

Department of Computer Science
University College London
University of London

Extracting Temporal Dependencies from Geospatial Time Series Data

Sarah Chisholm



Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
at the University of London

March 15, 2016

I, Sarah Chisholm confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright © 2016 by Sarah Chisholm
All rights reserved.

Abstract

In recent years, there have been significant advances in the technology used to collect data on the movement and activity patterns of humans and animals. GPS units, which form the primary source of location data, have become cheaper, more accurate, lighter and less power-hungry, and their accuracy has been further improved with the addition of inertial measurement units. The consequence is a glut of geospatial time series data, recorded at rates that range from one position fix every several hours (to maximise system lifetime) to ten fixes per second (in high dynamic situations). Since data of this quality and volume has only recently become available, the analytical methods to extract behavioural information from raw position data are at an early stage of development. An instance of this lies in the analysis of animal movement patterns. There are, broadly speaking, two types of animals: solitary animals and social animals. In the former case, the timing and location of instances of avoidance and association are important behavioural markers. In the latter case, the identification of periods and strengths of social interaction is a necessary precursor to social network analysis. In this dissertation, we present two novel analytical methods for extracting behavioural information from geospatial time series, one for each case. For solitary animals, a new method to detect avoidance and association between individuals is proposed; unlike existing methods, assumptions about the shape of the territories or the nature of individual movement are not needed. For social individuals, we have made significant progress in developing a method to test for cointegration; this measures the extent to which two non-stationary time series have a stationary linear relationship between them and can be used to assess whether a pair of animals move together. This method has more general application in time series analysis; for example, in financial time series analysis.

Acknowledgements

I would like to extend thanks to the many people, who so generously contributed to the work presented in this thesis.

Special mention goes to my supportive supervisor, Steve Hailes. My PhD has been a challenging experience and I thank Steve wholeheartedly, not only for his tremendous support, but also for giving me a lot of valuable advice along the way.

Similar, profound gratitude goes to Steffen Grünewälder, who has been a truly dedicated mentor. I am particularly indebted to Steffen for his constant support with the mathematical part of my work, and for the encouraging chats we had when I felt overwhelmed.

I am also hugely appreciative to Alan Wilson, for his advice on the ecology side of things.

Special mention goes to Sarah Power, Giulia Di Tomaso, Alessandra Staglianò, Venus Shum and Rae Harbird, for the lunches, coffees, walks and chats that were absolutely necessary for my sanity. To my Capoeira friends, especially Molinha and Zumbi, for keeping my mind off my thesis for at least four hours a week and giving me an amazing amount of energy.

Finally, but by no means least, thanks go to mum, dad, Daniel, Kim and of course Michele for sticking with me and believing in me, throughout a few difficult years. I would have never managed it without you. You are the most important people in my world and I dedicate this thesis to you.

Contents

I	Introduction	12
1.1	Contributions	15
II	Association and Avoidance	17
2	Test of Temporal Associations	18
2.1	Literature Review on Testing for Avoidance	19
2.2	Method	21
2.3	Simulations	27
3	Application of the Associations Test	35
3.1	Leopards	36
3.2	African Wild Dogs	37
3.3	Contributions in the Associations Test	40

III	Cointegration	43
4	Literature Review on Cointegration	44
4.1	Contributions	45
4.2	Mathematical and Statistical Background Information	46
4.3	Literature Review	53
4.4	Frequentist Cointegration	54
4.5	Bayesian Cointegration	56
5	Test of Cointegration	59
5.1	Overview	60
5.2	Model Description	61
5.3	Identification	63
5.4	Prior Distributions	64
5.5	Posterior Probability Density Function	65
5.6	Cointegration Tube	70
5.7	Numerical Integration	75
6	Simulations of Cointegration Test	96
6.1	Data Generating Process	97
6.2	Results	99
6.3	Application	100

7	Discussion on the Cointegration Test	102
7.1	Discussion	103
7.2	Future Work	103
IV	Conclusions	105
V	Appendix	110
A	Results of the Avoidance Test	111
A.1	Tables of Results of Simulation Test	112
B	Proofs to Chapter 4.2	116
B.1	Proof that $\text{Cov}(\epsilon_t, \epsilon_{t-k}) = \frac{\sigma^2 \phi^k}{1-\phi^2}$	117
B.2	Solving Maximisation using Lagrange Multiplier	117
C	Proofs to Chapter 5	119
C.1	Proof that $p(\mathbf{y}_1 \mathbf{y}_2) \neq 0$	120
C.2	Proof of $p(\boldsymbol{\theta} \mathbf{y}_2) = p(\boldsymbol{\theta})$	121
C.3	Proof of $p(\alpha, \beta \boldsymbol{\theta}, \mathbf{y}_2) = p(\alpha, \beta)$	121
C.4	Expansion of $p(\mathbf{E} = \mathbf{y}_1 - \alpha\mathbf{1} - \beta\mathbf{y}_2 \boldsymbol{\theta})$	123
C.5	Computation of $p(\epsilon_t \epsilon_{t-1}, \boldsymbol{\theta})$	127
C.6	Proof that $\mathbb{E}(E_t) = 0 \forall t$	128

C.7	Computation of Σ	128
C.8	Computation of C_2	129
C.9	Computation of μ_α^* and μ_β^*	130
C.10	Computation of $\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix}$	131
C.11	Calculation of Cointegration Relationship	132
C.12	Calculation of $ \text{Var}(E_1) - \text{Var}(E_t) $	134
C.13	Calculation of the bounds on $\ \Sigma(\tilde{\theta}_i) - \Sigma(\theta_i)\ _{\text{OP}}$	134
C.14	Computation of $\prod_{j=1}^T \frac{ \lambda_{i,j} - e_i }{ \lambda_{i,j} } \leq \prod_{j=1}^T \frac{ \tilde{\lambda}_{i,j} }{ \lambda_{i,j} } \leq \prod_{j=1}^T \frac{ \lambda_{i,j} + e_i }{ \lambda_{i,j} }$	135
C.15	Calculation of upper bound on $\ \tilde{\Lambda}_i - \Lambda_i\ $	137
C.16	Proof that Λ^{-1} is Positive Definite	138
C.17	Upper bound on $\left\ \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right\ _{\text{OP}}$	138
C.18	Upper bound on $\left \tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i \right $	139
C.19	Upper bound on $\left \sqrt{\frac{\det \tilde{\Sigma}_i}{\det \Sigma_i}} \right $	142
D	Results of the Cointegration Test	143
D.1	Tables of Results of the Cointegration Test	144
	References	148

List of Figures

2.1	Example of Significant Geographic Locations	26
2.2	Examples of Time Series Plots of Distance Between Leopard Dyads	27
2.3	Example of Observed Movement of Two Neighbouring Leopards	28
2.4	Example of Simulated Movement	29
2.5	Effect of the Association Time on the Test's Accuracy	31
2.6	Effect of the Association Distance and Observation Length	32
2.7	Effect of the Length of the Observation Period When No Association	33
3.1	Examples of Plots Showing the p-values for the Leopard Data	38
3.2	Examples of Time Series Plots of Distance Between Two Packs of Wild Dog	39
3.3	Examples of Plots Showing the p-values for the Wild Dog Data	40
5.1	Posterior Heat Map and Cointegration Tube	62
5.2	Sketch of a Cointegration Confidence Interval	71
5.3	Examples of Cointegrated and Not Cointegrated	73

5.4	Examples of Cointegrated and Not Cointegrated	74
5.5	Flowchart of the Chain of Upper Bounds	77
6.1	Heat Map of Cointegration Test Results	99

List of Tables

5.1	Comparison of Upper Bound with Corners Difference A	94
5.2	Comparison of Upper Bound with Corners Difference B	95
A.1	Effect of the Association Time Inside Association Distance	112
A.2	Effect of the Association Time Outside Association Distance	112
A.3	Effect of the Association Distance Inside the Association Distance	113
A.4	Effect of the Association Distance Outside the Association Distance	113
A.5	Effect of the Observation Length Inside the Association Distance	114
A.6	Effect of the Observation Length Outside the Association Distance	114
A.7	Effect of the Observation Length Without Association	115
D.1	Cointegration Test Results for $\sigma_1^2 = 0.01$	144
D.2	Cointegration Test Results for $\sigma_1^2 = 0.11$	144
D.3	Cointegration Test Results for $\sigma_1^2 = 0.21$	145
D.4	Cointegration Test Results for $\sigma_1^2 = 0.31$	145

D.5	Cointegration Test Results for $\sigma_1^2 = 0.41$	146
D.6	Cointegration Test Results for $\sigma_1^2 = 0.51$	146
D.7	Cointegration Test Results for $\sigma_1^2 = 0.61$	146
D.8	Cointegration Test Results for $\sigma_1^2 = 0.71$ and higher	147

Part I

Introduction

Methods for collecting data on the movement of individuals have advanced dramatically over the last two decades, with GPS and Inertial Measurement Units becoming smaller, lighter, more energy efficient and more accurate than ever before. These developments enable the detailed tracking of multiple individuals over long periods of time. To make the most of these technological advances, methods to analyse large amounts of data efficiently are essential.

The interactions of animals is one area of application in which vast amounts of data are collected, yet efficient methods to analyse them are scarce. There are, roughly speaking, two categories of animal interaction: that of solitary animals and that of social animals. One method for each of these areas is proposed in this thesis.

When analysing the interaction between solitary animals, the quantification of association or avoidance between territorial conspecifics would advance our understanding of animal ecology and, in the long term, the impact of changing environments. Existing forms of such tests are predicated on assumptions about the shape of the individuals' territory and boundaries [48, 19]. Perhaps for that reason, these methods are not often employed, and avoidance is instead inferred from circumstantial evidence.

In this dissertation a method to detect avoidance and association between solitary animals is presented, which does not make assumptions about the shape, size or use of the territory. The method relies purely on the disassociation of individuals by permutations of blocks (e.g. days) of the observed data. It is possible to create up to $N!$ permutations to which to compare the observed data. This enables the assessment of interaction over multiple inter-individual distances, for each of which the observed number of times the individuals were within a particular distance of each other, is compared to the expected number.

Part II starts with a comprehensive discussion of the state of the art in testing for avoidance and association in Section 2.1. The details of the proposed method are presented in Section 2.2 and it is applied to data collected from GPS collars on leopards (*Panthera pardus*) and African wild dogs (*Lycaon pictus*) in Chapter 3. This part concludes with a discussion of

the advantages and limitations of the suggested method in Section 3.3.

When analysing interactions between gregarious animals, social network analysis has received much attention in recent years. For example affiliation networks have been used in [25, 18, 64], in which the individual animals are represented by nodes and an affiliation between a dyad (something that consists of two elements or parts) is suggested by edges/links between the corresponding nodes.

To build such an affiliation network, a measure of association between dyads is needed. This measure depends on the kind of association in question. For example, when investigating the social behaviour of bottlenose dolphins in [47] all members of a school were assumed associated, because inter-school associations were of interest. When considering social dynamics within baboon troops, grooming behaviour and spatial affiliation were analysed in [32] and shown to give the same outcomes. However, when considering other animals the social interaction might not be as obvious, for example sheep, or fish. In these cases, a characteristic called cointegration could be used to find relationships between the movement of dyads.

Time series are said to be stationary when the time series has a constant mean, constant variance and time-independent covariance. This is a strong restriction on animal movement, which is not always fulfilled [53]. When testing for dependencies between two or more time series the Pearson product-moment correlation coefficient is often one of the first statistics considered. However, correlation only makes sense if the individual series are stationary and the relationship between the two time series is linear [14]. In practice there are many time series that are non-stationary, such as the aggressive communication of the hermit crab *Calcinus tibicen* [53], exchange rates [41], population and employment [36], electricity consumption [2], gas prices [27], maize prices [1], dissent rates on the High Court of Australia [51] and hemispheric temperature [42]. In these cases, a characteristic called cointegration could be used to find relationships between non-stationary time series. Cointegration describes a property of two or more time series that are individually

non-stationary but for which a linear relationship of the time series is stationary.

Part III starts with a detailed literature review on cointegration in Chapter 4, with some background information on stationarity, Bayesian statistics and norms (needed for the understanding of the suggested method). The proposed method is detailed in Chapter 5 and applied to synthetic data in Chapter 6. This part is concluded with a discussion of the advantages and limitations of the proposed method (see Chapter 7).

The dissertation is concluded with a discussion of the introduced methods in Part IV.

1.1 Contributions

In this dissertation, two novel analytical methods for extracting behavioural information from geospatial time series are presented, one for solitary and one for social animals. For solitary animals, a new method to detect avoidance and association between individuals is proposed. This method creates permutations of the observed data to disassociate dyads from each other. As a consequence, it is possible to create up to $N!$ permutations to which to compare the observed data. This enables the assessment of interaction over multiple inter-individual distances, for each of which the observed number of times the individuals were within that distance of each other, is compared to the expected number of times.

For gregarious individuals, a method to test for cointegration has been developed; this measures the extent to which two non-stationary time series have a stationary linear relationship between them and can be used to assess whether a dyad of animals move together. A novel method of testing for cointegration is proposed in this dissertation. To the best of our knowledge, this is the first method that tests whether the observed time series are cointegrated, rather than whether they converge towards being cointegrated at a future point in time.

This is done in three steps. First the posterior probability distribution of the cointegration

parameters is calculated analytically. Simultaneously, a cointegration tube is estimated, which describes the combinations of the relevant parameters that suggest cointegration. Finally the posterior probability distribution is integrated over the cointegration tube to ascertain what proportion of the posterior probability distribution lies within the cointegration tube, and therefore suggests cointegration.

Part II

Association and Avoidance

Chapter 2

Test of Temporal Associations

2.1 Literature Review on Testing for Avoidance

In this chapter, we introduce a general statistical framework to establish whether individuals or groups are more or less often in close proximity of each other than expected by chance. Existing forms of such tests are predicated on assumptions about the shape of the individuals' territory and boundaries [48, 19]. Perhaps for that reason, these methods are often not employed, and avoidance is instead inferred from circumstantial evidence. For example, Jackson *et al.* comment in [38] that *'judging by the intensity of use of core areas, the large amount of overlap among individuals, and the relatively small total home areas, it is remarkable that the tagged cats managed to remain on average > 2km apart. This implies that the Langu cats actively avoided one another, while sharing the same area'*. There is no explanation as to why an average of 2km could not have occurred purely by chance; rather, an absence of contacts is seen as evidence of active avoidance.

A test for dynamic interactions was first suggested by MacDonald *et al.* [48]; this is based on the application of a quadrivariate normal distribution to the co-ordinates of the two target individuals. Dunn describes a similar approach that employs a multivariate Ornstein-Uhlenbeck model rather than a multivariate normal model [19]. Both tests require that the utilization of each range is distributed about a single centre of activity and violation of this assumption, which has no obvious biological basis, can produce large errors [17].

Delgado *et al.* [16] proposed a functional response in which social behaviour is assumed to depend on proximity to other individuals. As detailed by the authors, the null model is supposed to account for all factors influencing movement behaviour apart from conspecifics. In their method they suggest a null model that is calculated from movements in a totally random direction with the same step length as the observed movement. This means that specific habitat areas with higher or lower chances of being visited have to be specifically incorporated into the null model. As an example a particularly dense areas of the habitat might be uncomfortable, or an area without any hiding possibilities. If not

included in the model, these areas could increase false positive results, because both the focal individual, as well as conspecifics might be more likely to be in a particular area of their habitat and therefore the observed distances between them would be smaller than expected by chance.

Elbroch *et al.* [20] suggested a generalized linear model to test for predictive power of a number of factors on the number of spatial associations observed. These factors included number of elk in the study area and mean genetic relatedness between interacting individuals. This is an interesting approach to understand what factors influence associations, yet does not easily extend to testing whether individuals avoid each other or seek each others' proximity.

Doncaster suggested the first non-parametric test in [17]. This compares the empirical distribution function of the N paired separations with that of the complete set of N^2 separations. For this, a critical separation is chosen within which presence of dynamic interaction is of interest. However, the correct value of this separation might not be easy to estimate and N would have to be very large to permit an analysis over multiple different separations. Furthermore, the significance test depends on the independence of successive data points and is only valid for fixed ranges [17].

In this chapter, we propose a method that creates perturbations of blocks (e.g. days) of the observed data to obviate the need for independence between measurements. As a consequence, it is possible to create up to $D!$ (where D is the number of blocks in the observation period) permutations to which to compare the observed data. This enables the assessment of interaction over multiple inter-individual distances, for each of which the observed number of times the individuals were within that distance of each other, is compared to the expected number of times.

One of the main advantages of this method is that specific geographic areas that are visited less or more often by the individuals do not have to be included manually. Instead they are automatically considered in this approach, since the frequency of each location is exactly

the same for the null model (the permutations) as it is for the observed movement.

2.2 Method

Dynamic interactions can be measured in two ways, as defined by Doncaster in [17]. The first is termed ‘*static interaction*’, which describes a spatial overlap of home ranges. The second characterises dependencies between individuals’ movements. This study examines the latter, ‘*dynamic interactions*’.

As Doncaster describes in [17] ‘*Dependency in the movements of two individuals (dynamic interaction) [...] can be expressed in terms of probability. Are the animals more likely to maintain a certain separation (positive dynamic interaction) or less likely (negative dynamic interaction) than is expected from the configuration and utilization of their ranges? At small separations in particular, does there exist a bond of attraction between them or do they respond to close contact by mutual repulsion?*’

The method described in this chapter does not assume any underlying distribution, nor a particular shape or usage of the individuals’ territories. It does not require independence of consecutive measurements, nor a constant time difference between the measurements. This test simply relies on the disassociation of the target individuals by using permutations.

The observation period is divided into blocks, such as days, or weeks. These blocks are then permuted to obtain the expected distances, if the two individuals/groups did not know of each other’s whereabouts. If days are permuted, this would imply that the individuals are equally likely to be at any location at, say, two in the afternoon as any other location that individual was at two in the afternoon during the observation period. If weeks are permuted, it would mean that at, say 2pm on a Monday, the permuted individual is as likely to be anywhere that individual has been at 2pm on a Monday during the observation period.

Following is a detailed description of the steps taken to test for association or avoidance using the proposed method:

First, the distance between two individuals is calculated for each measurement. The blocks (this could be days or weeks) are then permuted randomly 10,000 times, for one of the target individuals. Keeping blocks of days or weeks ensures that the diel/weekly movement stays intact in the permutations. The distance between the target individuals is calculated again, for each permutation.

The number of times the target individuals are within previously defined distances of each other is computed for the observed data as well as for each of the permuted sets. In Section 3.1, the application of this method to data collected from leopards, the intervals are chosen to be 0 – 20m, 20 – 40m, 40 – 80m, 80 – 160m, 160 – 320m and 320 – 640m. For each of these intervals, each permutation gives a count for that dyad. These 10,000 counts are the distribution of the null hypothesis, i.e. for the case in which movement patterns are independent of one another.

The null hypothesis is that there is no difference between the number of times the target individuals are within a certain distance interval in the permuted and observed time series. The two alternatives are that the individuals are (i) more often and (ii) less often in the distance examined than expected from the permutations.

A p-value is defined to be the probability of obtaining a result at least as extreme as the one that was observed, assuming that the null hypothesis is true [28]. Therefore the p-value in this case is the upper bound¹ on the proportion of permutations as extreme as the observation, or more extreme. Say the two target individuals were observed to be between 20 and 40m of each other five times, then the p-value for the null hypothesis vs the alternative that the individuals are less often between 20 and 40m of each other than

¹as we only have a sample of all possible scenarios

expected by chance lies between:

$$\frac{n_1}{n_{perm}} \leq p < \frac{n_1 + 1}{n_{perm}}$$

where n_{perm} is the number of permutations calculated and n_1 is the number of permutations in which the dyad was within that distance at most five times. The observed number of times the individuals were within that particular interval is then compared to the distribution created by the permutations. Observations lying in the $\frac{5}{2k}\%$ tail of the permutations will be regarded as evidence that the target individuals were less often in the distance interval than expected by chance. This percentage is calculated using the Bonferroni correction [50]; the 5% represent the significance level, which has to be divided by $2k$, where k is the number of distances tested and we test k distances for avoidance and k distances for association. The Bonferroni inequality balances out the effect of multiple testing.

If the p-value of a distance tested for is less than $0.05/2k$ this indicates that there is strong evidence against the null-hypothesis that the two individuals are within that particular distance as often as expected by chance. If this occurs when testing whether the individuals are more often within that particular distance, this would indicate that the individuals seek each others proximity. If it occurs when testing whether they are less often within the particular distance, this would indicate that the individuals avoid each other.

The pseudo-code for this method is as follows. The inputs are two $n \times 4$ matrices, y_1 and y_2 , an integer, $perm$, and a vector, $dist$. Each of the matrices corresponds to one individual. The first and second columns of each matrix should be the date and the time of the observation respectively, and the third and fourth columns are expected be the x and y coordinate of the individual. The integer should define how many permutations should be done, the default value is set to 10,000. The vector, $dist$, lists the distances that should be tested for association and avoidance, for example if $dist = (20, 40, 80, 160)$ then the algorithm will test whether the individuals are more or less often within 0-20m,

20-40m, 40-80m and 80-160m of each other. The output is a vector of distances that the individuals were observed to be more often and a vector of distances within which they were observed to be less often than expected by chance.

Algorithm 1 Pseudo-code for proposed method to test for avoidance and association

```

1: procedure AVOIDANCETEST( $v_1, v_2, perm \leftarrow 10,000, dist$ )
2:    $DistObs \leftarrow \sqrt{(v_1(:,3) - v_2(:,3))^2 + (v_1(:,4) - v_2(:,4))^2}$ 
                                      $\triangleright$  calculate distance between individuals
3:    $CountObs(1,1) \leftarrow sum(DistObs \leq dist(1))$ 
4:   for  $j_1$  from 2 to  $length(dist)$  do
5:      $CountObs(1, j_1) \leftarrow sum(DistObs \geq dist(j_1 - 1) DistObs \leq dist(j_1))$ 
                                      $\triangleright$  Count how often individuals were within  $dist$  of each other
6:   for  $i$  from 1 to  $perm$  do
                                      $\triangleright$  create null distribution through permutations
7:      $Days \leftarrow unique(v_2(:,1))$ 
                                      $\triangleright$  Get a vector of the days
8:      $v_{perm}(:,1) \leftarrow permute(Days)$ 
                                      $\triangleright$  Permute the days
9:      $v_{perm}(:,2:4) \leftarrow (v_2(:,2:4))$ 
                                      $\triangleright$  Add the time and location columns to the permuted matrix
10:     $v_{perm} \leftarrow sort(v_{perm}, \text{by days})$ 
                                      $\triangleright$  Sort the matrix by the order of the days
11:     $DistPerm \leftarrow \sqrt{(v_1(:,3) - v_{perm}(:,3))^2 + (v_1(:,4) - v_{perm}(:,4))^2}$ 
                                      $\triangleright$  Calculate Distance between original individual 1 and permuted individual 2
12:     $CountPerm(i,1) \leftarrow sum(DistPerm \leq dist(1))$ 
                                      $\triangleright$  Count how often permuted individuals were within  $dist$  of each other
13:    for  $j_2$  from 2 to  $length(dist)$  do
14:       $CountPerm(i, j_2) \leftarrow sum((DistPerm \geq dist(j_2 - 1))(DistPerm \leq dist(j_2)))$ 
15:    for  $j_3$  from 1 to  $length(dist)$  do
                                      $\triangleright$  Compare observed count for each distance with the distribution of the permuted data
16:       $p(j_3, 1) \leftarrow 2 \cdot length(dist) \cdot sum(CountPerm(:, j_3) \leq CountObs(1, j_3)) / perm$ 
                                      $\triangleright$  Calculate the p-value for each distance for avoidance
17:       $p(j_3, 1) \leftarrow 2 \cdot length(dist) \cdot sum(CountPerm(:, j_3) \geq CountObs(1, j_3)) / perm$ 
                                      $\triangleright$  Calculate the p-value for each distance for association
18:  return  $p$ 

```

This approach has the added benefit of ensuring that the same physical locations are visited at the same frequencies as the observed individuals and, because intra-day or week movement is unaffected, that diel or weekly patterns of movement and rest are maintained.

2.2.1 Graphical and Seasonal Associations

As Doncaster mentions in [17], ‘*A positive component is likely to arise particularly when the two animals have separate resting sites at which they regularly begin and end their cycles of activity.*’ The implication of Doncaster’s statement is that colocation, when associated with a geographic point, is of a different nature to colocation in featureless areas because it could occur by chance. The technique we propose here directly discounts chance interactions of this form. Assume that two individuals meet regularly at a waterhole each morning. Since we do not disrupt the diurnal cycle in permuting days, those individuals will meet regularly at that waterhole in the permuted time series as well. Consequently, to establish that a statistically significant interaction occurred in the observed data, the number of occurrences of observed colocation would need to be very high; much higher than might be accounted for by the number that occur by chance alone. Conversely, in areas in which few meetings occur by chance, a smaller number of meetings will be considered significant.

Doncaster’s comments, do however, point towards a need for care in the application of our test. If the waterhole is available only for part of the year, and permutations occur across the entire year, then the interactions could appear to be significant. Consequently, it is important to ensure that permutations occur only between days/weeks that are equivalent. For example, permuting days that have very different seasonal features may well lead to spurious results.

Such seasonal features could include seasons of drought, where for example a waterhole that both individuals generally use does not exist. Seasons of very high rain fall could also change the behavioural pattern, for example by forcing one or both of the individuals to find a different resting site. Such examples illustrate when care should be taken.

To investigate such dependencies in our case, the locations of the dyads that were significantly more often colocated were plotted (see Figure 2.1 for an example). To examine

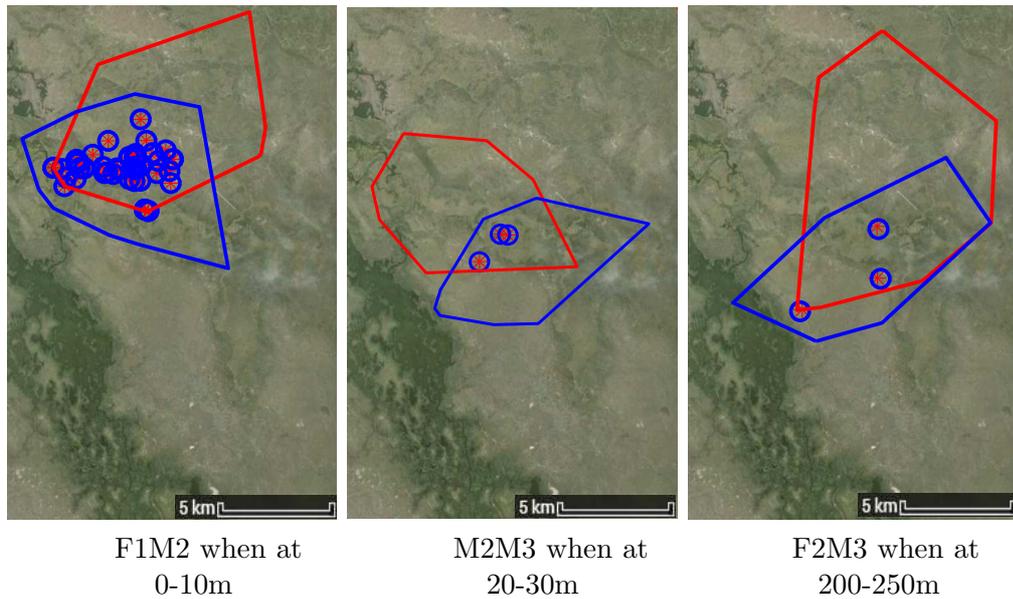


Figure 2.1: This figure shows examples of the geographic locations at which leopards are found when separated by a statistically significant distance. An estimate of the extent of each leopard's territory is given using the convex hull of the full set of observations.

possible seasonal effects, the distances between each dyad were plotted over time (see Figures 2.2 and 3.2 for examples). From these graphs it can be seen that there is no particular seasonal clustering of the small number of observed colocations.

When there are observation periods in which the location of the two observed individuals is not known in enough detail (when at least one of the two individuals' locations is recorded less than every six hours), that period is excluded from the analysis. This is shown in the time series plots, 2.2 and 3.2, by periods of missing data, such as that towards the beginning of plot M3M6 in Figure 2.2.

Other confounding factors, such as two individuals following a third conspecific or heterospecific that has not been fitted with a GPS collar, cannot be ruled out as possible explanations for an observed relationship. But this is simply a restatement of the truism that correlation and causation are different and that causations can generally not be tested for without a randomised experiment, which is not possible in observational studies.

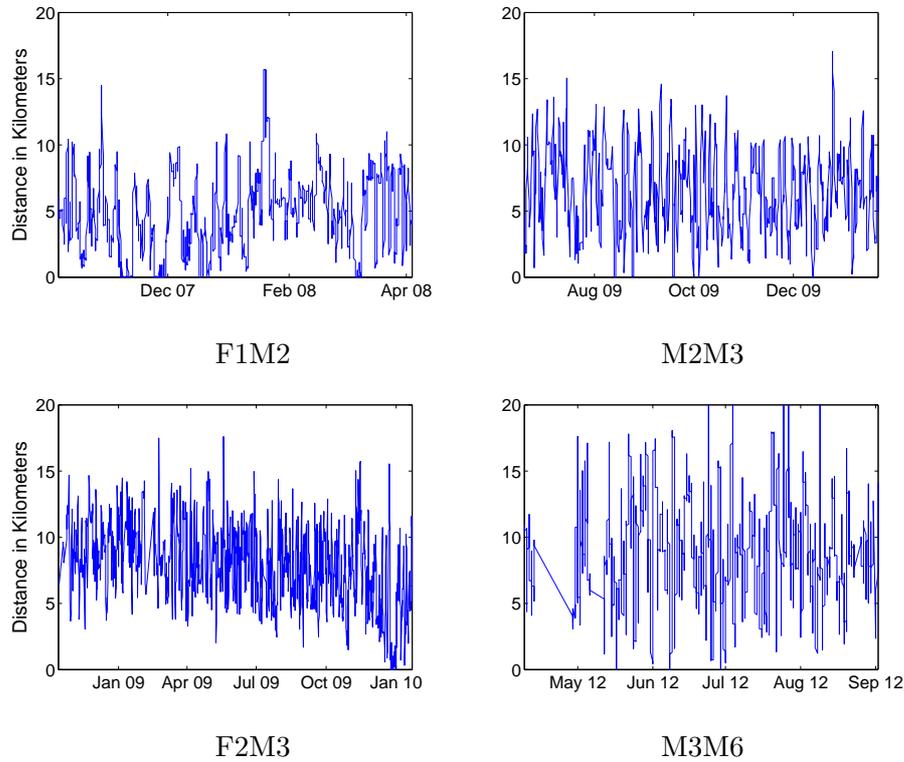


Figure 2.2: This figure shows examples of the time series plots referred to in Section 3.1. They show the distance between individuals in a dyad over time.

2.3 Simulations

To demonstrate the accuracy of this method, neighbouring leopard movements were simulated, with associations as well as without associations. The test was applied to these simulated movements and the proportion of correctly identified associations and non-associations were calculated.

Movements were simulated using simple random walk processes. These processes are defined by an equal probability of moving in any direction at each step (the direction is chosen from a uniform distribution on the range from 0 to 2π). The shape of the territory was assumed to be elliptical with radii of 9km and 4km. These parameters are roughly estimated from the observations collected on one of the leopards (randomly selected) used in Section 3.1. Figure 2.3 shows the areas visited by this individual (blue solid line) and

its neighbour (red dashed line) during the observation period.

The territories of both the simulated movements have the same size and shape. The second territory is shifted by 6km along the axis of the minor radius. Therefore the overlap is similar to that of the observed individuals. An example of the simulations is given in Figure 2.4. It is clear that the simulations are very different to the observed movement patterns, however mimicking the movement of the leopards is not the aim of these simulations. The method should be capable of finding associations between any sort of movement, as long as there are no major changes in the movement during the observation period.

10,080 simulations were run without any association between the dyad and 10,080 simulations in which the individuals seek each other's proximity if they are within a certain distance of each other (from here on referred to as the outer association distance). In the simulations in which there is no association between the individuals, both processes are simple random walks with elliptical boundaries. In the cases with association the general movement is again a simple random walk except when they are within the outer association distance of each other. In that case, they move directly towards each other and stay at one unit less than half the association distance from each other (this will be referred to as the inner association distance). The individuals stay within that distance of each other between 1 and 5 steps (this number of steps they stay within the inner association distance will be referred to as the association time).

Observed movements of neighbouring leopards

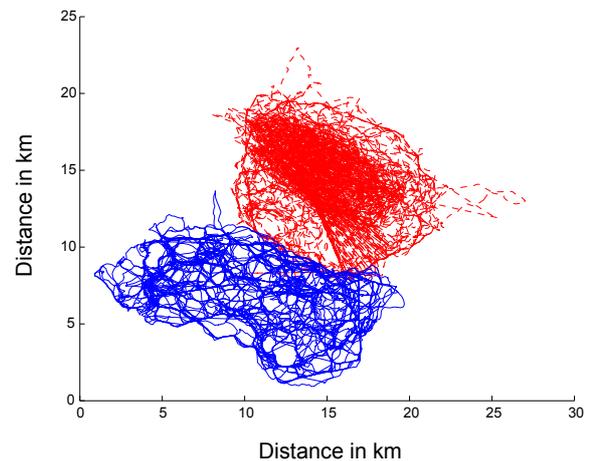


Figure 2.3: Movement of the individuals used for approximate territory size/shape/overlap and step size distribution.

Within the 10,080 simulations with associations, the association time is varied from 1 to 5 steps and the outer association distance is varied from 250 to 500 metres. For both the simulations with association as well as the simulations without association the observation period is varied from 100 to 350 days. For each of the simulations the method proposed in this thesis is applied and it is documented whether the method correctly detects an association in the distances containing the inner association distance, or correctly detects that there is no association.

For each simulation four distances are tested for associations. In the association scenario, one of these distances contains the inner association distance and should therefore test positive for a ‘more than association’, i.e. the individuals are more often within those distances than expected by chance. The distance that contains the outer association distance should test positive for a ‘less than association’, i.e. the individuals are less often within those distances, since they move straight to the

inner association distance, as soon as they are within the outer association distance. Simultaneously, the distance containing the inner association distance should test negative for a ‘less than association’ and the distance containing the outer association distance should test negative for a ‘more than association’.

The other two distances tested are greater than the outer association distance and should therefore test negative for both the ‘more than association’ as well as for the ‘less than association’. In the no association scenario, all of the distances should test negative for both the ‘more than association’ as well as for the ‘less than association’.

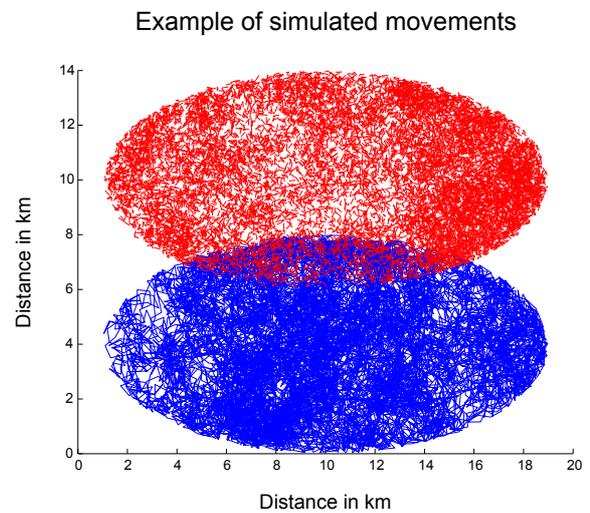


Figure 2.4: This graph shows an example of the simulations.

This means that the total number of distances tested for an association in each of the two scenarios is 40,320 ($= 4 \times 10,080$). In the scenario with association 30,240 distances should test negative for a ‘less than association’ (NLT - correctly identified as not being within those distances less often than expected by chance), whereas 10,080 should theoretically test positive for a ‘less than association’ (PLT - correctly identified as being less often in those distances). Simultaneously 30,240 should test negative for a ‘more than association’ (NMT - correctly identified as not being more often within those distances than expected), and 10,080 should test positive for a ‘more than association’ (PMT - correctly identified as being more often with those distances). However, for the association to present itself, the two individuals need to move within the outer association distance of each other by chance. Therefore, the number of distances that should test positive for a ‘less than’ or a ‘more than association’ is below 10,080. When there is no association, all 40,320 distances should test negative for a ‘more than association’ (NMT) and negative for a ‘less than association’ (NLT).

In the following two sections the results are detailed. First the results of the scenario in which there is an association is described in Section 2.3.1. This is followed by a discussion of the results of the scenario in which there is no association (Section 2.3.2).

2.3.1 Association Scenario

This section discusses the results in the scenario in which there is an association between the two simulated individuals. Overall, out of the 30,240 NLT distances tested, 27,899 (92%) were correctly identified as not being less often within close proximity of each other than expected by chance. Out of the 8,975 PLT distances tested, 39 (0%) were correctly identified as being less often within close proximity of each other than expected by chance. Out of the 30,240 NMT distances 29,761 (98%) were correctly identified as not being more often within close proximity of each other. And, out of the 8,975 PMT distances 7,250 (81%) were correctly identified as being more often within close proximity of each other

than expected.

Figure 2.5 shows the results broken down by the association time, i.e. by how many steps the individuals stay within the inner association distance of each other, before they go back to a random walk. The results are detailed in Tables A.1 and A.2 in the Appendix, with the number of simulations and correctly identified distances listed.

The NLT and NMT results are continuously very high, with false positives in less than 13% of cases. This demonstrates, that the test very rarely suggests that there is an association, when there is none. The NLT and NMT results should not be affected by the time spent within the inner association distance, since they check the distances that are outside the association distances. This can be seen in the results, as the NLT and NMT lines are close to horizontal.

The general upward trend of the PMT line is to be expected. The more time the individuals spend within a certain distance of each other, the higher the likelihood of the test detecting an association. The PLT is consistently around 0% suggesting that none of the individuals were detected as being within close proximity of each other more often than would be expected by chance.

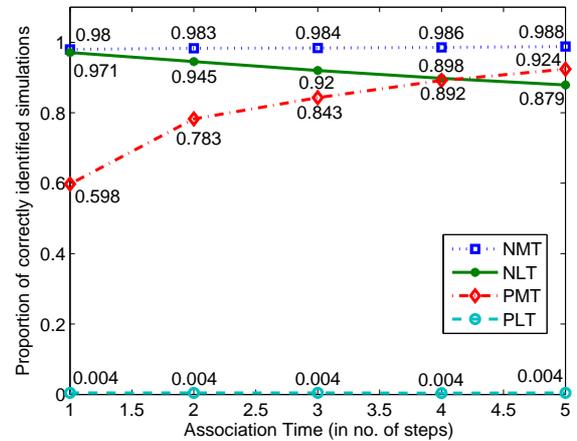


Figure 2.5: This figure details how many of the simulations were correctly classified as having a NLT association (correctly identified no less-than association), a NMT association (correctly identified no more-than association), a PLT association (correctly identified less-than association), and a PMT association (correctly identified more-than association) depending on the number of time steps (x-axis) spent in the inner association distance of each other.

This is most likely due to the fact that the ‘less than’ association is very weak. This association is a by-product of the ‘more than’ association built into the simulations. When the individuals are within the outer association distance of each other they move directly into the inner association distance of each other. Therefore it is expected that the individuals are less often in the outer association distance of each other. However, the step size of the individuals movement is 300m and the distance between the inner and outer association distance varies between 125m and 250m. Therefore the individuals would be very likely to move out of the outer association distance by chance.

The breakdown of the results with respect to the varying sizes of the association distance and lengths of observation period are presented in Figure 2.6 (i) and (ii) respectively. The absolute number of simulations and number of correctly identified simulations are listed in Tables A.3, A.4, A.5 and A.6 in the Appendix.

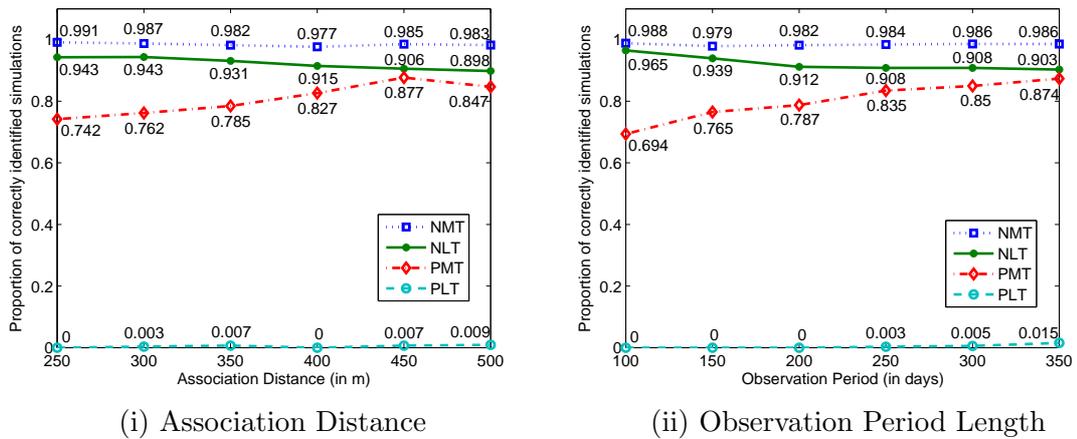


Figure 2.6: This figure shows how many of the simulations were correctly classified as having a NLT, NMT, PLT, and PMT association depending on (i) the size of the association distance and (ii) the length of the observation period.

The results are very similar to those detailed by the association time. False positive results (NMT and NLT) were suggested in less than 11% of the cases. The PMT results show that with a larger association distance, or with a longer observation period, the test is more likely to detect an association correctly. This is probably the case, because the individuals are more likely to be within the outer association distance of each other during

the simulations and therefore show an association more often than in the cases where the association distance is small or the observation period is short.

2.3.2 No Association Scenario

This section discusses the results in the scenario in which there is no association between the two individuals. In this case all 40,320 distances tested should indicate that the individuals were not less often within the distances tested of each other (NLT) and that they were not more often within those distances of each other (NMT).

Overall, out of the 40,320 NLT tests 39,915 (99%) were correctly identified as not being less often within the distances tested of each other than expected. And out of the 40,320 NMT tests 39,455 (98%) were correctly identified as not being more often within those distances of each other than expected by chance.

The break down of the results with respect to the length of observation period is presented in Figure 2.7 and detailed, including absolute number of distances tested and number of correctly identified significant distances, in Table A.7 in the Appendix.

The results are consistently high, with accuracies between 97% and 99%. This suggests that the method has a small type I error (false positive) between 1% and 3%.

In summary, the results demonstrate that false positives (type I error) are rare, which

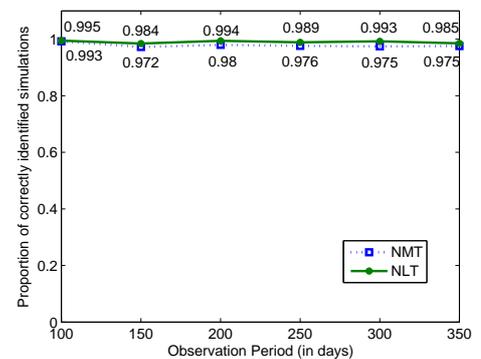


Figure 2.7: Observation Period Length Effect for no Association: This figure details how many of the simulations were correctly classified as having a negative less than (NLT) association and a negative more than (NMT) association depending on the length of the observation period.

means that the test rarely suggests that there is an association if there is none. When there is no association, between 88% and 99% of cases are correctly identified as not having an association. The results for PLT and PMT, i.e. the correct identification of an association, are consistently lower than the results for NLT and NMT (the correct identification of no association). We believe that a small false positive result is more important than a small false negative result, i.e. suggesting that the individuals do not show an association when there is one, is favoured over suggesting that the individuals show an association when there is none.

Chapter 3

Application of the Associations

Test

The method described in Chapter 2 has been applied to location data collected from eight resident neighbouring leopards and eight packs of African wild dogs in Northern Botswana. In the former case, each of the leopards was fitted with a GPS collar and the aim of the test was to identify whether the individual leopards avoided each other. For African wild dogs, one individual in each pack was fitted with a GPS collar. The purpose in this case was to determine whether neighbouring packs avoided each other or sought each other's proximity.

3.1 Leopards

Between 2007 and 2012 two female and six male leopards were fitted with GPS collars. Not all collars were fitted for the entire study period, therefore only periods of simultaneous tagging were used. The number of days each dyad was simultaneously tagged varied between 119 and 406. Locations were measured at least four times a day. As leopards are generally active at night and least active in the middle of the day [5] a day was considered to run from midday to midday for the purposes of permutation.

None of the dyads were shown to spend less time within close proximity of each other than they would by chance. This observation conflicts with the conclusions of studies suggesting that male leopards dynamically avoid one another [59, 38, 35] to reduce the likelihood of violent or fatal conflicts [11, 5]. The data demonstrate that not only do these leopards not actively avoid one another, they have no need to do so since they are highly unlikely to encounter one another by chance.

As expected, two of the six male-female dyads (F1M2 and F2M3) were significantly more often in close proximity to each other (F1M2: 0-160m and F2M3: 0-80m, 160-640m of each other) than expected by chance¹. This is most likely due to courtship and mating [5].

¹The following intervals were tested: 0 – 20m, 20 – 40m, 40 – 80m, 80 – 160m, 160 – 320m and 320 – 640m

More surprisingly, two of the five male-male dyads (M2M3 and M3M6) were highly significantly more often in close proximity of each other (both M2M3 and M3M6 between 0-80m of each other). The individuals in both of these dyads happen to be of similar size and similar weight and are in their prime². M2 and M3 are also the only males that were shown to be significantly more often in close proximity of the two females. Unfortunately M6 was only collared simultaneously with M3, so associations between M6 and the females, or M6 and M2 could not be tested.

None of the location plots (see Figure 2.1) or time series plots (see Figure 2.2) showed any particular geographic location as being the source of the significant proximities, and there was no obvious seasonal pattern (such as might happen if both leopards had a certain pattern of moving around their territories and would therefore be more likely to have a certain distance from each other more often than by chance).

For each dyad, the p-values per distance were plotted and the four plots belonging to F1M2, F2M3, M2M3 and M3M6 are shown in Figure 3.1.

3.2 African Wild Dogs

The eight packs of African wild dogs were collared between May 2011 and May 2014. As for the leopards, the wild dog packs were not all collared simultaneously. Therefore packs were only considered if they were tagged simultaneously for at least 100 days. The resulting 14 neighbouring pack dyads were collared for between 108 and 402 days.

African wild dogs are territorial and have mean annual home ranges of 739 square kilometres at this study site³, with the ranges of neighbouring packs overlapping. None of the dyads showed any significant distance patterns, neither being less often, nor more often in

²It can not be ruled out that they are related

³Unpublished: **Pomilia, M.A., McNutt, J.W. and Jordan, N.R.** Ecological predictors of African wild dog ranging patterns in northern Botswana.

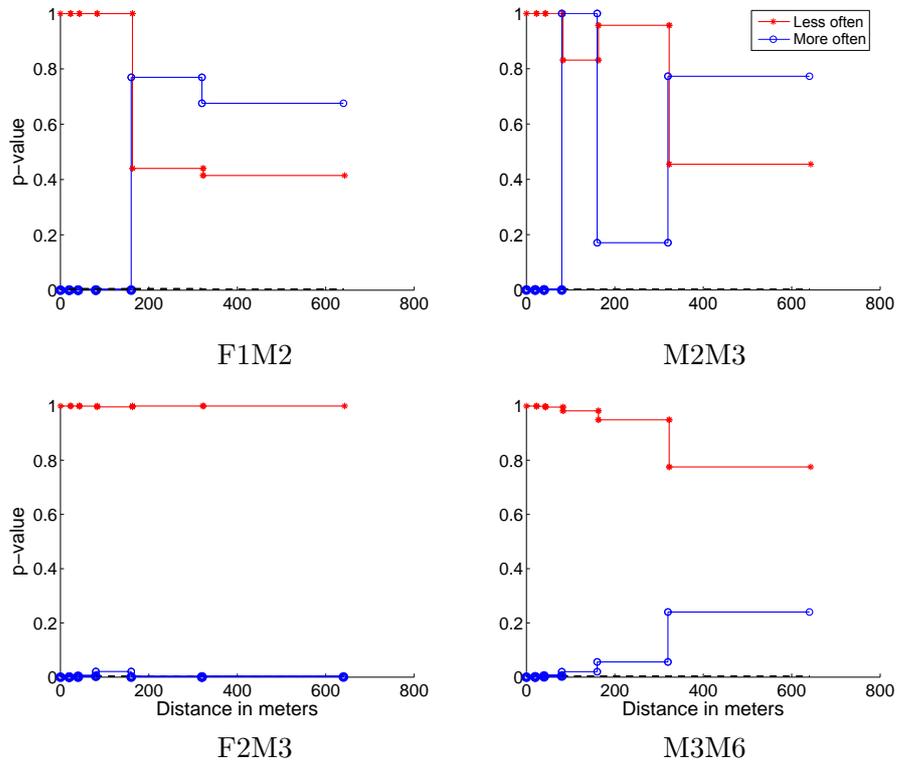


Figure 3.1: This figure shows four examples of the p-value plots mentioned in Section 3.1. The red line with stars gives p-values associated with the alternative hypothesis that the dyad is found less often within a given range of distances than would be expected by chance. The blue line with circles gives p-values associated with the hypothesis that dyads are found within a range more often than would be expected by chance. The black dashed line gives the level under which the results are assumed to be significant.

close proximity to each other over the distances measured⁴. Three of the p-value Figures, as described in Section 3.1, are shown in Figure 3.3.

In general, our results support the finding of previous work on mutual avoidance/attraction between neighbouring wild dog packs [49]. As previous data was acquired by VHF tracking collars, it was limited to relatively few near-simultaneous locations of neighbouring packs acquired by physically tracking the animals [49]. However, despite significant overlap between their ranges (ca. 35%; [57]), observed packs were seen to meet very rarely but it was not possible to determine whether this occurred by active avoidance or by random movement.

⁴The distance intervals used are 0-500m, 500m-1km, 1-1.5km, 1.5-2km

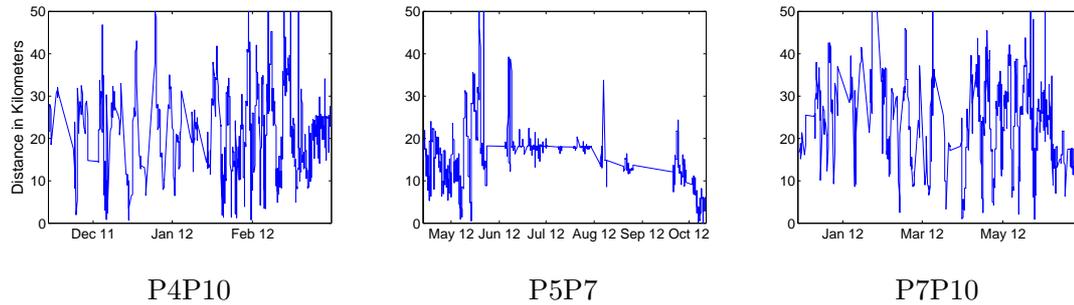


Figure 3.2: Examples of time series plots of distance between wild dog packs: This figure shows examples of the time series plots referred to in Section 3.2. They show the distance between packs in a dyad over time. The straight lines that are close to horizontal are due to the two individuals not being tagged simultaneously, in which case that period was not included in the analysis. This is particularly visible in (b).

In our study, using larger volumes of data acquired remotely using GPS radiocollars, we found no evidence of active spatial avoidance or association between neighbouring packs. As can be seen from the p-value plots in Figure 3.3, our close proximity counts could have happened by chance alone at all measured distances. Spatial interactions (though not necessarily direct interactions) at our measured scales were no more or less likely to occur than would be expected by chance. In fact our data suggest that only one dyad ever came within 250m of one another, suggesting that direct physical encounters were rare.

Although it is not yet clear by what mechanism wild dogs establish and maintain territories, there is strong evidence they do so based on chemical signalling using scent marks [37, 40]. It is possible that scent, which can be encountered without being simultaneously collocated, holds sufficient information to indicate the continued presence of a neighbouring pack and so may reduce the frequency and cost/benefit ratio of direct encounters. It would therefore be of great interest to investigate the temporal association/avoidance in more detail, particularly delayed association/avoidance (visiting areas in which another pack has recently been) of neighbouring packs, and indeed to assess the responses of wild dogs to direct and indirect (olfactory) inter-pack encounters.

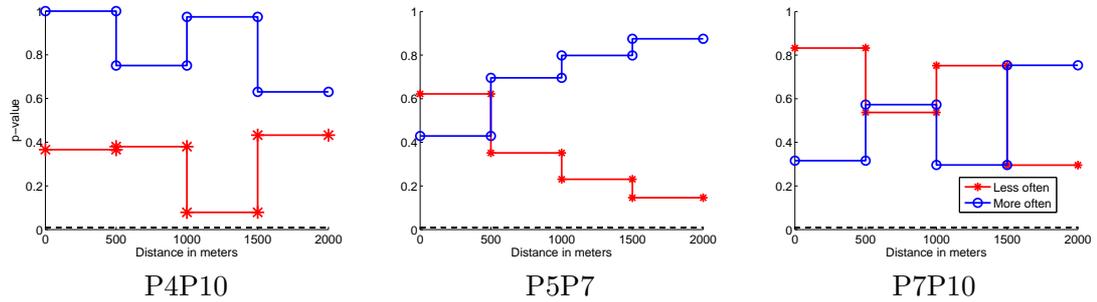


Figure 3.3: This figure shows four examples of the p-value plots mentioned in Section 3.2. The red line with stars gives p-values associated with the alternative hypothesis that the dyad is found less often within a given range of distances than would be expected by chance. The blue line with circles gives p-values associated with the hypothesis that dyads are found within a range more often than would be expected by chance. The black dotted line (not visible here, because it is so close to the x-axis) gives the level under which the results are assumed to be significant.

3.3 Contributions in the Associations Test

The rate of growth in the availability of GPS data from free-ranging animals has not been matched by progress in the development of mathematical techniques for analysing that data. In this thesis, a new method for detecting avoidance and association is presented. Unlike previous work, the method makes no assumption about the shape or size of the territories, nor about the way that individuals move. It relies purely on the disassociation of the individuals' movement through permutations.

Amongst other things, this new method permits the analysis of territorial behaviour in animals. When trying to detect avoidance and association, a simple heuristic would be to define a distance and time and to assume avoidance if the two observed individuals are never within that distance of each other for at least the defined amount of time. Choosing the relevant combination of distance and time is crucial, yet not well defined. The results can change dramatically when a different distance and time is chosen. The method proposed in Section 2.2 gives a simple way to calculate the distance/time combination from the observed movement and territory used for each dyad.

Both the presence and absence of positive spatial association between individuals or groups

are biologically interesting phenomena. In Chapter 3, the method was applied to data collected from GPS collars on individual leopards in which significant positive association was established between some male-male as well as male-female leopard dyads, and to African wild dogs, in which there was no significant dynamic interaction between packs.

For the leopards, two out of six male-female dyads were more often within close proximity of each other than would be expected by chance. This is most likely related to courtship and mating, and supports biological expectations. Interestingly, it was also shown that two out of five male-male dyads were more often within close proximity of each other. This observation is in opposition to conclusions from previous work [59, 35], but could be due to mutual evaluation, family relationships, or a range of unknown factors.

None of the wild dog packs were more or less often within close proximity of each other than would be expected by chance. It is possible that, although the movement patterns of individual packs bring neighbours into relatively close proximity, the risk and occurrence of direct encounters may be reduced by remote inter-pack information exchange, probably via fresh scent signals in these areas.

In using this method, it is important to ensure that data from both members of the dyad are as closely matched in time as possible in order to allow robust conclusions to be drawn on any spatial interactions between them. Temporal differences between compared locations within dyads do not preclude the use of the method, but in such circumstances it is essential to temper conclusions accordingly.

More generally, our method for movement and associations could be applied to epidemiological concerns. If individuals are more often within close proximity of each other than expected by chance, the transmission rate of diseases would be higher than that estimated using random movement models. The method could also be extended straightforwardly to include a time lag to determine whether individuals are more often in an area recently occupied by another animal than might be explained by chance. This could be important in cases of geo-located time-limited phenomena such as scent marking or the transmission

of parasites or infectious agents through the environment.

Part III

Cointegration

Chapter 4

Literature Review on Cointegration

When testing for dependencies between two or more time series, the Pearson product-moment correlation coefficient, commonly called ‘the correlation coefficient’, is often one of the first statistics considered. However, correlation only makes sense if the individual series are stationary and the relationship between the two time series is linear [14]. In practice there are a lot of time series that are non-stationary, such as exchange rates [41], population and employment [36], electricity consumption [2], gas prices [27], dissent rates on the High Court of Australia [51] and hemispheric temperature [42]. In these cases, a characteristic called cointegration, could be beneficial. Cointegration describes two or more time series that are individually integrated (i.e. non-stationary and therefore do not have a constant mean, variance or covariance), although a linear relationship of the time series is stationary.

In this dissertation a new fully Bayesian method to test for cointegration is presented. A Bayesian approach is advantageous for many reasons: it produces whole probability distributions for each unknown parameter and these distributions are valid for any sample size. Furthermore, it allows straightforward updates when more data becomes available, by using the posterior as the new prior distribution.

4.1 Contributions

The main advantage of the cointegration test proposed in this thesis is that it fully exploits the benefits of the Bayesian method. A cointegration tube is created, which describes the combinations of parameters that indicate cointegration of the two time series being compared. This allows us to test whether say 95% of the posterior lies within the cointegration tube, rather than using a point estimate, such as the maximum likelihood of the posterior.

Significant progress has been made in calculating a hard upper bound on the integration error. An upper bound on the error would lead to an upper and lower bound on the integration itself and therefore give a better understanding of how certain the result is.

The hard lower bound on the integration means that the integral is greater or equal to the lower bound. This means that if the lower bound suggests that 95% of the posterior probability density lies within the cointegration tube, then you can be sure that at least 95% lies within the cointegration tube.

Additionally this method is, to the best of my knowledge, the only one that tests whether the time series are cointegrated during the observation period. Other methods use only one parameter which is indicative of the time series becoming cointegrated at some future point in time [29, 39, 21]. The method we propose considers three parameters (introduced in Section 5.2) which allows us to test for cointegration of the observed time series themselves. This is particularly useful when change points are present in the data. Change points are not uncommon in long term observations. They have, for example, been used to study surface temperature [43] and oceanographic time series [44].

4.2 Mathematical and Statistical Background Information

Underpinning the idea of cointegration is stationarity. The joint distribution of any subset of the points of a stationary time series does not change when the subset is shifted in time. Therefore its mean, variance and auto-covariance do not depend on time. Most statistical forecasting methods are based on the assumption that the time series can be transformed into an approximately stationary time series through the use of transformations, such as removing trend or seasonality. As a result, this report starts with a detailed explanation of stationarity and its implications (Section 4.2.1). We then introduce some models used to represent stationary times series, called AR models (see Section 4.2.2). This is followed by an explanation of the new cointegration model in Chapter 5.

4.2.1 Stationarity

When analysing time series, an important simplifying assumption is that of stationarity. A realisation of a stochastic process, say $\mathbf{y} = (y_1, y_2, \dots, y_T)$, can be described by a T -dimensional probability distribution $p(\mathbf{y})$. The relationship between a realisation and a stochastic process is analogous to the relationship between a sample and a population in classical statistics.

From the realisation we cannot estimate all the first- and second-order moments, since we have T (means) + T (variances) + $T \times (T - 1)/2$ (covariances) = $\frac{1}{2}(T^2 + 3T)$ parameters to be estimated, yet only T observations. Therefore we need further assumptions. For example that of stationarity, as described in Chapter 3.1 of Hamilton's book on Time Series Analysis [30].

Definition 4.1. *A process is said to be **strictly stationary** if, for any values of j_1, j_2, \dots, j_n the joint distribution of $(\mathbf{Y}_t, \mathbf{Y}_{t+j_1}, \mathbf{Y}_{t+j_2}, \dots, \mathbf{Y}_{t+j_n})$ depends only on the intervals separating the dates (j_1, j_2, \dots, j_n) and not on the dates themselves $(t, t + j_1, \dots, t + j_n)$.*

Specifying the complete form of the distribution is impossible in most cases so attention is often focused on the first and second moments. As a result a weaker form of stationarity [30] is generally used.

Definition 4.2. *If neither the mean nor the auto-covariances depend on the date, then the process for \mathbf{Y}_t is said to be **covariance-stationary** or **weakly stationary**.*

If not specified differently, stationarity generally implies weak stationarity. The simplest example of a weakly stationary stochastic process is a sequence of uncorrelated random variables with constant zero mean and constant variance σ^2 . A process like this is known as white noise and will be denoted by $\boldsymbol{\eta}$ for the remainder of this dissertation. Because the variables in a white noise sequence are uncorrelated, the auto-covariances are all zero for a lag greater than 0.

Stationary processes are also often denoted as $I(0)$, which means that the time series is integrated of order zero. A time series, which is stationary after differencing d times, is said to be integrated of order d and denoted as $I(d)$. White noise is a typical example of a time series which is integrated of order zero and a random walk is an example of a time series integrated of order one, since its difference is white noise.

4.2.2 Autoregressive Processes

The model proposed in Chapter 5 uses an AR(1) process. This is defined as $\epsilon_t = \phi\epsilon_{t-1} + \eta_t$, where η is white noise, and is often also called a Markov process. For simplicity, and without loss of generality, this model is assumed to be a zero-mean process.

To study this process, note that it can be rewritten as

$$\epsilon_t = \phi^k \epsilon_{t-k} + \sum_{j=0}^{k-1} \phi^j \eta_{t-j}$$

which is only well-defined as $k \rightarrow \infty$ if $|\phi| < 1$. In that case it can be simplified to

$$\epsilon_t = \sum_{j=0}^{\infty} \phi^j \eta_{t-j} \quad (4.1)$$

Taking the expectation on both sides of Equation 4.1 gives $\mathbb{E}(\epsilon_t) = 0$. Taking the variances gives $Var(\epsilon_t) = \sum_{j=0}^{\infty} \phi^{2j} Var(\eta_{t-j}) = \frac{\sigma^2}{1-\phi^2}$, again requiring $|\phi| < 1$. Similarly, the covariance between ϵ_t and ϵ_{t-k} can be derived as:

$$Cov(\epsilon_t, \epsilon_{t-k}) = \frac{\sigma^2 \phi^k}{1-\phi^2} \quad (4.2)$$

as shown in Appendix B.1 which once more provides the same restriction.

Looking at all these results, we see that:

1. The single condition under which all the results can be derived is that $|\phi| < 1$.
2. None of the expressions depends on t , and therefore the AR(1) process must be weakly stationary provided $|\phi| < 1$. The requirement $|\phi| < 1$ is called the stationarity condition.

4.2.3 Bayes Factor

The Bayes Factor plays a similar role in Bayesian statistics to the likelihood ratio in frequentist statistics. Instead of comparing the maximum likelihood of two models, as is done in the likelihood ratio, the Bayes factor compares the integrated likelihoods [9]. This is a very familiar difference when comparing frequentist to Bayesian models. Where frequentist methods generally compare point estimates, the Bayesian methods investigate the whole probability distribution. This causes Bayesian methods to be generally higher in computational cost, but to retain more information.

The definition of the Bayes factor as per Bernardo and Smith [9], is:

Definition 4.3 (Bayes factor). *Given two hypotheses H_i and H_j corresponding to assumptions of alternative models, M_i and M_j , for data \mathbf{x} , the Bayes factor in favour of H_i (against H_j) is given by the posterior to prior odds ratio.*

$$B_{ij}(\mathbf{x}) = \frac{p(\mathbf{x}|M_i)}{p(\mathbf{x}|M_j)} = \left\{ \frac{p(M_i|\mathbf{x})}{p(M_j|\mathbf{x})} \right\} \bigg/ \left\{ \frac{p(M_i)}{p(M_j)} \right\}$$

4.2.4 Riemann Sum

The Riemann Sum is used in Section 5.7, so the basics of this method are explained in this section. The idea of the Riemann Sum is that the integral can be divided into small areas (intervals in the one dimensional case, squares in two dimensions, etc.). For the rest of this section, the one dimensional case will be considered; however, extensions to higher dimensions can be made easily.

Once the overall interval is divided into several small intervals, the value of the integral within each interval is estimated and an upper bound on the error is calculated. An estimate of the overall integral is then computed by adding up the estimated integrals for each interval and the upper bound on the error can be found similarly.

The integral within each interval is estimated by multiplying the length of the interval with the value of the function to be integrated over, at a certain point of the interval. One version of this, and the version used in this thesis, is that the left-endpoint of the integral (i.e. the point closest to $-\infty$), is used [61].

The formal definition is:

Definition 4.4 (One dimensional Riemann sum). *Let $f : D \rightarrow \mathbb{R}$ be a function defined on a subset, D , of the real line, \mathbb{R} . Let $I = [a, b]$ be a closed interval in D , and let $P = \{[x_0, x_1], [x_1, x_2], \dots, [x_{m-1}, x_m]\}$, be a partition of I , where $a = x_0 < x_1 < x_2 < \dots < x_m = b$. The Riemann sum of f over I with partition P is defined as*

$$S = \sum_{i=1}^m f(x_i^*)(x_i - x_{i-1}),$$

with $x_{i-1} \leq x_i^* \leq x_i$.

For the left Riemann sum $x_i^* = x_{i-1} \forall i$.

4.2.5 Operator Norm

The operator norm is used to calculate the effect of perturbations on matrices in Section 5.7.1.1. The information needed for this is summarised here. First the definition of the operator norm, as detailed in [3], is:

Definition 4.5 (Operator Norm). *Let $\mathcal{L}(V, W)$ denote the set of all the continuous linear operators from a normed space V to another normed space W . In the special case $W = V$,*

$\mathcal{L}(V)$ replaces $\mathcal{L}(V, V)$. If $L \in \mathcal{L}(V, W)$:

$$\|L\|_{V,W} = \sup_{0 \neq v \in V} \frac{\|Lv\|_W}{\|v\|_V}.$$

This norm is usually called the operator norm of L .

As shown in [3] the operator norm has the following properties:

$$\|L\|_{V,W} = \sup_{\|v\|_V=1} \|Lv\|_W$$

and

$$\|Lv\|_W \leq \|L\|_{V,W} \|v\|_V \quad \forall v \in V.$$

The spaces dealt with in this thesis are all of the form $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ and therefore $\|\cdot\|_{V,W}$ will be denoted by $\|\cdot\|_{OP}$ when referring to the operator norm, as long as the spaces V and W are clear from the context. The operator norm will generally be used when working with matrices. When the vector norm is not specified explicitly, the Euclidean vector norm will be used and will be denoted by $\|\cdot\|$.

4.2.5.1 Bounds on the Operator Norm of a Perturbed Inverse Matrix

It was shown by Dirk Ferus [22] that for invertible matrices G and F in finite Banach spaces V , W , and given $\|F - G\|_{OP} < \frac{1}{\|F^{-1}\|_{OP}}$, then

$$\|G^{-1} - F^{-1}\|_{OP} \leq \|G^{-1}\|_{OP} \|F - G\|_{OP} \|F^{-1}\|_{OP}. \quad (4.3)$$

A second result [22] says that for G and F as previously defined,

$$\|G^{-1}v\| \leq \left(\frac{1}{\|F^{-1}\|_{OP}} - \|G - F\|_{OP} \right)^{-1} \|v\|. \quad (4.4)$$

These two results will be used in Section 5.7.1.1, to calculate the effect of a small perturbation on the inverse of matrices.

4.2.5.2 Sub-Multiplicativity of the Operator Norm

The characteristic of a sub-multiplicative norm is the additional property that $\|AB\| \leq \|A\| \|B\|$. Here we show that the operator norm is a sub-multiplicative norm.

As defined at the beginning of Section 4.2.5, the operator norm is defined as $\|A\|_{\text{OP}} = \sup_{\|\mathbf{v}\|_V=1} \|A\mathbf{v}\|_W$ where $A \in \mathcal{L}(V, W)$. The second matrix is defined as $B \in \mathcal{L}(U, V)$. Then $AB \in \mathcal{L}(U, W)$ and

$$\|AB\|_{\text{OP}} = \sup_{\|\mathbf{u}\|_U=1} \|A(B\mathbf{u})\|_W.$$

This can be multiplied by 1 in the form of $\|B\mathbf{u}\|_V / \|B\mathbf{u}\|_V$

$$\|AB\|_{\text{OP}} = \sup_{\|\mathbf{u}\|_U=1} \frac{\|A(B\mathbf{u})\|_W}{\|B\mathbf{u}\|_V} \|B\mathbf{u}\|_V.$$

Since $\left\| \frac{B\mathbf{u}}{\|B\mathbf{u}\|_V} \right\| = 1$ and $\|A\|_{\text{OP}} \geq \|A\mathbf{v}\|$ when $\|\mathbf{v}\|_V = 1$ (as detailed in Section 4.2.5)

$$\begin{aligned} \|AB\|_{\text{OP}} &\leq \sup_{\|\mathbf{u}\|_U=1} \|A\|_{\text{OP}} \|B\mathbf{u}\|_V \\ &= \|A\|_{\text{OP}} \sup_{\|\mathbf{u}\|_U=1} \|B\mathbf{u}\|_V. \end{aligned}$$

which is, by definition,

$$\|AB\|_{\text{OP}} \leq \|A\|_{\text{OP}} \|B\|_{\text{OP}}$$

as desired.

4.3 Literature Review

As mentioned previously, cointegration describes two or more time series that are individually integrated (i.e. non-stationary and therefore do not have a constant mean, variance or covariance), although a linear relationship of the time series is stationary. This is common both in and outside biology and therefore represents a valuable technique, one for which advances are valuable in a wide variety of areas, such as animal behaviour [53], econometrics [1, 2, 27, 36, 41] and weather [42].

For a more detailed definition, let $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$ denote an $(n \times T)$ matrix, where row i represents the i th time series \mathbf{y}_i of length T . Each \mathbf{y}_i is integrated of order d (denoted by $I(d)$ for $d \geq 1$), this means that the time series has to be differenced d times before it becomes stationary. The rows of Y are cointegrated if there exists an $(n \times 1)$ vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$ such that

$$\boldsymbol{\beta}'Y = \beta_1\mathbf{y}_1 + \dots + \beta_n\mathbf{y}_n \sim I(d^*) \quad (4.5)$$

where $d^* < d$. Often $d = 1$ and $d^* = 0$, which means that the individual time series are integrated of order 1 and the resulting linear relationship is stationary. This will be assumed for the rest of the thesis.

The cointegrating vector $\boldsymbol{\beta}$ in Equation 4.5 is not unique since, for any scalar c , the linear combination $c\boldsymbol{\beta}'Y = \boldsymbol{\beta}^*Y$. Hence, some normalization assumption is required to uniquely identify $\boldsymbol{\beta}$. A typical normalization is

$$\boldsymbol{\beta} = (1, -\beta_2, \dots, -\beta_n)'$$

so that the cointegration relationship may be expressed as

$$\boldsymbol{\beta}'Y = \mathbf{y}_1 - \beta_2\mathbf{y}_2 - \dots - \beta_n\mathbf{y}_n \sim I(0)$$

or

$$y_{(1,t)} = \beta_2 y_{(2,t)} + \cdots + \beta_n y_{(n,t)} + \epsilon_t$$

where $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_T)' \sim I(0)$ [58].

As soon as more than two time series are being examined the cointegration tests can be approached from two angles:

- There is at most one cointegrating vector
- There are possibly $0 \leq r < n$ cointegrating vectors

The first case was originally considered by Engle and Granger [21] who developed a simple two-step residual-based testing procedure based on regression techniques. The second case was originally considered by Johansen [39] who developed a sophisticated sequential procedure for determining the existence of cointegration and for determining the number of cointegrating relationships based on maximum likelihood techniques. Most frequentist as well as Bayesian methods to test for cointegration have been based on either the Engle and Granger [10, 15, 24] or the Johansen approach [46, 60]. So far these methods have focused on finding suitable priors [8, 45], or suggesting varying numerical integration methods [7].

In this dissertation the focus will be on the Engle and Granger approach, as it is more flexible and allows testing of time series of any order of integration. We will first examine the Frequentist procedure of testing for cointegration in Section 4.4, which will be followed by a description of the Bayesian methods in Section 4.5, before describing the new method proposed in concluding with a discussion in Chapter 5.

4.4 Frequentist Cointegration

The null hypothesis in the Engle-Granger [21] procedure is no-cointegration, which is tested against the alternative hypothesis of cointegration. Engle and Ganger devised a

two-step procedure for determining if the $(n \times 1)$ vector β is a cointegrating vector:

- Run a regression of $\beta'Y = \epsilon$.
- Carry out a unit root test on the residuals ϵ . If the unit root hypothesis is rejected then conclude that the rows of Y are cointegrated.

Recall, that cointegration means that the linear relationship of the time series is stationary (see Section 4.3). The unit root test is a common way of testing individual time series for stationarity. Typical methods are the Dickey-Fuller Test, the Augmented Dickey-Fuller Test and the Phillips Perron test [30, 23, 56].

There are two cases to consider. In the first case, the proposed cointegrating vector β is pre-specified (not estimated). For example, economic theory may imply specific values for the elements in β such as $\beta = (1, -1)'$. The cointegrating residual is then readily constructed using the pre-specified cointegrating vector and any unit root test statistic may be used.

In the second case, the proposed cointegrating vector is estimated from the data and an estimate of the cointegrating residual $\hat{\beta}'Y = \hat{\epsilon}$ is formed. The pseudo-code for this approach is listed below. The inputs are the two time series, y_1 and y_2 , that are to be tested for cointegration and the output is the p-value corresponding to the null-hypothesis that the time series are not cointegrated vs. the alternative hypothesis that the time series are cointegrated.

Algorithm 2 Pseudo-code for Engle and Granger Test for Cointegration

- 1: **procedure** ENGLE AND GRANGER TEST FOR COINTEGRATION(y_1, y_2)
 - 2: Estimate α and β in $y_1 = \alpha + \beta y_2 + \epsilon$ through ordinary least squares
 - 3: $\hat{\epsilon} \leftarrow y_1 - \hat{\alpha} - \hat{\beta}y_2$
 - 4: Use any unit root test (for example the Augmented Dickey Fuller Test) to test whether $\hat{\epsilon}$ is stationary
-

Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector, yet the cointegration vector is rarely

known, which is why we will concentrate on the second case.

The hypotheses to be tested are

$$H_0 : \epsilon = \beta'Y \sim I(1) \quad (\text{no cointegration})$$

$$H_1 : \epsilon = \beta'Y \sim I(0) \quad (\text{cointegration}).$$

Since β is unknown, it must be estimated from the data to be able to use the Engle-Granger procedure. As mentioned at the beginning of this chapter, a normalization assumption must be made to identify β uniquely. A common normalization is to specify the first row in Y as the dependent variable and the rest as the explanatory variables. Then $Y = (\mathbf{y}_1, Y_2)'$ where $Y_2 = (\mathbf{y}_2, \dots, \mathbf{y}_n)'$ is an $(n-1) \times T$ matrix and the cointegrating vector is normalized as $\beta = (1, -\beta_2)'$.

For simplicity the rest of this method will be presented for $n = 2$ time series; however, all aspects can be easily extended to the more general case of n time series.

$$y_{(1,t)} = \alpha + \beta_2 y_{(2,t)} + \epsilon_t$$

The unit root test is then based on the estimated cointegrating residual

$$\hat{\epsilon}_t = y_{(1,t)} - \hat{\alpha} - \hat{\beta}_2 y_{(2,t)}$$

where $\hat{\alpha}$ and $\hat{\beta}_2$ are the least squares estimates of α and β_2 .

4.5 Bayesian Cointegration

A Bayesian approach is advantageous for many reasons: it produces whole probability distributions for each unknown parameter that are valid for any sample size and it allows straightforward updates when more data becomes available, by using the posterior as the

new prior distribution.

Most of the existing Bayesian cointegration methods follow the principle of either Johansen's [7, 8, 46] or Engle and Granger's [10, 15, 45] frequentist cointegration test. In this work we will focus on the methods based on Engle and Granger's approach, as the methods based on Johansen's work assume that the observed time series are integrated of order one and face an identification problem [15], which is not present in Engle and Granger's approach.

The assumption of first order integration in Johansen's method causes the differenced series to be stationary and allows the focus on a single matrix they call the coefficient matrix. The rank of this matrix then gives the number of cointegration vectors. However, as soon as the original time series are of a higher order of integration, the differenced series are not stationary and there is no way of identifying whether there are any cointegrating vectors.

Engle and Granger's approach [21] is considerably more flexible than the Johansen approach and allows expansion into higher orders of integration. There is no assumption that the original time series are of any particular order of integration. On the other hand, the main criticism of the Engle and Granger method is that the coefficients of the linear relationship need to be estimated before the resulting error is tested for stationarity and that therefore cointegration is tested for only one particular equation. A Bayesian approach would allow us to integrate over all possible coefficients during the calculation of the posterior probability, which means that cointegration is tested for all possible linear relationships and therefore overcomes the principal criticism of this approach.

The first Bayesian Cointegration test was suggested by Koop [45]. This test is based on the Engle and Granger approach [21] and compares three hypotheses (H_1 : the two time series follow random walks with drift and are cointegrated, H_2 : the two time series follow random walks and are not cointegrated and H_3 : the two time series do not follow random walks with drift and are not cointegrated). They specified the three hypotheses so they

are nested within each other and, consequently, use the bivariate odds ratio technique to decide which hypotheses to reject.

To calculate the odds ratios a multidimensional integration needs to be performed. Koop [45] uses a Monte Carlo method of integration, for which it is not possible to calculate hard bounds on the error [65].

De la Croix and Lubrano [15] test for cointegration by computing the posterior probability that a certain parameter is greater or equal to one, which is assumed to indicate that the time series are not cointegrated. This probability is then compared to a pre-assigned probability value like 5%. The posterior density is again computed using numerical integration. Focusing on that one parameter means that it is tested whether the time series would become cointegrated at a future point in time, not whether the observed time series are cointegrated.

One of the most recent Bayesian methods based on the Engle and Granger approach [21] is proposed by Bracegirdle and Barber in [10]. Similar to Koop's Bayesian approach Bracegirdle and Barber [10] suggest the likelihood ratio test to compare the model of no cointegration to the model of cointegration. The maximum likelihood is estimated using the Expectation-Maximisation algorithm; however, there is no guarantee that the algorithm converges to a maximum likelihood estimator. For multi-modal distributions it could converge to a local maximum.

Chapter 5

Test of Cointegration

5.1 Overview

In this chapter the new method to detect cointegration is introduced. This method is the only method, to the best of our knowledge, that detects whether the observed time series are cointegrated, as opposed to whether they will become cointegrated at a future point in time. This could be particularly useful when change points are present.

An example application domain that would benefit from this approach might be the analysis of position data from sheep with a neuro degenerative disease. One of the symptoms of this type of disease can be seen in gradual changes to the social behaviour of the affected animals. In such a situation, it might be possible to detect a slight change in the social behaviour by comparing several short periods, for which it would be important to detect whether the observed time series are cointegrated.

To achieve this, three steps are proposed: first, the posterior pdf of the relevant parameters (described in Section 5.2), given the observed data, is calculated analytically (see Section 5.5). Simultaneously, the combinations of the parameters is calculated that suggest cointegration in Section 5.6; these combinations are called the Cointegration Tube in the remainder of this dissertation. Finally, the posterior probability distribution is integrated over the Cointegration Tube to ascertain the proportion of the posterior probability that lies within the cointegration tube. If the proportion is above a predefined threshold, say 90%, then the two time series are assumed to be cointegrated.

The pseudo-code for the method is as follows. The input is the two time series, y_1 and y_2 , the maximum error for the numerical integration, E_{max} , that is accepted (default value set to 0.01), and the size of the cointegration tube parameter, ζ , which defines how wide the cointegration tube is (default value set to 0.75). By default, the prior distributions for ϕ , σ^2 and σ_1^2 , the cointegration parameters, are set to $\mathcal{N}(\mu_\phi, \sigma_\phi^2)$, $InvGamma(\tau, \kappa)$ and $InvGamma(\tau_1, \kappa_1)$, with default values for μ_ϕ , σ_ϕ^2 , τ , κ , τ_1 and κ_1 set to 0, 1, 1, 1, 1 and 1 respectively. However, as described in Section 5.4 these should be adapted depending

on any prior information about the data. The output is the proportion of the posterior that lies within the cointegration tube and therefore suggests cointegration.

Algorithm 3 Pseudo-code for proposed method to test for cointegration

```

1: procedure COINTEGRATIONTEST( $y_1, y_2, E_{max} \leftarrow 0.01, \zeta \leftarrow 0.75, p(\phi) \leftarrow$ 
 $\frac{1}{\sigma_\phi \sqrt{2\pi}} \exp\left\{-\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2}\right\}, p(\sigma^2) \leftarrow \frac{\tau^\kappa \sigma^{2(-\kappa-1)} e^{-\tau/\sigma^2}}{\Gamma(\kappa)}, p(\sigma_1^2) \leftarrow \frac{\tau_1^{\kappa_1} \sigma_1^{2(-\kappa_1-1)} e^{-\tau_1/\sigma_1^2}}{\Gamma(\kappa_1)}, \mu_\phi \leftarrow 0,$ 
 $\sigma_\phi^2 \leftarrow 1, \tau \leftarrow 1, \kappa \leftarrow 1, \tau_1 \leftarrow 1, \kappa_1 \leftarrow 1$ )
2:    $D_1 \leftarrow 0.2$ 
3:    $E \leftarrow \infty$ 
4:   while  $E > E_{max}$  do
5:      $S \leftarrow 0$ 
6:      $N \leftarrow 0$ 
7:      $E \leftarrow 0$ 
8:      $D_1 \leftarrow D_1/2$ 
9:     for  $\phi$  from -2 in  $D_1$  to 2 do
10:      for  $\sigma^2$  from 0 in  $D_1$  to 2 do
11:        for  $\sigma_1^2$  from 0 in  $D_1$  to 2 do
12:          Calculate  $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2, \phi, \sigma^2, \sigma_1^2)$  (see Section 5.5)
13:           $S_n \leftarrow D_1^3 p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2, \phi, \sigma^2, \sigma_1^2)$ 
14:           $N \leftarrow N + S_n$ 
15:          Calculate hard upper bound on the error of the integral within the
          current block (see Section 5.7.1.2)
16:           $E \leftarrow E + E_n$ 
17:          if  $\zeta \geq \max_{t,k} \left| \phi^k \left( (1 - \phi^{2(t-1)}) \sigma_1^2 - \sum_{i=0}^{t-2} \phi^{2i} \sigma^2 \right) \right|$  then
18:             $S \leftarrow S + S_n$ 
19:    $p \leftarrow S/N$ 
20: return  $p$ 

```

5.2 Model Description

This chapter focuses on the proposed method that detects cointegration between two time series, denoted by $\mathbf{Y}_1 = (Y_{1,1}, \dots, Y_{1,T})'$ and $\mathbf{Y}_2 = (Y_{2,1}, \dots, Y_{2,T})'$. Two time series are

cointegrated if a linear combination of them is stationary.

We assume the two time series are given by:

$$Y_{1,t} = A + BY_{2,t} + E_t \quad (5.1)$$

where $t \in \{1, \dots, T\}$ and $\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Omega\right)$. Noise in the system is modelled as

$$E_t = \Phi E_{t-1} + H_t, \quad (5.2)$$

with

$$\Phi \sim \mathcal{N}(\mu_\phi, \sigma_\phi^2), H_t \sim \mathcal{N}(0, S^2), E_1 \sim \mathcal{N}(0, S_1^2) \quad (5.3)$$

and $S^2 \sim \text{InvGamma}(\kappa, \tau)$, $S_1^2 \sim \text{InvGamma}(\kappa_1, \tau_1)$, $\mu_\phi \in \mathbb{R}$, $\sigma_\phi^2 \in [0, \infty[$ and $\kappa, \tau, \kappa_1, \tau_1 > 0$.

The process \mathbf{Y}_2 is described by a first-order Markov model

$$Y_{2,t} = \phi_y Y_{2,t-1} + W_t, \quad (5.4)$$

with $Y_{2,1} \sim \mathcal{N}(0, \sigma_y^2)$, $W_t \sim \mathcal{N}(0, \sigma_w^2)$, $\phi_y \in \mathbb{R}$ and $\sigma_y^2, \sigma_w^2 \in [0, \infty[$.

The linear relationship between the two time series is described in Equation 5.1. The intercept and the gradient of this relationship have a bivariate normal prior distribution. The noise in the system is described by E_t (Eq. 5.2) which follows a first-order Markov model. \mathbf{Y}_2 also follows a first-order Markov model (Eq. 5.4), which is a

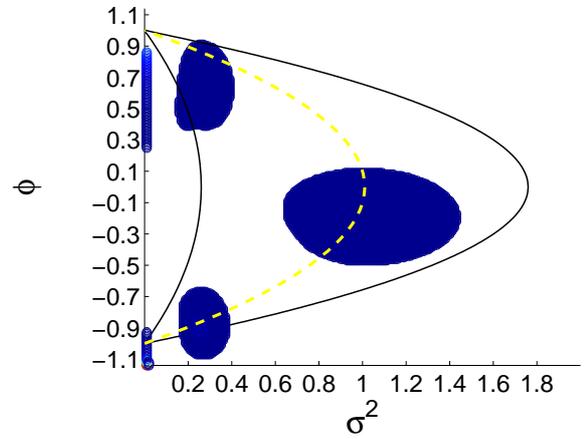


Figure 5.1: Example of a heat map of the posterior probability density overlaid with a Cointegration Tube. The white dashed line is the line of perfect cointegration and the yellow solid lines are the upper and lower bounds.

simple model often used for time series, when the following observation is expected to depend only on the previous one. It also allows for the possibility that the time series does not depend on the previous observations in which case ϕ_y would be 0.

In addition to the posterior probability density function a ‘cointegration tube’, such as in Figure 5.1, is needed. The combinations of S_1 , S and Φ that cause \mathbf{Y}_1 and \mathbf{Y}_2 to be cointegrated lie on a line in \mathbb{R}^3 . However the posterior mass of a line in a three dimensional space is 0.

Because of this, a margin ζ , as described in Section 5.6, is incorporated. This variable influences how close to ‘perfectly’ cointegrated the time series must be before they are assumed to be cointegrated. Summed up, the ‘cointegration tube’ depicts the area in which the particular value of $\Theta = (\Phi, S^2, S_1^2)$ indicates approximate cointegration of \mathbf{y}_1 and \mathbf{y}_2 . This is detailed in Section 5.6.

After the cointegration tube has been identified, the posterior is integrated over this area to determine what proportion of the posterior probability mass function lies within the cointegration tube, as is demonstrated in Section 5.7. This includes the significant progress in calculating a strict upper bound on the error of this integral.

5.3 Identification

An identification problem exists when \mathbf{y}_2 is close to constant over time. In that case the effect of A and B cannot be distinguished. Take the simple example of $\mathbf{y}_2 = \mathbf{1}$. Then $A + B\mathbf{1} = (A + a) + (B - a)\mathbf{1} = A^* + B^*\mathbf{1}$ for any $a \in \mathbb{R}$.

Similarly, when Ω , the prior covariance matrix of A and B is positive semidefinite rather than positive definite, there is an identification problem associated with A and B . This could happen if the prior variances as well as prior covariances are set to the same value, say 1 for example. Therefore the method requires a positive definite prior covariance

matrix for $\begin{pmatrix} A \\ B \end{pmatrix}$.

A further potential problem arises when the sum of two matrices is not invertible. This cannot occur when both matrices are positive definite, because the sum of positive definite matrices is again positive definite. In this thesis, when the sum of two matrices is required to be invertible, the individual matrices are shown to be positive definite.

5.4 Prior Distributions

For the remainder of the thesis, the following short form of the densities will be used, unless stated otherwise. Instead of for example $p(\mathbf{Y}_1 = \mathbf{y}_1)$ and $p(\Theta = \theta)$, $p(\mathbf{y}_1)$ and $p(\theta)$, will be used.

A prior for $\Theta = (\Phi, S^2, S_1^2)$, the vector of cointegration parameters, needs to be defined. It is assumed that Φ , S^2 and S_1^2 are independent, which results in $p(\theta) = p(\phi)p(\sigma^2)p(\sigma_1^2)$. For the individual priors the following distributions are chosen:

$$\begin{aligned}\Phi &\sim \mathcal{N}(\mu_\phi, \sigma_\phi^2), \\ S^2 &\sim \text{Inv-Gamma}(\kappa, \tau), \\ S_1^2 &\sim \text{Inv-Gamma}(\kappa_1, \tau_1)\end{aligned}\tag{5.5}$$

with $\mu_\phi \in]-\infty, \infty[$, $\sigma_\phi^2 \in [0, \infty[$ and $\kappa, \tau, \kappa_1, \tau_1 \in]0, \infty[$. The normal distribution for the prior of Φ is believed to be a good fit, as the prior should be symmetric around the most expected value¹, since we have no reason to believe that Φ should be skewed in any particular direction. We also believe that all values should be available, although it is much more likely that ϕ lies between -1 and 1 because if it were very large, the chances are small that people would even want to test for cointegration.

The Inverse Gamma distribution is used as prior for both S^2 and S_1^2 , since it is the most

¹which will most often probably be 0

common prior distribution chosen for variances and it is stated that for a simple two-level hierarchical model typically there is enough data to estimate the variance parameter well with any reasonable non-informative prior distribution [26].

This results in

$$\begin{aligned}
p(\boldsymbol{\theta}) &= p(\phi)p(\sigma^2)p(\sigma_1^2) \\
&= \frac{1}{\sigma_\phi\sqrt{2\pi}} \exp\left\{-\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2}\right\} \frac{\tau^\kappa \sigma^{2(-\kappa-1)} e^{-\tau/\sigma^2}}{\Gamma(\kappa)} \frac{\tau_1^{\kappa_1} \sigma_1^{2(-\kappa_1-1)} e^{-\tau_1/\sigma_1^2}}{\Gamma(\kappa_1)} \\
&= \frac{\tau^\kappa \sigma^{2(-\kappa-1)} \tau_1^{\kappa_1} \sigma_1^{2(-\kappa_1-1)}}{\sigma_\phi\sqrt{2\pi}\Gamma(\kappa)\Gamma(\kappa_1)} \exp\left\{-\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2} - \frac{\tau}{\sigma^2} - \frac{\tau_1}{\sigma_1^2}\right\}.
\end{aligned} \tag{5.6}$$

5.5 Posterior of $\boldsymbol{\theta}$ Given \mathbf{y}_1 and \mathbf{y}_2

To compute the posterior probability density function, Bayes' Theorem is initially used

$$p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2) = \frac{p(\boldsymbol{\theta}, \mathbf{y}_1|\mathbf{y}_2)}{p(\mathbf{y}_1|\mathbf{y}_2)} = \frac{p(\mathbf{y}_1|\boldsymbol{\theta}, \mathbf{y}_2)p(\boldsymbol{\theta}|\mathbf{y}_2)}{p(\mathbf{y}_1|\mathbf{y}_2)}. \tag{5.7}$$

This will always be defined since $p(\mathbf{y}_1|\mathbf{y}_2) \neq 0$, as shown in Appendix C.1.

Therefore

$$p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2) \propto p(\mathbf{y}_1|\boldsymbol{\theta}, \mathbf{y}_2)p(\boldsymbol{\theta}|\mathbf{y}_2).$$

with the normaliser

$$N = p(\mathbf{y}_1|\mathbf{y}_2) = \int_{\boldsymbol{\theta}} p(\mathbf{y}_1, \boldsymbol{\theta}|\mathbf{y}_2) d\boldsymbol{\theta} \tag{5.8}$$

which is equivalent to the numerator of Equation 5.7 integrated over Θ .

From the law of total probability it is known that

$$p(\mathbf{y}_1|\boldsymbol{\theta}, \mathbf{y}_2) = \int_{(\alpha, \beta)} p(\mathbf{y}_1, \alpha, \beta|\boldsymbol{\theta}, \mathbf{y}_2) d(\alpha, \beta)$$

taking the integral over all possible pairs of α and β . This can be rewritten as

$$p(\mathbf{y}_1|\boldsymbol{\theta}, \mathbf{y}_2) = \int_{(\alpha,\beta)} p(\mathbf{y}_1|\alpha, \beta, \boldsymbol{\theta}, \mathbf{y}_2)p(\alpha, \beta|\boldsymbol{\theta}, \mathbf{y}_2) d(\alpha, \beta).$$

As the following transformation can get slightly confusing the extended format of probabilities is used:

$$p(\mathbf{Y}_1 = \mathbf{y}_1|A = \alpha, B = \beta, \mathbf{Y}_2 = \mathbf{y}_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) = p(\mathbf{E} = \mathbf{y}_1 - \alpha\mathbf{1} - \beta\mathbf{y}_2|\boldsymbol{\Theta} = \boldsymbol{\theta}) \quad (5.9)$$

since $\mathbf{E} = \mathbf{Y}_1 - A - B\mathbf{Y}_2$ as defined in Equation 5.1, where $\mathbf{Y}_1, \mathbf{Y}_2, A, B, \boldsymbol{\Theta}, \mathbf{E}$ are random variables.

In addition $p(\boldsymbol{\Theta} = \boldsymbol{\theta}|\mathbf{Y}_2 = \mathbf{y}_2) = p(\boldsymbol{\Theta} = \boldsymbol{\theta})$, as shown in C.2, and

$p(A = \alpha, B = \beta|\boldsymbol{\Theta} = \boldsymbol{\theta}, \mathbf{Y}_2 = \mathbf{y}_2) = p(A = \alpha, B = \beta)$, as shown in Appendix C.3, from which follows:

$$p(\boldsymbol{\Theta} = \boldsymbol{\theta}|\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2) \propto p(\boldsymbol{\Theta} = \boldsymbol{\theta}) \int_{(\alpha,\beta)} p(\mathbf{E} = \mathbf{y}_1 - \alpha\mathbf{1} - \beta\mathbf{y}_2|\boldsymbol{\Theta} = \boldsymbol{\theta})p(A = \alpha, B = \beta) d(\alpha, \beta)$$

and, as shown in Appendix C.4, this is equivalent to

$$\begin{aligned} p(\boldsymbol{\Theta} = \boldsymbol{\theta}|\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2) &\propto \\ p(\boldsymbol{\Theta} = \boldsymbol{\theta}) &\int_{(\alpha,\beta)} p(E_1 = y_{(1,1)} - \alpha - \beta y_{(2,1)}|\boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &\prod_{s=2}^T p(E_s = y_{(1,s)} - \alpha - \beta y_{(2,s)}|E_{s-1} = y_{(1,s-1)} - \alpha - \beta y_{(2,s-1)}, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &p(A = \alpha, B = \beta) d(\alpha, \beta) \end{aligned} \quad (5.10)$$

Furthermore, the following is shown in Appendix C.5

$$p(E_s = y_{(1,s)} - \alpha - \beta y_{(2,s)}|E_{s-1} = y_{(1,s-1)} - \alpha - \beta y_{(2,s-1)}, \boldsymbol{\Theta} = \boldsymbol{\theta}) \sim \mathcal{N}(\phi\epsilon_{s-1}, \sigma^2).$$

The product of multiple normally distributed random variables is again normally distributed and therefore the mean and covariance matrix of E_t are $\mathbf{0}$ and Σ respectively, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \phi\sigma_1^2 & \cdots & \phi^{T-1}\sigma_1^2 \\ \phi\sigma_1^2 & \phi^2\sigma_1^2 + \sigma^2 & \cdots & \phi^T\sigma_1^2 + \phi^{T-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1}\sigma_1^2 & \phi^T\sigma_1^2 + \phi^{T-2}\sigma^2 & \cdots & \phi^{2(T-1)}\sigma_1^2 + \sigma^2 \sum_{i=0}^{T-2} \phi^{2i} \end{pmatrix}, \quad (5.11)$$

as shown in Appendix C.6 and C.7. Since Σ is a covariance matrix it is positive semi-definite and is therefore invertible.

We can rearrange $p(\epsilon_1 = y_{(1,1)} - \alpha - \beta y_{(2,1)} | \boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s = y_{(1,s)} - \alpha - \beta y_{(2,s)} | \epsilon_{s-1} = y_{(1,s-1)} - \alpha - \beta y_{(2,s-1)}, \boldsymbol{\theta})$, from Equation 5.10, into the matrix form as:

$$\begin{aligned} p(\epsilon_1 | \boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s | \epsilon_{s-1}, \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi)^T |\Sigma|}} \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}' \Sigma^{-1} \boldsymbol{\epsilon} \right\} \\ &= \frac{1}{\sqrt{(2\pi)^T |\Sigma|}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2)' \Sigma^{-1} (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2) \right\} \end{aligned}$$

where $|\Sigma|$ is the determinant of Σ and $\mathbf{1}$ is a $T \times 1$ vector of ones. This can be rewritten as

$$C_1 C_2 \exp \left\{ -\frac{1}{2} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix} \right\}$$

with

$$\begin{aligned} C_1 &= \frac{1}{(2\pi)^{T/2} |\Sigma|^{1/2}} \quad \text{and} \quad (5.12) \\ C_2 &= \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_1' \Sigma^{-1} \mathbf{y}_1 - \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + 2\alpha [c_\alpha + \mu_\alpha^* (\Lambda^{-1})_{11} + \mu_\beta^* (\Lambda^{-1})_{12}] \right. \right. \\ &\quad \left. \left. + 2\beta [c_\beta + \mu_\alpha^* (\Lambda^{-1})_{12} + \mu_\beta^* (\Lambda^{-1})_{22}] \right] \right\}, \end{aligned}$$

where

$$\mathbf{c} = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}_2' \end{pmatrix} \Sigma^{-1} (-\mathbf{y}_1) \quad (5.13)$$

and

$$\Lambda^{-1} = \begin{pmatrix} \mathbf{1}'\Sigma^{-1}\mathbf{1} & \mathbf{1}'\Sigma^{-1}\mathbf{y}_2 \\ \mathbf{y}_2'\Sigma^{-1}\mathbf{1} & \mathbf{y}_2'\Sigma^{-1}\mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}_2' \end{pmatrix} \Sigma^{-1} (\mathbf{1} \ \mathbf{y}_2), \quad (5.14)$$

as shown in Appendix C.8. N.B. that $\mathbf{1}'\Sigma^{-1}\mathbf{y}_2 = \mathbf{y}_2'\Sigma^{-1}\mathbf{1}$ and that the matrix is therefore symmetric and positive definite, since Σ^{-1} is positive definite.

This is the case $\forall \mu_\alpha^*$ and μ_β^* which can be chosen to be

$$\boldsymbol{\mu}^* = \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} = -\Lambda\mathbf{c} \quad (5.15)$$

which results in:

$$C_2 = \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_1'\Sigma^{-1}\mathbf{y}_1 - \boldsymbol{\mu}^{*'}\Lambda^{-1}\boldsymbol{\mu}^* \right] \right\}, \quad (5.16)$$

as detailed in Appendix C.9.

As previously described, the prior of A and B , was chosen to be bivariate normally distributed with mean $\boldsymbol{\mu} = \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}$ and covariance matrix Ω , and thus has the form

$$p(\alpha, \beta) = \frac{1}{\sqrt{(2\pi)^2|\Omega|}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix} \right\} \quad (5.17)$$

Therefore $p(\epsilon_1|\boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s|\epsilon_{s-1}, \boldsymbol{\theta}) p(\alpha, \beta)$ is equivalent to

$$C_1 C_2 \exp \left\{ -\frac{1}{2} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix} \right\} C_3 \exp \left\{ -\frac{1}{2} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix} \right\}$$

with

$$C_3 = \frac{1}{\sqrt{(2\pi)^2|\Omega|}}, \quad (5.18)$$

which is again independent of either A or B . The two parts inside the exponential functions

can be combined and rewritten as:

$$-\frac{1}{2} \left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} + \boldsymbol{\mu}^{*'} \Lambda^{-1} \boldsymbol{\mu}^* + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right]$$

with

$$\boldsymbol{\nu} := \begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} = (\Lambda^{-1} + \Omega^{-1})^{-1} [\Lambda^{-1} \boldsymbol{\mu}^{*'} + \Omega^{-1} \boldsymbol{\mu}] \quad (5.19)$$

as demonstrated in Appendix C.10. This can be rewritten as:

$$\begin{aligned} & \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \boldsymbol{\nu} + \boldsymbol{\nu}^{*'} (\Lambda^{-1} + \Omega^{-1}) \boldsymbol{\nu} \right] \right\} C_4 \\ &= C_4 \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix} \right] \right\} \end{aligned}$$

with

$$C_4 = \exp \left\{ -\frac{1}{2} \left[-\boldsymbol{\nu}' (\Lambda^{-1} + \Omega^{-1}) \boldsymbol{\nu} + \boldsymbol{\mu}^{*'} \Lambda^{-1} \boldsymbol{\mu}^* + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right] \right\} \quad (5.20)$$

This gives:

$$\begin{aligned} & p(\epsilon_1 | \boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s | \epsilon_{s-1}, \boldsymbol{\theta}) p(\alpha, \beta) \\ &= C_1 C_2 C_3 C_4 \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix} \right] \right\} \end{aligned}$$

where C_1 , C_2 , C_3 and C_4 as detailed in Equations 5.12, 5.16, 5.18 and 5.20 respectively, which are all independent of A and B . This means that

$$\begin{aligned} & \int_{(\alpha, \beta) \in \mathbb{R}^2} p(\epsilon_1 | \boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s | \epsilon_{s-1}, \boldsymbol{\theta}) p(\alpha, \beta) d(\alpha, \beta) \\ &= C_1 C_2 C_3 C_4 \int_{(\alpha, \beta) \in \mathbb{R}^2} \exp \left\{ -\frac{1}{2} \left[\begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha - \nu_\alpha \\ \beta - \nu_\beta \end{pmatrix} \right] \right\} d(\alpha, \beta) \\ &= C_1 C_2 C_3 C_4 \end{aligned}$$

as the pdf in the integral is a bivariate normal distribution over A and B , with mean ν_α and ν_β and covariance matrix $(\Lambda^{-1} + \Omega^{-1})^{-1}$ and therefore integrates to 1.

The form of $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2)$ follows:

$$\begin{aligned}
p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2) &= \frac{p(\boldsymbol{\theta}, \mathbf{y}_1|\mathbf{y}_2)}{N} = \frac{1}{N}p(\boldsymbol{\theta})C_1C_2C_3C_4 \\
&= \frac{1}{N}p(\boldsymbol{\theta})\frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left[\mathbf{y}'_1\Sigma^{-1}\mathbf{y}_1 - \boldsymbol{\mu}'\Lambda^{-1}\boldsymbol{\mu}^*\right]\right\} \frac{1}{\sqrt{(2\pi)^2|\Omega|}} \\
&\quad \exp\left\{-\frac{1}{2}\left[-\boldsymbol{\nu}'(\Lambda^{-1} + \Omega^{-1})\boldsymbol{\nu} + \boldsymbol{\mu}'\Lambda^{-1}\boldsymbol{\mu}^* + \boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}\right]\right\}
\end{aligned}$$

which is equivalent to

$$\frac{1}{N}p(\boldsymbol{\theta})\frac{1}{(2\pi)^{\frac{T}{2}+1}\sqrt{|\Sigma||\Omega|}} \exp\left\{-\frac{1}{2}\left[\mathbf{y}'_1\Sigma^{-1}\mathbf{y}_1 - \boldsymbol{\nu}'(\Lambda^{-1} + \Omega^{-1})\boldsymbol{\nu} + \boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}\right]\right\} \quad (5.21)$$

with $p(\boldsymbol{\theta})$, N , Σ , \mathbf{c} , Λ^{-1} , $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ as detailed in Equations 5.6, 5.8, 5.11, 5.13, 5.14, 5.17 and 5.19 respectively and Ω the prior covariance matrix of $\begin{pmatrix} A \\ B \end{pmatrix}$, as described in Section 5.4.

Now that the posterior for the cointegrating parameters given the data is known, the next step is to define the area in which these cointegrating parameters would indicate cointegration. Once this area is known, the proportion of the posterior consistent with \mathbf{y}_1 and \mathbf{y}_2 being cointegrated can be calculated by integrating the posterior over this area. In this report we call this area the ‘cointegration tube’ and it is defined and detailed in the next section.

5.6 Cointegration Tube

The definition of cointegration is based on weak stationarity². Weak stationarity itself is characterised by a constant mean and variance and a covariance that only depends on lag, not on the time point (see Section 4.2.1). These three constraints define a curved plane

²weak stationarity is an approximation of strict stationarity, which can only be tested in the rarest cases [30]

in a three dimensional space (see Appendix C.11) and the posterior probability mass of a curved plane in a three dimensional space is zero.

Even if two time series were perfectly cointegrated we would expect the posterior pdf not to lie perfectly on a curved plane. With more samples the posterior will concentrate around this plane and we expect that a suitably scaled tube around this plane will contain most of the posterior probability mass if the time series are in fact cointegrated. Such a tube could look similar to the sketch in Figure 5.2.

Because of this, a parameter ζ is introduced to allow for small deviations from perfect cointegration. ζ defines the width of what we call the cointegration tube. If say 90% of the probability mass of the combination of the three parameters lies within the cointegration tube, the null hypothesis that the two time series are not cointegrated is rejected. If less than that lies within the cointegration tube, the null hypothesis is not rejected.

To recap, the stationarity restrictions mentioned at the beginning of this section are that the linear combination of the two time series has a constant mean, constant variance and time independent covariance.

The constant mean does not create a restriction on the parameters, because $\mathbb{E}(E_1) = 0$, as defined in Equation 5.3. This means that $\mathbb{E}(E_t) = 0$ for all t , since $\mathbb{E}(E_t) = \phi\mathbb{E}(E_{t-1}) + \mathbb{E}(H_t) = \dots = \phi^{t-1}\mathbb{E}(E_1) = 0$, and therefore the mean will always be constant. Furthermore, variance is a special case of covariance and, therefore, will be dealt with as part of the restrictions associated with the constant covariance.

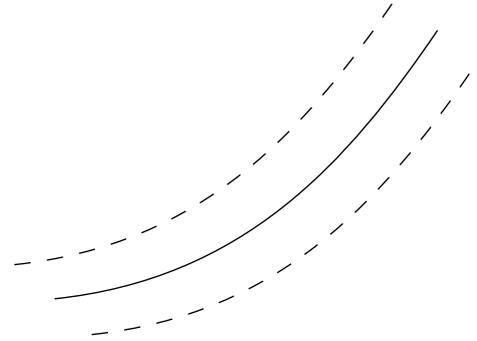


Figure 5.2: Sketch of the area around the line that describes perfect cointegration, which might include 95% of the posterior pdf.

The constraints on the parameters come from the need for constant covariance over time. If we allow the covariances to deviate slightly (ζ) from being perfectly constant we get:

$$\begin{aligned}
\zeta &\geq \max_{t,k} |\text{Cov}(E_1, E_{1+k}) - \text{Cov}(E_t, E_{t+k})| \\
&= \max_{t,k} |\text{Cov}(E_1, \Phi E_k + H_{k+1}) - \text{Cov}(E_t, \Phi E_{t+k-1} + H_{t+k})| \\
&= \max_{t,k} |\Phi \text{Cov}(E_1, \Phi E_{k-1} + H_k) - \Phi \text{Cov}(E_t, \Phi E_{t+k-2} + H_{t+k-1})| \\
&\quad \vdots \\
&= \max_{t,k} |\Phi^k \text{Cov}(E_1, E_1) - \Phi^k \text{Cov}(E_t, E_t)| \\
&= \max_{t,k} |\Phi^k (\text{Var}(E_1) - \text{Var}(E_t))|
\end{aligned}$$

since $\text{Var}(H_t) = S^2$, $\text{Var}(E_1) = S_1^2$ and $\text{Cov}(E_{t-k}, H_t) = 0$ (as detailed in Equation 5.3).

In Appendix C.12 we show that this is equivalent to

$$\Rightarrow \zeta \geq \max_{t,k} \left| \Phi^k \left((1 - \Phi^{2(t-1)}) S_1^2 - \sum_{i=0}^{t-2} \Phi^{2i} S^2 \right) \right| \quad (5.22)$$

The three defining parameters in this constraint are Φ , S^2 and S_1^2 . We say \mathbf{y}_1 and \mathbf{y}_2 are approximately cointegrated if Equation 5.22 is fulfilled. Figure 5.3 shows an example of heat maps of cointegrated and not cointegrated time series, with the cointegration line (white dashed line) and the boundaries of the cointegration tube (yellow solid lines).

The first is from cointegrated time series, the combination of (Φ, S^2, S_1^2) used to create the time series related to the second heat plot are 0.1 units away from the cointegration line, and the parameters linked to the third heat map are 0.2 units away from the cointegration line. A sketch of this set up is displayed in Figure 5.4(i).

The choice of ζ is crucial in the analysis. If it is chosen too large, the probability of a type I error³ increases. Conversely, if it is chosen too narrowly, the probability of a type

³the type I error is the incorrect rejection of the null hypothesis, i.e. rejecting no cointegration, although

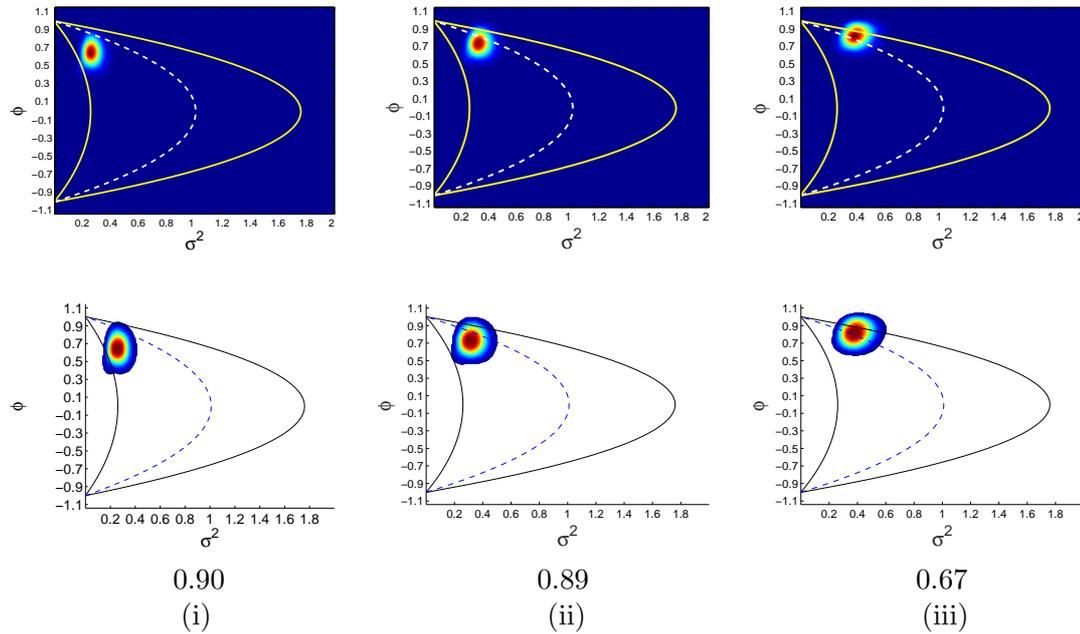


Figure 5.3: This plot shows the realisation of the Sketch in Figure 5.4(i). (i) The two plots on the left show an example of a heat map (top) and the top 95% of points that make up the heat map (bottom) for cointegrated time series. (ii) The two middle plots show the same graphs for time series that were created from parameters 0.1 units away from the cointegration line. (iii) The two plots on the right shows the graphs for time series created from parameters 0.2 units away from the cointegrated line. The numbers underneath the plots show the proportion of the posterior within the cointegration tube.

II error⁴ increases and small amounts of noise in the measurements could result in failing to reject the null hypothesis, even though \mathbf{y}_1 and \mathbf{y}_2 are cointegrated.

To get an estimate of an appropriate ζ cointegrated time series are simulated. The posterior probability density is then calculated for each of the simulations and the average over all the simulations is plotted on a heat map. The heat map is then overlaid with the boundaries implying the cointegration tube for various values of ζ , to find the ζ that approximates the cointegration tube the best. The combinations of $\Theta = (\Phi, S^2, S_1^2)$ inside this cointegration tube indicate approximate cointegration given ζ . An example of these plots is shown in Figure 5.3. Here the time series were simulated with $\phi = 0.704$, $\sigma^2 = 0.51$ and $\sigma_1^2 = 1.01$ and the cointegration tube was calculated using $\zeta = 0.75$. This was found

the time series are not cointegrated

⁴the type II error is the failure to reject a false null hypothesis.

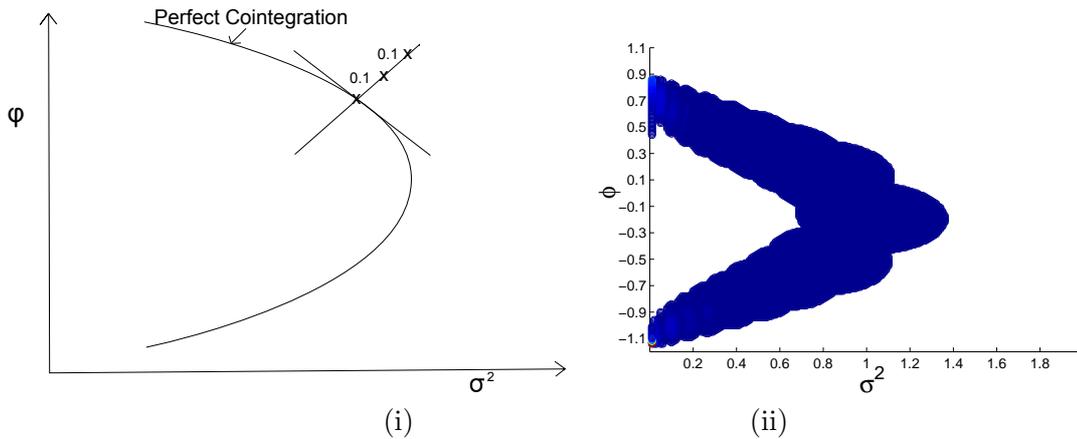


Figure 5.4: (i) This sketch shows the setup of the graphs in Figure 5.3. Time series are simulated that are cointegrated, i.e. the combination of the parameters used to create the time series, lie on the cointegration line. Two more pairs of time series are created which are 0.1 and 0.2 units away from the cointegration line and are therefore not cointegrated. (ii) This graph shows a collection of the top 90% of the heat maps of posterior densities of cointegrated time series with σ_1^2 fixed at 1.01.

to be the best fit given the observed time series are of length $T = 100$, and that the true prior distributions of ϕ , σ^2 and σ_1^2 are $\mathcal{N}(0, 1)$, $InvGamma(1, 1)$ and $InvGamma(1, 1)$ respectively.

For a further visualisation, σ_1^2 was fixed at 1.01 and the posterior density was calculated for 21 cointegrated combinations of ϕ and σ^2 . The collection of the top 90% of the heat maps of the posteriors is shown in Figure 5.4(ii).

To create the cointegrated time series, first two of the three parameters, Φ , S^2 and S_1^2 , are chosen and the third is calculated. In this thesis we chose S^2 and S_1^2 and calculated Φ , as described in Appendix C.11. This was done for all S^2 running from 0.01 to 2.01 in steps of 0.01. For each combination, the time series \mathbf{y}_1 and \mathbf{y}_2 are generated according to the data generating equations described in Section 5.2. Following this, the posterior probability is calculated as described in Section 5.5.

For each of the combinations the points that make up the top 90% of the heat map are selected and plotted on a figure. A collection of the resulting posterior densities is plotted in Figure 5.4(ii).

5.7 Numerical Integration

Now that we have the posterior probability density function of (Φ, S, S_1) given the data and the cointegration tube, the next step is to integrate the posterior over the cointegration tube.

The Riemann sum [62] is used for this integration, as an upper bound on the error can be calculated using no more than the first derivative. The one dimensional Riemann sum is defined in Section 4.2.4. The general idea is that the volume over which we wish to integrate is divided into sub-blocks. For each of the sub-blocks the area under the surface we wish to integrate over is estimated by multiplying the size of the sub-block by the value of the surface at a previously defined point of the sub-block. In our case this will be the corner closest to $(-\infty, -\infty, -\infty)$.

Once the estimates and the upper bounds on the errors for each sub-block have been calculated, the error is compared to a predefined threshold. If the upper bound on the error is greater than that threshold, the sub-block with the greatest error is divided into eight equally sized sub-blocks and the estimate and upper bound on the error is recalculated.

5.7.1 Error Estimates

The estimate of the integral per sub-block is the pdf at the point of the sub-block that is the closest to $(-\infty, -\infty, -\infty)$ multiplied by the size of the sub-block. Therefore an upper bound on the error (E_m) is the greatest possible change in the value of the pdf within the sub-block, again multiplied by the size of the sub-block.

$$\begin{aligned} E_m &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{a_3}^{b_3} f(\mathbf{x}) d\mathbf{x} - S \\ &= \sum_j^m (b_{1,j} - a_{1,j})^2 (b_{2,j} - a_{2,j})^2 (b_{3,j} - a_{3,j})^2 f'(\mathbf{c}), \end{aligned}$$

where m is the number of sub-blocks, S is the left-endpoint Riemann sum, $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ are the endpoints of the sub-block and \mathbf{c} is the point where the derivative of f is at its maximum between \mathbf{a} and \mathbf{b} [31], or equivalently where the norm between $f(\mathbf{a})$ and $f(\mathbf{c})$ is the greatest, within the sub-block.

It is not feasible to compute the second and third derivative of the posterior pdf, to be able to calculate the maximum first derivative within each sub-block. Instead an upper bound on the error of this estimate is calculated using perturbation theory [52]. It is added to and subtracted from the estimate to achieve the upper- and lower- bound on the integral, which in turn tells us what proportion of the posterior density lies within the cointegration tube.

5.7.1.1 Upper Bound on the Error

The bound on the effect a small change in $\boldsymbol{\theta}$ has on the posterior pdf $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2)$ is calculated for each sub-block. This will be indicated by

$$\left| p(\tilde{\boldsymbol{\theta}}_i|\mathbf{y}_1, \mathbf{y}_2) - p(\boldsymbol{\theta}_i|\mathbf{y}_1, \mathbf{y}_2) \right| \quad (5.23)$$

where $\tilde{\boldsymbol{\theta}}_i$ represents any point $\boldsymbol{\theta}$ within the i th sub-block and $\boldsymbol{\theta}_i$ stands for the corner point of the i th sub-block that is closest to $(-\infty, -\infty, -\infty)$.

As $p(\boldsymbol{\theta}|\mathbf{y}_1, \mathbf{y}_2) = p(\boldsymbol{\theta}, \mathbf{y}_1|\mathbf{y}_2)/N$, where $N = \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}, \mathbf{y}_1|\mathbf{y}_2)$ as described in Equations 5.7 and 5.8, Equation 5.23 can be rewritten as

$$\left| p(\tilde{\boldsymbol{\theta}}_i|\mathbf{y}_1, \mathbf{y}_2) - p(\boldsymbol{\theta}_i|\mathbf{y}_1, \mathbf{y}_2) \right| = \left| \frac{p(\tilde{\boldsymbol{\theta}}_i, \mathbf{y}_1|\mathbf{y}_2) - p(\boldsymbol{\theta}_i, \mathbf{y}_1|\mathbf{y}_2)}{N} \right|.$$

A lower bound on N is

$$\int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}, \mathbf{y}_1|\mathbf{y}_2) \geq \sum_i \mu(W_i) \left(p(\boldsymbol{\theta}_i, \mathbf{y}_1|\mathbf{y}_2) - \sup_{\tilde{\boldsymbol{\theta}}_i \in W_i} \left| p(\tilde{\boldsymbol{\theta}}_i, \mathbf{y}_1|\mathbf{y}_2) - p(\boldsymbol{\theta}_i, \mathbf{y}_1|\mathbf{y}_2) \right| \right)$$

where W_i represents the i th sub-block and $\mu(W_i)$ the area of the i th sub-block. Which means that calculating $\left|p(\tilde{\boldsymbol{\theta}}_i, \mathbf{y}_1 | \mathbf{y}_2) - p(\boldsymbol{\theta}_i, \mathbf{y}_1 | \mathbf{y}_2)\right|$ per sub-block will suffice to get a lower bound on N and together this will give the upper bound on $\left|p(\tilde{\boldsymbol{\theta}}_i | \mathbf{y}_1, \mathbf{y}_2) - p(\boldsymbol{\theta}_i | \mathbf{y}_1, \mathbf{y}_2)\right|$.

Because of the nested nature of the posterior it is necessary to first examine what effect a small change in $\boldsymbol{\theta}$ has on Σ and the prior $p(\boldsymbol{\theta})$. Similarly to the chain-rule for derivatives of nested functions, the effect has to then be calculated up the chain step by step.

Before going into the details of this a high-level overview of this chain is given and the main steps in the calculation of the upper bound on $\left|p(\tilde{\boldsymbol{\theta}}_i, \mathbf{y}_1 | \mathbf{y}_2) - p(\boldsymbol{\theta}_i, \mathbf{y}_1 | \mathbf{y}_2)\right|$. Figure 5.5 visualises the chain of upper bounds calculated in the following sections and their dependencies on each other.

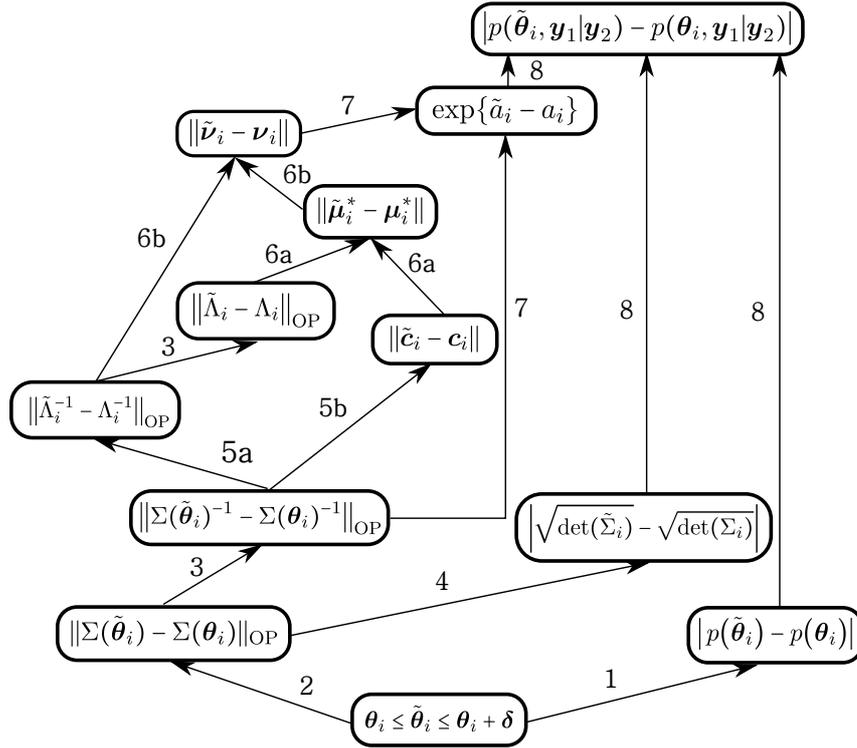


Figure 5.5: Flowchart of the Chain of Upper Bounds

1. The first upper bound on $|p(\tilde{\boldsymbol{\theta}}_i) - p(\boldsymbol{\theta}_i)|$ is fairly straight forward as it can be calculated directly from $\boldsymbol{\theta}_i \leq \tilde{\boldsymbol{\theta}}_i \leq \boldsymbol{\theta}_i + \delta$.

2. When dealing with matrix norms, such as $\|\Sigma(\tilde{\boldsymbol{\theta}}) - \Sigma(\boldsymbol{\theta})\|$, the operator norm, defined in Section 4.2.5, is used. In this case the definition of the operator norm was used to get $\|A\|_{\text{OP}} = \sup_{\|\mathbf{v}\|=1} \{\|A\mathbf{v}\|\} = \sup_{\|\mathbf{v}\|=1} \left\{ \sup_{\|\mathbf{w}\|=1} \{\mathbf{w}'A\mathbf{v}\} \right\}$. An upper bound is then calculated by first finding the maximum with respect to \mathbf{w} and then with respect to \mathbf{v} .
3. When calculating $\|\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\|_{\text{OP}}$ and $\|\tilde{\Lambda}_i - \Lambda_i\|_{\text{OP}}$ (see Sections 5.7.1.1.3 and C.15) the following two inequalities are used, for which the conditions are detailed in Section 4.2.5.

$$\begin{aligned} \text{(i)} \quad & \|G^{-1} - F^{-1}\|_{\text{OP}} \leq \|G^{-1}\|_{\text{OP}} \|F - G\|_{\text{OP}} \|F^{-1}\|_{\text{OP}} \\ \text{(ii)} \quad & \|G^{-1}v\| \leq \left(\frac{1}{\|F^{-1}\|_{\text{OP}}} - \|G - F\|_{\text{OP}} \right)^{-1} \|v\| \end{aligned}$$

These inequalities require F and G to be invertible and that $\|F - G\|_{\text{OP}} \leq \|F^{-1}\|_{\text{OP}}^{-1}$. These requirements put restrictions on the size of the sub-blocks which are detailed in Section 5.7.1.1.11.

4. For the upper bound on $|\sqrt{\det \tilde{\Sigma}_i} - \sqrt{\det \Sigma_i}|$ (see Section 5.7.1.1.4) the fact that the determinant of a matrix is equivalent to the product of its eigenvalues is used. How much the eigenvalues of $\tilde{\Sigma}$ can vary is determined using eigenvalue perturbation theory.
5. The upper bound on $\|\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\|_{\text{OP}}$ is then used to calculate the upper bound on both $\|\tilde{\Lambda}^{-1} - \Lambda^{-1}\|_{\text{OP}}$ (arrow 5a; see Section 5.7.1.1.5) and $\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\|$ (arrow 5b; see 5.7.1.1.6), where Λ is a 2×2 covariance matrix and \mathbf{c} is a 2×1 vector. Both depend on $\mathbf{y}_1, \mathbf{y}_2$ and Σ^{-1} and therefore present a large reduction in dimensionality from a $T \times T$ matrix and $T \times 1$ vector to a 2×2 matrix and a 2×1 vector. For the upper bound on $\|\tilde{\Lambda}^{-1} - \Lambda^{-1}\|_{\text{OP}}$ the distributive property of matrices and the sub-multiplicativity of the operator norm are used. The upper bound on $\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\|$ is calculated by first using the definition of the operator norm, as described in the next paragraph, which is followed by using the sub-multiplicativity of the operator norm.

To relate the upper bounds on the vectors to the previously calculated upper bounds on the matrices the definition of the operator norm is used. Let $A \in \mathbb{R}^{n \times n}$ and $a \in \mathbb{R}^{n \times 1}$. Further assume an upper bound on $\|Aa\|$ is needed and both $\|A\|_{\text{OP}}$ and $\|a\|$ are known. Then we can use the definition of the operator norm, defined in Section 4.2.5.

$$\|Aa\| = \left\| A \frac{a}{\|a\|} \right\| \|a\| \leq \sup_{\|u\|=1} \|Au\| \|a\| = \|A\|_{\text{OP}} \|a\|.$$

6. This method is used to calculate $\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\|$ (arrow 6a; see Section 5.7.1.1.7) from $\|\tilde{\Lambda} - \Lambda\|_{\text{OP}}$ and $\|\tilde{\boldsymbol{c}}_i - \boldsymbol{c}_i\|$, and $\|\tilde{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|$ (arrows 6b; see Section 5.7.1.1.8) from $\|\tilde{\Lambda}^{-1} - \Lambda^{-1}\|_{\text{OP}}$ and $\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\|$, where $\boldsymbol{\mu}^*$ and $\boldsymbol{\nu}$ are again 2×1 vectors dependent on Λ and \boldsymbol{c} , and Λ^{-1} and $\boldsymbol{\mu}^*$ respectively. The calculations cause further restrictions on the size of the sub-blocks, for example because $\tilde{\Lambda}$ needs to be invertible.
7. An upper bound on the exponential function of a difference, denoted by $\exp\{\tilde{a}_i - a_i\}$ (see Section 5.7.1.1.9), is calculated next. a , a scalar, depends on Σ^{-1} and $\boldsymbol{\nu}$ and $|\tilde{a}_i - a_i|$ therefore depends on $\|\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\|_{\text{OP}}$ and $\|\tilde{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|$. First the triangle rule is used then the Cauchy-Schwarz inequality, $|\langle x, y \rangle| \leq \|x\| \|y\|$ [54], and then again the definition of the operator norm is used.
8. Finally, the upper bound on the exponential function together with the upper bound on $|\sqrt{\det \tilde{\Sigma}_i} - \sqrt{\det \Sigma_i}|$ and $|p(\tilde{\boldsymbol{\theta}}_i) - p(\boldsymbol{\theta}_i)|$ are used to get the upper bound on $|p(\tilde{\boldsymbol{\theta}}_i, \boldsymbol{y}_1 | \boldsymbol{y}_2) - p(\boldsymbol{\theta}_i, \boldsymbol{y}_1 | \boldsymbol{y}_2)|$ (see Section 5.7.1.1.10). This is done by again first using the triangle rule followed by the sub-multiplicativity of the absolute value.

Following are the details of the calculations of each of the upper bounds.

5.7.1.1.1 Upper bound on the effect of a small change in $\boldsymbol{\theta}$ on $p(\boldsymbol{\theta})$: The size of the change in $\boldsymbol{\theta} = (\phi, \sigma^2, \sigma_1^2)'$ is defined by the size of the sub-blocks described at the beginning of Section 5.7. For the rest of this section let $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3)'$ denote the size of

the sub-block such that

$$\begin{aligned}\phi_i &\leq \tilde{\phi}_i \leq \phi_i + \delta_1, \\ \sigma_i^2 &\leq \tilde{\sigma}_i^2 \leq \sigma_i^2 + \delta_2, \text{ and} \\ \sigma_{1,i}^2 &\leq \tilde{\sigma}_{1,i}^2 \leq \sigma_{1,i}^2 + \delta_3.\end{aligned}\tag{5.24}$$

To quantify the effect a small change in θ_i has on $p(\theta_i)$ the absolute value of the difference between $p(\tilde{\theta}_i)$, the perturbed probability, and $p(\theta_i)$, the known value, is used, i.e.

$$\left| p(\tilde{\theta}_i) - p(\theta_i) \right|.$$

As detailed in Equation 5.6 $p(\theta) = \frac{\tau^\kappa \sigma^{-\kappa-1} \tau_1^{\kappa_1} \sigma_1^{-\kappa_1-1}}{\sigma_\phi \sqrt{2\pi} \Gamma(\kappa) \Gamma(\kappa_1)} \exp \left\{ -\frac{(\phi - \mu_\phi)^2}{2\sigma_\phi^2} - \frac{\tau}{\sigma} - \frac{\tau_1}{\sigma_1} \right\}$ which means that

$$\begin{aligned}&\left| p(\tilde{\theta}_i) - p(\theta_i) \right| \\ &= \left| \frac{\tau^\kappa \tau_1^{\kappa_1} (\tilde{\sigma}_i^2)^{-\kappa-1} (\tilde{\sigma}_{1,i}^2)^{-\kappa_1-1}}{\sigma_\phi \sqrt{2\pi} \Gamma(\kappa) \Gamma(\kappa_1)} \exp \left\{ -\frac{(\tilde{\phi}_i - \mu_\phi)^2}{2\tilde{\sigma}_\phi^2} \right\} \exp \left\{ -\frac{\tau}{\tilde{\sigma}_i^2} \right\} \exp \left\{ -\frac{\tau_1}{\tilde{\sigma}_{1,i}^2} \right\} \right. \\ &\quad \left. - \frac{\tau^\kappa \tau_1^{\kappa_1} (\sigma_i^2)^{-\kappa-1} (\sigma_{1,i}^2)^{-\kappa_1-1}}{\sigma_\phi \sqrt{2\pi} \Gamma(\kappa) \Gamma(\kappa_1)} \exp \left\{ -\frac{(\phi_i - \mu_\phi)^2}{2\sigma_\phi^2} - \frac{\tau}{\sigma_i^2} - \frac{\tau_1}{\sigma_{1,i}^2} \right\} \right|\end{aligned}$$

From Equation 5.24 it is known that $\sigma_i^2 \leq \tilde{\sigma}_i^2 \leq \sigma_i^2 + \delta_2$ which means that

$$(\sigma_i^2 + \delta_2)^{-\kappa-1} \leq (\tilde{\sigma}_i^2)^{-\kappa-1} \leq (\sigma_i^2)^{-\kappa-1}\tag{5.25}$$

since $\kappa > 0$ and therefore $-\kappa - 1 < -1$.

Furthermore, because the exponential function increases monotonically

$$\exp \left\{ -\frac{\tau}{\sigma_i^2} \right\} \leq \exp \left\{ -\frac{\tau}{\tilde{\sigma}_i^2} \right\} \leq \exp \left\{ -\frac{\tau}{\sigma_i^2 + \delta_2} \right\}.\tag{5.26}$$

Both Equations (5.25) and (5.26) are trivially adapted to give the bounds on $(\tilde{\sigma}_{1,i}^2)^{-\kappa_1-1}$

and $\exp\left\{-\frac{\tau_1}{\sigma_{1,i}^2}\right\}$. Additionally we define

$$f_1 := \begin{cases} \exp\left\{-\frac{(\phi_i - \mu_\phi)^2}{2\sigma_\phi^2}\right\} & \text{if } \mu_\phi < \phi_i \\ \exp\left\{-\frac{(\phi_i + \delta_1 - \mu_\phi)^2}{2\sigma_\phi^2}\right\} & \text{if } \mu_\phi > \phi_i + \delta_1 \\ 1 & \text{if } \phi_i \leq \mu_\phi \leq \phi_i + \delta_1 \end{cases}$$

which is an upper bound on $\exp\left\{-\frac{(\tilde{\phi}_i - \mu_\phi)^2}{2\sigma_\phi^2}\right\}$.

From this follows the upper bound on $\left|p(\tilde{\theta}_i) - p(\theta_i)\right|$

$$\left| \frac{\tau^\kappa \tau_1^{\kappa_1} (\sigma_i^2)^{-\kappa-1} (\sigma_{1,i}^2)^{-\kappa_1-1}}{\sigma_\phi \sqrt{2\pi} \Gamma(\kappa) \Gamma(\kappa_1)} \left(f_1 \exp\left\{-\frac{\tau}{\sigma_i^2 + \delta_2} - \frac{\tau_1}{\sigma_{1,i}^2 + \delta_3}\right\} - \exp\left\{-\frac{(\phi_i - \mu_\phi)^2}{2\sigma_\phi^2} - \frac{\tau}{\sigma_i^2} - \frac{\tau_1}{\sigma_{1,i}^2}\right\} \right) \right|.$$

5.7.1.1.2 Upper bound on the effect of a small change in θ on $\Sigma(\theta)$: To quantify the effect a small change in θ_i has on $\Sigma(\theta_i)$ we calculate an upper bound on the operator norm (introduced in Definition 4.5) of the difference between $\Sigma(\tilde{\theta}_i)$ and $\Sigma(\theta_i)$. From Equation 5.11 the form of $\Sigma(\theta_i)$ and therefore $\Sigma(\tilde{\theta}_i) - \Sigma(\theta_i)$ is known.

In Appendix C.13 it is shown that an upper bound on $\|\tilde{\Sigma} - \Sigma\|_{\text{OP}}$ is $\sqrt{\sum_{i=1}^T \sum_{j=1}^T a_{ij}^2}$, where a_{ij} is the element of $\Sigma(\theta + \delta) - \Sigma(\theta)$ on the i th row and the j th column.

5.7.1.1.3 Upper bound on the effect of a small change in θ on $\Sigma(\theta)^{-1}$: The next step is to get an upper bound on the effect a small change in θ has on the inverse of $\Sigma(\theta)$, i.e.

$$\left\| \Sigma(\tilde{\theta}_i)^{-1} - \Sigma(\theta_i)^{-1} \right\|_{\text{OP}}.$$

For this, the following inequality, based on a result by Ferus [22] which has been detailed in Equation 4.3, is used. It gives an upper bound on the operator norm of the difference

between two invertible matrices⁵.

$$\left\| \Sigma(\tilde{\boldsymbol{\theta}}_i)^{-1} - \Sigma(\boldsymbol{\theta}_i)^{-1} \right\|_{\text{OP}} \leq \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i)^{-1} \right\|_{\text{OP}} \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}} \left\| \Sigma(\boldsymbol{\theta}_i)^{-1} \right\|_{\text{OP}}. \quad (5.27)$$

An upper bound on $\left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}}$ has been calculated in Section 5.7.1.1.2. $\left\| \Sigma(\boldsymbol{\theta}_i)^{-1} \right\|_{\text{OP}}$ is also known, as it is just the inverse of the smallest eigenvalue of $\Sigma(\boldsymbol{\theta}_i)$ as detailed in Section 4.2.5.1.

To get $\left\| \Sigma(\tilde{\boldsymbol{\theta}}_i)^{-1} \right\|_{\text{OP}}$ a different result from Ferus [22] (detailed in Equation 4.4) is needed as well as the definition of the operator norm⁶

$$\begin{aligned} \|G^{-1}\|_{\text{OP}} &= \sup_{\|v\|=1} \|G^{-1}v\| \\ &\leq \left(\frac{1}{\|F^{-1}\|_{\text{OP}}} - \|G - F\|_{\text{OP}} \right)^{-1} \underbrace{\sup_{\|v\|=1} \|v\|}_{=1} \\ &= \left(\frac{1}{\|F^{-1}\|_{\text{OP}}} - \|G - F\|_{\text{OP}} \right)^{-1} \end{aligned}$$

which means that

$$\begin{aligned} \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i)^{-1} \right\|_{\text{OP}} &\leq \left(\frac{1}{\left\| \Sigma(\boldsymbol{\theta}_i)^{-1} \right\|_{\text{OP}}} - \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}} \right)^{-1} \\ &= \left(\min_{1 \leq j \leq T} |\lambda_{i,j}| - \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}} \right)^{-1} \end{aligned} \quad (5.28)$$

where $\lambda_{i,j}$ is the j th eigenvalue of $\Sigma(\boldsymbol{\theta}_i)$ and an upper bound on $\left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}}$ has been calculated in Section 5.7.1.1.2.

In summary this shows that an upper bound on $\left\| \Sigma(\tilde{\boldsymbol{\theta}}_i)^{-1} - \Sigma(\boldsymbol{\theta}_i)^{-1} \right\|_{\text{OP}}$ is

$$\left(\min_{1 \leq j \leq T} |\lambda_{i,j}| - \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}} \right)^{-1} \left\| \Sigma(\tilde{\boldsymbol{\theta}}_i) - \Sigma(\boldsymbol{\theta}_i) \right\|_{\text{OP}} \frac{1}{\min_{1 \leq j \leq T} |\lambda_{i,j}|} \quad (5.29)$$

⁵Since the size of the boxes used to get an upper bound on the integral is calculated to make sure that both Σ_i and $\tilde{\Sigma}_i$ are invertible (5.7.1.1.11) this result can be used

⁶The definitions of $\|A\|_{\text{OP}}$ is $\sup\{\|Av\| : v \in V \text{ with } \|v\| = 1\}$ (see Section 4.2.5)

where now all the individual components are known.

5.7.1.1.4 Upper bound on the effect of a small change in Σ on $\sqrt{\det(\Sigma)}$: For clarity a shorthand for $\Sigma(\boldsymbol{\theta}_i) =: \Sigma_i$ and $\Sigma(\tilde{\boldsymbol{\theta}}_i) =: \tilde{\Sigma}_i$ is introduced here. Similar abbreviations will be used in the following sections and should be intuitive from the context.

To quantify the effect a small change in Σ_i has on $\sqrt{\det(\Sigma_i)}$ the absolute value is used, since the determinant of a matrix is a scalar.

$$\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| = \left| \sqrt{\det(\Sigma_i)} \right| \left| \sqrt{\frac{\det(\tilde{\Sigma}_i)}{\det(\Sigma_i)}} - 1 \right|$$

The determinant of a matrix is equivalent to the product of the eigenvalues of that matrix, counting multiplicity [4]

$$\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| = \left| \sqrt{\det(\Sigma_i)} \right| \left| \sqrt{\prod_{j=1}^T \frac{\tilde{\lambda}_{i,j}}{\lambda_{i,j}}} - 1 \right|$$

where $\lambda_{i,j}$ represents the j th eigenvalue of Σ_i .

The definition of the absolute value is used:

$$\begin{aligned} & \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| \\ &= \max \left\{ \left| \sqrt{\det(\Sigma_i)} \right| \left(\sqrt{\prod_{j=1}^T \frac{\tilde{\lambda}_{i,j}}{\lambda_{i,j}}} - 1 \right), \left| \sqrt{\det(\Sigma_i)} \right| \left(1 - \sqrt{\prod_{j=1}^T \frac{\tilde{\lambda}_{i,j}}{\lambda_{i,j}}} \right) \right\} \end{aligned}$$

Since Σ_i is positive definite, all its eigenvalues are greater than 0 and therefore

$$\begin{aligned} & \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| \\ &= \max \left\{ \left| \sqrt{\det(\Sigma_i)} \right| \left(\sqrt{\prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|}} - 1 \right), \left| \sqrt{\det(\Sigma_i)} \right| \left(1 - \sqrt{\prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|}} \right) \right\} \end{aligned}$$

This allows the inequality $\prod_{j=1}^T \frac{|\lambda_{i,j}-e_i|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\lambda_{i,j}+e_i|}{|\lambda_{i,j}|}$ computed in the Appendix C.14 to be used to get:

$$\begin{aligned} & \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| \\ & \leq \max \left\{ \left| \sqrt{\det(\Sigma_i)} \right| \left(\sqrt{\prod_{j=1}^T \frac{|\lambda_{i,j}+e_i|}{|\lambda_{i,j}|}} - 1 \right), \left| \sqrt{\det(\Sigma_i)} \right| \left(1 - \sqrt{\prod_{j=1}^T \frac{|\lambda_{i,j}-e_i|}{|\lambda_{i,j}|}} \right) \right\} \end{aligned}$$

with $e_i = \left\| \tilde{\Sigma}_i - \Sigma_i \right\|_{\text{OP}}$ as detailed in C.14 and all other components are known.

5.7.1.1.5 Upper bound on the effect of a small change in Σ^{-1} on Λ^{-1} : As defined in Equation 5.14, Λ^{-1} is equivalent to $\begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \Sigma^{-1} (\mathbf{1} \ \mathbf{y}_2)$. Due to the distributive property of matrices, the operator norm of the difference is

$$\left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} = \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) (\mathbf{1} \ \mathbf{y}_2) \right\|_{\text{OP}}$$

and the sub-multiplicativity of the operator norm, as detailed in Section 4.2.5.2, gives an upper bound for which all the individual elements are known

$$\left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} \leq \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \right\|_{\text{OP}} \left\| \tilde{\Sigma}_i^{-1} - \Sigma_i^{-1} \right\|_{\text{OP}} \left\| (\mathbf{1} \ \mathbf{y}_2) \right\|_{\text{OP}}. \quad (5.30)$$

5.7.1.1.6 Upper bound on the effect of a small change in Σ^{-1} on \mathbf{c} : \mathbf{c} , as defined in Equation 5.13, can be decomposed into a product of the matrices $(\mathbf{1}, \mathbf{y}_2)$ and Σ^{-1} and the vector $-\mathbf{y}_1$:

$$\mathbf{c} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \Sigma^{-1} (-\mathbf{y}_1)$$

This is used to calculate an upper bound on the perturbation of \mathbf{c}_i

$$\begin{aligned} \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| &= \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \tilde{\Sigma}_i^{-1} (-\mathbf{y}_1) - \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \Sigma_i^{-1} (-\mathbf{y}_1) \right\| \\ &= \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) (-\mathbf{y}_1) \right\| \end{aligned}$$

Multiplying this equation by 1 in the form of $\|\mathbf{y}_1\|/\|\mathbf{y}_1\|$ lets us use the definition of the operator norm

$$\begin{aligned}\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| &= \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \frac{(-\mathbf{y}_1)}{\|\mathbf{y}_1\|} \right\| \|\mathbf{y}_1\| \\ &\leq \sup_{\|\mathbf{u}\|=1} \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \mathbf{u} \right\| \|\mathbf{y}_1\| \\ &= \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \right\|_{\text{OP}} \|\mathbf{y}_1\|\end{aligned}$$

As the operator norm is a sub-multiplicative norm (as detailed in Section 4.2.5.2) we can use the fact that $\|AB\|_{\text{OP}} \leq \|A\|_{\text{OP}}\|B\|_{\text{OP}}$ which results in

$$\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| \leq \left\| \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \right\|_{\text{OP}} \left\| \tilde{\Sigma}_i^{-1} - \Sigma_i^{-1} \right\|_{\text{OP}} \|\mathbf{y}_1\|$$

5.7.1.1.7 Upper bound on the effect of a small change in Λ and \mathbf{c} has on $\boldsymbol{\mu}^*$:

The next step is to estimate the effect of a perturbation on $\boldsymbol{\mu}^*$, which has been defined in Equation 5.15 to be $\boldsymbol{\mu}^* = -\Lambda\mathbf{c}$. From this follows

$$\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| = \left\| \tilde{\Lambda}_i \tilde{\mathbf{c}}_i - \Lambda_i \mathbf{c}_i \right\|$$

Zero can be added in the form of $-\Lambda_i \tilde{\mathbf{c}}_i + \Lambda_i \tilde{\mathbf{c}}_i$:

$$\begin{aligned}\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| &= \left\| \tilde{\Lambda}_i \tilde{\mathbf{c}}_i - \Lambda_i \tilde{\mathbf{c}}_i + \Lambda_i \tilde{\mathbf{c}}_i - \Lambda_i \mathbf{c}_i \right\| \\ &= \left\| (\tilde{\Lambda}_i - \Lambda_i) \tilde{\mathbf{c}}_i + \Lambda_i (\tilde{\mathbf{c}}_i - \mathbf{c}_i) \right\|\end{aligned}$$

and again in the form of $-(\tilde{\Lambda}_i - \Lambda_i) \mathbf{c}_i + (\tilde{\Lambda}_i - \Lambda_i) \mathbf{c}_i$

$$\begin{aligned}\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| &= \left\| (\tilde{\Lambda}_i - \Lambda_i) \tilde{\mathbf{c}}_i + \Lambda_i (\tilde{\mathbf{c}}_i - \mathbf{c}_i) - (\tilde{\Lambda}_i - \Lambda_i) \mathbf{c}_i + (\tilde{\Lambda}_i - \Lambda_i) \mathbf{c}_i \right\| \\ &= \left\| (\tilde{\Lambda}_i - \Lambda_i) (\tilde{\mathbf{c}}_i - \mathbf{c}_i) + \Lambda_i (\tilde{\mathbf{c}}_i - \mathbf{c}_i) + (\tilde{\Lambda}_i - \Lambda_i) \mathbf{c}_i \right\|\end{aligned}$$

Now the triangle rule is applied to get:

$$\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| \leq \left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) (\tilde{\mathbf{c}}_i - \mathbf{c}_i) \right\| + \|\Lambda_i (\tilde{\mathbf{c}}_i - \mathbf{c}_i)\| + \left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) \mathbf{c}_i \right\|.$$

For the final step a similar method to the one used in Section 5.7.1.1.6 is used. We first divide by the norm of the vector $\tilde{\mathbf{c}} - \mathbf{c}$, and \mathbf{c} respectively to get a unit length vector in the three norms in the equation above and then multiply by the same norm to keep the equation unchanged

$$\left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) \frac{\tilde{\mathbf{c}}_i - \mathbf{c}_i}{\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\|} \right\| \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| + \left\| \Lambda_i \frac{\tilde{\mathbf{c}}_i - \mathbf{c}_i}{\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\|} \right\| \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| + \left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) \frac{\mathbf{c}_i}{\|\mathbf{c}_i\|} \right\| \|\mathbf{c}_i\|.$$

This is of course less than or equal to the supremum taken over the norms

$$\begin{aligned} \|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| &\leq \sup_{\|\mathbf{u}\|=1} \left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) \mathbf{u} \right\| \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| + \sup_{\|\mathbf{u}\|=1} \|\Lambda_i \mathbf{u}\| \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| \\ &\quad + \sup_{\|\mathbf{u}\|=1} \left\| \left(\tilde{\Lambda}_i - \Lambda_i \right) \mathbf{u} \right\| \|\mathbf{c}_i\|. \end{aligned}$$

The definition of the operator norm gives the following upper bound

$$\|\tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\| \leq \left\| \tilde{\Lambda}_i - \Lambda_i \right\|_{\text{OP}} \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| + \|\Lambda_i\|_{\text{OP}} \|\tilde{\mathbf{c}}_i - \mathbf{c}_i\| + \left\| \tilde{\Lambda}_i - \Lambda_i \right\|_{\text{OP}} \|\mathbf{c}_i\|.$$

where an upper bound on $\left\| \tilde{\Lambda}_i - \Lambda_i \right\|_{\text{OP}}$ has been calculated in Appendix C.15, an upper bound on $\|\tilde{\mathbf{c}}_i - \mathbf{c}_i\|$ was calculated in Section 5.7.1.1.6 and $\|\Lambda_i\|$ and $\|\mathbf{c}_i\|$ are known values.

5.7.1.1.8 Upper bound on the effect of a small change in Λ^{-1} and μ^* has on

ν : From the definition of ν in Equation 5.19 the following is given

$$\begin{aligned}
& \|\tilde{\nu}_i - \nu_i\| \\
&= \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left[\tilde{\Lambda}_i^{-1} \tilde{\mu}^* + \Omega^{-1} \mu \right] - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \left[\Lambda_i^{-1} \mu^* + \Omega^{-1} \mu \right] \right\| \\
&= \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \tilde{\Lambda}_i^{-1} \tilde{\mu}^* - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* \right. \\
&\quad \left. + \left[\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right] \Omega^{-1} \mu \right\|
\end{aligned}$$

A similar method to the one in the previous sections is used to relate the required upper bound to upper bounds that have already been calculated. This is done by adding zero in useful forms, using the triangle rule to divide the norm of sums into the sum of norms and finally using the definition of the operator norm and the sub-multiplicativity of the operator norm.

Adding zeros in convenient forms gives

$$\begin{aligned}
& \|\tilde{\nu}_i - \nu_i\| \\
&= \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} + \Lambda_i^{-1} \right) \left(\tilde{\mu}^* - \mu^* + \mu^* \right) - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* \right. \\
&\quad \left. + \left[\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right] \Omega^{-1} \mu \right\| \\
&= \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \left(\tilde{\mu}^* - \mu^* \right) + \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \left(\tilde{\mu}^* - \mu^* \right) \right. \\
&\quad \left. + \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \mu^* + \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* \right. \\
&\quad \left. - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* + \left[\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right] \Omega^{-1} \mu \right\|
\end{aligned}$$

Now the triangle rule can be used to pull the Euclidean norm apart

$$\begin{aligned}
& \|\tilde{\nu}_i - \nu_i\| \\
& \leq \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) (\tilde{\mu}^* - \mu^*) \right\| \\
& \quad + \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} (\tilde{\mu}^* - \mu^*) \right\| + \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \mu^* \right\| \\
& \quad + \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \Lambda_i^{-1} \mu^* \right. \\
& \quad \left. + \left[\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right] \Omega^{-1} \mu \right\|
\end{aligned}$$

As was described at the beginning of Section 5.7.1.1, the definition of the operator norm, detailed in Section 4.2.5, is used to get

$$\begin{aligned}
& \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) (\tilde{\mu}^* - \mu^*) \right\| \\
& = \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \frac{\tilde{\mu}^* - \mu^*}{\|\tilde{\mu}^* - \mu^*\|} \right\| \|\tilde{\mu}^* - \mu^*\| \\
& \leq \sup_{\|u\|=1} \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) u \right\| \|\tilde{\mu}^* - \mu^*\| \\
& = \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \right\|_{\text{OP}} \|\tilde{\mu}^* - \mu^*\|.
\end{aligned}$$

Since the first norm is the operator norm the sub-multiplicativity, as detailed in Section 4.2.5.2, can be used to divide it into the product of two norms

$$\begin{aligned}
& \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) (\tilde{\mu}^* - \mu^*) \right\| \\
& \leq \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \left\| \left(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right) \right\|_{\text{OP}} \|\tilde{\mu}^* - \mu^*\|
\end{aligned}$$

Exactly the same method is used for the second and third norm in the sum, which culmi-

nates in

$$\begin{aligned}
& \|\tilde{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\| \\
& \leq \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \left[\left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} \|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\| + \|\Lambda_i^{-1}\|_{\text{OP}} \|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\| \right. \\
& \quad \left. + \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \|\boldsymbol{\mu}^*\| \right] + \left\| \left[\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right] \left(\Lambda_i^{-1} \boldsymbol{\mu}^* + \Omega^{-1} \boldsymbol{\mu} \right) \right\|
\end{aligned}$$

It has been shown in Section 5.3 that the sum of positive definite matrices is again positive definite. Ω is the prior covariance matrix of $\begin{pmatrix} A \\ B \end{pmatrix}$ and is required to be positive definite, as described in Section 5.3. In Appendix C.16 it is shown that Λ_i^{-1} is also positive definite. Therefore $\Lambda_i^{-1} + \Omega^{-1}$ is positive definite and invertible and subsequently it is shown in Section 5.7.1.1.11 that $\tilde{\Lambda}_i^{-1} + \Omega^{-1}$ is also invertible.

Given the invertibility of the sum we can use a theorem from Dirk Ferus [22], which has been detailed in Section 4.2.5.1. If G and F are invertible matrices in a finite dimensional Banach space, as long as $\|F - G\|_{\text{OP}} < 1/\|F^{-1}\|_{\text{OP}}$, then

$$\|G^{-1}\boldsymbol{v}\| \leq \left(1/\|F^{-1}\|_{\text{OP}} - \|G - F\|_{\text{OP}} \right)^{-1} \|\boldsymbol{v}\|.$$

This means that if we substitute G with $\tilde{\Lambda}_i^{-1} + \Omega^{-1}$ and F with $\Lambda_i^{-1} + \Omega^{-1}$ the following holds

$$\begin{aligned}
\left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} &= \sup_{\|\boldsymbol{v}\|=1} \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right) \boldsymbol{v} \right\| \\
&\leq \left(\frac{1}{\left\| \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}}} - \left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} \right)^{-1}.
\end{aligned}$$

An upper bound on $\|\tilde{\nu}_i - \nu_i\|$ is therefore

$$\begin{aligned} & \left(\frac{1}{\left\| (\Lambda_i^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}} - \left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} \right)^{-1} \\ & \left[\left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} \|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\| + \|\Lambda_i^{-1}\|_{\text{OP}} \|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\| + \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \|\boldsymbol{\mu}^*\| \right] \\ & + \left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \left\| \Lambda_i^{-1} \boldsymbol{\mu}^* + \Omega^{-1} \boldsymbol{\mu} \right\| \end{aligned}$$

for which upper bounds on $\|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}}$, $\|\tilde{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|$ and $\left\| \left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}}$ have been calculated in Sections 5.7.1.1.5, 5.7.1.1.7 and C.17 respectively, and the rest of the terms are known values. Since the operator norm of the inverse of a matrix is just the inverse of the smallest eigenvalue, in absolute terms, of that matrix, as shown in Section 4.2.5.1, $1/\left\| (\Lambda_i^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}$ is equivalent to the smallest eigenvalue, in absolute terms, of $\Lambda_i^{-1} + \Omega^{-1}$

5.7.1.1.9 Upper bound on the effect of a small perturbation on $\exp\{\tilde{a} - a\}$:

The next step is to calculate an upper bound on the absolute value of

$$\exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_1 \tilde{\Sigma}^{-1} \mathbf{y}_1 - \tilde{\boldsymbol{\nu}}' (\tilde{\Lambda}^{-1} + \Omega^{-1}) \tilde{\boldsymbol{\nu}} + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right] + \frac{1}{2} \left[\mathbf{y}'_1 \Sigma^{-1} \mathbf{y}_1 - \boldsymbol{\nu}' (\Lambda^{-1} + \Omega^{-1}) \boldsymbol{\nu} + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right] \right\}.$$

As the exponential function is always positive the absolute value is not needed. To explain the general idea, let $a_i := -\frac{1}{2} \left[\mathbf{y}'_1 \Sigma_i^{-1} \mathbf{y}_1 - \boldsymbol{\nu}'_i (\Lambda_i^{-1} + \Omega^{-1}) \boldsymbol{\nu} + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right]$.

Using this notation the perturbation is:

$$\exp\{\tilde{a}_i - a_i\} \leq \exp\{|\tilde{a}_i - a_i|\}$$

with

$$\begin{aligned}
& |\tilde{a}_i - a_i| \\
&= \left| -\frac{1}{2} \left[\mathbf{y}'_1 \tilde{\Sigma}_i^{-1} \mathbf{y}_1 - \tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right] \right. \\
&\quad \left. + \frac{1}{2} \left[\mathbf{y}'_1 \Sigma_i^{-1} \mathbf{y}_1 - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i + \boldsymbol{\mu}' \Omega^{-1} \boldsymbol{\mu} \right] \right| \\
&= \left| -\frac{1}{2} \mathbf{y}'_1 (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \mathbf{y}_1 + \frac{1}{2} \left[\tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i \right] \right|
\end{aligned}$$

The triangle rule can be used to get the following upper bound:

$$|\tilde{a}_i - a_i| \leq \frac{1}{2} \left| \mathbf{y}'_1 (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \mathbf{y}_1 \right| + \frac{1}{2} \left| \tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i \right|.$$

Now $\mathbf{y}'_1 (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \mathbf{y}_1$ can be interpreted as an inner product between \mathbf{y}'_1 and $(\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \mathbf{y}_1$ which would allow the use of the Cauchy-Schwarz inequality $|\langle x, y \rangle| \leq \|x\| \|y\|$ [54]. Together with the definition of the operator norm this gives the following upper bound

$$|\tilde{a}_i - a_i| \leq \frac{1}{2} \sum_{j=1}^T y_{1,j} \left\| (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \right\|_{\text{OP}} + \frac{1}{2} \left| \tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i \right|.$$

since $\|\mathbf{y}'_1\| \|\mathbf{y}_1\| = \sum_{j=1}^T y_{1,j}^2$.

Putting this back into the original equation results in

$$\begin{aligned}
& |\exp\{\tilde{a}_i - a_i\}| \\
&\leq \left| \exp \left\{ \frac{1}{2} \sum_{j=1}^T y_{1,j}^2 \left\| (\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}) \right\|_{\text{OP}} + \frac{1}{2} \left| \tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i \right| \right\} \right|.
\end{aligned}$$

for which upper bounds on $\|\tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\|_{\text{OP}}$ and $|\tilde{\nu}'_i (\tilde{\Lambda}_i^{-1} + \Omega^{-1}) \tilde{\nu}_i - \nu'_i (\Lambda_i^{-1} + \Omega^{-1}) \nu_i|$ have been calculated in Sections 5.7.1.1.2 and C.18 respectively.

5.7.1.1.10 Effect of a small change in $p(\theta)$, Σ and a have on $p(\mathbf{y}_1, \theta | \mathbf{y}_2)$: As is known from Equation 5.21 $p(\mathbf{y}_1, \tilde{\theta} | \mathbf{y}_2) = p(\theta) \exp \{a\} / \left((2\pi)^{\frac{T}{2}+1} \sqrt{\det(\Sigma) \det(\Omega)} \right)$ with a defined as in 5.7.1.1.9. Therefore

$$\begin{aligned} & \left| p(\mathbf{y}_1, \tilde{\theta}_i | \mathbf{y}_2) - p(\mathbf{y}_1, \theta_i | \mathbf{y}_2) \right| \\ &= \left| \frac{p(\tilde{\theta}_i) \exp \{ \tilde{a}_i \}}{(2\pi)^{\frac{T}{2}+1} \sqrt{\det(\tilde{\Sigma}_i) \det(\Omega)}} - \frac{p(\theta_i) \exp \{ a_i \}}{(2\pi)^{\frac{T}{2}+1} \sqrt{\det(\Sigma_i) \det(\Omega)}} \right| \end{aligned}$$

This can be rearranged to:

$$\left| \frac{\exp \{ a_i \}}{(2\pi)^{\frac{T}{2}+1} \sqrt{\det(\Omega)}} \right| \left| \frac{1}{\sqrt{\det(\tilde{\Sigma}_i)}} \right| \left| p(\tilde{\theta}_i) \exp \{ \tilde{a}_i - a_i \} - p(\theta_i) \sqrt{\frac{\det(\tilde{\Sigma}_i)}{\det(\Sigma_i)}} \right|$$

$\left| \sqrt{\det(\tilde{\Sigma}_i)} \right|^{-1}$ can be rewritten as $\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} + \sqrt{\det(\Sigma_i)} \right|^{-1}$ for which an upper bound is the inverse of a lower bound on $\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} + \sqrt{\det(\Sigma_i)} \right|$ which is $\left| \sqrt{\det(\Sigma_i)} \right| - \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right|$, provided $\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| < \left| \sqrt{\det(\Sigma_i)} \right|$ which is the case as detailed in Section 5.7.1.1.11. Therefore

$$\left| \sqrt{\det(\tilde{\Sigma}_i)} \right|^{-1} \leq \left(\left| \sqrt{\det(\Sigma_i)} \right| - \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| \right)^{-1}.$$

An upper bound on $\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right|$ has been calculated in Section 5.7.1.1.4 and $\left| \sqrt{\det(\Sigma_i)} \right|$ is known, as it is unperturbed.

Therefore, an upper bound on $\left| p(\tilde{\theta}_i) \exp \{ \tilde{a}_i - a_i \} - p(\theta_i) \sqrt{\det(\tilde{\Sigma}_i) / \det(\Sigma_i)} \right|$ can be found by using the triangle rule and the sub-multiplicativity of absolute values

$$\begin{aligned} & \left| p(\tilde{\theta}_i) \exp \{ \tilde{a}_i - a_i \} - p(\theta_i) \sqrt{\frac{\det(\tilde{\Sigma}_i)}{\det(\Sigma_i)}} \right| \\ & \leq \left(\left| p(\tilde{\theta}_i) - p(\theta_i) \right| + |p(\theta_i)| \right) \left| \exp \{ \tilde{a}_i - a_i \} \right| + |p(\theta_i)| \left| \sqrt{\frac{\det(\tilde{\Sigma}_i)}{\det(\Sigma_i)}} \right|. \end{aligned}$$

An upper bound on $\left| \sqrt{\det(\tilde{\Sigma}_i)/\det(\Sigma_i)} \right|$ has been shown to be $\sqrt{\prod_{j=1}^T |\lambda_{i,j} + e_i|/|\lambda_{i,j}|}$ in the Appendix C.19 with e_i defined as $e_i = \left\| \tilde{\Sigma}_i - \Sigma_i \right\|_{\text{OP}}$ in the Appendix C.14.

Summarising all of this, an upper bound on $\left| p(\mathbf{y}_1, \tilde{\boldsymbol{\theta}}_i | \mathbf{y}_2) - p(\mathbf{y}_1, \boldsymbol{\theta}_i | \mathbf{y}_2) \right|$ is

$$\left| \frac{\exp\{a_i\}}{(2\pi)^{\frac{T}{2}+1} \sqrt{\det(\Omega)}} \right| \left(\sqrt{\det(\Sigma_i)} - \left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| \right)^{-1} \left[\left(|p(\tilde{\boldsymbol{\theta}}_i) - p(\boldsymbol{\theta}_i)| + p(\boldsymbol{\theta}_i) \right) \exp\{\tilde{a}_i - a_i\} + p(\boldsymbol{\theta}_i) \sqrt{\prod_{j=1}^T |\lambda_{i,j} + e_i|/|\lambda_{i,j}|} \right]$$

for which upper bounds on $|p(\tilde{\boldsymbol{\theta}}_i) - p(\boldsymbol{\theta}_i)|$ and $|\exp\{\tilde{a} - a\}|$ have previously been calculated in 5.7.1.1.1 and 5.7.1.1.9 and the rest of the terms are unperturbed values and therefore known.

5.7.1.1.11 Size of boxes used to get an upper bound on the integral As mentioned previously, the three dimensional space over which we are integrating is divided into sub-blocks. Within each sub-block the integral is estimated and an upper bound on the absolute error is calculated. Dividing the space into sub-blocks reduces the error of the overall integral. The sub-blocks with an error greater than a certain threshold, are further divided into smaller sub-blocks and the estimate and upper bound on the error are recalculated.

The initial size of the sub-blocks depends on a series of restrictions that are necessary to be able to use the previously calculated upper bound on the error. These restrictions are that the sub-blocks need to be small enough such that:

- i) $\tilde{\Sigma}_i$ is invertible: $\left\| \tilde{\Sigma}_i - \Sigma_i \right\|_{\text{OP}} < 1/\left\| \Sigma_i^{-1} \right\|_{\text{OP}}$ - needed in Sections 5.7.1.1.3, C.14 and C.19
- ii) $\left| \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)} \right| < \left| \sqrt{\det(\Sigma_i)} \right|$ - needed in Section 5.7.1.1.10
- iii) $(\Omega^{-1} + \tilde{\Lambda}_i^{-1})^{-1}$ is invertible: $\left\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \right\|_{\text{OP}} < 1/\left\| (\Omega^{-1} + \Lambda_i^{-1})^{-1} \right\|_{\text{OP}}$ - needed in

Sections 5.7.1.1.8 and C.17

iv) $\tilde{\Lambda}_i$ is invertible: ($\|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} < 1/\|\Lambda_i\|_{\text{OP}}$) - needed in Section C.15

5.7.1.2 Accumulation of the Error in the Upper Bound

As described in Section 5.7.1.1, the posterior pdf relies on the three cointegration parameters, Φ , S^2 and S_1^2 , through a long chain of transformations. The upper bound on the error of the integral within each sub-block needs to be calculated up this chain, which means that at each step the error in the upper bound gets multiplied with various constants. In addition to this, one of the last steps involves taking the exponential function, which causes the error of the upper bound to explode.

To better understand which steps of the chain increase the error the most, the values along the chain are calculated for the point of the sub-block that is the closest to $(-\infty, -\infty, -\infty)$ as well as at the point that is the farthest from $(-\infty, -\infty, -\infty)$. This difference is compared to the upper bound on the error of the integral.

Table 5.1: This table compares the calculated upper bound (column two), to the difference between the values at the corner points closest and furthest away from $(-\infty, -\infty, -\infty)$ (column three) at various steps along the chain to calculate the upper bound on the error of the integral. Column four gives the relative difference between the upper bound and the difference between the corners. In this example the upper bound is numerically not ∞ .

Point in Chain	Upper Bound	Corner Difference	Relative Difference
$ p(\tilde{\Theta}_i) - p(\Theta_i) $	1.4	$1.1e - 14$	$1.2e + 14$
$\ \tilde{\Sigma}_i - \Sigma_i\ _{\text{OP}}$	$1.2e - 11$	$5.0e - 12$	2.4
$\ \tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\ _{\text{OP}}$	$3.8e - 10$	$4.5e - 13$	$8.5e + 02$
$\ \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)}\ $	$5.8e - 24$	$3.3e - 26$	$1.8e + 02$
$\ \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\ _{\text{OP}}$	$3.5e - 06$	$1.4e - 09$	$2.5e + 03$
$\ \tilde{\Lambda}_i - \Lambda_i\ _{\text{OP}}$	$2.8e - 07$	$1.4e - 13$	$2.0e + 06$
$\ \tilde{\mathbf{c}}_i - \mathbf{c}_i\ $	$5.6e - 07$	$2.0e - 10$	$2.8e + 03$
$\ \tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\ $	$6.7e - 05$	$2.2e - 13$	$3.1e + 08$
$\ \tilde{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\ $	$2.6e - 02$	$1.9e - 13$	$1.4e + 11$
$ \exp(\tilde{a}_i - a_i) $	$4.2e + 17$	$8.6e - 24$	$4.9e + 40$
$ p(\tilde{\Theta}_i, \mathbf{y}_1 \mathbf{y}_2) - p(\Theta_i, \mathbf{y}_1 \mathbf{y}_2) $	$6.3e - 21$	$3.3e - 51$	$1.9e + 30$

Tables 5.1 and 5.2 are two examples of the points in the chain where these values are calculated (column one), the calculated upper bound on the error of the integral at these points (column two), the difference between the point farthest away from $(-\infty, -\infty, -\infty)$ and the one closest to $(-\infty, -\infty, -\infty)$ (column three), and the relative differences by which the upper bound is greater than the difference between the two corners of the sub-block (column four).

Table 5.2: This table compares the calculated upper bound (column two), to the difference between the values at the corner points closest and furthest away from $(-\infty, -\infty, -\infty)$ (column three) at various steps along the chain to calculate the upper bound on the error of the integral. Column four gives the relative difference between the upper bound and the difference between the corners. In this example the upper bound is numerically ∞ .

Point in Chain	Upper Bound	Corner Difference	Relative Difference
$ p(\tilde{\Theta}_i) - p(\Theta_i) $	$4.0e + 09$	$6.6e - 89$	$6.1e + 97$
$\ \tilde{\Sigma}_i - \Sigma_i\ _{\text{OP}}$	$1.1e - 10$	$1.0e - 10$	1.04
$\ \tilde{\Sigma}_i^{-1} - \Sigma_i^{-1}\ _{\text{OP}}$	$1.7e - 05$	$4.1e - 09$	$4.2e + 03$
$\ \sqrt{\det(\tilde{\Sigma}_i)} - \sqrt{\det(\Sigma_i)}\ $	$1.1e - 106$	$5.0e - 110$	$2.2e + 03$
$\ \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\ _{\text{OP}}$	$1.6e - 01$	$3.7e - 05$	$4.3e + 03$
$\ \tilde{\Lambda}_i - \Lambda_i\ _{\text{OP}}$	$7.4e - 09$	$2.2e - 15$	$3.4e + 06$
$\ \tilde{\mathbf{c}}_i - \mathbf{c}_i\ $	$2.5e - 02$	$5.2e - 06$	$4.9e + 03$
$\ \tilde{\boldsymbol{\mu}}_i^* - \boldsymbol{\mu}_i^*\ $	$3.8e - 03$	$3.0e - 15$	$1.3e + 12$
$\ \tilde{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\ $	3.0	$2.2e - 15$	$1.4e + 15$
$ \exp(\tilde{a}_i - a_i) $	∞	0	∞
$ p(\tilde{\Theta}_i, \mathbf{y}_1 \mathbf{y}_2) - p(\Theta_i, \mathbf{y}_1 \mathbf{y}_2) $	NaN	0	NaN

In the first example (Table 5.1) the upper bound is calculated for the sub-block starting at $\phi = 0.704$, $\sigma^2 = 0.51$, and $\sigma_1^2 = 1.01$. In this example the upper bound exists and is not numerically infinity. However, the difference between the upper bound and the difference between the corners of the sub-block increases continuously and the upper bound is several magnitudes larger than the difference between the corners.

In the second example (Table 5.2) the upper bound is calculated for the sub-block starting at $\phi = -1$, $\sigma^2 = 0.01$, and $\sigma_1^2 = 0.01$. In this example the upper bound does not exist, because the one but last upper bound is numerically ∞ and is multiplied with a value that is numerically 0. The difference between the two corners is numerically zero for the one but last and last step.

Chapter 6

Simulations of Cointegration Test

To quantify the accuracy of the cointegration test proposed in the previous chapter (Chapter 5) cointegrated time series were simulated. These time series were then tested for cointegration using the above described method. The data generating process is described in Section 6.1. This is followed by a discussion of the results in Section 6.2. Section 6.3 describes how the method could be applied to real-world animal movement data.

6.1 Data Generating Process

As described in Section 5.6 there are three parameters, (Φ, S^2, S_1^2) , which are key to detecting whether time series are cointegrated or not. For perfect cointegration, these three parameters lie on a curved plane in a three dimensional space. This means that for any combination of two of the parameters, the value of the third parameter is nearly uniquely defined (if S^2 and S_1^2 are chosen, Φ can take up to 2 values), by the following equation (derived in the Appendix C.11).

$$\Phi = \pm \sqrt{1 - S^2/S_1^2}$$

For the simulations here S^2 and S_1^2 are varied from 0.01 to 1.01 in steps of 0.1. For each combination of S^2 and S_1^2 one or two values of Φ are calculated. When S^2 is equal to S_1^2 , then $\Phi = 0$. In all other cases there is a negative and a positive value of Φ . Since the posterior and the cointegration tube are symmetric, the simulations were run only for the cases in which ϕ is greater or equal to 0. In total 66 simulations were run.

Given these parameters, the assumptions described in Section 5.2 are used to generate the data. The details are explained in the remainder of this section.

Recall that cointegration describes two time series that have a stationary linear relationship. The linear relationship is defined by the following equations, first introduced in

Section 5.2.

$$Y_{1,t} = A + BY_{2,t} + E_t$$

$$E_t = \Phi E_{t-1} + H_t$$

where $t \in \{1, \dots, T\}$. The unknown variables in this relationship are assumed to have the following underlying distributions

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Omega\right)$$

$$H_t \sim \mathcal{N}(0, S^2)$$

$$E_1 \sim \mathcal{N}(0, S_1^2).$$

\mathbf{Y}_2 is assumed to follow a first-order Markov model

$$Y_{2,t} = \phi_y Y_{2,t-1} + W_t,$$

with $Y_{2,1} \sim \mathcal{N}(0, \sigma_y^2)$, $W_t \sim \mathcal{N}(0, \sigma_w^2)$, $\phi_y \in \mathbb{R}$ and $\sigma_y^2, \sigma_w^2 \in [0, \infty[$.

The first step is to draw A and B , the starting values for \mathbf{Y}_2 , \mathbf{E} and \mathbf{H} from the defined distributions. For this the necessary parameters are defined as follows:

$$\mu_\alpha = 1, \quad \mu_\beta = 1, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_y^2 = 1.$$

It would be useful to run the same test with varying values for these starting parameters. This would help understand the sensitivity of the method towards the prior information.

6.2 Results

As described in the previous section, 66 pairs of cointegrated time series were simulated. Each of these pairs were tested for cointegration using the method above. At a threshold of 24% (16/66) of the time series were correctly identified as being cointegrated.

To get a better understanding of the results, Figure 6.1 (i) shows a heat map of the positive values of ϕ , that make up the cointegrated tripple, given the values of σ^2 and σ_1^2 . Figure 6.1 (ii) then shows the proportion of the posterior within the cointegration tube for the simulated time series. See Tables D.1 to D.8 in the Appendix for the full set of results.

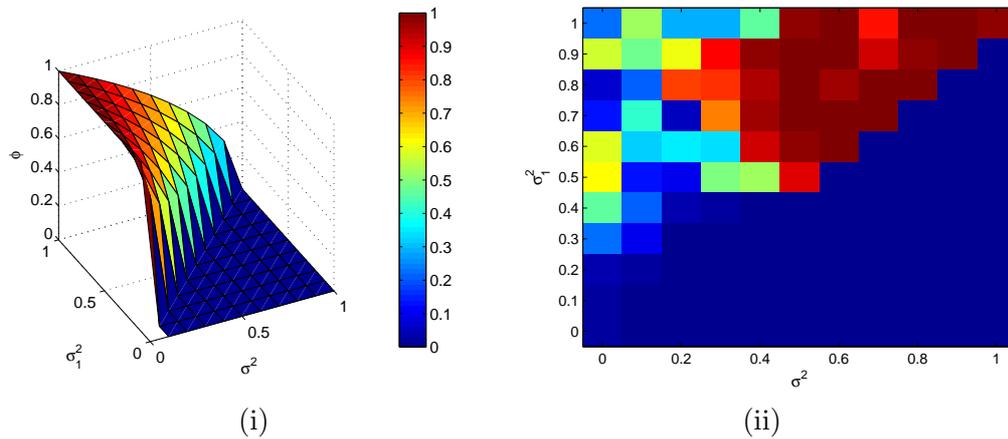


Figure 6.1: These plots show (i) the combinations of ϕ , σ^2 and σ_1^2 that cause cointegration and (ii) the proportion of the posterior that lies within the cointegration tube.

The test proposed in this thesis suggests that two time series are cointegrated if more than 90% of the posterior probability density lies within the cointegration tube. The two heat maps show that the test detects cointegration well when σ^2 and σ_1^2 are large. This is most likely because ζ , the parameter that defines the width of the cointegration tube, was primarily estimated from time series with $\sigma^2 = 0.51$ and $\sigma_1^2 = 1.01$. In the area around this point the test detects cointegration well.

Furthermore the shape of the posterior density when the time series are cointegrated needs to be examined at more detail. The average heat map in Figure 5.4 was created

with $\sigma_1^2 = 1.01$ and it shows that the cointegration tube is a lot wider for large values of σ^2 than for small ones, which means that it is unlikely that cointegrated time series with small values of σ^2 will be detected as being cointegrated.

It would be of great advantage to examine the cointegration tube in more detail in the future.

6.3 Application

As described in Chapter 1 the method to test cointegration proposed in Chapter 5 could be used to identify a change in behaviour of social animals. This change in behaviour could be an indicator of certain diseases, such as Huntington's disease [13] or Batten's [33].

To ascertain whether there has been a change in group dynamics or of the behaviour of certain individuals within a group, for example in a flock of sheep, social network analysis could be used. In such a social network the individuals are represented by nodes, and a link exists between two nodes if they are associated with one another. In this case, cointegration could be used as a measure of association, since it represents a level of confidence that animals move together.

A sample of social networks could be estimated from several samples of the sheep's movements, using the cointegration method proposed in Chapter 5. This sample could be compared to similar samples of social networks based on the behaviour of individuals at different stages of the development of symptoms. Significant differences between the groups' structures could be identified, which could help in the analysis of treatment efficacy in the long term.

In these situations it would be of particular interest whether the observed behaviour is cointegrated, as opposed to becoming cointegrated at a future point. To the best of our

knowledge, the method proposed in Chapter 5 is the only method that tests the observed time series.

Chapter 7

Discussion on the Cointegration Test

7.1 Discussion

In this dissertation a new method to test for cointegration is presented. A Bayesian approach is advantageous for many reasons: it produces whole probability distributions for each unknown parameter and these distributions are valid for any sample size. Furthermore, it allows straightforward updates when more data becomes available, by using the posterior as the new prior distribution.

The main advantage of the cointegration test proposed in this thesis is that it fully exploits the benefits of the Bayesian method. A cointegration tube is created, which describes the combinations of parameters that indicate cointegration of the two time series being compared. This allows us to test whether say 90% of the posterior lies within the cointegration tube, rather than using a point estimate, such as the maximum likelihood of the posterior.

In Appendix D, the results of the method proposed here are compared to the results of the traditional Engle and Granger approach. The Engle and Granger method correctly identifies 39 out of the 66 simulations, as opposed to 16 that are correctly identified by the method proposed here. Despite these results, it should be noted that the proposed method is likely to be improved greatly by further research into the shape of the cointegration tube.

The Bayesian method is simultaneously a significant advantage, because it has the potential to allow us to reason more fully about a cointegrated system, and a significant limitation, because prior distributions need to be chosen for the parameters of interest. Choosing a wrong prior could bias the result, however the more data is used the smaller the bias will be.

7.2 Future Work

To establish this method as a viable alternative to existing techniques two areas of work need to be addressed.

As described in Section 5.6, the cointegration tube is estimated by allowing small imperfections (ζ) in the stationarity constraint of the time independent covariance. The current shape of the so called cointegration tube, fits very well for some combinations of the cointegration parameters (see top right-hand corner of Figure 6.1 (ii) and tables D.6, D.7 and D.8), but does not fit particularly well in other areas. It would be of great advantage to study the shape of the posterior probability distributions resulting from cointegrated time series. This could greatly improve the accuracy of the method.

Section 5.7 describes the numerical integration used in the proposed method. The Riemann sum was used to allow the calculation of the hard upper bound on the error of the integration. However, the upper bound is larger than expected and future research into alternative numerical quadrature methods may be needed.

Part IV

Conclusions

Chapter 7

The interactions of animals is an area in which vast amounts of data are collected, yet efficient methods to analyse them are scarce. There are, roughly speaking, two categories of animal interaction: that of solitary animals and that of social animals. One method for each of these areas is proposed in this thesis.

When analysing the interaction between solitary animals, the quantification of association or avoidance between territorial conspecifics would advance our understanding of animal ecology and, in the long term, the impact of changing environments. Existing forms of such tests are predicated on assumptions about the shape of the individuals' territory and boundaries [48, 19].

In this dissertation, a new method for detecting avoidance and association is presented. Unlike previous work, the method makes no assumption about the shape or size of the territories, nor about the way that individuals move. It relies purely on the disassociation of the individuals' movement through permutations.

Amongst other things, this new method permits the analysis of territorial behaviour in animals. Both the presence and absence of positive spatial association between individuals or groups are biologically interesting phenomena. In Chapter 3, the method was applied to data collected from GPS collars on individual leopards in which significant positive association was established between some male-male as well as male-female leopard dyads, and to African wild dogs, in which there was no significant dynamic interaction detected between packs.

For the leopards, two out of six male-female dyads were more often within close proximity of each other than would be expected by chance. This is most likely related to courtship and mating, and supports biological expectations. Interestingly, it was also shown that two out of five male-male dyads were more often within close proximity of each other. This observation is in opposition to conclusions from previous work [59, 35], but could be due to mutual evaluation, family relationships, or a range of unknown factors.

Chapter 7

None of the wild dog packs were more or less often within close proximity of each other than would be expected by chance. It is possible that, although the movement patterns of individual packs bring neighbours into relatively close proximity, the risk and occurrence of direct encounters may be reduced by remote inter-pack information exchange, probably via fresh scent signals in these areas.

In using this method, it is important to ensure that data from both members of the dyad are as closely matched in time as possible in order to allow robust conclusions to be drawn on any spatial interactions between them. Temporal differences between compared locations within dyads do not preclude the use of the method, but in such circumstances it is essential to temper conclusions accordingly.

More generally, our method for movement and associations could be applied to epidemiological concerns. If individuals are more often within close proximity of each other than expected by chance, the transmission rate of diseases would be higher than that estimated using random movement models. The method could also be extended straightforwardly to include a time lag to determine whether individuals are more often in an area recently occupied by another animal than might be explained by chance. This could be important in cases of geo-located time-limited phenomena such as scent marking or the transmission of parasites or infectious agents through the environment.

When examining the interaction between social individuals, the relationships between non-stationary time series is important. Time series are said to be stationary when the time series has a constant mean, constant variance and time-independent covariance. This is a strong restriction on animal movement, which is not always fulfilled [53]. When testing for dependencies between two or more time series the Pearson product-moment correlation coefficient is often one of the first statistics considered. However, correlation only makes sense if the individual series are stationary and the relationship between the two time series is linear [14].

In practice there are many time series that are non-stationary, such as the aggressive

Chapter 7

communication of the hermit crab *Calcinus tibicen* [53], exchange rates [41], population and employment [36], electricity consumption [2], gas prices [27], maize prices [1], dissent rates on the High Court of Australia [51] and hemispheric temperature [42]. In these cases, a characteristic called cointegration could be used to find relationships between non-stationary time series. Cointegration describes a property of two or more time series that are individually non-stationary but for which a linear relationship of the time series is stationary.

In this thesis a new fully Bayesian method to test for cointegration is presented. A Bayesian approach is advantageous for many reasons: it produces whole probability distributions for each unknown parameter and these distributions are valid for any sample size. Furthermore, it allows straightforward updates when more data becomes available, by using the posterior as the new prior distribution.

The main advantage of the cointegration test proposed in this thesis is that it fully exploits the benefits of the Bayesian method. A cointegration tube is created, which describes the combinations of parameters that indicate cointegration of the two time series being compared. This allows us to test whether say 90% of the posterior lies within the cointegration tube, rather than using a point estimate, such as the maximum likelihood of the posterior.

Significant progress has been made in calculating a hard upper bound on the integration error. An upper bound on the error would lead to an upper and lower bound on the integration itself and therefore give a better understanding of how certain the result is. The hard lower bound on the integration means that the integral is greater or equal to the lower bound. This means that if the lower bound suggests that 90% of the posterior probability density lies within the cointegration tube, then you can be sure that at least 90% lies within the cointegration tube.

Additionally this method is, to the best of my knowledge, the only one that tests whether the time series are cointegrated during the observation period. Other methods use only one parameter which is indicative of the time series becoming cointegrated at some future point

Chapter 7

in time [29, 39, 21]. The method we propose considers three parameters (introduced in Section 5.6) which allows us to test for cointegration of the observed time series themselves. This is particularly useful when change points are present in the data.

Part V

Appendix

Appendix A

Results of the Avoidance Test

A.1 Tables of Proportions of Simulations Correctly Identified as having, or not having, an Association

Association Time (in steps)	No. correctly identified as PLT (%)	No. correctly identified as PMT (%)
1	1073/1795 (60%)	8/1795 (0%)
2	1405/1795 (78%)	8/1795 (0%)
3	1513/1795 (84%)	8/1795 (0%)
4	1601/1795 (89%)	7/1795 (0%)
5	1658/1795 (92%)	8/1795 (0%)

Table A.1: Effect of Association Time Within Distances Inside the Association Distance: This table details how many of the simulations were correctly classified as being less (PLT) or more (PMT) often within the distances tested than expected by chance. The results are divided into the number of time steps (first column) the individuals spent in the inner association distance of each other. In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Association Time (in steps)	No. correctly identified as NLT (%)	No. correctly identified as NMT (%)
1	5927/6048 (98%)	5872/6048 (97%)
2	5945/6048 (98%)	5717/6048 (95%)
3	5951/6048 (98%)	5564/6048 (92%)
4	5962/6048 (99%)	5431/6048 (90%)
5	5976/6048 (99%)	5315/6048 (88%)

Table A.2: Effect of Association Time Outside of the Association Distances: This table details how many of the simulations were correctly classified as not being less (NLT) or more (NMT) often within the distances tested than expected by chance. The results are divided into the number of time steps (first column) the individuals spent in the inner association distance of each other. In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Association Distance (in meters)	No. correctly identified as PLT (%)	No. correctly identified as PMT (%)
250	1046/1410 (74%)	0/1410 (0%)
300	1116/1465 (76%)	5/1465 (0%)
350	1173/1495 (78%)	10/1495 (1%)
400	1248/1510 (83%)	0/1510 (0%)
450	1350/1540 (88%)	10/1540 (1%)
500	1317/1555 (85%)	14/1555 (1%)

Table A.3: Effect of Association Distance Within Distances Inside the Association Distance: This table details how many of the simulations were correctly classified as being less (PLT) or more (PMT) often within the distances tested than expected by chance. The results are divided into the outer association distance (first column). In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Association Distance (in meters)	No. correctly identified as NLT (%)	No. correctly identified as NMT (%)
250	4996/5040 (99%)	4752/5040 (94%)
300	4976/5040 (99%)	4755/5040 (94%)
350	4949/5040 (98%)	4692/5040 (93%)
400	4923/5040 (98%)	4610/5040 (91%)
450	4965/5040 (99%)	4564/5040 (90%)
500	4952/5040 (98%)	4526/5040 (90%)

Table A.4: Effect of Association Distance within Distances Outside of the Association Distance: This table details how many of the simulations were correctly classified as not being less (NLT) or more (NMT) often within the distances tested than expected by chance. The results are divided into the outer association distance (first column). In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Observation Length (in days)	No. correctly identified as PLT (%)	No. correctly identified as PMT (%)
100	749/1080 (69%)	0/1080 (0%)
150	1121/1465 (77%)	0/1465 (0%)
200	1252/1590 (79%)	0/1590 (0%)
250	1298/1555 (83%)	5/1555 (0%)
300	1415/1665 (85%)	9/1665 (1%)
350	1415/1620 (87%)	25/1620 (2%)

Table A.5: Effect of Observation Length within distances inside the association distance: This table details how many of the simulations were correctly classified as being less (PLT) or more (PMT) often within the distances tested than expected by chance. The results are divided into the observation length (first column). In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Observation Length (in days)	No. correctly identified as NLT (%)	No. correctly identified as NMT (%)
100	4982/5040 (99%)	4863/5040 (96%)
150	4933/5040 (98%)	4734/5040 (94%)
200	4949/5040 (98%)	4595/5040 (91%)
250	4961/5040 (98%)	4578/5040 (91%)
300	4969/5040 (99%)	4578/5040 (91%)
350	4967/5040 (99%)	4551/5040 (90%)

Table A.6: Effect of Observation Length within distances outside the association distance: This table details how many of the simulations were correctly classified as not being less (NLT) or more (NMT) often within the distances tested than expected by chance. The results are divided into the observation length (first column). In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Observation Length (in days)	No. correctly identified as NLT (%)	No. correctly identified as NMT (%)
100	6670/6720 (99%)	6685/6720 (99%)
150	6535/6720 (98%)	6615/6720 (98%)
200	6585/6720 (98%)	6680/6720 (99%)
250	6560/6720 (98%)	6645/6720 (99%)
300	6550/6720 (97%)	6670/6720 (99%)
350	6555/6720 (98%)	6620/6720 (99%)

Table A.7: Effect of Observation Length when there is no association: This table details how many of the simulations were correctly classified as not being less (NLT) or more (NMT) often within the distances tested than expected by chance in the no association scenario. The results are divided into the observation length (first column). In each column the number before the forward-slash is the number of correctly classified distances, the number after the forward-slash is the number of distances that should be flagged, and the number in brackets is the proportion of correctly identified distances.

Appendix B

Proofs to Chapter 4.2

B.1 Proof that $\text{Cov}(\epsilon_t, \epsilon_{t-k}) = \frac{\sigma^2 \phi^k}{1-\phi^2}$

$$\begin{aligned}
\text{Cov}(\epsilon_t, \epsilon_{t-k}) &= \text{Cov} \left(\sum_{j=0}^{\infty} \phi^j \eta_{t-j}, \sum_{j=0}^{\infty} \phi^j \eta_{t-k-j} \right) \\
&= \text{Cov} \left(\sum_{j=0}^{\infty} \phi^j \eta_{t-j}, \sum_{l=k}^{\infty} \phi^{l-k} \eta_{t-l} \right) \\
&= \text{Cov} \left(\phi^k \eta_{t-k}, \eta_{t-k} \right) + \text{Cov} \left(\phi^{k+1} \eta_{t-k-1}, \phi \eta_{t-k-1} \right) \\
&\quad + \text{Cov} \left(\phi^{k+2} \eta_{t-k-2}, \phi^2 \eta_{t-k-2} \right) + \dots \\
&= \sigma^2 \left(\phi^k + \phi^{k+2} + \phi^{k+4} + \dots \right) \\
&= \frac{\sigma^2 \phi^k}{1-\phi^2}
\end{aligned}$$

□

B.2 Solving Maximisation using Lagrange Multiplier

The equation to be maximised is

$$\mathcal{L}(b_1, \dots, b_n, \mu) = \sqrt{\frac{b_1^2}{\lambda_1^2} + \dots + \frac{b_n^2}{\lambda_n^2}} + \mu \left(1 - \sqrt{b_1^2 + \dots + b_n^2} \right).$$

The partial derivatives are

$$\begin{aligned}
\frac{\delta \mathcal{L}}{\delta \mu} &= 1 - \sqrt{b_1^2 + \dots + b_n^2} \\
\frac{\delta \mathcal{L}}{\delta b_i} &= \frac{b_i}{\lambda_i^2 \sqrt{b_1^2/\lambda_1^2 + \dots + b_n^2/\lambda_n^2}} - \mu \frac{b_i}{\sqrt{b_1^2 + \dots + b_n^2}}
\end{aligned}$$

Setting the first partial derivative to 0 results in the constraint, i.e. $\sqrt{b_1^2 + \dots + b_n^2} = 1$.

This simplifies the partial derivative with respect to b_i to $\frac{\delta \mathcal{L}}{\delta b_i} = \frac{b_i}{\lambda_i^2 \sqrt{b_1^2/\lambda_1^2 + \dots + b_n^2/\lambda_n^2}} - \mu b_i$.

Setting this to 0 results in either $b_i = 0$ or $\mu = \frac{1}{\lambda_i^2 \sqrt{b_1^2/\lambda_1^2 + \dots + b_n^2/\lambda_n^2}}$. From the first derivative (i.e. the constraint) it is known that $\sqrt{b_1^2 + \dots + b_n^2} = 1$ which means that for at least one $j \in \{1, \dots, n\}$ $b_j \neq 0$. For that particular partial derivative, i.e. for $\frac{\delta \mathcal{L}}{\delta b_j}$ $\mu = \frac{1}{\lambda_j^2 \sqrt{b_1^2/\lambda_1^2 + \dots + b_n^2/\lambda_n^2}}$. Since there is only one μ , which means that, as long as $\lambda_i \neq \lambda_j$ $\forall i \neq j$, all other b_i s have to be 0. However, again due to the constraint $\sqrt{b_1^2 + \dots + b_n^2} = 1$, this means that $b_j = 1$.

Appendix C

Proofs to Chapter 5

C.1 Proof that $p(\mathbf{y}_1|\mathbf{y}_2) \neq 0$

First we will show that $p(\binom{\alpha}{\beta}|\mathbf{y}_2) = p(\binom{\alpha}{\beta})$, which will then be used to show that $p(\mathbf{y}_1|\mathbf{y}_2)$ is the integral of a product of two normal distributions, which by definition is > 0 everywhere [63].

For the first part we use the fact that the prior information for \mathbf{Y}_2 is independent of the prior information for $\binom{A}{B}$. For clarity of the proof we will abbreviate the prior information about \mathbf{Y}_2 (ϕ_y, σ_y^2 and σ_w^2) as γ .

Proof. Using Bayes' Theorem as well as the law of total probability

$$\begin{aligned} p(\binom{\alpha}{\beta}|\mathbf{y}_2) &= \frac{p(\mathbf{y}_2|\binom{\alpha}{\beta})p(\binom{\alpha}{\beta})}{p(\mathbf{y}_2)} = \frac{p(\binom{\alpha}{\beta})}{p(\mathbf{y}_2)} \int_{\gamma} p(\mathbf{y}_2, \gamma|\binom{\alpha}{\beta}) d\gamma \\ &= \frac{p(\binom{\alpha}{\beta})}{p(\mathbf{y}_2)} \int_{\gamma} p(\mathbf{y}_2|\gamma, \binom{\alpha}{\beta})p(\gamma|\binom{\alpha}{\beta}) d\gamma \end{aligned}$$

This is equivalent to $p(\binom{\alpha}{\beta})/p(\mathbf{y}_2) \int_{\gamma} p(\mathbf{y}_2|\gamma)p(\gamma|\binom{\alpha}{\beta}) d\gamma$, since \mathbf{Y}_2 given γ is independent of $\binom{A}{B}$, as defined in Section 5.2. This in turn is equivalent to $p(\binom{\alpha}{\beta})/p(\mathbf{y}_2) \int_{\gamma} p(\mathbf{y}_2|\gamma)p(\gamma) d\gamma$, because $\binom{A}{B}$ is independent of γ by assertion (Section 5.2). From this follows

$$p(\binom{\alpha}{\beta}|\mathbf{y}_2) = \frac{p(\binom{\alpha}{\beta})}{p(\mathbf{y}_2)} \int_{\gamma} p(\mathbf{y}_2, \gamma) d\gamma = \frac{p(\binom{\alpha}{\beta})}{p(\mathbf{y}_2)} p(\mathbf{y}_2) = p(\binom{\alpha}{\beta})$$

□

Using the law of total probability $p(\mathbf{y}_1|\mathbf{y}_2) = \int_{\binom{\alpha}{\beta}} p(\mathbf{y}_1, \binom{\alpha}{\beta}|\mathbf{y}_2)$. Using Bayes' Theorem this can be rewritten as $\int_{\binom{\alpha}{\beta}} p(\mathbf{y}_1|\binom{\alpha}{\beta}, \mathbf{y}_2)p(\binom{\alpha}{\beta}|\mathbf{y}_2)$. We have just shown that the second part is equivalent to $p(\binom{\alpha}{\beta})$, which is known, from Section 5.2, to be a bivariate Normal, which is by definition never 0. Additionally $p(\mathbf{y}_1|\binom{\alpha}{\beta}, \mathbf{y}_2)$ also follows a Gaussian distribution and is $\neq 0$. Therefore the product of the two can never be 0, and neither can the integral over this product.

C.2 Proof of $p(\boldsymbol{\theta}|\mathbf{y}_2) = p(\boldsymbol{\theta})$

For clarity of the proof we will abbreviate the prior information about \mathbf{Y}_2 (ϕ_y , σ_y^2 and σ_w^2) as $\boldsymbol{\gamma}$.

Proof. Using Bayes' Theorem

$$p(\boldsymbol{\theta}|\mathbf{y}_2) = \frac{p(\mathbf{y}_2|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y}_2)} = \frac{p(\boldsymbol{\theta})}{p(\mathbf{y}_2)} \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2, \boldsymbol{\gamma}|\boldsymbol{\theta}) d\boldsymbol{\gamma} = \frac{p(\boldsymbol{\theta})}{p(\mathbf{y}_2)} \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2|\boldsymbol{\gamma}, \boldsymbol{\theta})p(\boldsymbol{\gamma}|\boldsymbol{\theta}) d\boldsymbol{\gamma}$$

This is equivalent to $p(\boldsymbol{\theta})/p(\mathbf{y}_2) \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2|\boldsymbol{\gamma})p(\boldsymbol{\gamma}|\boldsymbol{\theta}) d\boldsymbol{\gamma}$, since \mathbf{Y}_2 given $\boldsymbol{\gamma}$ is independent of $\boldsymbol{\Theta}$, as defined in Section 5.2. This in turn is equivalent to $p(\boldsymbol{\theta})/p(\mathbf{y}_2) \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2|\boldsymbol{\gamma})p(\boldsymbol{\gamma}) d\boldsymbol{\gamma}$, because $\boldsymbol{\Theta}$ is independent of $\boldsymbol{\gamma}$ by assertion (Section 5.2). From this follows

$$p(\boldsymbol{\theta}|\mathbf{y}_2) = \frac{p(\boldsymbol{\theta})}{p(\mathbf{y}_2)} \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2, \boldsymbol{\gamma}) d\boldsymbol{\gamma} = \frac{p(\boldsymbol{\theta})}{p(\mathbf{y}_2)} p(\mathbf{y}_2) = p(\boldsymbol{\theta})$$

□

C.3 Proof of $p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{y}_2) = p(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Proof. Similarly to C.2 Bayes' Theorem and the law of total probability is used

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{y}_2) = \int_{\boldsymbol{\xi}} p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}|\boldsymbol{\theta}, \mathbf{y}_2) d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{y}_2) p(\boldsymbol{\xi}|\boldsymbol{\theta}, \mathbf{y}_2) d\boldsymbol{\xi}$$

where $\boldsymbol{\xi} = (\mu_\alpha, \mu_\beta, \Omega)$ is the prior information for $\binom{A}{B}$. Since $\binom{A}{B}$ given $\boldsymbol{\xi}$ is independent of $\boldsymbol{\Theta}$ and $\boldsymbol{\xi}$ is independent of $\boldsymbol{\Theta}$ (see Chapter 5), this is equivalent to

$$\int_{\boldsymbol{\xi}} p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\xi}, \mathbf{y}_2) p(\boldsymbol{\xi}|\mathbf{y}_2) d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}|\mathbf{y}_2) d\boldsymbol{\xi} = p(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{y}_2)$$

This can be rewritten, using Bayes' Theorem, as

$$\frac{p(\alpha, \beta, \mathbf{y}_2)}{p(\mathbf{y}_2)} = \frac{p(\mathbf{y}_2 | \alpha, \beta) p(\alpha, \beta)}{p(\mathbf{y}_2)} \quad (\text{C.1})$$

Now $p(\mathbf{y}_2 | \alpha, \beta)$ can be rewritten, using the law of total probability as

$$p(\mathbf{y}_2 | \alpha, \beta) = \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2, \boldsymbol{\gamma} | \alpha, \beta) d\boldsymbol{\gamma} = \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2 | \boldsymbol{\gamma}, \alpha, \beta) p(\boldsymbol{\gamma} | \alpha, \beta) d\boldsymbol{\gamma}$$

and because \mathbf{Y}_2 given $\boldsymbol{\gamma}$ is independent of (A, B) this is

$$p(\mathbf{y}_2 | \alpha, \beta) = \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2 | \boldsymbol{\gamma}) p(\boldsymbol{\gamma} | \alpha, \beta) d\boldsymbol{\gamma} \quad (\text{C.2})$$

Using Bayes' Theorem again

$$p(\boldsymbol{\gamma} | \alpha, \beta) = \frac{p(\boldsymbol{\gamma}, \alpha, \beta)}{p(\alpha, \beta)} = \frac{p(\alpha, \beta | \boldsymbol{\gamma}) p(\boldsymbol{\gamma})}{p(\alpha, \beta)} \quad (\text{C.3})$$

and again using the law of total probability

$$p(\alpha, \beta | \boldsymbol{\gamma}) = \int_{\boldsymbol{\xi}} p(\alpha, \beta, \boldsymbol{\xi} | \boldsymbol{\gamma}) d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} p(\alpha, \beta | \boldsymbol{\xi}, \boldsymbol{\gamma}) p(\boldsymbol{\xi} | \boldsymbol{\gamma}) d\boldsymbol{\xi}$$

Similarly to the previous calculation we use that (A, B) given $\boldsymbol{\xi}$ is independent of $\boldsymbol{\gamma}$ and that $\boldsymbol{\xi}$ is independent of $\boldsymbol{\gamma}$ and therefore

$$p(\alpha, \beta | \boldsymbol{\gamma}) = \int_{\boldsymbol{\xi}} p(\alpha, \beta | \boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} p(\alpha, \beta, \boldsymbol{\xi}) d\boldsymbol{\xi} = p(\alpha, \beta).$$

Plugging this back into equation C.3 gives

$$p(\boldsymbol{\gamma} | \alpha, \beta) = \frac{p(\alpha, \beta) p(\boldsymbol{\gamma})}{p(\alpha, \beta)} = p(\boldsymbol{\gamma})$$

which in turn can be plugged into equation C.2 giving

$$p(\mathbf{y}_2 | \alpha, \beta) = \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2 | \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) d\boldsymbol{\gamma} = \int_{\boldsymbol{\gamma}} p(\mathbf{y}_2, \boldsymbol{\gamma}) d\boldsymbol{\gamma} = p(\mathbf{y}_2).$$

This can be inserted into equation C.1 proving that

$$p(\alpha, \beta | \boldsymbol{\theta}, \mathbf{y}_2) = \frac{p(\mathbf{y}_2) p(\alpha, \beta)}{p(\mathbf{y}_2)} = p(\alpha, \beta)$$

□

C.4 Expansion of $p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\theta})$

Proof. First $p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta})$ is transformed into

$$\int_{\eta_T} \dots \int_{\eta_2} p\left(E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \dots, E_T = y_{1,T} - \alpha - \beta y_{2,T}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\eta_2 \dots d\eta_T$$

by applying the law of total probability $T - 1$ times. The definition of conditional probability, i.e. $p(A, B) = p(A|B)p(B)$, gives:

$$\int_{\boldsymbol{\eta}_{2:T}} p\left(E_T = y_{1,T} - \alpha - \beta y_{2,T} | \mathbf{E}_{1:(T-1)} = \mathbf{y}_{1,1:(T-1)} - \alpha \mathbf{1} - \beta \mathbf{y}_{2,1:(T-1)}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) p\left(\mathbf{E}_{1:(T-1)} = \mathbf{y}_{1,1:(T-1)} - \alpha \mathbf{1} - \beta \mathbf{y}_{2,1:(T-1)}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\boldsymbol{\eta}_{2:T}$$

As described in Equation 5.9 the probability of \mathbf{Y}_1 given A , B and \mathbf{Y}_2 is the same as the probability of \mathbf{E} given A , B and \mathbf{Y}_2 . E_t also depends on $\Phi \in \boldsymbol{\Theta}$, E_{t-1} and H_t as described in Equation 5.2, however it does not depend on any E s further in the past, or on any previous H s or Y_2 s. Therefore, simplifying the conditional probabilities and repeating the

law of total probability, gives:

$$\begin{aligned} & p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &= \int_{\eta_T} p(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, H_T = \eta_T, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ & \quad \int_{\boldsymbol{\eta}_{2:(T-1)}} p(\mathbf{E}_{1:(T-1)} = \mathbf{y}_{1,1:(T-1)} - \alpha \mathbf{1} - \beta \mathbf{y}_{2,1:(T-1)}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\boldsymbol{\eta}_{2:(T-1)} d\eta_T \end{aligned}$$

and using the definition of conditional probability again gives:

$$\begin{aligned} & p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &= \int_{\eta_T} p(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, H_T = \eta_T, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ & \quad \int_{\eta_{T-1}} p(E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1} | E_{T-2} = y_{1,T-2} - \alpha - \beta y_{2,T-2}, H_{T-1} = \eta_{T-1}, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ & \quad \int_{\boldsymbol{\eta}_{2:(T-2)}} p(\mathbf{E}_{1:(T-2)} = \mathbf{y}_{1,1:(T-2)} - \alpha \mathbf{1} - \beta \mathbf{y}_{2,1:(T-2)}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\boldsymbol{\eta}_{2:(T-2)} d\eta_{T-1} d\eta_T \end{aligned}$$

Repeating these steps another $T - 2$ times results in

$$\begin{aligned} & p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &= \int_{\eta_T} p(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, H_T = \eta_T, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ & \quad \dots \int_{\eta_2} p(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ & \quad p(E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\eta_2 \dots d\eta_T \end{aligned}$$

Now the definition of conditional probability can be used on $p(E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta})$ to get

$$p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T}, \boldsymbol{\Theta} = \boldsymbol{\theta}) p(\mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}).$$

However, since E_1 does not depend on any H s, as defined in Equation 5.3, $p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T}, \boldsymbol{\Theta} = \boldsymbol{\theta})$ can be simplified to

$$p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}).$$

The same definition of conditional probability can be used on $p(\mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta})$ to show that:

$$\begin{aligned} p(\mathbf{H}_{2:T} = \boldsymbol{\eta}_{2:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}) &= p(H_2 = \eta_2 | \mathbf{H}_{3:T} = \boldsymbol{\eta}_{3:T}, \boldsymbol{\Theta} = \boldsymbol{\theta}) p(\mathbf{H}_{3:T} = \boldsymbol{\eta}_{3:T} | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &\quad \vdots \\ &= p(H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \dots p(H_T = \eta_T | \boldsymbol{\Theta} = \boldsymbol{\theta}) \end{aligned}$$

which results in

$$\begin{aligned} p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) &= \int_{\eta_T} p(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, H_T = \eta_T, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &\quad \dots \int_{\eta_2} p(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &\quad p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) p(H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\eta_2 \dots p(H_T = \eta_T | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\eta_T \end{aligned}$$

The inner most product $p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) p(H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta})$ can be rewritten as

$$p(E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta})$$

which results in the following for the inner most integral:

$$\begin{aligned} &\int_{\eta_2} p(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &\quad p(E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\eta_2 \\ &= \int_{\eta_2} p(E_2 = y_{1,2} - \alpha - \beta y_{2,2}, E_1 = y_{1,1} - \alpha - \beta y_{2,1}, H_2 = \eta_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) d\eta_2 \end{aligned}$$

According to the law of total probability this is equivalent to

$$\begin{aligned} & p\left(E_2 = y_{1,2} - \alpha - \beta y_{2,2}, E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \\ &= p\left(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) p\left(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \end{aligned}$$

This can now be brought outside all the integrals, as it does not depend on $\boldsymbol{\eta}$ any more:

$$\begin{aligned} & p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\ &= p\left(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) p\left(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \\ & \quad \int_{\eta_T} p\left(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, H_T = \eta_T, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \\ & \quad \dots \int_{\eta_3} p\left(E_3 = y_{1,3} - \alpha - \beta y_{2,3} | E_2 = y_{1,2} - \alpha - \beta y_{2,2}, H_3 = \eta_3, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \\ & \quad p\left(H_3 = \eta_3 | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\eta_3 \dots p\left(H_T = \eta_T | \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\eta_T \end{aligned}$$

NB that $p(\eta_t | \boldsymbol{\theta}) = p(\eta_t | \epsilon_{t-1}, \boldsymbol{\theta})$, since η_t given $\boldsymbol{\theta}$, does not depend on any other variable, and therefore

$$\begin{aligned} & \int_{\eta_3} p\left(E_3 = y_{1,3} - \alpha - \beta y_{2,3} | E_2 = y_{1,2} - \alpha - \beta y_{2,2}, H_3 = \eta_3, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \\ & \quad p\left(H_3 = \eta_3 | E_2 = y_{1,2} - \alpha - \beta y_{2,2}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\eta_3 \\ &= \int_{\eta_3} p\left(E_3 = y_{1,3} - \alpha - \beta y_{2,3}, H_3 = \eta_3 | E_2 = y_{1,2} - \alpha - \beta y_{2,2}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) d\eta_3 \\ &= p\left(E_3 = y_{1,3} - \alpha - \beta y_{2,3} | E_2 = y_{1,2} - \alpha - \beta y_{2,2}, \boldsymbol{\Theta} = \boldsymbol{\theta}\right) \end{aligned}$$

This can again be pull outside of the integrals. This process can be repeated until the last integral, $\int_{\eta_T} \dots d\eta_T$, is evaluated.

$$\begin{aligned}
& p(\mathbf{E} = \mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2 | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\
&= p(E_1 = y_{1,1} - \alpha - \beta y_{2,1} | \boldsymbol{\Theta} = \boldsymbol{\theta}) \\
& \quad p(E_2 = y_{1,2} - \alpha - \beta y_{2,2} | E_1 = y_{1,1} - \alpha - \beta y_{2,1}, \boldsymbol{\Theta} = \boldsymbol{\theta}) \\
& \quad \dots p(E_T = y_{1,T} - \alpha - \beta y_{2,T} | E_{T-1} = y_{1,T-1} - \alpha - \beta y_{2,T-1}, \boldsymbol{\Theta} = \boldsymbol{\theta})
\end{aligned}$$

□

C.5 Computation of $p(\epsilon_t|\epsilon_{t-1}, \boldsymbol{\theta})$

Since $H_t \sim \mathcal{N}(0, S^2)$ and $E_1 | \boldsymbol{\Theta} \sim \mathcal{N}(0, S_1^2)$ as defined in Equation 5.3, the distribution of E_t given E_{t-1} and $\boldsymbol{\Theta}$ is again a normal distribution, since a linear combination of normally distributed random variables is again normal. Due to equation 5.2, i.e. $E_t = \Phi E_{t-1} + H_t$, the expected value of E_t given E_{t-1} and $\boldsymbol{\Theta}$ is

$$\mathbb{E}(E_t | \epsilon_{t-1}, \boldsymbol{\theta}) = \phi \epsilon_{t-1}$$

Similarly, given that E_{t-k} is independent of H_t , for $k > 0$,

$$\begin{aligned}
\text{Var}(E_t | \epsilon_{t-1}, \boldsymbol{\theta}) &= \phi^2 \text{Var}(E_{t-1} | \epsilon_{t-1}) + \text{Var}(H_t | \boldsymbol{\theta}) + \phi \text{Cov}(E_{t-1}, H_t) \\
&= 0 + \sigma^2 + 0
\end{aligned}$$

for all $t > 1$.

C.6 Proof that $\mathbb{E}(E_t) = 0 \forall t$

Proof. As $E_1|\Theta \sim \mathcal{N}(0, S_1^2)$ and $E_t = \Phi E_{t-1} + H_t$, as defined in Equations 5.2 and 5.3:

$$\mathbb{E}(E_t) = \mathbb{E}(\Phi E_{t-1} + H_t) = \Phi \mathbb{E}(E_{t-1}) = \dots = \Phi^{t-1} \mathbb{E}(E_1) = 0.$$

□

C.7 Computation of Σ

Again $\epsilon_1|\theta \sim \mathcal{N}(0, \sigma_1^2)$ and $\epsilon_t = \phi \epsilon_{t-1} + \eta_t$ are the main underlying equations used in this proof, as well as the fact that η_t is independent of $\epsilon_{t-k} \forall t, k$. From these it is easy to see that:

$$\text{Var}(\epsilon_1) = \sigma_1^2$$

$$\begin{aligned} \text{Var}(\epsilon_t) &= \text{Cov}(\epsilon_t, \epsilon_t) = \mathbb{E}(\epsilon_t^2) = \mathbb{E}((\phi \epsilon_{t-1} + \eta_t)^2) = \phi^2 \mathbb{E}(\epsilon_{t-1}^2) + 2\phi \mathbb{E}(\epsilon_{t-1} \eta_t) + \mathbb{E}(\eta_t^2) \\ &= \phi^2 \mathbb{E}(\epsilon_{t-1}^2) + \sigma^2 = \phi^{2(t-1)} \sigma_1^2 + \sum_{k=0}^{t-2} \phi^{2k} \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(\epsilon_t, \epsilon_1) &= \text{Cov}(\phi \epsilon_{t-1} + \eta_t, \epsilon_1) = \text{Cov}(\phi \epsilon_{t-1}, \epsilon_1) + \text{Cov}(\eta_t, \epsilon_1) = \phi \text{Cov}(\epsilon_{t-1}, \epsilon_1) \\ &= \dots = \phi^{t-1} \text{Var}(\epsilon_1) \end{aligned}$$

$$\begin{aligned} \text{Cov}(\epsilon_{t+k}, \epsilon_t) &= \text{Cov}(\phi \epsilon_{t+k-1} + \eta_{t+k}, \epsilon_t) = \text{Cov}(\phi \epsilon_{t+k-1}, \epsilon_t) + \text{Cov}(\eta_{t+k}, \epsilon_t) \\ &= \phi \text{Cov}(\epsilon_{t+k-1}, \epsilon_t) = \dots = \phi^k \text{Var}(\epsilon_t) \end{aligned}$$

Which fully defines the covariance matrix of ϵ_t , Σ .

C.8 Computation of C_2

The initial equation to start with is

$$\begin{aligned} p(\epsilon_1 | \boldsymbol{\theta}) \prod_{s=2}^T p(\epsilon_s | \epsilon_{s-1}, \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi)^T |\boldsymbol{\Sigma}|}} \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \right\} \\ &= \frac{1}{\sqrt{(2\pi)^T |\boldsymbol{\Sigma}|}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2) \right\} \end{aligned}$$

The choice of $C_1 = \frac{1}{\sqrt{(2\pi)^T |\boldsymbol{\Sigma}|}}$ is clear, however to achieve C_2 the part within the exponential function has to be rewritten:

$$\begin{aligned} \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} &= (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_1 - \alpha \mathbf{1} - \beta \mathbf{y}_2) \\ &= \mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 - \mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \alpha \mathbf{1} - \mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \beta \mathbf{y}_2 - (\alpha \mathbf{1})' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 + (\alpha \mathbf{1})' \boldsymbol{\Sigma}^{-1} \alpha \mathbf{1} + (\alpha \mathbf{1})' \boldsymbol{\Sigma}^{-1} \beta \mathbf{y}_2 \\ &\quad - (\beta \mathbf{y}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 + (\beta \mathbf{y}_2)' \boldsymbol{\Sigma}^{-1} \alpha \mathbf{1} + (\beta \mathbf{y}_2)' \boldsymbol{\Sigma}^{-1} \beta \mathbf{y}_2 \\ &= \mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 + \alpha c_\alpha + \beta c_\beta + \alpha c_\alpha + \alpha^2 (\Lambda^{-1})_{11} + \alpha \beta (\Lambda^{-1})_{12} + \beta c_\beta + \alpha \beta (\Lambda^{-1})_{21} + \beta^2 (\Lambda^{-1})_{22} \\ &= \mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + 2c_\alpha \alpha + 2c_\beta \beta \end{aligned}$$

where $(\Lambda^{-1})_{11} = \mathbf{1}' \boldsymbol{\Sigma}^{-1} \mathbf{1}$, $(\Lambda^{-1})_{22} = \mathbf{y}_2' \boldsymbol{\Sigma}^{-1} \mathbf{y}_2$, $(\Lambda^{-1})_{12} = (\Lambda^{-1})_{21} = \mathbf{1}' \boldsymbol{\Sigma}^{-1} \mathbf{y}_2$, $c_\alpha = -\mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{1}$ and $c_\beta = -\mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_2$ as listed in equations 5.14 and 5.13.

Therefore

$$\begin{aligned} &\exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + 2c_\alpha \alpha + 2c_\beta \beta \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y}_1 - \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + 2c_\alpha \alpha + 2c_\beta \beta + 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \right] \right. \\ &\quad \left. - \frac{1}{2} \left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \right] \right\} \end{aligned}$$

Which is just $C_2 \exp \left\{ -\frac{1}{2} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix} \right\}$ with

$$C_2 = \exp \left\{ -\frac{1}{2} \left[\mathbf{y}'_1 \Sigma^{-1} \mathbf{y}_1 - \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} + 2\alpha [c_\alpha + \mu_\alpha(\Lambda^{-1})_{11} + \mu_\beta(\Lambda^{-1})_{12}] \right. \right. \\ \left. \left. + 2\beta [c_\beta + \mu_\alpha(\Lambda^{-1})_{12} + \mu_\beta(\Lambda^{-1})_{22}] \right] \right\}$$

$\forall \mu_\alpha$ and μ_β .

C.9 Computation of μ_α^* and μ_β^*

We have two equations and two unknowns, μ_α^* and μ_β^* :

$$\begin{aligned} c_\alpha + \mu_\alpha^*(\Lambda^{-1})_{11} + \mu_\beta^*(\Lambda^{-1})_{12} &\stackrel{!}{=} 0 \\ \wedge \quad c_\beta + \mu_\alpha^*(\Lambda^{-1})_{12} + \mu_\beta^*(\Lambda^{-1})_{22} &\stackrel{!}{=} 0 \end{aligned}$$

Which can be rewritten in matrix format as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} + \begin{pmatrix} (\Lambda^{-1})_{11} & (\Lambda^{-1})_{12} \\ (\Lambda^{-1})_{12} & (\Lambda^{-1})_{22} \end{pmatrix} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} + \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix}$$

Which can easily be solved for $\begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix}$

$$\begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} = -\Lambda \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix}$$

C.10 Computation of $\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix}$

Combining the two inner parts of the exponential function gives:

$$\begin{aligned} & \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix} + \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix} \\ &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \\ & \quad - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \end{aligned}$$

The identity matrix can be inserted into the combined mixed element, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}$, in form of $(\Lambda^{-1} + \Omega^{-1})(\Lambda^{-1} + \Omega^{-1})^{-1}$, (this is possible, since both Λ and Ω are square positive definite matrices and therefore their sum is again positive definite as described in Observation 7.1.3 on p.430 in Horn's book called Matrix Analysis [34]) which results in:

$$\begin{aligned} & \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1})(\Lambda^{-1} + \Omega^{-1})^{-1} \left[\Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \right] \\ &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} \end{aligned}$$

with $\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} := (\Lambda^{-1} + \Omega^{-1})^{-1} \left[\Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} \right]$.

Therefore

$$\begin{aligned} & \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \alpha - \mu_\alpha^* \\ \beta - \mu_\beta^* \end{pmatrix} + \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \alpha - \mu_\alpha \\ \beta - \mu_\beta \end{pmatrix} \\ &= \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix}' \Lambda^{-1} \begin{pmatrix} \mu_\alpha^* \\ \mu_\beta^* \end{pmatrix} + \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} - 2 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}' (\Lambda^{-1} + \Omega^{-1}) \begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix}. \end{aligned}$$

C.11 Calculation of Relationship between the Cointegration Parameters given that the two time series are Cointegrated

As shown in Section 5.6 two time series are cointegrated if a linear relationship of the two time series is stationary, and therefore, if $\text{Cov}(E_1, E_{1+k}) - \text{Cov}(E_t, E_{t+k}) = 0 \forall t, k$. It has also been shown that

$$\text{Cov}(E_1, E_{1+k}) - \text{Cov}(E_t, E_{t+k}) = \Phi^k \left((1 - \Phi^{2(t-1)})S_1^2 - \sum_{i=0}^{t-2} \Phi^{2i} S^2 \right).$$

Assume first that $t = 2$. Then we have

$$0 = \text{Cov}(E_1, E_{1+k}) - \text{Cov}(E_2, E_{2+k}) = \Phi^k \left((1 - \Phi^2)S_1^2 - S^2 \right).$$

Since $t = 2$, k can either be 0, or 1. If $k = 0$ the equation simplifies to

$$0 = (1 - \Phi^2)S_1^2 - S^2$$

which results in

$$\Phi = \pm \sqrt{1 - S^2/S_1^2}.$$

When $k = 1$ there are two solutions

$$\Phi = 0 \quad \text{and} \quad \Phi = \pm \sqrt{1 - S^2/S_1^2}.$$

However, since the equation needs to hold for all k , the only solution is

$$\Phi = \pm \sqrt{1 - S^2/S_1^2}.$$

Furthermore, note that $1 - \Phi^{2(t-1)} = \sum_{i=0}^{t-2} \Phi^{2i}(1 - \Phi^2)$ since

$$\begin{aligned} & \left(1 + \Phi^2 + \Phi^4 + \dots + \Phi^{2(t-2)}\right) (1 - \Phi^2) \\ &= \left(1 - \Phi^2 + \Phi^2 - \Phi^4 + \Phi^4 - \dots - \Phi^{2(t-2)} + \Phi^{2(t-2)} - \Phi^{2(t-1)}\right) \\ &= \left(1 - \Phi^{2(t-1)}\right). \end{aligned}$$

Given this, it is easily shown that

$$(1 - \Phi^{2(t-1)})S_1^2 - \sum_{i=0}^{t-2} \Phi^{2i}S^2 = 0$$

is equivalent to

$$\frac{S^2}{S_1^2} = \frac{1 - \Phi^{2(t-1)}}{\sum_{i=0}^{t-2} \Phi^{2i}} = 1 - \Phi^2$$

and therefore the same solution holds for all t , i.e.

$$\Phi = \pm \sqrt{1 - S^2/S_1^2}.$$

This indicates that S_1^2 needs to be greater than S^2 , as we would get complex results otherwise. This makes sense, since the variance of the white noise process \mathbf{H} would overpower the variance of the process \mathbf{E} , which is the process being tested for stationarity.

C.12 Calculation of $|\text{Var}(E_1) - \text{Var}(E_t)|$

E_t is defined recursively as $\Phi E_{t-1} + H_t$ in Equation 5.2. Therefore we have

$$\begin{aligned}
|\text{Var}(E_1) - \text{Var}(E_t)| &= |\text{Var}(E_1) - \text{Var}(\Phi E_{t-1} + H_t)| \\
&= |\text{Var}(E_1) - (\Phi^2 \text{Var}(\Phi E_{t-2} + H_{t-1}) + S^2)| \\
&\quad \vdots \\
&= |\text{Var}(E_1) - \Phi^{2(t-1)} \text{Var}(E_1) - \sum_{i=0}^{t-2} \Phi^{2i} S^2| \\
&= |(1 - \Phi^{2(t-1)}) S_1^2 - \sum_{i=0}^{t-2} \Phi^{2i} S^2|
\end{aligned}$$

since $\text{Var}(H_t) = S^2$, $\text{Var}(E_1) = S_1^2$ and $\text{Cov}(E_{t-k}, H_t) = 0 \forall k > 0$.

C.13 Calculation of the bounds on $\|\Sigma(\tilde{\theta}_i) - \Sigma(\theta_i)\|_{\text{OP}}$

From the definition of the operator norm, detailed in Definition 4.5, it is known that

$$\|A\|_{\text{OP}} = \sup_{\|\mathbf{v}\|=1} \{\|A\mathbf{v}\|\} = \sup_{\|\mathbf{v}\|=1} \left\{ \sup_{\|\mathbf{w}\|=1} \{\mathbf{w}'A\mathbf{v}\} \right\}.$$

An upper bound on this can be found by calculating the maximum with respect to \mathbf{w} for each individual part of the sum, and then the full sum, including the results of the previous maxima, with respect to \mathbf{v} , each with the respective restrictions. For this the Lagrangian method for constrained maximisation is used [55].

Let the element of A in the i th row and the j th column be denoted by a_{ij} . Then an upper bound on the maximum of $\mathbf{w}'A\mathbf{v} = v_1 \sum_{j=1}^T a_{j1}w_j + v_2 \sum_{j=1}^T a_{j2}w_j + \dots + v_T \sum_{j=1}^T a_{jT}w_j$ with the constraint that $\|\mathbf{w}\| = 1$ is done by finding the maxima of

$$\mathcal{L}_i := \sum_{j=1}^T a_{ji}w_j - \lambda_i(w_1^2 + w_2^2 + \dots + w_T^2 - 1)$$

for $1 \leq i \leq T$. The values of w_{ji} that maximise these equations are easily shown to be $w_{ji} = a_{ji} / \sqrt{\sum_{j=1}^T a_{ji}^2}$.

Updating $\mathbf{w}'A\mathbf{v}$ with these values for the w_{ji} 's gives $v_1\sqrt{\sum_{j=1}^T a_{j1}^2} + \dots + v_T\sqrt{\sum_{j=1}^T a_{jT}^2}$ and therefore the final equation to maximise is

$$\mathcal{L} := \sum_{i=1}^T v_i \sqrt{\sum_{j=1}^T a_{ji}^2} - \lambda(v_1^2 + v_2^2 + \dots + v_T^2 - 1).$$

This is done in the same fashion as the previous maximisation and results in $v_i = \sqrt{\sum_{j=1}^T a_{ji}^2 / \sum_{i=1}^T \sum_{j=1}^T a_{ji}^2}$. This results in an upper bound on $\|A\|_{\text{OP}}$ being $\sqrt{\sum_{i=1}^T \sum_{j=1}^T a_{ji}^2}$.

And therefore an upper bound on $\|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$ is the square root of the sum of the square of all the elements of $\Sigma(\boldsymbol{\theta} + \boldsymbol{\delta})_i - \Sigma(\boldsymbol{\theta})_i$.

C.14 Computation of $\prod_{j=1}^T \frac{|\lambda_{i,j}-e_i|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\lambda_{i,j}+e_i|}{|\lambda_{i,j}|}$

The Bauer-Fike theorem on eigenvalue perturbation [6] gives an absolute upper bound for the difference between a known eigenvalue of a matrix to the eigenvalue of a perturbation of the same matrix. The matrix has to be complex-valued and diagonalizable, which is the case for the covariance matrix Σ [12], as it is symmetric and real. The theorem states:

Theorem C.1 (Bauer-Fike theorem). *$A \in \mathbb{C}^{n \times n}$ diagonalizable matrix, $V \in \mathbb{C}^{n \times n}$ non-singular eigenvector matrix such that $A = V\Lambda V^{-1}$. Let μ be an eigenvalue of $A + \delta A$; then an eigenvalue $\lambda \in \sigma(A)$ exists such that $|\lambda - \mu| \leq \kappa_p(V)\|\delta A\|_p$ where $\kappa_p(V) = \|V\|_p\|V^{-1}\|_p$ is the so called condition number in p-norm.*

The matrix p-norm referred to is $\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$, corresponding to the p-norm for vectors. Therefore this theorem is consistent for the operator norm when using the euclidean vector norm. Additionally, because Σ is a real symmetric matrix it can be

decomposed as $\Sigma = V\Lambda V'$ [12]. In the notation used so far this surmounts to:

$$|\tilde{\lambda}_{i,j} - \lambda_{i,j}| \leq \|V_i\|_{\text{OP}} \|V_i'\|_{\text{OP}} \|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$$

where V_i is the eigenvector matrix associated with the unperturbed matrix and is therefore known, and an upper bound on $\|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$ is given in Sections 5.7.1.1.2. Therefore

$$\lambda_{i,j} - e_i \leq \tilde{\lambda}_{i,j} \leq \lambda_{i,j} + e_i$$

where $e_i = \|V_i\|_{\text{OP}} \|V_i'\|_{\text{OP}} \|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$.

Additionally we can show that the operator norm of V is equal to 1 and that therefore $e_i = \|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$: Let $V = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ which is a basis and therefore orthogonal. Without loss of generality, assume V is an orthonormal basis. If they are not, it can easily be induced by dividing the eigenvectors by their norm. Similarly, let \mathbf{x} be a unit length vector in the space spanned by V . Then $\mathbf{x} = \sum_i a_i \mathbf{v}_i$ for some $\mathbf{a} = (a_1, \dots, a_n)$. The euclidean norm of \mathbf{x} is then

$$1 = \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_i a_i^2 \langle \mathbf{v}_i, \mathbf{v}_i \rangle} = \sqrt{\sum_i a_i^2}.$$

The operator norm is defined in Definition 4.5 and from this we know that

$$\|V\|_{\text{OP}} = \sup_{\|\mathbf{x}\|=1} \|V\mathbf{x}\| = \|V\hat{\mathbf{x}}\|,$$

for a point $\hat{\mathbf{x}}$ with $\|\hat{\mathbf{x}}\| = 1$. From the previous calculation we know that

$$V\hat{\mathbf{x}} = (\mathbf{v}_1, \dots, \mathbf{v}_n) \sum_i \hat{a}_i \mathbf{v}_i = \sum_i \hat{a}_i (\mathbf{v}_1, \dots, \mathbf{v}_n) \mathbf{v}_i.$$

Since V is orthonormal $V\mathbf{v}_i$ is a vector of zeros with a one at the i th point. Therefore

$\sum_i \hat{a}_i V \mathbf{v}_i = (a_1, \dots, a_n)'$ and finally this gives

$$\|V\|_{\text{OP}} = \|(a_1, \dots, a_n)'\| = \sqrt{\sum_i a_i^2} = 1.$$

And therefore

$$\prod_{j=1}^T \frac{|\lambda_{i,j} - e_i|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\lambda_{i,j} + e_i|}{|\lambda_{i,j}|}. \quad (\text{C.4})$$

as long as $\lambda_{i,j} \geq e_i$, with $e_i = \|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$, which is the case, since in Section 5.7.1.1.11 it is shown that $\|\tilde{\Sigma}_i - \Sigma_i\|_{\text{OP}}$ is smaller than the smallest Eigenvalue of Σ_i .

C.15 Calculation of upper bound on $\|\tilde{\Lambda}_i - \Lambda_i\|$

Following the same method as in Section 5.7.1.1.3 it is known that, since $\|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} < 1/\|\Lambda_i\|_{\text{OP}}$, as detailed in Section 5.7.1.1.11,

$$\|\tilde{\Lambda}_i - \Lambda_i\|_{\text{OP}} \leq \|\Lambda_i\|_{\text{OP}} \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \|\tilde{\Lambda}_i\|_{\text{OP}}$$

and

$$\|\tilde{\Lambda}_i\|_{\text{OP}} \leq \left(\frac{1}{\|\Lambda_i\|_{\text{OP}}} - \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \right)^{-1} = \left(\min_{1 \leq j \leq 2} |l_{i,j}| - \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \right)^{-1}$$

where $l_{i,j}$ is the j th eigenvalue of Λ_i^{-1} . So in total

$$\|\tilde{\Lambda}_i - \Lambda_i\|_{\text{OP}} \leq \|\Lambda_i\|_{\text{OP}} \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \left(\min_{1 \leq j \leq 2} |l_{i,j}| - \|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} \right)^{-1}$$

of which an upper bound on $\|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|$ has been calculated in Section 5.7.1.1.5.

C.16 Proof that Λ^{-1} is Positive Definite

As detailed in Equation 5.14 Λ^{-1} is defined as

$$\Lambda^{-1} = \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \Sigma^{-1} (\mathbf{1} \ \mathbf{y}_2).$$

Σ has been shown to be positive definite in Section 5.5. This means that

$$z' \Lambda^{-1} z = z' \begin{pmatrix} \mathbf{1}' \\ \mathbf{y}'_2 \end{pmatrix} \Sigma^{-1} (\mathbf{1} \ \mathbf{y}_2) z = s' \Sigma^{-1} s > 0,$$

which means that Λ^{-1} is also positive definite.

C.17 Upper bound on $\left\| \left(\tilde{\Lambda}^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}}$

For this the two results from Dirk Ferus [22] as detailed in Equations 4.3 and 4.4 are used.

For the first theorem $\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1}$ and $\left(\Lambda_i^{-1} + \Omega^{-1} \right)^{-1}$ are required to be invertible. This is the case, since both Λ_i^{-1} and Ω are positive definite (see Sections 5.3 and 5.5) and therefore their sum is also positive definite and invertible. Additionally the size of the sub-blocks is restricted such that $\left(\tilde{\Lambda}_i^{-1} + \Omega^{-1} \right)^{-1}$ is also invertible (see Section 5.7.1.1.11).

Furthermore, it is necessary that $\|\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}\|_{\text{OP}} < 1/\|\Lambda^{-1} + \Omega^{-1}\|_{\text{OP}}$, which is a requirement for both results used here. This is again included as a constraint to the size of each sub-block in Section 5.7.1.1.11.

Hence,

$$\begin{aligned} & \left\| \left(\tilde{\Lambda}^{-1} + \Omega^{-1} \right)^{-1} - \left(\Lambda^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \\ & \leq \left\| \left(\tilde{\Lambda}^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \left\| \left(\tilde{\Lambda}^{-1} + \Omega^{-1} \right) - \left(\Lambda^{-1} + \Omega^{-1} \right) \right\|_{\text{OP}} \left\| \left(\Lambda^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \\ & = \left\| \left(\tilde{\Lambda}^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}} \left\| \left(\Lambda^{-1} + \Omega^{-1} \right)^{-1} \right\|_{\text{OP}} \end{aligned}$$

Here $\left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}$ is known and $\left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}}$ has been previously calculated in Equation 5.30. To be able to get an upper bound on $\left\| (\tilde{\Lambda}^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}$ the second result by Ferus [22], detailed in Equation 4.4, is used

$$\begin{aligned} \left\| (\tilde{\Lambda}^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}} &\leq \left(\frac{1}{\left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}} - \left\| (\tilde{\Lambda}^{-1} + \Omega^{-1}) - (\Lambda^{-1} + \Omega^{-1}) \right\|_{\text{OP}} \right)^{-1} \\ &= \left(\frac{1}{\left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}} - \left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}} \right)^{-1} \end{aligned}$$

where $1/\left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}$ is the smallest absolute eigenvalue of $\Lambda^{-1} + \Omega^{-1}$ and an upper bound on $\left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}}$ has been calculated in Equation 5.30.

Overall this gives:

$$\begin{aligned} &\left\| (\tilde{\Lambda}^{-1} + \Omega^{-1})^{-1} - (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}} \\ &\leq \left(\frac{1}{\left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}}} - \left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}} \right)^{-1} \left\| \tilde{\Lambda}^{-1} - \Lambda^{-1} \right\|_{\text{OP}} \left\| (\Lambda^{-1} + \Omega^{-1})^{-1} \right\|_{\text{OP}} \end{aligned}$$

C.18 Upper bound on $\left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right|$

This calculation is very long and tedious is is purely based on adding zeros in convenient forms, such as $-\tilde{\nu}'_i(\Lambda_i^{-1} + \Omega^{-1})\tilde{\nu}_i + \tilde{\nu}'_i(\Lambda_i^{-1} + \Omega^{-1})\tilde{\nu}_i$, to expand the sum into non-perturbed parts and the norm of the difference between perturbed and non-perturbed variables, which have already been calculated previously.

Once the sum has been expanded sufficiently, the triangle rule, as well as the sub-multiplicativity of the operator norm and absolute value, are used to achieve the final upper bound.

The details of the initial transformation are not included fully, only the first couple of lines

and the final result are presented, to give an idea of the method used.

$$\begin{aligned}
& \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right| \\
&= \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\nu_i + \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\nu_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right| \\
&= \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) + \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\nu_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right| \\
&= \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) - \tilde{\nu}'_i(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) + \tilde{\nu}'_i(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) \right. \\
&\quad \vdots \\
&= \left| (\tilde{\nu}'_i - \nu'_i)(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i) + \nu'_i(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i) \right. \\
&\quad + (\tilde{\nu}'_i - \nu'_i)(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) + \nu'_i(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) \\
&\quad \left. + \left((\tilde{\nu}'_i - \nu'_i)(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}) + \nu'_i(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1}) + (\tilde{\nu}'_i - \nu'_i)(\Lambda_i^{-1} + \Omega^{-1}) \right) \nu_i \right|
\end{aligned}$$

Now the triangle rule can be applied to achieve

$$\begin{aligned}
& \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right| \\
&\leq \left| (\tilde{\nu}'_i - \nu'_i)(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i) \right| + \left| \nu'_i(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i) \right| \\
&\quad + \left| (\tilde{\nu}'_i - \nu'_i)(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) \right| + \left| \nu'_i(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i) \right| \\
&\quad + \left| (\tilde{\nu}'_i - \nu'_i)(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})\nu_i \right| + \left| \nu'_i(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})\nu_i \right| + \left| (\tilde{\nu}'_i - \nu'_i)(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right|
\end{aligned}$$

The product of two vectors, which is equivalent to the norm of a row vector multiplied with a matrix and multiplied with a column vector, can be interpreted as an inner product

and therefore the Cauchy-Schwarz inequality, $|\langle x, y \rangle| \leq \|x\| \|y\|$ [54], can be used to get

$$\begin{aligned} & \left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right| \\ & \leq \|(\tilde{\nu}'_i - \nu'_i)\| \|(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i)\| + \|\nu'_i\| \|(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})(\tilde{\nu}_i - \nu_i)\| \\ & \quad + \|(\tilde{\nu}'_i - \nu'_i)\| \|(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i)\| + \|\nu'_i\| \|(\Lambda_i^{-1} + \Omega^{-1})(\tilde{\nu}_i - \nu_i)\| \\ & \quad + \|(\tilde{\nu}'_i - \nu'_i)\| \|(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})\nu_i\| + \|\nu'_i\| \|(\tilde{\Lambda}_i^{-1} - \Lambda_i^{-1})\nu_i\| \\ & \quad + \|(\tilde{\nu}'_i - \nu'_i)\| \|(\Lambda_i^{-1} + \Omega^{-1})\nu_i\| \end{aligned}$$

Now the definition of the operator norm can be used again to separate the norm of a product of a matrix and a vector. The following transformation is used:

$$\|Aa\| = \left\| A \frac{a}{\|a\|} \right\| \|a\| \leq \sup_{\|u\|=1} \|Au\| \|a\| = \|A\|_{\text{OP}} \|a\|.$$

Applying this culminates in the following upper bound on

$$\left| \tilde{\nu}'_i(\tilde{\Lambda}_i^{-1} + \Omega^{-1})\tilde{\nu}_i - \nu'_i(\Lambda_i^{-1} + \Omega^{-1})\nu_i \right|:$$

$$\begin{aligned} & \| \tilde{\nu}_i - \nu_i \|^2 \| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \|_{\text{OP}} + 2 \| \nu_i \| \| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \|_{\text{OP}} \| \tilde{\nu}_i - \nu_i \| + \| \nu_i \|^2 \| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \|_{\text{OP}} \\ & + \| \tilde{\nu}_i - \nu_i \|^2 \| \Lambda_i^{-1} + \Omega^{-1} \|_{\text{OP}} + 2 \| \nu_i \| \| \Lambda_i^{-1} + \Omega^{-1} \|_{\text{OP}} \| \tilde{\nu}_i - \nu_i \| \end{aligned}$$

where an upper bound on $\| \tilde{\Lambda}_i^{-1} - \Lambda_i^{-1} \|_{\text{OP}}$ and $\| \tilde{\nu}_i - \nu_i \|$ have been calculated in Sections 5.7.1.1.5 and 5.7.1.1.8 respectively and the rest of the terms are known, unperturbed values.

C.19 Upper bound on $\left| \sqrt{\frac{\det \tilde{\Sigma}_i}{\det \Sigma_i}} \right|$

The determinant of a matrix is just the product of its eigenvalues and therefore

$$\left| \sqrt{\frac{\det \tilde{\Sigma}_i}{\det \Sigma_i}} \right| = \left| \sqrt{\prod_{j=1}^T \frac{\tilde{\lambda}_{i,j}}{\lambda_{i,j}}} \right| = \sqrt{\prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|}}$$

because Σ_i and $\tilde{\Sigma}_i$ are positive definite, as detailed in Sections 5.5 and 5.7.1.1.11, and therefore have all positive eigenvalues.

Additionally from Equation C.4 it is known that $\prod_{j=1}^T \frac{|\tilde{\lambda}_{i,j}|}{|\lambda_{i,j}|} \leq \prod_{j=1}^T \frac{|\lambda_{i,j} + e_i|}{|\lambda_{i,j}|}$ which results in

$$\left| \sqrt{\frac{\det \tilde{\Sigma}}{\det \Sigma}} \right| = \sqrt{\prod_{i=1}^T \frac{|\tilde{\lambda}_i|}{|\lambda_i|}} \leq \sqrt{\prod_{i=1}^T \frac{|\lambda_i + e_i|}{|\lambda_i|}}$$

with $e_i = \left\| \tilde{\Sigma}_i - \Sigma_i \right\|_{\text{OP}}$ as defined in C.14.

Appendix D

Results of the Cointegration Test

D.1 Tables of Proportions of Simulations Correctly Identified as having, or not having, an Association

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.01	0.01	0.0233	0.0010
0.9535	0.11	0.01	0.0266	0.7260
0.9759	0.21	0.01	0.0302	0.7935
0.9837	0.31	0.01	0.2115	0.8033
0.9877	0.41	0.01	0.4180	0.8078
0.9901	0.51	0.01	0.5646	0.8110
0.9918	0.61	0.01	0.5382	0.8136
0.9929	0.71	0.01	0.1263	0.8156
0.9938	0.81	0.01	0.0647	0.8173
0.9945	0.91	0.01	0.5278	0.8186
0.9950	1.01	0.01	0.2061	0.8198

Table D.1: Results of the proposed Cointegration Test and Engle and Granger method: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. It also gives the results In this table σ_1^2 is kept constant at 0.01.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.11	0.11	1.705e-07	0.0010
0.6901	0.21	0.11	0.0236	0.0010
0.8032	0.31	0.11	0.0890	0.0270
0.8554	0.41	0.11	0.1956	0.1244
0.8856	0.51	0.11	0.1164	0.2794
0.9054	0.61	0.11	0.2975	0.4157
0.9193	0.71	0.11	0.3767	0.5136
0.9296	0.81	0.11	0.1903	0.5851
0.9376	0.91	0.11	0.4388	0.6375
0.9440	1.01	0.11	0.4758	0.6761

Table D.2: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.11.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.21	0.21	5.165e-08	0.0010
0.5680	0.31	0.21	0.0094	0.0010
0.6984	0.41	0.21	0.0320	0.0010
0.7670	0.51	0.21	0.0903	0.0087
0.8098	0.61	0.21	0.3202	0.0330
0.8392	0.71	0.21	0.0457	0.0791
0.8607	0.81	0.21	0.7372	0.1435
0.8771	0.91	0.21	0.5513	0.2223
0.8900	1.01	0.21	0.2621	0.3091

Table D.3: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.21.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.31	0.31	2.022e-06	0.0010
0.4939	0.41	0.31	0.0273	0.0010
0.6262	0.51	0.31	0.4579	0.0010
0.7013	0.61	0.31	0.3127	0.0010
0.7506	0.71	0.31	0.6851	0.0049
0.7857	0.81	0.31	0.7499	0.0155
0.8120	0.91	0.31	0.8041	0.0353
0.8325	1.01	0.31	0.2700	0.0652

Table D.4: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.31.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.41	0.41	0.0021	0.0010
0.4428	0.51	0.41	0.4845	0.0010
0.5726	0.61	0.41	0.8356	0.0010
0.6500	0.71	0.41	0.8871	0.0010
0.7027	0.81	0.41	0.8720	0.0010
0.7412	0.91	0.41	0.8983	0.0041
0.7708	1.01	0.41	0.4196	0.0096

Table D.5: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.41.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.51	0.51	0.8310	0.0010
0.4049	0.61	0.51	0.8983	0.0010
0.5307	0.71	0.51	0.9125	0.0010
0.6086	0.81	0.51	0.9113	0.0010
0.6630	0.91	0.51	0.9084	0.0010
0.7036	1.01	0.51	0.9025	0.0010

Table D.6: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.51.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.61	0.61	0.9087	0.0010
0.3753	0.71	0.61	0.9210	0.0010
0.4969	0.81	0.61	0.8880	0.0010
0.5742	0.91	0.61	0.9169	0.0010
0.6293	1.01	0.61	0.9140	0.0010

Table D.7: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. In this table σ_1^2 is kept constant at 0.61.

Φ	σ_1^2	σ^2	Proportion within CT	Engle and Granger p-value
0	0.71	0.71	0.9166	0.0010
0.3514	0.81	0.71	0.9114	0.0010
0.4688	0.91	0.71	0.8386	0.0010
0.5450	1.01	0.71	0.7844	0.0010
0	0.81	0.81	0.9075	0.0010
0.3315	0.91	0.81	0.9033	0.0010
0.4450	1.01	0.81	0.9186	0.0010
0	0.91	0.91	0.9156	0.0010
0.3147	1.01	0.91	0.9135	0.0010
0	1.01	1.01	0.9035	0.0010

Table D.8: Results of the Cointegration Test: This table details the proportion of the posterior that was within the cointegration tube (CT - column 4) and the p-value from the Engle and Granger method (column 5) for the different combinations of ϕ , σ^2 and σ_1^2 which produce cointegrated time series. This table shows the results for σ_1^2 from 0.71, 0.81, 0.91 and 1.01.

Bibliography

- [1] B.O. Abidoye and M. Labuschagne. The transmission of world maize price to south african maize market: a threshold cointegration approach. *Agricultural Economics*, 45(4):501–512, 2014.
- [2] W. Ahmad, M.Z. Nain, and B. Kamaiah. On the role of the trend and cyclical components in electricity consumption and india’s economic growth: a cointegration and cofeature approach. *OPEC Energy Review*, 38(1):107–126, 2014.
- [3] K. Atkinson and W. Han. Linear operators on normed spaces. In *Theoretical Numerical Analysis*, volume 39 of *Texts in Applied Mathematics*, pages 51–113. Springer New York, 2009.
- [4] S. Axler. *Linear Algebra Done Right*, chapter 10. Trace and Determinant. Springer, 1997.
- [5] T.N. Bailey. *The African Leopard: Ecology and behavior of a solitary felid*. Columbia University Press, New York, 1993.
- [6] F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numerische Mathematik*, 2:137–141, 1960.
- [7] L. Bauwens and P. Giot. A gibbs sampling approach to cointegration. *Computational Statistics*, 13:339–368, 1997.
- [8] L. Bauwens and M. Lubrano. Identification restrictions and posterior densities in

- cointegrated gaussian var systems. CORE Discussion Papers 1994018, Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE), 1994.
- [9] Jos M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*, chapter 6. Remodelling, pages 377–426. John Wiley & Sons, Inc., 2008.
- [10] C. Bracegirdle and D. Barber. Bayesian conditional cointegration. In *29th International Conference on Machine Learning (ICML)*, 2012.
- [11] J.L. Brown. Optimal group size in territorial animals. *Journal of Theoretical Biology*, 95(4):793–810, April 1982.
- [12] D. Cherney, T. Denton, and A. Waldron. *Linear Algebra*, chapter 15 Diagonalizing Symmetric Matrices. UC Davis, 2013.
- [13] David Craufurd, Jennifer C Thompson, and Julie S Snowden. Behavioral changes in huntington disease. *Cognitive and Behavioral Neurology*, 14(4):219–226, 2001.
- [14] B. M. Damghani, D. Welch, C. O’Malley, and S. Knights. The misleading value of measured correlation. *Wilmott*, pages 64–73, 2012.
- [15] David de la Croix and Michel Lubrano. Are interest rates responsible for unemployment in the eighties? a bayesian analysis of cointegrated relationship with a regime shift. Discussion Papers (IRES - Institut de Recherches Economiques et Sociales) 1994015, Universit catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES), 1996.
- [16] M.d.M. Delgado, V. Penteriani, J.M. Morales, E. Gurarie, and O. Ovaskainen. A statistical framework for inferring the influence of conspecifics on movement behaviour. *Methods in Ecology and Evolution*, 5(2):183–189, 2014.
- [17] C.P. Doncaster. Non-parametric estimates of interaction from radio-tracking data. *Journal of Theoretical Biology*, 143(4):431 – 443, 1990.

- [18] Beatrice C Downing and Nick J Royle. *Animal Social Networks*. John Wiley & Sons, Ltd, 2013.
- [19] J.E. Dunn. A complete test for dynamic territorial interaction. In *Proceedings: Second International Conference on Wildlife Biotelemetry*, pages 159–169. International Conference on Wildlife Biotelemetry, 1979.
- [20] L.M. Elbroch, H.B. Quigley, and A. Caragiulo. Spatial associations in a solitary predator: using genetic tools and gps technology to assess cougar social organization in the southern yellowstone ecosystem. *acta ethologica*, pages 1–10, 2014.
- [21] R. F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, March 1987.
- [22] D. Ferus. *Analysis II Lecture Notes*. Technische Universitaet Berlin, 2004. <http://page.math.tu-berlin.de/~ferus/skripten.html>.
- [23] W.A. Fuller. *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley, 1996.
- [24] T. Furnston, S. Hailes, and A.J. Morton. A bayesian residual-based test for cointegration. *arXiv preprint arXiv:1311.0524*, 2013.
- [25] Thomas Furnston, A. Jennifer Morton, and Stephen Hailes. A significance test for inferring affiliation networks from spatio-temporal data. *PLoS ONE*, 10(7):e0132417, 07 2015.
- [26] A. Gelman. Prior distribution for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1(3):pp. 515–534, 2006.
- [27] S. S. Ghouri. Forecasting natural gas prices using cointegration technique. *OPEC review : energy economics and related issues*, 2006.
- [28] Steven N. Goodman. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine*, 130(12):995–1004, 1999.

- [29] C.W.J. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111 – 120, 1974.
- [30] J. D. Hamilton. *Time Series Analysis*, chapter 3 Stationary ARMA Processes. Princeton University Press, 1994.
- [31] J. He. *Math1432 Calculus II: Lecture Notes*. University of Houston, 2008. http://www.math.uh.edu/~jiwenhe/Math1432/lectures/lecture11_handout.pdf.
- [32] S.P. Henzi, D. Lusseau, T. Weingrill, C.P. van Schaik, and L. Barrett. Cyclicality in the structure of female baboon social networks. *Behavioral Ecology and Sociobiology*, 63(7):1015–1021, 2009.
- [33] Judith A. Hobert and Glyn Dawson. Neuronal ceroid lipofuscinoses therapeutic strategies: Past, present and future. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1762(10):945 – 953, 2006. Molecular Basis of {NCL}.
- [34] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.
- [35] Maurice G. Hornocker. An analysis of mountain lion predation upon mule deer and elk in the idaho primitive area. *Wildlife Monographs*, 21:pp. 3–39, 1970.
- [36] G. L. Hunt. Populationemployment models: Stationarity, cointegration, and dynamic adjustment. *Journal of Regional Science*, 46(2):205–244, 2006.
- [37] C. Jackson, J.W. McNutt, and P. Apps. Managing the ranging behaviour of wild dogs using translocated scent marks. *Wildlife Research*, 39:31–34, 2012.
- [38] R Jackson and G Ahlborn. Snow leopards (*panthera uncia*) in nepal - home range and movements. *National Geographic Research*, 5(2):161–175, 1989.
- [39] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254, 1988.

- [40] Neil R. Jordan, Krystyna A. Golabek, Peter J. Apps, Geoffrey D. Gilfillan, and John W. McNutt. Scent-mark identification and scent-marking behaviour in african wild dogs (*lycaon pictus*). *Ethology*, 119(8):644–652, 2013.
- [41] N. L. Joseph. Cointegration, error-correction models, and forecasting using realigned foreign exchange rates. *Journal of Forecasting*, 14(6):499–522, 1995.
- [42] R. K. Kaufmann and D. I. Stern. Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research - Atmospheres*, 107(D2):4012+, January 2002.
- [43] A. Khan, S. Chatterjee, D. Bisai, and N. Barman. Analysis of change point in surface temperature time series using cumulative sum chart and bootstrapping for asansol weather observation station, west bengal, india. *American Journal of Climate Change*, 3:83–94, 2014.
- [44] R. Killick, I. A. Eckley, K. Ewans, and P. Jonathan. Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering*, 37(13):1120 – 1126, 2010.
- [45] G. Koop. Cointegration tests in present value relationships: A bayesian look at the bivariate properties of stock prices and dividends. *Journal of Econometrics*, 49(12):105 – 139, 1991.
- [46] G.M. Koop, R. W. Strachan, H. Van Dijk, and M. Villani. Bayesian approaches to cointegration. In *The Palgrave Handbook of Theoretical Econometrics*, pages 871–898. Palgrave Macmillan, 2006.
- [47] David Lusseau, Karsten Schneider, OliverJ. Boisseau, Patti Haase, Elisabeth Slooten, and SteveM. Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

- [48] D.W. Macdonald, F.G. Ball, and N.G. Hough. The evaluation of home range size and configuration using radio tracking data. In C.J. Amlaner and D.W. Macdonald, editors, *A handbook on biotelemetry and radio tracking: proceedings of an International Conference on Telemetry and Radio Tracking in Biology and Medicine*. Pergamon Press, 1980.
- [49] Michael G. L. Mills and Martyn L. Gorman. Factors affecting the density and distribution of wild dogs in the kruger national park. *Conservation Biology*, 11(6):pp. 1397–1406, 1997.
- [50] D.F. Morrison. *Multivariate statistical methods*. McGraw-Hill series in probability and statistics. McGraw-Hill, 1990.
- [51] P. K. Narayan and R. Smyth. What explains dissent on the high court of australia? an empirical assessment using a cointegration and error correction approach. *Journal of Empirical Legal Studies*, 4(2):401–425, 2007.
- [52] jr. (originator) N.N. Bogolyubov. *Encyclopedia of Mathematics*, chapter Perturbation Theory. Cambridge University Press, 2011.
- [53] Neal Oden. Partitioning dependence in nonstationary behavioral sequences. In B.A. Hazlett, editor, *Quantitative Methods in the Study of Animal Behavior*, pages 203 – 220. Academic Press, 1977.
- [54] E.D. Solomentsev (originator). *Encyclopedia of Mathematics*, chapter Cauchy Inequality. Cambridge University Press, 2012.
- [55] I.B. Vapnyarskii (originator). *Encyclopedia of Mathematics*, chapter Lagrange multipliers. Cambridge University Press, 2010.
- [56] P. C. B. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):pp. 335–346, 1988.
- [57] A. Reich. *The behaviour and ecology of the African wild dog, Lycaon pictus, in the Kruger National Park*. PhD thesis, Yale University, 1981.

- [58] B.E. Sorensen. *Econ266 Cointegration: Lecture Notes*. University of Houston, 2005.
<http://www.uh.edu/~bsorensen/coint.pdf>.
- [59] P.E. Stander, P.J. Haden, II. Kagece, and II. Ghau. The ecology of asociality in namibian leopards. *Journal of Zoology*, 242(2):343–364, 1997.
- [60] J. H. Stock and Mark W. Watson. Testing for common trends. *Journal of the American Statistical Association*, 83(404):pp. 1097–1107, 1988.
- [61] G.B. Thomas and R.L. Finney. *Calculus and Analytic Geometry*, chapter 8. The Definite Integral (Continued). Addison-Wesley world student series. Addison-Wesley, 1996.
- [62] G.B. Thomas, M.D. Weir, and J. Hass. *Thomas' Calculus*, chapter 16 Surface Integrals. Pearson Education Inc, twelfth edition, 2010.
- [63] M.D. Weir, G.B. Thomas, and J. Hass. *Thomas' Calculus*, chapter 5.3 The Definite Integral. Always Learning. Addison-Wesley, 2009.
- [64] Tina Wey, Daniel T. Blumstein, Weiwei Shen, and Ferenc Jordn. Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour*, 75(2):333 – 344, 2008.
- [65] Xiaoyun Zhu, Yun Huang, and John Doyle. Soft vs. hard bounds in probabilistic robustness analysis. In *Decision and Control, 1996., Proceedings of the 35th IEEE Conference on*, volume 3, pages 3412–3417 vol.3, Dec 1996.