Cass Business School
CITY UNIVERSITY LONDON

This thesis is Submitted to the Department of Finance, Cass Business School, City University London

For the degree of Doctor of Philosophy in Finance

Development and Calibration of Relative Value Trading Models

Author:
Hugh Alexander Patience

Supervisor:
Prof. John Hatgioannides

October 2015

Contents

List of Figures

List of Tables

Acknowledgments

Firstly I would like to thank Professor John Hatgioannides for all the help and support that he has given me in the course of my studies. I benefitted greatly from his knowledge and guidance, quite simply it was an honour to have been his PhD student. Secondly I would like to thank my parents, for their unconditional love and support. Lastly I would like to thank Kwan, without your love and support this PhD would not have been possible.

Chapter 1

Introduction

Abstract

This thesis presents research into the development and calibration of relative value fixed income trading models. The first chapter provides some background into the models studied, chapters two and three focus on calibration problems relating to an earlier version of the model: the relative value Nelson Siegel and Svensson model (rv-NSS). Chapter four introduces a more advanced version of the model, the relative value Dynamic Nelson and Siegel model (rv-DNS). Chapter five draws overall conclusions and discusses avenues for further research.

Contributions to the literature:

- Chapter 2
  - Shows that Differential Evolution could be successfully applied to calibrate the rv-NSS
- Chapter 3
  - Compares the widest set of ridge regression estimators ever assembled
  - Modified (r-k) Class Ridge Regression (MCRR) did not specify how to estimate all of its parameters, two methods to address this were introduced
  - Improved Ridge Estimators (IRE) had convergence problems, chapter three tries to address these succeeding in the majority of conditions tested
  - Linearized Ridge Regression Estimator (LRRE) had estimation problems at the lowest volatility levels, an attempt was made to fix this
  - Although no one estimator dominated in every scenario tested the LRRE came closest to fulfilling that goal
- Chapter 4
  - Introduces a dynamic relative value trading model based on the Dynamic Nelson Siegel Model (DNS) introduced by Diebold and Li (2006)
  - This is the only relative value trading model based on the DNS
  - Successfully tests the model on simulated and real data

Overall the thesis successfully introduces a functioning relative value fixed income trading model based on the Nelson and Siegel approach.

1.1 Introduction

This thesis centers on three chapters which address the issue of relative value trading in the fixed income market using models based on the Nelson and Siegel approach. Chapters two and three tackle calibration problems relating to an earlier version of the model while the fourth chapter introduces an augmented version of the model that looks to incorporate some of the more recent additions to the literature. This chapter gives some background research into the underlying models and associated calibration problems. Each of the central empirical chapters was written to be self contained, so there will be some overlap with the material presented in this chapter and those later chapters. The final chapter summarizes the main findings and gives avenues for future research.

1.1.1 Nelson Siegel Yield Curve Model

To determine relative values, it is first necessary to model the underlying zero coupon yield curve. The model proposed by Nelson and Siegel (1987) and its extension by Svensson (1994) is a well known method for estimating the zero coupon rates which is popular with policy makers with nine out of thirteen central banks reportedly using the Nelson and Siegel and/or the Svensson variant BIS (2005).

The original Nelson-Siegel model consists of three basis functions, which when combined are sufficiently flexible to allow most commonly encountered yield curve shapes to be modelled. The Nelson-Siegel-Svensson (NSS) model includes an additional factor.

The NSS can be defined mathematically as follows:

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1-exp\left(-\frac{\tau}{\lambda_1}\right)}{\left(\frac{\tau}{\lambda_1}\right)} \right) + \beta_3 \left(\frac{1-exp\left(-\frac{\tau}{\lambda_1}\right)}{\left(\frac{\tau}{\lambda_1}\right)} - exp\left(\frac{\tau}{\lambda_1}\right) \right) + \beta_4 \left(\frac{1-exp\left(-\frac{\tau}{\lambda_2}\right)}{\left(\frac{\tau}{\lambda_2}\right)} - exp\left(\frac{\tau}{\lambda_2}\right) \right) \quad (1.1.1)$$

where:

$\tau =$ maturity of the instrument being modelled

$y(\tau) =$ zero coupon rate at the maturity indicated by τ

$\beta_1 =$ loading factor on the first basis function

$\beta_2 =$ loading factor on the second basis function

$\beta_3 =$ loading factor on the third basis function

$\beta_4 =$ loading on the fourth basis function

$\lambda_1 =$ Determines the shape of the second and third basis functions

$\lambda_2 =$ Determines the shape of the fourth basis function

The first three basis functions represent the NS model, the addition of the fourth factor then gives the NSS model. The basis functions are illustrated in figure 1.1.1. The first basis function, shown at the top of the picture is simply a constant, and can somewhat heuristically, be interpreted as the long term level of interest rates. The second basis function starts off at one when $\tau = 0$ and decays rapidly as τ increases, it can be interpreted as the curve's steepness. The third basis is shown at the botton of the graph, it starts off at zero and then rises to a maximum value as τ is increased, it then falls back to zero as τ continues to increase, it can be interpreted as a hump relating to the curvature of the curve. The similarity of these basis functions to the empirical factor loading found by Bliss (1997) and Litterman and Scheinkman (1991) was noted in Diebold and Li (2006). The additional basis function introduced by Svensson (1994), has the same mathematical structure as the third basis function but a

different lambda value and represents an additional hump located at a different point on the curve. This allows an even richer array of term structures to be modelled.

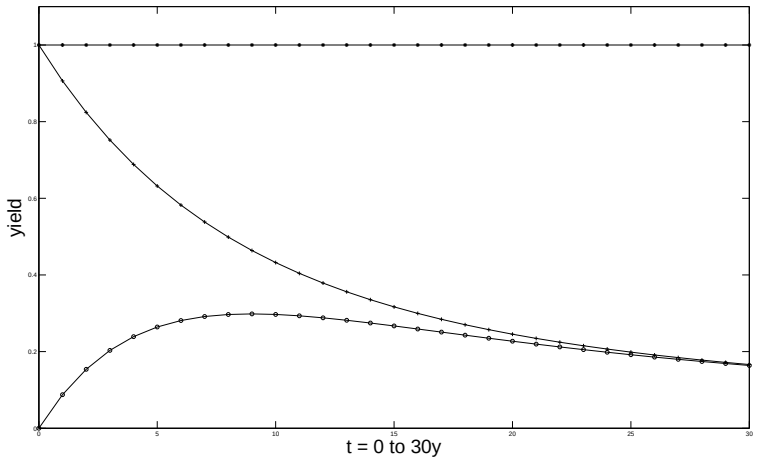


Figure 1.1.1: The Nelson-Siegel Basis Functions

1.2 Relative Value Fixed Income Strategies

Duarte et al. (2007) produced the seminal paper on relative value fixed income arbitrage. The paper looked at several commonly used fixed income relative value trading strategies using data from the US fixed income market from Dec 1988-Dec 2004. They wanted to test whether such strategies were true arbitrage, or are simply strategies that have highly negatively skewed PnL distributions. Meaning that they make small profits most of the time but occasionally experience catastrophic losses. Indeed the title of the article includes the phrase "Nickels in front of a Steamroller?".

As shown in Duarte et al. (2007) one class of fixed income relative value strategy with such a negatively skewed PnL distribution, at least for the data they

tested, is the swap spread. A swap spread is where one fixed rate is traded against another. As explained in Chincarini (2012), one high profile example of such a trade came from LTCM: they had several versions of these trades, across different yields and even markets, and initially they had quite some success with these types of trade. However, in August 1998 they had established a large short position in the 10y US swap spread, meaning they were receiving the 10y swap rate and paying the 10y treasury rate. The average of this spread between January 1991 and July 1998 was 43 basis points (bp) with a standard deviation of 8 bp. In January 1998 the spread had widened to 55 bp and so was approximately 2 standard deviations above the mean. LTCM established a short position expecting the spread to revert. However, by August 4, 1998 the spread had widened to 65.8 bp. There had only been 5 days over the January 1991 to July 1998 period in which the spread has been wider than this, including the all time high of 71 bp, and so LTCM most likely expected it to revert. However, in the wake of the Russian default, the spread widened to reach 97 bp on September 18 1998, causing LTCM huge losses. LTCM had several such swap spreads and combined they were LTCM's single biggest source of losses, 35% according to Lowenstein (2000).

Happily Duarte et al. (2007) also found that the yield curve type arbitrage considered in this thesis produced significant profits using the same data, even after accounting for transaction costs and even hedge fund fees. They note that there are several different versions of the these strategies but they share the common theme of identifying rich and cheap points on the curve. The profit is then realized by putting on a position that exploits those mispricings, this is then held until the mispricing reverses. They only focus on curvature trades: The example they give is of the 2 year swap rate being high relative to the model, where the model is based on the values of the 1y and 10y rates which are referred to as the key points. So they receive the 2 year swap rate whilst simultaneously taking offsetting long positions in the 1 year and 10 year swap rates, so leaving the overall position market neutral. Another archetypal example of a yield curve

trade, is the steepener trade where two maturities are traded against each other to capture changes in the steepness of the yield curve.

To fit the zero coupon curve they used a two dimensional Vasicek model which was set to match the key points exactly whilst matching the remaining points as well as possible. It should be noted that they ran their tests on the same data that they used for calibration, but the authors claim that lookahead bias should not arise in their model formulation. However, it doesn't seem that their arguments are completely convincing. It's at least conceivable that by calibrating their curve based on the spread between the curve and all the non key points for the whole data set, would mean that the resultant curve will encode some information about future relative value opportunities. They did claim in a footnote that they repeated the analysis with out of sample data and the results were very similar to those in the report (they did not however report those results). They use a trigger level of 10 bp to enter the trades (no mention is made of how this figure was derived so there may be scope for look ahead bias there too).

1.2.1 The Relative Value Nelson Siegel Svensson Model

The initial focus of this thesis was to build a model analogous to the one in Duarte et al. (2007) but instead basing it on the NSS model. The reason being is that the Vasicek model belongs to the affine arbitrage free class of models, and as shown by Duffee (2002) this class of models has a forecasting ability no better than a random walk. Whereas Diebold et al. (2006) and De Pooter (2007) showed that the NS model and models based on that approach tend to provide better forecasts. In addition the second and third NS basis functions mimic the steepener and curvature opportunities that the model is seeking to exploit. It was hoped that by basing the relative value model on these basis functions, it would be easier to identify and model those opportunities, as opposed to the Vasicek model which lacks such intuition. It was decided to try and model the

yield curve out to 30y to allow a richer array of relative value opportunities to be exploited. Duarte et al. (2007) used two key points to fit a 10y long section of the curve, as the goal here was to fit the whole yield curve out to 30y, it was decided to use four key points rather than two.

The original relative value Nelson Siegel Svensson model (rv-NSS) is based on the four suitably chosen key points that were chosen so that they are well spaced across the yield curve and correspond to the most liquid points on the yield curve. The Svensson model is used because it has four free parameters (the betas associated with each basis function) which can be used to match the four key points. The remaining points on the curve are then priced relative to these key points with lambdas in the NSS basis functions suitably calibrated to capture the historic relationship between the remaining rates and the key points.

The main assumptions underlying the model are:

- The key points 2, 5, 10, 30y represent the most liquid points on the curve
- The curve can be accurately modelled from the Nelson-Siegel basis functions
- These functions are calibrated to capture the historic relationships between different points on the curve
- These historic relationships are assumed to be stable meaning that they should not be updated unless there is some kind of regime change in the market

Analogously to the Duarte et al. (2007) paper, the central idea is that based on the current position of the key rates the model shows us where a non-key rate such as the 7y point should be priced based on its long term historical relationships to the key points. One can then observe the difference between where the model thinks the price should be based on their historical relationships and

where the 7y is currently trading in the market. This difference or spread represents how rich or cheap the 7y looks on an historical basis. By building up a time series of these spreads one can model how they evolve over time and use this as a basis for proprietary trading decisions, or alternatively market making decisions where one would look to skew their prices downwards when quoting for a price that appears rich and conversely upwards when a price appears cheap.

The third and fourth bullet points spell out the key assumptions that the model can capture the historical relationships between the key points and the rest of the curve accurately and that those relationships don't change. However regime changes, where there is a discrete and persistent change in interest rate dynamics can occur. Two prime examples of regime changes in US rates, which have been identified in the literature, are the Fed monetary policy experiment between October 1979 and October 1982 and the oil shock in the seventies. These regime changes were studied by Hamilton (1988), Gray (1996), and Garcia and Perron (1996) who found that using regime switching models that allow for discrete and persistent changes in the model parameters were best able to model data with these types of changes. If such a regime change occurred it would obviously no longer be reasonable to expect the spread to revert in the same time frame as before or even to the same level as before. However, such regime changes have been incorporated into the Dynamic Nelson and Siegel model which is introduced in section 1.2.3.

By assuming the key points are matched exactly it is also implicitly assumed that the key points are able to move independently of each other, i.e. a 1 bp movement in the 2y rate will not affect the values of the 5y, 10y and 30y rates which must all match their market observed values by design. However all the other non key rates will be affected by the movement, especially those closest to the 2y point.

The main objectives of the model are:

- A time series of the spreads can be used to make proprietary trading decisions
- Current rich/cheapness of a given point can be used to make market making decisions

1.2.2 The Calibration Problems

Models based on the Nelson and Siegel basis functions have some well documented calibration problems. Earlier papers did not identify the root causes of the problems instead reporting their symptoms which are broadly termed as "numerical instability".

Fabozzi et al. (2005) noted that the estimates for β_1, the long term level of rates, and β_2, slope in the long term level of rates, in the three factor NS model have unit roots, a finding echoed by Diebold and Li (2006).

Bolder and Streliksi (1999) noted that the model is non-linear in the λ values, they also identified that it has multiple local minima, as illustrated in figure 1.2.1, which means that multiple starting values are necessary for the optimization process. Indeed they feel that virtually all possible starting values would have to be tested to gain certainty over the results which obviously would be an onerous task computationally.

Similar to Bolder and Streliksi (1999), Cairns and Pritchard (2001) show that the Nelson and Siegel (1987) model has a calibration surface that exhibits multiple local minima, they note that this can cause parameter instability as the model jumps between minima on different calibration days.

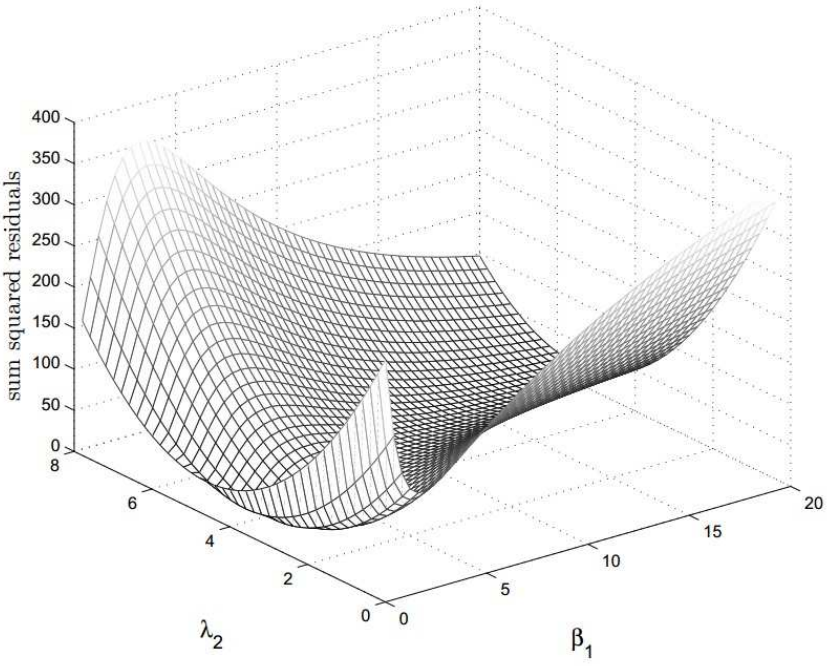


Figure 1.2.1: The Optimization Surface

Gurkaynak et al. (2006) report that the estimated parameters can be unstable often jumping from one day to the next despite there being little actual movement in the yield curve being estimated. The authors note that this instability was also observed by Anderson and Sleath (1999) who showed that even a small shift in the input data can produce notable changes in the estimated parameters and fitted yields. Building on the work of Waggoner (1996), Anderson and Sleath (1999) linked this to the condition number of the Svensson model. The condition number quantifies the sensitivity of a model's outputs, to a small change to the model's inputs. In linear regression, overly high sensitivity to model inputs is often caused by multicolliearity i.e. when there is a near linear relationship amongst two or more of the predictor variables. The condition number can be used to quantify the multicollinearity effect inherent in the model, Anderson and Sleath (1999) did not seek to address the multicollinearity problem nor did they even explicitly link the high condition number to multicollinearity however this was an important step in identifying the source of the numerical issues.

Diebold and Li (2006) did note that multicollinearity can be a problem when using the original Nelson and Siegel (1987) model. They addressed the estimation problems by fixing the value of λ a priori, before calibrating the remaining parameters, of course this solution is only as good as the assumption upon which it is based. Indeed Hurn et al. (2005) show that the assumed value for λ in Diebold and Li (2006) is not suitable for the UK gilts market.

De Pooter (2007) also noted the optimization problems, and also links these problems to multicollinearity. He notes that multicollinearity can arise in the NSS from the λ parameters being too similar (the problem was also noted by Hurn et al. (2005)). De Pooter (2007) addressed the problem by adding a constraint that the second lambda value must be such that its associated hump is located at a maturity at least 12 months shorter that the hump associated with the first λ. This is a somewhat ad-hoc solution and its performance relies on the assumption that the two humps can never be closer than 12 months apart

is always valid. De Pooter (2007) also shows that multicollinearity arises in NS/NSS type models as λ tends towards zero and as it tends towards infinity. This leads to non identification issues which can result in extreme parameter estimates.

Attempts to address these issues in the literature are: Gilli et al. (2010) who were the first to accurately identify the calibration problems as being attributable to two key factors: the non-convex optimization surface exhibiting multiple local minima, and the intrinsic multicollinearity problems. High correlation can be a symptom of multicolliearity and figure 1.2.2 illustrates how the level of correlation changes between the loading factors as the value of λ changes, the left graph illustrates this for β_1 and β_2, the middle for β_2 and β_3 and last one for β_3 and β_4. They went on to address the first problem by applying Differential Evolution, a population based optimization technique similar to genetic algorithms. Anneart et al. (2012) sought to address the multicollinearity problem (but not the optimization surface issues) by applying basic ridge regression to the three factor Nelson and Siegel (1987) model.

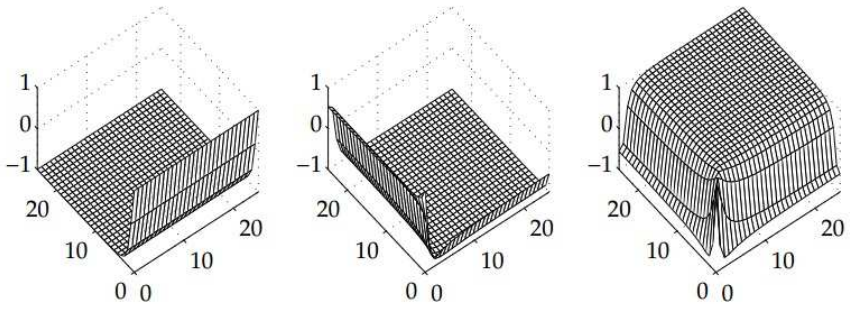


Figure 1.2.2: NSS Basis Function Correlations for Different Candidate λs

1.2.3 The Dynamic Nelson Siegel Model

There has been a spate of recent additions to the literature relating to Nelson Siegel style models which was prompted by the development of the Dynamic Nelson Siegel (DNS) model, introduced by Diebold and Li (2006). They employed a two step estimation procedure in which the λ value shown in equation (1.1.1) is fixed (note that there is only one λ in the paper as they were modelling the NS rather than the NSS model). The DNS extends the NS into a dynamic framework where the loading factors on the NS basis functions are given first order auto regressive dynamics AR(1). They also introduced a more general Vector Auto Regressive (VAR) version but this was not found to add significant value. Diebold et al. (2006) recast the DNS into a state space framework. To estimate the state space model parameters a Kalman filter was used to estimate the log-likelihood of the state space model, this log-likelihood was then maximized using a numerical optimization routine. This procedure had two key benefits, first it allowed all the parameters to be estimated in one step and secondly it allowed the λ to be estimated from the data. Although the optimal λ value was still assumed to be constant over the entire data set.

This assumption of a constant λ value has been criticized in the literature by Koopman et al. (2010) who feel it may be too restrictive, especially when considering long periods of time. Indeed they got statistically different estimates for the value of λ when testing over different time periods. They address this by estimating λ as an additional time varying parameter in the state space framework. To estimate changes in the λ value they initially used a step function to model the changes, but then went on to use a spline based method to allow smoother changes in the λ value, with the spline based method producing the best results. Similarly the problems of regime switching mentioned in section 1.2.1 could also cause the value of λ to change, Levant and Ma (2013) address this shortcoming by generalizing the DNS model into a Markov switching framework. They looked at a model that had two regimes with different λ parameters

in each regime (in addition they also looked at DNS based Markov switching models where the volatility switched in each regime). Xiang and Zhu (2013) also experimented with a regime switching DNS model with differing volatilities in each regime. They tested with up to three regimes, but found that two regimes worked best on the data they tested. Both authors reported improved forecasting abilities when this regime changing machinery is included in the DNS.

The DNS was also augmented by Christensen et al. (2011) who produced an arbitrage free version. Their work was later extended by Hevia et al. (2014) and Bandara and Munclinger (2012) who produced an arbitrage free version of the DNS model with Markov switching. The Markov switching versions again produced improved forecasts.

1.3 Plan of Thesis

Chapters two and three focus on the calibration problems relating to the rv-NSS. As described previously the rv-NSS shares the same calibration problems as the NSS model upon which it is based.

Chapter 2:

Gilli et al. (2010) showed how the non-convexity and multiple local minima problems could be addressed in the context of the NSS model by using differential evolution (DE). However the calibration procedure used in the rv-NSS is quite different to that of the NSS so the DE based optimization scheme had to be adapted. Chapter two shows how it was possible to adapt the DE based optimization procedure to fit the rv-NSS and as such represents a new contribution to the literature.

Chapter 3:

The only remaining calibration problem for the rv-NSS was multicollinearity. Ridge regression is known to produce more reliable parameter estimates when harmful levels of multicollinearity are present and Anneart et al. (2012) had already sought to address the multicollinearity problem in the context of the regular NS model by using ridge regression. This prompted the idea to apply ridge regression to the rv-NSS to address its multicollinearity issues. There has been a plethora of ridge regression estimators in the literature but it's not straightforward to identify which is best as their performance varies in response to the level of multicollinearity and error variance. There have been attempts to identify the best ridge estimators in the past using simulation based comparison studies. The largest of which was conducted by Clark and Troskie (2007). However since then several new estimators have been proposed. So it was decided to conduct a new simulation study, to extend the work of Clark and Troskie (2007) by including the new estimators (and some older ones). The simulation study in chapter three encompasses the widest range of ridge regression estimators ever assembled in one study, as such it represents a new contribution to the literature. The Modified (r-k) Class Ridge Regression (MCRR) introduced by Batah et al. (2009) did not specify how to estimate some of its parameter values, chapter three introduced two ways to address this issue and this represents a further contribution to the literature. In addition the Improved Ridge Estimators (IRE) of Liu et al. (2013) experienced convergence problems and chapter three tries to address these succeeding in the majority of conditions tested. The Linearized Ridge Regression Estimator (LRRE) of Gao and Liu (2011) also had estimation problems at the lowest volatility levels and an attempt was made to address this too. Although the solution did fix the LRRE's convergence issues at the lowest volatility levels, this was at the expense of its performance at every other volatility level. These amendments to the estimation procedures for the IRE type estimators and the LRRE represent further contributions to the literature. Although no estimator was found to be best in every scenario tested,

the LRRE was found to be best or joint best in the majority of the scenarios tested, and this is chapter three's final contribution to the literature.

Chapter Four:

The results of the simulation study in chapter three along with the results of chapter two meant that the two key calibration issues had now been addressed, and the rv-NSS could now be implemented effectively. However, at this stage there was a change of direction, I'd started researching Dynamic Nelson and Siegel models (DNS) which were first introduced by Diebold and Li (2006), this led me to think that a relative value model based on the DNS would have several advantages, over rv-NSS.

- The DNS incorporates the time series dynamics of the loading parameters internally
- The Kalman Filter used in estimation of the DNS, can help to alleviate the multicollinearity problems
- The DNS has regime switching variants and so can handle regime changes
- The DNS can be made arbitrage free

As detailed in chapter 4 some changes to the rv-NSS logic were necessary, but Chapter 4 details how the DNS can be developed into a fixed income relative value trading model. This represents a new contribution to the literature. The rv-DNS is also successfully calibrated and tested on both simulated and historical data, this also represents a new contribution to the literature.

Chapter 5:

Chapter 5 summarizes the results of chapters 2, 3 and 4 and draws conclusions based on those results. It also goes on to discuss potential avenues for further research/development.

Chapter 2

Calibration of the Relative Value Nelson Seigel and Svensson Model

2.1 Abstract

This chapter introduces the relative value Nelson Siegel and Svensson model (rv-NSS), like the standard Nelson-Siegel-Svensson (NSS) models upon which it is based it suffers from problems in its calibration when standard derivative based optimsiation techniques are employed. This is due to the fact that the optimisation problem is non-convex and features multiple local minima. This problem has already been addressed in Gilli et al. (2010) in the context of the NSS model. They used a population based optimisation method called Differential Evolution to tackle the problem and found that it produced superior results to the derivative based optimisation techniques with which it was compared. This chapter adapts their algorithm so that it can be used to calibrate the rv-NSS. It then tests this adapted model's performance against a more traditional derivative based method. The adapted model is found to outperform the derivative based optimiser.

2.2 Introduction

As explained in section 1.2.2 the non-convex nature of the optimization problem for models based on the NSS basis functions coupled with their non-linear nature, mean that traditional derivative based optimization methods will tend to struggle as they are inclined to get stuck in local minima. See figure 1.2.1 taken from Gilli et al. (2010) which illustrates the challenges posed when trying to calibrate NSS type models. This problem has already been addressed for the pure NSS model in the paper by Gilli et al. (2010). In that paper they found that Differential Evolution, a type of population based optimization method, was much better suited to the optimization of the NSS model than the derivative based methods they compared it to.

Chapter 2 looks to address these optimization problems in the context of the rv-NSS model. The idea is to adapt the Differential Evolution algorithm so that it could be applied to the rv-NSS model. Although at first blush the NSS and rv-NSS appear quite similar mathematically, the logic behind the models and crucially the calibration process is quite different. In the regular NSS model all parameters are used to try and fit all the yields equally well. In the rv-NSS model the betas are chosen to fit the key-rates exactly while the other yields are fitted using the lambda values as well as possible on a least squares basis. This meant that the differential evolution technique described in Gilli et al. (2010) could not be applied as given. Instead the algorithm was altered to fit with the differing calibration procedure for the rv-NSS model. As the multicollinearity was still the focus of ongoing research at that stage, no method had been found to calibrate the historical parameters. Instead the model will be calibrated to simulated data, where the simulated data is generated using assumed values for the historical parameters.

2.3 Literature Review

The modelling of the zero coupon yield curve is fundamental to fixed income. The zero coupon curve can't be observed directly but rather it must be inferred from the prices of fixed income assets traded in the market. The model proposed by Nelson and Siegel (1987) and its extension by Svensson (1994) is a well known method for estimating the zero coupon rates which is popular with policy makers with nine out of 13 central banks reportedly using the Nelson and Siegel and/or the Svensson variant BIS (2005).

The original Nelson-Siegel model consists of three parametric functions, which when combined are sufficiently flexible to allow most commonly encountered yield curve shapes to be modelled. The extension by Svensson (1994) simply adds another factor thus allowing an even richer array of term structures to be modelled. This chapter will focus on the Svensson formulation of the model.

The Nelson-Seigel-Svensson model will be used as the basis for a relative value model. The idea is that the assets at the most liquid points on the curve will be priced exactly, while the remaining assets are priced relative to these points.

As was highlighted by Gilli et al. (2010) the calibration of Nelson-Siegel-Svensson type models has two key problems:

- The optimization problem is non-convex and has multiple local minima
- For certain portions of the parameter space the model can be badly conditioned: causing the model to be overly sensitive to small perturbations of the input data, and parameter estimates with grossly inflated variances

Only the first problem will be addressed in this chapter, the second problem will be dealt with based on the results in chapter three.

2.4 The Model

2.4.1 The Nelson-Siegel-Svensson Model

The Nelson-Siegel-Svensson model can be defined mathematically as follows

$$\mathrm{y}(\tau) = \beta_1 + \beta_2 \left(\frac{1 - exp\left(-\frac{\tau}{\lambda_1}\right)}{\left(\frac{\tau}{\lambda_1}\right)} \right) + \beta_3 \left(\frac{1 - exp\left(-\frac{\tau}{\lambda_1}\right)}{\left(\frac{\tau}{\lambda_1}\right)} - exp\left(\frac{\tau}{\lambda_1}\right) \right) + \beta_4 \left(\frac{1 - exp\left(-\frac{\tau}{\lambda_2}\right)}{\left(\frac{\tau}{\lambda_2}\right)} - exp\left(\frac{\tau}{\lambda_2}\right) \right) \tag{2.4.1}$$

where:

$\tau =$ maturity of the instrument being modelled

$y(\tau) =$ zero coupon rate at the maturity indicated by τ

$\beta_1 =$ loading factor on the first basis function

$\beta_2 =$ loading factor on the second basis function

$\beta_3 =$ loading factor on the third basis function

$\beta_4 =$ loading on the fourth basis function

$\lambda_1 =$ Determines the shape of the second and third basis functions

$\lambda_2 =$ Determines the shape of the fourth basis function

The first factor is simply a constant, and can somewhat heuristically, be interpreted as the long term level of interest rates, the second can be interpreted as the curve's "steepness" (the difference between short and long term rates) while the third can be interpreted as a "hump"(such humps are often observed as the difference between an intermediate rate and both a longer term rate and a shorter term rate, e.g. the 20y rate relative to the 10y and 30y rates). The similarity of these factors to the empirical factor loading found by Bliss (1997) and Litterman and Scheinkman (1991) was noted in Diebold and Li (2006).

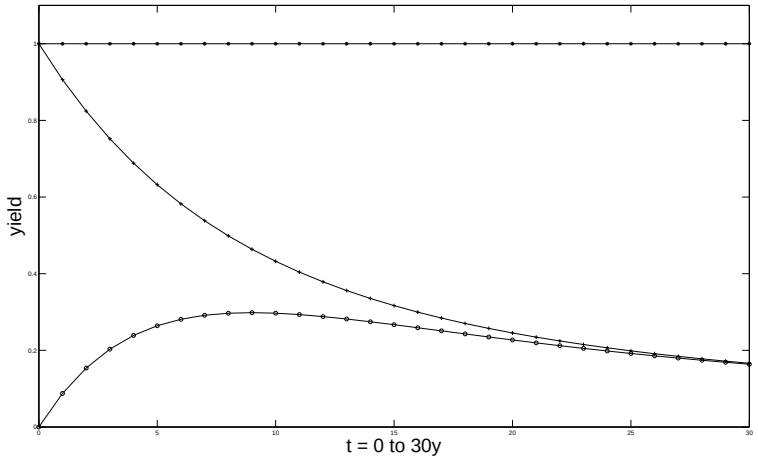


Figure 2.4.1: The Nelson-Siegel Basis Functions

A graph of the various basis function values versus the time maturity value is shown in figure 2.4.1. Further intuition on the role of the different factors can be gained when we consider the behaviours of their parameteric functions in the limit. The constant 1 is independent of time and so as noted before can be used to represent the long term level of rates. The second factor starts off at one when $\tau = 0$ and decays rapidly as τ increases, so it represents a short-term factor. The third factor starts off at zero and then rises to a maximum value as τ is increased, it then falls back to zero as τ continues to increase. The value of λ_1 in the third factor determines the location of its maximum value. Similarly

the fourth factor which has the same mathematical structure as the third factor but a different λ value (λ_2) allowing a second hump to be added to the curve.

2.4.2 The Relative Value Nelson Siegel Svensson Model

The relative value Nelson Seigel Svensson model (rv-NSS) is based on the four suitably chosen "key points" that should be chosen so that they are well spaced across the yield curve and correspond to the most liquid points on the yield curve. The remaining points on the curve are then priced relative to these key points with the NSS basis functions suitably calibrated to capture the historic relationship between the remaining rates and the key points.

The main assumptions underlying the model are:

- The key points: the 2y, 5y, 10y, 30y represent the most liquid points on the curve
- The curve can be accurately modelled from the Nelson-Siegel basis functions
- These functions are calibrated to capture the historic relationships between different points on the curve
- These historic relationships are assumed to be stable, meaning that they should not be updated unless there is some kind of regime change in the market, which means that the relationship between the key points and the other rates had permanently changed

The central idea is that based on the current position of the key rates, the model shows us where a non-key rate such as the 7y point should be priced based on its long term historial relationships to the key points. One can then observe the

difference between where the model thinks the price should be based on their historical relationships and where the 7y is currently trading in the market. This difference or "spread" between the model implied rate and the current rate represents how rich or cheap the 7y looks relative to the historical date the basis functions were calibrated to. By building up a time series of these spreads one can model how they evolve over time and use this as a basis for proprietary trading decisions, or alternately market making decisions where one would look to skew their prices downwards when quoting for a price that appears rich and conversely upwards when a price appears cheap.

Assumptions three and four spell out the key assumption that the model can capture the historical relationships between the key points and the rest of the curve accurately and that those relationships don't change. However as noted in Xiang and Zhu (2013) regime changes, where there is a discrete and persistent change in interest rate dynamics can occur. Notable examples in the literature of regime changes in US rates include, the Fed monetary policy experiment between October 1979 and October 1982 and the oil shock in the seventies. These regime changes were studied by Gray (1996), Hamilton (1988) and Garcia and Perron (1996) who found they were better able to model interest rates over these periods using regime changing models. If a regime change does occur and assumption four is violated, then it is no longer reasonable to expect rates to revert in the same time frames or even to the same level as before and so some form of recalibration would be required.

The main objectives of the model are:

- The time series of the spreads can be used to make proprietary trading decisions
- The current rich/cheapness of a given point can be used to make market making decisions on any products priced using the model

2.5 Model Calibration

2.5.1 Differential Evolution

As was reported in Gilli et al. (2010) several authors (Bolder and Streliksi (1999); Gurkaynak et al. (2006) and De Pooter (2007)) have reported "numerical problems" when calibrating NSS type models.

Gilli et al. (2010) identified the root causes of these numerical instabilities to be:

- The optimization problem is non-convex and has multiple local minima
- Large sections of the parameter space are ill-conditioned giving rise to unstable parameter estimates

Only the first problem will be addressed in this chapter. As a consequence of the non-convexity and multiple local minima derivative based optimisation methods such as trust-region-reflective algorithm in Matlab are ineffective. To tackle such problems Gilli et al. (2011) explain that so called population based methods are more appropriate.

There are two main reasons why this makes population based methods better suited to these type of problems:

- They allow members of the population to explore in up-hill directions of the search space
- They keep within their populations inferior solutions

This may seem counterintuitive at first, but by allowing the members of the population to explore up-hill they are able to find a way out of a local minima.

Keeping the inferior solutions in the population while this exploration takes place means that they are more likely to have enough time to find their way out. The net result is that population based methods are less likely to get stuck in local minima.

The population based methods which have been applied in the literature to the calibration of NSS type models are a genetic algorithm by Gimeno and Nave (2009) and Differential Evolution which was applied by Gilli et al. (2010) and Gilli et al. (2011). Both methods appear to provide good solutions when applied to NSS type models and arguably either could have been used to calibrate the rv-NSS, however as they are both population based methods there seemed little point in using both methods. In this chapter it was decided to follow the Differential Evolution approach as described in Gilli et al. (2010) and Gilli et al. (2011). The calibration procedure for the rv-NSS, as will be explained in the next section, is quite different to that of the regular NSS, this meant that Differential Evolution could not be applied exactly as given in Gilli et al. (2010) and Gilli et al. (2011) and changes had to be made.

2.5.2 Differential Evolution applied to Relative Value-Nelson Siegel and Svensson Model

The original NSS was intended to fit the whole yield curve as closely as possible in the least squares sense, it does this by varying all 6 input parameters until an optimal fit is found. The rv-NSS is different as it seeks to fit only the key points exactly and then to give the price of the other points on the curve based on the position of the key points, and their historical relationships to the points being priced. The calibration of the rv-NSS basis functions to historical data is subject to the ill-conditioning problems mentioned earlier. As this problem will not be addressed until the results of chapter three are available, the focus here is to produce an algorithm that allows the rv-NSS to be estimated accurately

given assumed values for the NSS-basis functions.

To ensure the key points are fitted exactly it was necessary to change the Differential Evolution algorithm. Analogously to Nelson and Siegel (1987) and Nyholm (2008) instead of all six parameters being used to fit all the points on the curve, as with regular Differential Evolution, the algorithm was changed so that only the λs vary. Conditional on the candidate λ values chosen the βs are then selected to ensure that the 2, 5, 10, 30y points are matched exactly. Given that the problem, conditional on the choice of the candidate λs, is linear and there are as many unknowns as knowns the problem becomes a system of simultaneous equations which can be solved via matrix inversion see equation (2.5.1). However the problem of selecting the optimal λs retains the numerical problems highlighted previously.

$$\begin{bmatrix} 2_y \\ 5_y \\ 10_y \\ 30_y \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-exp^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-exp^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - exp^{-\lambda_1\tau_2} & \frac{1-exp^{-\lambda_2\tau_2}}{\lambda_2\tau_2} - exp^{-\lambda_2\tau_2} \\ 1 & \frac{1-exp^{-\lambda_1\tau_5}}{\lambda_1\tau_5} & \frac{1-exp^{-\lambda_1\tau_5}}{\lambda_1\tau_5} - exp^{-\lambda_1\tau_5} & \frac{1-exp^{-\lambda_2\tau_5}}{\lambda_2\tau_5} - exp^{-\lambda_2\tau_5} \\ 1 & \frac{1-exp^{-\lambda_1\tau_{10}}}{\lambda_1\tau_{10}} & \frac{1-exp^{-\lambda_1\tau_{10}}}{\lambda_1\tau_{10}} - exp^{-\lambda_1\tau_{10}} & \frac{1-exp^{-\lambda_2\tau_{10}}}{\lambda_2\tau_{10}} - exp^{-\lambda_2\tau_{10}} \\ 1 & \frac{1-exp^{-\lambda_1\tau_{30}}}{\lambda_1\tau_{30}} & \frac{1-exp^{-\lambda_1\tau_{30}}}{\lambda_1\tau_{30}} - exp^{-\lambda_1\tau_{30}} & \frac{1-exp^{-\lambda_2\tau_{30}}}{\lambda_2\tau_{30}} - exp^{-\lambda_2\tau_{30}} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \tag{2.5.1}$$

In the full rv-NSS model we would then go on to find the optimal fit of the lambdas to the historical calibration data. However as mentioned previously the ill-conditioning of much of the lambda parameter space means that this is not practical currently and remains the focus of future research. In this chapter the problem of calibrating the rv-NSS model to simulated data is tackled.

2.6 Testing Procedure

To test the effectiveness of the "Altered Differential Evolution" (ADE) algorithm in fitting the rv-NSS model it is proposed that the ADE be used to fit the rv-NSS to simulated data.

Analogously to the testing procedure used in Gilli et al. (2010), to give a benchmark Matlab's lsqcurvefit function was used to fit the simulated data as well. Lsqcurefit is the function Matlab recommends for solving non-linear least squares type problems. The function uses the trust-region-reflective algorithm, which is a derivative based optimisation procedure which can handle constraints on the input parameters (the function can also be set to run using Levenberg-Marquardt algorithm but this option cannot handle such constraints and so was not chosen here).

2.6.1 Generation of Simulated Data

To simulate the data it was necessary to generate values for each of the rv-NSS model parameters. λ_1 was generated from a uniform distribution over the range $0.75 \leq \lambda_1 \leq 2.5$ and λ_2 over the range $2.5 \leq \lambda_2 \leq 5.5$. The permissible bounds on the λs were based on those suggested by Gilli et al. (2010) as those likely to result in acceptable levels of correlation between the NSS basis functions. They actually used $0 \leq \lambda_1 \leq 2.5$ for the bounds on λ_1 but in this study the range of values shown above worked better. They do not necessarily represent the λs one would expect to find when calibrating them to market data. To give the curves realistic shape it was decided to base the βs on the actual 2y, 5y, 10y and 30y rates observed in the euro swaps market between 2002 and 2011. This is also in line with the underlying model logic where all other points are priced relative to the key points according to the relationship dictated by the rv-NSS basis functions.

The method for simulating the data is summarised below

- Generate two λs from uniform distributions over the ranges $0.75 \leq \lambda_1 \leq 2.5$ and $2.5 \leq \lambda_2 \leq 5.5$
- Take actual 2, 5, 10, and 30y rates from a series of dates covering 2002 to 2011 in the euro swaps market, and solve for the β values using the known λ values and rates
- These simulated parameter values are used to generate a yield curve. The remaining maturities are generated using the same parameter values evaluated at their corresponding maturities

A total of 2,571 simulated curves were produced, corresponding to the number of trading days between 2002 and 2011, and each was estimated ten times by both optimisers. For lsqcurvefit this meant that 10 different pairs of initial guesses were used for each curve tested. The only prior assumption used for both optimisers was that the location of the two humps and hence the lambda values are located somewhere along the 30y term structure. Thus the initial guesses for the two lambda values were drawn from anywhere from 0.1-30y for lsqcurvefit (allowing zero to be chosen as an initial guess prevented lsqcurvefit from converging). The ADE's initial population was chosen from the full 30y yield curve. Both algorithms were constrained so that the final results would fall within the 30y term structure. The ADE as a population type algorithm generates an initial set of population parameters from the desired range, in this case the full 30y term structure. So running the ADE 10 times is equivalent to simply increasing the size of the population and/or the number of generations calculated in each optimisation, however for the sake of comparison it was re-run 10 times as well.

The comparison was made on a root mean square error basis (rmse):

$$RMSE = \sqrt{(\frac{1}{n}\Sigma_{i=1}^{n} y_s(i) - y_{est}(i))^2} \tag{2.6.1}$$

where n represents the number of maturities being estimated

y_s is the simulated rate

y_{est} is the estimation of the simulated point produced by the optimisers.

The simulated yield curves for each day are then passed into the ADE and lsqcurvefit and the esimated parameter values returned. As all the curve points are generated using the same parameter values the implicit assumption is at the point of observation they lie on their long term historic mean values. The practical upshot is that a perfect fit for all the rates is possible not just the key rates.

The best, worst and median estimates over the ten runs for each curve were collated.

2.7 Results and Discussion

2.7.1 Results of Calibration to Simulated Data

The calibration results are summarised in table 2.7.1 below:

Model	Median of Median RMSE (Original Priors)	Largest RMSE Value	Median Difference between Max and Min RMSE
ADE	1e-14	0.55	1e-14
Lsqcurvefit	0.37	46	0.66

Table 2.7.1: Summary Results of Calibrations (all results in basis points)

As can be seen from the table 2.7.1 Lsqcurvefit produced a median of medians error of 0.37 basis points (bp) which is good while ADE produced a median of medians RMSE of the order of $1e^{-14}$ bp which is virtually perfect. By design both algorithms fit the key rates exactly so the errors are distributed across the remaining points. As noted previously the data being fitted were originally generated from a rv-NSS model, therefore a perfect fit is possible for all points on the curve. Whether or not the lsqcurvefit median value is material would depend on the application, for a prop trading tool depending on the size of the signal from the proprietary trading model and its size relative to the rmse noise it may be acceptable, however for a market making tool it might not be. The largest individual observed rmse values were 46bp for lsqcurvefit and 0.55bp for ADE. The 46bp rmse value would be material for any conceivable application of the model.

Moreover when one considers the variation around these results by looking at the maximum and minimum rmse values produced over the ten random starts, one finds that the ADEs estimates to be far more stable than their lsqcurvefit

equivalents. Taking the median of the difference of the maximum and minimum rmse values for both methods, one finds that lsqcurvefit had a median spread between its highest and lowest rmse values of 0.66bp while the corresponding value for ADE was of the order $1e^{-14}$ bp, again virtually perfect.

These differences are also illustrated with the graphs shown in fig 2.7.1, which show the results produced when the optimisers were used to estimate simulated data constructed using rates observed in the market in 2002. The first column of the graph shows the results for lsqcurvefit, with the left arm signaling the lowest observed rmse value over the ten runs, the '.' the median rmse value and the right arm the largest observed rmse value. The middle graph shows the corresponding results for ADE. Clearly the ADE results are much more stable across the ten restarts. The graph in the final column shows the differences in the median values between the lsqcurvefit results and the ADE results across the ten restarts. Clearly ADE is more accurate than lsqcurvefit. The corresponding results for the remaining simulated data are shown in the appendices.

Although there are periods where the performance of lsqcurvefit is comparable to that of ADE it can be seen that over the whole test period lsqcurvefit can not be relied upon to produce the correct results and even when it does produce the correct median value the variance associated with that estimate is much higher than the corresponding ADE estimate.

A further set of results was obtained for lsqcurvefit with the initial estimates drawn from a much tighter range of values. This range was chosen to match the range of values from which the actual λ values underlying the simulated data were drawn. Thus it is assumed that the analyst has prior information. As might be expected for a derivative based optimisation procedure applied to a problem with multiple local minima, its performance is very sensitive to the initial estimates used. This proves to be the case with lsqcurvefit: with the tighter priors it is able to match ADE over a much wider set of dates see fig. 2.7.2. Remember however that the ADE results were produced without any

prior information other than we expect both λs to appear somewhere in the 30y term structure. Even with these tighter priors the highest observed rmse error for lsqcurvefit was of the order of 12bp which would almost certainly be material. Also even with these tight priors if the order of the priors is reversed so that the inital guess for λ_1 is produced from the range of values used to choose the true λ_2 and vice versa, lsqcurvefit produces very poor results at least as bad as running it without the prior information. Moreover there would still be the problem of producing reliable prior information in practice. Even if it could be done initially, problems would likely still arise if a regime shift in the market meant that these priors are no longer valid. By contrast ADE was able to produce these results without the need for these tighter priors. Results for several other years of simulated data are given in the appendices and the pattern is the same with ADE continuing to outperform lsqcurvefit.

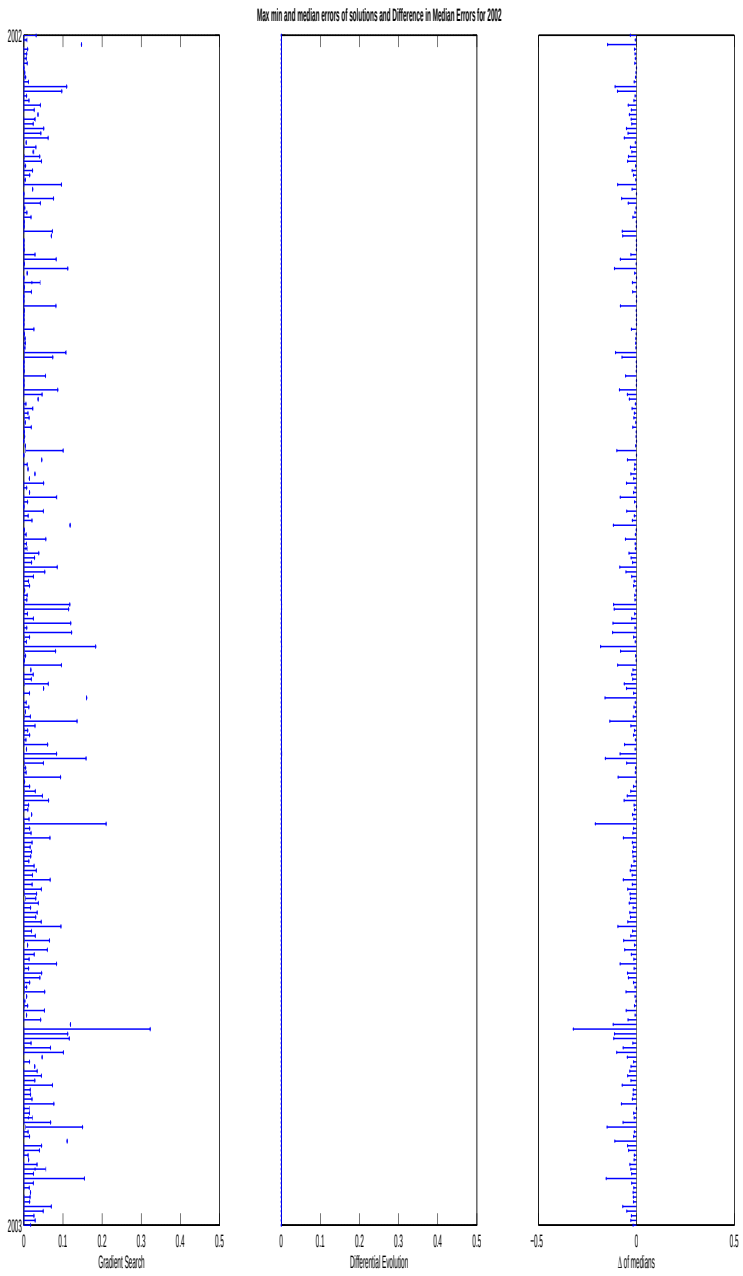


Figure 2.7.1: Max, Min and Median errors for 2002

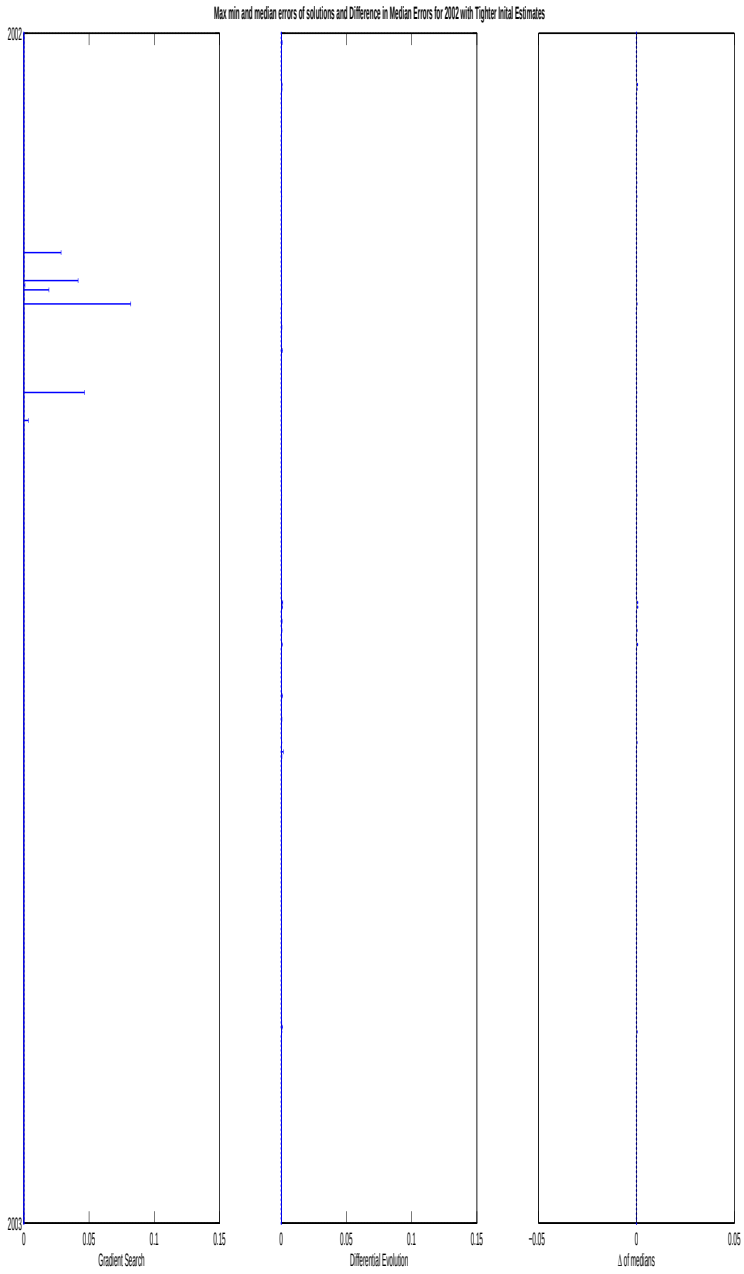


Figure 2.7.2: Max, Min and Median errors for 2002 with Tighter Initial Estimates

2.8 Conclusion

As was shown previously by Gilli et al. (2010) all NSS type models are problematic to calibrate because of their non-convex nature and multiple local minima. This chapter introduced the rv-NSS model which is subject to the same problems. To estimate the rv-NSS it was necessary to alter the DE algorithm that was given in Gilli et al. (2010) to produce the ADE. As can be seen from the results produced the ADE model clearly outperforms the derivative based optimisation methods employed by lsqcurvefit in estimating the rv-NSS, even when tighter priors are used for lsqcurvefit it is liable to produce significant errors. ADE however is able to produce reliable results without the need for any prior information. This now meant that the problems relating to the non-convex multi local minima nature of the rv-NSS optimisation surface had been addressed, however the multicollinearity problem still remained and this will be tackled in chapter three.

2.9 Appendices

2.9.1 Appendix 1: The Results of the Comparison for all the remaining Simulated Data

Figure 2.9.1: Max, Min and Median errors for 2003

Figure 2.9.2: Max, Min and Median errors for 2004

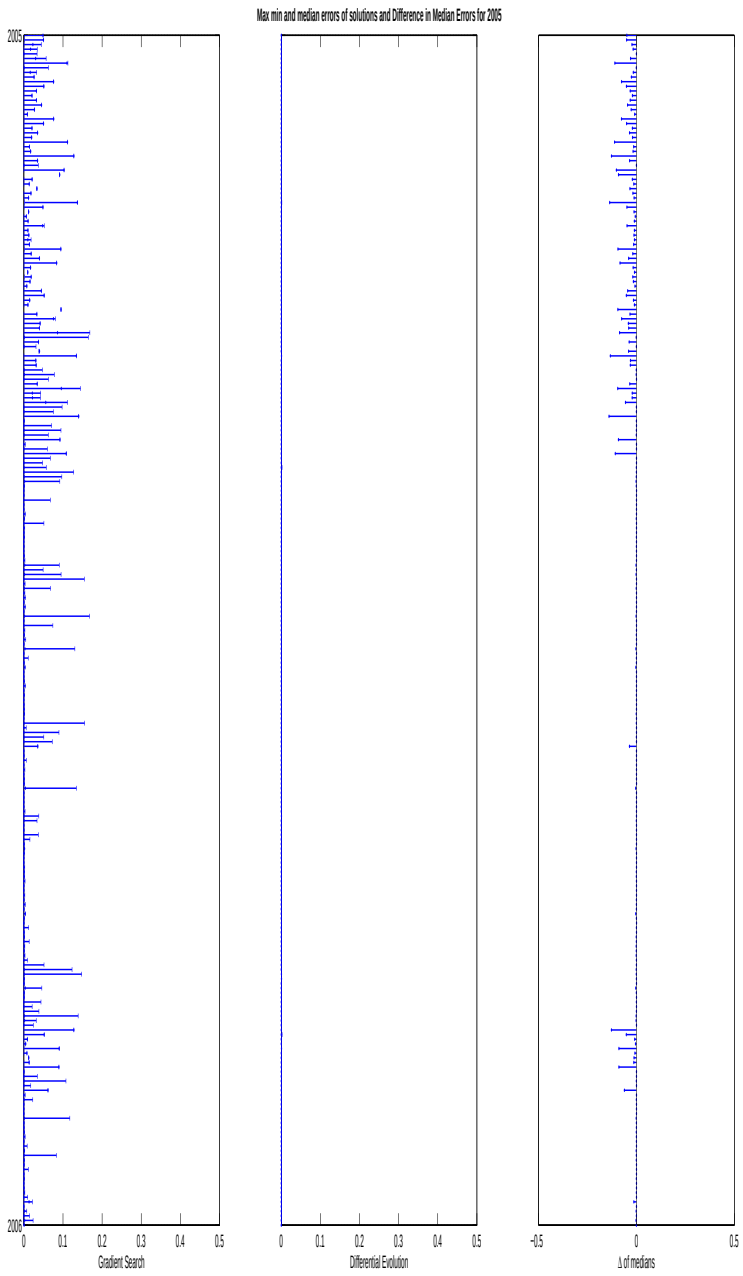


Figure 2.9.3: Max, Min and Median errors for 2005

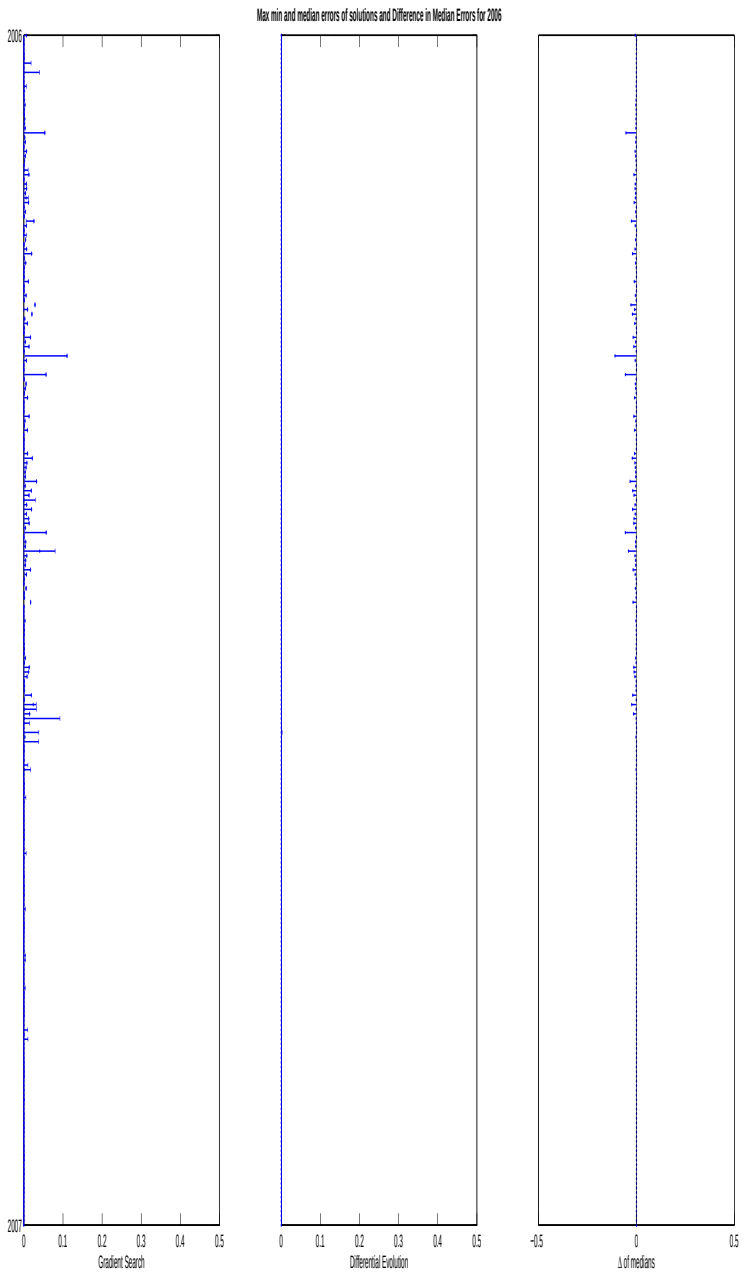


Figure 2.9.4: Max, Min and Median errors for 2006

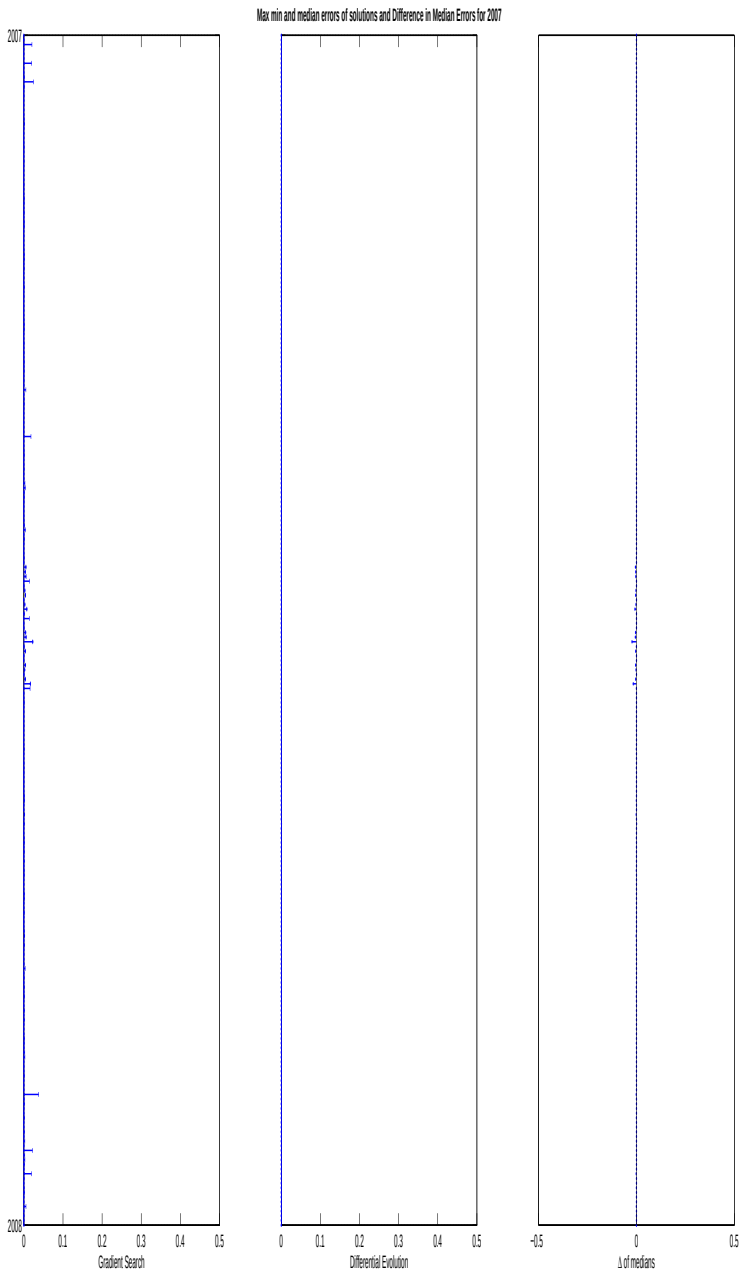


Figure 2.9.5: Max, Min and Median errors for 2007

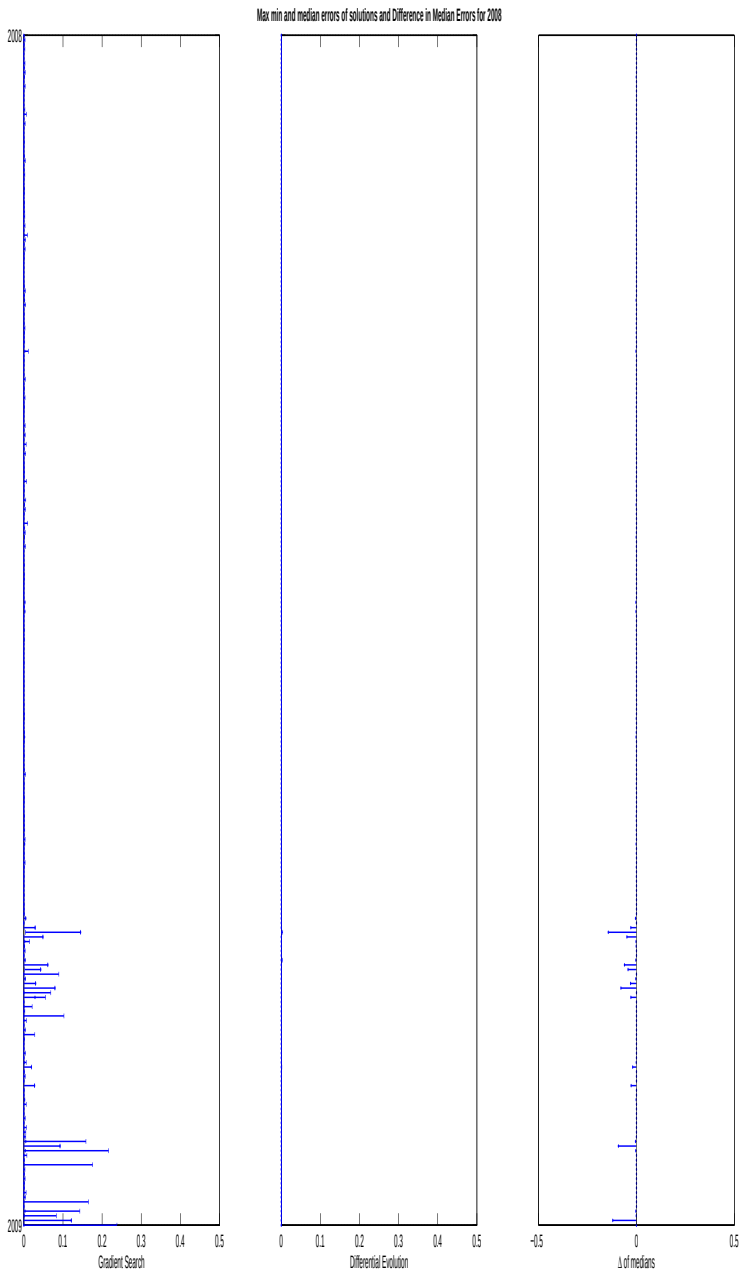


Figure 2.9.6: Max, Min and Median errors for 2008

Figure 2.9.7: Max, Min and Median errors for 2009

Figure 2.9.8: Max, Min and Median errors for 2010

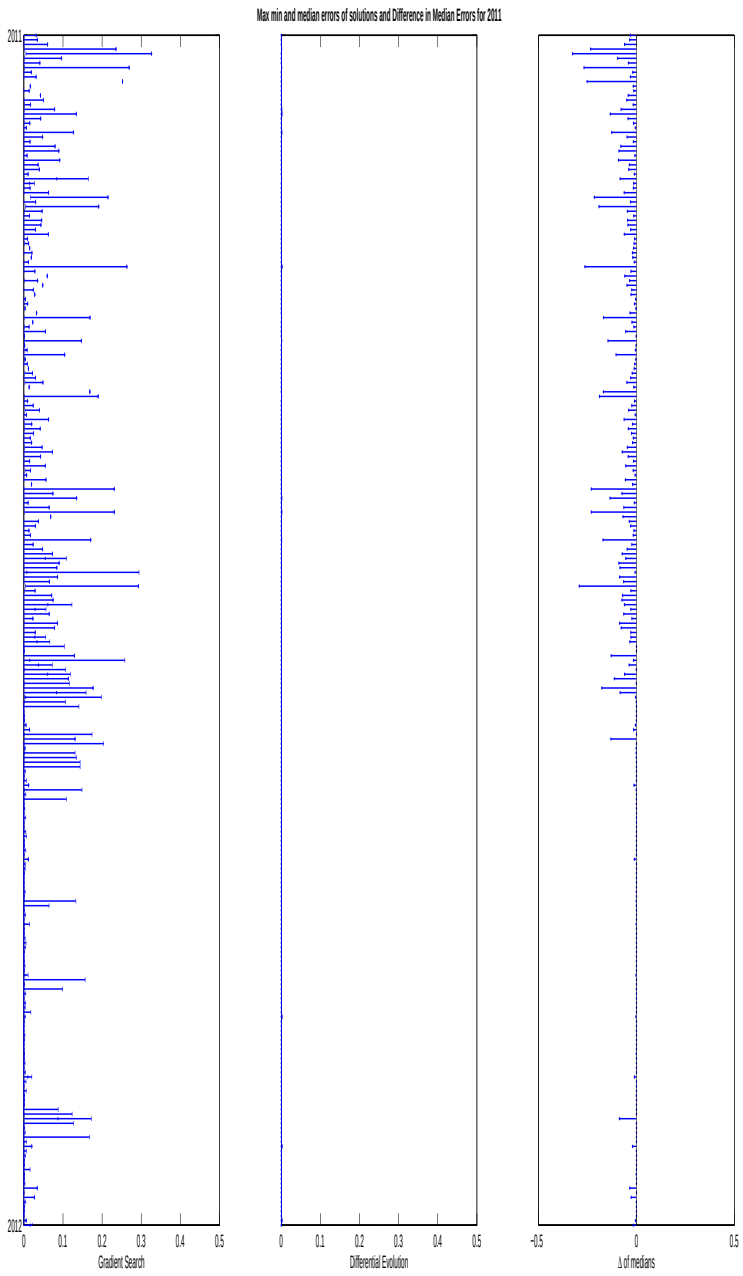


Figure 2.9.9: Max, Min and Median errors for 2011

2.9.2 Appendix 2: The Results of the Comparison with the Tighter Priors for all the remaining Simulated Data

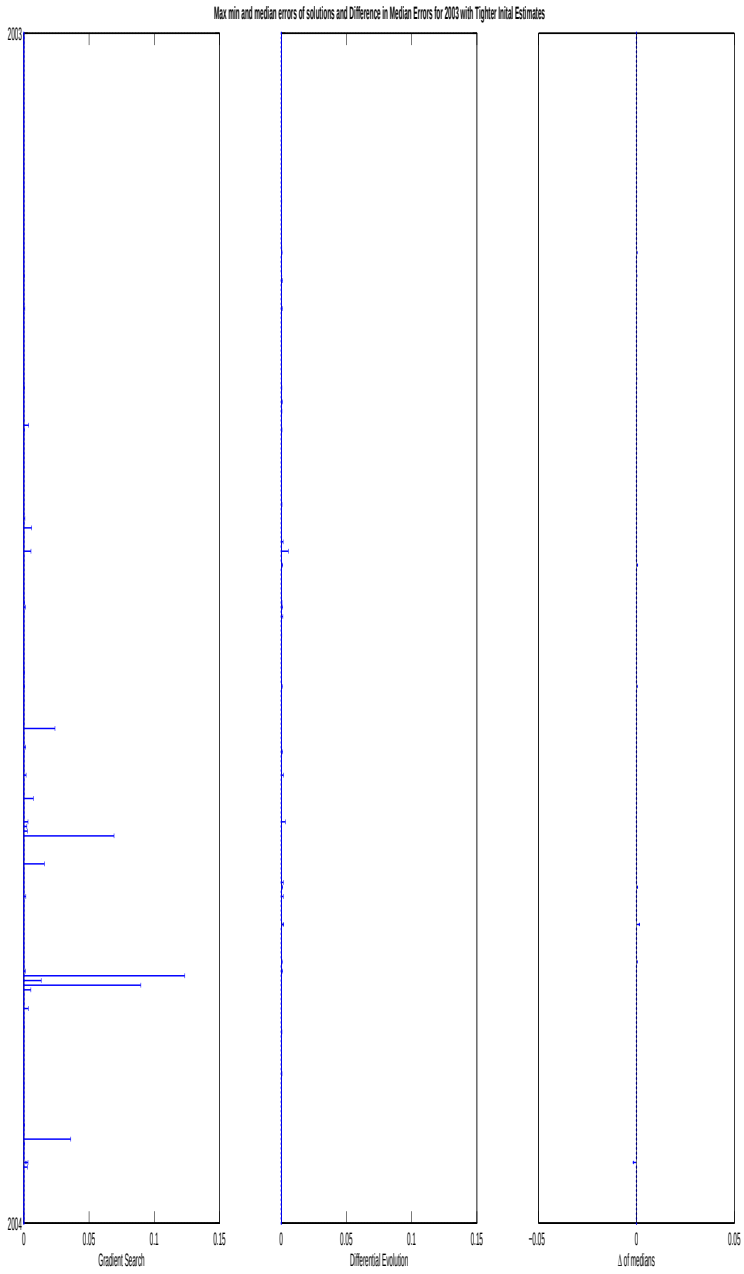


Figure 2.9.10: Max, Min and Median errors for 2003 with Tighter Initial Estimates

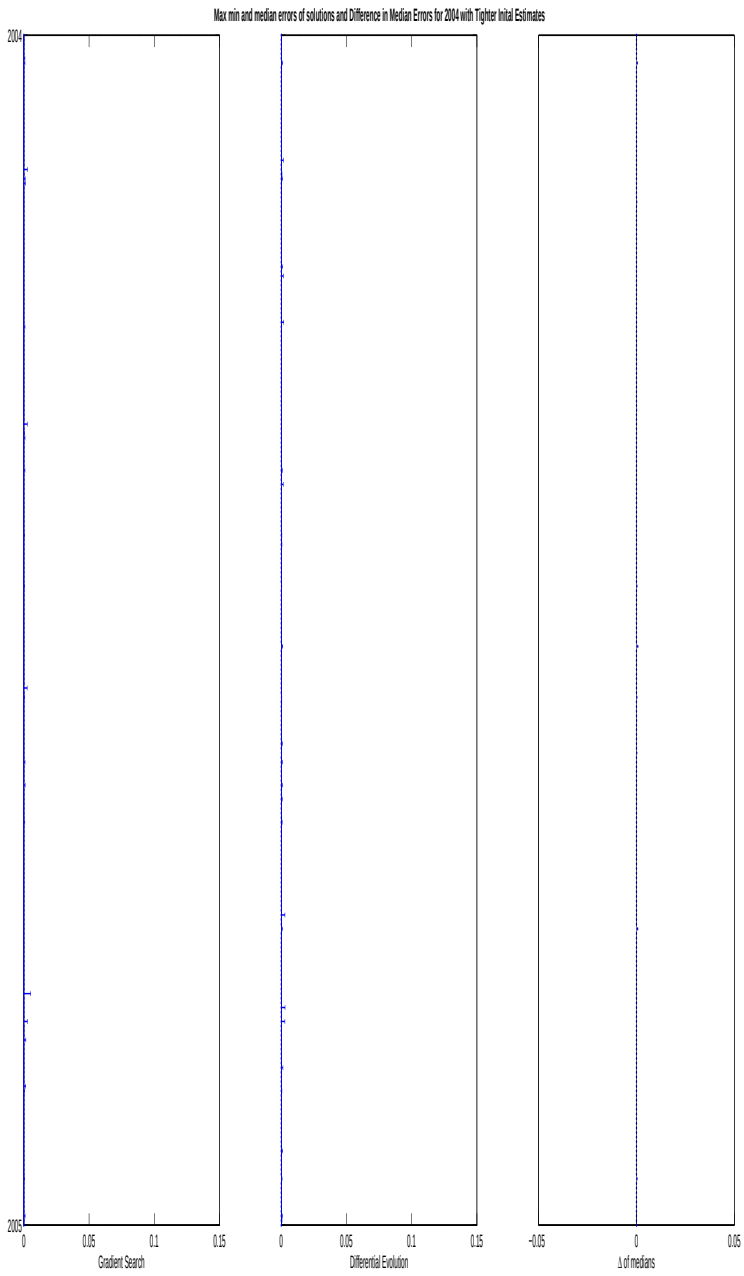


Figure 2.9.11: Max, Min and Median errors for 2004 with Tighter Initial Estimates

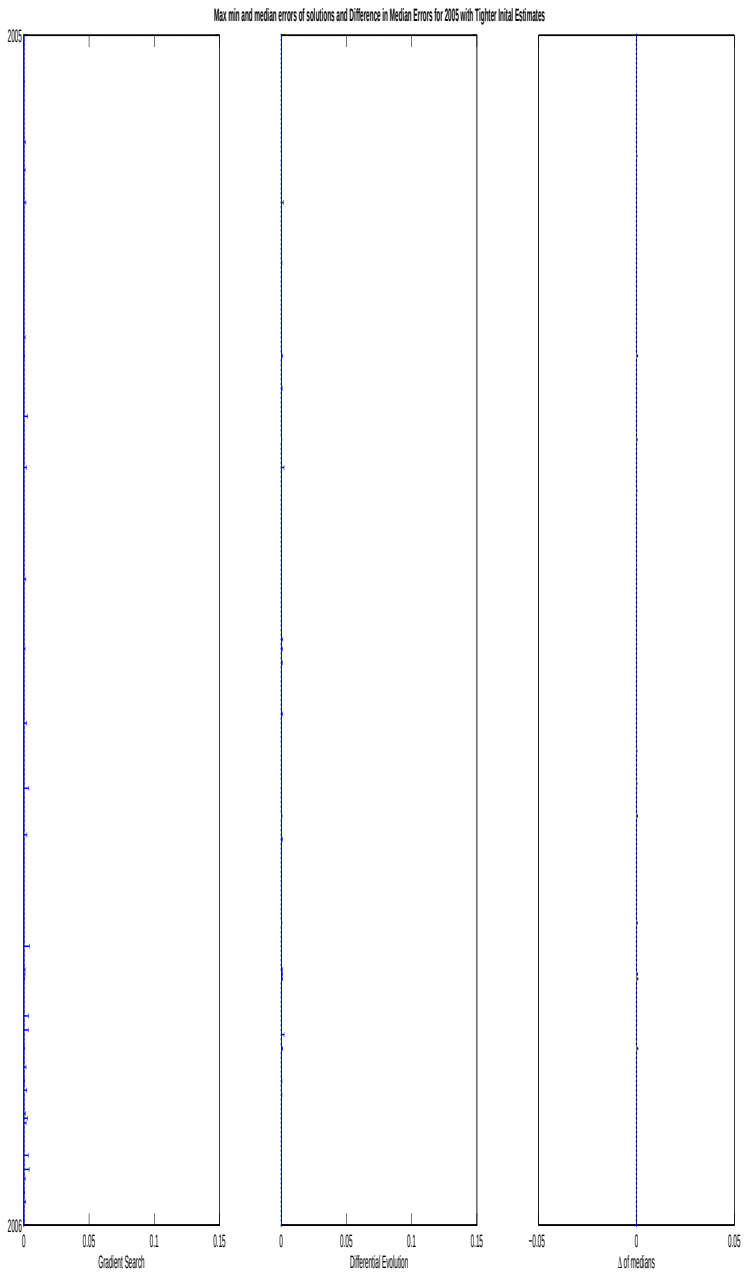


Figure 2.9.12: Max, Min and Median errors for 2005 with Tighter Initial Estimates

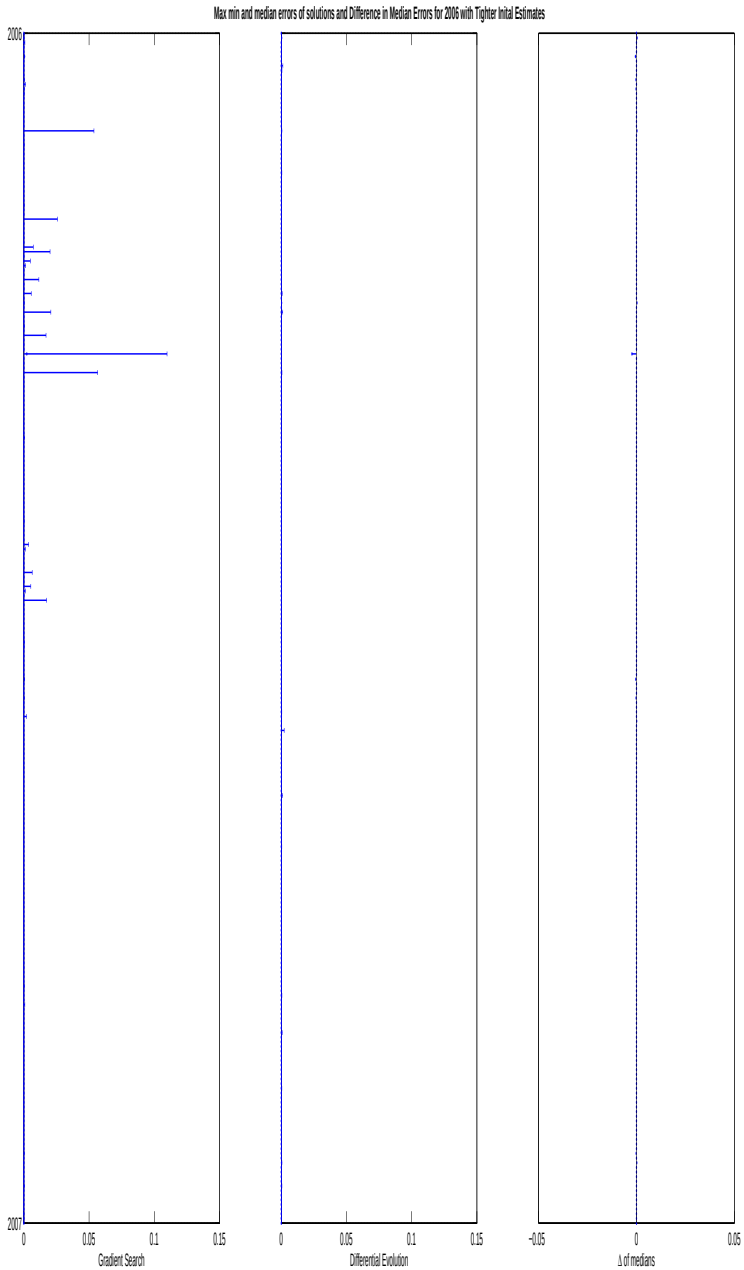


Figure 2.9.13: Max, Min and Median errors for 2006 with Tighter Initial Estimates

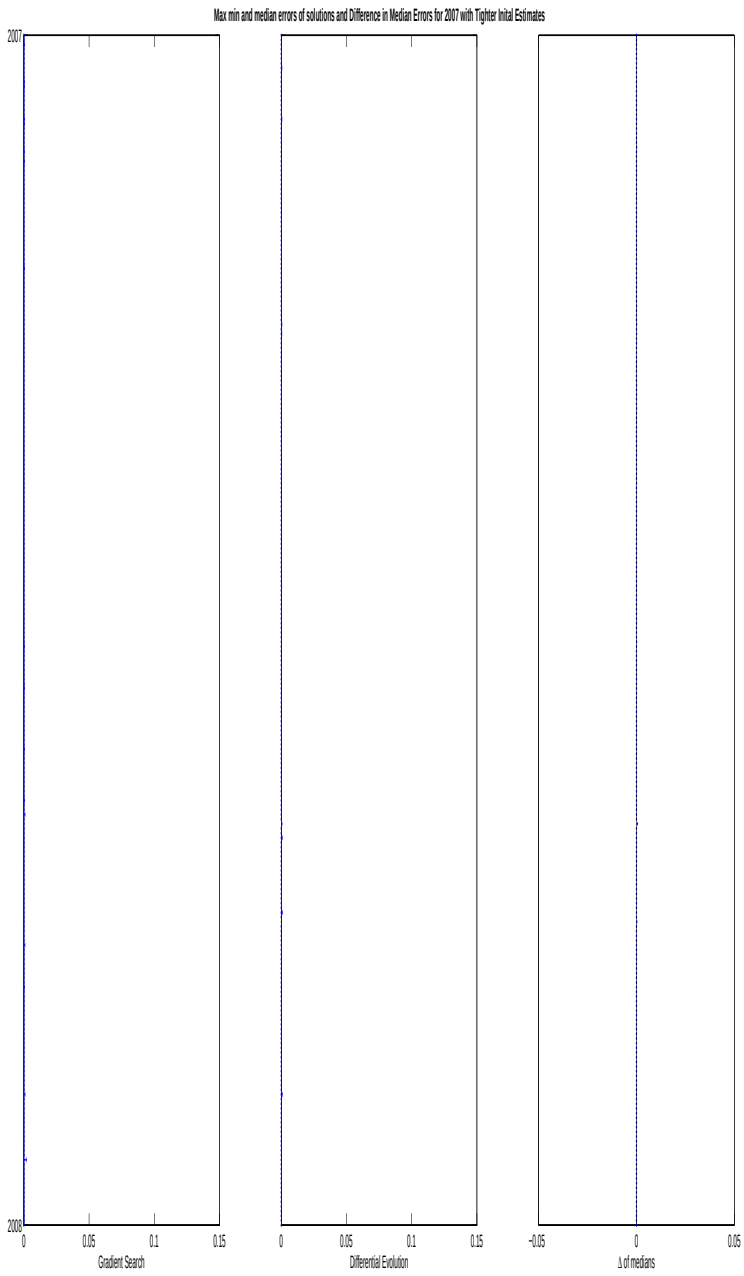


Figure 2.9.14: Max, Min and Median errors for 2007 with Tighter Initial Estimates

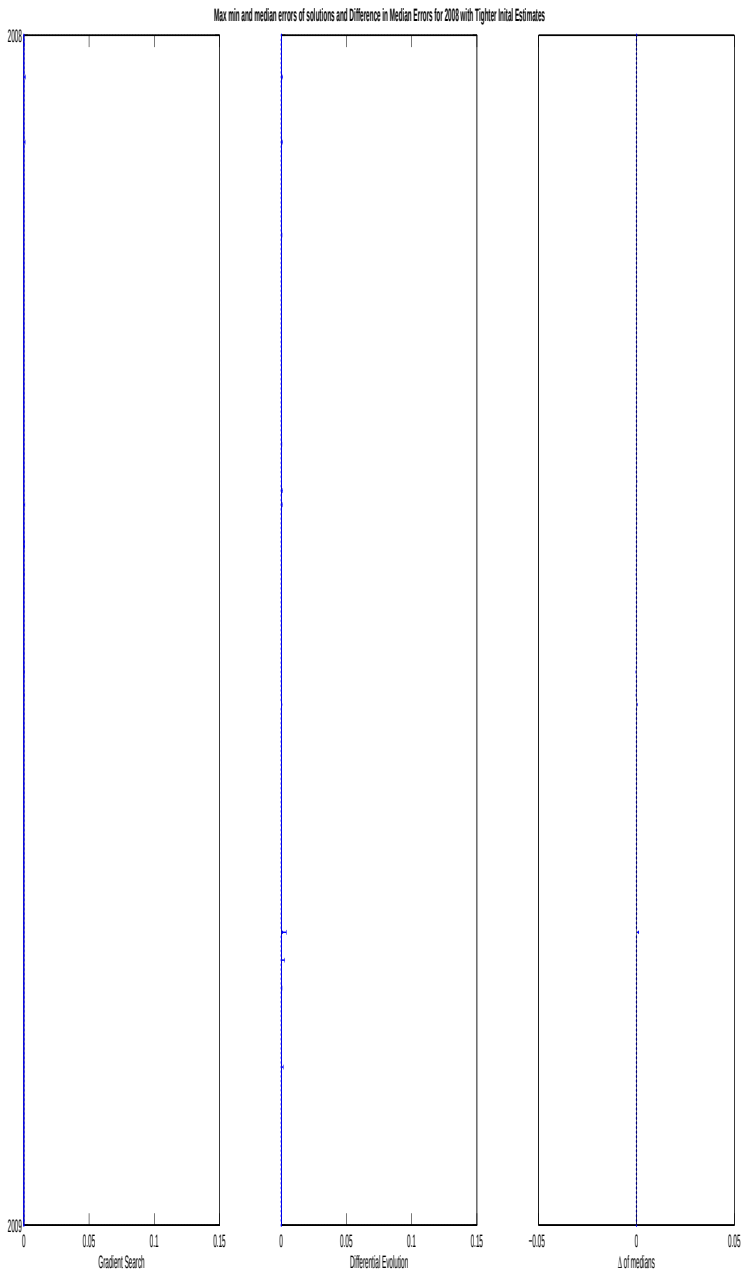


Figure 2.9.15: Max, Min and Median errors for 2008 with Tighter Initial Estimates

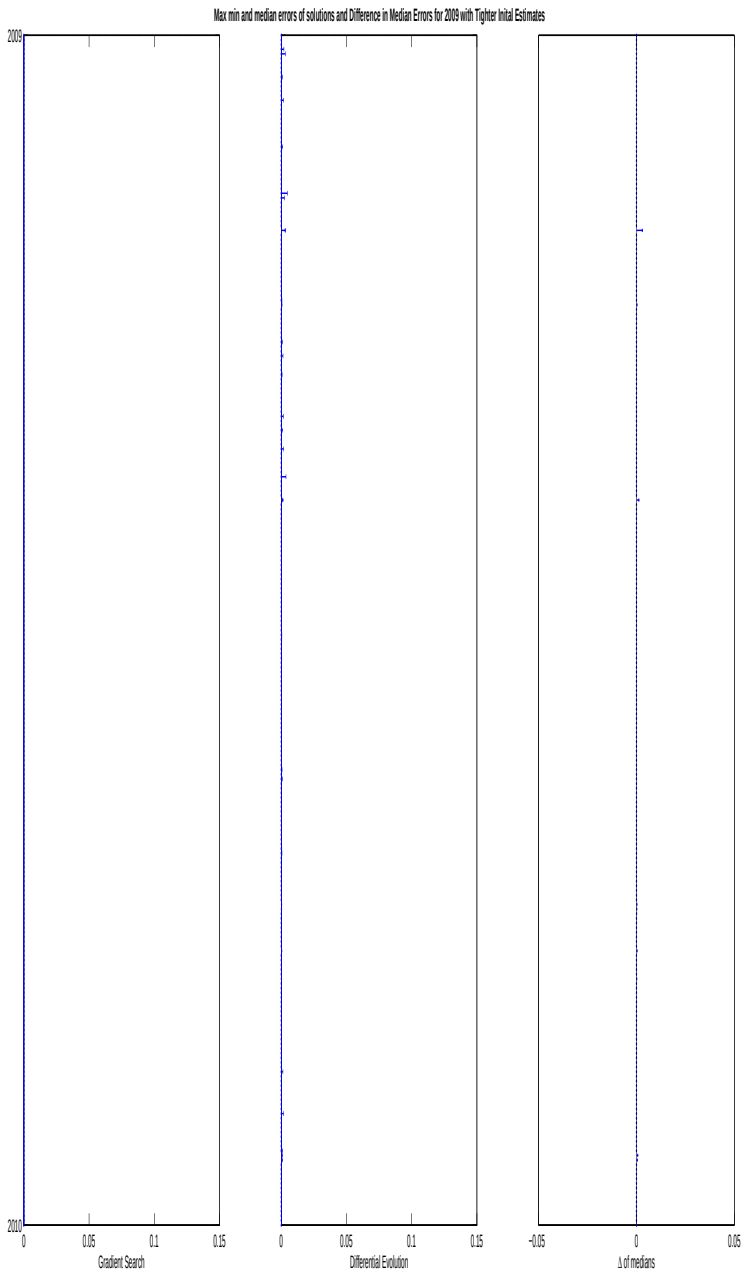


Figure 2.9.16: Max, Min and Median errors for 2009 with Tighter Initial Estimates

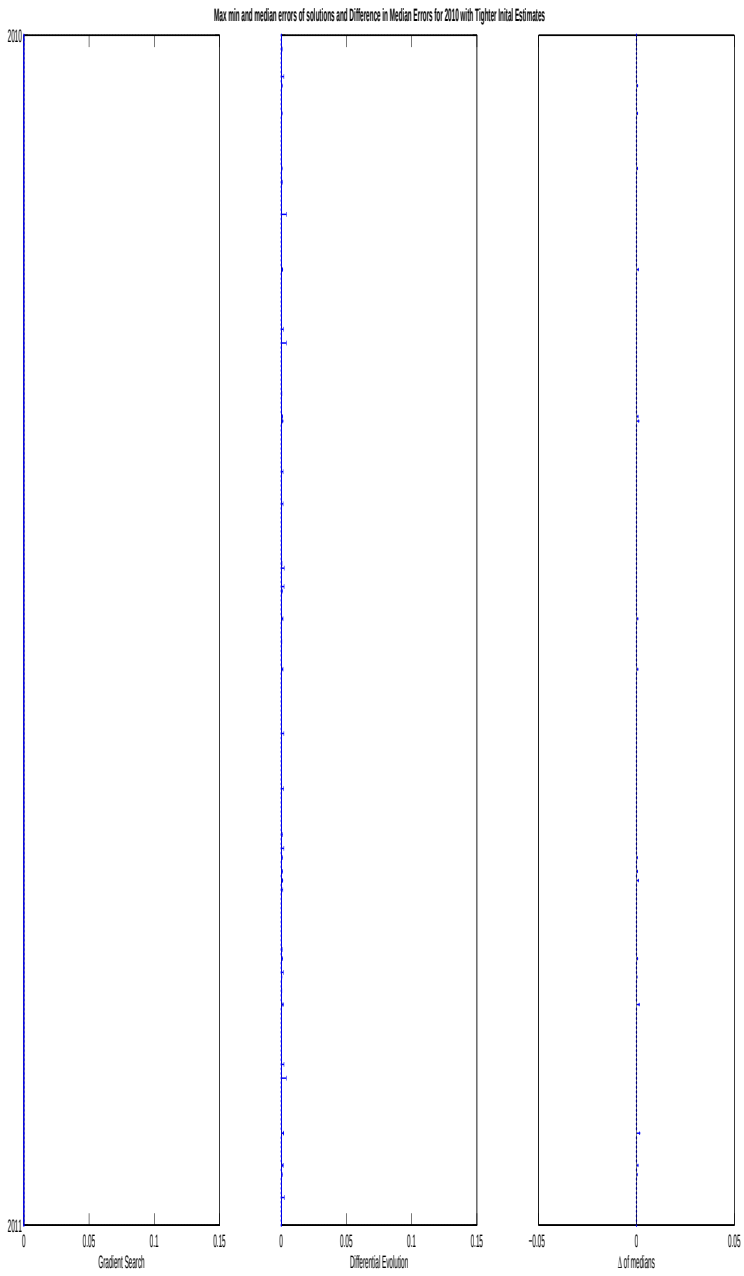


Figure 2.9.17: Max, Min and Median errors for 2010 with Tighter Initial Estimates

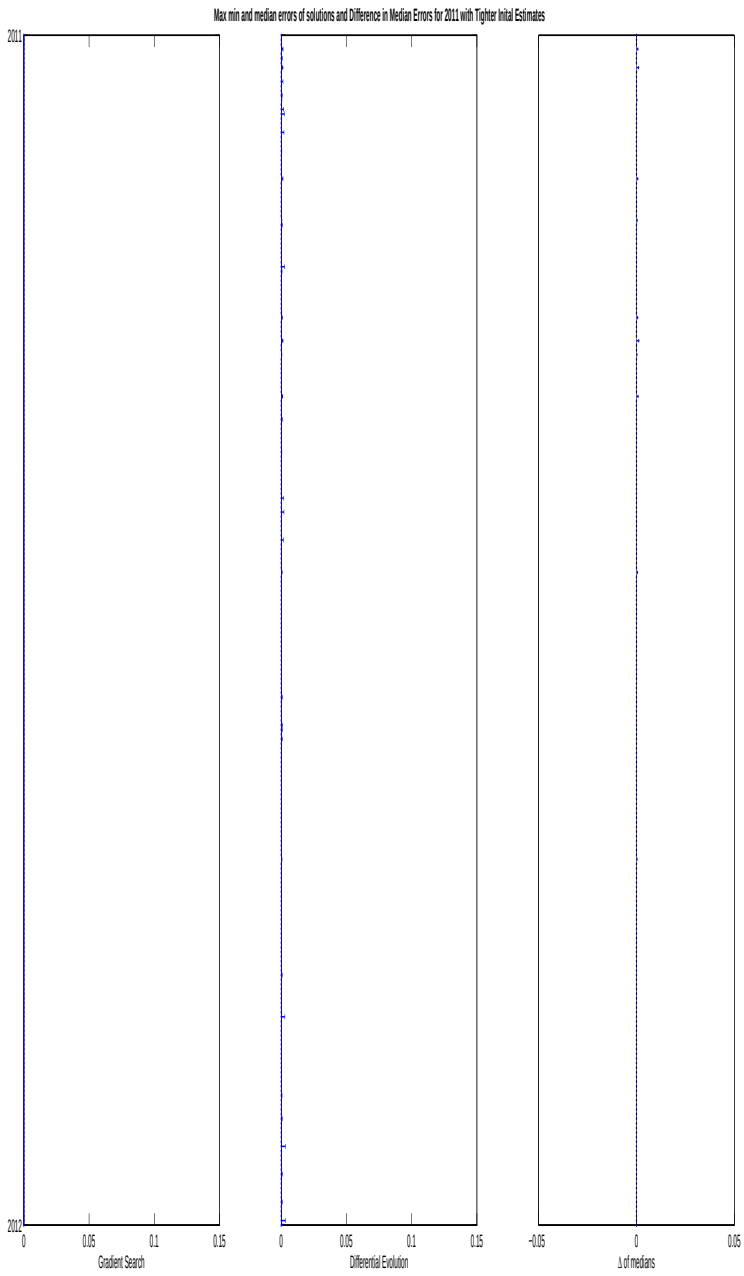


Figure 2.9.18: Max, Min and Median errors for 2011 with Tighter Initial Estimates

Chapter 3

A Simulation Study of Ridge Regression Estimators

3.1 Abstract

Clark and Troskie (2007) which extended the work of Kibria (2003) compared the performance of a set of ridge regression techniques empirically using Monte Carlo simulation. Each candidate ridge regression technique was judged on its mean square error (MSE) performance relative to OLS. This chapter seeks to expand their results by increasing the set of estimators considered to include some of the more recent additions to the ridge regression literature as well as some older ones. This represents the most comprehensive and up to date comparison in the literature to date. In addition some of the estimators considered such as the modified r-k class ridge regression (MCRR) introduced by Batah et al. (2009) did not specify how to estimate all the parameters in their model. This chapter addresses this shortcoming. Moreover the Improved Ridge Estimators (IRE) of Liu et al. (2013) experienced convergence problems and this chapter also tries to remedy these problems, with success in the majority of conditions tested. The Linearized Ridge Regression Estimator (LRRE) of Gao and Liu (2011) also had estimation problems at the lowest volatility levels and an attempt was made to fix this too. Although the solution did fix its convergence issues at the lowest volatility levels, this was at the expense of its performance at every other volatility level. However these amendments to the estimation procedures for the IRE type estimators and the LRRE represent a further contribution to the literature. No estimator is completely dominant under all conditions tested but the LRRE was the best under most scenarios considered.

3.2 Introduction

In chapter two it was shown that Differential Evolution could be adapted successfully to fit the rv-NSS model. However as discussed in section 1.2.2 the second source of the calibration problems is the inherent multicollinearity in models utilizing the NS basis functions. The problem gives rise to two undesirable consequences:

- The beta parameter estimates become unstable
- The problem becomes ill-conditioned

The first problem means that when applying OLS estimation, although the model retains its well known BLUE properties (Best, Linear, Unbiased, Estimator), so that the fit to the observed yields remains optimal in the error sum of squares (ESS) sense, the beta estimates are unreliable often taking large nonsensical values. Given that one of the chief outputs of the rv-NSS model is a time series of the model parameters, unstable parameter estimates are clearly a serious problem. The ill-conditioning which also arises from the multicollinearity means that even slight perturbations in the inputs can have large effects on the estimated parameters, this again means that the estimated parameters will be unreliable.

Anneart et al. (2012) showed how ridge regression can be applied to the NSS model to address the multicollinearity problem. Ridge regression reduces the multicollinearity but the price paid is that it introduces a bias into the estimate. Given that reliable parameter estimates is critical to successful implementation of the rv-NSS model, this made it imperative to find the best estimator possible. This prompted research into ridge regression estimators.

The problem is that although there have been many different ridge regression estimators suggested in the literature, no one estimator can be shown to be universally optimal, with estimator performance varying in response to changes in the level of error volatility and the level of multicollinearity present in the design matrix. To try and identify which estimator was best or at least best in a given level of error volatility/multicollinearity it was decided to conduct a simulation study, extending the work of Clark and Troskie (2007), encompassing the widest range of ridge regression estimators seen in the literature to date and incorporating the most recent estimators. The simulation study tested the estimators on five different design matrices with condition numbers ranging from over 700 (which represents severe multicollinearity) to 5 (which usually indicates no harmful multicollinearity is present). For each matrix eight different levels of error volatility were tested ranging from 0.0001 to 400. Each estimator was judged on the Mean Square Error (MSE) value it produced relative to OLS (the MSE is the variance of the parameter estimate plus its bias squared). Thus it quantifies the trade off between the variance and bias of each estimator providing a good measure upon which to compare the ridge regression estimators.

3.3 Theory

Multiple Linear Regression

The ordinary multiple linear regression model can be represented algebraically as:

$$Y = X\beta + e \tag{3.3.1}$$

- Y is a (n x 1) vector containing the observations of the response variable
- X is a (n x p) fixed matrix containing the observations of the explanatory variables
- β is a (p x 1) vector of unknown parameters
- e is a (n x 1) vector containing the unobserved errors which are assumed to be multivariate normal with $E(e) = 0$ and $Var(e) = \sigma^2$

The ordinary least squares (OLS) estimator is a well known solution to this problem that seeks to minimize the sum of squared residuals (SSR):

$$SSR = (y - X\hat{\beta})^T(y - X\hat{\beta}) \tag{3.3.2}$$

The resulting estimator is known to be BLUE (Best, Linear, Unbiased, Estimator). It can be expressed mathematically as

$$\beta_{OLS} = (X'X)^{-1}X'Y \tag{3.3.3}$$

Multicollinearity

Multicollinearity in the linear regression model refers to the situation in which one of the predictor variables can be expressed as a near linear combination of the others. The OLS estimator is derived under the assumption of no multicollinearity in the X matrix. When this assumption is violated it still retains its BLUE properties but its β estimates become unstable and their variance estimates inflated. These problems have been highlighted in the literature by: Hoerl and Kennard (1970) and Mayer and Wilkie (1973) amongst others:

Belsley (1991) page 1 succinctly summarizes these problems in the quote below:

> There are few statistical practitioners who have escaped the "collinearity problem" in their work, this certainly being true for those who have attempted to use linear regression to estimate models using nonexperimental data. Its symptoms are tell tale: high standard errors, low-t-statistics, nonsensical or overly sensitive parameter estimates.

In practice it is not simply a case of there either being multicollinearity present or not but rather a question of how much multicollinearity is present, see for example Mason et al. (1975).

The presence of strong multicollinearity gives rise to very small eigenvalues in the eigen-decomposition of the $(X'X)$ matrix. It is this that gives rise to the deleterious effects described above.

3.3.1 High standard errors/Low t-values

The reason these small eigenvalues are problematic can be best appreciated when one considers the mathematical expression for the variance of an OLS estimator:

$$Var = \sigma^2 Tr(X'X)^{-1} = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} \tag{3.3.4}$$

Where Tr() represents the trace of a matrix, and λ represents the eigenvalues of the $(X'X)$ matrix. As can clearly be seen a small eigenvalue will cause the variance of the estimators to be inflated.

Ill-Conditioning

The small eigenvalue also gives rise to the ill-conditioning problem. The degree of ill-conditioning can be assessed using the condition number κ:

$$\kappa(X) = \frac{\sigma_{max}(X)}{\sigma_{min}(X)} \tag{3.3.5}$$

Where $\sigma_{max}(X)$ and $\sigma_{min}(X)$ represent the maximal and minimal singular values of X. The singular values are the square roots of the eigenvalues of the $X'X$ matrix so that the condition number too is inflated by the presence of a small eigenvalue (relative to the largest eigenvalue).

Ridge Regression

Ridge type estimators seek to address these problems by adding an amount k to the diagonal of the $X'X$ matrix. By considering the basic Hoerl and Kennard (1970) ridge estimator an appreciation of how it works can be gained:

$$\hat{\beta_k} = (X'X + kI_p)^{-1}X'Y \tag{3.3.6}$$

When the value of the ridge parameter k is set to zero the estimator is the same as the OLS estimator, so the ridge estimator encompasses the OLS estimator. Conversely when you increase k towards infinity the estimator β_k shrinks towards zero (i.e. it is over shrunk). This shrinkage of the estimated parameters means that ridge regressions are part of the shrinkage class of estimators.

Under the ridge formulation the variance of the estimator becomes

$$Var = \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + k} \tag{3.3.7}$$

As k is increased, the variance of the estimator is reduced.

The price paid is that the estimator is biased with the bias for the Hoerl and Kennard (1970) model given in Gross (2003) as:

$$\mathbb{E}(\beta_k) = \beta - k(X'X + kI_p)^{-1}\beta \tag{3.3.8}$$

Strictly this expression for the bias only applies if k is fixed rather than chosen stochastically, as it is in practice, Gunst and Mason (1977).

So the more k is increased, the more biased the estimator becomes.

This trade off between reducing the variance and increasing the bias is best captured by the Mean Square Error Criterion. In the case of linear regression this can be expressed as:

$$MSE(\beta_k) = \mathbb{E}(\hat{\beta_k} - \beta)^2 \tag{3.3.9}$$

This can be shown to be equivalent to

$$MSE(\beta_k) = Var(\beta_k) + bias(\beta_k, \beta)^2 \tag{3.3.10}$$

The key is to find the value of k that provides the optimal trade off between variance and bias. Although optimal estimators can be derived, most rely on knowing the true β parameters and/or variance. In practice empirical approximations to these optimal parameters must be found. Worse, no one estimator works best in all situations so the goal is to identify which estimators are most likely to work best in a given situation.

3.3.2 Literature Review

There have been a vast number of papers looking at different ways to estimate the ridge parameter. The seminal ridge regression paper by Hoerl and Kennard (1970) included a graphical technique called the ridge trace. Golub (1979) introduced a computationally intensive technique based on generalized cross validation. Other methods to estimate the ridge parameter were suggested by McDonald and Galarneau (1975) and Hemmerelle (1975) amongst others. Hoerl et al. (1975) and Hemmerelle (1975) generalized the ridge parameter from a scalar to a vector of values, the estimation of these latter models was furthered in De Boer and Hafner (2005). Goldstein and Smith (1974) derive a generalized

ridge regression (k_i) such that each estimator β_i has a smaller MSE than its LS equivalent.

There have been several comparison papers that have sought to evaluate the relative performances of the different estimators proposed in the literature via computer simulation. Some of the more notable ones are: Gunst and Mason (1977) who compared Ridge Regression, Principal Components Regression (PCR), Least Squares Latent Root and James-stein estimators. They found that PCR performs best although ridge regression performs well for most of the scenarios considered. They attribute the relative underperformance of ridge regression to the difficulty in estimating k.

Dempster et al. (1977) use a factor experiment to study a variety of shrinkage estimators including ridge regression. No one estimator performed uniformly better but ridge regression with a Bayesian method for estimating k was found to be one of the better performers.

Wichern and Churchill (1978) tested Hoerl and Kennard (1970), Lawless and Wang (1976), Hoerl et al. (1975), McDonald and Galarneau (1975) and a graphical technique due to Vinvod (1976). Although no one rule was uniformly better than the others, they found the McDonald and Galarneau (1975) performed well when the correlation between the regressors is moderate to high and that multicollinearity as measured by the ratio of the largest to smallest eigenvalue is high.

Kibria (2003) used a similar scheme to compare four known estimators: Hoerl and Kennard (1970), Hoerl et al. (1975), Lawless and Wang (1976) and Hocking et al. (1976) and three new ones that they introduced. Again no one estimator was found to be optimal in all scenarios although they did find that one of their estimators, Geometric Mean (GM), to be one of the better performing estimators.

Clark and Troskie (2007) extended the work of Kibria (2003) and also introduced two new estimators. Once again no estimator dominated all the others for all of the situations examined.

This chapter aims to expand on the work of Clark and Troskie (2007) and to extend the comparison to include a wider array of models. The additional models include the modified class ridge regression model (MCRR) introduced by Batah et al. (2009) which combines the unbiased ridge regression model (URR) proposed by Crouse et al. (1995), the r-k model introduced by Baye and Parker (1984) and the PCA model introduced by Massy (1965) (this model was later extended by Marquardt (1970)). The MCRR itself was further generalized by Al-Gereari (2012) analogously to the way Marquardt (1970) generalized the PCR model but no methods were given to estimate the additional parameter and so this model was not included in this comparison. In addition this chapter will also include the jack-knifed ridge estimation technique of Singh et al. (1986)(JRR), the idea is to introduce the bias by introducing jack-knifing resampling techniques. Lastly the current chapter includes the Linearized Ridge Regression Estimators (LRRE) introduced by Gao and Liu (2011) and its generalizations the Improved Ridge Estimators introduced in Liu et al. (2013).

3.4 The Models

Following Kibria (2003) and Clark and Troskie (2007) the OLS estimator is put into its canonical form before introducing the estimators as this simplifies the exposition.

First it is necessary to consider the eigen-decomposition of the $X'X$ matrix as $X'X = V\Lambda V'$.

Canonical Form

$$Y = X^*\alpha + e \tag{3.4.1}$$

Where $X^* = XV$ and $\alpha = V'\beta$.

The OLS estimator is thus:

$$\hat{\alpha} = (\Lambda)^{-1}X^{*'}Y \tag{3.4.2}$$

and the generalized ridge regression is

$$\alpha(\hat{K}) = (\Lambda + K)^{-1}X^{*'} \tag{3.4.3}$$

Where K is a diagonal matrix with individual k_i on each diagonal element.

3.4.1 Ridge Parameter Estimators

In each case $\hat{\sigma^2}$ is the residual variance from the OLS regression. Most of the estimators up to the Kibria paper have a similar structure, with the main difference being how the αs are utilized:

3.4.2 Hoerl and Kennard

$$\hat{k}_{HK} = \frac{\hat{\sigma^2}}{max(\hat{\alpha}_i^2)} \tag{3.4.4}$$

As noted by Kibria (2003), (and Clark and Troskie (2007)), Hoerl and Kennard (1970) suggests setting the value of the denominator to the maximum α value.

3.4.3 Hoerl Kennard and Baldwin

$$\hat{k}_{HKB} = \frac{p\hat{\sigma^2}}{\sum_{i=1}^{p} \hat{\alpha}_i^2} \tag{3.4.5}$$

Hoerl et al. (1975) base their estimate on the harmonic mean of all the eigenvalues, p is the number of regressors.

3.4.4 Lawless and Wang

Most ridge estimators are derived under the assumption that $X'X = I_p$ i.e. the identity matrix. When this assumption is violated, Lawless and Wang (1976) show how the problem can be reparameterized to address the issue. Following the derivation given in Gross (2003), $y = Z\vartheta + \epsilon$, $Z = X(X'X)^{-\frac{1}{2}}$, $\vartheta = X(X'X)^{\frac{1}{2}}\beta$ under this reparameterization $Z'Z = I$. The denominator is thus $\hat{\vartheta}'\hat{\vartheta} = \hat{\beta}\Lambda\hat{\beta}$.

$$\hat{k}_{LW} = \frac{p\hat{\sigma^2}}{\hat{\beta}\Lambda\hat{\beta}} \tag{3.4.6}$$

3.4.5 Kibria

In a similar spirit to Hoerl et al. (1975), Kibria (2003) gave a series of estimators in which the αs were averaged in different ways, arithmetic (AM), geometric (GM), and median (MED):

$$\hat{k}_{AM} = \frac{1}{p}\sum_{i=1}^{p} \frac{\hat{\sigma^2}}{\hat{\alpha_i^2}} \tag{3.4.7}$$

$$\hat{k}_{GM} = \frac{\hat{\sigma^2}}{\Pi_{i=1}^{p} \hat{\alpha_i^2}} \tag{3.4.8}$$

$$\hat{k}_{MED} = Median \left\{ \frac{\hat{\sigma^2}}{\hat{\alpha_i^2}} \right\}, i = 1, 2, \ldots, p \tag{3.4.9}$$

3.4.6 Index of Informational Complexity (ICOMP) Ridge Type Estimator

Clark and Troskie (2007) present two new estimators for the ridge parameter based on Bozdogan's ICOMP Bozdogan (1998), which is an information criterion designed to be used in model selection. When applied to estimating the ridge parameter they show that

$$ICOMP(C_1(\widehat{\sum_{\hat{\theta}}})) = -2 \log L(\hat{\theta}) + dN_2 - N_1 \tag{3.4.10}$$

where

$L(\hat{\theta})$ is the log-likelihood function

$N_1 = \log |(\Lambda + kI_p)^{-1} \Lambda (|\Lambda + kI_p)^{-1}| = \sum_{i=1}^{p} \log \left(\frac{\lambda_i}{(\lambda_i + k)^2} \right)$

$N_2 = \log \left(\sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} \right) - \log(d)$

$\sum_{\hat{\theta}}$ is the estimated parameter covariance matrix

d is the rank of $\sum_{\hat{\theta}}$

As noted by Clark and Troskie (2007) this function must be solved numerically, in this study Matlab's fmincon function was used to solve for k. This k value is then $\hat{k}_{ICOMP}$.

3.4.7 K_S Ridge Type Estimator

Clark and Troskie (2007) also presented a second ridge estimator that is based on the ICOMP ridge type estimator and the Lawless and Wang estimator.

$$k_S = \begin{cases} k_{\text{LW}} \text{ if } k_{\text{ICOMP}} < k_{\text{LW}} \\ k_{\text{ICOMP}} \text{ otherwise} \end{cases} \tag{3.4.11}$$

3.4.8 Jackknife Ridge Regression Estimator

As shown in eqn. 3.3.8 the Ridge Regressors are biased. However Singh et al. (1986) show that this bias may be estimated using jack-knifing. The estimated bias is then subtracted from the original Ridge Regression estimate giving you the Jackknife Ridge (JRR) Regression Estimator

$$\hat{\beta_J}(K) = (I - K^2 A^{-2})\hat{\alpha}(K) \tag{3.4.12}$$

Where $A = (\Lambda + K)$

3.4.9 Modified (r-k) Class Ridge Regression Estimator

Batah et al. (2009) melded two previous estimators, the URR and the PCR, to create the Modified (r-k) Class Ridge Regression (MCRR). These earlier models are introduced first:

Unbiased Ridge Regression (URR) Estimator

Swindel (1976) generalized the Hoerl and Kennard (1970) model to incorporate prior information, this allowed shrinkage of the estimated beta parameters to values other than zero. Crouse et al. (1995) based their Unbiased Ridge Regression (URR) on this structure. They allowed their estimates to shrink towards the OLS estimates which are unbiased, creating what they called the URR.

$$\hat{\beta}(k, J) = V(\Lambda + kI_p)^{-1}V'(X'Y + kJ) \tag{3.4.13}$$

$J \sim N(\beta, \frac{\sigma^2}{k}I)$ for $k > 0$

$$\hat{k}_{URR} = \left\{ \begin{array}{ll} \dfrac{p\hat{\sigma^2}}{(\hat{\beta} - J)(\hat{\beta} - J) - \hat{\sigma^2}tr(X'X)^{-1}} & if(\hat{\beta} - J)(\hat{\beta} - J) - \hat{\sigma^2}tr(X'X)^{-1} > 0 \\ \dfrac{p\hat{\sigma^2}}{(\hat{\beta} - J)(\hat{\beta} - J)} & otherwise \end{array} \right\} \tag{3.4.14}$$

Principal Components Regression (PCR)

PCR due to Massy (1965) is not a ridge regression per se but it can also be used to treat multicollinearity, instead of adding the ridge parameter to ensure that the smallest eigenvalues do not cause a problem, the smallest eigenvalues and their associated eigenvectors are removed completely. So the eigenvector decomposition of $X'X$ becomes:

$$X'X = (V_r, V_{p-r}) \begin{pmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{p-r} \end{pmatrix} \begin{pmatrix} V_r' \\ V_{p-r}' \end{pmatrix} \tag{3.4.15}$$

Where V_r represents the matrix containing the ordered set of largest eigenvectors up to r, while V_{p-r} contains the remaining eigenvectors. Λ_r is an $r * r$ diagonal matrix containing the r largest eigenvalues in descending order, similarly Λ_{p-r} is a $(p-r) * (p-r)$ diagonal matrix containing the remaining eigenvalues in descending order.

Thus the decomposition is broken down into the first r principal components and the last $p-r$ principal components; which are discarded.

The PCR estimator can thus be expressed as:

$$\hat{\beta_{PC}}(r) = V_r(\Lambda_r)^{-1} X_r^{*'} Y \tag{3.4.16}$$

Examination of the MCRR expression below shows it to be a combination of the URR and PCR

$$\hat{\beta_{MCRR}}(r, K, J) = V_r(\Lambda_r + KI_r)^{-1} V_r^{'}(X'Y + kJ) \tag{3.4.17}$$

This is also a generalization of the r-k model introduced by Baye and Parker

(1984) allowing shrinkage towards J rather than 0 in the case of the r-k model.

$$\hat{\beta_{r-k}}(r,k) = V_r(\Lambda_r + kI_r)^{-1}V_r^{'}(X'Y) \tag{3.4.18}$$

Note that to facilitate comparisons with the URR model, both these models were tested using k_{URR}.

Unfortunately Batah et al. (2009) does not specify how to estimate the number of principal components to retain. So to make the estimator operational a suitable method to estimate the number of principal components to retain had to be found. The literature on how to estimate the number of principal components to retain is vast, but one of the most common is to base the choice on the condition number, see eqn: 3.3.5. The principal components are removed in order of magnitude, starting with the smallest, until the condition number is less than 10, see Belsley (1991) (p129). This can be considered the benchmark method for selecting which principal components to remove. However this is a rather simplistic approach and has been criticized by Jollife (1982) on the basis that the eigenvectors with smaller eigenvalues may still have predictive power and so removing them will increase the bias see also Mason and Gunst (1985).

To try and find a better method that balances the trade off between variance and bias inherent in these ridge type models, a method was trialled where the selection of the principal components is based on the theoretical MSE for each estimator. The approach can be summarized in the steps below:

- Calculate all possible combinations of the principal components
- Calculate the theoretical MSE (to estimate the bias, the OLS estimate was used as an estimator for the actual β value, on the basis that it is unbiased)
- Select the principal component set that yields the lowest theoretical MSE

Clearly the technique depends on how well the estimators of the unknown variables approximate their theoretical counterparts.

For all the estimators in this class (PCA, r-k, MCRR) both techniques were tried in the study. The estimators where the principal components set as above are simply denoted as PCA, r-k, and MCRR. Those that base their choice on reducing the condition number below the threshold value are denoted: PCA CI, r-k CI and MCRR CI respectively.

3.4.10 Linearized Ridge Regression Estimator

The original Liu estimator introduced by Liu (2003), combined the ORR with the contraction estimator of Mayer and Wilkie (1973). The contraction estimator is another type of shrinkage estimator, as can be seen from the equation below, it shrinks each of the β parameters by the same amount. The Liu estimator has the advantage that it is linear in the d parameter, unlike ridge regression estimators which are non linear in k which makes k harder to estimate. Despite this advantage the Liu estimator has been criticized by Gruber (1998) who questioned whether if gave any improvement over the shrinkage estimators upon which it is based.

$$\begin{aligned} \hat{\beta_{cont}}(\rho) &= \frac{1}{1+\rho} V\hat{\alpha} \\ \hat{\beta_{cont}}(d) &= dV\hat{\alpha} \end{aligned} \tag{3.4.19}$$

where $d = \frac{1}{1+\rho}$

$$\hat{\beta_{Liu}}(d) = (\Lambda + I)^{-1}(\Lambda + dI)\hat{\alpha} \tag{3.4.20}$$

Liu and Gao (2011) derived a transformation between the optimal parameters

of their generalized version of the Liu estimator and those of the Generalized Ridge Regression estimator (GRR) see eqn: 3.4.3. As opposed to regular ridge regression where there is only one k value and all eigenvalues are treated equally, GRR allows for more than one value of k to be used, this means that a larger amount can be added to smaller eigenvalues and a smaller amount added to larger ones. The downside is that there are more parameters to estimate. Given this link they refer to their generalized Liu model as the Linearized Ridge Regression Estimator as by solving for its optimal parameters which is a linear problem, the optimal parameters for the GRR (which is a non-linear problem) could be found. Liu and Gao (2011) solved for the optimal parameters under the Prediction Error Sum of Squares Criterion (PRESS). Gao and Liu (2011) then solved the optimal parameters under the MSE criterion, the optimal parameters found under the MSE criterion are shown below.

$$\hat{\beta_{LRRE}}(D) = V(\Lambda + I)^{-1}(\Lambda + D)V'\hat{\beta_{OLS}} \tag{3.4.21}$$

D is diagonal with the optimal $d_i = d_i*$
$d_i^* = \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$

Clearly this relies on knowing α and σ Gao and Liu (2011) recommend starting with the OLS estimates of these parameters and then iterating until convergence.

The theoretical underpinnings of this model are strong with Gao and Liu (2011) showing that this estimator has optimal MSE properties within what they term the Generalized Shrinkage Estimators (GSE) class of estimators.

$$\hat{\beta_{GSE}}(A) = VAV'\hat{\beta_{OLS}} \tag{3.4.22}$$

This GSE class includes OLS, Hoerl and Kennard (1970), and the Liu type estimators proposed by Liu (1993) and Liu (2003) as well as PCR Massy (1965).

3.4.11 Improved Ridge Estimators

Liu et al. (2013) introduce three estimators, collectively termed, Improved Ridge Estimators (IRE), that have an even lower theoretical MSE than the LRRE. The cost is that these models are more highly parameterized.

The basic form of the estimators is:

$$\hat{\beta_{IRE}}(K, D) = Q(K + D)P_1'Y \tag{3.4.23}$$

Q and D are found from the singular value decomposition (SVD) of X

$$X = P(\Delta, 0)'Q' \tag{3.4.24}$$

P is an $n * n$ matrix

Q is a $p * p$ matrix

Δ is a diagonal matrix with the p largest singular values in descending order.

In each case below $\hat{\theta} = \sum_{j=1}^{p} \delta_j \hat{\alpha_j}$

Ordinary IRE (OIRE)

The first IRE estimator is the OIRE. In this case D is chosen such that:

$$D = \varrho 1_p 1_p' \tag{3.4.25}$$

$$\hat{\varrho^*} = \frac{\sum_{j=1}^{p} \hat{\sigma^2}(\delta_j \hat{\alpha_j^2} - \hat{\theta}\hat{\alpha_j})/(\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2})}{\sum_{j=1}^{p} (\hat{\sigma^2} + \hat{\theta}\delta_j\hat{\alpha})^2/(\hat{\sigma^2} + \delta_j^2\hat{\alpha_j^2}) - (p\hat{\sigma^2} + p\hat{\theta^2})} \tag{3.4.26}$$

$$\hat{k_j^*} = \frac{\delta_j \hat{\alpha_j^2}}{(\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2})} - \hat{\varrho^*} \frac{(\hat{\sigma^2} + \hat{\theta}\delta_j \hat{\alpha_j})}{(\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2})} \tag{3.4.27}$$

Type-I Generalized IRE (GIREI)

$$\hat{D^}= \begin{cases} \hat{d} 1_p' & \text{if } n > 2p, \\ \hat{\varrho} 1_p 1_p' & \text{if } n \leq 2p, \end{cases} \tag{3.4.28}$$

$$\hat{d_j^*} = \frac{\hat{\sigma^2}(\hat{\theta}\hat{\alpha_j} - \delta_j \hat{\alpha_j^2})}{(p\hat{\sigma^2} + \hat{\theta^2})(\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2}) - (\hat{\sigma^2} + \hat{\theta}\delta_j \hat{\alpha_j})^2} \tag{3.4.29}$$

$$\hat{k_j^*} = \frac{\hat{\sigma^2}(p\delta_j \hat{\alpha_j^2} - \hat{\theta}\hat{\alpha_j})}{(p\hat{\sigma^2} + \hat{\theta^2})(\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2}) - (\hat{\sigma^2} + \hat{\theta}\delta_j \hat{\alpha_j})^2} \tag{3.4.30}$$

Type-II Generalized IRE (GIREII)

$$\hat{D} = \begin{cases} 1_p \hat{d'} & \text{if } n > 2p, \\ \hat{\varrho} 1_p 1_p' & \text{if } n \leq 2p, \end{cases} \tag{3.4.31}$$

$$\hat{d_j^*} = \frac{1}{p} \left(\frac{\sum_{i=1}^p \hat{\alpha_i}}{\hat{\sigma^2} + \sum_{i=1}^p \delta_i^2 \hat{\alpha_i^2}} \delta_j \hat{\alpha_j} - \hat{k_j^*} \right) \tag{3.4.32}$$

$$\hat{k_j^*} = \frac{\delta_j \hat{\alpha_j^2}}{q\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2}} - \left(\sum_{i=1}^p \frac{\hat{\alpha_i}}{q\hat{\sigma^2} + \delta_j^2 \hat{\alpha_i^2}} / \sum_{i=1}^p \frac{1}{q\hat{\sigma^2} + \delta_i^2 \hat{\alpha_i^2}} \right) \frac{\delta_j \hat{\alpha_j}}{q\hat{\sigma^2} + \delta_j^2 \hat{\alpha_j^2}} \tag{3.4.33}$$

where $q = (p-1)/p$

Analogously to the LRRE parameters Liu et al. (2013) recommend using the OLS estimates for the unknown parameters and then iterating "until convergence". However in a private communication the corresponding author said that they did not check the convergence properties of these estimators. Through some trial and error it was found that 100 iterations seems to give reasonable results, however convergence in not guaranteed in some cases, the estimator can diverge away from the correct solution. The LRRE also experienced some convergence problems at lower volatility levels.

To try and remedy this, a second method for selecting the optimal parameters was trialled. At each iteration the theoretical MSEs were calculated substituting in the OLS estimates for the true parameter values, (the OLS estimates were used to calculate the bias on the basis that OLS is an unbiased estimator). The parameters chosen were those that produced the lowest MSE value over all the iterations rather than just taking the last value from the iteration process. The rationale being: to still be able to find parameter estimates even when divergences occur and to find the optimal parameter estimates in the MSE sense. Both the parameters based purely on the iterative procedure and those conditioned on their theoretical MSE were tested in the study for both the LRRE and IRE type estimators.

3.5 Testing Methodology

The basic template for comparing ridge estimators has not changed much since the seminal paper of McDonald and Galarneau (1975) but as this study is to extend the results of Clark and Troskie (2007) their methodology is adopted.

The X variables are simulated as shown below:

$$\hat{X}_{ij} = \begin{cases} \sqrt{1-\rho_1^2}Z_{ij} + \rho_1 Z_{i6} & \text{for j=1,2,3,} i = 1, 2, \ldots, n \\ \sqrt{1-\rho_2^2}Z_{ij} + \rho_2 Z_{i6} & \text{for j=4,5,} i = 1, 2, \ldots, n \end{cases} \tag{3.5.1}$$

- Generate the X variables using 3.5.1 with rho pairs (0.99,0.99),(0.99,0.1), (0.9,0.9),(0.9,0.1),(0.7,0.3)
- Simulate the Y variable using the standard β_{act} X and different levels of error volatility i..e. e $\sim$N$(0, \sigma^2)$ σ^2=(0.0001,0.25,1,25,50,75,100,400)
- Standardize the X and Y variables
- Set β_{act} equal to the eigenvector corresponding to the largest eigenvalue from the standardized X matrix
- Evaluate each estimator using the standardized variables and then convert back to the same scale as the original variables
- Calculate the average MSE for each estimator, this is averaged over 5000 simulations of the error variables
- Divide the average MSE of each estimator by the OLS average MSE. A ratio more than 1 means the estimator underperformed OLS while a ratio less than 1 shows how much it outperformed OLS

The βs are set equal to the largest eigenvalue of the standardized X matrix because as is explained in Newhouse and Oman (1971) this value for β minimizes the MSE value for the estimator (strictly this only holds if the explanatory variables are fixed). The scaling chosen is also optimal (or close to optimal) for MSE estimation, Belsley (1991).

3.6 Results

In the following matrices 1-5, correspond to the matrices generated using the ordered rho pairs given in the methodology section. The original estimators shall refer to those estimators already considered in Clark and Troskie (2007) while the new estimators shall refer to those estimators not included in that report. The discussions refer to the full results which may be found in the appendices to this chapter. Included at the beginning of each section is a summary table that summarizes the performance of the estimator that performed best for each ρ pair over the majority of volatility levels tested.

3.6.1 Population Size 30

Summary

		Summary Results for n=30							
		σ^2							
(ρ_1, ρ_2)	Best Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	LRRE	0.465	0.001	0.001	0.001	0.001	0.001	0.001	0.001
$(0.99, 0.10)$	LRRE	0.472	0.007	0.007	0.008	0.007	0.007	0.006	0.006
$(0.90, 0.90)$	LRRE	8.418	0.053	0.032	0.021	0.019	0.019	0.018	0.016
$(0.90, 0.10)$	LRRE	2.626	0.112	0.083	0.069	0.064	0.064	0.062	0.057
$(0.70, 0.30)$	LRRE	1.993	0.179	0.157	0.143	0.134	0.129	0.127	0.121

Table 3.6.1: Best Estimator for Population Size 30

Table 3.6.1 shows the estimator that performed best in the majority of the scenarios for each of the ordered rho pairs.

Original Estimators

For matrices 1 and 2 (having the highest combination of correlations and hence condition numbers), for $\sigma^2 \geq 1$, k_S seems to be best and for $\sigma^2 \geq 50$ these results are matched by k_{LW} which is not too surprising as the k_S defaults to k_{LW} when $k_{ICOMP} < k_{LW}$. Indeed k_S and k_{LW} were able to match the performance of the best estimator LRRE at these highest volatilities. At the lowest volatilities k_{GM} tends to produce the best results. k_{AM} performs well for higher volatilities giving results comparable to k_S, k_{LW}, but at the lowest volatility level tested it seriously underperforms giving the worst results of the estimators tested. When tested on matrix 1 with intermediate volatilities $\sigma^2 = 0.25$ and $\sigma^2 = 1$ k_{AM} outperforms k_S, k_{LW}, but for matrix 2 k_S outperforms k_{AM}, this is because k_{ICOMP} outperforms in these scenarios and k_S can switch to k_{ICOMP}. Overall the performance of k_{ICOMP} was much better for matrix 2.

For matrices 3 and 4 k_{AM} gives the best performance at higher volatilities giving performances close to the LRRE results, however, at the lowest volatility levels k_{AM} still performs poorly. k_{GM} continues to dominate at the lowest volatility levels.

For the matrix with the lowest condition number, matrix 5, there is comparatively little multicollinearity, k_{HKB} is best or close to best of the original estimators for $\sigma^2 \leq 1$. After that k_{GM} and k_{AM} are best for higher volatilities, indeed k_{AM} outperforms LRRE for most of the high volatility scenarios but as volatility decreases so does k_{AM}'s out-performance so that for $\sigma^2 \leq 50$ LRRE is better. However both k_{GM} and k_{AM} struggle at lower volatility levels, with k_{AM} once again performing particularly poorly at low levels of volatility. k_S and k_{LW} perform consistently well across all the volatilities considered.

New Estimators

For the new estimators, JRR performs pretty consistently across all the volatilities and correlation combinations but only manages to reduce the MSE to circa 60% of the OLS value, a similar reduction to that achieved by k_{HK}. It seems that the bias reduction offered by the jack-knifing is being offset by an increase in the variance of the estimator. This criticism has been levelled at the jack-knifed ridge estimator previously see Gross (2003) p143.

The URR performs pretty consistently too providing good improvements over the OLS model for all combinations of volatility and correlation, indeed for the highest correlation and lowest volatility it is the best estimator. It is also consistently better than the JRR estimator.

As noted in section 3.4.9 the PCA based estimators appear as the PCR which chooses the components so as to minimize the theoretical MSE and the PCR CI which removes the eigenvectors in reverse order of their associated eigenvalue size until the condition number of the matrix is reduced below 10. For the correlation combinations associated with the highest condition number the PCR CI performs better than PCR, indeed for the matrix with the highest condition number (matrix 1) it is one of the best estimators in the study. For the test matrices with lower condition numbers PCR CI can only match the OLS result and is outperformed by PCR. The performance of the PCR CI for the matrix with the highest condition number demonstrates that the technique can be successful when the appropriate condition number threshold is applied. However getting the right threshold does require some judgment and often auxiliary regressions, Belsley (1991), which are hard to automate. It's little surprise that PCR CI performed no better than OLS for the matrices with lower condition numbers (matrices 3, 4 and 5) as their condition numbers are already below the condition number threshold of 10 to begin with, so all the principal components will be included, hence the PCR CI estimate will be the same as the OLS estimate.

As with the pure PCA estimators above, the r-k and MCRR estimators which also incorporate principal components regression as part of their estimation, also appear in two versions. The r-k and MCRR base their choice of principal components to minimize their theoretical MSE value, the r-k CI and MCRR CI remove principal components until the condition number of the matrix is reduced below 10. For the matrices 1 and 2 the CI versions performed best. Indeed the CI versions were amongst the best estimators in the whole study for matrix 1 (apart from the lowest volatility level where r-k CI underperformed). However as the condition number of the matrices drops so does the outperformance of the CI versions. For matrices 3-5 the performance of the two versions of the estimator are broadly similar. However unlike the PCA based estimators above, the CI versions of these estimators are able to outperform OLS for all matrices considered for all but the lowest volatility levels.

As can be seen from table 3.6.1 the LRRE is consistently amongst the best or close to the best of the estimators considered. The only scenarios in which the LRRE is not best or close to best was at the lowest volatility level for matrices 3, 4 and 5. To address this the choice of $\hat{\beta}$ was conditioned on the theoretical MSE it produces, although the theoretical MSE itself relies on estimated parameters it was hoped that it would reduce the incidence of large MSE values from being generated. This had the desired effect in that the performance of the LRRE was improved at the lowest volatility levels but at the cost of its performance being compromised at every other volatility level.

The improved ridge estimators' performance was qualitatively similar to the LRRE but they are consistently worse than the LRRE (and some of the other estimators considered). This is surprising as they can be shown to be theoretically superior to the LRRE, Liu et al. (2013) (in this paper the LRRE is called the CRE and is derived and expressed differently, but the two however can be shown to be mathematically equivalent). This theoretical advantage is thanks in part to their higher parameterization. However it seems that this high parame-

terization may be a double edged sword as the increased number of parameters to be estimated comes with associated higher estimation errors and these seem to offset the theoretical edge of the improved ridge models. It is possible that these models may benefit from larger numbers of iterations commensurate with their higher level of estimated parameters and given suitably high level of iterations it would catch and perhaps even surpass the LRRE. However as was previously noted convergence of these estimators is not guaranteed and warrants further investigation.

Similar to the PCA type models the IRE type models come in two forms: IRE MSE which seeks to base the choice of $\hat{\beta}$ on the choice that minimized the theoretical MSE. The other version simply uses the value that is observed after the last iteration. In contrast to the LRRE version the MSE versions of the IRE estimators perform better than the results generated purely via iteration in the majority of scenarios considered although at higher volatility levels the improvement tended to be modest.

3.6.2 Population Size 50

Summary

		Summary Results for n=50							
		σ^2							
(ρ_1, ρ_2)	Best Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	PCR CI	1.930	0.003	0.002	0.001	0.001	0.001	0.001	0.001
$(0.99, 0.10)$	LRRE	0.032	0.006	0.006	0.007	0.006	0.006	0.006	0.005
$(0.90, 0.90)$	LRRE	1.915	0.065	0.045	0.036	0.034	0.033	0.031	0.028
$(0.90, 0.10)$	LRRE	3.627	0.098	0.076	0.070	0.068	0.065	0.063	0.056
$(0.70, 0.30)$	LRRE	1.048	0.339	0.252	0.214	0.204	0.197	0.190	0.173

Table 3.6.2: Best Estimator for Population Size 50

Table 3.6.2 shows the estimator that performed best in the majority of the scenarios for each of the ordered rho pairs.

Original Estimators

The pattern is similar to that above for matrix 1 and matrix 2: k_S and k_{LW} continuing to do well at the highest volatilities. However k_{GM} dominates them at the lowest levels of volatility. The pattern also repeats for k_{AM}. It performs well for the higher volatilities, outperforming k_S and k_{LW} at the intermediate volatility levels $\sigma^2 = 0.25$ and $\sigma^2 = 1$ for matrix 1, but again underperforming k_S at those volatility levels for matrix 2 as k_S benefits from the improved performance of k_{ICOMP}. Again it is one of the worst performing estimators at the lowest volatility level and in many cases is the single worst estimator at this level of volatility.

The performances of the original estimators for matrices 3-5 are broadly similar to those for population size 30. k_{AM} continues to outperform LRRE at the highest volatility levels, k_{GM} also outperforms LRRE at the highest volatilities but still under performs k_{AM}. However LRRE is still the better at the lower volatilities and still manages to be the best in the majority of the volatility levels tested for matrix 5, so is still the best estimator for this matrix.

New Estimators

Both JRR and URR again produce consistent results, with the URR dominating the JRR, producing results that approach those of the best estimators for almost all the scenarios considered.

PCR CI is again very good when applied to matrix 1, it marginally outperforms LRRE for population size 50, having been marginally outperformed by LRRE for population size 30. Overall the results for the PCA based estimators: PCR, PCR CI, r-k, r-k CI, MCRR and MCRR CI follow the same pattern for the population of 50 observations as they did for the population of 30.

As can be seen from Table 3.6.2 LRRE continues to be the best or close to best in all scenarios, the only exceptions are at the lowest volatility level, where again it is outperformed by LRRE MSE. The other IRE estimators follow a similar pattern to LRRE but continue to underperform the LRRE in all the scenarios considered.

3.6.3 Population Size 100

Summary

		Summary Results for n=100							
		σ^2							
(ρ_1, ρ_2)	Best Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	LRRE	0.465	0.001	0.001	0.001	0.001	0.001	0.001	0.001
$(0.99, 0.10)$	LRRE	0.472	0.007	0.007	0.008	0.007	0.007	0.006	0.006
$(0.90, 0.90)$	LRRE	8.418	0.053	0.032	0.021	0.019	0.019	0.018	0.016
$(0.90, 0.10)$	LRRE	2.626	0.112	0.083	0.069	0.064	0.064	0.062	0.057
$(0.70, 0.30)$	LRRE	1.993	0.179	0.157	0.143	0.134	0.129	0.127	0.121

Table 3.6.3: Best Estimator for Population Size 100

Table 3.6.3 shows the estimator that performed best in the majority of the scenarios for each of the ordered rho pairs.

The results again follow the pattern set from the results run on the populations of 30 and 50, so it would seem that there is little difference between the results computed on the population sizes tested. Overall the results for the original estimators follow a very similar pattern to that found in Clark and Troskie (2007).

3.7 Conclusions

This chapter tests and compares the largest set of ridge type estimators and as such represents a new contribution to the literature.

In addition one of the estimators, MCRR, was proposed without a method to estimate one of its key parameters. Two methods were found to address this shortcoming, this also represents a new contribution to the literature.

Although no one estimator dominated in all the scenarios considered, the LRRE came closest to fulfilling this goal. It only struggled at the lowest volatility levels. As described in the text it was possible to address the problems the LRRE experienced at the lowest volatility level by conditioning the β estimate on its theoretical MSE value, but this was done at the expense of compromising the LRRE's performance in all the remaining scenarios considered, so no complete solution was found.

The PCA based estimators also performed well especially MCRR so it may be that combining PCA techniques with LRRE could produce improved results. This is despite the fact that LRRE was shown to be optimal in the class of estimators that included PCR, but as was demonstrated by the performance of the LRRE versus the IRE type estimators theoretical edge does not always translate into empirical results. Indeed there are a plethora of ways to choose which principal components to retain in a PCA regression, testing them all would be outside the scope of this chapter, however further investigation may yield additional performance gains for these techniques.

The URR is also worth a mention, despite being one of the older estimators it provided good estimates in all scenarios tested.

The performance of the IRE type estimators was a little underwhelming given

their theoretical edge, they experienced problems with their convergence which the current chapter tried to address by conditioning the choice of the optimal $\hat{\beta}$ on its theoretical MSE value. This was successful in the majority of the scenarios tested but the improvement was small at the higher volatility levels tested. A more thorough study of these estimators' convergence properties and or better estimates of its theoretical MSE using for example the bias estimator from Gross (2003) p142, may allow these estimators to realize more of their theoretical edge.

Given that no one estimator dominated in all scenarios it might be possible to develop a composite estimator by selecting the most appropriate estimator for the given condition number of the design matrix and error volatility, this is left as future work.

These results meant that the last calibration problem relating to the rv-NSS had been addressed and it was now possible to calibrate and test the model. However as will be seen in chapter four the direction of the research changed at this point to focus on a dynamic version of the model based on the Dynamic Nelson and Siegel model (DNS) which was first introduced by Diebold and Li (2006).

3.8 Appendices

3.8.1 Appendix 1: Condition Numbers for each Population Size

Condition Numbers			
		n	
(ρ_1, ρ_2)	30	50	100
$(0.99, 0.99)$	724.951	336.765	287.586
$(0.99, 0.1)$	196.553	272.875	197.219
$(0.9, 0.9)$	52.284	34.705	40.548
$(0.9, 0.1)$	29.179	29.671	17.09
$(0.7, 0.3)$	10.439	5.055	6.049

Table 3.8.1: Condition Numbers by Correlation Pair and Population Size

Note that the condition number shown here is the ratio of the highest to lowest eigenvalues, rather than the ratio of highest to lowest singular values used in the rest of the chapter. This was done to facilitate comparisons with the appendices of Clark and Troskie (2007) who use this alternate definition of condition number.

3.8.2 Appendix 2: Simulation Results for Population n=30

Eigenvalues of X'X and the β values for n=30					
Correlation Structure	(0.99,0.99)	(0.99,0.1)	(0.9,0.9)	(0.9,0.1)	(0.7,0.3)
λ_1	4.950	3.183	4.423	2.763	2.477
λ_2	0.018	1.044	0.208	1.030	1.050
λ_3	0.014	0.738	0.178	0.971	0.863
λ_4	0.011	0.019	0.107	0.142	0.372
λ_5	0.007	0.016	0.085	0.095	0.237
β_1	0.447	-0.550	-0.457	-0.573	0.548
β_2	0.447	-0.549	-0.448	-0.578	0.527
β_3	0.446	-0.553	-0.44	-0.556	0.568
β_4	0.448	-0.179	-0.454	0.065	0.091
β_5	0.448	-0.241	-0.438	-0.157	0.302

Table 3.8.2: Eigenvalues and Beta Values for Population Size 30

Simulation Results for n=30									
		σ^2							
(ρ_1, ρ_2)	Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	HK	1.000	0.587	0.564	0.551	0.557	0.559	0.559	0.558
	HKB	0.998	0.354	0.311	0.277	0.282	0.283	0.283	0.280
	LW	1.000	0.415	0.179	0.011	0.005	0.003	0.002	0.001
	ICOMP	0.873	0.469	0.450	0.415	0.423	0.426	0.426	0.421
	S	0.873	0.309	0.133	0.009	0.004	0.003	0.002	0.001
	GM	0.651	0.176	0.146	0.068	0.059	0.053	0.050	0.040
	AM	52.214	0.131	0.086	0.010	0.006	0.004	0.003	0.002
	MED	0.688	0.265	0.181	0.130	0.131	0.130	0.130	0.128
	JRR	1.000	0.673	0.613	0.563	0.569	0.570	0.570	0.567
	URR	0.431	0.176	0.178	0.170	0.177	0.180	0.179	0.174
	PCR	0.973	0.807	0.811	0.806	0.809	0.810	0.809	0.811
	PCR CI	0.476	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	R-K	2.132	0.178	0.178	0.170	0.177	0.180	0.179	0.174
	R-K CI	2.170	0.003	0.001	0.001	0.001	0.001	0.001	0.001
	MCRR	0.439	0.170	0.172	0.165	0.172	0.174	0.173	0.168
	MCRR CI	0.476	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	LRRE	0.465	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	LRRE MSE	0.645	0.486	0.486	0.478	0.485	0.487	0.487	0.486
	OIRE	1.479	0.294	0.145	0.149	0.162	0.164	0.163	0.159
	OIRE MSE	1.587	0.097	0.072	0.137	0.153	0.156	0.155	0.152
	GIREI	1.000	0.902	0.998	0.480	0.389	0.346	0.317	0.261
	GIREI MSE	1.000	0.703	0.600	0.373	0.325	0.299	0.279	0.239
	GIREII	1.029	0.487	0.359	0.233	0.239	0.237	0.231	0.216
	GIREII MSE	1.046	0.438	0.305	0.187	0.193	0.192	0.189	0.179
$(0.99, 0.1)$	HK	1.000	0.59	0.541	0.525	0.52	0.523	0.522	0.521
	HKB	0.998	0.320	0.257	0.218	0.212	0.214	0.212	0.212

	LW	0.999	0.369	0.154	0.020	0.014	0.012	0.011	0.009
	ICOMP	0.892	0.100	0.089	0.085	0.083	0.086	0.084	0.086
	S	0.892	0.088	0.058	0.015	0.012	0.011	0.010	0.009
	GM	0.728	0.062	0.043	0.020	0.018	0.016	0.016	0.014
	AM	101.610	0.373	0.142	0.016	0.010	0.007	0.006	0.004
	MED	0.772	0.196	0.118	0.020	0.016	0.014	0.013	0.012
	JRR	1.000	0.633	0.533	0.463	0.454	0.457	0.455	0.454
	URR	0.988	0.162	0.151	0.149	0.145	0.147	0.146	0.146
	PCR	0.944	0.813	0.816	0.817	0.809	0.811	0.814	0.812
	PCR CI	0.474	0.022	0.022	0.022	0.022	0.021	0.022	0.022
	R-K	0.933	0.163	0.151	0.148	0.144	0.146	0.145	0.145
	R-K CI	0.474	0.025	0.020	0.017	0.017	0.017	0.017	0.017
	MCRR	0.934	0.156	0.147	0.146	0.141	0.144	0.143	0.143
	MCRR CI	0.474	0.023	0.020	0.018	0.018	0.017	0.017	0.017
	LRRE	0.472	0.007	0.007	0.008	0.007	0.007	0.006	0.006
	LRRE MSE	0.639	0.488	0.490	0.491	0.486	0.489	0.487	0.485
	OIRE	0.845	0.055	0.103	0.158	0.154	0.156	0.153	0.150
	OIRE MSE	1.031	0.038	0.085	0.152	0.148	0.150	0.147	0.144
	GIREI	1.000	0.974	0.961	0.360	0.302	0.280	0.265	0.236
	GIREI MSE	1.000	0.539	0.485	0.292	0.262	0.249	0.238	0.219
	GIREII	0.853	0.213	0.160	0.143	0.137	0.138	0.133	0.131
	GIREII MSE	0.792	0.192	0.156	0.138	0.131	0.132	0.127	0.126
$(0.9, 0.9)$	HK	1.000	0.874	0.709	0.568	0.564	0.562	0.564	0.561
	HKB	1.000	0.649	0.475	0.300	0.292	0.289	0.291	0.285
	LW	1.000	0.861	0.654	0.102	0.070	0.059	0.054	0.043
	ICOMP	0.998	0.640	0.594	0.545	0.543	0.542	0.545	0.540
	S	0.998	0.639	0.539	0.096	0.068	0.058	0.054	0.043
	GM	0.996	0.306	0.244	0.113	0.099	0.091	0.087	0.074
	AM	1.024	0.811	0.380	0.065	0.044	0.035	0.031	0.021
	MED	0.998	0.361	0.324	0.145	0.136	0.133	0.133	0.129

	JRR	1.000	0.919	0.790	0.593	0.582	0.578	0.580	0.573
	URR	1.014	0.197	0.187	0.182	0.181	0.181	0.183	0.179
	PCR	1.000	0.828	0.817	0.813	0.812	0.812	0.815	0.814
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.018	0.275	0.217	0.183	0.180	0.180	0.183	0.178
	R-K CI	1.018	0.275	0.217	0.183	0.180	0.180	0.183	0.178
	MCRR	1.014	0.193	0.182	0.177	0.176	0.176	0.179	0.174
	MCRR CI	1.014	0.197	0.187	0.182	0.181	0.181	0.183	0.179
	LRRE	8.418	0.053	0.032	0.021	0.019	0.019	0.018	0.016
	LRRE MSE	1.064	0.502	0.493	0.493	0.491	0.491	0.493	0.489
	OIRE	0.999	0.229	0.236	0.218	0.205	0.200	0.201	0.189
	OIRE MSE	0.998	0.182	0.207	0.208	0.197	0.193	0.194	0.183
	GIREI	1.000	0.932	1.008	0.428	0.344	0.310	0.294	0.248
	GIREI MSE	1.000	0.633	0.586	0.350	0.298	0.274	0.266	0.230
	GIREII	0.978	0.449	0.376	0.297	0.271	0.258	0.254	0.231
	GIREII MSE	0.976	0.450	0.379	0.268	0.245	0.235	0.232	0.212
$(0.9, 0.1)$	HK	1.000	0.905	0.759	0.576	0.565	0.564	0.563	0.566
	HKB	1.000	0.706	0.521	0.304	0.289	0.286	0.283	0.281
	LW	1.000	0.842	0.630	0.156	0.125	0.115	0.107	0.095
	ICOMP	0.999	0.539	0.453	0.366	0.356	0.356	0.355	0.355
	S	0.999	0.538	0.440	0.147	0.121	0.114	0.106	0.095
	GM	0.998	0.484	0.316	0.117	0.100	0.093	0.088	0.077
	AM	56.142	2.452	1.063	0.116	0.072	0.057	0.049	0.034
	MED	0.998	0.628	0.412	0.135	0.113	0.108	0.102	0.092
	JRR	1.000	0.937	0.804	0.564	0.546	0.543	0.540	0.537
	URR	1.000	0.537	0.406	0.231	0.219	0.218	0.216	0.214
	PCR	1.064	0.847	0.828	0.818	0.818	0.818	0.818	0.819
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.063	0.491	0.370	0.210	0.198	0.197	0.195	0.194
	R-K CI	0.999	0.498	0.372	0.210	0.198	0.197	0.195	0.194

	MCRR	1.063	0.481	0.346	0.216	0.206	0.206	0.204	0.201
	MCRR CI	1.000	0.537	0.406	0.231	0.219	0.218	0.216	0.214
	LRRE	2.626	0.112	0.083	0.069	0.064	0.064	0.062	0.057
	LRRE MSE	1.094	0.527	0.504	0.497	0.492	0.491	0.490	0.491
	OIRE	96.090	0.266	0.425	0.242	0.216	0.210	0.204	0.197
	OIRE MSE	51.462	0.224	0.298	0.228	0.207	0.203	0.198	0.192
	GIREI	1.001	1.165	1.080	0.396	0.323	0.294	0.278	0.248
	GIREI MSE	1.000	0.652	0.583	0.324	0.281	0.262	0.252	0.231
	GIREII	2.356	0.750	0.445	0.275	0.247	0.238	0.232	0.221
	GIREII MSE	2.897	0.510	0.356	0.257	0.233	0.226	0.220	0.210
$(0.7, 0.3)$	HK	1.000	0.959	0.875	0.597	0.585	0.579	0.578	0.572
	HKB	1.000	0.842	0.678	0.347	0.327	0.318	0.315	0.306
	LW	1.000	0.922	0.790	0.273	0.232	0.215	0.208	0.190
	ICOMP	1.000	0.833	0.746	0.606	0.598	0.594	0.593	0.589
	S	1.000	0.833	0.732	0.270	0.231	0.215	0.208	0.190
	GM	1.036	1.325	0.750	0.192	0.154	0.139	0.132	0.116
	AM	2043.436	6.822	2.632	0.244	0.138	0.104	0.088	0.062
	MED	1.014	1.741	0.883	0.233	0.190	0.176	0.168	0.152
	JRR	1.000	0.982	0.912	0.627	0.608	0.599	0.596	0.587
	URR	1.000	0.725	0.515	0.250	0.236	0.231	0.231	0.229
	PCR	1.021	0.847	0.827	0.829	0.829	0.829	0.827	0.822
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.020	0.741	0.647	0.234	0.213	0.205	0.202	0.196
	R-K CI	1.000	0.748	0.649	0.234	0.213	0.205	0.202	0.196
	MCRR	1.021	0.577	0.422	0.240	0.228	0.223	0.223	0.218
	MCRR CI	1.000	0.725	0.515	0.250	0.236	0.231	0.231	0.229
	LRRE	1.993	0.179	0.157	0.143	0.134	0.129	0.127	0.121
	LRRE MSE	1.042	0.535	0.519	0.514	0.510	0.504	0.502	0.493
	OIRE	5.249	1.009	0.916	0.316	0.272	0.251	0.243	0.217
	OIRE MSE	5.249	0.542	0.424	0.286	0.259	0.243	0.237	0.213

GIREI	1.000	1.058	1.022	0.407	0.341	0.305	0.290	0.253
GIREI MSE	1.000	0.649	0.603	0.355	0.309	0.280	0.270	0.238
GIREII	5.432	0.774	0.521	0.361	0.322	0.298	0.286	0.250
GIREII MSE	4.496	0.562	0.422	0.341	0.308	0.285	0.275	0.240

Table 3.8.3: Simulation Results for Population Size 30

3.8.3 Appendix 3: Simulation Results for Population n=50

Eigenvalues of X'X and the β values for n=50					
Correlation Structure	(0.99,0.99)	(0.99,0.1)	(0.9,0.9)	(0.9,0.1)	(0.7,0.3)
λ_1	4.906	2.994	4.22	2.879	2.229
λ_2	0.033	1.076	0.265	0.986	1.002
λ_3	0.026	0.902	0.217	0.855	0.857
λ_4	0.021	0.017	0.177	0.182	0.471
λ_5	0.015	0.011	0.122	0.097	0.441
β_1	-0.448	0.573	-0.444	0.561	-0.534
β_2	-0.448	0.574	-0.443	0.554	-0.542
β_3	-0.447	0.574	-0.448	0.547	-0.548
β_4	-0.447	0.039	-0.46	0.253	-0.151
β_5	-0.447	-0.104	-0.441	0.122	-0.312

Table 3.8.4: Eigenvalues and Beta Values for Population Size 50

Simulation Results for n=50									
		σ^2							
(ρ_1, ρ_2)	Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	HK	1.000	0.655	0.580	0.560	0.558	0.558	0.557	0.556
	HKB	0.999	0.417	0.337	0.284	0.280	0.280	0.279	0.277
	LW	1.000	0.607	0.313	0.021	0.01	0.007	0.005	0.003
	ICOMP	0.954	0.479	0.460	0.415	0.409	0.409	0.407	0.402
	S	0.954	0.427	0.230	0.017	0.009	0.006	0.005	0.003
	GM	0.822	0.204	0.166	0.080	0.068	0.062	0.058	0.046
	AM	127.167	0.210	0.120	0.019	0.011	0.008	0.007	0.004
	MED	1.219	0.298	0.230	0.135	0.131	0.131	0.131	0.128
	JRR	1.000	0.747	0.648	0.572	0.567	0.567	0.566	0.563
	URR	0.814	0.163	0.164	0.166	0.166	0.166	0.165	0.165
	PCR	0.933	0.808	0.808	0.810	0.807	0.809	0.808	0.805
	PCR CI	1.930	0.003	0.002	0.001	0.001	0.001	0.001	0.001
	R-K	3.490	0.167	0.165	0.166	0.166	0.166	0.165	0.165
	R-K CI	4.590	0.006	0.003	0.001	0.001	0.001	0.001	0.001
	MCRR	0.808	0.157	0.158	0.161	0.161	0.161	0.160	0.159
	MCRR CI	1.930	0.003	0.002	0.001	0.001	0.001	0.001	0.001
	LRRE	1.836	0.003	0.002	0.002	0.002	0.002	0.002	0.001
	LRRE MSE	0.695	0.476	0.477	0.480	0.479	0.479	0.479	0.477
	OIRE	14.406	0.232	0.155	0.165	0.161	0.157	0.154	0.156
	OIRE MSE	6.929	0.097	0.083	0.154	0.152	0.149	0.146	0.148
	GIREI	1.000	0.935	1.027	0.467	0.372	0.325	0.298	0.245
	GIREI MSE	1.000	0.703	0.609	0.368	0.313	0.280	0.262	0.225
	GIREII	2.409	0.506	0.379	0.256	0.243	0.230	0.223	0.210
	GIREII MSE	2.421	0.458	0.333	0.211	0.201	0.191	0.186	0.178
$(0.99, 0.1)$	HK	1.000	0.595	0.530	0.518	0.515	0.515	0.514	0.515
	HKB	0.998	0.320	0.247	0.210	0.205	0.205	0.203	0.203

	LW	0.999	0.368	0.154	0.018	0.012	0.010	0.009	0.008
	ICOMP	0.851	0.067	0.052	0.057	0.055	0.054	0.053	0.052
	S	0.851	0.060	0.041	0.013	0.010	0.009	0.009	0.008
	GM	0.656	0.052	0.038	0.017	0.015	0.014	0.013	0.012
	AM	111.554	0.440	0.172	0.017	0.010	0.008	0.006	0.004
	MED	0.473	0.192	0.118	0.017	0.013	0.012	0.011	0.01
	JRR	1.000	0.629	0.515	0.446	0.438	0.438	0.436	0.436
	URR	0.996	0.180	0.148	0.140	0.137	0.137	0.135	0.135
	PCR	0.811	0.809	0.800	0.808	0.809	0.809	0.807	0.805
	PCR CI	0.028	0.018	0.019	0.018	0.018	0.018	0.018	0.018
	R-K	0.809	0.179	0.147	0.139	0.136	0.136	0.134	0.134
	R-K CI	0.028	0.022	0.019	0.015	0.015	0.015	0.015	0.015
	MCRR	0.808	0.173	0.143	0.137	0.133	0.133	0.132	0.132
	MCRR CI	0.028	0.023	0.020	0.016	0.015	0.015	0.016	0.016
	LRRE	0.032	0.006	0.006	0.007	0.006	0.006	0.006	0.005
	LRRE MSE	0.488	0.478	0.472	0.481	0.479	0.478	0.477	0.478
	OIRE	0.650	0.048	0.064	0.138	0.137	0.134	0.131	0.133
	OIRE MSE	0.515	0.031	0.054	0.132	0.131	0.129	0.126	0.128
	GIREI	1.000	0.937	1.033	0.391	0.317	0.283	0.263	0.231
	GIREI MSE	1.000	0.598	0.497	0.298	0.262	0.243	0.229	0.210
	GIREII	0.565	0.212	0.144	0.112	0.109	0.108	0.106	0.105
	GIREII MSE	0.468	0.183	0.138	0.109	0.106	0.104	0.102	0.102
$(0.9, 0.9)$	HK	1.000	0.933	0.808	0.576	0.563	0.560	0.561	0.565
	HKB	1.000	0.765	0.574	0.314	0.296	0.292	0.292	0.292
	LW	1.000	0.922	0.780	0.161	0.114	0.097	0.088	0.071
	ICOMP	0.999	0.717	0.648	0.567	0.558	0.555	0.557	0.563
	S	0.999	0.717	0.634	0.153	0.111	0.096	0.087	0.070
	GM	0.997	0.402	0.304	0.136	0.117	0.107	0.102	0.087
	AM	1.012	1.736	0.760	0.106	0.069	0.055	0.047	0.031
	MED	1.001	0.467	0.370	0.166	0.151	0.146	0.143	0.141

	JRR	1.000	0.966	0.867	0.606	0.586	0.580	0.580	0.581
	URR	1.003	0.195	0.180	0.176	0.172	0.172	0.174	0.176
	PCR	1.000	0.830	0.816	0.813	0.811	0.808	0.810	0.814
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.016	0.428	0.278	0.180	0.173	0.172	0.173	0.174
	R-K CI	1.016	0.428	0.278	0.180	0.173	0.172	0.173	0.174
	MCRR	1.003	0.193	0.176	0.172	0.168	0.169	0.171	0.171
	MCRR CI	1.003	0.195	0.180	0.176	0.172	0.172	0.174	0.176
	LRRE	1.915	0.065	0.045	0.036	0.034	0.033	0.031	0.028
	LRRE MSE	1.033	0.503	0.491	0.491	0.485	0.484	0.485	0.487
	OIRE	1.036	0.233	0.240	0.228	0.207	0.202	0.200	0.192
	OIRE MSE	1.038	0.206	0.217	0.215	0.198	0.195	0.194	0.187
	GIREI	1.000	0.926	1.037	0.465	0.371	0.329	0.306	0.259
	GIREI MSE	1.000	0.728	0.632	0.376	0.317	0.290	0.274	0.240
	GIREII	1.059	0.607	0.433	0.289	0.281	0.260	0.251	0.218
	GIREII MSE	1.027	0.544	0.431	0.302	0.273	0.260	0.252	0.223
$(0.9, 0.1)$	HK	1.000	0.949	0.847	0.574	0.565	0.559	0.560	0.557
	HKB	1.000	0.808	0.623	0.309	0.293	0.284	0.281	0.272
	LW	1.000	0.915	0.766	0.196	0.153	0.134	0.123	0.100
	ICOMP	0.999	0.601	0.501	0.368	0.364	0.359	0.358	0.353
	S	0.999	0.601	0.499	0.182	0.148	0.132	0.122	0.100
	GM	1.000	0.735	0.448	0.140	0.116	0.106	0.098	0.079
	AM	1.107	5.134	2.130	0.203	0.114	0.086	0.070	0.043
	MED	1.031	1.113	0.600	0.169	0.135	0.122	0.113	0.095
	JRR	1.000	0.974	0.882	0.568	0.550	0.539	0.536	0.526
	URR	1.000	0.551	0.422	0.213	0.207	0.202	0.200	0.193
	PCR	1.000	0.836	0.822	0.814	0.813	0.811	0.816	0.818
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.000	0.499	0.442	0.204	0.195	0.190	0.188	0.181
	R-K CI	1.000	0.507	0.443	0.204	0.195	0.190	0.188	0.181

	MCRR	1.000	0.463	0.343	0.198	0.195	0.192	0.191	0.183
	MCRR CI	1.000	0.551	0.422	0.213	0.207	0.202	0.200	0.193
	LRRE	3.627	0.098	0.076	0.070	0.068	0.065	0.063	0.056
	LRRE MSE	1.052	0.511	0.501	0.491	0.493	0.488	0.490	0.484
	OIRE	1.105	0.193	0.224	0.231	0.231	0.218	0.210	0.189
	OIRE MSE	1.105	0.164	0.193	0.218	0.221	0.210	0.203	0.184
	GIREI	1.000	0.918	1.021	0.477	0.401	0.345	0.316	0.257
	GIREI MSE	1.000	0.698	0.579	0.372	0.334	0.297	0.278	0.234
	GIREII	1.025	0.534	0.420	0.329	0.299	0.283	0.274	0.241
	GIREII MSE	1.058	0.426	0.338	0.265	0.262	0.244	0.235	0.206
$(0.7, 0.3)$	HK	1.000	0.985	0.949	0.626	0.606	0.596	0.592	0.580
	HKB	1.000	0.935	0.825	0.403	0.370	0.354	0.345	0.324
	LW	1.000	0.968	0.899	0.396	0.342	0.315	0.298	0.262
	ICOMP	1.000	0.962	0.913	0.756	0.744	0.738	0.735	0.727
	S	1.000	0.962	0.897	0.396	0.342	0.315	0.298	0.262
	GM	1.001	2.405	1.365	0.290	0.222	0.195	0.178	0.146
	AM	1.334	12.283	5.117	0.451	0.246	0.180	0.146	0.093
	MED	1.004	3.460	1.696	0.358	0.278	0.248	0.228	0.192
	JRR	1.000	0.997	0.975	0.689	0.661	0.644	0.636	0.615
	URR	1.000	0.902	0.763	0.282	0.261	0.252	0.247	0.240
	PCR	1.000	0.927	0.862	0.832	0.832	0.831	0.831	0.824
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.000	0.847	1.033	0.276	0.233	0.216	0.207	0.191
	R-K CI	1.000	0.855	1.036	0.276	0.233	0.216	0.207	0.191
	MCRR	1.000	0.724	0.574	0.258	0.246	0.239	0.235	0.226
	MCRR CI	1.000	0.902	0.763	0.282	0.261	0.252	0.247	0.240
	LRRE	1.048	0.339	0.252	0.214	0.204	0.197	0.190	0.173
	LRRE MSE	1.013	0.612	0.551	0.520	0.515	0.509	0.506	0.490
	OIRE	1.015	5.972	2.519	0.417	0.335	0.297	0.277	0.227
	OIRE MSE	1.015	0.779	0.495	0.316	0.298	0.278	0.264	0.222

GIREI	1.009	0.954	0.941	0.452	0.396	0.350	0.322	0.257
GIREI MSE	1.005	0.623	0.581	0.400	0.361	0.325	0.301	0.244
GIREII	1.011	1.023	0.641	0.419	0.375	0.342	0.321	0.261
GIREII MSE	1.011	0.690	0.486	0.393	0.359	0.329	0.310	0.252

Table 3.8.5: Simulation Results for Population Size 50

3.8.4 Appendix 4: Simulation Results for Population n=100

Eigenvalues of X'X and the β values for n=100					
Correlation Structure	(0.99,0.99)	(0.99,0.1)	(0.9,0.9)	(0.9,0.1)	(0.7,0.3)
λ_1	4.912	3.082	4.382	2.665	2.360
λ_2	0.027	1.002	0.202	1.014	0.898
λ_3	0.023	0.883	0.170	0.972	0.820
λ_4	0.020	0.017	0.137	0.193	0.531
λ_5	0.017	0.016	0.108	0.156	0.390
β_1	-0.448	-0.562	-0.446	0.568	-0.529
β_2	-0.447	-0.561	-0.449	0.579	-0.528
β_3	-0.447	-0.561	-0.452	0.575	-0.500
β_4	-0.447	-0.167	-0.444	0.106	-0.294
β_5	-0.447	-0.165	-0.444	0.005	-0.323

Table 3.8.6: Eigenvalues and Beta Values for Population Size 100

Simulation Results for n=100									
		σ^2							
(ρ_1, ρ_2)	Estimator	0.0001	0.25	1	25	50	75	100	400
$(0.99, 0.99)$	HK	1.000	0.778	0.624	0.557	0.557	0.555	0.555	0.556
	HKB	0.999	0.521	0.391	0.284	0.282	0.279	0.279	0.279
	LW	1.000	0.771	0.492	0.034	0.016	0.010	0.008	0.004
	ICOMP	0.977	0.502	0.479	0.404	0.405	0.400	0.402	0.400
	S	0.977	0.493	0.343	0.026	0.013	0.009	0.007	0.004
	GM	0.822	0.239	0.200	0.093	0.083	0.076	0.071	0.055
	AM	221.390	0.326	0.170	0.031	0.019	0.014	0.011	0.006
	MED	2.150	0.311	0.289	0.140	0.139	0.138	0.137	0.137
	JRR	1.000	0.844	0.715	0.574	0.571	0.568	0.567	0.567
	URR	0.571	0.155	0.158	0.154	0.159	0.157	0.158	0.159
	PCR	1.038	0.806	0.808	0.799	0.802	0.803	0.802	0.807
	PCR CI	0.796	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	R-K	13.926	0.166	0.161	0.155	0.159	0.157	0.158	0.159
	R-K CI	14.160	0.013	0.005	0.001	0.001	0.001	0.001	0.001
	MCRR	0.579	0.150	0.153	0.149	0.154	0.152	0.153	0.154
	MCRR CI	0.796	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	LRRE	0.763	0.002	0.002	0.002	0.002	0.002	0.002	0.001
	LRRE MSE	0.715	0.471	0.474	0.471	0.473	0.472	0.472	0.474
	OIRE	3.137	0.602	0.224	0.172	0.160	0.151	0.150	0.146
	OIRE MSE	2.811	0.116	0.088	0.152	0.147	0.142	0.142	0.138
	GIREI	1.000	0.897	0.937	0.640	0.512	0.435	0.388	0.283
	GIREI MSE	1.000	0.784	0.659	0.459	0.394	0.348	0.320	0.248
	GIREII	1.672	0.546	0.480	0.254	0.248	0.237	0.232	0.210
	GIREII MSE	1.693	0.527	0.431	0.207	0.203	0.194	0.191	0.175
$(0.99, 0.1)$	HK	1.000	0.788	0.617	0.530	0.524	0.522	0.521	0.519
	HKB	0.999	0.497	0.341	0.221	0.216	0.213	0.210	0.207

	LW	1.000	0.687	0.389	0.038	0.024	0.019	0.016	0.012
	ICOMP	0.958	0.092	0.076	0.062	0.065	0.065	0.061	0.057
	S	0.958	0.092	0.070	0.023	0.018	0.016	0.015	0.012
	GM	0.895	0.090	0.061	0.026	0.023	0.021	0.020	0.016
	AM	304.182	1.244	0.486	0.041	0.023	0.017	0.013	0.008
	MED	1.282	0.244	0.192	0.030	0.021	0.018	0.017	0.014
	JRR	1.000	0.824	0.645	0.462	0.453	0.448	0.446	0.441
	URR	0.998	0.212	0.167	0.140	0.139	0.137	0.136	0.133
	PCR	1.056	0.809	0.806	0.808	0.805	0.809	0.810	0.811
	PCR CI	2.495	0.025	0.024	0.024	0.024	0.024	0.023	0.023
	R-K	1.053	0.220	0.170	0.139	0.139	0.137	0.135	0.132
	R-K CI	2.495	0.049	0.031	0.020	0.019	0.019	0.018	0.018
	MCRR	1.054	0.196	0.158	0.135	0.135	0.134	0.132	0.129
	MCRR CI	2.495	0.041	0.028	0.020	0.019	0.019	0.019	0.019
	LRRE	2.427	0.010	0.008	0.008	0.008	0.008	0.008	0.007
	LRRE MSE	0.893	0.481	0.478	0.483	0.480	0.480	0.479	0.477
	OIRE	2268.456	0.935	0.225	0.164	0.155	0.148	0.141	0.131
	OIRE MSE	11.364	0.039	0.037	0.144	0.145	0.140	0.135	0.126
	GIREI	1.000	0.877	0.936	0.579	0.465	0.397	0.355	0.266
	GIREI MSE	1.000	0.708	0.537	0.391	0.341	0.306	0.285	0.231
	GIREII	2.804	0.349	0.232	0.124	0.129	0.125	0.120	0.114
	GIREII MSE	2.838	0.250	0.198	0.120	0.123	0.119	0.115	0.109
$(0.9, 0.9)$	HK	1.000	0.969	0.902	0.577	0.563	0.561	0.559	0.557
	HKB	1.000	0.871	0.705	0.330	0.307	0.300	0.296	0.288
	LW	1.000	0.965	0.890	0.233	0.160	0.127	0.109	0.073
	ICOMP	1.000	0.757	0.671	0.548	0.535	0.533	0.532	0.530
	S	1.000	0.757	0.671	0.215	0.152	0.123	0.106	0.071
	GM	1.028	0.455	0.347	0.166	0.140	0.131	0.124	0.098
	AM	2063.271	3.869	1.492	0.176	0.110	0.085	0.072	0.043
	MED	0.999	0.537	0.403	0.207	0.175	0.165	0.160	0.148

	JRR	1.000	0.991	0.940	0.627	0.600	0.592	0.586	0.576
	URR	0.998	0.207	0.180	0.164	0.164	0.165	0.165	0.164
	PCR	1.045	0.834	0.815	0.808	0.805	0.806	0.805	0.804
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.041	0.610	0.349	0.175	0.168	0.167	0.166	0.163
	R-K CI	0.998	0.610	0.349	0.175	0.168	0.167	0.166	0.163
	MCRR	1.039	0.202	0.174	0.159	0.159	0.161	0.160	0.159
	MCRR CI	0.998	0.207	0.180	0.164	0.164	0.165	0.165	0.164
	LRRE	2.139	0.071	0.042	0.026	0.026	0.026	0.025	0.022
	LRRE MSE	0.992	0.505	0.487	0.478	0.477	0.478	0.477	0.475
	OIRE	4.504	0.693	0.296	0.252	0.224	0.214	0.205	0.185
	OIRE MSE	4.544	0.227	0.176	0.205	0.198	0.197	0.191	0.176
	GIREI	1.000	0.910	0.894	0.689	0.541	0.469	0.417	0.306
	GIREI MSE	1.000	0.788	0.655	0.482	0.409	0.373	0.341	0.266
	GIREII	2.569	0.610	0.472	0.272	0.291	0.295	0.291	0.248
	GIREII MSE	2.253	0.612	0.484	0.255	0.267	0.268	0.264	0.225
$(0.9, 0.1)$	HK	1.000	0.977	0.925	0.591	0.575	0.569	0.568	0.563
	HKB	1.000	0.900	0.754	0.343	0.317	0.304	0.300	0.286
	LW	1.000	0.958	0.871	0.278	0.218	0.189	0.173	0.137
	ICOMP	1.000	0.760	0.634	0.426	0.417	0.411	0.410	0.404
	S	1.000	0.760	0.634	0.263	0.212	0.186	0.171	0.137
	GM	1.124	1.513	0.806	0.197	0.155	0.138	0.127	0.099
	AM	5334.459	13.833	5.059	0.407	0.211	0.150	0.118	0.068
	MED	1.009	2.549	1.159	0.254	0.193	0.167	0.152	0.121
	JRR	1.000	0.994	0.954	0.610	0.582	0.568	0.563	0.546
	URR	1.000	0.776	0.592	0.245	0.227	0.217	0.214	0.206
	PCR	1.022	0.843	0.827	0.811	0.813	0.816	0.818	0.816
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.022	0.682	0.524	0.221	0.204	0.194	0.191	0.182
	R-K CI	1.000	0.756	0.545	0.222	0.204	0.195	0.191	0.182

	MCRR	1.022	0.667	0.508	0.224	0.212	0.205	0.203	0.195
	MCRR CI	1.000	0.776	0.592	0.245	0.227	0.217	0.214	0.206
	LRRE	1.272	0.146	0.105	0.092	0.090	0.087	0.084	0.074
	LRRE MSE	1.023	0.516	0.501	0.491	0.494	0.493	0.494	0.486
	OIRE	2.248	0.269	0.301	0.255	0.254	0.243	0.238	0.209
	OIRE MSE	2.236	0.206	0.241	0.234	0.240	0.233	0.230	0.203
	GIREI	1.000	0.984	1.028	0.562	0.463	0.397	0.361	0.279
	GIREI MSE	1.000	0.760	0.616	0.423	0.377	0.335	0.313	0.252
	GIREII	1.790	0.923	0.631	0.325	0.313	0.299	0.292	0.246
	GIREII MSE	1.670	0.565	0.420	0.296	0.293	0.282	0.276	0.234
$(0.7, 0.3)$	HK	1.000	0.991	0.971	0.654	0.616	0.600	0.591	0.577
	HKB	1.000	0.960	0.884	0.440	0.389	0.367	0.353	0.324
	LW	1.000	0.982	0.942	0.466	0.384	0.346	0.322	0.265
	ICOMP	1.000	0.971	0.928	0.750	0.731	0.722	0.717	0.707
	S	1.000	0.971	0.927	0.465	0.384	0.346	0.322	0.265
	GM	1.002	3.625	2.049	0.378	0.274	0.233	0.209	0.156
	AM	1.150	23.962	10.139	0.820	0.419	0.291	0.229	0.128
	MED	1.006	6.187	2.910	0.470	0.339	0.288	0.262	0.205
	JRR	1.000	0.999	0.990	0.720	0.673	0.652	0.639	0.610
	URR	1.000	0.811	0.539	0.235	0.219	0.214	0.213	0.208
	PCR	1.000	0.926	0.859	0.821	0.828	0.830	0.829	0.825
	PCR CI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	R-K	1.000	2.839	2.482	0.387	0.281	0.243	0.225	0.193
	R-K CI	1.000	2.840	2.482	0.387	0.281	0.243	0.225	0.193
	MCRR	1.000	0.774	0.487	0.227	0.216	0.213	0.212	0.205
	MCRR CI	1.000	0.811	0.539	0.235	0.219	0.214	0.213	0.208
	LRRE	1.031	0.351	0.251	0.207	0.199	0.192	0.186	0.164
	LRRE MSE	1.008	0.616	0.548	0.523	0.520	0.515	0.511	0.495
	OIRE	1.008	5.893	3.412	0.475	0.369	0.331	0.310	0.250
	OIRE MSE	1.008	1.256	0.716	0.332	0.317	0.303	0.292	0.241

GIREI	1.000	0.948	1.068	0.590	0.488	0.437	0.398	0.300
GIREI MSE	1.000	0.737	0.641	0.480	0.421	0.389	0.361	0.278
GIREII	1.006	1.283	0.816	0.471	0.433	0.404	0.381	0.296
GIREII MSE	1.006	0.995	0.646	0.437	0.409	0.384	0.363	0.284

Table 3.8.7: Simulation Results for Population Size 100

Chapter 4

Relative Value Dynamic Nelson and Siegel Model

4.1 Abstract

This paper asks whether the Dynamic Nelson Siegel model can be used as the basis of a relative value trading model. It introduces a relative value trading system based on the Dynamic Nelson-Siegel term structure model and illustrates its use on some sample data. The model seeks to identify relative value opportunities in the steepness and curvature of the yield curve being tested. The model was tested on empirical and simulated versions of the US Treasury zero coupon curve: it was found to be profitable on both data sets. Although based on the Dynamic Nelson and Siegel model it utilizes different outputs from the Kalman filtering process to produce a trading signal and a threshold value for the trading signal that determines where the trades should be entered and exited and as such it represents a new contribution to the literature on Dynamic Nelson Siegel Models.

4.2 Introduction

The results of the simulation study in chapter three along with the results of chapter two meant that the two key calibration issues had now been addressed, and the rv-NSS could now be implemented effectively.

However, as part of my ongoing research, I'd started looking into Dynamic Nelson and Siegel models (DNS) which were first introduced by Diebold and Li (2006), this led me to think that a relative value model based on the DNS would have some distinct advantages, compared to the rv-NSS.

Firstly the rv-NSS only produces estimates of the parameter values at each observation point, thus the modelling of the time series dynamics must be done externally. The DNS model incorporates modelling of the time series of the parameter estimates meaning that all the parameters can be estimated simultaneously and only one round of calibration is necessary.

Secondly the DNS utilizes a Kalman filter to estimate its log-likelihood. This can help alleviate the multicollinearity problems as the filter knows what the previous estimate was and uses this as a basis to calculate the new parameter estimates making wild changes in the parameter estimates less likely.

Thirdly the standard rv-NSS does not account for the possibility of regime changes and would rely on the analyst observing that a regime change had taken place. This would then mean a recalibration of the model. By basing the rv-DNS on one of the regime switching versions mentioned in section 1.2.3, such regime shifts could be handled internally by the rv-DNS.

Fourthly as also mentioned in section 1.2.3 the DNS can be made arbitrage free, this structure could also be applied to the rv-DNS. In contrast the rv-NSS does not consider whether its curves are arbitrage free.

So it was decided that chapter 4 would focus on developing and testing an rv-DNS, for the first attempt it was decided to keep the model as parsimonious as possible so it was decided not to use the Markov switching or arbitrage free machinery mentioned above. However once the concept is validated these generalizations could likely be added.

To adapt the model logic of the rv-NSS to a dynamic version some changes had to be made. Originally it was planned to use the dynamic version of the NSS model (DNSS) but as shown by Christensen et al. (2009), it is not possible to have an arbitrage free version of the DNSS. Arbitrage freeness can only be re-established with the addition of another basis function. This meant that if the dynamic relative value model could be made arbitrage free, then either the five factor version introduced by Christensen et al. (2009) must be considered or the three factor DNS version must be considered. To reduce the risk of over-fitting and to keep the model as parsimonious as possible, the three factor version was chosen as the basis for the relative value model. Also as explained previously, the rv-NSS model is designed to match some key curve points, even with three factors it should be possible to match three key points, sufficient for steepener and curvature strategies. It was hoped that this could be done in the rv-DNS framework by adapting the DNS model to use constrained Kalman filtering (see for example Simon (2009)). Ultimately this proved fruitless as, when the number of loading parameters being estimated matches the number of restrictions, essentially no filtering takes place and the model breaks down. So the key point matching had to be dropped.

However the Kalman filter can be used to produce a prediction of where the model thinks the curve should be based on the previous estimate from the filtering process. This can then be used to calculate the model implied levels of steepness and curvature. By subtracting these from the market observed levels of steepness and curvature, we get a spread that quantifies the richness or

cheapness of the strategies relative to the market implied ones. Thus when the spread is suitably low a long position should be established and vice versa. To determine what constitutes a suitably low level, some measure of the predicted standard deviation of the observed error is required. Happily the Kalman filter provides this too. This prior predicted standard deviation can be used to calculate the standard deviation of the candidate strategies. The idea of using the Kalman filter outputs in this way was based on an idea presented in Chan (2013) who used a Kalman filter as part of a pairs trading strategy. Thus although the rv-DNS is based on the DNS it differs in the outputs it uses from the Kalman filtering process and the way in which it utilizes those outputs. Chapter four introduces the rv-DNS and tests it on simulated and empirical data.

4.3 Literature Review

In order to develop and evaluate relative value strategies in fixed income it is first necessary to have a good term structure model. As noted in Niu and Zeng (2012) term structure models can be thought of as belonging to two classes:

- Financial theory based affine arbitrage-free class
- Statistically based reduced form class

The general framework for the first class was developed by Duffie and Kan (1996) and Dai and Singleton (2000). Specific examples of these types of model include: Vasicek (1977), Cox et al. (1985) and Hull and White (1990) even the Heath et al. (1992) can be shown to be a member of this class. The arbitrage free nature of these models is appealing both on theoretical grounds and to practitioners when pricing derivative securities. However the main disadvantage is that the forecasting ability of these models has been shown to be no better than that of a random walk, Duffee (2002).

The model proposed by Nelson and Siegel (1987) (NS) and its extension by Svensson (1994) termed the Nelson Siegel and Svensson model (NSS), is a well known example of the second class of models which is popular with policy makers with nine out of thirteen central banks reportedly using the Nelson and Siegel and/or the Svensson variant according to report published by the Bank of International Settlements (BIS) BIS (2005).

The original NS model consists of three parametric functions. As noted by Diebold and Li (2006), each parametric function is similar to one of the level, slope and curvature factors found in statistical factor analyses of the yield curve. This gives the NS model intuitive appeal.

The weighted combination of these parametric functions are sufficiently flexible to allow most commonly encountered yield curve shapes to be modelled. The extension by Svensson (1994) simply adds another parametric function (a second curvature factor), thus allowing an even richer array of term structures to be modelled. The primary advantages of the model are that it is parsimonious, and as will be discussed can be used as the basis for good forecasts. One of the main criticisms of these models however has been that they do not explicitly enforce no arbitrage, although as shown by Coroneo et al. (2011) the parameters of no-arbitrage models are not statistically different from those obtained from the NS model so it can be regarded as being arbitrage free at least in the case of the US market data that they tested.

There has been renewed interest in the NS type models which was precipitated by Diebold and Li (2006) who introduced the Dynamic Nelson-Siegel (DNS) model which can be used for forecasting. The DNS extends the NS model by recasting it into a dynamic framework where the loadings applied to each of the NS parametric functions are given AR(1) dynamics. They employed a two step estimation procedure based on ordinary least squares estimation with a fixed lambda value, which was chosen beforehand to represent the point where maximum curvature was thought most likely to occur. Diebold et al. (2006)

reformulated this model into a state space framework. This allowed the parameters to be estimated in a one step process using the Kalman filter to estimate the log-likelihood of the state space model, and this log-likelihood was then maximized using derivative based optimization techniques to find the maximum log-likelihood estimates of the parameters. In this new one step estimation procedure it was also possible to estimate lambda from the data. Further their paper showed that the model has good forecasting properties in the longer term. De Pooter (2007) also showed that the DNS (and some generalizations thereof) have good forecasting properties.

The question over the arbitrage freeness of the DNS was fully addressed by Christensen et al. (2011) they combined the affine arbitrage free class of models with the DNS model to produce an arbitrage free Dynamic Nelson-Siegel model (AFDNS). This was done initially with continuous time dynamics and later extended to discrete dynamics by Niu and Zeng (2012). The AFDNS retains the level, slope and curvature factors of the NS/DNS but adds an additional factor that ensures that the model is arbitrage free. As shown in Christensen et al. (2011) the differences between the DNS and AFDNS models are most apparent for the maturities in excess of ten years. In addition Christensen et al. (2009) showed that the dynamic Svensson model cannot be re-cast into an arbitrage free form, however by adding a fifth parametric function (an additional slope factor) they were able to produce a dynamic Svensson like model that is arbitrage free.

Xiang and Zhu (2013) and Levant and Ma (2013) criticize the assumption that the lambda parameter is constant in DNS models, as empirical evidence suggests that lambda may vary over time in response to changes in the business cycle, monetary policy and times of market stress. Although lambda is estimated in the Diebold et al. (2006) estimation procedure it is still assumed to be constant. Two different methods to address this problem have been developed. Firstly Koopman et al. (2010) extends Diebold et al. (2006) so that lambda is also a

time varying parameter as well as the loading parameters. A second approach has been to generalize the DNS model into a Markov switching framework. Several papers have gone in this direction, Levant and Ma (2013) looks at two models, one with different lambda parameters and one with different volatility parameters in each regime. They use a Kalman filter to estimate the parameters and a Kim filter to estimate the regimes. Xiang and Zhu (2013) introduce a regime switching model which can have up to three volatility regimes, although they find that two regimes seem to work best. They estimate their model using Reverse Jump Markov Chain Monte Carlo. They also find that their model offers superior out of sample forecasting results as compared to the original DNS model. Hevia et al. (2014) and Bandara and Munclinger (2012) introduce arbitrage free Markov Switching versions of the DNS model, Hevia et al. (2014) also show that its forecasting performance is superior to that of the original single regime DNS model. Lastly Christensen (2013) uses a Markov switching model to handle the very current problem of rates being close to zero by having one of the switching states be a zero bound state.

This paper introduces a relative value dynamic Nelson-Siegel model (rv-DNS). The relative value strategies belong to one of two classes:

- Steepener
- Curvature

The original DNS model was used as the basis for the relative value model, also the same data set and optimization method were employed to keep the new model as consistent with the literature as possible. The simpler AR(1) version rather than the VAR(1) version has been used because as shown in Diebold et al. (2006) it does not make a big difference to the results, at least on this data set, and moreover keeps the model as parsimonious as possible. It was decided not to use the no-arbitrage version because as noted before, the arbitrage free and

the original version produce results that are statistically indistinguishable from each other, but the original version does it in a more parsimonious way. Also the main focus of this study will be on maturities up to ten years but as previously mentioned the biggest differences between the DNS and AFDNS only become apparent after the ten year point. Lastly making the model arbitrage free may remove or at least attenuate some of the very relative value opportunities the model is trying to capture.

As will be seen, although based on the DNS, the rv-DNS differs in the outputs it takes from the Kalman filtering process and the way those outputs are used. Rather than taking the fully filtered posterior estimates and trying to predict the whole yield curve, it takes the partially filtered estimates based on the previous day's data and looks for deviations of the steepener and curvature strategies away from their mean value.

4.4 The Underlying Models

4.4.1 The Nelson-Siegel Model

The NS model can be defined mathematically as follows

$$y(\tau) = \beta 1 + \beta 2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta 3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{4.4.1}$$

The first factor is simply a constant, and can be interpreted as the long term level of interest rates, the second can be interpreted as the curve's steepness while the last can be interpreted as the curvature of the yield curve. The similarity of these factors to the empirical factor loading found by Bliss (1997) and Litterman and Scheinkman (1991) was noted in Diebold and Li (2006).

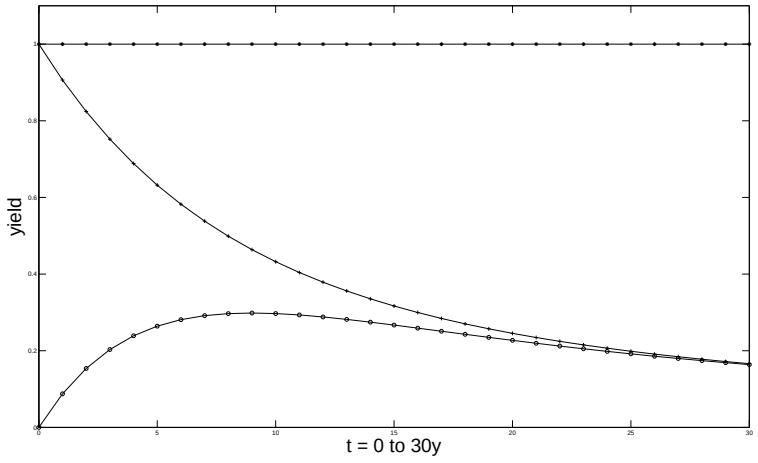


Figure 4.4.1: The Nelson-Siegel Basis Functions

Further intuition on the role of the different factors can be gained when we consider the behaviors of their parametric functions in the limit. The constant 1 is independent of time and so as noted before can be used to represent the

long term level of rates. The second factor starts off at one when $\tau = 0$ and decays rapidly as τ increases, so it represents a short term factor. The third factor starts off at zero and then rises to a maximum value as τ is increased, it then falls back to zero as τ continues to increase. The value of λ in the third factor determines the location of its maximum value.

4.4.2 The Dynamic Nelson Siegel Model

Following broadly the exposition given in Diebold et al. (2006), who recast the DNS model in state space form and as stated in the introduction using the assumption that the loading parameters follow an AR(1) processes, so that:

$$y(\tau)_t = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{4.4.2}$$

So the loading factors are now time varying. As in Diebold et al. (2006) L_t, S_t and C_t are the time varying Level, Slope and Curvature parameters, which represents the value of the driving AR(1) processes at each point in time.

State Space Representation

The measurement equation is:

$$\begin{bmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-exp^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-exp^{-\lambda\tau_1}}{\lambda\tau_1} - exp^{-\lambda\tau_1} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-exp^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-exp^{-\lambda\tau_N}}{\lambda\tau_N} - exp^{-\lambda\tau_N} \end{bmatrix} \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} \epsilon_t(\tau_1) \\ \vdots \\ \epsilon_t(\tau_N) \end{bmatrix} \tag{4.4.3}$$

This can be expressed in matrix form as:

$$\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{E}_t \tag{4.4.4}$$

Where $\mathbf{E}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{H})$ and $\mathbf{H}$ is a diagonal matrix with identical entries on the diagonal.

The transition equation is:

$$\begin{bmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{bmatrix} + \begin{bmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{bmatrix} \tag{4.4.5}$$

Which can also be expressed in matrix form as:

$$\mathbf{(X_t - \mu) = A(X_{t-1} - \mu) + \eta_t} \tag{4.4.6}$$

Where $\eta_t \sim \mathbf{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{Q}$ is a diagonal matrix with differing values on the diagonal.

In order for the Kalman filter (which is introduced in the next section) the measurement and transition errors must be orthogonal to each other and to the initial state.

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim WN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H & 0 \\ 0 & Q \end{pmatrix} \right]$$

$$\mathbb{E}(\mathbf{X}_0 \epsilon') = \mathbf{0}$$
$$\mathbb{E}(\mathbf{X}_0 \eta') = \mathbf{0}$$

4.4.3 Kalman Filter

The Kalman filter has two stages

- Prior step where the model parameters are estimated using the information available up to the previous time step
- Update step where information from the latest observation is incorporated into the model

Prior Step

The transition equation can be expressed as:

$$\mathbf{X_t = AX_{t-1} + (I - A)\boldsymbol{\mu} + \boldsymbol{\eta_t}} \tag{4.4.7}$$

Thus the predicted state estimate is

$$\mathbf{X_{t|t-1}} = \mathbb{E}[\mathbf{X_t}]_{\mathbf{t-1}} = \mathbf{AX_{t-1}} + \mathbf{(I - A)}\boldsymbol{\mu} \tag{4.4.8}$$

The predicted covariance estimate is

$$\mathbf{P}_{t|t-1} = \mathbb{E}[(\mathbf{X}_t - \mathbf{X}_{t|t-1})((\mathbf{X}_t - \mathbf{X}_{t|t-1})']_{t-1} = \mathbf{AP}_{t-1}\mathbf{A}' + \mathbf{Q} \tag{4.4.9}$$

Update Step

Measurement Residual

$$\mathbf{e}_t = \mathbf{Y}_t - \mathbf{BX}_{t|t-1} \tag{4.4.10}$$

Measurement Covariance

$$\mathbf{S}_t = Var(\mathbf{e}_t) = \mathbf{BP}_{t|t-1}\mathbf{B}' + \mathbf{H} \tag{4.4.11}$$

Optimal Kalman Gain

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{B}'\mathbf{S}_t^{-1} \tag{4.4.12}$$

Posteriori State Estimate

$$\mathbf{X}_{t|t} = \mathbf{X}_{t|t-1} + \mathbf{K}_t\mathbf{e}_t \tag{4.4.13}$$

Posteriori Covariance Estimate

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{B})\mathbf{P}_{t|t-1} \tag{4.4.14}$$

4.5 Relative Value Dynamic Nelson Siegel Model

4.5.1 Relative Value Dynamic Nelson Siegel Model Logic

The rv-DNS like the DNS uses the Kalman filter to estimate the unobservable system parameters, however the rv-DNS utilizes the outputs from the Kalman filter in a different manner to create the signals used for the trading system.

The measurement error equation 4.4.10 represents the deviation of the observed yield curve from the prior model estimate which is based on the previous Kalman filter estimate. From this the corresponding relative value measurement errors from the Steepener and Curvature strategy can be calculated as shown below. This can then be subtracted from the empirically observed steepness and curvature to give the spread between them which tells us how rich or cheap the relative value strategy is relative to the model's estimate. To determine what constitutes a significant spread the measurement covariance eqn 4.4.11, which also is based on the previous Kalman filter estimate, is used. The measurement covariance gives the covariance of the error between the empirical and model yield curves. This can be used to calculate the corresponding spread covariance also shown below. Taking the square root of the spread covariance gives the spread standard deviation which is used as the threshold value to decide when to take a position. The idea is simply to buy when the spread is below the negative of the threshold value and to take an offsetting position when the spread moves back above the positive value of the threshold and vice versa for short positions. E.g. if the spread between the empirical and model curve steepness is found to be higher than than the threshold value a flattener position would

be established with an offsetting position taken when the spread between the empirical and model steepness moves back below the negative of the threshold value. The idea to use the Kalman filter outputs in this way was based on an idea presented in Chan (2013) who used it as part of a pairs trading strategy.

4.5.2 rv-DNS signal

The following section presents these ideas mathematically. Analogously to Diebold and Li (2006) the Steepener position is taken to be the difference between the longest and shortest maturities. Similarly the Curvature position is taken to be twice the mid maturity rate minus the shortest and longest rates.

The model estimate of the curve based on the priori information

$$\mathbf{Y_{pred}}_t = \mathbf{B}\mathbf{X}_{t|t-1} \tag{4.5.1}$$

The relative value strategies are then calculated for the estimated and actual yield curves

Steepener

$$\mathrm{St}_t = Y(120) - Y(3) \tag{4.5.2}$$

Curvature

$$\mathrm{Crv} = 2Y(24) - Y(120) - Y(3) \tag{4.5.3}$$

The signal is then simply the difference between the observed and predicted steepener and curvature positions.

$$\mathrm{P}_{signal} = \mathrm{P}_{act} - \mathrm{P}_{pred} \tag{4.5.4}$$

Where P (Position) is used generically to represent either the steepener or curvature trading position.

To calculate the threshold value, the well known formula for calculating the variances of linear combinations of random variables is applied to the measurement covariance equation 4.4.11.

Variance Steepener

Applying the well known formula for the variance of two variables:

$$\mathrm{Var(St)} = \mathrm{Var}(Y(120)) + \mathrm{Var}(Y(3)) - 2\mathrm{Cov}(Y(3), Y(120)) \tag{4.5.5}$$

Variance Curvature

Similarly for three variables

$$\begin{aligned}\mathrm{Var(Crv)} = 4\mathrm{Var}(Y(24)) + \mathrm{Var}(Y(3)) + \mathrm{Var}(Y(120)) - 4\mathrm{Cov}(Y(3), Y(24)) \\ -4\mathrm{Cov}(Y(120), Y(24)) + 2\mathrm{Cov}(Y(3), Y(120))\end{aligned} \tag{4.5.6}$$

4.5.3 Relative Value Dynamic Nelson Siegel Trading Rules

The main assumptions underlying the model are:

- The curve can be accurately modelled from the Nelson-Siegel basis functions
- These functions are accurately calibrated to capture the relationships between different points on the curve
- The AR(1) structure accurately captures the dynamic relationship between the factors
- The Kalman filtered estimates are optimal, i.e. the data satisfies the linear/Gaussian assumptions of the model
- The model assumes that any (large) observed deviations from the predicted curve values will revert at a later time point

The model trading rules for short positions are:

- If the signal is $>$ standard deviation of the strategy error then a short position is put on
- Each time the signal is $>$ standard deviation of the strategy error, an additional position is put on
- All open positions are closed the first time that signal $<$ -standard deviation of the strategy error by taking an offsetting long position

and analogously

The model trading rules for long positions are:

- If the signal is $<$ -standard deviation of the strategy error then a long position is put on
- Each time the signal is $<$ -standard deviation of the strategy error, an additional position is put on
- All open positions are closed the first time that signal $>$ standard deviation of the strategy error by taking an offsetting short position

4.6 Testing Methodology

The model will be tested via means of a simulation study and an out of sample test conducted on empirical data.

To test the model out of sample a walk-forward-analysis (WFA) which was first introduced by Pardo (2008) was conducted. Rather than just having one optimization and out of sample test set WFA breaks the data set into a number of subperiods. These subperiods allow the model to be repeatedly optimized and tested out of sample, the procedure is illustrated in fig: 4.6.1, which was taken from Wiecki (2012). The advantages of this form of testing are that it exposes the model to a variety of different market conditions and assesses its out of sample performance in those differing market conditions. Also by testing the model repeatedly out of sample it gives a more robust test compared to traditional out of sample testing where typically there is only one out of sample testing period, thus the WFA results are less likely to be the result of chance. The WFA represents a stern test of any trading strategy. When deciding how to split the data there are two conflicting goals: as many walk forward tests as possible should be included, so the data should be split as small as possible, but each subset must be large enough to produce robust parameter estimates and sufficient trades for analysis. It was found that splitting the data into six subperiods (called periods 1-6) of fifty eight months each best satisfied these

conflicting requirements. So the model would be calibrated on periods 1 and 2 then tested on period 3. It would then be walked forward calibrated on periods 2 and 3 and tested on period 4 and so on.

In the simulation study 1,000 realizations of the data were generated using the state space model shown in section 4.4.2 using the results from the optimization step see section 4.9. The parameters used for the simulation study are different to those used in the out of sample testing discussed above. The parameters used in the simulation study were produced from calibration to periods 3, 4, 5 and 6 individually. These parameters were then used to simulate those same periods, and the strategies were then run with the same parameters used in the simulation, as such they can be thought of as in sample. The simulation study will test that the model is working as intended and will also serve as a benchmark for the out of sample results.

In addition it was decided to estimate the value of lambda in each of the calibration stages of the simulation study and the WFA, as this would allow its stability across the periods to be observed. The calibrated lambda value will still remain constant at the value from its corresponding calibration stage during each testing period of the WFA.

To illustrate the model function a time series of the rv-DNS trading signal against the threshold value will be presented, this should show whether it is functioning as intended. In addition the profit and loss (PnL) will be calculated in basis points (bp) according to the rules in the previous section. The base test is that the strategies should be profitable and a break-down of the PnLs will also be provided to give a fuller picture of how the strategies are performing including the Sharpe Ratios (SR) to assess the strategies' risk adjusted performance. The testing was performed in Matlab and included the use of the Kalman Filter toolbox by Murphy (2004).

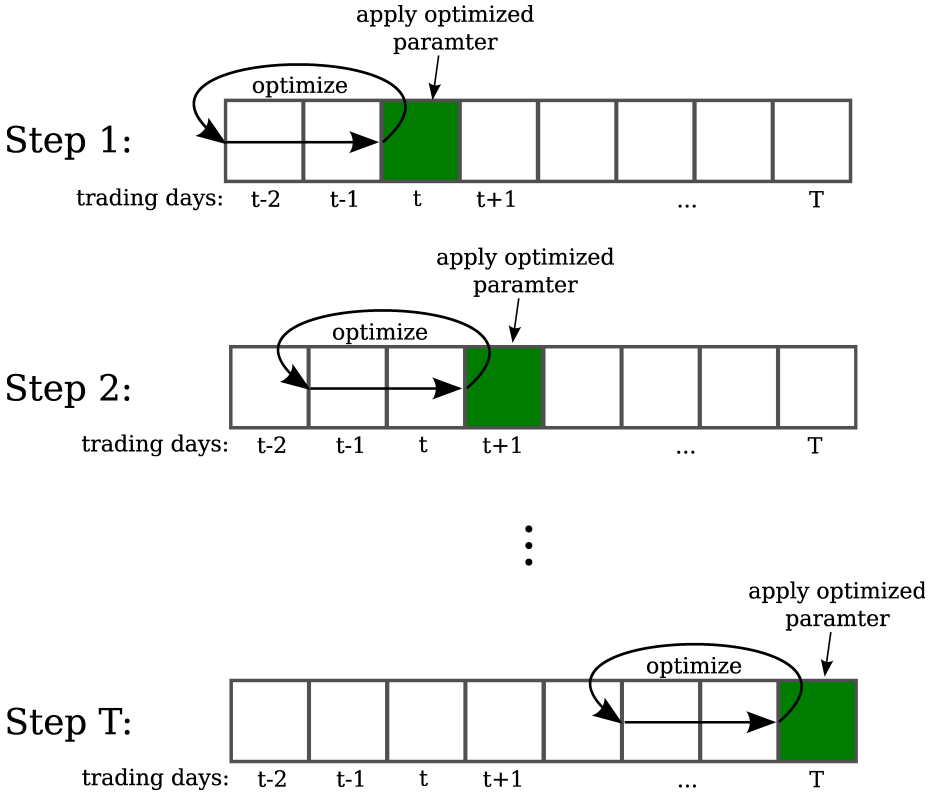


Figure 4.6.1: WFA methodology

4.7 Empirical Data

The data set is virtually the same as that used in Diebold et al. (2006), (an additional month was added to facilitate the WFA). The data set consists of US treasury yields for the period from January 1972 until January 2001, for the maturities 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. They were derived from the bid/ask average quotes observed at the end of each month using the Fama and Bliss (1987) approach. The data is shown in fig 4.7.1.

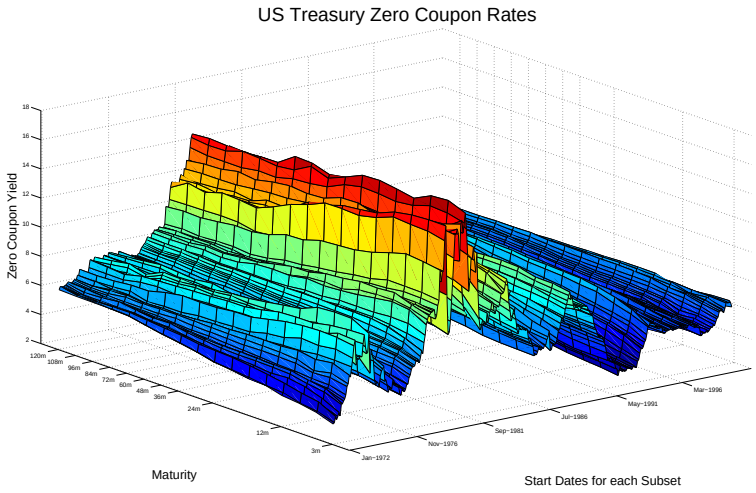


Figure 4.7.1: US Zero Coupon Curve 1972-2000

In accordance with the WFA methodology the data was then subdivided into 6 periods: Period 1=(Jan72-Nov76), Period 2=(Dec76-Sep81), Period 3=(Oct81-Jul86), Period 4=(Aug86-May91), Period 5=(Jun91-Mar96) and Period 6=(Apr96-Jan01).

US Treasury Zero Coupon Yields Descriptive Statistics							
Maturity (months)	Mean	Stdev	Max	Min	$\rho(1)$	$\rho(12)$	$\rho(24)$
3	6.85	2.70	16.02	2.73	0.970	0.700	0.417
6	7.08	2.71	16.48	2.89	0.972	0.719	0.449
9	7.20	2.68	16.39	2.98	0.972	0.726	0.466
12	7.30	2.61	15.82	3.11	0.971	0.729	0.480
15	7.41	2.55	16.04	3.29	0.973	0.737	0.499
18	7.48	2.54	16.23	3.48	0.974	0.744	0.511
21	7.54	2.52	16.18	3.64	0.975	0.747	0.521
24	7.56	2.48	15.65	3.78	0.975	0.745	0.527
30	7.65	2.40	15.40	4.04	0.975	0.755	0.545
36	7.72	2.38	15.77	4.20	0.977	0.761	0.557
48	7.86	2.32	15.82	4.31	0.977	0.765	0.570
60	7.93	2.29	15.01	4.35	0.980	0.779	0.589
72	8.05	2.26	14.98	4.38	0.980	0.786	0.600
84	8.08	2.22	14.98	4.35	0.980	0.768	0.589
96	8.14	2.20	14.94	4.43	0.982	0.793	0.616
108	8.18	2.21	15.02	4.43	0.982	0.794	0.616
120 (level)	8.14	2.17	14.93	4.44	0.982	0.771	0.599
Slope	1.29	1.46	4.06	-3.51	0.929	0.410	0.055
Curvature	0.12	0.72	3.17	-1.84	0.788	0.259	0.110

Table 4.7.1: Summary Statistics for Empirical US Zero Coupon Yields

4.8 Results

4.9 Maximum Likelihood Estimation of the Optimal Parameters

There were eleven values to be optimized for each calibration of the data sets: the AR(1) parameter, mean level and variance for all three loading parameters L_t, S_t and C_t from equation 4.4.5. In addition the error variance on the observations (which is assumed to be identical for all observations) and the value of lambda from equation 4.4.3.

The optimal parameters were found using the MARSS (Multivariate Autoregressive State-Space Modelling) package in R provided by Holmes et al. (2015).

The package allows maximum likelihood estimates of the parameters of linear multivariate autoregressive state space models with Gaussian errors through application of the EM (Expectation Maximization) algorithm. The package performed excellently and was able to find robust estimates on much smaller data sets than any other method tested. However the package as presented is not able to estimate the lambda parameters. Thus it was necessary to augment the algorithm with a separate step to update the lambda values between each step of the EM algorithm. To find the optimal value, the log-likelihood was differentiated numerically with respect to lambda using the numDeriv package Gilbert and Varadhan (2015), the value of lambda that set this derivative to zero was then found using the R function unitroot. Estimating lambda in this way was based on an idea mentioned in Coroneo et al. (2014). It was also possible to estimate the standard errors for the lambda estimate using the hessian function from numDeriv and taking the inverse of its negative value (as the log-likelihood is being maximized) and then taking the square root of the result. The various parameter estimates are listed in table 4.9.1 onwards, each estimate is listed with its standard error in brackets below it.

4.9.1 Maximum Likelihood Estimation of the Parameters for each Subperiod

Parameter Estimates Subset 1										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.973	0	0	7.98	L_t	0.031	0	0	0.097	0.0147
	(0.038)			(10.07)		(0.022)			(0.008)	(0.003)
S_t	0	0.919	0	-0.99	S_t	0	0.490	0		
		(0.050)		(1.38)			(0.070)			
C_t	0	0	0.606	0.04	C_t	0	0	1.2152		
			(0.118)	(0.46)				(0.119)		
Parameter Estimates Subset 2										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	1.018	0	0	4.78	L_t	0.120	0	0	0.135	0.0187
	(0.026)			(8.80)		(0.036)			(0.006)	(0.003)
S_t	0	0.883	0	0.69	S_t	0	0.881	0		
		(0.060)		(1.09)			(0.092)			
C_t	0	0	0.689	0.93	C_t	0	0	2.889		
			(0.095)	(0.74)				(0.173)		
Parameter Estimates Period 3										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.966	0	0	8.10	L_t	0.251	0	0	0.133	0.024
	(0.042)			(14.17)		(0.049)			(0.006)	(0.004)
S_t	0	0.733	0	-2.86	S_t	0	0.422	0		
		(0.083)		(0.92)			(0.067)			
C_t	0	0	0.845	-1.40	C_t	0	0	1.097		
			(0.064)	(0.96)				(0.126)		
Parameter Estimates Period 4										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.874	0	0	8.56	L_t	0.077	0	0	0.080	0.006
	(0.072)			(4.82)		(0.028)			(0.003)	(0.002)
S_t	0	0.972	0	-2.26	S_t	0	0.114	0		
		(0.037)		(2.83)			(0.034)			
C_t	0	0	0.794	-0.12	C_t	0	0	0.464		
			(0.082)	(0.49)				(0.075)		
Parameter Estimates Period 5										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.963	0	0	6.45	L_t	0.063	0	0	0.062	0.004
	(0.037)			(7.41)		(0.025)			(0.001)	(0.001)
S_t	0	0.994	0	1.79	S_t	0	0.089	0		
		(0.017)		(8.39)			(0.030)			
C_t	0	0	0.918	-1.74	C_t	0	0	0.597		
			(0.052)	(1.64)				(0.079)		
Parameter Estimates Period 6										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.951	0	0	5.44	L_t	0.054	0	0	0.082	0.003
	(0.049)			(5.95)		(0.023)			(0.004)	(0.001)
S_t	0	0.982	0	1.52	S_t	0	0.071	0		
		(0.051)		(3.68)			(0.027)			
C_t	0	0	0.857	0.39	C_t	0	0	0.210		
			(0.078)	(0.51)				(0.051)		

Table 4.9.1: Results of the Maximum Likelihood Estimation for Periods 1-6

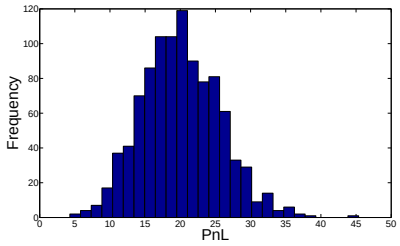
4.9.2 Maximum Likelihood Estimation of the Parameters for the WFA

Parameter Estimates Periods 1 and 2										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.997	0	0	6.02	L_t	0.077	0	0	0.123	0.017
	(0.040)			(4.37)		(0.021)			(0.005)	(0.002)
S_t	0	0.910	0	-0.02	S_t	0	0.720	0		
		(0.037)		(0.32)			(0.059)			
C_t	0	0	0.681	0.28	C_t	0	0	2.130		
			(0.070)	(0.46)				(0.107)		
Parameter Estimates Periods 2 and 3										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.974	0	0	10.71	L_t	0.204	0	0	0.135	0.021
	(0.020)			(8.24)		(0.031)			(0.005)	(0.002)
S_t	0	0.921	0	-1.26	S_t	0	0.728	0		
		(0.036)		(1.12)			(0.060)			
C_t	0	0	0.802	-0.02	C_t	0	0	2.051		
			(0.056)	(0.27)				(0.110)		
Parameter Estimates Periods 3 and 4										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.959	0	0	8.70	L_t	0.168	0	0	0.111	0.016
	(0.020)			(4.88)		(0.028)			(0.004)	(0.002)
S_t	0	0.896	0	-2.48	S_t	0	0.284	0		
		(0.039)		(0.98)			(0.038)			
C_t	0	0	0.845	-0.78	C_t	0	0	0.720		
			(0.048)	(0.53)				(0.071)		
Parameter Estimates Periods 4 and 5										
A	L_{t-1}	S_{t-1}	C_{t-1}	μ	Q	L_t	S_t	C_t	λ	H
L_t	0.975	0	0	7.42	L_t	0.070	0	0	0.067	0.005
	(0.030)			(9.71)		(0.019)			(0.001)	(0.001)
S_t	0	0.982	0	-2.07	S_t	0	0.096	0		
		(0.025)		(4.57)			(0.022)			
C_t	0	0	0.918	-0.90	C_t	0	0	0.561		
			(0.037)	(0.96)				(0.057)		

Table 4.9.2: Results of the Maximum Likelihood Estimation for WFA Calibration Periods

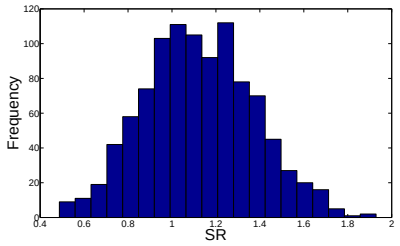
As can be seen the L_t, S_t and C_t all show high persistence particularly the first two. Examination of the estimated lambda values in table 4.9.1 reveals that there is some variation between periods. It's hard to discern a definite pattern from only 6 values but the estimated lambda values for periods 2 and 3 are significantly higher than the others. This may indicate the presence of a regime change in the market over those two periods, that reverted back from period 4 onwards. A similar pattern is observed in the WFA calibration results but is distributed slightly differently due to the differing bucketing of the data.

4.9.3 Simulation Study Results



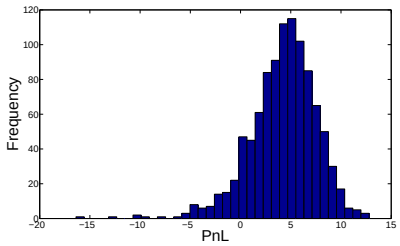
Histogram/Descriptive Stats.
of PnL for Period 3

Statistic	Value
Max	4,536 bp
min	429 bp
Mean	2,008 bp
Std	555 bp
Skew	0.29
Ex. Kurtosis	0.30
Percent Positive	100%



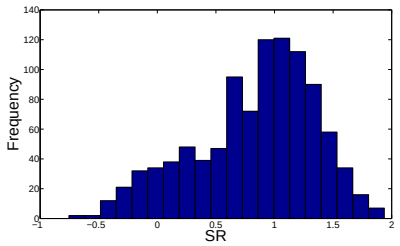
Histogram/Descriptive Stats.
of SR for Period 3

Statistic	Value
Max	1.93
min	0.49
Mean	1.12
Std	0.25
Skew	0.17
Ex. Kurtosis	-0.23
Percent Positive	100%



Histogram/Descriptive Stats.
of PnL for Period 4

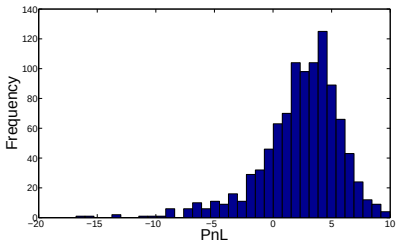
Statistic	Value
Max	1,280 bp
min	-1,638 bp
Mean	423 bp
Std	330 bp
Skew	-0.95
Ex. Kurtosis	2.93
Percent Positive	91.1%



Histogram/Descriptive Stats.
of SR for Period 4

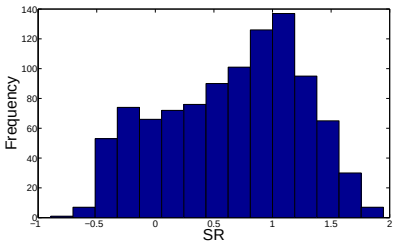
Statistic	Value
Max	1.94
min	-0.75
Mean	0.83
Std	0.52
Skew	-0.56
Ex. Kurtosis	-0.28
Percent Positive	91.1%

Figure 4.9.1: Results of Simulation Study for the Steepener Strategy



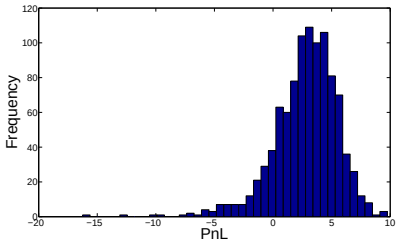
Histogram/Descriptive Stats.
of PnL for Period 5

Statistic	Value
Max	997 bp
min	-1,681 bp
Mean	240 bp
Std	343 bp
Skew	-1.24
Ex. Kurtosis	3.05
Percent Positive	81.2%



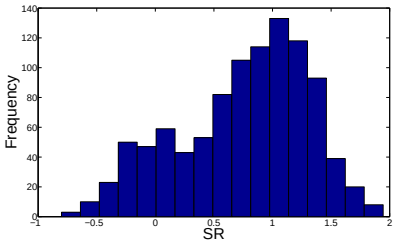
Histogram/Descriptive Stats.
of SR for Period 5

Statistic	Value
Max	1.95
min	-0.89
Mean	0.65
Std	0.59
Skew	-0.25
Ex. Kurtosis	-0.91
Percent Positive	81.2%



Histogram/Descriptive Stats.
of PnL for Period 6

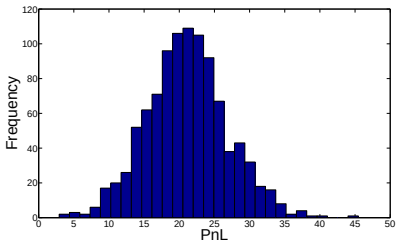
Statistic	Value
Max	975 bp
min	-1,625 bp
Mean	279 bp
Std	276 bp
Skew	-1.22
Ex. Kurtosis	4.23
Percent Positive	86.8%



Histogram/Descriptive Stats.
of SR for Period 6

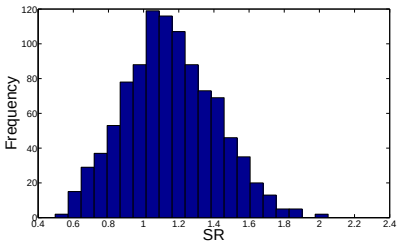
Statistic	Value
Max	1.94
min	-0.80
Mean	0.75
Std	0.55
Skew	-0.47
Ex. Kurtosis	-0.56
Percent Positive	86.8%

Figure 4.9.2: Results of Simulation Study for the Steepener Strategy



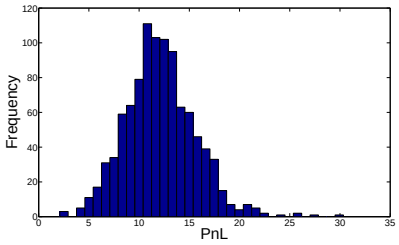
Histogram/Descriptive Stats.
of PnL for Period 3

Statistic	Value
Max	4,547 bp
min	292 bp
Mean	2,112 bp
Std	577 bp
Skew	0.15
Ex. Kurtosis	0.35
Percent Positive	100%



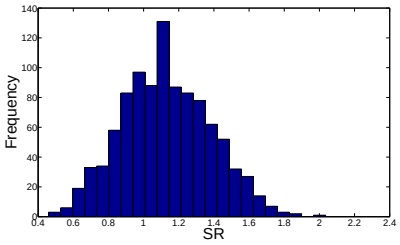
Histogram/Descriptive Stats.
of SR for Period 3

Statistic	Value
Max	2.05
min	0.50
Mean	1.15
Std	0.26
Skew	0.20
Ex. Kurtosis	-0.21
Percent Positive	100%



Histogram/Descriptive Stats.
of PnL for Period 4

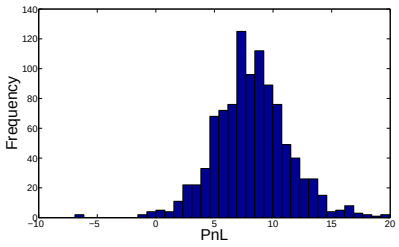
Statistic	Value
Max	3,034 bp
min	210 bp
Mean	1,211 bp
Std	346 bp
Skew	0.48
Ex. Kurtosis	1.33
Percent Positive	100



Histogram/Descriptive Stats.
of SR for Period 4

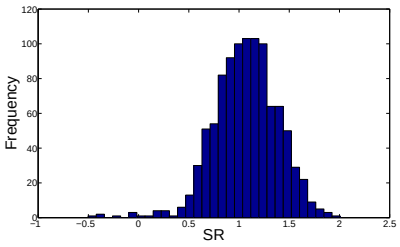
Statistic	Value
Max	2.04
min	0.46
Mean	1.13
Std	0.25
Skew	0.16
Ex. Kurtosis	-0.27
Percent Positive	100

Figure 4.9.3: Results of Simulation Study for the Curvature Strategy



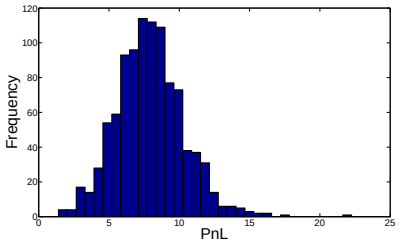
Histogram/Descriptive Stats.
of PnL for Period 5

Statistic	Value
Max	1,996 bp
min	-691 bp
Mean	816 bp
Std	317 bp
Skew	0.04
Ex. Kurtosis	1.30
Percent Positive	99.3%



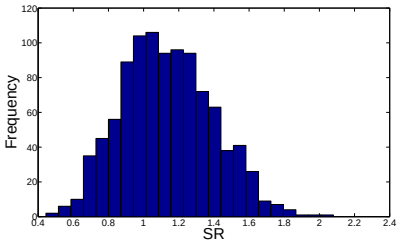
Histogram/Descriptive Stats.
of SR for Period 5

Statistic	Value
Max	2.01
min	-0.50
Mean	1.07
Std	0.32
Skew	-0.50
Ex. Kurtosis	1.56
Percent Positive	99.3%



Histogram/Descriptive Stats.
of PnL for Period 6

Statistic	Value
Max	2,226 bp
min	141 bp
Mean	792 bp
Std	240 bp
Skew	0.52
Ex. Kurtosis	1.58
Percent Positive	100%



Histogram/Descriptive Stats.
of SR for Period 6

Statistic	Value
Max	2.08
min	0.44
Mean	1.13
Std	0.26
Skew	0.27
Ex. Kurtosis	-0.18
Percent Positive	100%

Figure 4.9.4: Results of Simulation Study for the Curvature Strategy

4.9.4 Steepener Trading System Out of Sample Results for Empirical Data

Performance Summary: Period 3 Calibrated on Periods 1 & 2			
Total Net Profit	1,660 bp	Open Position PnL	683 bp
Number of Closed Trades	14	Number of Open Trades	1
Percent Profitable	100.0 %		
Sharpe Ratio	1.16		
Max Drawdown	0 bp		

Performance Summary: Period 4 Calibrated on Periods 2 & 3			
Total Net Profit	685 bp	Open Position PnL	292 bp
Number of Closed Trades	2	Number of Open Trades	1
Percent Profitable	100.0%		
Sharpe Ratio	0.46		
Max Drawdown	0 bp		

Performance Summary: Period 5 Calibrated on Periods 3 & 4			
Total Net Profit	3,682 bp	Open Position PnL	0 bp
Number of Closed Trades	14	Number of Open Trades	0
Percent Profitable	100.0%		
Sharpe Ratio	0.46		
Max Drawdown	0 bp		

Performance Summary: Period 6 Calibrated on Periods 4 & 5			
Total Net Profit	143 bp	Open Position PnL	(136) bp
Number of Closed Trades	14	Number of Open Trades	1
Percent Profitable	67%		
Sharpe Ratio	0.42		
Max Drawdown	127 bp		

Table 4.9.3: Steepener Strategy Performance Summary for Empirical Data

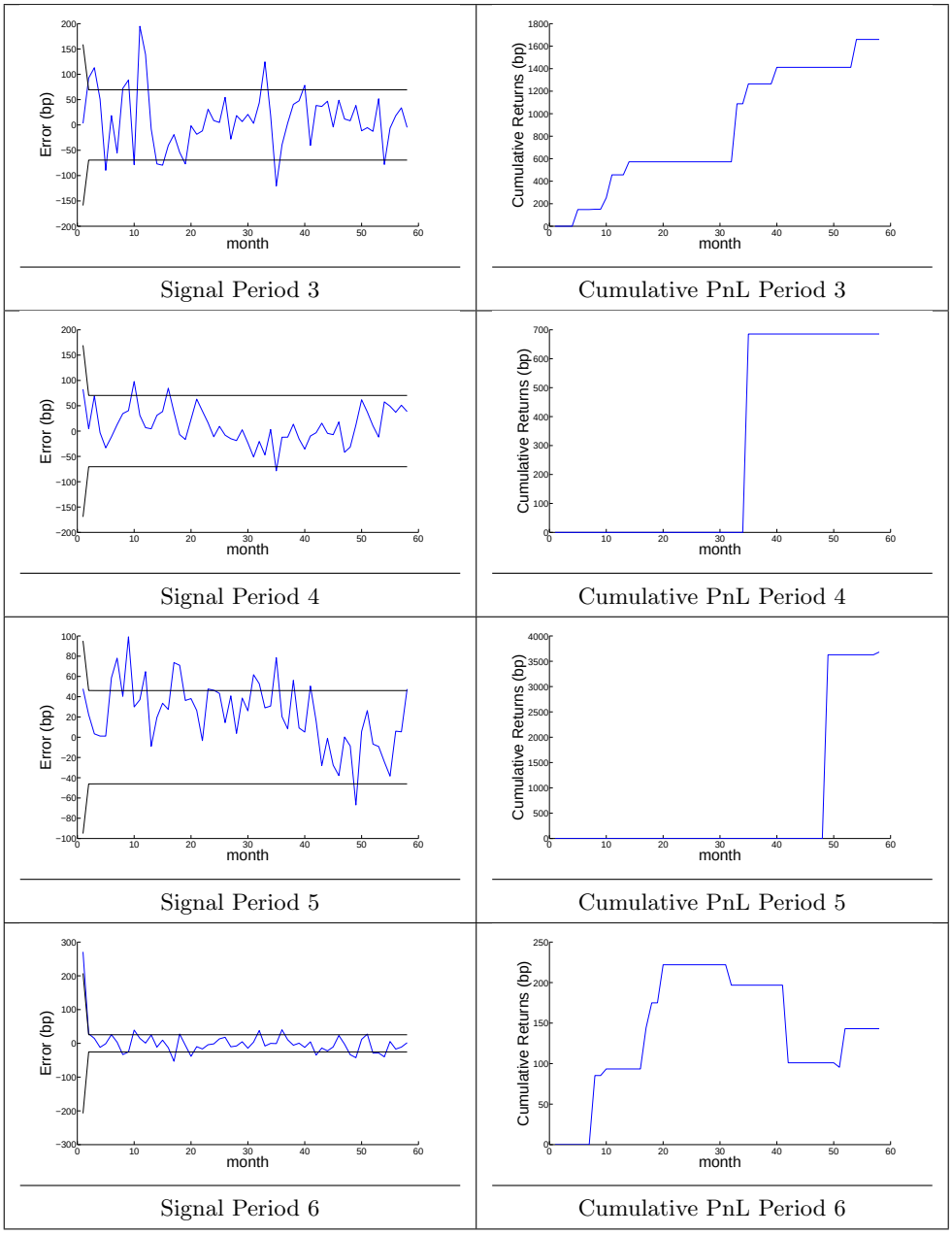


Figure 4.9.5: Steepener Signal and Cumulative PnL from application to Empirical Data

4.9.5 Curvature Trading System Out of Sample Results for Empirical Data

Performance Summary: Period 3 Calibrated on Periods 1 & 2			
Total Net Profit	836 bp	Open Position PnL	181 bp
Number of Closed Trades	7	Number of Open Trades	1
Percent Profitable	100.0 %		
Sharpe Ratio	0.75		
Max Drawdown	0 bp		

Performance Summary: Period 4 Calibrated on Periods 2 & 3			
Total Net Profit	410 bp	Open Position PnL	3 bp
Number of Closed Trades	5	Number of Open Trades	1
Percent Profitable	100.0%		
Sharpe Ratio	0.87		
Max Drawdown	0 bp		

Performance Summary: Period 5 Calibrated on Periods 3 & 4			
Total Net Profit	2,589 bp	Open Position PnL	(118) bp
Number of Closed Trades	12	Number of Open Trades	4
Percent Profitable	100.0%		
Sharpe Ratio	0.47		
Max Drawdown	0 bp		

Performance Summary: Period 6 Calibrated on Periods 4 & 5			
Total Net Profit	379 bp	Open Position PnL	(281) bp
Number of Closed Trades	7	Number of Open Trades	3
Percent Profitable	100%		
Sharpe Ratio	0.73		
Max Drawdown	0 bp		

Table 4.9.4: Curvature Strategy Performance Summary for Empirical Data

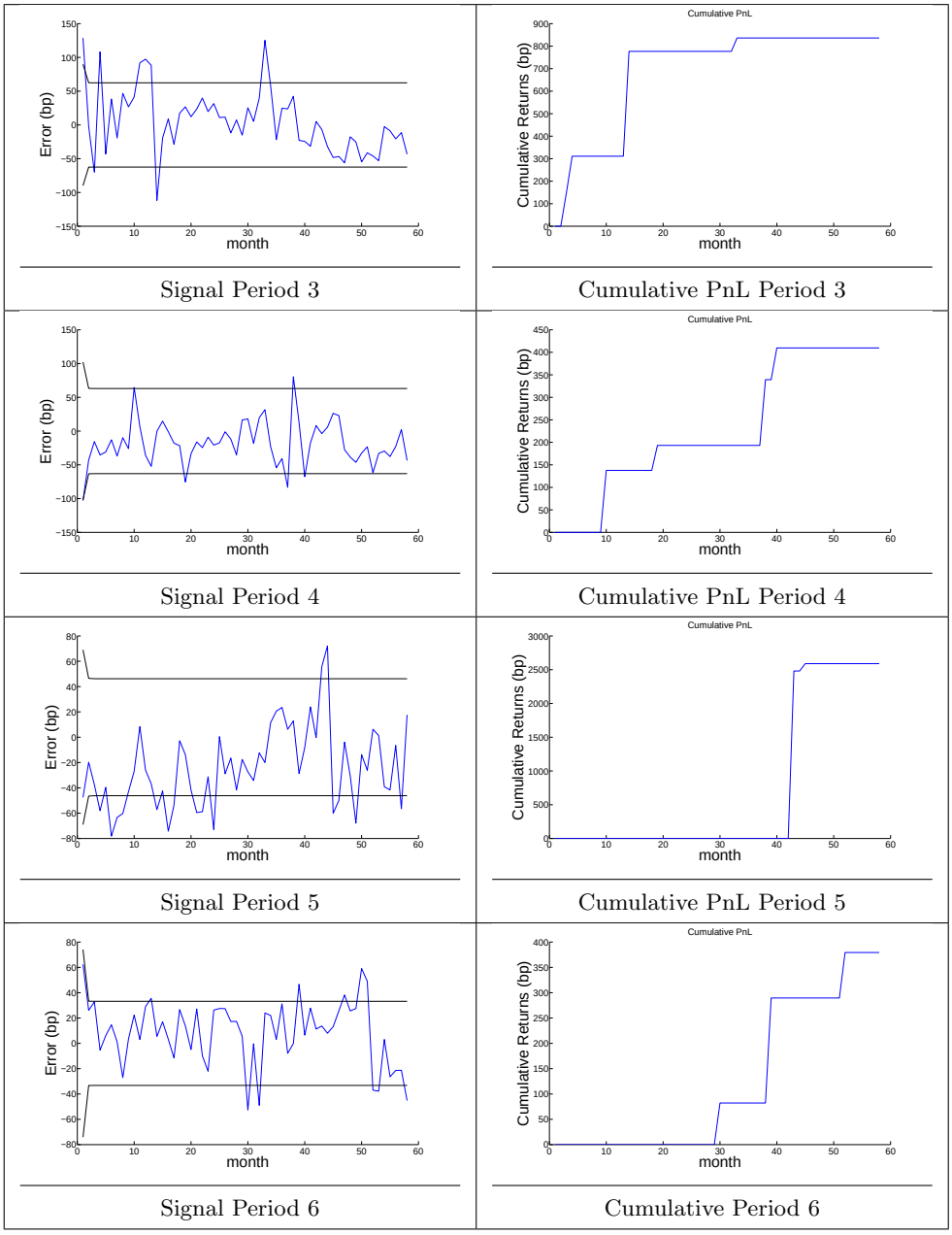


Figure 4.9.6: Curvature Signal and Cumulative PnL from application to Empirical Data

4.10 Discussion

The results of the simulation show that the strategy works well producing a positive PnL 100% of the time for the Curvature strategy for every test period apart from period 5 where it is 99.3% profitable. The Steepener strategy produced 100% profitable trades in the simulation of period 3, and produces a respectable: 91.1%, 81.2% and 86.8% profitability in periods 4, 5 and 6 respectively.

When applied to the simulated data the rv-DNS produced a Sharpe ratio of 1.12 for the Curvature strategy which can be considered good, and a Sharpe ratio of 0.84 Steepener strategy which is still reasonable. The Curvature strategy performed best on virtually every performance measure for the simulated data sets (although the results for period 3 were only marginally better than the Steepener results). These results are all the more impressive when one considers that they were generated by simple trading rules using outputs directly available from the Kalman filtering process, no additional tuning parameters were used. Moreover there was no attempt to finesse the PnL/Sharpe ratios by including stop losses/take profits, optimal position sizing (such as the Kelly Criterion) or modifying the strategy to only take trades with positive carry. Similarly effects that might have a deleterious effect, such as slippage or bid/offer spreads were not considered either. This was all done intentionally as the goal was to test whether the rv-DNS trading signal could produce the desired results without the PnL being polluted by secondary effects. So these results reflect the raw performance of the signal produced by the rv-DNS.

From the results of the simulation study it is evident that every subperiod except subperiod 3 produces a PnL distribution which is leptokurtic, this is especially true for the Steepener strategy. This echoes the results of Duarte et al. (2007), which was discussed in section 1.2, they also found that the results of curvature strategies are leptokurtic, although they were considering the kurtosis

of the returns rather than the cumulative PnL. Similarly every period except period 3 for the curvature strategy exhibited positive skewness again echoing the results of Duarte et al. (2007). However, the distribution of the PnLs from the Steepener strategy in every period except period 3 exhibits significant negative skewness. This may in part be explained by the different trading strategy used in this chapter rather than an intrinsic property of this type of strategy. The leptokurtic/negative skew effects are at least partly the result of large PnLs being produced when the signal spends a significant amount of time crossing the entry level and so generating several trades in one direction. When the signal finally does cross the opposite barrier all these trades exit at once. If the exit produces a positive PnL the PnL will be positive (and most often large as there are several trades) and vice versa. For example the best PnL seen in the simulation study was for the Curvature strategy in period 3. It had a raw PnL of 4,547 bp, and although it consists of 27 entries of trades there are only 11 exits, one of which sees 11 trades close for a combined profit of 2,257 bp. PnLs produced this way are volatile relative to their peers and so are penalized by the Sharpe ratio. For example the best PnL previously discussed has a relatively modest SR of 0.9. So the Sharpe ratios probably give a better measure of the system performance. Of course the strategy could be amended to prevent such PnLs from being produced: the most obvious solution being to limit the number of open trades in any one direction at any one time. Note also that the highest Sharpe ratios are generated by PnLs that make lots of trades with different entries and exits and so these would not be affected by a cap on the number of trades in any one direction. Therefore the system performance could undoubtedly be boosted and the results as presented represent a conservative estimate of the system performance, however the focus of this study is to examine the performance of the signal in its basic form so no such adjustments were made.

As noted previously the Nelson and Siegel basis functions are similar to the empirical loading factors found by Bliss (1997) and Litterman and Scheinkman

(1991) and in particular the second and third Nelson and Siegel basis functions can be thought of as proxies for the steepness and curvature of the curve. There does indeed seem to be some correlation between the level of the AR(1) parameters for the loading function on these basis functions and the profitability of the strategy. Generally the lower the level of the AR(1) parameter the more profitable the strategy tends to be, this is most apparent for the mean values produced by the Steepener strategy in the WFA. This is logical as the strategy is predicated on mean reversion and those loading factors with the lower AR(1) parameters tend to be more mean reverting. Similarly the volatility of the error process for the corresponding AR process also contributes to the profitability of the strategy, with larger volatilities producing larger deviations for the model to revert from. This is exemplified by the average values of the Curvature strategy for the WFA in periods 3 and 6. Although both these periods have similar values for their AR(1) parameter, period 3 produces the superior PnL thanks to its larger volatility. However given the relatively small number of parameter values tested it's hard to be certain about these patterns at this stage. A full sensitivity analysis for all the model parameters is left as future work but such insights could be used to identify market conditions which are favorable and vice versa.

The out of sample results produced positive PnLs for both strategies in every period tested. Moreover they produced average Sharpe ratios of 0.63 for the Steepener Strategy and 0.71 for the Curvature strategy, which represents an underperformance relative to the simulated Sharpe ratios, but are still respectable. In addition all the results produced in the out of sample testing (apart from subperiod 5 which the out of sample results outperformed the simulation results) were covered by the distribution of results in the simulation study. Thus the out of sample results are comparable to results produced when the data satisfies all the assumptions upon which the model is based and the optimal filtering parameters are used.

The Steepener produced an average Sharpe Ratio in the out of sample testing, that was closer to its simulated counterpart than the Curvature average was, although this was largely due to a strong performance in the first out of sample testing period. The Steepener strategy also tended to produce better out of sample PnLs than the Curvature strategy in all but the last test period. The Steepener strategy performed well in period 3 producing a PnL and Sharpe ratio close to their simulated counterparts. Also its PnL figure for period 4 was towards the top of the distribution of simulated PnLs although this was based on very few trades resulting in a Sharpe ratio which was less impressive at 0.46 but still within one standard deviation of the mean of the Sharpe ratio for the simulated data for that period. The Steepener strategy's PnL outperformed its simulated counterpart in period 5 and this will be discussed separately. Its worst performance came in period 6 both relative to the other out of sample Steepener results and the out of sample Curvature result for period 6. However the Steepener strategy also performed poorly in period 6 in the simulation study, and the out of sample result was less than one standard deviation from the mean of the simulation study, similarly its Sharpe ratio is within one standard deviation of its simulated counterpart so viewed in that context the results seem reasonable. The underperformance of the Steepener strategy in period 6 (both out of sample and simulated), is likely linked to the volatility of the autoregressive process for the Steepener loading parameter in this period which is very low so there simply is not much movement for the strategy to capture.

In contrast, for the Curvature strategy the out of sample PnLs were towards the lower end of the distribution of the simulated results (with the exception of subperiod 5). As noted above they underperform the Steepener strategy out of sample with the exception of period 6. This in tandem with the fact that the number of trades they produced were also toward the lower end of the distribution of simulated results, especially for period 4, may suggest that using the parameters from the previous two periods may attenuate the system performance at least for the curvature strategy. A sensitivity analysis of the strategies

to incorrect filtering parameters could be performed using simulated data and this would help to help answer this question, however this is left as future work. That said as we have seen raw PnL is not the best measure of system performance and the curvature strategies do produce respectable Sharpe ratios out of sample which are stable across the subperiods (excluding subperiod 5 which will be discussed below). Moreover as noted previously the PnLs and Sharpe ratios (and indeed number of trades produced) were all within the distribution generated by the simulation study and so could have been produced even when optimal parameters were used in the filtering process.

The out of sample PnLs for period 5 were larger than their simulated counterparts. These results were another example of a PnL where a large number of trades were built up in one direction and then exited all at once. This is reflected in the Sharpe ratios for period 5, which are the worst for the Curvature Strategy and close to worst for the Steepener strategy despite having by far the largest PnL.

Potential improvements to the trading logic include: the previously discussed cap on the maximum number of trades in any one direction and the use of the sensitivity analysis to identify favorable market conditions. Also the decision to use the strategy error as given from the filtering process was essentially arbitrary and motivated by the desire to keep the model simple, however an additional parameter could be incorporated to allow this value to be scaled and so tuned to optimize the system performance. In addition application of basic money management techniques such as optimal position sizing, stop losses/take profit levels should allow more of the systems potential to be realized.

In terms of potential improvements to the underlying DNS model structure. Given that the calibration results suggest a possible regime change in the lambda values, it would be interesting to augment the model into one of the Markov switching models discussed in the introduction as to allow it to detect and adapt to a regime change during a trading period. Also the model currently assumes

that the observation covariance matrix should be diagonal and that the value on that diagonal should be constant. This may be overly restrictive especially given the focus here on relative value trading where differences in measurement error volatility across maturities may be an important source of signal. It would be better to at least permit the values across the diagonal to change. Going further one could model a full covariance matrix to capture the correlations in the measurement errors. Similarly the decision to use the regular DNS model rather than the arbitrage free one as the basis for the rv-DNS although well motivated was ultimately a judgment call. Given that it would be relatively straightforward to extend the rv-DNS model presented here to an arbitrage free version (AF rv-DNS) and it would be interesting to test the arbitrage free version to see what effect it has on the results. This is left as future work.

The initial results from the simulations and out of sample testing are encouraging, the next step would be to test the system on more recent data and/or higher frequency data. The current ultra low rate regime would likely necessitate some changes to the model, this problem has already been tackled in the context of the pure DNS by Christensen (2013) and so this could serve as the basis to allow the rv-DNS to incorporate a zero bound. This however would be at the expense of higher parameterization putting extra stress on the calibration process. Moving to higher frequency data would have to be carefully considered too, as noted in Meucci (2009), the Ornstein-Uhlenbeck process, which is the discrete time analogue of the AR(1) processes used in the rv-DNS, behaves pretty much like a Brownian motion, as at smaller time scales the effect of the random component dominates the mean reverting component. This problem of higher frequencies could be addressed either by modifying the rv-DNS model or testing intermediate frequencies such as weekly or fortnightly to find which ones still allow the mean-reverting dynamics to take effect while still allowing more trades to be generated. Again this is all left as future work.

Lastly it must be emphasized again that aside from the calibration of the un-

derlying DNS model, no attempt was made to fine tune the trading parameters to finesse the PnL/Sharpe ratios, indeed the trading system based on the signal produced was kept intentionally simple so that the power of the raw system signal could be tested. So the results can be considered as promising at the very least.

4.11 Conclusion

This paper introduces a rv-DNS trading model and illustrates its use, showing that the DNS can be used as the basis for a profitable relative value trading model. Although based on the regular DNS model it takes the outputs from the Kalman filtering process and applies them in a different way to produce a trading signal. These signals are then used to enter and exit Steepener and Curvature positions as dictated by the trading rules. The rv-DNS was tested out of sample using a walk forward analysis: The system produced positive out of sample results in every period tested and also produced respectable Sharpe ratios. All the trading rules employed were straightforward and did not incorporate any additional manipulation of the data or the signal itself, thus at a minimum these results must be considered as highly encouraging and hence the model provides a good platform for further development.

Chapter 5

Conclusions and Directions for Future Research

5.1 Conclusions

In chapter two it was shown that Differential Evolution could be adapted successfully to fit the rv-NSS model, it contrasted the results with those obtained from a derivative based optimizer and was shown to produce far superior results. An additional round of testing was conducted in which prior information was assumed for the parameter values, effectively constraining the range of starting parameter values to be smaller. This did improve the results for the derivative based method but it was still outperformed by the Differential Evolution based method when used to calibrate the rv-NSS. These results extend the work of Gilli et al. (2010) and represents a new contribution to the literature.

Clark and Troskie (2007) which extended Kibria (2003), compared a wide range of ridge regression estimators, chapter three extends these results by including the newest ridge estimators in the literature into the comparison study as well as some older ones. As such this work represents a new contribution to the literature.

The goal of chapter three was to identify the best performing ridge regression estimator, ideally one that would be best in all the scenarios considered. The study showed that no one estimator performed best in every scenario, but it was possible to identify one estimator: the Linearized Ridge Regression Estimator (LRRE) Gao and Liu (2011) that performed best for most of the scenarios considered. This again represents a new contribution as the best estimator differs from that found in the previous studies. The only place that this model struggled was for the very lowest volatility level considered and when the level of multicollinearity was medium to low. In such situations the LRRE model tended to underperform OLS (the benchmark model for the tests). Indeed as none of the other models were able to beat OLS in such situations it's reasonable to let the estimator default to the OLS estimate in such situations.

The IRE type estimators of Liu et al. (2013) experienced convergence problems and chapter three addresses those by conditioning the choice of the estimated beta parameter on the theoretical MSE value it produces, with success in the majority of conditions tested. This complements the work of Liu et al. (2013) and represents a new contribution to the literature. As noted before the LRRE of Gao and Liu (2011) also had estimation problems at the lowest volatility levels and an attempt was made to fix this too, also by conditioning the estimated beta on the theoretical MSE. Although the solution did fix its convergence issues at the lowest volatility levels, this was at the expense of its performance at every other volatility level. However this work still extends Gao and Liu (2011) and represents a further contribution to the literature.

In addition the modified r-k class ridge regression (MCRR) introduced by Batah et al. (2009) did not specify how to estimate the number of principal components to retain in their model, chapter 3 examined two different ways to address this shortcoming. The first method based on reducing the condition number below a threshold level of 10 MCRR CI, the second was chosen to minimize the theoretical MSE value MCRR (the other estimators that incorporated PCA: r-k and PCR also had the number of principal components to retain determined in the same way to facilitate a fair comparison). The MCRR CI version performed best when applied to matrices with the highest condition numbers, however when applied to the remaining matrices the performance of the two versions of the estimator tended to converge. This work extends Batah et al. (2009) and represents the final contribution to the literature. Overall these results should prove interesting to anyone using ridge regression.

Chapter four saw the introduction of the rv-DNS. It tested the model on simulated and empirical data, the model was found to work as intended producing positive results. The rv-NSS returned an average Sharpe ratio of 1.12 for the simulated Curvature strategy which can be considered good and 0.84 for the simulated Steepener strategy which is still respectable. The out of sample re-

sults from the WFA produced an average Sharpe ratio of 0.71 for the Curvature strategy and 0.63 for the Steepener strategy which represents an underperformance relative to the simulated results but are still reasonable. To my knowledge there are no other relative value trading models based on the DNS and as such the rv-DNS represents a new contribution to the literature. The new model proposed will likely interest academics while the successful application should interest practitioners.

Overall these three empirical chapters have shown how it is possible to successfully develop and calibrate a fixed income relative-value trading model based on the Nelson Siegel approach.

5.2 Directions for Future Research

The results of chapters two and three mean that it is now possible to produce plausible estimates of the rv-NSS parameters. Although this was not included in this thesis, as the focus of the research switched to the rv-DNS, it would still be interesting to calibrate and test the rv-NSS to compare and contrast the results. One obvious comparison to make would be with the results produced by Duarte et al. (2007) as the rv-NSS is a close analogue of the Duarte et al. (2007) model. It would be interesting to see if the reported superior forecasting abilities of the NSS compared to the Vasicek model translated into superior trading performance for the rv-NSS as hoped.

Moreover it would also be interesting to compare the rv-DNS to the Duarte et al. (2007). Right now there are sufficient differences that would invalidate a direct comparison of the results.

- Duarte et al. (2007) produces their results in sample
- they consider a different set of maturities
- they consider a different set of dates
- they only consider curvature trades
- their trading rules differ
  - they take one position at a time
  - they then hold it for 12 months
  - whereas the rv-DNS takes a position every time a signal is generated
  - the rv-DNS uses the same signal for entry and exit
  - they use a predefined trigger level
  - rv-DNS trigger level is based on the outputs from the Kalman filtering process

However it would not be difficult to address these differences allowing a comparison to be made. If the rv-DNS results are found to be better than those of the rv-NSS and Duarte et al. (2007) this would show that the modelling of the dynamics internally adds value. Such a comparison would likely interest academics and practitioners alike.

Another potential application of the rv-NSS is in HFT. As the rv-NSS bases its update of the curve on the movement of the pillar points, it might be possible to base a High Frequency Trading (HFT) strategy on it. As the pillar points move the models estimates of the non pillar points will move, it could in theory be possible to trade the change in relative value produced. For example as the pillar

points change the rv-NSS estimate of the seven year rate will change relative to the rate currently quoted in the market. If the spread has narrowed, then in theory it would be possible to short the current seven year rate against the pillar points and vice versa assuming you had a lower latency than the person quoting the seven year rate. The simple nature of the calculations involved should help to achieve the low latencies required.

The results of chapter 4 while perfectly reasonable in their own right likely represented a conservative estimate of the rv-DNS performance. The system as presented was kept intentionally straightforward so that the performance of the raw signal could be examined. Now that chapter four has demonstrated that the signal is genuine, the next step would be to build a system that optimizes the PnL/Sharpe ratios produced, as discussed in chapter 4 there are a number of ways in which this can be done:

Firstly it was apparent that the strategies' performance was sensitive to the level of the AR(1) parameter, with smaller parameter values (being more mean reverting) producing better PnLs. Similarly the volatility of the AR process also influenced the PnL produced with larger volatilities meaning there is more movement to capture. By conducting a full sensitivity analysis of the model to its inputs, it would be possible to identify which market conditions are favorable and vice versa. By restricting trading to the more favorable periods overall performance should be enhanced.

Secondly the threshold value could be scaled and tuned to optimize the PnL/Sharpe ratios produced.

Thirdly optimal position sizing (perhaps based on how favorable the current market conditions are) and similar money management techniques such as stop losses/profit targets could be used to let the system realize more of its potential.

The results showed that some large PnLs could be produced when the model

spends a significant amount of time criss-crossing one of the thresholds building up several open positions before finally crossing the opposite threshold closing out all the open positions. These PnLs could be positive or negative but even when positive their volatile nature means that the associated Sharpe ratio tends to be modest. The most obvious way to prevent this would be to cap the number of trades which can be open at any one time. However it would be better to try to identify what causes these PnLs to be produced in this way. From visual inspection of the signals vs threshold graphs one possible explanation is that the level of volatility is changing, with low volatility causing the signal to cluster around the threshold initially, followed by a higher volatility period in which it then crosses the other threshold. If this is the case then Markov switching DNS models, with high and low volatility regimes such as Levant and Ma (2013) and Xiang and Zhu (2013) could be used to extend the rv-DNS, to address this problem.

In terms of improvements to the underlying DNS structure:

The assumption that the observation covariance matrix should be diagonal with a constant on the diagonal may be overly restrictive. Given the strategies rely on the relative movement of different points on the curve, differences in the observation volatilities could be an additional source of signal.

The decision to use the regular DNS rather than the arbitrage free version was based primarily on the desire for the model to be as parsimonious as possible. By developing and testing the AF rv-DNS the veracity of this decision can be tested.

Lastly there was some evidence of a regime change in the value of the lambda parameter during the calibration to the empirical data. The results of the previously suggested sensitivity analysis should allow the effect of this volatility to be quantified. Using Markov switching type models this time with differing lambdas in each regime could allow the model to handle such changes in the

lambda value internally, as was done by Levant and Ma (2013) for the DNS. More generally Koopman et al. (2010) showed how lambda could be made a time varying parameter of the model like the betas in the DNS, this also could be used as a basis to extend the rv-DNS.

Aside from improvements to the model structure, the other avenue for future research is in terms of the data sets to which the model is applied. A logical next step would be to test the model on more recent data. Given the current ultra low rate regime this may necessitate the incorporation of the zero bound version of the DNS presented by Christensen (2013). Going further it would be interesting to apply the model to higher frequency data. Such an extension may not be straightforward because as shown by Meucci (2009) at very short time intervals the effect of the Brownian motion dominates the mean reversion part of the process (he showed this in the context of the Ornstein-Uhlenbeck process which is the continuous time analogue of the AR(1) process used in chapter 4). So moving to shorter time frames may mean that the model needs to be adapted or that some intermediate frequency such as weekly or fortnightly can be found that allows the frequency to be increased while still allowing the model to function as intended.

Bibliography

Al-Gereari, F. (2012), 'Modified marquardt estimator', *International Journal of Research and Reviews in Applied Sciences* **13**, 29–40.

Alexander, C. (2001), *Market Models A Guide to Financial Data Analysis*, Wiley, Chichester, England.

Anderson, N. and Sleath, J. (1999), 'New estimates of u.k. real and nominal yield curves', Bank of England Quarterly Bulletin.

Anneart, J., Claes, A., De Ceuster, M. and Zhang, H. (2012), 'Estimating the yield curve using the nelson-siegel model: A ridge resgression appoach', International Review of Economics and Finance, Forthcoming.

Aronson, D. (2006), *Evidence Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*, Wiley, Chichester, England.

Bandara, W. and Munclinger, R. (2012), Term structure modeling with structural breaks: A simple arbitrage free approach. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract/id=1974033.

Batah, F., Ozkale, M. and Gore, S. (2009), 'Combining unbiased ridge and principal component regression estimators', *Communications in Statistics-Theory and Methods* **38**, 2201–2209.

Baye, M. and Parker, D. (1984), 'Combining ridge and principal component regression: a money demand illustration', *Communications in Statistics-Theory and Methods* **13**, 197–205.

Belsley, A., D. (1991), *Conditioning Diagnostics*, John-Wiley and Sons, New York, USA.

Bernadell, C., Coche, J. and Nyholm, K. (2005), 'Yield curve prediction for the strategic investor', European Central Bank, Working Paper Series No. 472.

BIS (2005), 'Zero-coupon yield curves: Technical documentation', Bank for International Settlements, Basle.

Bliss, R. (1997), 'Movements in the term structure of interest rates', Economic Review Federal Reserve Bank of Atlanta. 82.

Bolder, D. and Streliksi, D. (1999), 'Yield curve modelling at the bank of canada', Bank for Canada Technical Report.

Bozdogan, H.and Haughton, D. (1998), 'Informational complexity criteria for regression models', *Computational Statistics and Data Analysis* **28**, 51–76.

Cairns, A. and Pritchard, D. (2001), 'Stability of descriptive models for the term structure of interest rates with applications to german market data', *British Actuarial Journal* **7**, 467–507.

Chan, E. (2008), *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*, Wiley, Chichester, England.

Chan, E. P. (2013), *Algorithmic Trading: Winning Strategies and their Rationale*, Wiley, New Jersey, USA.

Chincarini, L. B. (2012), *The Crisis of Crowding: Quant Copycats, Ugly Models, and the New Crash Normal*, John Wiley and Sons, New Jersey.

Christensen, J. (2013), A regime-switching model of the yield curve at the zero bound. Federal Reserve Bank of San Francisco, Working Paper Series.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2009), 'An arbitrage-free generaliszed nelson-siegel term structure model', *The Econometrics Journal* **12**, 33–64.

Christensen, J. H. E., Diebold, F. X. and Rudebusch, G. D. (2011), 'The affine arbitrage-free class of nelson-siegel term structure models', *Journal of Econometrics* **164**, 4–20.

Christensen, J. M. (2012), Nelson siegel and affine processes- and their use in term structure modeling. MSc. Finance Thesis Aarhus School of Business.

Clark, A. E. and Troskie, C. G. (2007), 'Ridge regression-a simulation study', *Communications in Statistics-Simulation and Computation* **35**, 605–619.

Coroneo, L., Giannone, D. and Modugno, M. (2014), 'Unspanned macroeconomic factors in the yield curve', Federal Reserve Board, Washington, D.C. 57.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2008), 'How arbitrage-free is the nelson siegel model?', Bank of Canada Working Paper.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2011), 'How arbitrage-free is the nelson-siegel model?', *Journal of Empirical Finance* **18**, 393–407.

Cox, J. C., Ingersoll, J. and Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**, 385–407.

Crouse, R., Jin, C. and Hanumara, R. (1995), 'Unbaised ridge estimation with prior information and ridge trace', *Communications in Statistics-Theory and Methods* **24**, 2341–2354.

Dai, Q. and Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *The Journal of Finance* **55**, 1943–1978.

De Boer, P. and Hafner, C. (2005), 'Ridge regresion revisited', *Statistica Neerlandica* **59**, 498–505.

De Pooter, M. (2007), Examining the nelson-siegel class of term structure models. Tinbergen Institute Discussion Paper TI 2007/043/4.

Dempster, A. P., Schatzoff, M. and Wermuth, N. (1977), 'A simulation study of alternatives to ordinary least squares', *Journal of the American Statistical Association* **72**, 77–91.

Diebold, X. F. and Li, C. (2006), 'Forecasting the term structure of govenrment bond yields', *The Journal of Econometrics* **130**, 337–364.

Diebold, X. F., Rudebusch, G. D. and Aruoba, S. B. (2006), 'The macroeconomy and the yield curve: a dynamic latent factor approach', *The Journal of Econometrics* **131**, 309–338.

Duarte, J., LongStaff, F. A. and Yu, F. (2007), 'Risk and return in fixed-income arbitrage: Nickels in front of a steamroller?', *The Review of Financial Studies* **20**, 769–811.

Duffee, G. (2002), 'Term premia and interest rate forecasts in affine models', *The Journal of Finance* **57**, 405–443.

Duffie, D. and Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical Finance* **6**, 379–406.

Fabozzi, F., Martelli, L. and Priaulet, P. (2005), 'Predictibility in the shape of the term structure of interest rates', *Journal of Fixed Income* **15**, 40–53.

Fama, E. and Bliss, R. (1987), 'The information in long-maturity forward rates', *American Economic Review* **77**, 680–692.

Gao, F. and Liu, X. (2011), 'Linearized ridge regression estimator under the mean square error criterion in a linear regression model', *Communications in Statistics-Simulation and Computation* **40**, 1434–1443.

Garcia, R. and Perron, P. (1996), 'An analysis of the real interest rate under regime shifts', *The Review of Economics and Statistics* **78**, 111–125.

Gilbert, P. and Varadhan, R. (2015), *Accurate Numerical Derivatives*. R package version 2014.2-1.
URL: *https://cran.r-project.org/web/packages/numDeriv/numDeriv.pdf*

Gilli, G., Maringer, D. and Schumann, E. (2011), *Numerical Methods and Optimization in Finance*, Academic Press, Dan-Diego, USA.

Gilli, M., Grosse, S. and Schumann, E. (2010), Calibrating the nelson-siegel-svensson model. WPS-031.

Gimeno, R. and Nave, J. (2009), 'A generic algorithm estimation of the termstructure of interest rates', *Computational Statistica and Data Analysis* **53**, 2236–2250.

Goldstein, M. and Smith, A. (1974), 'Ridge type estimators for regression analysis', *Journal os the Royal Statistical Society B* **36**, 284–291.

Golub, G. (1979), 'Cross-validation as a method for choosing a good ridge parameter', *Technometrics* **21**, 215–223.

Gray, S. F. (1996), 'Modeling the conditional distibution of interst rates as a regime-switching process', *Journal of Financial Economics* **42**, 27–62.

Gross, J. (2003), *Linear Regression*, Springer, Heidelberg, Germany.

Gruber, M. (1998), *Improving Efficiency by Shrinkage*, Marcel Dekker, New York, USA.

Gunst, R. and Mason, R., L. (1977), 'Biased estimation in regression: an evaluation using mean squared error', *Journal of American Statistical Association* **72**, 616–628.

Gurkaynak, R., Sack, B. and Wright, J. H. (2006), 'The us treasury yield curve: 1961 to present', Federal Reserve Board Finance and Economics Discusssion Series. 28.

Hamilton, J. D. (1988), 'Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates', *Journal of Economic Dynamics and Control* pp. 385–423.

Heath, D., Jarrow, R. A. and Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology', *Econometrica* **60**, 77–105.

Hemmerelle, W. J. (1975), 'An explicit solution for generalised ridge regression', *Technometrics* **17**, 309–314.

Hevia, C., Gonzalez-Rozada, M., Sola, M. and Spagnolo, F. (2014), 'Estimating and forecasting the yield curve using a markov switching dynamic nelson-siegel model', *Journal of Applied Econometrics (forthcoming)* . Available at http://onlinelibrary.wiley.com/doi/10.1002/jae.2399/abstract.

Hocking, R., Speed, F. and Lynn, M. (1976), 'A class of biased estimators in linear regression', *Technometrics* **18**, 425–438.

Hoerl, A. E., Kennard, R. and Baldwin, K. (1975), 'Ridge regression: Some sumulations', *Communications in Statistics-Theory and Methods* **4**, 105–123.

Hoerl, A. E. and Kennard, R. W. (1970), 'Ridge regression: Application for non-orthogonal problems', *Technometrics* pp. 55–67.

Holmes, E., Ward, E. and Wills, K. (2015), *Multivariate Autoregressive State-Space Modeling*. R package version 3.9.
URL: *https://cran.r-project.org/web/packages/MARSS/MARSS.pdf*

Hull, J. and White, A. (1990), 'Pricing interest-rate derivative securities', *Review of Financial Studies* **3**, 573–592.

Hurn, A., Lindsay, K. and Pavlov, V. (2005), 'Smoothe estimation of yield curves by laguerre functions', Queensland University of TEchnology Working Paper.

Investopedia (n.d.), 'Understanding the sharpe ratio', http://www.investopedia.com/articles/07/sharperatio.asp.

Jollife, I. (1982), 'A note on the use of princiapl components in regression', *Applied Statistics* **31**, 300–303.

Jorgensen, T. (n.d.), 'Ml optimizer', http://www.econ.ku.dk/phdstudent/jorgensen/code.htm.

Kibria, G. B. M. (2003), 'Performance of some new ridge regression estimators', *Communications in Statistics-Simulation and Computation* **32**, 419–435.

Koopman, S. J., Mallee, M. I. P. and Van der Wel, M. (2010), 'Analysing the term structure of interest using the dynamic nelson-sigel model with time-varying parameters', *Journal of Business and Economic Studies* **28**, 329–343.

Lawless, J. and Wang, P. (1976), 'A simulation study of ridge and other regression estimators', *Communications in Statistics-Theory and Methods A* **5**, 307–323.

Levant, J. and Ma, J. (2013), A dynamic nelson-siegel yield curve model with markov switching. Available at https://www.aeaweb.org/aea/2014conference/program/retrieve.php?pdfid=938.

Litterman, R. and Scheinkman, J. (1991), 'Forecasting the term structure of govenrment bond yields', *Journal of Fixed Income* **1**, 54–61.

Liu, K. J. (1993), 'A new class of biased estimate in linear regression', *Communications in Statistics-Theory and Methods* **22**, 393–402.

Liu, K. J. (2003), 'Using liu-type estimator to combat collinearity', *Communications in Statistics-Theory and Methods* **32**, 1009–1020.

Liu, X. and Gao, F. (2011), 'Linearized ridge regression estimator inlinear regression', *Communications in Statistics-Theory and Methods* **40**, 2182–2192.

Liu, X., Gao, F. and Yu, Z.-F. (2013), 'Improved ridge estimators in a linear regression model', *Journal of Applied Statistics* **40**, 209–220.

Lowenstein, R. (2000), *When Genius Failed: The Rise and Fall of Long-Term Capital Management*, Random House, New York.

Marquardt, D. (1970), 'Generalised inverses, ridge regression, biased linear estimation and nonlinear estimation', *Technometrics* **23**, 591–612.

Mason, R. and Gunst, R. (1985), 'Selecting princiapl componnets in regression', *Statistics and Probability Letters* **3**, 299–301.

Mason, R., L., Gunst, R. and Webster, J., T. (1975), 'Regression analysis and problems of multicollinearity', *Communications in Statistics* **4**, 277–292.

Massy, W. F. (1965), 'Principal component regression in exploratory statistical research', *Journal of the American Statistical Associatio* **60**, 234–256.

Mayer, L. and Wilkie, T. (1973), 'On biased estimation in linear models', *Technometrics* **23**, 497–508.

McDonald, D. W. and Galarneau, D., I. (1975), 'A monte carlo evaluation of some ridge-type estimators', *Journal of the American Statistical Association* **70**, 407–416.

Meucci, A. (2009), Review of statistical arbitrage, cointegration and multivariate ornstein-uhlenbeck. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract/id=1404905.

Murphy, K. (2004), 'Kalman filter toolbox for matlab', http://www.cs.ubc.ca/ murphyk/Software/Kalman/kalman.html.

Nelson, C. R. and Siegel, A. F. (1987), 'Parsimonious modelling of yield curves', *The Journal of Business* **60**, 473–489.

Newhouse, J. and Oman, S. D. (1971), 'An evaluation of ridge estimators', Rand Corporation Report. R-716-PR.

Niu, L. and Zeng, G. (2012), The discrete-time framework of the arbitrage-free nelson-siegel class of term structure models. Available at SSRN: http://ssrn.com/abstract=2015858.

Nyholm, K. (2008), *Strategic Asset Allocation in Fixed-Income Markets-A Matlab-Based User's Guide*, Wiley Finance, Chichester, England.

Pardo, R. (2008), *The Evaluation and Optimization of Trading Strategies*, Wiley, Chichester, England.

Rebonato, R., Sukhdeep, M., Joshi, M., L-D., B. and Nyholm, K. (2005), 'Evolving yield curves in the real-world measures: a semi-parametric approach', *The Journal of Risk* **7**(3), 29–61.

Simon, D. (2009), 'Kalman filtering with state constraints: a survey of linear and nonlinear algorithms', *IET Control Theory and Applications* pp. 1–16.

Singh, B., Chaube, T. and Dwivedi, T. (1986), 'An almost unbiased ridge estimator', *Sankhya Ser.B* **48**, 342–346.

Svensson, L. (1994), 'Estimating and interpreting forward interest rates: Sweden 1992-1994', IMF Working Paper. WP/94/114.

Swindel, B. (1976), 'Good ridge estimators based on prior information', *Communications in Statistics-Theory and Methods* **A5**, 1065–1075.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**, 177–188.

Vinvod, H. (1976), 'Applications of ridge regression methods to a study of bell systems scale economies', *Journal of the American Statistical Association* **71**, 835–841.

Waggoner, D. (1996), 'The robustness of yield and forward rate curve extraction methods', Federal Reserve Bank of Atlanta, Working Paper.

Wichern, D. W. and Churchill, G. A. (1978), 'A comparison of ridge estimators', *Technomatrics* **20**, 147–157.

Wiecki, T. (2012), 'Parameter optimization in algorithmic finance part i: Walk forward optimization', http://blog.quantopian.com/parameter-optimization.

Xiang, J. and Zhu, X. (2013), 'A regime switching nelson-siegel term structure model and interest rate forecasts', *Journal of Financial Econometrics* **11**, 522–555.

Xiao, J. Y. (2001), 'Term structure estimation for u.s. corporate bond yields', *RiskMetrics Journal* **2**, 19–34.