frac{K-m}{v}}^{\infty} \\
&= mN\left(\frac{m-K}{v}\right) + \frac{v}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{K-m}{v}\right)^2} && (55) \\
\mathbb{E}\{1_{\{S_T \geq K\}}\} &= \int_{-\infty}^{\infty} 1_{\{m+v\,\eta \geq K\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\eta^2} d\eta \\
&= N\left(\frac{m-K}{v}\right) && (56)
\end{aligned}$$

A call option:

$$
\begin{aligned}
c_t &= e^{-r(T-t)}\mathbb{E}_t^Q\left\{max(S_T - K, 0)\right\}\\
c_t &= e^{-r(T-t)}\mathbb{E}_t^Q\left\{1_{\{S_T \geq K\}}(S_T - K)\right\}\\
c_t &= e^{-r(T-t)}\left(\mathbb{E}_t^Q\left\{1_{\{S_T \geq K\}}S_T\right\} - K\mathbb{E}_t^Q\left\{1_{\{S_T \geq K\}}\right\}\right)\\
c_t &= e^{-r(T-t)}\left(mN\left(\frac{m-K}{v}\right) + \frac{v}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{K-m}{v}\right)^2} - KN\left(\frac{m-K}{v}\right)\right)\\
c_t &= e^{-r(T-t)}\left(\frac{v}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{K-m}{v}\right)^2} + \left[S_t^\lambda e^{(r-\delta)(T-t)} - K\right]N\left(\frac{m-K}{v}\right)\right) \qquad (57)
\end{aligned}
$$

Futures contract: The present value of a futures contract maturing in T, entered into in the past, to buy steel for K will have the present value $c_{t,T}$

$$c_{t,T} = e^{-r(T-t)}\mathbb{E}_t^Q\{S_T^\lambda - K\} = e^{-r(T-t)}(\mathbb{E}_t^Q\{S_T^\lambda\} - K) \tag{58}$$

$$c_{t,T} = e^{-r(T-t)}([S_t^\lambda e^{(r-\delta)(T-t)}] - K) = S_t^\lambda e^{-\delta(T-t)} - Ke^{-r(T-t)} \tag{59}$$

If $\delta = r$, then the difference between the price of the contract maturing at T and the agreed price K, is discounted back at r. S_t^λ is the price of steel delivered in T, but the price S_t^λ is discounted by the rate δ because the cash flow is postponed. K is the price agreed at t for delivery at T. So for contracts traded and not yet entered into the present value is (59). In t, $c_{t,T} = 0$ because the payoff in T is symmetric as opposed to an option where there is a price to pay at date t. This observation gives

$$F_{t,T} = S_t e^{(r-\delta)(T-t)} \tag{60}$$

$F_{t,T}$ will not change after date t for the parties entering into the contract but take the constant value $F_{t,T} = K$. The contracts at the exchange are quotes on $F_{t,T}$.

Here it is also clear where the futures relationship in (25) comes from, and the motivation to solve for δ and treat it as a residual, as it is the only parameter not (easily) observable. In this model the interest rate is constant. Interest rates in the market are stochastic, another source of movements in the futures term structure.

32 Appendix: Contract Specifications

Steel Rebar Contract Specifications

- **Product**
  Steel Rebar
- **Contract Size**
  10 tons/lot
- **Price Quotation**
  (RMB) Yuan/ton
- **Minimum Price Fluctuation**
  1 Yuan/ton
- **Daily Price Limit**
  Within 3% above or below the settlement price of the previous trading day
- **Contract Series**
  Monthly contract of the recent 12 months
- **Trading Hours**
  9:00 a.m. to 11:30 a.m., 1:30 p.m. to 3:00 p.m. (the Beijing Time)
- **Last Trading Day**
  The 15th day of the delivery month (If it is a public holiday, the Last Trading Day shall be the 1st business day after the holiday)
- **Delivery Period**
  The 5 consecutive business days after the last trading day
- **Grade and Quality Specifications**
  Standard Products: As specified in Steel for Reinforcement of Concrete–Part 2: Hot-rolled Ribbed Bar, GB1499.2-2007, with designation of HRB400 or HRBF400 with a diameter of 16mm, 18mm, 20mm, 22mm, 25mm.
  Substitute Products: As specified in Steel for Reinforcement of Concrete– Part 2: Hot-rolled Ribbed Bar, GB1499.2-2007, with designation of HRB335 or HRBF335 with a diameter of 16mm, 18mm, 20mm, 22mm, 25mm.
- **Delivery Venue**
  SHFE Certified Delivery Warehouse
- **Minimum Trade Margin**
  5% of contract value
- **Minimum Warranted Delivery Size**
  300 tons
- **Settlement Type**
  Physical Delivery
- **Contract Symbol**
  RB
- **Exchange**
  SHFE

Steel Rebar Warehousing Fees

- **Warehouse Rental, yuan per ton**
  0.15 per day
- **Warehouse entry and exit fee, yuan per ton**
  Special line: 18
  Dock: 15
  Self-pick up: 15