Modelling probabilities of corporate default

Cole van Jaarsveldt

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

July 15, 2019

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy to the University of Cape Town. It has not before been submitted for any degree or examination.

Signed by candidate

Cole van Jaarsveldt

July 15, 2019

Abstract

"If I have seen further it is by standing on the shoulders of giants." – Isaac Newton

This dissertation follows, scrupulously, the probability of default model used by the National University of Singapore Risk Management Institute (NUS-RMI). Any deviations or omissions are noted with reasons related to the scope of this study on modelling probabilities of corporate default of South African firms. Using our model, we simulate defaults and subsequently, infer parameters using classical statistical frequentist likelihood estimation and one-world-view pseudo-likelihood estimation. We improve the initial estimates from our pseudo-likelihood estimation by using Sequential Monte Carlo techniques and pseudo-Bayesian inference. With these techniques, we significantly improve upon our original parameter estimates. The increase in accuracy is most significant when using few samples which mimics real world data availability.

Acknowledgements

I would like to thank my parents for their unwavering emotional and financial support during the last eight years. I would also like to thank Obeid Mahomed for his technical insight and guidance during this study. Additionally, I would like to thank Matthew Molyneux and Heleen Goussard from RisCura for their assistance, and for providing the data around which the model could be built. Finally, I would like to thank Emma Platts for her assistance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A corporate bond is a long-term debt instrument (with maturity greater than one year) that is issued by corporations to raise capital for various projects or potential expansions. The bonds are purchased by individuals or corporations that want higher returns than are achievable by purchasing "risk-free" government bonds. The fairly recent Greek government-debt crisis, however, has shown us that government bonds can no longer be regarded as being devoid of risk. This instrument securitizes a loan that can be sold on the secondary market. The bond is backed by the expected future earnings of the corporation and often the physical assets of the corporation are used as collateral. This is not always the case, with some companies relying on their reputation alone, and with others simply issuing personal guarantees to the purchasers of the bond. Despite their reputations, the collateral provided, or the personal guarantees, corporations often find themselves defaulting on their obligations.

When evaluating the credit risk posed by a debt instrument there are three factors of interest: the probability of the firm defaulting on its obligations over different incremental horizons; the total exposure or level of exposure when the firm defaults (total outstanding debt at default); and the loss-given-default or the recovery rate of the debt once the firm has defaulted expressed as a percentage of total exposure at the default event. The level of exposure at default is known at all times and is therefore not a relevant covariate. The loss-given-default is therefore the sole value of interest once a firm has defaulted. Since the first Basel Capital Accord, a significant amount of research has been done trying to model the loss-given-default of corporate debt. The research thus far has not yielded clear predictive power with numerous papers, including the work done in Bastos (2010), showing bimodal distributions in the recovery rates. The work done in Bastos (2010) involves fitting a regression with an appropriate link function to restrict the recovery rate, r, to the unit interval such that: $r \in [0, 1]$. This method is seemingly contrived, with the work first done in Breiman *et al.* (1984) and repeated in Bastos (2010) using re-

gression trees to better model recovery rates showing more promising results. The predictive power of these models have yet to be fully demonstrated. The pricing of corporate debt therefore primarily relies on a probability of default model.

Models that predict the probability of a firm defaulting on its obligations can be broadly grouped into two categories: structural models and reduced-form models. The almost legendary work done in Merton (1974) is an example of a structural model. Structural models can be candidly described as models that attempt to model a firm's balance sheet and impose some modellable dynamics on the assets, equity, or liabilities of the firm in order to make default predictions. Reduced-form models, contrarily, impose a hazard rate that directly influences a firm's probability of default and does not delve into the accounting of the firm except for potential explanatory variables. The original predictors of the creditworthiness of firms were the accounting ratios that have been known and studied for centuries. Gauging a firm's creditworthiness using often conflicting ratios necessitated the development of more complex and descriptive models that incorporated multiple measures into a single model. The earliest reduced-form models done in Beaver (1966) and Altman (1968) attempted to reconcile ratio analysis with a more modern statistical technique such as discriminant analysis. The output of the discriminant analysis was a single discriminant score which allowed firms to be grouped and ranked according to their probability of default. These models, while an improvement on earlier ratio analysis, did nothing to predict when the firms would ultimately fail.

Later work done in Ohlson (1980) and Zmijewski (1984) attempted to predict when a firm would default using regression with appropriate link functions such as logit and probit functions. These binary response models were appropriate for determining whether a firm would fail or survive. This work made progress in being able to predict a firm's probability of default over the next time interval, but did not offer predictions beyond this first time interval. The work done in Campbell *et al.* (2008) attempted to extend this work by using the multiple logit model to predict the probability of default over multiple horizons. The work done previously used static variables and ratios to predict the probability of firms defaulting in the next time increment and paid no attention to earlier variables and ratios when assessing default probabilities. The work done in Shumway (2001) and later reviewed in Chava and Jarrow (2004) attempted to remedy this by incorporating previous observations that could now be used to establish trends in variables that is crucial to our work. They also introduced further improvements to existing default models such as using monthly increments instead of annual increments, grouping firms based on their industries and using common market variables in addition to standard accounting ratios as variables. This work, while revolution-

ary in default prediction, failed to adequately account for firms that left the sample of active firms through events other than default such as mergers or acquisitions. Events such as these were merely treated as censored data. Finally, the work done in Duffie *et al.* (2007) incorporated all of this previous work and adequately dealt with these non-default exits by imposing two independent Poisson processes on a firm's probability of exit (from the sample of active firms) by default, and a firm's probability of exit by non-default.

Exit by default and exit by a non-default event are broad definitions and significantly contribute to the difficulty in obtaining data for the building of any probability of default model. Briefly, a default exit event has taken place when a firm fails to make a coupon or nominal repayment for a corporate bond and a non-default exit event has occurred when an event such as a merger or an acquisition of the firm has occurred. The work done in Duan *et al.* (2012) follows the work done in Duffie *et al.* (2007) with the exception of the treatment of the descriptive variables. In Duffie *et al.* (2007) the descriptive variables used to model the probability of default over potentially any horizon are assumed to follow a high-dimensional autoregressive process. This means that should we wish to model the probability of default of a firm over a horizon of k months, we would have to take the present values of the descriptive variables and model the change in these variables over $(k-1)$ intervals using our high-dimensional autoregressive assumptions. Clearly the error in prediction would increase dramatically the further into the future we try model the values of the descriptive variables. The work done in Duan *et al.* (2012) does not follow this technique, but rather uses the present value of the variables to predict the probability of default over different horizons without having to explicitly model the dynamics of the variables.

Before giving a brief overview of the contents of this paper, we feel that it should be explicitly stated that the primary objective of this paper is to demonstrate the efficiency of the estimation methodology used in NUS-RMI (2016), rather than demonstrate something that is economically meaningful. We simulate exits using the same underlying structure that underpins the estimation methodology in the hopes of demonstrating the improved convergence rate in the Sequential Monte Carlo parameter estimation technique used in NUS-RMI (2016), when compared to standard likelihood estimation techniques. We purposefully show the slower convergence of the one-world-view pseudo-likelihood estimation when compared against the classical statistical frequentist likelihood estimation to demonstrate the improved convergence when the one-world-view pseudo-likelihood estimation technique is used in conjunction with the Sequential Monte Carlo technique.

The variables of interest to our model are introduced in the proceeding chapter. The subsequent chapters follow the work done in NUS-RMI (2016) which in turn is based primarily on the work done in Duan *et al.* (2012). It is, however, noted in NUS-RMI (2016) that the extension of the work done in Duan *et al.* (2012) to more than one hundred economies using the same variables deemed to be significant based on the study done on United States firms is most definitely erroneous. Future updates to the work done in NUS-RMI (2016) will attempt to remedy this erroneous extension. We continue this potential error by using the analogous 12 variables to predict defaults for South African firms. Most of the variables require little explanation, with the exception of the Distance-to-default calculation that follows the work done in Merton (1974) and the calculation of a firms idiosyncratic volatility that follows the work done in Shumway (2001). In Duan *et al.* (2012) these variables are calculated using monthly updated inputs, while in NUS-RMI (2016) these variables are calculated using daily updated inputs with all the observations in the previous business year (approximately 250 trading days) being used with variables being coded as missing if there are not enough valid observations. This luxury is not extended to our work given the scope of our study and the significantly smaller synthesized data set.

In Chapter 3 we establish the modelling framework for our survival and exit probabilities. We additionally describe a few nuances in the implementation of the discrete version of the probabilities. This becomes necessary as the independent Poisson processes are such that a joint jump can occur within the same discrete time interval. Quite simply, should a joint jump occur, this will be treated as a default exit which is in keeping with the work done in Duan *et al.* (2012) and NUS-RMI (2016). Having established the treatment of survivals, default exits, and non-default exits, we introduce the dependence of the hazard rate function on the descriptive variables described in the preceding chapter for our reduced-form model. Understanding this dependence becomes vital in Chapter 4 and Chapter 5 where we describe how we simulate our exit data and estimate our parameters through various likelihood techniques respectively. The simulation of the default and non-default exits relies heavily on inverse transform sampling. The simulation of the exits in Chapter 4 is related to the likelihood estimation of the parameters in Chapter 5, where, ideally, we would arrive at the same parameters through various likelihood estimation techniques that were used to originally simulate the exits using the techniques described in detail in Chapter 4.

The parameters in Tables 4.1 and 4.2 for the Nelson-Siegel function, first introduced in Nelson and Siegel (1987) as Equation (4.10), as well as the parameters in Tables 5.2 and 5.4 are obtained directly from the NUS-RMI website. The institute

regularly posts updates to their paper, NUS-RMI (2016), on their website and it is this paper and these parameters that are the focus of this study. For the estimation of these parameters, information from the South African economy up until 31 October 2018 was used and then the parameters were calibrated on 7 November 2018. The reader will notice discrepancies between the date the parameters were calibrated and the dates for the associated variables throughout this study. This is not a problem as the survivals, default exits, and non-default exits are simulated using this combination of parameters and variables which nullifies the time discrepancies.

In Chapter 4 we describe how the default and non-default exits are simulated using the modelling framework described in the previous chapter along with a brief description of the hazard rate model and its relation to our exit simulation model. The simulation of our default and non-default exits relies on inverse transform sampling. This method is initially described for the homogeneous hazard rate case before it is extended to the non-homogeneous case and finally related to our simulations through a Riemann sum analogy. The reader might ponder the treatment of a firm that has been simulated to both default exit and non-default exit. This is in keeping with the original modelling framework presented in Duan *et al.* (2012) and in our study where the processes that cause the default and non-default exits are assumed to be independent Poisson processes. The number of default exits and non-default exits simulated using the parameters from Tables 4.1 and 4.2 in Equation (4.10) is displayed in the proceeding chapter after the variable dependence of the hazard rate function is made clear. The treatment of such a firm with regards to the likelihood estimation of default and non-default exit parameters is discussed in the next chapter.

The default and non-default parameter estimation techniques are described in Chapter 5. Before describing the manner in which the parameters are estimated at each horizon, we briefly describe how the default exits and non-default exits are simulated for the basic likelihood estimation and the pseudo-likelihood estimation to show the relationship between simulation and estimation. For the basic likelihood estimation, the same set of variables is used for every simulation and this motivates our description of this likelihood estimation technique as the classical statistical frequentist's approach to estimation. Contrarily, for the pseudo-likelihood estimation, a different set of variables is used for every simulation which motivates this as the many firms, but one scenario and therefore one-world-view of estimation. The simulation techniques for both the basic likelihood and pseudo-likelihood estimation are outlined before the parameter estimation is described. Along with the simulation technique that is made specific from the general out-

line in the previous chapter, we display the number of simulated default and non-default exits in Tables 5.1 and 5.7. The differences and similarities between the basic likelihood and pseudo-likelihood estimation of parameters is discussed with the pseudo-likelihood estimation output laying the foundation for the next chapter.

The work done in Duan and Fulop (2013) and repeated in NUS-RMI (2016) using a Sequential Monte Carlo algorithm and the subsequent pseudo-Bayesian inference is the focus of Chapter 6. We use the same simulated variables used in the previous chapter for the pseudo-likelihood estimation, while using a different pseudo-likelihood estimation function that takes the information over all horizons into account. The reader can note that the pseudo-likelihood probabilities in Equations (6.3) and (6.4) are decomposable into horizon-specific probabilities that are equivalent to Equations (5.13) and (5.14) respectively. These pseudo-likelihood functions are critical in assigning weights to generated particles in the reweighting step, and calculating the probabilities of acceptance for new particles in the move step. The work done in NUS-RMI (2016) progresses the algorithm by stepping forward in time, but since all of our simulated data is taken to be from the same time period, we progress the algorithm by simply adding more data in fixed increments. Once the algorithm has progressed through all of the available data, the final parameter estimates are inferred using a pseudo-Bayesian inference technique from Duan and Fulop (2013) and NUS-RMI (2016). We then compare the errors from the pseudo-likelihood estimation, the least squares fit of the Nelson-Siegel function, and the Sequential Monte Carlo technique. To ensure that the results are not merely an artefact of this one simulation we run this numerous times to create means and Monte Carlo error bounds to ensure the effectiveness of the technique.

The final two chapters summarise our results and go into detail about potential extensions that can be made should we further our work in this field. There are a large number of aspects to this work that can be reviewed and redone. Whether it be the variables selected to drive the hazard rate function, the structure of the hazard rate function, the use of actual data in our parameter estimates, other possible candidates besides the Nelson-Siegel function to model the long-term behaviour of our parameters, or the numerous improvements and extension that can be made to the Sequential Monte Carlo algorithm, we believe this is a very promising field of research and look forward to further developments.

Chapter 2

Input Variables

2.1 Variables of Interest

An inconvenience encountered when building any model is the frequency with which potential descriptive variables are updated. Stock indices and market capitalizations are updated almost continuously, whereas accounting ratios such as the current ratio or the debt-equity ratio are updated quarterly or even annually. We will use a fundamental time increment of one month $\left(\Delta t = \frac{1}{12}\right)$ over which variables are updated. The descriptive variables being used in our model can be divided into two groups: common variables ($W(n)$) that are used for all firms in the South African economy; and firm-specific variables ($U_i(n)$) that are used for individual firms. The variables are collectively denoted as: $X_i(n) = [W(n), U_i(n)]$. The index n is used to denote the value of a variable at time point n with the subscript i being used to denote the variable values for firm i. In Duan *et al.* (2012) the current value of the variable as well as the annual average of the variable (referred to as the level) and the current value minus the annual average of the variable (referred to as the trend) are considered as explanatory variables. The level of a variable serves to show what the effective value of the variable has been for the previous year while disregarding large outliers which would misrepresent the firm's financial status. The trend of a variable serves to take into account the firm's momentum in a sense. All other things equal, by considering two firms with different trends in their debt-equity ratio, we could logically conclude that the firm that is taking on more debt recently would be more prone to defaulting on its obligations.

Before discussing the variables in more detail, we feel that it would benefit the reader if we reiterate here that the variables chosen to predict exits of South African firms from the sample of active firms are almost certainly at least partly incorrect. This is also stated in NUS-RMI (2016) from which we obtained the same twelve variables to simulate the exits so we know for certain that it is these variables that drive the exit process.

Two common variables were chosen in Duan *et al.* (2012) as they had significant predictive ability as proxies for the general state of the economy. The common variables are:

- stock index annual simple return; and
- 3-month short-term interest rate.

In the work done in NUS-RMI (2016) on the South African economy, the stock index used is the MSCI South Africa Index, and the 3-month interest rate used is the 91-day South African treasury bill tender rate. For both the stock index and the interest rate the current values are used as the descriptive variables in the model.

Ten firm-specific variables were chosen in Duan *et al.* (2012) as having the most significant predictive power. These ten variables are:

- level of distance-to-default;
- trend of distance-to-default;
- level of the ratio of cash and cash equivalents to total assets;
- trend of the ratio of cash and cash equivalents to total assets;
- level of the ratio of net income to total assets;
- trend of the ratio of net income to total assets;
- level of the firm's relative size;
- trend of the firm's relative size;
- current value of the firm's market-to-book asset ratio; and
- current value of the firm's idiosyncratic volatility.

Most of the above variables are self-explanatory, while others are explained in detail below. It is noted in Duan *et al.* (2012) that the momentum effect of trend improves the predictive capabilities of the model for the distance-to-default, the ratio of cash and cash equivalents to total assets, the ratio of net income to total assets, and the relative size of the firm. It is also noted that the addition of the level and trend of the final two variables does not significantly improve the predictive power of the model.

2.2 Distance-to-Default

Our distance-to-default calculation follows closely the work done in Merton (1974). Merton published his work soon after the Black-Scholes model was first published and he made the first attempt to model corporate debt assuming default was possible. The Merton model has the simplifying assumptions that the firm is solely funded by equity and a zero coupon bond with principal L and maturity T. Owing to the limited liability of the firm, the equity of the firm at time t, E_t, can be thought of as an option on the market value of the assets at time t, V_t, of the firm with strike equal to principal amount of the bond, L. The market value of the assets of the firm are further assumed to follow geometric Brownian motion:

$$\frac{dV_t}{V_t} = \mu dt + \sigma dB_t. \tag{2.1}$$

It should be noted that it is difficult to model the drift, μ, of the geometric Brownian motion, but, fortunately, we can circumvent this problem, with σ and dB_t being the volatility of the market value of the assets and standard Brownian motion respectively. Using the Black-Scholes pricing formula we arrive at an expression for the equity value of the firm under risk-neutral dynamics:

$$E_t = V_t N(d_{t+}) - Le^{-r(T-t)} N(d_{t-}), \tag{2.2}$$

with N being the cumulative normal distribution function, r being the "risk-free" rate, $(T-t)$ being the time to maturity, and with $d_{t\pm}$ being defined as:

$$d_{t\pm} = \frac{ln\left(\frac{V_t}{L}\right) + \left(r \pm \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{2.3}$$

The calculation of L has been the subject of a significant amount of research. We follow the approach of Crosbie and Bohn (2003) which uses the standard KMV assumptions to set $(T-t)$ to one year and to calculate L as the sum of the firm's total short-term debt and one half of the firm's long-term debt. Unfortunately, this calculation of L vastly underestimates the values for distance-to-default for financial firms as most of the financial firm's liabilities are not classified as either short-term or long-term debt. A solution to this problem as stated in Duan (2010) involves including a fraction, δ, of other liabilities in our calculation of L. Other liabilities is simply calculated as the firm's total liabilities minus the firm's short-term debt and long-term debt.

We now have the tools to construct our log-likelihood function that we can maximize to find our optimal parameters in line with work done in Duan (1994) and Duan (2000). The optimal percentage of our other liabilities to include, δ, can be

optimised along with our drift term, μ, and our volatility, σ. To begin, the equity values should be viewed as transformed data using Equation (2.2). We can therefore write the likelihood of the equity values as the product of the likelihood of the unobserved market value of the assets and the Jacobian of this transformation. An additional level of complexity is added in Duan *et al.* (2012) where the market values of the firm's assets, V_t, are standardised by the book values of the firm's assets, A_t. This added level of complexity serves to scale the effect of large investments made by the firm so as not to significantly distort the series of market values from which we attain our parameter estimates. By applying Itō's Lemma to Equation (2.1) it can be shown that $ln(V_t)$ is normally distributed with mean $\left(\mu - \frac{\sigma^2}{2}\right)t$ and variance $\sigma^2 t$. We arrive at the following log-likelihood function:

$$\begin{aligned}\mathcal{L}(\mu, \sigma, \delta) = &- \frac{n-1}{2} ln(2\pi) - \frac{1}{2}\sum_{t=2}^{n} ln(\sigma^2 h_t) \\ &- \sum_{t=2}^{n} ln\left(\frac{\hat{V}_t(\sigma, \delta)}{A_t}\right) - \sum_{t=2}^{n} ln[N(\hat{d}_{t+}[\hat{V}_t(\sigma, \delta), \sigma, \delta])] \\ &- \frac{1}{2\sigma^2}\sum_{t=2}^{n} \frac{1}{h_t}\left(ln\left(\frac{\hat{V}_t(\sigma, \delta)}{A_t} \times \frac{A_{t-1}}{\hat{V}_{t-1}(\sigma, \delta)}\right) - \left(\mu - \frac{\sigma^2}{2}\right)h_t\right)^2.\end{aligned} \tag{2.4}$$

In Equation (2.4) above a few key points must be noted. The h_t represent the fraction of the year between the observations E_{t-1} and E_t. The dependence of $\hat{V}_t$ on σ and δ follows from Equation (2.2) where $\hat{V}_t$ is inferred from solving the equation with respect to $\hat{V}_t$ by means of some root-finding algorithm. The dependence of $\hat{d}_{t+}$ on $\hat{V}_t(\sigma, \delta)$, σ, and δ follows from Equation (2.3). The dependence of $\hat{V}_t$ and $\hat{d}_{t+}$ on δ follows as a result of the dependence of L on δ. The above equation should look familiar to most with the exception of a few terms.

The third term above in Equation (2.4) was originally not included in the work published in Duan (1994) and was later added in the follow-up paper, Duan (2000), where it is noted that if the market value of the assets were directly observable then the term would disappear, as the derivatives with respect to all the parameters would be zero. This term, however, is not directly observable and we are dealing with transformed data which leads us to have to infer the value from observed data.

The fourth term above in Equation (2.4) follows from the Jacobian of the derivative with respect to V_t of the inverse transformation which is analogous to the inverse of Equation (2.2). With the cumulative normal distribution being non-negative, the absolute value can be dropped when finding the log-likelihood. The derivation involving some nuances is done in more detail in Appendix A.

By referring to Appendix B you can find the derivation of an expression for

our maximal μ, which we then substitute into Equation (2.4) above to simplify our maximization problem from three variables to two variables:

$$\begin{aligned}\bar{\mathcal{L}}(\sigma,\delta) = &-\frac{n-1}{2}ln(2\pi) - \frac{1}{2}\sum_{t=2}^{n} ln(\sigma^2 h_t) \\ &- \sum_{t=2}^{n} ln\left(\frac{\hat{V}_t(\sigma,\delta)}{A_t}\right) - \sum_{t=2}^{n} ln[N(\hat{d}_{t+}[\hat{V}_t(\sigma,\delta),\sigma,\delta])] \\ &- \frac{1}{2\sigma^2}\left(\sum_{t=2}^{n}\frac{1}{h_t}\left[ln\left(\frac{\hat{V}_t(\sigma,\delta)}{A_t}\times\frac{A_{t-1}}{\hat{V}_{t-1}(\sigma,\delta)}\right)\right]^2\right. \\ &\left. - \frac{1}{\sum_{t=2}^{n} h_t}\left[ln\left(\frac{\hat{V}_n(\hat{\sigma},\hat{\delta})}{A_n}\times\frac{A_1}{\hat{V}_1(\hat{\sigma},\hat{\delta})}\right)\right]^2\right).\end{aligned} \tag{2.5}$$

As noted in NUS-RMI (2016), a number of stability problems are encountered when trying to get estimates for δ from Equation (2.5) above. Due to the scope of this course and the stability problems experienced in NUS-RMI (2016) with their far superior data set, the addition of financial firms in our calculations has been excluded to allow for us to set the δ parameter to zero. This is analogous to the standard KMV assumptions already mentioned above which is clearly a special case of the work done in Duan (2010) with δ set to zero. The likelihood function is now a function of one variable that is maximized using the previous year's implied asset values from Equation (2.2) to yield a σ value at each month end. This follows the work done in Duan *et al.* (2012) and NUS-RMI (2016) which differs from the work done in Duffie *et al.* (2007) in that the optimal σ estimated in Duffie *et al.* (2007) uses all available implied asset values to calculate the optimal sigma for each firm. This inappropriately peers into the future in a sense and we rather use the rolling window approach as in Duan *et al.* (2012) and NUS-RMI (2016). Our expression for distance-to-default at time t is now:

$$DTD_t = \frac{ln\left(\frac{V_t(\hat{\sigma})}{L}\right) + \left(\mu - \frac{\hat{\sigma}^2}{2}\right)(T-t)}{\hat{\sigma}\sqrt{T-t}}. \tag{2.6}$$

By observing the derivation of our maximal μ in Appendix B it can be seen that the value is highly sensitive to changes in the implied market value of the assets from almost a year previously. The value of μ could change drastically, solely as a result of the moving window as opposed to present adverse occurrences in the firm's balance sheet. This problem is circumvented by making the following approximation:

$$\mu \approx \frac{\hat{\sigma}^2}{2}. \tag{2.7}$$

Finally, our distance-to-default value calculated at each month end using our optimal sigma from Equation (2.5) is simplified by substituting Equation (2.7) into Equation (2.6) which yields:

$$DTD_t = \frac{ln\left(\frac{V_t(\hat{\sigma})}{L}\right)}{\hat{\sigma}\sqrt{T-t}}, \tag{2.8}$$

at each time point t.

2.3 Relative Size

The relative size of a firm in the South African market is calculated as the natural logarithm of the ratio of the firm's market capitalization to the median market capitalization of all the firms on the Johannesburg Stock Exchange (JSE). It might seem obvious that larger firms are less likely to default should market conditions deteriorate, with the larger firms having more diversified revenue streams and more sources of capital should it run into liquidity problems. Since the work done in Duan *et al.* (2012) on United States firms covers the period from 1991 to 2011, there is an additional variable added to take into account the bailing out of large firms during the financial crisis in 2007 and 2008. We mention this variable here as it is noted in Duan *et al.* (2012) that the largest firms in economies are very likely to get bailed out during financially difficult times, which would in turn confound the effects of these two variables with the largest firms being noted as simply being "too big to fail". The inclusion of the level and the trend of the firm's relative size was found to be significant in the work done in Duan *et al.* (2012).

2.4 Idiosyncratic Volatility

The calculation of the firm's idiosyncratic volatility, or SIGMA, follows the work done in Shumway (2001) that also made earlier contributions to predicting the probability of a firm defaulting. The work done in Duan *et al.* (2012) calculates SIGMA by regressing the monthly returns of the firm's market capitalization against the monthly returns of the selected economy stock index, which in turn follows directly from Shumway (2001). It is postulated in Shumway (2001) that the more variable a firm's market capitalization returns are relative to a proxy for the returns from the economy as a whole, the more likely a firm is to default. This work is extended in NUS-RMI (2016) by regressing the daily market capitalization returns of a firm against the daily returns of a stock index in the economy. We follow this

method exactly. It is postulated in NUS-RMI (2016) that this provides a better representation of the firm's idiosyncratic risk. The stock index used in the South African economy is the MSCI South Africa index, and the SIGMA for a firm is calculated as the standard deviation of the residuals of this regression of the daily returns over the last business year.

It may be beneficial to the reader if, at this point, we clarify the distinction between the volatility of the market value of the assets, σ, introduced earlier in Section 2.2 and the idiosyncratic volatility, SIGMA, introduced here. The volatility of the market value of the assets of the firm, σ, is merely a variable over which Equation (2.5) is maximized to yield an optimal $\hat{\sigma}$ that is then used to calculate one of our independent variables: distance-to-default. The idiosyncratic volatility, SIGMA, that can also be referred to as the idiosyncratic standard deviation of each firm's market capitalisation returns, is one of our independent variables that is calculated in accordance with the work done in Shumway (2001), Duan *et al.* (2012) and NUS-RMI (2016). The idiosyncratic volatility merely assigns a single value to how erratic the returns of the firm are relative to the returns of proxies for the general state of the economy which we then use as an independent variable. We attempt to quantify the firm-specific risk that works independently and in addition to the variables that attempt to quantify the macroeconomic risk.

Chapter 3

Model Framework

The modelling framework outlined in this chapter follows the work done in NUS-RMI (2016) exactly, with numerous equations below featuring in that paper. We begin by postulating that at each discrete time point, m, the firm will face three distinct possibilities during the next discrete time increment, Δt, from time point m until time point $\left(m + \frac{1}{12}\right)$: the firm can survive, the firm can have a default exit, or the firm can have a non-default exit. Information about firm i, in the form of the descriptive independent variables ($X_i(m)$) discussed in the previous chapter, is known at time m. We denote $p_i(m, n)$ as the conditional probability at time point m that firm i defaults before $\left(n + \frac{1}{12}\right)$ conditioned on the firm surviving until time n with $\bar{p}_i(m, n)$ being the corresponding probability of non-default exit. Now, at time point m, the probability of survival until time point $\left(n - \frac{1}{12}\right)$ of firm i only for it to default in the next increment, Δt, is:

$$\mathbb{P}_{t=m}(\tau_i = n, \tau_i < \bar{\tau}_i) = p_i\left(m, n - \frac{1}{12}\right) \prod_{j=m}^{n-\frac{2}{12}} [1 - p_i(m, j) - \bar{p}_i(m, j)], \tag{3.1}$$

with τ_i being the default time of firm i measured in years and $\bar{\tau}_i$ being the non-default exit time measured in years.

We will be modelling these probabilities as independent Poisson processes and therefore we define the forward intensity for default exit of firm i observed at time point m for the interval n to $\left(n + \frac{1}{12}\right)$ to be $h_i(m, n)$ with $m \leq n$. The analogous forward intensity for non-default exits is $\bar{h}_i(m, n)$. This independence assumption is out of necessity as modelling two dependent Poisson processes is an unnecessary complication. While this independence assumption is erroneous it is not entirely true as both forward intensities are functions of exactly the same variables. They are clearly related through these common variables. The assumption that they are driven by the same twelve variables rather limits possible extensions. Having defined the forward intensities, the conditional default probability becomes a simple

expression in terms of its forward intensity:

$$p_i(m,n) = 1 - exp[-\Delta t h_i(m,n)]. \tag{3.2}$$

We have discretized the time steps into months, and therefore joint jumps can occur in a single time step. We solve this problem by recognising these as defaults and therefore we have:

$$\bar{p}_i(m,n) = exp[-\Delta t h_i(m,n)] \times (1 - exp[-\Delta t \bar{h}_i(m,n)]). \tag{3.3}$$

Finally, we have the probability that in an interval the firm does not exit by default or by a non-default exit event:

$$\mathbb{P}_{t=m}\left(\tau_i, \bar{\tau}_i > n + \frac{1}{12} \,\middle|\, \tau_i, \bar{\tau}_i > n\right) = exp(-\Delta t[h_i(m,n) + \bar{h}_i(m,n)]). \tag{3.4}$$

We must still relate our forward intensities to our input variable vector, $X_i(m)$. Our forward intensities must be positive to ensure that our conditional probabilities are, at least, non-negative. To this end, we impose the following functional form on our forward intensities:

$$h_i(m,n) = exp[\beta(n-m) \cdot Y_i(m)], \tag{3.5}$$

where β is the coefficient vector that is a function of the difference in time between the start of the forward period, n, and the date of observation, m, in years. Note that $\bar{h}_i(m,n)$ and $\bar{\beta}$ are defined analogously. The necessity for the forward intensities to be positive is clear, but as a result of the choice for functional form, the exit probabilities are extremely sensitive to even minute changes in variables or parameters. This has an inconvenient consequence of creating stability issues during likelihood estimation, curve fitting and Sequential Monte Carlo estimation. We have simply redefined our descriptive variables such that $Y_i(m) = [1,\, X_i(m)]$ as this will allow a possible non-zero intercept. Finally, we introduce the notation:

$$H[\beta(n-m), X_i(m)] = exp[\beta(n-m) \cdot Y_i(m)], \tag{3.6}$$

for simplicity. This clarifies the parameter dependence of the forward rate. This is the forward rate for default with the forward rate for non-default exit being analogously defined. Equation (3.2) can now be restated as:

$$p_i(m,n) = 1 - exp(-\Delta t H[\beta(n-m), X_i(m)]). \tag{3.7}$$

Using this and the techniques introduced in Duan *et al.* (2012) and NUS-RMI (2016) we can get estimates for our default and non-default exit probabilities.

Chapter 4

Exit Simulation

Default data and default prediction is of critical importance to banks and other financial institutions. Unfortunately, these institutions do not readily share this information even for academic purposes. We have therefore simulated default exits and non-default exits, and used classical likelihood estimation as well as pseudo-likelihood estimation to recover the parameters for numerous horizons. We have the luxury of knowing the exact underlying structure that drives the exits as a result of the simulations. Given sufficient simulations, the parameters should be exactly recoverable through standard likelihood estimation. Note that below we showcase how default exits are simulated, but non-default exits are simulated in an analogous manner by simply replacing the β parameters with the analogous $\bar{\beta}$ parameters.

The first arrival time (default) for each of our independent Poisson processes is distributed exponentially. The cumulative distribution function (for $t \geq 0$) is given by:

$$F(t) = \mathbb{P}(\tau \leq t) = 1 - exp(-\lambda t), \tag{4.1}$$

with λ being the hazard rate of the exponential distribution. It is a well-known result that exponential random variable realisations, from a distribution with a constant or homogeneous hazard rate as in Equation (4.1), can be generated using the inverse transform method. This method yields exponentially distributed random variables from:

$$\tau = -\frac{ln(U)}{\lambda}, \tag{4.2}$$

where $U \sim \mathcal{U}(0, 1]$ is a uniformly distributed random variable. The simulation of arrival times from a non-homogeneous Poisson process is slightly more complex and needs to be done numerically. The cumulative distribution function (for $t \geq 0$) of a non-homogeneous exponentially distributed random variable is given by:

$$F(t) = \mathbb{P}(\tau \leq t) = 1 - exp\bigg(- \int_0^t \lambda(s) ds \bigg), \tag{4.3}$$

with $\lambda(s)$ being the non-homogeneous hazard rate. Assuming we have the underlying functional form of $\lambda(s)$, we would simulate realisations of τ numerically by solving:

$$\int_0^{\tau} \lambda(s)ds = -ln(U), \tag{4.4}$$

for τ. The above integral can be approximated using a left end-point Riemann sum:

$$\int_0^{\tau} \lambda(s)ds \approx \sum_{s=0}^{\tau-\Delta s} \lambda(s)\Delta s. \tag{4.5}$$

Fortunately, the model used by NUS-RMI (2016) is piece-wise homogeneous, with arrival times only possible at discrete month ends with homogeneous hazard rates present throughout each month. The work done in Duan *et al.* (2012) uses non-homogeneous hazard rates, which are less restrictive and will not necessitate discretization of variables. They could do this as it was a single paper studying the effectiveness of the model. Whereas, the NUS-RMI output monthly credit reports which could not be done without some kind of restrictive discretization. This has an unintended consequence of being slow to react to some variable changes. This piece-wise homogeneous hazard rate is analogous to the left end-point Riemann sum approximation of the non-homogeneous hazard rate described above in Equation (4.5). By equating Equations (4.4) and (4.5) above we arrive at a method for simulating defaults under our current model. With the Δs being fixed at $\frac{1}{12}$, or one month, defaults can therefore be simulated with the minimum τ such that:

$$\sum_{s=0}^{\tau-\Delta s} \lambda(s)\Delta s \geq -ln(U), \tag{4.6}$$

with the firm not defaulting when we have that:

$$\sum_{s=0}^{l-\Delta s} \lambda(s)\Delta s < -ln(U), \tag{4.7}$$

for a chosen boundary l. For our purposes, and following the work done in NUS-RMI (2016), we choose our boundary to be such that $l = 5$ years. For Section 5.1, in the proceeding chapter, default exits are simulated as the minimum τ such that:

$$\sum_{s=0}^{\tau-\Delta s} \Delta s H(\beta(s), X) \geq -ln(U), \tag{4.8}$$

with the firm not defaulting if:

$$\sum_{s=0}^{l-\Delta s} \Delta s H(\beta(s), X) < -ln(U), \tag{4.9}$$

with $l = 5$. Non-default exits are simulated in an analogous manner. The reader may have noted that the time dependence of the descriptive variables, X, has been dropped. As we are dealing with simulated data for firms at the same point in time, the dependence of the variables on time is an unnecessary notational inclusion. The time discrepancy between the parameters from the NUS-RMI website and the variables being used has already been discussed, but since the default and non-default exits are being simulated using this combination of parameters and variables the discrepancy is no longer relevant and the time dependency of the descriptive variables becomes unimportant. The variables in Tables 4.1 and 4.2 are used in Equation (4.10) below to generate the corresponding parameters:

Parameters	ρ_0	ρ_1	ρ_2	d
β_1	-4.0281	0.3448	1.1863	0.7021
β_2	0.0000	-0.1667	-0.2699	4.4145
β_3	0.0000	-8.3249	1.8207	6.2694
β_4	0.0000	-0.6894	0.3116	2.4672
β_5	0.0000	-0.3757	-0.0176	4.9106
β_6	0.0000	-0.3941	0.0215	0.9962
β_7	0.0000	-0.7489	-0.1777	0.8112
β_8	0.0000	-25.5749	10.1193	7.0136
β_9	0.0000	-12.3448	-7.5062	10.9382
β_{10}	0.0000	0.1256	-0.7029	10.9222
β_{11}	0.0000	-1.0498	1.9239	2.9040
β_{12}	0.0000	-0.4579	0.6863	1.7284
β_{13}	0.0000	1.1512	-14.0019	6.5642

Tab. 4.1: Default exit Nelson-Siegel function parameters

Parameters	ρ_0	ρ_1	ρ_2	d
$\bar{\beta}_1$	-5.4772	1.3235	3.4825	3.2454
$\bar{\beta}_2$	0.0000	-0.0344	0.4477	4.0782
$\bar{\beta}_3$	0.0000	12.2746	33.0122	1.8945
$\bar{\beta}_4$	0.0000	-0.0352	-0.0096	0.1573
$\bar{\beta}_5$	0.0000	-0.0004	-0.0583	1.9683
$\bar{\beta}_6$	0.0000	0.0804	-0.6667	1.0600
$\bar{\beta}_7$	0.0000	-0.1671	0.1351	3.3980
$\bar{\beta}_8$	0.0000	-15.6943	48.1537	32.3263
$\bar{\beta}_9$	0.0000	-2.0170	-9.8158	5.7312
$\bar{\beta}_{10}$	0.0000	0.0124	-0.5438	15.0020
$\bar{\beta}_{11}$	0.0000	-0.5746	0.4618	3.1899
$\bar{\beta}_{12}$	0.0000	0.0938	-0.1475	6.7029
$\bar{\beta}_{13}$	0.0000	2.5993	-0.6251	1.5453

Tab. 4.2: Non-default exit Nelson-Siegel function parameters

$$\begin{aligned}\beta(t;\rho_0,\rho_1,\rho_2,d) =&\rho_0+\rho_1\left(\frac{1-exp(-t/d)}{t/d}\right)\\ &+\rho_2\left(\frac{1-exp(-t/d)}{t/d}-exp(-t/d)\right),\end{aligned} \tag{4.10}$$

with $t = \{0, \frac{1}{12}, ..., \frac{59}{12}\}$. It should be noted that the parameters can be extrapolated beyond 5 years. For the purposes of our work and the work done in NUS-RMI (2016) this was not exercised or tested.

Chapter 5

Parameter Estimation

5.1 Basic Likelihood Estimation

We demonstrate here the classical statistical frequentist's approach that estimates parameters using maximum likelihood estimation. This is classical likelihood estimation in which there is one set of variables corresponding to a single firm with many possible scenarios. EOH Holdings is a technology services company that is listed on the JSE. We have the required data from the beginning of January 2016 to the end of April 2018 to construct the variables used in our model for this firm. We let X be the constructed variables for EOH Holdings for the end of April 2018 and we use these variables to simulate defaults up until the end of April 2023 - a period of 5 years. The variables constructed from this data in combination with the parameters used have created a rather favourable view of the default and other exit prospects of this firm. Our default exit times, τ_i, are therefore simulated as the minimum τ_i such that:

$$\sum_{t=0}^{\tau_i - \frac{1}{12}} \Delta t H(\beta(t), X) \geq -ln(U_i), \tag{5.1}$$

with the firm not defaulting if:

$$\sum_{t=0}^{l - \frac{1}{12}} \Delta t H(\beta(t), X) < -ln(U_i), \tag{5.2}$$

with l being the chosen boundary that we selected as 5 years in keeping with the work done in NUS-RMI (2016). The subscript i on the default times, τ_i, and the uniformly distributed random variables, U_i, is to denote the simulation number. Our non-default exit times, $\bar{\tau}_i$, are simulated as the minimum $\bar{\tau}_i$ such that:

$$\sum_{t=0}^{\bar{\tau}_i - \frac{1}{12}} \Delta t H(\bar{\beta}(t), X) \geq -ln(U_i^*), \tag{5.3}$$

with the firm not defaulting if:

$$\sum_{t=0}^{l - \frac{1}{12}} \Delta t H(\bar{\beta}(t), X) < -ln(U_i^*), \tag{5.4}$$

with l being defined as above. It must be noted that the uniformly distributed random variables used for each default exit simulation and each non-default exit simulation are independent. The sample size for both default exits and non-default exits starts at 10 and increases exponentially with base 10 up to 100000 simulations. Using the techniques described above, as well as the values in Tables 4.1 and 4.2 in conjunction with Equation (4.10) and the variables described above from EOH Holdings, we are able to simulate the default and non-default exits in Table 5.1 below.

Sample size	Total default exits	Total non-default exits
10	0	3
100	1	5
1000	35	79
10000	320	703
100000	3461	6995

Tab. 5.1: Number of exits per sample size

Were the variables and parameters from the same time period we could unambiguously state, given this simulated data, that the EOH Holdings has a 3.5% chance of defaulting within the next 5 years and a 7.0% chance of non-default exiting within the same time period. We could simply increase the number of simulations to get a more accurate estimate, but this would translate to more computing time. These percentages are a rather favourable view based on the observations made in NUS-RMI (2016) on global trends of defaults and other exit strategies. This will also become clear in Section 5.2. The default and non-default exit simulations and subsequent inference is a difficult tight-rope to walk. If there are too few defaults and other exits there will be poor, or simply incorrect, estimates for the parameters at some horizons, and if there are too many defaults and other exits the variables and possibly the parameters are a poor reflection of reality. Fortunately,

we can simulate enough scenarios to get adequate defaults and other exits at all horizons.

The parameters at each horizon, $t = \{0, \frac{1}{12}, ..., \frac{59}{12}\}$, can be estimated separately using probabilities of the form:

$$\begin{aligned}\mathbb{P}_l^{\beta(t)}(\beta(t);\tau_i,\bar{\tau}_i,X) &= 1_{\{min(\tau_i,\bar{\tau}_i)>l\}}exp[-\Delta t H(\beta(t),X)] \\ &+ 1_{\{t<(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}exp[-\Delta t H(\beta(t),X)] \\ &+ 1_{\{t=(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}(1-exp[-\Delta t H(\beta(t),X)]) \\ &+ 1_{\{t\leq(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}exp[-\Delta t H(\beta(t),X)] \\ &+ 1_{\{t>min[(\tau_i-\frac{1}{12}),(\bar{\tau}_i-\frac{1}{12})]\}},\end{aligned} \tag{5.5}$$

for default exits and probabilities of the form:

$$\begin{aligned}\mathbb{P}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau_i,\bar{\tau}_i,X) &= 1_{\{min(\tau_i,\bar{\tau}_i)>l\}}exp[-\Delta t H(\bar{\beta}(t),X)] \\ &+ 1_{\{t<(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}exp[-\Delta t H(\bar{\beta}(t),X)] \\ &+ 1_{\{t<(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}exp[-\Delta t H(\bar{\beta}(t),X)] \\ &+ 1_{\{t=(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}(1-exp[-\Delta t H(\bar{\beta}(t),X)]) \\ &+ 1_{\{t\geq(\tau_i-\frac{1}{12}),t>(\bar{\tau}_i-\frac{1}{12})\}},\end{aligned} \tag{5.6}$$

for non-default exits. The same variables, X, are used throughout for the statistical frequentist's approach with the index, i, being used to indicate the simulation number for default and non-default exit times. The likelihood function, Equation (5.7), is maximized separately for each horizon, t, to yield the default exit parameters:

$$\mathcal{L}_l^{\beta(t)}(\beta(t);\tau,\bar{\tau},X) = \prod_{i=1}^{I}\mathbb{P}_l^{\beta(t)}(\beta(t);\tau_i,\bar{\tau}_i,X), \tag{5.7}$$

with I being the total number of simulations, τ being the vector of default exit times, and $\bar{\tau}$ being the vector of non-default exit times. The likelihood function, Equation (5.8), is similarly maximized separately for each horizon, t, to yield the non-default exit parameters:

$$\mathcal{L}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau,\bar{\tau},X) = \prod_{i=1}^{I}\mathbb{P}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau_i,\bar{\tau}_i,X), \tag{5.8}$$

with the variables defined as above. In summation, we use a set of uniform random numbers to simulate default exits using Equations (5.1) and (5.2), and another

set of independent uniform random numbers to simulate non-default exits using Equations (5.3) and (5.4). We use these exits in Table 5.1 in Equations (5.5) and (5.7) to estimate default exit parameters, and in Equations (5.6) and (5.8) to estimate non-default exit parameters. The cumulative absolute errors for all $780(13 \times 60)$ parameters for default exits and non-default exits are displayed in Figures 5.1 and 5.2 respectively. It is clear from these figures that the convergence to the parameters that originally generated the exits is quicker for non-default exits than it is for default exits. This can simply be explained by the fact that far more non-default exits are generated than default exits which mirrors the observations made in NUS-RMI (2016). There is no doubt that should we simulate more exits, the errors in Tables 5.3 and 5.5 would decrease further with more accurate estimates of the parameters generating the exits being possible. This is also clearly evident from the trends observable in Figures 5.1 and 5.2. As already briefly alluded to in the previous chapter, we have now shown that the parameters are almost exactly recoverable, given that we know the underlying generating process. This is important for when we extend this to pseudo-likelihood estimation in the next section.

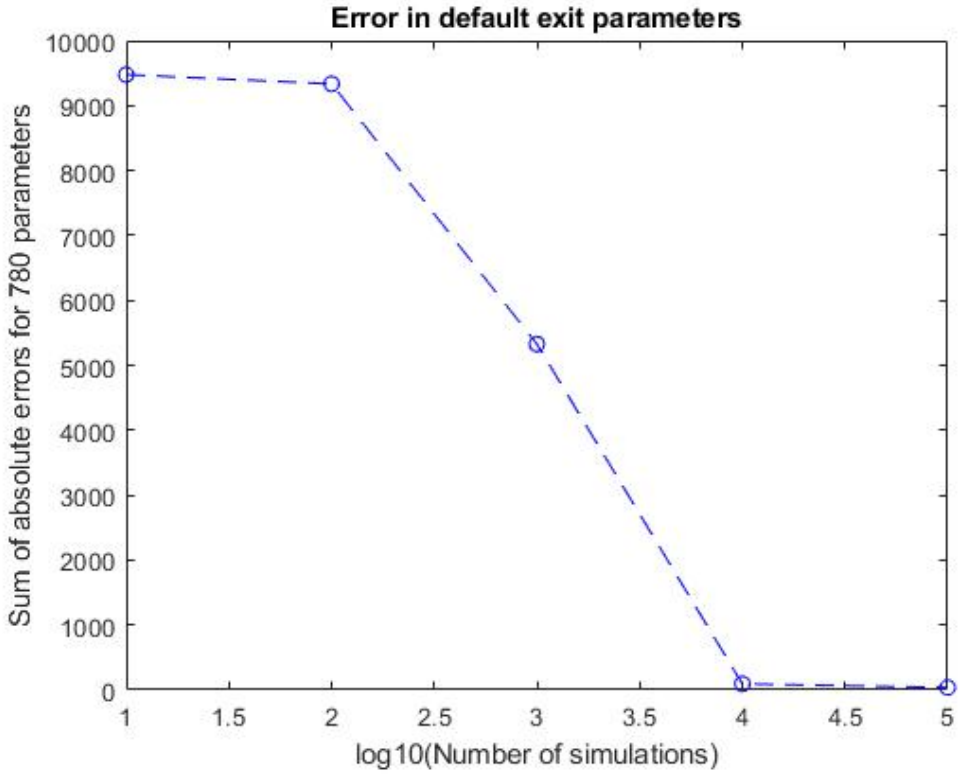

Fig. 5.1: Error in basic likelihood default parameter estimation

Horizon	1	12	24	36	48	60
β_1	-3.6833	-3.4947	-3.5812	-3.6839	-3.7591	-3.8107
β_2	-0.1667	-0.1749	-0.1793	-0.1801	-0.1783	-0.1747
β_3	-8.3249	-7.6241	-6.9451	-6.3444	-5.8118	-5.3390
β_4	-0.6894	-0.5305	-0.4060	-0.3172	-0.2531	-0.2062
β_5	-0.3757	-0.3442	-0.3137	-0.2869	-0.2631	-0.2420
β_6	-0.3941	-0.2521	-0.1685	-0.1216	-0.0933	-0.0751
β_7	-0.7489	-0.4977	-0.3385	-0.2458	-0.1890	-0.1521
β_8	-25.5748	-23.3676	-21.2232	-19.3211	-17.6319	-16.1300
β_9	-12.3448	-12.1391	-11.9093	-11.6754	-11.4387	-11.2006
β_{10}	0.1256	0.0926	0.0603	0.0315	0.0060	-0.0167
β_{11}	-1.0498	-0.6535	-0.3545	-0.1531	-0.0195	0.0674
β_{12}	-0.4579	-0.2266	-0.0884	-0.0167	0.0191	0.0357
β_{13}	1.1512	0.1833	-0.6882	-1.3967	-1.9679	-2.4239

Tab. 5.2: Select default exit parameters

Horizon	1	12	24	36	48	60
β_1	0.1088	0.0848	0.1531	0.1170	0.2034	0.0032
β_2	0.0010	0.0007	0.0012	0.0009	0.0016	0.0004
β_3	0.0212	0.0260	0.0045	0.0066	0.0006	0.0529
β_4	0.0136	0.0056	0.0083	0.0046	0.0052	0.0008
β_5	0.0045	0.0074	0.0049	0.0045	0.0046	0.0033
β_6	0.0045	0.0004	0.0014	0.0003	0.0009	0.0005
β_7	0.0021	0.0053	0.0017	0.0013	0.0018	0.0012
β_8	0.3322	0.3202	0.3990	0.4325	0.4333	0.0774
β_9	0.0809	0.2111	0.2252	0.2340	0.2534	0.1493
β_{10}	0.0014	0.0016	0.0007	0.0002	0.0001	0.0001
β_{11}	0.0166	0.0128	0.0081	0.0018	0.0004	0.0003
β_{12}	0.0068	0.0001	0.0010	0.0002	0.0002	0.0003
β_{13}	0.0040	0.0004	0.0090	0.0238	0.0313	0.0259

Tab. 5.3: Select default exit parameters absolute error with 100000 simulations

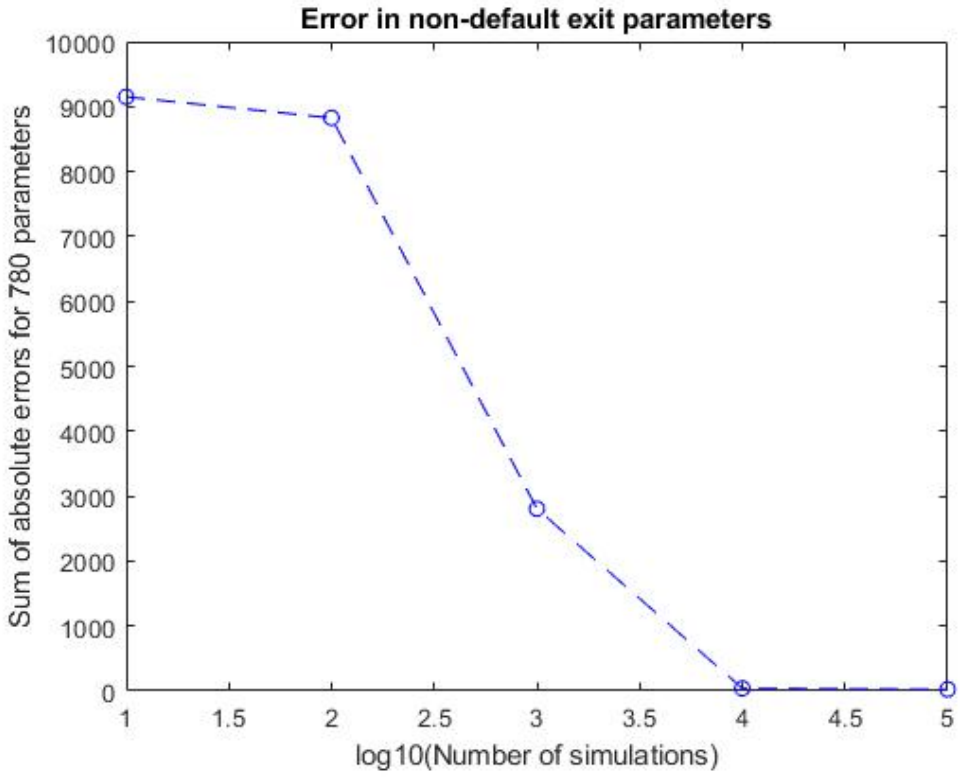

Fig. 5.2: Error in basic likelihood non-default parameter estimation

Horizon	1	12	24	36	48	60
$\bar{\beta}_1$	-4.1537	-3.9159	-3.7772	-3.7243	-3.7279	-3.7677
$\bar{\beta}_2$	-0.0344	0.0126	0.0499	0.0763	0.0943	0.1060
$\bar{\beta}_3$	12.2746	15.5544	16.4840	16.0262	14.9573	13.6840
$\bar{\beta}_4$	-0.0352	-0.0076	-0.0037	-0.0024	-0.0018	-0.0014
$\bar{\beta}_5$	-0.0004	-0.0103	-0.0155	-0.0174	-0.0175	-0.0168
$\bar{\beta}_6$	0.0804	-0.1117	-0.1618	-0.1569	-0.1382	-0.1187
$\bar{\beta}_7$	-0.1671	-0.1312	-0.1013	-0.0788	-0.0617	-0.0487
$\bar{\beta}_8$	-15.6943	-14.8039	-13.8658	-12.9611	-12.0889	-11.2481
$\bar{\beta}_9$	-2.0170	-2.5700	-3.0318	-3.3730	-3.6165	-3.7815
$\bar{\beta}_{10}$	0.0124	-0.0039	-0.0202	-0.0352	-0.0488	-0.0612
$\bar{\beta}_{11}$	-0.5746	-0.4445	-0.3380	-0.2590	-0.2002	-0.1564
$\bar{\beta}_{12}$	0.0938	0.0784	0.0641	0.0519	0.0415	0.0328
$\bar{\beta}_{13}$	2.5993	1.8345	1.3120	0.9822	0.7667	0.6207

Tab. 5.4: Select non-default exit parameters

Horizon	1	12	24	36	48	60
$\bar{\beta}_1$	0.0043	0.1267	0.0462	0.0057	0.0043	0.1729
$\bar{\beta}_2$	0.0003	0.0000	0.0002	0.0008	0.0006	0.0002
$\bar{\beta}_3$	0.0239	0.1024	0.0123	0.0805	0.0106	0.0515
$\bar{\beta}_4$	0.0004	0.0002	0.0000	0.0000	0.0000	0.0000
$\bar{\beta}_5$	0.0000	0.0001	0.0000	0.0002	0.0002	0.0001
$\bar{\beta}_6$	0.0007	0.0013	0.0013	0.0009	0.0017	0.0013
$\bar{\beta}_7$	0.0015	0.0001	0.0009	0.0006	0.0005	0.0000
$\bar{\beta}_8$	0.0561	0.2872	0.2601	0.0593	0.0352	0.2212
$\bar{\beta}_9$	0.0207	0.0207	0.0583	0.0382	0.0446	0.0202
$\bar{\beta}_{10}$	0.0001	0.0000	0.0002	0.0005	0.0004	0.0009
$\bar{\beta}_{11}$	0.0067	0.0123	0.0073	0.0031	0.0030	0.0023
$\bar{\beta}_{12}$	0.0010	0.0008	0.0012	0.0006	0.0005	0.0006
$\bar{\beta}_{13}$	0.0280	0.0226	0.0307	0.0107	0.0083	0.0060

Tab. 5.5: Select non-default exit parameters absolute error with 100000 simulations

5.2 Pseudo-likelihood Estimation

We now demonstrate the pseudo-likelihood function estimation technique that, while similar to the classical statistical frequentist's approach, has a key difference from what was discussed above. Instead of having a single set of variables representing a single firm and a large number of scenarios that we, of course, had to synthetically generate, we now have a large number of different variables, each representing a single firm, with a single scenario per firm. This can be seen as a one-world-view or real world situation from which parameters can be estimated. Collecting 100000 firms' worth of data for any length of time and then constructing our variables, assuming all of the observations are valid, would be an immense undertaking and far exceeds the scope of this study. To circumvent this problem, the data from Aspen Phamacare Holdings Limited (APN), Harmony Gold (HAR), African Rainbow Minerals Limited (ARI), Caxton and CTP Publishers and Printers (CAT), Pick n Pay (PIK), Anglo American plc (AGL), Cullinan Holdings (CUL), Mr Price Group Limited (MRP), Nedbank Group (NED), and Crookes Brothers Limited (CKS) was collected for one whole year from the beginning of January 2016 to the end of December 2016. We use the end of December observations and treat the set of 10 firm-specific variables as being from a multivariate normal distribution with its own underlying correlation structure. This resulted in the parameter estimates that appear in Table 5.6 below.

X	APN	HAR	ARI	CAT	PIK	AGL	CUL	MRP	NED	CKS
$X^{(1)}$	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158
$X^{(2)}$	0.0736	0.0736	0.0736	0.0736	0.0736	0.0736	0.0736	0.0736	0.0736	0.0736
$X^{(3)}$	3.0004	3.1663	5.4380	4.8303	5.2258	1.5492	2.5902	6.4563	**-34.4032**	5.1255
$X^{(4)}$	0.2054	-0.3844	-0.2859	-0.1923	0.2207	0.0787	-0.4424	-0.0605	4.0848	-0.1434
$X^{(5)}$	0.0561	0.1107	0.2484	0.1285	0.1441	0.4174	0.0404	0.2252	0.2907	0.2807
$X^{(6)}$	0.0025	0.0264	-0.0085	-0.0530	0.0188	0.0512	0.0009	-0.0266	0.0322	-0.0449
$X^{(7)}$	-0.0361	-0.2759	0.1166	0.0794	0.0617	0.1362	-0.0331	0.4596	0.1329	0.2954
$X^{(8)}$	0.0276	0.1247	-0.0156	0.0031	0.0108	0.0073	0.1059	-0.0093	-0.0046	-0.0051
$X^{(9)}$	1.9165	3.4687	3.9485	0.6977	-1.1197	-0.8036	1.9593	2.7274	3.5322	2.4463
$X^{(10)}$	0.0280	-0.3540	-0.2087	-0.2024	-0.0181	-0.0108	-0.4531	-0.1426	0.0803	-0.1623
$X^{(11)}$	1.3340	2.6924	4.9942	1.1893	1.2641	4.0972	0.7572	7.7950	**12.2155**	10.8280
$X^{(12)}$	0.0377	0.0373	0.0186	0.0271	0.0237	0.1418	0.0431	0.0294	0.0194	0.0197

Tab. 5.6: End of December 2016 variables for 10 firms

The actual variable values are rather inconsequential as we merely needed a distribution from which we could generate realisations. The two variable values in bold for Nedbank Group above were set to 5 as they were from a financial firm that our paper was not meant to use in our model. These two values are extreme and caused the simulation of far too many defaults, which, while beneficial to parameter estimation, is unrealistic. Finally the realisations for the final two variables generated were forced to be positive as they are the only two firm-specific variables that have to be positive. It is from this distribution that realisations are generated. We now let X_i be the constructed variables for firm i for the end of December 2016, and we can now use these variables in conjunction with our parameters generated from Tables 4.1 and 4.2 and Equation (4.10) to simulate exits up until December 2021. Our default exit times, τ_i, are therefore simulated as the minimum τ_i such that:

$$\sum_{t=0}^{\tau_i - \frac{1}{12}} \Delta t H(\beta(t), X_i) \geq -ln(U_i), \tag{5.9}$$

with the firm not defaulting if:

$$\sum_{t=0}^{l - \frac{1}{12}} \Delta t H(\beta(t), X_i) < -ln(U_i), \tag{5.10}$$

with l being the chosen boundary that we selected as 5 years as before and i being the observation or firm number. The non-default exit times, $\bar{\tau}_i$, are therefore simulated as the minimum $\bar{\tau}_i$ such that:

$$\sum_{t=0}^{\bar{\tau}_i - \frac{1}{12}} \Delta t H(\bar{\beta}(t), X_i) \geq -ln(U_i^*), \tag{5.11}$$

with the firm not defaulting if:

$$\sum_{t=0}^{l - \frac{1}{12}} \Delta t H(\bar{\beta}(t), X_i) < -ln(U_i^*), \tag{5.12}$$

with l being described as above. The sample size for both default exits and non-default exits starts at 10 and increases exponentially with base 10 up to 100000 simulations. Using the techniques described above as well as the values in Tables 4.1

and 4.2, in conjunction with Equation (4.10) and the simulated variables described above, we are able to simulate the default and non-default exits in Table 5.7 below.

Sample size	Total default exits	Total non-default exits
10	0	5
100	7	34
1000	57	258
10000	735	2732
100000	7583	27502

Tab. 5.7: Number of exits per sample size

It is important, at this point, to take note of the differences in Tables 5.1 and 5.7. It is clear to see that at every sample size and across both default exits and non-default exits there are more simulated exits. This is easily explained by the relatively favourable variables for EOH Holdings when compared to randomly simulated variables that can exhibit all manner of unfavourable characteristics that will most definitely increase the likelihood of the default and non-default exits based on our simulation technique. The parameters at each horizon, $t = \{0, \frac{1}{12}, ..., \frac{59}{12}\}$, can also be estimated separately using probabilities of the form:

$$\begin{aligned}\mathbb{P}_l^{\beta(t)}(\beta(t);\tau_i,\bar{\tau}_i,X_i) &= 1_{\{min(\tau_i,\bar{\tau}_i)>l\}}exp[-\Delta t H(\beta(t),X_i)]\\ &+ 1_{\{t<(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}exp[-\Delta t H(\beta(t),X_i)]\\ &+ 1_{\{t=(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}(1-exp[-\Delta t H(\beta(t),X_i)])\\ &+ 1_{\{t\leq(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}exp[-\Delta t H(\beta(t),X_i)]\\ &+ 1_{\{t>min((\tau_i-\frac{1}{12}),(\bar{\tau}_i-\frac{1}{12}))\}},\end{aligned} \tag{5.13}$$

for default exits and probabilities of the form:

$$\begin{aligned}\mathbb{P}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau_i,\bar{\tau}_i,X_i) &= 1_{\{min(\tau_i,\bar{\tau}_i)>l\}}exp[-\Delta t H(\bar{\beta}(t),X_i)]\\ &+ 1_{\{t<(\tau_i-\frac{1}{12})\leq(\bar{\tau}_i-\frac{1}{12}),\tau_i\leq l\}}exp[-\Delta t H(\bar{\beta}(t),X_i)]\\ &+ 1_{\{t<(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}exp[-\Delta t H(\bar{\beta}(t),X_i)]\\ &+ 1_{\{t=(\bar{\tau}_i-\frac{1}{12})<(\tau_i-\frac{1}{12}),\bar{\tau}_i\leq l\}}(1-exp[-\Delta t H(\bar{\beta}(t),X_i)])\\ &+ 1_{\{t\geq(\tau_i-\frac{1}{12}),t>(\bar{\tau}_i-\frac{1}{12}))\}},\end{aligned} \tag{5.14}$$

for non-default exits. For the pseudo-likelihood technique different variables, X_i, are used for each simulation with the parameters being used to simulate the defaults being those used for the basic likelihood estimation. The likelihood function,

Equation (5.15), is maximized separately for each horizon, t, to yield the default exit parameters:

$$\mathcal{L}_l^{\beta(t)}(\beta(t);\tau,\bar{\tau},X) = \prod_{i=1}^{I} \mathbb{P}_l^{\beta(t)}(\beta(t);\tau_i,\bar{\tau}_i,X_i), \tag{5.15}$$

with I being the total number of simulations, τ being the vector of default exit times, $\bar{\tau}$ being the vector of non-default exit times and X being the matrix of variables for all firms at the end of December 2016. The likelihood function, Equation (5.16), is similarly maximized separately for each horizon, t, to yield the non-default exit parameters:

$$\mathcal{L}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau,\bar{\tau},X) = \prod_{i=1}^{I} \mathbb{P}_l^{\bar{\beta}(t)}(\bar{\beta}(t);\tau_i,\bar{\tau}_i,X_i), \tag{5.16}$$

with the variables defined as above. In summation, we use a set of uniform random numbers to simulate default exits using Equations (5.9) and (5.10), and another set of independent uniform random numbers to simulate non-default exits using Equations (5.11) and (5.12). We use these exits in Table 5.7 in Equations (5.13) and (5.15) to estimate default exit parameters, and in Equations (5.14) and (5.16) to estimate non-default exit parameters. The cumulative absolute errors for all $780(13 \times 60)$ parameters for default exits and non-default exits are displayed in Figures 5.3 and 5.4 respectively.

It is clear from these figures that the convergence to the parameters that originally generated the exits is quicker for non-default exits than it is for default exits which mirrors Figures 5.1 and 5.2 above. This can simply be explained by the fact that, like above in Table 5.1, there are far more non-default exits generated than default exits in Table 5.7 which, again, mirrors the observations made in NUS-RMI (2016). There is no doubt that should it be possible to simulate more exits, the errors in Tables 5.8 and 5.9 would decrease further with more accurate estimates of the parameters generating the defaults being possible. This is also clearly evident from the trends observable in Figures 5.3 and 5.4.

By comparing Figures 5.1 and 5.3 it can easily be seen that the convergence for the pseudo-likelihood function is not only slower than for the basic likelihood function, but it is also non-monotonic. This is to be expected owing to the diverse set of explanatory variables for all 100000 firms. The same observation can be made with regards to Figures 5.2 and 5.4. The non-monotonic convergence of Figures 5.3 and 5.4 can be explained by the lack of default and non-default exits at certain horizons, coupled with the varying explanatory variables. This causes the parameters

at certain horizons to grow without bound and drastically increases the cumulative error. The same observation can be made about Figures 5.1 and 5.2 above, but since the variables are all identical and more default and non-default exits occur at each horizon, this non-monotonic convergence is not observed there. The zeros observable in Tables 5.8 and 5.9 below are a result of the initial guesses for the minimization functions being the parameters that were originally used to simulate the default and non-default exits. This, in combination with the varying firm variables and the tolerances of the minimization function, has lead the initial guess to be accepted as the maximum likelihood parameters which is accepted as being reasonable. The tolerances could be adjusted, but the intent is to demonstrate the relationship between simulation and estimation.

We postulate that if we were able to increase the number of observations further to several millions or even billions the differences would become increasingly minute. If it were possible to simulate infinitely many firms all of the idiosyncratic risk would be diversified away and the effect of the firm-specific variables would be akin to random noise in the model, with the exit probabilities being solely driven by the common variables that serve as proxies for the general state of the economy. By referring to Figures 5.3 and 5.4, it is clear that despite us knowing the exact underlying structure that underpins the exit simulation technique and mimicking these assumptions in our estimation methodology, we still have significant errors at all sample sizes. From these figures, the necessity for a more accurate estimation technique is clear.

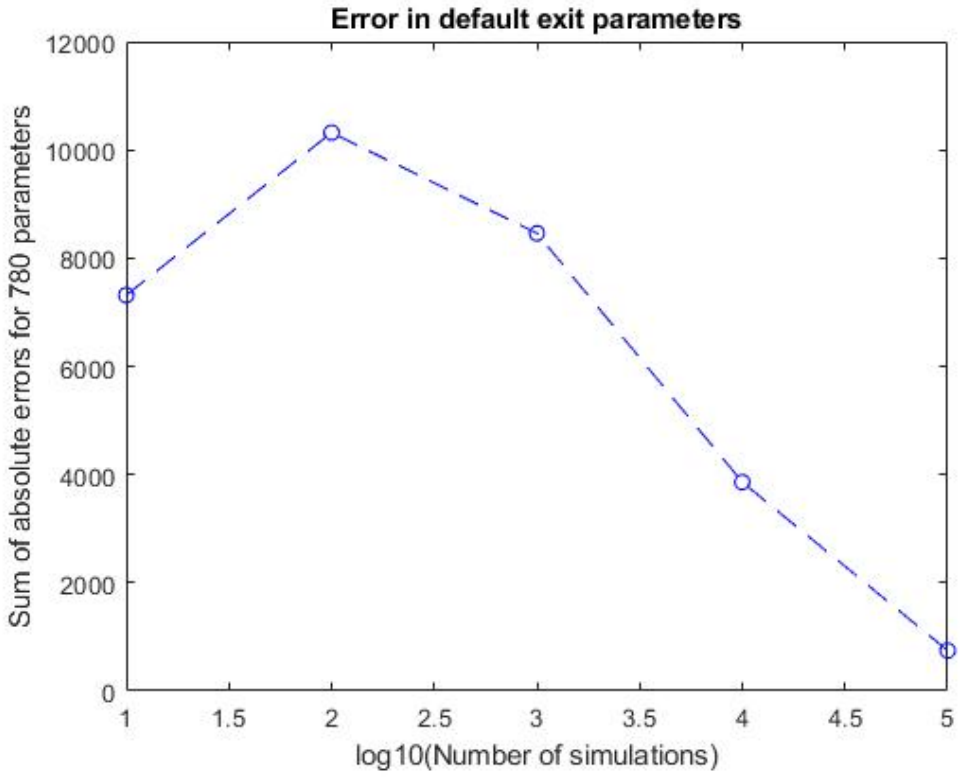

Fig. 5.3: Error in pseudo-likelihood default parameter estimation

Horizon	1	12	24	36	48	60
β_1	0.0000	0.5146	0.6928	0.2732	0.8105	1.8726
β_2	0.0000	0.0004	0.0182	0.0013	0.0101	0.0347
β_3	0.0000	4.7946	0.7722	4.0603	0.7747	1.3806
β_4	0.0000	0.1505	0.0655	0.0127	0.1353	0.1932
β_5	0.0000	0.1109	0.1724	0.2204	0.1094	0.0523
β_6	0.0000	0.0319	0.0196	0.0094	0.0112	0.0207
β_7	0.0000	0.0769	0.0571	0.0182	0.0490	0.1579
β_8	0.0000	1.9779	7.1779	3.5894	1.4467	10.3549
β_9	0.0000	3.0100	6.5898	4.2814	2.4349	15.6135
β_{10}	0.0000	0.0729	0.0030	0.0191	0.0013	0.0109
β_{11}	0.0000	0.0068	0.0486	0.0074	0.0019	0.0141
β_{12}	0.0000	0.1653	0.0153	0.0008	0.0059	0.0075
β_{13}	0.0000	0.0085	0.2167	0.2086	0.2905	0.2573

Tab. 5.8: Select default exit parameters absolute error with 100000 simulations

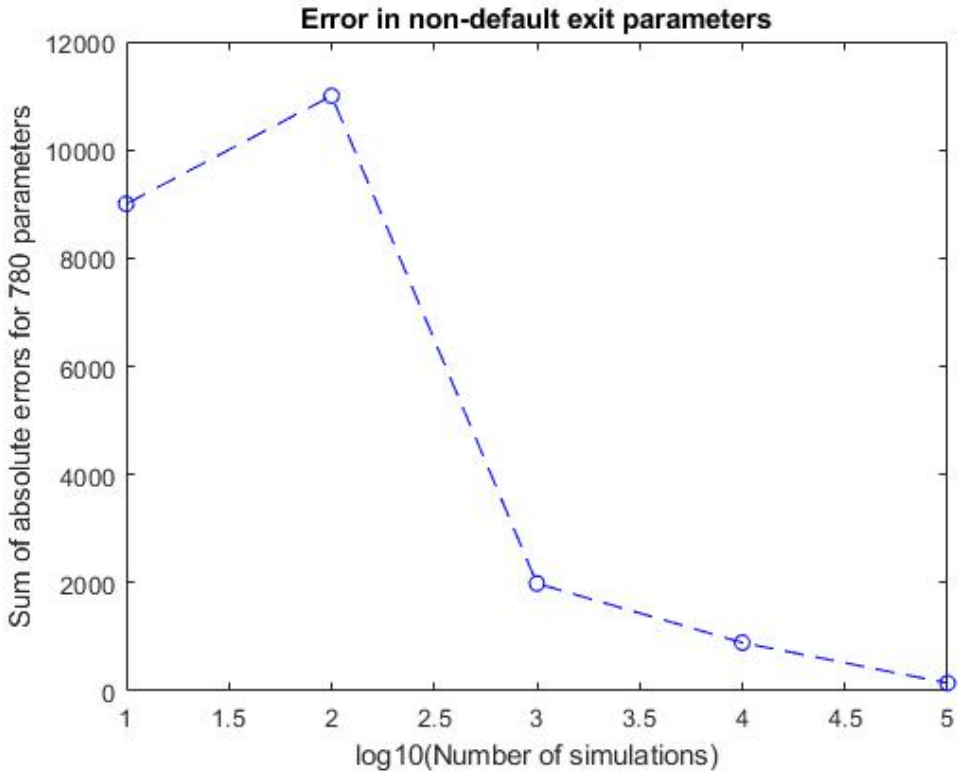

Fig. 5.4: Error in pseudo-likelihood non-default parameter estimation

Horizon	1	12	24	36	48	60
$\bar{\beta}_1$	0.0000	0.0000	0.0000	0.0000	0.0882	0.6381
$\bar{\beta}_2$	0.0000	0.0000	0.0000	0.0000	0.0450	0.0394
$\bar{\beta}_3$	0.0000	0.0000	0.0000	0.0000	0.8718	12.8695
$\bar{\beta}_4$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005
$\bar{\beta}_5$	0.0000	0.0000	0.0000	0.0000	0.0029	0.0038
$\bar{\beta}_6$	0.0000	0.0000	0.0000	0.0000	0.0158	0.0713
$\bar{\beta}_7$	0.0000	0.0000	0.0000	0.0000	0.0041	0.0089
$\bar{\beta}_8$	0.0000	0.0000	0.0000	0.0000	0.3791	1.8885
$\bar{\beta}_9$	0.0000	0.0000	0.0000	0.0000	0.6266	6.7842
$\bar{\beta}_{10}$	0.0000	0.0000	0.0000	0.0000	0.0042	0.0028
$\bar{\beta}_{11}$	0.0000	0.0000	0.0000	0.0000	0.2351	0.1614
$\bar{\beta}_{12}$	0.0000	0.0000	0.0000	0.0000	0.0004	0.0104
$\bar{\beta}_{13}$	0.0000	0.0000	0.0000	0.0000	0.1401	0.0504

Tab. 5.9: Select non-default exit parameters absolute error with 100000 simulations

Chapter 6

Sequential Monte Carlo Pseudo-Bayesian Estimation

Parameter smoothing is done in Duan *et al.* (2012) where a least squares fit of the maximum pseudo-likelihood parameter estimates for each sequence of β and $\bar{\beta}$ is done to the Nelson-Siegel function, Equation (4.10). It seems to a natural extension to impose some functional form to model the time dynamics of the parameters. We assume there is some underlying structure, but by adding additional structure we will most certainly introduce large errors at numerous horizons. It also seems natural to assume the long term stability of the parameters, but there is little evidence to support this conjecture that adds an addition potential source of error. We follow this method that is repeated in NUS-RMI (2016). Both papers only allow ρ_0 to be non-zero for β_1 and $\bar{\beta}_1$. This has the advantage of reducing the number of parameters that need to be estimated from 1560 $(2 \times 13 \times 60)$ to only 80 $(2 \times (12 \times 3 + 1 \times 4))$ parameters. An additional advantage is that this allows the default and non-default exit parameters to be extrapolated to periods over 5 years, with added risk of nonsensical parameters the further forward they are extrapolated. While this is the final output of the work done in Duan *et al.* (2012), we will use these 80 parameter estimates as the means in our prior distributions, $\pi(\theta)$, that are then used to generate an initial particle cloud that will serve as the starting point for our algorithm. Our chosen prior distributions are normal distributions, with our d parameters being from truncated normal distributions to ensure that $d > 0$, which ensures long-term stability of the parameters.

Without loss of generality the following demonstrates how the default exit parameters are estimated using our algorithm, with the non-default exit parameters being estimated in an analogous manner. The work done in NUS-RMI (2016) follows closely the work done in Duan and Fulop (2013). The algorithm follows the sequential resampling routine done in Chopin (2002), but we did not include the tempering sequence first done in Del Moral *et al.* (2006), and repeated in Duan and

Fulop (2013) and NUS-RMI (2016). We felt that the tempering sequence would be an unnecessary addition and complication given the scope of the study, but we have intentions to include it in future work on the subject. We initialise our, $K = 1000$, particles, each of 40 parameters, with equal weighting which follows that done in NUS-RMI (2016) exactly: $\left(\theta^{(k,0)} \sim \pi(\theta), \omega^{(k,0)} = \frac{1}{K}\right)$. The standard deviations for the priors are set to equal 0.1. The standard deviations for the priors in our work are significantly smaller than those used in NUS-RMI (2016). We do this because our likelihood function is extremely sensitive to even the smallest changes in the parameter values, and all but a few particles generated cause the weights in the first reweighting step to equal essentially zero. Once the algorithm has stepped through the first set of data, the standard deviations are increased as the particles are now generated around a far more likely area. We would like to, at this point, introduce some notation. We deal with pseudo-posterior distributions at each step of the algorithm. A difference between the work done in NUS-RMI (2016) and that done here is that while the algorithm steps through time in NUS-RMI (2016), our algorithm simply steps through batches of 10 samples from the same time point. The pseudo-posterior distributions are of the form:

$$\gamma_q(\theta) \propto \prod_{r=10}^{q} \mathcal{L}_{(r,l)}(\theta)\pi(\theta). \tag{6.1}$$

Each pseudo-likelihood function in Equation (6.1) above is of the form:

$$\mathcal{L}_{(r,l)}(\theta) = \prod_{i=(r-9)}^{r} \mathbb{P}_l^{\beta}(\beta; \tau_i, \bar{\tau}_i, X_i), . \tag{6.2}$$

Each probability in Equation (6.2) above is of the form:

$$\begin{aligned}
\mathbb{P}_l^{\beta}(\beta; \tau_i, \bar{\tau}_i, X_i) =& 1_{\{min(\tau_i,\bar{\tau}_i)>l\}} exp\left[-\Delta t \sum_{j=0}^{l-\frac{1}{12}} H(\beta(j), X_i) \right] \\
&+ 1_{\{\tau_i \leq \bar{\tau}_i, \tau_i \leq l\}} exp\left[-\Delta t \sum_{j=0}^{\tau_i - m - \frac{2}{12}} H(\beta(j), X_i) \right] \\
&\times \left(1 - exp\left[-\Delta t H\left(\beta\left(\tau_i - m - \frac{1}{12} \right), X_i \right) \right] \right) \\
&+ 1_{\{\bar{\tau}_i < \tau_i, \bar{\tau}_i \leq l\}} exp\left[-\Delta t \sum_{j=0}^{\bar{\tau}_i - m - \frac{1}{12}} H(\beta(j), X_i) \right].
\end{aligned} \tag{6.3}$$

The analogous probabilities for non-default exits are of the following form:

$$\begin{aligned}\mathbb{P}_l^{\bar{\beta}}(\bar{\beta};\tau_i,\bar{\tau}_i,X_i) =& 1_{\{min(\tau_i,\bar{\tau}_i)>l\}} exp\left[-\Delta t \sum_{j=0}^{l-\frac{1}{12}} H(\bar{\beta}(j),X_i)\right] \\ &+ 1_{\{\tau_i\leq\bar{\tau}_i,\tau_i\leq l\}} exp\left[-\Delta t \sum_{j=0}^{\tau_i-m-\frac{2}{12}} H(\bar{\beta}(j),X_i)\right] \\ &+ 1_{\{\bar{\tau}_i<\tau_i,\bar{\tau}_i\leq l\}} exp\left[-\Delta t \sum_{j=0}^{\bar{\tau}_i-m-\frac{2}{12}} H(\bar{\beta}(j),X_i)\right] \\ &\times\left(1-exp\left[-\Delta t H\left(\bar{\beta}\left(\bar{\tau}_i-m-\frac{1}{12}\right),X_i\right)\right]\right).\end{aligned} \tag{6.4}$$

Upon examining Equation (6.1) one can deduce the following recursive relationship:

$$\gamma_q(\theta) \propto \mathcal{L}_{(q,l)}(\theta)\gamma_{(q-10)}(\theta). \tag{6.5}$$

6.1 Reweighting Step

Having initialised the seed for the algorithm, we progress from particles representing $\gamma_{(q-10)}(\theta)$ to particles representing $\gamma_q(\theta)$ by using the importance sampling principle:

$$\theta^{(k,q)} = \theta^{(k,(q-10))}, \tag{6.6}$$

$$\begin{aligned}\omega^{(k,q)} &= \omega^{(k,(q-10))} \times \frac{\gamma_q(\theta^{(k,q)})}{\gamma_{(q-10)}(\theta^{(k,q)})} \\ &= \omega^{(k,(q-10))}\mathcal{L}_{(q,l)}(\theta).\end{aligned} \tag{6.7}$$

As we step through this algorithm it is clear that the weights will inevitably be concentrated on a few particles. To avoid the pseudo-posterior distribution being represented by only a handful of particles that will make any inference questionable, a resample step and move step is performed whenever the efficient sample size drops below some critical value, B. The efficient sample size (ESS) is defined as:

$$ESS = \frac{\left(\sum_{k=1}^{K}\omega^{(k,q)}\right)^2}{\sum_{k=1}^{K}\left(\omega^{(k,q)}\right)^2}. \tag{6.8}$$

Our critical value is chosen to be half the size of the number of particles representing our pseudo-posterior distribution such that $B = 500$. Therefore, should a set of particles and weights arise that represent the current pseudo-posterior distribution with $ESS < B$, the following steps are performed.

6.2 Resampling Step

For each k, the particle is resampled such that:

$$\theta^{(k,q)} = \theta^{(j,q)}, \tag{6.9}$$

with probability:

$$\mathbb{P}\Big(\theta^{(k,q)} = \theta^{(j,q)}\Big) = \frac{\omega^{(j,q)}}{\sum_{k=1}^{K} \omega^{(k,q)}}, \tag{6.10}$$

with $j \in \{1, 2, ..., K\}$. This step ensures that the new particle cloud is concentrated around a more likely area, with the particles that have higher weights being translated into numerous repetitions of those particles. To not be left with the same problem as before, where the distribution is represented by only a few particles, the move step is performed.

6.3 Move Step

For each k, a new potential particle, θ^{*k}, is sampled from our Metropolis-Hastings kernel, $\mathcal{M}_q(\cdot|\theta^{(k,q)})$. We have chosen the same distributions that serve as our prior distributions, with the means being changed to the current particle value ($\theta^{(k,q)}$):

$$\theta^{*k} \sim \mathcal{M}_q(\cdot|\theta^{(k,q)}). \tag{6.11}$$

The standard deviations of these distributions have been increased to 0.3 to allow for a more diverse particle set than the one generated by the prior distributions. This new particle, if selected, will leave the pseudo-posterior distribution unaltered. To decide whether the new particle is accepted, we compute the probability of accepting the new particle, α, which is referred to as the acceptance weight:

$$\alpha = min\left\{1, \frac{\gamma_q(\theta^{*k})\mathcal{M}_q(\theta^{(k,q)}|\theta^{*k})}{\gamma_q(\theta^{(k,q)})\mathcal{M}_q(\theta^{*k}|\theta^{(k,q)})}\right\}. \tag{6.12}$$

Finally, the particle is reset such that:

$$\theta^{(k,q)} = \theta^{*k}, \tag{6.13}$$

with probability:

$$\mathbb{P}\left(\theta^{(k,q)} = \theta^{*k}\right) = \alpha. \tag{6.14}$$

This algorithm is rather elegant in that the Metropolis-Hasting kernel from which new particles are drawn is continuously updated with the current particle cloud. This ensures that should a particular sample be a poor representation of our target distribution, the move step will shift the concentration of the particles.

6.4 Pseudo-Bayesian Inference

Once the algorithm has run through for different sample sizes, the different pseudo-posterior means can be calculated and form our new estimate. That is, for $q = 10, 100, 1000, 10000, 100000$, our new estimate for our parameters is:

$$\hat{\theta}_q = \frac{1}{\sum_{k=1}^{K} \omega^{(k,q)}} \sum_{k=1}^{K} \omega^{(k,q)} \theta^{(k,q)}. \tag{6.15}$$

The variables and estimates for the pseudo-likelihood estimation technique discussed in the previous chapter were used as starting points for the work presented in this chapter. This can be made clearer to the reader by comparing Figures 5.3 and 5.4 with Figures 6.1 and 6.2 respectively. In Figures 6.1 and 6.2 below, we compare the cumulative error over all 780 parameters for the different estimation techniques for default exits and non-default exits respectively. It is clear that the work done in Duan *et al.* (2012), whereby a least squares fit of Equation (4.10) is done to the parameters, that is repeated here, does significantly improve upon the error in our estimation of the parameters at all sample sizes.

This result, however, should be taken with a pinch of salt as the parameters used to simulate the default exits and non-default exits are from the NUS-RMI website where they are already made to fit the Nelson-Siegel function with the parameters in Tables 4.1 and 4.2 respectively. This least squares fit merely serves to produce an initial seed from which the Sequential Monte Carlo algorithm, used in Duan and Fulop (2013), can start, but we felt the need to include the reduction in error from this fit in our results so as not to mislead the reader as to the scale of the effectiveness of our algorithm in reducing the errors. That being said, it is clear from Figures 6.1 and 6.2 below that, especially at small sample sizes, the algorithm makes a significant improvement upon the estimate we arrive at by simply fitting the Nelson-Siegel function to our pseudo-likelihood parameter estimates.

The significance of this result is that, while the cumulative errors converge after a large number of simulations, the data available from insurance companies, banks,

and the financial sector as a whole is difficult to come by, and using this technique with a smaller data set clearly improves upon the result from likelihood estimation techniques. The reader might argue that this is as a result of the original functional form of the parameters, but consider that this algorithm was tested on only one functional form. This could be extended to any number of realistic functional forms for the default exit and non-default exit parameters. The functions do not even necessarily need to be of the same type for default exits and non-default exits, which makes sense given the vastly different factors that may drive the assumed to be independent processes.

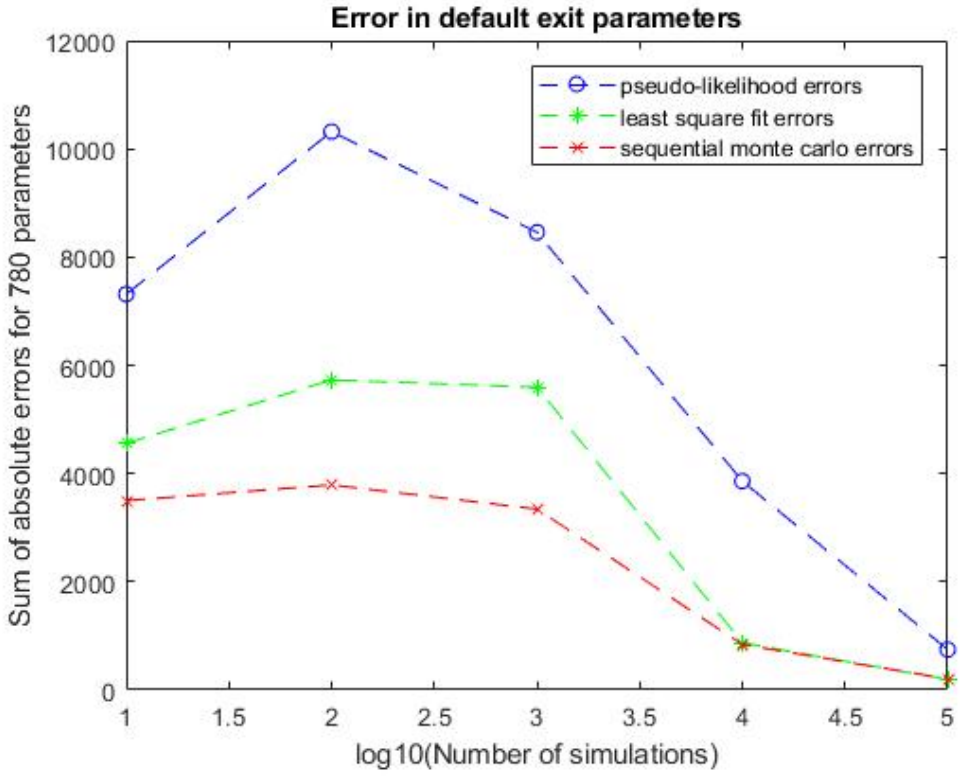

Fig. 6.1: Comparison of errors for default parameter estimation

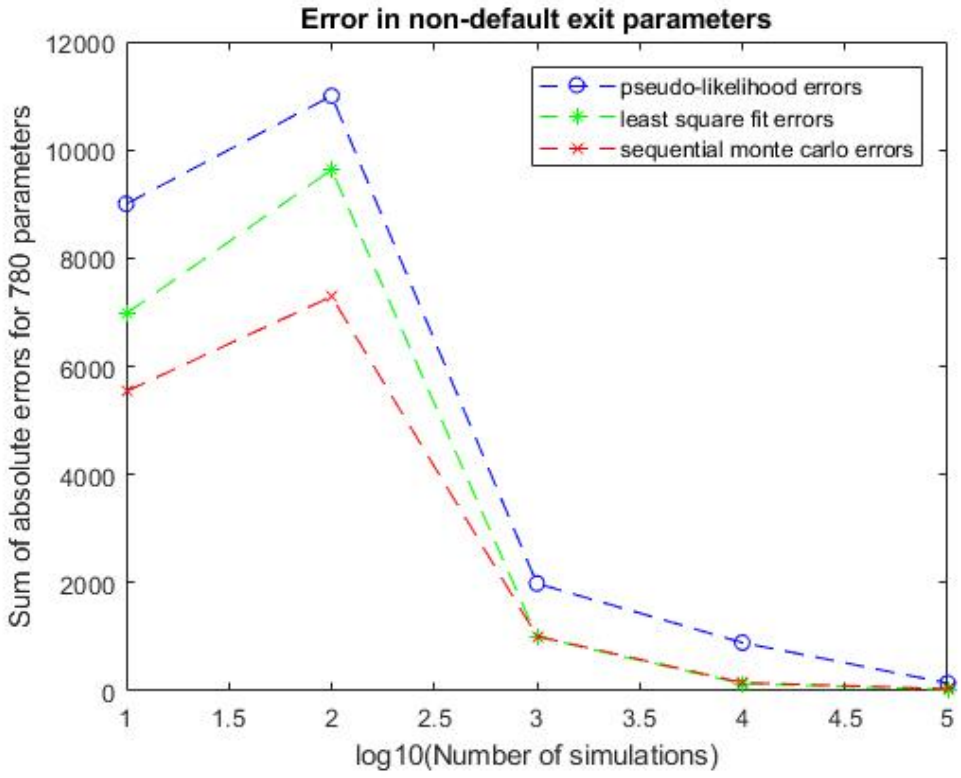

Fig. 6.2: Comparison of errors for non-default parameter estimation

6.5 Monte Carlo Simulation

While the above result is highly favourable, we must ensure that it is not merely an artefact of this single set of simulations. We do this by simulating additional variables and exits a number of times and creating Monte Carlo error bounds about the averages of the simulations. In order to simulate 50 independent events such as that depicted above, the simulation was left running for approximately 4 days. We believe the results support our claim that the Sequential Monte Carlo algorithm significantly improves upon the pseudo-likelihood estimation at all sample sizes. The algorithm also significantly improves upon the result from the least squares fit of the Nelson-Siegel function at small sample sizes. These results are most evident for the default exit simulation, with the error bars indicating a 3 standard deviation error bound around the means. For Figure 6.3 below, the error bounds for the algorithm are almost impossible to discern, and the cumulative errors for this method are far lower than that of the pseudo-likelihood estimation technique and the least squares fit with their respective error bounds also far exceeding those of our algorithm.

This result is less conclusive in Figure 6.4 for the non-default exit parameter estimation, but several favourable conclusions can still be drawn. The same trend observable in Figure 6.3 below can also be seen in Figure 6.4 in that the Sequential Monte Carlo algorithm drastically and visibly improves upon the cumulative error for all 780 parameters from the pseudo-likelihood estimation technique at all sample sizes. This difference is far less pronounced between the least squares fit of Equation (4.10) to the parameters and the algorithm that uses this fit as an initial seed. There is, however, a visible improvement upon the least squares fit at small sample sizes. The trend observable in both Figure 6.3 and Figure 6.4 is that as the sample sizes are increased, and the number of default and non-default exits also increase based on Table 5.1 and Table 5.7, the cumulative errors for all methods begin to converge to zero rather quickly. This is as expected, but the crux of our result is the significant improvement in the estimation of the parameters at small sample sizes, which is what we would have to deal with in the industry. We are satisfied with this result and look forward to future improvements and results.

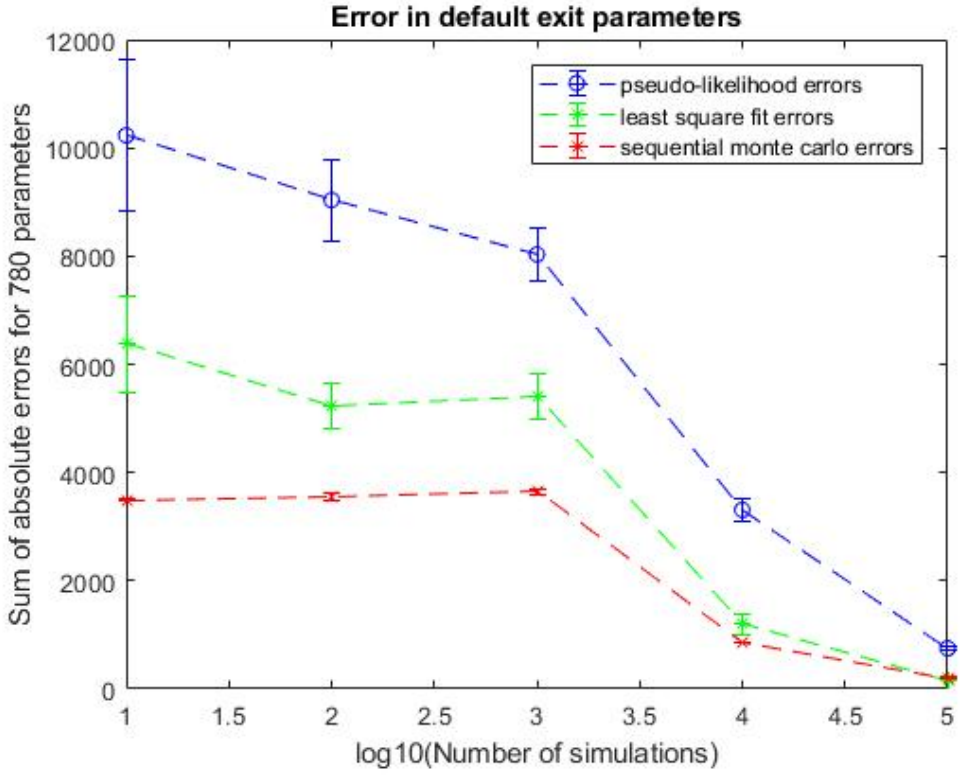

Fig. 6.3: Monte Carlo errors for default parameter estimation

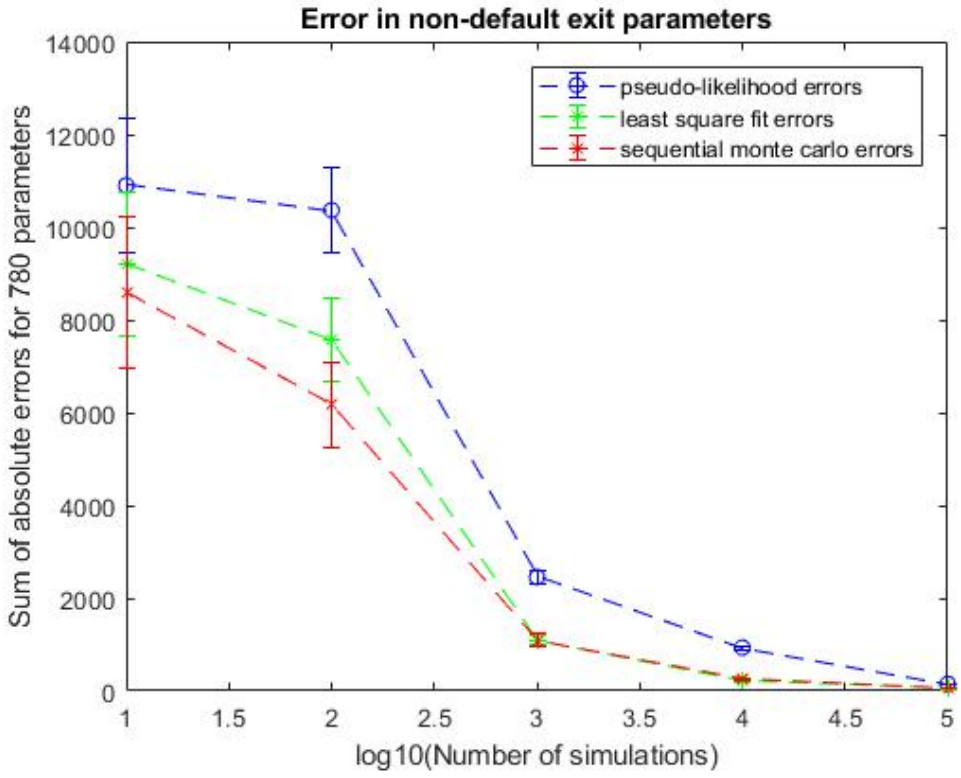

Fig. 6.4: Monte Carlo errors for non-default parameter estimation

Chapter 7

Conclusion

We believe that our paper has demonstrated our primary objective in that the forward intensity approach to modelling a firm's probability of default over multiple time increments using numerous firm-specific and macroeconomic variables used in NUS-RMI (2016) yields promising results. This method can clearly be extended to portfolio risk assessment as well as the pricing of corporate debt. While not demonstrated here owing to the limitations of the simulated data, the model can be updated monthly with new data which can continually improve the accuracy of the predictive ability of the model. The assumptions used in this model are somewhat restrictive, but we believe that they are far less so than the high dimensional autoregressive assumptions needed to model probabilities of default used in Duffie *et al.* (2007). We feel that the reader would benefit from a few remarks about the nature of our results.

By comparing Figures 5.1 and 5.2 with Figures 5.3 and 5.4 respectively, the reader can clearly see the slower convergence of the one-world-view pseudo-likelihood estimation when compared against the classical statistical frequentist likelihood estimation. The initial improvement in our estimates by least squares fitting Equation (4.10) to the parameters seen in Figures 6.1 and 6.2 can be explained by the few default and non-default exits at small sample sizes and the parameters used to initially simulate the exits already fitting Equation (4.10) based on the output from the NUS-RMI website. The small number of exits causes the parameter estimates to grow without bound at certain horizons, where default exits or non-default exits have not occurred, leading to completely erroneous estimates that are very much improved upon by the least squares fit of the Nelson-Siegel function.

The algorithm further improves upon the initial improvement from the least squares fit, with the most evident improvement being at small sample sizes which unambiguously reveals its importance in the real world setting where default data is scarce. The Monte Carlo simulation results in Figures 6.3 and 6.4 reaffirms our conclusions with the non-monotonic convergence seen in Figures 6.1 and 6.2 merely

being an artefact of that particular simulation with assumingly extreme variable values. Being able to collect data for 100000 firms over any length of time is a Herculean task, but our results from the Sequential Monte Carlo algorithm reveal improved convergence at small sample sizes. This will allow more accurate estimation of default and non-default exit parameters with a smaller data set leading to improved estimates for a firm's probability of default.

Chapter 8

Future Developments

Should we further our research in this topic, there are numerous avenues and extensions we wish to explore. As noted in the introduction, the variables used in our hazard rate function are those found to be relevant in the work done in Duan *et al.* (2012) for firms in the United States economy. This was then extended to over one hundred separate economies in various stages of development throughout the world in the work done in NUS-RMI (2016). Our work uses the same variables used in NUS-RMI (2016), but, ideally, we would first put a great deal of work into establishing which variables are most relevant in predicting default in the South African economy. We would, obviously, require a large amount of data from companies within the South African economy that are listed on the JSE. If we were to acquire this data, an additional extension of the work would be classifying what constitutes a default exit and what constitutes a non-default exit. With actual data from the economy there would be no need to make any erroneous assumptions about the dynamics of the variables in order to simulate realisations of these variables.

The choice for the functional form of the hazard rate is clear, owing to the necessity for the hazard rate to be non-negative at all times to induce positive probabilities of defaults or non-default exits. This has an unintentional side-effect of making the probabilities (and likelihood functions) extremely sensitive to even the smallest changes in either the parameter values or the variable values. This has a significant effect on the reweighting step as the weights for certain combinations of parameters approach the lower limit for non-zero numbers rather quickly and make further inference rather difficult. We would therefore like to explore less extreme functional forms.

The motivation for the use of the Nelson-Siegel function to model the time dynamics of the default exit and non-default exit parameters is not clear in the work done in Duan *et al.* (2012) and NUS-RMI (2016). The parameters do display regularity over time, but the Nelson-Siegel function, first introduced in Nelson and Siegel (1987), was originally used to model the term structure of interest rates. As

such, we would like to explore other functional forms that could better capture the dynamics of the these parameters. The necessity for a functional form is clear when it comes to potential extrapolations and interpolations of the work, but we conjecture that this may lead to significant errors in the South African market with its 'developing' designation in the work done in NUS-RMI (2016) and the relatively few firms listed on the JSE when compared to the United States economy or other larger, developed economies.

We would additionally like to review the treatment of missing and extreme values within our sample. We have greatly oversimplified the treatment of these variables, owing to the scope of this study. Different levels of winsorization are explored in the work done in NUS-RMI (2016), with this not being necessary for our work, owing to the possibly erroneous assumption that the descriptive variables are multivariate-normally distributed. The normality assumption in conjunction with the imposed upper limit on the sample size, greatly reduces the probability of exceptionally large outliers that would need to be winsorized. Additionally, the treatment of missing variables and their classification has been greatly simplified due to the relatively small simulated data set. The work done in Duan *et al.* (2012) and NUS-RMI (2016) use different frequencies of data when calculating the SIGMA and DTD variables, but both have clear definitions for classifying variables as missing. The classification and treatment of these missing variables would undoubtedly add an additional level of complexity to this study.

Finally, as already mentioned in the previous Chapter 6, we have greatly simplified the Sequential Monte Carlo algorithm from its presentation and use in NUS-RMI (2016). We would like to include the tempering sequence in our algorithm that was initially introduced in Del Moral *et al.* (2006). We suspect that this will assist with our weighting of particles and would serve the same purpose as quasi-random numbers in better distributing weights and particle significance. We would additionally like to run the resampling step and move step multiple times within a single iteration until the efficient sample size has moved above its critical value. This would create a more diverse particle set with better weighting much earlier in the algorithm. Finally, we would like to vary the standard deviation of the Metropolis-Hastings kernel distribution more throughout the execution of the algorithm. This would increase the range of particles that are given an opportunity to be selected as potential candidates to represent the underlying distribution. These are all inclusions that we believe would further improve our results.

Bibliography

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* **23**(4): 589–609.

Bastos, J. (2010). Forecasting bank loans loss-given-default, *Journal of Banking & Finance* **34**(10): 2510–2517.

Beaver, W. (1966). Financial Ratios as Predictors of Failure, *Journal of Accounting Research* **4**: 71–111.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*, Taylor & Francis Group.

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

Chava, S. and Jarrow, R. (2004). Bankruptcy Prediction with Industry Effects, *Review of Finance* **8**: 537–569.

Chopin, N. (2002). A sequential particle filter method for static models, *Biometrika* **89**(3): 539–551.

Crosbie, P. and Bohn, J. (2003). Modeling Default Risk, *Moody's KMV Company* .

Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers, *Journal of the Royal Statistical Society: Series B* **68**(3): 411–436.

Duan, J. (1994). Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance* **4**(2): 155–167.

Duan, J. (2000). Correction: Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance* **10**(4): 461–462.

Duan, J. (2010). Clustered Defaults, *National University of Singapore Working Paper* .

Duan, J. and Fulop, A. (2013). Multiperiod Corporate Default Prediction with the Partially-Conditioned Forward Intensity, *National University of Singapore Working Paper* .

Duan, J., Sun, J. and Wang, T. (2012). Multiperiod corporate default prediction - A forward intensity approach, *Journal of Econometrics* **170**(1): 191–209.

Duffie, D., Saita, L. and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* **83**(3): 635–665.

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance* **29**(2): 449–470.

Nelson, C. and Siegel, A. (1987). Parsimonious Modeling of Yield Curves, *Journal of Business* **60**(4): 473–489.

NUS-RMI (2016). NUS-RMI Credit Research Initiative Technical Report Version: 2016 Update 1, *Global Credit Review* **6**(1): 49–132.

Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* **18**(1): 109–131.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *Journal of Business* **74**(1): 101–124.

Zmijewski, M. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research* **22**(Supplement): 59–82.

Appendix

Appendix A

By applying Itō's Lemma to Equation (2.1) we obtain that $ln(V_t)$ is normally distributed with mean $\left(\mu - \frac{\sigma^2}{2}\right)t$ and variance $\sigma^2 t$. We therefore have the density for $ln(V_t)$:

$$f_{V_t}(v) = \frac{1}{\sqrt{2\pi\sigma^2 t}} exp\left(-\frac{\left(v - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right).$$

We now use the method of transformations to find the density of:

$$E_t = h(V_t) = V_t N(d_{t+}) - Le^{-r(T-t)}N(d_{t-}).$$

We can do this as $h(v)$ is increasing for all v such that $f_{V_t}(v) > 0$, which a derivation similar to the calculation of the Black-Scholes delta will verify. We therefore know that $h(\cdot)$ is one-to-one and therefore an inverse exists such that $v = h^{-1}(u)$. We can now calculate the density of E_t using the method of transformations and Jacobians. It must be noted that $h^{-1}(\cdot)$ is analogous to the inverse of Equation (2.2). We therefore have:

$$f_{E_t}(u) = f_{V_t}[h^{-1}(u)]\left|\frac{dh^{-1}(u)}{du}\right|.$$

From which it follows that:

$$f_{E_t}(u) = \frac{1}{\sqrt{2\pi\sigma^2 t}} exp\left(-\frac{\left(v - \left(\mu - \frac{\sigma^2}{2}\right)t\right)^2}{2\sigma^2 t}\right)\frac{1}{N(d_{t+})}.$$

Appendix B

Taking the partial derivative of Equation (2.4) with respect to μ we get:

$$\frac{\partial}{\partial\mu}\mathcal{L}(\mu,\sigma,\delta) = -\frac{1}{2\sigma^2}\sum_{t=2}^{n}\frac{1}{h_t}\left(-2h_t ln\left(\frac{\hat{V}_t(\sigma,\delta)}{A_t}\times\frac{A_{t-1}}{\hat{V}_{t-1}(\sigma,\delta)}\right) + 2\left(\mu - \frac{\sigma^2}{2}\right)h_t^2\right).$$

If $(\hat{\mu},\hat{\sigma},\hat{\delta})$ is an optimal solution to the maximization problem, then by setting the above equation equal to zero we can determine that:

$$\hat{\mu} = \frac{\hat{\sigma^2}}{2} + \frac{1}{\sum_{t=2}^{n} h_t} ln\left(\frac{\hat{V}_n(\hat{\sigma},\hat{\delta})}{A_n}\times\frac{A_1}{\hat{V}_1(\hat{\sigma},\hat{\delta})}\right).$$

We can substitute this into Equation (2.4) to arrive at our new maximization problem in Equation (2.5) of two variables.