Modelling Credit Spreads in an Illiquid South African Corporate Debt Market

Samantha Jones

Thesis Presented for the Degree of
Doctor of Philosophy

in the Department of Mathematics and Applied Mathematics
Faculty of Science
University of Cape Town

Supervised by
Doctor Henri Laurie, Doctor Ebrahim Fredericks,
Emeritus Professor Ronald Becker and Doctor Brett Dugmore

24 January 2019

Modelling Credit Spreads in an Illiquid South African Corporate Debt Market

Samantha Jones

Abstract

The South African debt market suffers from severe illiquidity, as is common in most emerging markets. Infrequent trading leads to out-of-date market prices and stale, unreliable credit spreads. Since the coverage of the South African debt market by credit ratings agencies is poor, meaningful credit spreads become even more important in gauging credit worth. The illiquidity of corporate vanilla bonds traded on the Johannesburg Stock Exchange and the ensuing adverse effects on their credit spreads are rigourously illustrated. Lack of data poses a serious problem when modelling any system and this analysis provides motivation for the necessity of a framework that addresses the statistical complications that incomplete data sets present. A new model, which is a distinctive modification of the well-known mean-reverting Ornstein-Uhlenbeck or Vasicek process, is introduced. This innovative approach creates a mathematically and intuitively sound relationship between the credit spread process and that of the stock price of the bond issuer. This key feature is used in a Bayesian methodology to impute missing credit spread data for calibration, for more meaningful inference. On sparse simulated data and market observed credit spread time series, the model proves to deliver an improved quality of the estimations, with probabilities that are now statistically founded. Even on complete credit spread time series, the model is shown to have some merit over the traditional model in terms of goodness of fit, giving further credence to its validity and explanatory power.

Plagiarism Declaration

1. I know that plagiarism is a serious form of academic dishonesty.
2. I have read the document about avoiding plagiarism, am familiar with its contents and have avoided all forms of plagiarism mentioned there.
3. Where I have used the words of others, I have indicated this by the use of quotation marks.
4. I have referenced all quotations and properly acknowledged ideas borrowed from others.
5. I have not and shall not allow others to plagiarise my work.
6. I declare that this is my own work.

Signed: Signed by candidate

To Kyle

Acknowledgements

I would like to thank each of my supervisors, Professor Becker, Dr Laurie, Dr Dugmore and Dr Fredericks for their invaluable guidance and dedication to the completion of this thesis.

Thanks go to Prescient Securities and Sygnia Asset Management for providing the data required for this research.

Many thanks as well to Gideon van der Linde, Emlyn Flint and Christopher Holdsworth for their insightful input and discussions.

Financial aid provided by the National Research Foundation, the University of Cape Town and the African Institute of Mathematical Sciences during the completion of this thesis is greatly appreciated.

Contents

List of Abbreviations

ACF Autocorrelation Function.

ADF Augmented Dickey-Fuller (test).

AIC Akaike Information Criterion.

AR Autoregressive (process).

BIC Bayesian Information Criterion.

CBOE Chicago Board Options Exchange.

CDS Credit Default Swap.

CIR Cox-Ingersoll-Ross (process).

EKF Extended Kalman filter.

FBA Full Bayesian Approach (imputation method within Gibbs sampler).

FFBS Forward Filtering Backward Sampling (method).

FRA Forward Rate Agreement.

FRN Floating Rate Note.

GBM Geometric Brownian Motion.

HLCD Ho-Lee Constant Drift (model).

ICB Industry Classification Benchmark.

KMV Kealhofer-Merton-Vasicek (model).

LM Liquidation Measure.

LVCF Last Value Carried Forward.

MAPE Mean Absolute Percentage Error.

MCMC Markov Chain Monte Carlo.

MI Multiple Imputation (imputation method).

MICE Multiple Imputation by Chained Equations.

MLE Maximum Likelihood Estimation/Estimator.

NACC Nominal Annual Compounded Continuously.

NACS Nominal Annual Compounded Semi-Annually.

OTC Over-the-Counter.

OU Ornstein-Uhlenbeck (model).

OUVLTM Ornstein-Uhlenbeck Variable Long Term Mean (model).

RB Rendleman-Barrter (model).

RMSE Root Mean Square Error.

SA South Africa(n).

SDE Stochastic Differential Equation.

SE Standard Error.

TFR Trade Frequency Ratio.

TIN Total Issued Nominal.

TP Time series of Traded Points.

TTN Total Traded Nominal.

USD US Dollars.

YTM Yield-to-maturity.

ZAR South African Rands.

ZTR Zero Trade Ratio.

1. Introduction

The sanctity of measured credit worth of corporate bonds in the South African debt market has long been under scrutiny. Quantification of credit risk of South African corporate debt has been difficult to formalise in any consolidated, meaningful framework, stemming from the pervasive and extreme lack of liquidity that plagues the debt market. Inadequate assessment of credit worth leads to a multitude of complications, all emanating from this inconsistency between the poorly gauged risk and the true credit risk. Credit instruments could also be mis-priced, leading to arbitrage opportunities.

Corporate debt is deemed risky as it is compared to South African government issued debt, which is government backed and considered to be essentially devoid of any credit risk. In fact, the credit spread of a corporate debt instrument is defined as the difference between that instrument's yield and the yield of a government instrument of the same type with the same duration (maturity, cash flow dates and coupon size). This spread serves to quantify the extra return offered due to the higher credit risk level, so more credit risk implies higher credit spreads and vice versa. However, infrequent trading means that less information is at hand to assess the market's current valuation of the risk to return profile of the debt instrument. The spread above the government instruments yield is uncertain during times when the instrument is not traded, since no new information about its credit worth has been built into its yield. Spreads thus become stale and out of date and lose their reliability as relevant credit worth monitors.

The listed corporate bond market (on the Johannesburg Stock Exchange or JSE) in South Africa is notoriously illiquid, with only government issuances trading on a daily basis (at best) and some corporate issuances not trading at all. The South African debt market is also still largely an over-the-counter (OTC) traded market. The debt derivative market, predominantly credit default swaps (CDS), interest rate swaps and forward-rate agreements (FRAs), are traded in large size in this way. Unlike the case where assets are exchange-traded, with trade information (such as price and volume) being publicly available, only the counter-parties to the OTC contract are privy to the particulars of the trade. Since the market-makers of these OTC trades are traditionally the large banks in South Africa, a consideration would be to source trade data directly from them to infer credit worth. But releasing such information to be used in the general market puts them at potential risk of revealing sensitive intellectual property, such as proprietary macro-economic views and credit models, and is likely the reason that this avenue has been unsuccessful thus far.

One could argue that credit ratings, compiled by the well-known and accredited agencies, Moody's, Standard and Poor's and Fitch, provide a viable alternative in assessing the risk in investing in debt from a specific corporate issuer. The rationale behind these ratings is that the better the credit rating, the less likely the issuer is to default on the repayment of its debt and so the investor's risk is reduced. In a market where every debt instrument is rated by a credit rating agency and ratings are updated regularly for changes in market climates, this argument would be acceptable. But to develop a credit rating for a particular issuance from a specific corporate issuer means an investigation into the unique fundamentals of that issuer as well as an understanding of the capital structure of the issuer. This process is costly in terms of time and resources and as such, issuers are be reluctant to absorb this cost to get their individual

issuances rated and thereafter continuously monitored (which would imply an annuity fee). In addition, access to this information is priced at a premium, typically in US Dollars (USD). To an average South African institutional issuer, the appeal of a rated issuance most often does not outweigh this substantial, contracted expense and so has led to less demand for the service. The rating agencies are subsequently deterred from actively seeking business in South African, reducing coverage even further - an unfortunate chicken and egg scenario. Furthermore, the rating of an issuance is expected to decrease uncertainty and so lower yields. Currently, only about 40% (RisCura Solutions (2015)) of the South African listed debt market has a credit rating (on an issuer level) provided by one of the aforementioned agencies. Thus it seems that the expense of attaining a rating outweighs the rating's reduction in the cost of the debt in the local market.

The illiquidity alluded to previously leads to an acute issue - problem with trying to statistically or mathematically investigate fragmentary, incomplete data is in drawing material and plausible results from such analyses. The application of any technique to data that is incomplete is going to produce outcomes that are most often baseless and biased. Data sets with missing points are penalised in traditional analyses as they contain fewer sample points, thus less information from which to postulate conclusions. The adverse effects missing data introduce into quantitative analysis are confirmed in a plethora of studies; Bennett (2001), Walton (2009), Dong *et al.* (2013) and Kwak *et al.* (2017) are but a few that refer to the decrease in quality of results when data is incomplete and go on to investigate ways to address missing data. The problem is anecdotally described by Honaker *et al.* (2010): "Most common methods of statistical analysis require rectangular data sets with no missing values, but data sets from the real (political) world resemble a slice of Swiss cheese with scattered missingness throughout."

The existing knowledge on modelling credit risk and default probabilities is extensive and well-documented in literature. Most of this research can be identified, to some degree, as based on either of the two primary approaches that have developed over the years; these being structural models and hazard rate/reduced form/default intensity models.

The groundbreaking research on structural models was done by Black *et al.* (1973) and Merton (1974). In this framework, it is assumed that default is signalled when the issuer's firm value drops below the values of its liabilities or some other predetermined threshold level. Black *et al.* (1976) elaborated on the model by examining different threshold signal assumptions. Stochastic interest rates were introduced into the Merton framework by Shimko *et al.* (1993). Longstaff *et al.* (1995) modelled the term structures of credit spreads in this setting, finding a negative correlation between interest rates and credit spreads, and Nielson *et al* (1996) explored a two-factor Merton model for interest rate and credit uncertainty. Geske (1977) extended the model by introducing compound options to refine the threshold by accounting for leveraged liabilities and further studies of this approach were discussed by Leland (1994) and Leland *et al.* (1996).

The weakness of structural models lies in determining default signal thresholds that represent real-world default scenarios. In reality, default may occur before the firm value reduces to the level of modelled liabilities. As such, these models need to be simulated longer for the threshold to be reached if the threshold is not realistically defined. As a result, structural models are prone to produce credit spreads that are lower than levels observed in the market, which is known as the credit spread puzzle, discussed in Jones *et al.* (1985), Amato *et al.* (2003), Tsuji (2005) and more recently refuted by Feldhütter *et al.* (2018). Nonetheless, structural models are still popular in the literature, as seen in Delianedis *et al.* (1998) and

Huang *et al.* (2012). A comparison of risk management methods using portfolios gauged by credit ratings versus Merton type credit spreads was performed by Nickell *et al.* (2001). The Merton based approach was found to not understate the portfolios' risks, as the credit ratings methodology did, another feather in the Merton model's cap. Kealhofer (2003i) and Kealhofer (2003ii) developed the backbone of the well-known Kealhofer-Merton-Vasicek or KMV model, which is used by Moody's in its credit rating system (see Crosbie *et al.* (2003)). It is a Merton based approach that also factors in credit cycles in calculating the 'distance-to-default', a concept that has become synonymous with structural models.

More recently, reduced form models for generating probabilities of default have come into favour due to their tractability. Default is treated as unpredictable and modelled by a point process governed by some hazard rate probability. Madan *et al.* (1998), Lando (1994), Jarrow *et al.* (1994) and Duffee (1999) assumed that the value of a defaultable contingent claim is discounted using a rate which includes this hazard rate added to the risk free rate. This assumption is the common thread throughout all reduced-form models. Jarrow *et al.* (1997) proposed a reduced form model where bankruptcy follows a Markov chain in credit ratings, whilst Schönbucher (1997) used a two-factor reduced form model to discuss the pricing of a few credit derivatives. In the work of Duffie et al. (1999), a defaultable version of the Heath-Jarrow-Morton model (Heath et al. (1992)) was developed and some derivatives on defaultable claims were priced. Modelling of risk premia of default intensities based on default events was studied by Driessen (2005). Saita *et al.* (2006) used a latent-factor intensity approach and empirically found that corporate debt compensated adequately for the risk measured.

The issue with using structural or hazard-rate models in illiquid markets is that both those frameworks require some degree of credit data for calibration. In most cases, CDS spreads are extracted from the market or credit ratings are utilised. Furthermore, additional information is most often needed for effective calibration, such as default recovery rates or losses given default, default transition probabilities or credit curves. Such data is severely limited in the South African market, partly due to illiquidity on exchange traded debt as mentioned before, but also due to the South African market being small and not as developed as those where default information is rich, where these models are easily applied. Without making a multitude of input data assumptions, which runs the risk misinformation, these models simply cannot be implemented successfully based on the current range of existing data.

Academic coverage of modelling credit spreads and default intensities in illiquid markets is growing but still relatively limited. The acute lack of trade observed predominantly in emerging markets was confirmed by Bernales *et al.* (2014), who used the Kalman filter to create complete data for Chilean government bonds, with the goal of improving measured risk metrics. Indeed, Nashikkar *et al.* (2011) stated that the direct cause of the ubiquitous problem of liquidity estimation was the lack of trade of corporate bonds, but the authors pointed out that the growth in CDS markets had abated this issue. Duffie *et al.* (2001) looked at calculating default intensities in terms of the conditional distribution of the issuer's assets based on accounting data, addressing the delay and infrequent publishing of such data. Credit spreads were then determined directly from this incomplete accounting information. Cross-sections of CDS spreads were examined using fundamental versus market models by Das *et al.* (2009), where the authors suggested that financials-based model could be used when trading was minimal. An extension was introduced by Shumway (2001), who included market-driven factors in forecasting bankruptcy and replaced missing firm data with that of a value-weighted US equity index. Campbell *et al.* (2008) claimed that using this model with different factors

and filling in missing data with cross-sectional means was more effective. By using Multiple Imputation to account for missing data whilst determining hazard rates via Shumway's model, Zhou (2007) furthers this research. The works of He *et al.* (2012) and He *et al.* (2014) consider the concept of debt rollover which arises from decreased liquidity, and ultimately leads to default occurring sooner.

Another approach was proposed by Fontana *et al.* (2010), where a reduced form credit spread model was used in conjunction with the Kalman filter in Expectation Maximisation for parameter estimation. Cortazar *et al.* (2012) developed a mechanism of jointly fitting a multi-factor Vasicek model to the Chilean market to obtain credit spreads for the partitioned different risk classes, using the Extended Kalman filter (EKF), as was done in the case of government debt by Cortazar *et al.* (2007). The incompleteness here refers to points in time across a range of bonds in the risk classes - the authors note that the EKF does not require the number of bond prices being jointly fitted to at each time point to be the same as at any other time point.

The current mechanism employed by the JSE to account for missing credit spread data is that of *Last Value Carried Forward*[1] (LVCF). The method is as its name describes; it imputes missing spreads by setting them equal to the previous traded spread level, carrying the static level forward until the next trade. LVCF fills in the missing spreads from a distribution that is different to the actual data, proposing a singular, static outcome of the credit spread with probability 1. LVCF estimates are biased and the uncertainty of these estimates are seriously underestimated (Lachin (2016)). Molnar *et al.* (2008) referred to LVCF as "the most inappropriate analytical technique" and Shoop (2015) concluded that usage of LVCF is statistically flawed and recommends that alternative techniques should be employed.

The primary aim of this research is to improve upon the current scheme of LVCF in the estimation of South African credit spreads and in so doing produce meaningful forecasting of credit spreads with reliable, more statistically reflective uncertainties using a suitable multiple imputation technique. Furthermore, the aim is to rely solely on publicly available information in this estimation framework, so as to mitigate the dependence on proprietary data from banks or credit ratings agencies. The model will thus allow unrestricted access to a credit worth quantification tool across the market.

The contributions that this research makes to the existing body of knowledge covering the determination of credit spreads in a thinly traded market are multi-faceted. Firstly, an enlightening and rigourous investigation into illiquidity of South African bond market is performed, which has not been seen before in published literature. Additionally, an innovative model to estimate credit spreads, also new to the literature, that follows the intuition that the credit spread must reflect the current credit worth is introduced. The new model postulates multiple outcomes of credit spread when data is missing using external data, and offers a probabilistically superior reflection of the possibilities of the credit spread to that of LVCF. The true uncertainty of the estimation is also captured by this model, which LVCF fails to do. The new model is mathematically grounded as it fills in missing data using samples from the same distribution as the observed credit spread data. The originality of the model comes from the manner in which it imputes the missing data; a unique link between the credit spread process and a driving exogenous factor counteracts the spreads being non-reactive, due to no trade, during times of credit crises.

There are a multitude of aspects and factors that need to be taken into consideration when developing

[1] Also known as Last Observation Carried Forward

any credit spread model. With the added complication of missing data, the scope of the task becomes unlimited. This research focusses on building a solid foundation that can be nuanced and grown, using core assumptions and market data. The following points detail assumptions that were made in this regard:

- In the empirical study, only the credit spreads of pure corporate vanilla bonds are analysed, but this limitation is not hugely significant as vanilla bonds still represent a significant share of the listed debt market.
- The modelling of the credit spreads does not take into account the term, seniority or level of guarantee of the bonds.
- The decomposition of the spreads into additional factors over and above credit, such as liquidity, interest rate risk, the states of the local and global economic environments and taxation, is not directly addressed.
- The framework is developed around a single exogenous factor.

The thesis is structured as follows:

- Chapter 2 investigates the degree of illiquidity in the South African corporate debt market and the subsequent impact on credit spreads, as well as an analysis into the general empirical nature of credit spreads,
- Chapter 3 introduces the new model which connects the credit spread to the exogenous information,
- Chapter 4 covers the problem of handling missing data whilst using the new model,
- Chapter 5 applies the new model with the missing data method to market bonds and
- Chapter 6 describes the interesting computational aspects encountered in building and generating the analysis.

2. Empirical Analysis of Credit Spreads

To model credit spreads, the empirical behaviour of observed time series plays a key role in identifying the most suitable direction to follow. This chapter concentrates on the analysis and understanding of historical credit spread data.

Firstly, credit spread indices from a variety of global markets and the credit spread of a US bond are investigated to justify the widespread assumption of mean reversion found in the literature. Then focus is placed on the South African bond market, where the lack of trade of corporate bonds is demonstrated using metrics developed to highlight the illiquidty in various ways. The detrimental effect of infrequent trade on the credit spread time series and term structures of the spreads is then explored. A set of the most liquid South African corporate bonds over the period from January 2005 to March 2018 are identified and their historical spreads analysed to affirm that it is reasonable to model credit spreads as mean reverting.

2.1 The Nature of Credit Spreads

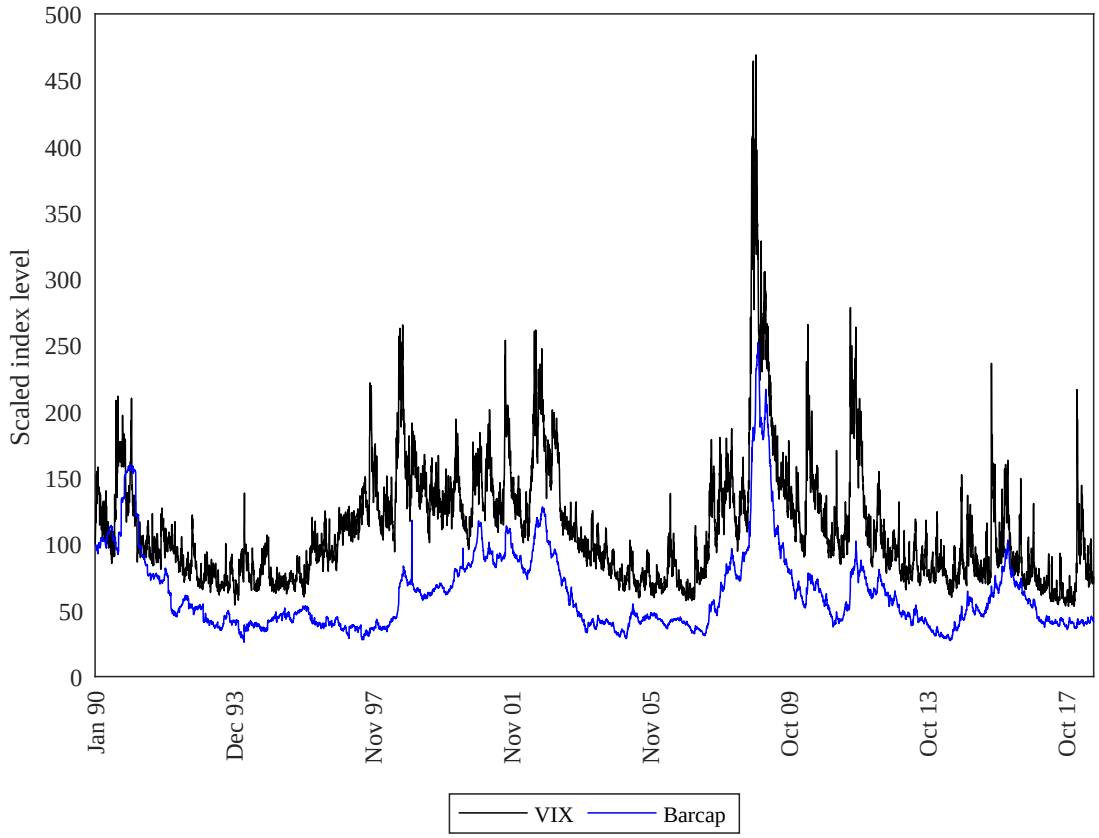


Figure 2.1: The VIX (implied volatility) and Barcap (credit spread) indices.

Credit spreads, like implied volatilities, are risk metrics. They incorporate premia from variety of systemic and idiosyncratic risk sources, such as liquidity, credit, tax and general market sentiment. They deflate when the risks in the underlying are low and increase when these risks increase. Figure 2.1 plots the CBOE Volatility index (VIX), which estimates the market implied volatility using options traded on the Standard

& Poor's (S&P) 500 index over roughly a 28 year period. Also shown is the Barclay's US Corporate High Yield Index (Barcap). The index measures the yields in developed market corporate bonds which have non-investment grade credit ratings. Both indices have been re-based to 100 for purposes of comparison. Figure 2.1 displays an observable degree of co-movement of the implied volatility index and the credit spread

	Correlation	Regression			Co-integration		
		β_0	β_1	Adjusted R^2			
Significance level					1%	5%	10%
Value	0.7478	4.8246	0.5182	0.5591	1	1	1
p-value		0.0000	0.0000		0.0010	0.0010	0.0010
Test statistic					-12.4492	-12.4492	-12.4492
Critical value					-3.9008	-3.3385	-3.0463

Table 2.1: Calculated correlation, regression and co-integration results from the analyses of the VIX and Barcap time series.

index. Table 2.1 shows that the correlation of the indices was 0.7748 and a linear regression of the Barcap index on the VIX yielded an adjusted R^2 of 0.5591 and low coefficient p-values. The indices were also co-integrated at the 1%, 5% and 10% significance levels.

Implied volatility is commonly modelled as a mean reverting process (in stochastic volatility models like Heston (1993) for example). The similarity in evolution between the VIX and Barcap indices gives merit to the assumption that credit spreads can be modelled as mean reverting as well. Indeed, the mean-reverting natures of credit spreads and default intensities have been well-documented and studied in literature and the usage of such processes is common practice. Reduced-form hazard rate models for modelling probability of default, such as those proposed by Carr *et al.* (2010), Duffie *et al.* (1999), Duffee (1999), Jaskowski *et al.* (2012) and Schönbucher (1997), commonly assume that the dynamics of the default intensity follow a mean-reverting square root Cox-Ingersoll-Ross (CIR) process (Cox *et al.* (1985)). Anderson (2008) chose to model the default process using the Ornstein-Uhlenbeck (OU) model (Uhlenbeck *et al.* (1930)) and found strong evidence of mean-reversion exhibited by the CDS spreads of North American and European Energy and Media sectors. A hybrid mean-reverting model for credit spreads and credit spread returns was introduced by O'Donoghue *et al.* (2014), where the Black–Karasinski model (Black *et al.* (1991)) was employed to model the credit spreads and OU models the spread returns. Prigent *et al.* (2001) also found significant mean-reversion in higher rated index credit spreads and suggest modelling these credit spread indices using OU with jumps. And the work of Bhanot (2005) elaborated on the drivers of the mean reversion observed in corporate bond index spreads.

Further examples of credit indices or spread time series are plotted in Figure 2.2 and are described as follows:

- CDX Emerging Market CDS 5 Year spread (CDX EM CDS 5Y): a Markit® index composed of fifteen sovereign entities across Latin America, Eastern Europe, the Middle East, Africa and Asia that participate in CDS market trading, with a term of 5 years.
- iTraxx European CDS 5 Year spread (ITRX EUR CDS 5Y): a Markit® iTraxx corporate CDS index consisting of 5 year CDS spreads of 125 investment grade entities within Europe.
- iTraxx Japan CDS 5 Year spread (ITRX Japan CDS 5Y): a Markit® iTraxx corporate CDS index consisting of 5 year CDS spreads of 40 liquid, investment grade entities in Japan.

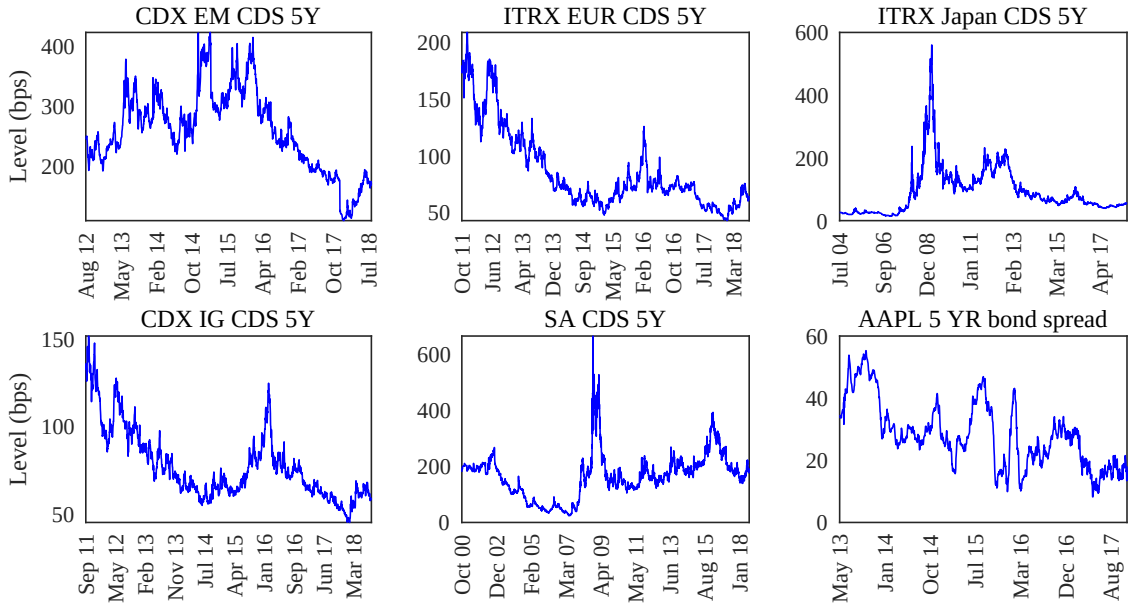


Figure 2.2: The evolutions of the CDS spread indices and the spread of the AAPL bond.

- iTraxx IG CDS 5 Year spread (CDX IG CDS 5Y): a Markit® iTraxx corporate CDS index consisting of 5 year CDS spreads of 125 of the most liquid, investment grade entities in North America.
- SA CDS 5 year spread (SA CDS 5Y): the 5 year CDS spread of the sovereign state of the Republic of South Africa. This value represents the aggregate credit spread of SA government bonds with a term to maturity of 5 years. Note that this oscillates from a low of 25 bps in June 2007 to a high of 6.6% at the time of the subprime crisis in late 2008.
- Credit spread of an Apple (with share price ticker AAPL) 5 year bond (AAPL 5YR bond spread): the yield of a 5 year Apple bond which matured in March 2018 taken over the 5 year US treasury yield.

Mean reversion by definition implies that a time series is stationary, meaning that the full joint probability distribution is not altered by a shift in time. The Augmented Dickey-Fuller (ADF) test assumes that the null hypothesis is the presence of a unit root in the time series, indicating that the time series is non-stationary (more detail in Appendix A.2.8). The ADF test is applied to the CDS index and bond spread time series and the outcome of these tests is shown in Table 2.2.

Time series	p-value	t-statistic	Critical value
BarCap US Corp HY10Y	0.2227	-2.1662	-2.8618
CDX EM CDS 5Y	0.2434	-2.1201	-2.8646
ITRX EUR CDS 5Y	0.1004	-2.5668	-2.8645
ITRX Japan CDS 5Y	0.1898	-2.2478	-2.8637
ITRX IG CDS 5Y	**0.0435**	**-2.9205**	**-2.8645**
SA CDS 5Y	**0.0480**	**-2.8799**	**-2.8633**
APPL 5Y bond	0.3058	-1.9791	-2.8647

Table 2.2: The results of the ADF test applied to the time series of the CDS spread indices and the spread of the AAPL bond. The red figures show the rejection of the null hypothesis in favour of stationarity.

The ITRX IG CDS 5Y and SA CDS 5Y indices proved to be stationary at the 5% significance level, with the ITRX EUR CDS 5Y index at the cusp of being stationary at the 10% level.

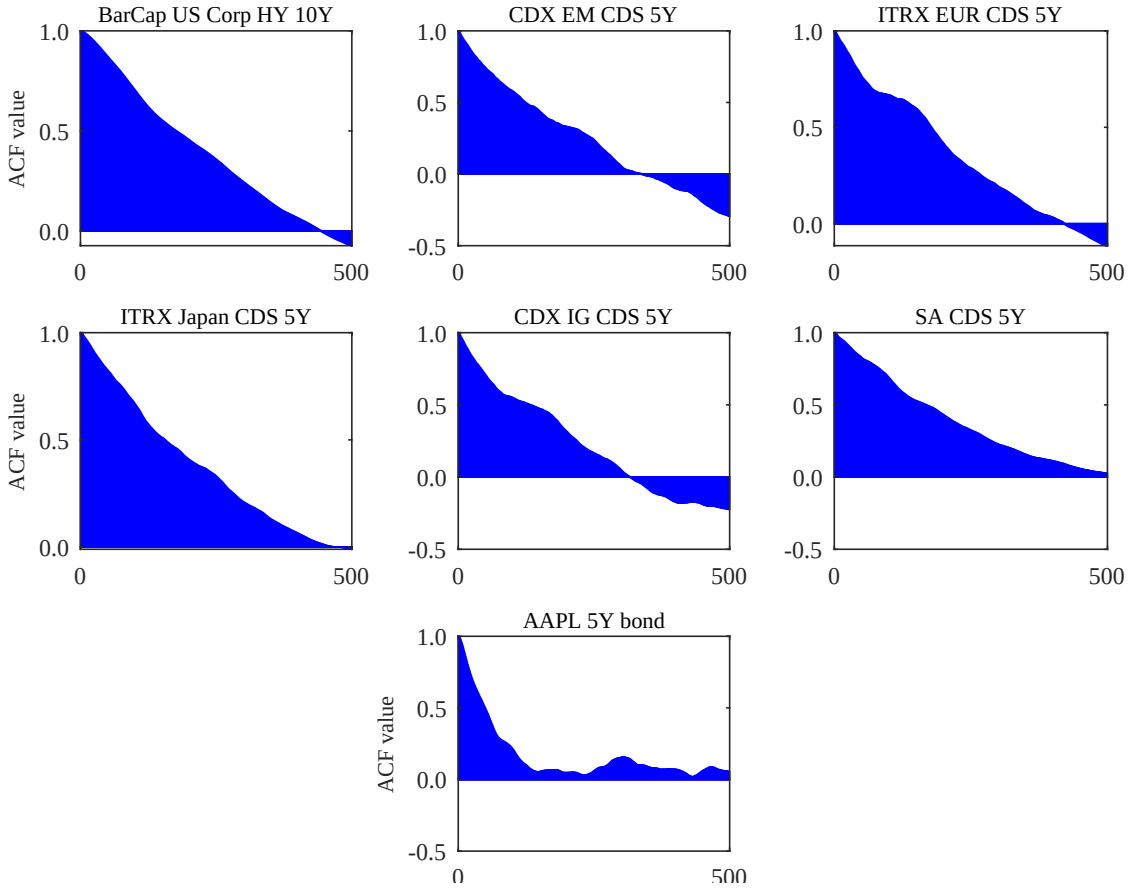


Figure 2.3: Autocorrelation function plots calculated from the time series of the CDS spread indices and the spread of the AAPL bond.

Mean reversion processes should also have quickly decaying autocorrelations. The values of the autocorrelation function (ACF), described in more detail in Appendix B.6.3, for the 7 spread time series are displayed in Figure 2.3. The ACF values for all the spreads decrease relatively rapidly and all attain a zero value before the 500^{th} lag.

Now, consider the following stochastic processes, defined by the accompanying SDEs, to model a credit spread denoted $x(t)$:

- The Ho-Lee model with constant drift (HLCD),

$$dx(t) = \mu dt + \sigma dW^{\mathbb{P}}(t),$$

  where the traditional Ho-Lee model is given by

$$dx(t) = \mu(t)dt + \sigma dW^{\mathbb{P}}(t).$$

- The Rendleman-Bartter (RB) model, which is geometric Brownian motion [1] (GBM) for interest rates,

$$dx(t) = \mu x(t)dt + \sigma x(t)dW^{\mathbb{P}}(t).$$

[1] See Appendix A.2.5 for further details on geometric Brownian motion.

- The OU model (detailed in Appendix A.2.6),

$$dx(t) = \theta(\mu - x(t))dt + \sigma dW^{\mathbb{P}}(t).$$

- The Lognormal OU model,

$$d\ln x(t) = \theta(\mu - \ln x(t))dt + \sigma dW^{\mathbb{P}}(t).$$

All of these models are widely used in interest rate modelling. Observe that the latter two models are mean-reverting, whilst the first two are not.

These models are calibrated to the 7 time series of credit spreads using Maximum Likelihood Estimation (MLE). The resultant parameter estimates, maximum likelihood values and corresponding values of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), given by equations (A.66) and (A.68) in Appendices A.2.10 and A.2.11, are shown in Table 2.3.

Time series	Model	μ	θ	σ	Maximum $\mathcal{L}$ ($\times 10^3$)	AIC ($\times 10^3$)	BIC ($\times 10^3$)
BarCap US Corp HY 10Y	HLCD	0.0004		1.9231	5.6835	-11.3631	-11.3491
	RB	1.7995		1.9232	5.6824	-11.3607	-11.3467
	OU	4.9553	0.2953	1.9237	5.6859	-11.3658	-11.3448
	Log OU	1.5123	0.3636	0.3470	**7.0607**	**-14.1154**	**-14.0943**
CDX EM CDS 5Y	HLCD	0.0001		128.6324	-5.1017	10.2075	10.2181
	RB	8265.8387		128.6742	-5.0987	10.2014	10.2120
	OU	250.5938	1.6909	128.8622	-5.0995	10.2049	10.2208
	Log OU	5.4648	1.2070	0.4645	**-4.9411**	**9.8882**	**9.9041**
ITRX EUR CDS 5Y	HLCD	0.0002		47.0821	-4.2905	8.5849	8.5958
	RB	1092.6317		47.0852	-4.2881	8.5802	8.5910
	OU	74.2057	1.3243	47.1045	-4.2868	8.5796	8.5959
	Log OU	4.2668	1.1953	0.4758	**-3.9314**	**7.8688**	**7.8851**
ITRX Japan CDS 5Y	HLCD	2.0098		88.4515	-11.2199	22.4438	22.4562
	RB	3914.9204		88.4638	-11.2173	22.4385	22.4509
	OU	95.6052	0.6939	88.5107	-11.2174	22.4407	22.4593
	Log OU	4.3827	0.3540	0.6051	**-8.5452**	**17.0963**	**17.1149**
CDX IG CDS 5Y	HLCD	0.0001		32.9216	-3.7102	7.4244	7.4353
	RB	531.2981		32.9242	-3.7082	7.4204	7.4313
	OU	71.6881	1.8346	32.9532	-3.7056	7.4172	7.4335
	Log OU	4.2517	1.7097	0.3964	**-3.5332**	**7.0724**	**7.0888**
SA CDS 5Y	HLCD	0.6967		111.3467	-15.0398	30.0836	30.0964
	RB	6200.7695		111.3591	-15.0370	30.0779	30.0907
	OU	165.3508	0.9371	111.4503	-15.0357	30.0773	30.0966
	Log OU	4.9659	0.3639	0.5021	**-13.0667**	**26.1394**	**26.1587**
AAPL bond	HLCD	0.0000		0.2048	3.3505	-6.6971	-6.6870
	RB	0.0000		0.2049	3.3472	-6.6903	-6.6803
	OU	0.2622	1.7915	0.2052	**3.3526**	**-6.6991**	**-6.6840**
	Log OU	0.0002	0.1483	0.9882	3.0812	-6.1563	-6.1412

Table 2.3: Calculated parameter estimates, maximum likelihoods, AIC and BIC values obtained through fitting the models to each spread series. The numbers in red indicate the maximum likelihood and smallest AIC and BIC attained for the time series.

The standard errors of the parameter estimates are shown in Table 2.4 for the purposes of determining confidence intervals if required.

Index	Model	μ	θ	σ
BarCap US Corp HY 10Y	Merton	0.0013	-	0.0150
	RB	0.0212	-	0.0150
	OU	0.0276	0.0592	0.0150
	Log OU	0.0147	0.0614	0.0002
CDX EM CDS 5Y	Merton	0.2126	-	2.3862
	RB	3.3757	-	2.3870
	OU	1.8383	0.2019	2.3904
	Log OU	0.5245	0.1471	0.0289
ITRX EUR CDS 5Y	Merton	0.0717	-	0.8046
	RB	1.1380	-	0.8047
	OU	0.6995	0.1947	0.8050
	Log OU	0.2786	0.1826	0.0080
ITRX Japan CDS 5Y	Merton	0.0932	-	1.0458
	RB	1.4791	-	1.0459
	OU	1.2563	0.1961	1.0465
	Log OU	0.1943	0.0966	0.0101
CDX IG CDS 5Y	Merton	0.0499	-	0.5602
	RB	0.7923	-	0.5602
	OU	0.4140	0.1577	0.5607
	Log OU	0.2776	0.1353	0.0057
SA CDS 5Y	Merton	0.1049	-	1.1780
	RB	1.6662	-	1.1782
	OU	1.2180	0.1442	1.1791
	Log OU	0.3842	0.0799	0.0518
AAPL bond	Merton	0.0004	-	0.0043
	RB	0.0061	-	0.0043
	OU	0.0032	0.3200	0.0043
	Log OU	0.0011	0.0252	0.0112

Table 2.4: Standard errors of calculated parameter estimates.

The mean reversion estimates in Table 2.3 appear to be centred around unity. The higher the value of θ, the greater the degree of mean reversion, implying a tighter oscillation around the equilibrium level. θ levels closer to 0 indicate a data series that is dominated by its stochastic nature.

For the the CDS spread time series, the Log OU model consistently attained the largest likelihood value in the parameter estimation and subsequently produced the lowest AIC and BIC levels. The OU model achieved the best fit to the time series of the credit spread of the AAPL bond, with the the highest maximum likelihood and smallest AIC and BIC values. The OU and Log OU models are penalised by the AIC and the BIC for having 3 model parameters to estimate as opposed to 2, as HLCD and RB do. Still, these mean reverting models proved to be best suited to the spread time series. These results indicate that there is a degree of mean reversion in empirical credit spread time series, with either normal or lognormal distributions.

From this exploratory investigation into a range of market credit spreads (indices or market bonds), it was found that there is some evidence of a mean reverting nature. Tests for stationary time series and autocorrelations supported, whilst not all of the time, the hypothesis of mean reversion. Mean reverting

models also showed to fit the data better than their non-mean reverting counterparts. Based on this analysis and bolstered by literature, it would not be unreasonable to model credit spreads as some kind of mean reverting process.

2.2 The South African Case

The aim of this section is to:

- illustrate the degree of the illiquidity in the South African corporate pure vanilla bond market,
- demonstrate the negative impact of the low liquidity on the credit spreads of the bonds,
- and investigate the statistical nature of the credit spreads of the most liquid of these bonds.

2.2.1 Data Description

JSE daily bond turnover and mark-to-market data, as well as the contractual static data required for pricing, was sourced from the beginning of January 2005 until 29 March 2018 and provided by Prescient Securities. Daily government zero coupon yield curves for this period were calculated by the JSE and provided by Prescient Securities. All equity share price data was sourced from Bloomberg, courtesy of Sygnia Asset Management.

The investigation considers pure vanilla corporate bonds, which fulfilled the following criteria:

- issued by a single non-government entity
- nominal in nature
- semi-annual coupon payments
- fixed coupon rates
- no callable features
- no early redemptions or split maturities
- no inflation linked features
- no credit linked features
- no asset swap features

The bonds are classified in terms of industry (sector), where the issuer of each bond is assigned an industry classification under the Industry Classification Benchmark (ICB) scheme. For issuers that were listed on the JSE, their existing ICB classification was used.

For non-listed issuers, the main revenue generating business of the entity was researched and an industry classification assigned accordingly.

The rationale behind limiting the analysis to those of pure vanilla fixed rate bonds comes from the sake of brevity in this study and from the scope of the types of bonds that were traded over the time period. Bonds with a floating coupon rate, known as floating-rate notes (FRNs), have become increasingly popular in the last few years. Future research analysing FRNs as well might provide even richer insight into the illiquidity phenomenon.

2.2.2 Determining Liquidity and Credit Spreads

2.2.2.1 Measuring Liquidity

Bid-offer spreads are typically used to estimate the liquidity of securities (as done in Fleming (2003) and Lybek *et al.* (2002)), with narrowing spreads indicating more trade and vice versa. But historical bid-offer spreads for listed debt in South Africa are not recorded on any data platform and as such a history of such spreads would need to be sourced directly from relevant market-makers.

Even for one of the most liquid government bonds, the R186, Pitsillis and Taylor (2015) reported that bid-offer spread data assimilation was difficult due to the fact that 80% of its trade occurred on a request-for-quote basis, whereby a broker would provide an informal quote of the bid-offer spread via electronic chat or telephone. Even if bid-offer spreads were indeed available, Bao *et al.* (2011) finds that the illiquidity of US corporate bonds exceeds the levels implied by bid-offer spreads.

Abdi *et al.* (2017) develop a model to estimate bid-offer spreads of instruments when liquidity in these market drops. Dick-Nielsen *et al.* (2012) and Schestag *et al.* (2016) also propose more intricate techniques to ascertain levels of liquidity and the drivers thereof. For the purposes of this study, where lack of trade is prominent, simpler approaches are more than sufficient in highlighting the degree and depth of trade.

Thus, the following quantitative measures (see Appendix D.4 for calculations) will be used:

1. Trade Frequency Ratio (TFR): the ratio of trade days over the term of the bond.
2. Liquidation Measure (LM): the number of days it would take to liquidate an assumed nominal amount based on historical trade data.
3. Zero Trade Ratio (ZTR): on an issuer level, the number of bonds that did not trade at all relative to the number of bonds issued by the issuer.

The measures estimate how often bonds traded, the volume of trades through time and the liquidity of issuers as a whole, respectively.

2.2.2.2 Calculating Credit Spreads

The theoretical definition of a credit spread is the difference between an instrument's yield and the yield of the government equivalent with the same term to maturity, coupon pay dates and coupon size. Credit spreads can be calculated using different approaches, resulting in different spreads. These are:

1. Compare the closing yield of corporate bond with that of its companion bond (a government issued bond with ideally the same maturity, coupon pay dates and coupon size as its corporate counterpart). The difference in these two yields serves as a credit spread, which will be called the quoted spread. The JSE employs this method to determine the daily spreads published.
2. Calculate the constant value that needs to be added to the government zero coupon bond yield curve in order for the sum of the discounted cash flows using the shifted government curve to match the market quoted bond price on that day. This value is called the Z spread.

3. Calculate the yield of the risk-free version of the corporate bond. The difference between the yield of the corporate and its risk-free counterpart will be called the nominal spread. This method synthesises the scenario typical of more developed markets, where each corporate bond has an identical (or nearly) government counterpart.

For specific detail on the different credit spread calculations, see Appendix D.5.

2.2.2.3 Comparison of the Different Credit Spread Calculation Methods

The quoted, Z and nominal spreads are all subject to disadvantages in usage and interpretation.

For the quoted spread, the government companion bond may not have the same maturity and cash flow dates nor coupon size as the corporate bond. The spread captures this duration mismatch (different coupon pay dates and maturity, or different coupon rates) and as such does not reflect the true spread. In markets that are more developed and fluid, it may be possible to find a government bond that has the cash flow specifications for every corporate bond, but unfortunately this is not the case in the South Africa.

The Z spread determines the credit spread directly from the government zero yield curve, thus avoiding any mismatch in duration, but is subject to the quality of the construction of that curve. In addition, the Z spread assumes that relative level of the credit spread against the government curve is constant i.e. that the term structure of credit spreads is flat.

The nominal spread makes no assertions about the shape of the credit spread term structure, only about the spread level at a certain term, but is also dependent on the accuracy of the zero curve in determining the risk-free price.

	IBL46	NBK2A	ABS3
Coupon rate (%)	7.25	10.55	8.45
Coupon frequency	2	2	2
Maturity	15 Jan 2020	15 Sep 2015	03 Apr 2011
Companion bond	R207	R157	R153
Coupon rate (%)	7.25	13.5	13
Coupon frequency	2	2	2
Maturity	15 Jan 2020	15 Sep 2015	31 Aug 2010

Table 2.5: Maturity, coupon rate and frequency of IBL46, NBK2A and ABS3 and their government companion bonds, as per JSE data.

Table 2.5 shows the cash flow details (maturity date, coupon rate and frequency) of IBL46, NBK2A and ABS3 (Investec, Nedbank and ABSA bonds), and their companion bonds. Figure 2.4 plots the quoted, Z and nominal spreads (all NACS) for these 3 bonds. These bonds were chosen to illustrate the drawbacks of the various spreads as well as understand the differences between them. In particular, the degree to which the duration mismatch between a bond and its companion impacts the quoted spread will be demonstrated.

There was no mismatch in duration between IBL46 and its companion bond R207, as they had the same coupon schedules and maturity. Thus the quoted and nominal spreads should have been the same, as the theoretical risk-free bond used in the nominal spread calculation would have been the same as the R207 in

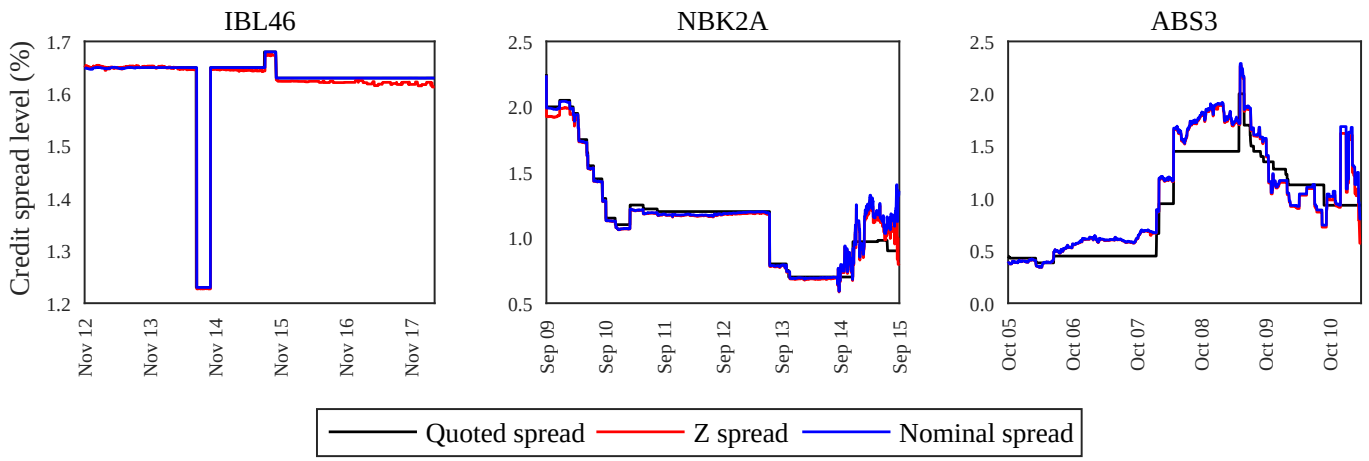


Figure 2.4: The Quoted, calculated Z and calculated nominal spreads of IBL46, NBK2A and ABS3, plotted from the date of issuance until maturity. If the bond was issued/matured before/after the period under analysis, the range of the plot started/ended at the first/last date in the period.

terms of cash flow profile. Thus it is expected that the all-in price and yield of the theoretical risk free bond would have equalled those of the R207. The plot of for IBL46 in Figure 2.4 confirms that the quoted spread and the nominal spread were nearly identical for the period plotted. The Z spread is slightly less accurate than the nominal spread.

NBK2A and its companion, R157, had the same maturity and hence cash flow schedule, but were mismatched on the coupon size by 2.95%. Thus even though they matched on term, the lower coupon rate of NBK2A would result in a lower all-in price and higher yield for the risk-free bond than the R157. The yield difference between NBK2A and its risk-free counterpart would be smaller as a result, giving a reduced credit spread. This deviation is shown in the NBK2A plot in Figure 2.4, where the nominal spread is only marginally below the quoted spread until August 2014.

From August 2014, this bond also displays an interesting example of the dependence of the Z and nominal spreads on the yield curve and the inputs used in its construction. Up until 21 August, NBK2A's companion bond was the R157 and this bond was also included as an input for the yield curve construction. On 22 August, the R159 (the longer dated R157 split leg with maturity 15 September 2016) replaced the R157 as NBK2A's companion bond and also replaced it as the input into the yield curve generation. Thus the quoted spread was completely mismatched in terms of duration.

The yield curve was no longer using a bond with the same term as NBK2A in its construction, thus a full duration mismatch was also introduced to the the nominal and Z spreads. On 29 August, the R158 (the middle leg of the split with maturity the same as NBK2A) replaced the R159 as the companion bond for NBK2A, but the R159 remained as the input for longer term point on the yield curve. Hence the mismatch was still inherent in the nominal and Z spreads until NBK2A matured in September 2015, but the quoted spread was only mismatched on coupon size. The duration mismatches during this time lead to the higher path errors between the calculated and quoted spreads.

The companion to ABS3 until 11 August 2009 was R153, which matured 7 months before its corporate counterpart, resulting in a term mismatch of 0.589 years. There was also a large difference of 4.55% in the

coupon rate. The R154, which had the same maturity as R153, became the companion bond of ABS3 on 11 August and was replaced by the R155, whose maturity was 31 August 2011, on 31 August 2010. Thus there was a duration mismatch throughout the entire bond's life time. This disparity manifested in the Z and nominal spreads being notably different to the quoted spread in absolute level and degree of fluctuation, evident in the ABS3 plot Figure 2.4. ABS3 is a telling example of the effect the dissimilarity between a corporate bond and its government companion can have on the estimates of its credit spread.

Since the nominal spread produced the expected results in the case of IBL46, it is the preferred calculation methodology of the credit spread, as the stability and quality of the input curve is a variable factor that has the potential to be improved. The Z spread assumes a constant credit spread term structure and the quoted spread is subject to the the depth of government issuances. The nominal spread will be calculated and used as the credit spread for the bonds in all investigations to follow.

2.2.3 Assessing Corporate Bond Liquidity

Between January 2005 and March 2018, 404 instruments fulfilling the pure vanilla corporate bond criteria were listed on the JSE (at one point or another). These instruments were issued by 86 different entities spanning a range of industries, terms, guarantee types and capital structure placements.

Table 2.6 shows the variety of issuers of pure vanilla corporate bonds across the different ICB industry classifications, as well as the number of bonds in each industry and the total nominal amount issued for these bonds in billions of South African Rands (ZAR bn). This table gives a glimpse of the trading activity and helps to identify differing levels of liquidity across the market.

Industry	Number of issuers	Number of bonds	Total Issued Nominal (ZAR bn)
Basic Materials	9	13	11.8310
Consumer Goods	7	15	12.5180
Financials	17	179	137.2364
Health Care	1	2	0.6250
Industrials	10	30	18.1910
Oil and Gas	1	1	1.0000
Telecommunications	1	5	10.3000
Utilities	1	1	0.3800
Special Purpose Vehicles	8	21	7.5774
Parastatals	13	95	286.0736
Municipalities	5	20	20.4773
Foreign Basic Materials	1	1	0.6957
Foreign Financials	7	13	10.0046
Foreign Government	3	6	2.8935
Foreign Telecommunications	1	1	0.0300
Foreign Utilities	1	1	0.5000
Total	86	404	520.3335

Table 2.6: Pure vanilla bonds per assigned industry: number of issuers, issuances and total nominal issued as calculated from JSE data.

The industries with the largest number of issuers and issuances were Parastatals and Financials, with total

nominal amounts issued being in the hundreds of billions. Both these industries are characterised by entities that have stable business models and so have the ability to issue large amounts of debt. The remaining industries with noteworthy sizes were Industrials, Special Purpose Vehicles (SPVs), Basic Materials, Consumer Goods, Telecommunications and Municipalities.

Pure vanilla bonds from the first 8 industries in Table 2.6 will be investigated, excluding Parastatals and Municipalities (as these are either wholly or partly owned by the South African government), Special Purpose Vehicles and any foreign issuers. The first 8 industries issued 246 pure vanilla bonds from 47 issuers with a total nominal of ZAR 192.0814 bn. Details of these 246 bonds are given in Table D.5 in Appendix D.

Figure 2.5 shows the trade frequency ratios for all 246 bonds, ranked from highest to lowest, against the liquidation measure, total traded nominal issued and total nominal traded. The most frequently traded bond according to TFR was a Mobile Telecommunications Networks Holdings or MTN bond, MTN01, with a TFR of 62.64%. There were 8 bonds that did not trade at all during the period under analysis, even though they were all alive for a minimum of two years from between 2005 to 2018. The issuers of these bonds were FirstRand Bank, Investec Bank, Kagiso Sizanani Capital, Nedbank, Northham Platinum and Standard Bank (SSN006).

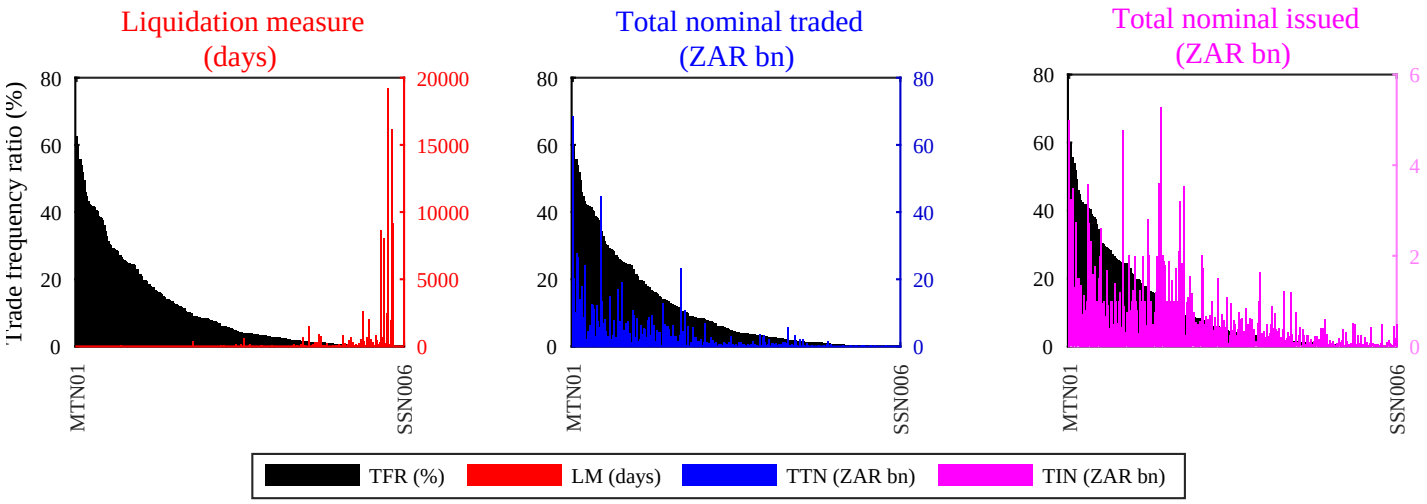


Figure 2.5: Trade frequency ratio (TFR) ranked from highest to lowest plotted against the liquidation measure (LM), total traded nominal (TTN) and total issued nominal (TIN) for the 246 vanilla corporate bonds.

The first and second plots in Figure 2.5 show that generally speaking, the bonds with high trade frequency ratios corresponded to larger volumes of trade, with lower liquidation measure values and larger total nominal traded levels. There didn't seem to be the such a significant link between frequency of trade and the total nominal issued shown in the final plot. Bonds with similar issuance sizes to the most liquid bonds showed much reduced tade frequency and volume.

On an issuer level, Figure 2.6 plots the average LM with the average TFR, the number of bonds issued by each entity as well as the entity's zero trade ratio (ZTR). The highest average TFR of 36.59% was achieved by MTN and had an average LM of 12 days. Harmony Gold Mining obtained the lowest LM of 3 days, with an average TFR of 25.14%. Considering the number of bonds issued, the 5 big banks (ABSA, FirstRand,

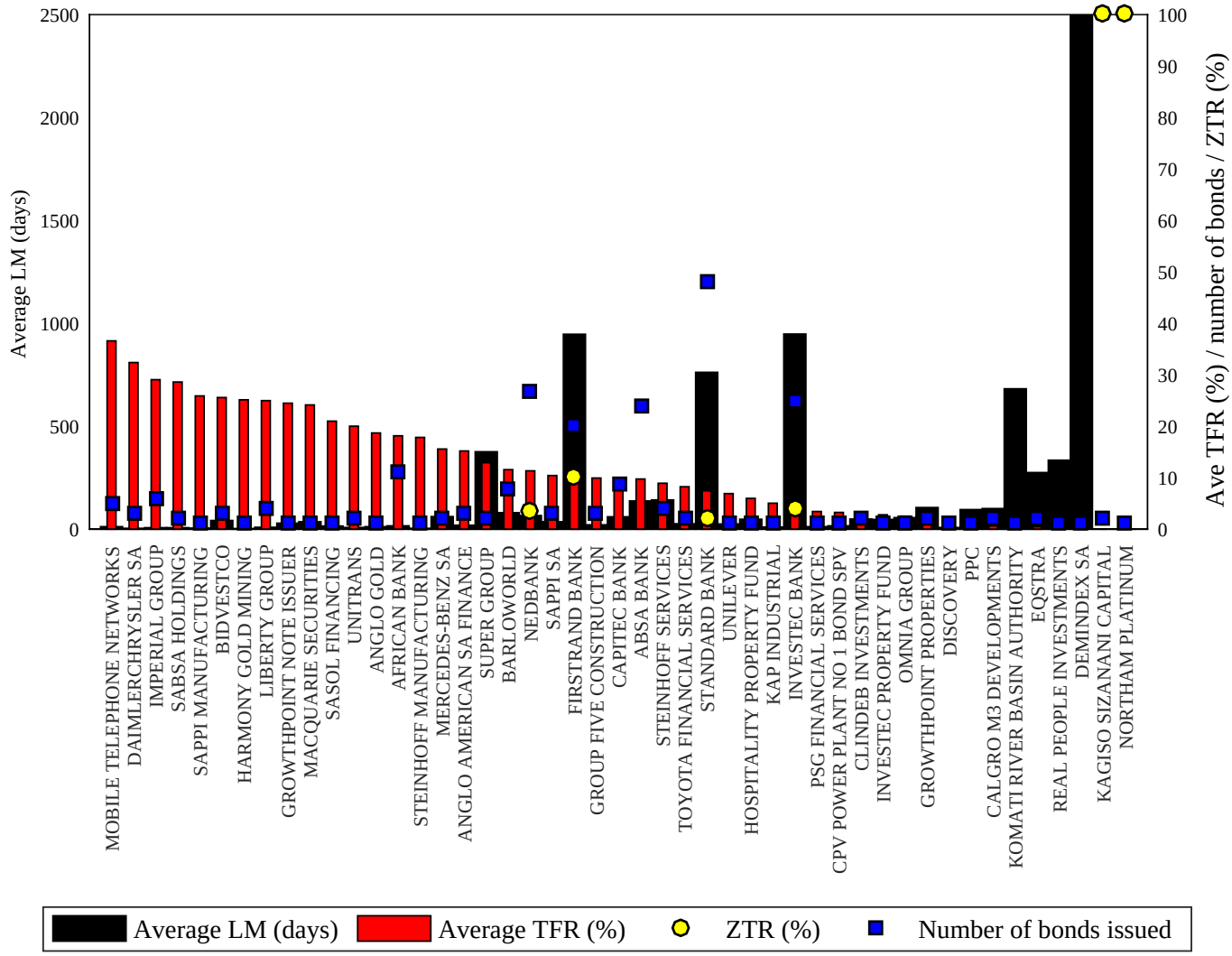


Figure 2.6: The calculated average LM and TFR of the bonds issued by the 47 entities. The blue squares represent the number of bonds each entity issued and the yellow dots are the ZTR of each issuer.

Investec, Nedbank and Standard Bank) all issued over 20 bonds, with the highest number (48) issued by Standard Bank. The next highest number of bonds issued, after the 5 big banks, was those of African Bank and Capitec Bank, which issued 11 and 9 bonds respectively.

Four of these banks with issuance sizes over 20 bonds had positive ZTR levels, meaning that a certain percentage of the bonds issued were not traded at all during the period. The accompanying average LM for three of those issuers (FirstRand, Investec, and Standard Bank) is significantly large. None of the bonds issued by the 2 left most entities, Kagiso Sizanani (issued 2 bonds) and and Northam Platinum (issued 1 bond), traded at all. Hence they have a ZTR of 100%, a zero average TFR. Calculation of LM in these cases is nonsensical, hence not depicted in the figure.

Table 2.7 presents the top 15 most liquid bonds according to their trade frequency ratios. MTN01 not only had the highest TFR but the highest TTN as well and an LM of only 1 day. NBK2A had the next highest TFR, but also a higher LM of 3 days and lower TTN. All of the bonds were traded at least 40% of the time, but some not in large size, resulting in some high liquidity measures (such as IPL4 and ABL8A

Bond	Issuer	Industry	TFR (%)	LM (days)	TTN (ZAR bn)	TIN (ZAR bn)
MTN01	MTN	Telecommunications	62.64	1	6,867.68	5.00
NBK2A	Nedbank	Financials	59.80	3	2,023.58	3.24
BID01	Bidvestco	Industrials	55.71	6	1,026.23	1.50
SBS1	Standard Bank	Financials	55.68	2	2,772.21	3.50
MTN02	MTN	Telecommunications	53.98	3	2,660.58	1.30
SBS9	Standard Bank	Financials	51.94	5	1,422.84	2.75
IPL4	Imperial Group	Industrials	49.40	16	370.39	1.50
IPL6	Imperial Group	Industrials	45.95	4	1,789.10	1.50
NBK3A	Nedbank	Financials	44.89	9	870.46	1.27
ABS3	Absa Bank	Financials	43.14	2	2,431.87	1.36
ABL8A	African Bank	Financials	42.53	20	212.93	0.73
NBK9A	Nedbank	Financials	42.00	9	464.00	1.14
CBL11	Capitec Bank	Financials	41.92	11	413.37	0.80
LGL02	The Liberty Group	Financials	41.84	7	637.93	1.00
FRX15	FirstRand Bank	Financials	41.40	5	1,275.31	3.58

Table 2.7: The 15 bonds found to have the highest trade frequency ratios (TFR), as well as their issuer, industry, liquidation measure (LM), total traded nominal (TTN) and total issued nominal (TIN).

with 16 and 20 days respectively).

The Financials sector was an area of high trade activity, with two thirds of the top 15 most liquid bonds originating from that industry. Banks in particular dominated, with 9 of the bonds being issued by either Nedbank, Absa, FirstRand or Standard Bank. MTN showed another liquid bond, MTN02, that traded frequently and in significant size, with a TFR of 53.98% and LM of 3 days.

2.2.4 Infrequent Trading and Credit Spread Evolutions

How did the general lack of liquidity shown in Section 2.2.3 impact the credit spreads of these bonds? Credit spreads only change to reflect current market sentiment of credit worth when bonds are traded. Little or no trade implies that credit spreads are not being adjusted to include the current outlook on credit worth. Lesmond *et al.* (1999) postulate that the reason for infrequent trading is that there is no current meaningful information about the bond or issuer that justifies the cost of transacting and that trade is only reasonable when when enough cumulative fundamental data has collected. With the most liquid vanilla corporate bond in South Africa only trading about 62% of the time, it can be argued that most of the credit spread time series observed and used in the corporate vanilla debt market are not reliable measures of credit worth, as large portions of these time series are out of date.

To demonstrate how infrequent trading leads to stale credit spreads, Figure 2.7 shows the LVCF evolutions of the nominal credit spreads of the 4 most liquid bonds as per TFR in Table 2.7, over the final 6 months of their lives. The 6 month period was chosen merely for a clear graphical display. The days on which the bonds traded during these periods are indicated by the vertical grey lines and the credit spread levels resultant from these trades are the red dots.

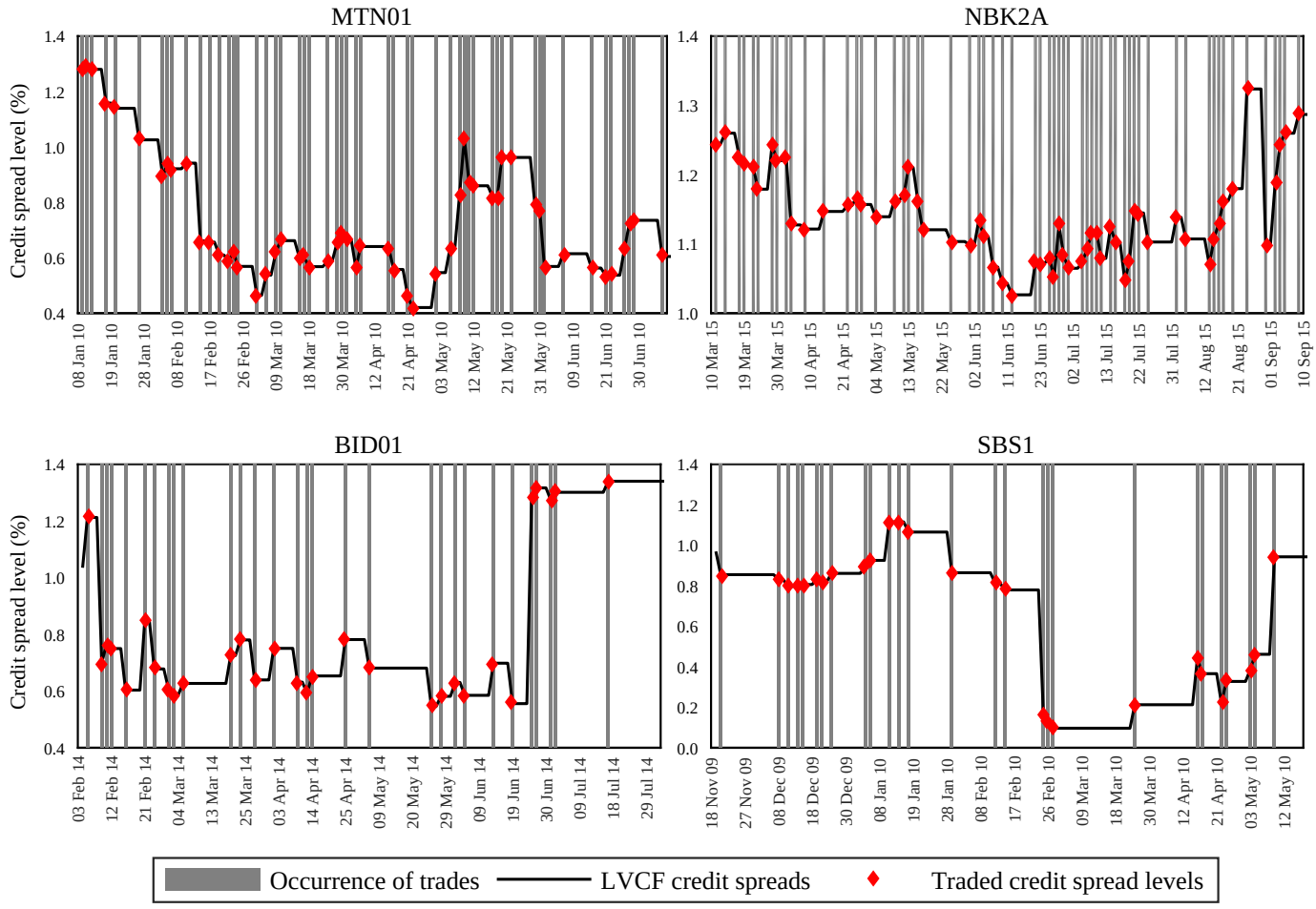


Figure 2.7: The LVCF credit spread time series (black lines), days of trade (vertical grey lines) and market observed credit spreads (red dots) of MTN01, BID01, NBK2A and SBS1 for the final 6 months before their maturity.

There were many distinct periods of zero trade for all the bonds, indicated by the large spaces between the of vertical grey lines. If the bond traded everyday, the grey lines would be consecutive and equally spaced. During all of these periods of no trade, the credit spreads reflecting the true credit worth of the bonds were unknown. The LVCF time series showed credit spreads that stayed constant at the previous traded levels. When subsequent trades occurred, the credit spreads were often observed to have jumped significantly to the new traded level.

Such sudden, drastic variations imply that during the periods where the bonds were not traded, the market perceived risk of the bonds or their issuers was changing, but these changes only manifested in the credit spreads when the bonds traded again. Given that the credit spreads of the 4 most liquid bonds were compromised by illiquidity, it follows that the spreads for the remaining bonds were even less reliable and explanatory as credit risk measures, since they experienced lower levels of trade.

2.2.5 Infrequent Trading, Credit Spread Term Structures and Seniority

A further consequence of the illiquidity of corporate bonds can be observed in the irregularity of the credit spread term structures derived from the LVCF time series, as well as anti-intuitive relative LVCF credit spread levels in terms of seniority. FirstRand Bank and Nedbank will be used as examples as they were the

only issuers that had 5 bonds alive at the same time for sufficiently long periods.

The LVCF credit spreads of the 5 FirstRand bonds are shown in Figure 2.8, for the period beginning with the oldest bond's issuance date until March 2018. The pertinent bond details are given in Table 2.8 (with term to maturity as at mid February 2014, the first day on which all 5 bonds were live). The table is ordered by term to maturity. Since all bonds were senior, no investigation on seniority was not possible.

Common consensus is that the term structure of credit spreads for higher quality issuers should be upward sloping and lower quality issuers should be downward sloping, even humped. Bedendo *et al.* (2007) and Trück *et al.* (2004) showcased this difference in term structure shapes between investment grade and non-investment grade bonds in the American and European debt markets respectively. Fons (1994) made assertions on the differences in the slope of the term structure between lower and higher credit quality firms, saying a lower credit quality implies wider spreads that decrease as maturity increases and a higher credit quality usually means narrow spreads in the short term that increase with term.

Bond	TFR (%)	Seniority	Term to mat. (years)
FRX18	8.93	Senior	4.16
FRX19	38.08	Senior	5.75
FRX24	9.97	Senior	10.83
FRX26	24.83	Senior	12.64
FRX31	14.08	Senior	17.03

Table 2.8: Calculated TFR, seniority and term to maturity (as at mid February 2014) of the FirstRand bonds.

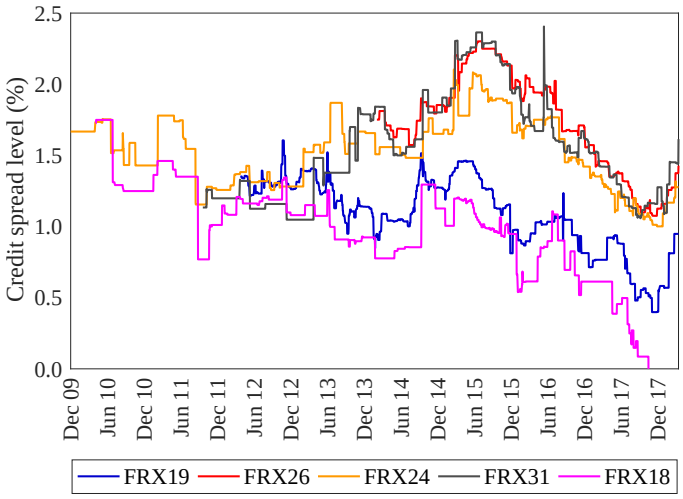


Figure 2.8: LVCF credit spreads of the FirstRand bonds from December 2013 to March 2018.

The overall levels of the credit spreads shown in Figure 2.8 were in line with the expectation of an upward sloping term structure; the bonds with shorter term to maturity (FRX18 in pink and FRX19 in blue) had the lowest credit spreads most of the time, with the longest dated bonds (FRX26 in red and FRX31 in black) having had the highest spreads for most of the period.

But there were instances where illiquidity caused disruptions. Term structure plots for February 2014 and March 2015 in Figure 2.9 show curves produced from the bonds' spread levels that are reasonable and continuously upward sloping.

However, in September 2014, the spreads of the shorter dated, more frequently traded bonds, FRX19 and FRX26, increased above the levels of FRX24 and FRX31 spreads, their respective longer dated but less liquid bonds. Because they were traded around this time, the spreads of FRX19 and FRX26 reflected the increase in uncertainty of South African banks as a result of the African Bank crisis in August 2014. But the spreads of FRX24 and FRX3 did not change accordingly until the bonds were traded weeks later. The

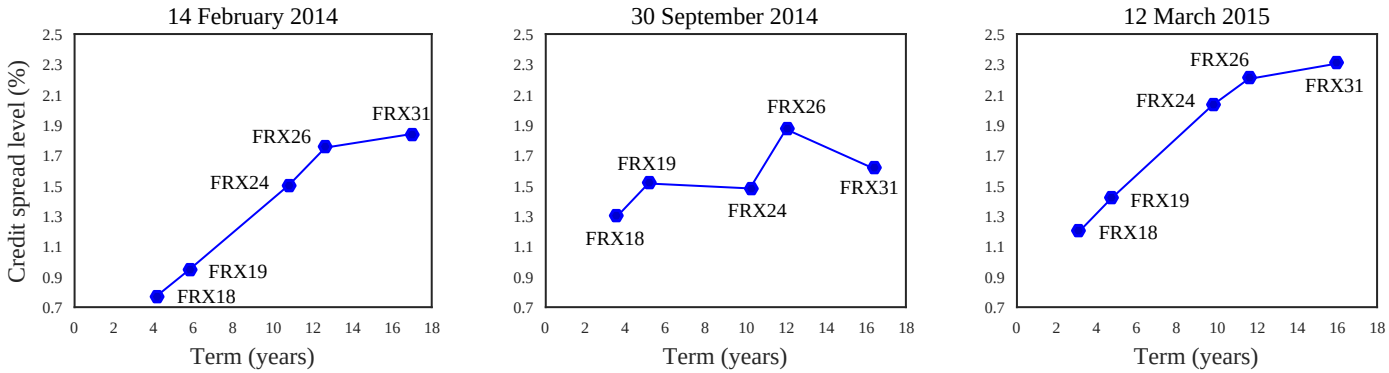


Figure 2.9: Term structures constructed from the LVCF credit spreads of FirstRand bonds as at 14 February 2014, 30 September 2014 and 12 March 2015.

result was a double humped curve with dampened 10 and 17 year points.

Replicating the analysis for Nedbank, the details of the bonds chosen are shown in Table 2.9 and their LVCF credit spread evolutions over the period beginning with the oldest bond's issuance (NBK3A on 9 September 2009) until 12 March 2015 plotted in Figure 2.10.

Bond	TFR (%)	Seniority	Term to mat. (years)
NBK6A	12.55	Senior	1.39
NBK2A	59.80	Subordinate	1.79
NBK9A	42.00	Senior	2.32
NBK3A	44.89	Subordinate	5.78
NBK11A	13.57	Senior	7.00

Table 2.9: Calculated TFR, seniority and term to maturity (as at end November 2014) of the Nedbank bonds.

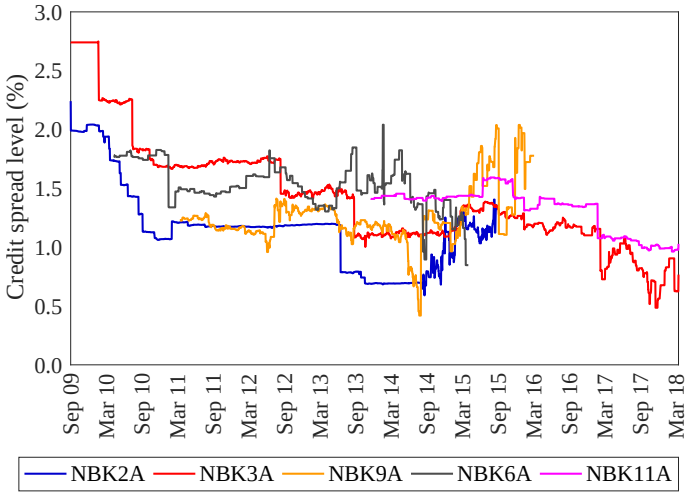


Figure 2.10: LVCF credit spreads of the Nedbank bonds from September 2009 to March 2018.

Following the reasonable assumption that Nedbank was a firm of relatively good credit quality at the time, one would expect that the shorter dated bonds (NBK6A and NBK2A) to have had consistently lower credit spreads than the longer dated bonds (NBK3A and NBK11A). However, Figure 2.10 shows that NBK3A (in red) did not have a consistently high credit spread, with a significant tightening of its spread being exhibited from August 2012 onwards. Similarly, NBK6A had the highest credit spreads for a significant period (June 2013 to June 2014), even though it was one of the shorter dated bonds.

Conversely, an assumption of poor credit quality for Nedbank implies a downward sloping term structure for the period. Thus the spreads of the short term bonds should have generally been higher than those of longer dated bonds. Again, there are instances where this behaviour was not observed. From January 2011 to June 2012 the spread of NBK6A, the shortest dated bond, was below that of NBK3A and above

it in June 2013 to June 2014. The 2 year bond, NBK9A, also exhibited spreads that were lower than those of longer 7 year bond, NBK11A between September 13 and April 2015 but higher from April 2015 to September 2015).

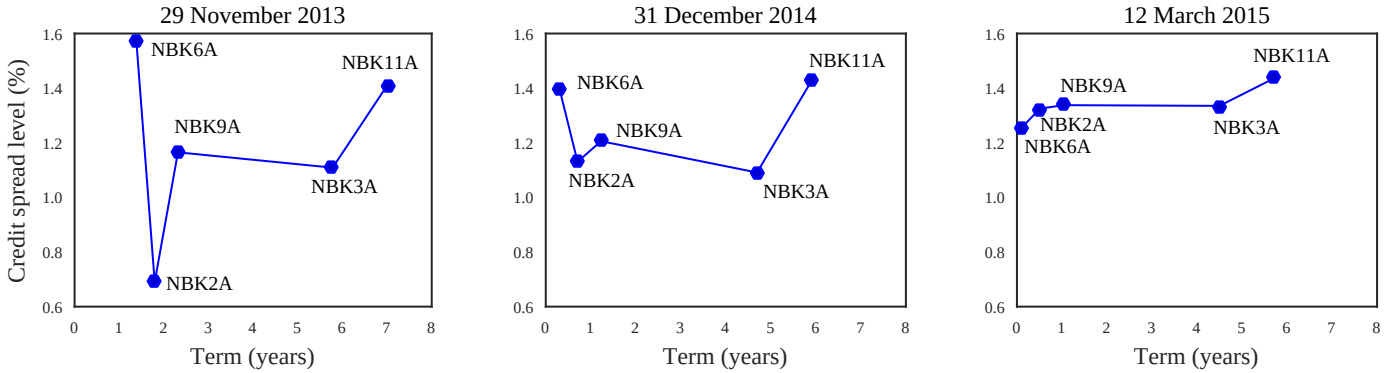


Figure 2.11: Term structures constructed from the LVCF credit spreads of Nedbank bonds as at 29 November 2013, 31 December 2014 and 12 March 2015.

Figure 2.11 illustrates the inconsistent relative nature of the shorter and longer dated bonds. It is clear that in November 2013 and December 2014, there was no discernible upward or downward slope to the line connecting the five credit spreads. In November 2014, the 2 shortest dated bonds, NBK6A and NBK2A had the highest and lowest spreads respectively and in December 2014, the two longest dated bonds, NBK3A and NBK11A, also had the lowest and highest spread levels. It was only in March 2015 that a seemingly reasonable curve emerged.

The poor quality of the term structures arising from the FirstRand and Nedbank LVCF credit spreads supports the findings of Greeff (2004), who showed that for high credit quality bonds, there was no statistically significant upward or downward sloping trend when fitting a linear regression to the term structure of South African corporate bonds for two periods examined in 2003.

Based on seniority, subordinate bonds are seen as more risky, as senior bonds take preference in default. Consequently, NBK2A and NBK3A should have had larger credit spreads than the senior bonds NBK6A, NBK9A and NBK11A, as they were subordinate. But from November 2013 to March 2015, these bonds had two of the lowest credit spreads, only experiencing an increase to the levels of NBK9A and NBK6A in mid March. The term to maturity and seniority of the bonds were analysed in isolation, without taking the other factors into account in the explanation of the spread levels. This way of analysis could be flawed, as a spread could be at a certain level due to a combination of factors, not just one in seclusion. A ceteris paribus approach would be ideal, but that would demand a set of bonds for each factor, keeping all the other factors the same e.g. bonds with the different seniorities but matching durations.

But considering the spreads of NBK2A and NBK6A in terms of term and seniority confirms that the above conjectures are not derailed by analysing the factors separately and that the LVCF credit spread time series are indeed implausible. NBK6A had a shorter term to maturity than NBK2A and was senior, so in the scenario where Nedbank is assumed to have been an investment grade issuer, the term structure should have been upward sloping. Hence the spread of NBK6A should have been somewhat lower than that of NBK2A, due to its reduced risk from being shorter dated and from being senior. It is obvious that this is not so from from Figure 2.11 where in November 2013 and December 2014, NBK6A's spread was

substantially larger than the spread of NBK2A.

Reversing the assumption of high investment quality to poor investment quality for Nedbank means a downward sloping term structure. So NBK6A is expected to have had a higher credit spread than that of NBK2A, which would then have been reduced to factor in less risk due to its seniority. The degree to which these netted each other out is unclear and would require a detailed analysis on empirical data into the absolute magnitude of risk carried by each unit of measurement for term to maturity and level of seniority. This section aims to show that such analysis is not possible because of inherently flawed and sparse historical credit spreads. NBK2A's subordination suggests a higher spread, which should have been closer, but still below, that of NBK6A, resulting in a flatter term structure at the short end. But NBK2A's spread was larger than that of NBK6A in March 2015 as shown in Figure 2.11, contradicting this reasoning.

A short synopsis on credit spreads in South Africa and implementation of Basel II/III regulation: it would be interesting and important to note the change in the spread dynamics due to the introduction of these regulations. However, the implementation of these regulatory changes was "effectively staggered over 10 years" and it is nearly impossible to try to "isolate the effect on spreads from B3 *(Basel III)* over a period that has seen effectively two economic cycles and numerous sovereign rating changes, combined with the growth of the debt capital market"[2].

2.2.6 The Nature of the Most Liquid SA Bond Credit Spreads

The same tests and models that were used on the credit spread indices in Section 2.1 are now applied to the 15 most liquid bonds given in Table 2.7. Figure 2.12 shows the evolutions of the credit spreads of these bonds, comprised only of observed spread levels on days of trade (not LVCF). The outcome of the ADF

Bond	p-value	t-statistic	Critical value
MTN01	0.8549	-0.6554	-2.8672
NBK2A	**0.0388**	**-2.9657**	**-2.8654**
BID01	0.4813	-1.5829	-2.8650
SBS1	0.5771	-1.3665	-2.8664
MTN02	0.2900	-2.0150	-2.8648
SBS9	**0.0010**	**-4.1909**	**-2.8654**
IPL4	0.4991	-1.5428	-2.8657
IPL6	0.9940	0.8117	-2.8661
NBK3A	**0.0278**	**-3.0899**	**-2.8651**
ABS3	0.4608	-1.6295	-2.8674
ABL8A	0.8666	-0.6036	-2.8678
NBK9A	**0.0023**	**-3.9666**	**-2.8679**
CBL11	**0.0140**	**-3.3339**	**-2.8678**
LGL02	0.9990	2.2778	-2.8679
FRX15	0.4140	-1.7351	-2.8666

Table 2.10: Results of the ADF test applied to the credit spreads of the 15 most liquid bonds. The red figures show the rejection of the null hypothesis in favour of stationary time series.

test applied to the credit spread time series of each bond is presented in Table 2.10. 5 out of the 15 bonds were stationary, again indicated by the statistics highlighted in red.

[2]Elena Ilkova of RMB Global Markets Fixed Income and Credit Research, via an email communique.

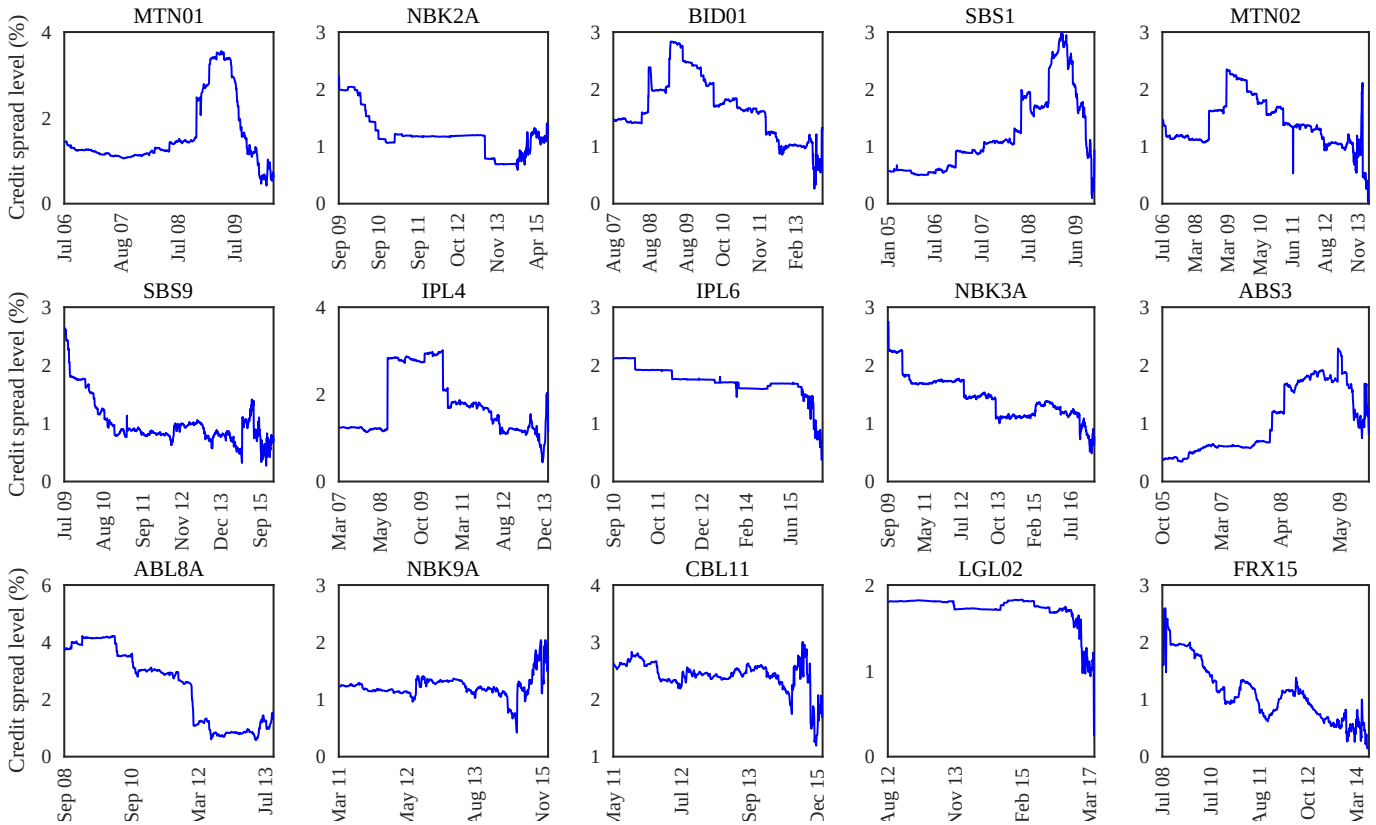


Figure 2.12: The credit spread evolutions of the 15 most liquid SA corporate bonds.

Figure 2.13 plots of the autocorrelation function values for the credit spread time series of the bonds and shows that the ACF values decayed sufficiently rapidly to support the assumption of stationary time series.

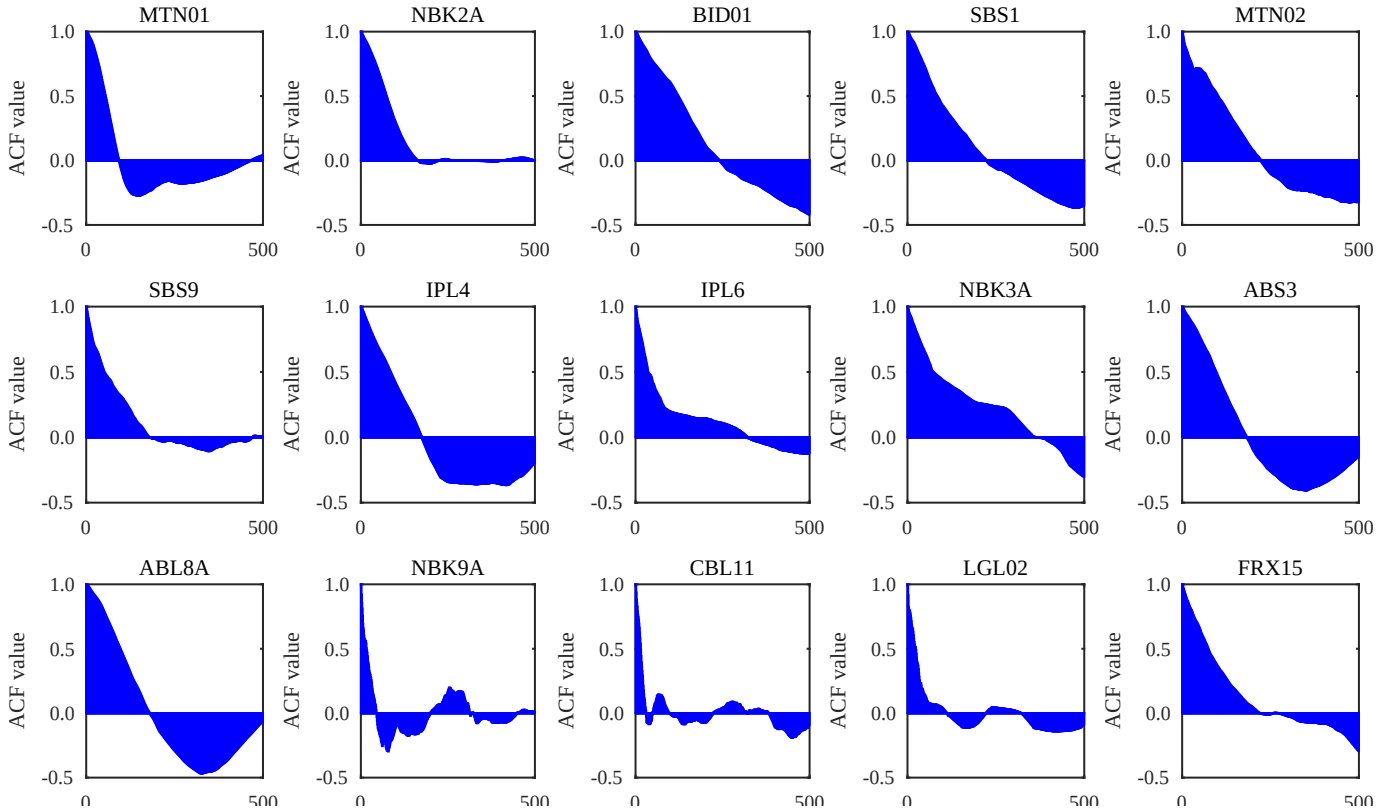


Figure 2.13: Autocorrelation function values up to the 500$^{\text{th}}$ lag of the 15 most liquid bonds.

The HLCD, RB, OU and Log OU models are now calibrated to the credit spread time series of each bond and the results are given in Table 2.11. The OU model had the highest likelihood value for 13 of the bonds, whilst the remaining 2 bond's maximum likelihood was attained using the Log OU model. The corresponding AIC levels for the models with the highest likelihoods were also the lowest.

The BIC penalised the OU model for 8 of the bonds, with lower BIC values obtained by the HLCD model for these instances. But in the majority of those cases, the estimate for the drift component (μ) from the HLCD model was close to or equal to zero, implying that these spreads did not exhibit any significant trend. Thus the spreads could have possibly been oscillating around some long term mean value.

The models used for calibration and inference in the following chapters will be the OU and (indirectly) Log OU models, as it justified to assume that credit spreads are mean reverting and these mean-reverting models were shown to have fit the time series of credit indices and credit spreads of a number of bonds (local and offshore) aptly.

Bond	Model	μ	θ	σ	Maximum $\mathcal{L}$ ($\times 10^3$)	AIC ($\times 10^3$)	BIC ($\times 10^3$)
MTN01	HLCD	0.0000		0.8142	0.8245	-1.6436	**-1.6361**
	RB	0.1194		0.8146	0.8228	-1.6416	-1.6327
	OU	0.9603	0.3408	0.8146	**0.8248**	**-1.6437**	-1.6304
	Log OU	0.0000	0.5424	0.6100	0.7824	-1.5589	-1.5455
NBK2A	HLCD	0.0000		0.4257	1.7480	-3.4920	-3.4824
	RB	0.0000		0.4259	1.7458	-3.4877	-3.4781
	OU	1.0974	1.4843	0.4255	**1.7527**	**-3.4995**	**-3.4851**
	Log OU	0.0844	1.5209	0.4186	1.6423	-3.2786	-3.2642
BID01	HLCD	0.0000		0.6757	1.4119	-2.8197	**-2.8100**
	RB	0.2107		0.6761	1.4099	-2.8158	-2.8061
	OU	1.6410	0.7454	0.6767	**1.4131**	**-2.8202**	-2.8056
	Log OU	0.4440	2.4126	0.8015	0.8164	-1.6268	-1.6122
SBS1	HLCD	0.0704		0.7268	1.0256	-2.0469	**-2.0379**
	RB	0.3349		0.7273	1.0237	-2.0434	-2.0341
	OU	1.3061	0.6275	0.7275	**1.0265**	**-2.0470**	-2.0331
	Log OU	0.0733	1.8055	1.0359	0.7496	-1.4933	-1.4794
MTN02	HLCD	0.0000		0.8751	1.2631	-2.5222	**-2.5122**
	RB	0.2019		0.8753	1.2616	-2.5192	-2.5092
	OU	1.2730	1.4858	0.8780	**1.2653**	**-2.5246**	-2.5096
	Log OU	0.0001	0.1312	1.2676	0.5750	-1.1439	-1.1290
SBS9	HLCD	0.0000		0.6784	1.2781	-2.5523	-2.5427
	RB	0.0000		0.6785	1.2768	-2.5496	-2.5399
	OU	0.9045	2.7748	0.6787	**1.2874**	**-2.5689**	**-2.5544**
	Log OU	0.0000	5.1159	1.0391	0.9615	-1.9169	-1.9025
IPL4	HLCD	0.0000		0.8391	1.0107	-2.0173	-2.0078
	RB	0.3397		0.8396	1.0090	-2.0140	-2.0045
	OU	1.7538	0.7044	0.8403	1.0119	-2.0177	-2.0034
	Log OU	0.4828	0.7914	0.4950	**1.0405**	**-2.0749**	**-2.0606**
IPL6	HLCD	0.0000		0.4100	1.4866	-2.9692	**-2.9598**
	RB	0.0000		0.4102	1.4850	-2.9659	-2.9565
	OU	0.0000	0.1314	0.4097	**1.4877**	**-2.9693**	-2.9553
	Log OU	0.0000	0.2149	0.4342	1.0214	-2.0367	-2.0226
NBK3A	HLCD	0.0000		0.3749	1.8477	-3.6913	-3.6816
	RB	0.0000		0.3750	1.8461	-3.6881	-3.6784
	OU	1.2268	1.0968	0.3742	**1.8541**	**-3.7021**	**-3.6875**
	Log OU	0.1767	0.9709	0.3452	1.6134	-3.2207	-3.2061
ABS3	HLCD	0.0774		0.6933	0.7657	-1.5275	-1.5187
	RB	0.3181		0.6939	0.7640	-1.5239	-1.5151
	OU	1.1373	0.8514	0.6945	0.7671	-1.5281	-1.5150
	Log OU	0.0938	0.6648	0.5451	**0.9757**	**-1.9455**	**-1.9323**
ABL8A	HLCD	0.0000		0.6162	0.7420	-1.4799	**-1.4714**
	RB	0.0000		0.6165	0.7411	-1.4781	-1.4696
	OU	0.0004	0.1831	0.6147	**0.7437**	**-1.4814**	-1.4686
	Log OU	0.0000	0.2707	0.4224	0.5611	1.1161	-1.1033
NBK9A	HLCD	0.1119		0.7830	0.6078	-1.2115	-1.2030
	RB	0.4190		0.7838	0.6061	-1.2082	-1.1997
	OU	1.2383	7.2293	0.7977	**0.6155**	**-1.2251**	**-1.2123**
	Log OU	0.1972	7.8351	0.6898	0.5947	-1.1834	-1.1706
CBL11	HLCD	0.0000		0.8975	0.5356	-1.0671	-1.0586
	RB	0.2122		0.8982	0.5342	-1.0643	-1.0558
	OU	2.3874	5.2567	0.9099	**0.5412**	**-1.0764**	**-1.0636**
	Log OU	0.8614	5.3578	0.4397	0.4627	-0.9194	-0.9066
LGL02	HLCD	0.0000		0.5026	0.8388	-1.6729	**-1.6651**
	RB	0.0000		0.5030	0.8373	-1.6707	-1.6621
	OU	0.0000	0.1529	0.5023	**0.8395**	**-1.6730**	-1.6602
	Log OU	0.0000	0.3002	0.7257	0.3729	-0.7399	-0.7271
FRX15	HLCD	0.0000		0.8150	0.8171	-1.6301	**-1.6210**
	RB	0.1155		0.8153	0.8157	-1.6274	-1.6182
	OU	0.9163	1.0859	0.8171	**0.8188**	**-1.6316**	-1.6179
	Log OU	0.0000	0.8904	0.9864	0.6855	-1.3650	-1.3513

Table 2.11: Calculated parameter estimates, maximum likelihoods, AIC and BIC values obtained through fitting the models to each bond's credit spread time series. The numbers in red indicate the maximum likelihood attained for the time series and the smallest AIC and BIC.

2.2.7 Recent South African Credit Crises

2.2.7.1 African Bank

The African Bank (with share price ticker ABL) crisis in August 2014 is a recent and revealing example of the rare event of a corporate institution defaulting on its debt obligations and declaring insolvency in South Africa. It was at one stage the largest provider of unsecured loans in the South African market. However, in May 2013, the bank was already reporting significant losses due to the failure of collections on these loans, attributing the poor performance to the aggressive uptake of unsecured loans in the market over preceding years (Reuters *et al.* (2013)). A rights issue was also offered in December 2013 as the bank was not meeting the required solvency levels at that time. Additionally, it is reported that the bank did not reduce its rate of unsecured lending to the retail market (Reuters (2014)), resulting in the bank finally declaring such large losses in August 2014 that it was not able to continue its operations or make good on any repayment of its debt without additional capital raising. The bank was subsequently bailed out by the South African Reserve Bank, which placed it under curatorship and split the debt book into 'bad' and 'good', the 'good' debt intending to be part of the new structure of the bank.

African Bank issued 11 pure vanilla bonds from January 2005 to March 2018. These bonds, their maturities, seniority, trade frequency ratios, liquidity measures, total traded and total issued nominals are shown in Table 2.14, ordered by maturity date.

Bond	Maturity	Seniority	TFR (%)	LM (days)
ABL2	18 Sep 06	Senior	30.37	4
ABL3	12 Jul 07	Senior	27.22	8
ABL4	31 Aug 10	Senior	19.81	7
ABL5	11 Aug 11	Senior	8.55	28
ABL6	18 Jun 12	Senior	8.79	16
ABL7	18 Feb 13	Senior	17.89	6
ABL8A	19 Sep 13	Senior	42.53	20
ABL11A	29 Sep 14	Senior	25.80	29
ABL10A	15 Mar 15	Senior	8.40	20
ABK2	24 May 18	Senior	5.80	14
ABK1	07 Nov 18	Senior	4.40	25
Average	-	-	18.14	16

Table 2.12: African Bank's pure vanilla bonds between January 2005 and March 2018 as well as their maturities, seniority, calculated trade frequency ratios and calculated liquidation measures.

ABL8A was the most liquid bond, with a TFR of 42.53% and was one of the top 15 liquid bonds identified in Table 2.7 and some of the bonds achieved LMs of only a few days. The average trade frequency ratio and average liquidation measure of 18.14% and 16 days respectively show reasonably liquidity, relative to the averages of the other issuers in Figure 2.6. All of the bonds traded during their lifetimes, resulting in a zero trade ratio of 0%, which 4 of the 5 big banks could not achieve.

Did the credit spreads of the African bank bonds alive during the default express the increasing risk in the months preceding the event? The analysis of liquidity for the bank in Table 2.14 shows that there was some liquidity, so perhaps there was a degree translation of the market's rising angst over the bank's

deterioration into the spreads of its bonds.

There were 2 bonds that were alive during the default in August 2014, ABL10A and ABL11A. Figure 2.14 shows their LVCF credit spreads, the days when they traded and the traded credit spread level, plotted against the African Bank share price from each bond's issue date to maturity.

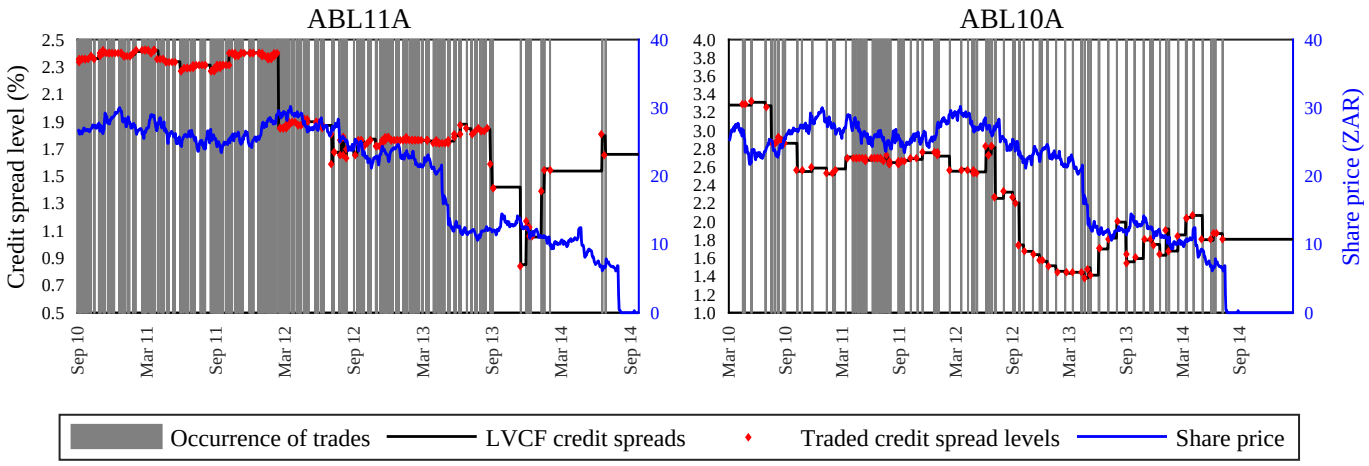


Figure 2.14: LVCF credit spread time series (black lines), days of trade (vertical grey lines) and market observed credit spreads (red dots) of ABL11A and ABL10A from issuance to maturity. African Bank's share price is shown in blue.

ABL11A experienced a decent number of trades for the first three years of its life, indicated by the greater density of vertical lines, but the frequency of trades greatly diminished in the later part of 2013. Post late 2013, the spread showed prolonged periods of zero movement and large shocks when a trade did occur, as was the case for MTN01, NBK2A, BID01 and SBS1, observed in Figure 2.7. Most notably, the credit spread did not widen to any degree representative of the financial turmoil the bank was experiencing in the time leading up to the crisis. The severity of its situation was abundantly evident by the decline in the share price from May 2013 onwards, but no equivalent reaction was reflected in the credit spread when the bond was subsequently traded after May 2013.

ABL10A also showed credit spreads that did not increase to levels representative of the bank's credit risk in the period leading up to the default. The bond traded at fairly regular intervals and its spread did experience some volatility from June 2013 until August 2014. But at default, the spread moved up by only 43 bps to 1.81%, which was still far below its highest value of 3.31% in May 2010, from its low of 1.38% in May 2013.

This example demonstrates why low liquidity means poor quality of credit spread time series. Bonds experienced some liquidity in the time leading up to the default - the shock in the share price in April 2013 did indeed register in the credit spreads, with ABL11A's spread experiencing an increase and the spread of ABL10A spiking. But after that, the credit spreads of the bonds failed to react to the escalating risk of default that was so clearly present in the deterioration of the share price.

2.2.7.2 Steinhoff

Steinhoff International is the holdings company for business units that operate in various retail areas across 3 continents. It is a dual listed share, trading primarily on the Frankfurt Stock Exchange with a secondary listing on the JSE. Locally domiciled debt instruments were issued under the subsidiary called Steinhoff

Services. On 6 December 2017, after it was announced that accounting regularities had been uncovered and that the CEO had resigned, the Steinhoff share price plummeted, the yields of bonds issued in Europe spiked (Bonorchis *et al.* (2018)) and were later downgraded to junk status by Moody's. It has been alleged that the holding company used entities belonging to the former CEO's associates to conceal losses, leading to an investigation into possible fraudulent activities by the CEO (Bloomberg (2018)). Early in 2018, the group said that the issues were predominantly in the central European business, where revenue figures had been overstated for years. As at the time of writing, PwC were performing a forensic audit to determine whether these figures were intentionally inflated. SA parliament was told by the new management that they could not be sure when the investigation would be complete (Cronje (2018)).

At their time of issue, the ZAR denominated instruments were 'cross guaranteed' (Thompson (2018)), meaning that they were backed by other units under the Steinhoff Holdings umbrella[3]. The Euro denominated bonds were only guaranteed by the holding company in Europe. The implication of this cross-guarantee of the local bonds was that investors could lay claim to the assets of the larger Steinhoff entity, some more profitable than others, in the case of default. To settle this cross guaranteed South African debt, Steinhoff embarked on a capital raising effort, by selling assets across the groups (Bonorchis *et al.* (2018)) and freezing dividend payments (Cronje (2018)). On 22 Feburary 2018, Steinhoff Services redeemed 11 of the existing 12 debts at par, after holders of these 11 local instruments voted in favour of this proposal, whilst investors holding the 12^{th} instrument (an FRN with code SHS34) voted against this decision (Crotty (2018)). However, this instrument was redeemed by Steinhoff on 6 March 2018.

Early redemption of the debt meant that the investors holding Steinhoff bonds were repaid their initial face value, but lost out on the subsequent coupons that were due to them until the maturity of the bonds. Essentially Steinhoff exercised a non-existing call option on their bonds, which had never been priced as having any optionality (explored in Bosman *et al.* (2009)). So although the initial investment value was recouped, Steinhoff defaulted in paying to the investors the outstanding interest on these loans.

Steinhoff Services issued 4 pure vanilla bonds between January 2005 and March 2018. These bond as well as their maturities, seniority, TFR and LM, are shown in Table 2.13.

Bond	Maturity	Seniority	TFR (%)	LM (days)
SHS05	29 Jun 17	Senior	24.88	17
SHS19	10 Sep 17	Senior	1.30	72
UTR40	10 Sep 17	Senior	0.40	112
SHS25	29 Jun 20	Senior	9.10	368
Average	-	-	8.92	142

Table 2.13: Steinhoff Services' pure vanilla bonds between January 2005 and March 2018: their maturities and seniority, as well as their calculated trade frequency ratios and calculated liquidation measures.

SHS25 was the only bond alive in December 2017 and only matured 2.5 years later. The LVCF credit spread time series for SHS25, as well as the days of trade, the market observed credit spreads and Steinhoff share price are plotted in Figure 2.15 from its date of issuance until the end of March 2018. Also plotted is the

[3]Steinhoff International Holdings, Steinhoff Investment Holdings, Steinhoff Africa Holdings, Ainsley Holdings and Pepkor Holdings

LVCF quoted credit spread time series, as calculated by the JSE.

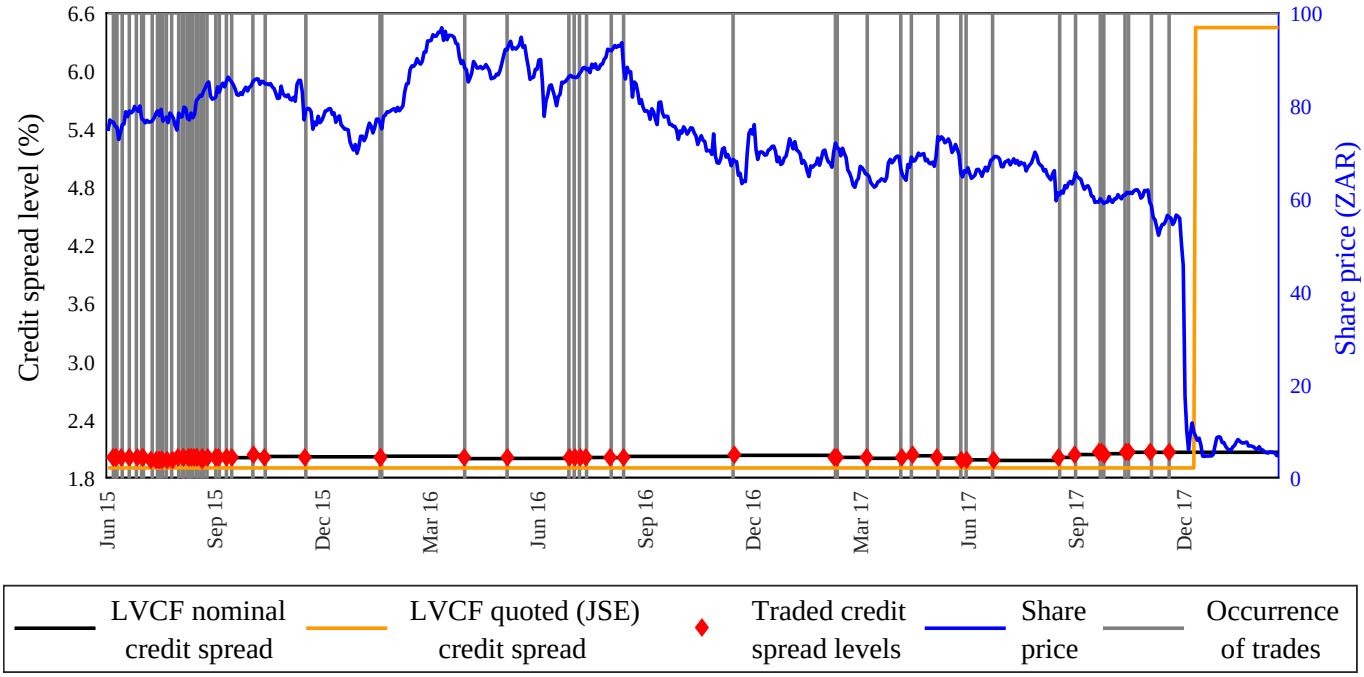


Figure 2.15: LVCF credit spread time series (black line), days of trade (vertical grey lines) and market observed credit spreads (red dots) of SHS25 from date of issuance to March 2018. Steinhoff's share price is shown in blue and the JSE quoted LVCF time series in yellow.

Figure 2.15 shows that SHS25 was traded quite frequently in the first 6 months after it was issued, but after that trade was sparse. The credit spread experienced low volatility, hovering around the 2% level for the period. The bond stopped trading on 23 November 2017, so the credit spread remained unchanged. Planting (2017) confirmed that the Steinhoff crisis reinforces how illiquid the bond market is in South Africa and the negative ramifications that this lack of trade has on pricing risk in a corporate debt instruments.

The quoted spread is shown in Figure 2.15 to illustrate an approach the JSE uses to address the marking-to market of illiquid corporate bonds. The JSE has an informal practice in place, where if a broker sees a bid/offer yield that is less/greater than the current mark-to-market yield on a trade size of ZAR5m or more, the broker can inform the JSE, on instruction from their client, to manually override the mark-to-market value to this bid/offer yield level. This non-standard policy has always been allowed, but the trade size necessary for the pricing intervention was reduced to ZAR1m in 2016. This insertion of non-traded pricing information, particularly with the current lowered trade value requirement, opens up the possibility of market manipulation and mis-pricing. Any bid/offer could be placed with a broker in small size at the time of the debt market's close of day, reported to the JSE and then cancelled, all with the intention of repricing an instrument to favour the client's portfolio. A bond would be valued at a price that is not representative of its true market value, as it is not a traded price, where 2 independent, distinct parties agree on its worth.

When the share price crashed in early December, the quoted spread was ramped up by 4.5%, even though no trades occurred after late November. In theory, the jump in the quoted spread is indeed the behaviour expected given the circumstances, but the change is only justified when traded by separate market participants and not set to the bid or offer value suggested by a singular party.

3. The Ornstein-Uhlenbeck Variable Long Term Mean (OUVLTM) Model

Chapter 2 gave compelling evidence of the flaws in the current credit spread quantification via LVCF, resultant from infrequent trading. In periods of no trade, market conditions that could affect the credit worth of an issuer/issuance are changing. But the lack of trade leads to spreads that are unreflective of any changes as they are simply set to the previous traded level.

This chapter introduces a new model, the Ornstein-Uhlenbeck Variable Long Term Mean (OUVLTM) model, which is an adaptation of the popular OU model documented in Appendix A.2.6. The new model is novel in that it assumes that the mean reversion equilibrium level is no longer constant, but is linked to one or more exogenous factors. These factors are prudently chosen for the material information they contained in estimating spreads when no trades have occurred. The OUVLTM model agrees with the sentiment of Ruiz *et al.* (2012), who proposed that default events must be included when using mean reversion to model credit spreads so that the mean reversion level is not only driven by the spread evolution, but also relevant external factors. The OU model uses the credit spread time series only to calibrate and forecast and so is limited by the quality of the credit spread data. The new OUVLTM model uses the exogenous factor(s) in addition to the credit spread time series, incorporating meaningful information into the estimations and predictions, resulting in far greater explanatory power.

A review of the literature reveals a definitive inverse relationship between credit risk and the stock price of the issuer, supporting the usage of the stock price of the issuer as an appropriate exogenous factor. Although the connection is not shown in a contingent claims framework, there is merit behind this assumption. Indeed, the base assumption of the class of structural (Merton (1974)) models is that default is triggered by decreases in the firm asset value. Hung *et al.* (2006) reported a negative correlation between a Taiwanese equity index and credit spreads in that market. Using a stochastic volatility Merton model with jumps, Zhang *et al.* (2009) also found a strong connection between high yield (large credit spread) bonds and their distressed equity value. Madan *et al.* (1998) presented a direct inverse relationship between the arrival rate of default and the issuer's equity value, relativised by the money market account. A multiple factor credit model was developed by Longstaff *et al.* (2008), citing firm-specific, industry and general market default events as the primary explanatory factors of credit variance. Furthering this work, the systemic risk of credit spreads of US bonds was demonstrated by Bhar *et al.* (2011) to be driven by the S&P 500 index, its implied volatility index, the VIX, and long-term interest rates. Finally, a key assumption made by Carr *et al.* (2010) in their study was that the stock price falls to zero in the event of default.

Taking the external factor of the OUVLTM model to be the log stock price, this chapter goes on to derive the theory underpinning the model. The behaviour of the model relative to the traditional OU model is investigated to assess reasonability. Convergence of the expected values and variances of numerically simulated OUVLTM paths to the derived analytical solutions is shown. The mathematics needed for the estimation of the OUVLTM parameters is developed, and the point estimates and posterior distributions of the OUVLTM and OU model parameters are compared in terms of bias and goodness of fit measures,

using simulated data sets.

3.1 Model Description

Consider the credit spread $\tilde{x}(t)$ of a bond issued by an entity with the stock price denoted $s(t)$. Then the dynamics of $\tilde{x}$ under OUVLTM are given by the following SDE:

$$d\tilde{x}(t) = \theta(\tilde{\mu}(t) - \tilde{x}(t))dt + \sigma_x dW_x^{\mathbb{P}}(t), \tag{3.1}$$

where θ is the constant rate of mean reversion, σ_x is the constant instantaneous volatility and $W_x^{\mathbb{P}}(t) \sim N(0,t)$ is a standard Wiener process with respect to the real world measure $\mathbb{P}$.

The model is original in that it assumes the mean reversion level is no longer constant but a function of the log of the stock price of the issuer. The time dependent equilibrium, $\tilde{\mu}(t)$, over time 0 to t, is given by

$$\tilde{\mu}(t) = \tilde{\mu}(0) - \eta t - \sigma_s W_s^{\mathbb{P}}(t), \tag{3.2}$$

where η is defined as

$$\eta = \mu_s - \frac{1}{2}\sigma_s^2. \tag{3.3}$$

μ_s and σ_s are the constant drift and instantaneous volatility parameters of the GBM SDE at (A.16) assumed to describe $s(t)$:

$$ds(t) = \mu_s s(t)dt + \sigma_s s(t) dW_s^{\mathbb{P}}(t).$$

$W_s^{\mathbb{P}}(t) \sim N(0,t)$ is a standard Wiener process with respect to the real world measure $\mathbb{P}$, with

$$dW_x^{\mathbb{P}} dW_s^{\mathbb{P}} = \rho_{sx} dt,$$

where ρ_{sx} is the constant correlation of the $\tilde{x}(t)$ and $s(t)$ Wiener processes over dt. In discrete time, $\rho_{sx} = 0$ for disjointed intervals.

The solution for the OUVLTM process $\tilde{x}(t)$ is given by

$$\begin{aligned}\tilde{x}(t) = \tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1 - e^{-\theta t}) - \eta\left(t - \frac{1 - e^{-\theta t}}{\theta}\right) \\ - \sigma_s \int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u) + \sigma_x \int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u).\end{aligned} \tag{3.4}$$

$\tilde{x}(t)$ is normally distributed, with mean

$$\mathbb{E}[\tilde{x}(t)] = \tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1 - e^{-\theta t}) - \eta\left(t - \frac{1 - e^{-\theta t}}{\theta}\right) \tag{3.5}$$

and variance

$$\mathrm{Var}[\tilde{x}(t)] = \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x\sigma_s\rho_{sx}}{2\theta}(1 - e^{-2\theta t}) - \frac{2\sigma_s}{\theta}(\sigma_s + \sigma_x\rho_{sx})(1 - e^{-\theta t}) + \sigma_s^2 t. \tag{3.6}$$

The risk-neutral dynamics for OUVLTM, necessary for pricing of contingent claims (derivatives), can be obtained via Girsanov's theorem (Appendix A.1.5.3) and a change in measure from the real world $\mathbb{P}$ to the unique equivalent martingale or risk-neutral measure $\mathbb{Q}$. However, the scope of this research remains within the realm of the real world, where real world time series are observed, analysed and modelled. As such, the risk-neutral dynamics of the OUVLTM model are not required here.

3.1.1 Model Derivation

The OU model for the credit spread, $x(t)$, of a bond assumes a static long term mean reversion or equilibrium level μ_x as seen in the OU SDE at (A.29) in Appendix A.2.6,

$$dx(t) = \theta(\mu_x - x(t))dt + \sigma_x dW_x^{\mathbb{P}}(t),$$

where θ is the constant rate of mean reversion, μ_x is the constant long-term mean, σ_x is the constant instantaneous volatility and $W_x^{\mathbb{P}}(t)$ is the same Wiener process as in (3.1).

The OUVLTM model indirectly links the credit spread to the change in the natural logarithm of the stock price via through the variable long term mean $\tilde{\mu}$ via

$$\tilde{\mu}(t) = \tilde{\mu}(0) + \ln\left(\frac{\ln s(0)}{\ln s(t)}\right). \tag{3.7}$$

In Appendix A.2.5 it is shown that $\ln s(t)$ is normally distributed with solution

$$\ln s(t) = \ln s(0) + \left(\mu_s - \frac{1}{2}\sigma_s^2\right)t + \sigma_s W_s^{\mathbb{P}}(t)$$

$$\Rightarrow \ln s(t) \sim \mathcal{N}\left(\ln s(0) + \left(\mu_s - \frac{1}{2}\sigma_s^2\right)t, \sigma_s^2 t\right).$$

Hence $\tilde{\mu}$ can be expressed as

$$\tilde{\mu}(t) = \tilde{\mu}(0) - \eta t - \sigma_s W_s^{\mathbb{P}}(t).$$

The mean reversion level is thus a scaled inverse replica of the log of the stock price, with

$$\mathbb{E}[\tilde{\mu}(t)] = \tilde{\mu}(0) - \eta t \quad \text{and} \quad \text{Var}[\tilde{\mu}(t)] = \sigma_s^2 t. \tag{3.8}$$

Applying Itô's formula to $e^{\theta t}\tilde{x}(t)$ gives:

$$\begin{aligned}
d\left(e^{\theta t}\tilde{x}(t)\right) &= \theta e^{\theta t}\tilde{x}(t)dt + e^{\theta t}d\tilde{x}(t) \\
&= \theta e^{\theta t}\tilde{x}(t)dt + e^{\theta t}(\theta(\tilde{\mu}(t) - \tilde{x}(t))dt + \sigma_x dW_x^{\mathbb{P}}(t)) \\
&= \theta e^{\theta t}\tilde{\mu}(t)dt + e^{\theta t}\sigma_x dW_x^{\mathbb{P}}(t) \\
\Rightarrow e^{\theta t}\tilde{x}(t) &= x(0) + \theta\int_0^t \tilde{\mu}(u)e^{\theta u}du + \sigma_x\int_0^t e^{\theta u}dW_x^{\mathbb{P}}(u) \\
\Rightarrow \tilde{x}(t) &= \tilde{x}(0)e^{-\theta t} + \theta e^{-\theta t}\int_0^t \tilde{\mu}(u)e^{\theta u}du + \sigma_x e^{-\theta t}\int_0^t e^{\theta u}dW_x^{\mathbb{P}}.
\end{aligned} \tag{3.9}$$

The integral in the second term of (3.9) can be written as

$$\int_0^t \tilde{\mu}(u)e^{\theta u}du = \int_0^t \left[\tilde{\mu}(0)e^{\theta u} - \eta u e^{\theta u} - \sigma_s e^{\theta u}W_s^{\mathbb{P}}(u)\right]du. \tag{3.10}$$

Now

$$\int_0^t \tilde{\mu}(0)e^{\theta u}du = \frac{\tilde{\mu}(0)}{\theta}e^{\theta u}\Big|_0^t = \frac{\tilde{\mu}(0)}{\theta}\left[e^{\theta t}-1\right]$$

and

$$\int_0^t \eta u e^{\theta u}du = \frac{\eta}{\theta^2}\left[\theta u-1\right]e^{\theta u}\Big|_0^t = \frac{\eta}{\theta^2}\left[(\theta t-1)e^{\theta t}+1\right].$$

Using

$$\begin{aligned} d\left(e^{\theta t}W^s(t)\right) &= \theta e^{\theta t}W_s^{\mathbb{P}}(t)dt + e^{\theta t}dW_s^{\mathbb{P}}(t) \\ \Rightarrow e^{\theta t}W_s^{\mathbb{P}}(t) &= \int_0^t \theta e^{\theta u}W_s^{\mathbb{P}}(u)du + \int_0^t e^{\theta u}dW_s^{\mathbb{P}}(u), \end{aligned}$$

shows that

$$\begin{aligned} \int_0^t e^{\theta u}W^s(u)du &= \frac{1}{\theta}\left[e^{\theta t}W_s^{\mathbb{P}}(t) - \int_0^t e^{\theta u}dW_s^{\mathbb{P}}(u)\right] \\ &= \frac{1}{\theta}\int_0^t \left[e^{\theta t}-e^{\theta u}\right]dW_s^{\mathbb{P}}(u), \end{aligned}$$

since

$$W_s^{\mathbb{P}}(t) = \int_0^t dW_s^{\mathbb{P}}(u).$$

So the integral at (3.10) is equal to

$$\int_0^t \tilde{\mu}(u)e^{\theta u}du = \frac{\tilde{\mu}(0)}{\theta}\left[e^{\theta t}-1\right] - \frac{\eta}{\theta^2}\left[(\theta t-1)e^{\theta t}+1\right] - \frac{\sigma_s}{\theta}\int_0^t \left[e^{\theta t}-e^{\theta u}\right]dW_s^{\mathbb{P}}(u),$$

which yields the following for the second term of (3.9)

$$\theta e^{-\theta t}\int_0^t \tilde{\mu}(u)e^{\theta u}du = \tilde{\mu}(0)(1-e^{-\theta t}) - \eta\left(t-\frac{1-e^{-\theta t}}{\theta}\right) - \sigma_s\int_0^t \left(1-e^{\theta(u-t)}\right)dW_s^{\mathbb{P}}(u).$$

And finally, the solution for $\tilde{x}(t)$ given at (3.4) is found to be

$$\begin{aligned} \tilde{x}(t) = \tilde{x}(0)e^{-\theta t} &+ \tilde{\mu}(0)(1-e^{-\theta t}) - \eta\left(t-\frac{1-e^{-\theta t}}{\theta}\right) \\ &- \sigma_s\int_0^t \left(1-e^{\theta(u-t)}\right)dW_s^{\mathbb{P}}(u) + \sigma_x\int_0^t e^{\theta(u-t)}dW_x^{\mathbb{P}}(u). \end{aligned}$$

From (3.4), it follows directly that the mean of $\tilde{x}(t)$ is equal to (3.5):

$$\mathbb{E}[\tilde{x}(t)] = \tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1-e^{-\theta t}) - \eta\left(t-\frac{1-e^{-\theta t}}{\theta}\right).$$

The variance at (3.6) is determined as follows:

$$
\begin{aligned}
\mathrm{Var}[\tilde{x}(t)] &= \mathbb{E}\left[\tilde{x}(t) - \mathbb{E}[\tilde{x}(t)]\right]^2 \\
&= \mathbb{E}\left[-\sigma_s \int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u) + \sigma_x \int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u)\right]^2 \\
&= \sigma_s^2 \mathbb{E}\left[\left(\int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u)\right)^2\right] + \sigma_x^2 \mathbb{E}\left[\left(\int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u)\right)^2\right] \\
&- 2\sigma_s\sigma_x \mathbb{E}\left[\int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u) \int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u)\right] \\
&= \sigma_s^2 \mathbb{E}\left[\left(\int_0^t \left(1 - e^{\theta(u-t)}\right)^2 du\right)\right] + \sigma_x^2 \mathbb{E}\left[\left(\int_0^t e^{2\theta(u-t)} du\right)\right] \\
&- 2\sigma_s\sigma_x\rho_{sx} \mathbb{E}\left[\int_0^t \left(1 - e^{\theta(u-t)}\right) e^{\theta(u-t)} du\right] \quad \text{(using Itô Isometry and Theorem A.2)} \\
&= \sigma_s^2 \left[u - \frac{2}{\theta} e^{\theta(u-t)} + \frac{1}{2\theta} e^{2\theta(u-t)}\right]\Bigg|_0^t + \frac{\sigma_x^2}{2\theta} e^{2\theta(u-t)}\Bigg|_0^t - 2\sigma_s\sigma_x\rho_{sx}\left[\frac{1}{\theta} e^{\theta(u-t)} - \frac{1}{2\theta} e^{2\theta(u-t)}\right]\Bigg|_0^t \\
&= \sigma_s^2 \left[t - \frac{2}{\theta}(1 - e^{-\theta t}) + \frac{1}{2\theta}(1 - e^{-2\theta t})\right] + \frac{\sigma_x^2}{2\theta}(1 - e^{-2\theta t}) \\
&- 2\sigma_s\sigma_x\rho_{sx}\left[\frac{1}{\theta}(1 - e^{-\theta t}) - \frac{1}{2\theta}(1 - e^{-2\theta t})\right] \\
&= \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x\sigma_s\rho_{sx}}{2\theta}(1 - e^{-2\theta t}) - \frac{2\sigma_s}{\theta}(\sigma_s + \sigma_x\rho_{sx})(1 - e^{-\theta t}) + \sigma_s^2 t \quad \text{(grouping exponent terms).}
\end{aligned}
$$

3.2 Illustrative Simulated Examples

Table 3.1 clarifies the notation that will be used to describe the simulations going forward, as well as the default parameter values used to generate the simulated paths. These specific parameter values were

Process		x	s	$\tilde{\mu}$	$\tilde{x}$
Description		Credit spread	Stock price	Variable long term mean	Credit spread
Model		OU	GBM	OUVLTM	OUVTLM
Default parameter values	Δt	$1/365$	$1/365$	$1/365$	$1/365$
	Path length	1,000	1,000	1,000	1,000
	Starting value	2.5	12	2.5	2.5
	μ_s		0.08	0.08	
	σ_s		0.2	0.2	
	μ_x	3			
	θ	1			1
	σ_x	0.4			0.4

Table 3.1: Notation description and chosen default parameters used in simulations.

chosen to represent real world scenarios as closely as possible. Δt is the equal spacing between business days, approximated as 252 days in one calendar year. Most of the analysed credit spread time series are about 1,000 points long or less. And Tables 2.3 and 2.11 give suitable proxies for the general levels of the OU and OUVLTM model parameters, based on those observed for market bonds in Table 2.11 in Chapter 2.

Assuming uncorrelated Wiener processes for each dt interval ($\rho_{sx} = 0$), one sample path is generated for x, s, $\tilde{\mu}$ and $\tilde{x}$. The left plot in Figure 3.1 shows the paths of $\ln(s)$ in black and $\tilde{\mu}$ in pink. The right plot adds x and $\tilde{x}$ in red and blue respectively. As desired, the variable long term mean is a scaled inverse of the

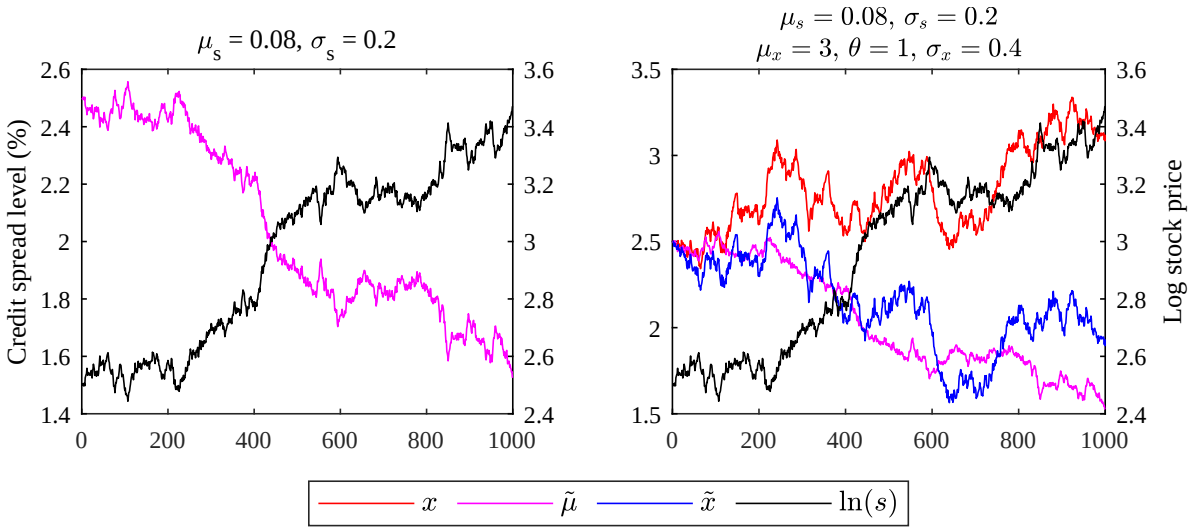


Figure 3.1: Generated sample paths for x, $\ln(s)$, $\tilde{\mu}$ and $\tilde{x}$ using the default model parameter values.

log stock price as shown in the left plot. x is observed to tend towards its mean reversion level of 3, whilst $\tilde{x}$ oscillates around $\tilde{\mu}$ with the same noise as x. The upward trend of the log stock price is precisely reflected by the downward sloping $\tilde{\mu}$, which then translates into a general inverse co-movement of $\tilde{x}$ and $\ln(s)$.

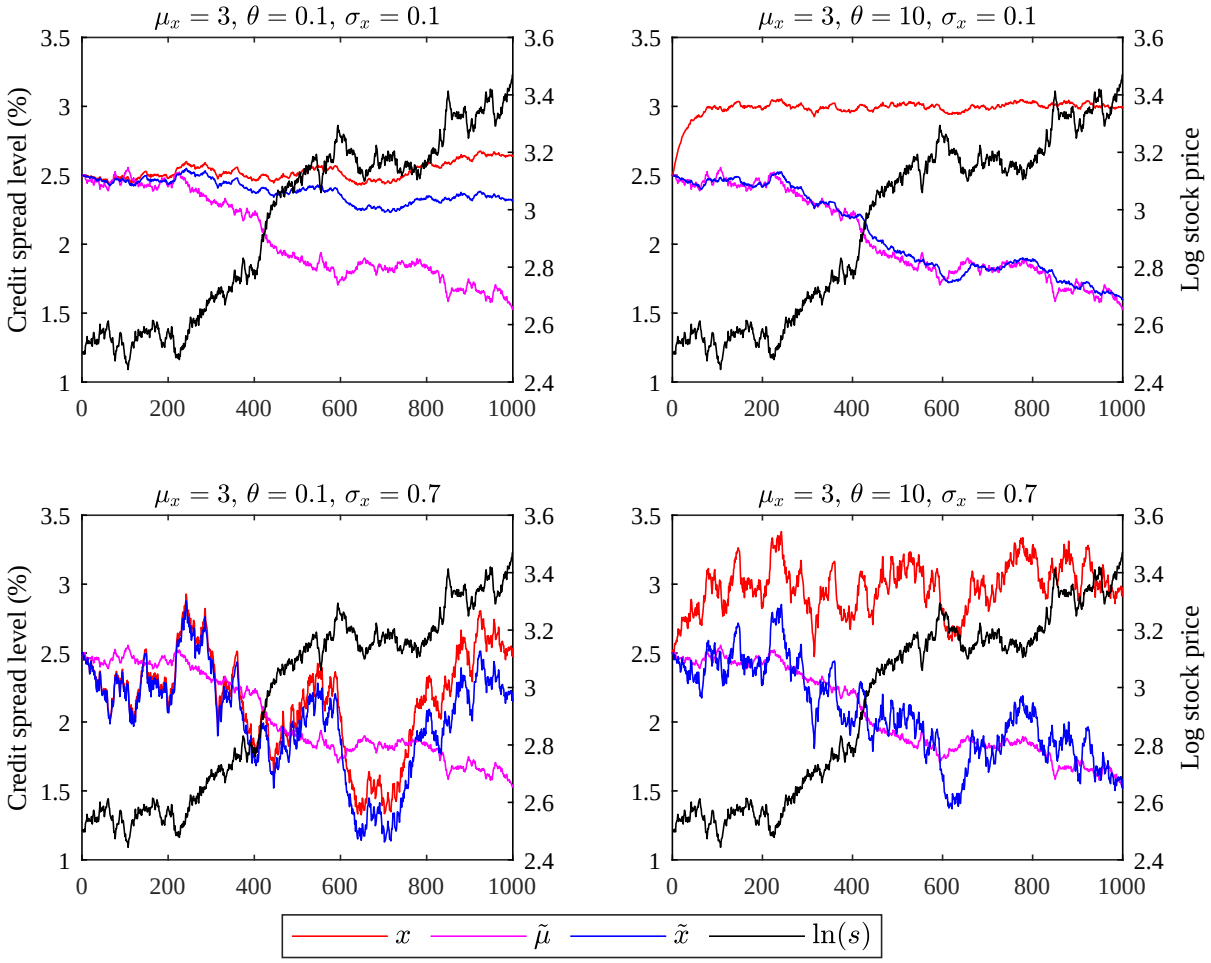


Figure 3.2: Generated sample paths for x, $\ln(s)$, $\tilde{\mu}$ and $\tilde{x}$, where θ and σ_x are perturbed simultaneously.

Figure 3.2 shows the sample paths generated using simultaneously perturbed values of θ and σ_x. The increase of θ from 0.1 to 10 results in $\tilde{x}$ pulling away from x towards its variable long term mean, as the

mean reversion dominates the stochastic noise. x behaves as expected, reaching its static equilibrium point of 3 much quicker when θ is high. Raising the volatility from 0.1 to 0.7 shows $\tilde{x}$ clearly becoming more volatile, thus the deviations from $\tilde{\mu}$ increase for both levels of θ. By simulating 1,500 paths for $\tilde{x}$, given the

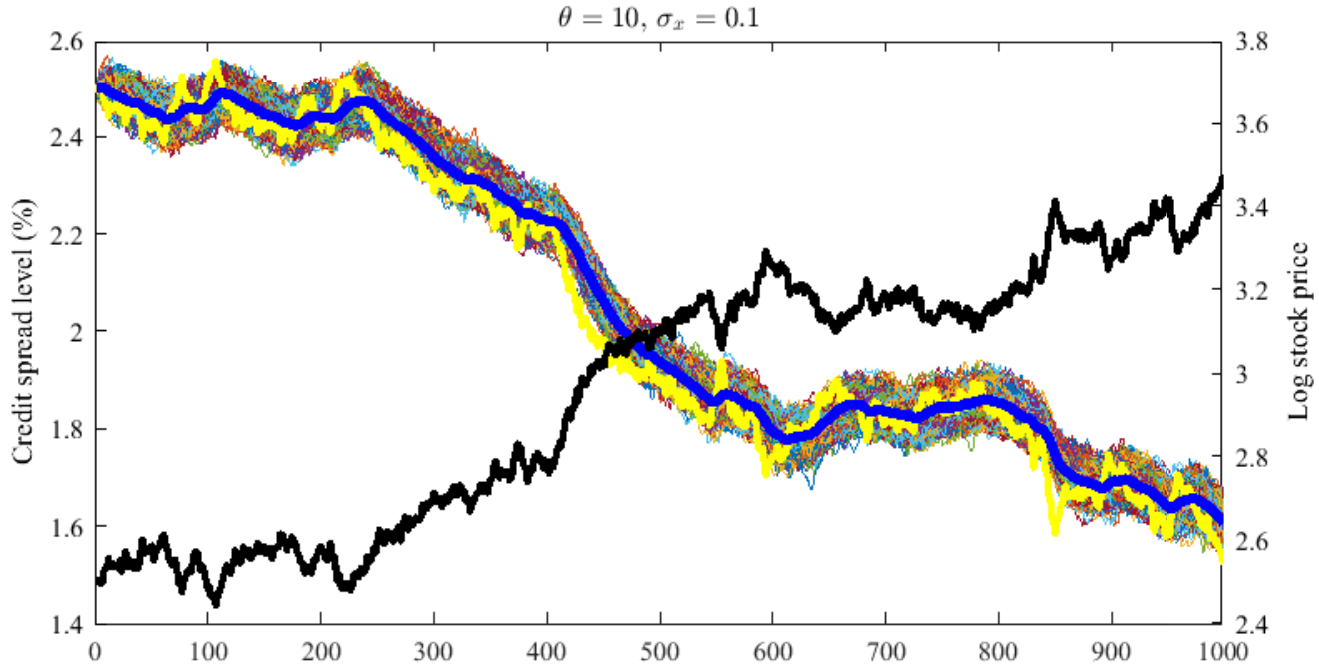


Figure 3.3: Illustration of the capability of the OUVLTM model to produce credit spreads that inversely track localised and general changes and trends in the stock price. The thick yellow line is $\tilde{\mu}$, the thick blue line is the mean of the sample paths and the thick black line is $\ln(s)$.

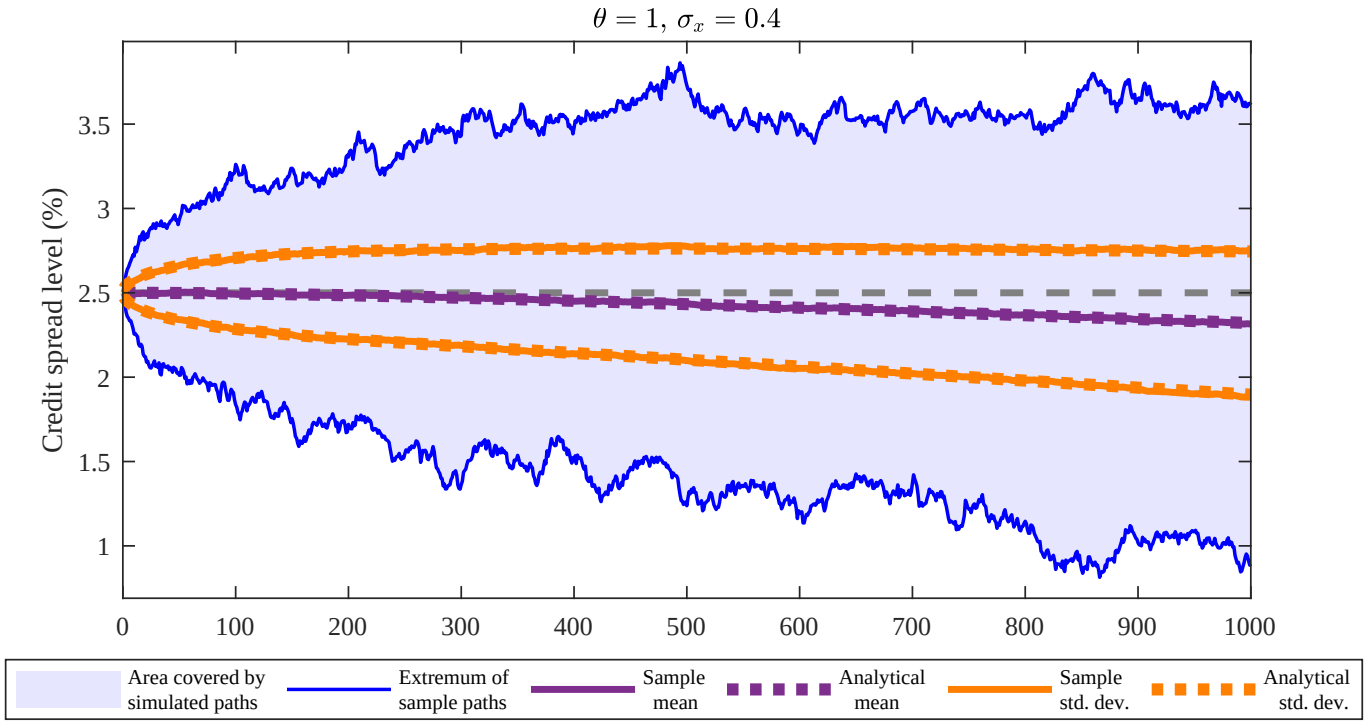


Figure 3.4: Sample mean and standard deviations plotted against the analytical results using the default model parameters.

same $\ln(s)$ and $\tilde{\mu}$ sample path presented in Figures 3.1 and 3.2, Figure 3.3 demonstrates how the OUVLTM model can be utilised to generate credit spreads that behave exactly as anticipated. $\tilde{\mu}$ follows the inverse

general trend of $\ln(s)$ and also reflects any localised changes in $\ln(s)$. At a high degree of mean reversion, the expected credit spread (blue line) follows $\tilde{\mu}$ (yellow line) and so retains the inverse relationship to the general trends and localised deviations of the stock price (black line).

A comparison of numerical and analytical expected values and standard deviations will give an indication of convergence of the model, where the sample mean and standard deviation of the numerically simulated OUVLTM paths should converge to the analytical values postulated at (3.5) and (3.6). Figure 3.4 confirms that both the expected value and standard deviation over many sample paths are equal to the analytical levels at each point in time. The mean of the multiple paths at each point form the continuous purple line. The analytical expected value, as determined from (3.5), is the dotted purple line which sits on top of the sample mean. Similarly, the orange dotted lines are the analytical standard deviation above and below the analytical mean. These lie on top of the continuous orange lines which are the sample standard deviations (calculated from (3.6) above and below the sample mean. The light blue shaded area shows the coverage of the simulated multiple paths, bounded by the blue lines which are the maximum and minimum of the sample paths at each point, just to give some context. Also observe that the analytical and numerical expected values are downward sloping when compared to the horizontal dashed grey line, a product of the positive trending stock prices.

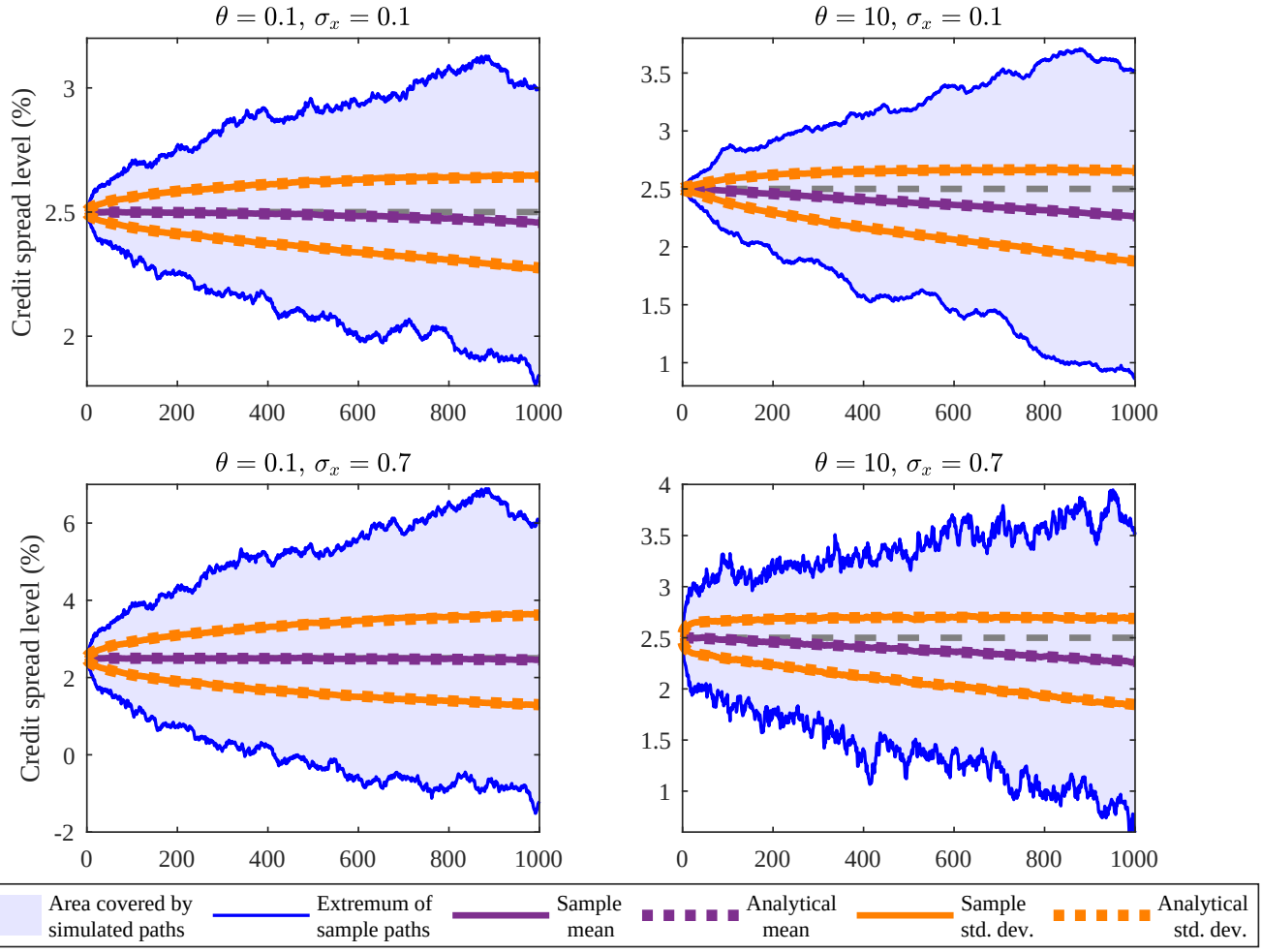


Figure 3.5: Sample mean and standard deviations of the paths plotted against the analytical results when θ and σ_x are simultaneously perturbed.

Figure 3.5 shows the same analysis for the perturbed θ and σ_x states. In each case, convergence of the sample to analytical statistic is shown, demonstrating that the OUVLTM model is robust in parameter

change. The high θ cases show more negatively sloped expected values than the low θ cases, implied by the greater distance from the horizontal dashed grey line. This result makes sense, as higher θ values means closer tracking of the downward sloping mean reversion by $\tilde{x}$. The lower volatility cases also exhibit further deviation from the horizontal grey lines than those of high volatility, showing that more noise leads to a decrease in distinguishable trend.

3.2.1 The Log OUVLTM model

The high volatility, low mean reversion perturbations in the previous section show that, like OU, credit spreads can be negative under OUVLTM. To resolve this issue, the log OUVLTM model can be used instead.

Assume that $\tilde{x}(t) = \ln(\tilde{y}(t))$, where $\tilde{x}(t)$ is modelled using OUVLTM. From (3.5), it is deduced that

$$\tilde{x}(t) = \ln(\tilde{y}(t)) \sim N\left(\tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1 - e^{-\theta t}) - \eta\left(t - \frac{1 - e^{-\theta t}}{\theta}\right), \Sigma_{\tilde{x}}\right),$$

where $\Sigma_{\tilde{x}} = \mathrm{Var}[\tilde{x}(t)]$ is defined at (3.6).

The strictly positive process $\tilde{y}(t) = \exp(\tilde{x}(t))$ then follows the dynamics of the log OUVLTM model; $\tilde{y}(t)$ is thus defined as the log OUVLTM process.

The expected value and variance of $\tilde{y}(t)$ are then calculable by using (A.12) and (A.13), giving

$$\mathbb{E}(\tilde{y}(t)) = \exp\left(\tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1 - e^{-\theta t}) - \eta\left(t - \frac{1 - e^{-\theta t}}{\theta}\right) + \frac{1}{2}\Sigma_{\tilde{x}}\right) \tag{3.11}$$

and

$$\mathrm{Var}[\tilde{y}(t)] = \exp\left(2\left(\tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1 - e^{-\theta t}) - \eta\left(t - \frac{1 - e^{-\theta t}}{\theta}\right)\right) + \Sigma_{\tilde{x}}\right)\left(\exp(\Sigma_{\tilde{x}}) - 1\right). \tag{3.12}$$

Hence one can model $\tilde{x}$ and easily retrieve the mean and variance of $\tilde{y}(t)$ using (3.11) and (3.12).

3.3 Parameter Estimation Theory

Parameter estimation of the models is performed using two methods. Point estimates for models parameters are obtained using Maximum Likelihood Estimation (MLE). Full posterior distributions of the parameters are generated using Gibbs sampling, a well-known Markov Chain Monte Carlo (MCMC) method stemming from Bayesian statistics (see Appendix B). Described in more detail in Appendix B.5.3, the algorithm forms the parameter posterior distributions via sequential sampling from the conditional probability densities of the parameters. Bayesian techniques produce distributions of parameter estimates as well as allowing prior knowledge of the parameters to be incorporated, which MLE does not do.

This section presents the mathematics necessary to estimate the parameters of OUVLTM. The likelihood function is derived and solutions for the maximum likelihood estimates are calculated. The standard errors of these point estimates are also determined. Then the conditional posterior distributions of the OUVLTM parameters needed in Gibbs sampling are also derived.

3.3.1 Maximum Likelihood Estimation (MLE)

3.3.1.1 Deriving the OUVLTM Likelihood Function

The Euler discretisation of the OUVLTM SDE at (3.1) for $0 \leq t_1 < t_i < t_n$ where $i = 1, ..., n$, is

$$\begin{aligned} d\tilde{x}(t) &\approx \tilde{X}_i - \tilde{X}_{i-1}, \\ dW_x^{\mathbb{P}}(t) &\approx W_i^x - W_{i-1}^x = \Delta t \epsilon_i, \;\; \epsilon_i \sim N(0,1), \\ dt &\approx t_i - t_{i-1} = \Delta t, \end{aligned} \tag{3.13}$$

where $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_n)$ is the discretised OUVLTM process.

The conditional density of the normally distributed $\tilde{X}_i$ for $i = 2, ..., n$ is

$$\begin{aligned} f(\tilde{X}_i|\tilde{X}_{i-1}; \mu_s, \sigma_s, \theta, \sigma_x) &= \frac{1}{\sqrt{2\pi \text{Var}[\tilde{X}_i]}} \exp\left[\frac{-(\tilde{X}_i - \mathbb{E}[\tilde{X}_i])^2}{2\text{Var}[\tilde{X}_i]}\right] \\ &= \frac{1}{\sqrt{2\pi V}} \exp\left[\frac{-\left(\tilde{X}_i - \tilde{X}_{i-1}e^{-\theta\Delta t} - \tilde{\mu}_{i-1}(1 - e^{-\theta\Delta t}) + \eta\left(\Delta t - \frac{1-e^{-\theta\Delta t}}{\theta}\right)\right)^2}{2V}\right]. \end{aligned}$$

Here, η is defined at (3.3). V is defined as the variance of $\tilde{X}_i$ - essentially the per step or discretised version of (3.6) - given by

$$V := \text{Var}[\tilde{X}_i] = \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x\sigma_s\rho_{sx}}{2\theta}(1 - e^{-2\theta\Delta t}) - \frac{2\sigma_s}{\theta}(\sigma_s + \sigma_x\rho_{sx})(1 - e^{-\theta\Delta t}) + \sigma_s^2\Delta t. \tag{3.14}$$

Taking the natural logarithm of the conditional density for $\tilde{X}_i$ gives

$$\ln f(\tilde{X}_i|\tilde{X}_{i-1}; \mu_s, \sigma_s, \theta, \sigma_x) = -\frac{1}{2}\ln(2\pi V) - \frac{1}{2V}\left(\tilde{X}_i - \tilde{X}_{i-1}e^{-\theta\Delta t} - \tilde{\mu}_{i-1}(1 - e^{-\theta\Delta t}) + \eta\left(\Delta t - \frac{1 - e^{-\theta\Delta t}}{\theta}\right)\right)^2.$$

Now the product of the individual conditional densities over $i = 2, ..., n$ yields the likelihood, thus the sum of the individual log conditional densities over $i = 2, ..., n$ gives the log likelihood for $\tilde{X}$ as follows

$$\begin{aligned} \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \theta, \sigma_x) &= \prod_{i=2}^{n} f(\tilde{X}_i|\tilde{X}_{i-1}; \mu_s, \sigma_s, \theta, \sigma_x) \\ \Rightarrow \ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \theta, \sigma_x) &= \sum_{i=2}^{n} \ln f(\tilde{X}_i|\tilde{X}_{i-1}; \mu_s, \sigma_s, \theta, \sigma_x). \end{aligned}$$

Thus

$$\begin{aligned} \ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \theta, \sigma_x) = &-\frac{n-1}{2}\ln(2\pi V) \\ &-\frac{1}{2V}\sum_{i=2}^{n}\left(\tilde{X}_i - \tilde{X}_{i-1}e^{-\theta\Delta t} - \tilde{\mu}_{i-1}(1 - e^{-\theta\Delta t}) + \eta\left(\Delta t - \frac{1 - e^{-\theta\Delta t}}{\theta}\right)\right)^2. \end{aligned} \tag{3.15}$$

3.3.1.2 Calculating the MLEs of the OUVLTM Model

The MLEs of θ and σ_x, denoted as $\hat{\theta}$ and $\hat{\sigma}_x$, are the values that maximise the log likelihood of $\tilde{X}$ at (3.15). Mathematically, the optimisation is expressed as

$$\underset{0<\theta<\infty,\sigma_x>0}{\arg\max} \ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \theta, \sigma_x) \tag{3.16}$$

$$= \underset{0<\theta<\infty,\sigma_x>0}{\arg\max} \left(-\frac{n-1}{2}\ln(2\pi V) - \frac{1}{2V}\sum_{i=2}^{n}\left[\tilde{X}_i - \tilde{X}_{i-1}e^{-\theta\Delta t} - \tilde{\mu}_{i-1}(1-e^{-\theta\Delta t}) + \hat{\eta}\left(\Delta t - \frac{1-e^{-\theta\Delta t}}{\theta}\right)\right]\right)^2,$$

with V defined at (3.14) and

$$\hat{\eta} = \hat{\mu}_s - \frac{1}{2}\hat{\sigma}_s^2,$$

which is explained in Appendix A.2.5.2. Closed form solutions for $\hat{\theta}$ and $\hat{\sigma}_x$ are analytically intractable using the likelihood in its current form, but determining stable estimates using (3.16) is computationally straightforward.

Nevertheless, a reparametrisation of the model is proposed: let

$$\begin{aligned} \phi &= e^{-\theta\Delta t}, \\ \zeta &= \frac{1-e^{-\theta\Delta t}}{\theta} \quad \text{and} \\ \tilde{A}_i &= \tilde{X}_i - \tilde{\mu}_i, \quad i = 1, ..., n. \end{aligned} \tag{3.17}$$

Then the sum in the second term of (3.15) is transformed as follows

$$\begin{aligned} &\tilde{X}_i - \phi\tilde{X}_{i-1} - \tilde{\mu}_{i-1}(1-\phi) + \eta\left(\Delta t - \frac{1-\phi}{\theta}\right) \\ &= \tilde{X}_i - \phi(\tilde{X}_{i-1} - \mu_{i-1}) - \tilde{\mu}_{i-1} + \eta\Delta t - \eta\zeta \\ &\approx \tilde{X}_i - \phi(\tilde{X}_{i-1} - \tilde{\mu}_{i-1}) - \tilde{\mu}_i - \eta\zeta \\ &= \tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\zeta. \end{aligned}$$

The reparamaterisation simplifies the parameters of the linear autoregressive AR(1) OUVLTM process, such that

$$\tilde{A}_i = \phi\tilde{A}_{i-1} + \eta\zeta + \sqrt{V_A\Delta t}\epsilon_i, \quad \epsilon_i \sim N(0,1). \tag{3.18}$$

Here ϕ represents the degree of autocorrelation.

In continuous time, the mean of $\tilde{a}(t) = \tilde{x}(t) - \tilde{\mu}(t)$ is derived to be

$$\begin{aligned} \mathbb{E}[\tilde{a}(t)] &= \mathbb{E}[\tilde{x}(t) - \tilde{\mu}(t)] = \mathbb{E}[\tilde{x}(t)] - \mathbb{E}[\tilde{\mu}(t)] \\ &= \phi\tilde{x}(0) + \tilde{\mu}(0)(1-e^{-\theta t}) - \eta\left(t - \frac{1-e^{-\theta t}}{\theta}\right) - \tilde{\mu}(0) + \eta t \\ \Rightarrow \mathbb{E}[\tilde{a}(t)] &= \phi\tilde{a}(0) + \eta\left(t - \frac{1-e^{-\theta t}}{\theta}\right). \end{aligned}$$

Hence in discrete time,

$$\mathbb{E}[\tilde{A}_i] = \phi\tilde{A}_{i-1} + \eta\zeta. \tag{3.19}$$

Similarly, the continuous time variance is

$$\begin{aligned}
\mathrm{Var}[\tilde{a}(t)] &= \mathrm{Var}[\tilde{x}(t) - \tilde{\mu}(t)] = \mathrm{Var}[\tilde{x}(t)] + \mathrm{Var}[\tilde{\mu}(t)] - \mathrm{Cov}[\tilde{x}(t), \tilde{\mu}(t)] \\
&= \mathrm{Var}[\tilde{x}(t)] + \sigma_s^2 t - \mathbb{E}[([\tilde{x}(t) - \mathbb{E}[[\tilde{x}(t)])(\tilde{\mu}(t) - \mathbb{E}[\tilde{\mu}(t)])] \\
&= \mathrm{Var}[\tilde{x}(t)] + \sigma_s^2 t - \mathbb{E}\left[\left(-\sigma_s \int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u) - \sigma_x \int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u)\right)\left(-\sigma_s \int_0^t dW_s^{\mathbb{P}}(u)\right)\right] \\
&= \mathrm{Var}[\tilde{x}(t)] + \sigma_s^2 t - \mathbb{E}\left[\sigma_s^2 \int_0^t \left(1 - e^{\theta(u-t)}\right) du - \sigma_x \sigma_s \rho_{sx} \int_0^t e^{\theta(u-t)} du\right] \\
&\qquad\qquad \text{(using Itô Isometry and Theorem A.2)} \\
&= \mathrm{Var}[\tilde{x}(t)] + \sigma_s^2 t - \sigma_s^2 \left[u - \frac{1}{\theta} e^{\theta(u-t)}\right]\Big|_0^t - \sigma_x \sigma_s \rho_{sx} \left[\frac{1}{\theta} e^{\theta(u-t)}\right]\Big|_0^t \\
&= \mathrm{Var}[\tilde{x}(t)] + \sigma_s^2 t - \sigma_s^2 \left[t - \frac{1}{\theta}(1 - e^{\theta t})\right] - \sigma_x \sigma_s \rho_{sx} \frac{1}{\theta} \left[1 - e^{\theta t}\right] \\
&= \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x \sigma_s \rho_{sx}}{2\theta}(1 - e^{-2\theta t}) - \frac{2\sigma_s}{\theta}(\sigma_s + \sigma_x \rho_{sx})(1 - e^{-\theta t}) + \sigma_s^2 t + \sigma_s^2 t \\
&- \sigma_s^2 t + \frac{\sigma_s}{\theta}(\sigma_s - \sigma_x \sigma_s \rho_{sx})(1 - e^{\theta t}) \\
&= \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x \sigma_s \rho_{sx}}{2\theta}(1 - e^{-2\theta t}) - \frac{\sigma_s}{\theta}(\sigma_s + \sigma_x \rho_{sx})(1 - e^{-\theta t}) + \sigma_s^2 t.
\end{aligned}$$

The discrete time variance, denoted $V_{\tilde{A}}$, follows as

$$V_{\tilde{A}} = \frac{\sigma_s^2 + \sigma_x^2 + 2\sigma_x \sigma_s \rho_{sx}}{2\theta}(1 - \phi^2) - \frac{\sigma_s}{\theta}(\sigma_s + \sigma_x \rho_{sx})(1 - \phi) + \sigma_s^2 \Delta t. \tag{3.20}$$

The difference between the variance of $\tilde{X}$ and $\tilde{A}$ shown in (3.14) and (3.20) manifests in the second term by a factor of 2, that is

$$V_{\tilde{A}} = V + \frac{\sigma_s}{\theta}(\sigma_s + \sigma_x \rho_{sx})(1 - \phi).$$

The conditional density now takes the form

$$f(\tilde{X}_i | \tilde{X}_{i-1}; \mu_s, \sigma_s, \theta, \sigma_x) = \frac{1}{\sqrt{2\pi V_{\tilde{A}}}} \exp\left[\frac{-\left(\tilde{A}_i - \phi \tilde{A}_{i-1} - \eta\zeta\right)^2}{2V_{\tilde{A}}}\right], \quad i = 2, ..., n,$$

leading to the modified log likelihood function for the model given by

$$\ln \mathcal{L}(\tilde{X} | \mu_s, \sigma_s, \phi, V_{\tilde{A}}) = -\frac{n-1}{2} \ln(2\pi V_{\tilde{A}}) - \frac{1}{2V_{\tilde{A}}} \sum_{i=2}^{n} \left(\tilde{A}_i - \phi \tilde{A}_{i-1} - \eta\zeta\right)^2. \tag{3.21}$$

The limits specified for the new parameters at (3.17) follow either from the limits set by the original parameters or from intuition about the characteristics of the underlying process itself. ϕ must be greater than 0 since $e^{-\theta \Delta t}$ can never be below zero. ϕ must also be less than 1 since θ must be greater than 0. $V_{\tilde{A}} > 0$ since $\sigma_x^2 > 0$. Using the following Taylor expansion of ζ around θ, observe that

$$\zeta = \frac{1 - e^{-\theta \Delta t}}{\theta} \approx \frac{1}{\theta}\left(1 - \left(1 - \Delta t \theta + \mathcal{O}(\Delta t^2)\right)\right) \approx \Delta t, \tag{3.22}$$

so $\zeta > 0$.

Substituting the result for ζ at (3.22) into (3.21) gives

$$\ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}}) = -\frac{n-1}{2}\ln(2\pi V_{\tilde{A}}) - \frac{1}{2V_{\tilde{A}}}\sum_{i=2}^{n}\left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2. \tag{3.23}$$

Closed form solutions for approximations of the MLEs of the new parameters, $\hat{\phi}$ and $\hat{V}_{\tilde{A}}$, can now be determined using (3.23) via

$$\begin{aligned}
\frac{\partial \ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}})}{\partial \phi}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} &= \left[-\frac{1}{V_{\tilde{A}}}\sum_{i=2}^{n}(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t)(-\tilde{A}_{i-1})\right]\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = 0,\\
\frac{\partial \ln \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}})}{\partial V_{\tilde{A}}}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} &= \left[-\frac{n-1}{2V_{\tilde{A}}} + \frac{1}{2V_{\tilde{A}}^2}\sum_{i=2}^{n}(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t)^2\right]\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = 0,
\end{aligned}$$

yielding

$$\hat{\phi} = \frac{\sum_{i=2}^{n}\tilde{A}_{i-1}(\tilde{A}_i - \eta\Delta t^2)}{\sum_{i=2}^{n}\tilde{A}_{i-1}^2}, \tag{3.24}$$

$$\hat{V}_{\tilde{A}}(\hat{\phi}) = \frac{1}{n-1}\sum_{i=2}^{n}\left(\tilde{A}_i - \hat{\phi}\tilde{A}_{i-1} - \eta\Delta t\right)^2. \tag{3.25}$$

The MLE for θ follows as

$$\hat{\theta}(\hat{\phi}) = -\frac{1}{\Delta t}\ln(\hat{\phi}). \tag{3.26}$$

(3.20) shows that $V_{\tilde{A}}$ is a quadratic function of σ_x and can be expressed in terms of σ_x as follows:

$$B\sigma_x^2 + C\sigma_x + D - \hat{V}_{\tilde{A}} = 0, \tag{3.27}$$

where

$$B = \frac{1-\hat{\phi}^{\,2}}{2\hat{\theta}}, \quad C = \frac{\hat{\sigma}_s\rho_{sx}\hat{\phi}}{\hat{\theta}}(1-\hat{\phi}) \quad \text{and} \quad D = \hat{\sigma}_s^2\left(B - \frac{1}{\hat{\theta}}(1-\hat{\phi}) + \Delta t\right).$$

Thus

$$\hat{\sigma}_x(\hat{\phi}, \hat{\sigma}_s) = \frac{-C \pm \sqrt{C^2 - 4B(D - \hat{V}_{\tilde{A}})}}{2B}. \tag{3.28}$$

When $\rho_{sx} = 0$, $C = 0$ and so (3.28) reduces to

$$\hat{\sigma}_x(\hat{\phi}, \hat{\sigma}_s) = \sqrt{\frac{\hat{V}_{\tilde{A}} - D}{B}}. \tag{3.29}$$

Maximum Likelihood Estimation is consistent and asymptotically normal i.e.

1. an estimate $\hat{\pi}$ of the 'true' parameter value π of the sample distribution, tends to π as the sample size n tends to infinity, and
2. $\sqrt{n}(\hat{\pi} - \pi) \to N(0, \sigma_\pi)$, where σ_π is the asymptotic variance of of $\hat{\pi}$

respectively.

3.3.1.3 Determining the Standard Errors of the OUVLTM MLEs

The standard errors of the estimates found in the Maximum Likelihood optimisation are determined using the method given by Enders (2010), set out as:

1. Determine the second partial derivative of the likelihood (or log likelihood) function with respect to the parameter.
2. Multiply the second derivative by -1.
3. Compute the inverse or reciprocal.
4. Take the square root.

The second partial derivatives of the OUVLTM log likelihood function at (3.23) with respect to ϕ and $V_{\tilde{A}}$ are

$$\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial \phi^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = \frac{\partial}{\partial \phi}\left[-\frac{1}{V_{\tilde{A}}}\sum_{i=2}^{n}(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t)(-\tilde{A}_{i-1})\right]\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = -\frac{1}{\hat{V}_{\tilde{A}}}\sum_{i=2}^{n}\tilde{A}_{i-1}^2,$$
$$\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial V_{\tilde{A}}^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = \frac{\partial}{\partial V_{\tilde{A}}}\left[-\frac{n-1}{2V_{\tilde{A}}} + \frac{1}{2V_{\tilde{A}}^2}\sum_{i=2}^{n}(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t)^2\right]\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = -\frac{n-1}{2\hat{V}_{\tilde{A}}^2}.$$

From (3.25), observe that

$$(n-1)\hat{V}_{\tilde{A}} = \sum_{i=2}^{n}(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t)^2.$$

This result is substituted into the expression for $V_{\tilde{A}}$ to obtain the second partial derivative.

The ensuing standard errors of $\hat{\phi}$ and $\hat{V}_{\tilde{A}}$, using Enders' Steps (2) to (4), are

$$\hat{\phi}^{SE} = \sqrt{\frac{\hat{V}_{\tilde{A}}}{\sum_{i=2}^{n}\tilde{A}_{i-1}^2}} \quad \text{and} \tag{3.30}$$
$$\hat{V}_{\tilde{A}}^{SE} = \sqrt{\frac{2\hat{V}_{\tilde{A}}^2}{n-1}}. \tag{3.31}$$

The chain rule is used to determine the second partial derivatives for θ and σ_x as follows:

$$\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial \theta^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = \left[\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial \phi^2}\frac{\partial \phi}{\partial \theta} + \frac{\partial \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial \phi}\frac{\partial^2 \phi}{\partial \phi \partial \theta}\right]\frac{\partial \phi}{\partial \theta}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}}$$
$$\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial \sigma_x^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} = \left[\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial V_{\tilde{A}}^2}\frac{\partial V_{\tilde{A}}}{\partial \sigma_x} + \frac{\partial \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial V_{\tilde{A}}}\frac{\partial^2 V_{\tilde{A}}}{\partial V_{\tilde{A}}\partial \sigma_x}\right]\frac{\partial V_{\tilde{A}}}{\partial \sigma_x}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}}.$$

From (3.17) and (3.20), it can be seen that

$$\frac{\partial \phi}{\partial \theta} = -\Delta t e^{-\theta\Delta t} \Rightarrow \frac{\partial}{\partial \phi}\left(\frac{\partial \phi}{\partial \theta}\right) = \frac{\partial}{\partial \phi}(-\Delta t e^{-\theta\Delta t}) = 0,$$
$$\frac{\partial V_{\tilde{A}}}{\partial \sigma_x} = \frac{1}{\theta}\left[(\sigma_x + \sigma_s\rho_{sx})(1 - e^{-2\theta\Delta t}) - \sigma_s\rho_{sx}(1 - e^{-\theta\Delta t})\right]$$
$$\Rightarrow \frac{\partial}{\partial V_{\tilde{A}}}\left(\frac{\partial V_{\tilde{A}}}{\partial \sigma_x}\right) = \frac{\partial}{\partial V_{\tilde{A}}}\left(\frac{1}{\theta}\left[(\sigma_x + \sigma_s\rho_{sx})(1 - e^{-2\theta\Delta t}) - \sigma_s\rho_{sx}(1 - e^{-\theta\Delta t})\right]\right) = 0.$$

So the second partial derivatives become

$$\begin{aligned}
\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial\theta^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} &= \frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial\phi^2}\left(\frac{\partial\phi}{\partial\theta}\right)^2\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}}, \\
&= -\frac{\Delta t^2\hat{\phi}^2}{\hat{V}_A}\sum_{i=2}^{n}\tilde{A}_{i-1}^2 \\
\frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial\sigma_x^2}\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} &= \frac{\partial^2 \ln \mathcal{L}(\tilde{X}|\mu_s,\sigma_s,\phi,V_{\tilde{A}})}{\partial V_{\tilde{A}}^2}\left(\frac{\partial V_{\tilde{A}}}{\partial\sigma_x}\right)^2\Big|_{\hat{\phi},\hat{V}_{\tilde{A}}} \\
&= -\frac{(n-1)}{2\hat{V}_{\tilde{A}}^2}\left(\frac{1}{\theta}\left[(\sigma_x+\sigma_s\rho_{sx})(1-e^{-2\theta\Delta t})-\sigma_s\rho_{sx}(1-e^{-\theta\Delta t})\right]\right)^2 \\
&= -\frac{(n-1)}{2\hat{V}_{\tilde{A}}^2}\left(\frac{\hat{V}_{\tilde{A}}+w^2-\mathcal{H}(\sigma_s)}{\sigma_x}\right)^2,
\end{aligned}$$

where

$$w^2 = \frac{\sigma_x^2}{2\theta}(1-e^{-2\theta\Delta t})$$

and

$$\mathcal{H}(\sigma_s) = \sigma_s^2\left(\Delta t + \frac{1}{2\theta}\left(1-e^{-\theta\Delta t}\right)^2\right).$$

A rearrangement of the expression for $V_{\tilde{A}}$ at (3.20) and grouping allows the $\frac{\partial V_{\tilde{A}}}{\partial\sigma_x}$ term to be rewritten in terms of $V_{\tilde{A}}$, w^2 and $\mathcal{H}(\sigma_s)$. Interestingly, the w^2 term is simply the variance of the OU model.

Finally, application of Steps (2) to (4) give the standard errors for $\hat{\theta}$ and $\hat{\sigma}_x$:

$$\hat{\theta}^{SE} = \sqrt{\frac{\hat{V}_{\tilde{A}}}{\Delta t^2\hat{\phi}^2\sum_{i=2}^{n}\tilde{A}_{i-1}^2}} \quad \text{and} \tag{3.32}$$

$$\hat{\sigma}_x^{SE} = \sqrt{\frac{2\hat{V}_{\tilde{A}}^2\hat{\sigma}_x^2}{(n-1)(\hat{V}_{\tilde{A}}+\hat{w}^2-\mathcal{H}(\hat{\sigma}_s))^2}}. \tag{3.33}$$

3.3.2 Gibbs Sampling

3.3.2.1 Deriving the Conditional Posterior Distributions of the OUVLTM Parameters

As detailed in Appendix B.5.3 and mentioned at the beginning of Section 3.3, Gibbs sampling uses the conditional posterior distributions of the model parameters to generate posterior distributions of estimates under the Clifford-Hammersley theorem (Hammersley (1970) and Besag (1974)), which essentially proves that the joint distribution can be fully specified by a set of conditional distributions if certain conditions are met. Consider the reparametrised OUVLTM model described at (3.17) with the likelihood function (3.23). By using conjugate priors (explained in depth in Appendix B.4.1), the full conditional distributions for ϕ and $V_{\tilde{A}}$ are be derived for the reparametrised model in Theorem 3.1. Hoff (2009) derives the full conditionals of the mean and variance parameters for the typical univariate normal distribution in a similar manner.

Theorem 3.1. *Normal/inverse gamma priors are conjugates for the likelihood of a normal distribution, such as the reparametrised OUVLTM model. Hence the following prior distributions for ϕ and and $V_{\tilde{A}}$ are*

assumed:

$$p(\phi) \sim N(\mu_{\phi 0}, \tau_{\phi 0}^2) \quad \textit{and} \tag{3.34}$$

$$p(V_{\tilde{A}}) \sim \mathcal{IG}\left(\nu_0, \psi_0\right). \tag{3.35}$$

The full conditional distributions for ϕ and $V_{\tilde{A}}$ are then given by

$$p(\phi|V_{\tilde{A}}, \tilde{X}) \sim N(\mu_\phi, \tau_\phi^2) \quad \textit{and}$$
$$p(V_{\tilde{A}}|\phi, \tilde{X}) \sim \mathcal{IG}\left(\nu_n, \psi_n\right).$$

where

$$\mu_\phi = \frac{\frac{\mu_{\phi 0}}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \tilde{A}_{i-1}(\tilde{A}_i - \eta\Delta t)}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \tilde{A}_{i-1}^2}, \qquad \tau_\phi^2 = \frac{1}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \tilde{A}_{i-1}^2},$$
$$\nu_n = \frac{n-1}{2} + \nu_0 \qquad \textit{and} \qquad \psi_n = \frac{1}{2}\sum_{i=2}^n \left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2 + \psi_0.$$

Proof. The conditional posterior distribution for ϕ will be derived first, hence assume that $V_{\tilde{A}}$ is known and ϕ is unknown.

Taking the exponent of (3.23) gives

$$\mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}}) \propto \exp\left[-\frac{1}{2V_{\tilde{A}}}\sum_{i=2}^n \left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2\right].$$

Similarly, (3.34) yields

$$p(\alpha) \propto \exp\left[-\frac{1}{2\tau_{\alpha 0}^2}(\alpha - \mu_{\alpha 0})^2\right].$$

Thus using Bayes' rule at (B.1), the conditional distribution for ϕ can be written as

$$\begin{aligned} p(\phi|V_{\tilde{A}}, X) &\propto \mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}})p(\phi) \\ &\propto \exp\left[-\frac{1}{2V_{\tilde{A}}}\sum_{i=2}^n \left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2\right]\exp\left[-\frac{1}{2\tau_{\phi 0}^2}(\phi - \mu_{\phi 0})^2\right]. \end{aligned} \tag{3.36}$$

Consider now the terms in the exponents, omitting the $-\frac{1}{2}$ multiplier for the moment:

$$\begin{aligned} &\frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2 + \frac{1}{\tau_{\phi 0}^2}(\phi - \mu_{\phi 0})^2 \\ &= \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \left(\tilde{A}_i^2 - 2\tilde{A}_i\eta\Delta t + \eta^2\Delta t^2 + \phi^2\tilde{A}_{i-1}^2 - 2\phi\tilde{A}_i\tilde{A}_{i-1} + 2\phi\tilde{A}_{i-1}\eta\Delta t\right) + \frac{1}{\tau_{\phi 0}^2}\left(\phi^2 - 2\phi\mu_{\phi 0} + \mu_{\phi 0}^2\right) \\ &= \phi^2\left(\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \tilde{A}_{i-1}^2\right) - 2\phi\left(\frac{\mu_{\phi 0}}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \tilde{A}_{i-1}(\tilde{A}_i - \eta\Delta t)\right) + \frac{\mu_{\phi 0}^2}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^n \left(\tilde{A}_i - \eta\Delta t\right)^2 \\ &= a_\phi\phi^2 - 2b_\phi\phi + c_\phi, \end{aligned}$$

where

$$a_\phi = \frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}} \sum_{i=2}^{n} \tilde{A}_{i-1}^2, \quad b_\phi = \frac{\mu_{\phi 0}}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}} \sum_{i=2}^{n} \tilde{A}_{i-1}(\tilde{A}_i - \eta\Delta t) \quad \text{and} \quad c_\phi = \frac{\mu_{\phi 0}^2}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}} \sum_{i=2}^{n} \left(\tilde{A}_i - \eta\Delta t\right)^2.$$

Hence

$$\begin{aligned} p(\phi|V_{\tilde{A}}, \tilde{X}) &\propto \exp\left[-\frac{1}{2}(a_\phi \phi^2 - 2b_\phi)\right] \\ &= \exp\left[-\frac{1}{2}a_\phi\left(\phi^2 - \frac{2b_\phi}{a_\phi} + \frac{b_\phi^2}{a_\phi^2}\right) + \frac{b_\phi^2}{2a_\phi}\right] \quad \text{(by completing the square)} \\ &\propto \exp\left[-\frac{1}{2}a_\phi\left(\phi - \frac{b_\phi}{a_\phi}\right)^2\right] \\ &= \exp\left[-\frac{(\phi - \frac{b_\phi}{a_\phi})^2}{2/a_\phi}\right]. \end{aligned} \tag{3.37}$$

By referring to (A.6), (3.37) is identified as being normally distributed with mean and variance given respectively by

$$\mu_\phi = \frac{b_\phi}{a_\phi} = \frac{\frac{\mu_{\phi 0}}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^{n} \tilde{A}_{i-1}(\tilde{A}_i - \eta\Delta t)}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^{n} \tilde{A}_{i-1}^2} \qquad \text{and} \qquad \tau_\phi^2 = \frac{1}{a_\phi} = \frac{1}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}}\sum_{i=2}^{n} \tilde{A}_{i-1}^2}.$$

Thus the result for $p(\phi|V_{\tilde{A}}, \tilde{X})$ is proved.

Now assume that ϕ is known and $V_{\tilde{A}}$ is unknown. Again from (3.23), it is observed that

$$\mathcal{L}(\tilde{X}|\mu_s, \sigma_s, \phi, V_{\tilde{A}}) \propto \frac{1}{V_{\tilde{A}}^{\left(\frac{n-1}{2}\right)}} \exp\left[-\frac{1}{2V_{\tilde{A}}}\sum_{i=2}^{n}\left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2\right],$$

but now including the $V_{\tilde{A}}$ term outside the exponent.

The prior of $V_{\tilde{A}}$ is of the form

$$p(V_{\tilde{A}}) \propto \frac{1}{V_{\tilde{A}}^{(\nu_0+1)}} \exp\left[\frac{-\psi_0}{V_{\tilde{A}}}\right]$$

as implied by (3.35).

Thus

$$\begin{aligned} p(V_{\tilde{A}}|\phi, \tilde{X}) &\propto \frac{1}{V_{\tilde{A}}^{\left(\frac{n-1}{2}\right)}} \exp\left[-\frac{1}{2V_{\tilde{A}}}\sum_{i=2}^{n}\left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2\right] \frac{1}{V_{\tilde{A}}^{(\nu_0+1)}} \exp\left[-\frac{\psi_0}{V_{\tilde{A}}}\right] \\ &= \frac{1}{V_{\tilde{A}}^{\left(\frac{n-1}{2}+\nu_0+1\right)}} \exp\left[-\frac{1}{V_{\tilde{A}}}\left(\frac{1}{2}\sum_{i=2}^{n}\left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2 + \psi_0\right)\right]. \end{aligned} \tag{3.38}$$

When compared to (A.15), the conditional distribution of $V_{\tilde{A}}$ at (3.38) is confirmed as inverse gamma, with

shape and scale parameters given by

$$\nu_n = \frac{n-1}{2} + \nu_0 \qquad \text{and} \qquad \psi_n = \frac{1}{2}\sum_{i=2}^{n}\left(\tilde{A}_i - \phi\tilde{A}_{i-1} - \eta\Delta t\right)^2 + \psi_0$$

respectively.

□

3.4 Parameter Estimation of Simulated Processes

3.4.1 Maximum Likelihood Estimation

Using the same 1,500 sample path sets for $\ln(s)$, x and $\tilde{x}$ as Section 3.2, the parameters of each model are now estimated using MLE, where the mathematical theory for MLE of the OU model is developed in Appendix A.2.6, in particular Sections A.2.6.1, A.2.6.2 and A.2.6.3. Likewise, the theory for fitting GBM using MLE is derived in Appendix A.2.5.2, as calibration of the stock price is needed in the OUVLTM model.

The OUVLTM estimates are compared to those of OU, concentrating on bias and uncertainty in the form of standard errors at (3.30), (3.32), (A.47) (A.51). Goodness of fit is also considered, determined using the Root Mean Square Error (RMSE) given at (A.64) and the Mean Absolute Percentage Error (MAPE), given at (A.65) in Appendix A.2.9.

The simulated paths for OUVLTM and OU may be similar but are not the same, as they were generated using different models, so comparisons of calculated maximum likelihood, AIC and BIC are not appropriate. These measures can only be compared when using the same data.

The parameter estimates of both models obtained via MLE are summarised in Table 3.2 and Figure 3.6 plots the percentiles of the 1,500 estimates for μ_x, θ and σ_x. Note there is neither the α nor μ parameter in the OUVLTM model, hence only results $\phi, V_{\tilde{A}}, \theta$ and σ_x are given. Finally, the means and medians of the resultant 1,500 RMSEs and MAPEs are presented for both models in Table 3.3.

	Parameter	Mean estimate	Median estimate	Mean SE	Bias of mean estimate
OUVLTM	ϕ	0.9939	0.9946	0.0003	-0.0021
	$V_{\tilde{A}}$	6.39E-04	6.38E-04	2.86E-05	6.10E-06
	θ	1.5441	1.3616	0.0796	0.5441
	σ_x	0.4022	0.4021	0.0090	0.0022
OU	α	0.0234	0.0209	0.0008	0.0115
	ϕ	0.9920	0.9930	0.0003	-0.0040
	v	0.0251	0.0251	0.0006	4.02E-07
	μ_x	113.8163	2.9524	0.0163	110.8163
	θ	2.0261	1.7823	0.0702	1.0261
	σ_x	0.4008	0.4008	0.0090	0.0008

Table 3.2: Calculated means and medians of the 1,500 OUVLTM and OU parameter estimates via MLE, as well as the calculated means of the standard errors produced from the 1,500 estimates and the calculated mean biases.

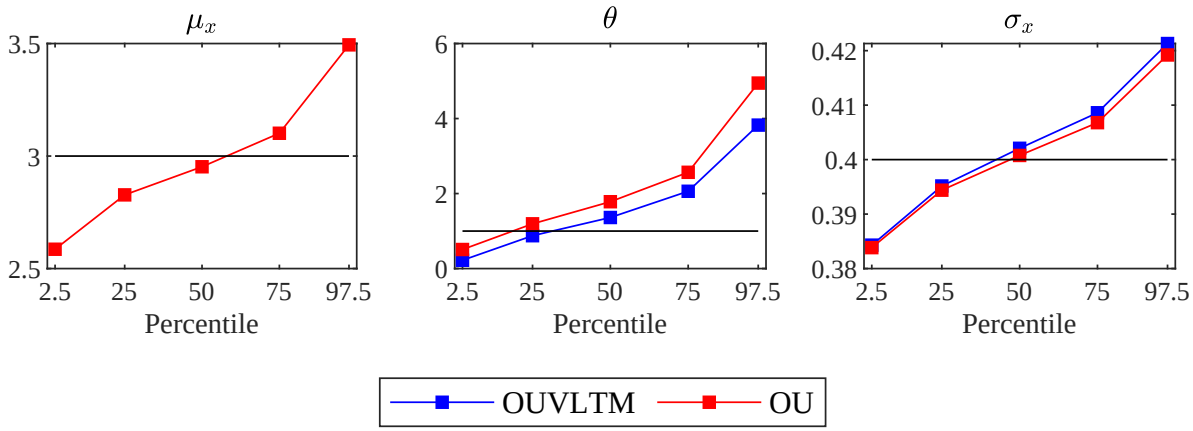


Figure 3.6: Percentiles of the 1,500 OU μ_x, θ, σ_x estimates and the 1,500 OUVLTM θ, σ_x estimates obtained via MLE.

	Goodness of fit test	OUVLTM	OU
Mean	RMSE	0.0295	0.0300
	MAPE	0.0452	0.0464
Median	RMSE	0.0104	0.0157
	MAPE	0.0343	0.0396

Table 3.3: Calculated means and medians of the 1,500 RMSE and MAPE values using the 1,500 OUVLTM and OU MLE parameter estimates.

To compare the OUVLTM variance $V_{\tilde{A}}$ with the OU variance v^2, the mean and median values of v^2, as shown in Table 3.4, are calculated by squaring the v estimates in Table 3.2. (A.50) is used to determine the standard error of v^2.

Parameter	Mean estimate	Median estimate	Mean SE	Mean bias
v^2	6.32E-04	6.32E-04	2.83E-05	1.94E-08

Table 3.4: Conversion of the results for v to v^2 for comparison to $V_{\tilde{A}}$.

The following observations were made about the MLE results:

- Both the OUVLTM and OU models overestimated[1] θ, but the OUVLTM model manages to reduce bias quite well. Figure 3.6 also exhibits that the OUVLTM estimates are lower than those of OU across all percentiles.
- The estimation of σ_x by OU was marginally better than OUVLTM. The OU bias was lower; 0.0008 vs 0.0022 for σ_x and 1.94E-08 vs 6.1E-06 for the variance parameters. The overestimation of σ_x by OUVLTM was slight, but was more apparent at higher percentiles, as seen in Figure 3.6.
- The average standard errors of the θ estimates were comparable (0.07 vs 0.08) and the average standard errors were equal for σ_x.
- The OUVLTM model produced noticeably smaller RMSEs and MAPEs values than the OU model.

The substantially high mean for OU's μ_x, in Table 3.2, was caused by its inverse exponential relationship with ϕ. Small increases in ϕ, especially those bringing it very close to 1, result in drastic increases in μ_x

[1]OU is known to do so, as discussed in Ball *et al.* (1996), Yu *et al.* (2001) and Tang *et al.* (2009).

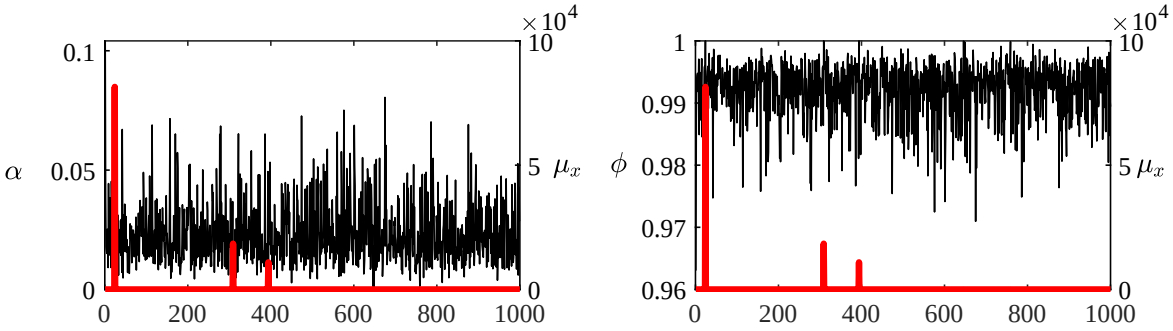


Figure 3.7: OU MLE estimates for μ_x relative to the OU α and ϕ MLEs.

and extreme positive outliers.

Figure 3.7 shows the hugely inflated μ_x estimates calculate from levels of α and ϕ that were not noticeably out of the ordinary. The median of μ_x was more in line with expectation in this case. The figure also shows that these outliers lay above the 97.5^{th} percentile, thus the probability of their occurrence was low.

3.4.2 Gibbs Sampling

Obtaining posterior distributions for the model parameters is achieved using Gibbs sampling of the conditional posterior distributions of the parameters. The conditional posterior distributions of the parameters of the OU and GBM models are developed in Appendix A.2.6.4 and A.2.5.4 respectively. Theorem thmconddist derived the OUVLTM parameters' conditional posterior distributions. The length of chains of parameter draws is set to 50,000 points, with a burn-in of 5,000 to eliminate any initial noise (refer to Appendix B.6.1 for a discussion on burn-in).

To put things into perspective, for the 1,500 MLE point estimates of the 4 OUVLTM parameters and the 6 OU parameters, there are now 1,500 posterior distributions, each posterior containing 45,000 elements. There is a lot of information to encapsulate and present, so the following points aim to provide clarity:

- The priors used for $\alpha, \phi, v/V_{\tilde{A}}$ are given in Table 3.5. A more detailed discussion on the choice of prior is documented later in this section.

Hyperparameters	$\mu_{\alpha 0}$	$\mu_{\phi 0}$	$\tau^2_{\alpha 0}$	$\tau^2_{\phi 0}$	ν_0	ψ_0
Prior values	0.5	0.5	1	1	1	0.01

Table 3.5: The values of the hyperparameters of the prior distributions used in the Gibbs sampling.

- The median of each posterior distribution is taken as the representation of the point estimate of the posterior distribution and is the equivalent of the MLE point estimate. The median is used instead of the mean as the median of a distribution is less affected by any outliers that the distribution contains.
- The standard deviations, skewnesses and excess kurtoses of the posteriors are recorded to assess the characteristics of the posteriors for this first presentation of Gibbs sampling results.

- Bias is taken as the mean of the medians of each posterior distribution minus the true parameter value.
- Standard errors for each element in each posterior are determined, resulting in 1,500 standard error distributions with 45,000 elements. Again the medians of these are taken as the representative point estimates.
- Convergence diagnostics for all the produced posterior distributions are documented in Appendix C.

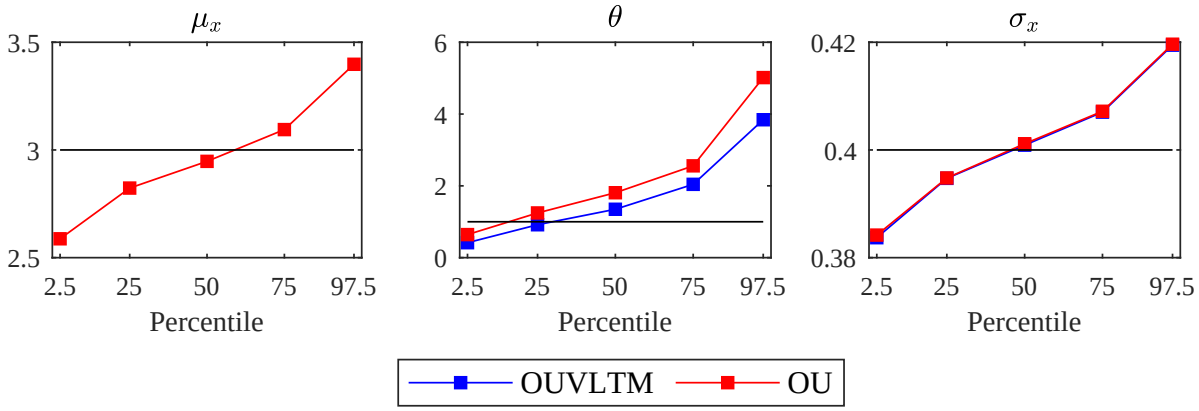


Figure 3.8: Percentiles of the 1,500 OU and OUVLTM point estimates obtained via Gibbs sampling.

The descriptive statistics of the posterior distributions for the OUVLTM and OU models are contained in Table 3.6 and the percentiles of the posterior point estimates are plotted in Figure 3.8. Lastly, the goodness of fit tests using these point estimates are given in Table 3.7.

	Parameter	Mean of estimate medians	Median of estimate medians	Mean of SE medians	Mean of standard deviations	Mean of skewnesses	Mean of excess kurtoses	Bias of mean of estimate medians
OUVLTM	ϕ	0.9938	0.9947	0.0003	0.0028	-0.3220	2.9020	-0.0021
	$V_{\tilde{A}}$	6.34E-04	6.34E-04	2.84E-05	2.85E-05	0.1792	3.0610	8.85E-07
	θ	1.5701	1.3481	0.0794	0.7109	0.3294	2.9107	0.5444
	σ_x	0.4010	0.4009	0.0090	0.0090	0.1121	3.0246	0.0006
OU	α	0.0239	0.0211	0.0008	0.0097	0.2421	2.8322	0.0115
	ϕ	0.9918	0.9928	0.0003	0.0034	-0.2446	2.8322	-0.0040
	v	0.0252	0.0252	0.0006	0.0006	0.1122	3.0240	2.18E-05
	μ_x	2.9648	2.9471	0.0069	48.3045	120.7468	20049.9249	-0.0073
	θ	2.0689	1.8084	0.0703	0.8626	0.2535	2.8400	1.0264
	σ_x	0.4012	0.4012	0.0090	0.0090	0.1128	3.0244	0.0008

Table 3.6: Descriptive statistics for the medians of the generated OUVLTM and OU parameter posteriors. Also shown are the means of the medians of the standard error posteriors, as well as the calculated bias of the mean of the estimate medians.

The OUVLTM parameter estimates, SEs and biases were fairly similar to the MLE results for ϕ and θ. For OU, the results were the almost identical for all parameters except for μ_x. Barring μ_x, the average skewness and excess kurtosis of the medians of the posterior distributions were not exceptionally far away from those of a normal distribution. The mean of the posterior medians of the OU parameter μ_x was substantially smaller the the MLE estimate for μ_x, leading to greatly reduced bias. Even taking the mean of the posterior

	Goodness of fit test	OUVLTM	OU
Mean	RMSE	0.0295	0.0300
	MAPE	0.0452	0.0460
Median	RMSE	0.0109	0.0155
	MAPE	0.0354	0.0395

Table 3.7: Calculated means and medians of the 1,500 RMSE and MAPE values using the Gibbs sampled parameter point estimates.

means, instead of the mean of the medians, gave a significantly improved estimate of 3.3841. The Gibbs sampler also experienced spikes in μ_x values, but these were toned down when the larger distribution was taken into account. However, the turbulence within the chains containing spikes was still reflected in the high standard deviation, skewness and kurtosis excess levels across the posteriors.

Gibbs sampling produced OUVLTM $V_{\tilde{A}}$ and σ_x posterior means and medians that were less biased than the MLE values, comparing Tables 3.2 and 3.6. In fact, the bias of the σ_x posterior estimate is slightly lower than the MLE and Gibbs sampler σ_x bias values for OU. Consequently, the percentiles of σ_x in Figure 3.8 for OUVLTM and OU were almost indistinguishable. Figure 3.8 also highlights that for OUVLTM, the Gibbs sampler, like MLE, did not overestimate θ as much as OU does.

Table 3.7 shows RMSEs and MAPEs means and medians that were in the same regions as the MLE values. And it shows that once again, the values for OUVLTM were smaller than those of OU, indicative of a calibration capability that is on a par with OU.

Now, the standard deviation of a posterior distribution of a parameter and the standard error of the parameter estimate are both measures of the estimates' deviation from the true parameter value. Thus one would expect them to be equal, but this is not necessarily the case. Table 3.6 shows posterior standard deviations that were much larger than the average standard errors calculated from the posterior estimates, except for $v, V_{\tilde{A}}$ and both OUVLTM and OU's σ_x. The difference arises due to the choice of prior used in the sampling.

Consider the OUVLTM ϕ parameter for example, that has the following standard error and posterior standard deviation,

$$\hat{\phi}^{SE} = \sqrt{\frac{\hat{V}_{\tilde{A}}}{\sum_{i=2}^{n} \tilde{A}_{i-1}^2}} \quad \text{and} \quad \tau_\phi = \sqrt{\frac{1}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_{\tilde{A}}} \sum_{i=2}^{n} \tilde{A}_{i-1}^2}},$$

obtained from (3.30) and Theorem 3.1 respectively. It is evident that the choice of the prior standard deviation $\tau_{\phi 0}$ is critical in determining the overall standard deviation of the posterior. However, an appropriate prior could align the standard error and posterior standard distribution to follow the intuition that they should be equivalent. For notational ease, the maximum likelihood estimate $\hat{V}_{\tilde{A}}$ is rewritten as V_{MLE} and let a draw from the Gibbs sampler chain $V_{\tilde{A}}$ be denoted by V_G. Let the prior used in the Gibbs sampling equal the standard error of the parameter estimate, that is

$$\tau_{\phi 0} = \hat{\phi}^{SE} = \sqrt{\frac{V_{MLE}}{\sum_{i=2}^{n} \tilde{A}_{i-1}^2}}.$$

Substituting $\tau_{\phi 0}$ into the expression for the posterior standard deviation yields:

$$\begin{aligned}\tau_\phi = \sqrt{\frac{1}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{V_G}\sum_{i=2}^{n}\tilde{A}_{i-1}^2}} &= \sqrt{\frac{V_G\tau_{\phi 0}^2}{V_G + \tau_{\phi 0}^2\sum_{i=2}^{n}\tilde{A}_{i-1}^2}} \\ &= \sqrt{\frac{V_G\frac{V_{MLE}}{\sum_{i=2}^{n}\tilde{A}_{i-1}^2}}{V_G + \frac{V_{MLE}}{\sum_{i=2}^{n}\tilde{A}_{i-1}^2}\sum_{i=2}^{n}\tilde{A}_{i-1}^2}} \\ &= \sqrt{\frac{V_G V_{MLE}}{(V_G + V_{MLE})\sum_{i=2}^{n}\tilde{A}_{i-1}^2}}.\end{aligned}$$

Thus only when $V_{MLE} \approx V_G$, the standard error of the MLE and standard deviation of the posterior are roughly equal. The Gibbs sampler is rerun on the 1,500 paths, now using hyperparameters set to the MLE estimates and standard errors of each path. The results are given in Table 3.8.

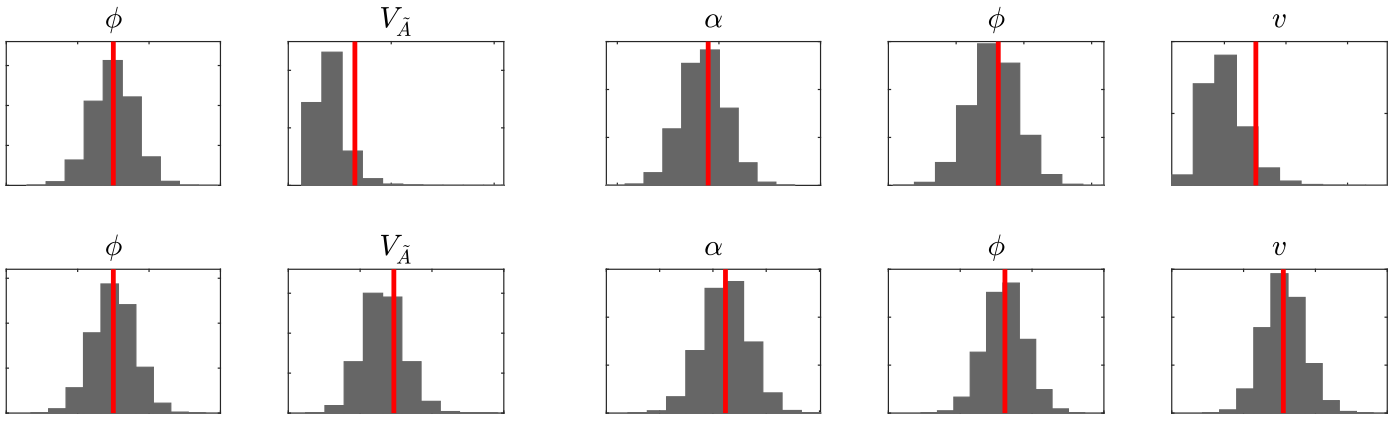
	Parameter	Mean of estimate medians	Median of estimate medians	Mean of SE medians	Mean of standard deviations	Mean of skewnesses	Mean of excess kurtoses	Bias of mean of estimate medians
OUVLTM	ϕ	0.9939	0.9946	0.0003	0.0003	-0.0174	3.0086	-0.0021
	$V_{\tilde{A}}$	6.33E-04	6.33E-04	2.83E-05	2.84E-05	0.1796	3.0602	8.85E-07
	θ	1.5444	1.3614	0.0793	0.0786	0.0183	3.0086	0.5444
	σ_x	0.4006	0.4005	0.0090	0.0090	0.1120	3.0238	0.0006
OU	α	0.0234	0.0209	0.0008	0.0006	0.0023	2.9991	0.0115
	ϕ	0.9920	0.9930	0.0003	0.0002	-0.0047	3.0043	-0.0040
	v	0.0251	0.0251	0.0006	0.0006	0.1113	3.0244	-3.79E-06
	μ_x	2.9927	2.9525	0.0073	5.0434	1.4358	201.9333	-0.0073
	θ	2.0264	1.7825	0.0702	0.0765	0.0063	3.0031	1.0264
	σ_x	0.4008	0.4007	0.0090	0.0090	0.1111	3.0226	0.0008

Table 3.8: Calculated results of the Gibbs sampler run using MLE standard error priors instead of uninformed priors.

Relative to the uninformative priors used originally (Table 3.5), the MLE priors produce posterior distribution estimates for the OUVLTM and OU model parameters with mostly lowered bias levels. The posteriors appeared to be more Gaussian-like, with less skewness and excess kurtosis values closer to 3. The similarities of the standard errors calculated using the posterior values and posterior standard deviations are also evident in Table 3.8, with the exception of OU's μ_x.

Again this difference was due to the few extreme values of μ_x already mentioned. The outliers within a Gibbs sampler chain for μ_x caused its elevated standard deviation, but the corresponding shocks in standard errors were not as potent and so were neutralised across the distribution. Using informed priors is obviously preferable as the standard of the sampling is improved. But in situations where the supposed informed prior becomes unreliable (such as in small data sets), using the MLE standard errors limits the Gibbs sampler's ability to explore the parameter domain sufficiently.

The posteriors of the variance parameters of the OUVLTM and OU models ($V_{\tilde{A}}$ and v) produced by the Gibbs sampler emerged looking normally distributed, even though the conditional posterior distributions that the algorithm sampled from are inverse gamma (proven in Theorems 3.1 and A.4). The length of sample path is the reason for this behaviour. Figures 3.9a and 3.9b show the posterior distributions of the



(a) OUVLTM (b) OU

Figure 3.9: Posterior distributions of the the parameters for the OUVLTM and OU models. The top plots show the posteriors from a short sample path of length of 20 and the bottom plots display the posteriors for sample path of the default length of 1,000. The vertical red lines are the corresponding MLE values.

OUVLTM and OU parameters for a short path (n=20) on the top and those of a sample path of default path length 1,000 at the bottom. The vertical red lines are the corresponding MLE values as per each model.

For the short path, the posteriors of $V_{\tilde{A}}$ and v were clearly asymmetrical with extended positive tails, as expected from an inverse gamma distribution. But for the longer sample path data set, the inverse gamma nature of the posteriors was much less apparent. The distributions were more symmetrical and tend towards normality, as predicted by the *Central Limit Theorem*. There was a semblance of left tilted distributions though when looking at the consistently positive skewness levels of the $V_{\tilde{A}}, v$ and σ_x in Tables 3.6 and 3.8. The posteriors of the other parameters remained Gaussian-like, as anticipated by the assumption that they were normally distributed, irrespective of the sample length.

Overall, the comparison of MLE versus Gibbs sampling as parameter estimation techniques reveals that the Bayesian method provides a richer understanding of the sample data. It takes prior knowledge of parameters into account and produces a range of intuitive estimates of the model parameters and their uncertainties, in a probabilistic framework that is shown to moderate extreme behaviour.

4. Missing Data

This chapter describes how missing data[1] is handled, the typical mathematical and statistical methods and then the methodologies employed specifically for this research.

But first, a practical demonstration, using the same simulated data sets introduced the Chapter 3, is performed. The sample paths are taken and reduced to form shorter or incomplete paths, through the removal of a varying number of data points. The removal mechanism is random and deletes points based on a replica of historical market trades. The lack of frequency of trades was discussed at length in Chapter 2. To mimic the behaviour of historical market trades sensibly, it is also crucial to take into account the homogeneity of the trade days - were the trades evenly spaced or did they cluster? The parameters of these incomplete data sets for OUVLTM and OU are then estimated, again using MLE and Gibbs sampling. Observing their estimate bias and goodness of fit, when compared to the complete data, provides a clear illustration as to the extent of the loss of statistical significance and power in the presence of missing data.

The technique used to account for the missing data in this study is a multiple imputation Bayesian method. The approach involves simultaneously "filling in" the missing observationsin the Gibbs algorithm whilst sampling the model parameters. Imputing the missing values of the paths results in a new set of complete paths, whose estimates, uncertainties and goodness of fit are judged relative to those of the incomplete and original paths.

4.1 Profiling the Clustering of Market Trades

The agglomeration of trades is taken here to mean, in informal terms, the number of times 1 singular trade occurred during the period, the number of times 2 trades occurred consecutively and so forth. So for example, for instruments that are traded daily, one would say that the number of times n trades occurred consecutively in the period is equal to one, where n is the number of days in the period.

Figure 4.1 shows the number of trades that happened in 1 day, 2 day, 3 day batches etc. for two liquid bonds and two illiquid South African bonds. MTN01, ABS3 and ABL6 were introduced in Chapter 2 and CBL01 was a Capitec Bank bond that matured in May 2011, with TFR = 3.45%. To interpret these graphs, consider MTN01 for example. The first bar indicates that 65 single day trades were observed, meaning there were 65 instances of a trade happening that was not preceded or followed by another trade. The second bar shows the the bond was traded two days in a row 43 times, so 43 of the trades were either preceded or followed by another trade. The logic follows for rest of the bars.

Exponential decay type functions are evident in Figure 4.1, where the number of times a bond traded n consecutive days fell off rapidly with increasing n. For the more liquid bonds the decay was slower, implying that there were more instances of the bond trading for multiple days in a row. The less liquid bonds predominantly traded for only one or two days, with very few, or no, instances of multiple consecutive

[1]Missing data here refers to the case where credit spread levels were unobtainable and unknown, whereas other meanings of missing data include data that was taken but lost e.g. clinical trials

trades.

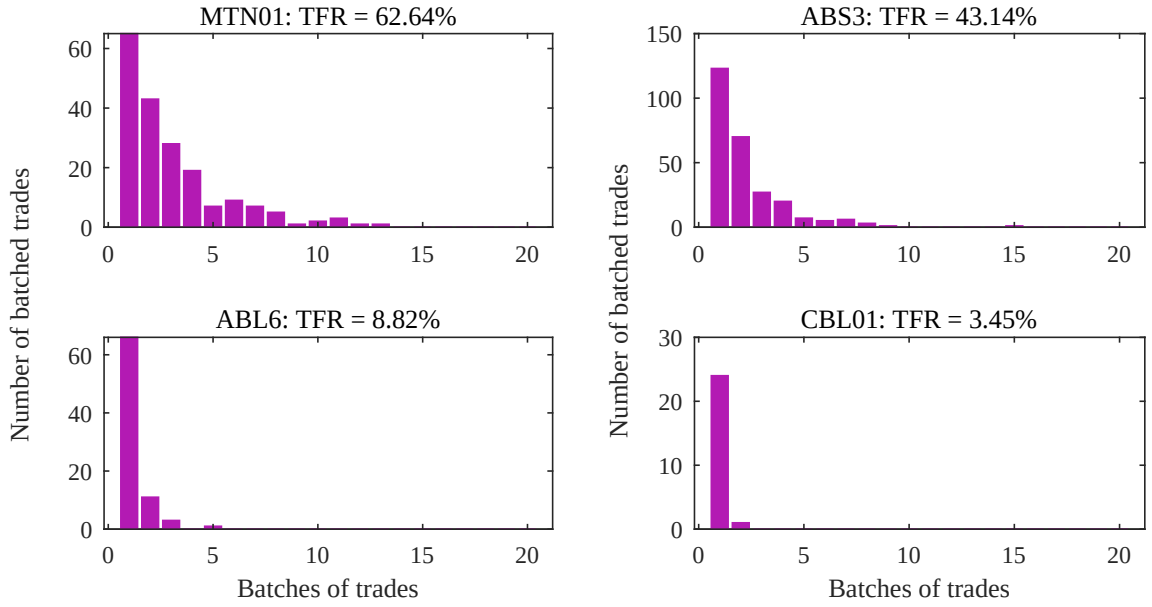


Figure 4.1: The clustering of trades for South African bonds with higher to lower liquidity. Liquid bonds show more consecutive trades and illiquid bonds tend to trade in isolation.

To replicate this behaviour pattern of missing trades, 10%, 50% and 90% of the points from the simulated data sets are removed. A $1,000 \times 1$ uniform distribution is then generated such that the probabilities across the domain (0,1) are equal so as to get an unbiased and perfectly uniform distribution. Figure 4.2 shows this distribution on the right, with a singular draw of 1,000 uniform random variables for comparison on the left. The details of this calculation are documented in Chapter 6.

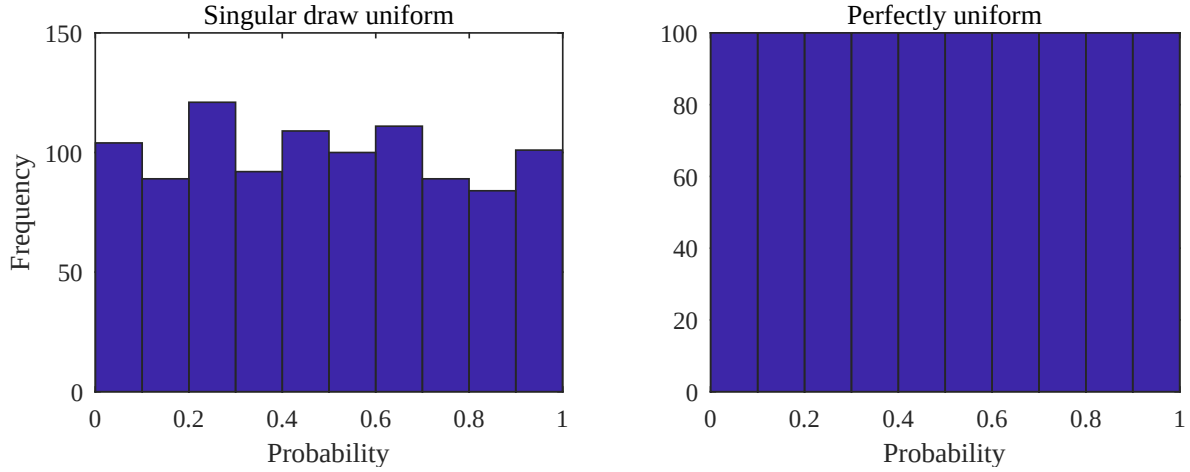


Figure 4.2: The perfectly uniform distribution used in determining omission of points versus a singular uniform distribution draw.

The omitted points are determined by the comparison of the omission rate, $\lambda = 10\%, 50\%, 90\%$, against each element in the $1,000 \times 1$ uniform distribution. If an element is less than λ, the corresponding point in the sample path is removed, and if it is greater, the point is left in the in the data set. The result is recorded by three binary series of length 1,000, termed inclusion indicators and denoted by K^{λ} (see (4.1), where 0 indicates an omission (representative of the days of no trade).

Figure 4.3 shows the analogous plots as Figure 4.1 for each K^{λ} series corresponding to $\lambda = 10\%, 50\%, 90\%$. The K^{λ} series are seen to replicate the bundling of the market bond trades well. As λ increases (liquidity decreases), the number of occurrences of multiple consecutive trades decreases dramatically, with the same exponential decay behaviour.

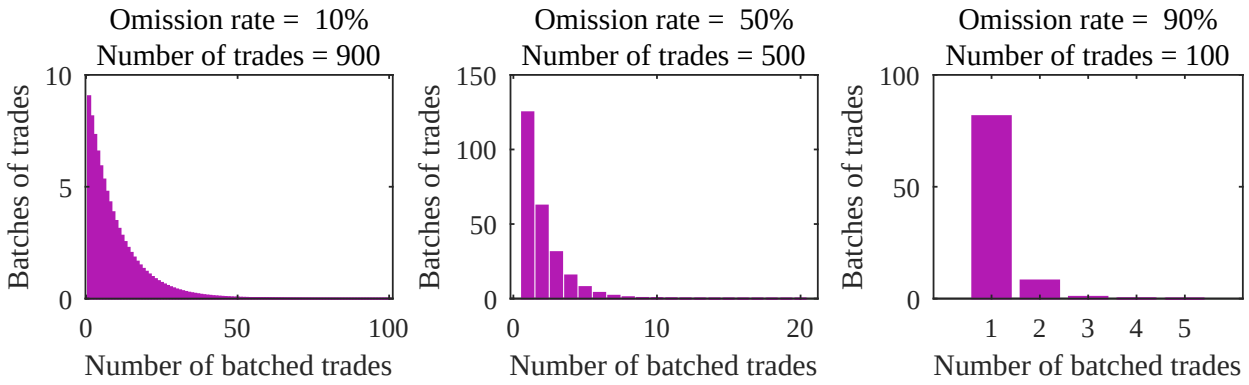


Figure 4.3: The clustering of simulated trades for omission rates of $\lambda = 10\%, 50\%$ and 90%.

4.2 Parameter Estimation of Incomplete Simulated OU and OUVLTM Processes

Applying K^{λ} for $\lambda = 10\%, 50\%, 90\%$, yields 3 additional sample path data sets for OUVLTM and OU to that of the complete data set, which is represented by zero omission rate, $\lambda = 0\%$. The lengths of these data sets are 1,000, 900, 500 and 100 respectively - the process shortens the original complete sample paths by the degree specified by λ.

4.2.1 Maximum Likelihood Estimation

λ		0%		10%		50%		90%	
Model	Parameter	Mean estimate	Median estimate	Mean estimate	Median estimate	Mean estimate	Median estimate	Mean estimate	Median estimate
OUVLTM	ϕ	0.9939	0.9946	0.9933	0.9941	0.9882	0.9897	0.9434	0.9493
	$V_{\tilde{A}}$	0.0006	0.0006	0.0007	0.0007	0.0013	0.0013	0.0060	0.0058
	θ	1.5441	1.3616	1.6014	1.4045	2.0826	1.8018	3.8074	3.3575
	σ_x	0.4022	0.4021	0.4100	0.4098	0.4567	0.4556	0.5664	0.5613
OU	α	0.0234	0.0209	0.0263	0.0234	0.0462	0.0410	0.2268	0.2019
	ϕ	0.9920	0.9930	0.9910	0.9921	0.9842	0.9860	0.9225	0.9326
	v	0.0251	0.0251	0.0265	0.0265	0.0354	0.0353	0.0759	0.0755
	μ_x	113.8163	2.9524	6712.6350	2.9527	120.0079	2.9556	11148.2870	2.9474
	θ	2.0261	1.7823	2.1531	1.8966	2.7792	2.4538	5.2901	4.5013
	σ_x	0.4008	0.4008	0.4091	0.4088	0.4566	0.4550	0.5649	0.5633

Table 4.1: Calculated means and medians of the maximum likelihood estimates of the complete and incomplete sample path sets, reduced by $\lambda = 10\%, 50\%$ and 90%.

The OUVLTM and OU models are now calibrated using maximum likelihood estimation to the incomplete ($\lambda = 10\%, 50\%$ and 90%) sample paths. The means and medians of the 1,500 maximum likelihood estimates

are presented in Table 4.1, where the complete case ($\lambda = 0\%$) results from Section 3.4.1 are included as well. All the means and medians of the parameter estimates (except for μ_x) for both models became less accurate and deviated more from the true parameter value as the omission rate increased. Table 4.2 shows that the mean biases increased as a result. This growing divergence is also evident in Figures 4.4 and 4.5.

The observations about θ and σ_x from the complete case still hold here: the biases of the OUVLTM θ were still less than those of OU as omission rates rose and the estimates of OUVLTM σ_x gave marginally higher biases than OU for all omission rates.

λ		0%		10%		50%		90%	
Model	Parameter	Mean SE	Bias of mean	Mean SE	Bias of mean	Mean SE	Bias of mean	Mean SE	Bias of mean
OUVLTM	ϕ	0.0003	-0.0021	0.0003	-0.0027	0.0006	-0.0079	0.0030	-0.0526
	$V_{\tilde{A}}$	0.0000	0.0000	0.0000	0.0001	0.0001	0.0006	0.0008	0.0053
	θ	0.0796	0.5441	0.0836	0.6014	0.1100	1.0826	0.2063	2.8074
	σ_x	0.0090	0.0022	0.0098	0.0100	0.0154	0.0567	0.0504	0.1664
OU	α	0.0008	0.0115	0.0009	0.0144	0.0016	0.0343	0.0076	0.2150
	ϕ	0.0003	-0.0040	0.0003	-0.0050	0.0006	-0.0118	0.0026	-0.0736
	v	0.0006	0.0000	0.0006	0.0013	0.0011	0.0103	0.0053	0.0508
	μ_x	0.0163	110.8163	0.0722	6709.6350	0.0214	117.0079	0.3609	11145.2870
	θ	0.0702	1.0261	0.0738	1.1531	0.0975	1.7792	0.1850	4.2901
	σ_x	0.0090	0.0008	0.0096	0.0091	0.0144	0.0566	0.0397	0.1649

Table 4.2: Calculated mean standard errors and biases of the MLE parameter estimate means of the complete and incomplete sample path sets, reduced by $\lambda = 10\%, 50\%$ and 90%.

Table 4.2 also gives the mean standard errors for each maximum likelihood estimate at each λ. Like the biases, the standard errors became larger as the omission rate increased, again evidence of a decrease in estimation precision.

Following suit, the mean and median RMSE and MAPE values, shown in Table 4.3 increased with omission rate. The error values for OUVLTM were below those of OU at each level of omission, as was the case using the complete sample paths.

Model		OUVLTM				OU			
λ		0%	10%	50%	90%	0%	10%	50%	90%
Mean	RMSE	0.0295	0.0301	0.0339	0.0407	0.0300	0.0313	0.0367	0.0564
	MAPE	0.0452	0.0457	0.0490	0.0549	0.0464	0.0474	0.0516	0.0601
Median	RMSE	0.0104	0.0109	0.0125	0.0153	0.0157	0.0165	0.0196	0.0242
	MAPE	0.0343	0.0352	0.0380	0.0426	0.0396	0.0401	0.0441	0.0472

Table 4.3: Calculated mean and median RMSE and MAPE levels resulting from the MLE estimates of the complete and incomplete OUVLTM and OU sample path sets, reduced by $\lambda = 10\%, 50\%$ and 90%.

The maximum likelihood estimation of μ_x in the OU model produced extreme outliers for the incomplete sample paths as was seen for the complete sample paths, implied by the hugely inflated means and mean biases in Tables 4.1 and 4.2. However, the medians weren't significantly affected by the omission of points and remained relatively unchanged as λ increased. The similar medians of the maximum likelihood estimates of μ_x for the complete and incomplete data sets suggest that the distributions of the estimates only

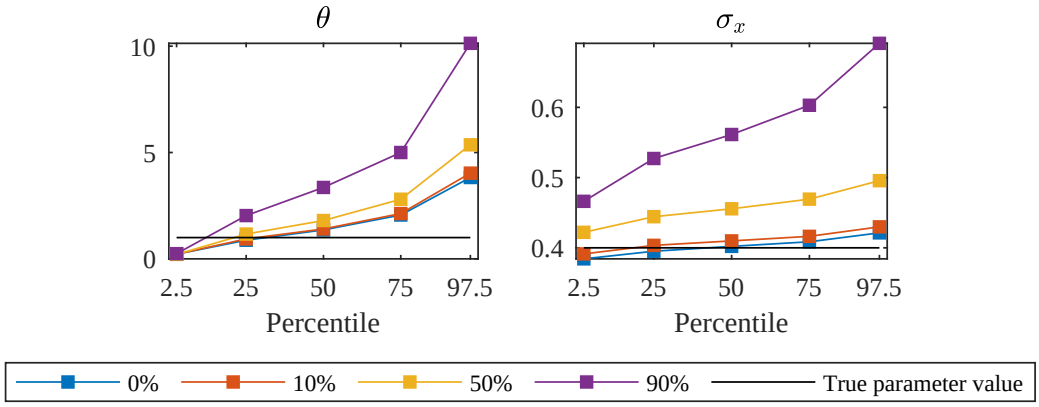


Figure 4.4: Percentiles of the 1,500 MLE parameter estimates for the complete and incomplete OUVLTM sample paths.

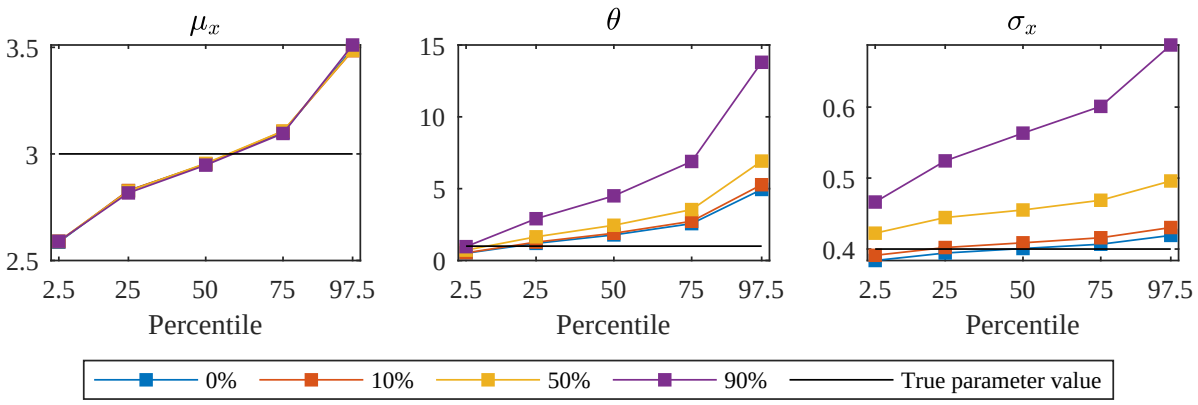


Figure 4.5: Percentiles of the 1,500 MLE parameter estimates for the complete and incomplete OU sample paths.

differed materially by their dissimilar sized outliers.

Figure 4.5 confirms that these outliers for μ_x were above the 97.5^{th} percentile for all the incomplete data sets and that the estimate percentiles values are almost identical for each omission rate. The unchanging distribution of μ_x can be explained by considering the fact that

$$\mu_x = \frac{\alpha}{1-\phi}.$$

The α maximum likelihood estimates are plotted against the $1-\phi$ estimates corresponding to each λ in Figure 4.6. The graphs show a distinct linear relationship between α and $1-\phi$, with constant slope representing μ_x.

The slopes of the plots are around 2.8 for all omission rates. The reason that the slopes are less than the median values is again due to the length of the data sets. The bottom graph in Figure 4.6 plots the maximum likelihood estimates of α against $1-\phi$ for sample paths of length 20,000. The slope using these longer paths is seen to be much closer to true μ_x value of 3.

To summarise, as the number of points removed increased (as the omission rate, λ, grew), the maximum likelihood estimates generally incurred more bias, produced larger standard errors and yielded greater RMSE and MAPE values.

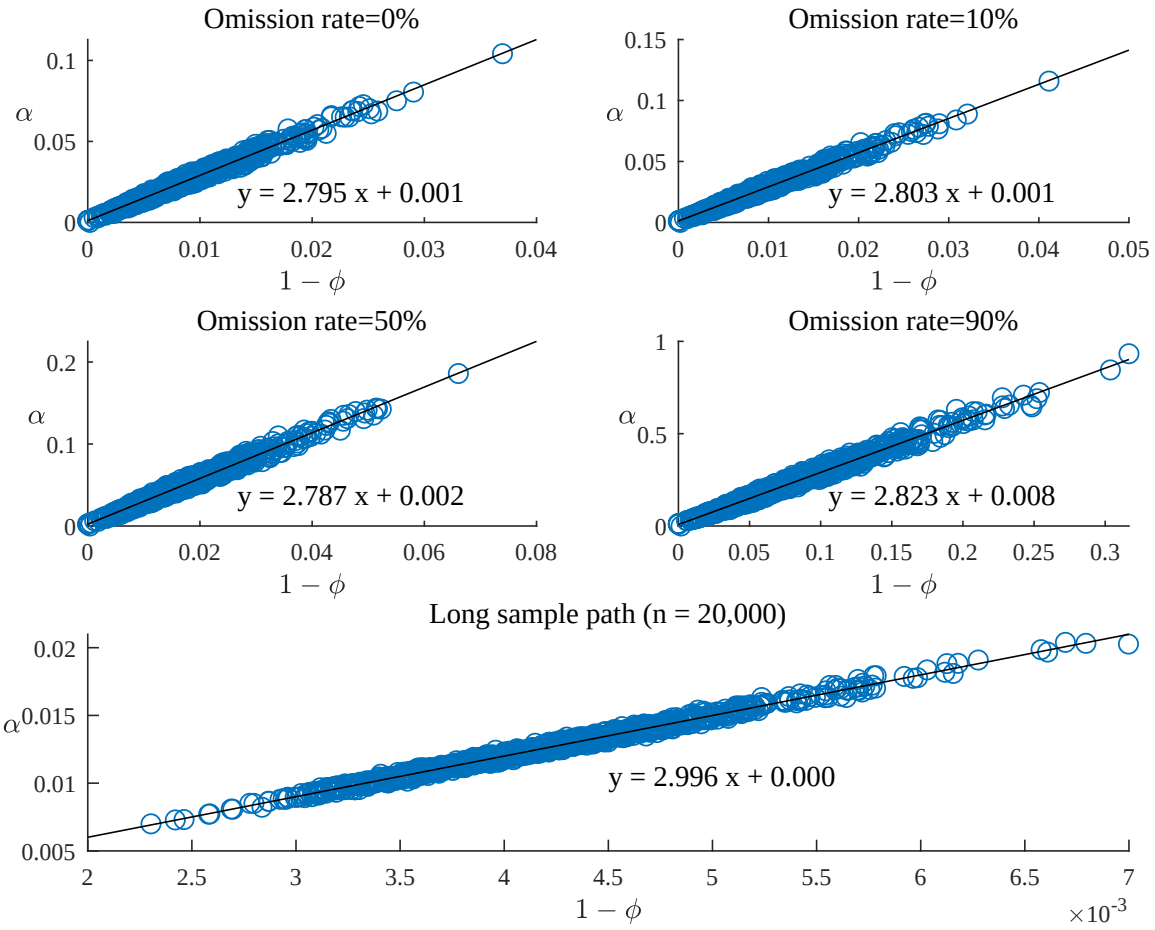


Figure 4.6: α plotted against $1-\phi$ for the complete and incomplete OU sample paths, reduced by $\lambda = 10\%, 50\%$ and 90%. The last chart is the same plot, just for a much longer sample path length.

4.2.2 Gibbs Sampling

As before, the the conditional posterior distributions of the OUVLTM and OU model parameters are repeatedly sampled to form posterior distributions via the Gibbs algorithm. Now posterior distributions for the models' parameters using the 3 incomplete data sets are generated.

The mean and the median of the medians of resultant parameter posterior distributions for the incomplete data sets are contained in Table 4.4, as are those of the complete data set ($\lambda = 0\%$) from Section 3.4.2. Similar to the MLE applied to the incomplete data sets, the mean and median estimates of all parameter posterior distributions generated using Gibbs sampling for both models became less accurate as the omission rate increased. The median estimates for θ that the Gibbs sampler produced using both models were, however, higher than those obtained in the MLE case. As the omission rate increased, the relative differences between the MLE and Gibbs estimates grew as well.

μ_x again was the exception, where the mean and median (of the median) posterior values of μ_x did not deteriorate as the omission rates increased - the explanation given for the MLE case applies here as well. The means of the median posterior μ_x values were also noticeably lower than the sizable MLE means seen in Table 4.1 at the higher omission rates (as was observed in the complete data analysis).

The relatively consistent estimates for μ_x and the rising deviations of the other parameter estimates from

λ		0%		10%		50%		90%	
Model	Parameter	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians
OUVLTM	ϕ	0.9938	0.9947	0.9931	0.9941	0.9876	0.9893	0.9404	0.9474
	V	0.0006	0.0006	0.0007	0.0007	0.0013	0.0013	0.0060	0.0059
	θ	1.5701	1.3481	1.6451	1.4113	2.1754	1.8686	4.0100	3.4889
	σ_x	0.4010	0.4009	0.4093	0.4088	0.4572	0.4559	0.5692	0.5637
OU	α	0.0239	0.0211	0.0269	0.0238	0.0474	0.0418	0.2350	0.2084
	ϕ	0.9918	0.9928	0.9908	0.9920	0.9838	0.9858	0.9196	0.9291
	v	0.0252	0.0252	0.0265	0.0265	0.0355	0.0354	0.0765	0.0760
	μ_x	2.9648	2.9471	2.9662	2.9490	2.9653	2.9481	2.9530	2.9364
	θ	2.0689	1.8084	2.2022	1.9230	2.8512	2.4965	5.4886	4.7488
	σ_x	0.4012	0.4012	0.4095	0.4092	0.4575	0.4560	0.5704	0.5683

Table 4.4: Calculated means and medians of the medians of the generated parameter posterior distributions for the complete and incomplete OUVLTM and OU sample paths, reduced by $\lambda = 10\%, 50\%$ and 90%.

their true values can be seen in Table 4.5 and are observable in Figures 4.7 and 4.8.

λ		0%	10%	50%	90%
OUVLTM	ϕ	-0.0022	-0.0029	-0.0084	-0.0557
	V	0.0000	0.0001	0.0006	0.0054
	θ	0.5701	0.6451	1.1754	3.0100
	σ_x	0.0010	0.0093	0.0572	0.1692
OU	α	0.0120	0.0150	0.0355	0.2231
	ϕ	-0.0042	-0.0052	-0.0122	-0.0765
	v	0.0000	0.0014	0.0103	0.0514
	μ_x	-0.0352	-0.0338	-0.0347	-0.0470
	θ	1.0689	1.2022	1.8512	4.4886
	σ_x	0.0012	0.0095	0.0575	0.1704

Table 4.5: Calculated biases of the means of the generated parameter posterior distribution medians for the complete and incomplete OUVLTM and OU sample paths, reduced by $\lambda = 10\%, 50\%$ and 90%.

A change in bias of μ_x is only really noticeable for $\lambda = 90\%$, whilst the biases of the other parameters distinctly rose with the omission rate. Table 4.5 also shows that the bias of the OUVLTM θ estimates were still less than those of OU for each non-zero omission rate. The biases of σ_x of OUVLTM were also smaller than the OU σ_x biases, but the reduction is minimal.

The means of the standard error posterior medians and standard deviations of the parameter posteriors are given in Table 4.6. The standard errors of μ_x were lower than those produced by MLE, as the very high ϕ values that caused the shocks in μ_x were being averaged out in each posterior. So reasonable deterioration of the SE levels was clear as omission rates increased. The same can be seen for the other OU parameters and the OUVLTM parameters, where the mean of the median of the standard error posteriors also increased with omission rate. In addition, the Gibbs sampler SEs for these parameters were very similar to those produced by MLE. Like the standard errors, the mean standard deviations of the parameter posterior distributions also increased with omission rate. Again the mean standard deviation of the μ_x posteriors

λ		0%		10%		50%		90%	
Model	Parameter	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations
OUVLTM	ϕ	0.0003	0.0028	0.0003	0.0031	0.0006	0.0056	0.0030	0.0270
	V	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0008	0.0009
	θ	0.0794	0.7109	0.0834	0.7457	0.1102	0.9880	0.2065	1.8928
	σ_x	0.0090	0.0090	0.0097	0.0097	0.0154	0.0146	0.0506	0.0419
OU	α	0.0008	0.0097	0.0009	0.0108	0.0016	0.0192	0.0076	0.0929
	ϕ	0.0003	0.0034	0.0003	0.0038	0.0006	0.0067	0.0027	0.0324
	v	0.0006	0.0006	0.0006	0.0006	0.0011	0.0011	0.0054	0.0055
	μ_x	0.0069	48.3045	0.0073	28.2317	0.0098	48.3118	0.0217	38.9391
	θ	0.0703	0.8626	0.0739	0.9085	0.0972	1.1917	0.1849	2.3291
	σ_x	0.0090	0.0090	0.0097	0.0097	0.0145	0.0146	0.0401	0.0426

Table 4.6: Calculated means of the medians of the generated standard error posteriors and the calculated mean of the standard deviations of the generated parameter estimate posterior distributions for the complete and incomplete OUVLTM and OU sample paths, reduced by $\lambda = 10\%, 50\%$ and 90%.

reflected the presence of extreme outliers, as expected.

Gibbs sampling produced the same results for the goodness of fit tests as MLE, shown in Table 4.7, where the mean and median RMSE and MAPE values rose with omission rate. Furthermore, OUVLTM still achieved smaller fit errors than OU at all omission rates.

Model		OUVLTM				OU			
λ		0%	10%	50%	90%	0%	10%	50%	90%
Mean	RMSE	0.0295	0.0305	0.0347	0.0421	0.0300	0.7886	3.3998	6.1047
	MAPE	0.0452	0.0462	0.0500	0.0563	0.0460	0.1106	0.1469	0.1586
Median	RMSE	0.0109	0.0114	0.0133	0.0181	0.0155	0.0318	0.0394	0.0451
	MAPE	0.0354	0.0363	0.0398	0.0469	0.0395	0.0591	0.0636	0.0670

Table 4.7: Calcualted mean and median RMSE and MAPE levels resulting from the medians of the generated parameter posteriors of complete and incomplete OUVLTM and OU sample path sets, reduced by $\lambda = 10\%, 50\%$ and 90%.

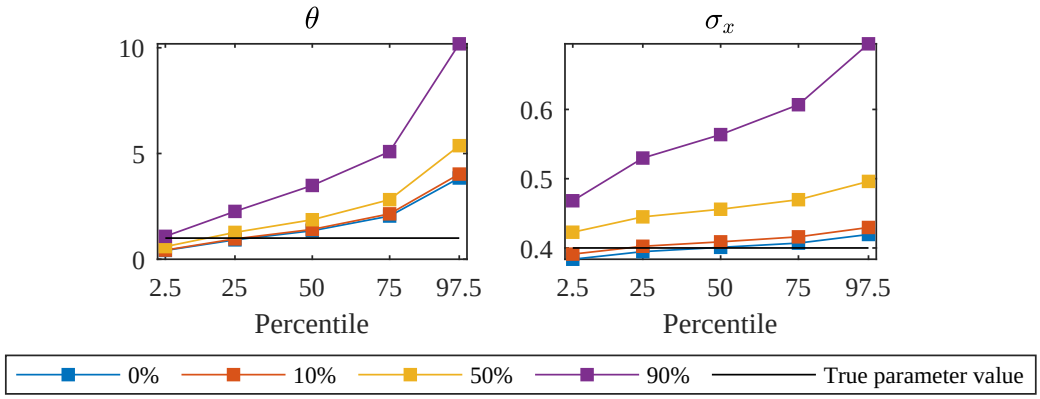


Figure 4.7: Percentiles of the 1,500 parameter posterior medians for the complete and incomplete OUVLTM sample paths.

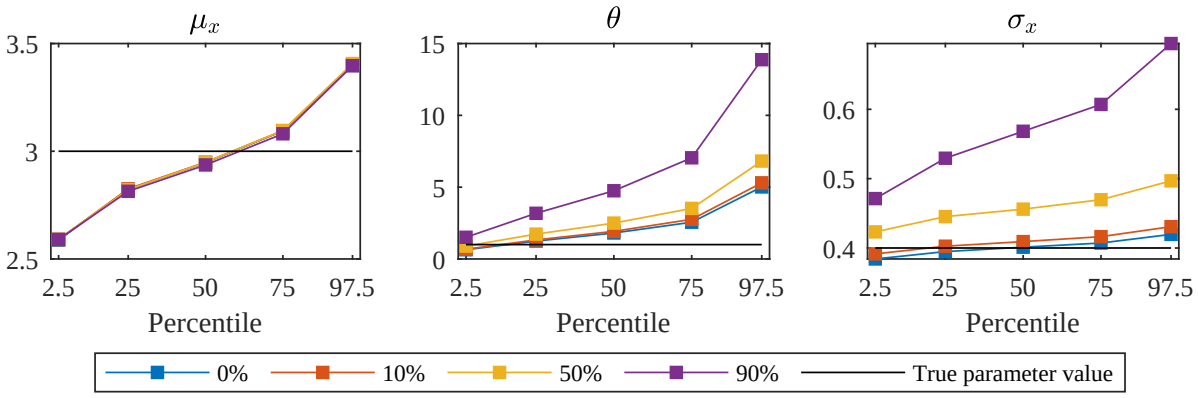


Figure 4.8: Percentiles of the 1,500 parameter posterior medians for the complete and incomplete OU sample paths.

To summarise, the implementation of MLE and Gibbs sampling to calibrate OUVLTM and OU to shortened samples, which are obtained by omission of sample points, is subject to:

- a decrease in accuracy of the model parameter estimates reflected in increased biases,
- an increase in uncertainty as indicated by heightened standard errors of the parameter estimates and increased parameter standard deviations, when Gibbs sampling was used, and
- higher RMSE and MAPE values and so a reduction in goodness of fit (but less so for OUVLTM than OU).

4.3 Accounting for Missing Data

Statistical methods handling missing data prevalent in most fields have received much attention in the last few decades, primarily due to the advent of computational programs and software packages allowing for their convenient implementation. Simpler techniques are being replaced by more sophisticated and robust ways of accounting for data that is missing.

Rubin (1996) stated that the the purpose of missing data methods is to appropriately handle missing data so as to provide valid statistical inference and not to just predict the missing data as accurately as possible. The objectives of methods that impute missing data are [2]:

1. to minimise bias,
2. to maximise available information,
3. and, most critically, to obtain appropriate estimates of uncertainty.

Recent advances in imputation methods have lead to improvements in the areas where the more rudimentary techniques fair poorly.

4.3.1 Notation

Sections 4.3.1 and 4.3.2 closely follow Gelman *et al.* (2004).

[2] According to a UCLA seminar available at https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/

The complete data set X is partitioned into its observed and missing components i.e. $X = (X_{obs}, X_{mis})$. The inclusion indicator, $K_i, i = 1, .., n$, is a series of binary variables given by the following:

$$K_i = \begin{cases} = 1, & X_i \text{ is observed} \\ = 0, & X_i \text{ is missing} \end{cases} \tag{4.1}$$

and governed by the parameter λ.

Given the model and inclusion indicator distribution parameters π and λ, the joint distribution of X and K is

$$p(X, K|\pi, \lambda) = p(X|\pi)p(K|X, \lambda). \tag{4.2}$$

λ is the parameter of the model used to detail the inclusion indicator and are characteristic of the data collection process that (for example, λ could represent the amount of data that is censored from an experiment).

By integrating over the distribution of the missing data, the distribution of the observed data (X, K) is determined:

$$p(X_{obs}, K|\pi, \lambda) = \int p(X_{obs}, X_{mis}|\pi)p(K|X_{obs}, X_{mis}, \lambda)dX_{mis}. \tag{4.3}$$

4.3.2 Missing Data Mechanism

$p(K|X, \lambda) = p(K|(X_{obs}, X_{mis}, \lambda)$ in (4.2) and (4.3) defines the missing data mechanism and gives the relationship between the probability that data is missing to the data itself. Rubin (1976) stated the three general types of missing data models to be:

1. missing completely at random (MCAR): the missing data mechanism is completely independent of X (both X_{obs} and X_{mis}) hence it is only dependent on λ.

$$p(K|X, \phi) = p(K|(X_{obs}, X_{mis}, \lambda) = p(K|\lambda).$$

2. missing at random (MAR): the distribution of the missing data mechanism does not depend on the missing data but only on the observed data i.e. $p(K|(X_{obs}, X_{mis}) = p(K|(X_{obs})$ and so (4.3) becomes

$$p(X_{obs}, K|\pi, \lambda) = p(K|X_{obs}, X_{mis}, \lambda) \int p(X_{obs}, X_{mis}|\pi)dX_{mis} = p(K|X_{obs}, X_{mis}, \lambda)p(X_{obs}|\pi)$$

   The missing data mechanism is ignorable if it is MCAR or MAR and π and λ are distinct and independent in the prior distribution. Bayesian analyses on the model parameter π need only look at the observed data likelihood $p(X_{obs}|\pi)$ in this case:

$$\mathcal{L}(X_{obs}|\pi) = \int p(X_{obs}, X_{mis}|\pi)dX_{mis} = \mathcal{L}(X_{obs}|\pi)$$

   as opposed to

$$\mathcal{L}(X_{obs}|\pi, \lambda) = \int p(X_{obs}, X_{mis}, K|\pi, \lambda)dX_{mis} = \mathcal{L}(X_{obs}, M|\pi, \lambda).$$

3. missing not at random (MNAR): the most general and complex missing data mechanism where the probability of a missing value is dependent on observed and missing data ((4.3) holds).

4.3.3 Methods to Handle Missing Data

This section gives an overview of the most commonly used methods to address missing data.

4.3.3.1 Deletion Techniques

Complete Case Analysis is also known as listwise deletion and is the default method for dealing with missing data in most statistical packages. It involves the deletion of all cases that have missing values. If the missing data mechanism is not MCAR, the analysis will be biased. If MCAR does hold, the analysis is still less efficient than an analysis that uses the full data (Chen (2013)). In terms of credit spread time series, this method equates to simply using the observed traded levels and not the LVCF time series.

Available Case Analysis selectively deletes cases based on the type of analyses and results. A case may be deleted when it is missing a value for one particular analysis only, but may be included when all values are present. It is also known as pairwise deletion.

Although these methods are easy to understand and implement, their power is reduced by the decrease in the effective sample size (since the number of observations used in the analysis are reduced). In addition, large biases are most often incurred. However, if the proportion of data missing is small, then deletion methods result in estimates that are plausible if the missing data are MCAR (Little *et al.* (1987) and Schafer (1999)).

4.3.3.2 Single Imputation Methods

To improve the statistical power and overcome the disadvantages of deletion methods, single imputation techniques 'fill in' the missing data points in some deterministic manner.

Hot-deck imputation is the imputation of missing data by substituting in values imputed directly from the data set for the missing data points. **Last Value Carried Forward** (LVCF) is such a method. Another common hot-deck imputation approach is to use the mean value of the observed data points to fill in the missing data. Like LVCF, the mean substitution approach suffers from overly reduced variation, as each missing value is replaced with a constant estimate. Interpolation is also frequently used, where the missing values are interpolated from the observed values. **Cold-deck imputation** imputes the missing data from a similar data set available, and not from the data set being analysed itself.

Regression imputation is another single imputation method, where the missing values are imputed by a regression model obtained by regressing the observed values on other variables. There is no residual variance as there is no error term included in the imputation of the missing data. Stochastic regression adds the average regression variance to the imputations so as to bring in an error.

Expectation-Maximisation is a technique that is applied to missing data in state space models, where there are model parameters to estimate, as well as unobserved latent variables. The algorithm is comprised of an iterative procedure, whereby proposed model parameter and missing data point values are checked to see if they are the most probable, until convergence is reached.

Single imputation techniques do not take the uncertainty arising from the imputation of the missing data into account. As such, the accuracy of calculated estimates is likely to be erroneously inflated as the between-imputation variability was not included. Single imputation may be valid if the proportion of missing data is small. But as this proportion increases, so should the variances around the parameter estimates. When estimating multiple parameters jointly, even low proportions of missing data can have an adverse effect (Schafer (1999)).

4.3.3.3 Multiple Imputation Techniques

For data sets with more missing points, methods which incorporate the uncertainty due to imputation are superior in statistical quality. The methods are grouped together by their characteristic of generating more than one imputation for the missing data.

A popular technique that is used to address the issue of incomplete data sets is that of Multiple Imputation (MI), proposed by Rubin (1987) and summarised by Schafer (1999). This statistical method involves the filling in of the missing data several times using imputation model specifications, such as sampling multiple times from posterior distributions generated from the observed points. Certain analyses is then invoked, assuming some model for the data, and the results are combined. More formally the following three processes are completed:

1. Imputation phase: the missing data is represented by m simulated versions, resulting in m data sets.
2. Analysis phase: each of the m imputed data sets is analysed, leaving m analyses results. Each of these analyses is unbiased if the data are missing-at-random (MAR).
3. Pooling phase: the results of the m analyses are combined to give one final output, taking the variance due to the multiple imputations into account.

Included in the final pooled variance of the parameter estimates is the uncertainty due to filling in of the missing data and the typical estimation uncertainty. So the pooled variance is larger than the variance of a single imputation, thus incorporating the uncertainty introduced by the imputations themselves.

Forward Filtering Backward Sampling or FFBS (see Section A.2.13) is commonly used to sample a time series when data points are missing. The forward recursion is executed using the Kalman filter (see Section A.2.12) and when a data point i is missing, Durbin *et al.* (2012) show that $R_i \to \infty$ in (A.82) and hence $K_i \to 0$. This result equates to predicting the filtered value at i, as the expected value of the model, which for OU is

$$\phi X_{i-1} + \alpha \quad \text{from (A.34)}$$

and for OUVLTM

$$\phi \tilde{A}_{i-1} + \eta\zeta \quad \text{from (3.19).}$$

Setting the missing point to the deterministic expected value is a single imputation of that missing value, which is exactly what MI does, but MI does it m times and accounts for the variance in the pooling phase. Similarly, the smoothed state in the backward Kalman recursion is equal to the predicted state from the forward Kalman filter when a point is missing. The cumulative errors are scaled by the total variances after the predicted point is added i.e. the current error over variance due to point i does not contribute to the smoothed state.

Multiple Imputation by Chained Equations (MICE), also known as Fully Conditional Specification (FCS) or Sequential Regression Multivariate Imputation (SRMI), MICE is a different version of MI which allows for the specification of the of the imputation model for each variable with missing data. A straight forward explanation of the method is provided in White *et al.* (2011):

"Initially, all missing values are filled in by simple random sampling with replacement from the observed values. The first variable with missing values, x_1 say, is regressed on all other variables $x_2, ..., x_k$, restricted to individuals with the observed x_1. Missing values in x_1 are replaced by simulated draws from the corresponding posterior predictive distribution of x_1. Then, the next variable with missing values, x_2 say, is regressed on all other variables $x_1, x_3, ..., x_k$, restricted to individuals with the observed x_2, and using the imputed values of x_1. Again, missing values in x_2 are replaced by draws from the posterior predictive distribution of x_2. The process is repeated for all other variables with missing values in turn: this is called a cycle. In order to stabilise the results, the procedure is usually repeated for several cycles (e.g. 10 or 20) to produce a single imputed data set, and the whole procedure is repeated m times to give m imputed data sets."

This method is useful in large data sets with many variables of varying types and thus is more flexible than MI, as each variable is modelled according to its own distribution (Azur (2011)).

In the **Maximum Likelihood Imputation**, the likelihood function of all the data is set to the product of the likelihood function of the observed data and that of the missing data. Conditioning on the unobserved data points, the marginal density of the observed points can be found. This function is then optimised using the observed data to infer model parameter estimates.

Bayesian inference lends itself quite naturally to address missing data with no further development of new methods needed. Missing data are treated unknowns, as are the model parameters. The unknown quantities have an associated probability distribution and so a joint distribution of the model parameters and the missing data can be specified. This approach is loosely named the **Full Bayesian Approach** (FBA) or the Bayesian Joint Model and is described in Hoff (2009) and Tsay (2002).

FBA differs from MI as it can be thought of as a singular phase algorithm, whilst MI has two phases. FBA simultaneously imputes all the model parameters and missing data from the joint distribution, whereas MI first imputes the missing data and then analyses the imputed data sets to determine estimates for the model parameters. By construction, the joint probability of the model parameters and missing data captures the cross-sectional uncertainty of all the unknown variables.

The Clifford-Hammersley theorem allows for the joint distribution of X_{obs}, π and X_{mis} to be specified by their conditional distributions. The complete conditionals for $p(\pi, X_{mis}|X_{obs})$ are $p(\pi|X_{obs}, X_{mis})$ and $p(X_{mis}|X_{obs}, \pi)$.

Practically, the Gibbs sampler is a intuitive way to numerically sample from the model parameters and missing data. The iterative part of the Gibbs sampler algorithm now includes an additional sampling step:

$$
\begin{aligned}
&\text{Sample } \pi_1^{i+1} \sim p(\pi_1|\pi_2^i, ..., \pi_q^i, X_{mis}^i, X_{obs}) \\
&\text{Sample } \pi_2^{i+1} \sim p(\pi_2|\pi_1^{i+1}, \pi_3^i ..., \pi_q^i, X_{mis}^i, X_{obs}) \\
&\vdots \\
&\text{Sample } \pi_q^{i+1} \sim p(\pi_q|\pi_1^{i+1}, ..., \pi_{q-1}^{i+1}, X_{mis}^i, X_{obs}) \\
&\text{Sample } X_{mis}^{i+1} \sim p(X_{mis}|\pi_1^{i+1}, ..., \pi_q^{i+1}, X_{obs})
\end{aligned}
$$

The choice of the method that will be used to address incomplete credit spread data in this research is dependent on the form of the credit spreads themselves. Credit spreads of bonds are univariate time series of varying lengths, contingent on the life spans of the bonds and the number of missing observations. Hence methods that allow for multi-dimensional data are not needed in this setting. Available Case Analysis, regression, stochastic regression and the Maximum Likelihood Model all fall into this category. For Expectation-Maximisation, the OUVLTM and OU models would need to be expressed in state space form, which is an unnecessary complication that only produces a basic single imputation.

Multiple imputation methods are preferred over single imputation or deletion techniques. Methods that produce posterior distributions for the parameters, as opposed to single point estimates, are also favoured. FBA fulfills both these criteria by producing multiple instances of the missing observations whilst calculating the posterior distributions for the unknown model parameters. The conditional distributions of the missing data points, given the observed data and known model parameter values, are sampled at each iteration in the Gibbs sampler, generating multiple complete imputed time series, or a posterior time series distribution so to speak.

4.3.4 Deriving the Conditional Posterior Distribution of Missing Data

To employ the FBA strategy in the calibration of data with missing elements, the conditional posterior distributions of all the unknown variables are required. In this case, the unknowns are the model parameters and the missing data. The conditional posterior distributions for the parameters of the OUVLTM, OU and GBM have already been discussed and are derived in Section 3.3.2 and in Appendices A.2.6.4 and A.2.5.4 respectively. This section derives the conditional posterior distribution of missing data in the autoregressive OUVLTM and OU models.

Consider the AR(1) process defined by

$$Z_i = \phi Z_{i-1} + \xi_i \quad \xi_i, \sim N(0, V),$$

where $V = \mathbb{E}[\xi_i \xi_i^T]$.

Tsay (2002) shows using iterative regressions that the least squares estimate of a missing data point, Z_m, from this AR(1) process is equal to

$$\bar{Z}_m := \mathbb{E}[Z_m] = \frac{\phi(Z_{m-1} + Z_{m+1})}{1 + \phi^2} \tag{4.4}$$

and has variance

$$\tau_{Zm}^2 := \text{Var}[Z_m] = \frac{V}{1 + \phi^2}.$$

The dependence of $\bar{Z}_m$ on both the previous and subsequent observed points is valid due to the reversibility of AR(1) processes.

To determine the conditional posterior distribution of Z_m, assume first a prior distribution for Z_m given by

$$p(Z_m) \sim N(\mu_{Z0}, \sigma^2_{Z0}) \propto \exp\left[-\frac{1}{\sigma^2_{Z0}}(Z_m - \mu_{Z0})^2\right].$$

The likelihood function of Z_m is

$$\mathcal{L}(Z_m|Z, \phi, V) = \frac{1}{\sqrt{2\pi\tau^2_{Zm}}} \exp\left[-\frac{1}{2\tau^2_{Zm}}(Z_m - \bar{Z}_m)^2\right]$$

and so the conditional posterior distribution of Z_m is:

$$p(Z_m|Z, \phi, V) \propto \exp\left[-\frac{1}{2\sigma^2_{Z0}}(Z_m - \mu_{Z0})^2\right] \exp\left[-\frac{1}{2\tau^2_{Zm}}(Z_m - \bar{Z}_m)^2\right].$$

The terms in the exponents, excluding the $-\frac{1}{2}$ multiple, are added which gives

$$\begin{aligned}
&\frac{1}{\sigma^2_{Z0}}(Z_m - \mu_{Z0})^2 + \frac{1}{\tau^2_{Zm}}(Z_m - \bar{Z}_m)^2 \\
=&Z_m^2\left(\frac{1}{\sigma^2_{Z0}} + \frac{1}{\tau^2_{Zm}}\right) - 2Z_m\left(\frac{\mu_{Z0}}{\sigma^2_{Z0}} + \frac{\bar{Z}_m}{\tau^2_{Zm}}\right) + \frac{\mu^2_{Z0}}{\sigma^2_{Z0}} + \frac{\bar{Z}^2_m}{\tau^2_{Zm}} \\
=&a_{Zm}Z_m^2 + b_{Zm}Z_m + c_{Zm},
\end{aligned}$$

with

$$a_{Zm} = \frac{1}{\sigma^2_{Z0}} + \frac{1}{\tau^2_{Zm}}, \qquad b_{Zm} = \frac{\mu_{Z0}}{\sigma^2_{Z0}} + \frac{\bar{Z}_m}{\tau^2_{Zm}} \qquad \text{and} \qquad c_{Zm} = \frac{\mu^2_{Z0}}{\sigma^2_{Z0}} + \frac{\bar{Z}^2_m}{\tau^2_{Zm}}.$$

Utilising the same analysis seen in Theorems 3.1, A.4 and A.3, the conditional posterior distribution of Z_m is deduced to be normal, with mean and variance given by

$$\mu_{Zm} = \frac{b_{Zm}}{a_{Zm}} = \frac{\bar{Z}_m\sigma^2_{Z0} + \mu_{Z0}\tau^2_{Zm}}{\sigma^2_{Z0}\tau^2_{Zm}} = \frac{\mu_{Z0}V + \sigma^2_{Z0}(1+\phi^2)\bar{Z}_m}{V + \sigma^2_{Z0}(1+\phi^2)} \tag{4.5}$$

and

$$\sigma^2_{Zm} = \frac{1}{a_{Zm}} = \frac{\sigma^2_{Z0}\tau^2_{Zm}}{\sigma^2_{Z0} + \tau^2_{Zm}} = \frac{\sigma^2_{Z0}V}{V + \sigma^2_{Z0}(1+\phi^2)} \tag{4.6}$$

respectively.

4.3.5 Application of FBA to Incomplete Simulated Data

Both OUVLTM and OU are AR(1) processes, as shown at (3.18) and (A.34), shifted by a constant. The conditional posterior distribution of missing points in an AR(1) process was shown to be normally distributed with mean and variance given by (4.5) and (4.6). Hence the missing data that was removed from the simulated OUVLTM and OU sample paths can be imputed by the FBA engine in the Gibbs, sampling estimates for the model parameters concurrently.

There are now 7 sets of results, those of

- the original complete sample path producing Tables 3.6 and 3.7, and Figure 3.8 in Section 3.4.2 (referred to as 'complete'),
- the 3 incomplete data sets, reduced by $\lambda = 10\%, 50\%$ and 90% giving Tables 4.4 to 4.7 and Figures 4.8 from Section 4.2.2 ('incomplete') and
- the 3 previously reduced but now complete data sets, generated through FBA imputation, yielding Tables 4.8 to 4.11 and Figures 4.12 and 4.8 shown later in this section ('FBA').

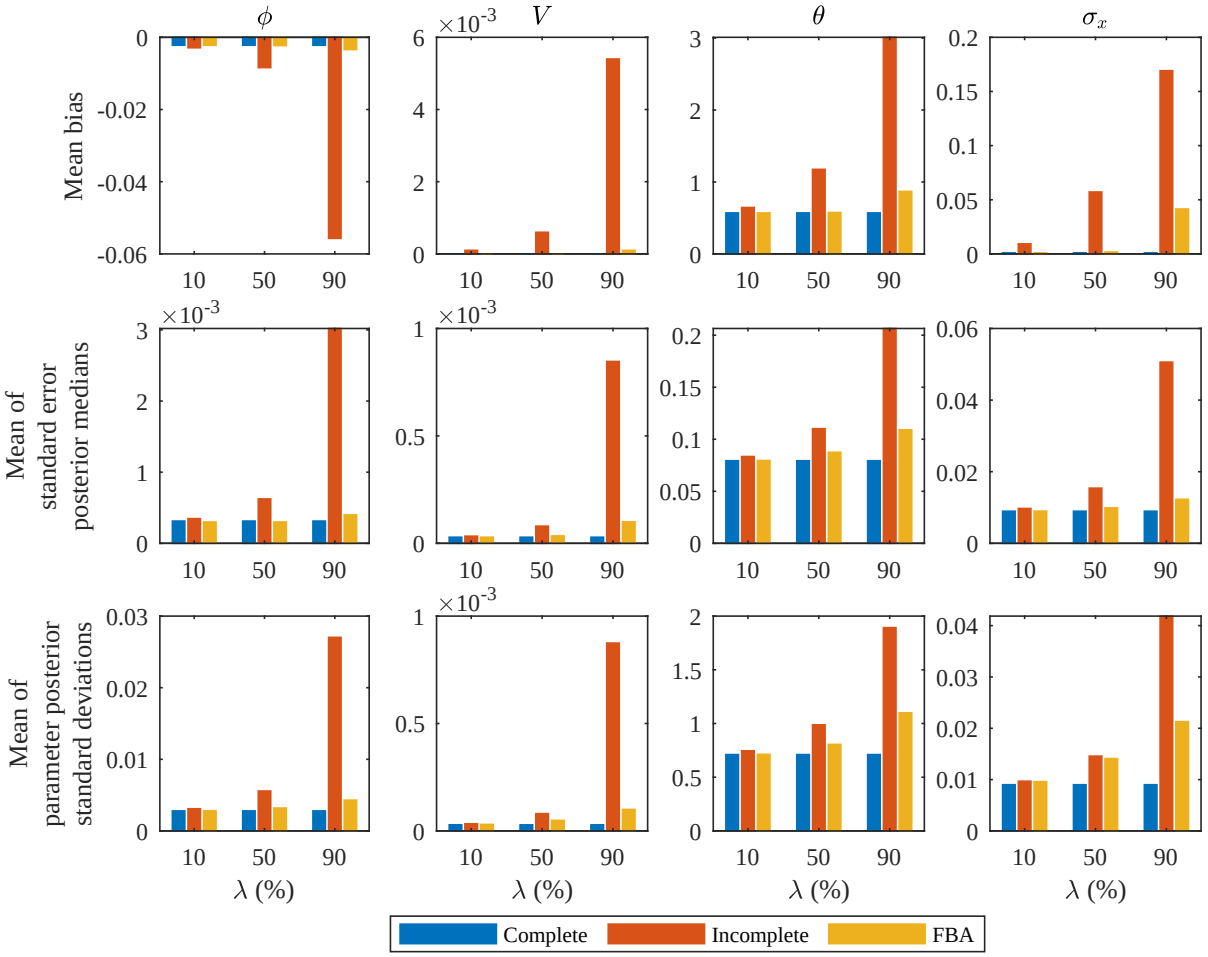


Figure 4.9: Graphical comparisons of biases, standard error and standard deviation point estimates for the complete, incomplete and FBA cases for the OUVLTM parameters.

For effective comparison between these results, consider Figures 4.9, 4.10 and 4.11. These figures bring together the pertinent results from the 7 data sets to easily observe any contrasts. The first 2 figures show the mean bias, means of the standard error posterior medians and mean standard deviations of the parameter posteriors, for each of the parameters in OUVLTM and OU respectively, for the different rates of omission. Figure 4.11 shows the mean and median RMSE and MAPE values achieved for the 7 analyses, again for the various λs. Naturally, the statistic for the complete case is the same for each omission rate.

The bias information contained in Tables 3.6, 4.5 and 4.9 for the parameters of OUVLTM for the complete case, the 3 incomplete cases and the 3 FBA cases, are plotted relative to each other in the top row of Figure 4.9. As expected from Section 4.2.2, all the incomplete parameter biases worsened significantly, implying reduced parameter estimate accuracy as the omission rates increased. The FBA case showed drastic

improvement in the biases, being much closer to the complete case bias. Thus FBA managed to successfully ameliorate the estimation of the parameters.

The middle row of Figure 4.9 draws the means of the standard error posteriors from Tables 3.6, 4.6 and 4.10 together. Each incomplete standard error was higher than those of the complete and FBA cases, with magnitude proportional to the omission rate. The accuracy of the FBA parameter estimates is deduced to be far greater, as the FBA standard errors were much reduced from those of the incomplete cases. However, they were still greater than the standard errors of the complete case, the deviation a product of the imputation of the missing points.

The means of the standard deviations of the parameter posteriors from Tables 3.6, 4.6 and 4.10 are shown in the final row of Figure 4.9. The posterior standard deviations mirrored the relative behaviour of the standard errors for each OUVLTM parameter quite well, although the absolute levels differed in magnitude. FBA again proved to substantially reduce the uncertainty of the incomplete cases, with mean standard deviations lower than those of the incomplete cases for all the OUVLTM parameters.

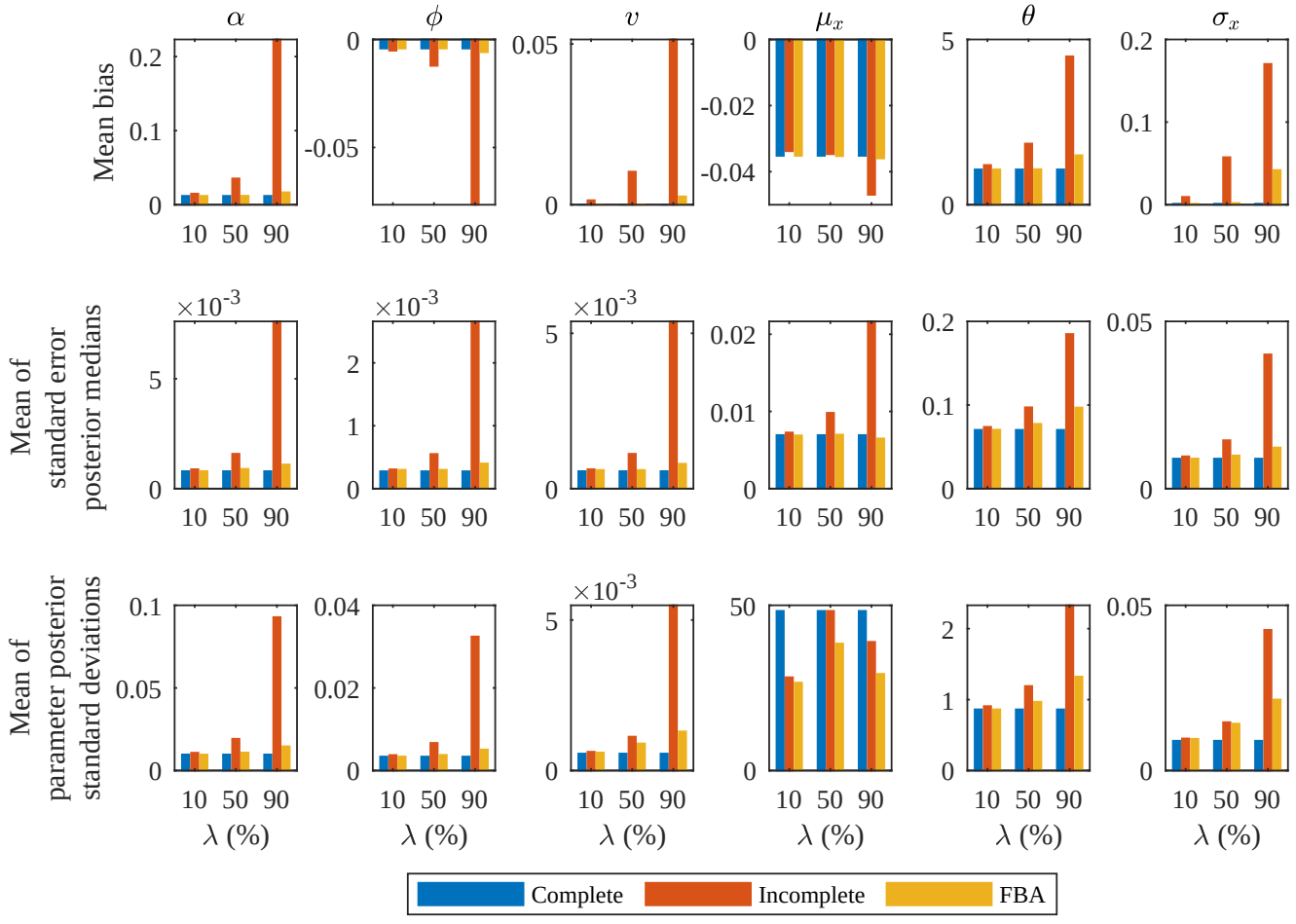


Figure 4.10: Graphical comparisons of biases, standard error and standard deviation point estimates for the complete, incomplete and FBA cases for the OU parameters.

Figure 4.10 is the OU version of Figure 4.9, where there are 6 model parameters. The observations made for OUVLTM apply to OU as well. So the mean bias, standard error and standard deviation point estimates of incomplete cases degenerate with increasing omission rate and FBA provided much less bias and uncertainty, with lowered standard errors and standard deviations. μ_x again did not conform to the behaviour exhibited by the other parameters, as was established earlier in this chapter.

The mean and median RMSE and MAPE values, obtained from the parameter estimates produced from calibrating OUVLTM and OU to the incomplete data sets and then using FBA in the calibration to impute the missing points for the different rates of omission, are graphed in Figure 4.11. So the light blue and orange shows OUVLTM and OU for the incomplete cases, the FBA cases for OUVLTM are in yellow and purple and the unchanging complete case for OUVLTM and OU are in dark blue and black.

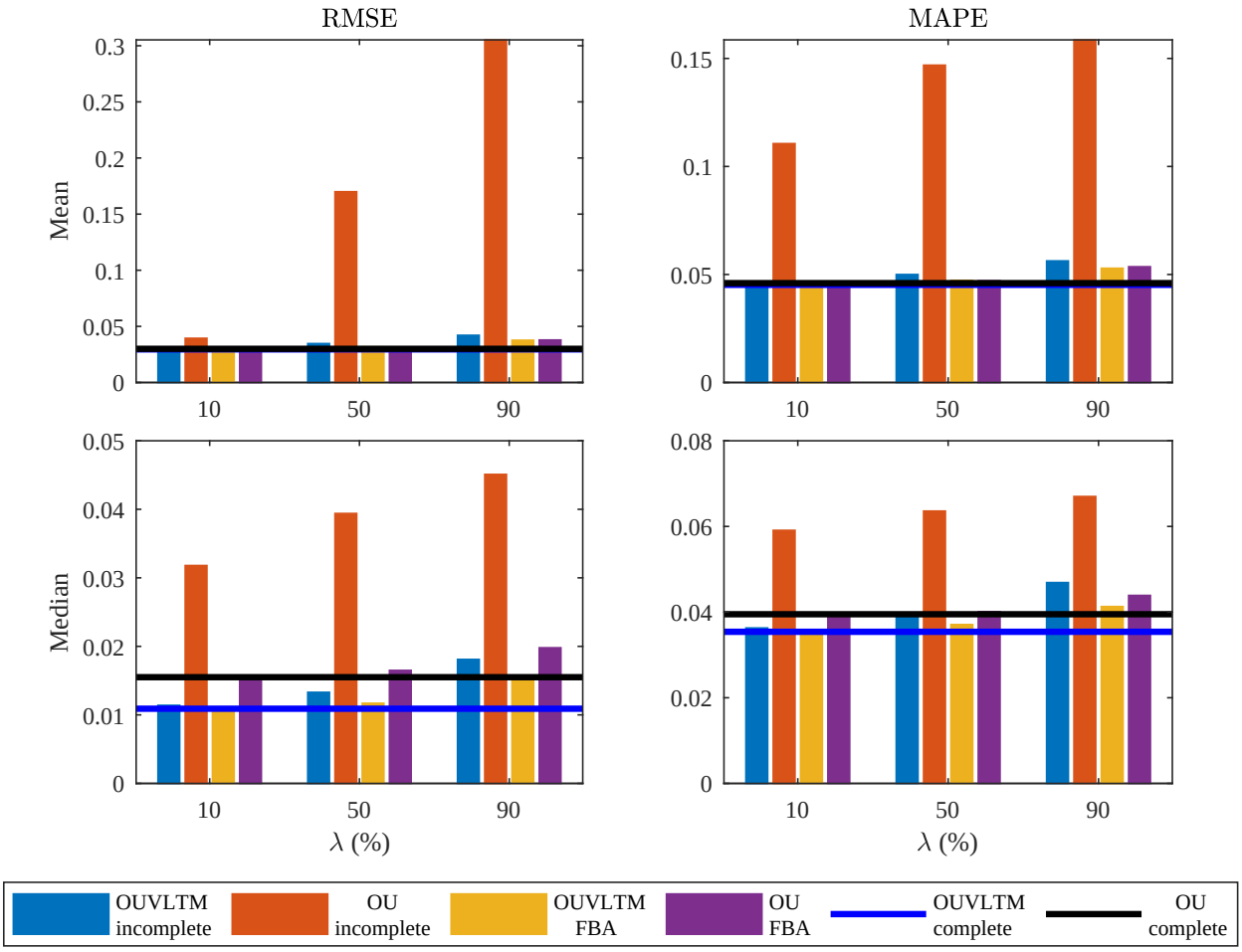


Figure 4.11: Mean and median RMSE and MAPE values for the incomplete cases (OUVLTM in light blue and OU in orange) and the FBA cases (OUVLTM in yellow and OU in purple). The complete case values for OUVLTM and OU are the horizontal dark blue and black lines respectively.

Observe that the mean RMSE and MAPE levels of the complete case for OUVLTM and OU were close (0.0295 vs 0.03 and 0.0452 vs 0.0460), hence the black lines lie on top of the darker blue line in the top graphs of Figure 4.11. The reduced path errors produced by OUVLTM were more pronounced when the medians were considered, as observed in the bottom plots.

For both mean and median RMSE and MAPEs, the OU incomplete cases for all omission rates produced the highest path error levels. On the other hand, the OUVLTM incomplete cases RMSE and MAPEs proved to not exceed those of the complete case excessively, particularly observable for the path error means in Figure 4.11.

Both models experienced an improvement in the mean and median RMSE and MAPEs at each omission rate via FBA. Visually, all the yellow bars are less than the light blue bars and all the purple bars are lower than the orange bars, indicative of the reduced path errors that FBA attained. Furthermore, a compelling

observation is that OUVLTM FBA gave mean and median RMSE and MAPE levels that were smaller than those shown by OU FBA for all omission rates, where the differences ranged from being marginal to clearly observable.

The same presentation of results that was shown in Sections 3.4 and 4.2 is now given for FBA for the sake of completeness.

λ		0%		10%		50%		90%	
Model	Parameter	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians	Mean of estimate medians	Median of estimate medians
OUVLTM	ϕ	0.9938	0.9947	0.9938	0.9946	0.9926	0.9937	0.9873	0.9894
	V	0.0006	0.0006	0.0006	0.0006	0.0008	0.0008	0.0012	0.0012
	θ	1.5701	1.3481	1.5757	1.3532	1.8688	1.5923	3.2365	2.6952
	σ_x	0.4010	0.4009	0.4018	0.4016	0.4415	0.4408	0.5489	0.5434
OU	α	0.0239	0.0211	0.0240	0.0210	0.0287	0.0253	0.0509	0.0427
	ϕ	0.9918	0.9928	0.9918	0.9929	0.9902	0.9914	0.9823	0.9854
	v	0.0252	0.0252	0.0252	0.0252	0.0277	0.0277	0.0344	0.0342
	μ_x	2.9648	2.9471	2.9647	2.9441	2.9640	2.9438	2.9589	2.9486
	θ	2.0689	1.8084	2.0768	1.8041	2.4976	2.1746	4.5075	3.7159
	σ_x	0.4012	0.4012	0.4021	0.4016	0.4420	0.4411	0.5503	0.5488

Table 4.8: Calculated means and medians of the medians of the FBA generated parameter posterior distributions for the OUVLTM and OU sample paths for the differing omission rates.

λ		0%	10%	50%	90%
OUVLTM	ϕ	-0.0022	-0.0023	-0.0034	-0.0088
	V	0.0000	0.0000	0.0001	0.0006
	θ	0.5701	0.5757	0.8688	2.2365
	σ_x	0.0010	0.0018	0.0415	0.1489
OU	α	0.0120	0.0121	0.0168	0.0390
	ϕ	-0.0042	-0.0042	-0.0059	-0.0137
	v	0.0000	0.0001	0.0026	0.0092
	μ_x	-0.0352	-0.0353	-0.0360	-0.0411
	θ	1.0689	1.0768	1.4976	3.5075
	σ_x	0.0012	0.0021	0.0420	0.1503

Table 4.9: Calculated biases of the means of the FBA generated parameter posterior distribution medians for the OUVLTM and OU sample paths.

λ		0%		10%		50%		90%	
Model	Parameter	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations	Mean of SE medians	Mean of standard deviations
OUVLTM	ϕ	0.0003	0.0028	0.0003	0.0028	0.0003	0.0032	0.0004	0.0043
	V	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
	θ	0.0794	0.7109	0.0796	0.7132	0.0875	0.8057	0.1091	1.0997
	σ_x	0.0090	0.0090	0.0090	0.0096	0.0099	0.0141	0.0123	0.0213
OU	α	0.0008	0.0097	0.0008	0.0097	0.0009	0.0109	0.0011	0.0147
	ϕ	0.0003	0.0034	0.0003	0.0034	0.0003	0.0038	0.0004	0.0051
	v	0.0006	0.0006	0.0006	0.0006	0.0006	0.0009	0.0008	0.0013
	μ_x	0.0069	48.3045	0.0069	26.6053	0.0070	38.4273	0.0065	29.2521
	θ	0.0703	0.8626	0.0705	0.8632	0.0775	0.9691	0.0970	1.3240
	σ_x	0.0090	0.0090	0.0090	0.0096	0.0099	0.0142	0.0123	0.0215

Table 4.10: Calculated means of the medians of the standard error FBA posteriors and the calculated means of the standard deviations of the FBA generated parameter posterior distributions for the OUVLTM and OU sample paths.

Model		OUVLTM				OU			
λ		0%	10%	50%	90%	0%	10%	50%	90%
Mean	RMSE	0.0295	0.0295	0.0317	0.0377	0.0300	0.0300	0.0300	0.0369
	MAPE	0.0452	0.0452	0.0473	0.0529	0.0460	0.0460	0.0471	0.0526
Median	RMSE	0.0109	0.0109	0.0117	0.0154	0.0155	0.0155	0.0165	0.0198
	MAPE	0.0354	0.0355	0.0371	0.0433	0.0395	0.0395	0.0401	0.0435

Table 4.11: Calculated mean and median RMSE and MAPE levels resulting from the medians of the FBA generated parameter posteriors of the OUVLTM and OU sample path sets.

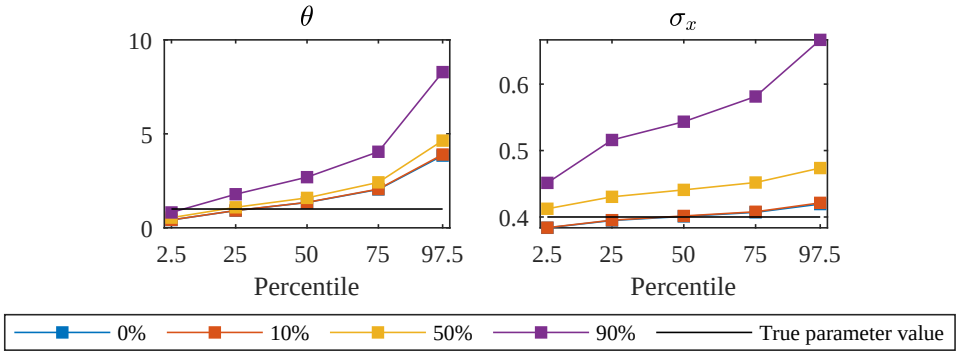


Figure 4.12: Percentiles of the 1,500 FBA parameter posterior medians for the OUVLTM data.

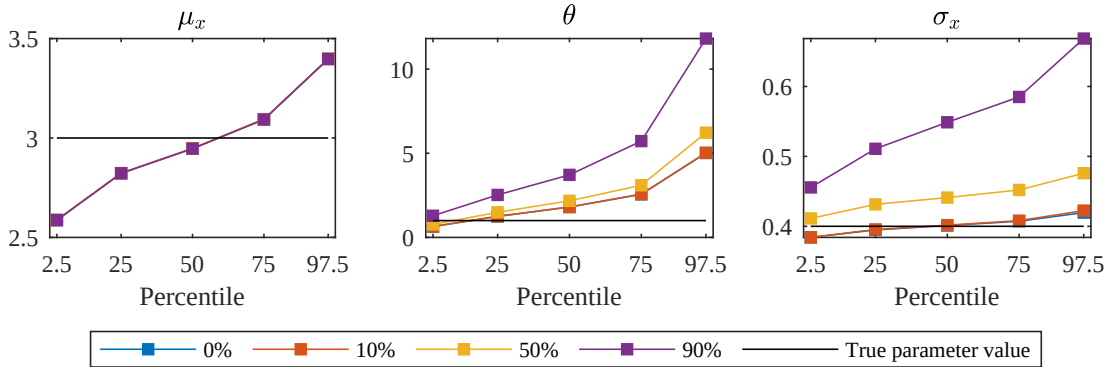


Figure 4.13: Percentiles of the 1,500 FBA parameter posterior medians for the OU data.

In final summary of parameter estimation using Gibbs sampling whilst imputing missing data through FBA:

- FBA improved the biases of the θ and σ_x parameters for both OUVLTM and OU, relative to those of the incomplete cases, indicating an increase in estimation quality.
- The parameter posterior standard deviations and medians of standard error posteriors were reduced for both models by FBA compared to the incomplete cases. So FBA yielded a reduction in parameter uncertainty.
- The mean and median RMSE and MAPE values were also diminished by FBA from the incomplete cases, again a testament to the quality of estimation of FBA.
- OUVLTM exhibited smaller RMSE and MAPEs than those of OU, evidence of OUVLTM's superior fit and its successful enhancement of the OU model.

5. Fitting to Market Credit Spreads

In this chapter, the modelling of credit spreads using the new OUVLTM methodology is explored. Both the new model and OU are calibrated to the credit spreads of numerous market bonds, some with daily liquidity. The performance of the models is assessed based on metrics such as maximum likelihood, AIC and BIC (when MLE is used), as well as standard errors of the estimates and of the estimate posterior distributions (in Gibbs sampling scenarios). Furthermore, both in-sample and out-of-sample goodness of fit is measured via RMSE and MAPE. More specifically, the time series of the credit spreads are divided to form in-sample and out-of-sample data sets. The in-sample data sets are used to obtain the parameter estimates of the models. Forecasts using these estimates are then calculated for the out-of-sample-period. Two daily traded bonds from the US market, representing the liquid bonds, are analysed, as well as 9 South African bonds with varying illiquidity.

5.1 Application to Credit Spreads of Liquid Market Bonds

The OUVLTM model should be applicable to complete data sets and provide sensible solutions without any missing data imputation. FBA is the engine which performs data imputation, so when data sets are complete, using FBA is not necessary. So just MLE on the complete data sets is carried out, where (3.16) and (A.39) apply for maximum likelihood estimation under the OUVLTM and OU models. The US bonds that are studied are the AAPL 5YR bond introduced in Chapter 2 and a General Motors (with share price ticker GM) 10YR bond maturing in May 2023.

5.1.1 In-Sample Testing

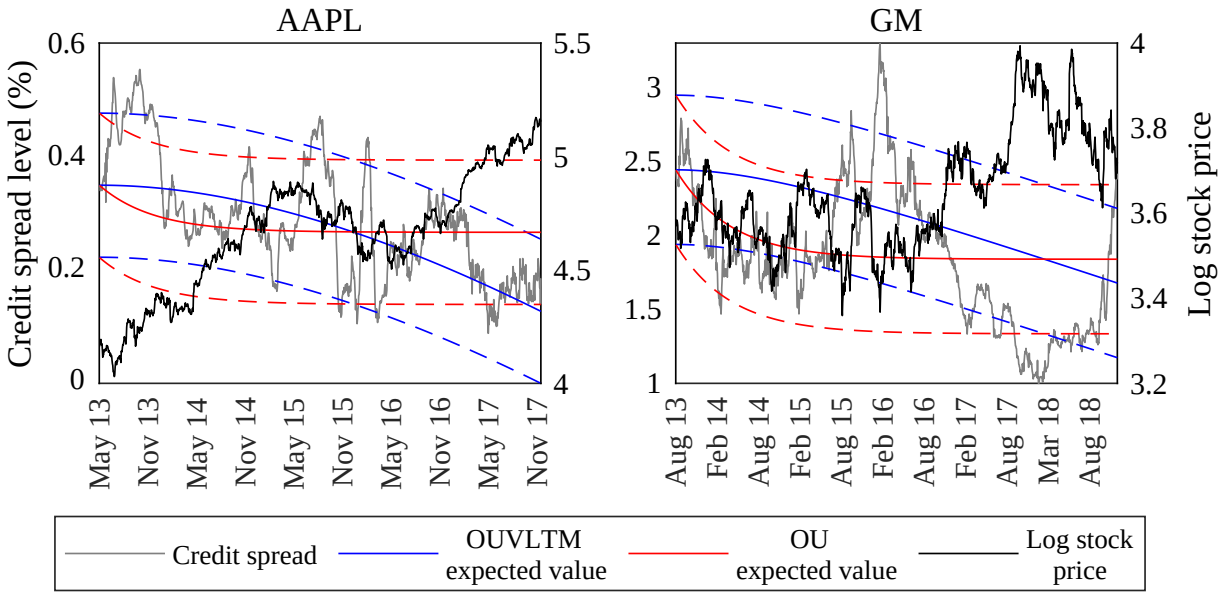


Figure 5.1: Credit spreads of the AAPL and GM bonds, the log stock prices and the expected values as postulated by OUVLTM (blue) and OU (red). The dashed lines showing the respective prediction intervals.

The time series of the AAPL and GM bonds covered periods of between 4 and 5 years. Figure 5.1 shows their daily credit spreads and log stock prices for these periods. The expected values resultant from MLE fitting OUVLTM and OU to the credit spreads and stock price data are also plotted, in blue and red respectively. The dashed lines are the 95% prediction intervals around the expected values for each model. For normally distributed models, as are OUVLTM and OU, the prediction intervals around the expected values are defined as

$$\mathbb{E}[x_{\text{model}}] \pm z\sqrt{\text{Var}[x_{\text{model}}]}$$

and for a 95% prediction interval, $z = 1.96$.

OUVLTM produced expected value paths that were distinctly downward sloping, because their associated equilibrium levels were inversely based on the trend of the log stock price. For both bonds, the stock prices were increasing over time and so the log stock prices' general trend was positive.

Table 5.1 presents the maximum likelihood function values and corresponding AIC and BIC levels attained for each model. The estimates for the original model parameters μ_x, θ and σ_x (OUVLTM has just θ and σ_x) are also given.

Bond	Model	μ_x	θ	σ_x	Maximum $\mathcal{L}$	AIC	BIC
AAPL	OUVLTM	-	3.1523	0.2054	3,354.38	-6,704.76	-6,694.67
	OU	0.2622	1.7915	0.2052	3,352.57	-6,699.15	-6,684.02
GM	OUVLTM	-	1.1138	0.8018	2,106.97	-4,209.95	-4,199.54
	OU	1.8361	1.5115	0.8017	2,108.21	-4,210.42	-4,194.81

Table 5.1: The results of the maximum likelihood estimation for OUVLTM and OU applied to the credit spreads of the AAPL and GM bonds.

For the AAPL credit spread, OUVLTM clearly achieved the superior fit, with the higher maximum likelihood and lower AIC and BIC values. The maximum likelihood value of OUVLTM for the GM credit spread was less than that of OU and its AIC level was only fractionally larger than the AIC of OU. But the OUVLTM BIC was more negative, thus taking the length of the time series into account saw OUVLTM as the superior model.

Standard errors of the parameter estimates for both models are shown in Table 5.2. The lower SE of OUVLTM's θ estimate relative to OU's estimate implies that the former was a more accurate reflection of true mean reversion value with less uncertainty. The errors for the σ_x estimates of OUVLTM and OU were the equal for both bonds.

Bond	Model	μ_x	θ	σ_x
AAPL	OUVLTM	-	0.2471	0.0043
	OU	0.0032	0.3200	0.0043
GM	OUVLTM	-	0.1502	0.0155
	OU	0.0126	0.1836	0.0155

Table 5.2: Calculated standard errors of the maximum likelihood parameter estimates applied to the credit spreads of the AAPL and GM bonds.

The goodness of fit tests, RMSE and MAPE, determined for the maximum likelihood estimates of the models are summarised in Table 5.3. AAPL showed reduced path errors for OUVLTM relative to OU, again

illustrating that OUVLTM was the more suitable model in this case. On the other hand, OU attained smaller path errors than OUVLTM for the credit spread of the GM bond, but only marginally so. This slight difference is due to the OU model's expected value following the observed credit spread in Figure 5.1 closely enough during the first 2 years of the period to outweigh the deviation seen later, as measured by the MAPE. In appearance, the OUVLTM expected value fitted the trend of the GM credit spread more intuitively but was penalised by the MAPE for it generally larger relative deviations.

Bond	AAPL		GM	
Model	OUVLTM	OU	OUVLTM	OU
RMSE	0.0070	0.0094	0.2097	0.1988
MAPE	0.2588	0.3190	0.2199	0.2190

Table 5.3: Calculated RMSE and MAPE levels produced from the OUVLTM and OU MLEs for the AAPL and GM bonds' credit spreads.

In fitting the models to these market bonds, a scaling factor, β, has been incorporated to relax the mean reversion level movements relative to the stock price from being restricted to unity. This updates (3.7) to

$$\tilde{\mu}(t) = \tilde{\mu}(0) + \beta \ln\left(\frac{\ln s(0)}{\ln s(t)}\right) \tag{5.1}$$

and (3.2) becomes

$$\tilde{\mu}(t) = \tilde{\mu}(0) - \beta\eta t - \sigma_s W_s^{\mathbb{P}}(t) \tag{5.2}$$

leading to the modified solution for OUVLTM as

$$\begin{aligned}\tilde{x}(t) = \tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1-e^{-\theta t}) - \beta\eta\left(t - \frac{1-e^{-\theta t}}{\theta}\right) \\ - \sigma_s \int_0^t \left(1 - e^{\theta(u-t)}\right) dW_s^{\mathbb{P}}(u) + \sigma_x \int_0^t e^{\theta(u-t)} dW_x^{\mathbb{P}}(u)\end{aligned} \tag{5.3}$$

with mean now given by

$$\mathbb{E}[\tilde{x}(t)] = \tilde{x}(0)e^{-\theta t} + \tilde{\mu}(0)(1-e^{-\theta t}) - \beta\eta\left(t - \frac{1-e^{-\theta t}}{\theta}\right). \tag{5.4}$$

The variance stays equal to (3.6) under this alteration.
The optimal value for β is found again by maximising the likelihood function. Figure 5.2 plots the parameter values and the likelihood function value of the OUVLTM model against β, fitted to the AAPL bond. The parameter estimates seen in Table 5.1 coincide with the maximum likelihood value, which was achieved for $\beta = 0.206$. Similarly, the optimal β for the GM bond was found to be 5.1.

5.1.2 Out-of-Sample Testing

For the out-of-sample testing, a long and short forecast period are chosen for each bond. The final year (roughly) is taken as the long forecast period for AAPL and the short period is assumed to be the last 2 months of the time series. The long forecast period for GM is the end of the credit spread time series approximately equalling 2 years, whilst the short period is chosen as roughly the final 6 months. Lengthier forecast periods for GM have been chosen as the time series is longer than that of AAPL.

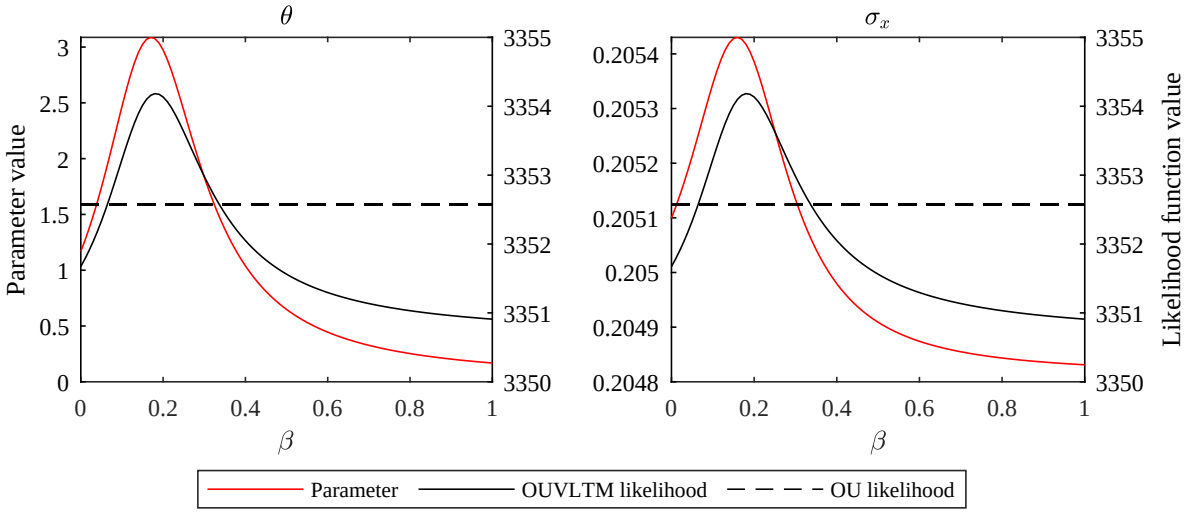


Figure 5.2: OUVLTM parameter and likelihood function values for varying β applied to the credit spread of the AAPL bond.

OUVLTM and OU are calibrated to data before the forecast periods and the parameter estimates produced by the MLE are used to determine the expected values and prediction intervals. Specifically, the GBM parameter estimates for the stock price yield a variable mean reversion level, $\tilde{\mu}$, for each time point in the forecast period, as per (3.5). During the forecasting period, $\tilde{\mu}$ can also be calculated directly from the change in the log stock price as given by (5.1), where the θ and σ_x obtained in the MLE calibration are used to determine the expected credit spread path. But the credit spread level based on this $\tilde{\mu}$ is simply the daily realisation implied by the crystalised market stock price and has no forecasting ability. It is a useful tool, however, to ascertain the potential error incurred by assuming that the stock price is a GBM.

The RMSE and MAPE goodness of fit tests for the bonds' out-of-sample forecasting periods are given in Table 5.4. The daily updated OUVLTM expected value (denoted OUVLTM*) is shown in addition to the expected values of OUVLTM and OU. Figure 5.3 graphs the expected values produced by OUVLTM, OUVLTM* and OU, in blue, purple and red respectively - AAPL are the top plots and the bottom plots are for GM. Their associated prediction intervals are given by the dashed lines of the same colour.

		Long forecast period			Short forecast period		
Bond	Model	OUVLTM	OUVLTM*	OU	OUVLTM	OUVLTM*	OU
AAPL	RMSE	0.0076	0.0037	0.0123	0.0013	0.0013	0.0019
	MAPE	0.5087	0.3382	0.6486	0.1999	0.1946	0.2528
GM	RMSE	0.2991	0.1140	0.4703	0.1202	0.0656	0.0685
	MAPE	0.3907	0.1941	0.4930	0.1090	0.1143	0.1414

Table 5.4: Calculated RMSE and MAPE levels produced from the OUVLTM and OU MLEs from the out of sample estimation periods.

The out-of-sample long and short OUVLTM expected values for the credit spread of AAPL had lower RMSE and MAPE values than OU. The OUVLTM* did noticeably better than OUVLTM in terms of RMSE and MAPE for the long forecast period, but the improvement was marginal in the short period case. The top plots in Figure 5.3 illustrate the deviations; the significant slope change for the long period whilst the short case showed very similar expected values.

OUVLTM outperformed OU for the long forecast period of GM, incurring less error with lower RMSE

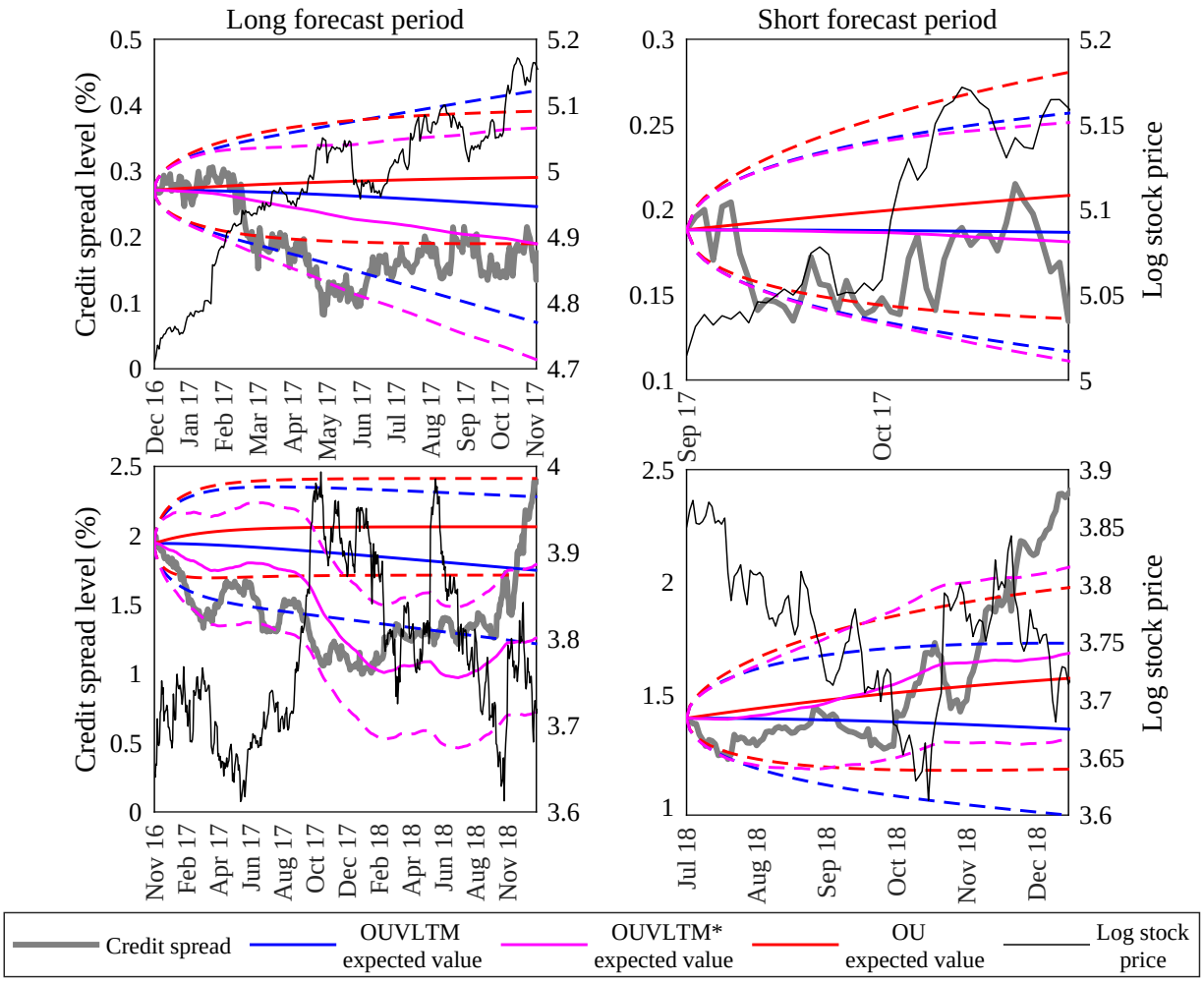


Figure 5.3: Long and short period forecasts for the AAPL (top) and GM (bottom) bonds.

and MAPE values. OUVLTM* attained even lower sample path errors, with the plot for this forecast period in Figure 5.3 (bottom left) showing how well OUVLTM* was able to track the credit spread. This result is primarily because GM is a very good example of the intuitive inverse relationship between credit spreads and the (log) stock price. For the short forecast period, the RMSE of OUVLTM was higher than those of OU and OUVLTM*, but did have the lowest MAPE. This particular forecast period saw a plunge in the stock price, which was picked up by OUVLTM*, manifesting in a bump up in the OUVLTM* spread at the same time.

The credit spread levels for both bonds were within the OUVLTM prediction intervals noticeably more than the OU prediction levels, following their more accurate expected values. The exception was GM's short forecast period, where the OU prediction cone was tilted upwards towards the increasing credit spread.

This section demonstrates that fitting OUVLTM to complete credit spreads produced meaningful results and so provides a valuable alternative perspective on the prediction of credit spreads to OU, that is also as statistically powerful.

5.2 Application to Illiquid South African Bonds

It is common knowledge that the smaller the data set, the more bias and uncertainty is incurred. Thus estimation using the time series of traded credit spread points (referred to henceforth as TP) is highly likely to be misleading. LVCF has also been established as exhibiting bias and severely understating uncertainties. It is preferable to consider as much information as is available with the goal of representing estimation uncertainties appropriately, circumventing the polarised views that TP and LVCF offer.

The 9 SA bonds selected to be analysed cover a diverse range of degrees of liquidity (indicated by TFR) with issuers from different sectors. The bonds were also alive over a variety of different periods between 2005 and 2018. Gibbs sampling is used to fit the OUVLTM and OU models to the TP and LVCF time series and are anticipated to produce undeniably inconsistent outcomes. The FBA engine within the Gibbs sampler is then applied to the TP data, yielding complete time series with the same number of points as the LVCF time series. For ease of notation, calibration to TP and LVCF time series will be termed TP and LVCF and FBA denotes the parameter estimation and missing data imputation applied to the TP time series.

Assessment of the performance of the different models applied to the various time series focusses on

1. the medians of the estimate posterior distributions,
2. the posteriors of the standard errors of the parameter estimates,
3. the standard deviations of the parameter estimate posteriors,
4. RMSEs and MAPEs using the parameter estimate medians fit to the TP time series (in-sample goodness of fit) and
5. out-of-sample goodness of fit testing using RMSE and MAPE.

The forecast periods assumed for the bonds in the out-of-sample testing are all approximately equal to one year. Calibration to the preceding data using OUVLTM and OU allows for the calculation of expected values and prediction intervals for the forecast periods, from which path errors are calculated.

The expectation is that FBA gives results that temper those produced by TP and LVCF, a correction of extremes so to speak. FBA accounts for variance due to imputation and so its uncertainty should be higher than the unrealistic, muted LVCF levels. In addition, FBA should reduce the uncertainty of TP to a certain extent, depending on the number of missing points. Using FBA to calibrate the OUVLTM model will also hopefully refine the calibration using FBA on OU, in terms of uncertainty (standard errors and posterior standard deviations) and goodness of fit. More explanatory information is being utilised by OUVLTM, where the assumed relationship between the stock price and credit spread is ideally being harnessed to the enhance predictive power. OUVLTM makes use of more explanatory information, so ideally is harnessing the assumed relationship between the credit spread and the log stock price to enhance predictive power.

The bonds whose credit spreads are analysed are listed in Table 5.5, along with their TFR and LM levels. The optimal β and the number of days in the forecast period for the bonds are also provided. The

remainder of the table contains the performance assessment information described in points 1 to 5 above, for the OUVLTM model. Table 5.6 shows the same information resultant from using OU. In addition, the in- and out-of-sample expected values and prediction intervals (dashed lines) are graphically presented for each bond in Figures 5.6 to 5.14.

Summary of observations made from Tables 5.5 and 5.6 and Figures 5.6 to 5.14:

- The FBA parameter estimate and standard error medians and the parameter posterior standard deviations lay between those of TP and LVCF for both models, evidence of the moderating ability of FBA.
- In terms of parameter estimates, OUVLTM FBA yielded slightly lower medians for σ_x than OU. There were some extreme TP and LVCF medians for θ that OUVLTM produced, which were not as acute using OU (see ABS10 for example). Thus there were a few raised OUVLTM FBA θ medians observed.
- As expected, TP attained the highest standard error medians whilst LVCF produced the lowest values for both OUVLTM and OU, confirmation of hypothesis that TP overstates uncertainty and LVCF understates it.
- OUVLTM obtained standard error medians of θ that were lower than the OU values, for some bonds significantly so (for example, FRX16). The OUVLTM standard error medians for σ_x were either equal to or less than those of OU.
- Due to the swings exhibited by the OUVLTM θ medians, the corresponding posterior standard deviations were high. In these cases, the standard deviations of the OUVLTM TP, LVCF and FBA posteriors were all higher than their OU counterparts. The σ_x posterior standard deviations for OUVLTM and OU were more or less equal across the bonds.
- The in-sample fitting showed that generally the MAPEs of OUVLTM were lower than the OU MAPEs (SBS9, CBL11, ABS10, NBK6A and MTN05). There were some instances where OUVLTM had higher RMSEs but lower MAPEs (MTN01, IPL3 and FRX16). But the most compelling result was that for all bonds, OUVLTM FBA achieved lower MAPE levels than OU FBA. So taking trends into account, OUVLTM FBA surpassed OU FBA in terms of goodness of fit.
- The goodness of fit of OUVLTM is shown to be even better in the out-of-sample fitting. The RMSE and MAPEs for all the OUVLTM calibrations were below the corresponding OU errors, barring IPL3. Some of the RMSE and MAPE levels for this bond were higher when calibrated to OUVLTM than OU. But again, the MAPE of OUVLTM FBA was less than the MAPE of OU FBA for this bond.
- There was a distinct difference in the in-sample expected values and prediction intervals between OUVLTM and OU, evident in the figures for ABS10, CBL11, NBK2A and MTN05. In most cases, the OUVLTM in- and out-of-sample expected values were characterised by a negative slope during the respective periods, as a result the general upward trends across the stock prices for those periods. There was indeed a noticeable link of the OUVLTM expected values to an equilibrium level that was inversely related to the stock price.
- IPL3 showed an in-sample OUVLTM FBA expected value that was essentially flat. The stock price during that period looked almost mean-reverting to some degree itself, with no discernible long term trend - an upward trend was followed by a downward trend after which another upward trend occurred.

- Apart from the most liquid bond MTN01, OUVLTM FBA produced more accurate expected values than OU FBA. For MTN01, OU FBA performed just as well as OUVLTM FBA.

A further demonstration of the intuitive power of the OUVLTM model is evident in the in-sample fitting of both models to the African Bank bonds used in Section 2.2.7.1, plotted in Figures 5.4 and 5.5. Whilst the OU model fitted the observed credit spreads better for both bonds, the OUVLTM model produced rising credit spread estimates consistent with the falling share price, meeting the expectation of that the credit risk of the bonds was increasingly significantly as the share price deteriorated.

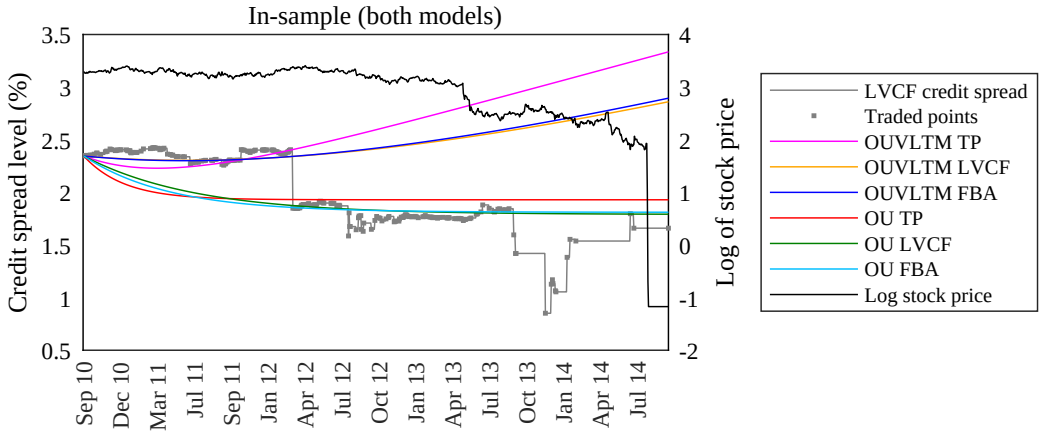


Figure 5.4: ABL11A

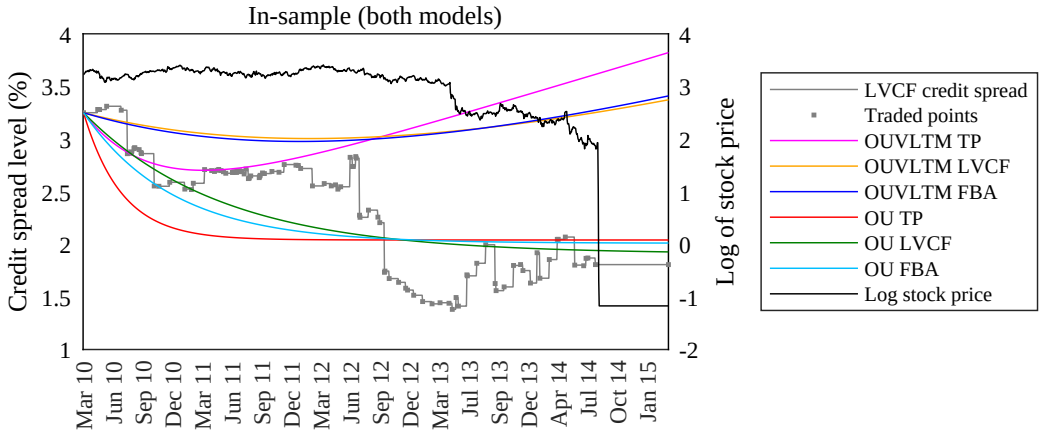


Figure 5.5: ABL10A

Bond			MTN01	SBS9	CBL11	ABS10	IPL3	NBK6A	MTN05	FRX16	BID02
TFR (%)			62.64	51.94	41.92	37.49	25.34	12.55	5.99	4.05	1.59
LM (days)			1	5	11	11	3	19	33	31	115
β			1	0.7	0.6	0.5	1	0.9	1.3	0.6	3
Forecast period (days)			255	244	245	243	252	252	248	256	246
Median estimates	θ	TP	0.8881	5.4984	13.8249	17.4075	1.1987	25.6925	11.7228	16.5480	28.6089
		LVCF	0.5713	3.0818	4.1320	3.8052	0.2685	7.2083	2.2271	1.7672	1.8746
		FBA	0.6424	3.5598	5.4298	5.3495	0.3833	17.9753	3.5365	3.8358	2.6362
	σ_x	TP	0.9718	0.8502	1.2177	0.4003	0.8719	1.0778	0.5472	1.3808	1.7315
		LVCF	0.8136	0.6788	0.8983	0.2867	0.6072	0.6521	0.3076	0.6336	0.6880
		FBA	0.8732	0.7572	0.9949	0.3191	0.7529	0.9030	0.3913	0.8882	0.8994
Median standard errors of estimates	θ	TP	0.3007	0.1601	0.3134	0.1892	0.4785	0.6600	0.7553	1.5804	2.4306
		LVCF	0.1931	0.0912	0.1472	0.0828	0.1474	0.1647	0.1503	0.1756	0.1312
		FBA	0.2077	0.1017	0.1631	0.0926	0.1835	0.2340	0.1923	0.2531	0.1726
	σ_x	TP	0.0300	0.0221	0.0426	0.0127	0.0388	0.0849	0.0860	0.2554	0.8314
		LVCF	0.0182	0.0115	0.0180	0.0049	0.0112	0.0131	0.0080	0.0135	0.0178
		FBA	0.0195	0.0128	0.0199	0.0054	0.0139	0.0181	0.0101	0.0189	0.0233
Standard deviation of posteriors	θ	TP	0.5389	1.2110	3.2911	3.4496	0.7491	7.0212	4.3815	6.8253	11.3686
		LVCF	0.3458	0.6807	1.2323	1.0588	0.1876	1.7738	0.8013	0.6922	0.5205
		FBA	0.3792	0.7740	1.4233	1.2567	0.2497	2.9139	1.1055	1.1133	0.7332
	σ_x	TP	0.0277	0.0201	0.0390	0.0116	0.0321	0.0695	0.0623	0.1609	0.4628
		LVCF	0.0182	0.0115	0.0182	0.0049	0.0112	0.0134	0.0080	0.0135	0.0179
		FBA	0.0210	0.0189	0.0275	0.0078	0.0228	0.0319	0.0137	0.0314	0.0319
In-sample	RMSE	TP	0.8486	0.0693	0.0677	0.0054	0.4510	0.0313	0.0440	0.1462	0.5728
		LVCF	0.7940	0.0418	0.0617	0.0050	0.3653	0.0309	0.0243	0.0977	0.2368
		FBA	0.8081	0.0483	0.0635	0.0051	0.3817	0.0312	0.0314	0.1216	0.3141
	MAPE	TP	0.2786	0.1700	0.0827	0.0482	0.3178	0.0941	0.2484	0.5663	0.4088
		LVCF	0.2823	0.1501	0.0775	0.0460	0.2732	0.0930	0.1918	0.5038	0.2018
		FBA	0.2811	0.1553	0.0791	0.0464	0.2823	0.0939	0.2152	0.5389	0.2601
Out-of-sample	RMSE	TP	2.5666	0.0299	0.4601	0.0296	0.0580	0.0624	0.0368	0.0848	0.2876
		LVCF	2.7671	0.0307	0.4912	0.0301	0.0367	0.0740	0.0412	0.1069	0.6493
		FBA	1.3939	0.0305	0.4885	0.0320	0.0513	0.0618	0.0429	0.0975	0.4065
	MAPE	TP	1.7325	0.2498	0.3793	0.1948	0.1319	0.1729	0.3576	1.4785	0.4466
		LVCF	1.8011	0.2524	0.3936	0.1992	0.0803	0.1936	0.3689	1.6386	0.6705
		FBA	1.2642	0.2518	0.3924	0.2112	0.1229	0.1715	0.3734	1.5778	0.5306

Table 5.5: Calculated median parameter estimates, median standard errors and standard deviations of the generated posterior distributions using OUVLTM on the traded point (TP) time series, the LVCF time series and then the application of FBA to the TP time series. In- and out-of-sample RMSE and MAPE values are also shown.

Bond			MTN01	SBS9	CBL11	ABS10	IPL3	NBK6A	MTN05	FRX16	BID02
Median estimates	μ_x	TP	1.3685	0.9044	2.3851	1.0857	1.2732	1.4518	0.7907	0.5066	1.2895
		LVCF	1.3617	0.8779	2.2739	1.0107	1.5994	1.4876	0.8158	0.5264	1.5245
		FBA	1.3691	0.8975	2.2995	1.0327	1.5146	1.5036	0.8540	0.6134	1.7457
	θ	TP	0.9283	4.8757	10.5343	2.8833	1.6618	16.8478	6.0671	14.0512	11.7388
		LVCF	0.5989	2.7526	2.7144	1.0386	0.4875	4.9998	1.4164	1.3781	0.8114
		FBA	0.6836	3.2340	3.8409	1.4594	0.7407	12.3085	5.3647	3.6116	1.5723
	σ_x	TP	0.9710	0.8516	1.2234	0.3931	0.8716	1.0605	0.5424	1.3758	1.7516
		LVCF	0.8140	0.6788	0.8975	0.2859	0.6076	0.6508	0.3079	0.6341	0.6888
		FBA	0.8740	0.7573	0.9933	0.3195	0.7544	0.8881	0.3920	0.8905	0.9032
Median standard errors of estimates	μ_x	TP	0.0298	0.0095	0.0124	0.0069	0.0273	0.0172	0.0283	0.0511	0.1615
		LVCF	0.0236	0.0069	0.0109	0.0048	0.0160	0.0058	0.0067	0.0115	0.0198
		FBA	0.0237	0.0071	0.0102	0.0044	0.0161	0.0051	0.0064	0.0100	0.0187
	θ	TP	0.3597	0.4229	0.3529	0.2041	0.5182	0.7469	0.9608	3.4122	3.2846
		LVCF	0.2327	0.2460	0.1712	0.0981	0.1654	0.1927	0.2005	0.3641	0.1913
		FBA	0.2504	0.2743	0.1895	0.1089	0.2065	0.2677	0.3789	0.5404	0.2507
	σ_x	TP	0.0274	0.0200	0.0378	0.0109	0.0319	0.0598	0.0566	0.1434	0.3575
		LVCF	0.0182	0.0115	0.0180	0.0048	0.0112	0.0131	0.0080	0.0135	0.0178
		FBA	0.0196	0.0128	0.0199	0.0054	0.0139	0.0181	0.0149	0.0189	0.0234
Standard deviation of posteriors	μ_x	TP	35.2844	0.1083	0.2682	1.8779	11.5746	1.9909	3.0384	1.6542	4.0775
		LVCF	14.8397	0.1089	17.1571	2.4959	64.9119	0.0815	0.8702	6.1739	3.9715
		FBA	19.1489	0.1017	0.1770	0.2876	13.5587	0.0346	0.0926	0.2401	2.0151
	θ	TP	0.5693	1.1692	3.0570	1.5329	0.9677	5.9211	3.5146	6.8454	9.3889
		LVCF	0.3577	0.6566	1.0459	0.5490	0.2943	1.5880	0.8066	0.6529	0.5398
		FBA	0.3990	0.7391	1.2685	0.6720	0.4005	2.3592	1.9527	1.2125	0.8986
	σ_x	TP	0.0277	0.0202	0.0387	0.0110	0.0322	0.0675	0.0623	0.1646	0.5256
		LVCF	0.0183	0.0116	0.0182	0.0049	0.0112	0.0133	0.0080	0.0136	0.0180
		FBA	0.0210	0.0189	0.0275	0.0076	0.0226	0.0298	0.0411	0.0315	0.0324
In-sample	RMSE	TP	0.6719	0.0724	0.0697	0.0183	0.2766	0.0386	0.0416	0.1349	0.4851
		LVCF	0.6712	0.0414	0.0752	0.0189	0.2409	0.0355	0.0321	0.0852	0.2568
		FBA	0.6696	0.0496	0.0745	0.0199	0.2675	0.0381	0.0393	0.0980	0.3134
	MAPE	TP	0.3048	0.2018	0.0817	0.1040	0.4312	0.1035	0.2578	0.7484	0.3283
		LVCF	0.3069	0.1693	0.0943	0.1022	0.4292	0.1013	0.2267	0.6877	0.3144
		FBA	0.3073	0.1832	0.0919	0.1040	0.4621	0.1064	0.2507	0.8080	0.3683
Out-of-sample	RMSE	TP	3.2199	0.0657	0.5957	0.1540	0.0372	0.0901	0.0744	0.2407	0.9104
		LVCF	3.2341	0.0584	0.6119	0.1017	0.0522	0.0982	0.0607	0.2135	0.9612
		FBA	1.5223	0.0474	0.6126	0.0399	0.0503	0.0807	0.0681	0.2562	0.6171
	MAPE	TP	1.9473	0.3994	0.4376	0.5236	0.0879	0.2124	0.4878	2.4140	0.7923
		LVCF	1.9517	0.3746	0.4437	0.4303	0.1041	0.2255	0.4429	2.2487	0.8148
		FBA	1.5832	0.3325	0.4440	0.2507	0.1234	0.1987	0.4669	2.4886	0.6515

Table 5.6: Calculated median parameter estimates, median standard errors and standard deviations of the generated posterior distributions using OU on the traded point (TP) time series, the LVCF time series and then the application of FBA to the TP time series. In- and out-of-sample RMSE and MAPE values are also shown.

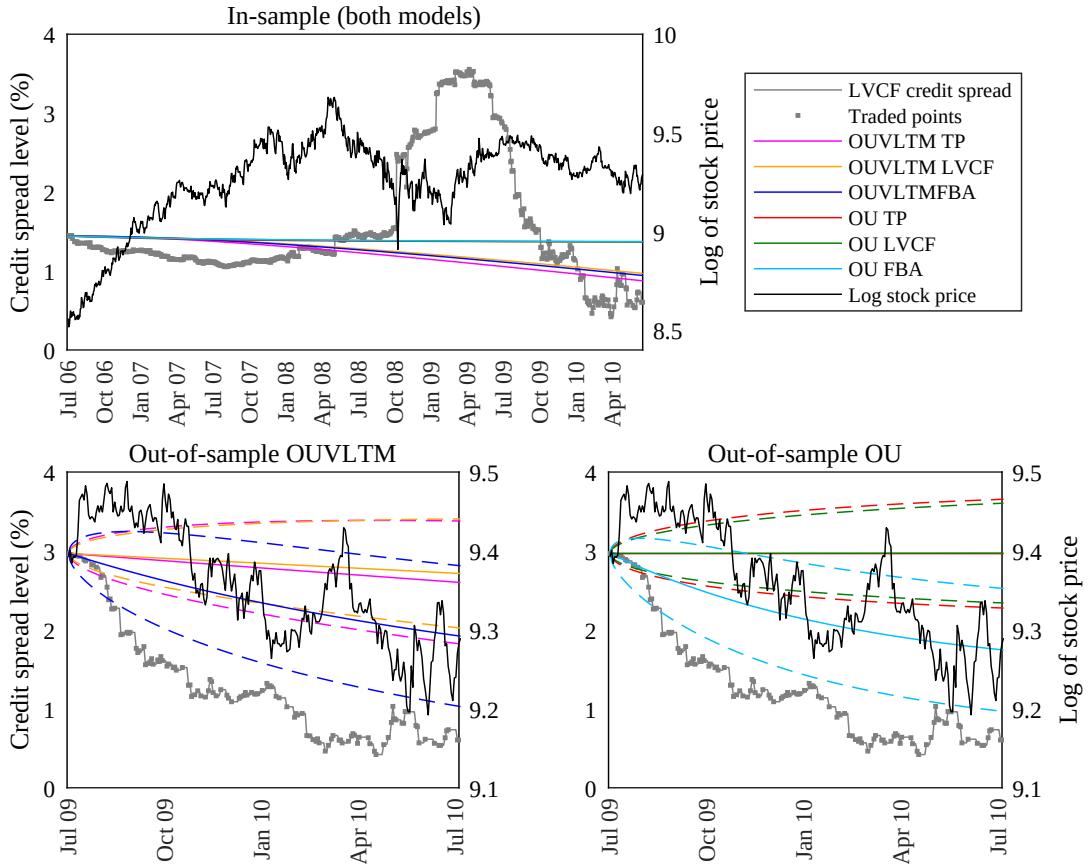


Figure 5.6: MTN01

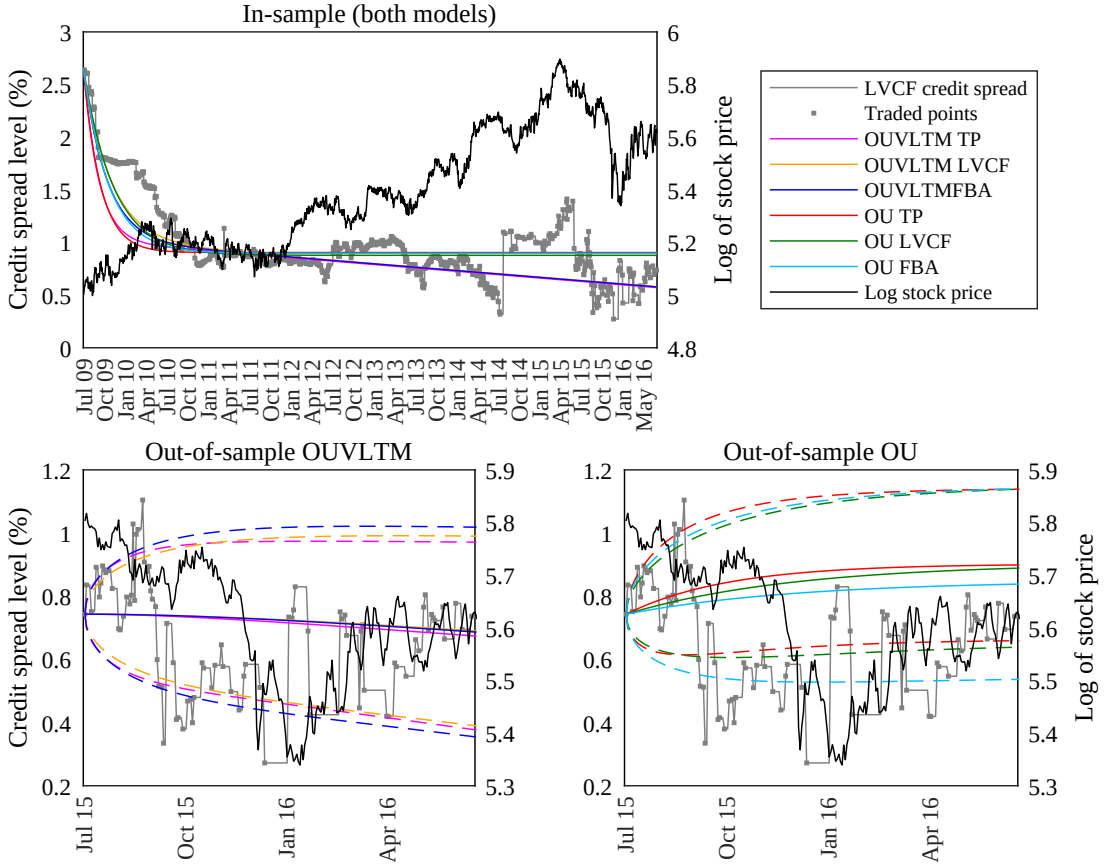


Figure 5.7: SBS9

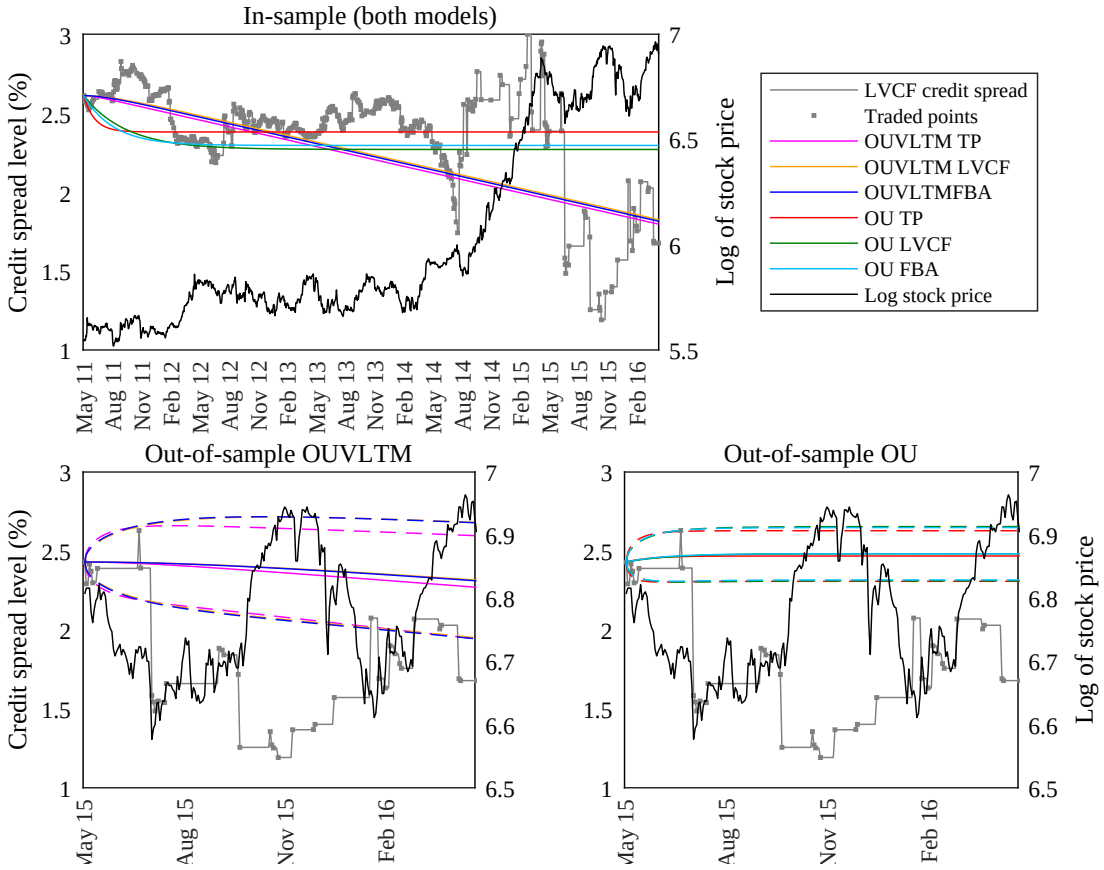


Figure 5.8: CBL11

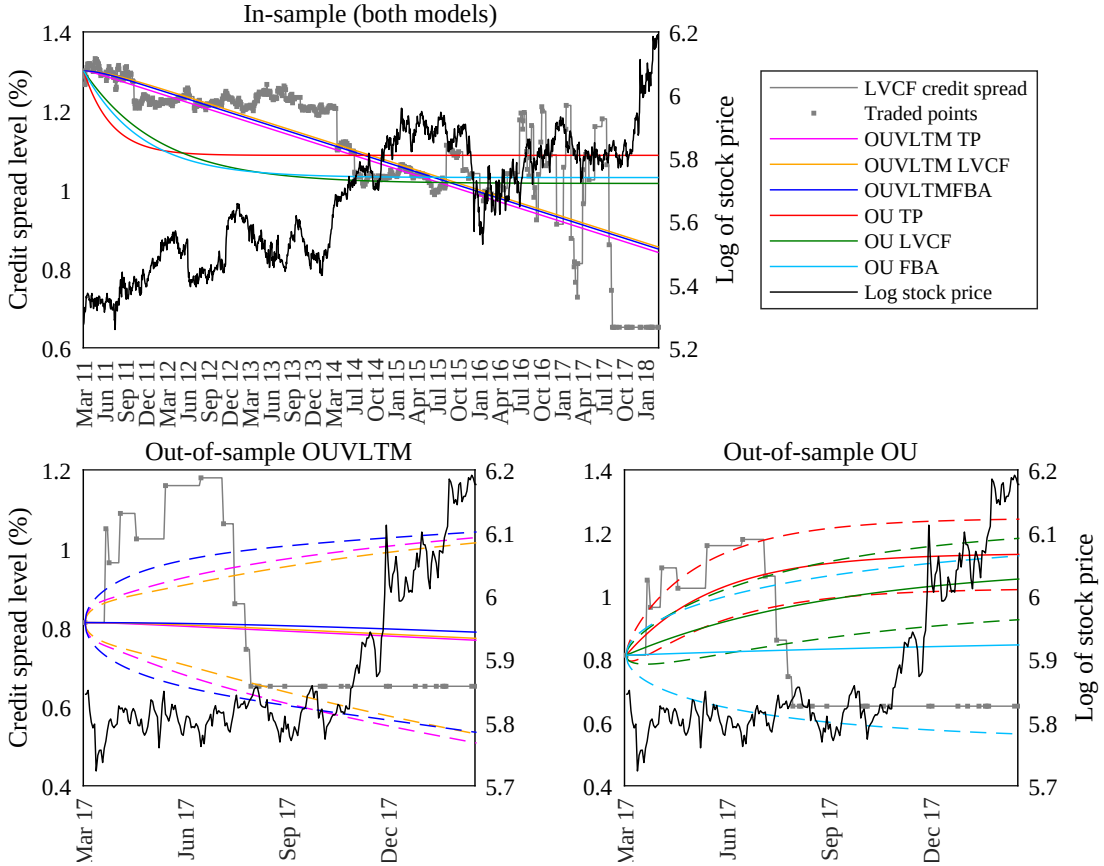


Figure 5.9: ABS10

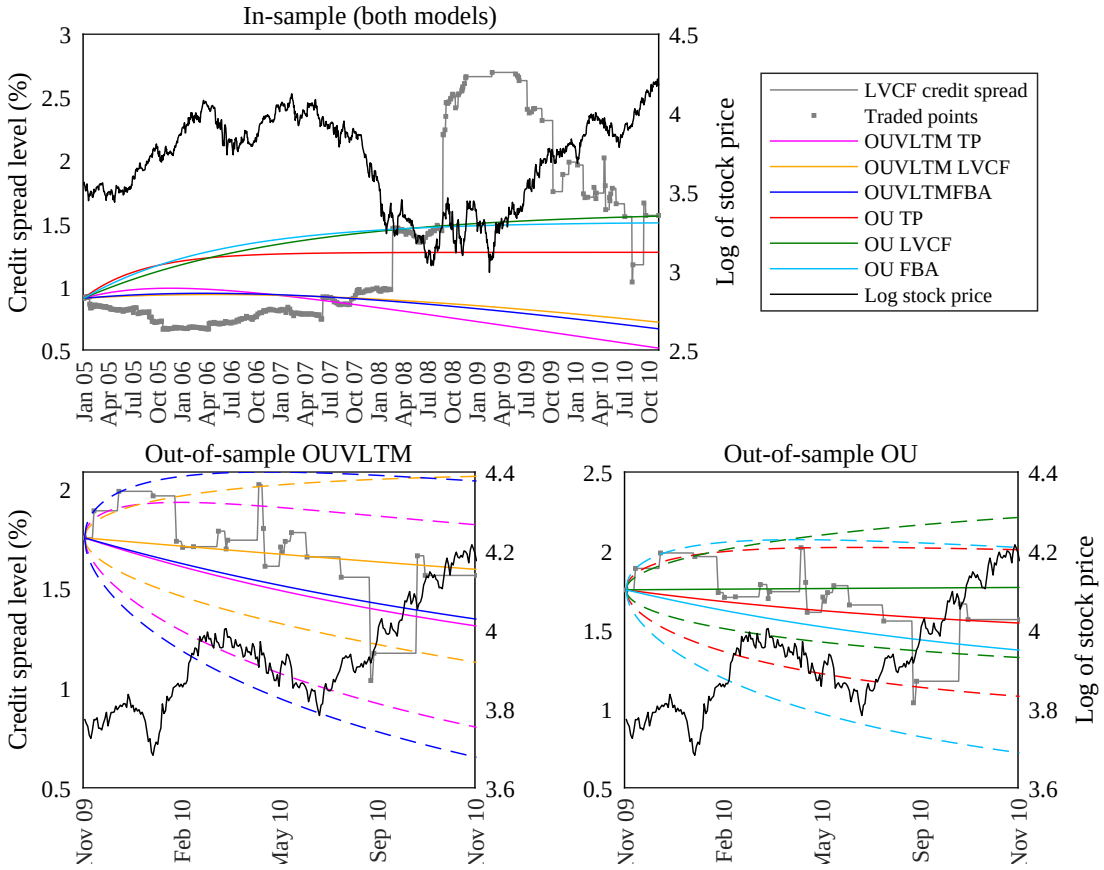


Figure 5.10: IPL3

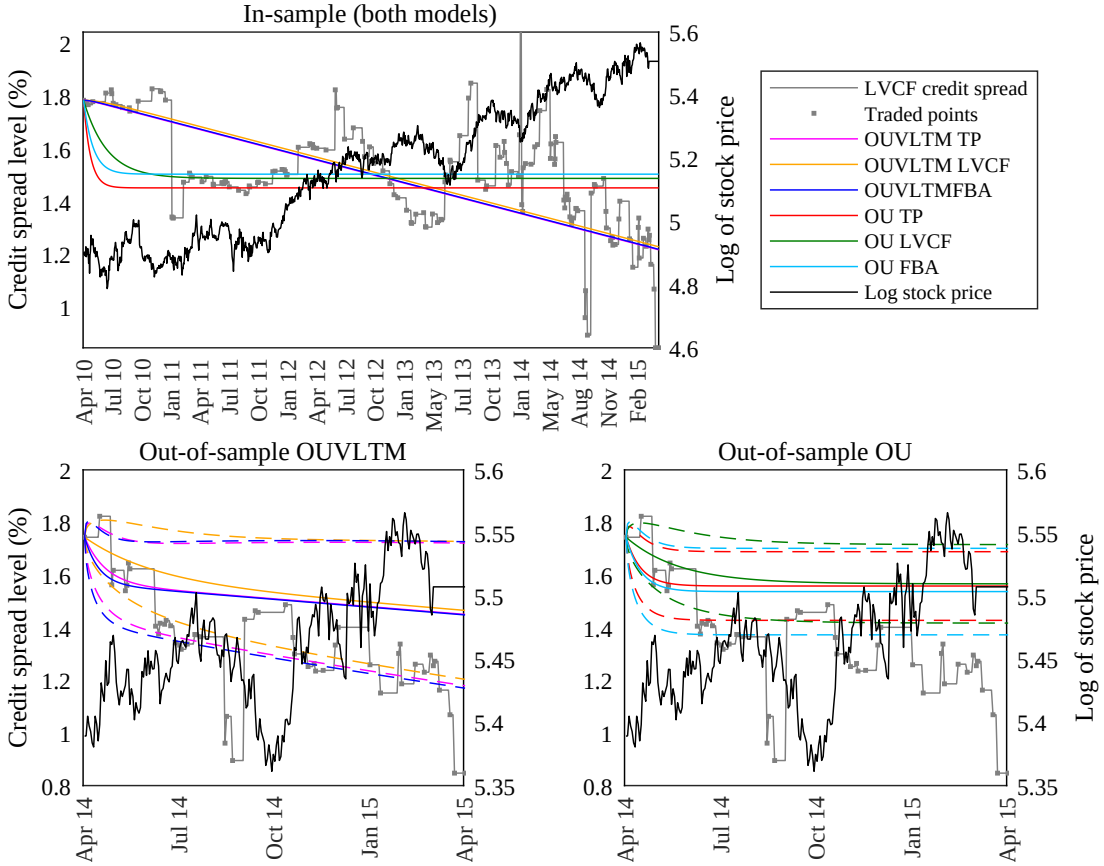


Figure 5.11: NBK6A

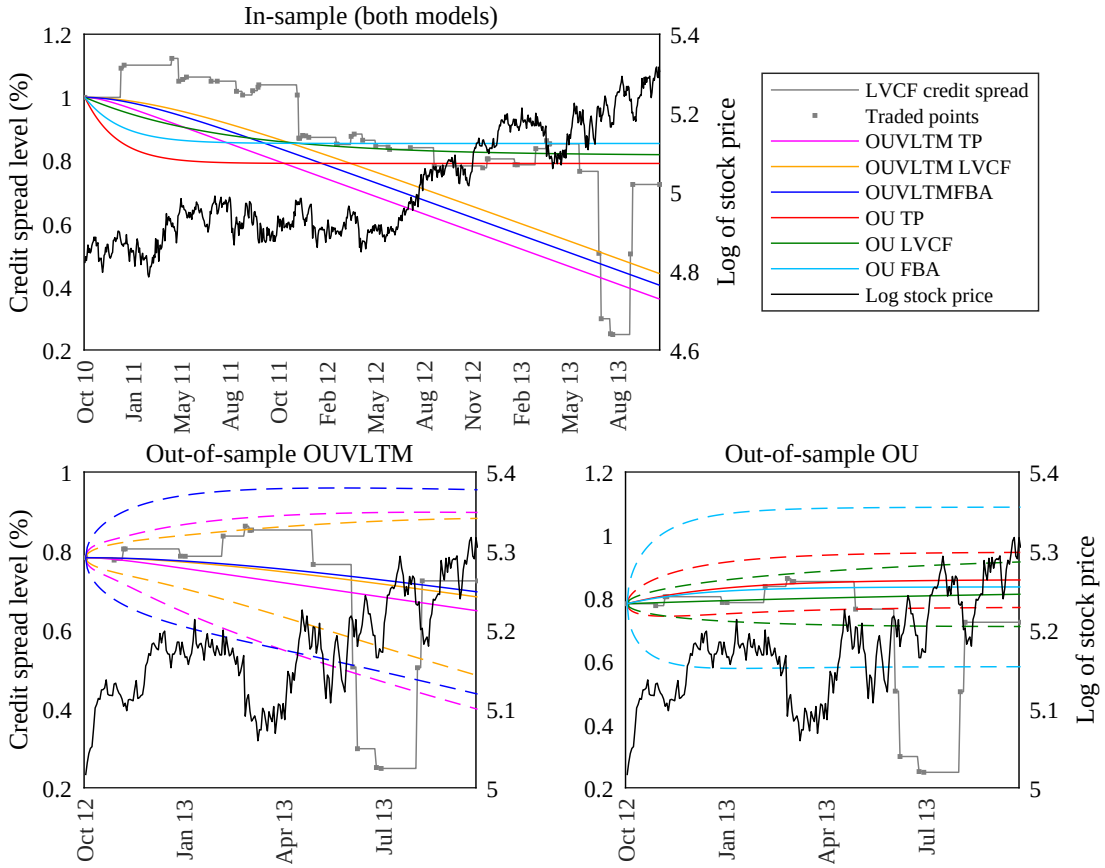


Figure 5.12: MTN05

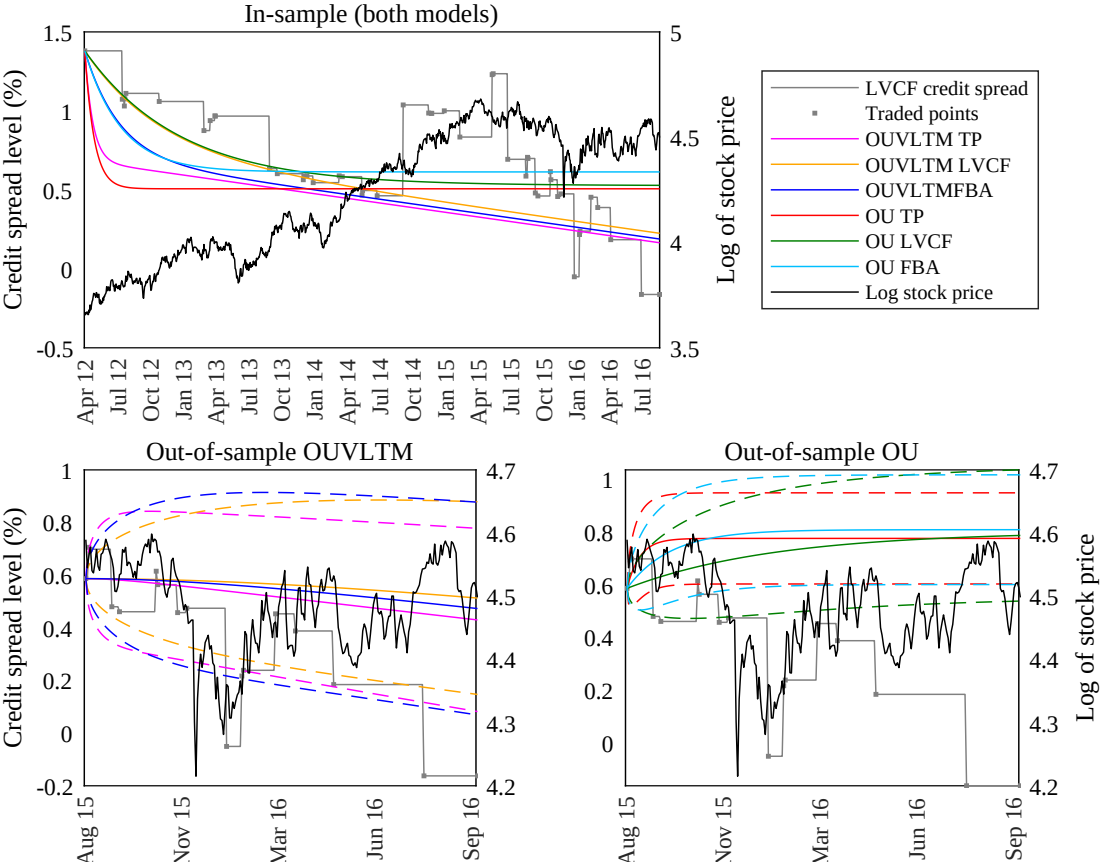


Figure 5.13: FRX16

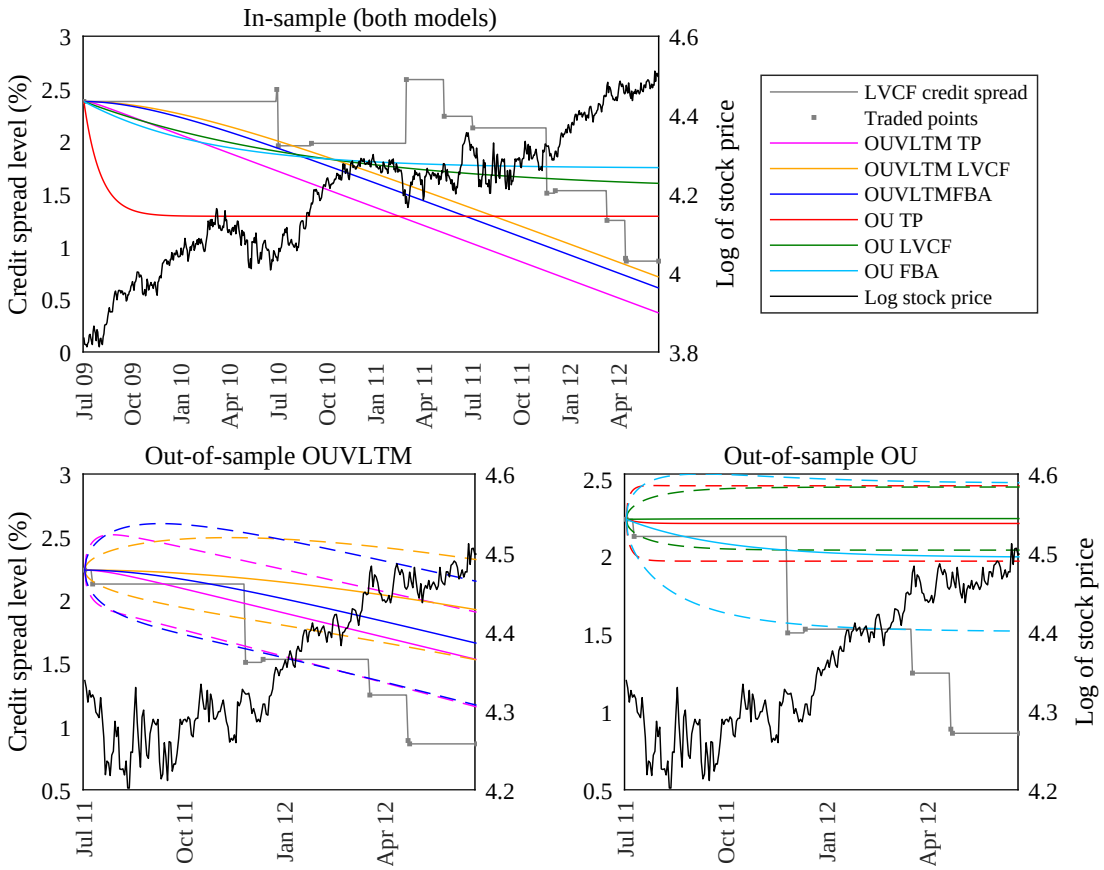


Figure 5.14: BID02

6. Computational Implementation

The computational requirements to produce the content in Chapters 3 to 5 were extensive. The details of how the theory put forward was numerically executed are presented in this chapter. Additionally, any areas that exhibited particularly interesting computational challenges are discussed.

All programming and analyses were carried out in Matlab using a Dell Precision M6600 machine with 16 GB of RAM and 2.7 GHz of RAM.

6.1 Testing for Stationarity

To determine whether a time series was stationary or not, the built-in Matlab function `adftest` from the Econometrics toolbox was used. This function returns a logical value of 0, indicating the rejection of the null hypothesis (existence of a unit root) in favour of stationarity. The function also returns the test statistic, its p-value and the critical value (refer to Appendix A.2.8 for more detail).

Additionally, the function allows the user to specify the model variant, which was chosen to be `ARD`: the test for the null follows an autoregressive model with drift variant

$$x(t) = x(t-1) + \rho_1 \Delta x(t-1) + \rho_2 \Delta x(t-2) + ... + \rho_p \Delta x(t-p) + \epsilon(t)$$

against an AR(1) process with constant non-zero drift c and $\phi < 1$

$$x(t) = c + \phi x(t-1) + \rho_1 \Delta x(t-1) + \rho_2 \Delta x(t-2) + ... + \rho_p \Delta x(t-p) + \epsilon(t).$$

For time series to be consistent with the mean-reverting OU model, $c > 0$ and $\phi < 1$. So `adftest` rejecting the null hypothesis in favour of the alternative model implied that the OU model was a reasonable representation of the time series dynamics.

6.2 Simulation of Stochastic Differential Equations

All stochastic models were simulated using Euler's method. An example of Euler discretisation was already seen at (3.13) in Section 3.3. In general, Euler's method describes a stochastic process $y(t)$ by the difference approximations for the process itself, the Wiener process noise component $W(t)$ and t, (assuming constant model parameters):

$$\begin{aligned} dY(t) &\approx Y(t_{i+1}) - Y(t_i) = Y_{i+1} - Y_i = \Delta Y_i, \\ dW(t) &\approx W(t_{i+1}) - W(t_i) = W_{i+1} - W_i = \Delta W_i, \\ dt &\approx t_{i+1} - t_i = \Delta t, \end{aligned}$$

where $\Delta W_i \approx \sqrt{\Delta t}\epsilon_i \;\; \epsilon_i \sim N(0,1)$.

For the OUVLTM, $\tilde{\mu}$, OU and GBM processes, the Euler discretised equations used in the numerical simulations were as follows:

$$\begin{aligned}
\tilde{X}(i+1) &= \tilde{X}(i) + \theta(\tilde{\mu}(i) - \tilde{X}(i))\Delta t + \sigma_x\sqrt{\Delta t}\epsilon_x(i),\\
\tilde{\mu}(i+1) &= \tilde{\mu}(i) - (\mu_s - \frac{1}{2}\sigma_s^2)\Delta t - \sigma_s\sqrt{\Delta t}\epsilon_s(i),\\
X(i+1) &= X(i) + \theta(\mu_x - X(i))\Delta t + \sigma_x\sqrt{\Delta t}\epsilon_x(i),\\
S(i+1) &= S(i)(1 + \mu_s S(i)\Delta t + \sigma_s S(i)\sqrt{\Delta t}\epsilon_s(i).
\end{aligned}$$

The simulated sample paths in Chapters 3 and 4 for $\tilde{X}$ and X were generated using the same vector of normal random variates (initialised from the same seed) to enable precise comparisons of the processes.

6.3 Numerical Parameter Estimation

6.3.1 Likelihoods and Maximum Likelihood Estimation

Numerical calculation of the likelihood functions in Section 3.3 and Appendices A.2.6 and A.2.5 took advantage of Matlab's ability to operate on whole matrices and arrays, as shown in the Matlab function `likelihood_OUVLTM` for the likelihood of OUVLTM given at (3.23).

```
function L = likelihood_OUVLTM(n,x,y,s_pars,dt,tildemu)

    zeta = dt;
    eta = spars(1,1)-0.5*spars(1,2)^2;
    L = -0.5*(n-1)*log(2*pi*y(2))-...
    sum((x(2:end,1)-y(1).*x(1:end-1,1)-(1-y(1)).*tildemu(1:end-1,1)-eta*zeta).^2)/(2*y(2));

end
```

In this function the input `s_pars` are the estimates from the GBM process for the log stock price. Applying operations to the array `x` as a whole (such as using the multiplication operator `.*`), as opposed to working on each element of `x` individually, reduces the computational time is significantly.

Similarly, the likelihoods of OU and GBM (assuming the input stock price time series was lognormally distributed) were coded in the functions `likelihood_OU` and `likelihood_GBMlog`:

```
function L = likelihood_OU(n,x,y)

    L = -0.5*(n-1)*log(2*pi*y(3)^2)-...
              sum((x(2:end,1)-y(2).*x(1:end-1,1)-y(1)).^2)/(2*y(3)^2);

end
```

and

```
function L = likelihood_GBMlog(n,x,y,dt)

    L = -0.5*(n-1)*log(2*pi*dt*y(2)^2)-sum(log(x(2:end,1))+...
            ((log(x(2:end,1)./x(1:end-1,1))-(y(1)-0.5*y(2)^2)*dt).^2)./(2*dt*y(2)^2));

end
```

where the log GBM likelihood was determined using (A.14).

The maximisations at (3.16), (A.39) and (A.20), to obtain the optimal parameter estimates was achieved by using the Matlab function `fmincon` from the Matlab Optimisation toolbox. Since this function finds the constrained minimum of a function of several variables, the negative of the likelihood was taken - finding the maximum of a positive function is the same as finding the minimum of its negative. The following line of code executed the optimisation using `fmincon` on the OUVLTM model.

```
[par_est, maxL]=fmincon(@(y)-likelihood_OUVLTM(n,x,y,s_pars,dt,tildemu),...
                                                    [0.5,0.5],[],[],[],[],[0,0],[1,ul]);
```

The starting values for the OUVLTM parameters ϕ and $V_{\tilde{A}}$ were set to $[0.5, 0.5]$, with constraints

$$0 < \phi < 1 \quad \text{and} \quad 0 < V_{\tilde{A}} < \texttt{ul},$$

where `ul` was set to some large number to proxy infinity. The optimisations for OU and GBM were executed similarly.

6.3.2 Conversions to Original Model Parameters

Both OUVLTM and OU were reparametrised at (3.17), (3.20) and (A.36), simplifying the likelihood functions and the parameters to be estimated. The function `convert_pv_to_ts` was used to convert estimates of OUVLTM's ϕ and $V_{\tilde{A}}$ to estimates for θ and σ_x.

```
function [theta, sigma_x] = convert_pv_to_ts(n,phiV,sigma_s,rho_sx,dt)

    m = length(phiV);
    theta = -log(phiV(:,1))./dt;

    B = (1-phiV(:,1).^2)./(2*theta);
    C = -(sigma_s.*rho_sx.*(1-phiV(:,1)).^2)./theta;
    D = (sigma_s^2).*(A-(1-phiV(:,1))./theta+dt);

    sigma_x = zeros(m,1);
    parfor i = 1:m
        [sigma_x(i,1),~] = fmincon(@(y) quadfunc(y,B(i,1),C(i,1),D(i,1),phiV(i,2)),...
                        0.5,[],[],[],[],0,ul);
    end

end
```

This function again made use of Matlab's matrix operation ability to convert vectors of ϕ and $V_{\tilde{A}}$ to vectors of θ and σ_x. Since σ_x cannot be negative, solving (3.28) becomes a constrained optimisation. The function used `fmincon` to minimise the objective function `quadfunc`, equal to the terms on the left side of (3.27), under the constraint that $\sigma_x > 0$. `fmincon` only takes in and solves for scalars, thus the number of elements in the parameter vector were looped through to produce a vector for σ_x.

`parfor` is the for loop activator from the Matlab's Parallel Computing toolbox, which partitions the number of iterations in the loop according to the number of available processing cores of the machine. The partitions are then farmed out across the cores to be computed at the same time, resulting in a massive improvement in computational time. In order to use parallel computing however, the code needed to be written such that there is no dependence between arrays within the `parfor` loop.

The conversion for OU at (A.37) was a far simpler plug-and-play calculation.

6.3.3 Gibbs Sampler Algorithm

Figure 6.1 diagrammatically represents the process of Gibbs sampling to generate posterior distributions for the OUVLTM model parameters ϕ and $V_{\tilde{A}}$ (refer to Section 3.3.2 and Appendix B.5.3).

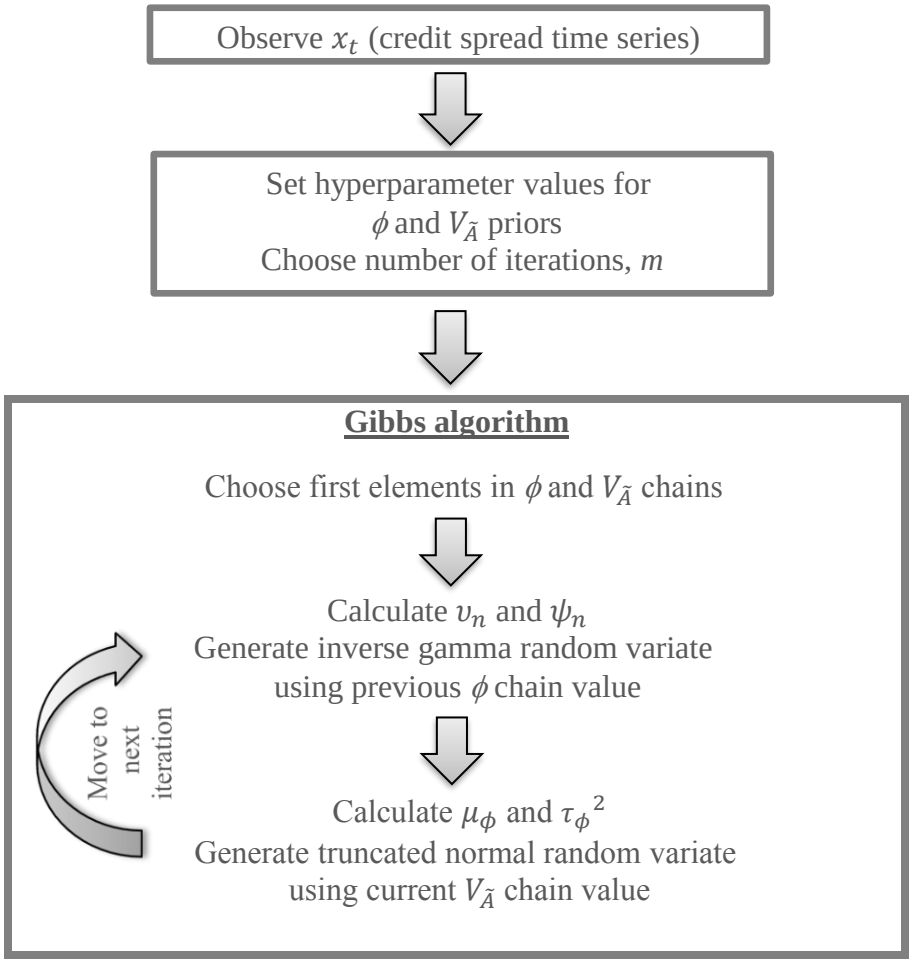


Figure 6.1: Representation of the Gibbs sampler algorithm.

The function that executed the algorithm numerically is `Gibbs_OUVLTM`.

```
function [chains_phi,chains_V] = Gibbs_OUVLTM(x,n,dt,eta,tildemu,iterations,burnin,sp)

    zeta = dt;
    mu0_phi = 0.5;
    t20_phi = 1;

    nu0 = 1;
    psi0 = 0.01;
    nun = nu0+(n-1);

    l_phi = 0;
    u_phi = 1;

    chains= zeros(iterations,2);
    for i = 1:iterations

        if i==1
            temp1 = sp;
        else
            temp1 = chains(i-1,1);
        end

        Aiminus1 = x(1:end-1,1)-tildemu(1:end-1,1);
        Ai = x(2:end,1)-tildemu(1:end-1,1);

        s2n = sum((Ai-temp1.*Aiminus1- eta*zeta).^2);
        psi2n = (nu0*psi0^2+s2n);
        dummy_V = gamrnd(nun/2,1/(psi2n/2));
        chains(i,2) = 1/dummy_V;

        a_phi = 1/t20_phi +(sum(Aiminus1.^2))/chains(i,2);
        b_phi = mu0_phi/t20_phi + (sum(Aiminus1.*(Ai-eta*zeta)))/chains(i,2);
        mu_phi = b_phi/a_phi;
        tau_rho = 1/a_phi;
        dummy_phi = trandn((l_phi-mu_phi)/sqrt(tau_phi),(u_phi-mu_phi)/sqrt(tau_phi));
        chains(i,1)=mu_phi+sqrt(tau_phi)*dummy_phi;

    end

    chains_phi = chains(burnin+1:iterations,1);
    chains_V = chains(burnin+1:iterations,2);

end
```

The hyperparameters that were used can be seen in Table 3.5 and the starting point `sp` of the ϕ chain was set to a reasonable yet non-specific value. The draw for $V_{\tilde{A}}$ utilised the built-in Matlab `gamrnd` function, which generates a random variate from the gamma distribution. This value was then transformed into an inverse gamma variate using

$$\frac{1}{V_{\tilde{A}}} \sim \mathcal{G}\left(\nu_n, \frac{1}{\psi_n}\right) \Longleftrightarrow V_{\tilde{A}} \sim \mathcal{IG}(\nu_n, \psi_n).$$

The `gamrnd` function takes the scale parameter of the distribution as the second input argument, which was inverted in order to generate an inverse gamma random variate.

So as to incorporate the limitations on the domain of ϕ, the random normal truncation function `trandn` (Botev (2016)) was employed. As per the author's instruction, if "one wishes to simulate a non-standard normal variate $z \sim N(m, s^2)$ conditional on $l < Z < u$, then first simulate $x =$ `trandn`$((l-m)/s, (u-m)/s)$ and set $z = m + sx$". `Gibbs_OUVLTM` also accounted for burn-in by discarding the points at the beginning of the chains, the number of which was set by the input `burnin`. The Gibbs sampler functions for GBM and OU follow the same structure as `Gibbs_OUVLTM`.

6.4 Gibbs Sampler Convergence Diagnostics

Autocorrelation function values for the model parameter chains were determined using the built-in `autocorr` function for $n - 1$ lags.

For the Gelman-Rubin convergence test, 10 more parameter chains were generated using 10 calls to the Gibbs sampler function (`Gibbs_OUVLTM` in the case of OUVLTM), where the starting values of ϕ were set to different uniform random variates. Each call to `Gibbs_OUVLTM` created a unique chain of parameter estimates, the result being 10 distinctive parameter posterior distributions, over and above the original posterior that was generated.

Storage of each of these chains made use of a 3-dimensional array; the first dimension indexing the iterations, the second the model parameter and the third the Gelman-Rubin chain.

The test was computed by execution of the following code.

```
numchains = 11;
chains_phigr = zeros(iterations,1,numchains);
chains_Vgr = zeros(iterations,1,numchains);
parfor k = 1:numchains-1
    sp = rand(1);
    [chains_phigr(:,:,k),chains_Vgr(:,:,k)] = ...
        Gibbs_OUVLTM(x,n,dt,eta,tildemu,iterations,burnin,sp);
end
chains_phigr(:,:,numchains) = chains_phi;
chains_Vgr(:,:,numchains) = chains_V;
ii = iterations - burnin;
GR_R = gettingR([chains_phigr,chains_Vgr],numchains,ii);
```

The function `gettingR` determined the scale reduction factors at (B.14) in Section B.6.4 for each parameter.

```
function GR_R = gettingR(chainsgr,numchains,ii)
    m1 = mean(chainsgr);
    m2 = mean(m1,2);
    sumsq = zeros(1,2);
    sumsq3 = zeros(1,2);
    for i = 1:numchains
        sumsq = sumsq + (m1(:,:,i)-m2).^2;
        sumsq2 = zeros(1,2);
        for j = 1:nn
            sumsq2 = sumsq2 + (chainsgr(j,:,i)-m1(:,:,i)).^2;
        end
        sumsq2 = sumsq2./(ii-1);
        sumsq3 = sumsq3+sumsq2;
    end
    B = ii.*sumsq./(numchains-1);
    W = sumsq3./numchains;
    ssq = (1-1/ii).*W+B./ii;
    GR_R= sqrt(ssq./W);
end
```

6.5 Missing Data

6.5.1 Creating the Inclusion Indicators and Incomplete Sample Paths

Obtaining the perfectly uniform distribution used to construct the inclusion indicator K^λ for the different λ values in Section 4.1 involved dividing the interval $(0, 1)$ into 1,000 equally spaced partitions, forming a series of values increasing from 0 to 1 in steps of size $1/1000 = 0.001$. This series was then repeatedly shuffled using Matlab's built-in function `randperm` to finally give a random series of independent but identically distributed uniform variates, `rr`.

The binary inclusion indicators for $\lambda = 10\%, 50\%$ and 90% were generated using one line of code for each value of λ:

```
K = rr > lambda;
```

Again, matrix operators were used; all 1,000 elements in `rr` were compared to the value of λ in the same calculation, resulting in the 1,000 element inclusion indicator vector, with

$$K^\lambda = \begin{cases} 1, & \texttt{rr} > \lambda \\ 0, & \texttt{rr} < \lambda \end{cases}$$

representing (4.1) from Chapter 4.

The 1,500 incomplete sample paths for the 3 omission rates,`x_red`, were obtained from the 1,500 complete sample paths for the OUVLTM and OU models, `x`, via

```
x_red = x(K==1,:);
n_red = sum(K);
```

Hence a point from the sample paths was removed if the corresponding K^{λ} element was 0. The number of points not removed, which gave the length of the reduced sample paths, `n_red`, was determined by summing the binary elements of the inclusion indicator vector.

6.5.2 Implementing FBA

In Section 4.3.5, imputation of missing data was introduced within the Gibbs sampler. All the missing values of the incomplete input time series, indicated by the inclusion indicator, were also sampled for each iteration in which the parameter draws were generated. This extension of the original OUVLTM Gibbs sampler, referred to generally in this thesis as FBA, can be seen in the schematic representation in Figure 6.2.

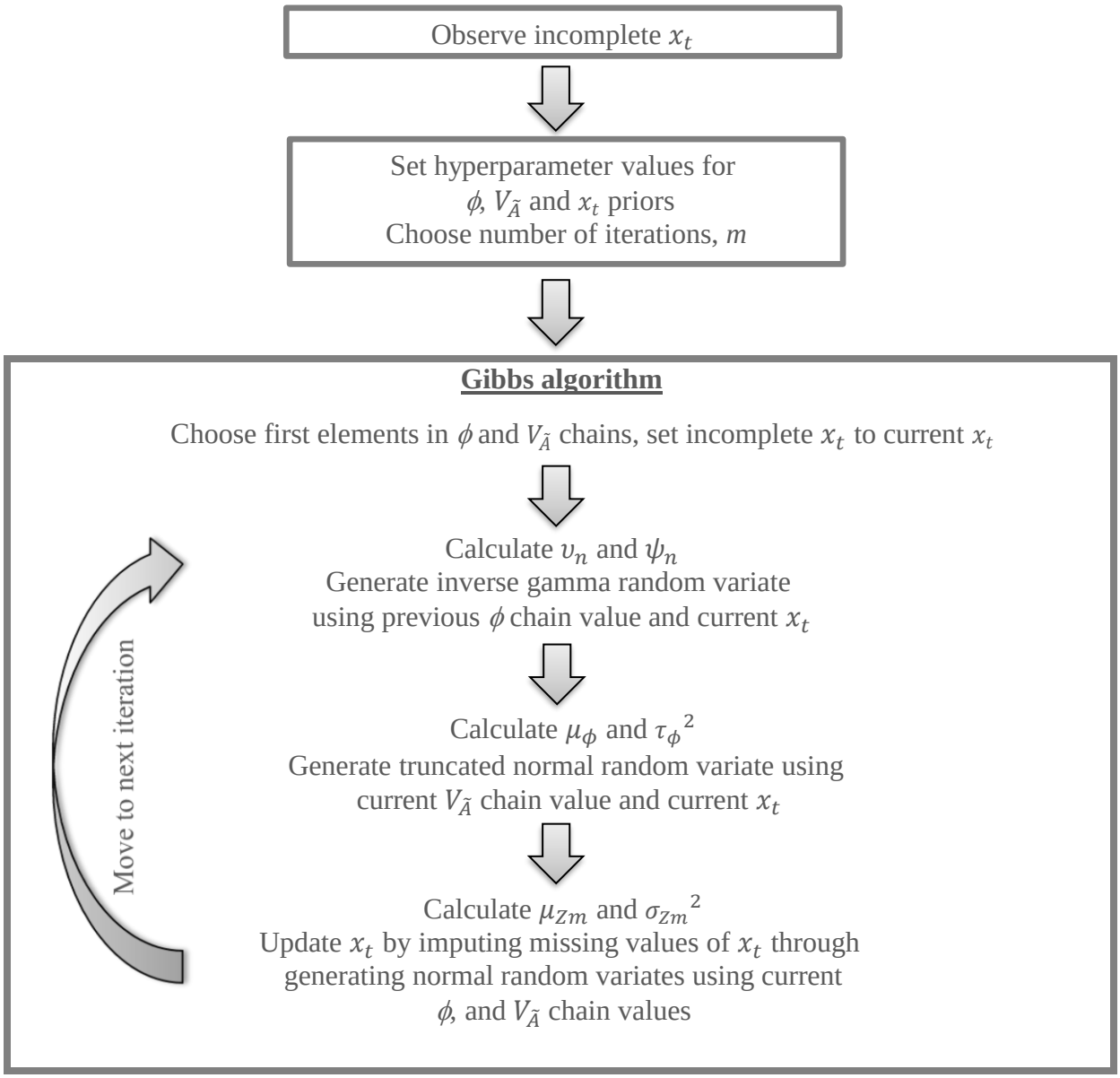


Figure 6.2: Representation of the Gibbs sampler algorithm implementing FBA to impute missing data points.

The function `Gibbs_OUVLTM_FBA` executed the FBA inclusion into the Gibbs sampler for OUVLTM.

```
function [chains_phi,chains_V] = ...
    Gibbs_OUVLTM_FBA(x,n,K,dt,eta,tildemu,iterations,burnin,xleft,xright,sp)

    zeta = dt;
    mu0_rho = 0.5;
    t20_rho = 1;

    nu0 = 1;
    psi0 = 0.01;
    nun = nu0+(n-1);

    mu0_mx = 0.5;
    t20_mx = 1;

    l_phi = 0;
    u_phi = 1;

    chains= zeros(iterations,2);
    for i = 1:iterations

        if i==1
            temp1 = sp;
        else
            temp1 = chains(i-1,1);
        end

        Aiminus1 = x(1:end-1,1)-tildemu(1:end-1,1);
        Ai = x(2:end,1)-tildemu(1:end-1,1);

        s2n = sum((Ai-temp1.*Aiminus1- eta*zeta).^2);
        psi2n = (nu0*psi0^2+s2n);
        dummy_V = gamrnd(nun/2,1/(psi2n/2));
        chains(i,2) = 1/dummy_V;

        a_phi = 1/t20_phi +(sum(Aiminus1.^2))/chains(i,2);
        b_phi = mu0_phi/t20_phi + (sum(Aiminus1.*(Ai-eta*zeta)))/chains(i,2);
        mu_phi = b_phi/a_phi;
        tau_rho = 1/a_phi;
        dummy_phi = trandn((l_phi-mu_phi)/sqrt(tau_phi),(u_phi-mu_phi)/sqrt(tau_phi));
        chains(i,1)=mu_phi+sqrt(tau_phi)*dummy_phi;

        barmx = (xleft+xright)*chains(i,1)/(1+chains(i,1)^2);
        a_mx = (chains(i,2)*mu0_mx+ t20_mx*(1+chains(i,1)^2)*barmx(K==0))/...
            (chains(i,2)+t20_mx*(1+chains(i,1)^2));
```

```
        b_mx = chains(i,2)* t20_mx/(chains(i,2)+ t20_mx^2*(1+chains(i,1)^2));
        xm_sample = a+sqrt(b)*randn(n-sum(K),1);
        x(K==0) = xm_sample;

    end

    chains_phi = chains(burnin+1:iterations,1);
    chains_V = chains(burnin+1:iterations,2);

end
```

From Figure 6.2 and `Gibbs_OUVLTM_FBA`, observe that the imputation of missing data was just one more set of calculations over and above those of the original Gibbs sampler, and was implemented in 5 additional lines of code. A random normal variate was generated for each of the `n-sum(K)` missing points and these samples were slotted into the data series where K^{λ} indicated that there were missing values (`K==0`).

`barmx` stored the calculated value of (4.4), which is a function of ϕ, the previous observed point and the next observed point. For each time point of the data series, the vectors `xleft` and `xright` stored the values of the last actual observed point and next actual observed point, at the same time point index. These vectors were determined in the function `xleftright`, executed before `Gibbs_OUVLTM_FBA`.

```
function [xleft,xright] = xleftright(n,q,x,K)

    xright=zeros(n,q);
    for i=n-1:-1:2
        if K(i+1,1)==1 &&  K(i,1)==0
            xright(i,:) = x(i+1,:);
        elseif K(i+1,1)==0 &&  K(i,1)==0
            xright(i,:) = xright(i+1,:);
        end
    end

    xleft=zeros(n,q);
    for i=2:n-1
        if K(i-1,1)==1 &&  K(i,1)==0
            xleft(i,:) = x(i-1,:);
        elseif K(i-1,1)==0 &&  K(i,1)==0
            xleft(i,:) = xleft(i-1,:);
        end
    end

end
```

7. Conclusion

Over the 13 year period 2005 to 2018, 246 pure vanilla bonds that traded on Johannesburg Stock Exchange were examined. The multiple approaches to determine the credit spreads of these bonds were discussed and the flaw of duration mismatch inherent in the current South African (SA) market methodology, due to limited market depth, was highlighted. Metrics to ascertain liquidity were introduced, since bid-offer spreads are not available on any public record in SA. These measures showed the severe degree of illiquidity across the 246 bonds, where the most liquid bond only traded 62% of the time and some bonds did not trade at all. The current method used to handle the missing data points due to lack of trade, Last Value Carried Forward (LVCF), was demonstrated to cause inconsistent and unstable terms structures and anti-intuitive relative spread levels of bonds of the same issuer but differing seniority. Two recent SA credit crises (African Bank and Steinhoff) showed that the credit spreads of their bonds did not react to reflect the calamities that their issuers were facing. This investigation provided compelling evidence to motivate the need for an alternative to the present LVCF scheme.

The general nature of credit spreads was explored, using a variety of credit spread indices covering a range of global markets, a liquid US bond and the 15 most liquid SA bonds. The spreads were shown to have displayed some stationarity and exhibited sufficiently decaying autocorrelations. These statistical tests justified the use of the mean reverting Ornstein-Uhlenbeck (OU) model to represent credit spreads, as is prominently seen in the literature. Mean reverting models were shown to fit all the spread time series better than non mean reverting models, once more proving that the assumption to use OU is sound.

The new Ornstein-Uhlenbeck Variable Long Term Mean (OUVLTM), an augmentation of traditional OU model not found in current literature, was introduced. The new model links the long term mean reversion level to the stock price, assumed to be a geometric Brownian motion (GBM). Using simulated sample paths, the OUVLTM model was illustrated to converge and exhibit the anticipated inverse relationship between credit spreads and stock prices. Closed form solutions for the maximum likelihood estimates of the OUVLTM parameters were developed through a reparametrisation, as were the associated standard errors. The conditional posterior distributions of the parameters were also derived so as to generate posterior distributions of parameter estimates using Gibbs sampling. The same theory was developed for the OU and GBM models. Using 1,500 simulated paths, OUVLTM was shown to reduce the bias of the mean reversion rate parameter and produce lower standard errors and parameter posterior standard deviations than OU. In addition, a slight improvement in goodness of fit was found, where OUVLTM resulted in marginally smaller Root Mean Square Errors (RMSEs) and Mean Absolute Percentage Errors (MAPEs). Thus, using simulated data, OUVLTM was found to increase estimation accuracy and reduce uncertainty relative to OU.

To imitate the intermittent observations in the credit spread time series of SA bonds, the simulated paths were shortened through the removal of 10%, 50% and 90% of the sample points. The removal procedure to form the incomplete sample paths was governed by a process that mimicked the historically observed clustering of market trades. For low liquidity bonds, it was observed that there was little to no clustering of trades, with only single or two consecutive trades seen and the more liquid bonds traded often in larger batches of consecutive days. The models were calibrated to the 3 incomplete data sets and both models

produced resulting parameter estimates that were much less accurate than the complete data set. More bias was incurred, with higher uncertainties and diminished goodness of fit, as the omission rate increased.

Multiple imputation of the missing sample points was built into the Gibbs sampler algorithm, altogether known as the Full Bayesian Approach (FBA). When compared with the Gibbs sampling results from the complete and incomplete data set calibrations, the FBA fitted to the incomplete data using OUVLTM and OU distinctly improved the parameter estimations of the incomplete samples. FBA produced biases that were more in line with those of the complete data and the standard errors and parameter posterior standard deviations were significantly diminished. However, the error statistics remained greater than those of the complete sample paths, evidence that the uncertainty due to the missing point imputations was indeed taken into account, leading to more realistic measurements of the total uncertainty. Furthermore, FBA applied to OUVLTM produced smaller RMSEs and MAPEs than FBA on OU, indicative of OUVLTM's superior goodness of fit when calibrated using FBA.

Fitting the models to the credit spreads of two US bonds with daily liquidity, and including a scaling factor, showed that the OUVLTM model was just as suitable, if not more so, than OU for complete credit spread time series. Lower standard errors were produced by OUVLTM and the model also displayed similar or better in- and out-of-sample (forecasting) goodness of fit than OU. The OUVLTM and OU parameters for the traded point time series and LVCF time series of a number of SA bonds of various liquidity were estimated. The traded point time series were also run through the FBA engine in the Gibbs sampler to generate a third set of estimation results. The latter were shown to moderate the results of the traded point and LVCF time series, where the parameter estimates and uncertainty levels lay in between those of the traded points and LVCF. The OUVLTM fit via FBA gave lower means of the standard error posterior medians, meaning increased confidence in the parameter estimates. In-sample testing also showed that OUVLTM FBA resulted in smaller MAPEs for all the bonds than OU FBA, implying OUVLTM FBA was the better model when trends were factored out. Except for one, whose RMSE was higher but MAPE was lower, the out-of-sample testing showed OUVLTM surpass OU to an even larger extent, with all RMSE and MAPE levels being smaller.

Overall, OUVLTM is shown to be a viable candidate to model credit spreads that enhances the estimation and forecasting power of the traditional OU model. When used in conjunction with FBA on credit spreads with missing observations, the model produces parameter estimation results that are more meaningful in terms of uncertainty than LVCF, whilst yielding estimates that provide increased forecasting precision over those generated using FBA on the OU model.

This research can be extended in a myriad of different ways:

- Broaden the empirical data set to include increasingly prevalent instruments, such as FRNs, as a new avenue of extracting liquidity information from the market.
- Finesse the OUVLTM model by considering non-normal log stock prices using Lévy processes.
- Compare the OUVLTM estimations with that of traditional Merton models.
- Create a multi-factor OUVLTM model, where drivers of the mean reversion element are not just the log stock price. Systemic risk, represented by the 5YR SA CDS spread or the JP Morgan Emerging Market

Bond Index for example, should be added as an exogenous factor. Other idiosyncratic determinants, such as the credit rating of the bond or issuer (if available) would be very interesting, especially if there is a credit rating migration.

8. References

Abdi, F and Ranaldo, A. (2017). A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices. *The Review of Financial Studies*. Vol. 30. Iss. 12. pp 4437 – 4480.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*. Vol. 19. Iss. 6. pp 716 – 723.

Amato, J.D. and Remolona, E.M. (2003). The Credit Spread Puzzle. BIS Quarterly Review, December 2003. Available at: https://ssrn.com/abstract=1968448 or http://dx.doi.org/10.2139/ssrn.1968448

Anderson, R.W. (2008). Some Determinants of the Price of Default Risk. Discussion paper, 615. Financial Markets Group, London School of Economics and Political Science, London, UK. Available at: http://eprints.lse.ac.uk/24435/1/dp615.pdf

Azur, M.J., Stuart, E.A., Frangakis, C. and Leaf, P.J. (2011). Multiple Imputation by Chained Equations: What Is It and How Does It Work? *International Journal of Methods in Psychiatric Research*. Vol. 20. Iss. 1. pp 40 – 49.

Ball, C. amd Torous, W. (1996). Unit Roots and the Estimation of Interest Rate Dynamics. *Journal of Empirical Finance*. Vol 3. pp 215 – 238.

Bao, J., Pan, J. and Wang, J. (2011). The Illiquidity of Corporate Bonds. *The Journal of Finance*. Vol 66. Iss. 3. pp 911 – 946.

Bedendo, M., Cathcart, L. and El-Jahel, L. (2007). The Shape of the Term Structure of Credit Spreads: An Empirical Investigation. *Journal of Financial Research*. Vol. 30. No. 2. pp 237 – 257.

Bennett, D.A. (2001). How Can I Deal with Missing Data in My Study? *Australian and New Zealand Journal of Public Health*. Vol 25. No. 5. pp 464 – 469.

Bernales, A., Beuermann, D.W. and Cortazar, G. (2014). Thinly Traded Securities and Risk Management. *Estudios de Economía*. Vol. 41. No. 1. pp 5 – 48.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Association Series B*. Vol. 36. pp 192 – 236.

Bhanot, K. (2005). What Causes Mean Reversion in Corporate Bond Index Spreads? The Impact of Survival. *Journal of Banking and Finance*. Vol. 29. No. 6. pp 1385 – 1403.

Bhar, R. and Handzic, N. (2011). A Multifactor Model of Credit Spreads. *Asia-Pacific Financial Markets*. Vol. 18. Iss. 1. pp 105 – 127.

Black, F. and Cox, J.C. (1976). Valuing Corporate Securities: Some Effects of the Bond Indenture Provisions. *Journal of Finance*. Vol. 31. Iss. 2. pp 351 – 367.
Black, F. and Karasinski, P. (1991). Bond and Option Pricing when Short Rates are Lognormal. *Financial Analysts Journal*. Vol. 47. No. 4. pp 52 – 59.

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*. Vol. 81. No. 6. pp 637 – 654.

Bloomberg. (2018). Why Investors are Piling into Steinhoff Again. BusinessTech article, 14 March 2018. Available at: https://businesstech.co.za/news/business/231495/why-investors-are-piling-into-steinhoff-again

Bogacka, B. (2016). Autoregressive Processes AR(p). *Time Series Module Lecture Notes*. Queen Mary University of London. Available at: http://www.maths.qmul.ac.uk/b̃b/TimeSeries/TS_Chapter4_5.pdf

Bolstad, W.M. (2010). *Understanding Computational Bayesian Statistics*. Wiley, New York.

Bonorchis, R., Khanyile, N. and Prinsloo, L. (2018). Local Steinhoff Unit Mulls Early Redemption of Bonds. Moneyweb article, 12 January 2018. Available at: https://www.moneyweb.co.za/in-depth/investigations/local-steinhoff-unit-mulls-early-redemption-of-bonds/

Bosman, P. and Jones, S. (2008). To Call or Not to Call: A Valuation Model for South African Callable Bonds. Cadiz Securities Research Piece. Personal Communication.

Botev, Z. I. (2016). The Normal Law Under Linear Restrictions: Simulation and Estimation via Minimax Tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Vol. 79. Iss. 1. pp 125 – 148.

Campbell, J.Y., Hilscher, J. and Szilagyi, J. (2008). In Search of Distress Risk. *Journal of Finance*. Vol. 63. Iss. 6. pp 2899 – 2939.

Carr P. and Wu, L. (2010). Stock Options and Credit Default Swaps: a Joint Framework for Valuation and Estimation. *Journal of Financial Econometrics*. Vol. 8. Iss. 4. pp 409 – 449.

Chen, F. (2013). Missing No More: Using the MCMC Procedure to Model Missing Data. SAS Global Forum. Paper 436. Available at: https://support.sas.com/resources/papers/proceedings13/436-2013.pdf

Cortazar, G., Schwartz, E.S. and Naranjo, L.F. (2007). Term Structure Estimation in Markets with Infrequent Trading. *International Journal of Finance and Economics*. Vol. 12. pp 353 – 369.

Cortazar, G., Schwartz, E.S. and Tapia, C. (2012). Credit Spreads in Illiquid Markets: Model and Implementation. *Emerging Markets Finance and Trade*. Vol. 48. pp 53 – 72.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rates. *Econo-*

metrica. Vol. 53. No. 2. pp 385 – 407.

Cronje, J. (2018). Steinhoff Announces Dividend Freeze, Creditor Waiver Proposals. Fin24 article, 2 February 2018. Available at: https://www.fin24.com/Companies/Retail/steinhoff-announces-dividend-freeze-creditor-waiver-proposals-20180202

Crosbie, P. and Bohn, J. (2003). Modelling Default Risk. Moody's KMV. Available at: http://www.defaultrisk.com/pp_model_35.htm

Crotty, A. (2018). Steinhoff's Early Bond Sale Plan Falls Short. Business Day article, 22 February 2018. Available at: https://www.businesslive.co.za/bd/companies/retail-and-consumer/2018-02-22-steinhoffs-early-bond-sale-plan-falls-short/

Das, S.R., Hanouna, P. and Sarin, A. Accounting-Based versus Market-Based Cross-Sectional Models of CDS Spreads. *Journal of Banking and Finance*. Vol. 33. pp 719 – 730.

Delianedis, G. and Geske, R. (2003). Credit Risk and Risk Neutral Default Probabilities: Information About Rating Migrations and Defaults. EFA 2003 Annual Conference Paper No. 962. Available at: https://ssrn.com/abstract=424301 or http://dx.doi.org/10.2139/ssrn.424301

Dick-Nielsen, J., Feldhütter, P. and Lando, D. (2012). Corporate Bond Liquidity Before and After the Onset of the Subprime Crisis. *Journal of Financial Economics*. Vol. 103. Iss. 3. pp 471 – 492.

Dong, Y. and Peng, C.J. (2013). Principled Missing Data Methods for Researchers. *Springerplus*. 2(1):222.

Driessen, J. (2005). Is Default Event Risk Priced in Corporate Bonds? *The Review of Financial Studies*. Vol. 18. Iss. 1. pp 165 – 195.

Duffee, G. (1999). Estimating the Price of Default Risk. *Review of Financial Studies*. Vol. 1. No. 1. pp 197 – 226.

Duffie, D. and Lando, D. (2001). Term Structures of Credit Spreads with Incomplete Accounting Information. *Econometrica*. Vol. 69. No. 3. pp 633 – 644.

Duffie, D. and Singleton, K. (1999). Modelling Term Structure of Defaultable Bonds. *Review of Financial Studies*. Vol. 12. No. 4. pp 687 – 720.

Durbin, J. and Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*. Second Edition. Oxford University Press.

Enders, C.K. (2010). *Applied Missing Data Analysis*. The Guildford Press. New York.

Feldhütter, P. and Schaefer, S.M. (2018). The Myth of the Credit Spread Puzzle. *The Review of Financial Studies*. Vol. 31. Iss. 8. pp 2897 – 2942.

Fleming, M.J. (2003). Measuring Treasury Market Liquidity. *FRBNY Economic Policy Review*. September 2003. Available at: www.newyorkfed.org/research/epr/03v09n3/0309flem.pdf

Fons, J.S. (1994). Using Default Rates to Model the Term Structure of Credit Risk. *Financial Analysts Journal*. Vol 50. Iss. 5. pp 25 – 32.

Fontana, C. and Runggaldier, W. (2010). Credit Risk and Incomplete Information: Filtering and EM Parameter Estimation. *International Journal of Theoretical and Applied Finance*. Vol. 13. No. 05. pp 683 – 715.

Franco, J.C.G. (2003). Maximum Likelihood Estimation of Mean Reverting Processes. Onward, Inc. Available at: http://www.investmentscience.com/Content/howtoArticles/MLE_for_OR_mean_reverting.pdf

Gelman, A. and Rubin, D.R. (1992). A Single Series from the Gibbs Sampler Provides a False Sense of Security. *Statisitical Science*. Vol. 7. No 4. pp 457 – 472.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. Second Edition. Chapman & Hall/CRC .

Gelman, A. and Shirley, K. (2011). Inference from Simulations and Monitoring Convergence. textitHandbook of Markov Chain Monte Carlo. Chapman & Hall/CRC. Brooks, S., Gelman, A., Jones, G.L. and Meng, X.L. (Eds). pp 163 – 174.

Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis*. Vol 12. No. 4. pp 541 – 552.

Geyer, C. (2011). Introduction to Markov Chain Monte Carlo. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC. Brooks, S., Gelman, A., Jones, G.L. and Meng, X.L. (Eds). pp 3 – 48.

Giet, L. and Lubrano, M. (2004). Bayesian Inference in Reducible Stochastic Differential Equations. Document de Travail No. 2004 – 57, GREQAM University d'Aix-Marseille. Available at: https://pdfs.semanticscholar.org/86c6/17a2852f1be91f677d3faec102a12e001105.pdf

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Greeff, P. (2004). Calculating Credit Spreads Using the BEASSA Zero Curve. RisCura Fixed Income Research Series. Available at: www.riscura.com/sites/default/files/case_study_docs/report_2004feb.pdf

Halls-Moore, M. (2013). Testing for Mean Reversion. Article Series. Available at: https://www.quantstart.com/articles/Basics-of-Statistical-Mean-Reversion-Testing.

Hammersley, J. and Clifford, P. (1970). Markov Fields on Finite Graphs and Lattices. Unpublished

Manuscript. Available at: http://www.statslab.cam.ac.uk/ grg/books/hammfest/hamm-cliff.pdf

He, Z. and Milbradt, K. (2014).Endogenous Liquidity and Defaultable Bonds. *Econometrica*. Vol. 82. Iss. 4. pp 1443 – 1508.

He, Z. and Xiong, W. (2012). Rollover Risk and Credit Risk. *The Journal of Finance*. Vol. 67. Iss. 2. pp 391 – 430.

Heath, D., Jarrow, R. and Morton, A. (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*. Vol. 60. Iss. 1. pp 77 – 105.

Heston, S.L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*. Vol. 6. Iss. 2. pp 327 – 343.

Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods*. Springer, New York.

Honaker, J. and King, G. (2010). What to Do about Missing Values in Time Series Cross-Section Data. *American Journal of Political Science*. Vol. 54. No. 2. pp 561 – 581.

Huang, J. and Huang, M. (2012). How Much of the Corporate-Treasury Yield Spread is due to Credit Risk? *The Review of Asset Pricing Studies*. Vol. 2. Iss. 2. pp 153 – 202.

Hull. J.C. (2009). *Options, Futures and Other Derivatives*. Seventh Edition. Wiley, New York.

Hung, K., Duan, C. and Yang, C.W. (2006). Rating, Credit Spread, and Pricing Risky Debt: Empirical Study on Taiwan's Security Market. *Annals of Economics and Finance*. Vol. 7. Iss. 2. pp 405 – 424.

Hunt, P.J. and Kennedy, J.E. (2004). *Financial Derivatives in Theory and Practice: Revised Edition*. Wiley, New York.

Jarrow, R.A, and Turnbull, S.M. (1995). Pricing Options on Financial Securities Subject to Credit Risk. *Journal of Finance*. Vol. L. No. 1 pp 53 – 85.

Jarrow, R.A., Lando, D. and Turnbull, S.M. (1997). A Markov Model for the Term Structure of Credit Risk Spreads. *The Review of Financial Studies*. Vol. 10. Iss. 2. pp 481 – 523.

Jaskowski, M. and McAleer, M. (2012). Estimating Implied Recovery Rates from the Term Structure of CDS Spreads. Tinbergen Institute Discussion Papers 13-005/III, Tinbergen Institute. Available at: http://eprints.ucm.es/17581/1/1228.pdf.

Johannes, M.S. and Polson, N. (2010). MCMC Methods for Continuous-Time Financial Econometrics. *Handbook of Financial Econometrics: Applications*. Vol. 2. pp 1 – 72.

Jones, C.S. (1998). Bayesian Estimation of Continuous-Time Finance Models. Simon School of Business,

University of Rochester. Working Paper. Available at: http://www-bcf.usc.edu/∼christoj/pdf/hfa7.pdf

Jones, P.E., Mason, S.P and Rosenfeld, E. (1985). Contingent Claims Valuation of Corporate Liabilities: Theory and Empirical Tests. *Corporate Capital Structures in the United States*. Edited by Benjamin M. Friedman. Chicago: University of Chicago Press. pp 239 – 261.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*. Vol. 82. Iss. 1. pp 35 – 45.

Kealhofer, S. (2003i). Quantifying Credit Risk I: Default Prediction. *Financial Analysts Journal*. Vol. 59 No. 1. pp 33 – 44.

Kealhofer, S. (2003ii). Quantifying Credit Risk II: Debt Valuation. *Financial Analysts Journal*. Vol. 59 No. 3. pp 78 - 92.

Kelly, F.P. (1979). *Reversibility and Stochastic Processes*. Wiley, New York.

Kwak, S.K. and Kim, J.H. (2017). Statistical Data Preparation: Management of Missing Values and Outliers. *Korean Journal of Anesthesiology*. Vol 70. No. 4. pp 407 – 411.

Lachin, J.M. (2016). Fallacies of Last Observation Carried Forward Analyses. *Clinical Trials*. Vol. 13. Iss. 2. pp 161 – 168.

Laine, M., Latva-Pukkila, N. and Kyrölä, E. (2014). Analysing Time-Varying Trends in Stratospheric Ozone Time Series using the State Space Approach. *Atmospheric Chemistry and Physics*. Vol. 14. pp 9707 – 9725.

Lando, D. (1994). Three Essays on Contingent Claims Pricing. Ph.D. Thesis, Cornell University. Available on request at: https://www.researchgate.net/publication/34551831_Three_essays_on_contingent_claims_pricing

Leland, H.E. (1994). Corporate Debt Value, Bond Covenants and Optimal Capital Structure. *Journal of Finance*. Vol. 49. Iss. 4. pp 1213 – 1252.

Leland, H.E. and Toft, K.B. (1996). Optimal Capital Structure, Endogenous Bankruptcy and the Term Structure of Credit Spreads. *The Journal of Finance*. Vol. 51. No. 3. pp 987 – 1019.

Lesmond, D.A., Ogden, J.P. and Trzcinka, C. (1999). A New Estimate of Transaction Costs. *Review of Financial Studies*. Vol. 12. Iss 5. pp 1113 – 1141.

Letham, B. and Rudin, C. (2012). Probabilistic Modelling and Bayesian Analysis. MIT OpenCourseWare Sloan School of Management Course Notes. Available at: http://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/lecture-notes/MIT15_097S12_lec15.pdf

Little, R.J.A and Rubin D.B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Longstaff, F.A. and Rajan, A. (2008). An Empirical Analysis of the Pricing of Collateralised Debt Obligations. *The Journal of Finance*. Vol 63. Iss. 2. pp 529 – 563.

Longstaff, F.A. and Schwartz, E.S. (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *Journal of Finance*. Vol. 50. No. 3. pp 789 – 819.

Lybek, T. and Sarr, A. (2002). Measuring Liquidity in Financial Markets. IMF Working Paper No. 02/232. Available at: www.imf.org/external/pubs/ft/wp/2002/wp02232.pdf

Madan, D. and Unal, H. (1998). Pricing the Risks of Default. *Review of Derivatives Research*. No. 2. pp 121 – 160.

Mastro, M. (2013). *Financial Derivative and Energy Market Valuation: Theory and Implementation in Matlab*. Wiley, New York.
McElreath, R. (2015). *Statistical Rethinking: A Bayesian Course with R Examples*. Chapman & Hall/CRC.

Merton, R.C. (1974). On the Pricing of Corporate Debt: the Risk Structure of Interest Rates. *Journal of Finance*. Vol 29. Iss. 2. pp 449 – 470.

Molnar, F.J. and Hutton, B. (2008). Does Analysis using "Last Observation Carried Forward" Introduce Bias in Dementia Research? *Canadian Medical Association Journal*. Vol. 179. Iss. 8. pp 751 – 753.

Moritz, S., Sardà, A., Bartz-Beielstein, T., Zaefferer, M. and Stork, J. (2015). Comparison of Different Methods for Univariate Time Series Imputation in R. Cornell University. Available at:
https://arxiv.org/abs/1510.03924

Nashikkar,A., Subrahmanyam, M.G. and Mahanti, S. (2011). Liquidity and Arbitrage in the Market for Credit Risk. *The Journal of Financial and Quantitative Analysis*. Vol. 46. No. 3. pp 627 – 656.

Nickell, P., Perraudin, W. and Varotto, S. (2001). Ratings versus Equity-Based Credit Risk Modelling: an Empirical Analysis. Bank of England working papers 132, Bank of England. Available at:
https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2001/ratings-versus-equity-based-credit-risk-modelling-an-empirical-analysis.pdf?la=en&hash=75FB2382C8E8741B26062D84F60405359060639F

Nielson, S.S and Ronn, E.I. (1996). The Valuation of Default risk in Corporate Bonds and Interest Rate Swaps. February 13, 1996. Available at: https://ssrn.com/abstract=7247

O'Donoghue, B., Peacock, M., Lee, J. and Capriotti, L. (2014). A Spread-Return Mean-Reverting Model for Credit Spread Dynamics. *International Journal of Theoretical and Applied Finance*. Vol. 17. Iss. 3. 1450017 (14 pages).

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Sixth Edition. Springer-Verlag Berlin Heidelberg.

Pitsillis, Z. and Taylor, D.R. (2015). The IOSCO Transparency Principle and Modelling the Bid-Ask Spread: Applications in the South African Bond Market. AIFMRM Technical Report 01 - 2015. Available at: http://aifmrm.co.za/Docs/AIFMRM_report_The_IOSCO_Transparency_Principle_and_Modelling_the_Bid-Ask_Spread.pdf

Planting, S. (2017). A Spotlight on Steinhoff's Corporate Debt - Mispricing Creates Risks. Moneyweb article, 12 December 2017. Available at: https://www.moneyweb.co.za/news/companies-and-deals/a-spotlight-on-steinhoffs-corporate-debt/

Prigent, J., Olivier, R. and Olivier, S. (2001). An Empirical Investigation in Credit Spread Indices. *Journal of Risk*. Vol. 3. No 1. pp 27 – 55.

RisCura Solutions (2015). Listed Market Credit Ratings Coverage. Internal report. Personal Communication.

Reuters (2014). African Bank 'Kept Lending to Indebted Customers'. Fin24 article, 21 August 2014. Available at: https://www.fin24.com/Companies/Financial-Services/African-Bank-kept-lending-to-indebted-customers-20140821

Reuters, I-Net Bridge. (2013). African Bank Suffers as Debtors Default. Fin24 article, 20 May 2013. Available at: http://www.fin24.com/Debt/News/African-Bank-suffers-as-debtors-default-20130520

Rubin, D.B. (1976). Inference and Missing Data. *Biometrika*, 63. pp 581 – 592.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse Surveys*. Wiley. New York.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*. Vol. 91. pp 473 – 489.

Ruiz, I. and Boca, P.D. (2012). Modelling Credit Spreads for Counterparty Risk: Mean Reversion is Not Needed. *Intelligence Risk*. No. 2. Available at: https://www.mocaxintelligence.com/wp-content/uploads/2016/02/iRuiz_Spreads_Mean_Reversion.pdf

Saita, L., Drucker, S., Goldberg, L., Grenadier, S., Nagel, S., Singleton, K. and Strebulaev, I. (2006). The Puzzling Price of Corporate Default Risk. Available at: http://web-docs.stern.nyu.edu/salomon/docs/Credit2006/LeandroSaita.pdf

Sanderson, L.B., Maré, E. and De Jongh D.C.J. (2017). Banking Regulations: An Examination of the Failure of African Bank Using Merton's Structural Model. *South African Journal of Science*. Vol 113. No. 7/8. pp 15 – 21.

Schestag, R., Schuster, P. and Uhrig-Homburg, M. (2016). Measuring Liquidity in Bond Markets. *Review of Financial Studies*. Vol. 29. Iss. 5. pp 1170 – 1219.

Schafer, J.L. (1999). Multiple Imputation: A Primer. *Statistical Methods in Medical Research.* Vol. 8. pp 3 – 15.

Schönbucher, P.J. (1997). Pricing Credit Risk Derivatives. Working Paper wp-10, London School of Economics, Financial Markets Group, July 1997b. Available at:
http://www.riskmania.com/pdsdata/PricingCreditRiskDerivatives.pdf

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics.* Vol. 6. No. 2. pp 461 – 464.

Shimko, D.C., Tejima, N. and van Deventer, D.R. (1993). The Pricing of Risky Debt when Interest Rates are Stochastic. *Journal of Fixed Income.* September (3). pp 58 – 65.

Shoop, S.J.W. (2015). Should We Ban the Use of 'Last Observation Carried Forward Analysis in Epidemiological Studies?'. *SM Journal of Public Health and Epidemiology.* 1(1):1004. Available at:
https://smjournals.com/public-health-epidemiology/download.php?file=fulltext/smjphe-v1-1004.pdf

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business.* Vol 74. No. 1. pp 101 – 124.

Tang, C.Y. and Chen, S.X. (2009). Parameter Estimation and Bias Correction for Diffusion Processes. *Journal of Econometrics.* Vol. 149. pp 65 – 81.
Thompson, W. (2018). Steinhoff's Early Redemption Signals Start of Divorce. Moneyweb article, 29 January 2018. Available at: https://www.moneyweb.co.za/in-depth/investigations/steinhoffs-early-redemption-signals-the-start-of-a-big-divorce/

Trück, S., Laub, M. and Rachev, S. (2004). The Term Structure of Credit Spreads and Credit Default Swaps - An Empirical Investigation. *Investment Management and Financial Innovations.* Vol. 1. Iss. 3. pp 8 – 30.

Tsay, R.S. (2002). *Analysis of Financial Time Series: Financial Econometrics.* Wiley.

Tsuji, C. (2005). The Credit-Spread Puzzle. *Journal of International Money and Finance.* Vol. 24. Iss. 7. pp 1073 – 1089.

Uhlenbeck, G.E. and Ornstein, L.S. (1930). On the Theory of the Brownian Motion. *Phys. Rev.* Vol. 36. Iss. 5. pp 823 – 841.

Walton, M.K. (2009). Addressing and Advancing the Problem of Missing Data. *Journal of Biopharmaceutical Statistics.* Vol 19. Iss. 6. pp 945 – 956.

White, I.R, Royston, P. and Wood, A.M. (2011). Multiple Imputation using Chained Equations: Issues and Guidance for Practice. *Statistics in Medicine.* Vol 30. Iss. 4. pp 377 – 399.

Yu, J. and Phillips, P.C. (2001). A Gaussian Approach for Estimating Continuous Time Models of Short Term Interest Rates. *The Econometrics Journal*. Vol. 4. pp 211 – 225.

Zhang, B.Y., Zhou, H. and Zhu, H. (2009). Explaining Credit Default Swap Spreads with the Equity Volatility and Jump Risks of Individual Firms. *The Review of Financial Studies*. Vol. 22. Iss. 12. pp 5099 – 5131.

Zhou, P. (2007). Forecasting Bankruptcy and Physical Default Intensity. Discussion paper, 614. Financial Markets Group, London School of Economics and Political Science, London, UK. Available at: http://eprints.lse.ac.uk/24434/

A. Mathematical Preliminaries

A.1 Stochastic Calculus

A.1.1 Probability Spaces and Filtrations

The probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$ is comprised of the set of all possible outcomes Ω, called the sample space, an increasing set of σ-algebras on Ω, or a filtration $\mathcal{F}$ with $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for $0 \leq s \leq t \leq T$, and a probability function $\mathbb{P}$ which maps an outcome in the sample space to the real number line (i.e. an event is assigned a real number between 0 and 1 by $\mathbb{P}$).

A probability space is said to be $\mathbb{P}$-complete if for each $B \subset A \in \mathcal{F}$ such that $P(A) = 0$, $B \in \mathcal{F}$.

A filtration satisfying the usual conditions has the following properties:

1. $\mathcal{F}$ is $\mathbb{P}$-complete.
2. $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets of Ω.
3. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_t$ (right-continuous).

If a measurable space or σ-field (Y, Σ), where Σ is a σ-algebra on the set Y, is defined, then a random variable X is a measurable function that maps the probability space to the state space, or $X : \Omega \to Y$.

A.1.2 Stochastic Processes and Brownian Motions

A stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of random variables $\{X_t\}$ indexed by a set τ, which for most purposes is the time horizon given by $[t_0, T]$. A continuous stochastic process has continuous indexation (commonly the continuous time scenario), whereas a discrete stochastic process is indexed by a finite set.

$\{X_t : t \in \tau\}$ is said to be adapted to the filtration $\mathcal{F}_t, t \in \tau$ if each X_t is $\mathcal{F}_t$-measurable. $\{X_t : t \in \tau\}$ is also said to be predictable if it $\mathcal{F}_t$-measurable as a map from $[t_0, T]$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (Hunt *et al.* (2004)).

Brownian motions (also known as Wiener processes) are a special class of stochastic processes. A continuous time Brownian motion $W(t)$ has the following defining characteristics:

1. $W(0) = 0$.
2. $W(t)$ is almost surely continuous.
3. $W(t) \sim N(0, t)$.
4. $W(s)$ and $W(t) - W(s)$ are independent random variables $\forall\ 0 < s < t$, with $(W(t) - W(s)) \sim N(0, t - s)$.

The Levy characterisation of Brownian motions states that they are almost surely continuous martingales with quadratic variation $[W(t)] = t$ and $W(0) = 0$.

A vector of n independent Brownian motions is a multidimensional or n-dimensional Brownian motion.

A.1.3 Markov Processes

A stochastic process $\{X_t\}$, adapted to the filtration $\mathcal{F}_t$, is said to be Markov with respect to this filtration if and only if

1. $\forall t \geq 0$ and for every bounded random variable $M \in \{X_u : u \geq t\}$,

$$\mathbb{E}[M|\mathcal{F}_t] = \mathbb{E}[M|X_t],$$

   is satisfied.

2. $\forall t, s \geq 0$ and for every bounded Borel function g, the following holds

$$\mathbb{E}[g(X_{t+u})|\mathcal{F}_t] = \mathbb{E}[g(X_{t+u})|X_t].$$

The Markov property is in essence the dependence of the future values of X on its current value, and not its previous values (Øksendal (2003)). Thus a Markov chain is a sequence of random variables for which at any state, the distribution of the current random variable or state, given all the previous states, is only dependent on the most recent state (Gelman *et al.* (2004)).

A stationary Markov chain is reversible (Kelly (1979)) if and only if there exists a collection of positive numbers $\gamma(j)\mathbb{P}(X_{t+1} = j|X_t = k)$ for any j, summing to one, that satisfy the detailed balance conditions

$$\gamma(j)\mathbb{P}(X_{t+1} = j|X_t = k) = \gamma(k)\mathbb{P}(X_{t+1} = k|X_t = j).$$

where stationary means that the joint probability distribution does not change when shifted in time.

A.1.4 Martingales

The stochastic process $X = \{X_t : t \in \tau\}$ is said to be a martingale with respect to $\mathcal{F}_t$ if and only if

1. $\mathbb{E}[|X|] < \infty$.

2. $\mathbb{E}[X_t|F_s] = X_s$ almost surely for $0 < s \leq t < \infty$.

If the second condition is changed to $\mathbb{E}[X_t|F_s] \leq X_s$ almost surely for $0 < s < t < \infty$, then X is called a supermartingale, whilst if $\mathbb{E}[X_t|F_s] \geq X_s$ almost surely for $0 < s < t < \infty$, then X is called a submartingale.

A.1.5 Change of Measure

A.1.5.1 Numéraires and Equivalent Martingale Measures

A numéraire is defined as a traded asset with a strictly positive price process.

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assets $S^0(t), ..., S^N(t)$, where $S^0(t)$ represents the risk-free assets and $S^1(t), ..., S^N(t)$ are the risky assets.

Suppose that $N(t)$ is a numéraire. A measure $\mathbb{Q}$ on the probability space is an Equivalent Martingale Measure (EMM) for $N(t)$ if and only if

1. $\mathbb{Q} \sim \mathbb{P}$.
2. $\bar{S}(t) = \frac{S(t)}{N(t)}$ is a $\mathbb{Q}$-martingale.

The equivalent martingale measure associated with the risk-free asset is called the risk-neutral measure $A(t)$.

An EMM on the market $\mathcal{M}$ is a probability measure on $(\Omega, \mathcal{F})$, equivalent to $\mathbb{P}$, such that all assets in the market are martingales under this measure.

If an EMM for a market $\mathcal{M}$ exists, then $\mathcal{M}$ is arbitrage-free.

Girsanov's Theorem presents a mechanism that facilitates the change from one probability space to another, using the Radon-Nikodym derivative (Hunt *et al.* (2004)).

A.1.5.2 Radon-Nikodym Derivative

Let μ and ν be σ-finite measures on some probability space $(\Omega, \mathcal{F})$. ν is said to be absolutely continuous with respect to μ if $\nu(A) = 0 \Rightarrow \mu(A) = 0$ on the set A and is written $\nu << \mu$. If $\nu << \mu$, then there exists a non-negative finite measurable function f such that

$$\nu(A) = \int_A f d\mu,$$

for any $A \in \mathcal{F}$.

The Radon-Nikodym derivative (or likelihood ratio) of ν with respect to μ is written as $f = \frac{d\nu}{d\mu}$. f is unique μ-a.e. If there exists a measurable function g such that $\nu(A) = \int_A g d\mu$ then $f = g$ μ-a.e.

Furthermore, if $\nu << \mu$ and $\mu << \nu$ (denoted $\nu \sim \mu$) then

$$\frac{d\nu}{d\mu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

A.1.5.3 Girsanov's Theorem

Let

$$dY(t, \omega) = \mu(t, \omega)dt + \sigma(t, \omega)dW(t)$$

be an m-dimensional Itô process in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, whose coefficients are dependent on time $t \in [0, T]$ and the state of the world ω, and where $W(t)$ is the associated m-dimensional $\mathbb{P}$-Brownian Motion. $\mu(t, \omega)$ is a q-dimensional drift vector and $\sigma(t, \omega)$ is the $(q \times m)$-dimensional volatility matrix.

Suppose that there exists m-dimensional vector process $\lambda(t, \omega)$, given a q-dimensional vector process $\beta(t, \omega)$,

such that

$$\sigma\lambda = \beta - \mu.$$

Let

$$\xi(t) = \exp\left(-\frac{1}{2}\int_{t_0}^{t} \|\lambda(u)\|^2 \, du + \int_{t_0}^{t} \lambda(u) dW(u)\right).$$

If $\mathbb{E}[\xi(T) = 1]$, then $\{\xi(t), 0 \leq t \leq T\}$ is a martingale and the measure $\mathbb{Q}$ defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \xi(T)$$

is equivalent to $\mathbb{P}$.

Then the dynamics under $\mathbb{Q}$ are

$$dY(t,\omega) = \beta(t,\omega)dt + \sigma(t,\omega)d\bar{W}(t)$$

where

$$\bar{W}(t) = W(t) - \int_{t_0}^{t} \lambda(u,\omega)du \tag{A.1}$$

is a $\mathbb{Q}$-Brownian Motion.

A.1.6 Stochastic Differential Equations

Many financial models are defined by differential equations with an added noise term, called stochastic differential equations, which take on the following form:

$$dS(t) = a(t, S(t))dt + b(t, S(t))dW(t), \tag{A.2}$$

where $S(0) = S_0$ and $W(t)$ is a Brownian motion.

The existence and uniqueness of these solutions is given by the following proposition (Glasserman (2004)).

Proposition: If $\mathbb{E}[\|S_0\|^2]$ is finite and there exists a constant K such that

1. $\|a(x,t) - a(y,t)\| + \|b(x,t) - b(y,t)\| \leq K\|x - y\|$ (Lipschitz condition),
2. $\|a(x,t)\| + \|b(x,t)\| \leq K(1 + \|x\|)$ (Linear growth condition),

for all $t \in [0,T]$ and all $x, y \in \mathbb{R}^k$, then $S(t)$ is a strong solution to (A.2). For all $t \in [0,T]$, $\mathbb{E}[\|S(t)\|^2] < \infty$ and if $S^*(t)$ is also a solution, $\mathbb{P}(S(t) = S^*(t)) = 1$ (Hunt *et al.* (2004)).

A.1.7 Itô Integrals and Processes

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathcal{F}_t, t \geq 0$ and a $d \times n$-dimensional $\mathcal{F}_t$-adapted process $b(t) : t \in [0,T]$, with

$$\mathbb{P}\left(\int_0^T \|b(u)\|^2 \, du < \infty\right) = 1, \tag{A.3}$$

where $\|\cdot\|$ denotes the square root of the sum of the squared components of b.

Assuming an n-dimensional Brownian motion $W = (W_1, ..., W_n)^T$ is defined on this probability space, then the Itô integral with respect to this Brownian Motion is well-defined and given by

$$\int_0^t b(s)^T dW(s)$$

where each row of the resulting d-dimensional vector is a stochastic integral.

The d-dimensional stochastic process $\{S(t), 0 \leq t \leq T\}$ is an Itô process if

$$S(t) = S(0) + \int_t^{t_0} a(s)ds + \int_t^{t_0} b(s)dW(s) \tag{A.4}$$

where $S(0)$ is measurable with respect to $\mathcal{F}_0$ and a is a d-dimensional adapted process satisfying

$$\mathbb{P}\left(\int_0^T |a_i(u)|du < \infty\right) = 1, \quad i = 1, ..., d$$

and b satisfies (A.3) (Glasserman (2004)).

If $S(t)$ is an Itô process as defined in (A.4) and the function $Y(t, S(t))$ is sufficiently smooth, then $X(t) = Y(t, S(t))$ is an Itô process. Itô's formula applied to $X(t)$ gives

$$\begin{aligned}
dX(t) &= \frac{\partial Y(t,S(t))}{\partial t}dt + \sum_{i=1}^n \frac{\partial Y(t,S(t))}{\partial s_i} dS_i(t) + \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 Y(t,S(t))}{\partial s_i \partial s_j} dS_i(t)dS_j(t) \\
&= \frac{\partial Y(t,S(t))}{\partial t}dt + \sum_{i=1}^n \frac{\partial Y(t,S(t))}{\partial s_i}(a_i(t)dt + b_i(t)dW(t)) \\
&+ \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 Y(t,S(t))}{\partial s_i \partial s_j} b_i(t)b_j(t)^T dt \\
&= \left(\frac{\partial Y(t,S(t))}{\partial t} + \sum_{i=1}^n \frac{\partial Y(t,S(t))}{\partial s_i} a_i(t) + \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 Y(t,S(t))}{\partial s_i \partial s_j} b_i(t)b_j(t)^T\right) \\
&+ \sum_{i=1}^n \frac{\partial Y(t,S(t))}{\partial s_i} b_i(t)dW(t).
\end{aligned} \tag{A.5}$$

If the processes S, a and b are scalar, then Itô's Identity becomes

$$dY(t,S(t)) = \left(\frac{\partial Y(t,S(t))}{\partial t}a(t) + \frac{1}{2}\frac{\partial^2 Y(t,S(t))}{\partial S^2}b(t)^2\right)dt + \frac{\partial Y(t,S(t))}{\partial S}b(t)dW(t).$$

A.1.7.1 Itô Isometry

Theorem A.1. *Let $0 = t_1 < t_2 < ... < t_n = T$. Consider the $\mathcal{F}_t$ adapted stochastic process $Y_t : 0 < t < T$ which is bounded and elementary and the Wiener process $W_t : 0 < t < T$ on $\mathcal{F}_t$.*

Then Itô Isometry states that

$$\mathbb{E}\left[\left(\int_0^T Y_t dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T Y_t^2 dt\right].$$

Proof. (taken from Øksendal (2003)) By definition,

$$\int_0^T Y_t dW_t = \sum_{i>0} Y_{t_i}[W_{t_i} - W_{t_{i-1}}],$$

where $i = 1, ..., n$.

Let $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ and $Y_i = Y_{t_i}$. Then

$$\mathbb{E}[Y_i Y_j \Delta W_i \Delta W_j] = \begin{cases} \mathbb{E}[Y_i^2][t_i - t_{i-1}], & i = j, \\ 0, & i \neq j. \end{cases}$$

Since Wiener processes have independent increments, ΔW_i is independent of ΔW_j for $i > j$. $\Delta W_i = W_{t_i} - W_{t_{i-1}}$ is also independent of $\mathcal{F}_t$ and so independent of the $\mathcal{F}_t$ adapted Y as well.

Thus

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T Y_t dW_t\right)^2\right] &= \sum_{i,j} \mathbb{E}[Y_i Y_j \Delta W_i \Delta W_j] \\ &= \sum_i \mathbb{E}\left[[Y_i^2][t_i - t_{i-1}]\right] \\ &= \mathbb{E}\left[\int_0^T Y_t^2 dt\right]. \end{aligned}$$

□

A.1.7.2 Extension of Itô Isometry

Itô Isometry can also be developed for stochastic integrals over multiple Wiener processes.

Theorem A.2. *Let $0 = t_1 < t_2 < ... < t_n = T$. Consider the $\mathcal{F}_t$ adapted stochastic processes $Y_t : 0 < t < T$ and $U_t : 0 < t < T$ which are bounded and elementary and the Wiener processes $W_t^Y : 0 < t < T$ and $W_t^U : 0 < t < T$ on $\mathcal{F}_t$.*

Then

$$\mathbb{E}\left[\left(\int_0^T Y_t dW_t^Y\right)\left(\int_0^T U_t dW_t^U\right)\right] = \rho_{YU}\mathbb{E}\left[\int_0^T Y_t U_t dt\right],$$

where ρ_{YU} is the correlation of the Wiener process over the same time interval and is assumed to be zero for disjoint intervals. i.e. ΔW_i^Y is independent of ΔW_j^U for $i \neq j$.

Proof. The proof follows closely that of Itô Isometry. Again, by definition,

$$\int_0^T Y_t dW_t^Y = \sum_{i>0} Y_{t_i}[W_{t_i}^Y - W_{t_{i-1}}^Y]$$

and

$$\int_0^T U_t dW_t^U = \sum_{j>0} U_{t_j}[W_{t_j}^U - W_{t_{j-1}}^U],$$

where $i, j = 1, ..., n$.

Letting $\Delta W_i^Y = W_{t_i}^Y - W_{t_{i-1}}^Y$, $\Delta W_j^U = W_{t_j}^U - W_{t_{j-1}}^U$, $Y_i = Y_{t_i}$ and $U_j = U_{t_j}$ gives

$$\mathbb{E}[Y_i U_j \Delta W_i^Y \Delta W_j^U] = \begin{cases} \rho_{YU} \mathbb{E}[Y_i U_i][t_i - t_{i-1}], & i = j, \\ 0, & i \neq j. \end{cases}$$

Again ΔW_i^Y is independent of ΔW_j^Y for $i > j$ and ΔW_i^U is independent of ΔW_j^U for $i > j$. ΔW_i^Y and ΔW_i^U are independent of Y and U as well.

Thus

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T Y_t dW_t^Y\right)\left(\int_0^T U_t dW_t^U\right)\right] &= \sum_{i,j} \mathbb{E}[Y_i U_j \Delta W_i^Y \Delta W_j^U] \\ &= \sum_i \mathbb{E}\left[[Y_i U_i][t_i - t_{i-1}]\right] \\ &= \mathbb{E}\left[\int_0^T Y_t U_t dt\right]. \end{aligned}$$

□

Theorems $A.1$ and $A.2$ are shown in Øksendal (2003) to extend from elementary functions to the family of functions defined by the Itô integral.

A.2 Statistical Models, Tests and Distributions

A.2.1 Definition of the Normal Distribution

A normal distribution on $\mathbb{R}$ of a random variable X with $\mathbb{E}[X] = \mu$, called the mean or expected value, and $\mathrm{Var}[X] = \sigma^2$, called the variance, is a statistical distribution with probability density function

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty. \tag{A.6}$$

The normal distribution of X, with mean μ and variance σ^2 is represented by the concise notation $X \sim N(\mu, \sigma^2)$.

The cumulative distribution function of a normally distributed variable is obtained by integrating (A.6) to give

$$\mathbb{P}(X \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) dy. \tag{A.7}$$

A standard normally distributed random variable has zero mean and a variance of 1. Any normally distributed variable can be transformed to a standard normal distribution by a change of variables.

Suppose $X \sim N(\mu, \sigma^2)$. A standard normal variable K is generated by letting $K = \frac{X-\mu}{\sigma}$. The proba-

bility density function of a standard normal distribution reduces from (A.6) to

$$f(k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2}, \quad -\infty < k < \infty. \tag{A.8}$$

Note that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k^2} = 1. \tag{A.9}$$

The moment generating function is by definition $\mathbb{E}[e^{tX}]$. By completing the square and using (A.9), the moment generating function for the normal distribution is determined to be

$$M_X(t) = \mathbb{E}[e^{tX}] = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right). \tag{A.10}$$

A.2.2 Likelihood of the Univariate Normal Distribution

Given (A.6) and that the n elements in the distribution are independent and identically distributed, the likelihood of the normal distribution is simply the product of the individual conditional densities, or

$$\mathcal{L}(x|\mu,\sigma) = \prod_{i=1}^{n} f(x_i|\mu,\sigma).$$

It is most often easier to work with natural logarithms when dealing with likelihoods, thus the likelihood can be written in terms of its natural logarithm, where the product of the conditionals now becomes a summation of the log conditionals as follows:

$$\begin{aligned} \ln \mathcal{L}(x|\mu,\sigma) &= \sum_{i=1}^{n} \ln f(x_i|\mu,\sigma) \\ &= \sum_{i=1}^{n} \left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i-\mu)^2\right) \\ \Rightarrow \ln \mathcal{L}(x|\mu,\sigma) &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \end{aligned} \tag{A.11}$$

A.2.3 Definition of the Lognormal Distribution

A lognormal distribution on $\mathbb{R}$ of the random variate Y is defined in terms of the normal distribution of the natural logarithm of Y. Suppose $X = \ln Y \sim N(\mu, \sigma^2)$, then the probability density function of the lognormal distribution is given by

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\ln y - \mu)^2}{2\sigma^2}\right), \quad 0 < y < \infty.$$

The mean of Y is calculated by standardising $\ln Y$ to $K = \frac{\ln Y - \mu}{\sigma} \sim N(0,1)$, and using the identity in (A.10). Thus

$$\mathbb{E}[Y] = e^{\mu + \frac{1}{2}\sigma^2}. \tag{A.12}$$

The variance of Y is determined by plugging (A.12) into the definition of variance

$$\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$

to get

$$\text{Var}[Y] = e^{2\mu+\sigma^2}(e^{2\sigma^2} - 1). \tag{A.13}$$

The log-likelihood of a lognormally distributed random variable, $X = \ln Y \sim N(\mu, \sigma^2)$, is given by

$$\ln \mathcal{L} = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\left[\ln y_i + \frac{(\ln y_i - \mu)^2}{2\sigma^2}\right]. \tag{A.14}$$

A.2.4 Inverse Gamma Distribution

The probability density function for a random variable x that follows the inverse gamma distribution is

$$f(x|k,\beta) = \frac{\beta^k}{\Gamma(k)x^{k+1}}\exp\left(-\frac{\beta}{x}\right) \tag{A.15}$$

where k is the shape parameter, β is the scale parameter and $\Gamma(c)$ is the gamma function evaluated at c.

A.2.5 Geometric Brownian Motion (GBM)

A geometric Brownian motion (GBM) is a continuous stochastic process $s(t)$ where $\ln s(t)$ is driven by a Brownian motion $W(t)$. The GBM $s(t)$ satisfies the SDE

$$ds(t) = \mu(t)s(t)dt + \sigma(t)s(t)dW(t) \tag{A.16}$$

where $\mu(t)$ is a deterministic drift coefficient, $\sigma(t)$ is a deterministic volatility function and $W(t)$ is a standard Wiener process under the real world measure $\mathbb{P}$ with $W(t) \sim N(0,t)$. Volatility can never be negative, hence $\sigma_s > 0$.

Letting $y(t) = \ln s(t)$ and applying Itô's lemma:

$$\begin{aligned} dy(t) &= \left[\frac{\partial y}{\partial t} + \frac{1}{2}\frac{\partial^2 y}{\partial s^2}\sigma(t)^2 s(t)^2\right]dt + \frac{\partial y}{\partial s}ds(t) \\ &= \left[0 - \frac{1}{2s(t)^2}\sigma(t)^2 s(t)^2\right]dt + \frac{1}{s(t)}(\mu(t)s(t)dt + \sigma(t)s(t)dW_t) \\ \Rightarrow dy(t) &= \left[\mu(t) - \frac{1}{2}\sigma(t)^2\right]dt + \sigma(t)dW(t). \end{aligned} \tag{A.17}$$

Integrating, given an initial value $S(0)$, yields

$$y(t) - y(0) = \int_{t_0}^{t}\left[\mu(u) - \frac{1}{2}\sigma(u)^2\right]du + \int_{t_0}^{t}\sigma(u)dW(u)$$

which is equivalent to

$$\ln s(t) = \ln s(0) + \int_{t_0}^{t}\left[\mu(u) - \frac{1}{2}\sigma(u)^2\right]du + \int_{t_0}^{t}\sigma(u)dW(u)$$

with solution

$$s(t) = s(0)\exp\left(\int_{t_0}^{t}\left[\mu(u) - \frac{1}{2}\sigma(u)^2\right]du + \int_{t_0}^{t}\sigma(u)dW(u)\right).$$

Finally it is observed that

$$\ln s(t) \sim N(\ln s(0) + \int_{t_0}^{t} \left[\mu(u) - \frac{1}{2}\sigma(u)^2\right] du, \int_{t_0}^{t} |\sigma(u)|^2 du).$$

Assuming that the drift and volatility coefficients are constant i.e. $\mu(t) = \mu$ and $\sigma(t) = \sigma$, and $t_0 = 0$, the solution reduces to

$$s(t) = s(0) \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\right). \tag{A.18}$$

$s(t)$ is lognormally distributed with mean $s(0)e^{\mu t}$ and variance $(s(0)e^{\mu t})^2(e^{\sigma^2 t} - 1)$. The lognormality of $s(t)$ implies the normality of $\ln s(t)$, where

$$\ln s(t) \sim N\left(\ln s(0) + (\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t\right).$$

A.2.5.1 Deriving the GBM Likelihood Function

As in the previous section, μ_s is denoted by μ and σ_s as σ in the following.

Under GBM, the stock price is lognormally distributed and $\ln s(t)$ is normally distributed (refer to Appendix A.2.5). Let $S = (S_1, S_2, ..., S_n)$ denote the discretised stock price process. Then conditional density of $\ln S_i$ is

$$f(\ln S_i | \ln S_{i-1}; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left[\frac{-(\ln S_i - \ln S_{i-1} - (\mu - \frac{1}{2}\sigma^2)\Delta t)^2}{2\sigma^2\Delta t}\right], \quad i = 2, ..., n$$

and the the natural logarithm of the conditional density is

$$\ln f(\ln S_i | \ln S_{i-1}; \mu, \sigma) = -\frac{1}{2}\ln(2\pi\sigma^2\Delta t) - \frac{1}{2\sigma^2\Delta t}\left(\ln S_i - \ln S_{i-1} - (\mu - \frac{1}{2}\sigma^2)\Delta t\right)^2, \quad i = 2, ..., n.$$

By taking the product of the individual conditional densities yields the likelihood, whilst the sum of the individual log conditional densities over $i = 2, ..., n$ gives the log likelihood:

$$\begin{aligned}
\mathcal{L}(\ln S|\mu, \sigma) &= \prod_{i=2}^{n} f(\ln S_i | \ln S_{i-1}; \mu, \sigma) \\
\ln \mathcal{L}(\ln S|\mu, \sigma) &= \sum_{i=2}^{n} \ln f(\ln S_i | \ln S_{i-1}; \mu, \sigma) \\
&= -\frac{n-1}{2}\ln(2\pi\sigma^2\Delta t) - \frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - (\mu - \frac{1}{2}\sigma^2)\Delta t\right)^2.
\end{aligned} \tag{A.19}$$

A.2.5.2 Calculating the MLEs of the GBM Model

The MLE optimisation problem for GBM can be written in terms of η and σ:

$$\max_{\sigma>0} \ln \mathcal{L}(\ln S|\eta, \sigma) = \max_{\sigma>0}\left(-\frac{n-1}{2}\ln(2\pi\sigma^2\Delta t) - \frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}(\ln S_i - \ln S_{i-1} - \eta\Delta t)^2\right) \tag{A.20}$$

since (3.3) holds. Numerical optimisation succeeds in determining $\hat{\eta}$ and $\hat{\sigma}$, but analytical solutions are also tractable via

$$\frac{\partial \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \eta}\Big|_{\hat{\eta},\hat{\sigma}} = \left[-\frac{1}{\sigma^2\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)(-\Delta t)\right]\Big|_{\hat{\eta},\hat{\sigma}} = 0$$
$$\frac{\partial \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \sigma}\Big|_{\hat{\eta},\hat{\sigma}} = \left[-\frac{n-1}{\sigma} + \frac{1}{\sigma^3\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)^2\right]\Big|_{\hat{\eta},\hat{\sigma}} = 0.$$

Hence $\hat{\eta}$ and $\hat{\sigma}$ are equal to

$$\hat{\eta} = \frac{1}{n-1}\sum_{i=2}^{n}(\ln S_i - \ln S_{i-1}) \text{ and} \tag{A.21}$$
$$\hat{\sigma}(\hat{\eta}) = \sqrt{\frac{1}{(n-1)\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \hat{\eta}\Delta t\right)^2} \tag{A.22}$$

respectively. $\hat{\mu}$ can be retrieved using

$$\hat{\mu} = \hat{\eta} + \frac{1}{2}\hat{\sigma}_x^2.$$

A.2.5.3 Determining the Standard Errors of the GBM Model

To obtain standard errors for the maximum likelihood estimates, the partial second derivatives of the log likelihood function with respect to η and σ as follows:

$$\frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \eta^2}\Big|_{\hat{\eta},\hat{\sigma}} = \frac{\partial}{\partial \eta}\left[-\frac{1}{\sigma^2\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)(-\Delta t)\right]\Big|_{\hat{\eta},\hat{\sigma}} = -\frac{n-1}{\hat{\sigma}^2}$$
$$\frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \sigma^2}\Big|_{\hat{\eta},\hat{\sigma}} = \frac{\partial}{\partial \sigma}\left[-\frac{n-1}{\sigma} + \frac{1}{\sigma^3\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)^2\right]\Big|_{\hat{\eta},\hat{\sigma}} = -\frac{2(n-1)}{\hat{\sigma}^2}.$$

As done in Section 3.3.1.3 for OUVLTM,

$$(n-1)\Delta t\hat{\sigma}^2 = \sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)^2$$

was substituted into the second partial derivative for σ.

The resulting standard errors are thus

$$\hat{\eta}^{SE} = \sqrt{\frac{\hat{\sigma}^2}{n-1}} \tag{A.23}$$
$$\hat{\sigma}^{SE} = \sqrt{\frac{\hat{\sigma}^2}{2(n-1)}}. \tag{A.24}$$

To get the standard error of $\hat{\mu}_x$, it has been established that the following is true:

$$\frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \mu^2}\Big|_{\hat{\eta},\hat{\sigma}} = \left[\frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \eta^2}\frac{\partial \eta}{\partial \mu} + \frac{\partial \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \eta}\frac{\partial^2 \eta}{\partial \eta \partial \mu}\right]\frac{\partial \eta}{\partial \mu}\Big|_{\hat{\eta},\hat{\sigma}}. \tag{A.25}$$

Since

$$\frac{\partial \eta}{\partial \mu} = 1 \Rightarrow \frac{\partial}{\partial \eta}\left(\frac{\partial \eta}{\partial \mu}\right) = 0,$$

(A.25) is now

$$\frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \mu^2}\Big|_{\hat{\eta},\hat{\sigma}} = \frac{\partial^2 \ln \mathcal{L}(\ln S|\eta,\sigma)}{\partial \eta^2}\Big|_{\hat{\eta},\hat{\sigma}} = -\frac{n-1}{\hat{\sigma}^2}.$$

Hence

$$\hat{\mu}^{SE} = \hat{\eta}^{SE} = \sqrt{\frac{\hat{\sigma}^2}{n-1}}.$$

A.2.5.4 Deriving the Conditional Posterior Distributions of the GBM Parameters

The presence of η in (3.21), which is used on Theorem 3.1 to obtain the conditional distributions of the OUVLTM parameters, requires the same Gibbs sampling of the GBM parameters for the stock price. Theorem A.3 derives the conditional posterior distributions for σ_s and $\eta = \mu_s - \frac{1}{2}\sigma_s^2$ (originally proved in continuous time by Jones (1998)).

Theorem A.3. *Assuming the following conjugate priors for η and σ_s^2*

$$p(\eta) \sim N(\mu_{\eta 0}, \tau_{\eta 0}^2) \quad \textit{and}$$
$$p(\sigma_s^2) \sim \mathcal{IG}\left(\nu_0, \psi_0\right).$$

Then the full conditional distributions for η and σ_s^2 are given by

$$p(\eta|\ln S, \sigma_s) \sim N(\mu_\eta, \tau_\eta^2) \quad \textit{and} \tag{A.26}$$
$$p(\sigma_s^2|\ln S, \eta) \sim \mathcal{IG}\left(\nu_n, \psi_n\right). \tag{A.27}$$

where

$$\mu_\eta = \frac{\frac{\mu_{\eta 0}}{\tau_{\eta 0}^2} + \frac{1}{\sigma_s^2}\sum_{i=2}^{n}(\ln S_i - \ln S_{i-1})}{\frac{1}{\tau_{\eta 0}^2} + \frac{n-1}{\sigma_s^2}\Delta t}, \qquad \tau_\eta^2 = \frac{1}{\frac{1}{\tau_{\eta 0}^2} + \frac{n-1}{\sigma_s^2}},$$
$$\nu_n = \frac{n-1}{2} + \nu_0 \qquad \textit{and} \qquad \psi_n = \frac{1}{2}\sum_{i=2}^{n}(\ln S_i - \ln S_{i-1} - \eta\Delta t)^2 + \psi_0.$$

Proof. Assume first that σ_s^2 is known and η is unknown. The prior of η is of the form

$$p(\eta) \propto \exp\left(-\frac{1}{2\tau_{\eta 0}^2}(\eta - \mu_{\eta 0})\right).$$

The exponent of (A.19) in Appendix A can be written in terms of η as

$$\mathcal{L}(S|\eta,\sigma) = \frac{1}{\sqrt{2\pi\sigma_s^2\Delta t}}\exp\left[-\frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)^2\right].$$

Thus

$$p(\eta|\ln S, \sigma_s^2) \propto \exp\left(-\frac{1}{2\tau_{\eta 0}^2}(\eta - \mu_{\eta 0})\right)\exp\left(-\frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}\left(\ln S_i - \ln S_{i-1} - \eta\Delta t\right)^2\right).$$

The terms in the exponent without the $-\frac{1}{2}$ add together as follows

$$\begin{aligned}
&\frac{1}{\tau_{\eta 0}^2}(\eta-\mu_{\eta 0})^2+\frac{1}{\sigma_s^2\Delta t}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1}-\eta\Delta t)^2\\
&=\frac{1}{\tau_{\eta 0}^2}\left(\eta^2-2\eta\mu_{\eta 0}+\mu_{\eta 0}^2\right)+\frac{1}{\sigma_s^2\Delta t}\left(\left(\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})\right)^2-2\eta\Delta t\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})+(n-1)\eta^2\Delta t^2\right)\\
&=\eta^2\left(\frac{(n-1)}{\sigma_s^2}\Delta t+\frac{1}{\tau_{\eta 0}^2}\right)-2\eta\left(\frac{1}{\sigma_s^2}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})+\frac{\mu_{\eta 0}}{\tau_{\eta 0}^2}\right)+\frac{1}{\sigma_s^2\Delta t}\left(\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})\right)^2+\frac{\mu_{\eta 0}^2}{\tau_{\eta 0}^2}\\
&=a_\eta\eta^2-2b_\eta\eta+c_\eta
\end{aligned}$$

where

$$a_\eta=\frac{(n-1)}{\sigma_s^2}\Delta t+\frac{1}{\tau_{\eta 0}^2},\quad b_\eta=\frac{1}{\sigma_s^2}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})+\frac{\mu_{\eta 0}}{\tau_{\eta 0}^2}\quad\text{and}\quad c_\eta=\frac{1}{\sigma_s^2\Delta t}\left(\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})\right)^2+\frac{\mu_{\eta 0}^2}{\tau_{\eta 0}^2}.$$

So the theory for η is proved:

$$p(\eta|\ln S,\sigma_s^2)\propto\exp\left[-\frac{(\eta-\frac{b_\eta}{a_\eta})^2}{2/a_\eta}\right]$$

which is normally distributed with mean

$$\mu_\eta=\frac{b_\eta}{a_\eta}=\frac{\frac{1}{\sigma_s^2}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1})+\frac{\mu_{\eta 0}}{\tau_{\eta 0}^2}}{\frac{(n-1)}{\sigma_s^2}\Delta t+\frac{1}{\tau_{\eta 0}^2}}$$

and variance

$$\tau_\eta^2=\frac{1}{a_\eta}=\frac{1}{\frac{(n-1)}{\sigma_s^2}\Delta t+\frac{1}{\tau_{\eta 0}^2}}.$$

Now assume η is known and σ_s is unknown. The assumption of an inverse gamma prior gives

$$p(\sigma_s^2)\propto\frac{1}{\sigma_s^{2(\nu_0+1)}}\exp\left[-\frac{\psi_0}{\sigma_s^2}\right].$$

Again, the following holds for the likelihood of $\ln S$ under GBM:

$$\mathcal{L}(S|\eta,\sigma)\propto\frac{1}{\sigma_s^{2\left(\frac{n-1}{2}\right)}}\exp\left[-\frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1}-\eta\Delta t)^2\right].$$

Hence the product of the prior and the likelihood functions give the conditional posterior of σ_s^2

$$\begin{aligned}
p(\sigma_s^2|\ln S,\eta)&\propto\frac{1}{\sigma_s^{2(\nu_0+1)}}\exp\left[-\frac{\psi_0}{\sigma_s^2}\right]\frac{1}{\sigma_s^{2\left(\frac{n-1}{2}\right)}}\exp\left[-\frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1}-\eta\Delta t)^2\right]\\
&=\frac{1}{\sigma_s^{2\left(\frac{n-1}{2}+\nu_0+1\right)}}\exp\left[-\frac{1}{2\sigma^2\Delta t}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1}-\eta\Delta t)^2+\psi_0\right].
\end{aligned}\tag{A.28}$$

(A.28) follows is an inverse-gamma distribution with shape and scale parameters as follows:

$$\nu_n=\frac{n-1}{2}+\nu_0\qquad\text{and}\qquad\psi_n=\frac{1}{2}\sum_{i=2}^{n}(\ln S_i-\ln S_{i-1}-\eta\Delta t)^2+\psi_0.$$

□

A.2.6 The Ornstein-Uhlenbeck Model

The Ornstein Uhlenbeck (OU) model describes a stochastic process introduced by Uhlenbeck *et al.* (1930) and is widely used to model interest rates. The process is stationary, Markovian and Gaussian and satisfies the following stochastic differential equation

$$dx(t) = \theta(\mu - x(t))dt + \sigma dW^{\mathbb{P}}(t), \tag{A.29}$$

where θ is the constant rate of mean reversion, μ is the constant long-term mean, σ is the constant instantaneous volatility and $W(t)$ is a standard Wiener process with respect to the real world measure $\mathbb{P}$. For the process to remain stationary $\theta > 0$. In addition, volatility can never be negative, hence $\sigma > 0$.

The solution to (A.29) can be found by applying Itô's formula to $e^{\theta t}x(t)$:

$$\begin{aligned}
d\left(e^{\theta t}x(t)\right) &= \theta e^{\theta t}x(t)dt + e^{\theta t}dx(t) \\
&= \theta e^{\theta t}x(t)dt + e^{\theta t}(\theta(\mu - x(t))dt + \sigma dW^{\mathbb{P}}(t)) \\
&= \theta e^{\theta t}\mu dt + e^{\theta t}\sigma dW^{\mathbb{P}}(t) \\
\Rightarrow e^{\theta t}x(t) &= x(0) + \mu\theta\int_0^t e^{\theta u}du + \sigma\int_0^t e^{\theta u}dW^{\mathbb{P}}(u) \qquad \text{(A.30)} \\
&= x(0) + \mu(e^{\theta t} - 1) + \sigma\int_0^t e^{\theta u}dW^{\mathbb{P}}(u) \\
\Rightarrow x(t) &= x(0)e^{-\theta t} + \mu(1 - e^{-\theta t}) + \sigma e^{-\theta t}\int_0^t e^{\theta u}dW(u). \qquad \text{(A.31)}
\end{aligned}$$

Now

$$\mathbb{E}[x(t)] = x(0)e^{-\theta t} + \mu(1 - e^{-\theta t}) \text{ and } \mathrm{Var}[x(t)] = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}) \tag{A.32}$$

where the latter is determined as follows:

$$\begin{aligned}
\mathrm{Var}[x(t)] &= \mathbb{E}[x(t) - \mathbb{E}[x(t)]]^2 \\
&= \mathbb{E}\left[x(0)e^{-\theta t} + \mu(1 - e^{-\theta t}) + \sigma e^{-\theta t}\int_0^t e^{\theta u}dW^{\mathbb{P}}(u) - x(0)e^{-\theta t} + \mu(1 - e^{-\theta t})\right]^2 \\
&= \mathbb{E}\left[\sigma e^{-\theta t}\int_0^t e^{\theta u}dW^{\mathbb{P}}(u)\right]^2 \\
&= \mathbb{E}\left[\sigma^2 e^{-2\theta t}\int_0^t e^{2\theta u}du\right] \quad \text{by Itô Isometry} \\
&= \frac{\sigma^2}{2\theta}e^{-2\theta t}e^{-2\theta u}\Big|_0^t \\
&= \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}).
\end{aligned}$$

As $t \to \infty$, the mean and variance reduce to their long-term counterparts, that is

$$x(t) \sim N\left(\mu, \frac{\sigma^2}{2\theta}\right).$$

A.2.6.1 Deriving the OU Likelihood Function

In the following, we write $\mu_x = \mu$ and $\sigma_x = \sigma$ for ease of presentation, as in isolation there is no ambiguity between parameters of OU and GBM, which was seen in previous sections.

Since $x(t)$ is normally distributed, it follows that the conditional density of the discretised OU process, where $X = (X_1, ..., X_n)$ is

$$f(X_i|X_{i-1};\mu,\theta,\sigma) = \frac{1}{\sqrt{2\pi \text{Var}(X)}} \exp\left[\frac{-(X_i - \mathbb{E}[X])^2}{2\text{Var}(X)}\right]$$

$$\Rightarrow f(X_i|X_{i-1};\mu,\theta,\sigma) = \frac{1}{\sqrt{\frac{\sigma^2\pi}{\theta}(1-e^{-2\theta\Delta t})}} \exp\left[\frac{-(X_i - X_{i-1}e^{-\theta\Delta t} - \mu(1-e^{-\theta\Delta t}))^2}{\frac{\sigma^2}{\theta}(1-e^{-2\theta\Delta t})}\right], \quad i = 2, ..., n.$$

The natural logarithm is given by

$$\ln f(X_i|X_{i-1};\mu,\theta,\sigma) = -\frac{1}{2}\ln\left(\frac{\sigma^2\pi}{\theta}(1-e^{-2\theta\Delta t})\right) - \frac{(X_i - X_{i-1}e^{-\theta\Delta t} - \mu(1-e^{-\theta\Delta t}))^2}{\frac{\sigma^2}{\theta}(1-e^{-2\theta\Delta t})}, \quad i = 2, ..., n.$$

By taking the product of the individual conditional densities yields the likelihood, whilst the sum of the individual log conditional densities over $i = 2, ..., n$ gives the log likelihood:

$$\begin{aligned}
\mathcal{L}(X|\mu,\theta,\sigma) &= \prod_{i=2}^{n} f(X_i|X_{i-1};\mu,\theta,\sigma) \\
\ln\mathcal{L}(X|\mu,\theta,\sigma) &= \sum_{i=2}^{n} \ln f(X_i|X_{i-1};\mu,\theta,\sigma) \\
&= -\frac{n-1}{2}\ln\left(\frac{\sigma^2\pi}{\theta}(1-e^{-2\theta\Delta t})\right) - \sum_{i=2}^{n}\left[\frac{(X_i - X_{i-1}e^{-\theta\Delta t} - \mu(1-e^{-\theta\Delta t}))^2}{\frac{\sigma^2}{\theta}(1-e^{-2\theta\Delta t})}\right].
\end{aligned} \tag{A.33}$$

A reparametrisation of the model, as put forward by Giet *et al.* (2004), simplifies the model parameters to be estimated and the OU AR(1) process can now be expressed as

$$X_i = \alpha + \phi X_{i-1} + v\sqrt{\Delta t}\epsilon_i, \quad \epsilon_i \sim N(0,1) \tag{A.34}$$

which is akin to the OUVLTM AR(1) process at (3.18) obtained from its own reparametrisation. (A.34) is normally distributed with mean α and variance v^2. As before, ϕ is representative of the degree of autocorrelation.

The conditional density of X_i is now written as

$$f(X_i|X_{i-1};\alpha,\phi,v) = \frac{1}{\sqrt{2\pi v^2}} \exp\left[\frac{-(X_i - \phi X_{i-1} - \alpha)^2}{2v^2}\right] \quad i = 2, ..., n \tag{A.35}$$

where

$$\begin{aligned}
\alpha &= \mu(1-e^{-\theta\Delta t}), \\
\phi &= e^{-\theta\Delta t}, \\
v &= \sqrt{\frac{\sigma^2}{2\theta}(1-e^{-2\theta\Delta t})}.
\end{aligned} \tag{A.36}$$

As in Section 3.3.1.2, $0 < \phi < 1$ and $v > 0$. In the scenario where the underlying process is a credit spread, $\alpha > 0$ follows from the assumption that the long term mean of credit spreads will always be positive. If they become negative then this indicates zero credit risk which, practically speaking, does not happen. The original parameter set can be retrieved using the inverse transformations given by:

$$\begin{aligned} \mu &= \frac{\alpha}{1-\phi}, \\ \theta &= -\frac{\ln \phi}{\Delta t}, \\ \sigma &= \sqrt{\frac{2\theta v^2}{(1-e^{-2\theta \Delta t})}}. \end{aligned} \tag{A.37}$$

The log of the conditional density of X_i is obtained by taking the natural logarithm of (A.35), resulting in

$$\ln f(X_i|X_{i-1};\alpha,\phi,v) = -\frac{1}{2}\ln(2\pi v^2) - \frac{1}{2v^2}[X_i - \phi X_{i-1} - \alpha]^2, \quad i = 2,...,n.$$

Again, the likelihood can be determined as the product of the individual conditional densities and the log likelihood is the sum of the individual log conditional densities for $i = 2,...,n$:

$$\begin{aligned} \mathcal{L}(X|\alpha,v,v) &= \prod_{i=2}^{n} f(X_{i+1}|X_i;\alpha,\phi,v) \\ \ln \mathcal{L}(X|\alpha,\phi,v) &= \sum_{i=2}^{n} \ln f(X_{i+1}|X_i;\alpha,\phi,v) \\ &= -\frac{n-1}{2}\ln(2\pi v^2) - \frac{1}{2v^2}\sum_{i=2}^{n}[X_i - \phi X_{i-1} - \alpha]^2. \end{aligned} \tag{A.38}$$

A.2.6.2 Calculating the MLEs of the OU Model

The MLEs of the OU model, denoted by $\hat{\alpha}, \hat{\phi}$ and $\hat{v}$, maximise the likelihood function at (A.38), or

$$\max_{\alpha>0,0<\phi<1,v>0} \ln \mathcal{L}(X|\alpha,\rho,v) = \max_{\alpha>0,0<\phi<1,v>0} \left(-\frac{n-1}{2}\ln(2\pi v^2) - \frac{1}{2v^2}\sum_{i=2}^{n}[X_i - \phi X_{i-1} - \alpha]^2 \right). \tag{A.39}$$

Numerical implementation of MLE means optimising the OU likelihood function in 3-dimensional space. However, reasonable estimates values are computed with relative ease.

Alternatively, Franco (2003) suggests a technique whereby $\hat{\alpha}(\hat{\phi})$ and $\hat{v}(\hat{\alpha},\hat{\phi})$ are substituted into the likelihood function at (A.38), so that it becomes a function of $\hat{\phi}$ only, thus reducing the optimisation problem to 1-dimension.

Analytical solutions for $\hat{\alpha}, \hat{\phi}$ and $\hat{v}$ can be found using

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \alpha}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = 0, \\ \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \rho}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \frac{1}{v^2}\sum_{i=2}^{n}(X_i X_{i-1} - \phi X_{i-1}^2 - \alpha X_{i-1})\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = 0, \\ \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} &= -\frac{n-1}{v^2} + \frac{1}{v^3}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)^2\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = 0. \end{aligned} \tag{A.40}$$

The solutions to (A.40) are

$$\hat{\phi} = \frac{(n-1)\sum_{i=2}^{n} X_i X_{i-1} \sum_{i=2}^{n} X_{i-1} - \sum_{i=2}^{n} X_i}{(n-1)\sum_{i=2}^{n} X_{i-1}^2 - (\sum_{i=2}^{n} X_{i-1})^2}, \tag{A.41}$$

$$\hat{\alpha}(\hat{\phi}) = \frac{1}{n-1}\sum_{i=2}^{n}[X_i - \hat{\phi}X_{i-1}], \tag{A.42}$$

$$\hat{v}(\hat{\alpha}, \hat{\phi}) = \sqrt{\frac{1}{n-1}\sum_{i=2}^{n}[X_i - \hat{\phi}X_{i-1} - \hat{\alpha}]^2}. \tag{A.43}$$

From (A.37), the MLEs for μ, θ and σ are found to be

$$\hat{\mu}(\hat{\alpha}, \hat{\phi}) = \frac{\hat{\alpha}}{1-\hat{\phi}}, \tag{A.44}$$

$$\hat{\theta}(\phi) = -\frac{\ln \hat{\phi}}{\Delta t}, \tag{A.45}$$

$$\hat{\sigma}(\hat{\theta}, \hat{v}) = \sqrt{\frac{2\hat{\theta}\hat{v}^2}{(1-e^{-2\hat{\theta}\Delta t})}}. \tag{A.46}$$

A.2.6.3 Determining the Standard Errors of the OU Model

The first partial derivatives given at (A.40) are used to obtain the second partial derivatives as follows

$$\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \alpha^2}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = \frac{\partial}{\partial \alpha}\left[\frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)\right]\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{n-1}{\hat{v}^2},$$

$$\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \phi^2}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = \frac{\partial}{\partial \phi}\left[\frac{1}{v^2}\sum_{i=2}^{n}(X_i X_{i-1} - \phi X_{i-1}^2 - \alpha X_{i-1})\right]\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{1}{\hat{v}^2}\sum_{i=2}^{n} X_{i-1}^2,$$

$$\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v^2}\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = \frac{\partial}{\partial v}\left[-\frac{n-1}{v^2} + \frac{1}{v^3}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)^2\right]\Big|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{2(n-1)}{\hat{v}^2}.$$

Again, the result from (A.43) for $\hat{v}$,

$$(n-1)\hat{v}^2 = \sum_{i=2}^{n}[X_i - \hat{\phi}X_{i-1} - \hat{\alpha}]^2,$$

is used in the last line. Thus the standard errors for the MLEs of α, ϕ and v are

$$\hat{\alpha}^{SE} = \sqrt{\frac{\hat{v}^2}{n-1}}, \tag{A.47}$$

$$\hat{\phi}^{SE} = \sqrt{\frac{\hat{v}^2}{\sum_{i=2}^{n} X_{i-1}^2}}, \tag{A.48}$$

$$\hat{v}^{SE} = \sqrt{\frac{\hat{v}^2}{2(n-1)}}. \tag{A.49}$$

The standard error of the maximum likelihood estimate of the variance $\hat{v}^2$ can be shown to equal

$$\hat{v}^{2\,SE} = \sqrt{\frac{2\hat{v}^4}{n-1}}. \tag{A.50}$$

The second partial derivatives for μ, θ and σ are obtained via the chain rule as before:

$$\begin{aligned}
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \mu^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left[\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \alpha^2}\frac{\partial \alpha}{\partial \mu} + \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \alpha}\frac{\partial^2 \alpha}{\partial \alpha \partial \mu}\right]\left.\frac{\partial \alpha}{\partial \mu}\right|_{\hat{\alpha},\hat{\phi},\hat{v}}, \\
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \phi^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left[\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \phi^2}\frac{\partial \phi}{\partial \theta} + \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \phi}\frac{\partial^2 \phi}{\partial \phi \partial \theta}\right]\left.\frac{\partial \phi}{\partial \theta}\right|_{\hat{\alpha},\hat{\phi},\hat{v}}, \\
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left[\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v^2}\frac{\partial v}{\partial \sigma} + \frac{\partial \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v}\frac{\partial^2 v}{\partial v \partial \sigma}\right]\left.\frac{\partial v}{\partial \sigma}\right|_{\hat{\alpha},\hat{\phi},\hat{v}}.
\end{aligned}$$

Now, using (A.36),

$$\begin{aligned}
\frac{\partial \alpha}{\partial \mu} = 1-\rho \Rightarrow \frac{\partial}{\partial \alpha}\left(\frac{\partial \alpha}{\partial \mu}\right) &= \frac{\partial}{\partial \alpha}(1-\rho) = 0, \\
\frac{\partial \phi}{\partial \theta} = -\Delta t e^{-\theta \Delta t} \Rightarrow \frac{\partial}{\partial \phi}\left(\frac{\partial \phi}{\partial \theta}\right) &= \frac{\partial}{\partial \phi}(-\Delta t e^{-\theta \Delta t}) = 0, \\
\frac{\partial v}{\partial \sigma} = \frac{\sigma(1-e^{-2\theta \Delta t})}{2\theta\sqrt{\frac{\sigma^2}{2\theta}(1-e^{-2\theta \Delta t})}} \Rightarrow \frac{\partial}{\partial v}\left(\frac{\partial v}{\partial \sigma}\right) &= \frac{\partial}{\partial v}\left(\frac{\sigma(1-e^{-2\theta \Delta t})}{2\theta\sqrt{\frac{\sigma^2}{2\theta}(1-e^{-2\theta \Delta t})}}\right) = 0.
\end{aligned}$$

Thus

$$\begin{aligned}
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \mu^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \alpha^2}\left(\frac{\partial \alpha}{\partial \mu}\right)^2\right|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{(n-1)(1-\hat{\phi})^2}{\hat{v}^2}, \\
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \theta^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \phi^2}\left(\frac{\partial \phi}{\partial \theta}\right)^2\right|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{\Delta t^2 \hat{\phi}^2}{\hat{v}^2}\sum_{i=2}^{n} X_{i-1}^2, \\
\left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial \sigma^2}\right|_{\hat{\alpha},\hat{\phi},\hat{v}} &= \left.\frac{\partial^2 \ln \mathcal{L}(X|\alpha,\phi,v)}{\partial v^2}\left(\frac{\partial v}{\partial \sigma}\right)^2\right|_{\hat{\alpha},\hat{\phi},\hat{v}} = -\frac{2(n-1)}{\hat{v}^2}\frac{\hat{\sigma}(1-e^{-2\hat{\theta}\Delta t})}{2\hat{\theta}\sqrt{\frac{\hat{\sigma}^2}{2\hat{\theta}}(1-e^{-2\hat{\theta}\Delta t})}} \\
&= -\frac{2(n-1)}{\hat{v}^2}\frac{\hat{v}^2}{\hat{\sigma}^2} = -\frac{2(n-1)}{\hat{\sigma}^2}.
\end{aligned}$$

And so finally

$$\hat{\mu}^{SE} = \sqrt{\frac{\hat{v}^2}{(n-1)(1-\hat{\phi}^2)}} \tag{A.51}$$

$$\hat{\theta}^{SE} = \sqrt{\frac{\hat{v}^2}{\Delta t^2 \hat{\phi}^2 \sum_{i=2}^{n} X_{i-1}^2}} \tag{A.52}$$

$$\hat{\sigma}^{SE} = \sqrt{\frac{\hat{\sigma}^2}{2(n-1)}}. \tag{A.53}$$

A.2.6.4 Deriving the Conditional Posterior Distributions of the OU Parameters

To use Gibbs sampling to estimate the parameters of the OU model, the conditional posterior distributions of the parameters at (A.36) in Appendix A are now determined in Theorem A.4.

Theorem A.4. *Assume that the prior distributions for α, ϕ and v^2 are*

$$p(\alpha) \sim N(\mu_{\alpha 0}, \tau_{\alpha 0}^2), \tag{A.54}$$

$$p(\phi) \sim N(\mu_{\rho 0}, \tau_{\phi 0}^2) \quad \textit{and} \tag{A.55}$$

$$p(v^2) \sim \mathcal{IG}(\nu_0, \psi_0). \tag{A.56}$$

Then the full conditional distributions for α, ϕ and v are given by

$$p(\alpha|\phi, v, X) \sim N(\mu_\alpha, \tau_\alpha^2), \tag{A.57}$$

$$p(\phi|\alpha, v, X) \sim N(\mu_\phi, \tau_\phi^2) \quad \textit{and} \tag{A.58}$$

$$p(v^2|\alpha, \phi, X) \sim \mathcal{IG}\left(\nu_n, \psi_n\right). \tag{A.59}$$

where

$$\mu_\alpha = \frac{\frac{\mu_{\alpha 0}}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^n (X_i - \phi X_{i-1})}{\frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}}, \qquad \tau_\alpha^2 = \frac{1}{\frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}},$$

$$\mu_\phi = \frac{\frac{\mu_{\phi 0}}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^n X_{i-1}(X_i - \alpha)}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^n X_{i-1}^2}, \qquad \tau_\phi^2 = \frac{1}{\frac{1}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^n X_{i-1}^2},$$

$$\nu_n = \frac{n-1}{2} + \nu_0 \qquad \textit{and} \qquad \psi_n = \frac{1}{2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2 + \psi_0.$$

Proof. The proof will proceed by deriving the results for α and ϕ, as these proofs follow the same procedure. Then, the result for v^2 will be proved.

Assume first that ϕ and v are known and α is unknown.

By taking the exponent of (A.38), it can be observed that

$$\mathcal{L}(X|\alpha, \phi, v) \propto \exp\left[-\frac{1}{2v^2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2\right].$$

From (A.54), it can also be seen that

$$p(\alpha) \propto \exp\left[-\frac{1}{2\tau_{\alpha 0}^2}(\alpha - \mu_{\alpha 0})^2\right].$$

Thus using Bayes' rule at (B.1), the conditional distribution for α can be written as

$$\begin{aligned} p(\alpha|\rho, v, X) &\propto \mathcal{L}(X|\alpha, \phi, v)p(\alpha) \\ &\propto \exp\left[-\frac{1}{2v^2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2\right] \exp\left[-\frac{1}{2\tau_{\alpha 0}^2}(\alpha - \mu_{\alpha 0})^2\right]. \end{aligned} \tag{A.60}$$

Consider now the terms in the exponents, omitting the $-\frac{1}{2}$ multiplier for the moment:

$$\begin{aligned} &\frac{1}{v^2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2 + \frac{1}{\tau_{\alpha 0}^2}(\alpha - \mu_{\alpha 0})^2 \\ &= \frac{1}{v^2}\sum_{i=2}^n \left(X_i^2 - 2\phi X_{i-1}X_i - 2\alpha X_i + \phi^2 X_{i-1}^2 + 2\alpha\phi X_{i-1} + \alpha^2\right) + \frac{1}{\tau_{\alpha 0}^2}\left(\alpha^2 - 2\alpha\mu_{\alpha 0} + \mu_{\alpha 0}^2\right) \\ &= \alpha^2\left(\frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}\right) - 2\alpha\left(\frac{\mu_{\alpha 0}}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^n (X_i - \phi X_{i-1})\right) + \frac{\mu_{\alpha 0}^2}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^n (X_i - \phi X_{i-1})^2 \\ &= a_\alpha \alpha^2 - 2b_\alpha \alpha + c_\alpha \end{aligned}$$

where

$$a_\alpha = \frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}, \quad b_\alpha = \frac{\mu_{\alpha 0}}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1}) \quad \text{and} \quad c_\alpha = \frac{\mu_{\alpha 0}^2}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1})^2.$$

Hence

$$\begin{aligned} p(\alpha|\phi, v, X) &\propto \exp\left[-\frac{1}{2}(a_\alpha \alpha^2 - 2b_\alpha)\right] \\ &= \exp\left[-\frac{1}{2}a_\alpha\left(\alpha^2 - \frac{2b_\alpha}{a_\alpha} + \frac{b_\alpha^2}{a_\alpha^2}\right) + \frac{b_\alpha^2}{2a_\alpha}\right] \quad \text{(by completing the square)} \\ &\propto \exp\left[-\frac{1}{2}a_\alpha\left(\alpha - \frac{b_\alpha}{a_\alpha}\right)^2\right] \\ &= \exp\left[-\frac{(\alpha - \frac{b_\alpha}{a_\alpha})^2}{2/a_\alpha}\right]. \end{aligned} \tag{A.61}$$

By referring to (A.6), it is obvious that (A.61) is normally distributed with mean and variance given respectively by

$$\mu_\alpha = \frac{b_\alpha}{a_\alpha} = \frac{\frac{\mu_{\alpha 0}}{\tau_{\alpha 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1})}{\frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}} \qquad \text{and} \qquad \tau_\alpha^2 = \frac{1}{a_\alpha} = \frac{1}{\frac{1}{\tau_{\alpha 0}^2} + \frac{n-1}{v^2}}.$$

Thus the result for $p(\alpha|\phi, v, X)$ is proved.

Next assume that α and v are known and ϕ is unknown.

Following a similar process for $p(\phi|\alpha, v, X)$ gives

$$p(\phi|\alpha, v, X) \propto \exp\left[-\frac{1}{2v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)^2\right]\exp\left[-\frac{1}{2\tau_{\phi 0}^2}(\phi - \mu_{\phi 0})^2\right].$$

Again, the terms in the exponents are considered, whilst the $-\frac{1}{2}$ multiplier is omitted:

$$\begin{aligned} &\frac{1}{v^2}\sum_{i=2}^{n}(X_i - \phi X_{i-1} - \alpha)^2 + \frac{1}{\tau_{\phi 0}^2}(\phi - \mu_{\phi 0})^2 \\ &= \frac{1}{v^2}\sum_{i=2}^{n}\left(X_i^2 - 2\phi X_{i-1}X_i - 2\alpha X_i + \phi^2 X_{i-1}^2 + 2\alpha\phi X_{i-1} + \alpha^2\right) + \frac{1}{\tau_{\phi 0}^2}\left(\phi^2 - 2\phi\mu_{\phi 0} + \mu_{\phi 0}^2\right) \\ &= \phi^2\left(\frac{1}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}X_{i-1}^2\right) - 2\phi\left(\frac{1}{v^2}\sum_{i=2}^{n}(X_{i-1}X_i - \alpha X_{i-1})\right) + \frac{\mu_{\phi 0}^2}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \alpha)^2 \\ &= a_\phi\phi^2 - 2b_\phi\phi + c_\phi, \end{aligned}$$

where

$$a_\phi = \frac{1}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i-1}^{n}X_{i-1}^2, \quad b_\phi = \frac{1}{v^2}\sum_{i=2}^{n}X_{i-1}(X_i - \alpha) \quad \text{and} \quad c_\phi = \frac{\mu_{\phi 0}^2}{\tau_{\phi 0}^2} + \frac{1}{v^2}\sum_{i=2}^{n}(X_i - \alpha)^2.$$

Using the same procedure as shown to obtain (A.61), the conditional distribution of ϕ can be takes the form

$$p(\phi|\alpha, v, X) \propto \exp\left[-\frac{(\phi - \frac{b_\phi}{a_\phi})^2}{2/a_\phi}\right].$$

It follows that $p(\phi|\alpha, v, X)$ is normally distributed with mean and variance given by

$$\mu_\phi = \frac{b_\rho}{a_\phi} = \frac{\frac{\mu_{\phi 0}}{\tau^2_{\phi 0}} + \frac{1}{v^2}\sum_{i=2}^n X_{i-1}(X_i - \alpha)}{\frac{1}{\tau^2_{\phi 0}} + \frac{1}{v^2}\sum_{i=2}^n X^2_{i-1}} \qquad \text{and} \qquad \tau^2_\phi = \frac{1}{a_\alpha} = \frac{1}{\frac{1}{\tau^2_{\phi 0}} + \frac{1}{v^2}\sum_{i=2}^n X^2_{i-1}}.$$

The proof for $p(\phi|\alpha, v, X)$ is complete.

Finally, assume that α and ϕ are known and v is unknown.

As before, it is observed from (A.38) that

$$\mathcal{L}(X|\alpha, \phi, v) \propto \frac{1}{v^{2\left(\frac{n-1}{2}\right)}} \exp\left[-\frac{1}{2v^2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2\right]$$

but now including the v^2 term outside the exponent.

The prior of v is of the form

$$p(v^2) \propto \frac{1}{v^{2(\nu_0+1)}} \exp\left[\frac{-\psi_0}{v^2}\right]$$

as implied by (A.56).

Thus

$$\begin{aligned} p(v^2|\alpha, \phi, X) &\propto \frac{1}{v^{2\left(\frac{n-1}{2}\right)}} \exp\left[-\frac{1}{2v^2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2\right] \frac{1}{v^{2(\nu_0+1)}} \exp\left[-\frac{\psi_0}{v^2}\right] \\ &= \frac{1}{v^{2\left(\frac{n-1}{2}+\nu_0+1\right)}} \exp\left[-\frac{1}{v^2}\left(\frac{1}{2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2 + \psi_0\right)\right]. \end{aligned} \tag{A.62}$$

When compared to (A.15), the conditional distribution of v^2 at (A.62) is clearly inverse gamma, with shape and scale parameters given by

$$\nu_n = \frac{n-1}{2} + \nu_0 \qquad \text{and} \qquad \psi_n = \frac{1}{2}\sum_{i=2}^n (X_i - \phi X_{i-1} - \alpha)^2 + \psi_0$$

respectively. When sampled, the square root is taken to get a value for v.

□

A.2.7 Autoregressive Processes

An autoregressive process is a time-varying stochastic process where the current state of the process is linearly linked to the previous state and a stochastic term (Bogacka (2016)).

More formally, an autoregressive process, X, of order p (denoted by $AR(p)$) satisfies

$$\begin{aligned} X(t) &= \beta_1 X(t-\Delta t) + \beta_2 X(t-2\Delta t) + ... + \beta_p X(t-p\Delta t) + W(t) \\ &= \sum_{i=1}^{p} \beta_i X(t-i\Delta t) + W(t), \end{aligned}$$

where $W(t)$ is a Wiener process with $W(t) \sim N(0, \sigma^2)$.

A process, denoted by $Y(t)$, with non-zero mean equal μ is given by

$$\begin{aligned} Y(t) - \mu &= \beta_1(Y(t-\Delta t) - \mu) + \beta_2(Y(t-2\Delta t) - \mu) + ... + \beta_p(Y(t-p\Delta t) - \mu) + W(t) \\ &= \sum_{i=1}^{p} \beta_i(Y(t-i\Delta t) - \mu) + W(t) \\ \Rightarrow Y(t) &= \delta + \sum_{i=1}^{p} \beta_i Y(t-i\Delta t) + W(t), \end{aligned} \tag{A.63}$$

where

$$\delta = \mu(1 - \beta_1 - ... - \beta_p).$$

From (A.63), an AR(1) process ,$Z(t)$, can be written as

$$Z(t) = \delta + \beta Z(t-\Delta t) + W(t).$$

A.2.8 Augmented Dickey-Fuller (ADF) Test for Stationarity

The Augmented Dickey-Fuller test is a method of determining the whether a time series is stationary. The null hypothesis for the test is the possible existence of a unit root and the alternative hypothesis is that the time series is stationary, since processes containing a unit root cannot be stationary. "It makes use of the fact that if a time series possesses mean reversion, then the next price level will be proportional to the current price level" (Halls-Moore (2013)).

The null hypothesis model is given by the linear lag model of order p

$$x(t) = c + \delta t + \phi x(t-1) + \rho_1 \Delta x(t-1) + \rho_2 \Delta x(t-2) + ... + \rho_p \Delta x(t-p) + \epsilon(t),$$

where c is the drift coefficient, $\Delta x(t) = x(t) - x(t-1)$ and $\epsilon(t)$ is the Gaussian noise process $N \sim (0, 1)$.

The null hypothesis is the existence of a unit root: $H_0 : \phi = 1$, tested against the alternative hypothesis, $H_1 : \phi < 1$. If ϕ =1, the time series diverges with an ever increasing mean and variance.

The test statistic is calculated as the sample proportionality constant over the standard error of the sample proportionality constant:

$$DF_\tau = \frac{\hat{\phi}}{SE(\hat{\phi})}.$$

The critical value is a standard t statistic given by

$$t = \frac{\hat{\phi} - 1}{SE(\hat{\phi})},$$

which uses ordinary least squares regression to obtain estimates of the AR(1) coefficient ($\hat{\phi}$) and $SE(\hat{\phi})$ in the alternative model.

If the test statistic is more negative (smaller) than the critical value taken at a specific significance level, then the null hypothesis of the existence of a unit root is rejected at that significance level (also indicated by the p-value of the regression being lower than the significance level) and the time series is concluded to be stationary.

The test allows for different models to be used: if $\delta = 0$ then there is no trend and if $c = 0$ then there is no drift.

A.2.9 Goodness of Fit Measures

A.2.9.1 Root Mean Square Error (RMSE)

The definition of the root mean square error of the parameter estimators $\hat{\pi}$, relative to the parameters being estimated π, is given by

$$\mathcal{E}_{RMS} = \sqrt{\mathbb{E}[(\hat{\pi} - \pi)^2]}.$$

Thus, the parameter estimators performance can be determined by calculating the RMSE between the observed or 'true' data X (of length n) and the imputed path generated using the estimators $\hat{X}$, using

$$\mathcal{E}_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - X_i)^2}. \tag{A.64}$$

The smallest RMSE indicates superior parameter estimation.

A.2.9.2 Mean Absolute Percentage Error (MAPE)

For data with strong trends, the mean absolute percentage error gives a more realistic view of the deviation between the true and imputed data (Moritz *et al.* (2015)). The measure is defined as

$$\mathcal{E}_{MAP} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{X}_i - X_i|}{|X_i|}. \tag{A.65}$$

A.2.10 Akaike Information Criterion (AIC)

The Akaike Information Criterion, or AIC, provides a way of assessing the goodness of fit of a model to a data set (Akaike (1974)). Thus the AIC can be used as a tool to choose the most suitable model for the data set from a selection of proposed models.

Suppose $\mathcal{L}$ is the maximum likelihood for some model, with k parameters, on a given data set. Then

the AIC of the model is calculated as

$$AIC = 2k - 2ln\mathcal{L}. \tag{A.66}$$

The superior model has the minimum AIC value.

If the size of the data set, n, is not much larger than k^2, then the AIC could potentially select a model with too many parameters. The AICc corrects the AIC for small data sets and is given by

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}. \tag{A.67}$$

A.2.11 Bayesian Information Criterion (BIC)

Closely related to AIC, the Bayesian Information Criterion (BIC) was developed by Schwarz (1978) in a Bayesian setting. Suppose $\mathcal{L}$ is the maximum likelihood for some model, with k parameters, on a given data set of size n. Then the BIC of the model is calculated as

$$BIC = \ln(n)k - 2ln\mathcal{L}. \tag{A.68}$$

The superior model has the minimum BIC value, the same as AIC. It also penalises models with more parameters, to a degree now determined by n.

A.2.12 The Kalman Filter

In real-world scenarios where the usage of models is afflicted by noisy or irregular data, smoothing or filtering techniques often prove useful in attaining estimates of the unknown variables (state space models where the latent variable is unknown, thus termed hidden Markov models). Such a method was developed by Kalman (1960).

Suppose that the variable under consideration to be modelled, x_k, depends on its previous state, x_{k-1}, perturbed by some noise factor, w_k, and satisfies

$$x_k = A_k x_{k-1} + B_k u_k + w_k. \tag{A.69}$$

A_k is the state transition model from state $k-1$ to state k, B_k is the control input model, u_k is the control input vector and $w_k \sim N(0, Q_k)$ is the noise of the process, with $Q_k = \mathbb{E}[w_k w_k^T]$. Control inputs are often not relevant in financial models and are mostly omitted in the pertaining literature, as will be done here as well. (A.69) represents the state transition equation.

The measurement equation, which relates x_k to the current observation, is given by

$$z_k = H_k x_k + v_k, \tag{A.70}$$

where H_k is the observation model and $v_k \sim N(0, R_k)$ is the measurement error, with $R_k = \mathbb{E}[v_k v_k^T]$.

A.2.12.1 Derivation of the Kalman Filter Algorithm

This section is based on content taken from Mastro (2013).

The Kalman filter algorithm utilises (A.69) and (A.70) to find the optimal weighting, called the Kalman gain, between the model predicted estimate and the observed value by finding the minimum variance of the mean square error between the true state, x_k, and the estimated state, $\hat{x}_k$. The error at time k is given by

$$e_k = x_k - \hat{x}_k. \tag{A.71}$$

The prediction estimate for the state at time step k, denoted by $\hat{x}_k^-$, is obtained using the transition model at (A.69) as follows:

$$\hat{x}_k^- = A_k \hat{x}_{k-1}. \tag{A.72}$$

Similarly, the observation model at (A.70) is used with the predicted estimate to determine the observed (measurement) estimate for the state at time step k:

$$\hat{z}_k^- = H_k \hat{x}_k^- = H_k A_k \hat{x}_{k-1}. \tag{A.73}$$

The innovation or measurement residual, i_k, is the difference between the actual observation at time k, z_k, and the predicted observation at (A.73):

$$i_k = z_k - \hat{z}_k^- = z_k - H_k \hat{x}_k^-.$$

The corrected estimate for the state variable at k, $\hat{x}_k$, is calculated by adding to the transition estimate the innovation residual, weighted by the Kalman gain K_k:

$$\begin{aligned} \hat{x}_k &= \hat{x}_k^- + K_k i_k \\ &= \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-) \\ \Rightarrow \hat{x}_k &= \hat{x}_k^-(I - K_k H_k) + K_k z_k. \end{aligned} \tag{A.74}$$

Substituting (A.70) into (A.74) yields

$$\hat{x}_k = \hat{x}_k^-(I - K_k H_k) + K_k(H_k x_k + v_k). \tag{A.75}$$

Now the error covariance matrix is given by

$$P_k = \mathbb{E}[e_k e_k^T] = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]. \tag{A.76}$$

Using the expression at (A.75) in (A.76) gives

$$\begin{aligned} P_k &= \mathbb{E}[(x_k - \hat{x}_k^-(I - K_k H_k) + K_k(H_k x_k + v_k)])(x_k - \hat{x}_k^-(I - K_k H_k) + K_k(H_k x_k + v_k))^T] \\ &= \mathbb{E}[((I - K_k H_k)(x_k - \hat{x}_k^-) - K_k v_k)((I - K_k H_k)(x_k - \hat{x}_k^-) - K_k v_k)^T]. \end{aligned} \tag{A.77}$$

The error between x_k and $\hat{x}_k^-$, written as

$$e_k^- = x_k - \hat{x}_k^-, \tag{A.78}$$

is uncorrelated with the measurement noise v_k, thus (A.77) becomes

$$\begin{aligned} P_k &= (I - K_k H_k)\mathbb{E}[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T](I - K_k H_k)^T + K_k \mathbb{E}[v_k v_k^T] K_k^T \\ &= (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T. \end{aligned} \tag{A.79}$$

The mean square errors are the diagonal elements of the error covariance matrix at (A.79), which are simply the variances (since the diagonal elements of a covariance matrix are the variances). The total mean square error is the sum of the diagonal mean square error terms. The sum of the diagonal elements of a square matrix is called the trace. Hence to find the minimum mean square error and obtain an expression for the optimal Kalman gain, the minimum of the trace needs to be determined.

First, expanding out the terms in (A.79) gives

$$P_k = P_k^- - K_k H_k P_k^- - P_k^- K_k^T H_k^T + K_k (H_k P_k^- H_K^T + R_k) K_k^T. \tag{A.80}$$

Now the trace of P_k, denoted by $\mathrm{Tr}(P_k)$, is given by

$$\begin{aligned} \mathrm{Tr}(P_k) &= \mathrm{Tr}(P_k^-) - \mathrm{Tr}(K_k H_k P_k^-) - \mathrm{Tr}(P_k^- K_k^T H_k^T) + \mathrm{Tr}(K_k (H_k P_k^- H_K^T + R_k) K_k^T) \\ &= \mathrm{Tr}(P_k^-) - 2\,\mathrm{Tr}(K_k H_k P_k^-) + \mathrm{Tr}(K_k (H_k P_k^- H_K^T + R_k) K_k^T), \end{aligned} \tag{A.81}$$

where the second line follows since

$$\mathrm{Tr}(K_k H_k P_k^-) = \mathrm{Tr}(P_k^- K_k^T H_k^T).$$

To find the optimal K_k, (A.81) is differentiated with respect to K_k and set to zero:

$$\frac{d\,\mathrm{Tr}(P_k)}{dK_k} = -2(H_k P_k^-)^T + 2K_k(H_k P_k^- H_K^T + R_k) = 0.$$

Solving for K_k yields the optimal Kalman gain as

$$\Rightarrow K_k = P_k^- H_k^T (H_k P_k^- H_K^T + R_k)^{-1}. \tag{A.82}$$

The measurement prediction covariance, S_k, is defined as

$$S_k := H_k P_k^- H_k^T + R_k. \tag{A.83}$$

Thus the Kalman gain can be rewritten as

$$K_k = P_k^- H_k^T (S_k)^{-1}. \tag{A.84}$$

Multiplying both sides of (A.84) by $S_k K_k^T$ yields

$$K_k S_k K_k^T = P_k^- H_k^T K_k^T. \tag{A.85}$$

Using (A.83) and then (A.85) in (A.80) results in the correction expression for the error covariance:

$$\begin{aligned} P_k &= P_k^- - K_k H_k P_k^- - P_k^- K_k^T H_k^T + K_k S_k K_k^T \\ &= P_k^- - K_k H_k P_k^- - P_k^- K_k^T H_k^T + P_k^- H_k^T K_k^T \\ &= P_k^- - K_k H_k P_k^- \\ \Rightarrow P_k &= (I - K_k H_k) P_k^-. \end{aligned} \tag{A.86}$$

The correction step of the Kalman filter is fully defined by (A.82), (A.74) and (A.86). The expression for the prediction of the state variable for the next time step is (A.72), with the corresponding error estimate given by (A.78). The remaining component to determine is the prediction of the error covariance.

Substituting (A.69) and (A.72) into (A.78) gives

$$\begin{aligned} e_k^- &= (A_k x_{k-1} + w_k) - A_k \hat{x}_{k-1} \\ &= A_k e_{k-1} + w_k. \end{aligned}$$

The prediction of the error covariance for the next time step is

$$P_k^- = \mathbb{E}[(e_k^-)(e_k^-)^T] = \mathbb{E}[(A_k e_{k-1} + w_k)(A_k e_{k-1} + w_k)^T]. \tag{A.87}$$

w_k and e_{k-1}^- are uncorrelated, as w_k only occurs between times $k-1$ and k, whilst e_{k-1}^- accumulates up until time $k-1$ only. Thus the expression for the prediction of the error covariance is

$$\begin{aligned} P_k^- &= A_k \mathbb{E}[(e_k^-)(e_k^-)^T] A_k^T + \mathbb{E}[w_k W_k^T] \\ \Rightarrow P_k^- &= A_k P_{k-1} A_k^T + Q_k. \end{aligned} \tag{A.88}$$

A.2.13 Forward Filtering Backward Sampling (FFBS)

The method of Forward Filtering Backward Sampling is often used in state space models to compute the probabilities of outcomes of the state space given a set of observations. The forward algorithm typically uses the Kalman filter to generate these probabilities given the first K observations. The backward algorithm uses backward recursion of the Kalman filter (the equations of which are shown in (A.89) as per Laine *et al.* (2014)) to calculate the probabilities of observing the observations after k

$$\begin{aligned} L_k &= A_k(I - K_k H_k), \\ r_k &= H_k(H_k P_k^- H_k^T + R_k) v_k + L_k^T r_{t+1}, \\ N_k &= H_k^T(H_k) P_k^- H_k^T + R_k) H_k + L_k^T N_{t+1} L_k, \\ \tilde{x}_k &= \hat{x}_k + P_k^- r_k \qquad \text{smoothed state}, \\ \tilde{P}_k &= P_k^- - P_k^- N_k P_k^- \quad \text{smoothed state covariance}. \end{aligned} \tag{A.89}$$

B. Bayesian Statistics and Inference

Sections B.2 to B.5.2 are almost entirely comprised of content taken directly from Gelman *et al.* (2004) and Letham *et al.* (2012).

B.1 Frequentist versus Bayesian Statistics

Statistical inference has been dominated by two ways of thinking, namely the Frequentist approach and Bayesian inference. These two schools of thought differ in their definitions of probability, i.e. to what the derived probability is being applied.

The Frequentist approach sees a certain amount of data and from the frequency of this data, derives some correct conclusion (hypothesis testing) for a certain given probability, using the likelihood function of the model. The hypothesis is either true or false, based on significance tests such as p-values. Underlying parameters are assumed to be fixed and deterministic *and* so do not have associated probability distributions, whilst the data is the unknown and described probabilistically.

Bayesian statistics assigns probabilities to the unknowns (parameters and missing data), whilst the observed data is taken to be fixed. These probabilities take into account prior knowledge of the unknown values and the conclusion is in the form of probability distributions. Thus there is dependence on the likelihood and the prior probability distribution.

Informally, "...Bayesian statistics starts from what has been observed and assesses possible future outcomes. Frequentist statistics starts with an abstract experiment of what would be observed if one assumes something, and only then compares the outcomes of the abstract experiment with what was actually observed. Otherwise the two approaches are compatible. They both assess the probability of future observations based on some observations made or hypothesised."[1]

B.2 Notation

Assume that the m parameters that govern a model proposed to describe some observed data set, denoted by $\boldsymbol{X} = (X_1, ..., X_n)$, are given by $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_m)$, which may be also be scalar. $\tilde{\boldsymbol{X}}$ denotes the unobserved data.

The probability density function in typical probability theory is usually denoted by $f(\boldsymbol{X})$, and in the Bayesian setting is given by $p(\boldsymbol{X})$. The conditional density function of $\boldsymbol{X}$ given $\boldsymbol{\pi}$ is written as $p(\boldsymbol{X}|\boldsymbol{\pi})$.

[1] user36160 (https://stats.stackexchange.com/users/36160/user36160), Bayesian and Frequentist reasoning in plain English, URL (version: 2013-12-13): https://stats.stackexchange.com/q/79605

B.3 Bayes' Rule

The conditional probability of $\boldsymbol{\pi}$ given $\boldsymbol{X}$ is obtained using the formula for conditional probabilities, which is the joint probability of $\boldsymbol{\pi}$ and $\boldsymbol{X}$ over the density of $\boldsymbol{X}$:

$$p(\boldsymbol{\pi}|\boldsymbol{X}) = \frac{p(\boldsymbol{\pi}, \boldsymbol{X})}{p(\boldsymbol{X})} = \frac{p(\boldsymbol{X}|\boldsymbol{\pi})p(\boldsymbol{\pi})}{p(\boldsymbol{X})}, \tag{B.1}$$

where

$$\begin{aligned} p(\boldsymbol{X}) &= \int_{\boldsymbol{\pi}} p(\boldsymbol{X}, \boldsymbol{\pi'})p(\boldsymbol{\pi'})d\boldsymbol{\pi'}, \\ p(\boldsymbol{X}) &= \sum_{\boldsymbol{\pi}} p(\boldsymbol{X}, \boldsymbol{\pi})p(\boldsymbol{\pi}) \end{aligned} \tag{B.2}$$

in the continuous and discrete settings respectively. (B.2) is known as the partition function and can also be thought of as the distribution of $\boldsymbol{X}$ before being observed, also known as the prior predictive distribution.

(B.1) is formally known as Bayes' rule and results in a formulation for the posterior distribution, $p(\boldsymbol{\pi}|\boldsymbol{X})$, in terms of the likelihood function, $p(\boldsymbol{X}, \boldsymbol{\pi})$, the prior distribution, $p(\boldsymbol{\pi})$ and the partition function at (B.2).

The unnormalised posterior distribution omits the partition function to give:

$$p(\boldsymbol{\pi}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{\pi})p(\boldsymbol{\pi}). \tag{B.3}$$

The posterior predictive distribution is the distribution of the predicted values of $\tilde{\boldsymbol{X}}$ given the observed data $\boldsymbol{X}$:

$$\begin{aligned} p(\tilde{\boldsymbol{X}}|\boldsymbol{X}) &= \int_{\pi} p(\tilde{\boldsymbol{X}}, \boldsymbol{\pi'}|\boldsymbol{X})d\pi' \\ &= \int_{\pi} p(\tilde{\boldsymbol{X}}|\boldsymbol{\pi'}, \boldsymbol{X})p(\boldsymbol{\pi'}|\boldsymbol{X})d\pi' \\ &= \int_{\pi} p(\tilde{\boldsymbol{X}}|\boldsymbol{\pi'})p(\boldsymbol{\pi'}|\boldsymbol{X})d\pi'. \end{aligned}$$

The last line follows from the fact that $\tilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ are conditionally independent given $\boldsymbol{\pi}$. In other words, conditioning on $\boldsymbol{\pi}$ makes new observations independent of previous observations and $\boldsymbol{\pi}$ contains all the information required about $\tilde{\boldsymbol{X}}$.

B.4 Choosing the Prior Distribution

The prior distribution encompasses any information about the distribution of $\boldsymbol{\pi}$ that may be known before observation.

In the instances where there is definite knowledge about the distribution of the model parameters, the prior may be chosen to embody such knowledge and is called an informative prior. An example of an informative prior for the volatility of a stock price using Geometric Brownian Motion is the Inverse-Gamma distribution, which satisfies the condition that volatility can never be negative and is positively skewed.

In many circumstances, however, information about the model parameters is nebulous and not much is known about their distributions before observation of the data. For these scenarios, an uninformative or objective prior can be chosen. One manner of developing an objective prior is to use the Principle of Indifference, where equal probabilities are given to all possible outcomes.

There are often times when the partition function is not integrable, which occurs frequently when using higher dimensional models. There are a reasonable number of ways to avoid having to deal with intractable partition functions; these include choosing specific prior distributions which lead to closed form posteriors, as well as numerical simulations.

B.4.1 Conjugate Priors

Choosing a particular set of priors, based on the likelihood of a model, is a technique to ensure a closed form posterior distribution and avoids having to integrate a potentially intractable partition function. This concept is formalised by a definition from Letham et al. (2012).

Definition B.1. Let $\mathcal{F}$ be a family of likelihood functions and $\mathcal{P}$ a family of prior distributions. $\mathcal{P}$ is a *conjugate prior* to $\mathcal{F}$ if for any likelihood function $f \in \mathcal{F}$ and for any prior distribution $p \in \mathcal{P}$, the corresponding posterior distribution p^* satisfies $p^* \in \mathcal{P}$.

A conjugate prior distribution's parameters and the parameters of the resulting posterior are called the hyperparameters. The prior hyperparameters are chosen to such that the distribution generated from those parameter values represents the known information about $\boldsymbol{\pi}$. The posterior hyperparameters can be determined explicitly as a result of the conjugacy.

Definition B.2 and Theorem B.3, also taken from Letham *et al.* (2012), layout a mechanism allowing for easy identification of conjugacy.

Definition B.2. The family of distributions $\mathcal{F}$ is an exponential family if every member of $\mathcal{F}$ has the form

$$p(X_i|\boldsymbol{\pi}) = h(X_i)g(\boldsymbol{\pi}) \exp[\chi(\boldsymbol{\pi})^T u(X_i)] \tag{B.4}$$

for some $h(\cdot), g(\cdot), \chi(\cdot)$ and $u(\cdot)$.

Theorem B.3. *If the likelihood model is an exponential family, then there exists a conjugate prior.*

Proof. Consider the likelihood of an identical and independently distributed random variable $X = (X_1, ..., X_n)$:

$$\begin{aligned} p(X|\boldsymbol{\pi}) &= \prod_{i=1}^{n} p(X_i|\boldsymbol{\pi}) = \prod_{i=1}^{n} h(X_i)g(\boldsymbol{\pi}) \exp[\chi(\boldsymbol{\pi})^T u(X_i)] \quad \text{from (B.4)} \\ &= \left[\prod_{i=1}^{n} h(X_i)\right] g(\boldsymbol{\pi})^n \exp\left[\chi(\boldsymbol{\pi})^T \sum_{i=1}^{n} u(X_i)\right]. \end{aligned}$$

Take the prior distribution to be

$$p(\boldsymbol{\pi}) = \frac{g(\boldsymbol{\pi})^\kappa \exp[\chi(\boldsymbol{\pi})^T \varphi]}{\int g(\boldsymbol{\pi}')^\kappa \exp[\chi(\boldsymbol{\pi}')^T \varphi] d\boldsymbol{\pi}'}, \tag{B.5}$$

where κ and φ are prior hyperparameters. Then the posterior is

$$\begin{aligned} p(\boldsymbol{\pi}|X) &\propto p(X|\boldsymbol{\pi})p(\boldsymbol{\pi}) \\ &= \left[\prod_{i=1}^{n} h(X_i)\right] g(\boldsymbol{\pi})^n \exp\left[\chi(\boldsymbol{\pi})^T \sum_{i=1}^{n} u(X_i)\right] \frac{g(\boldsymbol{\pi})^\kappa \exp[\chi(\boldsymbol{\pi})^T \varphi]}{\int g(\boldsymbol{\pi}')^\kappa \exp[\chi(\boldsymbol{\pi}')^T \varphi] d\boldsymbol{\pi}'} \\ &\propto g(\boldsymbol{\pi})^{n+\kappa} \exp\left[\chi(\boldsymbol{\pi})^T \left(\sum_{i=1}^{n} u(X_i) + \varphi\right)\right] \quad \text{omitting non-}\boldsymbol{\pi}\text{ terms,} \end{aligned}$$

which is in the same family as the prior at (B.5), with posterior hyperparameters being

$$n + \kappa \quad \text{and} \quad \sum_{i=1}^{n} u(X_i) + \varphi.$$

□

B.5 Markov Chain Monte Carlo Sampling

Only recently has it become possible to use computational methods to determine approximations to the posterior distribution in the cases where the partition function at (B.2) is not analytically tractable and conjugate priors are not used.

There are many algorithms with an assortment of variations that have been developed to numerically approximate the posterior distribution. The algorithmic class that will be focussed on and implemented here is that of Markov Chain Monte Carlo (MCMC). Markov Chain Monte Carlo sampling of the posterior is a general term referring to several unique simulation methods. All these methods, however, use sequential draws from approximate distributions (proposal distributions), looking to create a chain of draws whose distribution tends to the posterior (target) distribution when containing a large enough number of draws. The stationary distribution of the chain is then considered a good approximation to the posterior.

Since each accepted draw in the MCMC chain is dependent only on the previous draw (as the criteria for acceptance only looks at the current draw and the previous draw), the chain is Markov.

B.5.1 Metropolis-Hastings Algorithm

One of the more well-known MCMC sampling methods is the Metropolis-Hastings algorithm. This algorithm employs an acceptance/rejection rule to build a distribution that converges to the posterior when run long enough.

The Metropolis-Hastings algorithm is defined by the following process steps:

1. Choose starting values, $\boldsymbol{\pi_0}$, for $\boldsymbol{\pi}$. These values should lie somewhere in the intuitively expected ranges for the model parameters.

2. Assume some functional form for the prior distribution $p(\boldsymbol{\pi})$. The choice of prior distributions is discussed further in Section B.4.

3. For each iteration, indexed $i = 1, ..., q$, where q is the length of the Markov chain (or the number of iterations used in the MCMC run):

   (a) Sample $\boldsymbol{\pi^*}$ from the proposal or jumping density $J(\boldsymbol{\pi^*}, \boldsymbol{\pi^{i-1}})$, where $\boldsymbol{\pi^{i-1}}$ denotes the immediately previous value in the chain.

   (b) Calculate the acceptance ratio

$$\begin{aligned} a(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*}) &= \min\left(\frac{p(\boldsymbol{\pi^*}|\boldsymbol{X})J(\boldsymbol{\pi^*}, \boldsymbol{\pi^{i-1}})}{p(\boldsymbol{\pi^{i-1}}|\boldsymbol{X})J(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*})}, 1\right) \\ &= \min\left(\frac{p(\boldsymbol{X}|\boldsymbol{\pi^*})p(\boldsymbol{\pi^*})J(\boldsymbol{\pi^*}, \boldsymbol{\pi^{i-1}})}{p(\boldsymbol{X}|\boldsymbol{\pi^{i-1}})p(\boldsymbol{\pi^{i-1}})J(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*})}, 1\right), \end{aligned} \tag{B.6}$$

   where the posteriors are calculated as unnormalised using (B.3).

   (c) Set

$$\boldsymbol{\pi^i} = \begin{cases} \boldsymbol{\pi^*} & \text{with probability } a(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*}) \\ \boldsymbol{\pi^{i-1}} & \text{otherwise.} \end{cases}$$

   By generating a uniform random number r, the acceptance or rejection of the proposal draw is determined if $a(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*}) > r$ or $a(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^*}) < r$.

   (d) Repeat Steps 2-5 until stationary distribution and n have been reached.

The usefulness of this sampling technique in circumventing the need to calculate the partition function, $p(\boldsymbol{X})$, can be seen at (B.6), where the ratio of the posteriors causes the partitions to cancel out.

Theorem B.5 (taken from Letham *et al.* (2012)) shows that when the MCMC chain is run for long enough, simulated draws approximate the posterior distribution. In order to prove this theorem, the following is needed.

Theorem B.4. *Define the transition kernel $K(\boldsymbol{\pi}, \boldsymbol{\pi'})$ as the probability of transitioning from state $\boldsymbol{\pi}_i$ to state $\boldsymbol{\pi}_j$, where $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ form part of the Markov chain $\boldsymbol{\pi}$. If the distribution $\chi(\cdot)$ satisfies the detailed balance equation given by*

$$K(\pi_i, \pi_j)\chi(\pi_i) = K\pi_j, \pi_i)\chi(\pi_j), \quad \textit{for all } \pi_i, \pi_j, \tag{B.7}$$

then $\chi(\cdot)$ is the stationary distribution.

Proof. Suppose that $\chi(\cdot)$ satisfies (B.7). Then

$$\begin{aligned} \sum_i K(\pi_i, \pi_j)\chi(\pi_i) &= \sum_i K(\pi_j, \pi_i)\chi(\pi_j) \\ &= \chi(\pi_j)\sum_i K(\pi_j, \pi_i) \\ &= \chi(\pi_j) \\ \Rightarrow \chi(\cdot) &= \chi(\cdot)K(\cdot, \cdot). \end{aligned}$$

Thus $\chi(\cdot)$ is stationary, since it is invariant under application of the transition kernel. □

Theorem B.5. *If $J(\boldsymbol{\pi}, \boldsymbol{\pi'})$ is such that that the Markov chain $\pi_0, \pi_1, ...$ produced by the Metropolis-Hastings algorithm has a unique stationary distribution, then the stationary distribution is $p(\boldsymbol{\pi}|\boldsymbol{X})$.*

Proof. It needs to be shown that the posterior satisfies (B.7), which from Theorem B.4 means it is the stationary distribution.

The transition kernel arising from Metropolis-Hastings is equal to

$$\begin{aligned} K(\boldsymbol{\pi}, \boldsymbol{\pi}') &= \text{probability of proposing } \boldsymbol{\pi}' \times \text{probability of accepting } \boldsymbol{\pi}' \text{ given it was proposed} \\ &= J(\boldsymbol{\pi}, \boldsymbol{\pi}')a(\boldsymbol{\pi}, \boldsymbol{\pi}'). \end{aligned}$$

Now consider any $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ and suppose, without loss of generality, that a is less than or equal to 1 for the transition from $\boldsymbol{\pi}$ to $\boldsymbol{\pi}'$. Hence

$$J(\boldsymbol{\pi}, \boldsymbol{\pi}')p(\boldsymbol{\pi}|\boldsymbol{X}) \geq J(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X})$$

so that

$$a(\boldsymbol{\pi}, \boldsymbol{\pi}') = \frac{J(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X})}{J(\boldsymbol{\pi}, \boldsymbol{\pi}')p(\boldsymbol{\pi}|\boldsymbol{X})}$$

and $a(\boldsymbol{\pi}', \boldsymbol{\pi}) = 1$. Thus

$$\begin{aligned} K(\boldsymbol{\pi}, \boldsymbol{\pi}')p(\boldsymbol{\pi}|\boldsymbol{X}) &= J(\boldsymbol{\pi}, \boldsymbol{\pi}')a(\boldsymbol{\pi}, \boldsymbol{\pi}')p(\boldsymbol{\pi}|\boldsymbol{X}) \\ &= J(\boldsymbol{\pi}, \boldsymbol{\pi}')\frac{J(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X})}{J(\boldsymbol{\pi}, \boldsymbol{\pi}')p(\boldsymbol{\pi}|\boldsymbol{X})}p(\boldsymbol{\pi}|\boldsymbol{X}) \\ &= J(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X}) \\ &= J(\boldsymbol{\pi}', \boldsymbol{\pi})a(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X}) \quad \text{since } a(\boldsymbol{\pi}', \boldsymbol{\pi}) = 1 \\ &= K(\boldsymbol{\pi}', \boldsymbol{\pi})p(\boldsymbol{\pi}'|\boldsymbol{X}). \end{aligned}$$

□

B.5.2 Metropolis Algorithm

A specific case of the Metropolis-Hastings algorithm arises when the proposal density is symmetric, that is

$$J(\boldsymbol{\pi}^*, \boldsymbol{\pi}^{i-1}) = J(\boldsymbol{\pi}^{i-1}, \boldsymbol{\pi}^*) \text{ for all } i. \tag{B.8}$$

This specific scenario is known as the Metropolis algorithm. The procedure followed for Metropolis is identical to that used for the Metropolis-Hastings algorithm, except for the acceptance ratio at (B.6). The acceptance ratio for the Metropolis algorithm is as follows:

$$a(\boldsymbol{\pi}^{i-1}, \boldsymbol{\pi}^*) = \min\left(\frac{p(\boldsymbol{\pi}^*|\boldsymbol{X})}{p(\boldsymbol{\pi}^{i-1}|\boldsymbol{X})}, 1\right) = \min\left(\frac{p(\boldsymbol{X}|\boldsymbol{\pi}^*)p(\boldsymbol{\pi}^*)}{p(\boldsymbol{X}|\boldsymbol{\pi}^{i-1})p(\boldsymbol{\pi}^{i-1})}, 1\right). \tag{B.9}$$

There are no proposal densities present in (B.9) due to the symmetry of these densities, shown at (B.8), causing the density in the numerator to cancel out with the density in the denominator.

The proposal distribution for the MCMC draws acts as a candidate for the posterior; it provides the samples which will either be accepted or rejected by the algorithm. Thus the proposal density must support all possible values for $\boldsymbol{\pi}$ such that there is a positive probability of reaching all valid states.

Letham *et al.* (2012) propose four criteria that the proposal distribution must satisfy in order for the

MCMC sampler to be successful:

1. Efficient sampling
2. Easy computation of the acceptance ratio $a(\boldsymbol{\pi^{i-1}}, \boldsymbol{\pi^{*}})$
3. Moves from draw to draw must not be too large, which would lead to low acceptance and slow convergence
4. Moves from draw to draw must not be too small, which would lead to high acceptance and slow exploration of the parameter space

B.5.3 Gibbs Sampler

Many models are high in dimension, thus the posterior $p(\boldsymbol{\pi}|\boldsymbol{X})$ can become complex and is difficult to sample from. The Clifford-Hammersley theorem provides the means to uniquely specify a highly dimensional joint distribution in terms of lower dimensional conditionals (Johannes (2010)).

In the instances where the full conditional distributions of each model parameter,

$$p(\pi_j|\boldsymbol{X}, \pi_1, \pi_2, ..., \pi_{j-1}, \pi_{j+1}, ..., \pi_m),$$

for $j = 1, ..., m$, are available in closed form and can be directly sampled, a special case of the Metropolis-Hastings algorithm, known as the Gibbs sampler, can be applied.

Here, the proposal densities for the parameters are the known full conditional distributions and their usage results in the acceptance ratio at (B.6) always being equal to unity (in other words, every draw is accepted). McElreath (2015) describes the Gibbs sampler as being more 'efficient', since it requires far fewer samples to converge to the posterior distribution than a similar Metropolis-Hastings approach.

The closed form solutions of the full conditional distributions are typically made available through conjugacy, where a conjugate prior to the model likelihood is chosen for each parameter.

The Gibbs sampler algorithm is as follows:

1. Choose conjugate priors for each π_j.
2. Determine the corresponding conditional distribution for each π_j.
3. Repeat the following steps for $i = 1, .., q$ iterations until convergence has been reached:
   (a) Sample $\pi_1^{i+1} \sim p(\pi_1|\pi_2^i, ..., \pi_q^i, X_1, ..., X_n)$
   Sample $\pi_2^{i+1} \sim p(\pi_2|\pi_1^{i+1}, \pi_3^i ..., \pi_q^i, X_1, ..., X_n)$
   $\vdots$
   Sample $\pi_q^{i+1} \sim p(\pi_q|\pi_1^{i+1}, ..., \pi_{q-1}^{i+1}, X_1, ..., X_n)$
   (b) Set the current draw to be $(\pi_1^{i+1}, ..., \pi_q^{i+1})$.

B.6 Convergence of the MCMC Sampler

In usual Monte Carlo simulation, the sequential random values generated are preferably independent and identically distributed (i.i.d). MCMC does not maintain this characteristic, as the chain is Markovian and each draw is dependent on the previous one. Thus convergence for MCMC chains is typically slower than typical Monte Carlo.

The mixing of the chain, the degree to which the chain explores the parameter space, is a good indication of whether adequate convergence has been achieved. Poor mixing indicates slow convergence, whilst convergence is faster for a chain that mixes rapidly.

Numerous techniques have been developed to ascertain whether a MCMC chain has reached convergence. Convergence can be assessed initially by using qualitative or visual methods and then by subsequent more rigorous quantitative measures.

B.6.1 On the Usage of Burn-In

When using MCMC sampling, it is generally advised that discarding an initial portion of the chain eliminates the dependence on the starting values of the chain and supposedly quickens convergence. In most literature involving MCMC sampling, some reference to using burn-in is proposed. Indeed, Bolstad (2010), Gelman *et al.* (2011) and Gelman *et al.* (2004) all advocate the usage of a burn-in period, generally taken as half the length of the MCMC chain.

Geyer (2011), on the other hand, argues that burn-in is not required and a more intuitive mechanism must be used in its place to find suitable starting values. The rule of thumb given by the author is "any point you don't mind having in a sample is a good starting point". By running a series of preliminary simulations, the modeller can assess a suitable point from which to initialise the MCMC chain, perhaps even the last value of the previous run. Gelman *et al.* (2011) also mention the possibility of inefficiency due to the usage of burn-in, but for the sake of convergence are willing to accept an increase in model error.

B.6.2 Trace Plots

The simplest manner to gauge whether an MCMC chain has converged to a stationary distribution is by observing the evolution of the chain graphically, which is called a trace plot. If the chain values display asymptotic behaviour (at least at the end of the chain) to some equilibrium or oscillatory behaviour around that equilibrium, then the chain is assumed to have converged. In particular, one can graphically gauge the degree of mixing using a trace plot; mixing is considered poor if the trace plot looks either like a random walk (usually corresponding to a scaling that is too small and too many draws are accepted) or has discontinuous jumps with periods of low variability (typically from a scaling that is too large and too many draws are rejected).

B.6.3 Autocorrelations

Autocorrelation or cross-correlation is the correlation of a series of observations Z with itself, lagged by an increasing time distance. Autocorrelation is measured using the following function:

$$R_k = \frac{\mathbb{E}[Z_t - \mu]\mathbb{E}[Z_{t+k} - \mu]}{\sqrt{\mathbb{E}[(Z_t - \mu)^2]\mathbb{E}[(Z_{t+k} - \mu)^2]}} \tag{B.10}$$

where μ is the mean of the series of observations.

Due to the inherent Markovian structure of the MCMC chains, a certain level of autocorrelation is expected. For the chain to be experiencing adequate mixing however, the autocorrelations should decrease significantly as the lag increases. Zero autocorrelation means no dependence, which in turn means that every draw is giving a new piece of information and contributing to more a accurate picture of the posterior distribution.

B.6.4 Gelman-Rubin Test

The Gelman-Rubin test for convergence (Gelman *et al.* (1992)) uses multiple MCMC chains and analyses similarities across these chains. If there are dissimilarities between one or more chains, then convergence has failed to occur.

The test is comprised of two calculations, one where the 'within-chain' variance for each chain is determined and one calculation for the 'between-chain' variance. These variances are given respectively by

$$W = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{\pi}_{ij} - \bar{\boldsymbol{\pi}}_i)^2\right), \tag{B.11}$$

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{\boldsymbol{\pi}}_i - \bar{\bar{\boldsymbol{\pi}}})^2, \tag{B.12}$$

$$\bar{\bar{\boldsymbol{\pi}}} = \frac{1}{m}\sum_{i=1}^{m}\bar{\boldsymbol{\pi}}_i,$$

where m is the number of chains used and n is the length of the chains. The variance of the stationary distribution of $\boldsymbol{\pi}$ is then estimated to be

$$\boldsymbol{\sigma_\pi}^2 = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B. \tag{B.13}$$

If the starting points used for the MCMC chains were not elements of the stationary distribution, then (B.13) overestimates the true variance. W also underestimates the variance of each chain in the earlier iterations of the MCMC run as the sequences have not had enough time to explore the whole parameter space. But as $n \to \infty$, both W and B tend to the true variances.

A scale reduction factor, measuring the factor by which the current distribution would be reduced by if $n \to \infty$, is calculated as

$$\hat{R} = \frac{\boldsymbol{\sigma_\pi}}{\sqrt{W}}. \tag{B.14}$$

If $\hat{R}$ is high (Gelman *et al.* (2011) recommend any values above 1.1), then convergence is not certain.

C. Convergence Diagnostics of Generated Posterior Distributions

The results of testing for convergence to the target distribution of the Gibbs sampler chains making up the posterior distributions of the parameter estimates generated in Chapters 3, 4 and 5 are given in this appendix. Trace plots, autocorrelation function values and the Gelman-Rubin test (the reader is referred to Appendix B.6 for more information) will be used to ascertain convergence.

For convergence to be confirmed, trace plots must show the chains being well mixed, the autocorrelation function values must decrease quickly as the number of iterations increases and the calculated Gelman-Rubin scale reduction factor $\hat{R}$ must be below 1.1.

In Sections 3.4.2, 4.2.2 and 4.3.5, 1,500 posteriors of the 2 OUVLTM and 3 OU parameters are generated in the Gibbs sampler on the 1,500 simulated sample paths. Section 4.2.2 shows the generation of posterior for both models on 3 incomplete data sets, where 10%, 50% and 90% of the points were removed. FBA is used in Section 4.3.5 to form 3 complete data sets by the imputation of the missing 10%, 50% and 90%.

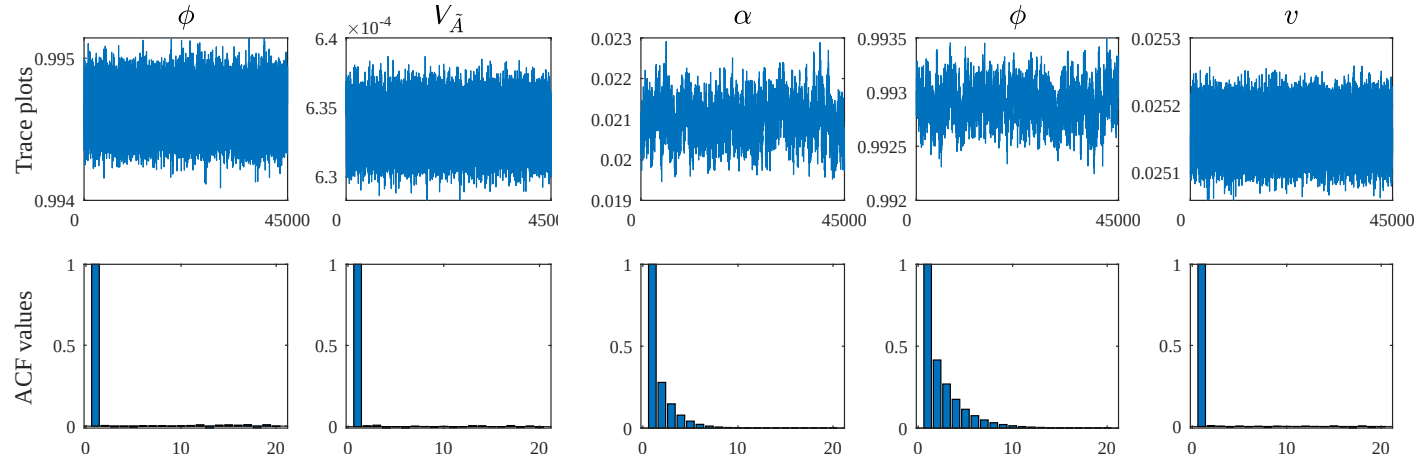
The results of the Gelman-Rubin test for these 7 data sets are displayed in Table C.1 and Figures C.1a to C.7b show the trace plots of the median posteriors and autocorrelation function values of the parameters.

Relevant posteriors	Model	OUVLTM		OU		
	λ	ϕ	$V_{\tilde{A}}$	α	ϕ	v
Section 3.4.2	0%	1.0385	1.0861	1.0364	1.0167	1.0719
Section 4.2.2	10%	1.0811	1.0923	1.0020	1.0893	1.0981
	50%	1.0907	1.0188	1.0096	1.0024	1.0863
	90%	1.0814	1.0493	1.0490	1.0253	1.0478
Section4.3.5	10%	1.0794	1.0975	1.0839	1.0910	1.0392
	50%	1.0089	1.0818	1.0939	1.0411	1.0195
	90%	1.0632	1.0448	1.0288	1.0770	1.0305

Table C.1: Calculated scale reduction factors of the Gelman-Rubin test for convergence on the median posteriors obtained on simulated data in Sections 3.4.2, 4.2.2 and 4.3.5.

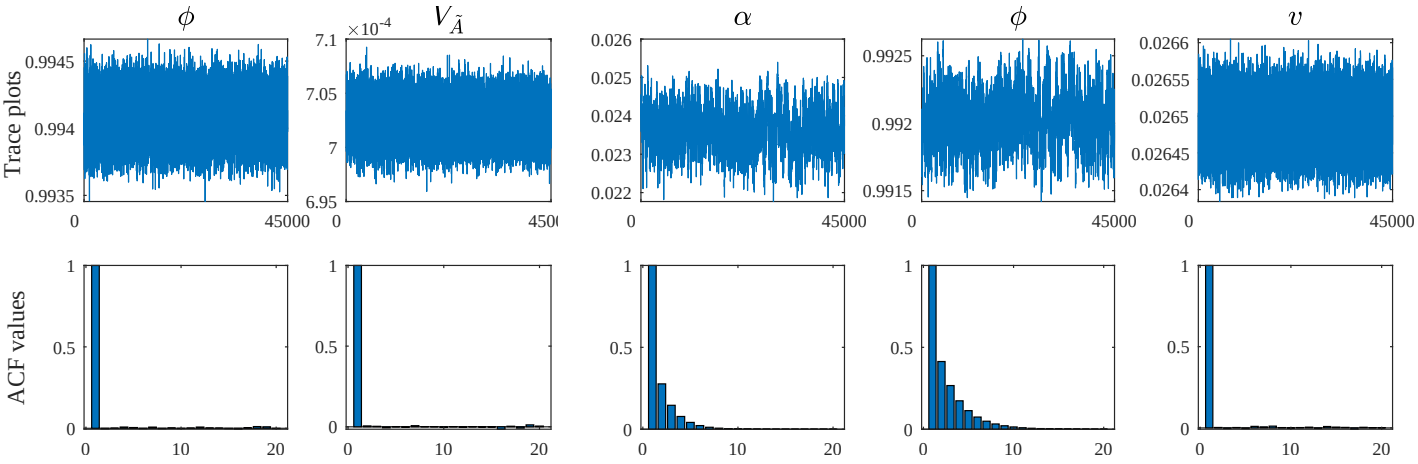
The trace plots exhibit good mixing, the ACF values decay satisfactorily and the $\hat{R}$ are all below 1.1, indicating that the posteriors achieved convergence.

For the OUVLTM and OU parameter posteriors distributions obtained from the TP time series, LVCF time series and FBA for the 9 bonds in Section 5.2, Table C.2 shows the autocorrelations function values at the 20^{th} lag and calculated $\hat{R}$ values. Convergence of all the posteriors is confirmed by the low ACF values at lag 20 and $\hat{R}$ values less than 1.1.



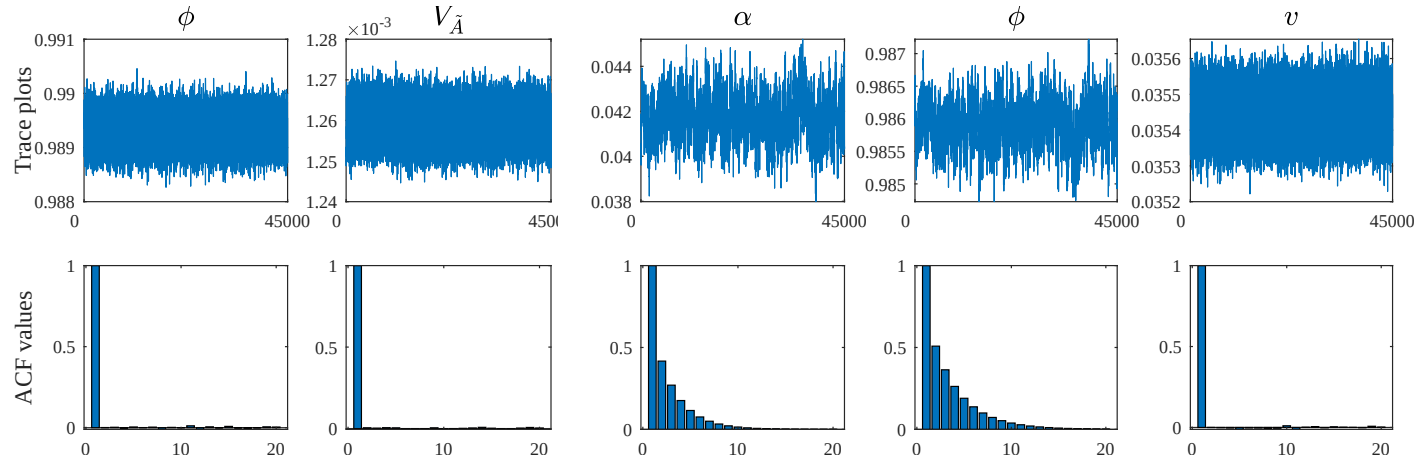
(a) OUVLTM (b) OU

Figure C.1: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 3.4.2.



(a) OUVLTM (b) OU

Figure C.2: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.2.2 on the incomplete data set where $\lambda = 10\%$.



(a) OUVLTM (b) OU

Figure C.3: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.2.2 on the incomplete data set where $\lambda = 50\%$.

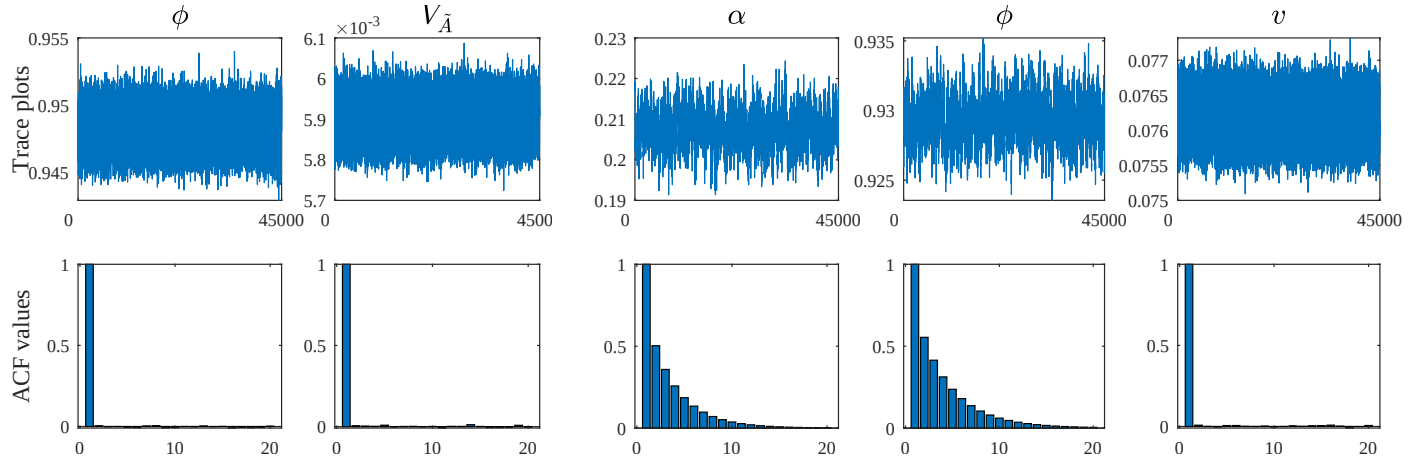


Figure C.4: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.2.2 on the incomplete data set where $\lambda = 90\%$.

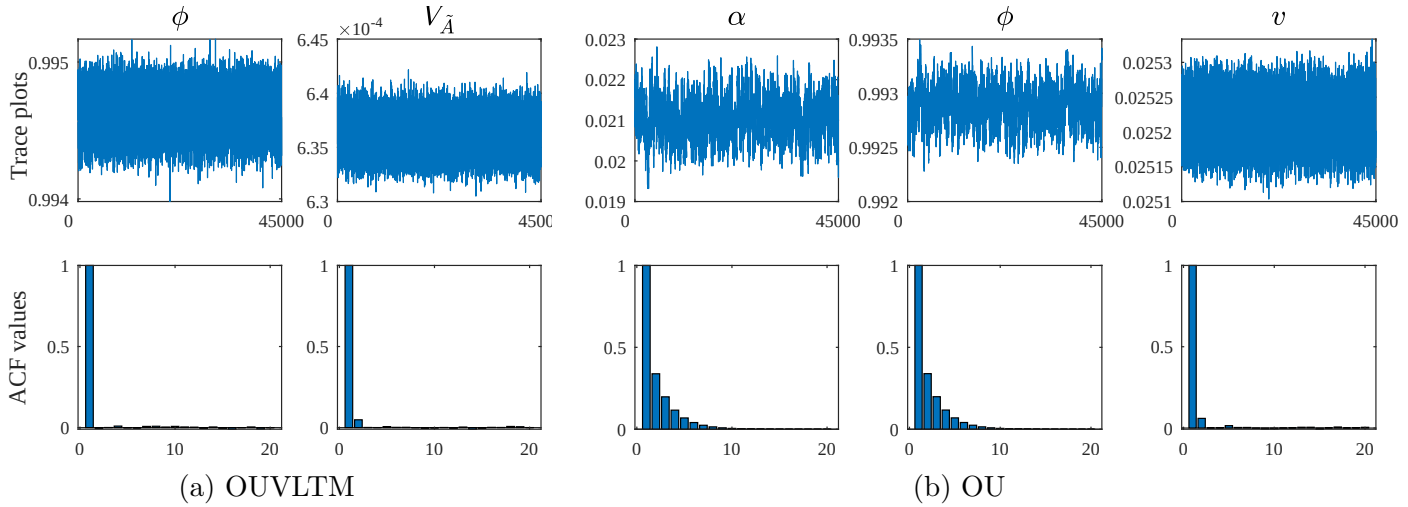


Figure C.5: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.3.5 on the FBA complete data set where $\lambda = 10\%$.

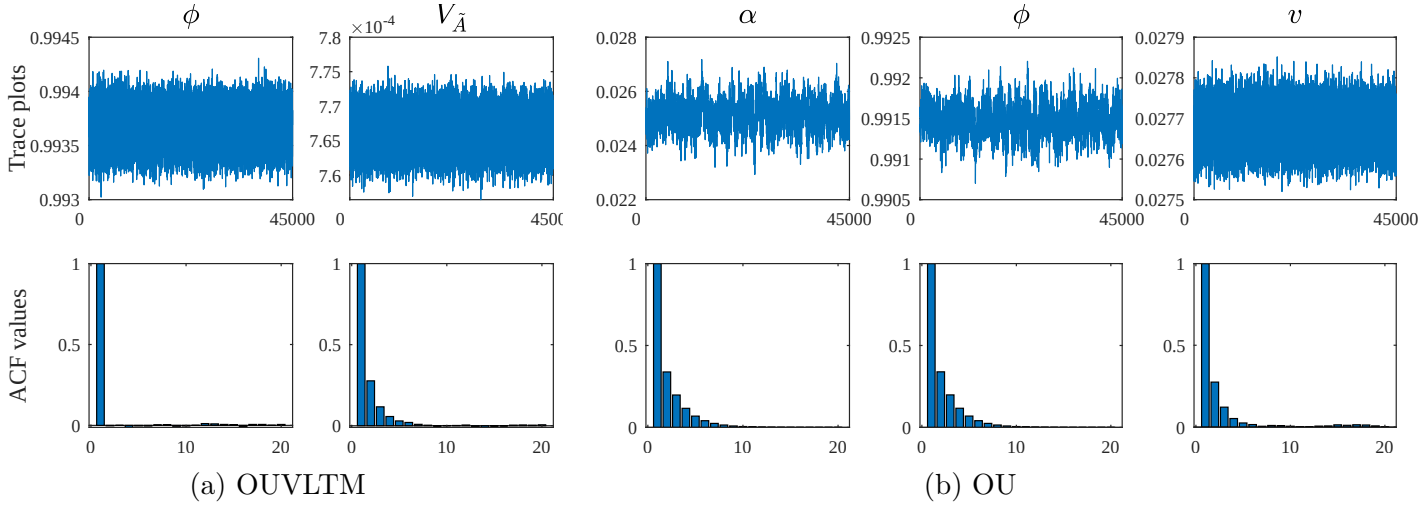
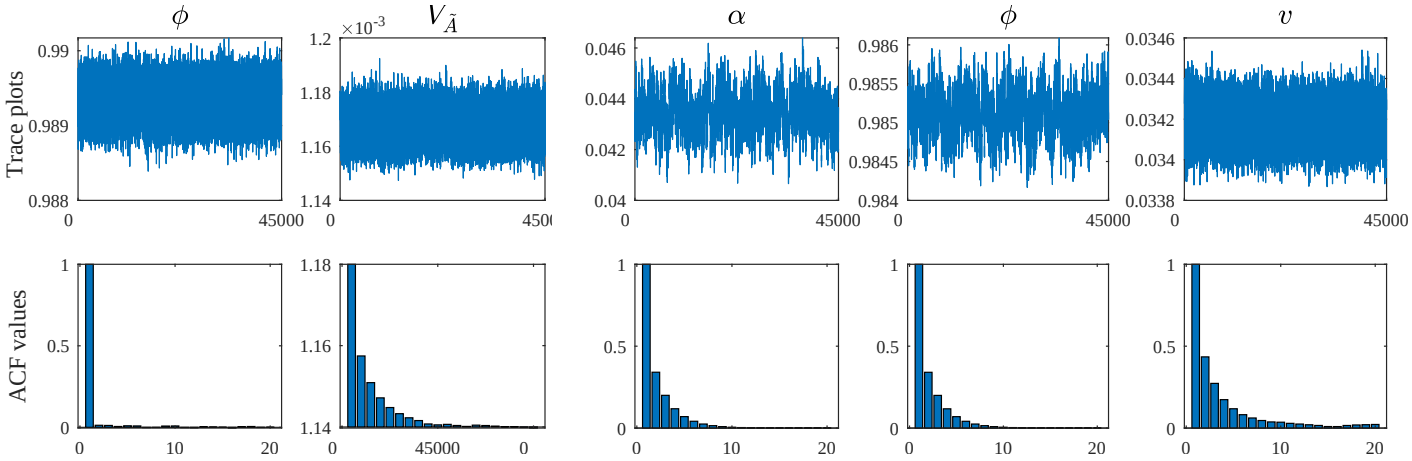


Figure C.6: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.3.5 on the FBA complete data set where $\lambda = 50\%$.



(a) OUVLTM (b) OU

Figure C.7: Trace plots and ACF values for median of the OUVLTM and OU posteriors in Section 4.3.5 on the FBA complete data set where $\lambda = 90\%$.

Bond	Statistic	ACF value at 20$^{\text{th}}$ lag					Gelman-Rubin scale reduction factor $\hat{R}$				
	Model	OUVLTM		OU			OUVLTM		OU		
	Parameter	ϕ	$V_{\tilde{A}}$	α	ϕ	v	ϕ	$V_{\tilde{A}}$	α	ϕ	v
MTN01	TP	0.0035	0.0062	0.0032	0.0010	0.0003	1.0089	1.0840	1.0907	1.0547	1.0935
	LVCF	0.0083	0.0008	0.0190	0.0072	0.0044	1.0692	1.0283	1.0900	1.0121	1.0884
	FBA	0.0069	0.0040	0.0039	0.0028	0.0016	1.0404	1.0895	1.0938	1.0795	1.0654
SBS9	TP	0.0011	0.0028	0.0007	0.0162	0.0062	1.0438	1.0163	1.0246	1.0736	1.0838
	LVCF	0.0070	0.0079	0.0034	0.0115	0.0053	1.0534	1.0989	1.0684	1.0776	1.0470
	FBA	0.0042	0.0061	0.0066	0.0120	0.0045	1.0812	1.0200	1.0574	1.0313	1.0579
CBL11	TP	0.0033	0.0066	0.0010	0.0168	0.0019	1.0486	1.0226	1.0961	1.0569	1.0991
	LVCF	0.0064	0.0028	0.0037	0.0052	0.0099	1.0714	1.0129	1.0345	1.0450	1.0472
	FBA	0.0092	0.0006	0.0131	0.0084	0.0011	1.0481	1.0244	1.0054	1.0252	1.0428
ABS10	TP	0.0053	0.0068	0.0187	0.0038	0.0063	1.0542	1.0111	1.0753	1.0272	1.0570
	LVCF	0.0014	0.0006	0.0029	0.0188	0.0038	1.0010	1.0529	1.0742	1.0414	1.0999
	FBA	0.0007	0.0068	0.0081	0.0195	0.0074	1.0308	1.0073	1.0903	1.0758	1.0471
IPL3	TP	0.0086	0.0048	0.0069	0.0052	0.0077	1.0868	1.0824	1.0387	1.0103	1.0385
	LVCF	0.0019	0.0017	0.0013	0.0194	0.0031	1.0401	1.0139	1.0841	1.0057	1.0663
	FBA	0.0087	0.0026	0.0069	0.0024	0.0090	1.0461	1.0508	1.0845	1.0716	1.0679
NBK6A	TP	0.0060	0.0038	0.0111	0.0132	0.0050	1.0823	1.0471	1.0626	1.0558	1.0150
	LVCF	0.0041	0.0026	0.0163	0.0108	0.0010	1.0217	1.0729	1.0380	1.0334	1.0240
	FBA	0.0075	0.0003	0.0088	0.0032	0.0051	1.0150	1.0241	1.0011	1.0554	1.0834
MTN05	TP	0.0005	0.0071	0.0016	0.0083	0.0083	1.0060	1.0572	1.0492	1.0921	1.0278
	LVCF	0.0027	0.0098	0.0114	0.0111	0.0010	1.0805	1.0705	1.0500	1.0589	1.0286
	FBA	0.0037	0.0023	0.0157	0.0054	0.0095	1.0820	1.0266	1.0381	1.0989	1.0805
FRX16	TP	0.0056	0.0087	0.0076	0.0164	0.0033	1.0299	1.0841	1.0783	1.0219	1.0105
	LVCF	0.0089	0.0055	0.0187	0.0065	0.0041	1.0727	1.0680	1.0176	1.0658	1.0765
	FBA	0.0076	0.0048	0.0057	0.0173	0.0002	1.0064	1.0157	1.0925	1.0796	1.0560
BID02	TP	0.0074	0.0096	0.0199	0.0142	0.0074	1.0100	1.0548	1.0777	1.0809	1.0718
	LVCF	0.0009	0.0048	0.0069	0.0174	0.0010	1.0389	1.0511	1.0610	1.0989	1.0837
	FBA	0.0012	0.0008	0.0045	0.0067	0.0092	1.0308	1.0578	1.0858	1.0499	1.0198

Table C.2: Calculated autocorrelation function values and Gelman-Rubin scale reduction factors for the posteriors generated on the TP time series, the LVCF time series and FBA of the 9 South African bonds' credit spreads in Section 5.2, using OUVLTM and OU.

D. Fixed Income Market in South Africa

The fixed income or debt market in South Africa is not as sophisticated and well traded as developed markets, and exhibits a few nuances. This section gives a brief overview of areas particular to the SA market, as well as the liquidity measures and calculation of credit spreads mentioned in Chapter 2.

D.1 Market Composition

Debt instruments are listed on the Interest Rate Market on the JSE and trade on a platform called Yield-X. Some of the main listed instruments are:

- Fixed coupon bonds (pure vanilla and others e.g. callable, broken coupons)
- Inflation linked bonds
- Floating rate notes
- Amortising bonds
- Bond futures and options

Some of the more exotic bonds (such as convertibles) are also listed on the main (equity) board of the exchange. There is also a large OTC market, where swaps, FRAs and exotic bonds (equity linked notes and structured products) are traded. Collaterilised bonds are also prevalent, where the bond repayment is guaranteed by some other entity or assets. As in most markets, debt instruments are issued in a capital tier structure, with the order of repayment in the event of default specified in the bond contract. For example, senior debt will be repaid before subordinated debt.

D.2 Yield Curves

A yield curve is a term structure of interest rates, derived by some type of interpolation or bootstrapping technique of the yields of identical bonds of differing term. The two most utilised types of yield curves are the government zero coupon yield curve and the swap curve. Both are released by the JSE on a daily basis, using the closing yields of the chosen instruments to construct them.

The government zero coupon yield curve is generated by stripping out the coupons of a set of government issued bonds and can be used to determine the price of a (risk-free) bond by discounting the cash flows of the bond down the curve and taking the some of these.

D.3 Bonds

A fixed coupon bond is a loan from the issuer to the investor. The issuer pays interest or coupons (usually semi-annually in SA on pre-specified dates) at pre-determined rates of interest (coupon rates) to the investor. The maturity of the bond is when the bond expires, and on that date the last coupon and the original amount (called the principal/nominal/face value) borrowed by the issuer are paid to the investor.

The majority of South African bonds are yield traded, meaning that the market is made on yields instead of actual bond prices and bid/offer spreads are quoted on yields.

Some bonds have split maturities, where tranches of the total bonds in issue expire at different consecutive times. Usually only seen in government bonds, the split maturities occur once over 3 calendar years. For split maturities, the first maturity (called the pricing redemption date) is used as the maturity for pricing purposes prior to any maturity date being reached.

The date on which the trade will be settled is called the settlement date. The South African listed debt market follows a t + 3 day convention. South Africa also follows the Modified Following Day Count convention, where if a cash flow date (coupon payment date) does not fall on a business day, the modified following day for that date is the next business day. If this business day is in the following month, the business day preceding the cash flow date is used (Investopedia).

D.3.1 Pricing

D.3.1.1 Compounding

The time value of money means that an investment today grows with time by earning interest (if the interest is a positive rate). Thus a cash flow in the future is worth more than its present value. Determining the future cash flow (FV) can be done by compounding the interest on the present value (PV) to the future specified date (called future-valuing). Taking a future cash flow, one can determine the present cash flow value by discounting the future value by the interest.

Interest can be compounded in different ways (where r is the interest rate and t is the future date):

- Simple: $FV = PV(1 + rt)$
- Quarterly: $FV = PV\left(1 + {}^{r}/_{4}\right)^{4t}$
- Semi-annual: $FV = PV\left(1 + {}^{r}/_{2}\right)^{2t}$
- m times per year: $FV = PV\left(1 + {}^{r}/_{m}\right)^{mt}$
- Continuously: $FV = PV\exp(rt)$ (limit as m approaches ∞)

D.3.1.2 Calculation

Coupon payments dates of a bond are deduced by iteratively counting backwards n periods from the maturity date, where n is 6 months for semi-annual coupons, 4 months for quarterly coupons etc. So the last coupon payment date would have been the maturity date, the second last coupon payment date would have been n periods before the maturity, the third last coupon payment would have been n periods before the second last payment and so forth.

The terms from the trade date to each of these coupon payment dates were determined (in years) by

$$T_i = \frac{t_i - t_0}{365} \text{ for } i = 1, \ldots, n,$$

where T_i is the term corresponding to the i^{th} coupon payment, t_i is the time that the i^{th} coupon payment occurs, t_0 is the trade date and n is the total number of coupon payments that occur.

The i^{th} cash flow amount, denoted by $C_i, i = 1, \ldots, n$ is determined as follows:

$$C_i = \begin{cases} \frac{\text{(nominal)(coupon rate)}}{\text{coupon frequency}} & i \leq n, \\ \frac{\text{(nominal)(coupon rate)}}{\text{coupon frequency}} + \text{nominal} & i = n \end{cases} \quad \text{for } i = 1, \ldots, n.$$

The zero rates corresponding to the coupon payments dates are interpolated from the yield curve in Figure D.1 at their term points.

The discounted cash flows, $\hat{C}_i, i = 1, \ldots, n$, are given by

$$\hat{C}_i = C_i \mathrm{e}^{-r_i T_i} \text{ for } i = 1, \ldots, n,$$

where r_i is zero rate corresponding to the i^{th} coupon payment.

The sum of the discounted cash flows is the all-in price of the bond, $P(t_0)$ on the trade date i.e.

$$P(t_0) = \sum_{i=1}^{n} \hat{C}_i.$$

To determine the all-in price of the bond, $P(t_s)$ on the settlement date, t_s, all-in price, $P(t_0)$, determined at the trade date, t_0, is future-valued using the following formula:

$$P(t_s) = P(t_0) e^{r_s t_s}. \tag{D.1}$$

The yield-to-maturity (YTM) of a bond is the expected rate of return at any point in time if the bond is held to maturity. It is determined using a root-finding technique to solve for the following non-linear equation:

$$P(t_s) = \left(\sum_{i=1}^{n} C_i e^{-y T_i} \right) e^{r_s T_s}. \tag{D.2}$$

Thus if the YTM is known, the bond price is equal to the sum of the cash flows discounted by this rate. Typically, the bond is issued at par i.e. YTM = coupon rate, thereafter YTM is determined by the market. If the yield is greater than the coupon, the bond is said to be trading at a discount. If the yield is less than the coupon, the bond is trading at a premium.

D.4 Liquidity Measures

D.4.1 Trade Frequency Ratio (TFR)

A reasonable starting point in measuring the trade activity of a bond is to look at the number of days that trades for that bond took place during the year period under investigation. The ratio of the number of trades during the period (excluding trades that occurred on the issue date of the bond as this constitutes the initial placement of the bond in the primary market) over the number of days the bond was live during the period, called the trade frequency ratio, gives a normalised measure of how often that bond traded. The trade frequency ratio is useful in comparison across bonds; a ratio of 100% means that the bond traded everyday of its lifetime and so could be said to have higher liquidity than a bond with a ratio of, say, 50%.

As a demonstration of how the trade frequency ratio is calculated, an example on a bond that is traded on the JSE will be performed. The required details needed to determine the measures of liquidity of the bond to be used in this example, NBK2A, are given in Table D.1.

Number of days live in period	1,505 days
Number of trades during period	900 days
Total nominal traded during period	ZAR20,253,757,291.00

Table D.1: Details of NBK2A: number of days the bond was live during the period (January 2005 to March 2018), number of trades of the bond that took place during the period and total nominal amount of the bond traded during the period.

The trade frequency ratio is then calculated as follows:

$$\text{Trade frequency ratio} = \frac{\text{Number of trades during period}}{\text{Number of days alive during period}} = \frac{\text{900 days}}{\text{1,505 days}} = 59.80\%.$$

D.4.2 Liquidation Measure

The trade frequency ratio is telling as to the frequency of the bond's trade, but gives no indication as to the depth or size of trades. So two bonds that have the same trade frequency ratio might not necessarily be of the same liquidity, as the one might have had a greater depth of trade. A more robust measure hence must take the volume of trades into account.

One typical method that is used to measure liquidity in equity markets is detailed in the following steps:

1. Determine the total value traded, or TVT, in the security over the historical period (excluding the amount traded on the issue date).
2. Calculate the daily average value traded (by dividing the total value traded by the number of days in the period), given by
$$AVT = \frac{TVT}{n}$$
where n is the number of days in the period. Essentially this step calculates the value that would have been bought and sold each day if the security had been traded everyday to match the total value traded during the entire period.
3. Assume that a position of certain value[1], N, needs to be bought or sold in the security. Further assume that participation in daily market trading is capped at a level given by C[2].
4. Calculate the position value as a percentage of the daily average value traded (determined in Step 2), weighted by the maximum participation in daily market trading C or
$$\frac{N}{C \times AVT}.$$

The ratio calculated in the final step is the expected number of days it would take to buy or sell the position stipulated in Step 3, based on historical trading volumes. The more days it is expected to take to buy or sell the position, the less liquid the security is expected to be.

[1] A nominal size of ZAR10,000,000 (ZAR10mn) is assumed in the case for corporate bonds. The choice of this nominal amount is arbitrary, as the resulting liquidity estimate is only used comparatively with other bonds and not in absolute terms. However, a trade notional of ZAR1mn is reasonable for the South African bond market.

[2] A value of 30% is assumed here.

This quantity will be called the liquidation measure and is a reflection of the traded volume of the security, normalised by the number of days the security has been in existence.

The Nedbank bond shown in the previous example will again be used to illustrate the manner in which the liquidation measure is calculated.

From Table D.1, the daily average nominal traded for NBK2A is

$$\begin{aligned}\text{Daily average nominal traded} &= \frac{\text{Total nominal traded during period}}{\text{Number of days live during period}} \\ &= \frac{\text{ZAR20,235,757,297.00}}{\text{1,505 days}} \\ &= \text{ZAR13,445,685.91 per day}\end{aligned}$$

The liquidation measure is obtained as follows

$$\begin{aligned}\text{Liquidation measure} &= \frac{\text{Position to be liquidated}}{(\text{Daily average nominal traded})(\text{Maximum daily market participation})} \\ &= \frac{\text{ZAR10,000,000}}{(\text{ZAR13,445,685.91 per day})(30\%)} = 2.48 \text{ days}\end{aligned}$$

which is rounded up to 3 days.

Bonds that experienced no zero trade throughout the period were not assigned liquidation measure values as their AVT would have been zero, resulting in a zero in the denominator in Step 4.

D.4.3 Zero Trade Ratio (ZTR)

Another approach in considering the liquidity of an issuer as a whole would be to determine the number of issuances that did not trade at all out of the total number of issuances of the issuer. So, for example, if an issuer had five bonds in issuance and two experienced zero trade during the period then:

$$\text{Zero trade ratio} = \frac{\text{Number of bonds with zero trade}}{\text{Total number of bonds in issuance}} = \frac{2}{5} = 40\%.$$

A zero trade ratio of 0% means that all the issuances issued by an entity were traded at least once during the period and a zero trade ratio of 100% means that none of the issuances traded. Thus a lower ratio implies that more issuances were traded and the issuer was more liquid.

The measures of liquidity introduced in this section will be able to quantitatively estimate how often bonds traded and how large these trades were through time. These metrics will be used together in assessing the degree of liquidity of the bonds under investigation.

D.5 Credit Spread Calculation Methods

Another type of spread that is commonly employed in foreign debt markets is the option-adjusted spread, which factors in the possibility of a change in cashflows of the bond due to embedded options (known as

callable or puttable bonds). Since the empirical investigation is being limited to pure vanilla bonds, the need to account for optionality is unnecessary.

D.5.1 The Quoted Spread

The quoted spread does not require any complex calculations to determine; it is simply the difference between two bonds' yields. The majority of bonds in the South African listed debt market are quoted and traded in terms of yield, not all-in price (yield is the market traded quantity). These yields are provided by the JSE for all pure vanilla government and corporate bonds on a daily basis. The JSE will release quoted spreads for any bond that has a designated companion bond on a daily basis, which is simply the difference between the corporate and companion government bond yields.

D.5.2 The Z Spread

The method to determine the Z spread of a bond on any day follows the same structure as valuing a bond when discounting its cashflows using a daily government curve. The steps of the algorithm are given below.

1. Determine the settlement date corresponding to the trade date of the bond.
2. Determine the payment dates and amounts of future cash flows of the bond.
3. From the government zero coupon bond yield curve, interpolate
   (a) the zero rate corresponding to the settlement date determined in Step 1.
   (b) the zero rates corresponding to the bond cash flow payment dates identified in Step 2.
4. Add an estimate of the credit spread [3] to each zero rate calculated in Step 3.
5. Discount each cash flow to the trade date using the corresponding zero rate + credit spread guess.
6. Determine the all-in price of the bond by calculating the future value of the sum of the discounted cashflows obtained in Step 5 to the settlement date, using the corresponding zero rate determined in Step 3(a).
7. If the all-in price obtained in Step 6 does not match the market quoted all-in price of the bond, then adjust the estimated credit spread.
8. Iterate Steps 4 to 7 until the calculated all-in price equals (within a certain error of tolerance) the market quoted all-in price.

A numerical example of the above algorithm is calculated for the ABSA bond, ABS3, on 12 May 2009. Table D.2 shows the bond specifications and all-in price on 12 May 2009 for ABS3 needed to determine the cash flow payment dates and amounts. The settlement date of 15 May 2009, three business days after the trade date of 12 May 2009, as well as the term to the settlement date from the trade date in years, are given in this table. Lastly, the table shows the zero rate corresponding to the settlement from the trade date of 8.16%, as interpolated (using the cubic spline method) at the 0.0082 year point on the government zero coupon yield curve, shown in Figure D.1.

[3] A reasonable first guess would be the quoted spread on that day.

Maturity	03 April 2011
Coupon rate (%)	8.45
Coupon frequency	2
All-in price (ZAR)	101.63
Settlement date	15 May 2009
Term to settlement date T_s (years)	0.0082
Zero rate to settlement date r_s (%)	8.16

Table D.2: Maturity, coupon rate and frequency and all-in price of ABS3, as well as the settlement date and corresponding zero rate, as interpolated from the yield curve, on 12 May 2009.

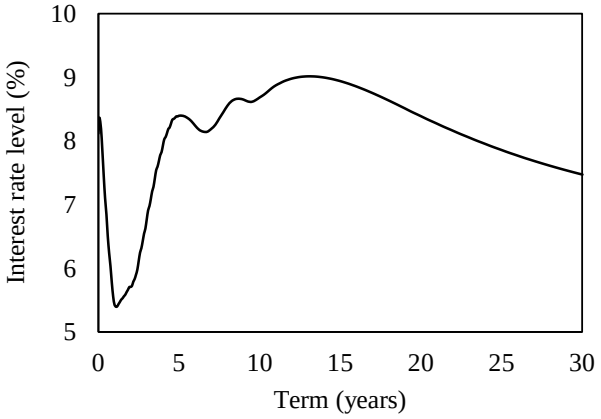


Figure D.1: Government zero coupon yield curve (NACC) as at 12 May 2009.

Since ABS3 matured on 3 April 2011 and paid coupons semi-annually, the coupon payments dates were deduced using the method described in Section D.3 and are shown in Table D.3. Also given are the numerical values calculated for cash flow payment dates and amounts (in ZAR using a nominal of ZAR100), the terms to each of the cash flows (in years), the zero rates to each cash flow payment date, the cash flow payments (in ZAR) discounted to the trade date and the all-in price (the sum of the discounted cash flows), for ABS3 as at 12 May 2009. The values in this table are the first set of values i.e. the first iteration in the algorithm.

Cash flow payment date	Time to cash flow payment (years)	Cash flow payment amount (ZAR)	Zero rate (%)	Discounted cash flows (ZAR)
05 Oct 09	0.4	4.225	7.17	4.11
06 Apr 10	0.9	4.225	5.60	4.02
04 Oct 10	1.4	4.225	5.50	3.91
04 Apr 11	1.9	104.225	5.70	93.56
			Sum	105.59

Table D.3: The first iteration of the algorithm: cash flow payment dates and amounts (assuming a nominal of ZAR100), the terms to each of the cash flows, the zero rates to each cash flow payment date, the discounted cash flow payments and the sum of the discounted cash flows for ABS3, on 12 May 2009.

The all-in price of the bond is determined by future-valuing the sum of the discounted cash flows in Table D.3 using (D.1) to get

$$P(t_s) = (105.59)e^{(0.0816)(0.0082)} = 105.66, \tag{D.3}$$

where r_s and t_s are from Table D.2. The all-in price of 105.66 does not equal the market all-in price on 12 May 2009 of 101.63 (shown in Table D.2). A constant amount (the Z spread) is thus added to each of the zero rates $r_i, i = 1, \ldots, n$ and r_s, the discounted cash flows then recalculated, summed and future valued to the settlement date and the result compared with the market all-in price. If the resultant all-in price does not equal the market all-in price, the amount added to the zero rates is adjusted, using a root-finding technique, and the all-in price is recalculated and again compared with the market all-in, until the calculated

price and market price are within a small error of tolerance of each other.

Cash flow payment date	Time to cash flow payment (years)	Cash flow payment amount (ZAR)	Zero rate (%)	Discounted cash flows (ZAR)
05 Oct 09	0.4	4.225	9.35	4.07
06 Apr 10	0.9	4.225	7.78	3.94
04 Oct 10	1.4	4.225	7.68	3.79
04 Apr 11	1.9	104.225	7.88	89.76
			Sum	101.56

Table D.4: The final iteration of the algorithm; the constant amount or Z spread added to the zero rates which matched the calculated all-in to the market all-in was 2.18% NACC.

The final iteration of this process for ABS3 is shown in Table D.4. A constant value of 2.18% (in NACC, when converted to NACS becomes 2.20%) added to the zero rates resulted in the calculated price at the settlement date, given by

$$P(t_s) = (101.56)e^{(0.0816)(0.0082)} = 101.63,$$

matching the market all-in price.

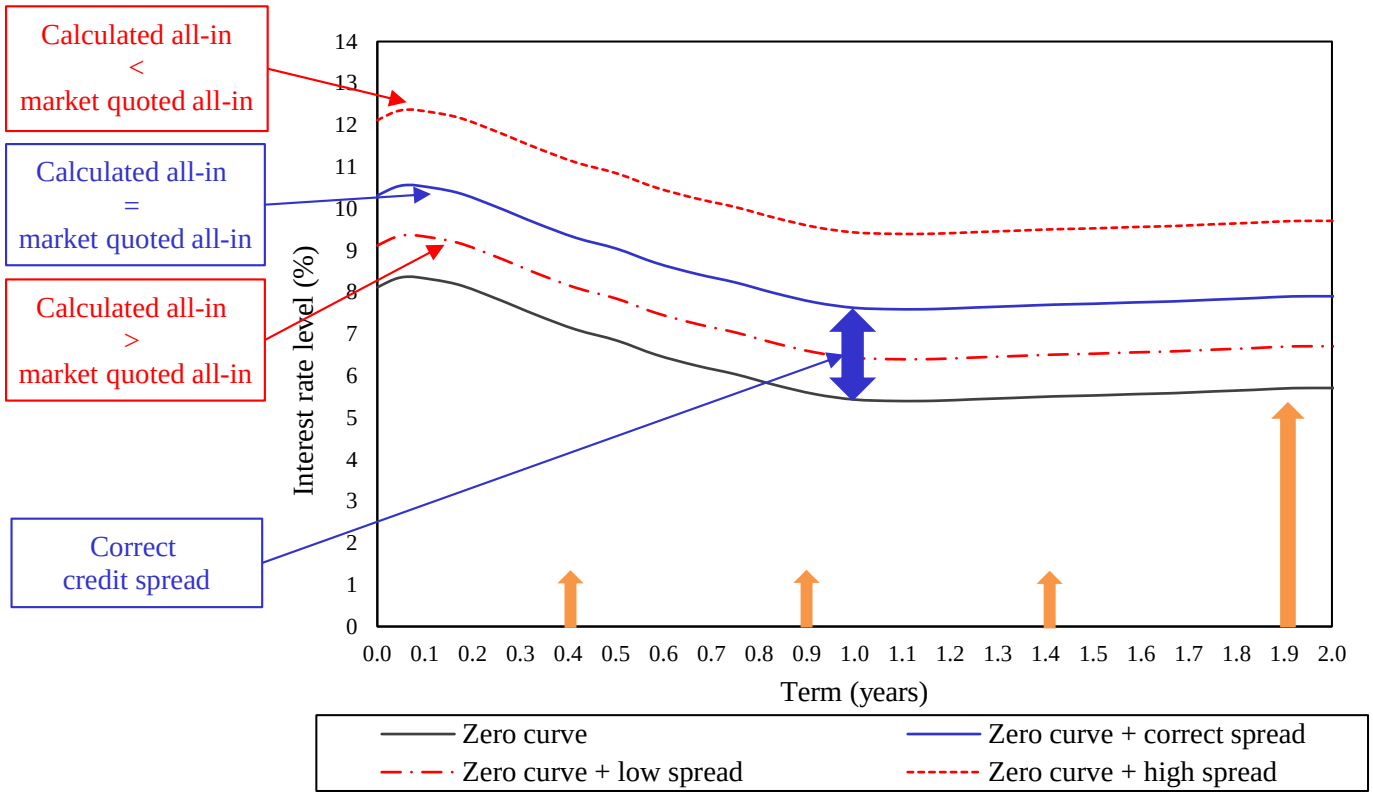


Figure D.2: Graphical representation of the algorithm to determine the Z spread of ABS3 on 12 May 2009. The orange arrows indicate the cash flow payments of the bond, at the term on which they occur (not to scale or representative of the actual size of the coupon payments). The spread above the government zero curve in black is adjusted (examples of which are the red curves) until the 'correct' spread is found which matches the calculated all-in to the market all-in (blue curve).

Figure D.2 shows a graphical representation of the Z spread calculation algorithm for ABS3 on 12 May 2009. The cash flows occurring in 0.4 years, 0.9 years, 1.4 years and 1.9 years are indicated by the orange

arrows. These cashflows are discounted down the different shifted curves (the red curves) and the sum of the cashflows (future valued to the settlement date) are compared with the market price, until the 'correct' spread is found which matches the discounted cash flows to the market price.

Essentially the algorithm perturbs the government zero coupon yield curve linearly until it gives a price for the bond that matches the market quoted price.

D.5.3 The Nominal Spread

The nominal spread for a corporate bond is determined as follows (the first three steps are identical to those of the Z spread algorithm):

1. Determine the settlement date corresponding to the trade date of the bond.
2. Determine the payment dates and amounts of future cash flows of the bond.
3. From the government zero coupon bond yield curve, interpolate
   (a) the zero rate corresponding to the settlement date determined in Step 1.
   (b) the zero rates corresponding to the bond cash flow payment dates identified in Step 2.
4. Discount each cash flow to the trade date using the corresponding zero rate determined in Step 3(b).
5. Determine the risk-free all-in price of the bond by calculating the future value of the sum of the discounted cashflows obtained in Step 4 to the settlement date, using the corresponding zero rate determined in Step 3(a).
6. Assume that the yield of the risk free bond is at a certain level.
7. Discount each cash flow to the trade date using the yield guessed in Step 6.
8. Future value the sum of the discounted cashflows obtained in Step 7 to the settlement date using the corresponding zero rate calculated in Step 3(a).
9. If the all-in price obtained in Step 8 does not match (within a certain error of tolerance) the risk-free all-in price of the bond determined in Step 5, then adjust the yield.
10. Iterate Steps 6 to 9 until the calculated all-in price equals the risk-free all-in price of the bond.
11. The calculated spread is the difference in the yield of the corporate bond and the risk-free yield determined in Step 10.

In essence, the nominal spread is the theoretically correct version of the quoted spread; pricing a theoretical riskless bond with the same duration and coupon schedule as the corporate bond in order to see how wide the spread is between their yields.

An example calculating the nominal spread for ABS3 on 12 May 2009 will be used to demonstrate the above algorithm. The information given in Tables D.2 and D.3 from the Z spread calculation example are relevant in the calculation of the nominal spread, as determining the risk-free all-price follows the same procedure as that of the first iteration of the Z spread algorithm.

(D.3) shows the resultant all-in calculated by discounting the cash flows down the government zero coupon yield curve (105.59) and is the hypothetical riskless version of ABS3.

The yield-to-maturity of this riskless ABS3 bond is calculated by solving for y in D.2:

$$105.59 = 4.225e^{-0.4y} + 4.225e^{-0.9y} + 4.225e^{-1.4y} + 104.225e^{-1.9y}.$$

The result is a yield-to-maturity of 5.78% (NACS). The market quoted yield-to-maturity of ABS3 on 12 May 2009 was 8.06% (NACS). Thus,

$$\text{Nominal spread(ABS3)} = y(\text{ABS3}) - y(\text{riskless ABS3}) = 8.06\% - 5.78\% = 2.27\%.$$

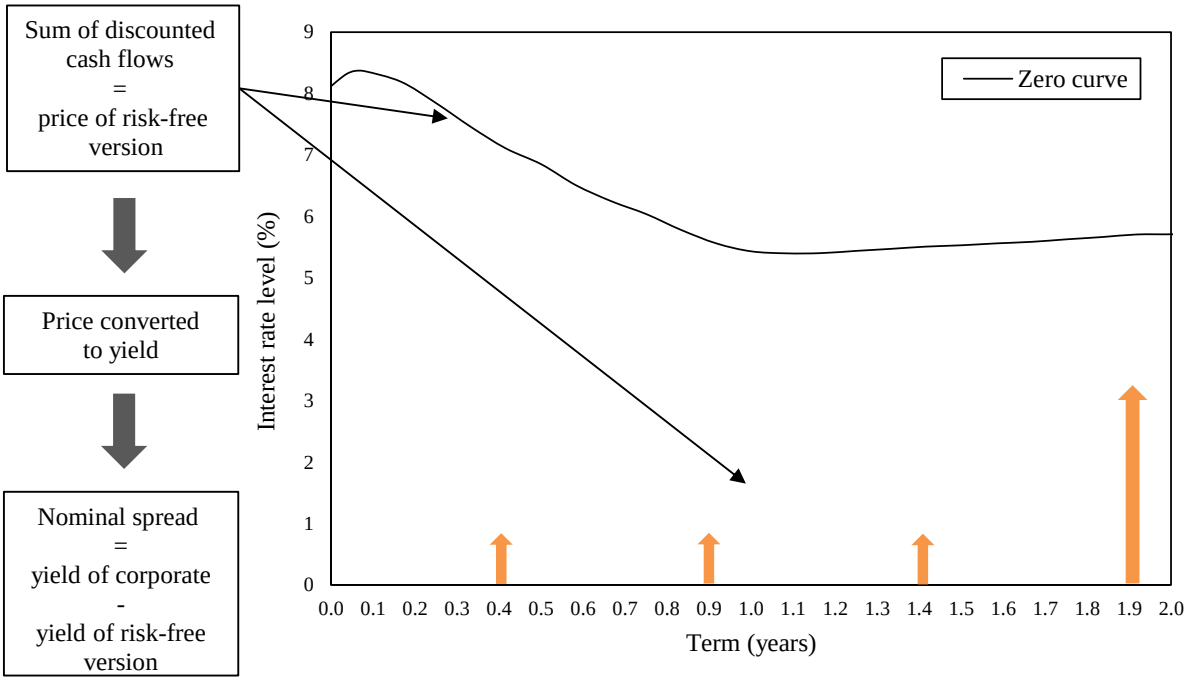


Figure D.3: Graphical representation of the algorithm to determine the nominal spread of ABS3 on 12 May 2009. The orange arrows indicate the cash flow payments of the bond, at the term on which they occur (not to scale or representative of the actual size of the coupon payments). The government zero curve is in black. The discounted cash flows using the government zero curve are summed to get the all-in price of the risk-free bond. The yield is determined, using a root-finding technique, and the nominal spread calculated as the difference between the bond's yield and this risk-free yield.

A graphical representation of the manner in which the nominal spread is determined is shown in Figure D.3. ABS3 on 12 May 2009 is again used. The cashflows are discounted down the government zero curve (in black) and the allin-price of the risk-free version is the sum of the cashflows (future valued to the settlement date). The yield of the risk-free version of the bond is obtained via an optimisation method and the nominal spread is the difference between the yields of the bond and its risk-less version.

D.6 List of Pure Vanilla Corporate Bonds

Bond	Issuer	Issue date	Maturity	TFR (%)	LM (days)
MTN01	MOBILE TELEPHONE NETWORKS HOLDINGS LTD	13-Jul-2006	13-Jul-2010	62.64	1
NBK2A	NEDBANK LTD	09-Sep-2009	15-Sep-2015	59.80	3
BID01	BIDVESTCO LTD	06-Aug-2007	06-Aug-2014	55.71	6
SBS1	STANDARD BANK OF SOUTH AFRICA LTD	24-Nov-2004	24-May-2010	55.68	2
MTN02	MOBILE TELEPHONE NETWORKS HOLDINGS LTD	13-Jul-2006	13-Jul-2014	53.98	3
SBS9	STANDARD BANK OF SOUTH AFRICA LTD	07-Jul-2009	07-Jul-2016	51.94	5
IPL4	IMPERIAL GROUP LTD	29-Mar-2007	29-Mar-2014	49.40	16
IPL6	IMPERIAL GROUP LTD	28-Sep-2010	28-Sep-2017	45.95	4
NBK3A	NEDBANK LTD	09-Sep-2009	09-Sep-2019	44.89	9
ABS3	ABSA BANK LTD	03-Oct-2005	03-Apr-2011	43.14	2
ABL8A	AFRICAN BANK LTD	19-Sep-2008	19-Sep-2013	42.53	20
NBK9A	NEDBANK LTD	23-Mar-2011	23-Mar-2016	42.00	9
CBL11	CAPITEC BANK	06-May-2011	06-May-2016	41.92	11
LGL02	THE LIBERTY GROUP LTD	13-Aug-2012	13-Aug-2017	41.84	7
FRX15	FIRSTRAND BANK LTD	14-Mar-2008	14-Mar-2015	41.40	5
SBS25	STANDARD BANK OF SOUTH AFRICA LTD	24-May-2012	24-May-2019	40.61	5
ABS5	ABSA BANK LTD	23-Apr-2007	01-May-2015	40.43	7
MTN04	MOBILE TELEPHONE NETWORKS HOLDINGS LTD	13-Jul-2010	13-Jul-2017	38.93	20
BEER01	SABSA HOLDINGS (PTY) LTD	19-Jul-2007	19-Jul-2012	38.50	4
FRX19	FIRSTRAND BANK LTD	04-Apr-2012	15-Nov-2019	38.08	7
ABS10	ABSA BANK LTD	17-Mar-2011	17-Mar-2018	37.49	11
DC02	DAIMLERCHRYSLER SA (PTY) LTD	02-Apr-2003	02-Oct-2008	36.10	1
BAW1	BARLOWORLD LTD	29-Jul-2004	29-Jul-2011	34.49	5
DC03	DAIMLERCHRYSLER SA (PTY) LTD	26-May-2006	26-May-2011	32.77	7
NBK16A	NEDBANK LTD	12-Feb-2015	12-Feb-2025	31.38	4
ABL2	AFRICAN BANK LTD	11-Sep-2003	18-Sep-2006	30.37	4
LGL04	THE LIBERTY GROUP LTD	14-Aug-2013	14-Aug-2020	30.05	13
IBL46	INVESTEC BANK LTD	15-Nov-2012	15-Jan-2020	29.43	27
ABS6	ABSA BANK LTD	18-Jun-2007	01-Jun-2020	29.27	6
CBL16	CAPITEC BANK	18-May-2012	18-May-2017	29.02	19
UTR01	UNITRANS SERVICES (PTY) LTD	17-May-2005	31-Aug-2010	28.60	13
DC01	DAIMLERCHRYSLER SA (PTY) LTD	25-Sep-2001	25-Sep-2007	28.26	3
ABL3	AFRICAN BANK LTD	12-Jul-2004	12-Jul-2007	27.22	8
MBSA01	MERCEDES-BENZ SOUTH AFRICA LTD	16-Apr-2012	16-Apr-2019	27.10	71
AA05	ANGLO AMERICAN SA FINANCE LTD	22-Mar-2012	22-Mar-2019	26.63	3
SMF2	SAPPI MANUFACTURING PTY LTD	25-Sep-2007	14-Oct-2011	25.89	5
ABL11A	AFRICAN BANK LTD	29-Sep-2010	29-Sep-2014	25.80	29
IPL3	IMPERIAL GROUP LTD	02-Dec-2003	30-Nov-2010	25.34	3
HAR1	HARMONY GOLD MINING COMPANY	08-Jun-2001	14-Jun-2006	25.14	3
SHS05	STEINHOFF SERVICES LTD	29-Jun-2012	29-Jun-2017	24.88	17
FRX26	FIRSTRAND BANK LTD	10-Feb-2014	01-Oct-2026	24.83	5
SPG1	SUPER GROUP LTD	25-Jun-2004	25-Jun-2008	24.63	4
E157	GROWTHPOINT NOTE ISSUER COMPANY (PTY) LTD	05-Oct-1982	01-Nov-2008	24.48	29
BAW2	BARLOWORLD LTD	02-Oct-2008	02-Oct-2015	24.40	15
MQB02	MACQUARIE SECURITIES SOUTH AFRICA LTD	31-May-2012	31-May-2017	24.14	36
IV01	INVESTEC BANK LTD	17-Jun-2000	31-Mar-2012	23.17	6
IPL1	IMPERIAL GROUP LTD	30-Aug-2001	14-Mar-2006	22.92	2
SBS2	STANDARD BANK OF SOUTH AFRICA LTD	18-Apr-2005	18-Apr-2008	21.65	10
MTN03	MOBILE TELEPHONE NETWORKS HOLDINGS LTD	13-Jul-2010	13-Jul-2015	21.44	5
SFL2	SASOL FINANCING (PTY) LTD	01-Sep-2003	01-Sep-2007	20.99	15
SMF1	SAPPI SOUTHERN AFRICA (PTY) LTD	27-Jun-2006	27-Jun-2013	19.91	26
ABL4	AFRICAN BANK LTD	23-Aug-2005	31-Aug-2010	19.81	7
GFC1	GROUP FIVE CONSTRUCTION (PTY) LTD	27-Feb-2007	27-Feb-2010	19.71	7
BID06	BIDVESTCO LTD	30-Jun-2014	30-Jun-2019	19.43	6
BEER02	SABSA HOLDINGS (PTY) LTD	28-Mar-2013	28-Mar-2018	18.71	9
AG01	ANGLO GOLD LTD	28-Aug-2003	28-Aug-2008	18.69	12
ABL7	AFRICAN BANK LTD	18-Feb-2008	18-Feb-2013	17.89	6
SHF01	STEINHOFF MANUFACTURING (PTY) LTD	08-Dec-2003	28-Feb-2008	17.83	4
BAW19	BARLOWORLD LTD	05-Dec-2013	05-Dec-2020	17.72	12
SBS38	STANDARD BANK OF SOUTH AFRICA LTD	29-Jan-2015	29-Jan-2025	16.88	3
FRX30	FIRSTRAND BANK LTD	09-Mar-2015	31-Jan-2030	16.56	6
LGL03	THE LIBERTY GROUP LTD	03-Oct-2012	03-Apr-2018	16.24	8
ABS11	ABSA BANK LTD	09-Apr-2014	09-Apr-2021	16.10	21
SBS4	STANDARD BANK OF SOUTH AFRICA LTD	16-Nov-2006	16-Nov-2021	16.00	20

Bond	Issuer	Issue date	Maturity	TFR (%)	LM (days)
IPL10	IMPERIAL GROUP LTD	20-May-2014	20-May-2021	15.67	8
IPL2	IMPERIAL GROUP LTD	05-Feb-2002	28-Feb-2008	15.04	6
NBK21A	NEDBANK LTD	21-Jul-2015	21-Jul-2027	14.64	56
ABS7	ABSA BANK LTD	11-Sep-2007	11-Sep-2026	14.25	7
NBK17A	NEDBANK LTD	22-Apr-2015	22-Apr-2026	14.25	19
FRX31	FIRSTRAND BANK LTD	21-Feb-2011	21-Feb-2031	14.08	9
SBS27	STANDARD BANK OF SOUTH AFRICA LTD	20-Feb-2014	20-Feb-2024	13.83	6
NBK11A	NEDBANK LTD	28-Nov-2013	28-Nov-2020	13.57	7
ABS9	ABSA BANK LTD	17-Mar-2011	17-Mar-2016	13.52	8
AA03	ANGLO AMERICAN SA FINANCE LTD	11-May-2010	11-May-2015	12.88	14
NBK6A	NEDBANK LTD	19-Apr-2010	19-Apr-2015	12.55	19
SBS20	STANDARD BANK OF SOUTH AFRICA LTD	15-May-2012	15-May-2026	12.50	8
TFS06	TOYOTA FINANCIAL SERVICES (SA) LTD	20-Mar-2007	20-Mar-2014	12.33	19
SBS34	STANDARD BANK OF SOUTH AFRICA LTD	19-Sep-2014	19-Sep-2024	12.12	10
SBS5	STANDARD BANK OF SOUTH AFRICA LTD	07-Dec-2006	07-Dec-2011	11.98	27
LGL06	THE LIBERTY GROUP LTD	04-Oct-2016	04-Oct-2022	11.80	8
NBK20A	NEDBANK LTD	01-Jun-2015	01-Jun-2026	11.53	8
UTR02	UNITRANS SERVICES (PTY) LTD	21-Nov-2007	21-Nov-2012	11.42	2
SBS19	STANDARD BANK OF SOUTH AFRICA LTD	23-Jun-2011	23-Jun-2021	10.50	13
FRX23	FIRSTRAND BANK LTD	22-Apr-2013	28-Feb-2023	10.35	9
SBS39	STANDARD BANK OF SOUTH AFRICA LTD	29-Jan-2015	29-Jan-2030	10.20	3
NBK30A	NEDBANK LTD	20-Feb-2017	20-Feb-2024	10.11	25
FRX24	FIRSTRAND BANK LTD	10-Dec-2009	10-Dec-2024	9.97	12
SHS25	STEINHOFF SERVICES LTD	29-Jun-2015	29-Jun-2020	9.10	368
SBS42	STANDARD BANK OF SOUTH AFRICA LTD	12-Nov-2015	12-Nov-2025	9.06	16
FRX32	FIRSTRAND BANK LTD	07-Mar-2017	31-Mar-2032	9.02	9
FRX18	FIRSTRAND BANK LTD	14-Apr-2010	14-Apr-2018	8.93	12
ABL6	AFRICAN BANK LTD	18-Jun-2007	18-Jun-2012	8.79	16
IBL49	INVESTEC BANK LTD	02-Apr-2013	02-Apr-2018	8.71	16
ABL5	AFRICAN BANK LTD	11-Aug-2006	11-Aug-2011	8.55	28
ABS17	ABSA BANK LTD	11-Nov-2015	11-Nov-2027	8.54	13
SBS56	STANDARD BANK OF SOUTH AFRICA LTD	12-Jun-2017	12-Jun-2022	8.42	5
ABL10A	AFRICAN BANK LTD	15-Mar-2010	15-Mar-2015	8.40	20
NBK29A	NEDBANK LTD	02-Aug-2016	31-Jul-2026	8.39	53
GFC2	GROUP FIVE CONSTRUCTION (PTY) LTD	27-Feb-2007	27-Feb-2012	8.31	32
SBS31	STANDARD BANK OF SOUTH AFRICA LTD	12-Jun-2014	12-Jun-2027	8.29	5
FRX27	FIRSTRAND BANK LTD	07-Mar-2017	07-Mar-2027	7.89	8
SSA01	SAPPI SOUTHERN AFRICA (PTY) LTD	28-Jun-2011	28-Jun-2016	7.75	47
NBK24A	NEDBANK LTD	19-Nov-2015	19-Nov-2027	7.61	23
SBK6	STANDARD BANK OF SOUTH AFRICA LTD	25-Feb-2005	01-Mar-2007	7.55	7
FRX20	FIRSTRAND BANK LTD	01-Oct-2013	01-Oct-2020	7.47	50
SBS13	STANDARD BANK OF SOUTH AFRICA LTD	22-Sep-2010	22-Sep-2017	7.36	30
BAW11	BARLOWORLD LTD	14-Jun-2011	01-Oct-2018	6.94	53
OMO1	UNILEVER SOUTH AFRICA (PTY) LTD	12-Sep-2003	12-Sep-2008	6.91	25
NBK27A	NEDBANK LTD	10-May-2016	10-May-2026	6.09	125
AA07	ANGLO AMERICAN SA FINANCE LTD	15-Apr-2014	15-Apr-2021	6.06	39
SBS41	STANDARD BANK OF SOUTH AFRICA LTD	12-Nov-2015	12-Nov-2022	6.04	71
MTN05	MOBILE TELEPHONE NETWORKS HOLDINGS LTD	28-Oct-2010	28-Oct-2013	5.99	33
HPF05	HOSPITALITY PROPERTY FUND LTD	17-Feb-2014	17-Feb-2017	5.98	48
ABK2	AFRICAN BANK LTD	04-Apr-2016	24-May-2018	5.80	14
BAW21	BARLOWORLD LTD	24-Mar-2015	24-Mar-2022	5.69	23
IBL87	INVESTEC BANK LTD	20-May-2016	20-May-2019	5.56	21
ABS16	ABSA BANK LTD	11-Nov-2015	11-Nov-2025	5.36	34
KAP007	KAP INDUSTRIAL HOLDINGS LTD	26-Oct-2016	26-Oct-2021	5.04	10
FRX17	FIRSTRAND BANK LTD	04-Apr-2012	15-Sep-2017	4.84	15
SBS7	STANDARD BANK OF SOUTH AFRICA LTD	26-Mar-2008	26-Mar-2013	4.79	212
SSN053	STANDARD BANK OF SOUTH AFRICA LTD	26-Oct-2017	26-Oct-2020	4.67	6
NBK22A	NEDBANK LTD	19-Nov-2015	19-Nov-2022	4.57	57
ABK1	AFRICAN BANK LTD	04-Apr-2016	07-Nov-2018	4.40	25
ABS12	ABSA BANK LTD	14-May-2015	14-May-2020	4.29	159
TFS84	TOYOTA FINANCIAL SERVICES (SA) LTD	28-Jul-2011	28-Jul-2016	4.16	33
IBL78	INVESTEC BANK LTD	25-Nov-2015	25-Nov-2022	4.09	609
FRX16	FIRSTRAND BANK LTD	04-Apr-2012	15-Sep-2016	4.05	31
MBSA02	MERCEDES-BENZ SOUTH AFRICA LTD	15-Apr-2014	15-Apr-2019	4.04	51
IBL54	INVESTEC BANK LTD	31-Jul-2013	31-Jul-2016	4.01	33
ABS13	ABSA BANK LTD	14-May-2015	14-May-2022	3.87	82
GRT17	GROWTHPOINT PROPERTIES LTD	17-Oct-2016	17-Oct-2023	3.85	177
NBK28A	NEDBANK LTD	02-Aug-2016	02-Aug-2023	3.84	20
NBK12A	NEDBANK LTD	19-Mar-2014	19-Mar-2021	3.67	65
CBL22	CAPITEC BANK	06-May-2013	06-May-2020	3.66	54
NBK14A	NEDBANK LTD	26-Jun-2014	25-Jun-2021	3.60	42

Bond	Issuer	Issue date	Maturity	TFR (%)	LM (days)
ABN15	ABSA BANK LTD	15-Oct-2007	15-Sep-2015	3.59	39
SMF3	SAPPI SOUTHERN AFRICA (PTY) LTD	30-Jun-2009	30-Jun-2012	3.59	36
ABS15	ABSA BANK LTD	11-Nov-2015	11-Nov-2022	3.52	36
NBK13A	NEDBANK LTD	19-Mar-2014	19-Mar-2024	3.47	75
CBL01	CAPITEC BANK	06-May-2008	06-May-2011	3.45	136
PSG01	PSG FINANCIAL SERVICES LTD	13-Oct-2006	13-Oct-2011	3.43	12
ABS14	ABSA BANK LTD	11-Nov-2015	11-Nov-2020	3.35	38
CPV01	CPV POWER PLANT NO 1 BOND SPV (RF) LTD	29-Apr-2013	30-Jun-2029	3.25	16
SSN017	STANDARD BANK OF SOUTH AFRICA LTD	30-Aug-2013	21-Dec-2026	3.23	11
IBL88	INVESTEC BANK LTD	20-May-2016	20-May-2021	3.21	12
NTC06	CLINDEB INVESTMENTS (PTY) LTD	17-Apr-2008	17-Apr-2011	3.06	50
CBL18	CAPITEC BANK	23-Aug-2012	23-Aug-2019	2.93	82
IBL45	INVESTEC BANK LTD	15-Nov-2012	21-Dec-2018	2.91	21
NTC01	CLINDEB INVESTMENTS (PTY) LTD	30-Nov-2006	28-Feb-2008	2.88	47
IPF06	INVESTEC PROPERTY FUND LTD	13-Apr-2012	13-Apr-2017	2.79	47
SBS6	STANDARD BANK OF SOUTH AFRICA LTD	26-Mar-2008	26-Mar-2011	2.79	25
CBL26	CAPITEC BANK	06-May-2016	06-May-2021	2.72	27
NBK23A	NEDBANK LTD	19-Nov-2015	19-Nov-2025	2.71	22
OMN01	OMNIA GROUP (PTY) LTD	28-Nov-2008	28-Nov-2011	2.53	56
IBL60	INVESTEC BANK LTD	17-Feb-2014	17-Feb-2024	2.43	60
NBK18A	NEDBANK LTD	01-Jun-2015	01-Jun-2020	2.39	56
IBL79	INVESTEC BANK LTD	27-Jan-2016	27-Jan-2019	2.39	22
SBS43	STANDARD BANK OF SOUTH AFRICA LTD	12-Nov-2015	12-Nov-2027	2.35	16
DSY03	DISCOVERY LTD	21-Nov-2017	21-Nov-2024	2.25	8
IBL28	INVESTEC BANK LTD	18-Aug-2011	18-Aug-2016	2.16	31
IBL50	INVESTEC BANK LTD	02-Aug-2013	02-Aug-2020	2.15	39
SBS37	STANDARD BANK OF SOUTH AFRICA LTD	29-Jan-2015	29-Jan-2020	2.02	5
CBL03	CAPITEC BANK	18-May-2009	18-May-2012	1.99	94
IBL39	INVESTEC BANK LTD	02-Apr-2012	02-Apr-2018	1.93	144
SSN021	STANDARD BANK OF SOUTH AFRICA LTD	18-Nov-2013	28-Feb-2031	1.92	58
IBL21	INVESTEC BANK LTD	15-Mar-2011	15-Mar-2016	1.92	86
ABN67	ABSA BANK LTD	22-Jan-2013	28-Feb-2023	1.85	13
GFC04	GROUP FIVE CONSTRUCTION (PTY) LTD	11-Apr-2012	11-Apr-2017	1.76	25
BAW8	BARLOWORLD LTD	15-Sep-2010	02-Oct-2017	1.70	151
IBL99	INVESTEC BANK LTD	21-Oct-2016	21-Oct-2019	1.67	699
SBS3	STANDARD BANK OF SOUTH AFRICA LTD	25-May-2006	25-May-2026	1.65	46
BID02	BIDVESTCO LTD	14-Jul-2009	14-Jul-2012	1.59	115
CBL12	CAPITEC BANK	06-Sep-2011	06-Sep-2018	1.52	24
NBK26A	NEDBANK LTD	10-May-2016	10-May-2023	1.47	40
SSN038	STANDARD BANK OF SOUTH AFRICA LTD	26-Jul-2016	26-Jan-2021	1.42	1534
SSN005	STANDARD BANK OF SOUTH AFRICA LTD	07-May-2012	31-Mar-2021	1.42	50
PPC004	PPC LTD	10-Jul-2014	30-Jun-2021	1.39	95
IB153B	INVESTEC BANK LTD	14-Jan-2008	31-Aug-2010	1.37	149
NBK25A	NEDBANK LTD	18-Feb-2016	17-Feb-2023	1.32	306
SHS19	STEINHOFF SERVICES LTD	13-Aug-2014	10-Sep-2017	1.30	72
SBN27	STANDARD BANK OF SOUTH AFRICA LTD	23-Aug-2007	15-Sep-2015	1.29	291
IBL55	INVESTEC BANK LTD	31-Jul-2013	31-Jul-2018	1.29	949
SBN21	STANDARD BANK OF SOUTH AFRICA LTD	27-Jun-2006	31-Aug-2010	1.24	204
SGL05	SUPER GROUP LTD	07-Apr-2008	07-Apr-2012	1.20	748
IBL01	INVESTEC BANK LTD	15-Jun-2009	15-Jun-2012	1.20	71
GRT21	GROWTHPOINT PROPERTIES LTD	23-Mar-2017	03-Apr-2024	1.18	33
IBL153	INVESTEC BANK LTD	28-Nov-2007	31-Aug-2010	1.16	136
SBS46	STANDARD BANK OF SOUTH AFRICA LTD	15-Feb-2016	15-Feb-2023	1.13	287
SBS50	STANDARD BANK OF SOUTH AFRICA LTD	31-Jan-2017	31-Jan-2022	1.03	27
SBS51	STANDARD BANK OF SOUTH AFRICA LTD	31-Jan-2017	31-Jan-2024	1.03	56
ABN27	ABSA BANK LTD	03-Apr-2008	31-Aug-2010	1.00	26
FRBN01	FIRSTRAND BANK LTD	04-Mar-2004	31-Aug-2010	0.99	29
BAW7	BARLOWORLD LTD	15-Sep-2010	02-Oct-2014	0.89	123
NBK19A	NEDBANK LTD	01-Jun-2015	01-Jun-2022	0.84	35
SBN38	STANDARD BANK OF SOUTH AFRICA LTD	03-Apr-2008	31-Aug-2010	0.83	89
NBK10A	NEDBANK LTD	25-Jul-2013	25-Jul-2016	0.80	133
ABN01	ABSA BANK LTD	15-Sep-2006	15-Sep-2011	0.80	203
ABS18	ABSA BANK LTD	26-Sep-2016	26-Sep-2023	0.79	80
CGR34	CALGRO M3 DEVELOPMENTS LTD	22-Sep-2017	21-Sep-2018	0.77	52
ABN79	ABSA BANK LTD	03-Jul-2013	21-Dec-2026	0.76	877
SBN40	STANDARD BANK OF SOUTH AFRICA LTD	21-Apr-2008	31-Aug-2010	0.68	82
BAW5	BARLOWORLD LTD	15-Sep-2010	15-Sep-2013	0.67	255
CBL14	CAPITEC BANK	01-Feb-2012	01-Feb-2019	0.58	97
EQS07	EQSTRA CORPORATION LTD	09-Apr-2013	09-Apr-2018	0.56	175
SSN041	STANDARD BANK OF SOUTH AFRICA LTD	13-Oct-2016	13-Oct-2023	0.55	456
KW01	KOMATI RIVER BASIN AUTHORITY	27-Oct-1997	31-Oct-2027	0.54	682

Bond	Issuer	Issue date	Maturity	TFR (%)	LM (days)
SSN009	STANDARD BANK OF SOUTH AFRICA LTD	02-Jul-2012	15-Sep-2017	0.54	357
IBL23	INVESTEC BANK LTD	21-Apr-2011	21-Apr-2014	0.54	304
ABN80	ABSA BANK LTD	04-Jul-2013	31-Mar-2021	0.51	88
CGR11	CALGRO M3 DEVELOPMENTS LTD	27-Mar-2012	28-Mar-2016	0.50	147
SBN39	STANDARD BANK OF SOUTH AFRICA LTD	03-Apr-2008	31-Aug-2010	0.50	84
IBL101	INVESTEC BANK LTD	24-May-2017	24-May-2022	0.47	280
ABN50	ABSA BANK LTD	10-May-2012	21-Dec-2014	0.46	42
ABN87	ABSA BANK LTD	29-Aug-2013	21-Dec-2026	0.44	539
SBS29	STANDARD BANK OF SOUTH AFRICA LTD	12-Jun-2014	12-Jun-2019	0.42	2642
EQS02	EQSTRA CORPORATION LTD	01-Jul-2010	01-Jul-2015	0.40	374
UTR40	STEINHOFF SERVICES LTD	10-Sep-2010	10-Sep-2017	0.40	112
IBL22	INVESTEC BANK LTD	15-Mar-2011	15-Mar-2018	0.40	712
IBL58	INVESTEC BANK LTD	13-Feb-2014	13-Feb-2017	0.40	48
SSN010	STANDARD BANK OF SOUTH AFRICA LTD	02-Jul-2012	15-Sep-2015	0.37	2057
SBN29	STANDARD BANK OF SOUTH AFRICA LTD	12-Sep-2007	15-Sep-2015	0.30	578
NBK15A	NEDBANK LTD	12-Feb-2015	11-Feb-2022	0.26	262
NBRN1	NEDBANK LTD	28-Jan-2010	31-Aug-2011	0.25	302
ABN76	ABSA BANK LTD	13-Jun-2013	21-Dec-2026	0.25	100
ABN56	ABSA BANK LTD	02-Jul-2012	15-Sep-2015	0.25	857
FRBN04	FIRSTRAND BANK LTD	15-Jun-2007	15-Sep-2015	0.24	469
FRX45	FIRSTRAND BANK LTD	14-Apr-2010	14-Apr-2045	0.20	141
RP018	REAL PEOPLE INVESTMENTS HOLDINGS (PTY) LTD	12-Feb-2013	12-Feb-2020	0.16	335
SSN012	STANDARD BANK OF SOUTH AFRICA LTD	15-Jan-2013	21-Dec-2018	0.15	8687
SBN31	STANDARD BANK OF SOUTH AFRICA LTD	18-Sep-2007	15-Sep-2015	0.15	732
SSN004	STANDARD BANK OF SOUTH AFRICA LTD	26-Apr-2012	31-Mar-2021	0.13	8104
SBN28	STANDARD BANK OF SOUTH AFRICA LTD	12-Sep-2007	31-Aug-2010	0.13	263
DP001	DEMINDEX RESOURCES CORPORATION (S.A.) (PTY) LTD	28-Feb-2007	28-Feb-2010	0.13	2500
IBL20	INVESTEC BANK LTD	15-Mar-2011	15-Mar-2014	0.13	19231
SBN43	STANDARD BANK OF SOUTH AFRICA LTD	17-Jul-2008	17-Jul-2018	0.08	164
FRBN06	FIRSTRAND BANK LTD	18-Nov-2009	15-Sep-2015	0.07	1942
FRBN07	FIRSTRAND BANK LTD	18-Nov-2009	15-Sep-2015	0.07	16178
SSN007	STANDARD BANK OF SOUTH AFRICA LTD	29-May-2012	15-Jan-2020	0.07	9192
FRBN02	FIRSTRAND BANK LTD	31-Mar-2004	28-Feb-2006	0.00	No trades
FRBN03	FIRSTRAND BANK LTD	22-Apr-2004	28-Feb-2008	0.00	No trades
IBL59	INVESTEC BANK LTD	17-Feb-2014	17-Feb-2019	0.00	No trades
KSB004	KAGISO SIZANANI CAPITAL (PTY) LTD	18-Jan-2008	01-Feb-2013	0.00	No trades
KSB005	KAGISO SIZANANI CAPITAL (PTY) LTD	29-Feb-2008	28-Feb-2013	0.00	No trades
NBK8A	NEDBANK LTD	23-Mar-2011	24-Mar-2014	0.00	No trades
NHM002	NORTHAM PLATINUM LTD	13-May-2016	12-May-2021	0.00	No trades
SSN006	STANDARD BANK OF SOUTH AFRICA LTD	15-May-2012	21-Dec-2014	0.00	No trades

Table D.5: The details of the 246 pure vanilla corporate bonds that were analysed in Chapter 2, ordered from highest to lowest TFR.

E. Glossary

Bid-offer spread (also known as **bid-ask spread**): The amount by which the selling (offer) price exceeds the buying (bid or ask) price.

Burn-in: In MCMC sampling, the number of initial draws that is discarded to eliminated the dependence on the starting values.

Cash flows: In terms of a bond, the series of expected coupon and bullet payments to be made by the issuer during the lifetime of the bond.

Central limit theorem: The theorem states that, in some circumstances, the addition of independent random variables, not necessarily normally distributed themselves, tends toward a normal distribution.

Clifford-Hammersley theorem: The theorem that proves that the joint distribution of the unknown variables can be fully specified by a set of conditional distributions of each of the unknown variables, if certain conditions are met.

Conjugate prior: The prior distribution relative to the likelihood that results in the posterior distribution being of the same family as the prior distribution.

Coupon: Fixed interest payments made by the issuer of the bond at pre-determined points in the lifetime of the bond.

Credit rating: The assessment of an entity's ability to fulfill its financial commitments such as the repayment of loans and the probability of the entity's default.

Credit spread: In terms of a bond, the spread above the government equivalent yield reflecting the credit worth of the issuer.

Default: Failure to fulfill a financial obligation, such as the repayment of a loan.

Default intensity (also known as **hazard rate**): Probability of default for a certain time period conditional on no earlier default. (Hull (2009)).

Duration: In terms of a bond, the time weighted present value of the cash flows.

Duration mismatch: Bonds who do not have the same maturity, cash flow dates or coupon size experience a duration mismatch.

Efficient market hypothesis: the postulation that asset prices fully reflect all available information.

Equivalent martingale measure (EMM): Risk-neutral probability measure such that each asset price is exactly equal to the discounted expectation of the share price under this measure.

Gaussian: Normally distributed.

Government zero coupon bond yield curve: The yield curve constructed using only government bonds, with the coupons stripped out.

Guarantee type: In terms of a bond, an indication of whether the bond is collateralised or not.

Hyperparameters: Conjugate prior distribution's parameters and the parameters of the resulting posterior.

Inclusion indicator: Binary series indicating whether data at a certain point is missing (0) or not (1).

Informative prior: A prior distribution reflecting specific, definitive information about a variable.

Interest rate market: The sector of the financial market dealing in the trading of interest rate and debt instruments.

Issuance: In terms of a bond, its release into the market.

Issuer: In terms of a bond, the entity issuing or selling the bond to investors.

Liquidation measure: Number of days it would take to buy or sell out of a given position in a bond.

Liquidity: Ability to convert an investment into cash freely and quickly, with little or no effect on the price of the instrument.

Markov: Property defined by the probability distribution of the next state depending only on the current state and not on the sequence of events that preceded it.

Mixing: The stationary distribution being reached quickly starting from an arbitrary position in MCMC sampling.

Modified following day count: Implied automatic change of a date (usually relating to a bond payment date) whereby if the date does not fall on a business day, the modified following day for that date is the next business day, unless this business day is in the next month, in which case the business day preceding the date is used.

Nominal: In terms of a bond, the number of bonds, usually expressed in ZAR.

Nominal spread: Theoretically correct version of the quoted spread, calculated as the difference between the yield of the bond and its risk-free counterpart.

Objective or uninformative prior: A prior distribution reflecting vague, general information about a variable.

Path errors: The goodness of fit of estimated parameters to the data as given by the Root Mean Square Error and the Mean Absolute Percentage Error.

Prediction interval: Range of values so defined that there is a specified probability that the value of a parameter lies within it.

Primary market: Part of the capital market that deals with issuing of new instruments, the first uptake of newly issued instruments.

Principle of indifference: If n possibilities are indistinguishable except for their names, then each possibility should be assigned a probability equal to $\frac{1}{n}$. (Wikipedia).

Prior distribution: Probability distribution reflecting beliefs about a certain variable before pertinant information is taken into account.

Posterior distribution: Probability distribution of an unknown quantity, treated as a random variable, conditional on the information obtained from observed data.

Pure vanilla corporate instrument: Standard contracts with no exotic features.

Quoted spread: Difference between the yield of a bond and the yield of its companion bond (government bond with same characteristics), as quoted by the JSE.

Risk-neutral measure: EMM associated with the risk-free asset.

Secondary market: Market where investors purchase instruments from other investors, rather than from the issuers themselves.

Seniority: In terms of bonds, order of repayment of bonds in the event of bankruptcy of the issuer - higher than subordinate.

Settlement date: Date on which trade settles, usually 3 days after the trade date.

Special Purpose Vehicles: Legal entity created to fulfill narrow, specific or temporary objectives.

Stationary: Property of a stochastic time series process whereby the joint probability distribution does not change when shifted in time.

Term structure: In terms of bonds, the relationship between different bonds' yields and maturities.

Term-to-maturity: Time between the current day and the maturity of the bond.

Yield-to-maturity: Total return expected on a bond if the bond is held until the end of its lifetime.

Z spread: Constant value that needs to be added to the government zero coupon bond yield curve in order for the sum of the discounted cash flows using the shifted government curve to match the market quoted bond price.

Zero trade ratio: Ratio of number of bonds that did not trade over total number of bonds issued.