FINANCIAL MATHEMATICS TEAM CHALLENGE

A collection of the five reports from the 2016 Financial Mathematics Team Challenge.

ACQuFRR

AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH

Preamble

FMTC 2016's Winning Team Leader, Ralph Rudd: "Of all of the work I have been involved in since I started my studies, this is probably the work that I am by far the most proud of. The team was phenomenal and this is the hardest any of us has ever worked."

One of the key aims of the FMTC is for South African postgraduate students in Financial and Insurance Mathematics to have the opportunity to focus on a topical, industry-relevant research project, while simultaneously developing links with international students and academics in the field. An allied purpose is to bring a variety of international researchers to South Africa to give them a glimpse of the dynamic environment that is developing at UCT in the African Institute of Financial Markets and Risk Management. The primary goal, however, is for students to learn to work in diverse teams and to be exposed to a healthy dose of fair competition.

The Third Financial Mathematics Team Challenge was held from the 30th of June to the 11th of July 2016. The challenge brought together five teams of Masters and PhD students from France, Australia, South Africa and the UK to pursue intensive research in Financial Mathematics. Each team worked on a separate research problem during the twelve days. Professional and academic experts from France, Switzerland, Australia, Austria, South Africa, and the UK individually mentored the teams; fostering teamwork and providing guidance. As they have in the past, the students applied themselves with remarkable commitment and energy.

This year's research included topical projects on *(a) credit risk in stock-based lending, (b) model risk criteria for recalibration, (c) polynomial models for market weights in stochastic portfolio theory, (d) XVA metrics for CCP optimisation, and (e) stochastic models for commodities*. These were either proposed directly by our industry partners or chosen from areas of current relevance to the finance and insurance industry. In order to prepare the teams, guidance and preliminary reading was given to them a month before the meeting in Cape Town. During the final two days of the challenge, the teams presented their conclusions and solutions in extended seminar talks. The team whose research findings were adjudged to be the best was awarded a floating trophy. Each team wrote a report containing a critical analysis of their research problem and the results that they obtained. This volume contains these five reports, and will be available to future FMTC participants. It may also be of use and inspiration to Masters and PhD students in Financial and Insurance Mathematics.

FMTC III was a great success, so 2017 and FMTC IV are already in the pipeline!

***David Taylor**, University of Cape Town*
***Andrea Macrina**, University College London & University of Cape Town*

Contents

[1] Winning team of the third Financial Mathematics Team Challenge

Stochastic Models for Commodities: a Case Study under the Financialisation Period

Team 1

YANNICK ARMENTI, University of Evry-Paris Saclay
CAMERON HUTCHISON, University of Cape Town
JIANG NANXUAN, University College London
DIRK VAN HEESWIJK, University of Cape Town

Supervisor:
GARETH PETERS, University College London

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

Traditionally, there are two basic types of participants in commodity markets: *hedgers* and *speculators*. Whilst hedgers seek to minimise and manage price risk associated with production, storage, refinement and manufacturing activities specifically related to the physical commodity; speculators instead seek to take on risk associated with taking future-looking positions in the value of a particular physical asset with the primary goal of making profit. Typically, the speculators are not concerned with receiving the physical deliverable and will be less concerned with the aspects of storage cost, inventory or convenience yield so far as they pertain to the true value of the commodity futures price. Instead, they often speculate on perceived short-term discrepancies between spot and future price dynamics which may not bear any real relationship between actual physical production, refinement or inventory management of the physical quantities.

To further illustrate the difference between these two primary market participants, consider the example of a *hedger*, who is a large coffee farmer who wishes to sell their produce at the highest possible price. Risks the farmer faces include unpredictable weather which could affect crop yields, as well as excess market supply that could drive prices down. In such an uncertain position, the farmer could seek to lock in a minimum return for their product by taking a short position in coffee futures. If coffee prices fall, they could then buy back the futures at a lower price than they previously had sold them. This would help to offset the loss from their cash crop and help minimise their risk. Of course, if prices rose, they would lose money on the futures transaction, but the idea is to use futures as a hedge.

Alternatively, *speculators* are typically represented by institutional investors, hedge funds and other financial market participants that are not directly linked to the production, refinement, storage or delivery of the raw commodity. Such market participants are required at some level to balance positions (long and short) in the commodity markets and offer liquidity to hedgers. Typically, they could take the opposite side of the hedger's futures transaction. In this manner, this market participant would bear the risk that prices are going to rise in hopes of generating a profit on the long futures position. Most likely, this type of speculator has no actual

stake in the business, other than futures trading.

There is also the possibility that a manufacturer, refiner or in a perishable commodity setting, a group like a commercial food producer in need of the raw product (such as a coffee processor), may also take the other side of the short hedger's trade to offset the risk of paying higher prices for the commodity. In this manner if the price of coffee rises, the commercial food producer could still capture a profit from the futures position, even though he or she would be paying more for the actual coffee.

Finally, there are also the so called "*E-locals*" who take very short term views on the commodity prices. These are typically individuals actively trading in commodity products throughout a session. They provide liquidity by constantly buying and selling throughout the trading session and are viewed as important participants in the market by shouldering risk.

Prior to the late 1990s and early 2000s period, there was a reasonable mix of both market participant types. However, at around the start of 2000 there was a sustained large increase in investment into commodity markets by institutional investors, as compared to traditional futures market participants such as hedgers (farmers, manufacturers etc.). This was primarily driven by the increasing demand from such institutional investors for commodity futures which were perceived at the time as an alternative asset class. In principle, the commodities futures were not strongly correlated to equity market indices. The primary mechanism by which such large volumes of institutional investors were able to enter into the commodity markets was driven by increasing indexation of commodity futures. This was attractive to institutional investors, which no longer had to worry about understanding intricacies related to harvest cycles, inventory levels, netback margin, freighting costs, weather or other important features that typically directly affect futures prices of particular commodities at particular maturities throughout the year. With indexation of commodities, institutional investors were able to abstract away many of these concerns and invest directly into commodity markets through an index. This obfuscated typical concerns associated with understanding the true relationship between the risk premium associated with taking positions in an ac-

tual futures contract and the associated aspects of convenience yield, storage cost, freighting costs or inventory levels, apart from just tracking an index of a large number of commodity contracts across a spectrum of commodity future types.

With the increasing investment in commodity indices, naturally an explosion in the development of a wide range of commodity indices resulted. The most popular of this wide universe of indices includes the Goldman Sachs Commodity Index (GSCI), the Dow Jones AIG Index (DJAIG), the Reuters-CRB Index, The Rodgers International Commodities Index, the Standard & Poor's Commodity Index and the DJ-UBS Commodity Index to name just a few.

For instance, the DJ DJ-UBSCI is composed of commodity futures contracts on physical commodities, traded on U.S exchanges. The only exceptions are aluminium, nickel and zinc which are traded in London (LME). This index is based upon relative trading activity of individual commodities. The components of this index include: aluminium, coffee, copper, corn, cotton, crude oil, gold, heating oil, lean hogs, live cattle, natural gas, nickel, silver, soybeans, soybean oil, sugar, unleaded gasoline, wheat and zinc. No single commodity may constitute less than 2% of the index and no related group of commodities (e.g., energy, precious metals, livestock or grains) may constitute more than 33% of the index.

With all this sudden index investment potential, there was an opportunity for very large institutional investors to exploit large trading positions in the commodity markets that previously had only been the domain of more specialised investor hedgers and speculators. With this massive increase in investments, came a significant and measurable change in relationships between commodity future indices and equity indices. This structural change in the price behaviour of commodities - especially those linked to indexes such as those referred to above - became known as the *financialisation* of commodities futures. The reason such a phenomenon arose is currently an active area of research, with papers such as Cheng and Xiong (2013) and Basak and Pavlova (2015) reporting such interesting relationships. For instance in the latter paper, the authors demonstrated that since institutional investors care about their performance relative to a commodity index (and not other traditional commodity specific considerations that hedgers have in mind when considering

individual commodities and individual contracts) the increased presence of institutional investors has increased both prices and volatilities of all commodity futures. This effect occurred more so for index futures than for non-index ones. The correlations amongst commodity futures, as well as in equity-commodity correlations, also increase, with higher increases for index commodities. Through their study design, they were also able to demonstrate mechanisms, incorporating aspects of storage. They showed how financial markets may transmit shocks not only to futures prices, but also to commodity spot prices and inventories. Commodity spot prices and inventories rise with financialisation. In the presence of institutional investors, shocks to any index commodity spill over to all storable commodity prices.

This leads us to finally state the key question to be studied in this paper:
Principally, if one considers a range of multi-factor stochastic models for commodity futures curve dynamics as build-around stochastic factors that allow one to reconstruct the model based spot prices for individual commodity futures stochastically over time, can one observe similar evidence for financialisation directly in individual model-based spot price dynamics when compared to equity index observed dynamics?

The remainder of this report addresses the research question in the following sections. The models to be used are introduced and their formulations are specified. Thereafter, the techniques that will be used for filtering, calibration and parameter estimation are described. The real-world data to be used in the investigation is characterised and simulated data is used to understand the behaviour of the model parameters as well as the performance of the filtering techniques. Results are then provided based on the numerical cases studies in which the models were fitted to the data. Finally, conclusions and recommendations for further work are given.

2 Commodity markets

Commodities began to move into the exchange market at a time when the price of commodities had a negligible relationship with market indices (see e.g. Tang and Xiong (2012)), but provided investors with a risk premium for investing in the commodities Falkowski (2011). With an unprecedented inflow of institutional funds, the trend changed in the late 1900's, from which until 2012 the price have

been closely interrelated with the market index. Now, a commodity market can be a physical open outcry or completely virtual electronic exchange for trading raw or primary products.

In the commodity market, there exist hard commodities (natural resources that must be mined or extracted) such as oil and gold, as well as soft commodities (agricultural products or livestock), such as coffee and pork bellies and energy can be traded. Perishable commodities (e.g. corn, sugar, coffee) refer to the commodities which may spoil after some time without proper handling and shipping conditions.

Commodity forward contracts and commodity future contracts are normally traded. A forward contract is a commitment to buy (sell) at a future date a given amount of a commodity or an asset at a price agreed on today. Properties of forward contracts include the following; traded over the counter (not on exchanges), custom tailored, settlement at maturity. On the other hand, a futures contract is an exchange-traded, standardised, forward-like contract that is marked to the market daily. Futures contract can be used to establish a long (or short) position in the underlying commodity or asset. Five key features of futures contract are: standardised contracts, traded on exchanges, guaranteed by the clearing house, gains or losses settled daily and margin account required as collateral to cover losses.

These futures contracts are traded on an exchange, such as NYMEX or the IPE. And the participants in future markets includes hedgers (e.g. farmers), speculators (people who take risk to make profit) and a small group of arbitragers. From a commodity contract, the buyer can avoid the risk of price spike and the seller can lock in a specific selling price. And the price predetermined at the fixed date is called the future price, while the current price in the market is called the spot price.

Key concepts in the commodity market that might affect the commodity price dynamics are as follows.

- **Convenience yield:** a convenience yield is the yield due to holding an underlying product or physical good, rather than the contract or derivative product. For example, Saudi Arabia has large oil reserves, so even if there is no oil in the market this country would still have enough oil.

- **Inventory:** it can help absorb price fluctuations, avoid disruption in the flow of commodities and construct the link between the present and the future.
- **Storage cost:** it is a fixed cost as long as storage capacities are not saturated. And deterioration and obsolescence need to be considered for storage cost. It includes handling costs, maintenance costs and financial costs.
- **Transportation cost:** the future price usually has a positive correlation with shipping cost, especially when there is a booming freight market Geman (2009).
- **Commodity price:** the commodity price observable normally is the future price.
- **Volume:** it represents the total amount of trading activity or contracts that have changed hands in a given commodity market for a single trading day and it measures the pressure or intensity behind a price trend.
- **Open interest:** open interest is the total number of outstanding contracts that are held by market participants at the end of each day, and it measures the flow of money into the future market.

Futures contract dynamics can exhibit two types of behaviour: contango and backwardation. Contango means futures prices increase with maturity while backwardation means futures prices decrease with maturity. In a contango, the spread between futures and spot prices is related to the cost of holding commodities over time (carrying charges). In a backwardation, the spread is related to the net storage cost ('pure storage cost minus convenience yield').

A commodity index is an index that tracks a basket of commodities to measure their performance. These indices are often traded on exchanges, allowing investors to gain easier access to commodities without having to enter the futures market. For example, the Dow Jones Commodity Index Series measures the commodity futures market, emphasising diversification and liquidity using a simple, transparent, equal-weighted approach. The index representing market beta is world-production weighted and is designed to be investable by including the most liquid

commodity futures, and provides diversification with low correlations to other asset classes.

In this paper, we focus on four commodities: coffee, cotton, soybeans and sugar. Parties can see the settlement price, based on the contract specification. For instance, if you buy or sell a soybean futures contract, you will see a ticker-tape handle that looks like this: $S8X@1383'6$. This is just like saying "Soybean (S) 2008 (8) November (K) at \$ 13.8375 per bushel (1383'6)." A trader buys or sells a soybean contract according to this type of quotation. Depending on the quoted price, the value of a commodities contract is based on the current price of the market multiplied by the actual value of the contract itself. In this case, the soybean contract is the equivalent of 5,000 bushels multiplied by our hypothetical price of 1383'6, as in:\$ $13.8375 \times 5,000 bushels =$ \$ $69,187.50$. Contract specifications for these four commodities are as follows:

- *Soybean.* Contract size: 5,000 bushels; tick size: .25 cent/bu ($12.50/contract).
- *Sugar.* Contract size: 112,000 pounds; price quote: .0001 cent/pound = $11.20.
- *Coffee.* Contract size: 37,500 pounds; tick size: .05 cent/pound = $18.75 per contract.
- *Cotton.* Contract size: 50,000 pounds net weight; tick size: $0.0001 = $5.

3 Modeling

Let $(\Omega, (\mathcal{F}_t)_{t\in\mathbb{R}^+}, \mathbb{Q})$ represent a stochastic pricing basis, such that all our processes are $(\mathcal{F}_t)$-adapted. The expectation under $\mathbb{Q}$ and the $(\mathcal{F}_t, \mathbb{Q})$ conditional expectation are denoted by $\mathbb{E}$ and $\mathbb{E}_t^{\mathbb{Q}}$ repsectively. We denote by $(r_t)_{t\in\mathbb{R}^+}$ a $\mathcal{F}_t$ progressive OIS rate process and by $\beta_t = e^{-\int_0^t r_s ds}$ the corresponding discount factor. Finally, we denote by $\mathbb{P}$ the historical probability measure.
We recall that the future price contract for maturity T is defined as the strike $F_{t,T}$ which gives a price of 0 for the payoff $(S_T - F_{t,T})$ at time T. By classical pricing theory arguments, we have:

$$\beta_t P_t = \mathbb{E}_t^{\mathbb{Q}}\Big(\beta_T \left(S_T - F_{t,T}\right)\Big) := 0.$$

Supposing that interest rate are deterministic, this leads to the future pricing formula:

$$F_{t,T} = \mathbb{E}_t^{\mathbb{Q}}\left(S_T\right). \tag{1}$$

3.1 Existing Models

3.1.1 Schwartz and Smith (2000)

One is the more generally accepted commodity models is that of Schwartz and Smith (2000). The authors propose a 2-factor model based on short-term deviation process $(\chi_t)_{t\in\mathbb{R}^+}$ and long run equilibrium process $(\xi_t)_{t\in\mathbb{R}^+}$ of log spot prices. They argue that this is able to correctly capture the dynamics of the spot prices, comparing the parameters to those of Gibson and Schwartz (1990) to show that the new model does include components such as the convenience yield.

The long run equilibrium process should reflect the exhaustion of existing supply of the commodity, cost of production technology and discovery of the commodity, inflation, as well as political and regulatory effects implying an increase in prices. The short-term deviation component should reflect, for example, short-term changes in demand. Because of these stochastic aspects of supply and demand, the convenience yield is implicitly defined through these processes. They define:

$$\ln S_t = \chi_t + \xi_t, \tag{2}$$

with the relevant processes defined by the following stochastic differential equations (SDE) under the risk neutral probability measure $\mathbb{Q}$:

$$\begin{cases} d\chi_t = (-\kappa\chi_t - \lambda_\chi)\,dt + \sigma_\chi\,dW_t^\chi \quad \text{and} \\ d\xi_t = (\mu^\xi + f(\tilde{\mathbf{x}}_t) - \lambda_\xi)\,dt + \sigma_\xi\,dW_t^\xi, \end{cases} \tag{3}$$

where W^χ, W^ξ are two Brownian motions with $d\left\langle W^\chi, W_t^\xi\right\rangle = \rho\,dt$ and λ^χ (resp. λ^ξ) denotes the risk premia of χ (resp. ξ). In particular, letting $\lambda = 0$ in both equations give the historical dynamic of both processes. Note that this assumption of having a constant premium over time will allow us to stay in the class of exponential affine

models and thus derive closed form solutions for derivatives prices. We refer to Cheridito et al. (2007) for more explanations on that subject.

The paper does not mention the $f(\tilde{\mathbf{x}}_t)$ term, but we would rather prefer to add it now, and its goal is to include in the model some external observable market variables namely covariates $\tilde{\mathbf{x}}_t$. Thus, we can add as many of covariates to our model and just aggregate them with a deterministic function f.

These above SDEs are easily solved noting that (χ_t) and (χ_t) are respectively Ornstein-Uhlenbeck and drifted Brownian motion processes which have the following closed form:

$$\begin{cases} \chi_t = \xi_0 e^{-\kappa t} - \frac{\lambda_\chi}{\kappa}(1 - e^{-\kappa t}) + \sigma_\chi \int_0^t e^{-\kappa(t-s)} d\, W_s^\chi, \quad \text{with } \xi_0 = x_0; \\ \xi_t = \xi_0 + (\mu^\xi - \lambda_\xi)t + \int_0^t f(\tilde{\mathbf{x}}_s) ds + \sigma_\xi\, W_t^\xi \quad \text{with } \xi_0 = y_0. \end{cases} \tag{4}$$

In order to derive a closed form solution for the futures price, we can use standard pricing derivation equations. Assuming market efficiency, the future price $F_{t,T} = \mathbb{E}_t^{\mathbb{Q}}(S_T)$ can be represented as a function u^T and we have $F_{t,T} = u^T(t, \chi_t, \xi_t)$ for all t. By Itô's lemma, we can derive the PDE pricing formula:

$$\begin{cases} \partial_t u + (-\kappa\chi_t - \lambda_\chi)\partial_x u^T + (\mu_\xi - \lambda_\xi)\partial_y u^T \\ \quad + \frac{\sigma_\chi^2}{2}\partial_{x^2}^2 u^T + \frac{\sigma_\xi^2}{2}\partial_{y^2}^2 u^T + \rho\sigma_\chi\sigma_\xi \partial_{x,y}^2 u^T = 0 \text{ on } [0,T) \times \mathbb{R} \times \mathbb{R} \\ u^T(T, \chi_T, \xi_T) = \exp(\chi_T + \xi_T). \end{cases} \tag{5}$$

This PDE is of the exponentially affine class (see Filipovic and Larsson (2014) for example) and thus has a solution of the form

$$F_{t,T} = u^T(t, \chi_t, \xi_t) = \exp\Big(B_0(\tau) + B_1(\tau)\xi_t + B_2(\tau)\chi_t\Big), \tag{6}$$

where we denote by $\tau := T - t$ the time to maturity of the contract.

Duffie and Kan (1996) establishes the following system of ODE's (by grouping terms) that solve the latter equations:

$$\begin{cases} \frac{dB_1(\tau)}{d\tau} = 0 \\ \\ \frac{dB_2(\tau)}{d\tau} = -\kappa B_2(\tau) \\ \\ \frac{dB_0(\tau)}{d\tau} = -(\mu_\xi - \lambda_\xi)B_1(\tau) + \lambda_\chi B_2(\tau) - \frac{1}{2}\sigma_\xi^2 B_1^2(\tau) - \frac{1}{2}\sigma_\chi^2 B_2^2(\tau) - \rho_{\chi\xi}\sigma_\chi\sigma_\xi B_1(\tau)B_2(\tau) \end{cases}$$

Using the terminal condition, $B_1(0) = B_2(0) = 1$, based on $F_{T,T} = S_T$ we see that $B_1(\tau)$ and $B_2(\tau)$ are easily solved, and are

$$\forall \tau \geq 0, B_1(\tau) = 1 \text{ and } B_2(\tau) = \exp(-\kappa\tau). \tag{7}$$

Substituting, solving for B_0 by integrating with respect to τ and using the fact that $B_0(0) = 0$, we obtain a closed-form solution:

$$B_0(\tau) = (\mu_\xi - \lambda_\xi)\tau - \frac{\lambda_\chi}{\kappa}\left(1 - e^{-\kappa\tau}\right) + \frac{\sigma_\chi^2}{4\kappa}(1 - e^{-2\kappa\tau}) + \frac{\sigma_\xi^2\tau}{2} + \frac{\rho_{\chi\xi}\sigma_\chi\sigma\xi}{\kappa}(1 - e^{-\kappa\tau}). \tag{8}$$

3.1.2 Yan (2002)

Yan (2002) differed from the model of Schwartz and Smith in the sense that the spot prices and convenience yields were modeled as stochastic processes as opposed to the long-run equilibrium and short-term deviation of Schwartz and Smith. The authors then furthered this through the inclusion of stochastic volatility in the form of square-root jump diffusion process, and stochastic interest rates. The model can be summarised by the following equations:

$$\begin{cases} \frac{dS_t}{S_{t-}} = (r_t - \delta_t - \lambda\mu_J)\,dt + \sigma_S\,dW_t^1 + \sqrt{V_{t-}}\,dW_t^2 + J(\omega)\,dN_t; \\ \ln(1+J) \sim \mathcal{N}\Big(\ln(1+\mu_J) - \frac{1}{2}\sigma_J^2, \sigma_J^2\Big); \\ dr_t = (\theta_r - \kappa_r r_t)\,dt + \sigma_r\sqrt{r_t}\,dW_t^3; \\ d\delta_t = (\theta_\delta - \kappa_\delta \delta_t)\,dt + \sigma_\delta\,dW_t^4; \\ dV_t = (\theta_V - \kappa_V V_t)\,dt + \sigma_V\sqrt{V_t}\,dW_t^5 + J_V(\omega)\,dN_t; \\ J_V \sim \mathcal{E}(\theta), \theta > 0. \end{cases}$$

In deriving a closed-form solution to the futures price, Yan (2002) notes that the jumps in the stochastic volatility do no influence the price. He justifies this through the effect of the jumps on the higher moments (which have no impact on the futures price).

3.1.3 Peters *et al.* (2013)

Peters et al. (2013) furthered the models of Yan (2002) and Schwartz and Smith (2000). First the authors allow the long and short stochastic factors to be mean reverting processes. Secondly they include a third stochastic process with stochastic volatility to allow for modelling of implied volatility smiles, and deterministic seasonality.

The log price $X_t = \ln S_t$ is there defined as $X_t = \chi_t + \xi_t + \theta_t + f(t)$, where the financial meaning of the processes (χ_t) and (ξ_t) have not changed. Denote by (θ_t) this latter component of the formula and f some stepwise function to handle the seasonality. Under $\mathbb{P}$, the SDEs governing this model are as follows:

$$\begin{cases} d\chi_t = -\beta\chi_t\,dt + \sigma_\chi\,dW_t^\chi; \\ d\xi_t = (\mu_\xi - \kappa_\xi\xi_t)\,dt + \sigma_\xi\,dW_t^\xi; \\ d\theta_t = \sqrt{V_t}\,dW_t^\theta \quad \text{and} \\ dV_t = (\mu_V - \kappa_V V_t)\,dt + \sigma_V\sqrt{V_t}\,dW_t^V. \end{cases}$$

Even if a stochastic volatility component appears in the equations of Yan (2002) and Peters et al. (2013), the authors note that it does not impact the future prices.

3.1.4 Conclusion

The two-factor model proposed by Schwartz and Smith (2000) captures many of the observed dynamics of commodity prices, whilst maintaining a closed form solution for the stochastic processes and futures price. Whilst the models of Yan (2002) and Peters et al. (2013) include additional factors such as stochastic volatility, the added complexity of the volatility components make the implementation of such models unsuitable given the time frame. Thus the two-factor model was selected as the best candidate for the task at hand.

4 Filtering and Calibration

In order to fit the models selected, the Kalman filtering framework was used. In order to carry out filtering and the subsequent calibration of the models, state space specification of the models is required. In this section, we provide the state space representation of the model to be used. Then the Kalman filter is introduced and extensions are discussed. The methodology for parameter calibration using the filtering techniques is also described.

4.1 Introduction to use of state space models

Under the affine exponential class of models for the futures price, the log-spot price of a commodity is a linear combination of *latent* processes which are unobservable in the market. In the above section, closed-form solutions for these were obtained.

Thanks to our Markovian setting, we can use the *filtering* technique to calibrate our models. Since we will not be projecting prices of futures contracts, calibration is not intended to be done on a day-by-day basis and thus an off-line approach is adopted. Indeed, this paper is written more with a statistical approach: we seek to find the parameters of our model that can be used to explain the time series observed in our data set. Given a model for the latent processes, the relevant futures pricing equations, and noting that the observed prices in our data set are discrete-time observations, we can formulate the problem of calibration through a state space representation of the model.

Find the best set of parameters Θ, through jointly estimating the latent states at each point in time, using the futures curve available at that time i.e.

$$\begin{cases} \boldsymbol{\alpha}_t = f_\Theta(t, \mathbf{x}_{t-1}, \boldsymbol{\epsilon}_t) \\ \mathbf{y}_t = g(t, \alpha_t, \boldsymbol{\epsilon}'_t) \end{cases} \tag{9}$$

where $\mathbf{y}_t$ is the vector of observed future prices at time t, $\mathbf{x}_t$ is the vector of the latent processes at time t, $(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}'_t)$ are some noise vectors, f is related to some SDE discretisation of the latent processes, and finally g is related to our pricing equation.

The first of the above equations is called the *transition* or *state equation* because it gives information about the distribution of the latent processes at time t knowing the state at time $t-1$. The second one is called the *measurement equation* and gives the relation of the latent process at time t and the observation at time t.

4.2 State space representation of Schwartz and Smith

Since the Schwartz and Smith model SDEs can be solved in closed-form, the discretisation could be applied directly to the conditional densities of the state processes. Due to the model being exponentially affine, we can express the log of the futures price in matrix notation as follows:

$$\mathbb{E}y_t(\tau) = lnF_{t,T};$$

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} B_0(\tau_1) \\ B_0(\tau_2) \\ \vdots \\ B_0(\tau_N) \end{bmatrix} + \begin{bmatrix} e^{-\kappa\tau_1} & 1 \\ e^{-\kappa\tau_2} & 1 \\ \vdots & \vdots \\ e^{-\kappa\tau_N} & 1 \end{bmatrix} \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} + \begin{bmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{bmatrix} \quad \text{and} \tag{10}$$

$$\mathbf{y}_t = \mathbf{B}_0(\tau) + \mathbf{Z}_t(\alpha_t) + (\epsilon_t). \tag{11}$$

This is the measurement equation.
The state equation can also be expressed in matrix notation as

$$\begin{bmatrix}\chi_t \\ \xi_t\end{bmatrix} = \begin{bmatrix}0 \\ \mu_\xi \Delta t\end{bmatrix} + \begin{bmatrix}e^{-\kappa\Delta t} & 0 \\ 0 & e^{-\gamma\Delta t}\end{bmatrix}\begin{bmatrix}\chi_{t-1} \\ \xi_{t-1}\end{bmatrix} + \begin{bmatrix}\eta_t^\chi \\ \eta_t^\xi\end{bmatrix}; \quad \text{and} \tag{12}$$

$$\boldsymbol{\alpha}_t = \mathbf{d} + \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t. \tag{13}$$

Based on the assumptions that the state evolution error and measurement error are both Gaussian, we have

$$\begin{bmatrix}\boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_t\end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}\end{bmatrix}\Big); \tag{14}$$

$$\boldsymbol{\alpha}_t = \begin{bmatrix}\chi_t \\ \xi_t\end{bmatrix} \tag{15}$$

$$\mathbf{d} = \begin{bmatrix}0 \\ \mu_\xi \Delta t\end{bmatrix}; \tag{16}$$

$$\mathbf{Z} = \begin{bmatrix}e^{-\kappa\Delta t} & 0 \\ 0 & 1\end{bmatrix} \quad \text{and} \tag{17}$$

$$\mathbf{Q} = \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa}(1 - e^{-2\kappa\Delta t}) & \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}(1 - e^{-\kappa\Delta t}) \\ \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}(1 - e^{-\kappa\Delta t}) & \sigma_\xi^2\Delta t \end{bmatrix}. \tag{18}$$

Alternative discretisation schemes, such as Euler-Maruyama, or Milstein, would result in the above formulation to change. In the case of a Milstein scheme, it is possible that one or both of the equations will become non-linear Platen. In such cases, an Extended Kalman Filter or Unscented Kalman filter would be needed to approximate these non-linear functions and resulting distributions.

4.3 The Kalman Filter

General Kalman filter is an approach used within the context of state space modelling to the optimal estimator for the state vector $\boldsymbol{\alpha}_t$ at time t, given the overall available information at time t: $\mathbf{y}_{0:t}$. In this section, we derive the steps necessary

in carrying out the filtering. Subsequently we look at the marginal likelihood function for the observations in addressing parameter estimation for the filter.

4.3.1 Derivation of the Filter

The state space model relates these vectors through the through the *measurement equation*:

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{d}_t + \varepsilon_t,$$

where $\mathbf{T}_t$ is an $N \times m$ matrix, $\mathbf{d}_t$ is an $N \times 1$ vector, and ε_t is an $N \times 1$ noise vector such that $\mathbf{E}[\boldsymbol{\varepsilon}_t] = 0$ and $\mathbf{Var}[\boldsymbol{\varepsilon}_t] = \mathbf{H}_t$
A *transition equation* is used to describe the evolution of the state vector through time:

$$\boldsymbol{\alpha}_t = \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \eta_t$$

. Here $\mathbf{T}_t$ is an $m \times m$ matrix, $\mathbf{c}_t$ is an $m \times 1$ vector $\boldsymbol{\eta}_t$ is an $m \times 1$ noise vector with $\mathbf{E}[\boldsymbol{\eta}_t] = 0$ and $\mathbf{Var}[\boldsymbol{\eta}_t] = \mathbf{Q}_t$.
It is assumed that the disturbances and initial state vector follow Gaussian distributions. We complete the specification therefore by assuming an initial state vector $\mathbf{a}_0 = \mathbf{E}[\boldsymbol{\alpha}_0]$, with covariance matrix $\mathbf{P}_0 = \mathbf{Var}[\alpha_0]$. Furthermore, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are uncorrelated with each other at all times t and also uncorrelated with the initial state vector, $\boldsymbol{\alpha}_0$.
Furthermore, we assume that the matrices $\mathbf{T}_t, \mathbf{c}_t, \mathbf{c}_t$ and $\mathbf{Z}_t$ are constant for t, allowing us to drop the subscript. For the ease of notation, we define the following:

$$\mathbf{a}_{t|t-1} = E[\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}];$$

and

$$\mathbf{P}_{t|t-1} = E[(\boldsymbol{\alpha}_t - E[\boldsymbol{\alpha}_t])(\boldsymbol{\alpha}_t - E[\boldsymbol{\alpha}_t])'|\boldsymbol{\alpha}_{t-1}]$$

Note also that $\mathbf{y}_{1:t}$ will be used to represent $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t$.
The filtering is carried out in two steps: a *prediction* step and an *update* step. The formulation of these equations is facilitated by the result that the product of multivariate Gaussian random variables is Gaussian.
In the prediction step, α_{t+1} is estimated, conditional on knowing $\mathbf{a}_t = E[\boldsymbol{\alpha}_t]$. We

estimate the mean and corresponding covariance:

$$\mathbf{a}_{t|t-1} = \mathbf{c} + \mathbf{T}\mathbf{a}_{t-1|t-1} \quad \text{and} \tag{19}$$

$$\mathbf{P}_{t|t-1} = \mathbf{T}\mathbf{P}_{t-1|t-1}\mathbf{T}' + \mathbf{H}. \tag{20}$$

In the update step, the observed values $\mathbf{y}_t$ are used to improve the estimate for the state vector:

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{Z}\mathbf{a}_{t|t-1} - \mathbf{d}) \quad \text{and} \tag{21}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{Z}\mathbf{P}_{t|t-1}. \tag{22}$$

Here, the Kalman gain is defined as

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \mathbf{G})^{-1}. \tag{23}$$

Based on these results, it is possible to derive the prediction error $\mathbf{v}_t$ and the corresponding variance, $\mathbf{F}_t$ i.e.

$$\mathbf{v}_t = \mathbf{P}_{t|t-1}\mathbf{Z}'(\mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \mathbf{G})^{-1}; \tag{24}$$

$$\mathbf{F}_t = \mathbf{H} + \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}_t'. \tag{25}$$

This allows us to express the Kalman gain as

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t. \tag{26}$$

The Kalman gain $\mathbf{K}_t$ has an intuitive interpretation. It represents the relative importance of the error term $\mathbf{v}_t$ with respect to the prior estimate $\mathbf{a}_{t|t-1}$ and is used to form the updated set of estimates. From the formulation above, one can see that the Kalman gain is proportional prediction covariance $\mathbf{P}_{t|t-1}$, resulting in a greater influence when updating. It is inversely proportional to the residual covariance $\mathbf{F}_t$, demonstrating that when the residual covariance is high, less weight should be given to the updating procedure.

The advantages of the Kalman filter include that it can be used for on-line estimation, as well as for prediction and smoothing. It can also be shown that the filter estimates the state vector optimally in a mean-square error sense.

4.3.2 Maximum Likelihood Estimation

The Kalman filter represents a dynamic Bayesian network and therefore allows for the use of recursive Bayesian estimation to derive the necessary marginal conditional likelihood. This likelihood can then be maximised with respect to the parameter vector $\boldsymbol{\Theta}$ in order to find an estimate, $\hat{\boldsymbol{\Theta}}$.
By the Markov assumption, we have independence of each state conditional on previous states i.e.

$$p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{t-1}) = p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}), \tag{27}$$

and similarly, we have independence of each observation conditional on the current state i.e.

$$p(\mathbf{y}_t|\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_t) = p(\mathbf{y}_t|\boldsymbol{\alpha}_t). \tag{28}$$

This allows us to express the joint distribution of all states in the model as:

$$p(\mathbf{y}_{1:t}; \boldsymbol{\alpha}_{1:t}) = p(\boldsymbol{\alpha}_0) \prod_{i=1}^{t} p(\mathbf{y}_i|\boldsymbol{\alpha}_i) p(\boldsymbol{\alpha}_i|\boldsymbol{\alpha}_{i-1}). \tag{29}$$

We can now express the prediction step of the Kalman filter in probabilistic form,

$$p(\boldsymbol{\alpha}_t|\mathbf{y}_{t-1}) = \int p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}) p(\boldsymbol{\alpha}_{t-1}|\mathbf{y}_{t-1}) d\boldsymbol{\alpha}_{t-1}, \tag{30}$$

and the update step,

$$p(\boldsymbol{\alpha}_t|\mathbf{y}_t) = \frac{p(\mathbf{y}_t|\boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t|\mathbf{y}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{t-1})} = \frac{p(\mathbf{y}_t|\boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t|\mathbf{y}_{t-1})}{\int p(\mathbf{y}_t|\boldsymbol{\alpha}_t) p(\boldsymbol{\alpha}_t|\mathbf{y}_{t-1}) d\boldsymbol{\alpha}_t}, \tag{31}$$

The denominator of this term, $p(\mathbf{y}_t|\mathbf{y}_{t-1})$ is the term that we can use to find the conditional likelihood,

$$L = \prod_{i=1}^{t} p(\mathbf{y}_i|\mathbf{y}_{i-1}) = \prod_{i=1}^{t} \int p(\mathbf{y}_i|\boldsymbol{\alpha}_i) p(\boldsymbol{\alpha}_i|\mathbf{y}_{i-1}) d\boldsymbol{\alpha}_i. \tag{32}$$

Under the usual Kalman filter, since the conditional distributions in the integral are Gaussian, we can find a closed-form for the distribution of $p(\mathbf{y}_t|\mathbf{y}_{t-1})$, which will be Gaussian too i.e.

$$\mathbf{y}_t|\boldsymbol{\alpha}_t \sim \mathcal{N}(\mathbf{T}_t\boldsymbol{\alpha}_t, \mathbf{Q}_t); \tag{33}$$

$$\boldsymbol{\alpha}_t|\mathbf{y}_{t-1} \sim \mathcal{N}(\mathbf{a}_{t|t-1}, \mathbf{P}_{t|t-1}). \tag{34}$$

Integrating over $\boldsymbol{\alpha}_t$, we have that

$$\mathbf{y}_t|\mathbf{y}_{t-1} \sim \mathcal{N}(\tilde{\mathbf{y}}_{t|t-1}, \mathbf{F}_t). \tag{35}$$

Based on this we can explicitly write down the log-likelihood for the Gaussian case i.e.

$$\ell(\boldsymbol{\Theta}|\mathbf{y}_{1,2,\ldots,T}) = -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{F}_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{v}_t'\mathbf{F}^{-1}\mathbf{v}_t. \tag{36}$$

If the Gaussian assumption is dropped, there may not necessarily be a closed-form for this conditional distribution. In such cases, a numerical method such as quadrature techniques or Monte Carlo methods could be used to evaluate the integral. An example of when this issue arises is when we have t-distributed observation errors, which will be dealt with further in the synthetic data case, in section.

In the case where maximum likelihood is to be carried out analytically, the following least-square regression procedure can be carried out to recursively find the gradient of the log-likelihood with respect to each of the parameters.

Our score function is of the following form:

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \sum_{t=1}^{T}\frac{\partial l_t(\theta)}{\partial \theta}. \tag{37}$$

Using the following two useful matrix identities, for the symmetric matrix A:

$$\frac{\partial|A|}{\partial x} = |A|tr\left[A^{-1}\frac{\partial A}{\partial x}\right]; \tag{38}$$

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1}; \tag{39}$$

we can write

$$\frac{\partial l_t(\theta)}{\partial \theta_i} = -\frac{1}{2}tr\left[[\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \theta_i}][I - \mathbf{F}_t^{-1}\mathbf{v}_t\mathbf{v}_t']\right] - \left(\frac{\partial \mathbf{v}_t}{\partial \theta_i}\right)^T \mathbf{F}_t^{-1}\mathbf{v}_t. \tag{40}$$

We therefore require the derivatives of $\mathbf{v}_t$ and $\mathbf{F}_t$ with respect to θ_i i.e.

$$\frac{\partial \mathbf{F}_t}{\partial \theta_i} = \frac{\partial \mathbf{Z}_t}{\partial \theta_i}\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{Z}_t\frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i}\mathbf{Z}_t' + \mathbf{Z}_t\mathbf{P}_{t|t-1}\frac{\partial \mathbf{Z}_t'}{\partial \theta_i} + \frac{\partial \mathbf{H}}{\partial \theta_i}; \tag{41}$$

$$\frac{\partial \mathbf{v}_t}{\partial \theta_i} = -\mathbf{Z}_t\frac{\mathbf{a}_{t|t-1}}{\partial \theta_i} - \frac{\partial \mathbf{Z}_t}{\partial \theta_i}\mathbf{a}_{t|t-1}. \tag{42}$$

Next we need derivatives of $\mathbf{a}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ with respect to θ_i i.e.

$$\frac{\partial \mathbf{a}_{t|t-1}}{\partial \theta_i} = \frac{\partial \mathbf{T}}{\partial \theta_i}\mathbf{a}_{t-1} + \mathbf{T}\frac{\partial \mathbf{a}_{t-1}}{\partial \theta_i} + \frac{\partial \mathbf{c}}{\partial \theta_i}; \tag{43}$$

$$\frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i} = \frac{\partial \mathbf{T}}{\partial \theta_i}\mathbf{P}_{t-1}\mathbf{T}' + \mathbf{T}\frac{\partial \mathbf{P}_{t-1}}{\partial \theta_i}\mathbf{T}' + \mathbf{T}\mathbf{P}_{t-1|t-1}\frac{\partial \mathbf{T}'}{\partial \theta_i} + \frac{\partial \mathbf{Q}}{\partial \theta_i}. \tag{44}$$

Finally, the derivatives of $\mathbf{a}_t$ and $\mathbf{P}_t$ with respect to θ_i are given by:

$$\begin{aligned}\frac{\partial \mathbf{a}_t}{\partial \theta_i} &= \frac{\partial \mathbf{a}_{t|t-1}}{\partial \theta_i} + \frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i}\mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + \mathbf{P}_{t|t-1}\frac{\partial \mathbf{Z}_t'}{\partial \theta_i}\mathbf{F}_t^{-1}\mathbf{v}_t \\ &\quad \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \theta_i}\mathbf{F}_t^{-1}\mathbf{v}_t + \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\frac{\partial \mathbf{v}_t}{\partial \theta_i} \quad \text{and}\end{aligned} \tag{45}$$

$$\begin{aligned}\frac{\partial \mathbf{P}_t}{\partial \theta_i} &= \frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i} - \frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i}\mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\frac{\partial \mathbf{Z}_t'}{\partial \theta_i}\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1} \\ &\quad +\mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \theta_i}\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\frac{\partial \mathbf{Z}_t}{\partial \theta_i}\mathbf{P}_{t|t-1} \\ &\quad -\mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t\frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i}.\end{aligned} \tag{46}$$

Despite having this analytical formulation for evaluating the gradient of the log-likelihood function, our calibrations were performed numerically, with the use of

Matlab routine `fmincon`.

4.4 The Extended Kalman Filter

The Kalman filter is specified in such a way that both the state space equation and the observation equation are linear in the state vector. In practice, it is common to find systems which are not linear in nature. In such cases, the Kalman filter is no longer optimal.

One approach to addressing this problem is provided by the *Extended Kalman Filter* (EKF). In this extension, a first order Taylor series expansion is used in order to linearise the state (and possibly the observation) equation. In this case, the estimators produced are no longer optimal and if the initial state vector is specified incorrectly, the filter may diverge quickly, due to the propagation of errors through the estimates. The EKF also tends to perform poorly when using state space equations that are highly non-linear or of high dimensionality.

4.5 The Unscented Kalman filter

An alternative approach to addressing the non-linearity issue was proposed by Julier and Uhlmann (1997) and developed further by Wan and Van Der Merwe (2000), in what has come to be known as the *Unscented Kalman Filter* (UKF). The UKF deals with the approximation inaccuracies of the EKF by making use of the *unscented transform*. The algorithm is described briefly below. In most applications of the Kalman filter to non-linear problems, the UKF has been shown to be preferred over the EKF.

The *unscented transform* (UT) operates by deterministically sampling a set of *sigma points* from the distribution of a random variable in such a way that a selection of the distributions moments are characterised by these points. The method that follows originates from Wan and Van Der Merwe (2000).

Suppose we have a random vector $\mathbf{x}$ and define $\mathbf{y} = f(\mathbf{x})$ where f is a non-linear function. We choose a set of sigma points $\mathcal{X}$, and corresponding weights w_i such that the following constraints are satisfied:

$$\sum_i w_i^m = 1 \qquad \sum_i w_i^c = 1;$$

$$\boldsymbol{\mu}_X = \sum_i w_i^m(\mathcal{X}_i) \qquad \boldsymbol{\Sigma}_X = \sum_i w_i^c(\mathcal{X}_i - \mu_X)(\mathcal{X}_i - \mu)^T,$$

where $\boldsymbol{\mu}_X$ is the mean of X, and $\boldsymbol{\Sigma}_X$ the corresponding covariance matrix. The above constraints do not lead to a unique solution. Julier and Uhlmann (1997) propose the following specifications for sampling these sigma points and calculating their corresponding weights:

$$\mathcal{X}_0 = \mu; \tag{47}$$

$$\begin{aligned} \mathcal{X}_i &= \mu + (\sqrt{(L+\lambda)P_x})_i \qquad i = 1, \ldots, L; \\ \mathcal{X}_i &= \mu + (\sqrt{(L+\lambda)P_x})_{i-L} \qquad i = L+1, \ldots, 2L. \end{aligned}$$

We define $\lambda = \alpha^2(L + \kappa) - L$ as a scaling parameter, where L is the dimension of $\mathbf{x}$ and α, κ are tuning parameters such that $\alpha \in [0, 1]$ and $\kappa \geq 2$, but usually set to 0 (Wan and Van Der Merwe, 2000). The notation $(\mathbf{A})_i$ is used to denote the i^{th} column vector of matrix $\mathbf{A}$.

The corresponding weights are then computed with the introduction of a third tuning parameter, β:

$$w_0^m = \frac{\lambda}{L+\lambda};$$

$$w_0^c = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta;$$

$$W_i^m = W_i^c = \frac{1}{2(n+\lambda)} \quad i = 1, \ldots, 2L.$$

For Gaussian distributions, the optimal choice of β is 2 (Wan and Van Der Merwe, 2000). These parameters can be tuned to capture higher moments of the distributions, such as skew and kurtosis, should this be necessary. Finally we propagate the sigma points through f to obtain estimates for the mean $\boldsymbol{\mu}_Y$ and covariance of Y (respectively $\boldsymbol{\mu}_Y$ and $\mathbf{P}_Y$):

$$\boldsymbol{\mu}_Y \approx \sum_{i=0}^{2L} w_i^m f(\mathcal{X}_i);$$

$$\mathbf{P}_Y \approx \sum_{i=0}^{2L} w_i^c (f(\mathcal{X}_i) - \boldsymbol{\mu}_Y)(f(\mathcal{X}_i) - \boldsymbol{\mu}_Y)^T.$$

Applying the UT in the context of the Kalman filter, the state space model is now defined with observation equation

$$\boldsymbol{\alpha}_t = F(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\epsilon}_t), \tag{48}$$

and state equation

$$\mathbf{y}_t = H(\boldsymbol{\alpha}_t, \boldsymbol{\eta}_t). \tag{49}$$

Functions F and H may be non-linear functions, and the error term is not necessarily additive in each case.
The random variable matrix on which we will be applying the UT is

$$\mathbf{x}_t = \begin{bmatrix} \boldsymbol{\alpha}_t^T & \boldsymbol{\epsilon}_t^T & \boldsymbol{\eta}_t^T \end{bmatrix}^T.$$

Initial conditions needed by the filter are:

$$\mathbf{a}_0 = \mathbb{E}[\boldsymbol{\alpha}_0]; \tag{50}$$

$$\mathbf{P}_0 = \mathbb{E}[(\boldsymbol{\alpha}_0 - \mathbf{a}_0)(\boldsymbol{\alpha}_0 - \mathbf{a}_0)^T]; \tag{51}$$

$$\mathbf{x}_0^* = \mathbb{E}[\mathbf{x}_0] = \begin{bmatrix} \mathbf{a}_0 & \mathbf{0} & \mathbf{0} \end{bmatrix}^T \quad \text{and} \tag{52}$$

$$\mathbf{P}_0^* = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_\epsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_\eta \end{bmatrix}. \tag{53}$$

Next, the sigma points are computed as

$$\mathcal{X}_t^* = \begin{bmatrix} \mathbf{x}_t^* & \mathbf{x}_t^* \pm \sqrt{(L+\lambda)\mathbf{P}_t^*}. \end{bmatrix} \tag{54}$$

The prediction step is given by

$$\mathcal{X}^{\alpha}_{t|t-1} = F(\mathcal{X}^{\alpha}_{t-1}, \mathcal{X}^{\epsilon}_{t-1}); \tag{55}$$

$$\mathbf{a}^{-}_{t} = \sum_{i=0}^{2L} w^{m}_{i} \mathcal{X}^{\alpha}_{i,t|t-1}; \tag{56}$$

$$\mathbf{P}^{-}_{t} = \sum_{i=0}^{2L} w^{c}_{i} (\mathcal{X}^{\alpha}_{i,t|t-1} - \mathbf{a}^{-}_{t})(\mathcal{X}^{\alpha}_{i,t|t-1} - \mathbf{a}^{-}_{t})^{T} \quad \text{and} \tag{57}$$

$$\mathcal{Y}_{t|t-1} = H(\mathcal{X}^{\alpha}_{t|t-1}, \mathcal{X}^{\eta}_{t-1}); \tag{58}$$

$$\mathbf{y}^{-}_{t} = \sum_{i=0}^{2L} w^{m}_{i} \mathcal{Y}_{i,t|t-1}. \tag{59}$$

Moreover, the update step is specified by

$$\mathbf{P}_{y_t y_t} = \sum_{i=0}^{2L} w^{c}_{i} (\mathcal{Y}_{i,t|t-1} - \mathbf{y}^{-}_{t})(\mathcal{Y}_{i,t|t-1} - \mathbf{y}^{-}_{t})^{T}; \tag{60}$$

$$\mathbf{P}_{\alpha_t y_t} = \sum_{i=0}^{2L} w^{c}_{i} (\mathcal{X}_{i,t|t-1} - \mathbf{a}^{-}_{t})(\mathcal{Y}_{i,t|t-1} - \mathbf{y}^{-}_{t})^{T}; \tag{61}$$

$$\mathbf{K}_{\alpha_t y_t} = \mathbf{P}_{\alpha_t y_t} (\mathbf{P}_{y_t y_t})^{-1}; \tag{62}$$

$$\mathbf{a}_{t} = \mathbf{a}^{-}_{t} + \mathbf{K}(\mathbf{y}_{t} - \mathbf{y}^{-}_{t}) \quad \text{and} \tag{63}$$

$$\mathbf{P}_{t} = \mathbf{P}^{-}_{t} - \mathbf{K}\mathbf{P}_{y_t y_t}\mathbf{K}^{T}. \tag{64}$$

Again, using this framework it is possible to construct a marginal likelihood which can be maximised with respect to the parameters in order to calibrate the model. A diagram in figure 1 shows the differences in methodology between actual sampling, using a linear (extended) Kalman filter and using the Unscented Transform for estimating the distribution of the transformed random variable, as taken from Wan and Van Der Merwe (2000).

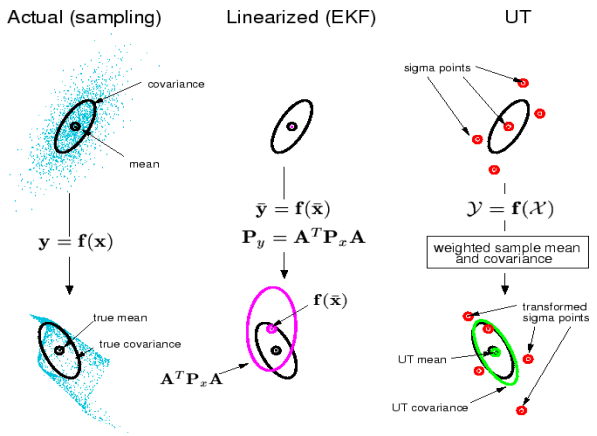


Figure 1: Explanation of the Unscented Kalman Filter

5 Data analysis

5.1 Description

For our study, we could retrieve the prices of futures contracts from the ICE Clear Europe CCP (Central Counterparty Clearing House) for the following commodities:

- coffee;
- cotton;
- soybean and
- sugar.

The given period of study starts from October 1984 until May 2016. This 30-year period of data encompasses thus, *inter alia*, the financialisation period, different financial crisis. The data not only contained the futures prices of the different contracts between 1984 and 2016 but also the *traded volume* and the *open-interest* on a day-by-day basis.

During this period, 165 contracts were traded for Coffee, the same number for Cotton, 190 contracts for Soybean and 128 for Sugar. The figures 2, 3 show the total number of contracts that could be traded day after day for each market referenced.

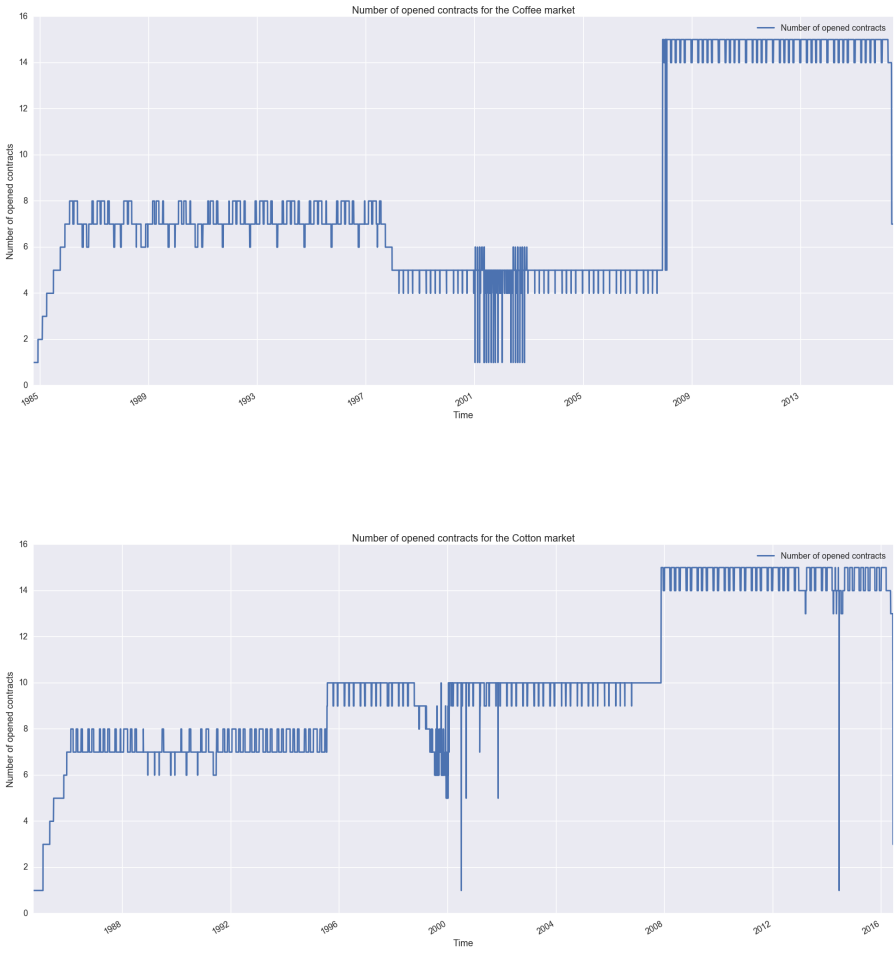


Figure 2: Number of contracts that were tradable during the period analysis for Coffee and Cotton

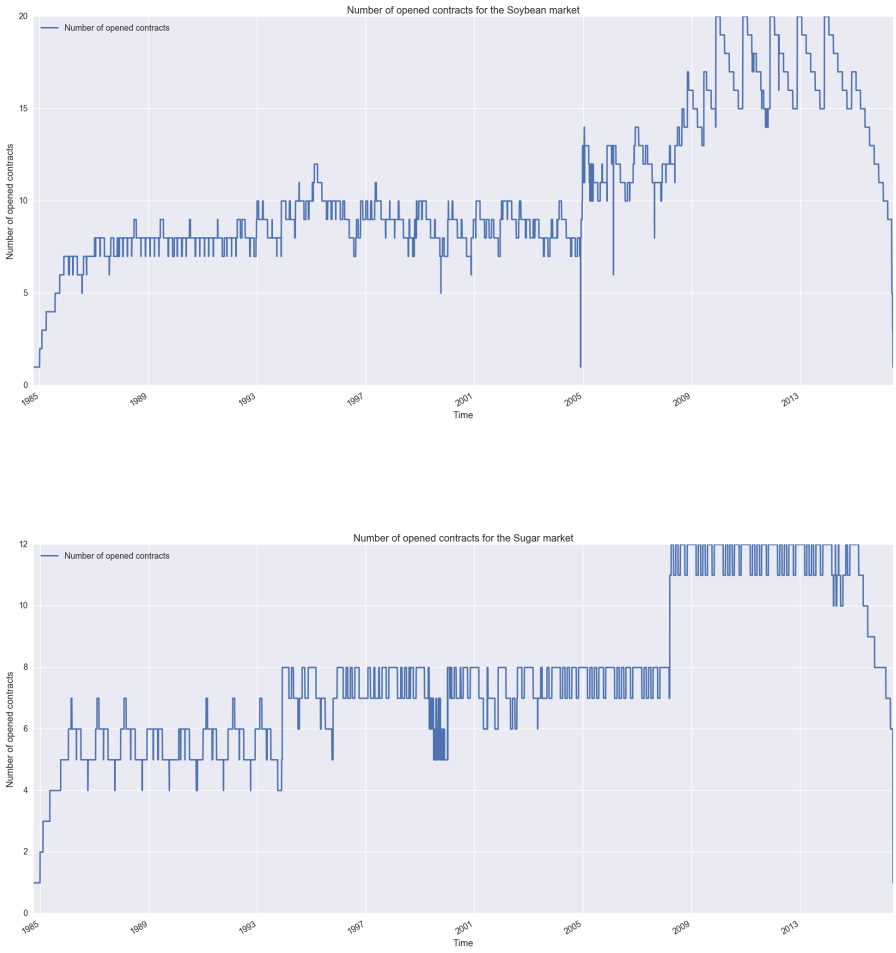


Figure 3: Number of contracts that were tradable during the period analysis for Soybean and Sugar

The graphs show that for a particular day, several maturities of future contracts can be traded. Because of the physical delivery option amongst others, the maturity can't be defined for a particular day. Thus the contract maturities are defined by only the month/year delivery. Moreover, each commodity has its own schedule of maturities:

- coffee: March, July, September, December.
- cotton: March, May, July, October, December.
- soybean: January, March, May, July, August, September, November.
- sugar: March, May, July, October.

Moreover, this graphs show the burst of the commodities market: between 1980's and now, the number of contracts have been multiplied by a factor of 10. The positive trend of volumes confirm that observation: except for soybean, we have an exponential increase of total traded volume during this period (see Figures 4, 5).

6 Numerical results

Data was run on the computational model developed by Ames et al. (2016). This computational model was based on the commodities model of Schwartz and Smith (2000), with the optional addition of gamma.

6.1 Simulated case studies

Synthetic data was simulated in order to study the behaviour of the filtering methods and calibration of parameters. In each case, a set of latent processes was simulated, given a set of parameter values. Thereafter, the filter was applied to the data (both regular Kalman Filter and Unscented Kalman Filter) to recover these latent state processes. The filter was also used in conjunction with numerical maximisation of the conditional log-likelihood to recover the parameter values.

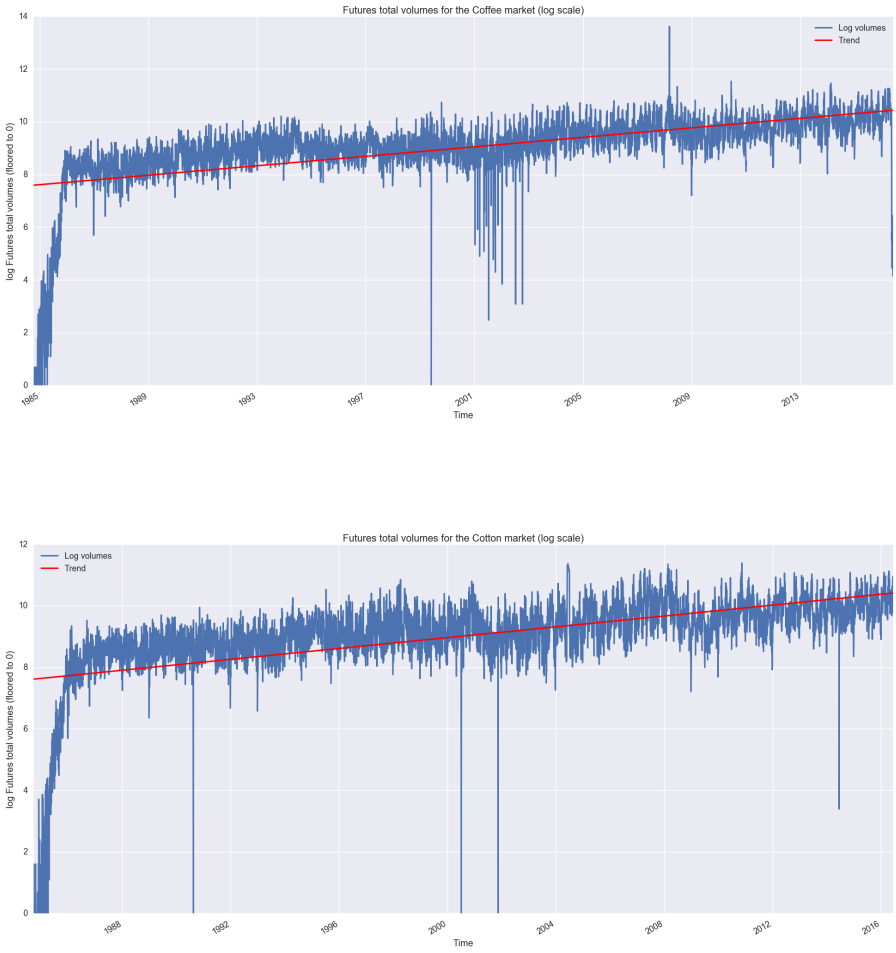


Figure 4: Total futures traded volume during the period analysis for Coffee and Cotton (log scale)

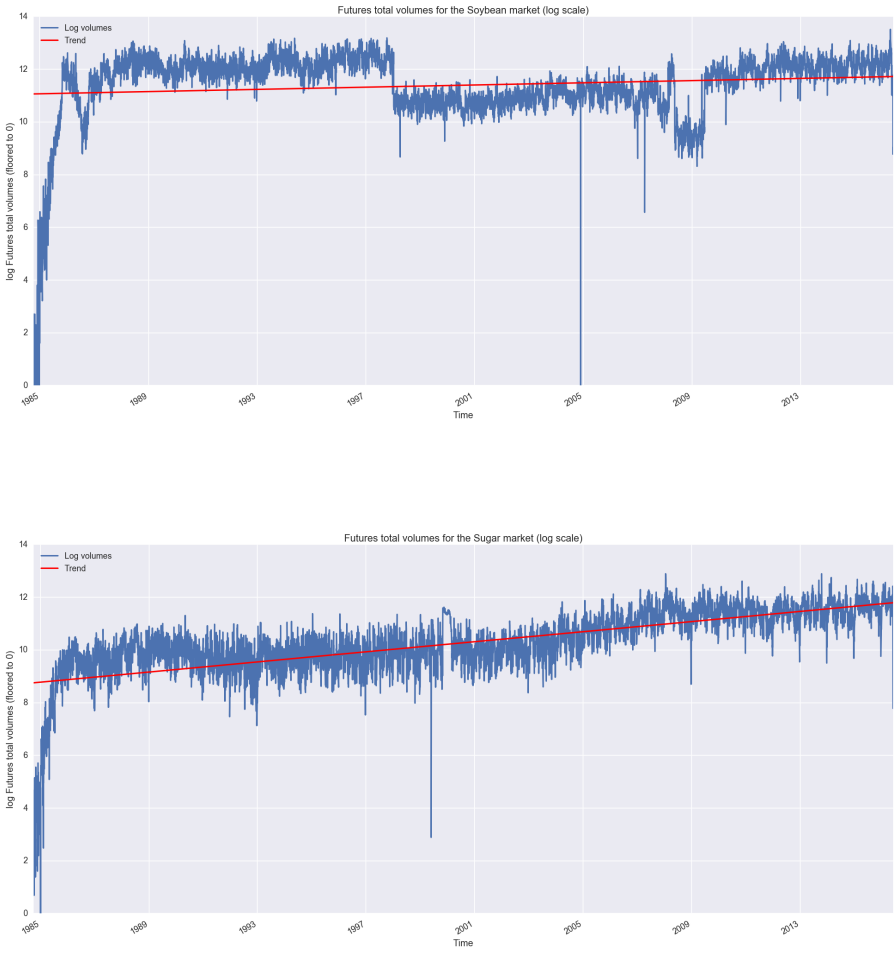


Figure 5: Total futures traded volume during the period analysis for Soybean and Sugar (log scale)

Table 1: Simulation parameters

Parameter	Value	Parameter	Value	Parameter	Value
κ	1.49	λ_χ	0.157	s_2	0.006
σ_χ	0.286	μ_ξ	- 0.0125	s_3	0.0003
σ_ξ	0.145	$\mu_\xi - \lambda_\xi$	0.01	s_4	0.00001
ρ	0.3	s_1	0.042	s_5	0.004

6.1.1 Schwartz and Smith (SS) model

The first data set simulated corresponded to the model specified by Schwartz and Smith (2000) as outlined in section 3.1.1. Using the state space formulation of this model in conjunction with a discretisation of the closed-form solution to the state vector, the state processes χ_t, ξ_t were simulated. Initial values for the processes were $(\chi_0 = 0.119, \xi_0 = 3.02)$ while the remaining model parameters were set as in Schwartz and Smith (2000). These parameters are specified in Table 1.
A total of 286 time points were used, with five initial contract maturities of $1, 5, 9, 13$ and 17 months respectively. The initial simulated observations (contracts) were then rolled forward and a new contract was introduced when a contract reached maturity. A time step of $\frac{7}{360}$ years was used.
After the futures prices were simulated using the latent processes, Gaussian noise ϵ_t was added to each point, such that $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$ with

$$\mathbf{G} = \mathbf{sI} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \end{bmatrix} \times \mathbf{I}$$

Figure 6 displays the simulated futures surface formed by the futures prices at different points in time, for each of the maturities. A cross section of this plot for a is given in figure 7, for a fixed maturity of one month.
In order to better understand the behaviour of model and its parameters, simulations were run to produce various futures price surfaces with perturbations to each of the parameters. The relationship of the parameter with the futures surface can be seen clearly from the diagrams below. The following observations were made:

- κ can be seen to influence speed of reversion to the process mean value. In the case where kappa is increased, we see that over time, the futures surface becomes flatter at a faster rate for a given maturity;

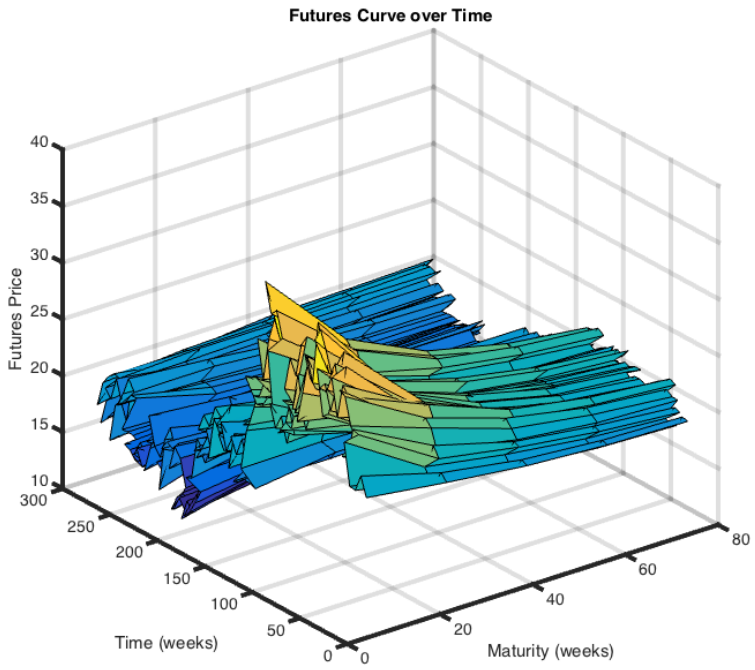


Figure 6: Simulated futures curve

- μ is the drift of the process over time; a positive value causes prices to rise over time, while a negative value causes prices to fall;
- σ_χ relates to the short-term volatility: when it is increased, we see an increase in volatilities, but only for shorter maturities;
- On the other hand σ_ξ reflects the long-term equilibrium volatility. Hence when increased, the volatility of the futures for longer maturities increases too.

λ_χ influences the convexity of the futures surface, with positive values representing a concave surface, or *backwardation* in the futures market. Negative values produce a convex surface, illustrating *contango*.
These behaviours assume a fixed set of parameters with only one parameter being perturbed. In reality, the effect of changing multiple parameters at the same time may not affect the futures prices in such a clear way.

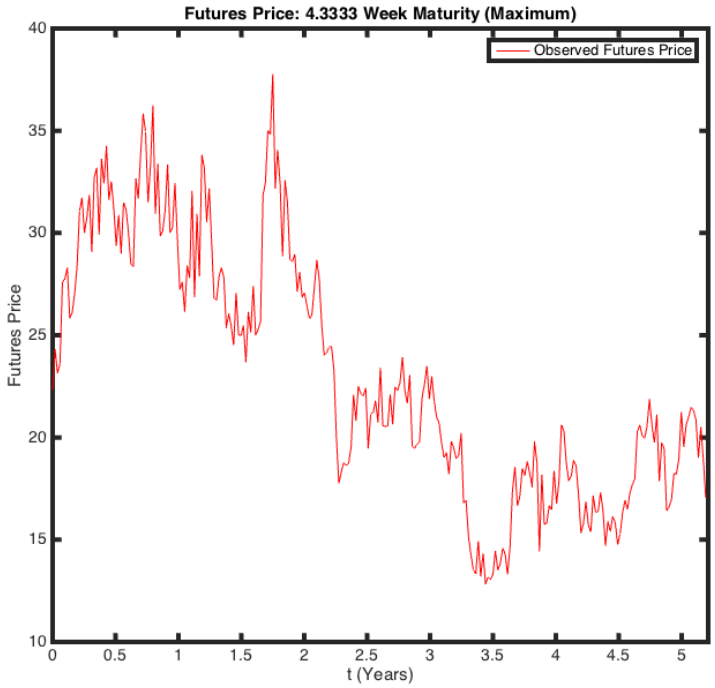


Figure 7: Simulated futures curve cross-section

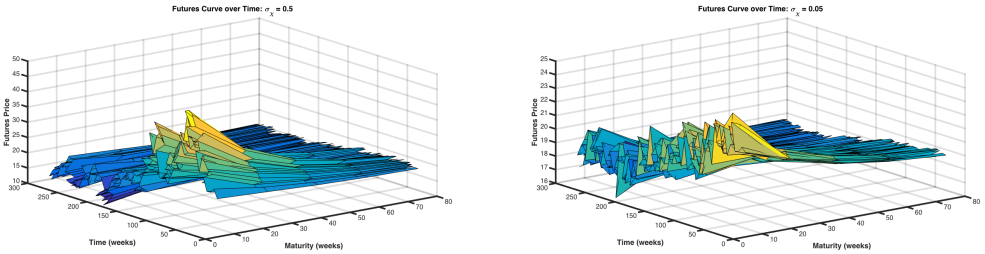


Figure 8: Simulated futures curves

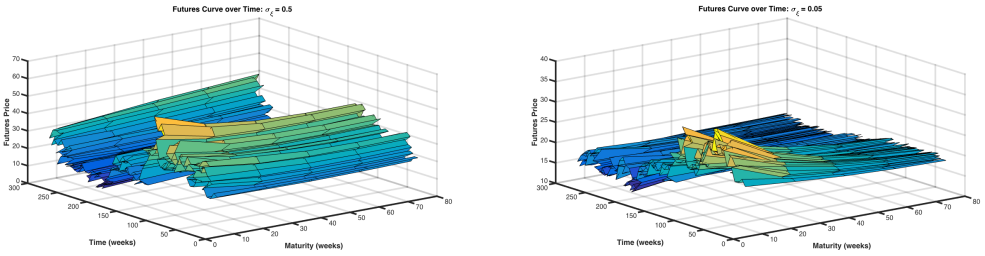


Figure 9: Simulated futures curves

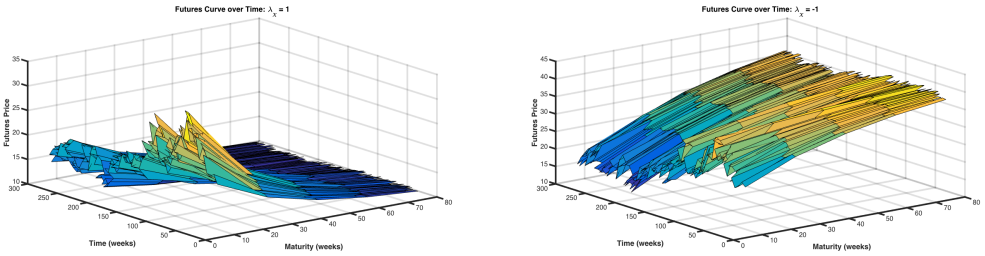
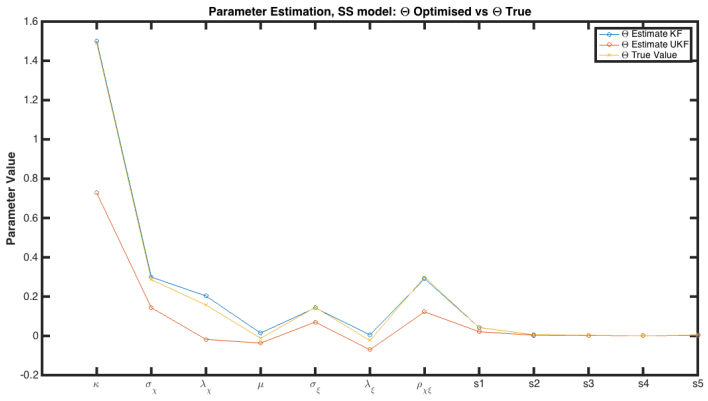


Figure 10: Above: simulated futures curve cross-section; below: parameter estimation comparison.



Based on this simulated data, the original parameters were estimated using `fmincon` in Matlab in conjunction with the log-likelihood function from the Kalman filter derived in section 4.

Figure 11 demonstrates the relative performance of the KF and UKF in estimating the parameters. It was found that the UKF did not perform better than the KF in this example, with greater percentage estimation errors demonstrated. Since the hessian matrix is known to be numerically unstable, error bounds for the parameters were not computed. While alternative solutions are available to deal with this, they were not explored.

In addition to estimating the parameters, a set of profile negative log-likelihood plots were constructed with the aim of identifying which parameters were most difficult to estimate.

From these profile log-likelihoods, one can see that κ and ρ are relatively flat. This confirms the difficulty of estimating the parameters and motivates why κ and ρ

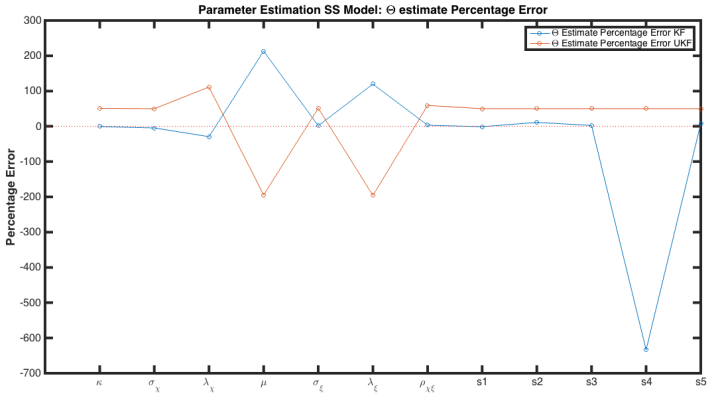
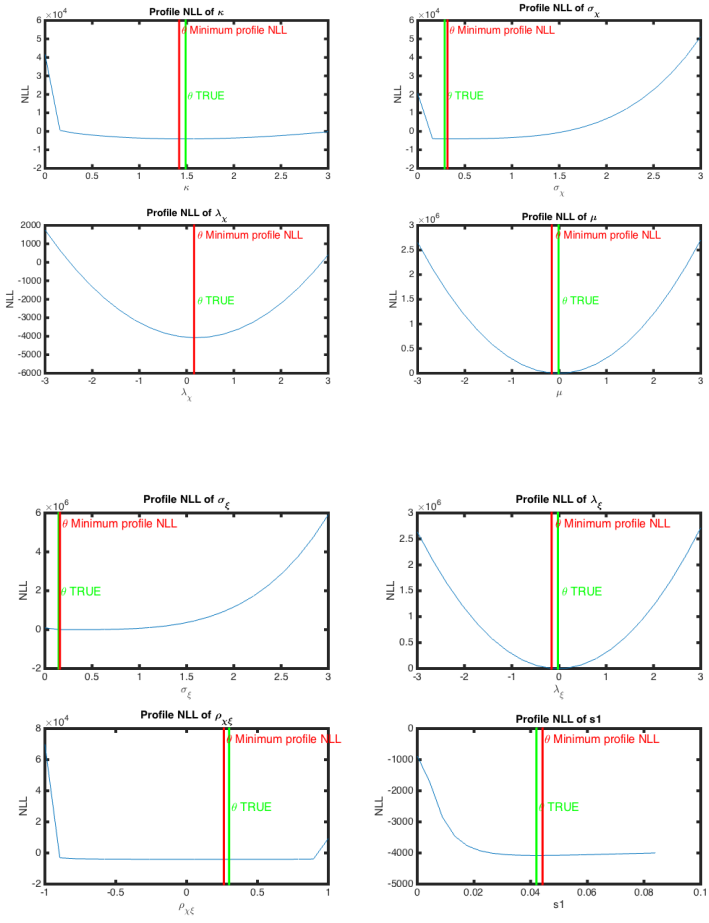
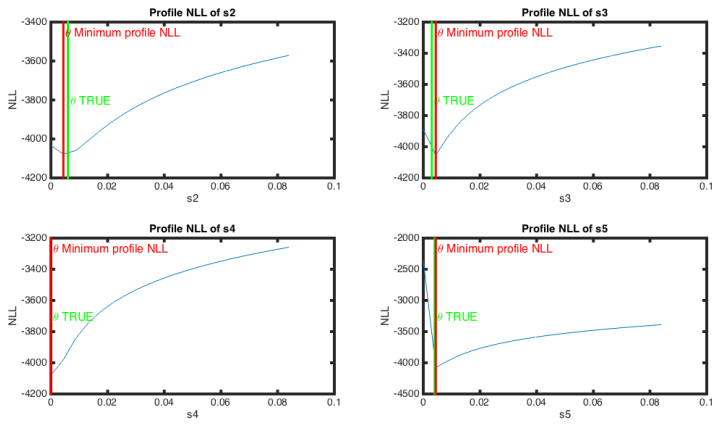


Figure 11: Relative performance of the KF and UKF.





were not estimated as closely (absolutely) by the filters as the other parameters)

6.1.2 SS Model with multiple of error variance

The second set of data simulated also used the SS model, however the observation error covariance matrix used was a scalar multiple of a diagonal matrix. In this case the diagonal elements of the matrix were specified to be the inverse of the associated traded volume for each observed price. This was motivated by a hypothesised inverse relationship between contract volumes traded and the variance of the price observed. This reduced the dimensionality of the estimation problem, since only the constant multiplier for this matrix was needed to be estimated. Constant volumes were used for the purpose of simulation, meaning that the diagonal matrix used was effectively an identity matrix.

Both the KF and UKF were used to estimate the values for the parameters once again. These are compared to the true values in figure 12 while figure 13 demonstrates the associated estimate error.

In this case the UKF performed well, accurately estimating all the parameters.

6.1.3 SS Model with t-distributed errors

Finally, a third case study was done in which a further extension was made: the observation errors were simulated to have a t-distribution with 10 degrees of freedom. This was motivated by the fact that in practice, error distributions tend to have fatter tails than that of a Gaussian distribution. Clearly in such a case, the

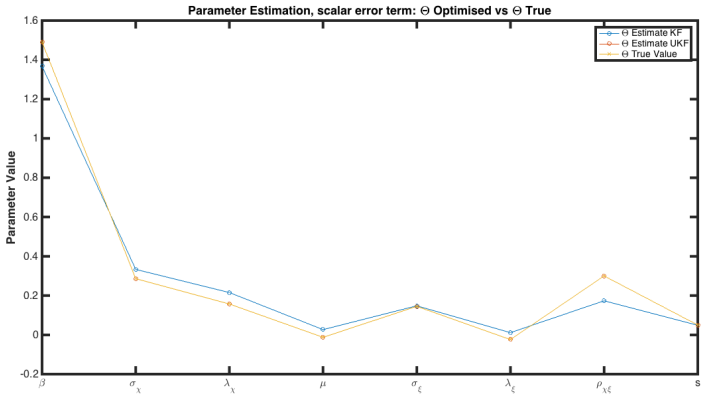


Figure 12: Estimation performance

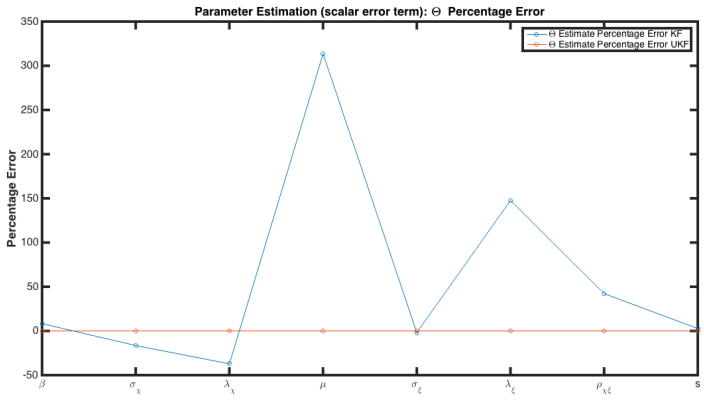


Figure 13: Estimation performance

KF will no longer provide optimal estimates for the parameters, however the Unscented Kalman filter can be used to provide a more appropriate approximation. The case in which the model observation error is t-distributed was outlined in section 4.3.2. The KF was still applied to be able to compare the filtering methods.
The futures price simulated surface is displayed in figure 14. Note the use of the log-scale. The resulting parameters from the KF and UKF estimation procedures given in figure 15 demonstrate that the UKF was again superior to the KF. In this case, the level of accuracy obtained was not to the same degree as the previous case study, where the observation errors were specified to be Gaussian. Correspondingly, the KF produced estimates that were further away from the true parameter

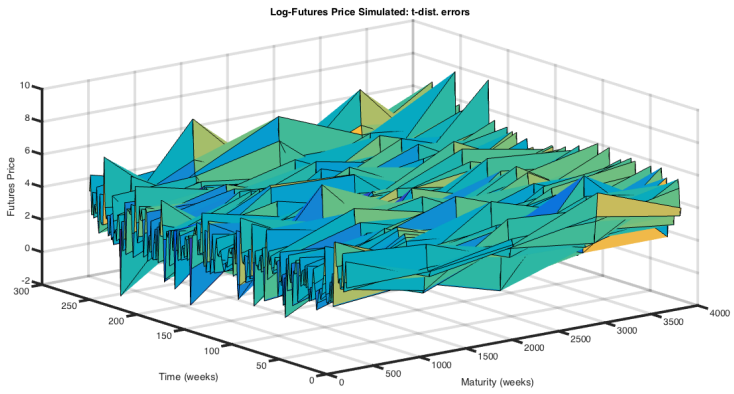


Figure 14: Simulated futures price surface example.

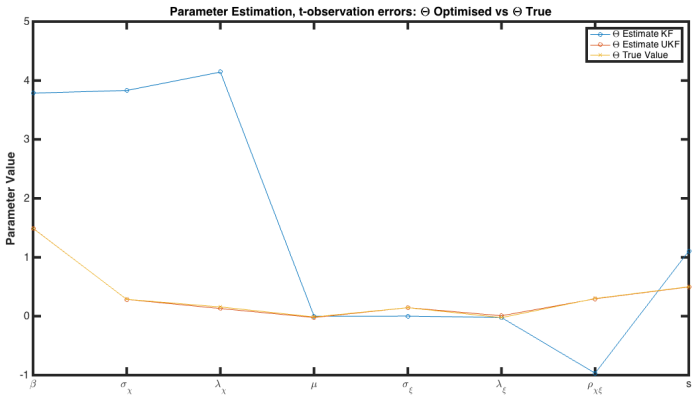


Figure 15: Estimation performance on estimated static parameters.

values than it did in the Gaussian case.

6.2 Real data

Having performed simulations on synthetic data, and established the performance of the model, filtering and calibration, real market data was then used as an input to the computational model.

Data was run on a year-on-year basis, across four commodities (coffee, cotton, sugar, soybeans).

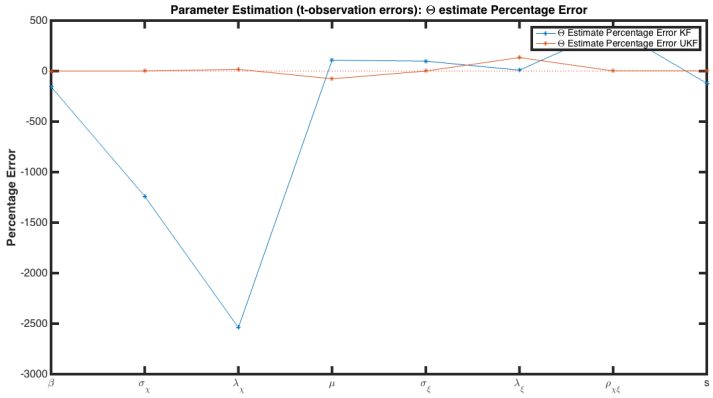


Figure 16: Estimation performance on static parameters.

6.2.1 Model adaptation

Due to changes in the structure of the input data, the computational model of Ames et al. (2016) was adapted slightly. This invovled changes to the way the Kalman filter handled the input contracts and times to maturity. In addition, the possibility of a varying number of contracts per day was coded for, along with the associated time to maturity for the respective contracts.
Raw data, as opposed to filtered data, was used as the input due to the number of gaps present in the filtered data. These gaps prevented the computational model from producing reasonable/accurate results.
A summary of the adaptation to the Kalman filter code of Ames et al. (2016) is as follows:

- Import raw data, with empty cells received as `NaN`;
- Identify the number of contracts per observation (day);
- Trim the data according to the above index;
- Size the required vectors, and matrices according to the index.

The results that were required from the computational model were stored in `Excel` files.

Table 2: Model Parameters

Coffee							
1995		2000		2005		2010	
β	1.03834	β	0.03396	β	0.33605	β	0.01405
σ_χ	0.16037	σ_χ	2.64929	σ_χ	0.18575	σ_χ	9.99997
λ_χ	-0.02298	λ_χ	-0.17657	λ_χ	-0.07882	λ_χ	1.82291
μ	-0.59372	μ	-0.14933	μ	0.05263	μ	0.28595
σ_ξ	0.32586	σ_ξ	2.22709	σ_ξ	0.25614	σ_ξ	9.48593
rmnu	-0.53515	rmnu	0.03218	rmnu	0.10886	rmnu	-1.44911
λ_ξ	-0.07388	λ_ξ	-0.99710	λ_ξ	0.43868	λ_ξ	-0.99964

6.2.2 Inputs and outputs

The inputs to the computational model were the commodity futures prices and the corresponding times to maturity for each of the contracts. Initialisation of the states involved setting the long-run equilibrium to the log of the first futures price, and the short-term deviation to 0.
The outputs retrieved from the model were that of the parameter estimates, latent states (both predicted and updated), and the log futures prices (both predicted and updated). The states would be used to estimate the spot prices for the commodities, with the log futures being used to validate the performance of the model by comparison to the actual observed data.

6.2.3 Gamma vs no gamma

In order to assess the impact of the inclusion of gamma in the model, both versions were run on the same set of raw data. This would allow for comparison of the same parameters and results, and would provide insight into the performance of different model specifications.

6.2.4 Parameter estimates

Estimated parameters for different years are shown in the Tables below (rmnu - risk neutral μ):
It was found that in general the year-on-year parameters did not differ significantly. In certain cases the parameters differ due to the way in which the Kalman filter

estimated the latent states. This will be discussed further in the following section. Any variance in the parameter estimates can be explained by the year-on-year way in which the data was analysed. Years with a greater volatility or prices movements (say a large drop) would exhibit different parameters to a year with more stable observed prices.

6.2.5 Latent processes

The performance of the computational model with regard to the latent processes is of particular importance. As the latent processes are used to estimate the log spot prices, consideration of the latent processes produced can allow for a better understanding of the movement of prices in a particular year. Furthermore, comparison between the results produced by the models with and without gamma would allow for further insight into the each of the commodities model.
In order to illustrate and discuss key observations of the latent processes produced by the computational models, plots for select commodities and years are shown below:

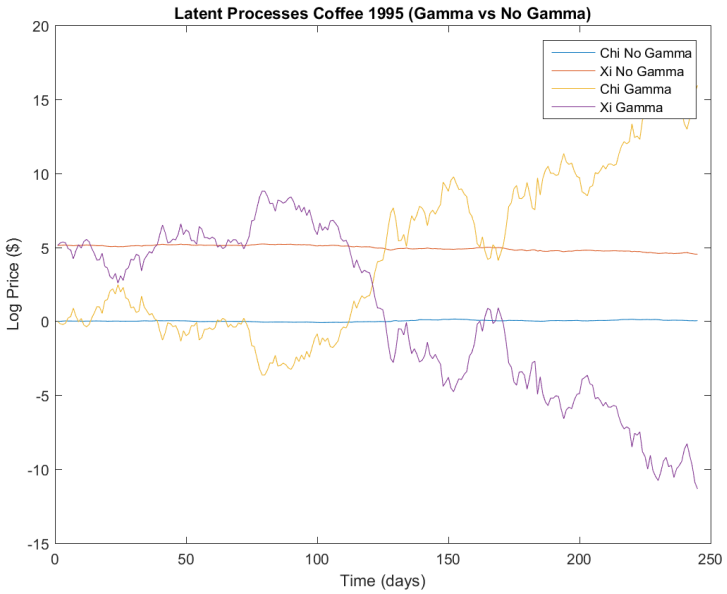


Figure 17: Coffee 1995 latent Gamma vs No Gamma

Figure 17 shows the difference in performance between the gamma and no gamma

models for particular coffee in 1995. It was observed that in this particular case, the latent processes for the gamma model behave as expected, with the long-run equilibrium tracking the log futures futures price and the short-term deviation having only a minor influence. The latent processes for the gamma model do not behave as expected. Both the long-run equilibrium and short-term deviation begin to deviate from their expected values within the first few observations, and this deviation is amplified over the rest of the remaining observations. One aspect to note is the ”mirroring” of the two processes (in the case of the gamma model). This is due to increases in the long-run equilibrium needing to be canceled out by the short-term deviation in order to correctly track the log futures price observed. This increase (or decrease) is then carried over to the next iteration leading to the ”blowing-up” of the short-term deviation and decoupling of the long-run equilibrium from the log of the futures price. Thus the latent processes are no longer performing as initially intended.

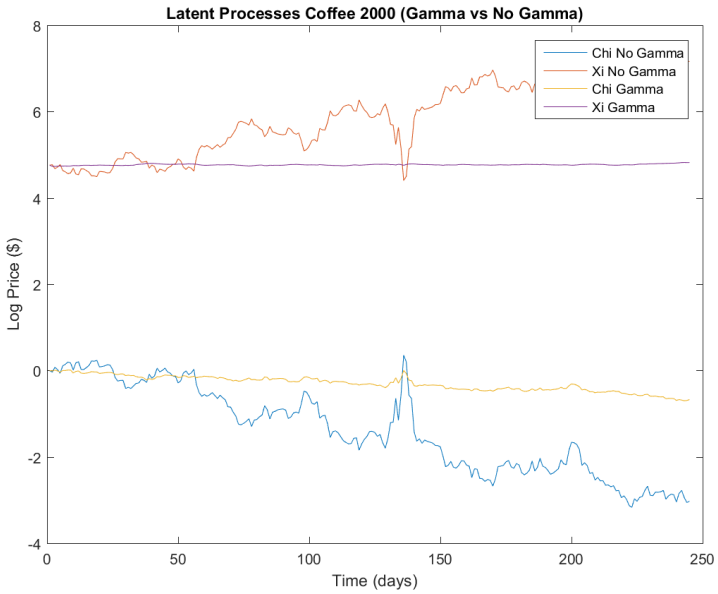


Figure 18: Coffee 1995 latent Gamma vs No Gamma

In the case of coffee 2000, the opposite behaviour was observed between the two models, with the the gamma model exhibiting unstable latent processes. It follows that in certain years the log futures prices observed may be better to suited either

the gamma or no gamma model.
It was noted during the filtering and calibration process that the initial latent process points have a large influence on the dynamics of each of the individual processes. In long-run equilibrium and short-term deviation components contributed equally to the log spot price (in the ”mirroring” seen above).

6.2.6 Observation

The predicted observations (log futures prices) produced by the model provided an insight into the performance of the model by comparison against the observed real world log futures prices. This was useful in understanding the general performance of the computational model, and the validity of the results it produced.
As with the above sections, select commodities and years have to chosen to illustrate key concepts and observations of the observation processes.
Figure 18 shows the log futures panels produced by the computational model (for the no gamma model) and those of the actual observed log futures panels. It can be seen that the model panels capture the real world panels reasonably well, indicating that the model and calibration is performing as desired. It was found that the model panels were less volatile than the actual observed panels. This can be explained
Plots of the predicted and updated log futures prices are shown against the actual observed prices. In the case of multiple contracts, the prices plotted are those corresponding to the contract with the highest traded volume on each day. This is due to these being the most liquid contracts, and thus can be considered to be the best estimate or most accurate prices.
The maximum and minimum futures prices are plotted to compare the model and actual price (for the highest traded volume) against the boundaries. It was seen from the results that in almost all cases the most liquid price was that of the minimum.
In comparing the performance of the gamma and no gamma models estimates of the log futures prices against the actual observed prices, it was found that there was very little between the two models. The performance of each of the models is dependent on the data, with certain subsets being more suited to one of the models.
The unscented Kalman filter was run on a subset of coffee data in order to assess

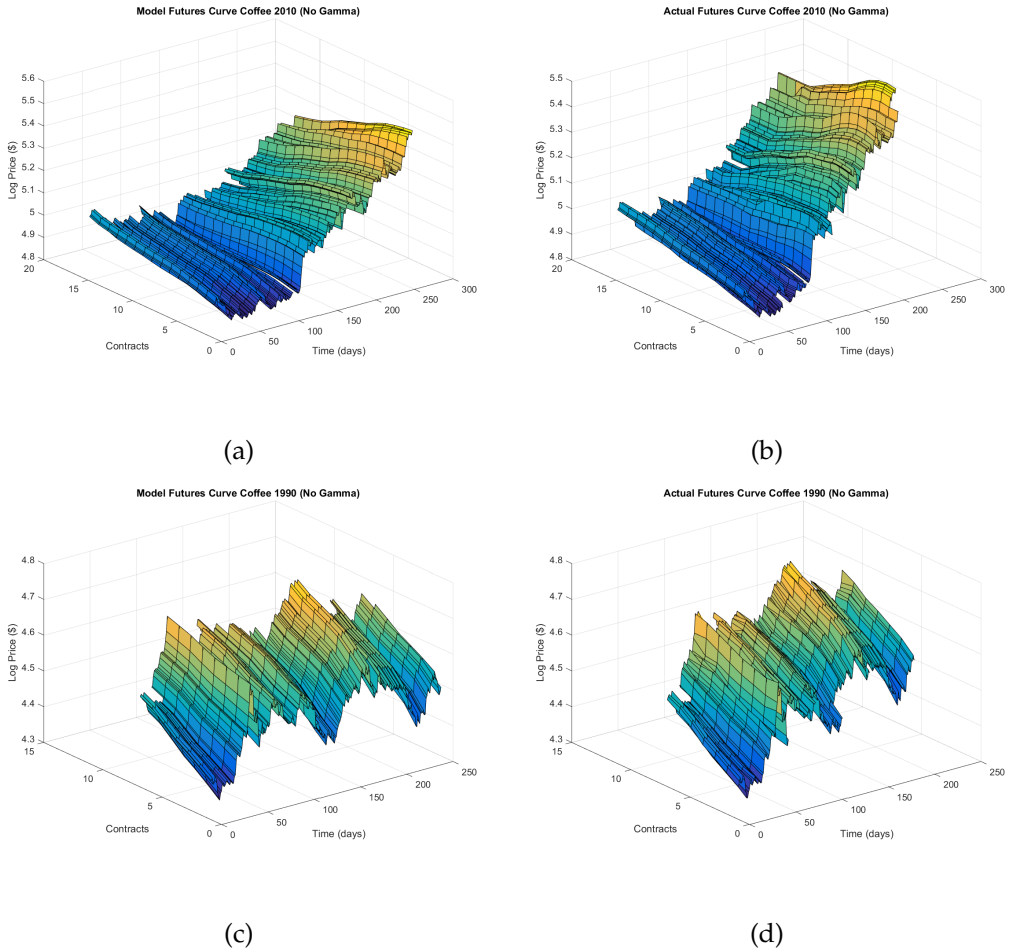


Figure 19: Log futures panels.

its performance on real data. It was found that it did not perform as well as the standard Kalman filter.

6.3 Cointegration and the Johansen Test

Cointegration is a time series analysis of two (or more) series, in order to determine the existence (or lack thereof) of a linear relationship between the two series.

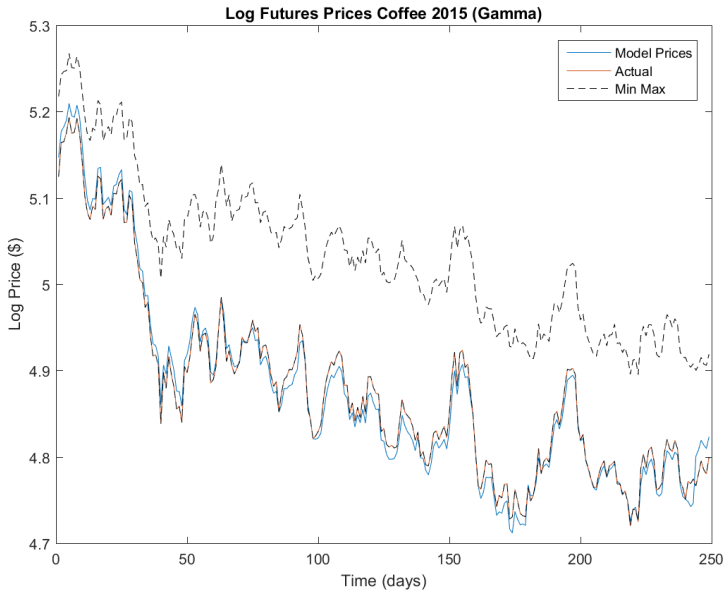


Figure 20

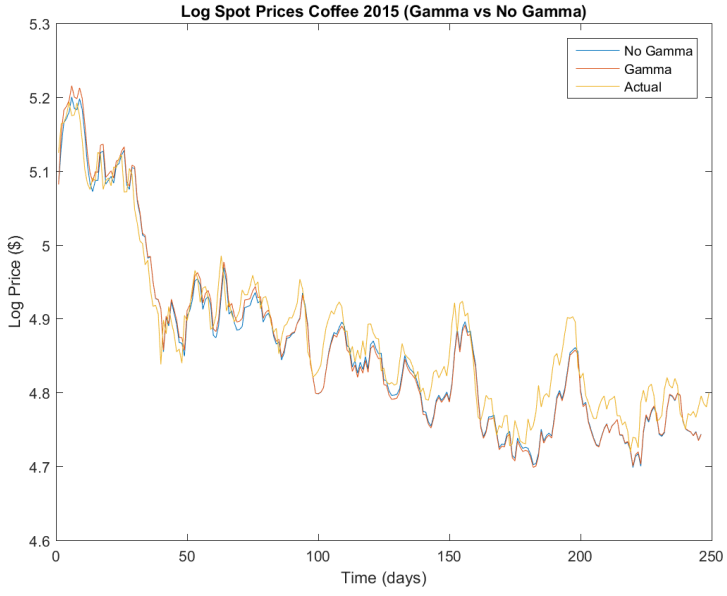


Figure 21

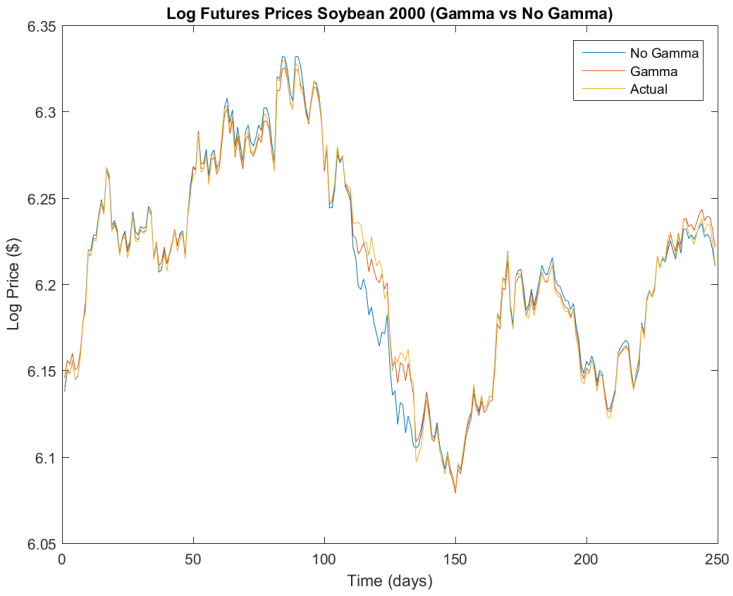


Figure 22

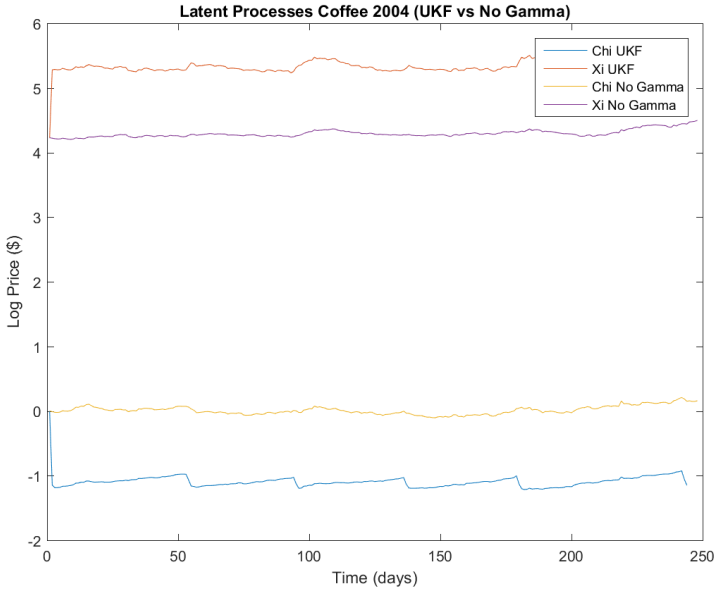


Figure 23

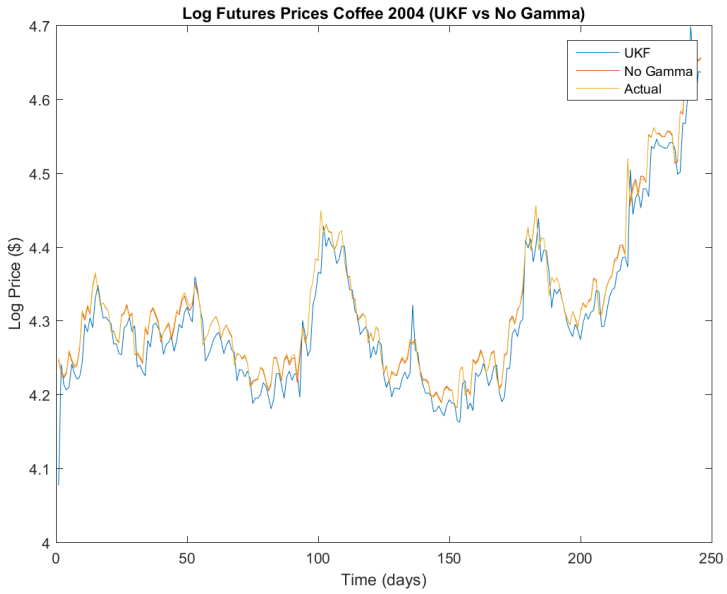


Figure 24

6.3.1 Vector-error correcting model

A vector-error correcting model allows for changes in a variable to be expressed in terms of the deviation from an equilibrium position of the variate and another covariate, with which the relationship is defined below (for the bivariate case).

$$\triangle x_{1t} = \alpha_1(x_{1,t-1} - \beta_1 x_{2,t-1}) + \epsilon_{1t} \tag{65}$$

The general form for the vector error correction model representation of an n-dimensional vector-autocorrelated process of order p is

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x_{t-1}} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma_i}\boldsymbol{\Delta}\mathbf{x_{t-1}} + \phi \mathbf{d_t} + \boldsymbol{\epsilon_t}. \tag{66}$$

Here $\boldsymbol{\beta}$ is the cointegrating matrix, $\boldsymbol{\alpha}$ is the loading matrix, and together they represent the trend $\mathbf{\Pi}$ Cox (2013). It is the rank of $\mathbf{\Pi}$ that the Johansen test aim to test. This is done through , by comparing the maximum likelihood functions against those of models with cointegration r_0 and r_1.

The null hypothesis $\mathcal{H}_0$: rank($\mathbf{\Pi}$) $= r_0$, with the alternative being $\mathcal{H}_1 : r_0 <$

$rank(\mathbf{\Pi}) < r_1$.

6.3.2 Testing for cointegration

The application of the Johansen test on the log spot prices and observed inidices on a year-on-year basis over the period around financialisation would allow for the determination of a cointegration relationship between the two (should one exist). Such a relationship would be a strong indicator of finicialisation, showing that impact of finicialisation on commodity spot prices.
In the case of the data, comparing individual commodities to individual indexes, the rejection of the r_0 hypothesis whilst failing to reject the r_1 hypothesis was considered to be a desirable outcome. This corresponds to the knowledge that there exists some cointegration between the series. Due to time constraints, the test was performed on the coffee and soybean commodities, and the AORD, FTSE, SPX, and NASDAQ indices. The reason for the above was the confidence in the performance of the model in the commodities, and the indices being some of the most frequently traded.
As data for the indices was only available from 1996, the cointegration test was performed over 1996 to 2016. It was found that for the majority of years the test failed to reject the null hypothesis for both the r_0 and r_1 case. For the years for which the test returned cointegration, plots of the normalised inidices and log spot prices are shown such that a visual comparison of the series can be made. It can be seen for these years, there is a clear trend between the movement of the index and log spot prices.
The corresponding cointegration parameters are shown in the table below:

6.3.3 Correlation test

Based on the results of the cointegration test, it was decided to check the correlation between the indices and commodities spot prices. This was a simple procedure and a consistent correlation over a period of years might indicate the presence of a relationship between the indices and commodity spot prices.
The correlation coefficients between coffee and soybean,and certain indices are shown over a period of years. It is evident from the table that there is no sustained correlation between the commodity spot prices and indices. Notably, the

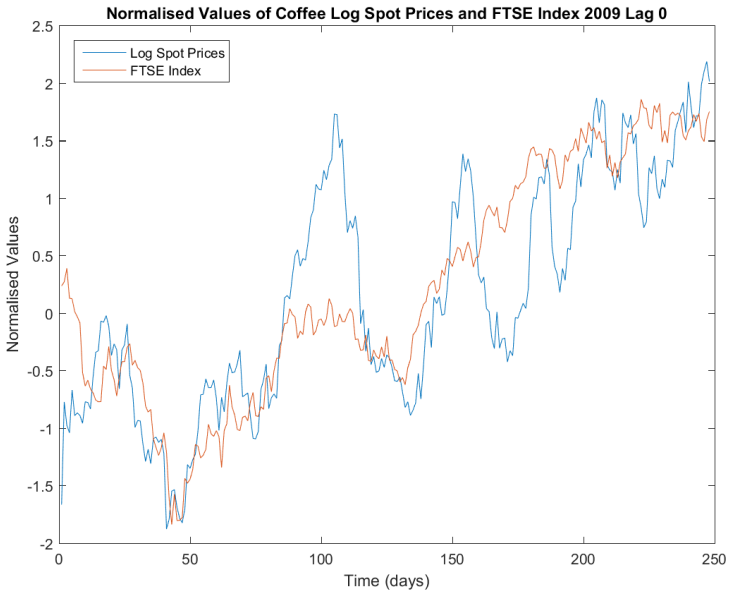


Figure 25

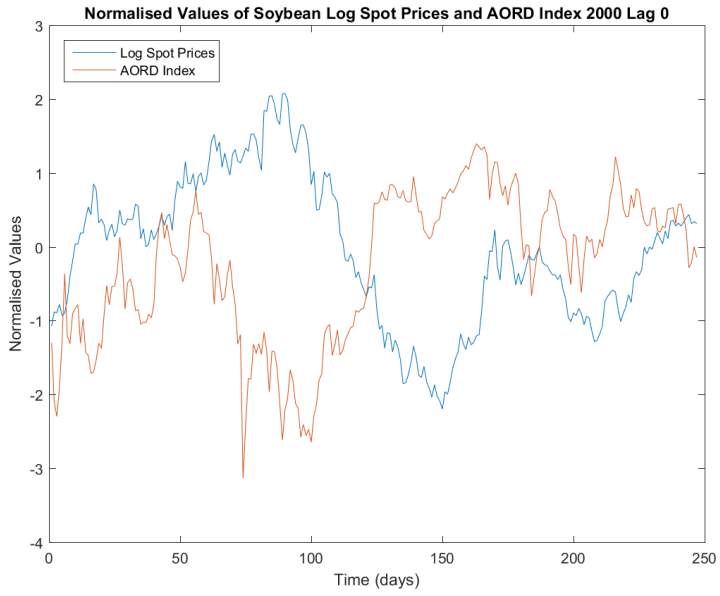


Figure 26

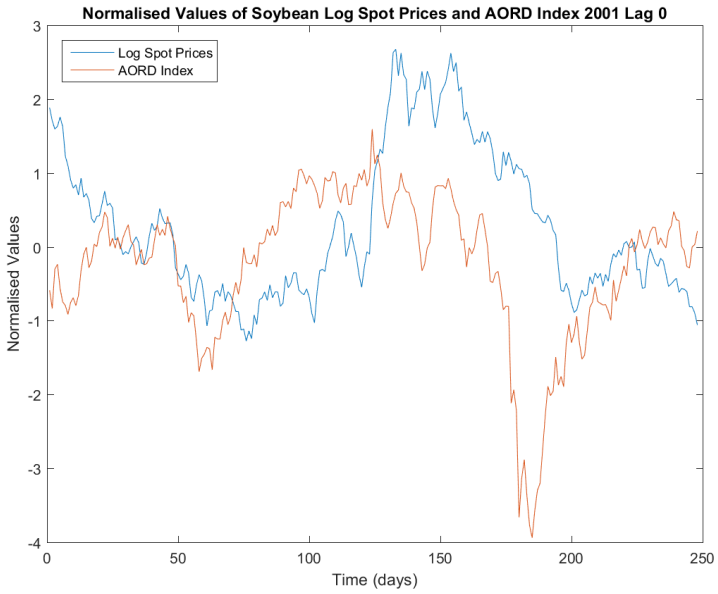


Figure 27

significant changes in the correlation year-on-year indicate that the two series are uncorrelated in general.

6.3.4 Autocorrelation of prices

In order to check for seasonality in the spot prices, the sample autocorrelation was determined on a year-on-year basis. The autocorrelation was performed using a lag of 251, (1 less than the number of observed futures prices per year). This was due to `Matlab`'s implementation of autocorrelation. The autocorrelation function (ACF) was plotted for each year in order to examine whether any seasonality was present in the prices.

There was no consistent pattern in the ACF for both coffee and soybean, indicating that seasonality may not have a large influence on futures prices. Some persistence was observed, however there was no explicit pattern observed that could be linked to seasonality.

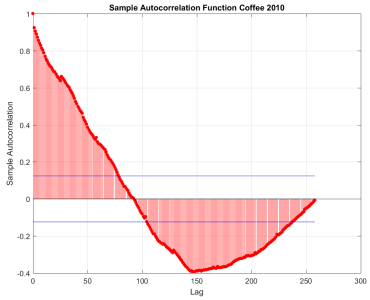
Differences between the ACF's of certain years, though this could be attributed to large movements in the prices around certain times and is not necessarily linked to any seasonality.

Table 3: Cointegration parameters for years in which a cointegration relationship was established

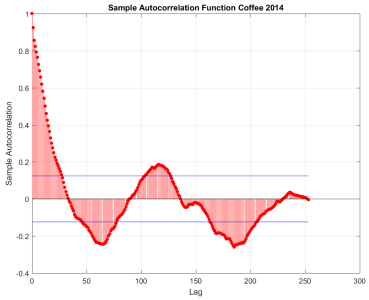
	Cointegration parameters			
Parameter	Coffee - FTSE 2009	Soy - AORD 2000	Soy - AORD 2001	Soy - AORD 2006
r_0				
c_1	0.001	0.000	-0.001	0.000
	3.105	0.426	0.327	3.520
r_1				
A	-0.004	0.000	0.002	0.003
	7.114	-6.768	-1.944	-3.089
B	19.722	16.296	-6.083	-15.414
	-0.003	0.014	0.010	0.006
c_0	-80.685	-144.382	5.693	64.971
c_1	0.003	0.000	0.000	0.003
	0.000	0.000	0.000	0.000

Table 4: Correlation coefficients between select commodities and indices

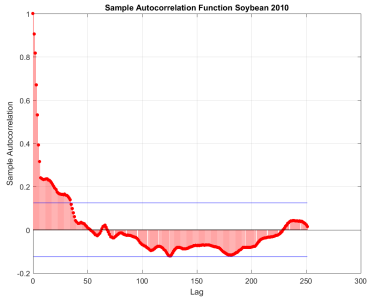
	Correlation coefficient			
Year	Coffee - FTSE	Coffee - SPX	Soy - AORD	Soy - FTSE
1996	0.2365	0.3396	-0.6187	-0.3922
1997	0.0469	-0.0592	-0.5363	-0.6746
1998	0.1618	-0.4320	0.5517	0.4252
1999	0.3530	0.1631	-0.4911	-0.5125
2000	0.0493	0.2409	-0.6202	-0.2208
2001	0.8266	0.7482	0.1244	-0.0558
2002	-0.3108	-0.2419	-0.9158	-0.9240
2003	0.0304	0.0260	0.6494	0.6530
2004	0.6871	0.7674	-0.8091	-0.5270
2005	-0.6914	-0.5564	-0.1335	-0.0207
2006	0.4066	0.5595	0.6686	0.5167
2007	-0.0309	0.1993	0.7630	0.1533
2008	0.7561	0.7013	0.8394	0.8504
2009	0.8125	0.8206	0.0736	0.0589
2010	0.3853	0.3258	0.3030	0.6724



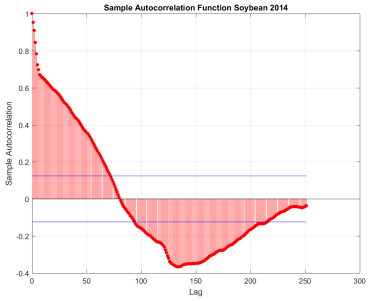
(a)



(b)



(c)



(d)

Figure 28

7 Conclusion and Recommendations

7.1 Conclusions

Based on the results of the modelling, calibration and filtering, and data analysis, the following conlusions can be drawn.
The two-factor commodities model of Schwartz and Smith (2000) was able to simulate the log futures and spot prices to a reasonable degree. The Kalman filter was able to correctly calibrate the model parameters in most cases, with the unscented Kalman filter not being able to perform as well given the real market data.
The lack of cointegration between the commodity log spot prices and the major indices suggests that the financialisation of commodities did not lead to the linking of commodity futures and indices. The significant increase investment can be attributed to the opportunity for increased exposure, with little consideration as to the possible relationship between the two classes (futures and indices). The lack of any correlation between futures and indices further supports this notion.

7.2 Recommendations

Due to the length of the period of financialisation, cointegration periods of longer than one year could be considered. This would allow for any long-run trends between indices and commodities spot prices to be identified.
In addition, cointegration between individual commodities and multiple indices could be considered. However the approach taken above allows for relationships between individual indices and commodities to be determined. This is valuable as depending the countries or areas of production for certain commodities, one may find that these commodities are cointegrated with indices closely related to the area or country. This is an area of extension that could be further researched.

Bibliography

Ames, M., Bagnarosa, G., Peters, G., Shevchenko, P., 2016. Oil price models: Explanatory note.

Basak, S., Pavlova, A., 2015. A model of financialization of commodities. Journal of Finance, forthcoming.

Cheng, I.-H., Xiong, W., 2013. The financialization of commodity markets. Tech. rep., National Bureau of Economic Research.

Cheridito, P., Filipović, D., Kimmel, R. L., 2007. Market price of risk specifications for affine models: Theory and evidence. Journal of Financial Economics 83 (1), 123–170.

Cox, J., 2013. Methodology for joint estimation of static and dynamic parameters in bayesian cvar models. Ph.D. thesis, University College London.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical finance 6 (4), 379–406.

Falkowski, M., 2011. Financialization of commodities. Contemporary Economics 5 (4), 4–17.

Filipovic, D., Larsson, M., 2014. Polynomial diffusions and applications in finance.

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

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

Julier, S. J., Uhlmann, J. K., 1997. New extension of the kalman filter to nonlinear systems. International Society for Optics and Photonics, pp. 182–193.

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

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

Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financial Analysts Journal 68 (5), 54–74.

Wan, E. A., Van Der Merwe, R., 2000. The unscented kalman filter for nonlinear estimation. In: Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000. Ieee.

Yan, X. S., 2002. Valuation of commodity derivatives in a new multi-factor model. Review of Derivatives Research 5 (3), 251–271.

XVA Metrics for CCP Optimisation

Team 2

DOROTA TOCZYDŁOWSKA, University College London
NDINAE MASUTHA, University of Cape Town
VUYO MAKHUVHA, University of Cape Town
DI WANG, University College London

Supervisor:
STÉPHANE CRÉPEY, Université d'Evry

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

In the aftermath of the financial crisis, the banking regulators undertook a number of initiatives to cope with counterparty credit risk (CCR). One major evolution is the introduction of central counterparties (CCPs), which are also known as clearing houses. In contrast to the traditional bilateral transactions, the clearing house stands between its member banks and serves as an intermediary during the completion of their transactions. Banks that use services provided by a CCP are called the clearing members of the CCP, with involved portfolios referred as their CCP portfolios. The main service provided by a CCP is to ensure collateralisation of transactions between its clearing members and the liquidation of the CCP portfolios of defaulted members.

In order to reduce counterparty risk, the CCP asks clearing members to meet specific collateralisation requirements. Apart from the variation and initial margins (VM and IM respectively) that are also required in bilateral trading, the clearing members need to contribute to a default fund (DF), which forms one additional layer of protection from extreme or systemic risk of default.

This report is concerned with challenging the current design of CCP collateralisation scheme in two directions as suggested in (Albanese (2015)). The first one is to challenge the "cover 2" EMIR rule which is commonly used for determining default fund, with a new method based on the so called economic capital (EC). We also compare the traditional IM-based allocation of the default fund with a CCP risk-measure, based allocation of the default fund that is obtained in our new method. The second one is to compare the traditional approach of funding initial margin through unsecured borrowing with a new one whereby a third party provides the IM in exchange of some service fee.

2 Introduction

In the aftermath of the financial crisis, the banking regulators undertook a number of initiatives to cope with counterparty credit risk (CCR). One major evolution is the introduction of central counterparties (CCPs), which are also known as clearing houses. In contrast to the traditional bilateral transactions, the clearing house stands between its member banks and serves as an intermediary during the completion of their transactions. Banks that use services provided by a CCP are called the clearing members of the CCP, with involved portfolios referred as their CCP portfolios. The main service provided by a CCP is to ensure collateralisation of transactions between its clearing members and the liquidation of the CCP portfolios of defaulted members.

In order to reduce counterparty risk, the CCP asks clearing members to meet specific collateralisation requirements. Apart from the variation and initial margins (VM and IM respectively) that are also required in bilateral trading, the clearing members need to contribute to a default fund (DF), which forms one additional layer of protection from extreme or systemic risk of default.

This report is concerned with challenging the current design of CCP collateralisation scheme in two directions as suggested in (Albanese (2015)). The first one is to challenge the "cover 2" EMIR rule which is commonly used for determining default fund, with a new method based on the so called economic capital (EC). We also compare the traditional IM-based allocation of the default fund with a CCP risk-measure, based allocation of the default fund that is obtained in our new method. The second one is to compare the traditional approach of funding initial margin through unsecured borrowing with a new one whereby a third party provides the IM in exchange of some service fee.

3 Basic Setup and Terminology

The variation margin (VM) is the mark-to-market value of the contracts exchanged between the clearing members daily. The initial margin (IM) for each individual clearing member, which is also updated daily, was previously computed using the scenario-based "SPAN" methodology, but at present the majority of CCPs use Value at Risk (VaR) methodologies. The most common of these is the Value at Risk of the one week profit and loss of the portfolio held by each clearing member.

Within Europe, the European Market Infrastructure Regulation (EMIR) "Cover 2" rule is used to determine the size of the default fund. This stipulates that the default size should at least be the greater of the largest exposure or the sum of the second and third largest exposures "under extreme but plausible market conditions" (European Parliament, 2012). This default fund is to be used when the loss of a defaulting member exceeds the sum of its variation margin and of its initial margin. The default fund contribution (DFC) of the defaulted member should be used up to cover this deficit, however if this is not sufficient the default fund contributions of other clearing members should be used. EMIR suggests that the default fund should be allocated to members in a manner proportional to their individual exposures. It is common practice that it is allocated proportionally to initial margins.

While margins and default fund are beneficial from a risk management perspective, those collateralisation requirements generate substantial costs for clearing members, e.g. costs for funding the margins and compensation to shareholders for putting their equities at risk into the pool of default fund. For a clearing member, the cost of the central clearing framework is mainly composed of two parts: the margin valuation adjustment (MVA) corresponding to the cost of funding its initial margin, and the capital valuation adjustment (KVA) which is the cost of remunerating shareholders for putting their capital at risk in the default fund. The cost of funding variation margin is typically negligible as variation margins track the mark-to-market of the portfolio on a daily basis. The credit valuation adjustment (CVA) of a clearing member, or losses on the default fund upon default of other members, is also typically small given the high level of collateralisation of

centrally cleared trading,

Banks want to pass their costs on to their clients, while the evaluation of these costs might be difficult at times. This difficulty is due to the fact that members do not know which IM model will be employed by the CCP in future, as well as the fact that the default fund size is dependent on the CCPs entire portfolio which is not known to an individual.

We consider a CCP with n members, labeled by $i \in N = \{1, 2, ..., n\}$. The following notation is used.

- τ_i: the default time of the ith member,
- $\mathbb{E}_t$: the conditional expectation unber the risk neutral measure $\mathbb{Q}$,
- $\beta_t = e^{-\int_0^t r_s ds}$: the discount factor based on a OIS (overnight index swap) rate process r,
- D_t^i: the cumulative contractual cash flows process of the ith member's CCP portfolio,
- $\Delta_t^i = \int_{\tau_i}^t e^{\int_{\tau_i}^s r_u du} dD_s^i$: the accumulated contractual cash flows of the member i unpaid between the default of the member i and later time t
- $P_t^i = \mathbb{E}_t[\int_t^T \beta_t^{-1} \beta_s dD_s^i]$: the mark-to-market value of the ith member's CCP portfolio, with final maturity time T of all portfolios.
- δ: the length of the liquidation period of a defaulted member's CCP portfolio, usually set as one week,
- $\mathbb{V}a\mathbb{R}$: value at risk.
- $\mathbb{ES}$: expected shortfall,
- W_t: Brownian motion under the risk-neutral measure $\mathbb{Q}$
- Φ : the standard normal cdf.

4 Model

We use the CCP model of Armenti and Crépey (2015). For our programming we enriched the CCP epython library of this paper by new functionalities, so that we can make comparisons between our suggested methods and those currently employed by the CCPs.

4.1 Default Times

We use a common shock model, the dynamic Marshall-Olkin (DMO) copula model (see Crépey and Song (2014)), for the default times τ_i of nine members. The main idea of DMO copula model is that defaults for different members can happen simultaneously with positive probabilities. This model can then be efficiently calibrated to both individual marginal and portfolio credit data.

Before we go through the DMO model in details, we need to understand what a "shock" is. Remember that our CCP has n members labeled by $i \in N = \{1, 2, ..., n\}$. Then shocks are simply pre-specified subsets of members, i.e., the singletons $\{1\}, \{2\}, ..., \{n\}$, for single defaults, and a small number of simultaneous defaults, e.g., the shock $\{1, 3, 7\}$ represents the event that whoever among the members 1, 3 and 7 is still alive defaults at that time.

For a given family $\mathcal{Y}$ of shocks, the arrival time η_Y of a shock $Y \in \mathcal{Y}$ is an independent standard exponential random variable with intensity function λ_Y. Hence, for each clearing member i, we have

$$\tau_i = \min_{\{Y \in \mathcal{Y}; i \in Y\}} \eta_Y, i \in N,$$

which means that the default time for member i is the first time that a shock Y such that $i \in Y$, is observed. As a consequence, the intensity λ_i of τ_i is given by

$$\lambda_i = \sum_{\{Y \in \mathcal{Y}; i \in Y\}} \lambda_Y.$$

Example 4.1. Consider a family of shocks $\mathcal{Y} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 3\}, \{2, 3\}, \{1, 2, 4, 5\}\}$ (with $n = 5$). The following process shows a possible default path in the model.

$t = 0.9:$	$\{3\}$	$1 \quad 2 \quad \textcircled{3} \quad 4 \quad 5$	$\tau_3 = 0.9$
$t = 1.4:$	$\{5\}$	$1 \quad 2 \quad 3 \quad 4 \quad \textcircled{5}$	$\tau_5 = 1.4$
$t = 2.6:$	$\{1,3\}$	$\textcircled{1} \quad 2 \quad 3 \quad 4 \quad 5$	$\tau_1 = 2.6$
$t = 5.5:$	$\{1,2,4,5\}$	$1 \quad \textcircled{2} \quad 3 \quad \textcircled{4} \quad 5$	$\tau_2 = \tau_4 = 5.5$

At time $t = 0.9$, shock $\{3\}$ happens and results in the default of member 3. This is the first time that a shock involving member 3 appears, hence the default time for member 3 is $\tau_3 = 0.9$. At $t = 1.4$, member 5 defaults as the consequence of the shock $\{5\}$ and $\tau_5 = 1.4$. Then at time 2.6, the shock $\{1, 3\}$ triggers the default of member 1 alone as member 1 has already defaulted. Finally, only member 2 and 4 default simultaneously at $t = 5.5$ since the other two have defaulted before.

4.2 Driving Asset

We consider a stylised swap for an underlying interest rate process S with strike rate $\bar{S}$ and maturity T as our CCP portfolios' driving asset. At discrete time points T_l: $T_1 < T_2 < ... < T_d = T$, the member either pays an amount of $h_l \bar{S}$ and receives $h_l S_{T_{l-1}}$ (long position) or generates the opposite cash flows (short position), where $h_l = T_l - T_{l-1}$. We also assume that the underlying S_t follows the standard Black-Scholes model with risk-neutral drift κ and volatility σ, i.e.,

$$dS_t/S_t = \kappa dt + \sigma dW_t.$$

Here W_t stands for the risk-neutral Brownian motion. By applying Itô's Lemma to $f(S) = ln(S)$, we obtain the conditional expected value of S_{t_2} for an arbitrary $t_2 > t_1 > 0$ given its value at t_1:

$$\mathbb{E}_{t_1}[S_{t_2}] = S_{t_1} e^{\kappa(t_2 - t_1)}. \tag{1}$$

For $t \in [T_0 = 0, T_d = T]$, we denote the smallest $T_l > t$ by T_{l_t}, i.e., $T_{l_t - 1} \leq t < T_{l_t}$.

Then the mark-to-market value of a short position in the swap is given by

$$\begin{aligned} P_t &= \text{discounted expected future cash flows} \\ &= \mathbb{E}_t[\beta_t^{-1}\beta_{T_{l_t}} h_{l_t}(\bar{S} - S_{T_{l_t-1}}) + \sum_{l=l_t+1}^{d} \beta_t^{-1}\beta_{T_l} h_l(\bar{S} - S_{T_{l-1}})] \\ &= \beta_t^{-1}\beta_{T_{l_t}} h_{l_t}(\bar{S} - \mathbb{E}_t[S_{T_{l_t-1}}]) + \sum_{l=l_t+1}^{d} \beta_t^{-1}\beta_{T_l} h_l(\bar{S} - \mathbb{E}_t[S_{T_{l-1}}]) \\ &= \beta_t^{-1}\beta_{T_{l_t}} h_{l_t}(\bar{S} - S_{T_{l_t-1}}) + \beta_t^{-1} \sum_{l=l_t+1}^{d} \beta_{T_l} h_l(\bar{S} - S_t e^{\kappa(T_{l-1}-t)}). \end{aligned}$$

We denote the second part of the formula by

$$P_t^\star = \beta_t^{-1} \sum_{l=l_t+1}^{d} \beta_{T_l} h_l(\bar{S} - S_t e^{\kappa(T_{l-1}-t)},) \tag{2}$$

which represents the current value of future cash flows beyond the next one.

The following parameters are used:

$$r = 2\%,\ S_0 = 100,\ \kappa = 12\%,\ \sigma = 20\%,\ h_l = 3 \text{ months},\ T = 5 \text{ years},$$

Moreover the nominal (Nom) of the swap is set so that each leg equals one at time 0. Figure 1 shows the resulting mark-to-market process of the swap viewed from a party with a long unit position in the swap, i.e., the process $(-P)$, where the increasing term structure of the underlying interest rate makes the swap in the money on average. This avoids the case that the mark-to-market process of the swap is zero and not give rise to any adjustments due to a flat interest rate.

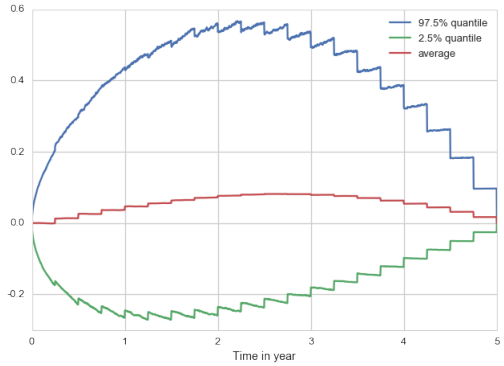


Figure 1: The mean and quantiles as a function of time of the mark-to-market process of the swap from the point of view of a party paying fix and receiving floating, calculated by Monte Carlo simulation of the process (S_t).

4.3 CCP Portfolios

We consider a CCP with 9 members chosen among 125 names of the CDX index as of 17 December 2017. The default times of those members are modeled by a DMO model with 5 common shocks, with intensities λ_Y calibrated to the corresponding CDS (credit default swap) and CDO (collateralised debt obligation) data. The following table shows the credit spreads $\sum_i$ and swap positions ω_i of these nine members. In particular, $P^i = \omega_i P$.

$\sum_i$	45	52	56	61	73	108	176	367	1053
ω_i	(9.20)	1.80	4.60	(1.00)	6.80	(0.80)	13.80	(8.80)	(7.20)

Table 1: (Top) Average 3 and 5 years CDS spread $\sum_i$, in basis points (bp). (Bottom) Short swap positions ω_i of each member, where parentheses mean negative numbers.

5 Default Fund

This section will describe the calculation of DF and its allocation to clearing members as default fund contributions (DFC). Currently, the default fund is calculated using the EMIR Cover 2 rule which prescribes that the default fund is at least the value of the maximum between the largest exposure and the sum of the second and third largest exposures faced by the CCP, where exposures are regulatory defined the way explained in (Armenti and Crépey, 2015, Section A). The default fund is then allocated propotionally to the IM of each clearing member. We regard these as rather ad hoc rules and seek a risk-measure based rule to evaluate the required size and/or allocation of the default fund. Namely, we will set the size of DF to an expected shortfall of the one year ahead profit and loss of the CCP.

5.1 CVA from the CCP's perspective

In order to size the default fund, we define CVA^{ccp}, the CVA of the CCP as a whole, as the total discounted expectation of losses beyond the margins of the clearing members when no default fund is assumed. That is,

$$\mathrm{CVA}_t^{ccp} = \sum_i \mathbb{1}_{\tau_i > t} \mathbb{E}_t[\beta_t^{-1} \beta_{\tau_i+\delta}(P^i_{\tau_i+\delta} + \Delta^i_{\tau_i+\delta} - P^i_{\tau_i} - IM^{i,\alpha}_{\tau_i})^+]$$

The following result is established by standard Gaussian computations in Armenti and Crépey (2015).

Lemma 5.1. *We have*

$$\mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} + \Delta^i_{s+\delta} - P^i_s - IM^{i,\alpha}_s)^+] = Nom \times |\omega_i| e^{-\kappa s} S_s A^\alpha_i(s),$$

where α is the confidence level used to set the IM and

$$A^\alpha_i(t) = (1-\alpha) \times \begin{cases} \left(e^{\sigma\sqrt{\delta}\Phi^{-1}(\alpha)} - e^{\sigma\sqrt{\delta}\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}}\right) e^{-\frac{\sigma^2}{2}\delta} \sum_{l=l_{s+\delta}}^d \beta_{T_l} h_l e^{\kappa T_l}, & \omega_i > 0 \\ -\left(e^{\sigma\sqrt{\delta}\Phi^{-1}(1-\alpha)} - e^{-\sigma\sqrt{\delta}\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}}\right) e^{-\frac{\sigma^2}{2}\delta} \sum_{l=l_{s+\delta}}^d \beta_{T_l} h_l e^{\kappa T_l}, & \omega_i \le 0 \end{cases}$$

∎

Proposition 5.1. *We have*

$$\beta_t \mathrm{CVA}_t^{ccp} = Nom \times \sum_i \mathbb{1}_{\tau_i > t} e^{-\kappa t} S_t \int_t^T \lambda_i e^{-\lambda_i (t-s)} |\omega_i| A^\alpha_i(s) ds.$$

Proof. We have

$$\begin{aligned}
\beta_t \mathrm{CVA}_t^{ccp} &= \sum_i \mathbb{1}_{\tau_i > t} \mathbb{E}_t[\beta_{\tau_i+\delta}(P^i_{\tau_i+\delta} + \Delta^i_{\tau_i+\delta} - P^i_{\tau_i} - IM^{i,\alpha}_{\tau_i})^+] \\
&= \sum_i \mathbb{1}_{\tau_i > t} \mathbb{E}_t[\mathbb{E}_{\tau_i -}[\beta_{\tau_i+\delta}(P^i_{\tau_i+\delta} + \Delta^i_{\tau_i+\delta} - P^i_{\tau_i} - IM^{i,\alpha}_{\tau_i})^+]] \\
&= \sum_i \mathbb{1}_{\tau_i > t} \mathbb{E}_t \int_t^T \mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} + \Delta^i_{s+\delta} - P^i_s - IM^{i,\alpha}_s)^+] \lambda_i e^{-\lambda_i (t-s)} ds \\
&= Nom \times \sum_i \mathbb{1}_{\tau_i > t} \mathbb{E}_t \int_t^T e^{-\kappa s} S_s \lambda_i e^{-\lambda_i (t-s)} |\omega_i| A^\alpha_i(s) ds,
\end{aligned}$$

by Lemma 5.1. We conclude the proof by (1). ∎

5.2 DF as Expected Shortfall

We propose to set the default fund at time t as the economic capital of the CCP in the sense of the expected shortfall of its one-year ahead loss and profit, i.e.

$$DF_t = \mathbb{ES}_t^\alpha\Big(\sum_{t\le\tau_i<t+1} \beta_{\tau_i+\delta}(P^i_{\tau_i+\delta} + \Delta^i_{\tau_i+\delta} - P^i_{\tau_i} - IM^i_{\tau_i})^+ - (\beta_t \mathrm{CVA}^{ccp}_t - \beta_{t+1}\mathrm{CVA}^{ccp}_{t+1})\Big) \quad (3)$$

The random variable under $\mathbb{ES}_t^\alpha$ represents the one year ahead loss and profit of the CCP without the default fund. Setting the size of the default fund to the expected shortfall of the loss and profit at level α ensures the sustainability of the CCP to an adverse yearly scenario of probability $(1-\alpha)$.

Figure 1 shows the default fund based on economic capital as a function of time for $\alpha = 70\%, 92\%$ and 97.5% (top to bottom). For too low quantile levels, the VaR does not even capture the tail of the distribution, which is only visible through the ES. The four curves in each graph show value at risks versus expected shortfalls, accounting or not for the CVA terms in (3). When the quantile levels increase, the impact of the CVA terms in (3) decreases, which confirms the intuition that the tail of the distribution should be driven by the counterparty default losses (denoted by L in the captions of the curves). Figure 2 shows the time-0 default fund based on economic capital as a function of α. The value $\alpha = 92.001\%$ calibrates the economic capital based default fund to a cover two default fund itself computed based on IM of quantile level 70%). In other terms, the size of the default fund determined by the cover 2 rule corresponds to a CCP robust to $\approx 100/8 = 12.5$-yearly adverse events. Figure 3 shows the time-0 default fund allocations based on initial margins, member incremental economic capital and member incremental KVA (top to bottom), for a quantile level of $\alpha = 92\%$. In the case of our portfolios driven by a single asset, the initial margins are simply proportional to the sizes $|\omega_i|$ of the positions of the members, whereas the member incremental economic capital or KVA allocations also incorporate the information of their credit spreads.

5.3 KVA from a CCP perspective

The KVA from a CCP perspective is an estimate of the cost for the CCP of the remuneration of the shareholders of the clearing members for their capital at risk

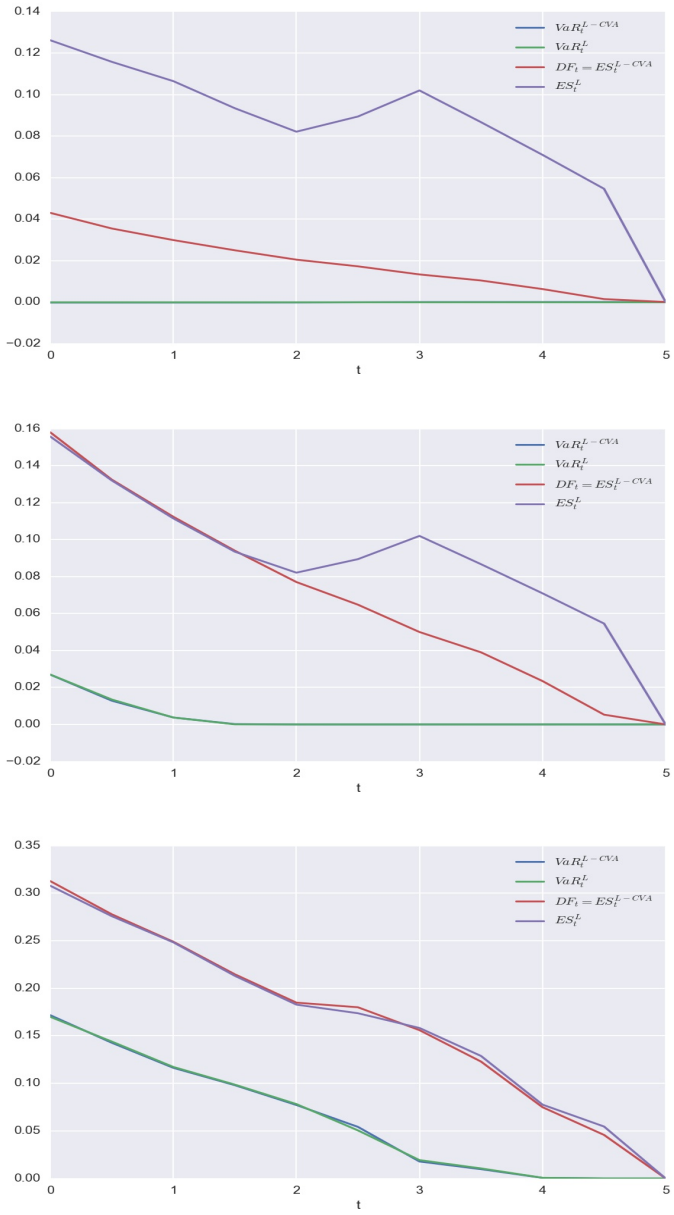


Figure 1: Default fund based on economic capital as a function of time for $\alpha = 70\%, 92\%$ and 97.5% (top to bottom). Four curves in each graph: value at risks versus expected shortfalls, accounting or not for the CVA terms in (3).

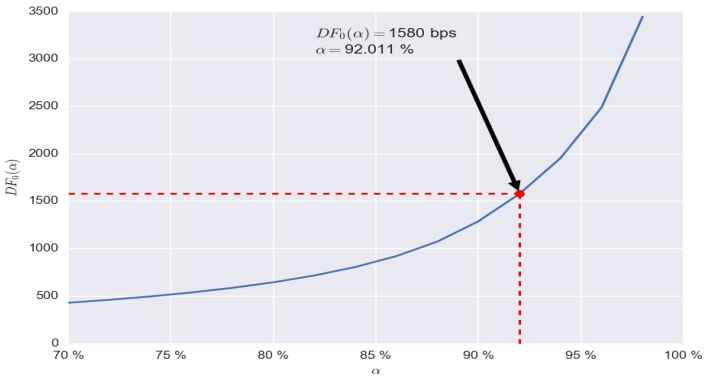


Figure 2: Time-0 default fund based on economic capital as a function of α.

in the default fund, namely (see (Albanese et al., 2016))

$$\mathrm{KVA}_t^{ccp} = h\mathbb{E}_t[\int_t^T e^{-(r+h)s}\mathrm{DF}_s ds],$$

for some hurdle rate h, taken in our numerics as 10%.

5.4 Allocation of the default fund by member incremental economic capital

Under this method, each clearing member would contribute an amount of the DF proportional to the incremental change in the economic capital attributable to the jth member. Hence

$$w_j = \frac{\mathrm{DF}^{ccp} - \mathrm{DF}^{ccp(-i)}}{n\mathrm{DF}^{ccp} - \sum_{j=1}^n \mathrm{DF}^{ccp(-j)}}$$

where:

$$DF^{ccp(-j)} = \mathbb{ES}_t^\alpha\Big(\sum_{t\le\tau_i<t+1, i\neq j} \beta_{\tau_i}(P^i_{\tau_i+\delta}+\Delta^i_{\tau_i+\delta}-P^i_{\tau_i}-IM^i_{\tau_i})^+-(\beta_t\mathrm{CVA}_t^{ccp(-j)}-\beta_{t+1}\mathrm{CVA}_{t+1}^{ccp(-j)})\Big)$$

is the value of the default fund without the j^{th} member and $\mathrm{CVA}_t^{ccp(-j)}$ is the CVA without the jth member, i.e.

$$\beta_t\mathrm{CVA}_t^{ccp(-j)} = Nom \times \sum_{i\neq j} \mathbb{1}_{\tau_i>t}\int_t^T A_i^\alpha(s)\lambda_i e^{-\lambda_i(t-s)}ds.$$

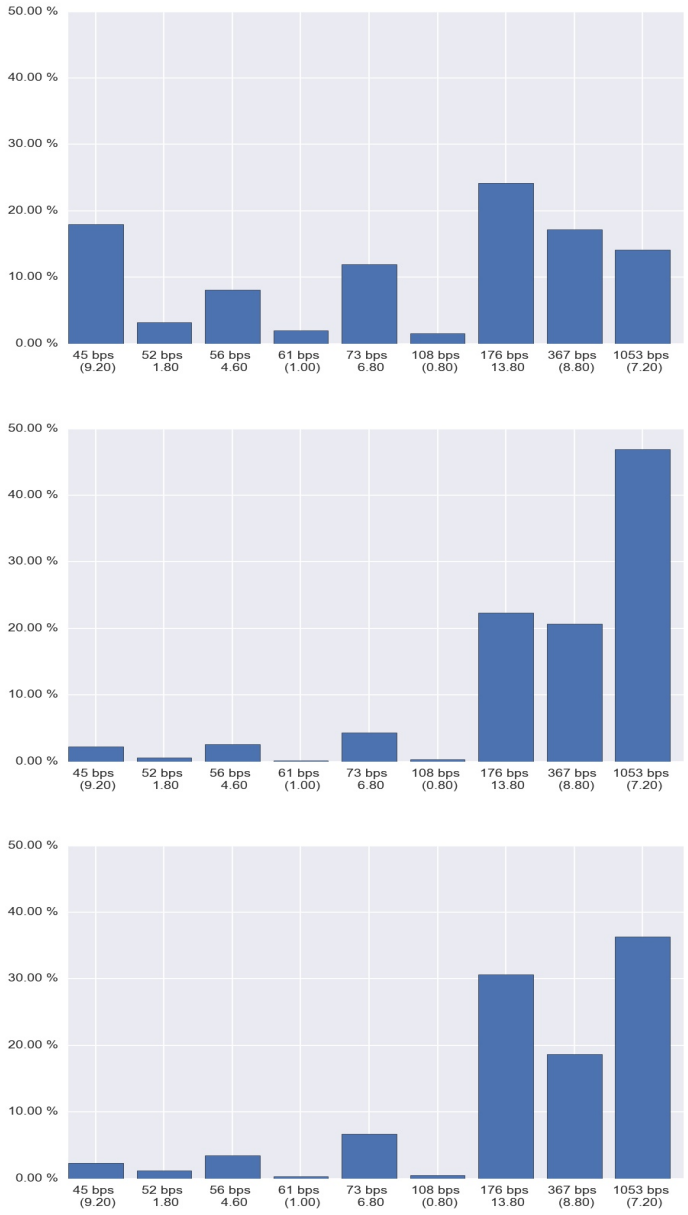


Figure 3: Default fund allocation based on initial margins, member incremental economic capital and member incremental KVA (top to bottom). Quantile level $\alpha = 92\%$.

An alternative is to allocate the DF proportionally to the incremental change to KVA attributable to their portfolio. In this case, for each clearing member j, their contribution is

$$\tilde{w}_j = \frac{\mathrm{KVA}^{ccp} - \mathrm{KVA}^{ccp(-i)}}{n\mathrm{KVA}^{ccp} - \sum_{j=1}^{n} \mathrm{KVA}^{ccp(-j)}}$$

where $\mathrm{KVA}^{ccp(-j)} = h\mathbb{E}[\int_t^T e^{-(r+h)s}\mathrm{DF}_s^{ccp(-j)}ds]$ is the value of the KVA without the jth member.

6 Funding the IM

MVA^i is calculated to determine the cost of funding the IM of the ith member. MVA^i is usually valued under the assumption that IM^i is funded through unsecured borrowing. However an alternative way to obtain IM, would be for an external party to post IM^i to the CCP in exchange for some premium payment from the clearing member. The third party would have rights to the initial margin at the maturity of transaction. But in the event of default, this third party would lose the minimum of the entire IM and the portfolio value after the liquidation period net the VM. Since the discounted expected value of this arrangement would be lower than the MVA, this funding method would reduce the costs for funding faced by clearing members.
The time 0 MVA when the IM is funded through borrowing is

$$\mathrm{MVA}_0^i \quad = \mathbb{E}[\int_0^{T_\wedge \tau_i} \beta_s \lambda_s^i (1 - R_i)\mathrm{IM}_s^i ds],$$

where R^i is the recovery rate (40% in our numerics) and IM^i is the initial margin posted by the i^{th} clearing member.
Under a third party arrangement we have instead

$$\widetilde{\mathrm{MVA}}_0^i \quad = \mathbb{E}[\mathbb{1}_{\tau_i < T}\beta_{\tau_i+\delta}(1 - R_i)(IM_{\tau_i}^i \wedge (P_{\tau_i+\delta}^i - P_{\tau_i}^i + \Delta_{\tau_i}^i)^+)].$$

By Gaussian computations (cf. (Armenti and Crépey, 2015, Equation (7.2))):

Lemma 6.1. *We have* $\beta_t IM_t^i = Nom \times |\omega_i| e^{-\kappa t} S_t B_i(t)$, *where*

$$B_i(t) = \begin{cases} \sum_{l=l_t+1}^{d} \beta_{T_l} h_l (e^{\sigma\sqrt{\delta}\Phi^{-1}(\alpha) + (\kappa - \frac{\sigma^2}{2})\delta} - 1) e^{\kappa T_{l-1}}, & \omega_i \geq 0 \\ \sum_{l=l_t+1}^{d} \beta_{T_l} h_l (1 - e^{\sigma\sqrt{\delta}\Phi^{-1}(1-\alpha) + (\kappa - \frac{\sigma^2}{2})\delta}) e^{\kappa T_{l-1}}, & \omega_i \leq 0. \end{cases}$$ ∎

Proposition 6.1. *We have*

$$\mathrm{MVA}_0^i = Nom \times S_0 \int_0^T (1-R_i)|\omega_i| B_i(s)\lambda_i e^{-\lambda_i s} ds.$$

Proof. We have

$$\mathrm{MVA}_0^i = \mathbb{E}\int_0^{T\wedge\tau_i} (1-R_i)\beta_s \lambda_s^i \mathrm{IM}_s^i ds = (1-R_i)\mathbb{E}\int_0^T \beta_s \lambda_i e^{-\lambda_i s}\mathrm{IM}_s^i ds,$$

which yields the result by Lemma 6.1 and (1). ∎

Lemma 6.2. *We have*

$$\mathbb{E}_s[\beta_{s+\delta}(P_{s+\delta}^i + \Delta_s^i - P_s^i)^+] = e^{-\kappa s} S_s |\omega_i| C_i(s)$$

where

$$C_i(s) = \begin{cases} \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} \Big(e^{-r\delta}\Phi(-y_-) - \Phi(-y_+)\Big), & \omega_i \geq 0 \\ \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} \Big(\Phi(y_+) - e^{-r\delta}\Phi(y_-)\Big), & \omega_i \leq 0 \end{cases}$$

with $y_\pm = \frac{r\delta}{\sigma\sqrt{\delta}} \pm \frac{1}{2}\sigma\sqrt{\delta}$.

Proof. We have (setting $Nom = 1$ in the proof to alleviate the notation)

$$\begin{aligned}
&(P_{s+\delta} - P_s + \Delta_s)^- = (P^\star_{s+\delta} - P^\star_s)^- \\
&= \Big(-\beta_{s+\delta}^{-1} S_{s+\delta} \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa(T_{l-1}-(s+\delta))} + \beta_s^{-1} S_s \sum_{l=l_s+1}^d \beta_{T_l} h_l e^{\kappa(T_{l-1}-s)} \Big)^- \\
&= \beta_{s+\delta}^{-1} \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} \big(e^{-\kappa(s+\delta)} S_{s+\delta} - e^{-r\delta} \frac{\sum_{l=l_s+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}}}{\sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}}} e^{-\kappa s} S_s \big)^+ \\
&= \beta_{s+\delta}^{-1} \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} \big(e^{-\kappa(s+\delta)} S_{s+\delta} - e^{-r\delta} e^{-\kappa s} S_s \big)^+
\end{aligned}$$

Hence, using the Black Scholes formula for a forward starting option (see e.g. (Crépey and Song, 2014, p.404)

$$\begin{aligned}
&\mathbb{E}_s[\beta_{s+\delta}(P_{s+\delta} - P_s + \Delta_s)^-] = \mathbb{E}_s\Big[\sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} \big(e^{-\kappa(s+\delta)} S_{s+\delta} - e^{-r\delta} e^{-\kappa s} S_s \big)^+ \Big] \\
&= \sum_{l=l_{s+\delta}+1}^d \beta_{T_l} h_l e^{\kappa T_{l-1}} e^{-\kappa s} S_s \big(\Phi(y_+) - e^{-r\delta}\Phi(y_-) \big).
\end{aligned}$$

The analog formula for $\mathbb{E}_s[(P_{s+\delta} - P_s + \Delta_s)^+]$ is deduced by call/put parity. The proof is concluded from $P^i = \omega_i P$. ∎

Proposition 6.2. *We have*

$$\widetilde{\mathrm{MVA}}_0^i = Nom \times S_0 \int_0^T (1-R_i)|\omega_i|\big(C_i(s)-A_i(s)\big)\lambda_i e^{-\lambda_i s} ds$$

Proof. Note that $x^+ {}_\wedge a = x^+ - (x-a)^+$. Hence (setting $R_i = 0$ in the proof to alleviate the notation),

$$\begin{aligned}
\widetilde{\mathrm{MVA}}_0^i &= \mathbb{E}[\mathbb{1}_{\tau_i<T}\beta_{\tau_i+\delta}(IM^i_{\tau_i} \wedge (P^i_{\tau_i+\delta} - P^i_{\tau_i} + \Delta^i_{\tau_i})^+)] \\
&= \mathbb{E}[\mathbb{1}_{\tau_i<T}\mathbb{E}_{\tau_i-}[\beta_{\tau_i+\delta}IM^i_{\tau_i} \wedge \beta_{\tau_i+\delta}(P^i_{\tau_i+\delta} - P^i_{\tau_i} + \Delta^i_{\tau_i})^+]] \\
&= \mathbb{E}[\mathbb{1}_{\tau_i<T}\xi^i_{\tau_i}] = \mathbb{E}[\int_0^T \lambda e^{-\lambda s}\xi^i_s ds]
\end{aligned}$$

where

$$\begin{aligned}
\xi^i_s &= \mathbb{E}_s[\beta_{s+\delta}IM^i_s \wedge \beta_{s+\delta}(P^i_{s+\delta} - P^i_s)^+] \\
&= \mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} - P^i_s + \Delta^i_s)^+] - \mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} - P^i_s + \Delta^i_s - IM^i_s)^+]
\end{aligned}$$

is a predictable process in our model, in which $\mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} - P^i_s + \Delta^i_s - IM^i_s)^+]$ is given by Lemma 5.1 and

$$\mathbb{E}_s[\beta_{s+\delta}(P^i_{s+\delta} + \Delta^i_s - P^i_s)^+]$$

is given by Lemma 6.2. ∎

7 Conclusion

In this work we considered two key capital and funding issues related to CCPs. Firstly, from the CCP perspective, we challenged the "cover two" EMIR rule for the sizing of the CCP default fund by an economic capital specification set as an expected shortfall of the one year ahead loss and profit of the CCP. We calibrated the quantile level in this expected shortfall to a cover two specifications, concluding that the size of the default fund determined by the cover 2 rule corresponds to a default fund sufficient to cover 12.5-yearly adverse events. We challenged in turn the classical IM-based allocation of the default fund, by using an allocation proportional to the incremental impact on the economic capital of the CCP (or on the ensuing KVA) of the clearing members. It turns out that the economic capital (or KVA) based allocation make a much better mix of market and credit information

than the one based on initial margins, which in the simple case of our toy model is equivalent to an allocation proportional to the size of the positions of the clearing members (irrespective of the credit risk of the clearing members).
Secondly, from a clearing member perspective, we compared the MVAs of two different strategies regarding initial margins: the classical one where the initial margin is unsecurely borrowed by the clearing member and a strategy where the clearing member delegates the posting of the initial margin to a specialist lender in exchange of a service fee.

Bibliography

Albanese, C. (2015). The cost of clearing. *Available at SSRN 2667518*.

Albanese, C., Caenazzo, S., and Crépey, S. (2016). Capital valuation adjustment and funding valuation adjustment. *Available at SSRN*.

Armenti, Y. and Crépey, S. (2015). Central clearing valuation adjustment. *arXiv preprint arXiv:1506.08595*.

Crépey, S. and Song, S. (2014). Counterparty risk modeling: Beyond immersion. *hal-00989062vl*.

Polynomial Models for Market Weights at Work in Stochastic Portfolio Theory: Theory, Tractable Estimation, Calibration and Implementation

Team 3

KAROL GELLERT, University of Technology, Sydney
MARIO GIURICICH, University of Cape Town
ALEX PLATTS, University of Cape Town
SHIVAN SOOKDEO, University of Cape Town

Supervisors:
CHRISTA CUCHIERO, University of Vienna
JOSEF TEICHMANN, ETH Zurich

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

1.1 Background and problem statement

Many models have been proposed for market weights of (large) indices in the last decades. They suffer, even though valuable and a good source of inspiration, from three drawbacks.

- Firstly, stock capitalisations are often assumed to be locally uncorrelated.
- Secondly, ranked market weights in such models are too stable in their dynamic behaviour.
- Thirdly, such models are often highly intractable (in estimation and calibration, as well as implementation and evaluation).

We propose, from the tractable class of polynomial models, a generic model for market weights which resolves, to a large extent, all three drawbacks mentioned above.

We also consider the problem of portfolio selection, in the context of a market which has a large number of different financial products. Often, such an action is performed by utility maximisation. However, it has been argued that (i) the choice of such a utility function is subjective and (ii) an investor's utility function is both time and circumstance dependent. Moreover, given a preference, non-trivial portfolios can only be selected after having estimated drift, and these portfolios depend critically on the drift estimators. Therefore, drift is a second ingredient for a standard portfolio selection procedure. In view of this, we opt for a preference-free way of selecting portfolios without drift estimation.

Stochastic Portfolio Theory, developed by Robert Fernholz, allows for such preference-free portfolio choices which can potentially beat the market portfolio almost surely over a sufficiently long time horizon. In the realm of stochastic portfolio theory, our objective is to use only market-observable quantities to develop portfolio strategies that give good pathwise performance compared to that of the entire market portfolio.

Moreover, the fact that stochastic portfolio theory does not require any drift estimation leads to more robustness relative to the more standard approaches to portfolio optimisation, such as mean-variance portfolio theory. Central to the construction of such good investment strategies, which can beat the market over some time period, are *functionally generated portfolios*. These are portfolios that depend on current market weights (i.e. weighted market capitalisations) in a simple way, and we aim to implement such strategies. Two crucial points are, however, firstly the assumption of no transaction costs and secondly that assets can be purchased in any amount.

To formalise, our goals in this report are as follows:

1. We aim to firstly calibrate a polynomial model to both the market weights and to the market capitalisations.

2. Secondly, we will implement functionally-generated portfolios and analyse the relative pathwise (potential) outperformance of these portfolios with respect to the market portfolio.

3. Third, with the help of the calibrated model we are able to analyse time horizons over which relative arbitrage appears. Therefore, for concrete applications, the calibration of a flexible model is key, even though the strategies do only depend on observables.

The report is organised as follows. Section 2 describes the theoretical framework for our problem. Section 3 contains an empirical analysis using market data from the MSCI World index. Section 4 delves into the procedures (and difficulties) when calibrating our model. Thereafter, Section 5 is dedicated to a simulation study implements our model in the context of stochastic portfolio theory, in that we derive portfolios that can potentially outperform the market. We then finish in Section 6 with discussions and conclusions.

All computational implementations are performed in Matlab R2016A. Appendix C provides some of our code.

2 Theoretical framework

2.1 Polynomial diffusions

Polynomial models form a large class of diffusion models. Examples of such models include the famous Black-Scholes model for stock prices, as well as various interest rate models such as the Vasicek and Cox-Ingersoll-Ross models. Polynomial models are useful because of their mathematical tractability (Cuchiero et al., 2012). Moments may be calculated explicitly and this is important, for example, for option pricing or model estimation by moment methods. Notice in particular that the model parameters appear linearly in the model characteristics, which allows for easier identification.

We aim, in this work, to construct a polynomial diffusion model that is firstly relatively simple to calibrate, secondly accounts for correlations between stocks and thirdly can be used to outperform the market within the context of Stochastic Portfolio Theory. We measure outperformance with respect to the market portfolio, in our case the MSCI. One specific model class considered in Stochastic Portfolio Theory which is a particular example of such polynomial diffusion model are *volatility stabilised models* introduced by Fernholz and Karatzas (2005) (Fernholz and Karatzas (2005)).

We consider a diffusion model for the $d-$dimensional weights process, $\{\boldsymbol{\mu}_t : t \geq 0\}$, defined on the unit simplex, Δ^d,

$$d\boldsymbol{\mu}_t = b\left(\boldsymbol{\mu}_t\right) dt + \sqrt{\Sigma\left(\boldsymbol{\mu}_t\right)} d\boldsymbol{W}_t, \tag{1}$$

where $\boldsymbol{W}_t$ is a $d-$dimensional vector of standard Brownian motions on a filtered probability space (with filtration $\mathbb{F}$, say) satisfying the usual conditions, and where the drift, $b : \Delta^d \to \Delta^d$ and the instantaneous covariance matrix $\Sigma : \mathbb{R}^d \to S_d^+$ (where S_d^+ is the space of $d-$dimensional positive semi-definite matrices) are certain functions. Indeed, for polynomial models,

- the drift is affine i.e. $b\left(\boldsymbol{\mu}_t\right) = \beta + B\boldsymbol{\mu}_t$, with $\beta \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d\times d}$, and
- Σ is a quadratic function.

Recall that one feature of our portfolio selection method is that we need not concern

ourselves with the drift and its estimation: we will show that for the purposes of the application of the model, the drift term is unnecessary. Therefore optimisation, in particular filtering, is not required when estimating parameters. With a single path of data, parameters can be easily calculated.

When the state space is the unit simplex, we have according to Lemma 2.2 and Proposition 6.6 of Filipović and Larsson (2016),

$$\Sigma(\boldsymbol{\mu}_t)_{ij} = -\gamma_{ij}\mu_t^i\mu_t^j, \quad \text{for } j \neq i, \quad \text{and} \tag{2}$$

$$\Sigma(\boldsymbol{\mu}_t)_{ii} = \sum_{j\neq i}\gamma_{ij}\mu_t^i\mu_t^j \tag{3}$$

on the unit simplex Δ^d for some $\gamma_{ij} \in \mathbb{R}_+$ such that $\gamma_{ij} = \gamma_{ji}$ for all i, j. Indeed, non-negativity of γ_{ij}, $i \neq j$, is a consequence of (2) and (3) in order to ensure that positive semi-definiteness of the instantaneous covariance is not violated. Moreover, if the γ_{ij}'s are non-negative, the instantaneous covariance matrix is indeed positive semi-definite.

We endeavour to calibrate this model, and then apply it in the context of stochastic portfolio theory. In particular, the instantaneous covariance matrix (or the integrated covariance matrix) needs to be estimated from the data. We discuss this in section 4.

2.2 Stochastic portfolio theory

2.2.1 Background and strong relative arbitrage

Stochastic portfolio theory (SPT) is a framework in which the usual assumptions underlying classical mathematical finance are not fully made. Rather, one nestles oneself within an empirical "data-driven" context and assumes only the weaker No Unbounded Profit with Bounded Risk (NUPBR) condition (Vervuurt, 2015). The aim of the theory is then to discover investment strategies which outperform the market (with market performance measured by some specified index) in a pathwise manner. In particular, the beauty of SPT lies within the fact that it avoids making any assumptions about the expected return of stocks (the "drift"), which are difficult to estimate (Fernholz, 2002).

First, we begin by introducing some notation. Let $\boldsymbol{S}_t = \left(S_t^1, S_t^2, ..., S_t^d\right)$ represent the market capitalisations of each stock in the market (and by assumption, each stock in the index) at time t. We define a strategy (representing the positions in each of the stocks) as an $\mathbb{F}$-progressively-measurable $d-$dimensional vector process $\pi := \left(\pi^1, \pi^2, ..., \pi^d\right)$ taking values in the $d-$dimensional hyperplane $H := \left\{x \in \mathbb{R}^d : \sum_{i=1}^d x^i = 1\right\}$. Notice that π defined in this way implies that our portfolio can take both long and short positions. If we rather let $\pi \in \Delta^d := \left\{x \in \mathbb{R}_+^d : \sum_{i=1}^d x^i = 1\right\}$, we constrain our portfolio to long only positions. Now, $\left\{\pi_t^i : t \geq 0\right\}$ is the process representing the proportion of an investor's current wealth in the i'th asset at time t.

We are now in a position to define the **wealth process**. We denote the wealth process of an investor, investing according to strategy π, by $\{V_t^\pi : t \geq 0\}$. It can be easily seen that it evolves as

$$\frac{dV_t^\pi}{V_t^\pi} = \sum_{i=1}^d \pi_t^i \frac{dS_t^i}{S_t^i}, \tag{4}$$

with $V_0^\pi = 1$ unit of currency. Notice that if we slightly re-arrange (4), $\frac{\pi_t^i V_t^\pi}{S_t^i}$ represents, at any time t, the number of units of asset i held in the portfolio. One particular strategy we consider (and we call such a strategy the market portfolio) is the market weights, i.e. $\boldsymbol{\mu}$, and if we employ this strategy to invest in the index, the implication is that at each time t

$$V_t^\mu = \frac{\sum_{i=1}^d S_t^i}{\sum_{i=1}^d S_0^i}. \tag{5}$$

Our aim in this section is to compare V^π for some strategy π with the market portfolio. So, we are in fact interested in analysing the **relative wealth process** (relative to the market portfolio), $\{Y_t^\pi : t \geq 0\}$ defined by

$$Y_t^\pi := \frac{V_t^\pi}{V_t^{\boldsymbol{\mu}}}, \tag{6}$$

with $Y_0^\pi = 1$.

Definition 1 (strong relative arbitrage over the market portfolio) If π is a trading strategy, and $\boldsymbol{\mu}$ is the market portfolio, then π exhibits strong relative arbitrage over the market portfolio, over the period $[0, T]$ for $T > 0$ if and only if

$$\mathbb{P}\left(Y_T^{\pi} > 1\right) = 1. \tag{7}$$

In order for us to meet our aim, we need to find π such that Definition (1) is satisfied. Before we proceed, we present a lemma.

Lemma 1: The relative wealth process given by (6) satisfies the following dynamics:

$$\frac{dY_t^{\pi}}{Y_t^{\pi}} = \sum_{i=1}^{d} \pi_t^i \frac{d\mu_t^i}{\mu_t^i}, \tag{8}$$

with $Y_0^{\pi} = 1$.
Proof: see Appendix A.

2.2.2 Functionally Generated Portfolios

Functionally Generated Portfolios (or strategies), which were first introduced in Fernholz (1998), generalise the idea of the market portfolio introduced above. An interesting feature of such portfolios is that the relative wealth process can be expressed without using stochastic integrals (Fernholz and Karatzas, 2009). This allows for easy, probability-free comparisons over given, fixed time-horizons.

We begin by considering the market weights $\boldsymbol{\mu}_t$ at every time point t. The essential idea is that certain real-valued functions defined on the unit simplex Δ^d can be used to generate portfolios (or equivalently, strategies). We investigate smooth functions of the market weights only.

We then use these functions to develop portfolios whose returns satisfy almost sure relationships relative to the market portfolio. Thereafter, these portfolios can be applied in situations where arbitrage may be possible (i.e. in the case where the market (or model) admits relative arbitrage). We now proceed in a similar

spirit as in Section 11 of Fernholz and Karatzas (2009). Suppose we have a function $\mathcal{G} : U \to (0, \infty)$, bounded and of class $C^2(U)$, where U is some open neighbourhood of Δ^d. We now define what is meant by the portfolio (or strategy) generated by $\mathcal{G}$.

Definition 2: (Portfolio generated by $\mathcal{G}$) The strategy $\pi^{\mathcal{G}}$, given by

$$\pi_t^{\mathcal{G},i}(\boldsymbol{\mu}_t) = \mu_t^i \left(\frac{D_i \mathcal{G}(\boldsymbol{\mu}_t)}{\mathcal{G}(\boldsymbol{\mu}_t)} + 1 - \sum_{j=1}^{d} \frac{\mu^j D_j \mathcal{G}(\boldsymbol{\mu}_t)}{\mathcal{G}(\boldsymbol{\mu}_t)} \right), \tag{9}$$

where D_i is the derivative with respect to the i'th vector component, is called the portfolio generated by $\mathcal{G}$.

Notice that the strategy $\pi^{\mathcal{G}}$ depends only on the market weights, and therefore this portfolio can be easily implemented and its associated wealth process can be observed over time, via only the evolution of the market weights. Moreover, it is encouraging to note that when one invests according to the strategy $\pi^{\mathcal{G}}$, tractable formulae for the relative wealth process to the market portfolio emerge. See the "master formula" (i.e. Equation (11.2) in Fernholz and Karatzas (2009)). In fact, by taking the exponential of Equation (11.2) in Fernholz and Karatzas (2009), we obtain the following expression for the relative wealth process of our strategy $\pi^{\mathcal{G}}$ to the market portfolio $\boldsymbol{\mu}$, that is

$$Y_t^{\pi^{\mathcal{G}}} = \frac{\mathcal{G}(\boldsymbol{\mu}_t)}{\mathcal{G}(\boldsymbol{\mu}_0)} \exp\left(\int_0^T g(s) \mathrm{d}s \right), \quad \text{with} \tag{10}$$

$$g(s) = -\frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{D_{ij} \mathcal{G}(\boldsymbol{\mu}_s)}{\mathcal{G}(\boldsymbol{\mu}_s)} \frac{\langle \mu^i, \mu^j \rangle_s}{\mathrm{d}s} \tag{11}$$

where D_{ij} is the second derivative, first with respect to the i'th and second with respect to the j'th component. Notice in (10) (for a concave function) that if the ratio $\frac{\mathcal{G}(\boldsymbol{\mu}_t)}{\mathcal{G}(\boldsymbol{\mu}_0)}$ drops, this can be offset to some extent by an increase in $\int_0^T g(s) \mathrm{d}s$. This is in line with the idea of volatility harvesting - that for assets which are liquid and volatile, the actions of rebalancing and diversifying (perhaps according to a particular strategy) can create excess growth of the portfolio $\pi^{\mathcal{G}}$.

Now, we consider specific functionals. For the purpose of this research, we suppose

that $\mathcal{G}$ is concave. Then, the portfolio $\pi^{\mathcal{G}}$ is long only (Fernholz and Karatzas, 2009). We consider three types, all of them from Fernholz and Karatzas (2009).

1. *Equally-weighted portfolio:* $\mathcal{G}(\boldsymbol{\mu}_t) \equiv \boldsymbol{F}(\boldsymbol{\mu}_t) \coloneqq \bar{c} + \left(\prod_{i=1}^{d} \mu_t^i\right)^{1/d}$ for any given $\bar{c} \in (0, \infty)$.

2. *Entropy-weighted portfolio:* $\mathcal{G}(\boldsymbol{\mu}_t) \equiv \boldsymbol{H}(\boldsymbol{\mu}_t) \coloneqq c - \sum_{i=1}^{d} \mu_t^i \ln \mu_t^i$ for any given $c \in (0, \infty)$.

3. *Diversity-weighted portfolio:* $\mathcal{G}(\boldsymbol{\mu}_t) \equiv \boldsymbol{G}_p(\boldsymbol{\mu}_t) \coloneqq \left(\sum_{i=1}^{d} (\mu_t^i)^p\right)^{1/p}$ for any given $p \in (0, 1)$.

According to Examples (11.1) and (11.2) in Fernholz and Karatzas (2009), each of the three functionally generated portfolios exhibit relative arbitrage over the market portfolio, over sufficiently long periods of time, when the market model satisfies several generic conditions (e.g. diversity[1] and ellipticity models). Karatzas and Fernholz provide guidance on how long these periods should be, however, we aim to empirically test this for our data set, which is subject of the following sections.

[1]Diversity models, in the context of the market weights, would have to satisfy two implicit conditions for there to be a strong drift present. Those conditions are $\max_i \left(\mu_t^i\right) < 1 - \delta$ for some small $\delta > 0$ and $\det(\tilde{Q}) \neq 0$ for $\tilde{Q}$ the matrix defined in Section 5.1 and Appendix B. Note, however, that our model we use in this report does not satisfy the former condition, in that μ_t^i can approach 1 infinitesimally.

3 Empirical analysis based on real data

3.1 Data description

We were provided with data from 301 stocks from the MSCI World Index. The daily data spanned 327 days, from 1 August 2006 to 31 October 2007, a relatively stable period for the market. We noted that most of the large-cap companies maintained their presence in the index, while many smaller cap companies "jumped" in and out. We furthermore noted that from day to day, but in particular from month to month, not many companies maintained their ranks in the index.

The data was a time-series of market capitalisations for each of the companies comprising the MSCI index, but we were also given stock prices for each index constituent, on a daily basis. Upon inspection of the market capitalisation data we noticed that it was not clean. In particular, data regarding the market capitalisations was missing in two ways: (i) sporadically, in that there were missing entries on certain days; and (ii) companies unjoining and joining the index. The latter was evident in that certain companies had long, consecutive runs of no data.

Given the prevalence of (ii), and the fact that companies were in the index for a very short period of time, we decided to remove 10 companies from our analysis, and therefore dealt with 291 stocks. This removed the problem of (ii) from our data set. We thereafter linearly interpolated for the missing values, and overcame (i). After this cleaning, we obtained our final data set to be used in our study.

After understanding the data and cleaning it to a certain extent, it became necessary to obtain a time series of (index) weights for the capitalisations, over the 327-day period. Let S_t^i denote the market capitalisation of the ith company at trading-day t (out of a set of d companies, say). For each day, the market weights[2] for the i'th company, μ_t^i, is given by

$$\mu_t^i = \frac{S_t^i}{\sum_{i=1}^d S_t^i} \tag{12}$$

i.e. $\boldsymbol{\mu}_t = \left(\mu_t^1, \mu_t^2, ..., \mu_t^d, \right)$ is constrained to the unit simplex, that is

[2] In our case, $d = 291$, which is the number of companies analysed.

$\boldsymbol{\mu}_t \in \Delta^d$ where $\Delta^d := \left\{ \boldsymbol{x} \in \mathbb{R}^d : x^j \in [0,1], \sum_{j=1}^d x_j = 1 \right\}$.

3.2 Fluctuation size: empirical study.

Empirical market data reflects the feature of concentration i.e. there are a few firms with large market capitalisations and many firms with much lower market capitalisations. One can illustrate this feature by plotting the log market capitalisaitons against the log rank of the firm in respect of market .

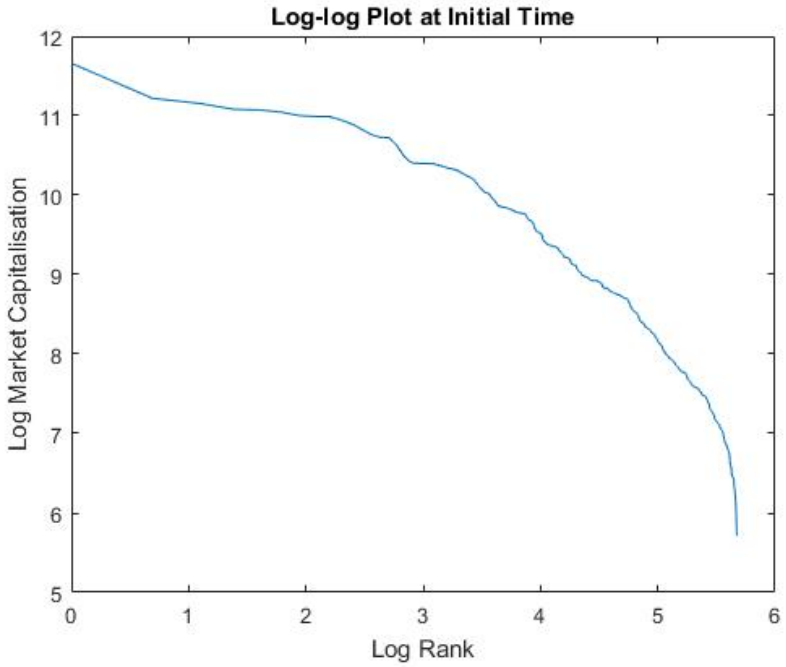


Figure 1: Empirical log-log plot.

Classical models do not reflect this market feature, but rather suggest linear decay of the curve.

Furthermore one can observe fluctuations in these log-log plots over time. Let $f_{t,i}$ be log market capitalisation at time t of the i^{th} firm. i.e.

$$f_{t,i} := \log(S_t^{(i)})$$

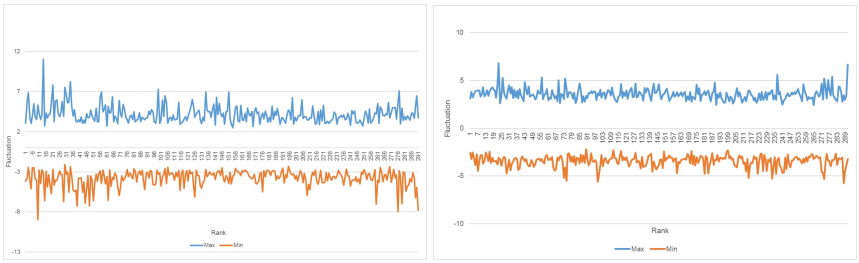
where $S_t^{(i)}$ is the ranked market capitalisation of the i'th firm.

We define the fluctuation as the change over time:

$$\Delta f_{t,i} := f_{t,i} - f_{t-1,i}$$

.

We propose that these fluctuations are driven by both jumps and discontinuities. A jump is defined as a change in a firm's market weight of magnitude larger than three standard deviations. After removing these jumps from the empirical data, the fluctuations are significantly reduced. This illustrated below by plotting the minimum and maximum of the fluctuations.



(a) With Jumps (b) Without Jumps

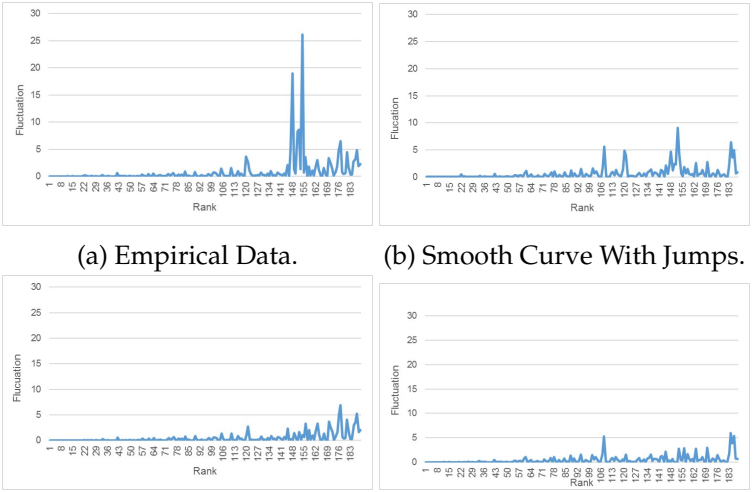
Figure 2: Effect of removing jumps.

In order to test the effect of discontinuities, a smooth parametric curve is fitted to the initial log-log plot. The effects of smoothing on fluctuations are illustrated in Figure 3 below.

3.3 Empirical in-sample study of functionally generated portfolios.

In order to check the robustness of our three postulated functions for use in our functionally-generated portfolios from an empirical perspective, it was necessary to test them in-sample for relative arbitrage. That is, we checked whether they produced relative outperformance over the period from which the data was collected, using Lemma (1). Observe that our analysis below is independent of the model for the market weights (1).

In implementing the procedure, it was necessary to discretise the stochastic integral represented by (8), and using the sample changes in market weights (that is,



(a) Empirical Data.

(b) Smooth Curve With Jumps.

(c) Non-smooth Curve without Jumps.

(d) Smooth Curve without Jumps.

Figure 3: Effect of removing discontinuities.

$\mu_{t+1}^i - \mu_t^i$) as an approximation to $d\mu_t^i$ for each of the i stocks. That is, the following discretisation scheme was applied:

$$Y_{t+1}^{\pi} = Y_t^{\pi} + Y_t^{\pi} \sum_{i=1}^{d} \pi_t^i \left(\boldsymbol{\mu}_t\right) \frac{\left(\mu_{t+1}^i - \mu_t^i\right)}{\mu_t^i} \tag{13}$$

with $Y_0^{\pi} = 1$ and for all times t, apart from the last, in the time grid.

In order to evaluate (13), it is necessary to compute $\pi_t^i \left(\boldsymbol{\mu}_t\right)$ by using Equation (9). After some algebra the strategies, at each time point t, are as follows:

- *Equally-weighted portfolio:* $\pi_t^i \equiv {}^{\boldsymbol{G}}\pi_t^i \left(\boldsymbol{\mu}_t\right) := \frac{1}{d}$;
- *Entropy-weighted portfolio:* $\pi_t^i \equiv {}^{\boldsymbol{H}}\pi_t^i := \mu_t^i \left(\frac{c - \ln(\mu_t^i)}{\boldsymbol{H}(\boldsymbol{\mu}_t)}\right)$;
- *Diversity-weighted portfolio:* $\pi_t^i \equiv {}^{\boldsymbol{G}_p}\pi_t^i := \frac{\mu_t^i}{\sum_{j=1}^{d} \left(\mu_t^j\right)^{1/p}}$.

Figure (4) shows the outperformance of our three strategies ${}^{\boldsymbol{G}}\pi^i$, ${}^{\boldsymbol{H}}\pi^i$ and ${}^{\boldsymbol{G}_p}\pi^i$ relative to the market portfolio. Notice that our strategies, at each time point, took long positions in each of the $d = 291$ stocks.

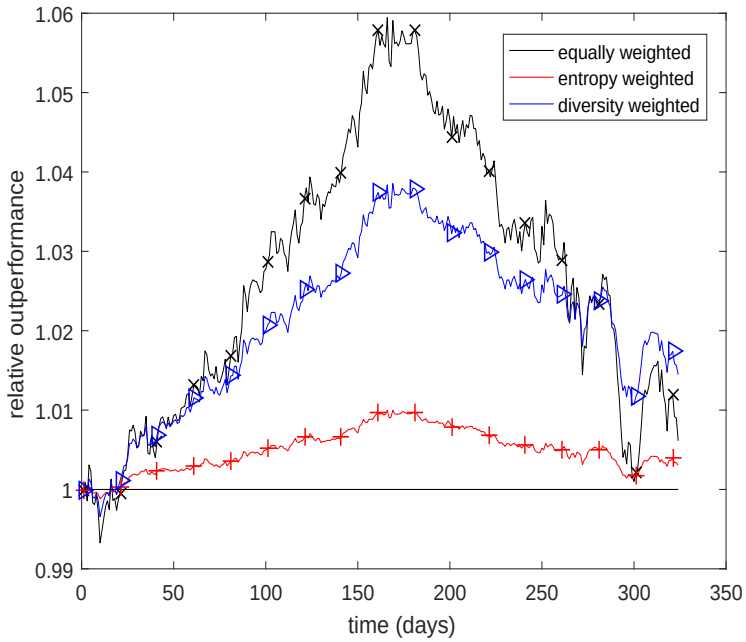


Figure 4: Relative in-sample outperformance of each of the three functionally generated portfolios.

It is clear from Figure (4) that from about 25 days since the beginning of the period of investigation, all three strategies outperform the market, with the equally-weighted portfolio giving the best outperformance after about 150 days. Also, the performance of the equally-weighted and diversity-weighted portfolios appears to be more volatile than that of the entropy-weighted portfolio - the volatile performance of the equally-weighted portfolio comes as no surprise, given the constant nature of the portfolio weights. Finally, over the 327-day period of investigation, all three portfolios appear to provide a relative arbitrage over the market portfolio, however, it would be difficult to say whether this phenomena would continue for longer than 327 days given the downward movement of the relative performance, of all three portfolios, from about the 160'th day. Even though our in-sample analysis provides some evidence of relative arbitrage portfolios, the true test of such a portfolio selection technique lies in testing it on simulated data, based on our polynomial diffusion calibrated to the time series of the market weights. We perform this in Section 5.4.

4 Model calibration and estimation

4.1 Integrated covariance matrix estimation

Our first goal in this section is to find a robust estimator for the integrated covariance of the market weights. This will then allow us to calibrate our model i.e. to specify its instantaneous covariance matrix. However, two problems arise in such an estimation - asynchronicity of the data and the fact that the estimate must be positive-semidefinite (see, for example, Boudt et al. (2012)). We keep these in mind when estimating the integrated covariance.

We do the estimation via two methods, the first being the classical route to estimating the integrated covariance, while the second having its roots in harmonic analysis.

4.1.1 Classical estimation

In using the classical integrated covariance estimator, it is necessary to have a complete data set. By that, we mean there must be no missing values in the market weight data. So, we use the naïve (interpolated) market weight data in calculating the estimator given by definition 3.

Definition 3: (Classical estimator for the integrated covariance matrix) The classical estimator for the integrated covariance over the time period $[0, T]$ for the ij'th element of the integrated covariance matrix, is

$$\widehat{\text{ICov}_{ij}} = \sum_{k=1}^{N} \Delta\mu_k^i \Delta\mu_k^j \tag{14}$$

where N is the number of observations in the time series and $\Delta\mu_k^i \coloneqq \mu_{t^k}^i - \mu_{t^{k-1}}^i$ for t^k an indexed point of the time grid.

4.1.2 Fourier transform non-parametric estimation

Notice that the classical integrated covariance estimator is no longer robust (or useful, in fact) if there is asynchronicity in the data. Therefore, we propose that an alternative methodology accommodating for such asynchronicities of such observations is used. We investigate the non-parametric method proposed by Malliavin et al. (2009), present it here, but apply it in the case of interpolated market weight data.

Definition 4: (Fourier coefficient) Let $g : [0, T] \to \mathbb{R}$ be a function. The k'th Fourier coefficient is given by

$$\mathcal{F}(g)(k) = \frac{1}{T}\int_0^T \exp\left(-i\frac{2\pi}{T}kt\right)g(t)\mathrm{d}t, \tag{15}$$

for $k \in \mathbb{Z}$.

Notice that if we set $k = 0$ in Equation (15), we simply obtain $\frac{1}{T}\int_0^T g(t)\mathrm{d}t$. This will be useful in relating the Fourier coefficient to the integrated covariance. Moreover, we can express (15) for differentiated quantities, that is

$$\mathcal{F}(\mathrm{d}g)(k) = \frac{1}{T}\int_{[0,T]} \exp\left(-i\frac{2\pi}{T}kt\right)\mathrm{d}g(t), \tag{16}$$

and if g is differentiable with derivative g', we can write

$$\mathcal{F}(\mathrm{d}g)(k) = \frac{1}{T}\int_0^T \exp\left(-i\frac{2\pi}{T}kt\right)g'(t)\mathrm{d}t. \tag{17}$$

Also, if we are given two functions on $\mathbb{Z}$, Φ and Ψ, we say that the Bohr convolution product *exists* if the following limit,

$$\left(\Phi *_B \Psi\right)(k) \coloneqq \lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N} \Phi(n)\Psi(k-n) \tag{18}$$

exists for all $k \in \mathbb{Z}$.

We can now apply the above Fourier transforms in the context of our model (1). Note that the conditions (H) on page 1986 of Malliavin et al. (2009) are satisfied. We are now in the context to apply Theorem 2.1 of Malliavin et al. (2009).

Theorem 1: Let $\boldsymbol{\mu}$ follow (1). Then

$$\frac{1}{T}\mathcal{F}\left(\Sigma^{ij}\right)(k) = \frac{1}{T^2}\int_0^T \exp\left(-i\frac{2\pi}{T}kt\right)\Sigma_t^{ij}\mathrm{d}t \tag{19}$$

$$= \left(\mathcal{F}(\mathrm{d}\mu^i) *_B \mathcal{F}(\mathrm{d}\mu^j)\right)(k) \tag{20}$$

$$= \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N}\mathcal{F}(\mathrm{d}\mu^i)(n)\mathcal{F}(\mathrm{d}\mu^j)(k-n). \tag{21}$$

Proof: see Malliavin et al. (2009) pp. 1987 and 1988.

In applying Theorem 1 above, notice that two approximations need to be made. Firstly, it is necessary to choose N, and the larger N is, the more accurate the approximation. Secondly, it is necessary to approximate $\mathcal{F}(\mathrm{d}\mu^i)$.

In approximating $\mathcal{F}$, consider its integrated form given by (22) below.

$$\mathcal{F}(\mathrm{d}\mu^i)(k) = \frac{1}{T}\int_{[0,T]}\exp\left(-i\frac{2\pi}{T}kt\right)\mathrm{d}\mu_t^i \tag{22}$$

Suppose that we have a time grid (not necessarily equally spaced) for the i'th component of $\boldsymbol{\mu}$, given by

$$0 = t_0^i < t_1^i < t_2^i < ... < t_{M_i}^i = T, \tag{23}$$

where M_i is the largest index in the time grid. Let the consecutive day-by-day change in market weights be denoted by

$$\Delta_\ell\mu^i \coloneqq \mu^i_{t^i_{\ell+1}} - \mu^i_{t^i_\ell}. \tag{24}$$

By applying the naïve Riemann sum approximation of the integral, we obtain an estimator for $\mathcal{F}$, given by

$$\widehat{\mathcal{F}}(\mathrm{d}\mu^i)(k) = \frac{1}{T}\sum_{\ell=0}^{M_i-1}\exp\left(-i\frac{2\pi}{T}kt_\ell\right)\Delta_\ell\mu^i. \tag{25}$$

Therefore, insert (25) into (21) to obtain an estimator for Equation (21) for any $N \in \mathbb{N}^+$ and for $k = 0$, that is

$$\frac{1}{T}\hat{\mathcal{F}}^N\left(\Sigma^{ij}\right)(0) \coloneqq \frac{1}{2N+1}\sum_{n=-N}^{N}\widehat{\mathcal{F}}(\mathrm{d}\mu^i)(n)\widehat{\mathcal{F}}(\mathrm{d}\mu^i)(-n) \tag{26}$$

$$= \frac{1}{2N+1}\sum_{n=-N}^{N}\left[\frac{1}{T}\sum_{\ell=0}^{M^i-1}\exp\left(-i\frac{2\pi}{T}nt_\ell\right)\Delta_\ell\mu^i.\right]\left[\frac{1}{T}\sum_{\ell=0}^{M^j-1}\exp\left(i\frac{2\pi}{T}nt_\ell\right)\Delta_\ell\mu^j\right]. \tag{27}$$

Hence, we see that we are able to use the estimator (27) to find estimates for the integrated covariance for our particular data set. In particular, our estimator for γ in general is given by

$$\gamma^{ij} = \frac{\int_0^T \Sigma^{ij}\left(\boldsymbol{\mu}_t\right)\mathrm{d}t}{\int_0^T \mu_t^i\mu_t^j\mathrm{d}t}. \tag{28}$$

This follows from Equations (3) and (4). Applying the Fourier estimator for the integrated covariance, we obtain the following estimator for γ.

$$\widehat{\gamma^{ij}} = \frac{\hat{\widehat{\mathcal{F}}}^N\left(\Sigma^{ij}\right)}{\sum_{\ell=1}^{M^i}\Delta_\ell\mu^i\Delta_\ell\mu^j\left(t_{\ell+1}-t_\ell\right)}. \tag{29}$$

4.2 Investigation into negative γ_{ij}'s

As mentioned, the instantaneous covariance matrix $\Sigma\left(\boldsymbol{\mu}_t\right)$ needs to be positive semi-definite. This requires each realised γ_{ij} to be non-negative. However, after applying the above estimators (29) and the corresponding one using the naive integrated covariance estimator, the realised γ_{ij}'s were observed to not all be non-negative and appeared to be distributed symmetrically around a non-negative number (see Figure (5)). There are two reasons for this phenomenon: one is the ubiquitous estimation error, which is due to sparsity of data, a second is correlation between the market weights, on which we concentrate in the sequel.

Note that positively correlated market weights lead to negative γ_{ij}'s. Intuitively,

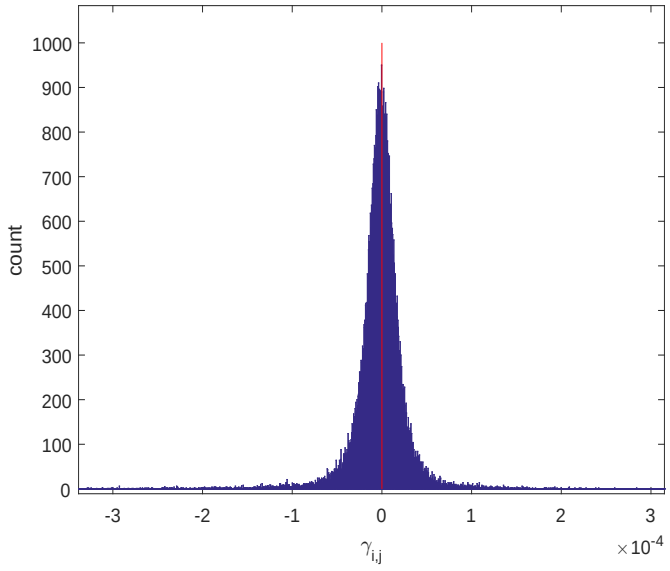


Figure 5: Histogram representing the realised gammas for each pair of stocks in the data set.

one would expect the correlation between corr$(\Delta\mu_k^i, \Delta\mu_k^j)$ $(k \in 1, 2, ..., N)$ to be negative, since when μ^i increases, μ^j (for $j \neq i$) must decrease so as to maintain $\sum_{i=1}^N \mu_k^i = 1$. However, if μ^i increases it is possible that some μ^j, for $j \neq i$ may increase while others decrease so that $\sum_{i=1}^N \mu_k^i = 1$. Essentially positive correlations in $\Delta\mu_k^i$ and $\Delta\mu_k^j$ are possible as long as there are negative correlations of a similar magnitude. Extending this intuition, it is more likely that $\Delta\mu_k^i$ and $\Delta\mu_k^j$ will be positively correlated if the underlying changes in market capitalisations ΔS^i and ΔS^j are positively correlated.

This intuition can be confirmed by Figure (6), where a relatively strong negative correlation between γ_{ij} and corr$(\Delta S^i, \Delta S^j)$ is visible. These findings indicate that in a highly positively correlated market it is unlikely that each γ_{ij} will be strictly positive and, as a result, it is likely that polynomial models do not fit to such a situation.

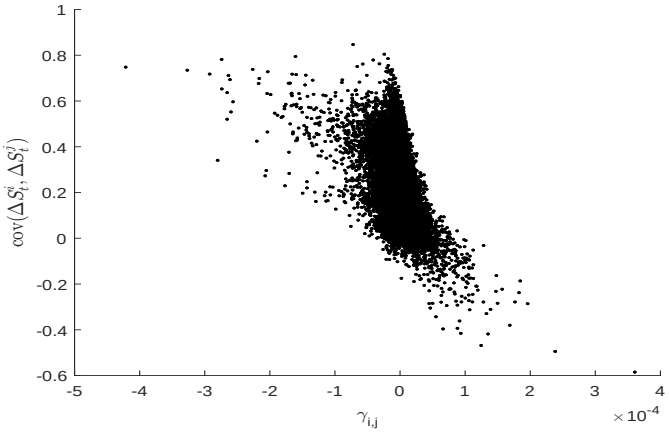


Figure 6: Plot of γ_{ij} versus corr$(\Delta S^i, \Delta S^j)$.

5 Simulation results

5.1 Simulation Set Up

A simulation study was performed to analyse how the model recovers the modelled gamma matrix. The focus is to verify whether a model with a smooth and strictly positive gamma matrix over the same time period produces similarly unstable gamma matrix. An unstable simulated gamma matrix similar to the market calibrated gamma matrix would suggest the instability is an artefact of the dimension of the calibration data, particularly a low ratio of the number of stocks to the number of days. To achieve this the simulation is seeded with a smoothed and strictly positive gamma matrix and we observe the gamma matrix obtained from the simulated trajectories.

As described in Section 2 we consider the following model

$$d\mu_t = b_t dt + \sqrt{\Sigma(\mu_t)} dW_t,$$

where W is a d-dimensional Brownian motion, $\Sigma(\mu)_{ii} = \sum_{i \neq j} \gamma_{ij}^{\mu} \mu^i \mu^j$ and $\Sigma(\mu)_{ij} = -\gamma_{ij}^{\mu} \mu^i \mu^j$ with $\gamma_{ij}^{\mu} = \gamma_{ji}^{\mu} \geq 0$ for all $i \neq j$ and b an (arbitrary not necessarily affine) drift process. We model the total capitalisation process $\bar{S} := \sum_{i=1}^{d} S^i$ by a Black &

Scholes model independent of μ of the form

$$d\bar{S}_t = \lambda \bar{S}_t dt + \bar{S}_t \sigma dB_t,$$

where B is a Brownian motion independent of W and $\lambda, \sigma \in \mathbb{R}$. Since $S^i = \mu^i \bar{S}$ we obtain the following dynamics (see Appendix B for the derivation) for $\log(S)$.

$$d \log S_t = (\frac{b_t}{\boldsymbol{\mu_t}} + \lambda \mathbf{1} - \frac{1}{2} \text{diag} \widetilde{Q}(S_t))dt + \sqrt{\widetilde{Q}(S_t)} d\widetilde{W}_t, \tag{30}$$

where $\frac{b_t}{\boldsymbol{\mu_t}}$ is the componentwise ratio, $\widetilde{W}$ is d-dimensional Brownian motion and

$$\widetilde{Q}^{ij}(S_t) = -\gamma_{ij} + \sigma^2, \quad i \neq j, \tag{31}$$

and

$$\widetilde{Q}^{ii}(S_t) = \sigma^2 + \sum_{i \neq j} \gamma_{ij} \frac{S_t^j}{S_t^i}, \tag{32}$$

For this above model a standard Euler scheme is implemented.

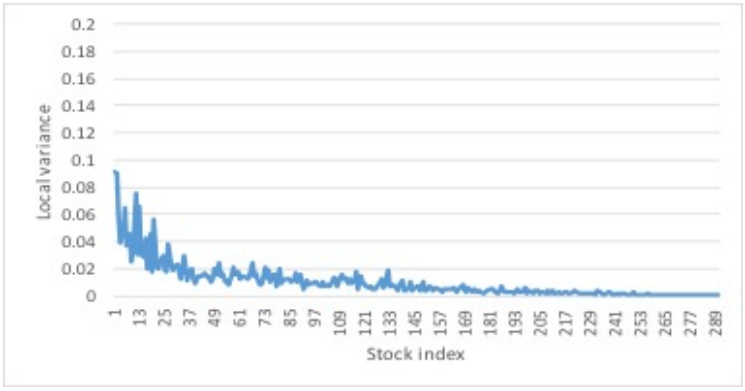
5.2 Stability Analysis

Additional analysis was performed on the instantaneous variance for each simulated stock. The motivation for the analysis is the intuition that the $\gamma_{ij} \frac{S_t^j}{S_t^i}$ term in (32) creates potentially unstable results as the ratio $\frac{S_t^j}{S_t^i}$ changes over time. The analysis looked at the pathwise behaviour of the variance terms $\widetilde{Q}^{ii}(S_t)$.

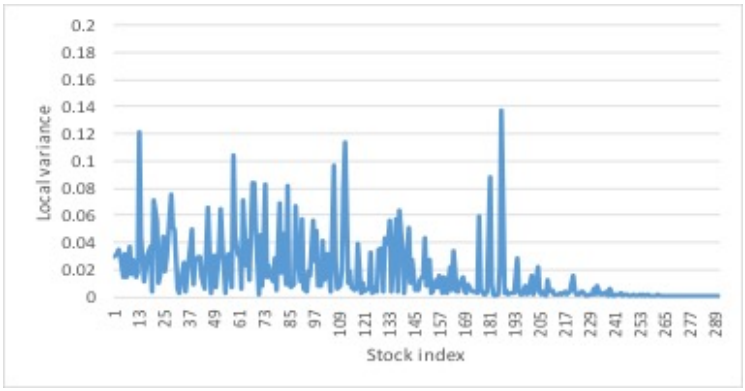
The analysis proceeds by examining the variance plot of $\widetilde{Q}^{ii}(S_t)$ at time $t = 0$ and $t = 1$. The plot is ranked based on initial market capitalisation with increasing order from left to right.

For the gamma matrix calibrated directly from the market (including "not allowed"

negative gamma entries), the variance plot indeed becomes progressively unstable over time as shown in figure (5.2) below.



(a) t = 0



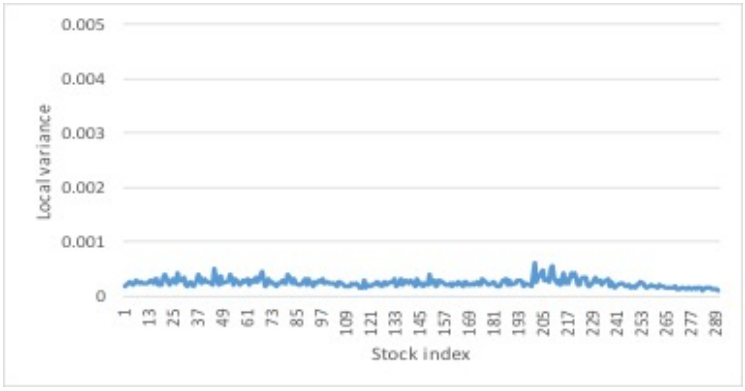
(b) t = 1 year

Figure 7: $\widetilde{Q}^{ii}$ using "directly" calibrated gamma.

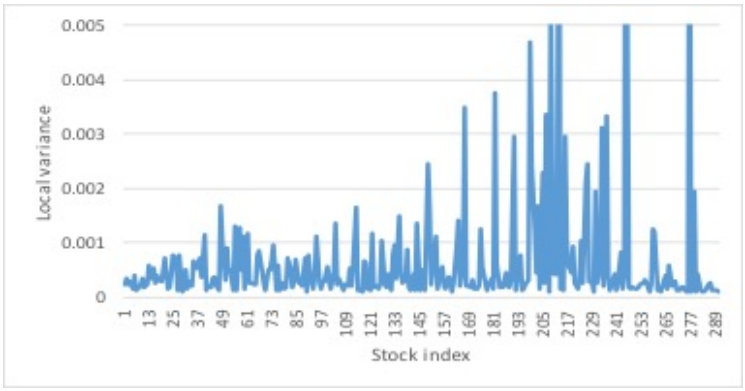
For the "idealised" gamma matrix discussed in (5.3.3), the instability is also observed, however, after a much longer simulation period.

The results shed light on the stability of the chosen model: on the one hand the form in (32) is key for the stylised fact of higher volatility for low caps, on the other hand this form is potentially causing instability, which becomes visible after rather long periods of 20 years.

(a) t = 0.



(b) t = 1 year.



(c) t = 20 years.

Figure 8: Instantaneous variance using "idealised" gamma.

5.3 Gamma calibration

The estimation of each γ_{ij} was made difficult by spurious correlations arising from the sparsity of data. In order to generate reasonable set of γ_{ij} it was necessary to attempt to mitigate the effects of these erroneous correlations. The first method sets all negative γ_{ij}equal to 0. The second method uses a set of assumptions to simplify the estimated correlations. The third method attempts to generate a parametric form of the integrated covariance matrix structure.

5.3.1 Method 1: removal of negative γ_{ij}

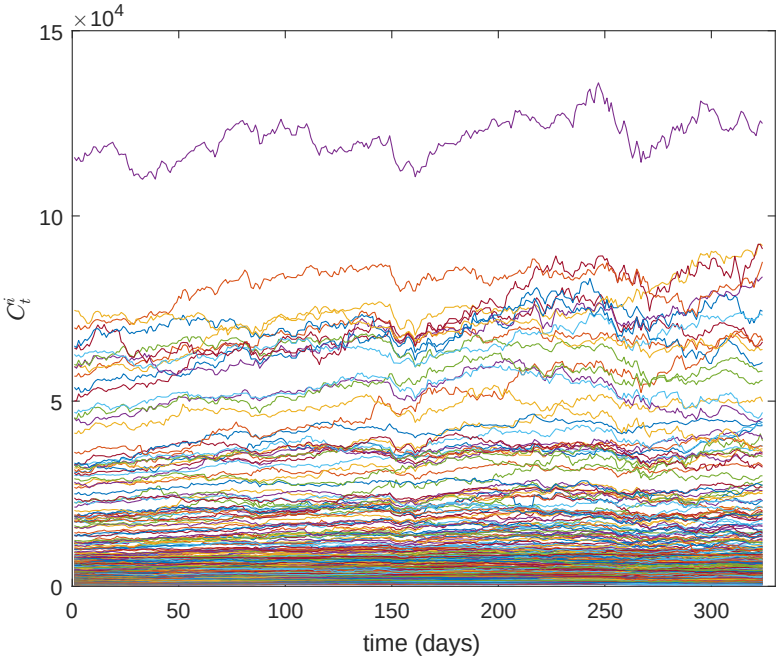
This model disallows any negative γ_{ij} if a calculated γ_{ij} is negative it is set to 0. However, when removing these Γ_{ij} one must be careful to conserve the relationship given by:

$$\int_0^t \Sigma_t^{ii} dt = \sum_{j \neq i}^{d} \sum_{t=1}^{N} \gamma^{ij} \mu_t^i \mu_t^j (t_n - t_{n-1}). \tag{33}$$

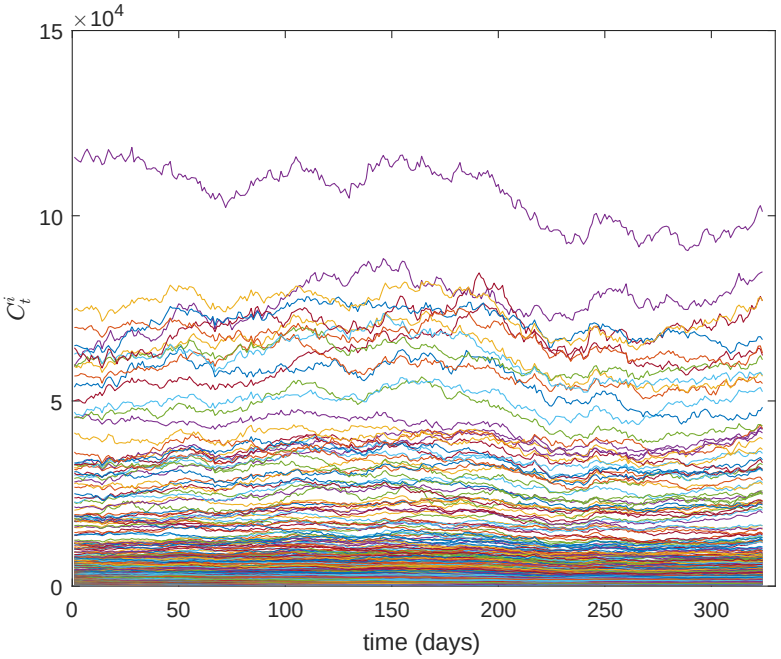
Each of the three methods ensure that this relationship is maintained.
In order to assess the validity of this model a simulation was run under for the same timespan as the market data (Figure 9b) and compared to the the trajectories of the market data (Figure 9a). These figures show that, while the volatility of each market capitalisation is roughly correct, the market capitalisations (especially the low capitalisations) are strongly correlated. This phenomena suggests that removing the negative γ_{ij} leads to an exaggerated correlation structure.
Figure (10) shows the average magnitude of the γ_{ij} as a function of market capitalisation. It is clearly visible that γ_{ij} decreases with decreasing capitalisation, implying an increasing correlation with decreasing capital (due to the fact that the instantaneous covariance for the market capitalisation is given by $\sigma^2 - \gamma_{ij}$).

5.3.2 Method 2: simplified correlation structure

This model was used for the majority of the simulations, as a result this section contains a more in depth validation. This model made several simplifying assump-



(a) Market portfolio capitalisation trajectories.



(b) Simulated portfolio capitalisation trajectories using Method 1 for γ_{ij} estimation.

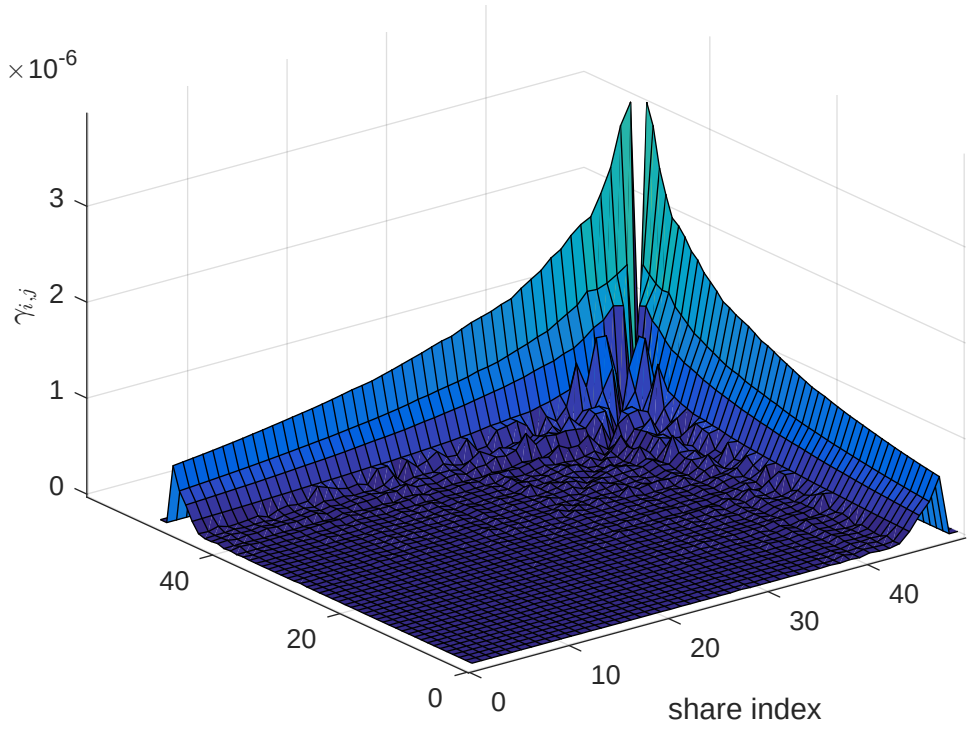


Figure 10: γ_{ij} matrix surface (where share index 0 represents the smallest capitalisation group and 50 is the largest).

tions in an attempt to generate a stable set of γ_{ij}. The assumptions required in the proceeding algorithm are as follows:

1. The market (S_t^i) can be broken into d overlapping markets of size $\tilde{d}^k$, $\tilde{S}_t^{(k)i} \quad k \in \{1, \ldots, d\}, i \in \{1, \ldots \tilde{d}^k$. Note that $\tilde{d}^k$ is not necessarily constant over each block.

2. Significant correlations exist only for stocks with similar market capitalisations - stocks with highly different capitalisations were assumed to be uncorrelated.

3. Each γ_{ij} could be most accurately represented by the mean value of each $\gamma_{ij}^{(k)}$ calculated for overlapping market blocks.

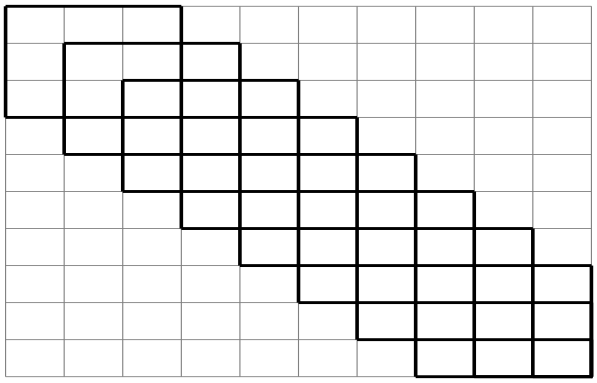


Figure 11: Illustration of overlapping block formation

4. For a suitably small market each off-diagonal element of the covariance matrix (of $\mu^{(k)i}$) should be negative (as an increase in the weight of $\tilde{\mu}^{(k)i}$ should lead to a decrease in the weights of $\mu^{(k)j},\ j \neq i$).

In order to understand the importance of these assumptions it is necessary to consider the implementation of the algorithm. The market capitalisation data is first separated into d overlapping markets. For each market block an integrated covariance matrix is calculated (using the Fourier method previously discussed with $N = 200$) and the $\gamma_{ij}^{(k)}$ are determined. Each γ_{ij} is then taken as the mean of $\gamma_{ij}^{(k)}$ over overlapping market blocks. All γ_{ij} that are not calculated in at least one block are set to 0. Any negative γ_{ij} are set to zero (as a result of assumption 4). Figure (11) above illustrates the overlapping blocks on which $\gamma_{ij}^{(k)}$ are calculated.

The validity of the calculated γ_{ij} was assessed based on its ability to replicate the observed market behaviour, the persistence of the γ_{ij} through a simulation and the stability of the simulation through time.

The simulated capitalisation trajectories can be seen in Figure (12). The simulation appeared to generate behaviour similar to that seen in the market data. There was, however, a weaker correlation between the high and low capitalisations (which is expected by construction). In other words, while the above described method produces positive γ_{ij} in each block, it neglects the correlation of small capitalisations and large capitalisations.

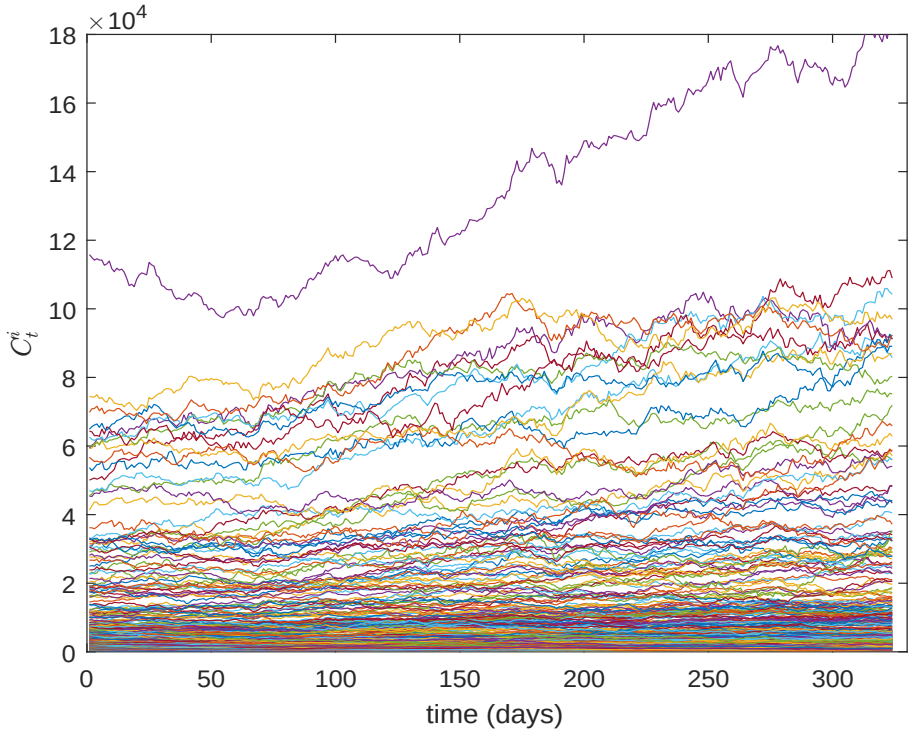


Figure 12: Simulated portfolio capitalisation trajectories using Method 2 for γ_{ij} calculation

In Figure 9a there are visible regions of stronger correlations which occur around the dips in capitalisation. These regions may be a remnant of the removed jumps, as these dips usually occurred immediately after a large upward jump in the capitalisations. The presence of the regions of high correlations may then be a result of macro market shifts and not fully represent the correlation structure during normal market conditions.

In order to analyse the stability of the calculated γ_{ij} capitalisation trajectories were simulated using the γ_{ij} calculated from the market data in the manner described above. A new set of γ_{ij}^2 was then estimated from the simulated trajectories and compared to (γ_{ij}^1). Figure 13 shows that when passed through a simulation each γ_{ij} tends to increase in magnitude. Despite the increase in γ_{ij}, the strong correlation between γ_{ij}^1 and γ_{ij}^2 indicates that γ_{ij} can be calculated from relatively sparse data

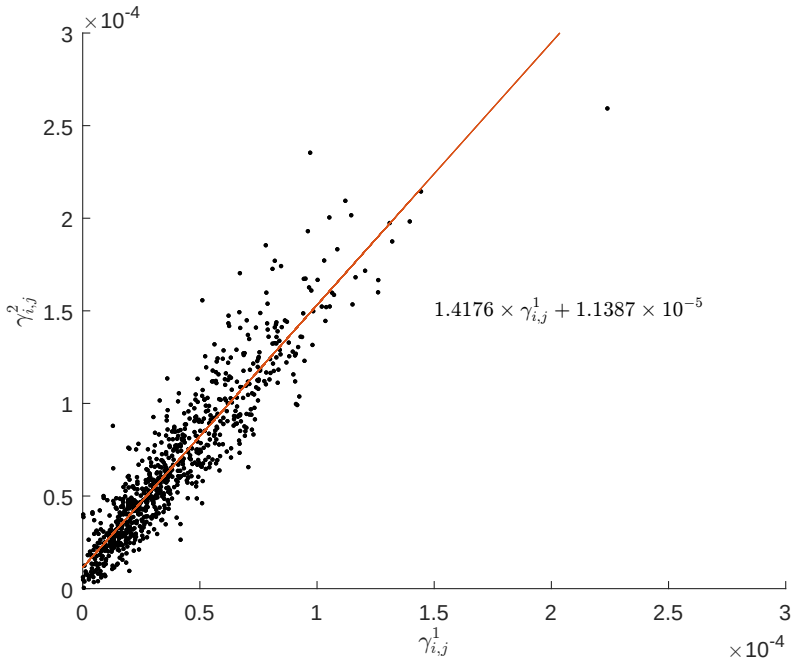


Figure 13: Analysis of γ_{ij} stability through simulations, with plotted line of best fit/

while maintaining a good estimate of the actual correlation structure.

As mentioned in Section (1), classical models tend to poorly model the progression of μ^i through time. Figure 14 shows that this model is capable of preserving the relationship between $\log(\mu^i)$ and $\log(\text{rank})$ over the duration of the market data. Moreover, the dynamics of the capital distribution curves exhibit a very reasonable behaviour, which is a very promising result.

Figure (15) shows the volatilities of several randomly selected stocks over the length of the simulation. As one would expect, the instantaneous volatilities display a small amount of correlation and appear stable over the length of the simulation.

An analysis of the initial and final instantaneous volatilities (shown in Figure 16) shows that there are several stocks that become slightly more volatile over the length of the simulation (10 years). This indicates that the simulation may be unstable for durations much greater than 10 years.

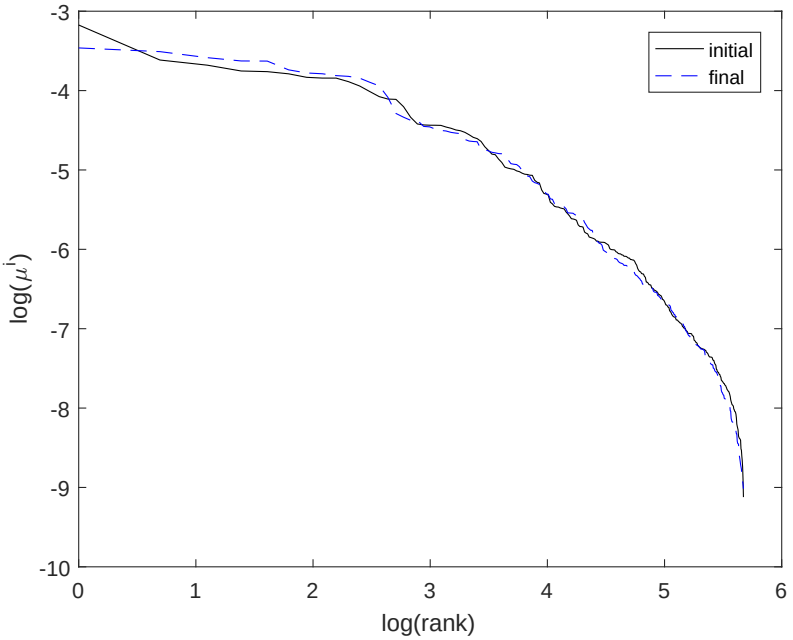


Figure 14: Capitalisation curves over time

5.3.3 Method 3: "idealised" gamma

An "idealised" gamma was created as one of the various calibration methods explored in this study. The "idealised" gamma matrix is created to impose two desirable features. The first feature is a market observed integrated variance smoothed across realised market capitalisation. This is achieved by fitting a quadratic polynomial to the integrated variances as a function of capital weight, see Figure (5.3.3).

The second desirable feature is a smooth and rapidly decreasing integrated covariance as a function of the difference between capitalisation size. To achieve this, an exponential function is projected perpendicularly from the diagonal of the integrated covariance matrix. Each off diagonal element is calculated as $\sum^* e^{-D}$ where D is a scaled perpendicular distance to the diagonal. $\sum^*$ represents the closest point on the diagonal. The off-diagonal elements are then recursively rescaled to maintain consistency between the diagonal and off-diagonal elements as described in (33). The "idealised" gamma matrix is shown in figure (5.3.3).

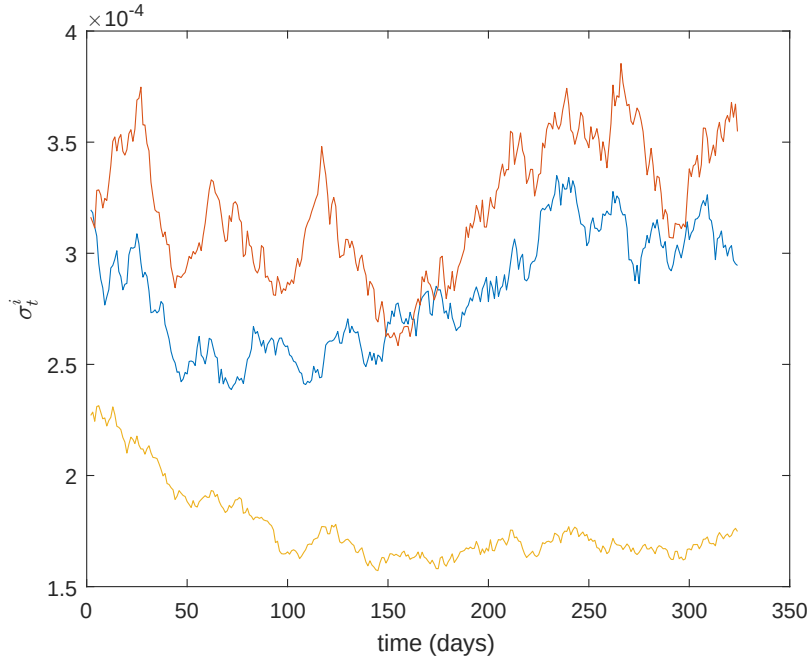


Figure 15: Volatility trajectories for randomly selected shares.

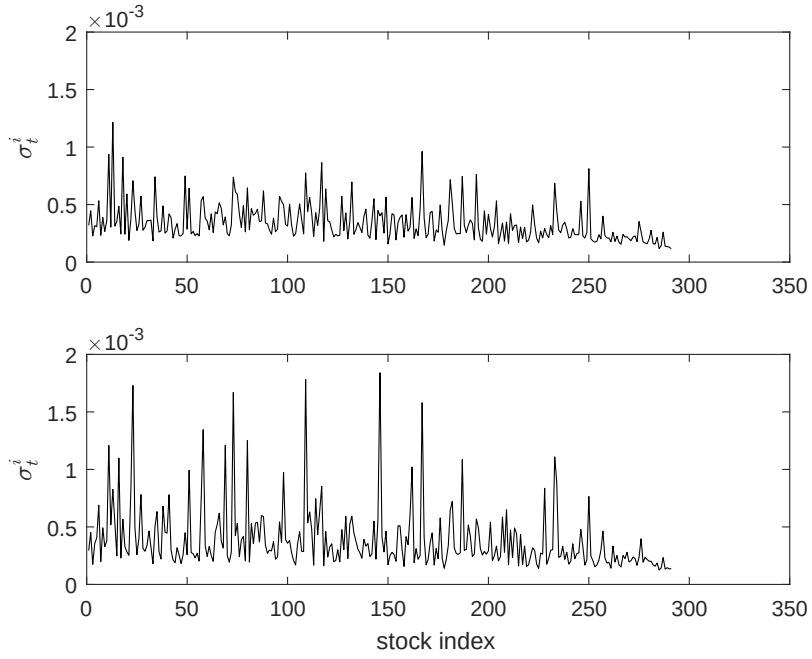


Figure 16: Comparison of share volatilities at t_0 and t_n.

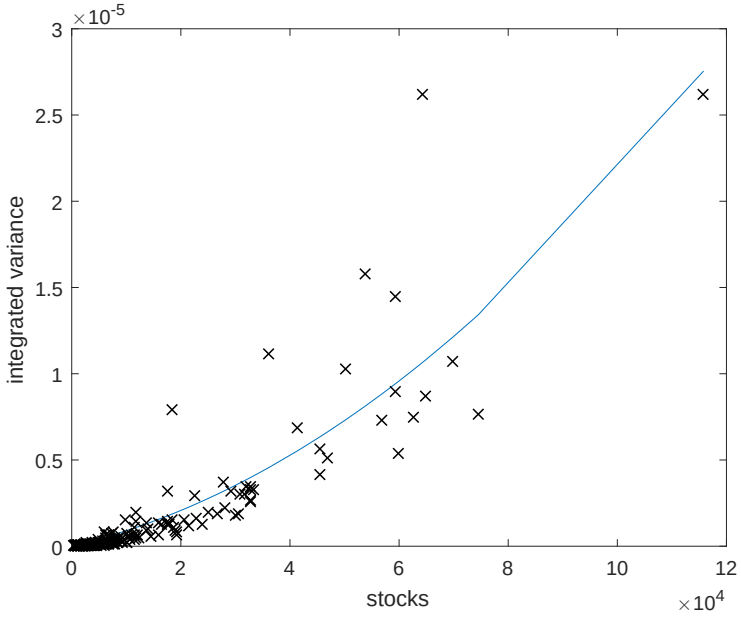


Figure 17: Quadratic polynomial fit to integrated variances.

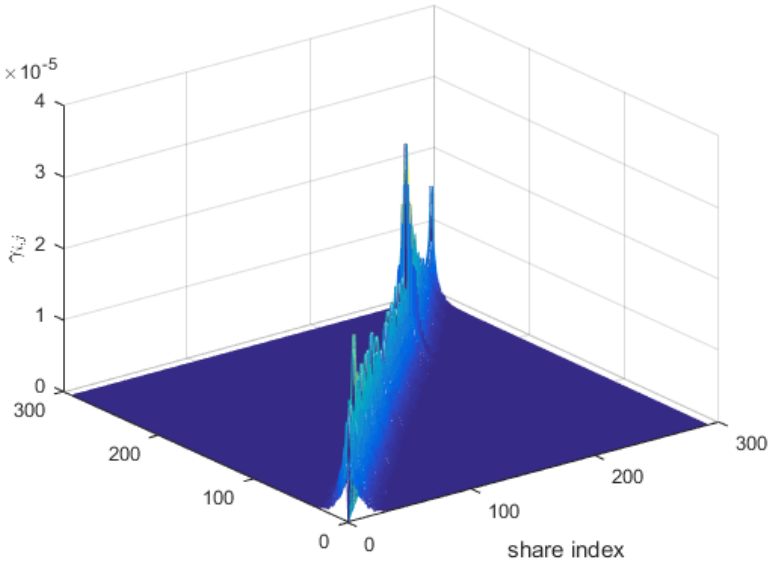


Figure 18: "Idealised" γ.

To further demonstrate the link between instability and the ratio $\frac{S_t^j}{S_t^i}$, as well as propose a hypothetical modelling approach, the simulation was adapted such that at each step the simulated stocks are reassigned to γ_{ij}'s based on the current state rank. This effectively relates the gammas to stocks by rank rather than by stock name. The intended effect of this change is to stabilise the ratios with respect to the corresponding initial state gammas.

The simulated results confirm this model is completely stable over time which further confirms the initial intuition regarding the impact of the ratio on model stability.

In terms of modelling, such a model modification appears mathematically intractable. It can be rationalised intuitively as follows. Instead of modelling individual stocks, it is a model of ranked states. The individual stocks are modelled as switching between the modelled states. When combined with the results discussed in Section (3.2) and the intuition that a change in rank is more likely when a jump occurs, this modelling approach can be intuited as energy state jumping where each energy state is modelled as a stochastic process. The trading strategy can incorporate rebalancing the portfolio to coincide with the rank changes.

5.4 Empirical study of functionally generated portfolios on simulated data

We undertook an analysis of the performance of our three functionally generated portfolios, using the polynomial model (1) which admits relative arbitrage (see Section 2 above) under certain conditions. Our simulation exercise was performed in order to ascertain when relative arbitrage of our portfolios over the market portfolio pervades.

We simulated the log market capitalisations for each stock using a Euler discretisation scheme (see Section 3), firstly considering the martingale case, and secondly incorporating a strong drift. The former case is an arbitrage-free model, while the latter admits arbitrage. Indeed, the drift we used in the latter case was an estimation of the drift of the entire market capitalisation, calculated by the mean of the log returns of each stock over the period of investigation. To formalise, when we assume no arbitrage we suppose that the log market capitalisations follow the

following process:

$$d\log(S_t) = -\frac{1}{2}\text{diag}\tilde{Q}(S_t)\,\mathrm{d}t + \sqrt{\tilde{Q}(S_t)}d\tilde{W}_t.$$

However, if arbitrage is admitted then the dynamics for $\ln(S_t)$ are assumed to follow

$$d\log(S_t) = \left(\lambda_1\mathbf{1} - \lambda_2\text{diag}\tilde{Q}(S_t)\right)\mathrm{d}t + \sqrt{\tilde{Q}(S_t)}d\tilde{W}_t.$$

where $\mathbf{1}$ is a d-dimensional vector of ones and λ_1, λ_2 are constants.

The top panel of Figure (19) shows the trajectories of the relative wealth process Y_t^π in the case when the model does not admit arbitrage, with respect to each of the three functionally generated portfolios. Since no arbitrage is allowed, we observe the relative wealth process fluctuating around one for the investigated simulation period of one year. However, we see that when we admit arbitrage, for the same seed relative arbitrage is evinced. This can be seen in the bottom panel of Figure (19).

The trajectories of the relative wealth process Y_t^π in the case when the model admits arbitrage, with respect to each of the three functionally generated portfolios, are shown in Figures (20) and (21).

It is clear from Figure (20) that relative arbitrage over the market portfolio is possible from 7 years and onwards, for both the diversity- and entropy-weighted portfolios. Nonetheless, it is evident from Figure (21) that the equally-weighted portfolio demonstrates relative arbitrage over the market portfolio from an earlier time, that is two years. In addition, a striking feature of such a portfolio is that it provides markedly better outperformance compared to the diversity- and entropy-weighted portfolios. However, we caution that the reason for the behaviour of the equally-weighted portfolio lies within the way we simulated our market capitalisations, not in that it is performing very well. In the simulation scheme, we specified that if a market capitalisation for a particular stock hits the barrier of one, it immediately drifts upwards[3]. Indeed this is an (implicit) unbounded profit. And since the equally-weighted portfolio gives constant weight to all stocks at all times, it stands to reason that it is most affected (in terms of attaining relative outperformance) by

[3]The reason for this imposition is so that we keep the assets nonnegative

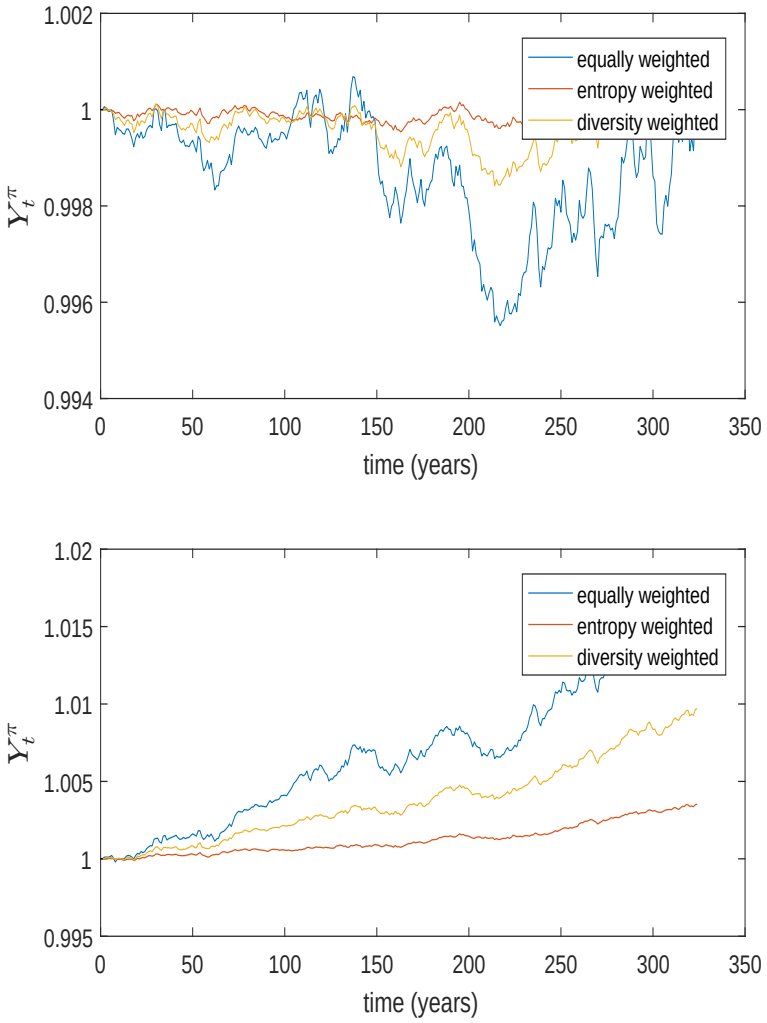


Figure 19: Trajectory of the relative wealth process for each of the diversity-weighted, entropy-weighted and equally-weighted portfolios, for the case when the model does not admit arbitrage (top panel) and does admit arbitrage (bottom panel). *(Erratum: the x-axis label should say "days" and not "years".)*

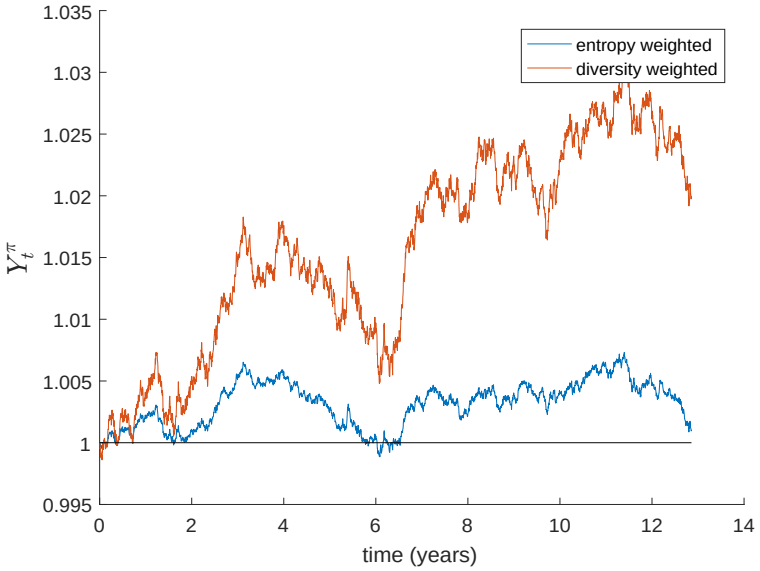


Figure 20: Trajectory of the relative wealth process for each of the diversity-weighted and entropy-weighted portfolios.

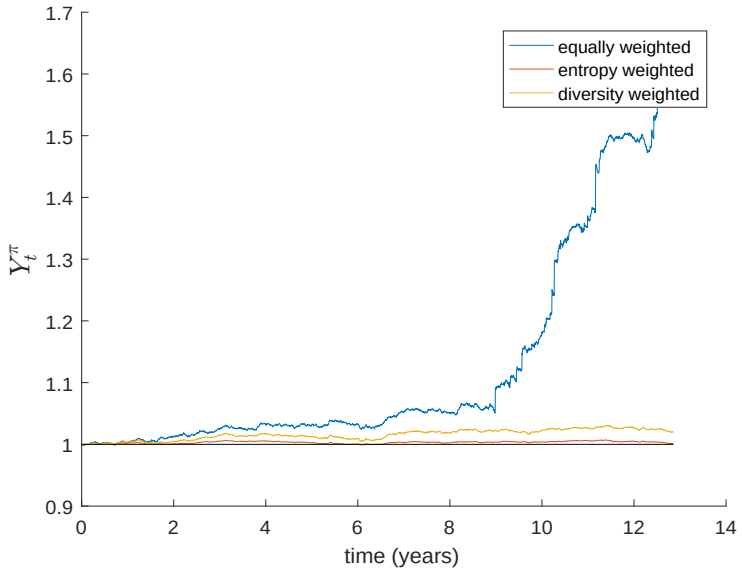


Figure 21: Trajectory of the relative wealth process for the diversity-weighted, entropy-weighted and equally-weighted portfolios.

the cap we enforced in the modelling. The remaining two portfolios are not much affected by this phenomenon enforced in the modelling of the market capitalisations.

We end this section by providing some comment on the practical implications of our portfolio strategies, and the reasons why they may be difficult to implement in practice. Firstly, we emphasise the assumption that the company stocks comprising the market index (i.e. the market portfolio) remain the same over time - this is not the case in practice. Indeed, the index contains many small cap stocks, and these will enter and leave the index as time progresses. So, we will not always be holding the same stocks. Secondly, losses on the investment in particular stocks is inevitable - some companies will default, and this risk will be magnified the longer the time period over which we invest and intend to realise the relative arbitrage over. Finally, all our analyses have ignored transaction costs: given that our portfolio strategy is dynamic over time, transaction costs can be expected to be quite large and will erode investment returns.

6 Discussion and conclusions

It was one goal of this work to calibrate a polynomial diffusion model's instantanous covariance for market weights of a large index based on relatively sparse daily data. This reflects a realistic situation and is therefore an important calibration task, since we can neither assume high frequency data nor considerably longer time series. Additionally we ask for a quick and direct calibration method whose complexity lies considerably below standard filtering techniques. This allows for easy model calibration and validation.

Here three methods have been proposed, each of which produces a realistic model calibration, which passes some statistical tests. Of course the methods can be refined given more implementation and computation time in the direction of Bayesian methods.

After having established the instantanous covariance structure of the polynomial model the normative requirements of stochastic portfolio theory come to play. Having certain qualitative assumptions, like the existence of relative arbitrages in the market model, the model can be used to actually construct and validate these arbitrages. More concretely, given a γ matrix, one can calculate which generating function should actually be used and over which time horizon relative arbitrages will appear.

Appendix A

Proof of Lemma 1

For any portfolio process $\{V_t : t \geq 0\}$, it is well-known that (for example, see Jeanblanc et al. (2009))

$$V_t = V_0 + \sum_{i=1}^{d} \int_0^t \phi_s^i dS_s^i \tag{A.1}$$

$$= \sum_{i=1}^{d} \phi_t^i S_t^i \tag{A.2}$$

for any trading strategy $\left\{\phi_t^i : t \geq 0\right\}$ and stock price process $\{S_t : t \geq 0\}$. In the current context, from (5) we have that

$$dV_t^\pi = \sum_{i=1}^{d} \frac{\pi_t^i V_t^\pi}{S_t^i} dS_t^i,$$

and also notice that our wealth process is self-financing. Moreover, $\bar{S}_t = V_t^\mu$. Now consider $\frac{\pi_t^i V_t^\pi}{S_t^i}$ as the trading strategy, and the discounted processes $\frac{V_t^\pi}{V_t^\mu}$ and $\frac{S_t^i}{V_t^\mu}$ (i.e. V_t^μ is the numerairé), and apply (A.1) to obtain

$$\begin{aligned}
\frac{V_t^\pi}{V_t^\mu} &= \frac{V_0^\pi}{V_0^\mu} + \sum_{i=1}^{d} \int_0^t \frac{\pi_s^i V_s^\pi}{S_s^i} d\left(\frac{S_s^i}{V_s^\mu}\right) \\
&= \frac{V_0^\pi}{V_0^\mu} + \sum_{i=1}^{d} \int_0^t \pi_s^i \frac{V_s^\pi}{V_s^\mu} \frac{1}{\mu_s^i} d\left(\mu_t^i\right)
\end{aligned}$$

by noticing that $\frac{S_t^i}{V_t^\mu} = \mu_t^i$. The result follows. ■

Appendix B

Derivation of model for log(S)

Consider

$$d\mu_t = b_t dt + \sqrt{\Sigma(\mu_t)} dW_t, \quad \text{for} \quad W_t^i \perp\!\!\!\perp W_t^j, \quad \text{for all} \quad i \neq j,$$

set

$$\bar{S}_t = \sum_{i=1}^{d} S_i$$

and define

$$d\bar{S}_t = \lambda \bar{S}_t dt + \sigma \bar{S}_t dB_t, \quad W_t^i \perp\!\!\!\perp B_t \quad \text{for all} i$$

We know that

$$S_t = \mu_t \bar{S}_t.$$

Thus applying the multidimensional Itô formula for Itô processes yields

$$\begin{aligned} dS_t &= \bar{S}_t d\mu_t + \mu_t d\bar{S}_t \\ &= \bar{S}_t (b_t dt + \sqrt{\Sigma(\mu_t)} dW_t) + \mu_t (\lambda \bar{S}_t dt \sigma \bar{S}_t dB_t) \\ &= \bar{S}_t (b_t dt + \sqrt{\Sigma(\mu_t)} dW_t) + \lambda S_t + \sigma S_t dB_t, \end{aligned}$$

so that we can write

$$dS_t = (\bar{S}_t b_t + \lambda S_t) dt + \sqrt{Q(S_t)} d\widetilde{W}_t$$

for some Brownian motion $\widetilde{W}$ and some matrix $Q(s)$ that we compute now. Note that

$$\begin{aligned} dS_t^i dS_t^j &= \sum_{k=1}^{d} (\sqrt{Q(S_t)})^{ik} d\widetilde{W}_t^k + \sum_{l=1}^{d} (\sqrt{Q(S_t)})^{jl} d\widetilde{W}_t^l \\ &= \sum_{k=1}^{d} (\sqrt{Q(S_t)})^{it} (\sqrt{Q(S_t)})^{jk} dt \\ &= Q(S_t)^{ij} dt \end{aligned}$$

and

$$dS_t^i = \bar{S}_t (\sum_{k=1}^{d} (\sqrt{\Sigma(\mu_t)})^{ik} dW_t^k) + \sigma S_t^i dB_t,$$

$$dS_t^j = \bar{S}_t (\sum_{k=1}^{d} (\sqrt{\Sigma(\mu_t)})^{ik} dW_t^k) + \sigma S_t^j dB_t.$$

Hence

$$dS_t^i dS_t^j = \bar{S_t^i} \bar{S_t^j} \sum_{k=1}^{d} (\sqrt{\Sigma(\mu_t)})^{ik} (\sqrt{\Sigma(\mu_t)})^{jk} dt + \sigma^2 S_t^i S_t^j dt$$

If $i \neq j$

$$\begin{aligned} Q(S_t))^{ij} &= \bar{S}_t \bar{S}_t (-\gamma_{ij} \mu_t^i \mu_t^j) + \sigma^2 S_t^i S_t^j \\ &= -\gamma_{ij} S_t^i S_t^j + \sigma^2 S_t^i S_t^j \end{aligned}$$

Also

$$\begin{aligned} (Q(S_t))^{ii} &= (\bar{S}_t)^2 (\sum_{j \neq i} \gamma_{ij} \mu_t^i \mu_t^j) + \sigma^2 (S_t^i)^2 \\ &= \sum_{j \neq i} \gamma_{ij} S_t^i S_t^j + \sigma^2 (S_t^i)^2 \end{aligned}$$

Consider now $\log(S_t)$. By Itô's formula, we get that

$$d \log(x_t) = \frac{1}{x} dx - \frac{1}{2} \frac{1}{x^2} (dx_t)^2$$

$$d\log(S_t^i) = \frac{1}{S_t^i} dS_t^i - \frac{1}{2}\frac{1}{(S_t^i)^2} Q^{ii}(S_t)dt,$$

so that we obtain

$$d\log(S_t) = \left(\frac{b_t}{\mu_t} + \lambda - \frac{1}{2}\text{diag}(\widetilde{Q})\right) dt + \sqrt{\widetilde{Q}} d\widetilde{W}_t,$$

where

$$(\widetilde{Q}(S_t))^{ij} = -\gamma_{ij} + \sigma^2 \quad , i \neq j$$

$$(\widetilde{Q}(S_t))^{ii} = \sigma^2 + \sum_{j \neq i} \gamma_{ij} \frac{S_t^j}{S_t^i}$$

Appendix C

Matlab code for some of the implementation procedures

Stochastic integration

```
function [ yt ] = Stochint( mc,type )
%discrete stochatic integration
mu=mc./repmat(sum(mc,1),size(mc,1),1);
if type==1
%equally weighted
Pii=ones(size(mu,1),1)/size(mu,1);
yt=ones(size(mu,2),1);
for n=2:length(yt)
yt(n)=yt(n-1)+yt(n-1)*sum((mu(:,n)-mu(:,n-1)).*Pii./
mu(:,n-1));
end
elseif type==2
%entropy
yt=ones(size(mu,2),1);
for n=2:length(yt)
c=1;
Hx=-sum(mu(:,n-1).*log(mu(:,n-1)));
Hxc=Hx+c;
```

```
Pii=mu(:,n-1)/Hxc.*(c-log(mu(:,n-1)));
yt(n)=yt(n-1)+yt(n-1)*sum((mu(:,n)-mu(:,n-1)).*Pii./
mu(:,n-1));
end
elseif type==3
%diversity weighted
p=0.5;
yt=ones(size(mu,2),1);
for n=2:length(yt)
Pii=(mu(:,n-1).^p)/sum(mu(:,n-1).^p);
yt(n)=yt(n-1)+yt(n-1)*sum((mu(:,n)-mu(:,n-1)).*Pii./
mu(:,n-1));
end
end
end
Not enough input arguments.
Error in Stochint (line 3)
mu=mc./repmat(sum(mc,1),size(mc,1),1);
```

Simulation study

```
function [ Stmat gam bmat gamCount sigMat] =
Stockgen( N,t,mc,D_size,t_sim,st )
%Master function: market data -> Sigma -> Gamma -> simulated paths
%bmat,gamcount,sigMat are only ouputs in order to generate plots.
Ngroup=ceil(size(mc,1)/D_size);
Lsize=size(mc,1)-(Ngroup-1)*D_size;
if Lsize<3
Ngroup=Ngroup-1;
ngadd=Lsize;
end
[~,I]=sort(mc(:,1));
mc=mc(I,:);
%remove jumps
Stmat=zeros(size(mc,1),length(t_sim));
```

```
dmc=-mc(:,1:end-1)+mc(:,2:end);
stdmat=repmat(std(dmc,[],2),1,size(dmc,2));
dmc(abs(dmc)>3*stdmat)=0;
mc2=zeros(size(mc));
mc2(:,1)=mc(:,1);
for n=2:size(mc2,2)
mc2(:,n)=mc2(:,n-1)+dmc(:,n-1);
end
mc=mc2;
%declerations
gamMat=zeros(size(mc,1),size(mc,1));
sigMat=zeros(size(mc,1),size(mc,1));
gamCount=zeros(size(mc,1),size(mc,1));
%loops through each block in diagonal
for P=1:size(mc)-1
fprintf('Calculating group %d.\n',P)
if P>=size(mc,1)-D_size
%determine current market block
mcur=mc((P-1)+1:size(mc,1),:);
%Calculate gamma and int Sig
[gamcur sigcur]=gamcalc(N,t,mcur);
%please in correct place in father gamma mat
gamMat((P-1)+1:size(mc,1),(P-1)+1:size(mc,1))...
=gamMat((P-1)+1:size(mc,1),(P-1)+1:size(mc,1))+...
gamcur.*(gamcur>0);
%count succesful additions to find average.
gamCount((P-1)+1:size(mc,1),(P-1)+1:size(mc,1))...
=gamCount((P-1)+1:size(mc,1),
(P-1)+1:size(mc,1))+1.*...
(gamcur>0);
%populate sigma
sigMat((P-1)+1:size(mc,1),(P-1)+1:size(mc,1))...
=sigMat((P-1)+1:size(mc,1),
(P-1)+1:size(mc,1))+sigcur.*...
(gamcur>0);
```

```
%not all necessary, these outputs were generated for
plotting
else
2
mcur=mc((P-1)+1:min((P-1)+1+D_size,size(mc,1)),:);
[gamcur sigcur]=gamcalc(N,t,mcur);
gamMat((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))...
=gamMat((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))+gamcur.*(gamcur>0);
gamCount((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))=...
gamCount((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))+1.*(gamcur>0);
sigMat((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))...
=sigMat((P-1)+1:min((P-1)+1+D_size),...
(P-1)+1:min((P-1)+1+D_size))+sigcur.*(gamcur>0);
end
end
%find average gam and int Sig
gamMat(gamCount~=0)=gamMat(gamCount~=0)./...
gamCount(gamCount~=0);
sigMat(gamCount~=0)=sigMat(gamCount~=0)./...
gamCount(gamCount~=0);
[Stmat bmat]=Euler_path(t_sim,gamMat,mc,sigMat);
if st==1
Stmat=Stmat(:,1:mesh_t:end);
end
gam=gamMat;
end
function [ gamdet Sig] = gamcalc(N,t,mc)
%Calculates gamma using fourier coefficients
t=t-t(1);
T=t(end);
```

```
%removes jumps again (does nothing in current state,...
%jumps already removed)
dmc=-mc(:,1:end-1)+mc(:,2:end);
stdmat=repmat(std(dmc,[],2),1,size(dmc,2));
dmc(abs(dmc)>3.5*stdmat)=0;
mc2=zeros(size(mc));
mc2(:,1)=mc(:,1);
for n=2:size(mc2,2)
mc2(:,n)=mc2(:,n-1)+dmc(:,n-1);
end
mu=mc2./repmat(sum(mc2),size(mc2,1),1);
Sig=zeros(size(mc,1),size(mc,1));
for I = 1:size(mu,1)
for J = I:size(mu,1)
ti=t;
tj=t;
muit=mu(I,:);
mujt=mu(J,:);
k1=find(muit==0);
3
k2=find(mujt==0);
ti(k1)=[];
muit(k1)=[];
tj(k2)=[];
mujt(k2)=[];
dmuit=-muit(1:end-1)+muit(2:end);
dmujt=-mujt(1:end-1)+mujt(2:end);
%Make more efficient
n = -N:N;
t1=1/T*sum(exp(-1i*2*pi/T*...
repmat(n',1,length(ti(2:end))).*repmat(ti(2:end),...
length(n),1)).*repmat(dmuit,length(n),1),2);
t2=1/T*sum(exp(1i*2*pi/T*...
repmat(n',1,length(tj(2:end))).*repmat(tj(2:end),...
length(n),1)).*repmat(dmujt,length(n),1),2);
```

```
t3=sum(t1 .*t2);
t4=T^2/(2*N+1)*t3;
Sig(I,J)=real(t4);
summ=0;
for k=1:length(muit)-1
summ=summ+muit(k)*mujt(k)*(t(k+1)-t(k));
end
gamdet(I,J)=-Sig(I,J)/summ;
end
end
Sig=Sig+Sig';
for n=1:size(Sig,1)
Sig(n,n)=Sig(n,n)/2;
end
gamdet=gamdet+gamdet';
for n=1:size(gamdet,1)
gamdet(n,n)=0;
end
gamdet(gamdet<0)=1d-9;
gamdet(gamdet==0)=gamdet(gamdet==0)+1d-9;
for I=1:size(gamdet,1)
sm=0;
for J=1:size(gamdet,1)
smm=0;
for k=2:length(t)
mui=mu(I,k);
muj=mu(J,k);
smm=smm+mui*muj*(t(k)-t(k-1));
end
sm=sm+gamdet(I,J)*smm*(I~=J);
end
gamdet(I,:)=gamdet(I,:)*Sig(I,I)/sm;
gamdet(:,I)=gamdet(I,:)';
4
end
```

```
end
function [St bmat] = Euler_path(te,game,mce,sige)
%Euler approximation
delt=te(2)-te(1);
sig=std(log(sum(mce(:,2:end))./sum(mce(:,1:end-1))));
delWt=sqrt(delt)*randn(size(game,1),length(te));
drift=0.000752133;
a=@(x) -1/2*diag(x);%+ones(size(diag(x)))*drift;
b=@(x) sqrtm(x);
S0=mce(:,1);
logSt=zeros(size(S0,1),length(te));
logSt(:,1)=log(S0);
bmat=[]
for n=2:length(te)
if any(n==(1:round(length(te)/100):length(te)))
fprintf('Euler: %f %%\n',100*n/length(te))
end
qmat=Qmat_fun(game,exp(logSt(:,n-1)),sig,sige,te);
bmat=[bmat;diag(qmat)'];
logSt(:,n)=logSt(:,n-1)+a(qmat)*delt+(b(qmat))*delWt(:,n-1);
%logSt(logSt(:,n)<log(1),n)=log(1); %floor long term
simulation
%stops numerical issues
end
St=exp(logSt);
end
function [ qmat ] = Qmat_fun( gam, St,sig,sigm,t )
%Q tilde determination
qmat=zeros(size(gam));
for I=1:size(gam,1)
qmat(I,I)=sig^2;
for J=1:size(gam,2)
if J~=I
qmat(I,J)=(-gam(I,J)+sig^2);
qmat(I,I)=qmat(I,I)+gam(I,J)*St(J)/St(I);
```

```
end
end
end
if all(eig(qmat) >= 0) ==0
error('Q must be positive definite')
end
end
Not enough input arguments.
Error in Stockgen_2 (line 4)
Ngroup=ceil(size(mc,1)/D_size);
```

Bibliography

Boudt, K., Cornelissen, J., Croux, C., 2012. Jump robust daily covariance estimation by disentangling variance and correlation components. Computational Statistics & Data Analysis 56 (11), 2993–3005.

Cuchiero, C., Keller-Ressel, M., Teichmann, J., 2012. Polynomial processes and their applications to mathematical finance. Finance and Stochastics 16 (4), 711–740.

Fernholz, E. R., 1998. Portfolio generating functions. Available at SSRN 139549.

Fernholz, E. R., 2002. Stochastic portfolio theory. In: Stochastic Portfolio Theory. Springer, pp. 1–24.

Fernholz, R., Karatzas, I., 2005. Relative arbitrage in volatility-stabilized markets. Annals of Finance 1 (2), 149–177.
URL http://dx.doi.org/10.1007/s10436-004-0011-6

Fernholz, R., Karatzas, I., 2009. Stochastic portfolio theory: an overview. Handbook of numerical analysis 15, 89–167.

Filipović, D., Larsson, M., 2016. Polynomial diffusions and applications in finance. Finance and Stochastics, 1–42.
URL http://dx.doi.org/10.1007/s00780-016-0304-4

Jeanblanc, M., Yor, M., Chesney, M., 2009. Mathematical methods for financial markets. Springer Science & Business Media.

Malliavin, P., Mancino, M. E., et al., 2009. A fourier transform method for nonparametric estimation of multivariate volatility. The Annals of Statistics 37 (4), 1983–2010.

Vervuurt, A., 2015. Topics in stochastic portfolio theory. Ph.D. thesis, University of Oxford.

Recalibration: a Criterion Based on Model Risk

Team 4

RALPH RUDD, University of Cape Town
CHRISTOPHER BAKER, University of Cape Town
QAPHELA MASHALABA, University of Cape Town
MELUSI MAVUSO, University of Cape Town

Supervisor:
ERIK SCHLÖGL, University of Technology, Sydney

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

Practitioners in the finance industry are mainly faced with at least four categories of model risk. The first model risk category regards the ability of the model to fit the market data; the second category of model risk regards how appropriately practitioners *use* their models, irrespective of the kind of fit their models provide. These two aspects of model risk−model fit risk and recalibration risk−are related and thus should not be considered or accounted for separately. Differing models generally respond disparately to the nature of their use: two distinct models could, for example, be recalibrated at the same frequency, but still respond to the recalibration to different extents. In dealing with model risk as it has been discussed above, it is essential to have in hand a tool or a notion which allows for comparison of model risk between a variety of models. With that tool in hand, it will be possible to answer interesting questions on model risk, particularly model risk relating to recalibration and model fit.

Recent work by Schlögl (2015) has attempted to classify the sources of model risk into four categories:

1. **Type 0:**
   The first type of model risk encapsulates parameter uncertainty - given a finite set of realizations of a model the parameters can only be estimated within a statistical error bound. Type 0 model risk also includes the risk owing to the sensitivity of the model to these parameter misspecifications.

2. **Type 1:**
   Type 1 model risk refers to the potential inability of the model to perfectly reprice or fit the full set of market observations. A simple example is the inability of the traditional Black-Scholes model to capture a volatility smile. Type 1 model risk also includes the calibration error or risk - the nonlinear nature of the necessary optimization problem preclude perfect recovery of the model parameters. This will lead to imperfect recovery of the market prices even if the the model perfectly captured the model dynamics.

3. **Type 2:**
   Type 2 model risk refers to the error induced by the necessary recalibration

of the model. This recalibration contradicts the model assumptions and becomes apparent when we move from one day to the next and consideration is given to two versions of history. The first version of history is that an existing model is recalibrated today and becomes a better description of future market dynamics. The second version of history is that an existing model is recalibrated today and becomes a worse description of future market dynamics than before. This second version of history is the Type 2 model risk.

4. **Type 3:**
   Finally, Type 3 model risk refers to the error which is due to the model not correctly describing the true empirical dynamics of the state variables.

For this project, we are interested in the model risk(s) that the model contradicts the market data in the sense of Type 1 risk, which we could think of as a model fit risk, and the Type 2 risk, which we could think of recalibration risk. 'Model risk' shall henceforth collectively refer to Type 1 model risk (model fit risk) and Type 2 model risk (recalibration risk)

In particular, our interest in model risk has three levels. The first regards how the Black Scholes model, the Heston stochastic volatility model, and the Derman-Kani local volatility compare in the context of model risk. Our interest in these particular models arises from their increasing ability to fit the volatility surface. We are also interested, for these models, in the optimal recalibration frequency which minimizes model risk. Lastly, we will also look at whether it is better or not to ignore data for options for which there are no trades or open interest from a model risk perspective.

The rest of the project proceeds as follows. Chapter 2 will give a substantive introduction into the model risk framework used in order to compare the abovementioned models from a model risk perspective. Chapter 3 will provide the details of how each of these models describe the evolution of a stock, and also give a brief explanation of how each of the models are calibrated. Chapter 4 and Chapter 5 contain the methodology and the results respectively, and Chapter 6 concludes the work.

2 Model Risk

In order to be able to compare models with regards to model risk, even those whose underlying probability laws may differ, we need to have a function which captures the notion of distance between density functions. The work done by Glasserman and Xu (2014) exemplifies how the relative entropy framework can be used to capture this notion of distance between density functions. A brief discussion of the relative entropy framework, the way in which Glasserman and Xu employ it, and the ideas we take from their approach follows.

Let f be the density relative to which all density distances are measured. The relative entropy of $\tilde{f}$ with respect to f is

$$\mathcal{R}(f, \tilde{f}) = E[m \log m] = \int \frac{\tilde{f}(x)}{f(x)} \log \frac{\tilde{f}(x)}{f(x)} f(x) dx,$$

where $m = \frac{\tilde{f}}{f}$ is a well-defined likelihood ratio. While $\mathcal{R}$ is not a metric by definition (it is not symmetric, for instance), it has some desirable properties which are part of the definition of a metric. In particular, $\mathcal{R}(f, \tilde{f}) = 0$ if and only if $f = \tilde{f}$ almost everywhere. Furthermore, $\mathcal{R} \geq 0$, since by Jensen's inequality, $E[m \log m] \geq E[m] \log E[m]$, where $m = \frac{\tilde{f}}{f}$ as before. Because m is a Radon-Nikodym derivative, $E[m] = 1$, and thus we conclude

$$E[m \log m] \geq E[m] \log E[m] = 0.$$

At this point, it is natural to ask the question: why do we use a measure which is not a metric to obtain a notion of distances between densities when one could easily use a metric? The advantage of the relative entropy framework is that it allows practitioners and researchers to measure risk in a way which is robust to model error (see Glasserman and Xu (2014)).

In Glasserman and Xu's work, their main objective was to provide a tool to make risk measurement robust to model error. This is where they needed a notion of distance of an arbitrary model relative to some nominal model. In particular, they were able to make risk measurement robust to the choice of model, given that the model had a relative entropy measure within some reasonable bound.

In our use of relative entropy, given a density $\tilde{f}$, which represents a proposed model for the underlying stock prices, and a nominal density f, which represents a base model for stock prices, $\mathcal{R}(f, \tilde{f})$ quantifies the level of model risk inherent in using $\tilde{f}$, relative to the risk of using f. Unlike Glasserman and Xu, our objective is to consider the model risk (type 1 and type 2) across different models and option maturities in the manner in which we described at the end of the introduction. In order to achieve this task, we need a justifiable choice of f which will play the role of a reference distribution.

One of the most preferable characteristics the reference distribution should possess is that it should make the least assumptions about the market prices. In practice, this implies sourcing information mainly from the relevant exchange. The suggestions on how to approach this problem have a fair amount of variation in the literature. Dupire (1992) extracted a model from the market for stock prices using the corresponding option prices and a model for stochastic volatility; Rubinstein (1994) extended this by also using stock price information. Buchen and Kelly (1996) used only the observed option prices information to infer the reference or "market" distribution. As we can see, the approach taken by Buchen and Kelly uses the least information in comparison to the other approaches. Following Feng and Schlögl (2016), we use the same method to obtain the reference distribution, f. This method relies on the Principle of Maximum Entropy (PME).

Let X be the underlying stock or index price at a future time T. We want to obtain $f(x)$, the density function for X. To use the PME, we need the entropy of $f(x)$, which is defined as

$$S(f) = \int_0^\infty f(x) \log f(x) dx.$$

$S(f)$ is the objective function we need to maximise so that we obtain f, the reference distribution implied by the market. In order for f to be a well-defined density, we require that $f(x) \geq 0$ and $\int_0^\infty f(x)dx = 1$. In addition to these two constraints, the $S(f)$ is maximised subject to

$$E[c_i(X)] = \int_0^\infty f(x)c_i(x)dx = d_i$$

for $i = 1, 2, ..., n$, where $c_i(x)$ are the discounted payoff functions associated with d_i, the observed option prices. These options are written on the same underlying

but have varying strikes. Formulated in this way, the task of finding f, the market implied reference distribution, reduces to an optimisation problem.

The resulting density f which is the solution to the optimisation problem is the reference distribution relative to which all density distances (or equivalenty, model risk) will be measured.

3 Models

As mentioned already in the Chapter 1, our model risk analysis will be centered around the Black-Scholes model, the Heston stochastic volatility model and the Derman-Kani model. In this section, we will give a brief description of how each of these models price European options on non-dividend paying stock. Thereafter, we will describe how each of these models were calibrated to market data.

3.1 Pricing in the Black-Scholes and Heston Model

3.1.1 Black-Scholes Model

The Black-Scholes (BS) model for European options on non-dividend paying stocks assumes that the stock price processes is described by Geometric Brownian Motion (GBM). The BS risk neutral dynamics of a stock price process in $(\Omega, \mathcal{F}, \mathbb{Q})$ are given by:

$$dS_t = rS_t dt + \sigma S_t dW_t \tag{1}$$

where r is the risk-free rate of interest, $\sigma > 0$ the constant volatility and W_t a $\mathbb{Q}$ Brownian motion. The corresponding BS price for a European option at time t on a non-dividend paying stock S_t with strike K and time-to-maturity τ is given by:

$$S_t N(d_1) - K \exp(-r\tau) N(d_2). \tag{2}$$

Here,

$$d_1 = \frac{\log(S_t/K) + (r + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}}, \qquad d_2 = d_1 - \sigma\sqrt{\tau}$$

and N is the cumulative distribution of the standard normal distribution.

3.1.2 Heston Model

In a risk-neutral setting $(\Omega, \mathcal{F}, \mathbb{Q})$, the Heston model for stock prices is given by

$$dS_t = rS_t dt + \sqrt{\lambda_t} S_t dW_t \qquad d\lambda_t = \kappa(\theta - \lambda_t)dt + \sigma\sqrt{\lambda_t} dB_t, \tag{3}$$

where r is the riskless rate, $\lambda(0) = \lambda_0 \geq 0$ the instantaneous variance, κ the mean reversion speed, θ the mean reversion level, and $\sigma > 0$ the non-stochastic volatility of the variance process. Note that W_t and B_t are $\mathbb{Q}-$ dependent Brownian motions.

Following Gschnaidtner and Escobar (2015), the price of a European call option at time t with strike price K and time-to-maturity τ, when stock price dynamics are given by the Heston model, is given by

$$C(S_t, K, \tau) = \frac{1}{\pi K^\alpha} \int_0^\infty \mathcal{R}e(G(v))dv$$

where

$$\begin{aligned}
G(v) &= \exp(-iv \ln K) \frac{\exp(-r\tau)\Phi_H(v - i(\alpha+1), \tau)}{\alpha^2 + \alpha - v^2 + i(2\alpha+1)v} \\
C(\tau, \phi) &= r\phi i\tau + \frac{\kappa\theta}{\sigma^2}\left[(\kappa - \rho\sigma\phi i - d)\tau - 2\ln\left(\frac{1 - g\exp^{-d\tau}}{1-g}\right)\right] \\
D(\tau, \phi) &= \frac{\kappa - \rho\sigma\phi i - d}{\sigma^2} \frac{1 - \exp(-dr)}{1 - gexp(-dr)} \\
d &= \sqrt{(\rho\sigma\phi i - \kappa)^2 + \sigma^2(\phi i + \phi^2)}
\end{aligned}$$

and $\Phi_H(\phi, \tau) = \exp(C(\tau, \phi) + D(\tau, \phi)\lambda_0 + i\phi x_0)$ is the characteristic function of the log price process associated with the Heston model.

3.2 Calibrating the Black-Scholes and Heston Model

The model calibration problem involves solving a non-linear constrained optimisation problem. This is true for both the Black-Scholes model and the Heston model.

3.2.1 Black-Scholes Model

In order to calibrate the Black-Scholes model, one needs to take as inputs S_0, $[c_i]$, $[K_i]$, $[\tau_i]$,$[r_i]$ where c_i is the market price of the option. Next, one solves for the constant volatility σ_i implied by the market prices that causes the Black-Scholes price

(in (2)) and the market price c_i to be equal. This is done by changing the σ_i until the the two prices agree. The parameter σ_i is subject to the constraint that $\sigma_i > 0$.

Once the implied volatilities, σ_i, are obtained for each instrument, the Black-Scholes model is calibrated by solving for the σ_{BS} that minimises the Relative Square Volatility Error (RSVE). In order to do this, the following expression is minimised :

$$\sum_{i=1} \left(\frac{\sigma_{\text{BS}} - \sigma_i}{\sigma_i}\right)^2. \tag{4}$$

RSVE is the chosen risk measure to minimise because Gschnaidtner and Escobar (2015) find that minimising the RSVE results in the most accurate recovered parameters for the Heston model. Since RSVE is used to calibrate the Heston model, it is used to calibrate the Black-Scholes model for consistency.

To ensure that the calibration is returning accurate model parameters (Θ), the calibration is conducted on generated data. Black-Scholes call prices are generated for a fixed volatility over a variety of strikes and terms to maturity. The implied volatilities (σ_i) for the Black-Scholes call prices are calculated and are in turn used to calculate the Θ that minimises the Relative Square Volatility Error (RSVE). The resulting Θ is compared to the volatility used to generate the Black-Scholes call prices. The σ_i and Θ are tightly packed around the input volatility, this is illustrated in Figure 1a.

3.2.2 Heston

The Heston model generates call prices by taking the following input model parameters: $\lambda_0, \kappa, \theta, \sigma,$ and ρ as well as market parameters S_0, r_i, τ_i and K_i. The input model parameters can be interpreted as follows: λ_0 is the instantaneous volatility of the stock price, κ is the mean reversion speed, θ is the mean reversion level, σ is the constant volatility of the of the variance process and ρ is the correlation between the Brownian motions that drive the variance process and the stock price process.

Thus, in order to calibrate the Heston model for a given trade date, it is necessary to solve for the model parameters. The first step in the calibration is to calculate the

constant Black-Scholes volatility implied by each option, σ_i. Next, given an initial parameter set Θ_{Heston} that is a vector that contains $\lambda_0, \kappa, \theta, \sigma,$ and ρ, call prices are generated for the (fixed) market parameters S_0, r_i, τ_i and K_i. The constant Black-Scholes volatility for each instrument ($\sigma_{H,i}$) is implied from the Heston-generated call prices. One solves for the model parameter set Θ_{Heston} by again minimising the following expression:

$$\sum_{i=1}(\frac{\sigma_{H,i} - \sigma_i}{\sigma_i})^2. \tag{5}$$

In (5), $\sigma_{H,i}$ is the Black-Scholes volatility implied by the Heston price for a set of initial model parameters Θ_{Heston} and the i^{th} set of market data S_0, r_i, τ_i and K_i.

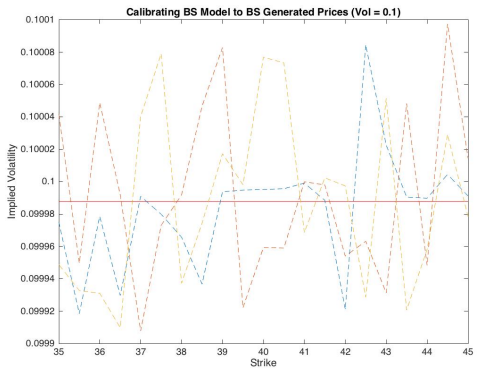
To do this, the MATLAB function *lsqnonlin* is used with the following specifications, in line with the Gschnaidtner and Escobar (2015) paper:

TolFun:	1e-10
TolX:	1e-5
MaxFunEvals:	10 000
MaxIter:	50 000

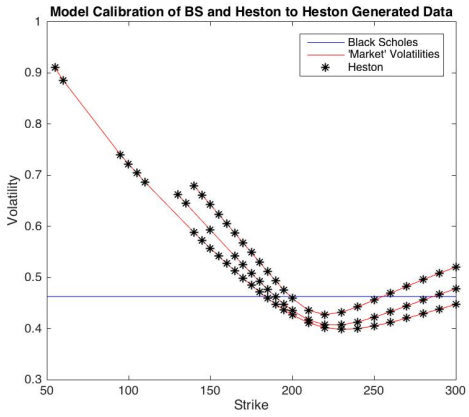
Furthermore, the following lower bound (l) and upper bound (u) constraints are placed on the elements of Θ_{Heston}, again in line with the constraints in Gschnaidtner and Escobar (2015):

l = (1e-5, 1e-5, 1e-5, 1e-5, -0.99)
u = (3, 20, 3, 3, 0.99)

To ensure that the calibration is returning accurate model parameters (Θ_{Heston}), the calibration is conducted on generated data. Heston call prices are generated for a given initial vector $\Theta_{Initial}$ for a variety of strikes and maturities. The implied volatilities ($\sigma_{H,i}$) for the Heston call prices are calculated and are in turn used to calculate the Θ_{Heston} that minimises the Relative Square Volatility Error (RSVE). The resulting Θ_{Heston} is compared to the set of initial parameters, $\Theta_{Initial}$. The Heston calibration is able to reproduce, exactly, the volatility smile implied by the initial parameters, $\Theta_{Initial}$. Furthermore, the Heston calibration is able to recover



(a) Error when calibrating to the Black-Scholes-generated 'market' volatilities



(b) Heston-generated 'market' volatilities

Figure 1: Black-Scholes and Heston volatility smiles compared to generated smiles

(to 7 decimal points) the initial parameters. The volatility 'smiles' resulting from the Heston and Black-Scholes calibration to Heston-generated data is illustrated in Figure 1b.

3.3 Derman-Kani Model

The pricing and calibration treatment of the Derman-Kani model is sufficiently different from the other two models that we should deal with it separately. The Derman-Kani model is a model for stock prices which extends the Black-Scholes model by allowing volatility to be a function of time, t and the stock price S_t at time t. However, instead of specifying a functional form for the volatility $\sigma(t, S_t)$, the Derman-Kani model uses option price data on the underlying stock to infer $\sigma(t, S_t)$. This model prices instruments in the context of a risk neutral binomial market-implied tree framework. A description of how the tree is constructed follows.

Suppose that the maturities of the options used to imply the tree are uniformly spaced Δt time units apart. Also suppose that we have thus far constructed the implied tree out to the n^{th} level or maturity, and that we wish to extend the tree to the $(n+1)^{th}$ level. Carrying out this task involves solving for n up-move probabilities at each of the n nodes at the n^{th} level; however, if we know the up-move

probabilities for each of the n nodes at the n^{th} level of the tree, we also have the corresponding down-move probabilities and there are thus only n independent unknowns. We also need to solve for the $(n+1)$ unknown prices at each of the nodes at the $(n+1)^{th}$ level of tree. If we have these two sets of information, we would have succeeded in extending the tree to the $(n+1)^{th}$ level.

As we can see from above, in order to obtain these two sets of information, we need to solve for $2n+1$ parameters. Since the implied binomial tree is risk neutral, the price of a forward, F_i, at the tree level n and node i is

$$F_i = p_i S_{i+1} + (1-p_i) S_i,$$

where S_{i+1} (S_i) is the *unknown* stock price at the tree level $n+1$ which results due to an up-move (down-move) from s_i, the *known* stock price at the tree level n. The probability of an up move from the tree level n and node i is p_i (we will use (n,i) to represent the level of the tree, and the node number on that level respectively. For example, $(5,4)$ represents the 4^{th} node from the bottom of the tree on the 5^{th} level or maturity). Supposing we also have r, a continuously compounded forward riskless interest rate at the n^{th} level, we obtain that $F_i = e^{r\Delta t} S_i$, for $i = 1, 2, ..., n$. Therefore, our first n equations are

$$e^{r\Delta t} S_i = p_i S_{i+1} + (1-p_i) S_i$$

for $i = 1, 2, ..., n$.

To be able to solve for the remaining $n+1$ unknown parameters, we first use n option prices (calls and puts). Let $C(s_i, t_{n+1})$ and $P(s_i, t_{n+1})$ respectively be the price of a call and put option which expire at tree level $n+1$ and struck at s_i, the price at the (n,i) position the tree. It is important to note that these options are priced at time 0, and that the inductive procedure to generate the tree implies together with the option price data the strikes for which the options (calls and puts) are priced exactly. Using the risk neutral binomial tree framework, we can price both call and put options using discounted expectation. For example, the price of a call option struck at K with maturity t_{n+1} is given by

$$C(K, t_{n+1}) = e^{-r\Delta} \sum_{j=1}^{n} \{\lambda_j p_j + \lambda_{j+1}(1-p_{j+1})\} \max(S_{j+1} - K, 0),$$

where the λ_j are the Arrow-Debreu prices at (n, j). There is a similar equation for pricing put options. The value for s_i informs us to obtain the market price for an option whose strike corresponds to s_i for $i = 1, ..., n$ and mature at time t_{n+1}. These prices make give us the expressions for $C(K, t_{n+1})$ and $P(K, t_{n+1})$ which make up the left hand side of the equations such as the one above. There are n such equations, each corresponding to the $K = s_i$ for $i = 1, 2, ..., n$. At this point we have $2n$ equations, but $2n + 1$ unknowns.

With the remaining degree of freedom, we fix one of the center nodes in order to ensure the Cox-Ross-Rubinstein (CRR) centering condition. Suppose that we are on the n^{th} level on the tree and we are interested in populating the $(n + 1)^{th}$ level of the tree. If the number of nodes at the $(n + 1)^{th}$ level of the tree is odd, put the center of the tree at that level to be the stock price today. On the other hand, if the number of nodes at the $(n + 1)^{th}$ level of the tree is even, the logarithmic CRR centering condition requires that we set the average of the natural logarithm of the innermost nodes (at the $(n + 1)^{th}$ level of the tree) to be equal to the natural logarithm of today's stock price.

The formulae to help us to extend the tree to the $(n + 1)^{th}$ level follow. The suggestion above is that the first part of the tree to populate at the $(n+1)^{th}$ level is the center. If $n + 1$ is odd, we put the center of the tree to be the spot price; if $n + 1$ is even, we put the upper of the central nodes to be

$$S_{i+1} = \frac{S[e^{r\Delta t}C(S, t_{n+1}) + \lambda_i S - \Sigma]}{\lambda_i F_i - e^{r\Delta t}C(S, t_{n+1}) + \Sigma}$$

where S is today's stock price and $\Sigma = \sum_{j=i+1}^{n} \lambda_j (s_j - F_j)$. From the centering condition, we can obtain S_i, the lower of the central nodes.

Once we have populated the central nodes, we can populate the nodes above the central nodes using the formula

$$S_{i+1} = \frac{S_i[e^{r\Delta t}C(s_i, t_{n+1}) - \Sigma] - \lambda_i s_i(F_i - S_i)}{[e^{r\Delta t}C(s_i, t_{n+1}) - \Sigma] - \lambda_i(F_i - S_i)},$$

where $\Sigma = \sum_{j=i+1}^{n} \lambda_j (s_j - F_j)$, and the lower nodes can be populated using the formula

$$S_i = \frac{S_{i+1}[e^{r\Delta t}P(s_i, t_{n+1}) - \Sigma] - \lambda_i s_i(F_i - S_i)}{[e^{r\Delta t}P(s_i, t_{n+1}) - \Sigma] - \lambda_i(F_i - S_i)},$$

where $\Sigma = \sum_{j=1}^{i-1} \lambda_j (s_j - F_j)$. Finally, the up probabilities at (n, i) can be calculated by

$$p_i = \frac{F_i - S_i}{S_{i+1} - S_i},$$

for $i = 1, 2, .., n$.

The output of this procedure is a tree of prices and up-down probabilities for the stock price process at the maturities $\Delta t, 2\Delta t, ..., T\Delta t$, where T is the number of periods in our tree. With these two trees, we can price options on the relevant stock.

Recalibrating the Derman-Kani model to the market involves reconstructing the tree to match the updated market information.

4 Methodology

This section details the methodology followed when processing data, estimating the MED and measuring model risk .

4.1 Data Processing

The data used contains information regarding options on stocks from both the NYSE and NASDAQ exchanges as well as interest rates observed in the market. The data relates to the period starting in January 2005 and ending in December 2008.

Initially, the data is in the form of an unmanageably large .csv file. Before processing the data it is necessary to split, using the Linux terminal, the single .csv file into multiple files that are smaller and thus more manageable.

The multiple .csv files are read into MATLAB, where they are filtered by: company of interest, call options and trading volume.

The data is filtered for trading volume so that implied volatilities inferred from option prices do not relate to stale prices. For this reason, only options where trades occur are considered i.e trading volume greater than zero.

Companies that did not declare/pay dividends for the year of 2005 are investigated. This is done because there is scant data relating to European options, for which analytical solutions exist. But since it is never optimal to exercise an American call option before maturity (provided the company pays no dividends), one may treat American call options as European call options. The companies chosen in the investigation are: Apple (AAPL) and Google (GOOG).

The data is filtered by option type in order to include only call options, for the reason mentioned above.

The resulting data tables are saved as company-specific .mat files relating to the option trade activity for 2005. The year 2005 is used because it is the first year of

available data.

In order to calibrate the market data to the models referred to in Section 3, it is necessary to process the .mat files into a form that suits the MATLAB calibration functions. The calibration functions take, as inputs, the stock price S_0 as a scalar, the implied market volatilities σ_i, the option strikes K_i, time to maturity τ_i and NACC interest rate r_i as column vectors. In order to extract market data for a given day, it is necessary to filter the data table for information on that specific day. Ultimately, the data is transformed from a company-specific table of option data for 2005 from a .mat file into:

$$S_0, [\sigma_i], [K_i], [\tau_i], [r_i]$$

These outputs are used to calibrate the three models under investigation.

4.2 Market Calibration

The process outlined in 3.2.1 is applied to market data. The calibrations do not succeed in every case. When the calibration process converges to an output (for either model), the model-implied volatilities fit the market volatility smile with varying degrees of accuracy. This is illustrated in Figure 3. In all (successful) cases, the constant Black Scholes-implied volatility is a horizontal line that intersects the market volatilities as well as the Heston-implied volatilies. The Heston-implied volatility smile captures the shape of the market smile more accurately than the Black Scholes 'smile'.

Upon examination of the parameters that result from calibrating the Heston parameters Θ_{Heston} to the observed market volatilities σ_i, it becomes clear that the individual parameters are not stable over the period of investigation. This is illustrated in Figure 3. It is clear that individual parameters oscillate between the upper and lower bounds described in Section 3.2.2. The observed instability may have arisen because each calibration of the Heston model begins from the same initial parameters. It is worth investigating whether the observed instability is reduced when the calibration takes the parameters, Θ_{Heston} from the most recent calibration

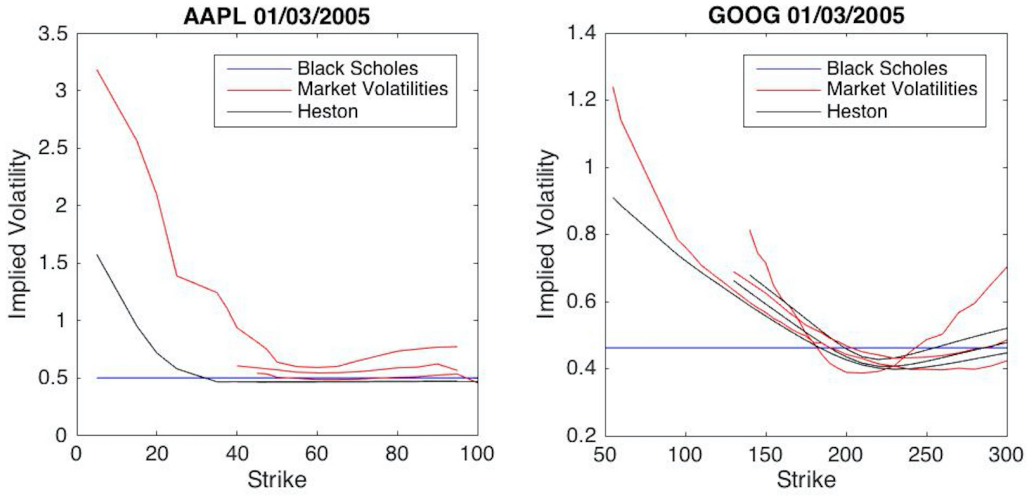


Figure 2: Volatility smiles implied by the Black Scholes and Heston calibrations

as the initial values for the new calibration. Alternatively, one could investigate the effect of using random initial values from which to begin the calibration.

Gschnaidtner and Escobar (2015) measure the difference between calibrated parameter sets and true initial parameter sets (for generated data) using the calibration error $\hat{\epsilon}$ which is defined as follows:

$$\hat{\epsilon} = \sum_{\psi \in \{\lambda_0, \kappa, \theta, \sigma, \rho\}} \hat{\epsilon_\psi} \tag{6}$$

where Gschnaidtner and Escobar (2015) define $\hat{\epsilon}_\psi$ as the sum of the absolute relative difference between elements of the known, true parameter set ψ^* and the calibrated parameter set $\hat{\psi}$, i.e.

$$\hat{\epsilon}_\psi = |\frac{\hat{\psi} - \psi^*}{\psi^*}| \tag{7}$$

The idea of using calibration error to assess the stability of Heston parameters is evaluated. It is decided not to use the sum of the absolute relative difference between calibration parameters calculated on consecutive days. This decision is made because individual Heston parameters frequently reach their lower bounds. Instead, the level of individual Heston parameters is observed over consecutive (daily) calibrations. This is illustrated in Figure 3. From Figure 3, it is clear that

parameters move between their respective upper and lower bounds, described in Section 3.2.2. Thus the (recalibrated) Heston parameters do not appear stable over time.

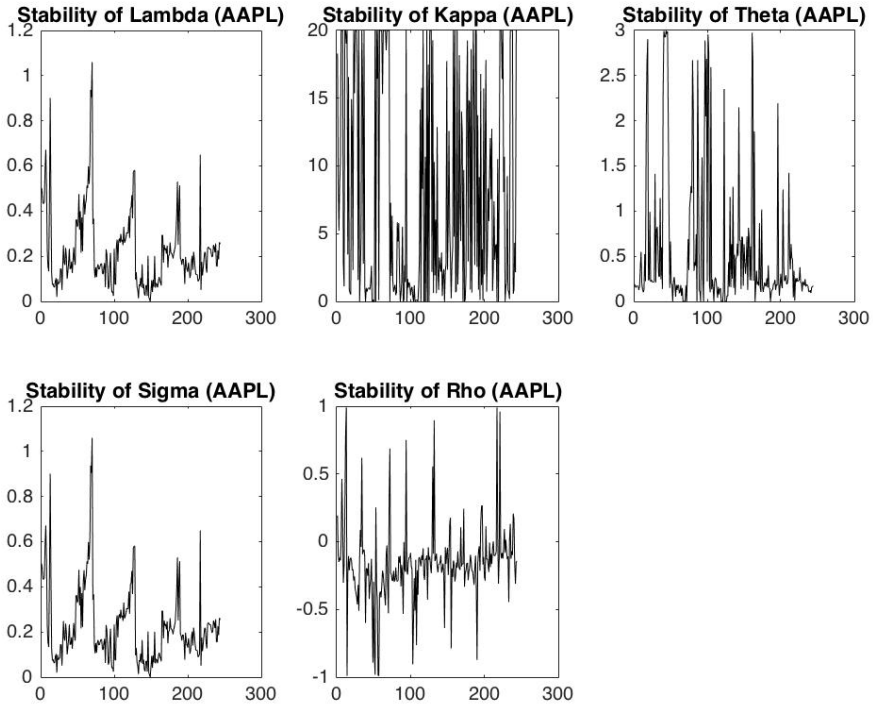


Figure 3: Evolution of individual Heston parameters

4.3 Implementing the Derman Kani Model

One of the main challenges in implementing the DK model using the recursive algorithm outlined in Chapter 3 is that the algorithm often produces negative path probabilities, which consequently allows arbitrage opportunities to exist. This problem is solved in various ways in practice. Derman and Kani (1994) suggest that if the stock price created on a node gives a probability that is negative or greater than one, then that price is ignored and a new price is calculated such that the logarithmic difference between adjacent nodes is equal to the previous level.

First consider the upper level of the tree. Note that a negative probability (or arbitrage opportunity) will be realised if and only if $S_{i+1} < F_i$ or $S_{i+1} > F_{i+1}$. Keeping the logarithmic distance between adjacent nodes the same as the previous level, we have the following equation:

$$\ln(S_{i+1}) - \ln(S_i) = \ln s_{i+1} - \ln s_i, \text{ for } i < n,$$

which implies that

$$S_{i+1} = S_i \frac{s_{i+1}}{s_i} \text{ for } i < n.$$

We also set

$$S_{n+1} = S_n \frac{s_n}{s_{n-1}}.$$

Using the same type of reasoning one obtains, for the lower part of the tree,

$$S_i = \begin{cases} S_{i+1} \frac{s_i}{s_{i+1}} \text{ if } i > 0 \\ S_1 \frac{s_0}{s_1} \text{ if } i = 0. \end{cases}$$

One drawback with this correction is that the model will no longer have an exact fit on the data it is calibrated to.

Other adjustments to this have been made by Barle and Cakici (1995). The two main adjustments are on determining the centre of the tree and choosing the strike price of the options used in populating the tree. Instead of using the current stock price as the centre for the tree, they propose compounding the current stock price by the risk-free rate over one period. They also propose using forward prices as the strike prices for the options used instead of the stock prices themselves. Brandt and Wu (2002) extend the model to deal with discrete dividends and a term structure of interest rates. Derman et al. (1996) have extended the model to construct an *Implied Trinomial Tree* to introduce more flexibility in the parametrisation of the tree.

4.4 Measuring Model Risk

We will measure model risk using a modification of the relative entropy framework outlined in Glasserman and Xu (2014). This method has the advantage that it can be applied universally to all models, and thus allows us to come up with a single number to measure the risk of a model in a way that is universal across models and all types of model risk.

4.4.1 Maximum Entropy Distributions

We will work in the following setting. Assume that on any given day $t \geq 0$ we have available to us the prices of finitely many call options $\{C(t, S_t, K_i, T) : i = 1, 2, \ldots, n\}$ written on different strikes $\{K_i : i = 1, 2, \ldots, n\}$ and maturity T. We assume that all these options are written on the same underlying asset $S = \{S_t : t \geq 0\}$.

In the theory of risk neutral pricing, the price of each option is the expected value (ignoring the discount factor), under the risk neutral distribution, of the discounted value of the option at maturity S_T. That is,

$$C(0, S_0, K_i, T) = \mathbb{E}(C(T, S_T, K_i, T)) = \mathbb{E}(\max(S_T - K_i, 0)), \tag{8}$$

where the expectation is taken under the risk neutral measure. To evaluate this expectation, we need the risk neutral distribution of S_T, the terminal stock price.

Since we have the option prices available, any sensible risk neutral distribution for S_T (model) that we choose should at least return back the option prices when formula (8) is applied to it. This requirement has implications on the model type and the number of parameters the model can have. Let us first consider distributions of S_T that satisfy (8). That is, we consider densities f such that

$$\int_0^\infty \max(S - K_i, 0) f(S)\, dS = C(0, S_0, K_i, T), \text{ for } i = 1, 2, \ldots, n. \tag{9}$$

The principle of maximum entropy states that in order to use no other prior information (than the already observed option prices), one should pick a density f_M that maximises the entropy I defined by

$$I(f) = -\int_0^\infty f(x) \log f(x)\, dx$$

for suitable densities f. This is equivalent to starting with a non-informative prior distribution for S_T, and constraining it to satisfy (9). We will call this distribution the *Market Implied Maximum Entropy Distribution* (MIMED). We will use this distribution as a basis for comparison of all other distributions of S_T.

Now let $g(\cdot;\theta)$ be a model density for S_T, that depends on a vector of parameters θ. Depending on the form and number of parameters θ, it may not be possible to choose θ such that

$$\int_0^\infty \max(S - K_i, 0) g(S;\theta)\, dS = C(0, S_0, K_i, T), \text{ for } i = 1, 2, \ldots, n. \tag{10}$$

In such a case, we choose θ to minimise some deviation between the model prices and the market prices. This imperfect fit will introduce an error, which we call a *Type 1 Error*. Quantifying this error consistently across different model classes is challenging due to the different parameter sets in each model and the difficulty in choosing a metric that is comparable across models.

The work of Glasserman and Xu (2014) give insight into a way of approaching this by introducing a pre-metric to compare distributions. Given a reference distribution f and another density g with the same support as f, one can define the *Relative Entropy* of g with respect to f as

$$\mathcal{R}(f, g) = \mathbb{E}(m \log m),$$

where $m = f/g$ is the likelihood ratio.
We will use $\mathcal{R}$ to measure this type of error as follows:

1. For the calibrated set of parameters θ_0, we generate the option prices that are implied by the model $g(\cdot, \theta_0)$ by the formula

   $$\tilde{C}_i := \int_0^\infty \max(S - K_i, 0) g(S, \theta_0)\, dS.$$

   These prices will in general not be equal to the market prices, unless the model perfectly fits the market data.

2. We can then find the Maximum Entropy Distribution implied by these model prices, g_M.

3. The Type 1 Error is then defined as the relative entropy between the MIMED f_M and g_M. That is

   $$\text{Type 1 Error} := \mathcal{R}(g_M(\cdot, \theta_0), f_M)$$

The main advantage of comparing $g_M(\cdot, \theta_0)$ to f_M, rather than directly comparing $g(\cdot, \theta_0)$ to f_M is that this gives a fairer comparison of the two distributions based on the option prices that does not take into account any further information contained in the model implied distribution. Also, it is computationally easier to compute the density of g_M than the density of $g(\cdot, \theta_0)$ directly for many models in finance.

As mentioned previously, the error in choosing a particular model of S_T comes from more than just Type 1 error. We are particularly concerned with the risk that the chosen model parameters, though being the best fit on one day, might not be suitable on another day. That is, upon recalibrating on another day, the set of parameters might need to change in response to the changes in market variables. We call this Type 2 Error and measure it as the relative entropy between the model implied MEDs with the two parameter sets:

$$\text{Type 2 Error} := \mathcal{R}(g_M(\cdot, \theta_0), g_M(\cdot, \theta_1)).$$

It is important to note that the parameter sets above must *exclude* the state variables, which are allowed to change. Thus for the Heston model, even though we obtain λ_0 from a calibration procedure, it is a state variable and is thus allowed to change without influencing the recalibration risk[1].

4.5 The MED Distribution

The Maximum Entropy Distribution (MED) contains only the information about an asset that can be inferred from option price data alone. It is the least committal distribution with respect to unknown or missing information. In kind, it accepts no "view" of the market dynamics or shape of the risk-neutral distribution.
The MED minimises the entropic "distance" to the uniform distribution under the constraint that it must correctly reproduce the market-observed option prices.
For a density $p(x)$ the entropy is given by

$$I(p) = -\int_0^\infty p(x) \log p(x)\, dx.$$

The considered constraints can be written as

$$\mathbb{E}[\phi_i(X)] = C_i, \quad i = 1, \ldots, M$$

[1]The fact that the thus obtained evolution of λ might not be consistent with the model dynamics assumed for λ (Type 3 model risk) is not the subject of the present analysis.

where $\phi_i(x)$ is the discounted payoff of the option and C_i is the observable option price. Following Buchen and Kelly (1996), this constrained optimisation problem leads to the following explicit representation of the density for the MED

$$p(x) = \exp(\sum_{i=1}^{M} \lambda_i \phi_i(x)) \tag{11}$$

where $\phi_0(x) = 1$ and $C_0 = 0$ enforce the probability constraint that the density must integrate to one and the λ_i's are the Lagrangian multipliers that need to be solved by the optimisation.

The maximum entropy obtained by the density described in Equation (11) is given by

$$F(\lambda_0, \ldots, \lambda_N) = -\sum_{i=0}^{M} \lambda_i C_i.$$

Defining $f_i(\lambda_0, \ldots, \lambda_N) = \mathbb{E}[\phi_i(X)] - C_i$, a simple calculation will show that $f_i = \nabla_i F$. Thus F reaches a minimum when $f = 0$, which simultaneously satisfies all the constraints.

Newton-Raphson is a standard numerical method that can be applied to solving $\nabla F = 0$, however it requires the Jacobian of F. In the appendix of Buchen and Kelly (1996) it is shown that the elements of the Jacobian are given by

$$J_{ij}(\lambda_0, \ldots, \lambda_M) = \text{cov}(\phi_i(X), \phi_j(X)).$$

The system to be solved via Newton-Raphson can then be written

$$\Lambda_{l+1} = \Lambda_l + \alpha J^{-1} \nabla F,$$

where l is the iteration index and $[\Lambda]_i = \lambda_i$.

To avoid computing the integrals inside the Jacobian numerically, some analysis can be done for the special case of only using European call option prices as constraints.

Firstly, the normalisation constant, λ_0, can be expressed as a function of the other

Lagrangian multipliers,

$$\begin{aligned}\exp(-\lambda_0) &= h(\Lambda) \\ &= \int_0^\infty \exp(\sum_{i=1}^{M} \lambda_i (x-k_i)^+)\, dx \\ &= \sum_{i=1}^{M} \frac{1}{a_i}\left(e^{a_i k_{i+1} - b_i} - e^{a_i k_i - b_i}\right) + k_1,\end{aligned}$$

where k_i is the sequence of market strikes for the considered call options with $k_{M+1} = \infty$ and

$$a_j = \sum_{i=1}^{j} \lambda_i$$

and

$$b_j = \sum_{i=1}^{j} \lambda_i k_i.$$

Noting that $\frac{\partial h(\Lambda)}{\partial \lambda_j} = C_j h(\Lambda)$ allows for the partial derivative to be taken directly to express the integral constraint in a closed form. This results in

$$\frac{\partial h(\Lambda)}{\partial \lambda_j} = \sum_{i=j}^{M} \left[\frac{(k_{i+1} - k_j)e^{a_i k_{i+1} - b_i} - (k_i - k_j)e^{a_i k_i - b_i}}{a_i} - \frac{e^{a_i k_{i+1} - b_i} - e^{a_i k_i - b_i}}{a_i^2} \right].$$

The elements of the Jacobian can now be determined by taking the partial derivative with respect to each λ_i again.

For numerical stability, good initial guesses for the Lagrangian multipliers are a necessity. For closely placed strikes, it is expected that the MED will closely resemble the implied risk-neutral distribution. The Breeden and Litzenberger (1978) equations can be used to approximate the risk-neutral density at each interior strike,

$$p_{BL}(k_i) = \frac{D}{k_{i+1} - k_{i-1}} \left[\frac{C_{i+1} - C_i}{k_{i+1} - k_i} - \frac{C_i - C_{i-1}}{k_i - k_{-1}} \right], \tag{12}$$

where D is the discount factor corresponding to the maturity we are considering. Noting that

$$\log p(x) \propto \sum_{i=1}^{M} \lambda_i (x - k_i)^+,$$

allows for the approximation

$$\sum_{i=1}^{j} \lambda_i (k_j - k_i) = \log p_{BL}(k_j).$$

Taking differences leads to an initial guess for each λ_i on the interior nodes,

$$\lambda_i = \frac{\log\left(\frac{p_{BL}(k_{i+1})}{p_{BL}(k_i)}\right)}{k_{i+1} - k_i} - \frac{\log\left(\frac{p_{BL}(k_i)}{p_{BL}(k_{i-1})}\right)}{k_i - k_{i-1}}.$$

This leads to a starting point for the numerical optimisation. Nonetheless, it is easy for the Jacobian matrix to become poorly conditioned.

Please note that although the methodology in this section appears correct to all consideration, the researchers could not stabilise the numerical algorithm in the time required. Finally, an adapted version of the algorithm provided by Mohammad-Djafari (1992) was used. This would need to be further investigated for future work on this topic.

Using the modified algorithm, determining MED distributions from model-implied option prices became very stable. Nonetheless, implying MED distributions from market option data remained problematic.

4.6 Filtering Market Data

As laid out in the problem statement, it is necessary to investigate filtering the market option prices to both remove the noise, most notably stale option prices, and to ensure that the fitting procedure for the MED proceeds successfully.

The first layer of filtering was simply on volume - no option prices were considered that traded less than 10% of the mean volume for a given maturity on a given day. This is sufficient for most cases, except as options become close to expiry. Then options that are deep in or out of the money are suddenly traded in large volumes as traders exit their positions. This leads to a very skewed volatility smile that the considered models cannot attempt to match and does not imply a proper density for the MED.

Although considerable work can be done to determine the optimal filters for noise versus signal in the market, an alternative approach was taken in this work.

First, data is filtered to only keep options trading within 20% of moneyness. Then these options are filtered on volume; they must trade at least 10% of the mean

Figure 4: The market implied MED and the calibrated Heston MED, using only market prices.

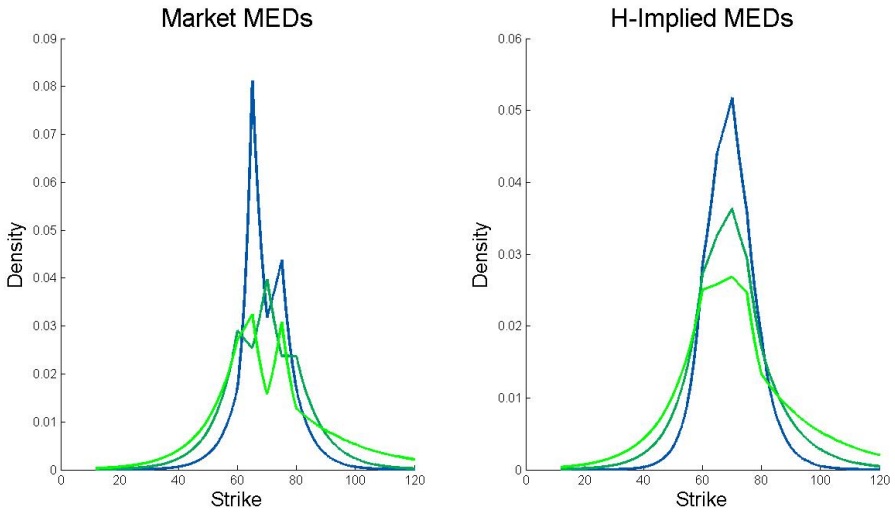
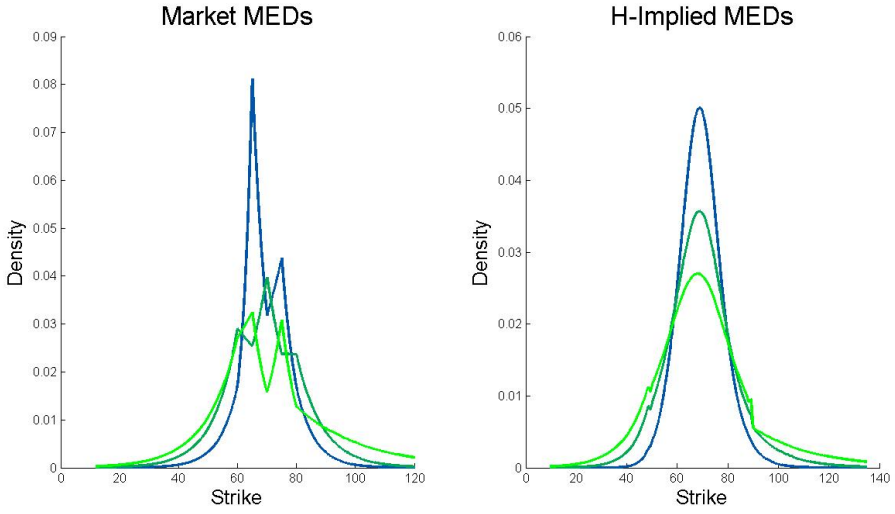


Figure 5: The market implied MED and density implied by the calibrated Heston, approximated with finely spaced strikes.



volume for the day. Finally, an MED fitting algorithm is run on the remaining smiles.

If the MED fitting fails, either owing to numerical instability or because the re-

sulting density is not a true density (does not integrate to one) that smile is not considered for calibration or entropy calculation purposes.
Using this procedure, over 90% of the daily data over the course of the year considered could be used for entropy calculations.
In Figure 4 the MED fitted to the market is shown next to the MED implied by the calibrated Heston model created by only repricing (with Type I error) the market prices available. In Figure 5 the market MED is again shown, but here the full implied distribution for the calibrated Heston is plotted - this is approximated by the MED using very finely spaced strikes.

5 Results

5.1 Optimal Calibration Frequency

We begin this final section with an overview of the completed work:

- From Glasserman and Xu (2014) we adopt the method of relative entropy as a measure of distance between distributions. For a given relative entropy bound the framework of Glasserman and Xu (2014) can be used to produce worst-case bounds for any functional risk measure.
- From Buchen and Kelly (1996) we adopt the Maximum Entropy Distribution as the distribution implied by the market with no information beyond the visible option prices.
- Using the error classification from Schlögl (2015), we can measure the aggregate error owing to model risk for different models and different calibration frequencies.

The above by itself is still insufficient to make decisions regarding model calibration. What is required is the relative entropy *budget* implied by the choice of model and calibration frequency. Following Schlögl (2015), we create an empirical distribution for the entropy over a year.
The time-series of aggregate error for daily recalibration over a year is shown in Figure 6. The resulting histogram with the 95% quantile is plotted in Figure 7. This quantile tells us that, for a given model, and given calibration frequency, the

aggregate error owed to model risk (as measured using relative entropy) will be below this value 95% of the time.

Figure 6: Aggregate error for AAPL over 2005 with daily recalibration.

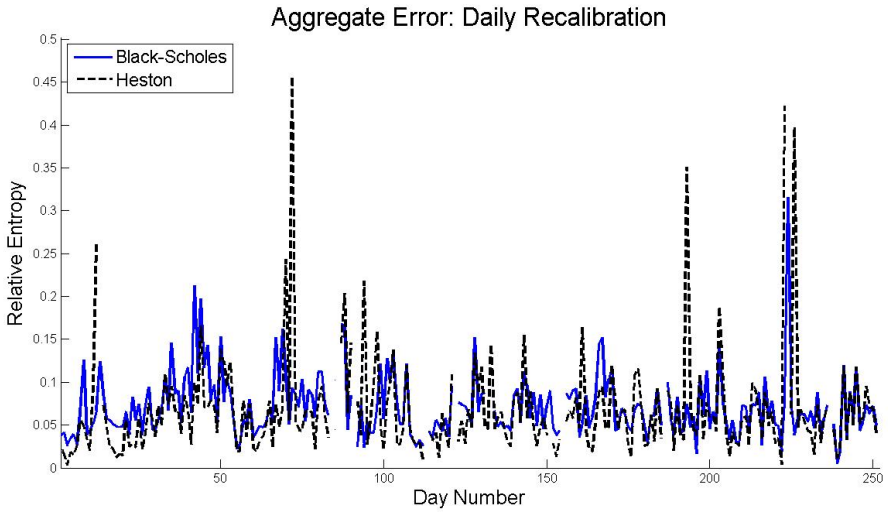
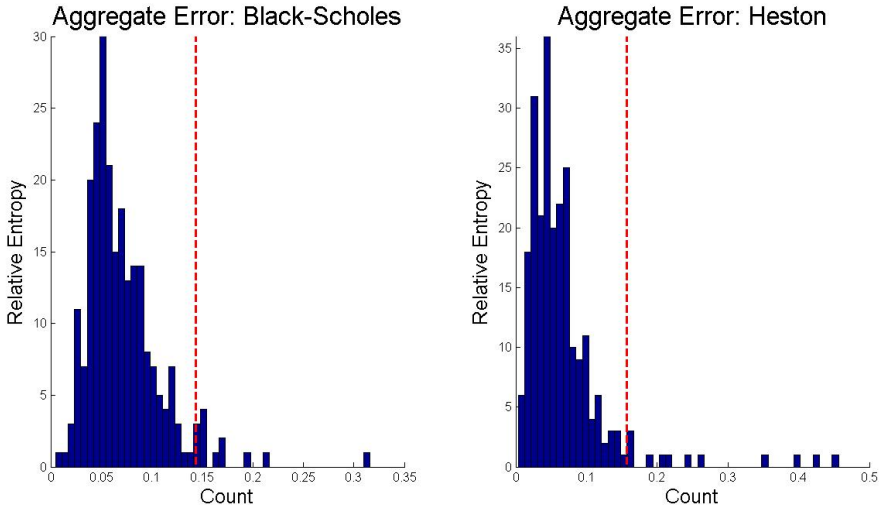


Figure 7: Histograms for relative entropy values over one year.



Thus, this number is the appropriate relative entropy bound at the 95% confidence level and can be used as a means of comparison across models and calibration frequencies. The results for GOOG over the year 2005 are summarised in Table 1.

Table 1: Relative entropy budget for GOOG over the 2005 trading year and at the 95% confidence level.

Frequency	**Model**	
	Black-Scholes	Heston
Annual	0.2334	0.1190
Semi-annual	0.2241	0.3571
Quarterly	0.1945	0.1686
Monthly	0.1685	0.1470
Weekly	0.1431	0.1820
Daily	0.1434	0.1572

Thus, Table 1 provides a unified number with which to compare model risk across models and calibration frequencies.
In our case, the Heston model does best when only calibrated once. We believe this is due to the parameter instability shown in the earlier section - this is exaggerating the Type II model error.
The semi-annual calibration frequency for the Heston model appears to be an outlier owing to an extreme jump in the parameters on the calibration day. For time-periods shorter than semi-annually the Heston model performs best when calibrated monthly.
This could indicate that the Heston model has a tendency to overfit the data and the lower calibration frequency leads to less noise being incorporated in the parameters.
The Black-Scholes model performs best when calibrated weekly. Daily calibration appears to be too frequent, again highlighting the influence of noise on the interplay between the Type I and Type II model error.

6 Conclusion

In this work we present a framework for measuring specifically the aggregation of Type I and Type II model risk across models and calibration frequencies.
This has two consequences:

- It provides a unified criteria for comparing models with regards to model risk

and calibration frequency. Given a specific model, it is possible to determine an optimal (from model risk perspective) calibration frequency.

- Once a model and calibration frequency have been selected, the corresponding relative entropy bound for a given confidence level can be used in the framework of Glasserman and Xu (2014) to determine worst-case bounds for a given risk measure.

Several improvements must be considered before applying this framework in the real world:

- A more stable MED estimation algorithm must considered.
- The aggregate error measure is heavily influenced by the quality of the model calibration. The Heston model requires a more robust calibration methodology than is considered here.

We would like to thank Prof Erik Schlögl for posing this problem and his mentorship and guidance throughout the FMTC.

Bibliography

Brandt, M. W. and Wu, T. (2002). Cross-sectional tests of deterministic volatility functions. *Journal of Empirical Finance*, 9(5):525–550.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

Buchen, P. W. and Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(01):143–159.

Derman, E. and Kani, I. (1994). The volatilities smile and its implied tree. *Goldman Sachs Quantitative Strategies Research Notes*.

Derman, E., Kani, I., and Chriss, N. (1996). Implied trinomial tress of the volatility smile. *The Journal of Derivatives*, 3(4):7–22.

Feng, Y. and Schlögl, E. (2016). Market Calibration of Risk-Neutral Measures and Model Risk. Work in progress.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.

Gschnaidtner, C. and Escobar, M. (2015). Parameters recovery via calibration in the heston model. a comprehensive review.

Mohammad-Djafari, A. (1992). A matlab program to calculate the maximum entropy distributions. In *Maximum Entropy and Bayesian Methods*, pages 221–233. Springer.

Schlögl, E. (2015). Toward quantifying model risk. QFRC, University of Technology Sydney.

Credit Risk in Stock-Based Lending

Team 5

ALEX BACKWELL, University of Cape Town
NICOLE HOLDER, University of Cape Town
YI XUE, University College London
QOBOLWAKHE DUBE, University of Cape Town

Supervisor:
THOMAS MCWALTER, University of Cape Town

African Collaboration for Quantitative Finance and Risk Research

Contents

1 Introduction

Lenders often demand collateral when extending credit. In this paper, *stock-based loans* (SBLs) – when a stock position is used as collateral – are examined.

In Section 2, we give an introduction to SBLs and develop a mathematical formalism in order to analyse them quantitatively. This involves introducing notation to capture the key variables of a prototypical SBL.

Section 3 builds on the formalism by developing an approach to measure the key risks inherent in an SBL. As will be explained below, an SBL is dependent on the price path of the equity collateral, as opposed to the solvency of the borrower. Taking the perspective of the lender, we propose a number of different equity modelling approaches, providing some rationale for each. This section also describes the parameterisation and implementation of our models and the SBL formalism. This is based on an example deal that we outline and proceed to analyse.

In Section 4, our results regarding the example deal are presented. Our empirical model – based directly on historical share data – is most able to present the potential losses and risks of an SBL. Traditional models do not give rise to these losses at typical parameter levels, which we take as tentative evidence that such losses are not very likely. Our final risk assessment is positive – the SBL, at prevailing parameters, appears an attractive prospect from the lender's perspective. In addition to this assessment, we also examine a few dynamics and effects of economic importance.

Section 5 concludes and suggests possible extensions.

2 Stock-based loans

2.1 Background

Stock-based loans (SBLs) are a type of loan where the borrower provides collateral for the loan in the form of an equity portfolio. The borrower is said to *pledge* the stock position that he or she owns as security against a borrowed amount over a fixed term.

The initial borrowed amount accumulates interest at some rate, typically a reference rate plus a spread. Dividends from the pledged stock, as they are received, are generally used to pay down the loan.

To protect the lender, the agreement involves a minimum *share cover ratio*: a certain ratio of the stock position value to the loan value, that is exceeded at initiation of the loan and must be exceeded for the life of the loan. If the minimum share cover ratio is *breached*, the lender obtains the right to liquidate the equity position in order to recoup the loan value at the relevant time. SBLs are *non-recourse*: the borrower is not liable beyond the equity they have pledged, and therefore if the lender fails to recoup their loan value from the equity liquidation, they suffer a loss. If the collateral value attained in liquidation exceeds the loan amount, this excess is returned to the borrower.

Typical SBLs provided by large banks involve an initial cover ratio (of the equity position to the loaned amount) in the region of three, and a minimum cover ratio in the region of two. In practice, when the minimum cover ratio is breached, the bank will decide, on a case by case basis, whether and in what fashion to liquidate the pledged equity.

The loans typically have maturities of two to six years. Because of this relatively short term, there is generally no explicit arrangement for the stock position to be adjusted (if, for instance, the cover ratio becomes close to the minimum) – instead, the deal is renegotiated upon breach or maturity.

An SBL provides a number of benefits to the lender. Because of the general liquidity of equity markets, the lender retains the essential and valuable option of liquidation in the event that the collateral value is in risk of no longer covering the loan position. This, for instance, obviates the need to undertake extensive due diligence on the borrower – because of the pledged equity, the particular state of the borrower becomes much less important.

A more typical corporate loan will generally involve a recourse option – in the event of a failure to pay, the lender has the right to recover their loan, to the extent possible, from the unwinding of the borrower's assets. Even though the lender has privilege over, for example, the borrower's equity holders, the unwinding and recovery process will typically be slow and expensive. While an SBL lender has no claim on the borrower's assets beyond the pledged equity, they can realistically hope to liquidate the pledged equity quickly and cheaply when appropriate.

The lender is naturally concerned about the factors affecting their ability to prevent any losses – for SBLs, these include the liquidity and value of the pledged equity, and the minimum cover ratio. Additionally, the lender will very possibly face regulatory capital requirements – these can be costly, and will likely rely on a wider set of variables.

From the point of view of the borrower, the negotiated terms – in particular, the interest rate spread – dictate the attractiveness of the SBL. Certain borrowers are happy to pledge equity in an effort to negotiate a favourable borrowing rate, because they intended to retain the equity in any case. Such a borrower would secure a loan and not give up the benefits of owning the stock – the dividends would contribute toward the loan repayment, and the equity would be returned at the conclusion of the loan (if minimum cover ratio is not breached).

SBLs have proved attractive borrowing vehicles for small and medium-sized enterprises, wealthy individuals and entities involved in Black Economic Empowerment (BEE) deals . There is are similarities between BEE deals – see West and West (2009) – and SBLs.

While we do not consider it in this paper, the borrower might have an option to repay the loan before the agreed maturity and reclaim the pledged equity. A few papers – see Zhang and Zhou (2009), Wong and Wong (2012) and Cai and Sun (2014) – have modelled SBLs from the borrower's point of view, analysing this American-style feature.

2.2 Formalism

Our goal in this paper is to examine the risk characteristics of an SBL from the perspective of the lender. This has not been attempted in the literature. In order to quantitatively model the potential outcomes, a few simplifying assumptions are

necessary.

Firstly, a mathematical analysis requires some unambiguous trigger for the equity liquidation (because the case-by-case discretion seen in practice is impossible to model). We will consider minimum cover ratios slightly smaller than typically used, but use breach of this minimum to indicate immediate liquidation of the pledged equity. The more typical minimum ratios (of approximately two) can thus be viewed a warning ratio, for the *possibility* of liquidation. Our findings will be prudent in this sense – a lender might feel that their ability to exercise discretion will improve the SBL risk characteristics.

Secondly, we assume the liquidity characteristics of the equity are predictable in the following sense: we assume that liquidation over a particular number of days will not adversely affect market prices. Traders generally have an accurate sense of the percentage of the daily volume that can be traded without incurring significant liquidity risk, and therefore estimating the number of liquidation days is straightforward. In the final section, we briefly discuss relaxing this assumption.

We now propose the following formalism for a typical SBL, based on the essential features of the description so far. The price process of the stock underlying the loan is denoted $\{S_t\}$. We let $\{D_t\}$ denote the *cumulative* concomitant dividend stream, initialised at zero at the start of the loan period (i.e., $D_0 = 0$). The amount loaned at time zero is written N_0, and this *loan* process evolves according to

$$dN_t = (r + s)N_t dt - dD_t, \tag{1}$$

where r denotes the short-term risk-free rate of interest and s the spread that the lender charges. We assume both of these quantities to be constant. Letting n_s denote the number of shares in the stock position underlying the loan (all other things equal, this has a one-to-one relationship with the initial cover ratio), and letting n_b denote the breach ratio, we define the *breach stopping time* τ with

$$\tau = \inf\{t \geq 0 | n_s S_t \leq n_b N_t\}.$$

This has the straightforward interpretation of the time the minimum cover ratio is breached. In the case that $\{\tau < T\}$ – where T denotes the agreed duration of the SBL – the bank is owed N_τ and is assumed to begin liquidating the shares. We assume that the position can be liquidated in n_d days - in other words, we assume that the price process $\{S_t\}$ does represent a sufficiently liquid price, provided we

do not attempt to liquidate any faster than over a period of n_d days. We define a *liquidation value* random variable $X^{(\tau)}$ with

$$X^{(\tau)} = \sum_{k=0}^{n_d-1} \left(S_{\tau+\frac{k}{250}}\right) \frac{n_s}{n_d} e^{-r\frac{k}{250}}.$$

Note that although $X^{(\tau)}$ is not measurable at the breach time, the separate cash flows have been discounted to time τ.

In the case that $\{\tau \geq T\}$, the final loan amount N_T is simply returned to the lender. The *net present value* of the loan is given by

$$\begin{aligned} NPV = -N_0 + \mathbb{I}_{\tau<T} \min(N_\tau, X^{(\tau)}) e^{-r\tau} \\ + \mathbb{I}_{\tau\geq T} N_T e^{-rT} + \int_0^{\min(\tau,T)} e^{-ru} dD_u. \end{aligned} \tag{2}$$

This construction is appropriate as it simply reflects the discounted cash flows of the SBL from the lender's perspective – after lending the initial amount, the lender receives a capital repayment (either upon breach or upon loan maturity), as well as the dividends throughout the life of the loan (which, as reflected above, pay down the loan). The potential variation in this random variable is the risk the lender faces in entering the SBL. Of particular significance is the term with the minimum, to which the lender is exposed if the breach occurs during the deal. In this case, the lender *might* recover the loan balance N_τ, but will lose part of it if $X^{(\tau)} < N_\tau$. We therefore define the following risk metrics; the *probability of breach* is given by

$$PB = \mathbb{P}[\tau < T],$$

where $\mathbb{P}$ is the real-world (or data-generating) probability measure. Expectations under this measure are used to define the *loss given breach* thus

$$LGB = \mathbb{E}[\max(N_\tau - X^{(\tau)}, 0) | \tau < T].$$

Note that, as reflected in Equation (2), if the liquidation value exceeds the amount owed, this *does not result in a profit* to offset losses from insufficient liquidation. It is only when $X^{(\tau)} < N_\tau$ that the lender does not receive the owed amount of $N_{\min(\tau,T)}$.

If, on the other hand, no breach occurs, the bank simply receives dividend payments against their loan, and then the rolled up loan balance at time T. Taking this possibility into account, we can then define an *expected loss*, without conditioning on a breach occurring, simply with

$$EL = PB \times LGB.$$

Also, importantly, we can consider the SBL value averaged over all possibilities by defining *expected net present value* with

$$ENPV = \mathbb{E}[NPV].$$

Note that while this accounts for time value of money (as the cash flows are discounted to time zero), it has made no adjustment for risk. For the risk they take on, the lender should demand that the parameters of the SBL are such that $ENPV$ is positive. A typical way to think about this would be to ensure that the spread s is large enough to offset the losses, on average, from a particular breach and liquidation risk. The $ENPV$ metric can be related to an *excess annual return* r^* by setting

$$1 + \frac{ENPV}{N_0} = e^{r^* \mathbb{E}[\min(\tau, T)]},$$

so that

$$r^* = \frac{\log(1 + \frac{ENPV}{N_0})}{\mathbb{E}[\min(\tau, T)]}. \tag{3}$$

In the absence of losses, reflected by a non-zero EL, we would have $r^* \approx s$. Due to some convexity adjustments in the expectations, we do not have $r^* = s$, but it is intuitive that we have an approximation: the lender has succeeded in charging the spread above the risk-free rate, and thus earns an excess return.
Note that we do no introduce a risk-neutral measure – these positions are not typically hedged, and our risk-measurement approach will involve attempting to get a handle on the real-world dynamics of the key quantities.

3 Risk management approach

3.1 Requirements and concerns

Continuing with our SBL risk analysis, we first point out that the position resembles an equity option more than a loan – losses are contingent on future equity

prices, rather than on the solvency of the borrower. Despite this, typical stock-based lenders do not delta-hedge their positions. Our analysis will thus continue under the real-world measure; a hedging approach is briefly discussed in Section 5.

Given that the position is not hedged, the primary risk metrics of an SBL relate to the real-world distribution of the NPV. Its expectation $ENPV$ is of course key: this is the value of the deal to the bank, adjusted for time value of money, on average. Other distributional characteristics are also relevant[1]. For instance, banks are obligated to hold capital based on the 99.5% Value-at-Risk (VaR) of these kinds of risky positions.

As mentioned above, the source of contingency of the SBL is the variation in the underlying equity. In order to produce risk metrics that faithfully represent the risks of an SBL, one is required to represent the different categories of contingency, and represent each category with a reasonable and realistic probability. As a foundation to our model selection, we propose the following categorisation of stock price path:

- paths that do not result in a breach – in this case the lender successfully earns the interest rate (with spread) without incurring any loss;
- paths that result in a breach but allow a full recovery of the loan amount: $\tau < T$ but $X^{(\tau)} \geq N_\tau$;
- paths that result in a breach and do not allow a full recovery of the loan amount – that is, $\tau < T$ but $X^{(\tau)} < N_\tau$;
- paths that result in a breach and a *large* loss – in this case $X^{(\tau)}$ is *significantly* smaller than N_τ.

The next subsection will propose a few modelling choices for the pledged equity price process $\{S_t\}$. The choices are motivated by the models' abilities to produce paths of each of the above categories. Reasonable parameterisation of these models (as well as parameter sensitivity and other analysis) will then provide a reasonable probability for each category and thus plausible SBL risk metrics.

[1] We are considering the case of a single entity – if a large portfolio of SBLs is held, risks indicated by these other characteristics will be mitigated by the aggregation and diversification of the many SBLs.

3.2 Model types

We take three broad approaches to modelling the equity price process. The first – drawing from standard risk-management texts such as Coleman (2011) – is *empirical* (or *historical*): returns are randomly drawn from an empirical return history. The particulars of our data are introduced in the next subsection.

The second approach is a direct model of the stock price process. In particular we opt for a Merton (1976) model, where the stock price process satisfies

$$\frac{dS_t}{S_{t^-}} = \mu dt + \sigma dW_t + (J_{(t)} - 1)dN_t,$$

where $\{W_t\}$ is a standard Brownian motion, $\{N_t\}$ is a Poisson process with intensity λ and $J_{(t)}$ are the jump sizes. These have a shifted log-normal distribution; in particular, $\log(J_{(t)}) \sim N(\alpha, \beta)$, independently for each jump. Note that μ, σ, λ, α and β are constants, and that the Poisson process and jump sizes are independent of the Brownian motion and each other.

The rationale behind such a choice is that jumps make allowance for significant, abrupt changes in stock price – significant drops in stock value could induce paths where losses are incurred by the lender. We will thus be in a position to capture such risks.

Our third approach is a model of firm value $\{V_t\}$, which is used to imply the equity value is the following simple manner:

$$S_t = \max(V_t - D, 0),$$

for a constant D. The firm value is modelled with a simple geometric Brownian motion thus

$$\frac{dV_t}{V_t} = \mu_V dt + \sigma_V dW_t,$$

for constants μ_V and σ_V. The rationale is based on the *structural* approach to credit risk – see, for instance, Black and Cox (1976) – where the equity value is considered a option on the firm's assets. Our equity to firm value relationship is extremely simple, but economically intuitive: the equity holders have the right to sell the firms assets, pay off the debt, and claim the difference (if it is positive). This simple relationship achieves two goals: firstly, the model offers an endogenous possibility of default (if V_t reaches the debt level, then $S_t = 0$) – this point is addressed

directly in Section 4.2 – and secondly, it induces a *leverage effect*. Although the returns of the firm value process have constant volatility, the truncation causes the stock-return volatility to increase as S_t becomes low (because S_t will have the same absolute variance as V_t, but the relative volatility, reflected in the returns process, will increase as the firm nears bankruptcy).

The direct equity model does not have an endogenous default component, but they can easily be embellished with an exponentially distributed time of default. This is implemented in Section 4.2.

While the jump diffusion and the leverage effect are a start to representing the different categories of stock price paths, we also introduce a *distressed regime* to some of our models. Hamilton (1989) is the seminal reference for general *regime-switching models*, where different parameter sets are allowed to prevail when the model is in different regimes. In order to breach the minimum cover ratio, the stock underlying a SBL would have suffered a downturn in value. This is not an implausible event – the equity market is largely unpredictable after all, and indeed all of our models would ascribe some probability to this outcome. Some breaches, though, would be caused by a stock in significant financial distress. Such stocks would likely exhibit different characteristics, and it would be prudent to make allowance for this. The implementation details of the distressed regime are given below.

3.3 Model parameters

We now consider an example SBL, based on market data and parameter levels used in practice. The stock underlying the loan is Brait. The loan size, N_0, is R300 million, but we normalise this by setting $N_0 = 100$. The duration is three years the initial cover ratio is 2.5, and breach ratio is 1.5. We set $S_0 = 100$, so that $n_s = 2.5$. Given the average daily volumes, and the pre-normalised loan size, we set a prudent liquidation period of $n_d = 10$ days[2]. Brait pays relatively small dividends – for all of our price paths, we assume two dividends a year, each paying 0.5% of the stock price at the relevant time.

We set $r = 0.06$ and $s = 0.04$ (s will be altered below) – these are approximate

[2]This requires the lender sell shares of approximately 10% of the daily volume, which we are told is a prudent assumption. In some cases, traders are comfortable selling 20% of daily volume.

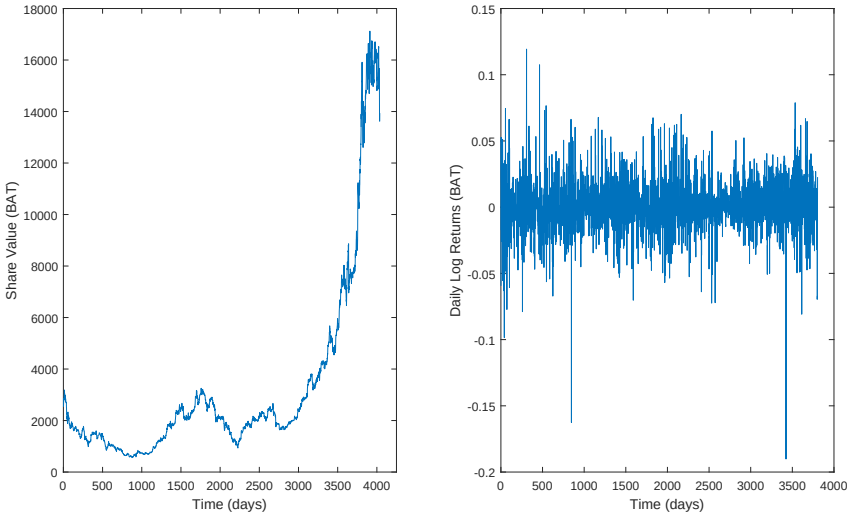


Figure 1: *Brait share price and log return history.*

levels of South African interest rates and prevailing SBL spreads.

We have a 16 year history of daily Brait returns (see Figure 1) – these are randomly sampled for our empirical approach. Note that our implementation is discrete, at a daily frequency. We therefore define $\Delta t = \frac{1}{250}$, which corresponds to approximately one business day.

Parameter estimation for Merton jump model is a difficult problem – the jumps are difficult to dissociate from Brownian innovations (see Aït-Sahalia et al. (2009) and Lee and Hannig (2010) for this problem in more general terms), and maximum likelihood[3] estimation is not robust. We therefore take jump parameters that other authors have estimated in a time-series analysis. We have reviewed a number of such papers – Hanson and Westman (2002), Khaldi et al. (2014), Ramezani and

[3]The density, and therefore the likelihood, is available in closed form – see Hanson and Westman (2002).

Zeng (2007) – and we set our jump parameters with $\lambda = 0.6$, $\alpha = 0$; and $\beta = 0.007$. λ and β were estimated in a reasonably similar region in the studies. Our β value is from the upper range of the estimated values (these parameters are altered in Section 4). α was very small and positive, and therefore we set it to zero (note that we did not consult papers that fitted these models cross-sectionally).
With this in mind, consider the log-stock $X_t = \log(S_t)$ in the Merton jump model:

$$dX_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t + j_{(t)}dN_t,$$

from Itó's lemma (see Crépey (2013) for the formula extended to jump diffusions). At our discretisation time step Δt, we therefore have

$$X_{t+\Delta t} = X_t + (\mu - \frac{\sigma^2}{2})\Delta t + \sigma(W_{t+\Delta t} - W_t) + \sum_{k=1}^{N_{t+\Delta t}-N_t} j_{(k)}.$$

Taking the jump parameters as above, we ensure that mean and variance of the log returns are consistent with the observed share price history. We therefore set σ to ensure

$$\sigma_X^2 = \mathbb{VAR}[\log(\frac{S_{t+\Delta t}}{S_t})] = \sigma^2\Delta t + \lambda\Delta t(\alpha^2 + \beta) \tag{4}$$

and μ to ensure

$$\mu_X = \mathbb{E}[\log(\frac{S_{t+\Delta t}}{S_t})] = (\mu - \frac{\sigma^2}{2} + \lambda\alpha)\Delta t,$$

where μ_X and σ_X^2 are the mean and variance of the return history.
Note that the above adjustment of the moments as a result of the jumps is different to the adjustment in no-arbitrage pricing – to avoid arbitrage, it is not the expected log return that needs to be restricted, it is the expectation of the future stock price.
Our third approach (which models firm value) is also difficult to estimate; while we want to infer parameters for the firm value process, our data pertains to the stock process, and the translation is not easily handled. We opt for a very simple but prudent parameterisation: we set $\mu_V = r$ (the firm, before leverage, grows no faster than the money market) and set σ_V as we set σ above (this will then be leveraged up when considering the equity). In the case of Brait, leverage is extremely low – we set D equal to one sixth of the firm value at initiation (corresponding to a debt-equity ratio of one to five).

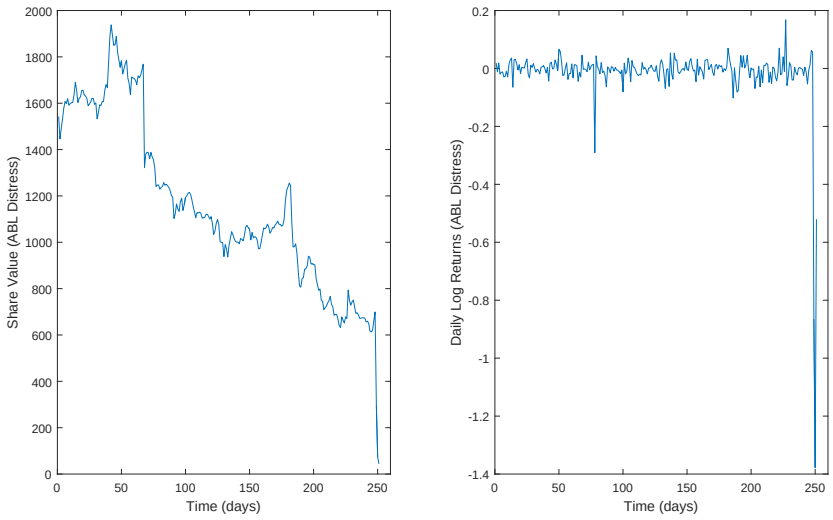


Figure 2: *African Bank share price and log return history.*

We also implement versions of our empirical and Merton jump models that include a simple distressed regime. An exponentially distributed stopping time (with intensity λ^*) is used as the distress indicator – for times after the realised stopping time, we draw daily returns instead from a one-year history of African Bank (see Figure 2). Our history is the last year before African Bank's default, which contains a number of large negative returns, and thus is suitable as a distressed distribution. For the Merton jump model, after each realised stopping time, we make the following parameter adjustments:

- we set $\mu = -\lambda^* \Delta t \alpha$ – this is consistent with an agnostic view on the stock future average movements (and note that this corresponds to the no-arbitrage drift restriction mentioned above – here we ensure that $\{S_t\}$ is a $\mathbb{P}$-martingale, when in distress);
- we use a different empirical volatility $\tilde{\sigma}_X$ – one estimated from the African

Bank data, excluding the final three days of extreme downturn – to set σ to be consistent with Equation 4;

- we double the Merton jump intensity λ.

These are very simple and heuristic adjustments[4]. Our motivation has been to give a reasonable approximation of a distressed stock environment without overparameterising the model. In a study where a more intricate estimation is feasible, many of these quantities – the jump intensity in the distressed regime, for example – could be estimated from specific data sets.

Similarly, the distress intensity λ^* could be estimated if one had a flexible and robust estimation procedure. In our case however, our aim is not to be overly focussed on the particular Brait and African Bank data. We set $\lambda^* = 0.015$ and as an initial reasonable (annual) distress intensity – having selected a non-distressed stock to initiate the loan, this seems a plausible intensity to account for the possibility of future distress. We will of course vary this parameter, and instead of being focussed on one value, will be able to examine the role of this (and other) parameters – we could for instance investigate what ranges of the parameter give rise to acceptable risk metrics.

We conclude with some final implementation details. We discretise the dynamic (in Equation 1) of the loan process $\{N_t\}$, and we also monitor the breach barrier on a daily, as opposed to continuous, frequency. We use a Monte Carlo sample size of 25000, and seed our pseudo-random number generator so that our comparisons do not rely heavily on sample variation.

4 Results

4.1 Basic implementation

Based on our initial parameter choices and on a large simulation of stock price and corresponding loan process paths, we present a number of metrics – defined above in Section 2.2 – for our different models in Table 1.

Firstly, we are able to obtain a first sense of the probabilities of breach (PB) – at the parameters of the example Brait deal, breach probability is relatively large ($> 20\%$)

[4]For background to our choices we reviewed a few papers – McEnally and Todd (1993), Hilscher et al. (2011), and Correia and Población (2015) – that address the characteristics of distressed equity.

Table 1: Results of initial parameters

	$ENPV$	PB	LGB	$\mathbb{E}[\tau \wedge T]$	r^*
Empirical	10.925	22.78%	0	2.689	3.86%
Empirical with Distress Regime	10.416	26.12%	1.231	2.636	3.76%
Merton Jump	10.884	23.50%	0	2.679	3.86%
Merton Jump with Distress Regime	10.832	24.40%	0	2.667	3.86%
Structural	8.978	49.46%	0	2.215	3.88%

for all our models. Recall that our minimum cover ratio is smaller than those often applied (because we take breach as an automatic decision to sell, rather than a warning for potential liquidation), so the breach probability would be at least as large in practice. The structural model estimates an even higher PB.

With regard to LGB, we have a value of zero for all models except the empirical-with-distress. This means that our models are clearly producing paths with downturns that cause breaches but, at our initial parameters, these are not sufficiently pronounced to cause losses to the lender (the liquidation value is sufficient to cover the loan amount). Despite our model constructions (in particular, the jumps, the distress parameters and leverage effect of the structural model), we are in fact not realising paths of all the relevant categories discussed in the previous section.

For the empirical-with-distress model, though, we do realise paths that result in losses. The LGD is non-negligible, but relatively small. It causes only minor decreases in $ENPV$ and annual excess return r^* – we now briefly discuss these metrics.

One might expect that $ENPV \approx$ (term of deal) $\times$ (spread) $= 3 \times 4$ – this is because $ENPV$ relates to a deal with a term of three years, and the four percent spread is charged on an initial principal of 100 (the deal, without losses, should therefore offer a 12% gross return). In other words, the lender is charging a spread of 4% for three years and (except in the empirical-with-distress case) is not incurring any losses. The $ENPV$ results are, however, less than this. The expected life of the deal – note that $\mathbb{E}[\tau \wedge T]$ is reported in Table 1 – reflects the fact that the deal does not always last the full three years, hence the lower $ENPV$ results. The excess return

Table 2: Incorporating possibility of default

	$ENPV$	PB	LGB	$\mathbb{E}[\tau \wedge T]$	r^*
Merton Jump	9.797	23.91%	4.671	2.665	3.51%
Merton Jump with Distress	9.662	25.66%	4.574	2.644	3.49%
Structural	7.928	57.20%	0	1.964	3.88%

metric, defined in Equation 3, corrects for this, and we then see that excess return in indeed in the region of 4%.
Further, we can then see that the losses incurred in the empirical-with-distress case only detract 0.1% from the excess return.

4.2 Incorporating default

A possible objection to the initial model choice and parameterisation is that it does not take account of credit rating data. According to its rating, Brait is supposed to have a 0.56% probability of *default* (rather than breach) in a year, and a 1.67% probability of default in three years (the term of our example deal).
One could argue that our models should reflect the possibility of default ($S_t = 0$) with at least the default probability implied by the credit rating. Presumably losses would be incurred at least some of these cases.
We therefore add an exponentially distributed jump-to-default to the Merton Jump models. The jump to default parameter, λ_d, is set to make probabilities of default consistent with the credit rating.
For the firm-value model, with our original specification of $D = 20$, there is only a 0.004% probability of default. This increases to 1.850% when the debt level is raised to 50, which is close to the credit rating probability. Although Brait is not geared this highly, we implement this variation (which ensures we are not falling foul of the credit rating probability) and investigate the effects, such as the increased leverage effect.
For the two Merton Jump models, and the firm-value model, we report revised estimates for our risk metrics in Table 2, taking into account the above. As one would expect, the inclusion of a jump to default decreases $ENPV$ and excess returns, and increases LGB in both Merton jump models. The expected loan life decreases of

Table 3: Proportions of different recovery (R) classes

	No breach	Full R	R>80%	50%<R<80%	R <50%
Merton Jump	76.09%	22.46%	0.016%	0.05%	1.38%
Merton Jump with Distress	74.34%	24.16%	0.012%	0.04%	1.45%
Structural	42.81%	57.20%	0%	0%	0%

course, but only marginally. The firm-value model, with its increased leverage, exhibits a very high probability of breach. However, none of these breaches give rise to a loss.

To elaborate on this, we sort the realised paths into a few categories (recall the discussion of the different categories of stock path contingencies) based on the recovery made from the liquidation. We denote the fraction of the loan recovered R, and report this categorisation in Table 3. For the Merton jump models, although we have a significant number of breaches, the only ones that incur a loss are the jumps to default – this is clear because virtually no paths breach and give rise to *small* losses. For the firm-value model, as seen in the LGB metric, all of the many breach paths give rise to full recovery.

The leverage effect, even at the artificially high debt level, does not give rise to sufficient variability to induce losses. Nor do the jumps in the Merton model, but the separate jump to default of course does. One could argue that this immediate jump to default is not very realistic, however. Typical defaults are not quite so abrupt, and very well may allow some recovery.

In conclusion, we have found so far that usual equity variation (arising from our distressed regime, our Merton jumps, parametrised at typical levels and increased during distress, and the leverage effect in the firm-value model), even making allowance for distress, does not give rise to losses in the SBL context. We of course do get losses from the jump to default, admittedly in a rather crude fashion.

We do not embellish the empirical model with a jump to default, because, at the assumed distress intensity, the probability of an *effective* default (which we define as the stock losing 95% of its value, i.e., $S_t < 5$) over three years is 3.020%. Because we sample *returns* (all of which are finite), we never attain $S_t = 0$, but we do have enough paths (with the credit rating in mind) to reflect the possibility of a

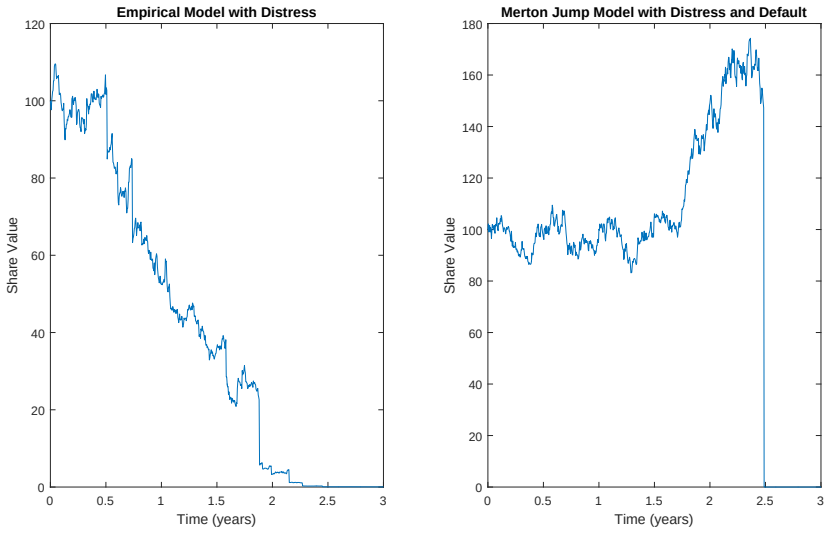


Figure 3: *Example defaults from our empirical and jump models.*

catastrophic downturn in stock price. Further, this is likely of a more realistic nature than instantaneous jump to default – see Figure 3 for contrasting examples of downturn: the empirical model can capture highly distressed periods of downturn (and can allow different degrees of such downturns), whereas the jump-to-default is not so much a downturn as an instantaneous, arbitrary stock price crash.

4.3 Parameter sensitivity

4.3.1 Merton jump model with distressed regime and default

The intention here is to verify that the Merton jump model's failure (even when embellished with a distressed regime) to produce losses is not due to our specific parameter choices (some of which were heuristic). The mean return is of particular interest: this parameter is notoriously difficult to estimate accurately, and we want to verify that our results are robust to drift changes.

Table 4: Sensitivity Analysis of Merton Jump Model with Distress Regime

Drift	Vol	Beta	Vol*	$ENPV$	PB	LGB	$\mathbb{E}[\tau \wedge T]$	r^*
$\mu - 3\%$	σ	β	σ^*	9.403	31.00%	3.735	2.576	3.49%
$\mu - 6\%$	σ	β	σ^*	9.170	36.96%	2.889	2.497	3.51%
$\mu - 3\%$	1.25σ	$1.25^2\beta$	$1.25\sigma^*$	8.627	40.85%	2.450	2.359	3.51%
$\mu - 3\%$	1.25σ	β	$1.25\sigma^*$	8.511	41.21%	2.630	2.348	3.48%
$\mu - 6\%$	1.5σ	β	σ^*	7.547	52.18%	1.854	2.094	3.48%

Table 5: Sensitivity of Structural Model

$ENPV$	PB	LGB	$\mathbb{E}[\tau \wedge T]$	r^*
5.712	73.70%	0	1.431	3.88%

In Table 4, we report our risk metrics at a variety of parameter levels. The breach probabilities, as one would expect, increase, as the parameters become worse (have a lower mean return or are more volatile), and this increase is significant. However, $ENPV$ decreases only slightly for each iteration of less favourable parameters. This decrease is mitigated by the fact that the expected life of the deal also decreases – when viewing the excess return r^*, which accounts for this, the effects from the parameter changes are very minor.

Note that LGB is *lower* when we consider worse parameter sets. This is odd at face value, but recall that losses were driven only by jump to default – this is true at the different parameter levels too. With the worse parameters, a breach before a default becomes more likely, and these paths give rise to a breach without a loss.

4.3.2 Firm-based model

As a final check on our firm value model, we report our metrics based on the higher debt level of 50 and a volatility increased by 50% – see Table 5. The probability of breach increases significantly, but the loss given breach remains at zero.

This firm-value approach has some attractive elements, in particular the endogenous possibility of default. This could allow a natural possibility of default. Thus this may be a good approach in a study that has more time to consider estimating the relevant parameters, including a jump component. The continuity of the paths, even with a leverage effect and higher volatility, prevents any losses from

Table 6: Varying λ^* in Empirical Model with Distress Regime

λ^*	PD	$ENPV$	PB	LGB	$\mathbb{E}[\tau \wedge T]$	r^*
0.015	2.93%	10.442	25.68%	1.237	2.642	3.76%
0.009	1.67%	10.685	24.65%	0.613	2.664	3.81%

Table 7: Varying jump into distress intensities for empirical model

λ^*	$ENPV$	VaR 99.5%	ES 99.5%	PD
0.075	3.638	-41.429	-46.389	14.46%
0.15	2.044	-46.399	-52.053	26.60%
0.225	0.980	-47.119	-52.928	36.54%
0.3	-0.122	-47.230	-51.352	45.49%

this model in our study.

4.3.3 Empirical model

We now consider different distress intensities for our empirical model. Table 6 considers a *calibration* of the (effective) default probability to that implied by the credit rating. Although this model is dependent on our specific data, this degree of freedom is available to customise the model.
Table 7 considers a modest spread of 2%, and varies λ^* to determine at what rate $ENPV$ becomes negative. This is an upper bound investigation of the distress intensity – at this relatively low spread, we confirm that the initial intensity needs to be increased significantly to make the SBL unattractive to the lender on average.

4.3.4 Spread

Table 8 documents the effects of manipulating the spread (with two models – the empirical model uses the initial distress intensity, and the Merton jump model results were simulated with a down-adjusted drift of $\mu - 3\%$). As the spread increases, expected net present value of course also increases, and there is a minor increase in probability of breach. This feedback dynamic – the spread inducing more breaches – was of interest at the outset of our study. We find that the dynanmic effect is minor – a 1% increase in the spread induces an *almost* 1% *increase* in the excess return (and so the relation is weakly sub-linear).

Table 8: Sensitivity of Spread Parameter

	s	$ENPV$	PB	$\mathbb{E}[\tau \wedge T]$	r^*	EL
Empirical with Distress	1%	2.361	21.52%	2.698	0.86%	0.284
Empirical with Distress	2%	5.012	22.90%	2.682	1.82%	0.305
Empirical with Distress	3%	7.710	24.66%	2.658	2.79%	0.296
Merton with Distress	1%	1.599	25.91%	2.642	0.60%	1.049
Merton with Distress	2%	4.061	27.84%	2.619	1.52%	1.207
Merton with Distress	3%	6.704	29.44%	2.593	2.50%	1.175

Table 9: Sensitivity of Minimum Cover Ratio

Model	n_b	PB	LGB	EL	$ENPV$	$\mathbb{E}[\tau \wedge T]$	r^*
Empirical w. Distress	1.1	10.17%	7.792	0.792	11.048	2.888	3.63%
Empirical w. Distress	1.2	13.34%	4.623	0.617	11.042	2.849	3.68%
Empirical w. Distress	1.3	16.93%	2.646	0.448	10.967	2.795	3.72%
Empirical w. Distress	1.4	21.15%	1.832	0.388	10.754	2.732	3.74%
Merton w. Distress	1.1	12.16%	11.392	1.385	10.454	2.875	3.46%
Merton w. Distress	1.2	15.69%	8.558	1.343	10.301	2.830	3.47%
Merton w. Distress	1.3	20.28%	5.977	1.212	10.142	2.765	3.49%
Merton w. Distress	1.4	25.01%	4.380	1.095	9.918	2.685	3.52%

We note that, according to our models so far, a 1% spread appears sufficient for the lender to cover their potential losses on average.

4.3.5 Minimum cover ratio

Table 9 confirms the basic intuition that as the breach cover ratio increases, the probability of breach increases and loss given breach decreases. We find that the net effect on expected loss (the product of these two metric) is a decrease. Because the expected loan life decreases, the $ENPV$ does not reflect this benefit from the increased minimum cover ratio, but the excess return of course does.

The lender would in fact enjoy two benefits if the breach ratio increases – the average loss decreases (reflected by a lower EL); and the risk from losses is *less concentrated* (because the probability of a breach is higher, as opposed to the losses being driven by a small probability of breach with a high loss given breach). There would

Table 10: Merton Jump Model with Distress Regime

n_b	PB	$PB1$	$PB2$	$PB3$
1.1	12.16%	1.02%	4.94%	6.60%
1.2	15.69%	1.66%	6.58%	8.21%
1.3	20.28%	2.57%	9.16%	9.88%
1.4	25.01%	4.41%	11.64%	11.18%

however likely be some soft costs – such as the lender's relationship with the borrower – that would need to be considered when looking to increase the minimum cover ratio, along with the likelihood of prepayment.
We note finally that, for the Merton model, volatility drives expected loss only when the minimum cover ratio is set to 1.

4.4 Overall risk assessment

4.4.1 Merton jump model with distressed regime

While the Merton jump model, at typical parameter levels, is unable to induce losses, there is every reason to think that it produces reliable probabilities of breach. These are of interest to the lender. We investigate PB as well as conditional probabilities of default in each of the three years in the deal, assuming no prior breach has occurred.
Table 10 presents some changes from our initial value of the breach ratio. Table 11 alters the spread. These give a rough sense of the effects of these two negotiated quantities. The first-year probability is relatively low – this is because the deal is initialised at a specified initial cover ratio (whereas the second year is not guaranteed to be initialised in this way). This is likely beneficial from a capital provisioning point of view (as this often considers a one-year horizon).

4.4.2 Empirical model with distressed regime

Before making a final risk assessment, we briefly probe our sensitivity the assumption for $n_d = 10$ liquidation days (in this particular case, the assumption seemed prudent, but might have to be examined in other instances). Table 12 confirms that the effect of a longer liquidation period is relatively minor, and we are not terribly

Table 11: Merton Jump Model with Distress Regime - n_b = 1.5

s	PB	$PB1$	$PB2$	$PB3$
1%	25.91%	6.27%	11.87%	10.26%
2%	27.84%	6.50%	12.99%	11.26%
3%	29.44%	7.10%	13.59%	12.05%
4%	31.00%	7.15%	14.42%	13.12%

Table 12: Number of days to liquidate and affect on $ENPV$ and LGB

n_d	$ENPV$	LGB
13	10.374	1.327
16	10.342	1.690

exposed to a revision in this parameter. Table 13 shows that effects on recovery rates is intuitive but minor.

As our empirical-with-distress model has been favoured as the most representative of the potential losses resulting from an SBL, we give a final risk assessment of our example deal. Table 14 reports a summary for the initial, and lower (credit-rating calibrated) distress intensity. If one felt justified in using the lower, calibrated intensity, the risk assessment is very positive: even at a strict Value-at-Risk threshold, the lender hardly incurs a loss. The large losses are in the extreme tail and, according to the Expected Shortfall, still of a manageable magnitude. Of course, the picture deteriorates for the initial intensity, which, for prudence, might be a preferable view. Even in this case, the proportion of the $N_0 = 100$ capital that is lost at the 99.5% percentile is not very large.

Such risk metrics would be helpful to a lender bound by capital requirement regulation.

Table 13: Liquidation effect on different recovery (R) classes

n_d	No breach	Full R	R>80%	50%<R <80%	R <50%
13	0.736	0.252	0.005	0.006	0.001
16	0.743	0.243	0.005	0.007	0.002

Table 14: Empirical Model with Distress Regime

λ^*	PD	VaR 99.5%	ES 99.5%
0.015	2.93%	-22.082	-33.988
0.009	1.67%	-1.988	-20.406

5 Discussion

Based on our abstract formalism of a prototypical SBL, we have conducted a risk assessment of an example of such a loan. The credit risk of the loan depends on the path of the pledged equity, and therefore we have implemented a few different approaches to modelling possible stock price paths.

At the general level, we have found that typical SBLs, with prevailing spread sizes and other details, have attractive risk characteristics from the lender's point of view. On average, the spread is more than sufficient to cover potential losses – this is reflected in our positive $ENPV$ estimates. Furthermore, we have provided estimates of other distributional characteristics of the SBL net present value. Our empirical model, with a distress component, proved the most suitable basis for these calculations, and we have examined different levels of our distress intensity.

In addition to our risk assessment, we have examined some of the features and dynamics involved in SBL risk management. Our direct stock price models were not able to capture the kinds of losses which with the lender is concerned. We interpret this as evidence – albeit very tentative – that such losses are not especially likely in the first place. These models, however, are able to produce reliable estimates of breach probability, which are of interest to the lender. Also of interest are the economic effects of some of the key SBL parameters, such as minimum cover ratio and spread size, which we have examined.

We conclude by listing some possible extensions to our work.

- A more careful consideration of our assumption of liquidation period – although we were given to understand that our assumption is relatively safe, there may be instances (e.g., a very large loan, or an illiquid stock) where volumes would need closer examination. Some simple correlations between recent volumes and returns could give a first indication of any potential effects of interest.

- Stochastic interest rates, with correlation to the pledged equity, could be added to our modelling. We do not expect this to have a significant impact on our results, but this might be of interest to a practitioner who is constrained by funding cost adjustments.
- An aspect we have not considered is the potential prepayment by the borrower, for which some SBL contracts allow. This would be related to stochastic rates – as a borrower might want to refinance if rates or spreads decrease – and could affect how a lender thinks about their capital management.
- Finally, we could consider hedging the stock risk inherent in the SBL. A lender is of course exposed to an equity downturn, and this is at least partially hedgeable. In our modelling framework, this would require some numerical methods, as the stopping times, amongst other aspects, do not allow easy analysis.

Bibliography

Aït-Sahalia, Y., Jacod, J., et al., 2009. Testing for jumps in a discretely observed process. The Annals of Statistics 37 (1), 184–222.

Black, F., Cox, J. C., 1976. Valuing corporate securities: Some effects of bond indenture provisions. The Journal of Finance 31 (2), 351–367.

Cai, N., Sun, L., 2014. Valuation of stock loans with jump risk. Journal of Economic Dynamics and Control 40, 213–241.

Coleman, T. S., 2011. A practical guide to risk management. Research Foundation Publications 2011 (3), 1–212.

Correia, R., Población, J., 2015. A structural model with explicit distress. Journal of Banking & Finance 58, 112–130.

Crépey, S., 2013. Financial Modeling. Vol. 10. Springer.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica: Journal of the Econometric Society, 357–384.

Hanson, F. B., Westman, J. J., 2002. Stochastic analysis of jump-diffusions for financial log-return processes (corrected), 169–183.

Hilscher, J., Campbell, J., Szilagyi, J., 2011. Predicting financial distress and the performance of distressed stocks. Tech. rep., Harvard University Department of Economics.

Khaldi, K., Djeddour, K., Meddahi, S., 2014. Comparison of jump-diffusion parameters using passage times estimation. Journal of Applied Mathematics 2014.

Lee, S. S., Hannig, J., 2010. Detecting jumps from Lévy jump diffusion processes. Journal of Financial Economics 96 (2), 271–290.

McEnally, R. W., Todd, R. B., 1993. Systematic risk behavior of financially distressed firms. Quarterly Journal of Business and Economics, 3–19.

Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3 (1-2), 125–144.

Ramezani, C. A., Zeng, Y., 2007. Maximum likelihood estimation of the double exponential jump-diffusion process. Annals of Finance 3 (4), 487–507.

West, G., West, L., 2009. The pricing of bee share purchase schemes. The Southern African Treasurer: Special Issue on Risk Management, 35–40.

Wong, T. W., Wong, H. Y., 2012. Stochastic volatility asymptotics of stock loans: Valuation and optimal stopping. Journal of Mathematical Analysis and Applications 394 (1), 337–346.

Zhang, Q., Zhou, X. Y., 2009. Valuation of stock loans with regime switching. SIAM Journal on Control and Optimization 48 (3), 1229–1250.