Multivariate financial econometrics: with applications to volatility modelling, option pricing and asset allocation

Williams, Julian

Award date:
2007

Awarding institution:
University of Bath

MULTIVARIATE FINANCIAL ECONOMETRICS: WITH APPLICATIONS TO VOLATILITY MODELLING, OPTION PRICING AND ASSET ALLOCATION

By
Julian Williams

University of Bath, School of Management
A thesis submitted for the degree of Doctor of Philosophy
Submitted December 2006 [1]

[1]This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Table of Contents

List of Tables

List of Figures

Short Abstract

In the last two decades financial time series econometrics has developed from a sub-discipline to a key driver in mainstream econometric theory. Much of the work has been based around univariate and bivariate model specifications. High-variate extensions are often complex and difficult to estimate, even if model parameterization achievable then the results are often difficult to interpret and visualize. This work focuses on the design and implementation of linear and non-linear multivariate models, their system identification and model robustness. The work sets out several new models and demonstrates their usage in the financial context.

Acknowledgements

This thesis is for my parents and my supervisor Prof. Christos Ioannidis, both of whom have shown immeasurable patience and without whom I could not have even thought to begin.

I would also like to Acknowledge the help and support of the Accounting and finance section in the University of Bath. With particular thanks, in no particular order to Dr Andreas Krause, Dr Richard Fairchild, Dr Philip Cooper, Dr Andros Gregoriou and Prof. Ania Zalewska. Great people, in a challenging and productive work environment.

I would like to thank my friends Kashif, Prateek and Ram, who have been with me during the last two and a half years.

Long Abstract

The need for aggressive analytical methodologies in the modern financial econometricians toolbox has never been greater than today. Data is now omnipresent and massive in length and breadth, in fact modern data-sets are now so large it is almost impossible to visualize them in their entirety and draw any significant ad-hoc conclusions. In conjunction with this explosion in data is the criticality of decision making based on that data. The growth in complex investment instruments has meant that traditionally robust strategies are often sub-optimal in the current trading environment, this thesis sets out 8 chapters demonstrating a wide variety of tools that encompass many of the common problems that face financial practitioners and academics.

0.0.1 The Multivariate Approach

The multivariate approach has often been shunned as too complex and too difficult to interpret. Limited to fundamental probabilistic assumptions, multivariate analysis has been ostensively left to extensions of the general linear models/covariance in disturbances approach in the empirical literature. Whilst in macro-economics the vector

auto-regression methods in conjunction with theoretical structural models are now the primary modelling archetype, the situation in finance is more acute. Multivariate models have until recently been shunned by financial practitioners in favour of detailed univariate models. A particular example of this apparent lack of interest is illustrated by the extensive use of univariate GARCH models but the lack of interest in MV-GARCH models. Multivariate volatility models such as the MV-ARCH/GARCH specifications are obviously very useful in the analysis of volatility but suffer from the large number of parameters required to estimate them (and hence heavy data requirements), and a lack of financial decisions tools to interpret the information provided when the models are parameterized. Part of the objective of this thesis in conjunction with presenting several new model specifications is to focus on practical methods of identifying multivariate interactions and presenting useful approaches to system identification, model selection and decision making.

0.0.2 The Approach of this Thesis

This thesis is divided into five main chapters and a brief final summary. Each chapter has an introduction and review of the appropriate literature.

Synopsis of Chapter 1

The first chapter offers a brief review of vector and matrix multivariate time series analysis, in conjunction with notation and some fundamental mathematical preliminaries required to understand the later chapters. Whilst the majority of the chapter

is a review of the current literature, the reader is drawn to some new specifications for multivariate recursive Kalman filter methodologies and random matrix processes. The chapter also review some of the continuous time literature in preparation for chapter 8.

Synopsis of Chapter 2

Reviews the MV-GARCH literature and demonstrates the effective use of these models in the finance context, emphasis is made on parameter estimation and restriction testing and two brief empirical studies are undertaken to demonstrate the use of these models in the finance context.

Synopsis of Chapter 3

Addresses the use of copula type marginal/joint distribution frameworks in multivariate modelling of financial time series, the chapter concentrates on monte-carlo applications and methodologies. A new empirical copula model is suggested and demonstrated which offers a flexible means of estimating and generating monte-carlo pathways from multivariate data.

Synopsis of Chapter 4

Offers some empirical evidence for non-linear adjustments in co-movement between the excess returns of biotechnologies stocks in the UK, in particular the chapter focuses on the response to news events in the returns of listed bio-pharmaceutical

companies. The empirical results show that the adjustment in co-movement after a news event is related to the magnitude of the shock of the event and interdependency of the companies involved.

Synopsis of Chapter 5

Focuses on volatility models and in particular a new approach to modelling dynamic covariation, by use of switching mixture models. The main model presented is a dynamic switching model which offers real time parameter estimation, even for very high-variate systems. The model is shown to be tractable even for systems of 100 variables and is therefore very suitable for use in equity and interest rate analysis.

Synopsis of Chapter 6

Extends much of the material covered in the first section of chapter one, the chapter introduces the first new model, the integrated VAR(r)-MV-GARCH(p,q) model. This model utilizes an underlying vector auto-regression model and integrates it with an adapted BEKK model of conditional covariance. We demonstrate a robust maximum likelihood estimator and restriction tests, with monte-carlo analysis of the performance of this estimator. An identification strategy using impulse responses in mean, variance and covariance is discussed and implemented.

Synopsis of Chapter 7

This chapter departs from most of the previous work by assuming significant deviations from conditional normality in asset price returns. The chapter begins by presenting a model of asset returns with no predefined conditional or unconditional probability distribution. And then presents a toolkit for navigating through the non-multi-normality of both unconditional and conditional asset returns. The chapter uses multivariate-moment arrays to capture non-linearities in the dependency structure an demonstrates a simple asset allocation method.

Synopsis of Chapter 8

The final chapter presents a continuous time model of stochastic covariation and it's application for the appropriate pricing of call options, which includes the volatility risk of the asset in conjunction with the rest of the market. The model is essentially two vector stochastic differential equations, whereby the volatility dynamics are modelled as a separate vector of Brownian motions. The chapter adds to the growing literature by first specifying a maximum likelihood approach to parameterizing such models in discrete time and second giving a systematic approach to hedging in the presence of stochastic covariation.

Chapter 1

Vector and Matrix Multivariate Stochastic Processes

1.1 Chapter Abstract

This chapter reviews the current literature and common results on vector and matrix dynamic processes and general discrete and continuous multivariate time series analysis, it begins with a review of notation, then proceeds to define the current state of the art in vector time series, the chapter concentrates on discrete time series, but the final section alludes to the useful contribution of multivariate continuous time models in financial engineering and multivariate rational expectations models. This chapter represents the bulk of the literature review and other chapters will feature only brief recaps of the relevant material.

1.2 Introduction and Focus

The current literature on multivariate time series is as vast as it is eclectic, this review does not intend to be over-arching in terms of coverage, but restricts itself to the main results and important findings which will be used in later chapters. To this extent this chapter bypasses much of the main literature on vector auto-regressions (VAR) models and keys in on the main results, in terms of identification, specification and robustness. This includes an extensive review of the Kalman filter approach and the state space vector as an unobserved variable, as this is useful in understanding the structure and techniques used in the bounded dynamic covariance model in chapter four. Whilst not strictly in theme, the review of univariate ARCH/GARCH models features some of the asymptotic theory developed in the last decade and a half on these models. Whilst the author recognizes that the variety of the univariate specification is vast, focusing on the properties of the basic model specification, it helps serves as a basis for understanding some of the effects of the difference in specifications of their multivariate extensions. To this end the review of multivariate ARCH/GARCH models includes the main specifications and a review of up and coming innovations in this area. Interested readers are drawn to the extensive section on modelling disturbances via copulas, whilst the introduction to this section is a review of the current literature, the section also introduces two new estimation methods, the first in regards to empirical copulas uses Latin hypercube stratification to construct

n-variate empirical copulas, the second section proposes a new method of constructing MV-ARCH type models with time invariant and time varying copulas, the section also suggests two empirical applications in analyzing derivatives and credit risk via copulas. The section on continuous time models restricts itself to a brief chronology of key literature in this area and the key results required when reading chapter 5, continuous time models are of prevalent importance in financial engineering. The brevity of this section does not do justice to their true importance, however an eclectic review is not in the scope of this work. In this respect the same accusation can be made at the review of estimation and optimization techniques, however lack of inclusion of large swathes of this fascinating literature is justified as our general focus is on those techniques directly applicable to time series modelling in finance.

This review has two major themes, first the specification and identification of the underlying deterministic model and second the probabilistic properties of the stochastic processes that act as disturbances on this underlying model. The first section reviews specific notation and several important mathematical underpinnings required to understand the following section, including rewriting the lag operator as a vectorized tensor, with singleton time dimension, a very brief review of the continuous functional mapping form of the multivariate central limit theorem is also presented here for completeness. The second broad section introduces multivariate general linear

models and presents a complete derivation of the vector ARMAX model, in our specific notation. The Kalman model is also reviewed in preparation for the basic model presented in chapter four, which utilizes an unobserved state space operator to describe conditional covariance. The Kalman filter section also introduces the concept of incorporating a decomposition of covariance in the gaussian objective function. The final section briefly surveys the continuous time literature and provides several derivations in preparation for chapter 8.

1.2.1 Notation and Common Results

The following basic mathematical notation nomenclature is used throughout this chapter,

{}	braces	specifies the elements of a set
[]	square brackets	concatenation of scalars, vectors and matrics
()	brackets	function ordering
$\|\mathbf{A}\|$	vertical bars (matrix)	determinant of matrix
$\|\mathbf{a}\|$	vertical bars (vector)	norm (normative distance from the center of the Euclidian ball)
$\|a\|$	vertical bars (scalar)	absolute
$< . >$	angle brackets	cross variation of two functions
$a_{i,j,t}$	lowercase with subscript	single index represents vertical position in column vector two indices represents the coordinates in a matrix n - indices represents the coordinates in an n dimensional array (tensor)
$\mathbf{a}$	bold lowercase	column vector
A_t^i	uppercase italics	continuous process, subscript represents continuous time superscript process index (i^{th} process out of k)
$\mathbf{A}$	uppercase bold	Matrix
$\mathbf{\Sigma}$	Greek Uppercase Bold	Coefficients matrix/non - negative Hermitian Matrix
∇	Nabla operator	Gradient vector of partial derivatives of a vector function
$\mathbf{e}$	bold e	vector of ones of appropriate length
$\mathbf{0}$	bold 0	vector/matrix of zeros of appropriate size
$\mathbb{R}$	hollow capital	set
$\mathcal{R}$	Caligraphic capital	array/tensor
$\mathcal{Y}_{t\|r}$	Caligraphic capital	array of the history of a vector or matrix process from $t-r$ to $t-1$
$\times$	Hadamard Product	Element by element multiplication of equivalent sized arrays (deviation from standard)
$\odot$	Khatri - Rao Product	Element by element Kronecker Product of arrays with one equal dimension
$\otimes$	Kronecker Product	See Kronecker Product
θ	bold lowercase theta	vector of parameters
$\mathbf{\Theta}_{\|g(\theta)}$	bold uppercase theta	set of available valid parameters conditioned on the constraint g (θ)
θ^*	theta star	true parameter vector
$\hat{\theta}$	theta hat	estimated parameter vector

θ	bold lowercase theta	vector of parameters
$\Theta_{\|g(\theta)}$	bold uppercase theta	$\left\{\begin{array}{l}\text{set of available valid parameters}\\ \text{conditioned on the constraint g}(\theta)\end{array}\right\}$
θ^*	theta star	true parameter vector
$\hat{\theta}$	theta hat	estimated parameter vector
$\min_{\theta}[f(\theta \mid g(\theta))]$	min	the constrained minimisation of the objective function f
$\mathfrak{L}(\theta)$	Fraktur capital L	the evaluation of the objective function $\Im$
T	upper case T	Transpose operator
$\mathbf{w}^{[1,..,m]}$	vector power	$\left\{\begin{array}{l}\text{the } m \text{ - way vector multiplication}\\ \text{of the column vector } \mathbf{w},\end{array}\right\}$ $\left\{\begin{array}{l}\text{results in a } m-\text{dimensional}\\ \text{array with } m^{\text{th}} \text{ order supersymmetry}\end{array}\right\}$
$\mathbf{H}^{(j)}$	uppercase H, with bracketed superscript	$\left\{\begin{array}{l}j^{\text{th}} \text{ evaluation of the Hessian}\\ \text{matrix of the function } \Im(\theta)\end{array}\right\}$
$\mathfrak{F}(x)$	Fraktur Capital F	Probability density function
$\mathfrak{H}(x)$	Fraktur Capital H	Cumulative density function
$\mathfrak{C}(x)$	Fraktur Capital C	Characteristic function of $\mathfrak{F}(x)$
$\mathsf{H}(x)$	Sans Serif H	Generic H-function
$\mathsf{L}(x)$	Sans Serif L	Laplace Transform
$\mathsf{M}(x)$	Sans Serif M	Mellin Transform
$\mathsf{F}(x)$	Sans Serif F	Fourier Transform

1.2.2 Introduction and Commentary

In their major text on stochastic processes Karatzas and Shreve (1991) [162] note that there are three main areas of study in discrete and continuous time stochastic processes for which there is a rigorous mathematical foundation, Brownian motion, Martingales and Stationarity. This chapter will look at all three of these concepts generally in reverse order, however it should be noted that because of their interwoven

nature much of the chapter will refer in both a forward and backward manner. For example when understanding non-stationarity, an understanding of Weiner processes and central limit theorems which is inextricably linked to Brownian motion. Markov processes are briefly discussed in discrete time and these are strongly linked to the Martingale measure, a fundamental principle of finance in relation to portfolio replication and arbitrage free pricing. As such the reader is suggested to keep the table of contents in mind whilst reading and use this to navigate these inter-relationships.

A major theme of this chapter is notation and the importance of understanding how a little knowledge of consistent notation can be a very powerful tool in quickly understanding various concepts and approaches. For example the extensive use of the vectorization *vec* and transpose $^{\mathrm{T}}$ operators in conjunction with n-dimensional structured arrays, (classical tensors). This is very helpful in that model specifications are no longer wedded to lag operators and deal simply in vectorized notational form, which is imperative in multivariate analysis in order to reduce functions to simple matrix, vector equations. As such stationarity conditions may be captured in a very concise and simple manner.

1.2.3 The Basic Multivariate Time Series Object

The basic multivariate time series object is an n length random vector process $\mathbf{y}_t$, by measuring time in discrete increments, $t \in \{1, 2, ..., T\}$, then the current realization

of $\mathbf{y}_t \in \mathbb{R}^n$ is,

$$\mathbf{y}_t = [y_{1,t}, \cdots, y_{n,t}]^{\mathrm{T}} \tag{1.2.1}$$

now consider the mapping of $\hat{\mathbf{y}}_{t|\Omega_{t|\{t-r,\ldots,t-1\}}}$ where $\Omega_{t|\{t-r,\ldots,t-1\}}$ is the set that contains all the necessary information to form an unbiased estimate of $\mathbf{y}_t$, $\hat{\mathbf{y}}_t$, the expected first moment deviation and second moment quadratic deviation will be,

$$E\left(\mathbf{y}_t - \hat{\mathbf{y}}_t \left|\Omega_{t|\{t-r,\ldots,t-1\}}\right.\right) = \mathbf{0} \tag{1.2.2}$$

$$E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_t \left|\Omega_{t|\{t-r,\ldots,t-1\}}\right.\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_t \left|\Omega_{t|\{t-r,\ldots,t-1\}}\right.\right)^{\mathrm{T}}\right) = \mathbf{\Sigma} \tag{1.2.3}$$

Where E_t is the expectations operator. If the process is in continuous time then the elements of $\mathbf{y}_t$, are now continuous time processes Y_t^i and as such no longer have specific co-ordinates therefore,

$$\mathbf{y}_t = \begin{bmatrix} Y_t^1 \\ Y_t^2 \\ \vdots \\ Y_t^n \end{bmatrix} \tag{1.2.4}$$

$$Y_t^n = f(\mathbf{y}_t, t) \tag{1.2.5}$$

The index is shifted to a superscript to denote the process to show that there is no discrete coordinate for any realization of $\mathbf{y}_t$.

1.2.4 Commonly used Matrix Operators

The basic multivariate time series object is a vector process, as such the major mechanism for manipulation is via matrix algebra and matrix differential calculus. Consider

a rectangular $m \times n$ matrix $\mathbf{A}$, the following operations are valid for any sized matrix,

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \tag{1.2.6}$$

$$\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{m,n} \end{pmatrix} \tag{1.2.7}$$

$$vec(\mathbf{A}) = \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \\ a_{1,2} \\ \vdots \\ a_{n,2} \\ \vdots \\ a_{m,n} \end{bmatrix} \tag{1.2.8}$$

For a multidimensional array, the *vec* operator reshapes the array into a column vector, starting with the first integer dimension and slicing the array dimension by dimension stacking each slice column-wise. Now consider the $n \times n$ non-negative Hermitian Matrix $\mathbf{\Sigma}$, i.e. $\mathbf{\Sigma} \in \mathbb{C}^{n \times n}$ the following operations are valid,

$$\mathbf{\Sigma} = [\sigma_{i,j}] \tag{1.2.9}$$

$$\mathbf{\Sigma}^{\frac{1}{2}} = chol(\mathbf{\Sigma}) \tag{1.2.10}$$

$$\mathbf{\Lambda} = \mathbf{\Sigma}^{\frac{1}{2}} \tag{1.2.11}$$

$$\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}} = \mathbf{\Sigma} \tag{1.2.12}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,n} \\ 0 & \lambda_{2,2} & \cdots & \lambda_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{n,n} \end{bmatrix} \tag{1.2.13}$$

For any square diagonal matrix the unique elements may be extracted column-wise by use of the *vech* operator, a lower (or upper) triangular matrix may also be formed, from an appropriate length vector. For any given square diagonal matrix, of size n, the corresponding vector of unique elements will be $\frac{1}{2}n(n+1)$ in length.

$$vech(\mathbf{A}) = \mathbf{a} \tag{1.2.14}$$

$$\mathbf{a} = \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{1,n} \\ a_{2,2} \\ \vdots \\ a_{2,n} \\ \vdots \\ a_{n-1,n-1} \\ a_{n,n-1} \\ a_{n,n} \end{bmatrix} \tag{1.2.15}$$

$$ivech(\mathbf{a}) = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 & 0 \\ a_{2,1} & a_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & 0 \\ a_{n} & a_{n} & \cdots & a_{n,n-1} & a_{n,n} \end{bmatrix} \tag{1.2.16}$$

1.2.5 Matricizing

Matricizing has two major purposes, first to convert higher dimensional arrays into two dimensional matrices and second to reshape vectors into appropriately sized matrices. Consider the higher dimensional array, $\mathcal{A}$, which is an $n \times m \times p$ array, there

are many different ways of matricizing $\mathcal{A}$, for example,

$$\underset{n\times m\times p\to nm\times p}{mat}\left(\underset{n\times m\times p}{\mathcal{A}}\right) = \underset{nm\times p}{\mathbf{A}} \tag{1.2.17}$$

$$\underset{n\times m\times p\to n\times mp}{mat}\left(\underset{n\times m\times p}{\mathcal{A}}\right) = \underset{n\times mp}{\mathbf{A}} \tag{1.2.18}$$

For a vector matricizing requires consistency in the dimensions of the vector and matrix respectively for example,

$$\underset{nm\times 1\to n\times m}{mat}\left(\underset{nm\times 1}{\mathbf{a}}\right) = \underset{n\times m}{\mathbf{A}} \tag{1.2.19}$$

Generating Multi-normals via Cholesky factorization

Cholesky factorization is an extremely useful tool, as the following relationship between a random vector of i.i.d. standard normals and a random vector of zero centred multi-normals, exists,

$$\varepsilon \sim N(\mathbf{0},\mathbf{I}) \tag{1.2.20}$$

$$\mathbf{I} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \tag{1.2.21}$$

$$\mathbf{u} = \boldsymbol{\Sigma}^{\frac{1}{2}}\varepsilon \tag{1.2.22}$$

$$\mathbf{u} \sim N(\mathbf{0},\boldsymbol{\Sigma}) \tag{1.2.23}$$

In order to generate a valid random covariance matrix, i.e. a non-negative Hermitian, which may be factorized via the Cholesky method, then this may be generated from

a random vector using the *ivech* transformation.

$$\underset{\frac{1}{2}n(n+1)\times 1}{\mathbf{a}} \sim N\left(\underset{\frac{1}{2}n(n+1)\times 1}{\mu}, \underset{\frac{1}{2}n(n+1)\times\frac{1}{2}n(n+1)}{\mathbf{\Omega}}\right) \tag{1.2.24}$$

$$\underset{n\times n}{\mathbf{A}} = ivech\left(\underset{\frac{1}{2}n(n+1)\times 1}{\mathbf{a}}\right) \tag{1.2.25}$$

$$\underset{n\times n}{\mathbf{\Sigma}} = \underset{n\times n}{\mathbf{A}}\ \underset{n\times n}{\mathbf{A}}^{\mathrm{T}} \tag{1.2.26}$$

The lower diagonal matrix, when multiplied by its transpose yields a non-negative matrix which is in essence a random covariance matrix. Please note the following, in general, if $\mathbf{\Lambda}$ is the Cholesky factorization of a matrix $\mathbf{\Sigma}$ and $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$, where $\mathbf{A}$ is a lower diagonal matrix, then under most circumstances $\mathbf{\Lambda} = \mathbf{A}^{\mathrm{T}}$, does NOT hold.

1.2.6 A Multivariate Central Limit Theorem

Multivariate central limit theorems are the subject of extensive and ongoing academic work, the following is a summary of results from work by Davidson (1994) [66] and articles by van Zanten (1998) [257], Karatzas and Shreve (1991) and Kuchler and Sorenson (1996) [174]. This section provides some key results and propositions, the interested reader is directed to Davidson (1994) and Karatzas and Shreve (1991,2000) for more detailed descriptions. Consider a random vector $\mathbf{y}_t$, where $\mathbf{y}_t \in \mathbb{R}^n$, the sequence of $\mathbf{y}_t$, $\{\mathbf{y}_1, ..., \mathbf{y}_\tau\}$, is said to obey a multivariate central limit theorem if the joint distribution of, $\mathbf{s}_\tau = \sum_{t=1}^{\tau} \mathbf{y}_t$ converges to the multivariate Gaussian distribution. A weighted summation of $\mathbf{y}_t$ obtains a multivariate central limit theorem by showing that for a process $\mathbf{y}_t$, the scalar sequence $\alpha^{\mathrm{T}}\mathbf{y}_t$, obeys an ordinary scalar

central limit for any arbitrary vector α. This means that a multivariate central limit convergence maybe framed in terms of the Cramer-Wold device, see Davidson (1994) and Kuchler and Sorenson (1996), and furthermore define the conditions under which random variables converge asymptotically to the multivariate central limit theorem.

The Cramer-Wold Device

For some arbitrary n length fixed column vector α, where $\alpha \neq \mathbf{0}$ the real valued process $\mathbf{s}_\tau \in \mathbb{R}^n$ converges in distribution to a random vector $\mathbf{s}$, iff $\alpha^{\mathrm{T}}\mathbf{s}_\tau \xrightarrow{D} \alpha^{\mathrm{T}}\mathbf{s}$, where $\xrightarrow{D}$ is the convergence in distribution operator. Now consider the $n \times \tau$ matrix $\mathbf{Y}$ where $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_\tau]^{\mathrm{T}}$, the covariance matrix $\mathbf{\Sigma} = \frac{1}{\tau}\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$, is the sample covariance matrix of the sequence of n-length vectors constituting $\mathbf{Y}$. For the cumulative summation sequence $\mathbf{S}_\tau$, where $\mathbf{S} = \left[\sum_{i=1}^{1}\mathbf{y}_i, \sum_{i=1}^{2}\mathbf{y}_i, ..., \sum_{i=1}^{\tau}\mathbf{y}_i\right]^{\mathrm{T}}$, if $\mathbf{\Sigma}_{\mathbf{s},\tau} = \frac{1}{\tau}\mathbf{S}^{\mathrm{T}}\mathbf{S}$, by construction of this matrix is positive semi-definite (non negative definite) and as such $\mathbf{\Sigma}_{\mathbf{s},\tau}$, maybe factorized as follows, $\mathbf{\Sigma}_{\mathbf{s},\tau} = \mathbf{C}_{\mathbf{s},\tau}\mathbf{\Lambda}_{\mathbf{s},\tau}\mathbf{C}_{\mathbf{s},\tau}^{\mathrm{T}} = \mathbf{A}_{\mathbf{s},\tau}\mathbf{A}_{\mathbf{s},\tau}^{\mathrm{T}}$, where $\mathbf{A}_{\mathbf{s},\tau} = \mathbf{C}_{\mathbf{s},\tau}\mathbf{\Lambda}_{\mathbf{s},\tau}^{\frac{1}{2}}$ and $\mathbf{C}_{\mathbf{s},\tau}\mathbf{C}_{\mathbf{s},\tau}^{\mathrm{T}} = \mathbf{C}_{\mathbf{s},\tau}^{\mathrm{T}}\mathbf{C}_{\mathbf{s},\tau} = \mathbf{I}$. The matrices $\mathbf{C}_{\mathbf{s},\tau}$ and $\mathbf{\Lambda}_{\mathbf{s},\tau}$ are respectively the eigenvector matrix and the diagonal, non-negative matrix of eigenvalues.

Multivariate CLT as a Weiner Process

The relationship between a zero centred multi-normal distribution and a zero centred vector of i.i.d. standard normals is key to defining a multivariate central limit theorem as a continuous time process. Consider a continuous time vector process W_t^i, from an n-length continuous vector process $W_t^i \in \mathbf{w}_t$, assuming that every innovation of

$\mathbf{w}_t$, is drawn from a zero centred multi-normal distribution then, $\mathbf{w}_t$, is an n-length continuous time vector process,

$$\mathbf{w}_t = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{w}_t^{\sigma} \tag{1.2.27}$$

$$\mathbf{w}_t^{\sigma} \sim N(\mathbf{0}, \mathbf{I}) \tag{1.2.28}$$

For any functional $\mathfrak{F}$ that has a continuous map onto the set of real numbers $\mathbb{R}^n$, then, $\mathfrak{F} : \mathbb{R}^n \to \mathbb{R}^n$, for some continuous mapping $\mathcal{C}$, whereby an n-length vector of random variables $\mathbf{x}$ is mapped onto the real space $\mathbb{R}^n$, then the following holds

$$\mathbf{w}_t = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{w}_t^{\sigma} \tag{1.2.29}$$

$$\mathbf{w}_t^{\sigma} \sim N(\mathbf{0}, \mathbf{I}) \tag{1.2.30}$$

$$\mathbf{x}_t = \begin{bmatrix} f^1\left(W_t^{1,\mathbf{s}}\right) \\ f^2\left(W_t^{2,\mathbf{s}}\right) \\ \vdots \\ f^1\left(W_t^{n,\mathbf{s}}\right) \end{bmatrix} \tag{1.2.31}$$

$$\sqrt{T}.\mathbf{x}_t \to \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{w}_t \tag{1.2.32}$$

Which implies that for a generic stochastic process $\mathbf{x}$ there is a mapping which will yield a process, that will converge to some probability measure $\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{w}_t$. for a sufficiently large T. The local martingale measure is discuss later in this chapter.

Some Common Matrix Decompositions

Other useful matrix decompositions commonly used in chapter six are, Eigenvector decomposition, the Jordan canonical form, the Schur decomposition and the Moore-Penrose Inverse these are defined as follows,

The Eigenvector problem, offers solutions to matrix polynomial equations of non-negative matrices functions and is of the following form,

$$\mathbf{Ax} = \lambda \mathbf{x} \tag{1.2.33}$$

The Jordan decomposition,

$$\mathbf{A} \in \mathbb{C}^{(n \times n)} \tag{1.2.34}$$

$$\mathbf{J} = \mathbf{X}^{-1}\mathbf{AX} \tag{1.2.35}$$

$$\mathbf{J} = \begin{bmatrix} j_{1,1} & 0 & \cdots & 0 \\ 0 & j_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & j_{n,n} \end{bmatrix} \tag{1.2.36}$$

The Schur decomposition

$$\mathbf{A} \in \mathbb{C}^{(n \times n)} \tag{1.2.37}$$

$$\mathbf{S}^*\mathbf{AS} = \mathbf{M} \tag{1.2.38}$$

$$\mathbf{S}^*\mathbf{S} = \mathbf{I} \tag{1.2.39}$$

The Least Squares Problem, With Rank deficient Matrices and the Pseudo an Moore Penrose Inverses

Consider the following problem, $\mathbf{Ax} = \mathbf{b}$, the solution to this problem, $\mathbf{x}_{LS}$, is the full rank least squares problem. In order to tackle this problem two definitions are required first the orthogonalization conditions for a matrix and second the rank of the matrix, for more context on these definitions see references Golub and Van Loan (1996) [112] and Horn and Johnson (1999) [146] for the specialist linear algebra applications

and Magnus and Neudecker (1998) [191] and Davidson (1994) for insight into the application to matrix differential calculus and stochastic limit theory respectively. The results used here are extensively utilized in 5.

Definition: Orthogonal Matrix

A matrix $\mathbf{Q} \in \mathbb{C}^{n \times n}$ is said to be orthogonal iff $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}$ In general the full rank least squares problem is satisfied when $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is an upper triangular matrix.

Matrix Products

In general there are four commonly used matrix products, the basic matrix product, the Hadamard product, the Kronecker Product and the Khatri-Rao(Bro) product, these are defined as follows, defining,

$$\mathbf{A} \equiv [a_{i,j}] \tag{1.2.40}$$

$$\mathbf{B} \equiv [b_{i,j}] \tag{1.2.41}$$

$$\mathbf{C} \equiv [c_{i,j}] \tag{1.2.42}$$

$$\mathbf{a}_i = [a_{i,1}, ..., a_{i,m}]^{\mathrm{T}} \equiv \underset{m \times n}{\mathbf{A}} = [\mathbf{a}_1, ..., \mathbf{a}_n] \tag{1.2.43}$$

defining the various products as,

$\underset{m \times n}{\mathbf{A}} \underset{n \times p}{\mathbf{B}} = \underset{n \times m}{\mathbf{C}}$	matrix product	$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{j,k}$
$\underset{m \times n}{\mathbf{A}} \times \underset{m \times n}{\mathbf{B}} = \underset{m \times n}{\mathbf{C}}$	Hadamard product	$c_{i,j} = a_{i,j} b_{i,j}$
$\underset{m \times n}{\mathbf{A}} \otimes \underset{p \times q}{\mathbf{B}} = \underset{mp \times nq}{\mathbf{C}}$	Kronecker product	$c_{i,j} = a_{i,j}\mathbf{B}$
$\underset{m \times n}{\mathbf{A}} \odot \underset{p \times n}{\mathbf{B}} = \mathbf{C}$	Khatri - Rao - Bro product	$\mathbf{C} = [\mathbf{a}_1 \otimes \mathbf{b}_1, ..., \mathbf{a}_n \otimes \mathbf{b}_n]$

In general only the first three matrix/vector products are commonly used, the Khatri-Rao product is very useful in defining the contravariant indexing structure of high order tensors, when systematic indexing is difficult (i.e. when the tensor is greater than rank 4), this in a programming sense allows element operations in tensor problems to be vectorized, for example for an array, $\mathcal{A}$, with 3 contravariant dimensions, $d = 3$ each, $n = 3$ elements in length, then there are a total of 27 elements to index, for element by element operations on this array systematic indexation requires a 3 indexing dimensions by 27 elements matrix,

$$\mathcal{A} = [a_{i,j,k}] \tag{1.2.44}$$

$$\mathbf{a} = vec(\mathcal{A}) \tag{1.2.45}$$

$$\mathbf{a} = [a_{1,1,1}, a_{1,1,2}, ..., a_{\mathbf{s}_l}, ..., a_{3,3,3}]^{\mathrm{T}} \tag{1.2.46}$$

Converting the subscripts into rows of an $n^d \times d$ matrix,

$$\mathbf{s}_l = [i, j, k] \tag{1.2.47}$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_{n^d} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_{27} \end{bmatrix} \tag{1.2.48}$$

$$\mathbf{n} = [1, 2, ..., n]^{\mathrm{T}} \equiv [1, 2, 3]^{\mathrm{T}} \tag{1.2.49}$$

$$\mathbf{e} = [1, 1, 1]^{\mathrm{T}} \tag{1.2.50}$$

Then for a $d = 3$, n length cubic array the elements are indexed by a matrix, $\mathbf{S}$, in terms of an n length unit column vector $\mathbf{e}$ and an n-length vector of natural integers,

$\mathbf{n} = [1, ..., n]^{\mathrm{T}}$ therefore,

$$\mathbf{S} = [(\mathbf{e} \otimes \mathbf{e}) \otimes \mathbf{n}, \mathbf{e} \otimes (\mathbf{n} \otimes \mathbf{e}), \mathbf{n} \otimes (\mathbf{e} \otimes \mathbf{e})] \tag{1.2.51}$$

More generally for a hyper-cubic with d dimensions each n in length then the indexing matrix is decomposed into the left and right terms of the Khatri-Rao product, the first being,

$$\mathbf{E} = [\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes ... \otimes \mathbf{e}_{d-1}, \mathbf{e}_1 \otimes \mathbf{n} \otimes ... \otimes \mathbf{e}_{d-1}, ..., \mathbf{e}_1 \otimes \mathbf{e}_2 \otimes ... \otimes \mathbf{e}_{d-2} \otimes \mathbf{n}] \tag{1.2.52}$$

the second is therefore,

$$\mathbf{N} = [\mathbf{e}_1, ..., \mathbf{e}_{d-1}, \mathbf{n}] \tag{1.2.53}$$

finally the indexing matrix, $\mathbf{S}$ is the product,

$$\mathbf{S} = \mathbf{E} \odot \mathbf{N} \tag{1.2.54}$$

This algorithm is very simple to program and provides a mechanism for indexing d dimensional arrays and ensures that the process of applying functions to d dimensional arrays a simple and tractable task. See Anderson and Bro (2000) [9] and Moravitz-Martin (2004) [200], for a more extensive review of tensor and matrix operations.

1.3 Covariance Stationary Vector ARMA processes

It is possible to impose a wide variety of data generating functions on the conditional form of $\mathbf{y}_t$, most common functional forms utilize a linear or non-linear autoregressive

framework, usually derived from the Euler scheme of some continuous time underlying process.

1.3.1 Basic Model Notation

An n length auto-regressive process with p lags AR(p) maybe written out as follows,

$$\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.1}$$

$$\mathcal{Y}_t = \left[\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}\right] \tag{1.3.2}$$

$$\mathbf{u}_t = \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t \tag{1.3.3}$$

$$\varepsilon_t \sim N\left(\mathbf{0}, \mathbf{I}\right) \tag{1.3.4}$$

$$\mathbf{\Sigma} = E\left(\mathbf{u}_t \mathbf{u}_t^{\mathrm{T}}\right) \tag{1.3.5}$$

Where $\mathbf{\Pi}_0$ is an $np \times n$ matrix, of coefficients and $\mathbf{c}$ is an n length vector of constants.

A vector moving average process MA(q), is similarly defined as follows,

$$\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{U}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.6}$$

$$\mathcal{U}_t = \left[\mathbf{u}_{t-1}, ..., \mathbf{u}_{t-q}\right] \tag{1.3.7}$$

$$\mathbf{u}_t = \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t \tag{1.3.8}$$

$$\varepsilon_t \sim N\left(\mathbf{0}, \mathbf{I}\right) \tag{1.3.9}$$

$$\mathbf{\Sigma} = E\left(\mathbf{u}_t \mathbf{u}_t^{\mathrm{T}}\right) \tag{1.3.10}$$

Putting these together then gives a V-ARMA(p,q) model as follows,

$$\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{\Pi}_1^{\mathrm{T}} vec\left(\mathcal{U}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.11}$$

Incorporating an exogenous component gives the V-ARMAX(p,q)

$$\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec(\mathcal{Y}_t) + \mathbf{\Pi}_1^{\mathrm{T}} vec(\mathcal{U}_t) + \mathbf{\Pi}_2^{\mathrm{T}} \mathbf{x}_t + \mathbf{u}_t \tag{1.3.12}$$

$$\mathbf{x}_t = \left[x_{1,t}, x_{2,t}, ..., x_{m,t}, \mathbf{e}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.13}$$

1.3.2 Specifying the Objective Function

Assuming that the disturbances $\mathbf{u}_t$ have a multi-normal distribution then the objective log-likelihood function is specified as follows,

$$\theta = \left[vec(\mathbf{\Pi}_0)^{\mathrm{T}}, vec(\mathbf{\Pi}_1)^{\mathrm{T}}, vec(\mathbf{\Pi}_2)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.14}$$

$$\mathfrak{F}(\theta, \mathbf{\Sigma}) = \frac{1}{\sqrt{2\pi^n |\mathbf{\Sigma}|}} \exp\left(-\tfrac{1}{2}(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t)\right) \tag{1.3.15}$$

Taking the logarithm of the multi-normal pdf yields,

$$\log \mathfrak{F}(\theta, \mathbf{\Sigma}) = -\tfrac{1}{2} n \log(2\pi) - \tfrac{1}{2} \log(|\mathbf{\Sigma}|) - \tfrac{1}{2} (\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t) \tag{1.3.16}$$

For a sample set of τ observations, $t \in [1, ..., \tau]$ the objective function is therefore,

$$\mathfrak{L}(\theta) = -\tfrac{1}{2} n\tau \log(2\pi) - \tfrac{1}{2} \tau \log(|\mathbf{\Sigma}|) - \tfrac{1}{2} \sum_{t=1}^{\tau} (\mathbf{y}_t - \hat{\mathbf{y}}_t | \theta)^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_t | \theta) \tag{1.3.17}$$

Stationarity Conditions for the AR(p) Process

A series is said to be stationary if the consequences of any given event ε eventually die out. For a univariate AR(p) process,

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... \varphi_p y_{t-p} + \varphi_c e \tag{1.3.18}$$

The roots of the polynomial,

$$0 = \varphi_1 z + \varphi_2 z^2 + ... + \varphi_p z^p + \varphi_c e \tag{1.3.19}$$

Must lie within the unit circle. Rewriting the vector AR(p) process,

$$\mathbf{y}_t = \underset{np\times n}{\mathbf{\Pi}^{\mathrm{T}}} vec\left(\mathcal{Y}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.20}$$

as the first step forward recursion of $\mathbf{y}_t$, then,

$$\xi_{t+1} = \mathbf{\Phi}\xi_t + \mathbf{v}_t \tag{1.3.21}$$

and,

$$\underset{np\times 1}{\xi_{t+1}} = vec\left(\mathcal{Y}_{t+1}\right) - \mu \tag{1.3.22}$$

$$\mathcal{Y}_{t+1} = \left[\mathbf{y}_t, \mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-(p+1)}\right] \tag{1.3.23}$$

$$\mu = \mathbf{c} + \mathbf{\Pi}^{\mathrm{T}} \mu \underset{np\times 1}{\mathbf{e}^{\mathrm{T}}} \tag{1.3.24}$$

$$\underset{np\times np}{\mathbf{\Phi}} = \begin{bmatrix} \underset{np\times n}{\mathbf{\Pi}^{\mathrm{T}}} \\ \underset{n(p-1)\times n(p-1)}{\mathbf{I}} \quad \underset{n(p-1)\times n}{\mathbf{0}} \end{bmatrix} \tag{1.3.25}$$

$$\underset{np\times 1}{\mathbf{v}_t} = \begin{bmatrix} \mathbf{u}_t \\ \underset{n(p-1)\times n}{\mathbf{0}} \end{bmatrix} \tag{1.3.26}$$

For an s-step ahead recursion,

$$\xi_{t+s} = \mathbf{v}_{t+s} + \mathbf{\Phi}\mathbf{v}_{t+(s-1)} + \mathbf{\Phi}^2\mathbf{v}_{t+(s-2)} + ... + \mathbf{\Phi}^{s-1}\mathbf{v}_{t+1} + \mathbf{\Phi}^s\xi_t \tag{1.3.27}$$

For stationarity the effects of any given innovation of $\mathbf{u}_t$ must die out after a finite number of steps. The common result is that for the AR(p) generating process of $\mathbf{y}_t$,

the matrix $\mathbf{\Phi}$, must have all its eigenvalues within the unit circle. For an extended proof with slightly different notation see Fuller (1976) [101] or Hamilton (1994) [135].

Stationarity Conditions for a MA(q) Process

The stationarity conditions of a moving average process actually turn out to be much simpler than the AR(p) counter part, in fact any MA(q) process is in fact covariance stationary. Consider the r^{th} auto-covariance matrix $\mathbf{\Omega}$,

$$\mathbf{\Omega}_r = E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_t\right)\left(\mathbf{y}_{t-r} - \hat{\mathbf{y}}_{t-r}\right)^{\mathrm{T}}\right) \tag{1.3.28}$$

$$\equiv \mathbf{u}_t \mathbf{u}_{t-r}^{\mathrm{T}} \tag{1.3.29}$$

In a departure from the normal notation in the literature please note that, $\mathbf{\Gamma}_{i,t}$ to denote the r^{th} rolling covariance matrix of the i^{th} process.

$$\mathbf{\Gamma}_{i,t} = vec\left(\mathcal{U}_{i,t}\right) vec\left(\mathcal{U}_{i,t}\right)^{\mathrm{T}} \tag{1.3.30}$$

$$\mathcal{U}_{i,t} = \left[u_{i,t-1}, ..., u_{i,t-r}\right] \tag{1.3.31}$$

$$\mathbf{\Gamma}_t = \left[\mathbf{\Gamma}_{1,t}, ..., \mathbf{\Gamma}_{n,t}\right] \tag{1.3.32}$$

If $\mathbf{\Sigma} = E\left(\mathbf{y}_t - \mathbf{c}\right)\left(\mathbf{y}_t - \mathbf{c}\right)^{\mathrm{T}}$ is the variance/covariance matrix of $\mathbf{u}_t$, then formulating $\mathbf{\Omega}_r$ in terms of the coefficients matrix $\mathbf{\Pi}$ and the covariance matrix,

$$\mathbf{y}_t = \underset{nq \times n}{\mathbf{\Pi}^{\mathrm{T}}} vec\left(\mathcal{U}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.33}$$

$$\mathbf{\Omega}_q = \mathbf{\Pi}^{\mathrm{T}} \mathrm{Z}_q \tag{1.3.34}$$

$$\underset{nq \times n}{\mathrm{Z}_q} = \left[\mathbf{\Sigma}, ..., \mathbf{\Sigma}\right]^{\mathrm{T}} \tag{1.3.35}$$

As $q \rightarrow \infty$ then if each row of $\mathbf{\Pi}$ is sequence of scalars and $\sum_{i=1}^{\infty} \pi_i < \infty$ then the MA(q) process will have ergodic mean and the variance will satisfy,

$$\mathbb{P}\left(\sum_{t=1}^{T} \mathbf{e}^{\mathrm{T}}\mathbf{y}_t < \infty\right) = 1 \tag{1.3.36}$$

$$\mathbb{P}\left(\sum_{t=1}^{T} \left(\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t\right) < \infty\right) = 1 \tag{1.3.37}$$

The process will also have bounded fourth moments and in contrast to the AR(p) process the eigenvalues of the matrix,

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Pi}^{\mathrm{T}} \\ \mathbf{I} \quad \mathbf{0} \end{bmatrix} \tag{1.3.38}$$

are not bounded for stationarity.

1.3.3 Non-Stationary Vector Processes

Whilst non-stationary time series is not the major theme of this work, the continuous time stochastic covariance model has some intriguing results for the asymptotic theory of non-stationary time series, in particular the local non-stationarity caused by the stochastic co-dependency, this section will very briefly describe the interrelationship between multi-variate discrete time processes with unit roots and multi-variate Wiener processes. This section derives the discrete time non-stationary process.

The development of multivariate non-stationary models and the local non-stationary co-integration models has in many respects been the most important break though in modern econometrics. Formally introduced by Granger (1980) [119] and Engle and

Granger (1987) [88], co-integration has an attractive appeal. A commonly observed feature of economic time series is that whilst particular indicators often exhibit univariate unit roots, ratios of indicators are often stationary. Reviews of non-stationary vector processes, vector unit roots and co-integration, see Maddala and Kim (1998) [190], specialist book on this area. Philips (1985) in his article Time series regression with [a] unit root[s], gives the first systematic methodological approach to regression analysis of non-stationary time series models. Consider the multivariate difference data generating process,

$$\mathbf{y}_t = \mathbf{\Phi}^{\mathrm{T}} vec\left(\mathcal{D}_t\right) + \mathbf{\Psi}^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.39}$$

$$\mathcal{D}_t = [\Delta \mathbf{y}_{t-1}, ..., \Delta \mathbf{y}_{t-p}] \tag{1.3.40}$$

$$\mathcal{Y}_t = [\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-q}] \tag{1.3.41}$$

The first difference operator Δ, operates over the vector process $\mathbf{y}_t$, and is defined as $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$. If we assume that $q = 1$ can rewrite this in terms of first differences as follows,

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \Delta \mathbf{y}_t \tag{1.3.42}$$

$$\Delta \mathbf{y}_t = \mathbf{\Psi}^{\mathrm{T}} vec\left(\mathcal{D}_t\right) + \mathbf{c} + \mathbf{u}_t \tag{1.3.43}$$

By implication a multivariate unit root, process is defined as a simple extension of the univariate equivalent as follows,

$$\mathbf{y}_t = \mathbf{\Psi}^{\mathrm{T}} vec\left(D_t\right) + \mathbf{I}\mathbf{y}_{t-1} + \mathbf{c} + \mathbf{u}_t \tag{1.3.44}$$

A review of underlying multivariate vector processes cannot be complete without reviewing the concepts of state space models and co-integration. The literature on Kalman filters and co-integration is vast, this section will cover some of the key points, which will be used briefly in chapter 3, the review will cover, much of the intuitive findings and again will cover co-integration from the standpoint of detecting unit roots and various tests

1.3.4 Models with Dynamic Coefficients: The Kalman filter approach

The approach of the discrete time state-space model forms the basis of the methodologies used in Chapters 4,5 and 6. In many respects this follows the deductive reasoning that led to the ARCH model, Engle (2003) [91] discusses the development of this model and explains how its derivation was a product of work on the Kalman filter as an attempt to explain the Friedman conjecture on inflation and business cycles. In this case the ARCH model described the evolution of a process with one unobserved time varying parameter, the volatility.

For the purposes of this thesis, the majority techniques will use inferred state spaces to model dynamic dependency in the first, second and order higher moments of discrete and continuous time stochastic processes. Appendix IV offers a new innovation in multivariate time series modelling using a random matrix recursive coefficients approach, this model is then utilized to model the interactions of a set of global market

indices and sheds new light in the decoupling of Hong Kong from the global market during the asian monetary crisis in the late 1990s. Before moving onto the more advance state space model, the chapter will derive the basic state-space model, concentrating on the recursion mechanism, which will be applied at several points in the remainder of this thesis.

The Kalman filter model, Kalman (1960) [159] has been extensively used in modelling economic and financial time series since being adapted from the signal processing literature in the late 1970s and early 1980s. In the continuous time literature the Kalman-Bucy filter, Bucy and Joseph (1968) [46], is a one dimensional linear filtering method, which is the direct analogue of the discrete time Kalman filter. By products of the methodology are the ARCH/GARCH family of volatility models and the random coefficients ARMA(p,q) models. The Kalman filter approach allows for exact identification of the maximum likelihood estimates of the model and is in general a very effective method of smoothing time series with Gaussian noise white noise. Because of the mathematical rigor applied to the recursion process in Kalman filter models, this approach maybe used to identify the exact maximum likelihood specification of vector-ARMAX processes, like those used in the previous section. Another attractive feature of the Kalman filter is the use of an unobserved variable of the state space operator, a concept which shall be used extensively in chapter 4, the recursion structure in chapter takes its inspiration from the Kalman filter approach, applied

to matrix autoregressive processes. The Kalman filter has found many applications in finance and the main theoretical underpinnings provided the foundations of Engle (1986) [92] univariate ARCH model. Earlier work on Kalman filters can be traced to Duncan and Horn (1972) [147] and Anderson and Moore (1979) [8]. Harvey (1989) [141]looks at the Kalman filter approach with non-normal disturbances and Maddala and Kim, (1997) and Hamilton (1994) [135] offer extensive derivations of the Kalman filter approach.

Derivation of the State Space Kalman Filter

This derivation follows closely the main Kalman filter literature and takes its direction and notation from Ristic, Arulampalam and Gordon, (2004) [224] which is the major engineering reference on the topic. For econometricians Harvey (1989), Hamilton (1994), Magnus and Neudecker (1998) [191] and Harvey and De Rossi (2006) [228] all offer useful derivations similar to the one presented here, however neglecting much of the useful properties of the recursion. Consider a vector equation system consisting of an unobserved state vector, ψ_t and an observed vector process $\mathbf{y}_t$, assuming linear updating then the following state space and mean equations characterize the model,

$$\underset{(m\times 1)}{\psi_{t+1}} = \underset{(m\times m)}{\mathbf{\Pi}_0}\ \underset{(m\times 1)}{\psi_t} + \underset{(m\times 1)}{\mathbf{v}_{t+1}} \tag{1.3.45}$$

$$\underset{(n\times 1)}{\mathbf{y}_t} = \underset{(n\times k)}{\mathbf{A}^{\mathrm{T}}}\ \underset{(k\times 1)}{\mathbf{x}_t} + \underset{(n\times m)}{\mathbf{\Psi}^{\mathrm{T}}}\ \underset{(m\times 1)}{\psi_{t+1}} + \underset{(n\times 1)}{\mathbf{u}_t} \tag{1.3.46}$$

Where, ψ, is an $m \times 1$, vector of state spaces, $\mathbf{x}$, is a $k \times 1$, vector of exogenous variables, $\mathbf{A}$, $\mathbf{\Pi}_0$ and $\mathbf{\Psi}$ are matrices of parameters, $\mathbf{u}_t$ and $\mathbf{v}_t$ are respectively $n \times 1$

and $m \times 1$, vector disturbance processes. The expectations of this process are defined below,

$$E\left(\mathbf{v}_t \mathbf{v}_\tau^{\mathrm{T}}\right) = \begin{cases} \underset{(m \times m)}{\mathbf{\Omega}} & t = \tau \\ \mathbf{0} & t \neq \tau \end{cases} \tag{1.3.47}$$

$$E\left(\mathbf{u}_t \mathbf{u}_\tau^{\mathrm{T}}\right) = \begin{cases} \underset{(n \times n)}{\mathbf{\Sigma}} & t = \tau \\ \mathbf{0} & t \neq \tau \end{cases} \tag{1.3.48}$$

$$E\left(\mathbf{v}_t \mathbf{u}_\tau^{\mathrm{T}}\right) = \mathbf{0} \tag{1.3.49}$$

Assuming that the state space is uncorrelated with the realizations of the vector process $\mathbf{y}_t$, then the following set of relations hold, see, Maddala and Kim (1997) [190], Hamilton (1994) [135]

$$E\left(\mathbf{v}_t \psi_1^{\mathrm{T}}\right) = \mathbf{0} \tag{1.3.50}$$

$$E\left(\mathbf{u}_t \psi_1^{\mathrm{T}}\right) = \mathbf{0} \tag{1.3.51}$$

$$t \in [1, \tau] \tag{1.3.52}$$

$$\psi_t = \mathbf{v}_t + \mathbf{\Pi}_0 \mathbf{v}_{t-1} + \mathbf{\Pi}_0{}^2 \mathbf{v}_{t-2} + ... + \mathbf{\Pi}_0{}^{t-2} \mathbf{v}_2 + \mathbf{\Pi}_0{}^{t-1} \psi_1 \tag{1.3.53}$$

$$E\left(\mathbf{v}_t \psi_\tau^{\mathrm{T}}\right) = \mathbf{0} \qquad \tau = [1, T] \tag{1.3.54}$$

$$E\left(\mathbf{u}_t \psi_\tau^{\mathrm{T}}\right) = \mathbf{0} \qquad \tau = [t-1, t-2, ..., 1] \tag{1.3.55}$$

$$E\left(\mathbf{v}_t \mathbf{y}_\tau^{\mathrm{T}}\right) = \mathbf{0} \qquad \tau = [t-1, t-2, ..., 1] \tag{1.3.56}$$

$$E\left(\mathbf{u}_t \mathbf{y}_\tau^{\mathrm{T}}\right) = E\left(\mathbf{u}_t \left(\mathbf{A}^{\mathrm{T}} \mathbf{x}_\tau + \mathbf{\Psi}^{\mathrm{T}} \psi_\tau + \mathbf{u}_\tau\right)^{\mathrm{T}}\right) \tag{1.3.57}$$

$$= \mathbf{0} \tag{1.3.58}$$

In the notation previously introduced the Kalman filter is derived as a recursion of the past projections of $\hat{\mathbf{y}}_t$ on $\mathbf{y}_t$.

$$\hat{\psi}_{t+1|t} \equiv E\left(\hat{\psi}_{t+1|t} \,|\mathcal{X}_t\right) \tag{1.3.59}$$

$$\mathcal{X}_t = [\mathbf{y}_t, \mathbf{y}_{t-1}, ..., \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_1] \tag{1.3.60}$$

The error of the projections of the state spaces are as follows,

$$\mathbf{P}_{t+1|t} \equiv E\left(\left(\psi_{t+1} - \hat{\psi}_{t+1|t}\right)\left(\psi_{t+1} - \hat{\psi}_{t+1|t}\right)^{\mathrm{T}}\right) \tag{1.3.61}$$

One of the benefits of the Kalman filter is that it allows the derivation of the exact recursion sequence for a vector process and as a consequence of this recursion it is possible to construct exact identification of the structure of the maximum likelihood estimates, starting at the first iteration,

$$\hat{\psi}_{1|0} = E(\psi_1) \tag{1.3.62}$$

$$\mathbf{P}_{1|0} = E\left(\left(\psi_1 - E\left(\hat{\psi}_1\right)\right)\left(\psi_1 - E\left(\hat{\psi}_1\right)\right)^{\mathrm{T}}\right) \tag{1.3.63}$$

Moving into the forward recursion of the state space operator, with vectorization, (Magnus and Neudecker (1998) [191], results in the following,

$$E(\psi_{t+1}) = \mathbf{\Pi}_0 (E(\psi_t)) \tag{1.3.64}$$

$$(\mathbf{I} - \mathbf{\Pi}_0) E(\psi_t) = \mathbf{0} \tag{1.3.65}$$

$$E\left(\psi_{t+1}\psi_{t+1}^{\mathrm{T}}\right) = E\left((\mathbf{\Pi}_0\psi_t + \mathbf{v}_{t+1})(\mathbf{\Pi}_0\psi_t + \mathbf{v}_{t+1})^{\mathrm{T}}\right) \tag{1.3.66}$$

$$= \mathbf{\Pi}_0 \left(E\left(\psi_{t+1}\psi_{t+1}^{\mathrm{T}}\right)\right) \mathbf{\Pi}_0{}^{\mathrm{T}} + E\left(\mathbf{v}_{t+1}\mathbf{v}_{t+1}^{\mathrm{T}}\right) \tag{1.3.67}$$

$$\mathbf{M}_{\sigma_\psi} = \mathbf{\Pi}_0 \mathbf{M}_{\sigma_\psi} \mathbf{\Pi}_0{}^{\mathrm{T}} + \mathbf{\Omega} \tag{1.3.68}$$

$$vec\left(\mathbf{M}_{\sigma_\psi}\right) = (\mathbf{I} - (\mathbf{\Pi}_0 \otimes \mathbf{\Pi}_0))^{-1} .vec(\mathbf{\Omega}) \tag{1.3.69}$$

$$vec\left(\mathbf{P}_{1|0}\right) = (\mathbf{I} - (\mathbf{\Pi}_0 \otimes \mathbf{\Pi}_0))^{-1} .vec(\mathbf{\Omega}) \tag{1.3.70}$$

Once the recursion of the state space operator is defined then the projections of $\mathbf{y}_t$ may be iterated,

$$E(\psi_t | \mathbf{x}_t, \mathcal{X}_t) = E(\psi_t | \mathcal{X}_t) \tag{1.3.71}$$

$$= \hat{\psi}_{t|t-1} \tag{1.3.72}$$

$$\hat{\mathbf{y}}_{t|t-1} \equiv E(\mathbf{y}_t | \mathbf{x}_t, \mathcal{X}_t) \tag{1.3.73}$$

$$E(\mathbf{y}_t | \mathbf{x}_t, \psi_t) = \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\psi_t \tag{1.3.74}$$

The law of iterated projections then guides the derivation of each step ahead projection of $\hat{\mathbf{y}}_t$,

$$\hat{\mathbf{y}}_{t|t-1} = \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\left(E\left(\psi_t \,|\mathbf{x}_t, \mathcal{X}_t\right)\right) \tag{1.3.75}$$

$$= \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1} \tag{1.3.76}$$

$$\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\psi_t + \mathbf{u}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1} \tag{1.3.77}$$

$$= \mathbf{\Psi}^{\mathrm{T}}\left(\psi_t - \hat{\psi}_{t|t-1}\right) + \mathbf{u}_t \tag{1.3.78}$$

The squared errors of the projections are then defined as simply,

$$E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right) = E\left(\left(\mathbf{\Psi}^{\mathrm{T}}\left(\psi_t - \hat{\psi}_{t|t-1}\right)\right)\left(\mathbf{\Psi}^{\mathrm{T}}\left(\psi_t - \hat{\psi}_{t|t-1}\right)\right)^{\mathrm{T}}\right) + E\left(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\right) \tag{1.3.79}$$

$$E\left(\mathbf{u}_t\left(\psi_t - \hat{\psi}_{t|t-1}\right)^{\mathrm{T}}\right) = \mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma} \tag{1.3.80}$$

The structure of the iteration sequence of $\hat{\psi}_{t|t}$, conditioned on the past projections of $\hat{\mathbf{y}}_t$ and the innovations of $\mathcal{X}_t$ is defined as,

$$\hat{\psi}_{t|t} = E\left(\psi_t \,|\mathbf{y}_t, \mathbf{x}_t, \mathcal{X}_t\right) \tag{1.3.81}$$

$$= E\left(\psi_t \,|\mathcal{X}_t\right) \tag{1.3.82}$$

$$\hat{\psi}_{t|t} = \hat{\psi}_{t|t-1} + \left(E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right)\right)\left(E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right)\right)^{-1}\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right) \tag{1.3.83}$$

Therefore,

$$
\begin{aligned}
E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right) &= E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\right. \\
&\quad \left.\left(\boldsymbol{\Psi}^{\mathrm{T}}\left(\psi_t - \hat{\psi}_{t|t-1}\right) + \mathbf{u}_t\right)^{\mathrm{T}}\right) && (1.3.84) \\
&= E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\left(\psi_t - \hat{\psi}_{t|t-1}\right)^{\mathrm{T}}\boldsymbol{\Psi}\right) && (1.3.85) \\
&= \mathbf{P}_{t|t-1}\boldsymbol{\Psi} && (1.3.86)
\end{aligned}
$$

and the projection of the state space operator is therefore,

$$
\begin{aligned}
\hat{\psi}_{t|t} &= \hat{\psi}_{t|t-1} && (1.3.87) \\
&\quad + \mathbf{P}_{t|t-1}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\boldsymbol{\Psi} + \boldsymbol{\Sigma}\right)^{-1}\left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t - \boldsymbol{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right)
\end{aligned}
$$

Updating the mean squared error of the state space vector, results in the following,

$$
\begin{aligned}
\mathbf{P}_{t|t} &= E\left(\left(\psi_t - \hat{\psi}_{t|t}\right)\left(\psi_t - \hat{\psi}_{t|t}\right)^{\mathrm{T}}\right) && (1.3.88) \\
&= E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\left(\psi_t - \hat{\psi}_{t|t-1}\right)^{\mathrm{T}}\right) && (1.3.89) \\
&\quad -E\left(\left(\psi_t - \hat{\psi}_{t|t-1}\right)\left(\psi_t - \hat{\psi}_{t|t-1}\right)^{\mathrm{T}}\right) \\
&\quad E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right)^{-1} \\
&\quad E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)^{\mathrm{T}}\right) \\
&= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\boldsymbol{\Psi} + \boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1} && (1.3.90)
\end{aligned}
$$

Forecasting the forward recursion of the state space vector,

$$\hat{\psi}_{t+1|t} = E\left(\psi_{t+1}\left|\mathcal{X}_t\right.\right) \tag{1.3.91}$$

$$= \mathbf{\Pi}_0\left(E\left(\psi_t\left|\mathcal{X}_t\right.\right)\right) + E\left(\mathbf{v}_{t+1}\left|\mathcal{X}_t\right.\right) \tag{1.3.92}$$

$$= \mathbf{\Pi}_0\hat{\psi}_{t|t} + \mathbf{0} \tag{1.3.93}$$

$$\hat{\psi}_{t+1|t} = \mathbf{\Pi}_0\hat{\psi}_{t|t-1} + \mathbf{\Pi}_0\left(\mathbf{P}_{t|t-1}\mathbf{\Psi}\left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)^{-1}\right. \tag{1.3.94}$$

$$\left.\left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t - \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right)\right)$$

$$\mathbf{K}_t = \mathbf{\Pi}_0\mathbf{P}_{t|t-1}\mathbf{\Psi}\left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)^{-1} \tag{1.3.95}$$

$$\hat{\psi}_{t+1|t} = \mathbf{\Pi}_0\hat{\psi}_{t|t-1} + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t - \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right) \tag{1.3.96}$$

and then update the projection error of the state space vector, for each step ahead,

$$\mathbf{P}_{t+1|t} = E\left(\left(\psi_{t+1} - \hat{\psi}_{t+1|t}\right)\left(\psi_{t+1} - \hat{\psi}_{t+1|t}\right)^{\mathrm{T}}\right) \tag{1.3.97}$$

$$= E\left(\left(\mathbf{\Pi}_0\psi_t + \mathbf{v}_{t+1} - \mathbf{\Pi}_0\hat{\psi}_{t|t}\right)\left(\mathbf{\Pi}_0\psi_t + \mathbf{v}_{t+1} - \mathbf{\Pi}_0\hat{\psi}_{t|t}\right)^{\mathrm{T}}\right) \tag{1.3.98}$$

$$= \mathbf{\Pi}_0 E\left(\left(\psi_t - \hat{\psi}_{t|t}\right)\left(\psi_t - \hat{\psi}_{t|t}\right)^{\mathrm{T}}\right)\mathbf{\Pi}_0{}^{\mathrm{T}} + E\left(\mathbf{v}_{t+1}\mathbf{v}_{t+1}^{\mathrm{T}}\right) \tag{1.3.99}$$

$$= \mathbf{\Pi}_0\mathbf{P}_{t|t}\mathbf{\Pi}_0{}^{\mathrm{T}} + \mathbf{\Omega} \tag{1.3.100}$$

The final update is therefore,

$$\mathbf{P}_{t+1|t} = \mathbf{\Pi}_0\left(\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{\Psi}\left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)^{-1}\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\right)\mathbf{\Pi}_0{}^{\mathrm{T}} + \mathbf{\Omega} \tag{1.3.101}$$

and then for s-steps ahead,

$$\mathcal{X}_t = \left[\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_1, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{x}_1\right] \tag{1.3.102}$$

$$\psi_{t+s} = \mathbf{\Pi}_0{}^s\psi_t + \mathbf{\Pi}_0{}^{s-1}\mathbf{v}_{t+1} + \mathbf{\Pi}_0{}^{s-2}\mathbf{v}_{t+2} + ... + \mathbf{\Pi}_0{}^1\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s} \tag{1.3.103}$$

The expectations of the s-steps is therefore,

$$E\left(\psi_{t+s}\left|\psi_t, \mathcal{X}_t\right.\right) = \mathbf{\Pi_0}^s \psi_t \tag{1.3.104}$$

$$\hat{\psi}_{t+s|t} \equiv E\left(\psi_{t+s}\left|\mathcal{X}_t\right.\right) \tag{1.3.105}$$

$$= \mathbf{\Pi_0}^s \hat{\psi}_{t|t} \tag{1.3.106}$$

and s-step error is therefore,

$$\psi_{t+s} - \hat{\psi}_{t+s|t} = \mathbf{\Pi_0}^s \left(\psi_t - \hat{\psi}_{t|t}\right) + \mathbf{\Pi_0}^{s-1}\mathbf{v}_{t+1} \tag{1.3.107}$$

$$+\mathbf{\Pi_0}^{s-2}\mathbf{v}_{t+2} + ... + \mathbf{\Pi_0}^1 \mathbf{v}_{t+s-1} + \mathbf{v}_{t+s}$$

$$\mathbf{P}_{t+s|t} = \mathbf{\Pi_0}^s \mathbf{P}_{t|t} \left(\mathbf{\Pi_0}^s\right)^{\mathrm{T}} + \mathbf{\Pi_0}^{s-1} \mathbf{\Omega} \left(\mathbf{\Pi_0}^{s-1}\right)^{\mathrm{T}} \tag{1.3.108}$$

$$+\mathbf{\Pi_0}^{s-2} \mathbf{\Omega} \left(\mathbf{\Pi_0}^{s-2}\right)^{\mathrm{T}} + ... + \mathbf{\Pi_0} \mathbf{\Omega} \mathbf{\Pi_0}^{\mathrm{T}} + \mathbf{\Omega}$$

$$\mathbf{y}_{t+s} = \mathbf{A}^{\mathrm{T}} \mathbf{x}_{t+s} + \mathbf{\Psi}^{\mathrm{T}} \psi_{t+s} + \mathbf{u}_{t+s} \tag{1.3.109}$$

$$\hat{\mathbf{y}}_{t+s|t} \equiv E\left(\mathbf{y}_{t+s}\left|\mathcal{X}_t\right.\right) \tag{1.3.110}$$

$$= \mathbf{A}^{\mathrm{T}} \mathbf{x}_{t+s|t} + \mathbf{\Psi}^{\mathrm{T}} \hat{\psi}_{t+s|t} \tag{1.3.111}$$

$$\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t} = \left(\mathbf{A}^{\mathrm{T}} \mathbf{x}_{t+s} + \mathbf{\Psi}^{\mathrm{T}} \psi_{t+s} + \mathbf{u}_{t+s}\right) - \tag{1.3.112}$$

$$\mathbf{A}^{\mathrm{T}} \mathbf{x}_{t+s|t} + \mathbf{\Psi}^{\mathrm{T}} \hat{\psi}_{t+s|t}$$

$$= \mathbf{\Psi}^{\mathrm{T}} \left(\psi_{t+s} - \hat{\psi}_{t+s|t}\right) + \mathbf{u}_{t+s} \tag{1.3.113}$$

which yields,

$$E\left(\left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t}\right)\left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|t}\right)^{\mathrm{T}}\right) = \mathbf{\Psi}^{\mathrm{T}} \mathbf{P}_{t+s|t} \mathbf{\Psi} + \mathbf{\Sigma} \tag{1.3.114}$$

From the recursion of the model, the maximum likelihood estimates for a Gaussian objective function is as follows,

$$\mathbf{y}_t \left| \mathbf{x}_t, \mathcal{X}_t \right. \sim N\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right), \left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)\right) \tag{1.3.115}$$

Which yields the likelihood function,

$$\begin{aligned} \mathfrak{L}_{\mathbf{y}_t|\mathbf{x}_t,\mathcal{X}_t}\left(\mathbf{y}_t \left| \mathbf{x}_t, \mathcal{X}_t \right.\right) &= (2\pi)^{\frac{-n}{2}} \left|\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right|^{\frac{1}{2}} \\ &\quad \exp\left(-\tfrac{1}{2}\left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right)^{\mathrm{T}}\right. \\ &\quad \left.\left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)\left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right)\right) \\ t &\in [1, ..., \tau] \end{aligned} \tag{1.3.116}$$

Summation over the sample period $[1, ..., \tau]$, the log likelihood function reverts to the familiar objective function maximization,

$$\begin{aligned} \mathfrak{L}(\theta) &= -\tfrac{1}{2}n\tau\log(2\pi) \\ &\quad -\tfrac{1}{2}\sum_{t=1}^{\tau}\log\left|\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right| + \\ &\quad \left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right)^{\mathrm{T}} \\ &\quad \left(\mathbf{\Psi}^{\mathrm{T}}\mathbf{P}_{t|t-1}\mathbf{\Psi} + \mathbf{\Sigma}\right)^{-1} \\ &\quad \left(\mathbf{y}_t - \mathbf{A}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Psi}^{\mathrm{T}}\hat{\psi}_{t|t-1}\right) \end{aligned} \tag{1.3.117}$$

$$\theta = \left[vec(\mathbf{\Psi})^{\mathrm{T}}, vec(\mathbf{A})^{\mathrm{T}}, vec(\mathbf{\Pi}_0)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.118}$$

1.3.5 Models with Recursive Coefficients State Space R-V-ARMAX models

The basic Kalman filter representation and model updating maybe extended to form a rich class of models with dynamic coefficients with random effects. The basic univariate discrete time recursive ARX model is as follows,

$$y_t = \mathbf{b}_t^{\mathrm{T}} \mathbf{z}_t + u_t \tag{1.3.119}$$

$$\mathbf{b}_t = \mathbf{\Pi}_0^{\mathrm{T}} \mathbf{b}_{t-1} + v_t \tag{1.3.120}$$

$$\mathbf{z}_t = \left[\mathbf{x}_t^{\mathrm{T}}, 1\right]^{\mathrm{T}} \tag{1.3.121}$$

Here the coefficients $\mathbf{b}_t$ exhibit a first order vector autoregressive process. This simplest model sets $\mathbf{\Pi}_0 = \mathbf{I}$, and treats each coefficient as a random walk. The multivariate extension of the basic state space representation introduces a practical use for random matrix processes,

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{m\times n}{\mathbf{A}_t^{\mathrm{T}}}\ \underset{m\times 1}{\mathbf{z}_t} + \underset{k\times n}{\mathbf{\Phi}^{\mathrm{T}}}\ \underset{k\times 1}{\psi_t} + \underset{n\times 1}{\mathbf{u}_t} \tag{1.3.122}$$

$$\underset{m\times n}{\mathbf{A}_t} = mat\left(\underset{mn\times 1}{\mathbf{a}_t}\right) \tag{1.3.123}$$

$$\underset{mn\times 1}{\mathbf{a}_t} = \underset{mn\times mn}{\mathbf{\Pi}_0^{\mathrm{T}}}\ \underset{mn\times 1}{\mathbf{a}_{t-1}} + \underset{mn\times 1}{\mathbf{v}_t} \tag{1.3.124}$$

$$\underset{k\times 1}{\psi_t} = \underset{k\times k}{\mathbf{\Pi}_1^{\mathrm{T}}}\ \underset{k\times 1}{\psi_{t-1}} + \underset{k\times 1}{\mathbf{w}_t} \tag{1.3.125}$$

$$\mathbf{z}_t = \left[vec\left(\mathcal{Y}_t\right)^{\mathrm{T}}, vec\left(\mathcal{X}_t\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.126}$$

There are now two state space vectors, $\mathbf{a}_1$ and ψ_t, (although they maybe treated as one under some specifications) the first is a matricized vector process for the

coefficients and the second is the adjustment coefficient (sometimes referred to as the scale coefficient in the signal processing literature, see Grewal and Andrews [123] for a review of univariate specifications). Each disturbance term $\mathbf{u}_t$,$\mathbf{v}_t$ and $\mathbf{w}_t$ are assumed to be drawn from a multi-normal distribution, with no autocorrelation, the overall disturbance is therefore,

$$E\left(\mathbf{u}_t\mathbf{u}_\tau\right) = \begin{cases} \mathbf{\Sigma}, \tau = t \\ \mathbf{0}, \tau \neq t \end{cases} \tag{1.3.127}$$

$$E\left(\mathbf{v}_t\mathbf{v}_\tau\right) = \begin{cases} \mathbf{\Omega}, \tau = t \\ \mathbf{0}, \tau \neq t \end{cases} \tag{1.3.128}$$

$$E\left(\mathbf{w}_t\mathbf{w}_\tau\right) = \begin{cases} \mathbf{\Xi}, \tau = t \\ \mathbf{0}, \tau \neq t \end{cases} \tag{1.3.129}$$

$$\underset{(n(1+m)+k)\times 1}{\xi_t} = \left[\mathbf{u}_t^{\mathrm{T}}, \mathbf{v}_t^{\mathrm{T}}, \mathbf{w}_t^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.130}$$

$$E\left(\xi_t\xi_t^{\mathrm{T}}\right) = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Xi} \end{bmatrix} = \underset{(n(1+m)+k)\times(n(1+m)+k)}{\mathbf{\Lambda}} \tag{1.3.131}$$

$$\xi_t = \mathbf{\Lambda}^{\frac{1}{2}}\varepsilon_t \tag{1.3.132}$$

$$\varepsilon_t \sim N\left(\underset{(n(1+m)+k)\times 1}{\mathbf{0}}, \underset{(n(1+m)+k)\times(n(1+m)+k)}{\mathbf{I}}\right) \tag{1.3.133}$$

Appendix IV uses the recursive approach to model stock market integration and response to cross market shocks. The Kalman filter approach, as an econometric tool in the modern finance literature, has been used recently to evaluate dynamic market models in Zalweska and Grout (2003) [124], where the dynamic coefficients are used to estimate the evolution of stock price beta's. The use of state space representations are used in the identification of volatility models in Javaheri (2005) and by Chen,

Huang and Lin (2005) [104] to analyze stock market bubbles. Relaxing the normality assumption in the estimation of the model has been achieved in two ways, first by maximum likelihood estimation under different distributional assumptions e.g. the multi-t distribution, see Harvey and De Rossi (2006) [228] and second by GMM estimation, i.e. setting objective theoretical moments from some system of disturbances and minimizing the parameter vector in terms of these moments over the sample, this is an appealing approach as it is possible to over-identify the variables in $\mathbf{x}$ and the length of the state-space vector ψ_t and begin to restrict both accordingly using Hansen's (1982) [137] χ^2 test. GMM estimation of random coefficient ARMA models will be covered in section 6.2.

1.3.6 Maximum Likelihood Estimation

In order to construct robustness and parameter restriction tests for the models outlined an understanding the implications of various parameter specifications and the asymptotic properties of maximum likelihood estimates of the k length parameter vector θ. Consider the following notation for a partitioned Jacobian matrix,

$$\mathbf{D}_j\left(f\left(\mathbf{z}\right)\right) = \left[\frac{\partial f\left(\mathbf{z}\right)}{\partial \mathbf{x}_j^{\mathrm{T}}}\right] \tag{1.3.134}$$

$$\mathbf{z} = \left[\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}}, ..., \mathbf{x}_n^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.135}$$

For a log likelihood function, $\mathfrak{L}_t(\theta) = \log \mathfrak{F}_t(\theta)$, where

$$t \in [1, 2, ..., \tau] \tag{1.3.136}$$

$$\theta = [\theta_1, ..., \theta_\gamma]^{\mathrm{T}} \tag{1.3.137}$$

The objective function for a given distribution is,

$$\mathfrak{F}(\theta) = \prod_{t=1}^{\tau} \mathfrak{F}_t(\theta) \tag{1.3.138}$$

$$\mathfrak{L}(\theta) = \log(\mathfrak{F}(\theta)) \equiv \sum_{t=1}^{\tau} \log(\mathfrak{F}_t(\theta)) \tag{1.3.139}$$

Taking the first derivative of the likelihood function gives the following gradient vector of partial derivatives,

$$\nabla \mathfrak{L}_t(\theta) = \left[\frac{\partial \mathfrak{L}(\theta)}{\partial \theta_1}, ..., \frac{\partial \mathfrak{L}(\theta)}{\partial \theta_\gamma}\right]^{\mathrm{T}} \tag{1.3.140}$$

Taking the second derivative and forming the hessian yields,

$$\mathfrak{H}_t(\theta) = \nabla^2 \mathfrak{L}_t(\theta) = \begin{bmatrix} \frac{\partial^2 \mathfrak{L}(\theta)}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial^2 \mathfrak{L}(\theta)}{\partial \theta_1 \partial \theta_\gamma} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathfrak{L}(\theta)}{\partial \theta_\gamma \partial \theta_1} & \cdots & \frac{\partial^2 \mathfrak{L}(\theta)}{\partial \theta_\gamma \partial \theta_\gamma} \end{bmatrix} \tag{1.3.141}$$

From the Hessian the information matrix, see Magnus and Neudecker (1998) [191] is defined as,

$$\Xi = \mathfrak{I}_t(\theta) = -E(\mathfrak{H}(\theta)) \tag{1.3.142}$$

The asymptotic information matrix is therefore,

$$\Xi_\infty = \mathfrak{I}_\infty(\theta) = \lim_{\tau \to \infty} \left(\frac{1}{\tau}\right) \mathfrak{F}_t(\theta) \tag{1.3.143}$$

The previous section derived the Kalman filter and it now provides us with the three basic specifications for maximum likelihood estimation, with a multi-normal/Gaussian objective function, first one with parameters affecting the conditional mean only, second one with overlapping parameters in mean and variance/covariance and finally a specification with separate parameters in mean and variance/covariance. The importance of the second and, in particular, third specifications will be shown in the next section on MV-ARCH/GARCH models. In their core text on the subject Magnus and Neudecker (1988,re 1999) [192] [191], bring together the work of Norden (1972) [211], Cramer (1986) [60], Holly (1988) [145] and Magnus and Neudecker (1980) [191], whilst the origins of this methodology dates back to Fisher (1966). The interested reader is directed to Srivastava (2002) [241] for a modern treatment, with accessible explanation and to Magnus and Neudecker (1999) [192] which is the main resource on this methodology. Consider the following general model, where μ_t and $\mathbf{\Sigma}_t$ are the conditional mean and conditional covariance, based on some set of parameters,

$$\mathbf{y}_t = \mu_t(\theta) + (\mathbf{\Sigma}_t(\theta))^{\frac{1}{2}} \varepsilon_t \tag{1.3.144}$$

$$\theta \in \mathbf{\Theta} \tag{1.3.145}$$

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}) \tag{1.3.146}$$

In mean only models

Now consider a model whereby the covariance matrix is $\mathbf{y}_t \in \mathbb{R}^n$ and is ergodic, i.e. $\mathbb{P}\left(\mathbf{y}_t^{\mathrm{T}}\mathbf{y}_t < \infty\right) = 1$.

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_0 \tag{1.3.147}$$

Therefore the covariance matrix $\boldsymbol{\Sigma}_t$ for each vector draw from the sample is identical, $\boldsymbol{\Sigma}_0$ and conditionally invariant. For a sample matrix of observations of $\mathbf{y}_t$, i.e. $\mathcal{Y}_\tau = \mathbf{Y}^{\mathrm{T}}$ where $t \in [1, ..., \tau]$,

$$\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_\tau]^{\mathrm{T}} \tag{1.3.148}$$

$$\mathbf{M} = [\mu_1(\theta^\mu), ..., \mu_\tau(\theta^\mu)]^{\mathrm{T}} \tag{1.3.149}$$

$$\mathbf{U} = \mathbf{Y} - \mathbf{M} \tag{1.3.150}$$

$$\mathfrak{L}(\theta^\mu) = -\tfrac{1}{2}\left(n\tau \log(2\pi) + n \log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathrm{T}}\right)\right) \tag{1.3.151}$$

$$\mathfrak{F}\left(\hat{\theta}^\mu\right) \triangleq \min\left(\tfrac{1}{2}\operatorname{tr}\left(\mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathrm{T}}\right)\right) \tag{1.3.152}$$

We can now set up a constrained quadratic optimization problem, by imposing the following constraints, $\mathbf{M}\alpha = \mathbf{0}$ and $\alpha^{\mathrm{T}}\boldsymbol{\Sigma}\alpha = 1$, where $\alpha \in \mathbb{R}^n$ is some unknown vector in the kernel of $\boldsymbol{\Sigma}$, the lagrangian function is therefore,

$$\mathfrak{G}(\theta^\mu) = \tfrac{1}{2}\operatorname{tr}\left(\mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^{\mathrm{T}}\right) - \mathbf{l}^{\mathrm{T}}\mathbf{M}\alpha - l\left(\alpha^{\mathrm{T}}\boldsymbol{\Sigma}\alpha - 1\right) \tag{1.3.153}$$

Where $\lambda = \left[\mathbf{l}^{\mathrm{T}}, l\right]^{\mathrm{T}}$, is a column vector of lagrange multipliers. The elements of the information matrix, $\boldsymbol{\Xi}$, are therefore,

$$\begin{aligned} \xi\left(\theta^{\mu}\right) &= \left(\frac{\partial \mu\left(\theta^{\mu}\right)}{\partial \theta_i^{\mu}}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\left(\theta^{\mu}\right) \frac{\partial \mu\left(\theta^{\mu}\right)}{\partial \theta_j^{\mu}} \\ &\quad + \tfrac{1}{2} \operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}\left(\theta^{\mu}\right)}{\partial \theta_i^{\mu}} \boldsymbol{\Sigma}\left(\theta^{\mu}\right) \frac{\partial \boldsymbol{\Sigma}^{-1}\left(\theta^{\mu}\right)}{\partial \theta_j^{\mu}} \boldsymbol{\Sigma}\left(\theta^{\mu}\right)\right) \end{aligned} \tag{1.3.154}$$

From the Cramér-Rao inequality, see Cramer (1986) and for modern applications see Williams (2006), the inverse of the information matrix $\boldsymbol{\Xi}^{-1}$, is utilized as follows,

$$\zeta = \left(\hat{\theta} - \theta\right) \tag{1.3.155}$$

$$\zeta \sim N\left(\mathbf{0}, \boldsymbol{\Xi}^{-1}\right) \tag{1.3.156}$$

Therefore, if $\varsigma = diag\left(\boldsymbol{\Xi}^{-1}\right)$, then $\sqrt{\varsigma_i}$ is the standard error of the i^{th} parameter.

1.3.7 Restriction Tests under Miss-specification

Another important use of the information matrix is in defining restriction tests, under model miss-specification. Consider the following example, if the density function of the model disturbances is drawn from a finite mixture distribution, i.e. $\mathbf{y}_t - \mu_t = \mathbf{u}_t$, where $\mathfrak{F}\left(\mathbf{u}_t\right) = \sum_{i=1}^{n} \alpha_i \Phi\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right)$, $\sum_{i=1}^{n} \alpha_i = 1$ and $\Phi(.)$, is the multivariate normal density function. Foutz and Srivastava (1979) [98], suggest that under mild conditions, e.g. $n = 2$, and $-\frac{1}{2}\left(-\log\left|\boldsymbol{\Sigma}_1\right| + \log\left(\left|\boldsymbol{\Sigma}_1\right| + \left|\boldsymbol{\Sigma}_2\right|\right)\right) > \alpha$, the following test maybe used to test for over-identified parameters.

Over Identification under Model Miss-Specification

Consider the following partitioned parameter vector $\theta = \left[\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}\right]^{\mathrm{T}}$, where θ_1 is the true parameter vector and θ_2 is a vector of over identified parameters. For example if $\theta = \left[\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}\right]^{\mathrm{T}}$ is the unrestricted parameter vector, formed by p true parameters and q, over-estimated parameters, if the evaluated log-likelihoods, over $t \in [1, ..., \tau]$, of the restricted and unrestricted models are, $\sum_{t=1}^{\tau} \mathfrak{L}_t(\theta) = \varrho$ and $\sum_{t=1}^{\tau} \mathfrak{L}_t(\theta_1) = \varrho_1$ the eigenvalues, $\lambda_i \in \{\lambda_1, ..., \lambda_{p+q}\}$, of the unrestricted information matrix $\Xi_u = \mathfrak{I}(\theta)$ maybe used to test the restriction of θ against θ_1. Foutz and Srivastava (1977) suggest the following likelihood ratio test,

$$2(\log(\varrho) - \log(\varrho_1)) = \varphi \tag{1.3.157}$$

Where $\varphi \sim \sum_{i=1}^{q} \lambda_i \chi^1(\nu = 1)$. Using this result, a hypothesis testing algorithm maybe designed for restrictions for over-identified models with possibly miss-specified distributional assumptions, in this case the density function being drawn from a mixture of zero centered multi-normals. If the distribution is drawn from a distribution where $\alpha = 0$, then the over-identification algorithm maybe defined as follows,

$$\begin{array}{ll} H_0 : \theta = \left[\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}\right]^{\mathrm{T}} & \varphi > \phi(\nu = q, P) \\ H_1 : \theta = \theta_1 & \varphi < \phi(\nu = q, P) \end{array} \tag{1.3.158}$$

Where ϕ is the inverse cumulative χ^2 density function $P \in [0, 1]$ and ν is the number of degrees of freedom. When $0 < \alpha < -\frac{1}{2}(-\log|\Sigma_1| + \log(|\Sigma_1| + |\Sigma_2|))$, then the

following test maybe employed,

$$\begin{array}{ll} H_0 : \theta = \left[\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}\right]^{\mathrm{T}} & \varphi > \sum_{i=1}^{q} \lambda_i \phi\left(\nu = 1, p\right) \\ H_1 : \theta = \theta_1 & \varphi < \sum_{i=1}^{q} \lambda_i \phi\left(\nu = 1, p\right) \end{array} \tag{1.3.159}$$

The eigenvalues of Ξ in effect weight the restrictions in order to account for the miss-specification of the density function.

Mean and Variance Models with Separate Parameters

We can now generalize the previous method to problems involving conditional parameterization of the covariance matrix, consider 1.3.147, now assuming that the covariance matrix is conditioned on some set of parameters independent of the mean equation parameters,

$$\mathbf{y}_t = \mu_t\left(\theta^{\mu}\right) + \left(\mathbf{\Sigma}_t\left(\theta^{\sigma}\right)\right)^{\frac{1}{2}} \varepsilon_t \tag{1.3.160}$$

$$\theta = \left[\left(\theta^{\mu}\right)^{\mathrm{T}}, \left(\theta^{\sigma}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{1.3.161}$$

$$\theta \in \mathbf{\Theta} \tag{1.3.162}$$

$$\varepsilon_t \sim N\left(\mathbf{0}, \mathbf{I}\right) \tag{1.3.163}$$

This is a specific case for various types of MV-GARCH models whereby the conditional variance/covariance matrix is based upon a different set of parameters to the mean model, however the residuals from the mean model feed directly into the volatility model. As such the normal parameterization of multivariate ARCH models is done in two steps, first a mean model is constructed with time invariant covariance

matrix and the residuals extracted. To these residuals a volatility model is then fitted, with the assumption that the mean of the residuals is now restricted to being unconditionally zero.

Mean and Variance Models with Overlapping Parameters

The final specification relies on models whereby the first and second moments are described by processes with parameters common to both. Consider,

$$\mathbf{y}_t \sim N\left(\mu_t\left(\theta\right), \mathbf{\Sigma}_t\left(\theta\right)\right) \tag{1.3.164}$$

where, $\theta = \left[\theta_{\mu}^{\mathrm{T}}, \theta_{\mathbf{\Sigma}}^{\mathrm{T}}, \theta_{\mu,\mathbf{\Sigma}}^{\mathrm{T}}\right]^{\mathrm{T}}$, here the subscripts represent the parameter belonging to the mean equation, θ_{μ}, the variance equation, $\theta_{\mathbf{\Sigma}}$, and those that overlap, $\theta_{\mu,\mathbf{\Sigma}}$, the maximum likelihood estimator of θ, is therefore obtained by,

$$0 = \mathbf{u}_t \mathbf{\Sigma}_t^{-1} \frac{\partial \mu_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mu}\right)}{\partial \theta_{\mu}} \tag{1.3.165}$$

$$\begin{aligned} 0 = {} & \tfrac{1}{2} \operatorname{tr}\left(\frac{\partial \mathbf{\Sigma}_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mathbf{\Sigma}}\right)}{\partial \theta_{\mu,\mathbf{\Sigma}}}\right) \\ & + \mathbf{u}_t \mathbf{\Sigma}_t^{-1} \frac{\partial \mu_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mu}\right)}{\partial \theta_{\mu,\mathbf{\Sigma}}} \\ & - \tfrac{1}{2} \frac{\partial \mathbf{\Sigma}_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mathbf{\Sigma}}\right)}{\partial \theta_{\mu,\mathbf{\Sigma}}} \mathbf{u}_t \end{aligned} \tag{1.3.166}$$

$$\operatorname{tr}\left(\frac{\partial \mathbf{\Sigma}_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mathbf{\Sigma}}\right)}{\partial \theta} \mathbf{\Sigma}_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mathbf{\Sigma}}\right)\right) = \mathbf{u}_t^{\mathrm{T}} \frac{\partial \mathbf{\Sigma}_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mathbf{\Sigma}}\right)}{\partial \theta} \mathbf{u}_t \tag{1.3.167}$$

where, the residuals are set out as,

$$\mathbf{u}_t = \mathbf{y}_t - \mu_t\left(\theta_{\mu,\mathbf{\Sigma}}, \theta_{\mu}\right) \tag{1.3.168}$$

The information matrix is therefore,

$$\Xi = \begin{bmatrix} \mathbf{F}_{\mu,\mu} & \mathbf{F}_{\Sigma,\mu,\mu} & \mathbf{0} \\ \mathbf{F}_{\mu,\mu\Sigma} & \mathbf{F}_{\mu\Sigma,\mu\Sigma} & \mathbf{F}_{\mu\Sigma,\Sigma} \\ \mathbf{0} & \mathbf{F}_{\Sigma,\mu\Sigma} & \mathbf{F}_{\Sigma,\Sigma} \end{bmatrix} \tag{1.3.169}$$

where,

$$\mathbf{F}_{\mu,\mu} = \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_\mu}\right)^{\mathrm{T}} \Sigma^{-1} \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_\mu}\right) \tag{1.3.170}$$

$$\begin{aligned}
\mathbf{F}_{\mu,\mu\Sigma} &= \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_\mu}\right)^{\mathrm{T}} \Sigma^{-1} \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_{\mu,\Sigma}}\right) \\
\mathbf{F}_{\mu\Sigma,\mu\Sigma} &= \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_{\mu,\Sigma}}\right)^{\mathrm{T}} \Sigma^{-1} \left(\frac{\partial \mu(\theta_\mu, \theta_{\mu,\Sigma})}{\partial \theta_{\mu,\Sigma}}\right) \\
&\quad + \tfrac{1}{2}\left(\left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_\Sigma}\right) vec(\Sigma)^{\mathrm{T}} \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_\Sigma}\right) vec(\Sigma)^{\mathrm{T}}\right) \\
\mathbf{F}_{\mu\Sigma,\Sigma} &= \tfrac{1}{2}\left(\left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_{\mu,\Sigma}}\right) vec(\Sigma)^{\mathrm{T}} \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_\Sigma}\right) vec(\Sigma)^{\mathrm{T}}\right) \\
\mathbf{F}_{\Sigma,\Sigma} &= \tfrac{1}{2}\left(\left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_\Sigma}\right) vec(\Sigma)^{\mathrm{T}} \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \left(\frac{\partial vec\Sigma(\theta_{\mu,\Sigma}, \theta_\Sigma)}{\partial \theta_\Sigma}\right) vec(\Sigma)^{\mathrm{T}}\right)
\end{aligned} \tag{1.3.171}$$

This is the generalized asymptotic information matrix for all multi-normal specifications and forms the basis of the restriction tests utilized hence forth. Neudecker and Liu (1993) [208] demonstrate that the asymptotic variance matrix, $\Theta = \frac{1}{\tau}\Xi_t$, is positive definite, therefore yielding positive standard errors for $\theta \in \mathbb{R}^d$, where d is the length of θ and $t \in [1, ..., \tau]$. This system then stands as the de-factor estimation for all specifications of the form,

$$\mathbf{y}_t \sim N(\mu_t, \Sigma_t | \theta) \tag{1.3.172}$$

i.e. for all linear and non-linear, multi-equation regression models.

1.3.8 Random Matrix Processes

In section 1.2.5 a random non-negative matrix was generated from a normally distributed random vector, using the *ivech* transformation and the outer product of a matrix with its transpose. (Note in this section the assumption is that all matrix products are normalized, i.e. if $diag\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right) :\neq \mathbf{e}$, then the product $\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ is redefined as $\left\{\mathbf{A} = \mathbf{B} \times \mathbf{q}\mathbf{q}^{\mathrm{T}}; \mathbf{B} = \mathbf{X}^{\mathrm{T}}\mathbf{X}; \mathbf{q} = \left[\sqrt{p_i}^{-1}\right]_{n\times 1}; \mathbf{p} = diag\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\right\}$). Now consider the zero centred sample of observations, generated from some covariance matrix, $\mathbf{\Sigma}$, $\mathcal{U}_{\{1,\ldots,\tau\}} = \{\mathbf{u}_1, \ldots, \mathbf{u}_\tau\}$, where $\mathbf{u}_t \sim N(0, \mathbf{\Sigma})$. The $\tau \times n$ data matrix $\mathbf{U}$, where $\mathbf{U} = \underset{\tau\times n}{mat}\mathcal{U}_{\{1,\ldots,\tau\}}$, is then a matrix of multi-normals, which may then be defined in terms of an $\tau \times n$ matrix of independent standard normals, $\mathbf{E}$ and the upper triangular (Cholesky) factorization, $\mathbf{\Sigma}^{\frac{1}{2}}$, of the covariance matrix as $\mathbf{U} = \mathbf{E}\mathbf{\Sigma}^{\frac{1}{2}}$, asymptotically as the number of observations tends to infinity then the estimated matrix, $\hat{\mathbf{\Sigma}}$, tends to the generating non-negative definite matrix $\mathbf{\Sigma}$, more specifically the following derivation (see Srivastava and Khatri (1979) [242] for more details) illustrates how the uncertainty in estimates of a covariance matrix maybe derived in terms of the uncertainty parameterizing the standard normal distribution.

Systematic Convergent Variability in Estimation of the Covariance Matrix

Consider a $\tau \times n$ data matrix $\mathbf{E}$ with elements drawn from a standard normal distribution with the following properties,

$$\mathbf{E} = [\varepsilon]_{\tau \times n} \tag{1.3.173}$$

$$[\varepsilon]_{\tau \times n} \sim N(0,1) \tag{1.3.174}$$

$$\mathbf{E} = [\varepsilon_1^c, ..., \varepsilon_n^c] \tag{1.3.175}$$

$$\mathbf{E} = \left[(\varepsilon_1^r)^{\mathrm{T}}, ..., (\varepsilon_\tau^c)^{\mathrm{T}} \right]^{\mathrm{T}} \tag{1.3.176}$$

$$\mathbf{e}^c \sim N\left(\mathbf{0}, \underset{\tau \times \tau}{\mathbf{I}}\right) \tag{1.3.177}$$

$$\mathbf{e}^r \sim N\left(\mathbf{0}, \underset{n \times n}{\mathbf{I}}\right) \tag{1.3.178}$$

$$\underset{n \times n}{\hat{\mathbf{I}}} = \mathbf{E}^{\mathrm{T}}\mathbf{E} \tag{1.3.179}$$

$$\hat{\mathbf{I}} \rightarrow \mathbf{I} : \{\tau \rightarrow \infty\} \tag{1.3.180}$$

Therefore defining a multi-normal data matrix $\mathbf{U}$ in terms of $\mathbf{E}$, recalling that $\mathbf{u} = \varepsilon\boldsymbol{\Sigma}^{\frac{1}{2}}$,

$$\mathbf{U} = \mathbf{E}\boldsymbol{\Sigma}^{\frac{1}{2}} \tag{1.3.181}$$

$$E\left(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\right) = \boldsymbol{\Sigma} \tag{1.3.182}$$

$$\boldsymbol{\Delta} = \mu\mu^{\mathrm{T}} \tag{1.3.183}$$

$$\mathbf{u}_t \sim N\left(\mu, \boldsymbol{\Sigma}\right) \tag{1.3.184}$$

$$\hat{\boldsymbol{\Sigma}} = \mathbf{U}^{\mathrm{T}}\mathbf{U} \tag{1.3.185}$$

$$\hat{\boldsymbol{\Sigma}} = \left(\left(\mathbf{E}\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}}\mathbf{E}\boldsymbol{\Sigma}^{\frac{1}{2}}\right) \tag{1.3.186}$$

$$\hat{\boldsymbol{\Sigma}} = \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{\mathrm{T}}\hat{\mathbf{I}}\boldsymbol{\Sigma}^{\frac{1}{2}} \equiv \sum_{t=1}^{\tau}\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}} \tag{1.3.187}$$

$$\hat{\boldsymbol{\Sigma}} \rightarrow \boldsymbol{\Sigma} : \{\tau \rightarrow \infty\} \tag{1.3.188}$$

Therefore given repeated draws of the data matrix $\mathbf{E}$, the estimates of the matrix $\hat{\boldsymbol{\Sigma}}$ will be distributed with a Wishart distribution $\hat{\boldsymbol{\Sigma}} \sim W\left(\boldsymbol{\Sigma}, \tau, \boldsymbol{\Delta}\right)$, more extensive coverage of the Wishart distribution and it's moments is given in Kollo and Van Rosen (2005) [170] and Muirhead (2005) [202].

1.3.9 Variability in the Quadratic Form

Consider an $n \times 1$ vector α, where $\alpha \neq \mathbf{0}$, for a non-negative matrix, $\hat{\mathbf{\Sigma}} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$, the distribution of $\alpha^{\mathrm{T}}\hat{\mathbf{\Sigma}}\alpha$, for a set of random draws of the data matrix $\mathbf{U}$ is as follows,

$$\alpha^{\mathrm{T}}\mathbf{\Sigma}\alpha = a^2 \tag{1.3.189}$$

$$\alpha^{\mathrm{T}}\hat{\mathbf{\Sigma}}\alpha = a\chi^2(\tau) \tag{1.3.190}$$

Where $\chi^2(\tau)$ is a χ^2 distribution with n degrees of freedom. If $\mu \neq \mathbf{0}$, then $\alpha^{\mathrm{T}}\hat{\mathbf{\Sigma}}\alpha$ is distributed with a non-central χ^2 distribution with re-centering parameter $\delta = trace\,(\mu\mu)^{\mathrm{T}}$. Given that the matrix $\hat{\mathbf{\Sigma}}$ is drawn from a distribution of matrices, much like the sample variance is drawn from a distribution of variances and the fact that the distribution of the scalar $\alpha^{\mathrm{T}}\hat{\mathbf{\Sigma}}\alpha$ is drawn from a χ^2 distribution the Wishart distribution is often thought of as a multivariate χ^2 distribution. However this is not strictly the case as to be considered a pure multivariate analogue each marginal distribution (single dimension projection of the multivariate density function) must be the univariate analogue, which in this case it is not. If each observation of the sample $\mathbf{u}_t$ is drawn is multi-normal then the sample covariance matrix is drawn from the Wishart distribution. If the distribution $\mathbf{u}_t \sim N(\mu, \mathbf{\Sigma})$, and $\mu = \mathbf{0}$, then the matrix $\hat{\mathbf{\Sigma}}$, is drawn from the central Wishart distribution. If $\mu \neq \mathbf{0}$, then the non-central

Wishart distribution, of $\hat{\mathbf{\Sigma}}$ is,

$$\mathfrak{F}\left(\hat{\sigma}_{i,j}\right) = \frac{\left|\hat{\sigma}_{i,j}\right|^{\frac{1}{2}(\tau-n-1)} \exp\left(-\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\frac{\hat{\sigma}_{i,j}}{2}\right)\right)}{2^{\frac{1}{2}n(n-1)}\left|\mathbf{\Sigma}\right|^{\frac{\tau}{2}}\Gamma_n\left(\frac{1}{2}\tau\right)} \tag{1.3.191}$$

$$\Gamma_n\left(\tfrac{1}{2}\tau\right) = \pi^{\frac{1}{4}n(n-1)}\prod_{k=1}^{n}\Gamma\left(\tfrac{1}{2}\left(\tau+1-k\right)\right) \tag{1.3.192}$$

$$\Gamma\left(\tfrac{1}{2}\left(\tau+1-k\right)\right) = \left(\tfrac{1}{2}\left(\tau+1-k\right)-1\right)! \tag{1.3.193}$$

where $\mathbf{\Xi}$ is a fixed non-negative matrix and maybe estimated via maximum likelihood. Extending the expectations operator from a vector to a matrix process.

$$\mathbf{\Psi}_t - \hat{\mathbf{\Psi}}_{t\left|\Omega_{t}\right|\{t-r,\ldots,t-1\}} = \mathbf{U}_t \tag{1.3.194}$$

$$E\left(\mathbf{U}_t\right) = \underset{n\times n}{\mathbf{0}} \tag{1.3.195}$$

$$\mathbf{U}_t \sim W\left(\mathbf{\Sigma},\tau\right) \tag{1.3.196}$$

This is very useful to as we can specify a matrix process for covariation, consider the simple stationary sequence of covariance matrices generated by the following process,

$$\mathbf{H}_t = \mathbf{A}\times\mathbf{H}_{t-1}+\mathbf{\Sigma}_t \tag{1.3.197}$$

$$\mathbf{\Sigma}_t \sim W\left(\mathbf{\Omega},t=v\right) \tag{1.3.198}$$

Whilst it is possible to think of the $\mathbf{\Sigma}_t$ term as analogous to a normal disturbance term, the stationarity conditions are somewhat different, first the diagonal elements of $\mathbf{\Sigma}_t$, are always positive by definition and as such the diagonal elements of the parameter matrix are bounded to $\mathbf{1} > diag\left(\mathbf{A}\right) > \mathbf{0}$, and to ensure $\mathbf{H}_t$ is non-negative, the parameter matrix $\mathbf{A}$, must also be non-negative. The degree of variation in the

covariance matrices, i.e. $E(\mathbf{\Sigma}_t - \mathbf{\Omega}|v)$, in 1.3.191 the variable t is the number of observations from the data matrix $\mathbf{U}$, and as $\tau \to \infty$, the matrix $\hat{\mathbf{\Sigma}} \to \mathbf{\Sigma}$. Now the generating matrix is $\mathbf{\Omega}$ and as the velocity of the convergence is dependent on the size of the determinant $|\mathbf{\Omega}|$ and the effective sample size $\tau = v$. The quadratic variability of the matrix is the second moment of the matrix process which is defined as follows,

$$E\left((\mathbf{\Sigma}_t - \mathbf{\Omega}) \otimes (\mathbf{\Sigma}_t - \mathbf{\Omega})\right) = \begin{bmatrix} E\left((\sigma_{1,1,t} - \omega_{1,1,t})(\mathbf{\Sigma}_t - \mathbf{\Omega})\right) & \cdots & E\left((\sigma_{1,n,t} - \omega_{1,n,t})(\mathbf{\Sigma}_t - \mathbf{\Omega})\right) \\ \vdots & \ddots & \vdots \\ E\left((\sigma_{n,1,t} - \omega_{n,1,t})(\mathbf{\Sigma}_t - \mathbf{\Omega})\right) & \cdots & E\left((\sigma_{n,n,t} - \omega_{n,n,t})(\mathbf{\Sigma}_t - \mathbf{\Omega})\right) \end{bmatrix}$$

Random matrix processes are usually defined in terms of the matricized vector process, however it is often necessary to ensure non-negativity in the random matrix, something which is often very difficult to build into a vector process without compromising the degree of variability in the resulting process.

Example of a Set of Bivariate Normal Random Draws

A simple empirical test of the Wishart distribution requires the generation of multiple sets of draws from a multivariate normal distribution, then comparing the inner product $\alpha\mathbf{\Sigma}\alpha$, to the anticipated χ^2 distribution using the Kolmogorov-Smirnov test. The distribution of the three unique elements of the estimated 2×2 matrix $\hat{\mathbf{\Sigma}}$, from 500 experiments, with sample sizes varying from 10 to 10,000 in length. The distribution of the scalar $\alpha^{\mathrm{T}}\mathbf{\Sigma}\alpha$

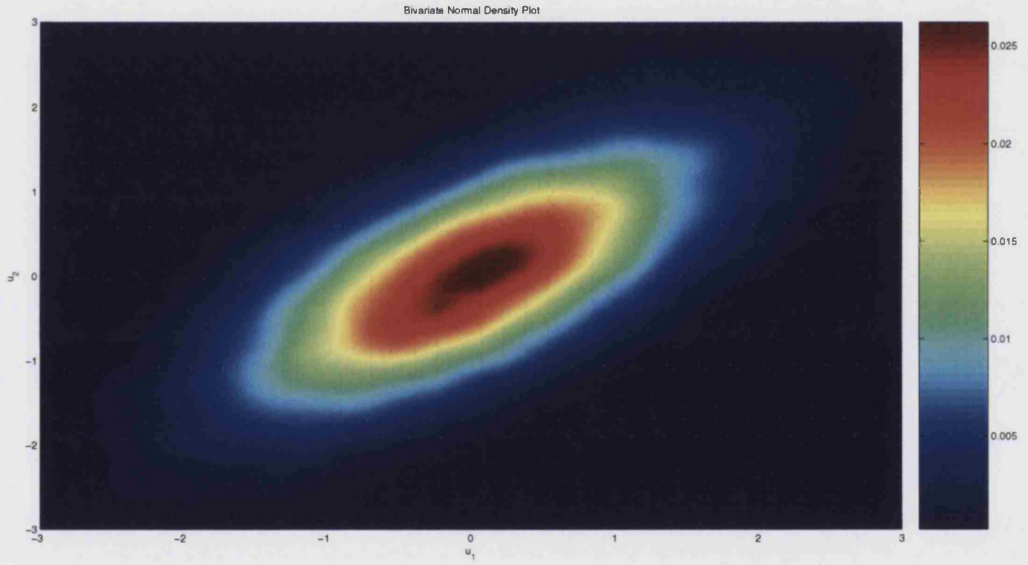

Figure 1.1: A Contour plot from one draw (10,000 observations) of a bivariate normal data matrix.

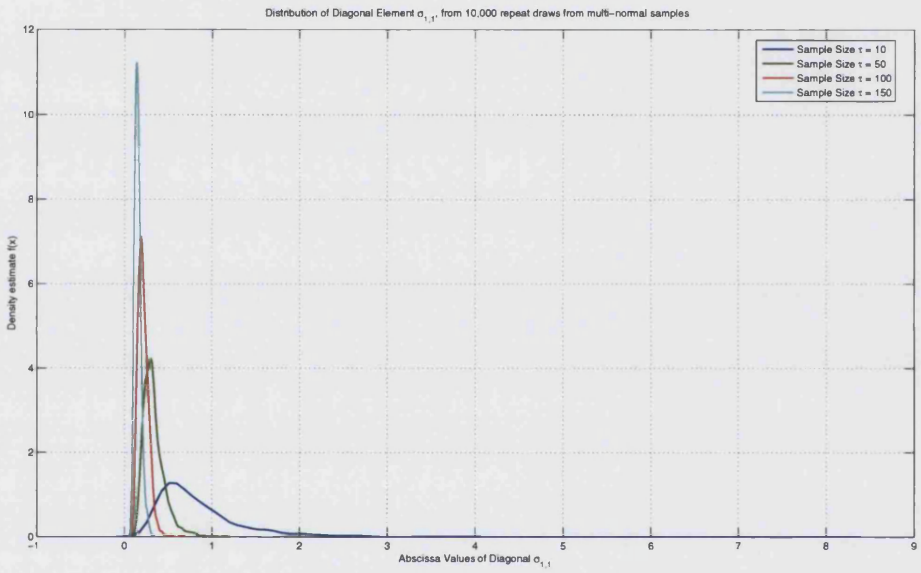
Distribution of Diagonal Element $\sigma_{1,1}$ from 10,000 repeat draws from multi-normal samples
Sample Size τ = 10
Sample Size τ = 50
Sample Size τ = 100
Sample Size τ = 150
Density estimate f(x)
Abscissa Values of Diagonal $\sigma_{1,1}$
0
2
4
6
8
10
12
−1
0
1
2
3
4
5
6
7
8
9

Figure 1.2:

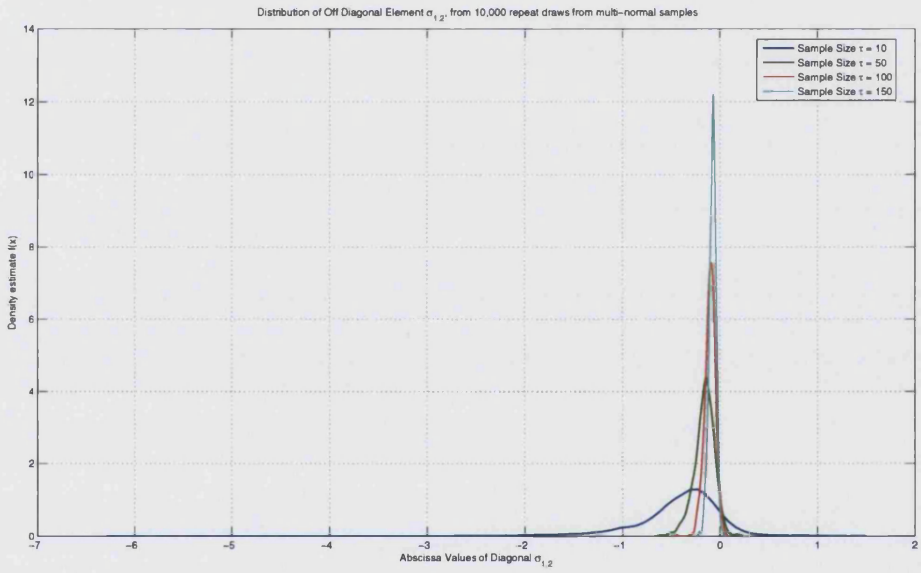
Distribution of Off Diagonal Element $\sigma_{1,2}$ from 10,000 repeat draws from multi-normal samples
Sample Size τ = 10
Sample Size τ = 50
Sample Size τ = 100
Sample Size τ = 150
Density estimate f(x)
Abscissa Values of Diagonal $\sigma_{1,2}$
0
2
4
6
8
10
12
14
−7
−6
−5
−4
−3
−2
−1
0
1
2

Figure 1.3:

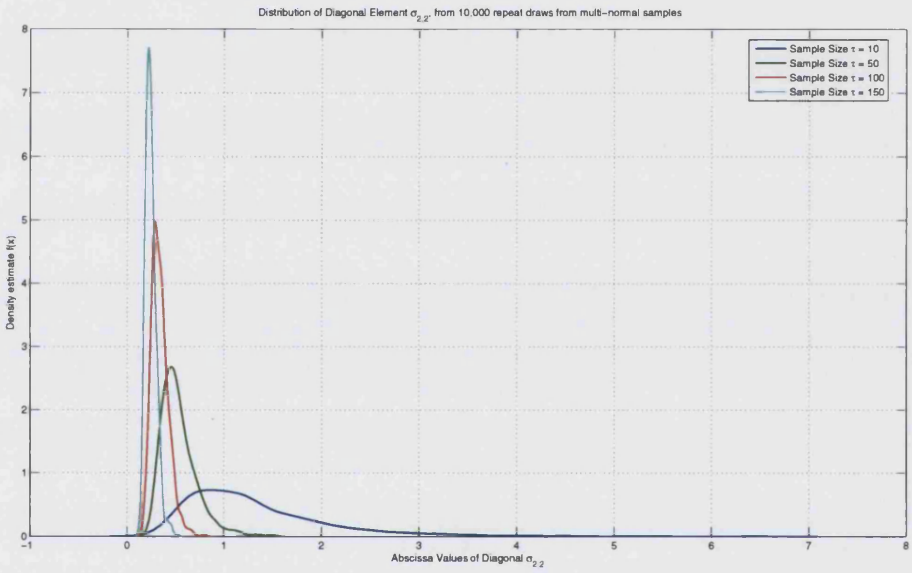
Distribution of Diagonal Element $\sigma_{2,2}$ from 10,000 repeat draws from multi-normal samples
Sample Size τ = 10
Sample Size τ = 50
Sample Size τ = 100
Sample Size τ = 150
Density estimate f(x)
Abscissa Values of Diagonal $\sigma_{2,2}$
0
1
2
3
4
5
6
7
8
−1

Figure 1.4:

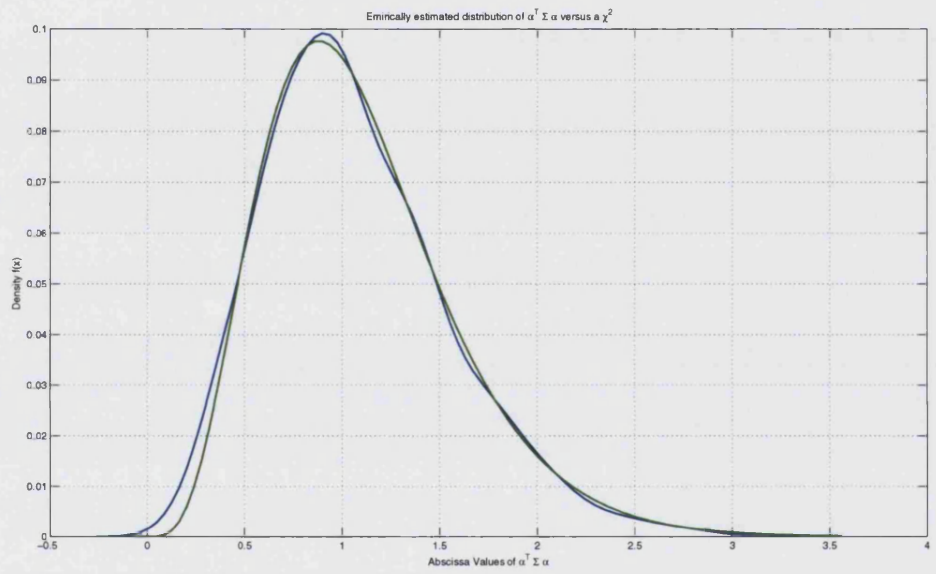

Figure 1.5: The empirically estimated distribution of $\alpha^{\mathrm{T}}\mathbf{\Sigma}\alpha$, blue, versus the anticipated $\chi^2(v)$ distribution, with appropriate degrees of freedom. This is taken from a simulation using 10,000 repeated experiments from a multi-normal data population, of length 10. The plots are very nearly convergent, demonstrate the Srivastava and Khatri proof, (1979).

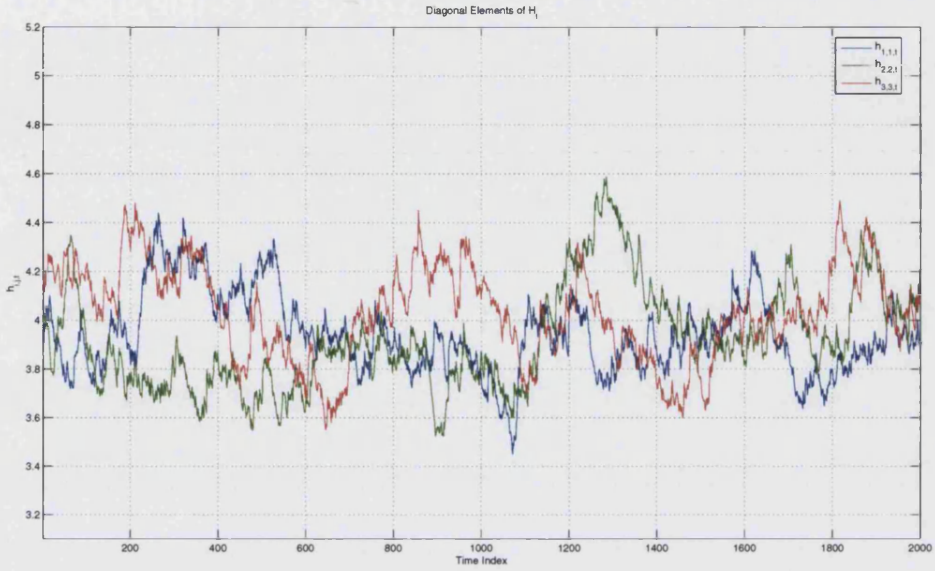

Figure 1.6: The evolution of the diagonal elements of the matrix process $\mathbf{H}_t$

Example of a Simple 3×3 Dynamic Wishart Matrix Process

Using the matrix process from 1.3.197, with $\mathbf{\Omega} = \mathbf{I}$ and $v = n + 1$, for a tri-variate system figures 1.6 and 1.7 illustrate the evolution of the diagonal and off diagonal elements.

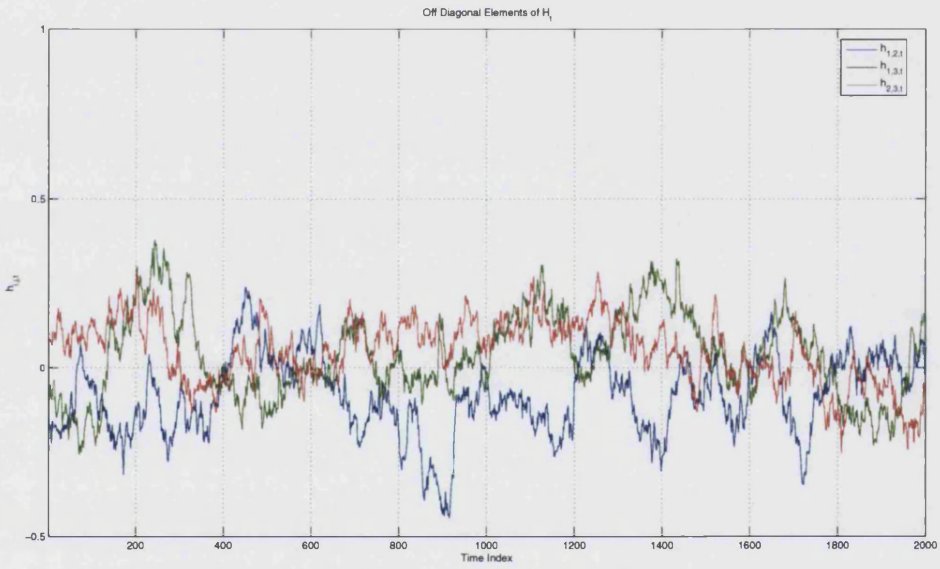

Figure 1.7: The evolution of the off-diagonal elements of the matrix process $\mathbf{H}_t$

1.3.10 Estimating a Wishart Autoregressive Process

The final part of this review of random matrix processes looks at estimating the model proposed in 1.3.197, consider an autoregressive matrix process whereby the

$$\mathbf{A} = ivech\left(\underset{\frac{1}{2}n(n+1)\times 1}{\mathbf{a}}\right) ivech\left(\underset{\frac{1}{2}n(n+1)\times 1}{\mathbf{a}}\right)^{\mathrm{T}} \tag{1.3.199}$$

$$\mathbf{\Sigma}_t \sim W(\mathbf{\Omega}, v) \tag{1.3.200}$$

$$\theta = \left[\mathbf{a}^{\mathrm{T}}, v\right]^{\mathrm{T}} \tag{1.3.201}$$

Therefore the maximum likelihood problem maybe specified in terms of the following constrained optimization problem,

$$\log\left(\mathfrak{L}\left(\theta^*\right)\right) \triangleq \max_{\theta}\left(\log\left(\prod_{t=1}^{\tau}\mathfrak{F}\left(\mathbf{H}_t - \mathbf{A}\times\mathbf{H}_{t-1}\,|\theta\right)\right)\right) \tag{1.3.202}$$

$$\equiv \max_{\theta}\left(\sum_{t=1}^{\tau}\log\left(\mathfrak{F}\left(\left(\mathbf{H}_t - \mathbf{A}\times\mathbf{H}_{t-1}\right)|\theta\right)\right)\right) \tag{1.3.203}$$

Where $\mathfrak{F}$ is the Wishart density function, from 1.3.191. Kollo and Van Rosen (2005) [170] and Srivastava (2004) [241] show that $\mathfrak{F}$ is twice differentiable and as such should have a global maxima for most problems. Using the vector of likelihood scores the expectation of the Hessian maybe computed, and the diagonal of the Cholesky factorization of this matrix provides the model standard errors.

1.3.11 Empirical Example: Estimating A Wishart Auto-regression of the Conditional Covariation of a Vector of Equity Returns

Consider a subset of raw equity returns, $\mathbf{y}_t$, chosen at random from the S&P 500 cross section. Assuming that the quadratic covariation of the returns is a Wishart Autoregressive process, then it is possible to postulate the following auto-regressive model of instantaneous autoregressive covariation,

$$\mathbf{y}_t = \mathbf{H}_t^{\frac{1}{2}} \varepsilon_t \tag{1.3.204}$$

$$\mathbf{H}_t = \mathbf{A} \times \mathbf{H}_{t-1} + \mathbf{\Sigma}_t \tag{1.3.205}$$

$$\mathbf{A} = ivech\left(\mathbf{a}\right) ivech\left(\mathbf{a}\right)^{\mathrm{T}} \tag{1.3.206}$$

$$\mathbf{\Sigma}_t \sim W\left(\mathbf{\Omega}, v\right) \tag{1.3.207}$$

Where $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{A}$ is non-negative definite, via $\mathbf{A} = ivech\mathbf{a}\mathbf{a}^{\mathrm{T}}$. Of course by not including all the other elements in the market, the model is incomplete however for practical purposes this model will be restricted to a 10-variate system. The price and return processes for the chosen assets are as follows, to visualize the variability in quadratic covariation, the evolution of the Eigenvector $\mathbf{zeta}_t$ of the matrix formed by the outer product $\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}$ is illustrated in figure 1.9, the largest Eigenvalue clearly dominates the rest, this is indicative of a very dense matrix.

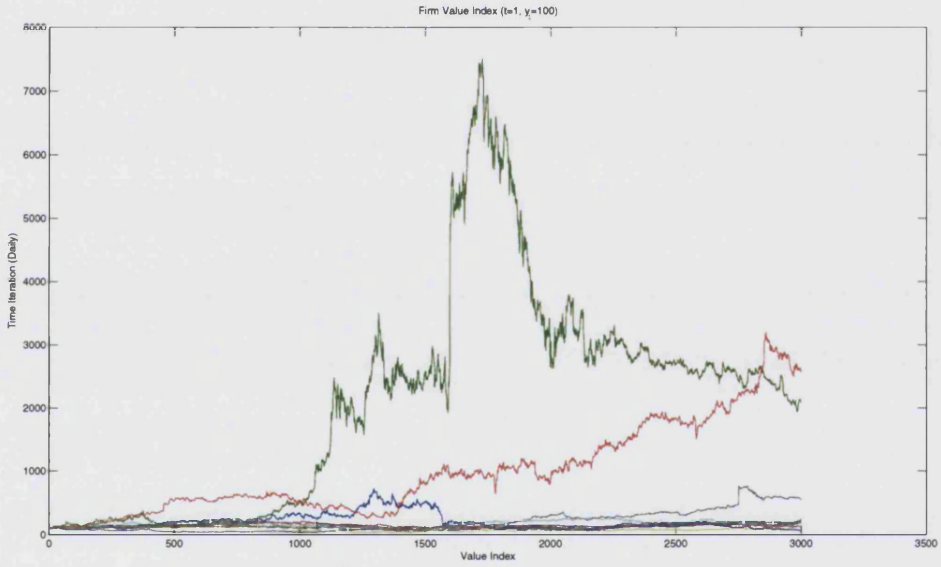

Figure 1.8: The Indexed value of the randomly chosen stocks from the S&P 500.

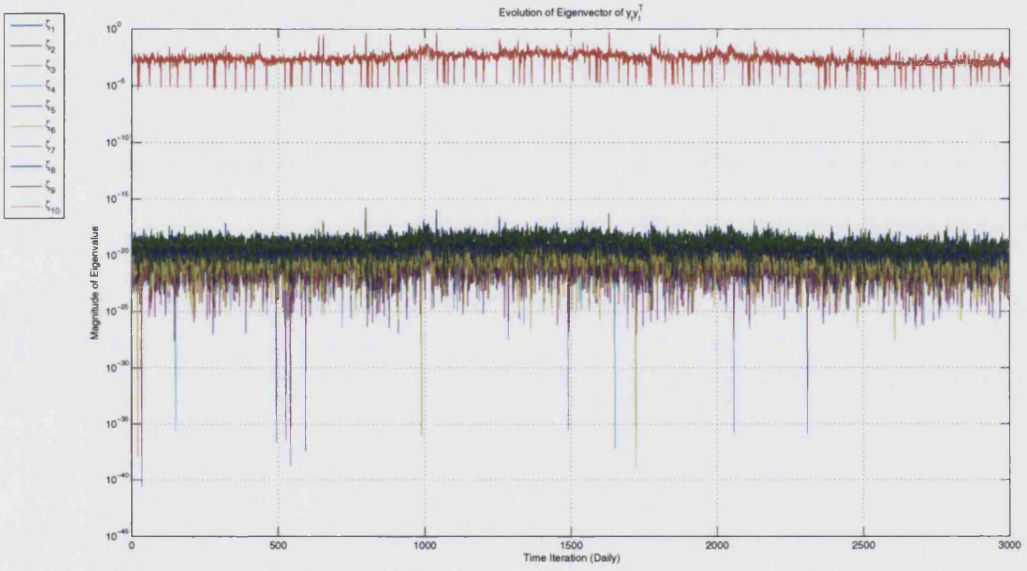

Figure 1.9: The Evolution of the Eigenvector of the instantaneous Quadratic Covariation of $\mathbf{y}_t$

Concluding Remarks on the Wishart Distribution and RMT

The most comprehensive review of the Wishart Distribution and its properties are in Kollo and von Rosen (2005) [170] and more formally in Muirhead (1982) [202], basic financial applications of the Wishart Auto-regression model maybe found in Gourieroux and Sufana (2004,2005) [118] and Gourieroux, Jasiak and Sufana (2004) [116]. The Wishart Distribution is also useful in testing the uniqueness of estimated covariance matrices, several tests based around the characteristics of the distribution to test for example an estimated covariance matrix $\mathbf{\Sigma} \approx \mathbf{I}$, which is very useful in testing the effectiveness of dimension reduction techniques such as principle component and factor analysis methods. Srivastava (2004) [241] reviews the Wishart distribution in association with a variety of variance reduction techniques. The approach to the derivation of 1.3.8 is the foundation to the approach taken to understand the asymptotic convergence of the most common matrix processes found in finance, the MV-GARCH family, to which the next section is devoted.

1.4 A Brief Review of Continuous Time Models and Applications

Up until a series of articles by Merton in the 1970s the finance/economic time series literature was dominated by discrete time dynamics. Discrete time systems have a great deal to offer. First and foremost they are easy to characterize and almost all data is discrete data. Second they exhibit memory effects which are of great importance in the finance realm. Third discrete iterative systems can exhibit complex dynamics which offer an explanation for a variety of observed patterns. However continuous time stochastic models have emerged from their use in thermodynamics and signal processing to become an extremely useful tool in financial engineering. This review focuses on the contribution of continuous time models in the asymptotic theory of discrete time models and the interrelationship between discretized continuous time data generating processes and observed data. This section briefly recaps some of the major articles and reviews the continuous time literature stating some of the major results. The first sub-section reviews some of the underlying probability theory and notation. The second section reviews the important finance literature on continuous time stochastic processes from 1970 to the present innovations in this field of research.

1.4.1 A Brief Overview of the Existing Continuous Time Literature

The basic split in the literature on continuous time models divides along the lines of pure mathematical concepts versus practical financial approaches. Karatzas and Shreve (1987) [162], Rogers and Williams (1980) [226] and Oksendale (1990) [212], present in descending order of complexity approaches to stochastic integrals and Brownian motion. These focus on the mathematical prerequisites and the financial uses emerge as an ancillary set of tools in addition to the technical concepts. The major financial tools such as optimal stopping and martingale measures are developed in detail in Musiela and Rutkowski (2002) [203], which presents very detailed coverage of arbitrage free pricing theory. Moving into the pure finance arena the literature again subdivides into two major themes, those dealing with the modelling of fixed income products and credit risk and those dealing with equity based modelling. Shreve (2000) [237] covers in some detail the modelling of both of these areas of research and builds on the mathematical concepts developed in Karatzas and Shreve (1987) [162]. Glassermann (2001) [109], Johannes and Polson (2003) [155], Yacine-Sahalia (2004,2005,2006) [3] extensively cover the relationship between continuous time models and discretely sampled data. Glassermann and Johannes and Polson focusing on monte-carlo methods and stratification of parameterized models. Glassermann covers a broad literature on monte-carlo methods and focuses on

methods utilizing Latin hypercube stratification. Yacine-Sahalia (2005) [3] focuses on determining the appropriate diffusion model to apply to continuous time data and looks at the difficult issue of parameterizing generic stochastic volatility models.

The Continuous Time Equity Based Literature

Away from the general modelling approaches the most eclectic literature is based around the specification and parameterization of equity diffusion models. The pricing of contingent claims of these models has been the major focus of the applied finance literature for the last thirty years. In these sections option pricing models and hedging strategies will be addressed, however the major focus will be on the dynamics of the price evolution model and the implications for the parameterization of these models. The basic Black and Scholes (1973) [31]model, treats the stock price dynamics as a simple geometric brownian motion, described by a stochastic differential equation. The resultant log-normal process has been the focus of considerable attention in the financial econometrics literature and a considerable literature, starting with Mandelbrot (1963) has developed rejecting this simple dynamics. However the attractiveness of the model, i.e. that it is a Martingale produces an arbitrage free pricing formula for European options, (those with a fixed strike date and no early exercise), has meant that it is still used extensively, despite the rejection of the core price process model. A major innovation has been in the pricing of American options by path-wise based methods. Beginning with McKean (1965) [158] who addressed the early exercise of

American options and the concept of optimal stopping through to very sophisticated treatment in Bank and Follmer (2003) [16]. The essential result is that an American Call option with constant strike price written on a non-dividend paying stock is equivalent to the corresponding European option. The difference, however, is in the early exercise conditions for American put options. American options are path-wise instruments and as such have no closed form expression, numerical techniques usually involve the discretization of the time parameter and the application of backward induction and simulation to find the price distribution of the option, examples of this approach maybe found in Geske and Johnson (1984) [105] and Geske and Roll (1984) [106].

Probability Spaces and Continuous Random Variables

This brief recap is extensively paraphrased from Williams (1991), Musiela and Rutkowski (2004) [203], Glasserman (2002) [109], Cyganowski *et al* (2002) [62], Steele (2000) [243], Musiela and Rutkowski (2004) [203] and Williams (2005) [263]. We mentioned very briefly during the generic derivation of the multivariate central limit theorem the formulation of a continuous time probability space, which describes functional likelihoods of diffusion process on a set of potential outcomes. We now formalize this notation as follows, consider a set Ω of possible co-ordinates in the k-dimensional Euclidean real space, $\mathbb{R}^k$, then for a set of functionals $\mathbb{F}$, is a σ-algebra which is a family of subsets of Ω, Øksendale (1997) [212], for any σ-algebra $\mathbb{F}$, on Ω, there is a

probability measure $\mathbb{P}$ such that $\mathbb{P} : \mathbb{F} \rightarrow [0,1]$, therefore for any element outside of Ω, $\emptyset$, the probability $\mathbb{P}(\emptyset) = 0$ and for the set of Ω, $\mathbb{P}(\Omega) = 1$. The tri-tuple $(\Omega, \mathbb{F}, \mathbb{P})$ is called the probability space. The probability space is said to be complete if $\mathbb{F}$, contains all of the subsets of Ω with $\mathbb{P}$ outer measure zero and therefore contains all possible zero probability elements of Ω. An important result from this notation is that for any function g whereby, $Y = g(X)$, then if $X, Y : \Omega \rightarrow \mathbb{R}^k$, then if $\mathbb{B}$ is the smallest subset of $\mathbb{F}$, for which, $\mathbb{P}(\mathbb{B}) = 1$, $\mathbb{B}$ is the Borel σ-algebra on Ω, as such a function g, is Borel measurable if $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$. Every random variable induces a probability measure $\mathfrak{F}$, on $\mathbb{R}^k$, where

$$\mathfrak{F}(\mathbb{B}) = \mathbb{P}\left(X^{-1}(\mathbb{B})\right) \tag{1.4.1}$$

Therefore $\mathfrak{F}$ is the distribution of X, furthermore, the expectation of $E(X)$ is,

$$E(X) := \int_{\Omega} X(u)\, d\mathbb{P}(u) = \int_{\mathbb{R}^k} x d\mathfrak{F}(x) \tag{1.4.2}$$

For any space $t \in T$, a parameterized collection of random variables is called a k-dimensional stochastic process assuming that $(X_t)_{t \in T}$ defined on some probability space, $(\Omega, \mathbb{F}, \mathbb{P})$, with co-ordinates in $\mathbb{R}^k$, the parameter space T is usually in the halfline, $[0, \infty)$. The conventional notation uses a variable u, which is the path of X_t, therefore,

$$u \rightarrow X_t(u) \equiv X_u \tag{1.4.3}$$

Therefore the evolution of a random process X may be thought of in terms of the evolution in time of some continuous set of *experiments* u.

1.4.2 Multivariate Itô and Stratonovich Integrals and Martingale Representations

The final part of this short review of the underlying mathematics of continuous time processes covers the multidimensional Itô formulation and introduces the global and local martingale measure. Stochastic calculus was introduced by the Japanese mathematician K. Itô in the 1940s, following Einstein's interpretation of Brownian motion forty years earlier. He found that certain processes most notably Wiener Brownian motions exhibited a smooth auto-covariance pseudo spectral density. The Itô integral approach is to solve the following stochastic integrals,

$$\mathbf{x}_u = \mathbf{x}_0 + \int_0^T \mu\left(t, \mathbf{x}_u\right) du + \int_0^T \sigma\left(t, \mathbf{x}_u\right) d\mathbf{w}_u \tag{1.4.4}$$

The Riemann-Steitjes integral for such an integral would normally be as follows, for a function $\sigma : [0,T] \times \Omega \to \mathbb{R}^k$, assuming that σ is continuous in t, for $u \in \Omega$, as follows,

$$S_{(T,\sigma)} = \sum_{j=1}^{n} \sigma\left(\tau_j^{(n)}, u\right)\left(\mathbf{w}_{t_{j+1}^{(n)}}(u) - \mathbf{w}_{t_j^{(n)}}(u)\right) \tag{1.4.5}$$

for any evaluation point, $\tau_j^{(n)} \in \left[t_j^{(n)}, t_{j+1}^{(n)}\right]$ with partitions, $0 = t_1^{(n)} < t_2^{(n)} < \ldots < t_{n+1}^{(n)} = T$, however with the inclusion of the Wiener Brownian motion term the limit, $\delta^{(n)} = \max_{1 \le j \le n}\left(t_{j+1}^{(n)} - t_j^{(n)}\right) \to 0$, as $n \to \infty$ does not exist given the sample path

irregularity, therefore there is no *path-wise* convergence limit. However there may be mean-square convergence of the pathway for a specific class of functions, therefore for an explicit set of evaluation points, $\tau_j^{(n)} = (1-\lambda)\, t_j^{(n)} + \lambda t_{j+1}^{(n)}$, the convergence of $\sigma(t,u) = W_t(u)$ will be,

$$\tfrac{1}{2}W_T^2 - \left(\tfrac{1}{2} - \lambda\right) T \tag{1.4.6}$$

Setting $\lambda = 0$, i.e. as the lower evaluation bound, the summation becomes tractable,

$$\sum_{j=1}^{n} W_{\tau_j}\left(W_{t_{j+1}} - W_{t_j}\right) = -\tfrac{1}{2}\sum_{j=1}^{n}\left(W_{t_{j+1}} - W_{\tau_j}\right)^2 + \tfrac{1}{2}\sum_{j=1}^{n}\left(W_{\tau_{j+1}} - W_{t_j}\right)^2 + W_T^2 - W_0^2 \tag{1.4.7}$$

therefore under expectations the properties of the stochastic function are,

$$E\left(\int_0^T W_t dW_t\right) = 0 \tag{1.4.8}$$

$$E\left(\left|\int_0^T W_t dW_t\right|^2\right) = \int_0^T E\left(|W_t|^2\right) dt = \tfrac{1}{2}T^2 \tag{1.4.9}$$

Which yields the famous Itô stochastic integral,

$$\int_0^T \sigma(t,u)\, dt = ms - \lim \sum_{j=1}^{n} \sigma\left(t_j^{(n)}, u\right)\left(\mathbf{w}_{t_{j+1}^{(n)}}(u) - \mathbf{w}_{t_j^{(n)}}(u)\right) \tag{1.4.10}$$

The Stratonovich Derivation

By contrast to the Itô form of the stochastic integral, the Stratonovich form uses the mid point of the partition interval, $\tau_j^{(n)} = \frac{1}{2}\left(t_j^{(n)} + t_{j+1}^{(n)}\right)$, for a given function with

convergent mean squared pathways, the Stratonovich derivation under the Riemann-Steitjes summation yields,

$$\int_0^T \sigma(t,u) \circ dW_t(u) = \lambda - \lim \sum_{j=1}^{n} \sigma\left(\tau_j^{(n)}, u\right)\left(\mathbf{w}_{t_{j+1}^{(n)}}(u) - \mathbf{w}_{t_j^{(n)}}(u)\right) \qquad (1.4.11)$$

Under the Stratonovich approach the integral, $\int_0^T W_t dW_t = \frac{1}{2}W_t^2 - \frac{1}{2}T$ is now altered to $\int_0^T W_t \circ dW_t = \frac{1}{2}W_t^2$, where $\circ$ is used to symbolically represent the integration via the Stratonovich stochastic integral equation. We can now rewrite the general stochastic integral in Stratonovich notation as,

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^T \mu(t, \mathbf{x}_u)\, du + \int_0^T \sigma(t, \mathbf{x}_u) \circ d\mathbf{w}_u \qquad (1.4.12)$$

Solving Stochastic Differential Equations

A brief explanation is now required to suggest under which circumstances either of these two integrals should be used. In general the properties of the Stratonovich integral lend themselves to tractably analytic solutions, for example if the function σ is a smooth "high signal" function as deterministic tricks maybe used to solve Stratonovich type differential equations, in general these are problems which involve signal extraction and filtration, Stratonovich SDEs are used extensively in signal processing and thermodynamics. The Itô integral lends itself to most general stochastic problems whereby the very nature of the noise process is the most prescient element of the analysis. In general in financial engineering solutions based on the Itô integral

are the most commonly used when signal extraction is not critical, but an understanding of the potential richness of the sample paths is, i.e. an understanding of the dynamics of the volatility of a process, see Cyganowski (2002) [62], Musiela and Rutkowski (2002) [203] and Rebonato (2002) [223] for more discussion on this area.

1.4.3 Stochastic Differential Equations and Stochastic Integrals

The motivation of reviewing this area of the literature is in defining a fixed set of rules that cover a variety of functional forms which maybe defined in conjunction with an n-dimensional Brownian motion object. Let $\mathbf{w}(t)$ be the vector of zero centred Brownian motions $\mathbf{w} = [W^1, W^2, ..., W^n]^{\mathrm{T}}$ where $W^i \sim N(\mathbf{0}, \sigma_i^2)$ feasibly valued on $\mathbb{R}^k$ and there is a filtered probability space $(\Omega, \mathbb{F}, \mathbb{P})$ such that $\mathbf{w}$ is a vector of continuous random variables each drawn from an independent and identically distributed vector of one dimensional Brownian motions, thus for any process γ that is an $\mathbb{R}^n$ valued process the following condition applies,

$$\mathbb{P}\left\{\int_0^T |\gamma_u|^2 \, du < \infty\right\} = 1 \tag{1.4.13}$$

Where $|\cdot|$ stands for the Euclidean norm in $\mathbb{R}^n$. We are therefore dealing with processes that exist exclusively within a framework that allows for symmetric covariant metrics. The general form of the Itô stochastic integral of γ with respect to $\mathbf{w}$, is

given as:

$$I_{\mathrm{t}}(\gamma) = \int_0^t \gamma_u . d\mathbf{w}_u \tag{1.4.14}$$

Now consider a k dimensional random vector $\mathbf{x}$ where $\mathbf{x} = [X^1, X^2, ..., X^k]^{\mathrm{T}}$ and let the i^{th} process of $\mathbf{x}$ be defined as,

$$X(t)^i = X_0^i + \int_0^t \alpha_u^i du + \int_0^t \beta_u^i dW_u \tag{1.4.15}$$

where α^i are real value adapted processes a and β^i are $\mathbb{R}^n$valued processes. For a given function $g = g(\mathbf{x}, t)$ the projection $g : \mathbb{R}^n \times [0, T] \to \mathbb{R}$ must implicitly hold. To incorporate the broader form of covariation, consider that the function g is in a broader class of functions $\mathcal{C}(\mathbb{R}^n, \mathbb{R})$. We can define the vector process in the Itô form as

$$dg(\mathbf{x}(t)) = \sum_{i=1}^k g_{x_i}(\mathbf{x}(t))\,\alpha(t)^i + \sum_{i=1}^k g_{x,i}(\mathbf{x}(t))\,\beta(t)^i . d\mathbf{w}(t) + \tfrac{1}{2}\sum_{i,j=1}^k g_{x_i,x_j}(\mathbf{x}(t))\,\beta(t)^i . \beta(t)^j dt \tag{1.4.16}$$

More generally if the processes X_i are in $\mathcal{S}^c(\mathbb{R})$ for $i = 1, ..., k$ then

$$g(\mathbf{x}(t)) = g(\mathbf{x}_0) + \sum_{i=1}^k \int_0^t g_{x_i}(\mathbf{x}_u)\, dX_u^i + \tfrac{1}{2}\sum_{i,j=1}^k \int_0^t g_{x_i,x_j}(\mathbf{x}_u) d\left\langle X^i, X^j \right\rangle_u \tag{1.4.17}$$

Where $\mathcal{S}^c(\mathbb{R})$ is the space of continuous cross-variation of the vector of k martingales. We can therefore show that the derivative vector is as follows,

$$dg(\mathbf{x}(t)) = \sum_{i=1}^k g_{x_i}(\mathbf{x}_u)\, dX(t)^i + \tfrac{1}{2}\sum_{i,j=1}^k g_{x_i,x_j}(\mathbf{x}_u)\, d\left\langle X^i, X^j \right\rangle_u \tag{1.4.18}$$

Consider a new noise vector $\mathbf{z}$ again real valued on$(\Omega, \mathbb{F}, \mathbb{P})$ that constitutes a set of Brownian motions drawn from multivariate normal distribution. The relationship

between $\mathbf{z}$ and the independent vector Brownian motion process $\mathbf{w}$ can be expressed via a matrix $\mathbf{\Sigma}$, i.e.$\mathbf{z} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{w}$ where $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}$ .Where $\mathbf{\Sigma}$ is a non-negative definite Hermitian matrix and $\mathbf{A}$ is a lower triangular matrix. Then $\mathbf{\Sigma}^{\frac{1}{2}}$ is the upper symmetric square root (Cholesky factorization) of $\mathbf{\Sigma}$. Therefore we can consider a subset of independent one dimensional processes that have zero co-variation and a transformation such that the transformed vector of these Brownian motions exhibits a fixed multi-normal quadratic co-variation, this concept is covered in significant detail in Shreve (2004) [237], covers this in considerable detail and this result is used extensively in chapter 8.

1.4.4 Multivariate Finance in Continuous Time: Stochastic Volatility Models

Continuous time models are used in both the academic and practitioner domains to model asset prices, exchange rates, interest rates and many macro-economic factors. It is interesting to note that continuous time models are fairly recent additions to the finance literature, generally originating from early 1970s, and two major contributions to the literature. In his article "Optimum consumption and portfolio rules in a continuous time model", Merton (1971) [197], effectively introduced continuous time stochastic processes into a financial context and in effect created the discipline of financial engineering. Black and Scholes (1973) [31] followed with their formulation of option prices and contingent claims. Merton (1991) [197] is the classic references for

the continuous time literature, in Sahalia and Johansen (2002) [1], several chapters are exclusively dedicated to continuous time processes, these include parameterization of continuous time models using discrete data in Ait-Sahalia (2006) [1], analysis of Markov Chain Monte-Carlo routines and Bayesian updating in continuous time models in Johannes and Polson (2003) [155] and non-stationary continuous time functions in Bandi and Philips (2002) [15]. This brief list shows that like a stochastic process the literature has taken a rich variety of paths. The answer to the question of as to why continuous time models are so important, is based around their analytical tractability. For almost any problem which requires optimal decision making based on some stochastic process, continuous time models are usually at the basis of the solutions to these problems. The major problems of continuous time models are the challenges of econometric estimation and empirical implementation, see Bandi and Philips (2002) [15]. However very recent advances in econometric theory in particular the novel use of GMM models in parameterizing continuous time models from discretely sampled data see Hansen and Scheinkman (1995) [138], Duffie and Glynn (1997) [78] and new approaches in maximum likelihood, see Yacine-Sahalia (2002) [1] and (2006) have revolutionized the use of even the most complex continuous time models. The most basic continuous time stochastic differential equation is defined as follows,

$$dS_t = \mu dt + \sigma dW_t \tag{1.4.19}$$

Here dS_t is a continuous time stochastic gradient described by a stochastic differential equation (SDE), on the real line, $\mathbb{R}$, between, [0,T], S_t is usually considered to be the logarithm of an asset price. W_t is a one dimensional brownian motion whereby each innovation of W_t is a Gaussian process, i.e. W_t is drawn from a continuous normal distribution. In general μdt is usually described as the drift term and σdW_t is described as the diffusion term. Logical extension of this simple process into the multivariate domain is as follows,

$$d\mathbf{x}_t = \mu dt + \sigma d\mathbf{w}_t \tag{1.4.20}$$

where $\mathbf{x_t}$ is is a $k+1$ length vector of continuous stochastic processes in the $k+1$-dimensional real Euclidean Space, where $\mathbf{x}_t$ is the *state* of the system in $\mathbb{R}^{k+1}$, $\mathbf{w}_t$ is a k length vector of continuous brownian motions drawn from a (multi)-normal distribution, and μ and σ are appropriately sized vectors or matrices, which reshape the individual terms to a k length column vector. In fact relaxing the constant drift and diffusion terms and formulate,

$$d\mathbf{x}_t = \mu\left(\mathbf{x}_t \left| \theta \right.\right) dt + \phi\left(\mathbf{x}_t \left| \theta \right.\right) d\mathbf{w}_t \tag{1.4.21}$$

where $\mu\left(\mathbf{x}_t \left| \phi \right.\right)$ and $\sigma\left(\mathbf{x}_t \left| \theta \right.\right)$ are vector/matrix functions of the current realization of $\mathbf{x}_t$, with parameter vector θ.

1.4.5 Stochastic Volatility Models

The state space vector $\mathbf{x}_t$, encapsulates the stock price and the volatility components in one vector equation, defining the vector stochastic differential equations as follows,

$$d\mathbf{x}_t = d\begin{bmatrix} \mu^s(\log(S_t)) \\ \mu^\sigma(\mathbf{y}_t) \end{bmatrix} dt + \begin{bmatrix} \phi^s(S_t, \mathbf{y}_t) & \\ \mathbf{0} & \phi^\sigma(\mathbf{y}_t) \end{bmatrix} d\begin{bmatrix} W_t^s \\ \mathbf{w}_t^\sigma \end{bmatrix} \tag{1.4.22}$$

Yields a generic stochastic volatility model, where $\mathbf{y}_t \in \mathbb{R}^k$ is the vector process of volatility components, and $\log(S_t) \in \mathbb{R}$ is the price process, the functions μ^s, μ^σ, ϕ^s and ϕ^σ are continuously differentiable and $\mu^s : \mathbb{R} \to \mathbb{R}$, $\mu^\sigma : \mathbb{R}^k \to \mathbb{R}^k$, $\phi^s : \mathbb{R}^{k+1\times 1} \to \mathbb{R}^{1\times k+1}$ and $\phi^\sigma : \mathbb{R}^{k\times 1} \to \mathbb{R}^{k\times k}$.

1.5 Concluding Remarks

This chapter has reviewed some of the current developments in vector processes in both discrete and continuous time. The chapter has also demonstrated several generic techniques such as the vector ARMA, ARMAX and state space Kalman filters and introduced a new type of matrix random effects model and demonstrated its use in modelling the dependency in international stock markets, see chapter appendix. These results form the basis for the second moment models in 5 and 6. The continuous time section introduces the generic stochastic volatility model and its relationship to standard vector processes and a set of results that will be extensively used in chapter 8.

1.6 Chapter Appendix: Kalman Filter Approach to Cross Market Contagion

Using the Multivariate Kalman filter approach a dynamic model of the same nine international financial markets is estimated. The model attempts to capture *in mean* cross variation between the financial market. The following simple single output state space model is estimated via Gaussian maximum likelihood, the model is a basic market model with time varying coefficients. This type of model offers an alternative method to the MV-ARCH approach to understanding financial market integration. By appropriate choice of global market changes in relationships between individual country indices maybe elucidated in a simple and convenient manner.

$$r_{i,t} = \psi_t^{\mathrm{T}} \mathbf{x}_t + u_t \tag{1.6.1}$$

$$\mathbf{x}_t = [r_{m,t}, 1]^{\mathrm{T}} \tag{1.6.2}$$

$$\psi_t = \mathbf{\Phi}^{\mathrm{T}} \psi_{t-1} + \mathbf{v}_t \tag{1.6.3}$$

where x_t, is the returns from the global stock market at t, i.e. and a constant to represent the excess returns. The algorithm used is the Kalman filter approach based on an algorithm by Grewal and Andrews (2001) [122] who present a convenient method for a recursive Kalman filter approach to this type of model specification. The global index is generated using a value weighted index of more than 20,000 equities from the nine markets, mostly computed using the market value of the index. The value and return series are as follows. The results show that the markets generally

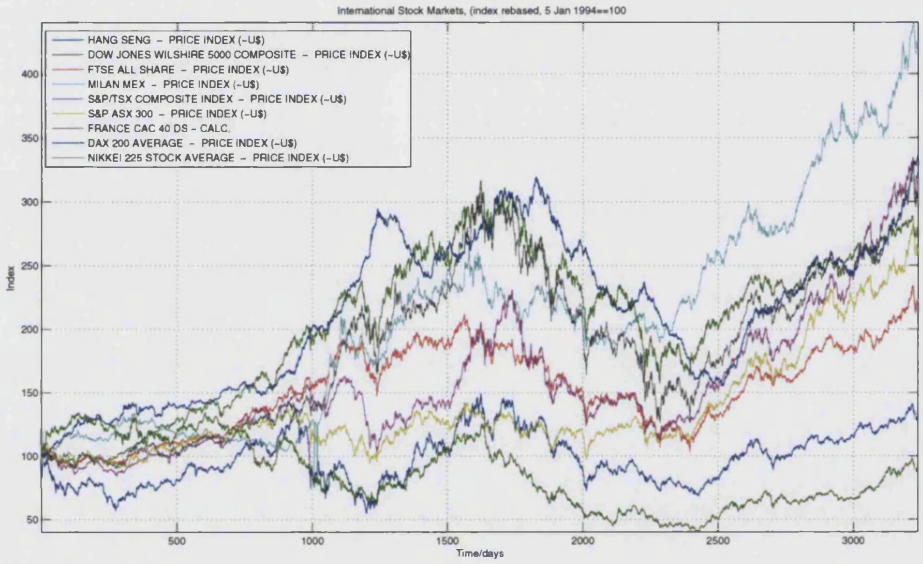

Figure 1.10: Re-based index for nine international stock markets, daily data from 05 Jan 1994 to 05 June 2006, where 05 Jan 1994 == 100, each of the indices listed has to contain at least 70% of the publicly listed companies in that country. Note 1: In keeping with trends with regards to mechanisms for raising capital, the number of components ranges from the DOW Wilshire 5000 and the FTSE All Share to the less well represented CAC 40 and MIB 30 indices for France and Italy respectively. Note 2: that the Milan Stock exchange value has be interpolated on the 19th June 2000, 3rd July 2000 and 19th July 2000, due to significantly erroneous data points. Note 3: The Hang Seng is the most unbalanced exchange in terms of company weighting, with HSBC holding accounting for on average 30% of the total index value.

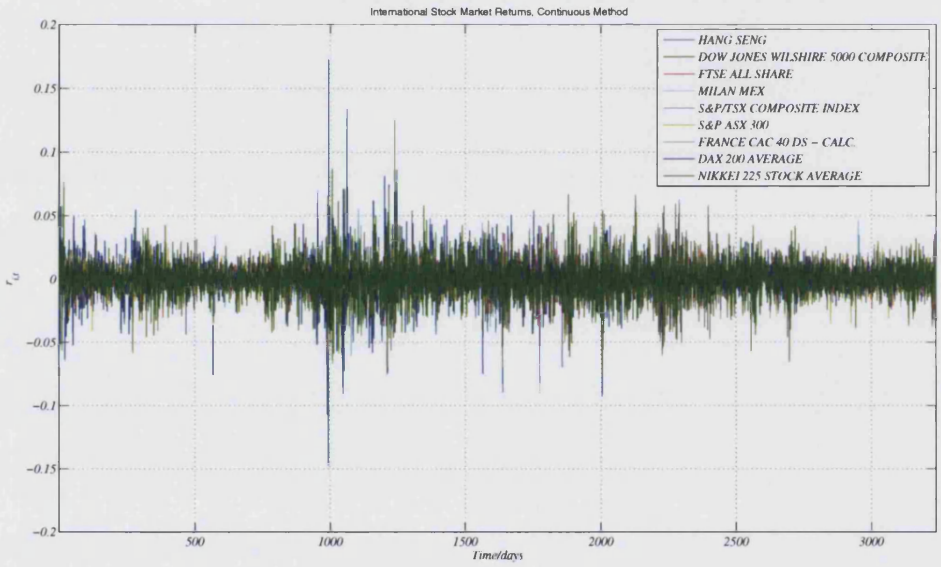

Figure 1.11: Daily International Stock Market Returns for nine global indices, returns computed using continuous method, $r_{i,t} = \log(p_{i,t}) - \log(p_{i,t-1})$, where $p_{i,t}$ is the index value of the i^{th} stock market index at time t

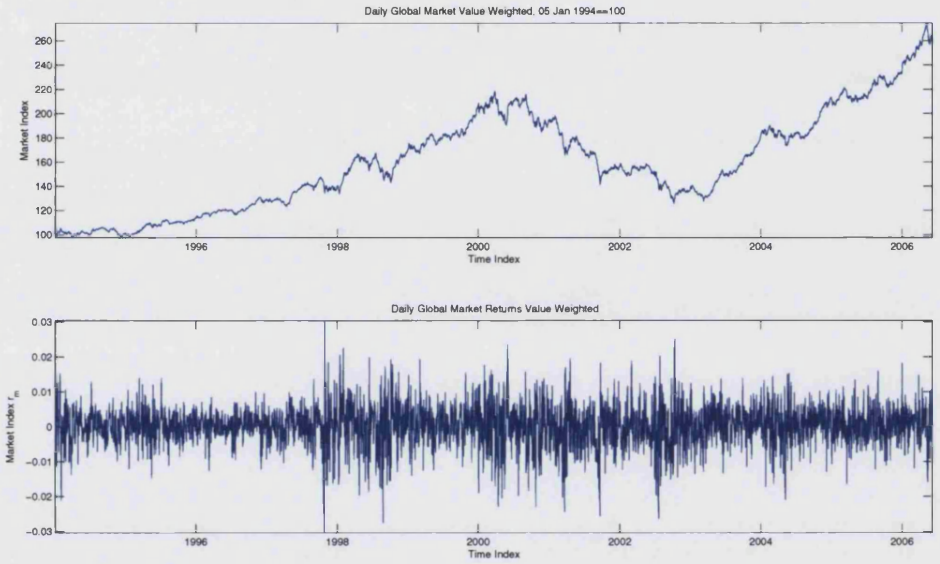

Figure 1.12: The Global Market Index, value weighted from more than 20,000, global equities. the value of S&P Wilshire 5000, (existing data daily 1974 - Present, this sample 05/01/1990 - 05/06/2006), which is the largest cross section of Equities in the states forms about $\frac{1}{4}$ of the stocks and around $\frac{1}{7}$ of the value weighting, which is in line with the NASDAQ Global Equity Market, which has available data from 1997 onwards. The market generally follows the widely accepted global business cycle, with the Hi Tech bubble leading to a correction at the end of 1999. the bull market reforms in 2003 and accelerates past the Hi Tech peak at the start of 2006. General analyst predictions (see Bloomberg^{TM}, Global Analyst report global market 2006) suggest another correction is due late 2006 early 2007, in response to high energy prices and interest rate rises in the US.

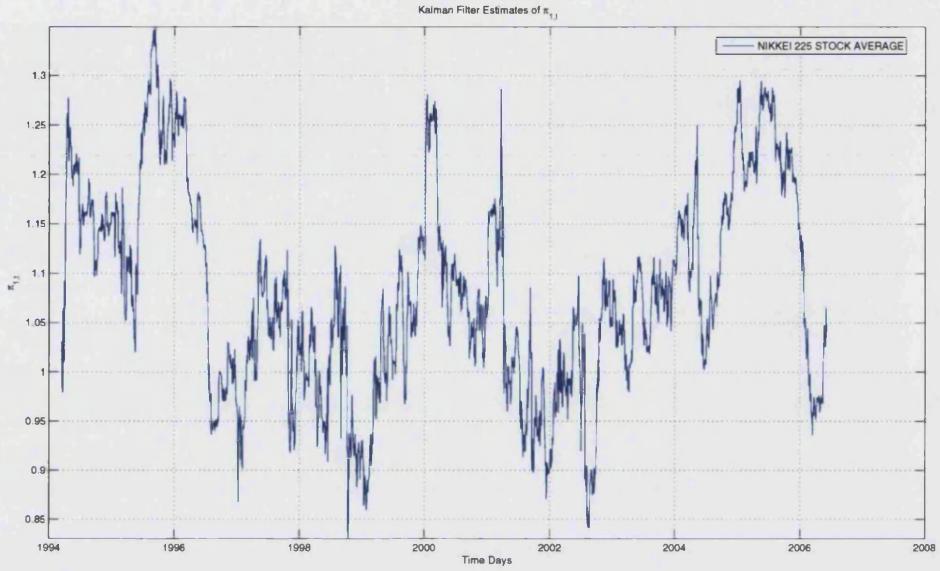

Figure 1.13: The Recursively Estimated $\pi_{2,t}$, coefficient for the broad base Nikkei of 225 Japanese Stocks, classified Japanese Titans. Over the sample period the NIKKEI 225 coefficient varies between 0.85 and 1.35, but in general is well above 1, indicating that the volatility of the NIKKEI is normally well above the global index.

converge to unity and that the excess returns normally vary around zero.

1.6.1 Evidence of Subsets within the Global Market

Sub groupings are a stylized facts in equity markets, the following overlays illustrate two observed groupings, the continental European indices and the Anglo saxon grouping of Australia, Canada the US and the UK.

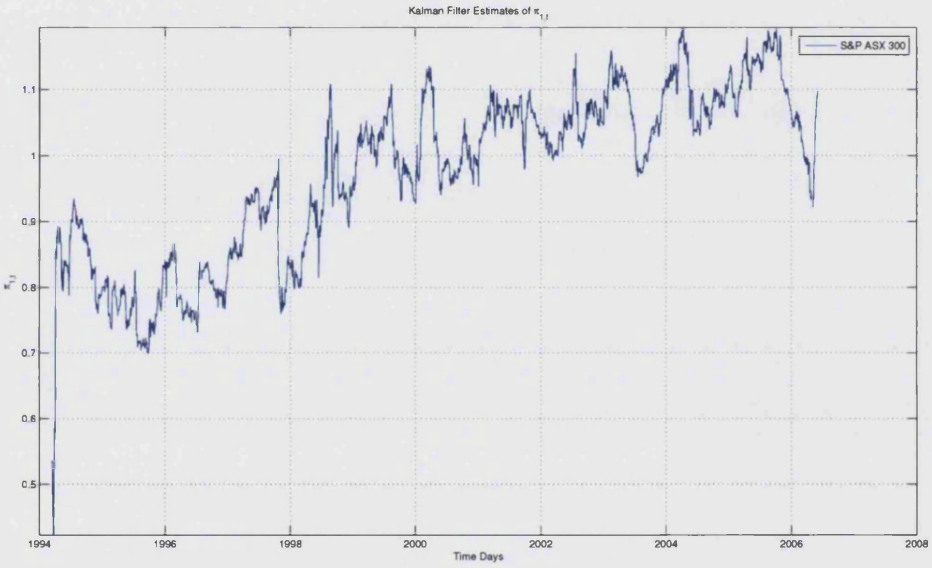

Figure 1.14: The Australian Stock Exchange, compiled by Standard and Poors, over the sample period the markets $\pi_{2,t}$ coefficient increases from 0.8 in 1994 to around 1.1 in 2006. The Asian crisis has some effect from late 1997 to mid 1998, however the index reverts to the long run trend rapidly.

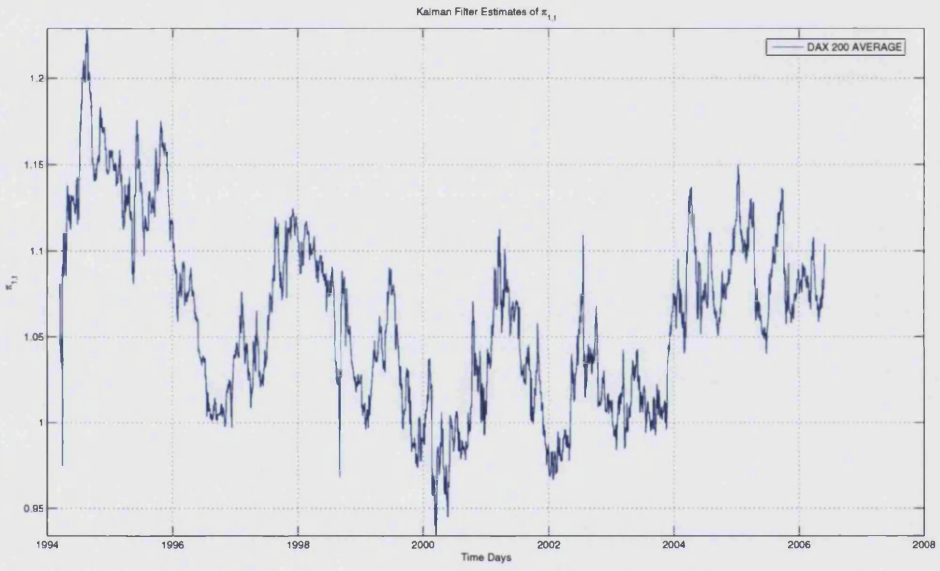

Figure 1.15: The DAX index of 200 publicly quoted German companies, is a semi stable mean reverting process around 1.05.

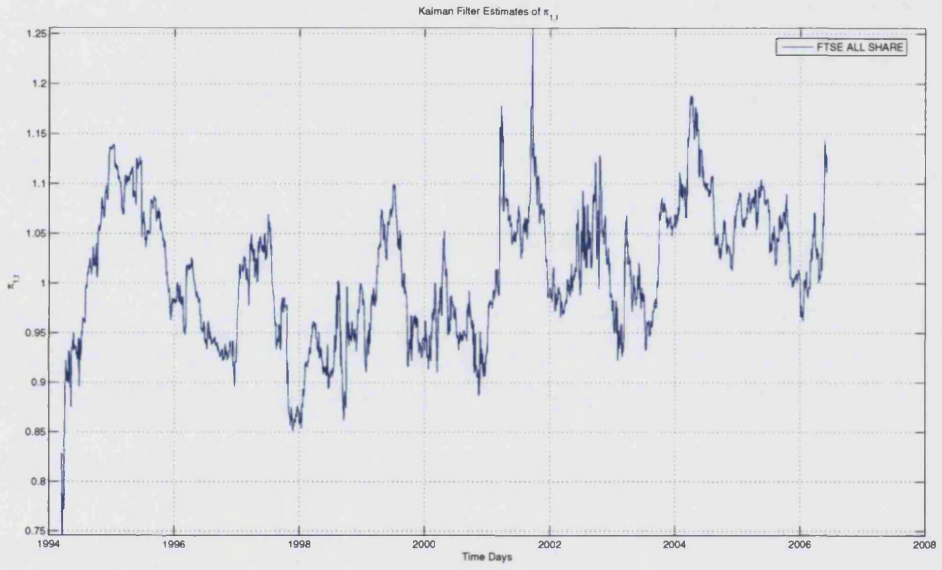

Figure 1.16: The Recursively Estimated $\pi_{2,t}$, coefficient for the FTSE all share of companies covering between 98-99% listed in the UK, (Source: FTSE.co.uk, 2006), the coefficient varies between 0.75 and 1.25 around the global market.

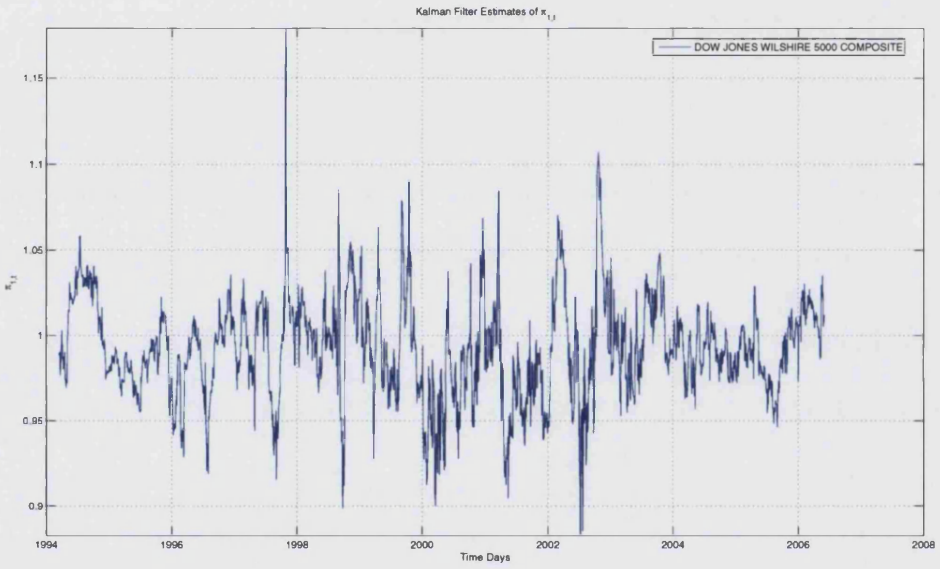

Figure 1.17: The DOW Wilshire 5000 Index, covers the largest cross section of US publicly listed companies. In keeping with US global dominance in terms of overall market capitalization, the DOW Wilshire 5000's $\pi_{2,t}$ coefficient is very tightly clustered around 1.

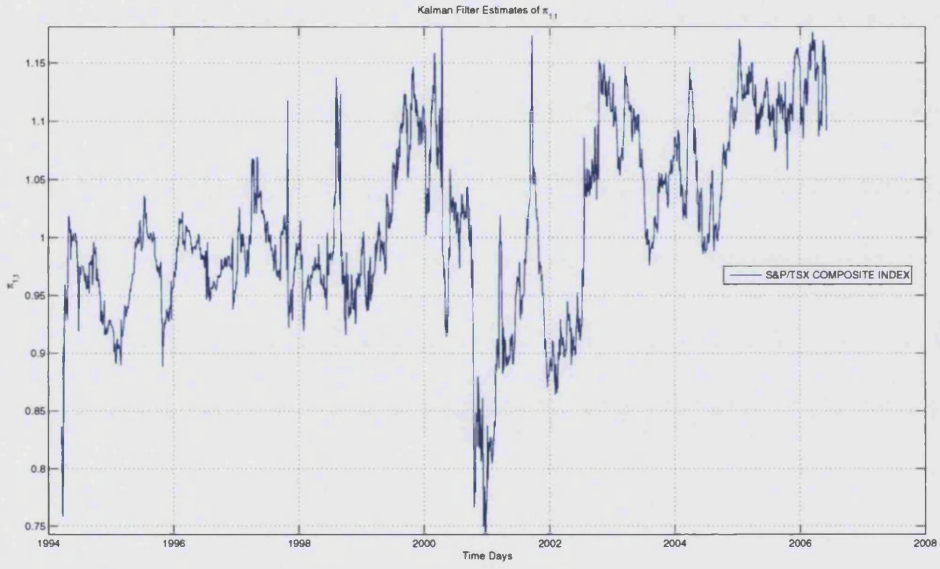

Figure 1.18: The TSX index of the 276 largest companies on the Toronto Stock Exchange, index is compiled by Standard and Poors and the $\pi_{2,t}$ coefficient whilst more variable than the DOW Wilshire 5000, it has less variability than the FTSE All Share.

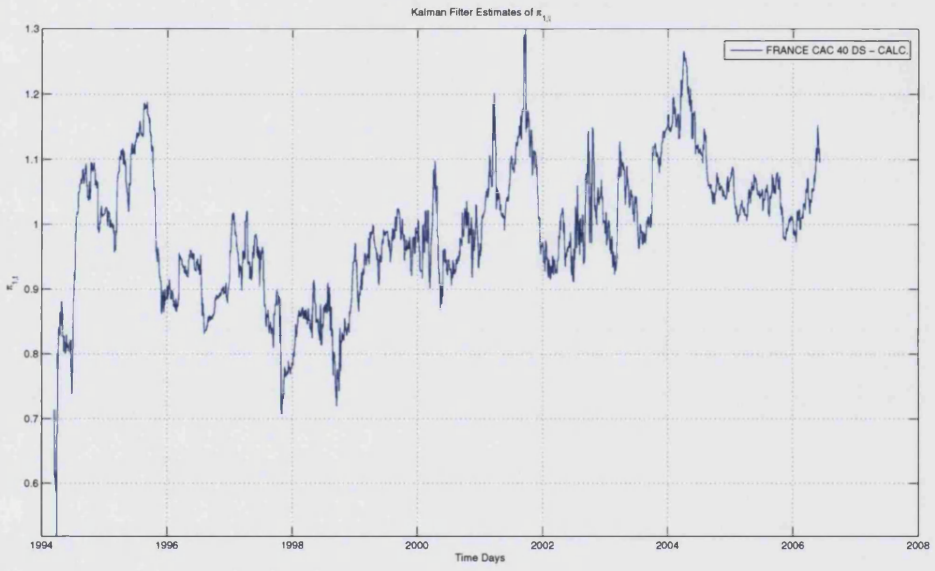

Figure 1.19: The CAC 40 of the largest companies listed on the Euronext Paris Exchange, the $\pi_{2,t}$ coefficient follows a similar pattern to the DAX in contrast to the FTSE All Share.

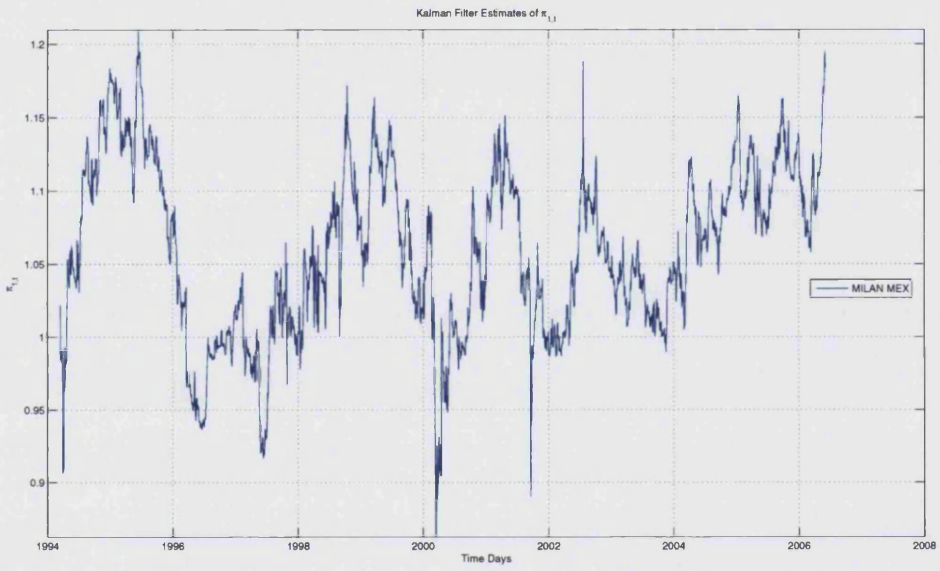

Figure 1.20: The MIB 30 of companies quoted on the Milan Stock Exchange, the $\pi_{2,t}$ coefficient follows very tightly to the pattern seen in the CAC and DAX indices.

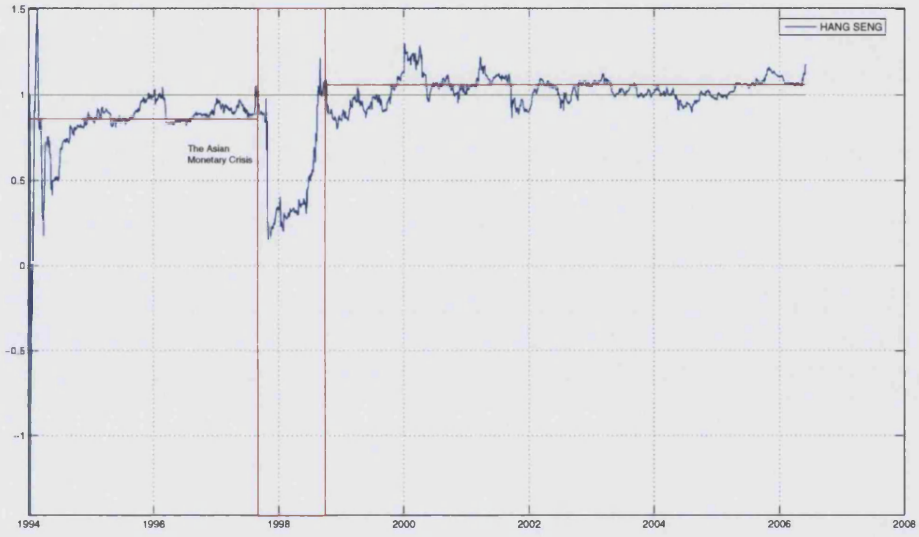

Figure 1.21: The Hang-Seng Index of 33 Companies floated on the Hong Kong Stock Exchange. The Asian monetary crisis that stretched from late 1997 to mid 1998, is clearly shown as a decoupling of the Hang Seng from the global equity index.

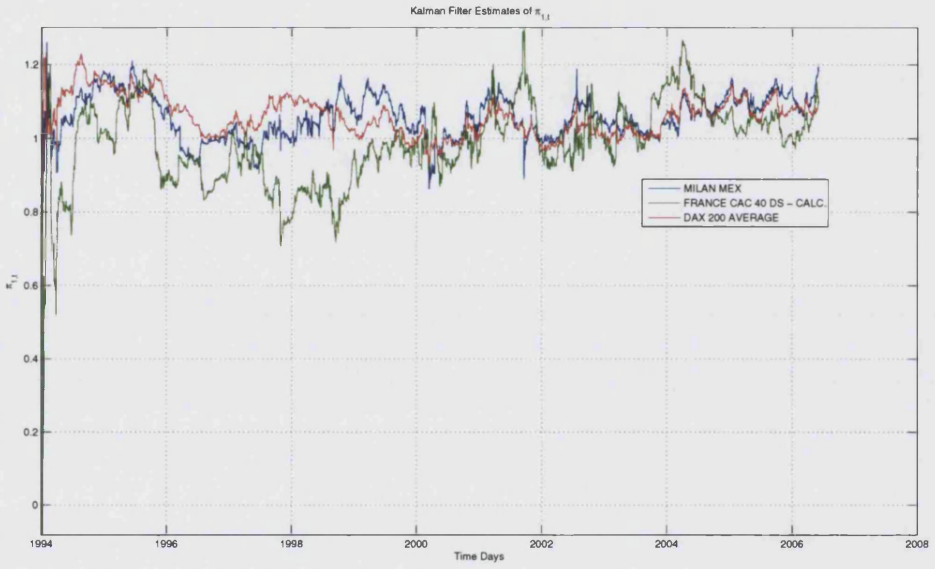

Figure 1.22: The Integration of the European Stock markets is illustrated in this overlay of the DAX, CAC and MIB

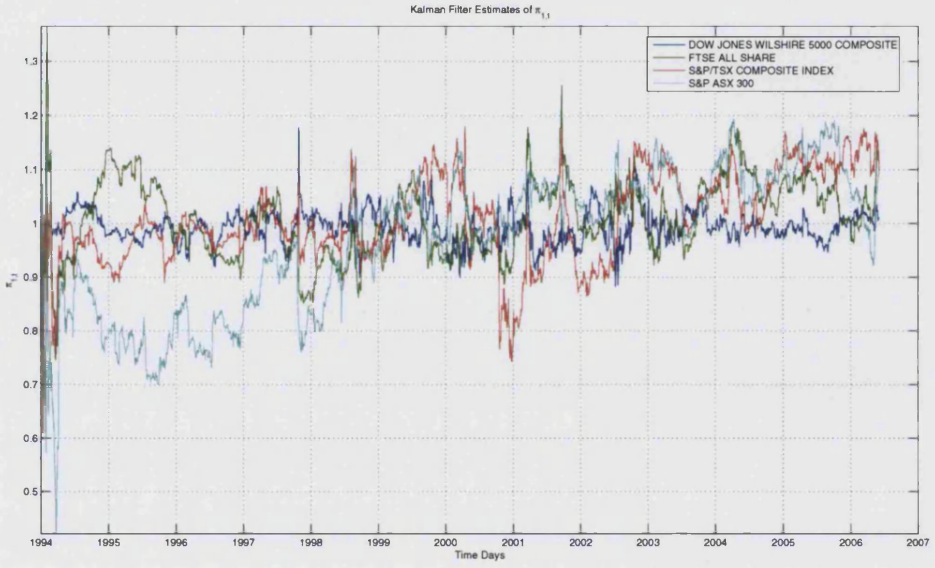

Figure 1.23: The Integration of the Anglo Saxon Stock markets is illustrated in this overlay of the DOW, FTSE, ASX and TSX indices.

1.6.2 Appendix II: Data Generating Simulations

Using the following global set of artificial parameter matrices,

$$\mathbf{\Pi}_0 = \begin{bmatrix} -0.2 & 0.9 \\ 0.57 & -0.3 \end{bmatrix} \tag{1.6.4}$$

$$\mathbf{\Pi}_1 = \begin{bmatrix} 2.31 & 0.6 \\ 1.05 & -6 \end{bmatrix} \tag{1.6.5}$$

$$\mathbf{\Pi}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.6.6}$$

$$\mathbf{\Pi}_3 = \begin{bmatrix} 0.5 \\ 0.9 \end{bmatrix} \tag{1.6.7}$$

The autoregressive coefficients are stationary but have one negative root in the kernel of $\mathbf{F}$,

$$\mathbf{z} = [0.46798, -0.96798]^{\mathrm{T}} \tag{1.6.8}$$

In general the recursive and dynamical systems literature relies on stable and semi stable chaotic systems to generate the deterministic variation, which often underlines many macroeconomic systems, such as the Philips curve, see Ljungqvist and Sargent (2004) [186], Stokey and Lucas (1989) [244]. For a review of deterministic linear and non-linear dynamical systems see the core text by Robinson (1998) [225].

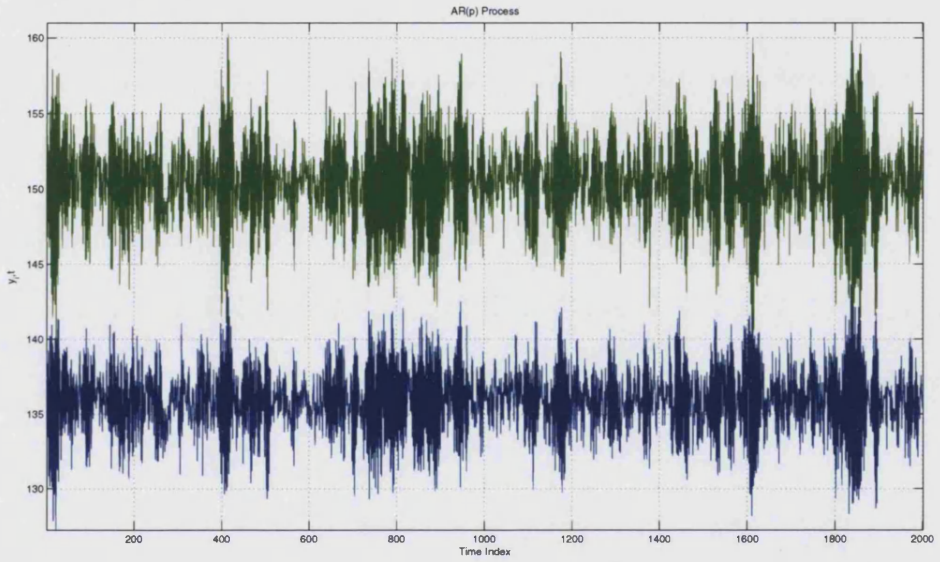

Figure 1.24: A Simple AR(p) model with negative roots, the chaotic effects are clearly visible and should not be confused for stochastic volatility or ARCH effects. $\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec(\mathcal{Y}_t) + \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t$

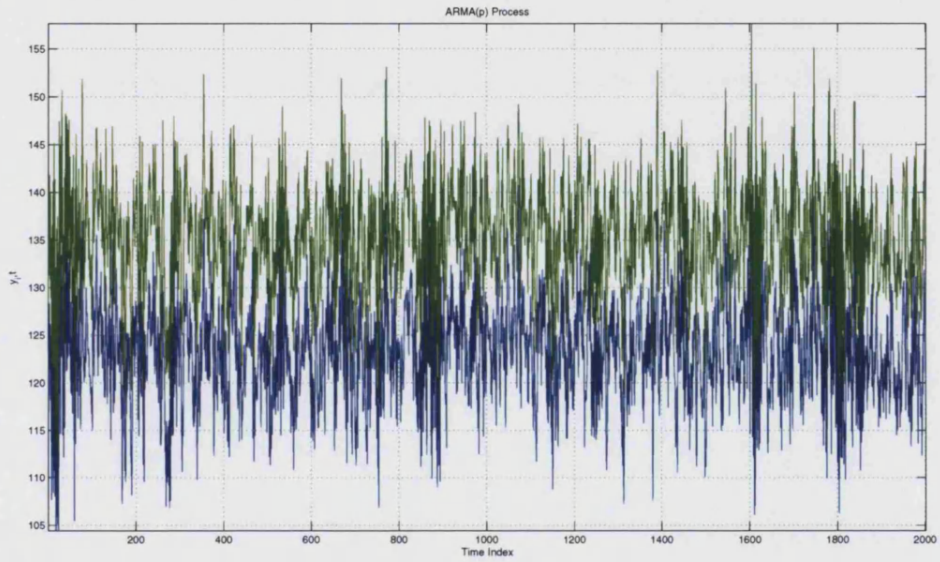

Figure 1.25: An ARMA(p,q) model again with negative roots, in the Kernel of the AR Coefficient Matrix, note that the MA term dampens the Chaotic effects. $\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{\Pi}_1^{\mathrm{T}} vec\left(\mathcal{U}_t\right) + \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t$

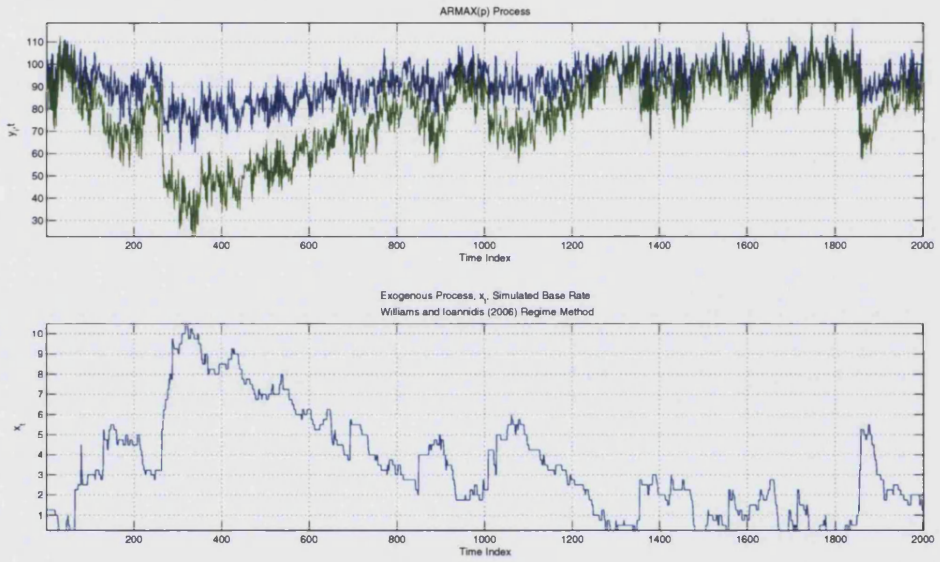

Figure 1.26: ARMAX(p,q) model with single exogenous factor, the variable x_t, corresponds to a simulated base rate using the method suggested in Williams and Ioannidis 2006c. $\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\,(\mathcal{Y}_t) + \mathbf{\Pi}_1^{\mathrm{T}} vec\,(\mathcal{U}_t) + \mathbf{\Pi}_2^{\mathrm{T}} \mathbf{x}_t + \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t$

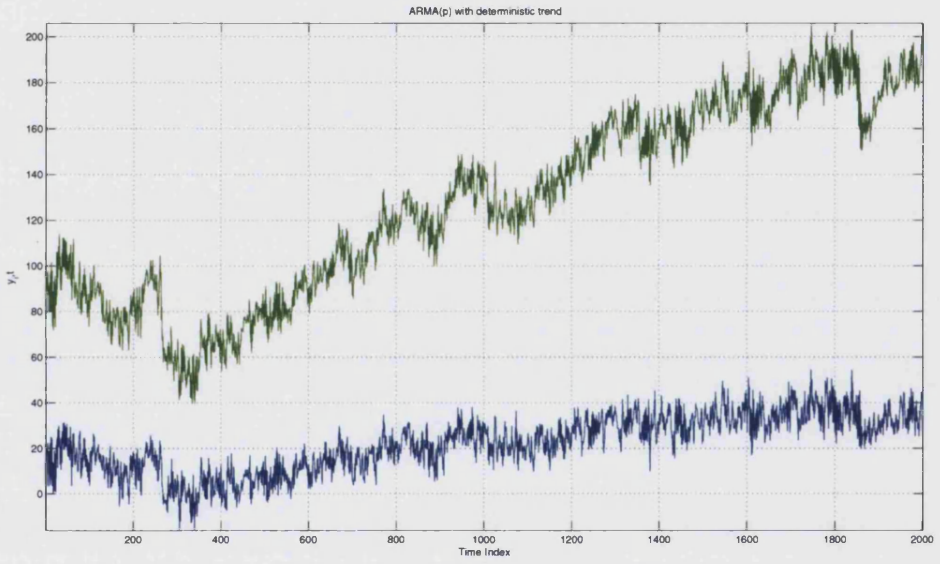

Figure 1.27: An ARMAX(p,q) model with deterministic common trend, $\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{\Pi}_1^{\mathrm{T}} vec\left(\mathcal{U}_t\right) + \mathbf{\Pi}_2^{\mathrm{T}} \mathbf{x}_t + \mathbf{\Pi}_3^{\mathrm{T}} t + \mathbf{\Sigma}^{\frac{1}{2}} \varepsilon_t$

Chapter 2

Non-Linear Models: MV-ARCH/GARCH Models A Review

2.1 Chapter Abstract

In the previous vector models the general assumption is that the objective function in the optimization tends to an unconditional or conditional multi-normal distribution, identification of parameter estimates then proceeds with assumption in mind. In this section the assumption of unconditional multi-normality is relaxed, by construction of an autoregressive specification of the covariance matrix.

2.2 Auto-Regressive models of Conditional Second Moments

The ARCH model of Engle (1982) [?], was quickly followed by the GARCH model of Bollerslev (1986) [32] and the EGARCH model of Nelson (1985) [207], these specifications characterized the volatility of filtered economic time series as an autoregressive

system. In essence they treated these time series as an infinite mixture model of conditional normals. At any given point the volatility parameter was conditioned on the past recursion of the disturbances.

$$y_t = \mathbf{b}^{\mathrm{T}}\mathbf{x}_t + u_t \tag{2.2.1}$$

$$u_t \sim N\left(0, \sigma_t^2\right) \tag{2.2.2}$$

$$\sigma_t^2 = f\left(\mathcal{U}_t, \mathcal{S}_t, \theta\right) \tag{2.2.3}$$

$$\mathcal{U}_t = \left[u_{t-1}^2, ..., u_{t-p}^2\right] \tag{2.2.4}$$

$$\mathcal{S}_t = \left[\sigma_{t-1}^2, ..., \sigma_{t-q}^2\right] \tag{2.2.5}$$

the conditional error in the univariate case is,

$$u_t = \sigma_t \varepsilon_t \tag{2.2.6}$$

$$\varepsilon_t \sim N\left(0, 1\right) \tag{2.2.7}$$

therefore the regression is specified as,

$$y_t = \mathbf{b}^{\mathrm{T}}\mathbf{x}_t + \sigma_t \varepsilon_t \tag{2.2.8}$$

The original ARCH specification was a simple autoregressive process, AR(p), the GARCH model extended this to an ARMA(p,q) representation, First the ARCH model,

$$\sigma_t^2 = \bar{\sigma}^2 + \alpha^{\mathrm{T}} vec\left(\mathcal{U}_t\right) \tag{2.2.9}$$

And second the GARCH type model

$$\sigma_t^2 = \bar{\sigma}^2 + \alpha^{\mathrm{T}} vec(\mathcal{U}_t) + \beta^{\mathrm{T}} vec(\mathcal{S}_t) \tag{2.2.10}$$

The parameter vectors are therefore respectively, for the ARCH model,

$$\theta = \left[\bar{\sigma}, \alpha^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.2.11}$$

And for the GARCH model,

$$\theta = \left[\bar{\sigma}, \alpha^{\mathrm{T}}, \beta^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.2.12}$$

The parameters must be bounded such that the following stationarity condition holds,

$$\mathbb{P}\left(\sum_{t=1}^{T} u_t^2 < \infty\right) = 1 \tag{2.2.13}$$

Other specifications have attempted to create non linearities in adjustment the most famous being the EGARCH of the Nelson (1991) [206], long memory effects, FI-GARCH, threshold transitions between volatility states TARCH and yet more have rejected the conditional normality assumption and utilized other objective distribution mixtures, e.g. the t-distribution or the asymmetric t-distribution, such as in the ARCD model of Hanssen (1994) [139]. See Gourieroux (2001) [113] for the definitive guide to univariate ARCH/GARCH models. In general the models focus on smoothed or asymmetric responses to shocks, the ARCD model permits deformation of the probability density mass, by allowing conditional freedom in two additional parameters. Appendix II.1 demonstrates the estimation criterion under maximum and quasi maximum likelihood.

2.2.1 Taking the Univariate Specification into the Multivariate Domain

The generic multivariate approach is along the same lines, except that there is a significant complication, the dynamic volatility is now a matrix process, the basic model set up is as follows,

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{u}_t \tag{2.2.14}$$

$$\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t) \tag{2.2.15}$$

$$\mathbf{\Sigma}_t = \mathfrak{H}(\mathcal{U}_t, \mathcal{S}_t, \theta) \tag{2.2.16}$$

$$\mathcal{U}_t = \left[\mathbf{u}_{t-1}^2, ..., \mathbf{u}_{t-p}^2\right] \tag{2.2.17}$$

$$\mathcal{S}_t = \left[\sigma_{t-1}^2, ..., \sigma_{t-q}^2\right] \tag{2.2.18}$$

Rearranging in terms of the factorized conditional covariance matrix $\mathbf{\Sigma}_t$,

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.19}$$

$$\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.20}$$

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}) \tag{2.2.21}$$

Multivariate extensions have been hamstrung by several problems, primarily as a result of the numbers of parameters in even a modest specification, first the large number of parameters mean that the degrees of freedom quickly runs out, the models are quite restrictive as the conditional covariance matrices must be non-negative definite. Computing the model is difficult, the amount of computation time required

explodes and as a result real time results and restriction tests are extremely difficult. Finally the inference from the parameter estimates is difficult to contextualize, it is hard to visualize the impact of a single parameter amongst a great many others and as such economic inference is hard to draw from such models. The first major multivariate extension of the basic ARCH/GARCH framework was put forward by Bollerslev, Engle and Wooldridge (1988) [40], they suggested using the *ivech* operator to formulate a vector process which could then be transformed into a non-negative Hermitian matrix. One of the simplest and most intuitive models was put forward by Bollerslev (1990), he assumed that the marginal components were univariate GARCH models coupled by a constant correlation matrix. The BEKK model of Engle and Kroner (1994) [89] was the first true Multivariate extension of the basic GARCH model, by use of the *ivech* transformation, parameter matrices could be converted into non-negative Hermitian matrices and as a result each of the conditional covariance matrices is non-negative definite. Other specifications attempted to reduce the number of parameters by decomposing the vector processes into a reduced number of underlying components, the OGARCH model of Alexander (1992) [6] utilized a principle component approach and then reconstructed the conditional covariance matrix from its factor loadings.

The Bollerslev, Engle and Wooldridge Model

The Bollerslev, Engle and Wooldridge, MV-GARCH(p,q) model utilizes the *vech* representation to ensure that $\mathbf{\Sigma}_t$ is non-negative,

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}{}_t\mathbf{x}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.22}$$

$$\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.23}$$

$$\mathbf{\Sigma}_t = \mathbf{\Lambda}_t\mathbf{\Lambda}_t^{\mathrm{T}} \tag{2.2.24}$$

$$\mathbf{\Lambda}_t = \mathbf{K} + \sum_{i=1}^{p}\mathbf{A}_k vech\left(\mathbf{u}_{t-i}\mathbf{u}_{t-i}^{\mathrm{T}}\right) + \sum_{j=1}^{q}\mathbf{B}_l vech\left(\mathbf{\Sigma}_{t-j}\right) \tag{2.2.25}$$

$$\mathcal{A} = [\mathbf{A}_1, ..., \mathbf{A}_p] \tag{2.2.26}$$

$$\mathcal{B} = [\mathbf{B}_1, ..., \mathbf{B}_q] \tag{2.2.27}$$

$$\theta = \left[vec\left(\mathbf{K}\right)^{\mathrm{T}}, vec\left(\mathcal{A}\right)^{\mathrm{T}}, vec\left(\mathcal{B}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.2.28}$$

$$\mathbf{K} = ivech\left(\mathbf{k}\right) \tag{2.2.29}$$

Give the assumption of conditional normality, the objective function is therefore,

$$\mathfrak{L}\left(\theta\right) = -\tfrac{1}{2}\sum_{t=1}^{T} n\log\left(2\pi\right) + \log\left(\left|\mathbf{\Sigma}_t\right|\right) + \mathbf{u}_t^{\mathrm{T}}\mathbf{\Sigma}_t^{-1}\mathbf{u}_t \tag{2.2.30}$$

Imposing the normal constraints on the parameters to ensure the following stationarity condition is met,

$$\mathbb{P}\left(\sum_{t=1}^{T}\mathbf{u}_t^{\mathrm{T}}\mathbf{u}_t < \infty\right) = 1 \tag{2.2.31}$$

The BEKK model of Engle and Kroner

The BEKK model of Engle and Kroner, is similar in many respects to the Bollerslev, Engle and Wooldridge Model, but the model specification allows for a greater degree of freedom in capturing the dynamics of the conditional covariation. In the Bollerslev, Engle and Wooldridge Model, forming the conditional covariance matrix by matrix multiplying $\mathbf{\Lambda}_t$ by its own transpose, the cross products in this formulation result in many of the subtle off main-diagonal dynamics being drowned out by the main-diagonal processes. In the BEKK model the only cross products are from the parameter matrices, therefore potential shocks in covariation, may propagate without being overly affected by the dynamics from the main diagonal.

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}{}_t\mathbf{x}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.32}$$

$$\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.33}$$

$$\mathbf{\Sigma}_t = \mathbf{K}\mathbf{K}^{\mathrm{T}} + \sum_{i=1}^{p}\mathbf{A}_i^{\mathrm{T}}\left(\mathbf{u}_{t-i}\mathbf{u}_{t-i}^{\mathrm{T}}\right)\mathbf{A}_i + \sum_{i=1}^{q}\mathbf{B}_i^{\mathrm{T}}\left(\mathbf{\Sigma}_{t-j}\right)\mathbf{B}_i \tag{2.2.34}$$

$$\theta = \left[vec\left(\mathbf{K}\right)^{\mathrm{T}}, vec\left(\mathcal{A}\right)^{\mathrm{T}}, vec\left(\mathcal{B}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2.2.35}$$

$$\mathbf{K} = ivech\left(\mathbf{k}\right) \tag{2.2.36}$$

Again parameter estimation is via maximum likelihood, using the log multi-normal density function. The basic BEKK specification is the true multivariate extension of the univariate GARCH(p,q) model, however the drawbacks mentioned previously are particularly acute in this specification. The numbers of parameters under extremely

modest specifications is very large and computation times are similarly large. To this extent several modifications to the BEKK have been proposed that keep the same basic structure but reduce the size parameter matrices by incorporating various restrictions, these include the Scalar BEKK and the Diagonal BEKK. Additionally the BEKK has been specified in terms of a multivariate students t distribution to incorporate more complex dependency dynamics.

Correlation based MV-GARCH

The correlation based multivariate GARCH takes the following approach, volatility is dictated by the diagonal processes and the direction and magnitude of dependency is a product of the co-volatilities and either a constant or dynamic correlation. The first model by Bollerslev (1992) [34], the constant correlation MV-GARCH, (CCC) proposed the following two step algorithm, first the basic setup is the same as previously,

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{u}_t \tag{2.2.37}$$

$$\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t) \tag{2.2.38}$$

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.39}$$

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}) \tag{2.2.40}$$

However the conditional covariance matrix is now defined as a Hadamard product of a vector of conditional standard deviations and a time invariant normalized non-negative Hermitian matrix, i.e. the separation of volatility from correlation,

$$\mathbf{\Sigma}_t = \mathbf{R} \times \sigma_t \sigma_t^{\mathrm{T}} \tag{2.2.41}$$

$$\sigma_t = [\sigma_{1,t}, ..., \sigma_{n,t}]^{\mathrm{T}} \tag{2.2.42}$$

The conditional standard deviations are the functional evaluation at time t of a vector univariate ARCH/GARCH processes. For example under the basic GARCH(p,q) specification,

$$\sigma_{i,t} = \sqrt{\bar{\sigma}_i + \alpha_i^{\mathrm{T}} vec\,(\mathcal{U}_t) + \beta_i^{\mathrm{T}} vec\,(\mathcal{S}_t)} \tag{2.2.43}$$

$$\mathcal{U}_t = [u_{i,t-1}^2, ..., u_{i,t-p}^2] \tag{2.2.44}$$

$$\mathcal{S}_t = [\sigma_{i,t-1}^2, ..., \sigma_{i,t-q}^2] \tag{2.2.45}$$

Of course there is no limitation on the specification of $\sigma_{i,t}$, any combination of the plethora of univariate models is valid.

Dynamic Correlation

Engle (2002) [92]and Engle and Sheppard (2003) [235] proposed a new class of correlation MV-GARCH the dynamic conditional correlation model (DCC), this model utilizes a two step method in the same way as the (CCC) however the correlations

are treated as dynamic and modelled as an autoregressive matrix process,

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{u}_t \tag{2.2.46}$$

$$\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t) \tag{2.2.47}$$

$$\mathbf{y}_t = \mathbf{\Pi}^{\mathrm{T}}\mathbf{x}_t + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.48}$$

$$\varepsilon_t \sim N(\mathbf{0}, \mathbf{I}) \tag{2.2.49}$$

$$\sigma_t = [\sigma_{1,t}, ..., \sigma_{n,t}]^{\mathrm{T}} \tag{2.2.50}$$

$$\sigma_{i,t} = \sqrt{\bar{\sigma}_i + \alpha_i^{\mathrm{T}} vec(\mathcal{U}_t) + \beta_i^{\mathrm{T}} vec(\mathcal{S}_t)} \tag{2.2.51}$$

Collecting the past observations of each the marginal variables from the vector process, a normalized variable $\epsilon_{i,t}$, with the volatility effects captured is then fed into the matrix process,

$$\mathcal{U}_t = [u_{i,t-1}^2, ..., u_{i,t-p}^2] \tag{2.2.52}$$

$$\epsilon_{i,t} = u_{i,t}\sigma_{i,t}^{-1} \tag{2.2.53}$$

$$\epsilon_t = [\epsilon_{1,t}, ..., \epsilon_{n,t}] \tag{2.2.54}$$

$$\epsilon_t \sim N(\mathbf{0}, \mathbf{R}_t) \tag{2.2.55}$$

$$\tag{2.2.56}$$

The vector process ϵ_t, is then defined in terms of a dynamic correlation matrix as follows,

$$\epsilon_t = \mathbf{R}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.57}$$

This formulation may be rewritten in the following manner which is probably more intuitive,

$$\begin{aligned}
\mathbf{\Sigma}_t &= \sigma_t \sigma_t^{\mathrm{T}} \mathbf{R}_t && (2.2.58)\\
\mathbf{u}_t &= \left(\sigma_t \sigma_t^{\mathrm{T}} \mathbf{R}_t\right)^{\frac{1}{2}} \varepsilon_t && (2.2.59)\\
\mathbf{y}_t &= \mathbf{\Pi}^{\mathrm{T}} \mathbf{x}_t + \left(\sigma_t \sigma_t^{\mathrm{T}} \mathbf{R}_t\right)^{\frac{1}{2}} \varepsilon_t && (2.2.60)
\end{aligned}$$

Engle (2000) specifies three different alternatives for the matrix process $\mathbf{R}_t$, first the RiskMetrics$^{\mathrm{TM}}$ exponential smoother,

$$\rho_{i,j,t} = \frac{\sum_{l=1}^{r} \gamma^l \epsilon_{i,t-l} \epsilon_{j,t-l}}{\sqrt{\sum_{l=1}^{r} \gamma^l \epsilon_{i,t-l}^2} \sqrt{\sum_{l=1}^{r} \gamma^l \epsilon_{j,t-l}^2}} \tag{2.2.61}$$

Alternatively the matrix process $\mathbf{R}_t$ maybe specified as the normalized form of some autoregressive non-negative matrix $\mathbf{Q}_t$ operating on the standardized residuals ϵ_t.

$$\begin{aligned}
\rho_{i,j,t} &= \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}} \sqrt{q_{j,j,t}}} && (2.2.62)\\
\mathbf{Q}_t &= [q_{i,j,t}] && (2.2.63)
\end{aligned}$$

treating $\mathbf{Q}_t$ as a conditional covariance matrix, using for example the BEKK representation,

$$\mathbf{Q}_t = \bar{\mathbf{Q}} + \sum_{l=1}^{p} \mathbf{A}_l^{\mathrm{T}} \varepsilon_{t-l} \varepsilon_{t-l}^{\mathrm{T}} \mathbf{A}_l + \sum_{k=1}^{q} \mathbf{B}_k^{\mathrm{T}} \mathbf{Q}_{t-k} \mathbf{B}_l \tag{2.2.64}$$

Or the MARCH model of Engle and Ding (2001),

$$\mathbf{Q}_t = \bar{\mathbf{Q}} \times \left(\mathbf{e}\mathbf{e}^{\mathrm{T}} - \sum_{l=1}^{p} \mathbf{A}_l - \sum_{k=1}^{q} \mathbf{B}_k\right) + \sum_{l=1}^{p} \mathbf{A}_l \times \varepsilon_{t-l} \varepsilon_{t-l}^{\mathrm{T}} + \sum_{k=1}^{q} \mathbf{B}_k \times \mathbf{Q}_{t-k} \tag{2.2.65}$$

Several authors have extended the basic methodology to incorporate various properties such as the incorporation of block dynamics in the correlation structure, using copulas to represent the interaction of the marginal variances with the dependency matrix. The major problem with this approach as noted in Alexander (2001) is the flattening of the objective function, which increases computation time and complicating restriction tests.

Dimension Reduction and Orthogonal/Factor GARCH

The final multivariate ARCH model considered, is the Orthogonal GARCH model of Alexander and Chibumba (1997) [6], Byström (2004) [47] and Alexander (2001) [5]. Factor analysis and principle component decomposition are both dimension reduction techniques which attempt to summarize a large number of variables in terms of linear combinations of those variables. This is a useful technique in multivariate volatility models as in general the number of parameters increases significantly with the number of variables. Consider the following decomposition, of the covariance matrix of multi-equation residuals from (2.2.22),

$$\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t) \tag{2.2.66}$$

$$\mathbf{u}_t = \mathbf{W}\mathbf{\Psi}_t^{\frac{1}{2}}\varepsilon_t \tag{2.2.67}$$

$$\mathbf{\Sigma}_t = \mathbf{W}\mathbf{P}_t\mathbf{W} \tag{2.2.68}$$

Where $\mathbf{W}$ is a matrix of factor weights, $\mathbf{\Psi}_t$, is the time varying covariance matrix of the principle components.

Computing the Principle Components

There are a variety of suggested approaches to factor GARCH models, this description is an amalgamation of the Alexander (2000) [5], Lanne and Saikkonen (2005) [176] and the Engle, Ng and Rothschild, (1990) [91] approach. Again consider a filtered set of residuals from some sample, $[1, ..., \tau]$, the $\tau \times n$ matrix $\mathbf{U}$, with columns $[\mathbf{u}_1, ..., \mathbf{u}_n]$ in the time dimension. The matrix $\mathbf{U}^{\mathrm{T}}\mathbf{U}$, is therefore an $n \times n$ matrix, normalizing by the diagonal, yields a matrix $\mathbf{R}$, with all positive eigenvalues which sum to n, therefore, $\mathbf{R}$ is a valid sample correlation matrix. Eigenvector decomposition of $\mathbf{R}$ yields the factor loading, i.e.,

$$\mathbf{wR} = v\mathbf{w} \tag{2.2.69}$$

$$v \in \{v_1, ..., v_n\} \tag{2.2.70}$$

$$\mathbf{w} \in \{\mathbf{w}_1, ..., \mathbf{w}_n\} \tag{2.2.71}$$

Where v_i is the i^{th} eigenvalue and $\mathbf{w}$ is the i^{th} eigenvector. The first principle component explains the maximum amount of the total variation in the cross section of the sample $\mathbf{U}$, the second explains any of the remaining variation and so on, the maximum number of principle components is equal to the number of variables, n.

Rewriting 2.2.69 as a matrix of eigenvectors and a diagonal matrix of eigenvalues,

$$\mathbf{V} = [v_{i,j} = v_i] \tag{2.2.72}$$

$$\mathbf{z} \in \{\mathbf{z}_1, ..., \mathbf{z}_n\} \equiv \mathbf{Z} \tag{2.2.73}$$

Therefore the $\tau \times p$ matrix of principle components is as follows,

$$\mathbf{P} = \mathbf{UW} \tag{2.2.74}$$

Selection of the optimal number of principle components maybe undertaken using Kaiser-Guttman rule, whereby the number of components used is equal to the number of Eigenvalues where $v_i > 1$, i.e. the variance of the principle component is at least greater than a single column from $\mathbf{U}$. An alternative approach has been suggested by Kapetanios and Marcellino (2005) [160], which looks at the explanatory power of each of the factors. In this method the principle components are used to create linear predictions of the sample, the largest eigenvalue of the correlation matrix of the residuals from this linear model then describes the degree of orthogonality introduced by the component, i.e. the degree of deformation of the euclidian ball of the residuals density. Kapetanios and Marcellino (2006a) [161] Kapetanios and Marcellino (2006b) [160] suggest that if the largest eigenvalue of this correlation matrix is greater than 4, then another component should be added. We can mathematically describe their procedure as follows, the optimal number of components p^* belongs to

the set of real natural positive integers from 1 to the number of variables n,

$$p_k^* \in \{1, ..., k, ..., n\} \tag{2.2.75}$$

correspondingly there is a set of components,

$$\mathbf{P}_k = [\mathbf{P}_1, ..., \mathbf{P}_k, ..., \mathbf{P}_n] \tag{2.2.76}$$

For each possible set of components the following linear model is estimated

$$\mathbf{U} = \mathbf{P}_k \mathbf{b}_k^{\mathrm{T}} + \mathrm{E}_k \tag{2.2.77}$$

Where E_k is the matrix of residuals from the multiple regression of $\mathbf{P}_k$ onto $\mathbf{U}$, computing the correlation matrix of E_k yields,

$$\mathrm{E}_k^{\mathrm{T}} \mathrm{E}_k = \Xi_k \tag{2.2.78}$$

$$\xi_k = \left(\left[\sqrt{\xi_{i,j}} \right]_{i=j} \right)_k \tag{2.2.79}$$

$$\mathbf{R}_k = \Xi_i \times \left(\xi_k \xi_k^{\mathrm{T}} \right)^{-1} \tag{2.2.80}$$

The eigenvector decomposition of $\mathbf{R}_k$, yields,

$$\mathbf{z}_k \mathbf{R}_k = v_{i,k} \mathbf{z}_k \tag{2.2.81}$$

The optimization problem suggested in Kapetanios and Marcellino (2005) is very simple iterate, k until the largest eigenvalue, $v_{i,k}$ is less than an exogenously imposed constant α,

$$p_k^* \triangleq \arg \min_{p_k} \left[\max \left[v_{i,k} \right] < \alpha \right] \tag{2.2.82}$$

The suggested threshold for the largest eigenvalue is $\alpha = 4$, this threshold is derived from Roy's maximum root test, applied to normalized matrices, see chapter appendix II.1. An alternative approach is to test to see the degree of significance to which the correlation matrix deviates from the identity matrix, given that the distribution of covariance matrices is defined by the Wishart distribution, therefore constructing a significance test for α, i.e. the construct a null hypothesis that the estimated matrix $H_0 : \mathbf{R}_k = \mathbf{I}$, see Takemura and Kuriki (2000) [246] and Muirhead2005 [202]. They propose a simple maximization test, to distinguish between two covariance matrices, see Chapter Appendix II.2, along with a set of sample critical values.

Model Selection with Orthogonal GARCH

In some ways the orthogonal GARCH approach is similar to the correlation GARCH model, i.e. a first stage that reduces the model to a simpler form and a second stage based around one of the standard models, Alexander (2000) [5] suggests modelling the dynamic covariance matrix of the principle components as an EWMA process is the same as that in Engle (2002) [83], suggests modelling the dynamics of the correlation structure in the Dynamic Correlation methodology. In general any of the models in this section maybe used to model the conditional covariances, the only restriction is generally computational, if the dimension reduction for the optimal number of principle components is very large (e.g. 500 variables to 15 principle components), then any of the models are reasonable to use and in general as the

unrestricted BEKK model offers the greatest latitude in dynamics of the covariance matrix, this would generally be the most appropriate choice. Factor GARCH offers an impressive list of features to recommend it for use, first the fact that the reduction in dimensions allows for a greater degree concavity in form of the objective function due to the parameter reduction, second a systematic means by which to assess the optimal number of components and third the ability to forecast mean and variance using the standard Kalman filter forward recursion methodology on a reduced set of variables.

Asymptotic Theory of the Multivariate ARCH/GARCH Specification

Very little literature exists on the asymptotic behavior of multivariate ARCH/GARCH models, in comparison to the large literature devoted to the asymptotic theory underlying the univariate specification. This section briefly surveys several key results which have been demonstrated in recent articles. The majority of asymptotic theory surrounds the convergence of the fourth moment structure of the process $\mathbf{u}_t$. Generally asymptotic behavior is determined by the structure of the matrix autoregressive process underlying the second moments. However, several generalizations as to the asymptotic distribution of the resultant process maybe made. Two general approaches are first to utilize results from infinite mixture models to approximate the moment structure and second to define the distribution of the quadratic form of the second moment model, then utilize this inference in defining the structure of

the result. Generally these approaches have been limited to applications relating to the most basic MV-GARCH framework, the Bollerslev CCC model. However several more recent studies have attempted to reconcile the asymptotic properties of MV-GARCH models and drawn inference as to the consistency of ML and QML estimation. The most comprehensive assessment of the asymptotic properties of vector ARCH models is Ling and McAleer (2003) [185]. Using the Bollerslev CCC model to describe the disturbances from an autoregressive filtration, they show the consistency conditions for parameter estimates and asymptotic normality of the QMLE estimators. Other work by Comte and Lieberman (2001) [54] suggests a multivariate central limit theorem approach which does not simply rely on the asymptotic normality assumptions extending univariate work by Bollerslev (1995), Boussama (2000) [42] and existing work on the consistency of Multivariate GARCH estimators by Jeantheau (2001) [152]. Consider the following process,

$$\mathbf{u}_t = \boldsymbol{\Sigma}_t^{\frac{1}{2}} \varepsilon_t \tag{2.2.83}$$

Where the n length random vector process $\mathbf{u}_t \in \mathbb{R}^n$ and as previously stated the process ε_t is an n length vector of independent standard normal variables under the normal Labesgue measures, i.e. $\varepsilon_t \sim N(\mathbf{0}, \mathbf{I})$, if $\mathbb{F}$ is the σ-field operating over the sample space $\mathbb{R}$, then stating the following tuple, $(\Omega, \mathbb{F}, \mathbb{P})$, which contains the complete Borel-σ field required to describe the evolution of $\mathbf{u}_t$, now consider the BEKK model from 2.2.32, in this model the matrix process is assumed to be an ARMA(p,q)

where non-negativity is preserved as the result of pre and post multiplication of the past recursions of the covariance matrix $\mathbf{\Sigma}_t$ and the innovations of $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$ with the time invariant parameter matrices, $\mathbf{A}_i$ and $\mathbf{B}_j$, for the parameter matrix. Consider the evolution of $\mathbf{\Sigma}_t$, over some sample period of length τ.

$$\mathbf{S}_\tau = \left[\left(vec\sum_{t=1}^{1}\mathbf{\Sigma}_t\right)^{\mathrm{T}}, \left(vec\sum_{t=1}^{2}\mathbf{\Sigma}_t\right)^{\mathrm{T}}, ..., \left(vec\sum_{t=1}^{\tau}\mathbf{\Sigma}_t\right)^{\mathrm{T}}\right]^{\mathrm{T}} \quad (2.2.84)$$

$$\hat{\mathbf{\Psi}}_t = \tfrac{1}{\tau}\mathbf{S}_\tau^{\mathrm{T}}\mathbf{S}_\tau \quad (2.2.85)$$

Comte and Lieberman (2001) demonstrate using the Jantheau consistency result, that the process, $\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t$, will have convergent fourth moment structure, under the BEKK(1,1) framework, that as $\tau \to \infty$, the matrix $\hat{\mathbf{\Psi}}_t$ does converge to some matrix $\mathbf{\Psi}$, i.e. $\frac{1}{\tau}\mathbf{S}_\tau^{\mathrm{T}}\mathbf{S}_\tau \to \mathbf{\Psi}_t$. Given this fourth moment convergence the estimator $\hat{\theta}$, from the MLE methodology demonstrated in 1.3.170, will show strong consistency.

2.2.2 Some Empirical Results for MV-GARCH models

This section details two empirical studies designed to test the effectiveness of the various MV-GARCH models outlined in the previous section. Given the parameter estimation difficulties inherent for high dimensional systems the dimensionality of the system is restricted to sets of $n = 4$. The first sample dataset utilizes stocks drawn randomly from the S&P 500 index, the second sample dataset utilizes market index data for 10 broad market indices from around the world, this data is commonly used in contagion and market integration analysis, data coverage is daily for 20 years. The

continuous time returns are computed and filtered using n univariate ARX(p) filters,

$$y_{i,t} = \lambda_1 y_{i,t-1} + ... + \lambda_2 y_{i,t-p} + \gamma^{\mathrm{T}} \mathbf{x}_t + u_i, t \tag{2.2.86}$$

$$\mathbf{x}_t = [x_{1,t}, x_{1,t}, ..., x_{m,t}]^{\mathrm{T}} \tag{2.2.87}$$

Where $\lambda_i \in \lambda$ and $\gamma_j \in \gamma$ are model parameters. The order of the AR process and inclusion tests on the library of exogenous parameters is undertaken using the generic likelihood ratio approach suggested in 1.3.157, to account for the possible miss-identification of the conditional variance. The residuals, $u_{i,t} \in \mathbf{u}_t$, are collected in the array $\mathcal{U}_t = [\mathbf{u}_t, \mathbf{u}_{t-1}, ..., \mathbf{u}_{t=1}]$ and finally a data matrix over the sample period $[1, ..., \tau]$, $\underset{(n\times 1)\times\tau\rightarrow\tau\times n}{\mathrm{mat}}(\mathcal{U}_\tau) = \mathbf{U}$, is constructed. The iterations of the evaluation of the likelihood functions are plotted as are the comparative evolution of the correlations, covariances and standard deviations. A comparative chart of the number of parameters for the respective models is illustrated for the MV-ARCH/GARCH specifications, covered in this review.

2.2.3 Dataset 1: Randomly Sampled Stocks from the S&P 500

A typical requirement in modern portfolio analysis is to analyze the conditional covariation of a set of assets, this section looks at parameterizing models over a variety of specifications, for a group of asset from the S&P 500 composite. For computational reasons random subsets of four assets were taken from the S&P 500 composite, filtered and the four MV-GARCH models are fitted. The possible exogenous variables

Table 2.1: Exogenous variables, used in the first stage ARX(p), filter, the PEM-ARX approach uses the 1.3.157 approach test variable inclusion/exclusion restrictions.

$x_{1,t}$: Dividend Yeild

$x_{2,t}$: Price to Earnings Ratio

$x_{3,t}$: Market to Book Ratio

$x_{4,t}$: Gearing

$x_{5,t}$: Results Dummy

$x_{6,t}$: 9/11 Dummy

$$x_{7,t}, ..., x_{12,t} : \left\{ \begin{array}{l} \text{risk free rates, overnight Deposit Rate to 1 year rate} \\ \text{(scaled to daily returns continuous method)} \end{array} \right\}$$

used in the ARX(p) model are defined in table 2.1.

*CCC Model Specification Refers to First Stage Univariate GARCH Model Specification,

**DCC Model Refers to Univariate GARCH Model specifications and the Multivariate MARCH Specification Underlying the Correlation Processes,

Table 2.2: A Comparison of Maximum Likelihoods Obtained with a Gaussian Objective Function

Model Specification	$p = 1, q = 1$	$p = 2, q = 1$	$p = 1, q = 2$	$p = 2, q = 2$
CCC*	−8547.76	−4355.24	−8213.36	−4314.54
DCC**	−8121.56	−4121.31	−8067.20	−3991.71
BEKK	−5939.01	−5933.63	−5877.54	−5874.51
O - GARCH**	−6401.01	−6399.11	−6387.21	−6383.99

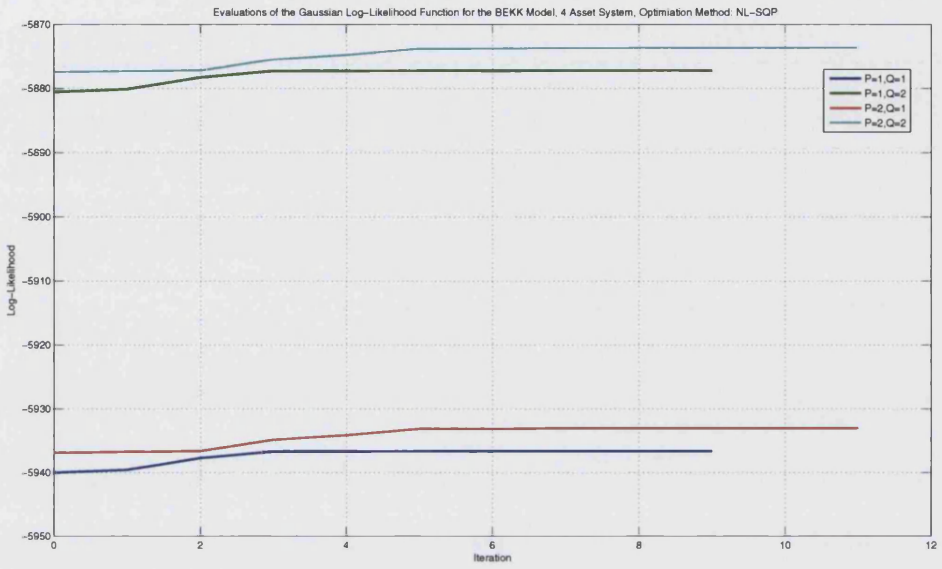

Figure 2.1: Comparison of the iterations of the evaluations of the BEKK model for different specifications, as the model specification increases the number of parameters the slope of the objective function flattens very quickly.

***O-GARCH, Specification Refers to the BEKK Model Specification Underlying the Principle Component Processes.

Table 2.2, illustrates the comparative log-likelihoods of the four models over a sample dataset, the BEKK model is clearly the optimal performer, from 20 sets of sub-samples from the S&P 500 the BEKK model consistently produced conditional covariance estimates that resulted in the largest evaluated likelihoods. However the difficulty in estimating the model becomes apparent when assessing the objective function, in comparison to that of the DCC model with MARCH type correlation dynamics. The

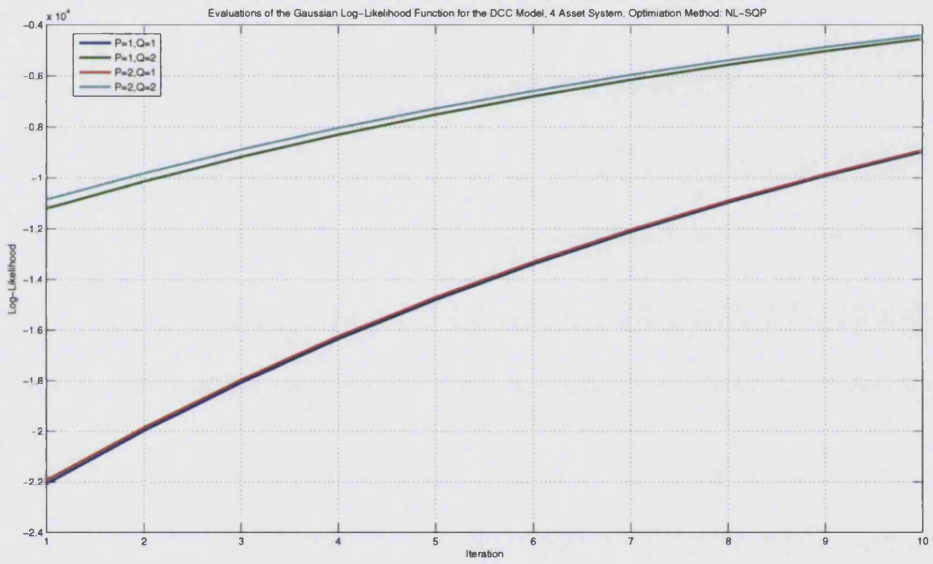

Figure 2.2: By comparison to the BEKK model the DCC with MARCH correlation structure, has a much steeper objective evaluation for the same parameter specification.

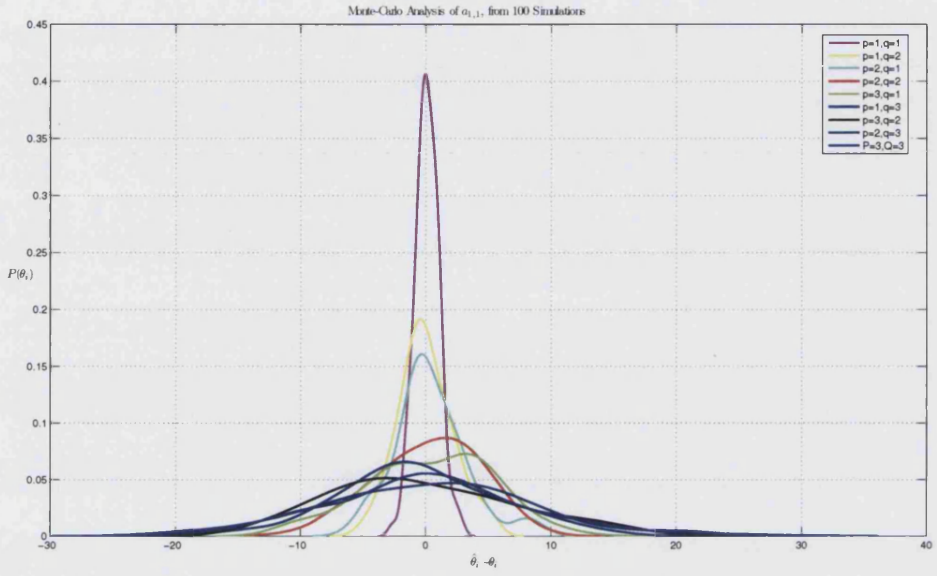

Figure 2.3: Monte-Carlo Analysis of the MLE estimation of $\hat{\theta}_i = a_{1,1}$, from the general BEKK(p, q) model, as a deviation from the true parameter, θ_i.

flat nature of the BEKK iterates is not in itself indicative of any particular deficiencies in the model estimation. However when undertaking monte-carlo studies the impact of this flatness emerges very quickly, the figure 2.3 indicates the joint error of one parameter from 100 monte-carlo experiments, as the lag operator is increased, the number of parameters jumps and the spread of the parameter estimates increases. Figures 2.4 to 2.7, illustrate the evolution of the variance/covariance matrix for some subsets of the S&P 500.

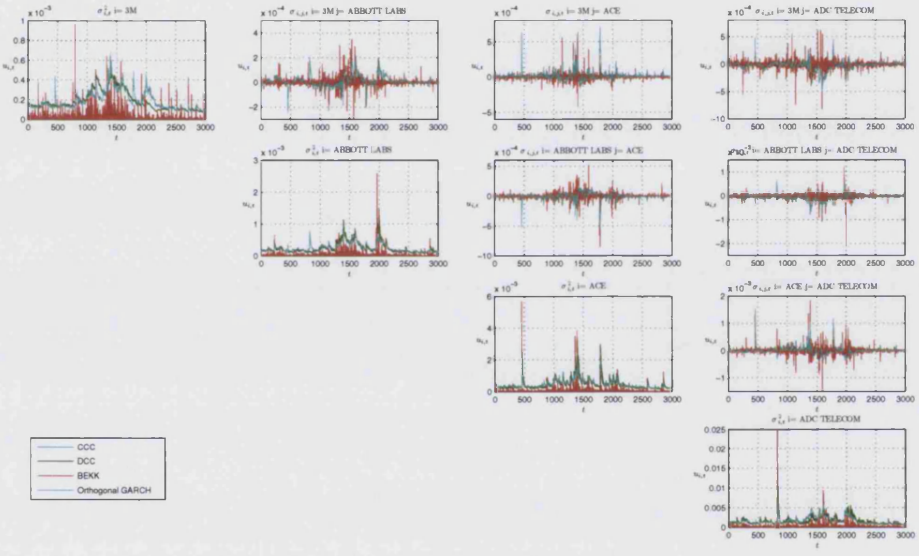

Figure 2.4: Subset 1 of four stocks from the S&P 500 Index.

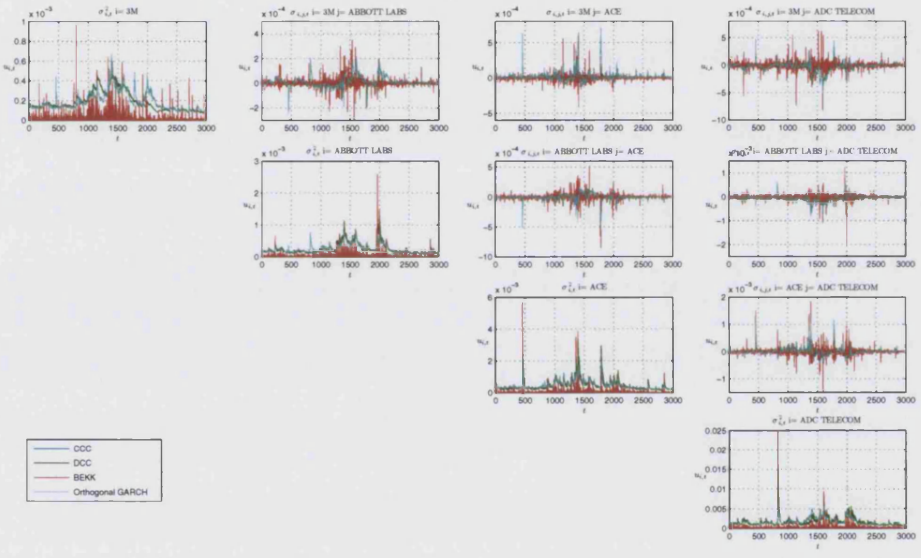

Figure 2.5: Subset 2 of four stocks from the S&P 500 Index.

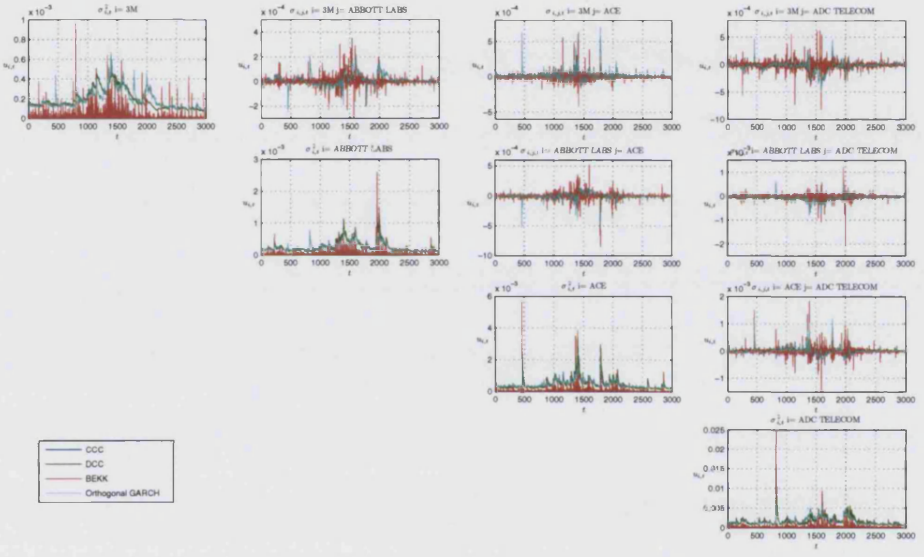

Figure 2.6: Subset 3 of four stocks from the S&P 500 Index.

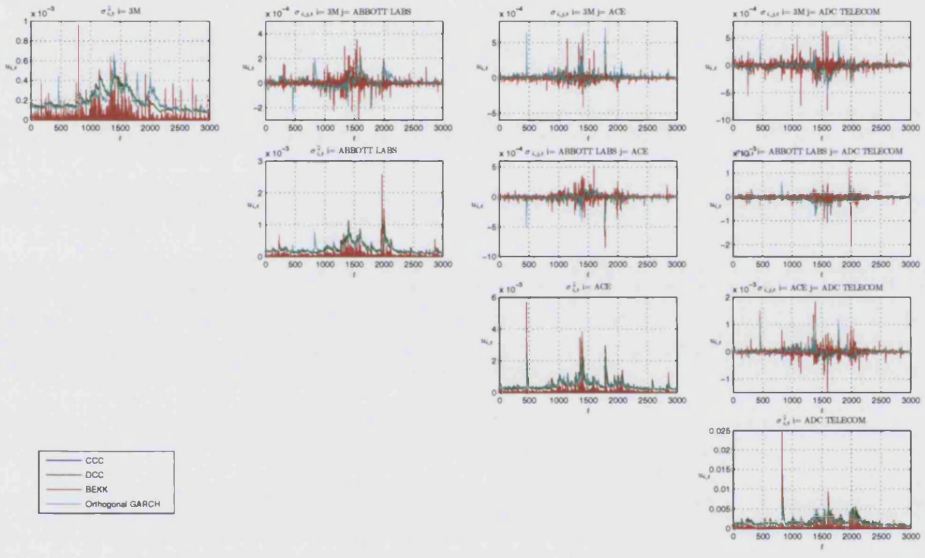

Figure 2.7: Subset 4 of four stocks from the S&P 500 Index.

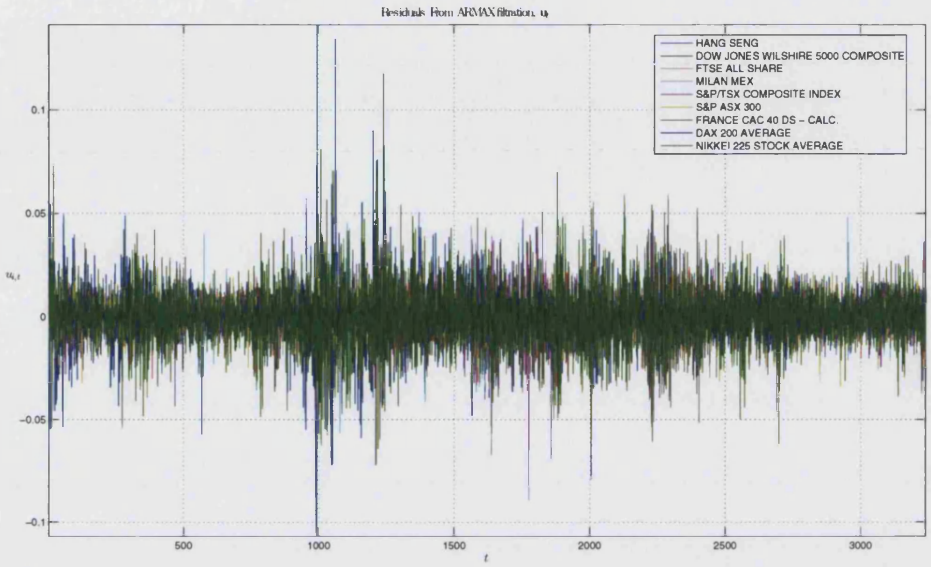

Figure 2.8: Residuals from ARMAX(p, q) filtration of nine global indices.

2.2.4 Dataset 2: Analysis of International Stock Market Indices

Another typical application of MV-GARCH models is in the analysis of the evolution of the quadratic covariation of global financial markets. Appendix IV, offers an alternative quicker approach using the recursive matrix Kalman filter suggested in 1.3.122, which models the dependency as a random matrix process. The first stage filter is a basic ARMAX(p, q), model as specified in 1.3.12 and uses the standard infinite order MA process approach. The filtered returns are as illustrated in figure 2.8. The evolution of the variance/covariance matrix for two sample 4×4 sets, is illustrated in

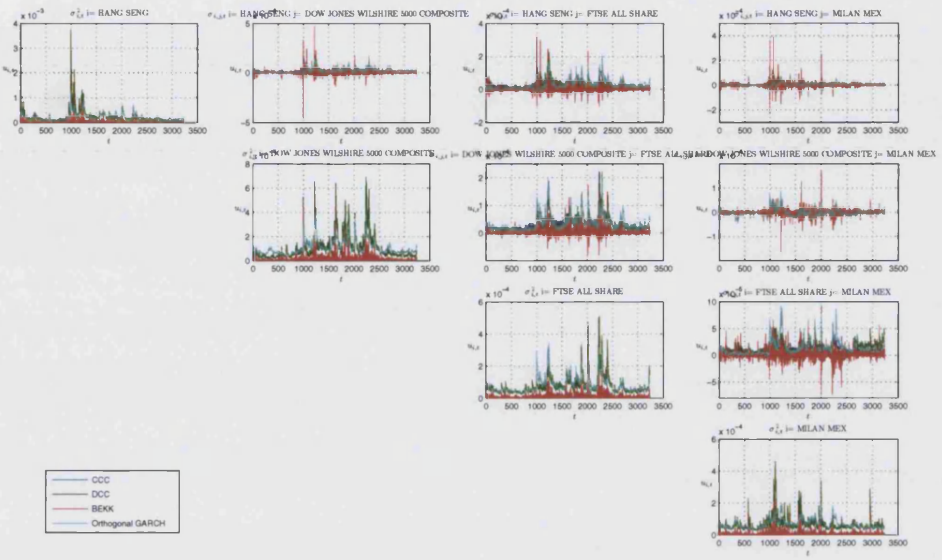

Figure 2.9: The evolution of the conditional covariance matrix, $\mathbf{\Sigma}_t$, for the Hang Seng (HK), Dow Jones (US), FTSE (UK) and Milan MEX (IT) Indices.

figures Table 2.3, illustrates the evaluated log-likelihoods from the optimization of a Gaussian objective function, using the covariance specifications for the four major model classes.

2.3 Chapter Appendix: Monte-Carlo Testing of Optimization Approaches

The following is a simple monte-carlo evaluation of a univariate GARCH(p,q) model. 10,000 observations are drawn 1000 times from the basic model and the parameters estimated, the distribution of the point estimates under maximum likelihood. The

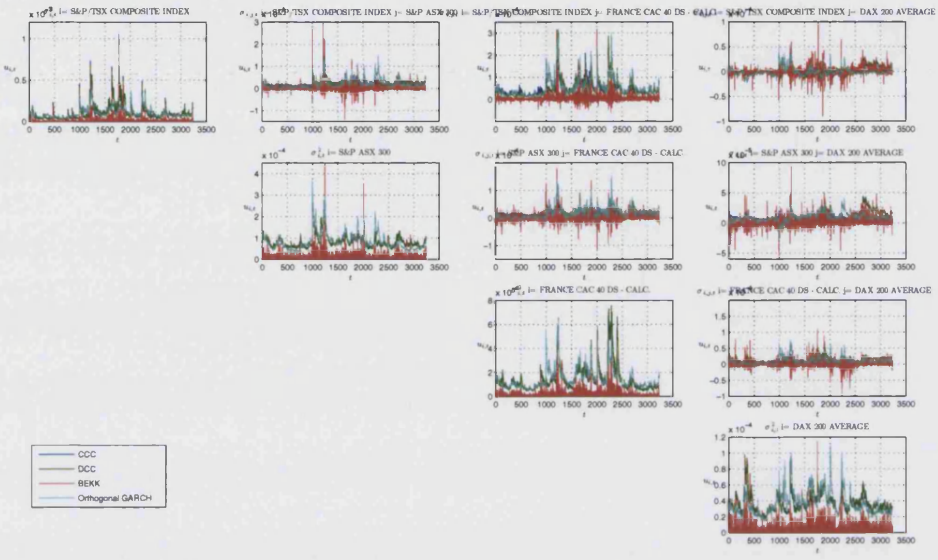

Figure 2.10: The evolution of the conditional covariance matrix, $\mathbf{\Sigma}_t$, for the TSX (Can), ASX (Aus), DAX (GER) and CAC (FR) indices.

Table 2.3: Table of evaluated likelihoods for various GARCH specifications.

Model Specification	$p = 1, q = 1$	$p = 2, q = 1$	$p = 1, q = 2$	$p = 2, q = 2$
CCC*	−108952	−108841	−108320	−108200
DCC**	−95457	−95298	−95362	−95130
BEKK	−88721	−88604	−88702	−88511
O - GARCH***	−99411	−98193	−99397	−98177

optimization algorithm used is the constrained SQP algorithm with BGFS Hessian updating, see Fletcher (1994), the parameter constraints are generated from the stationarity and consistency conditions, which are as follows,

$$\omega > 0 \tag{2.3.1}$$

$$a_1 + b_1 < 1 \tag{2.3.2}$$

$$a_1 > 0 \tag{2.3.3}$$

$$b_1 > 0 \tag{2.3.4}$$

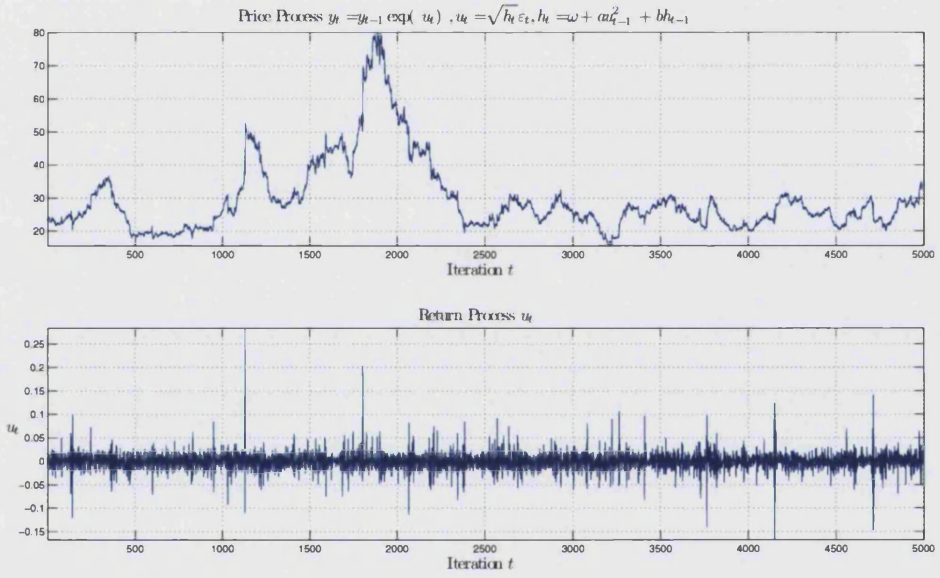

Figure 2.11: A single Price and Return process pathway for the ARCH(1,1) model, here the unconditional variance is imposed as $\omega = 0.001$ and the lag coefficients are respectively $a = 0.5, b = 0.4$.

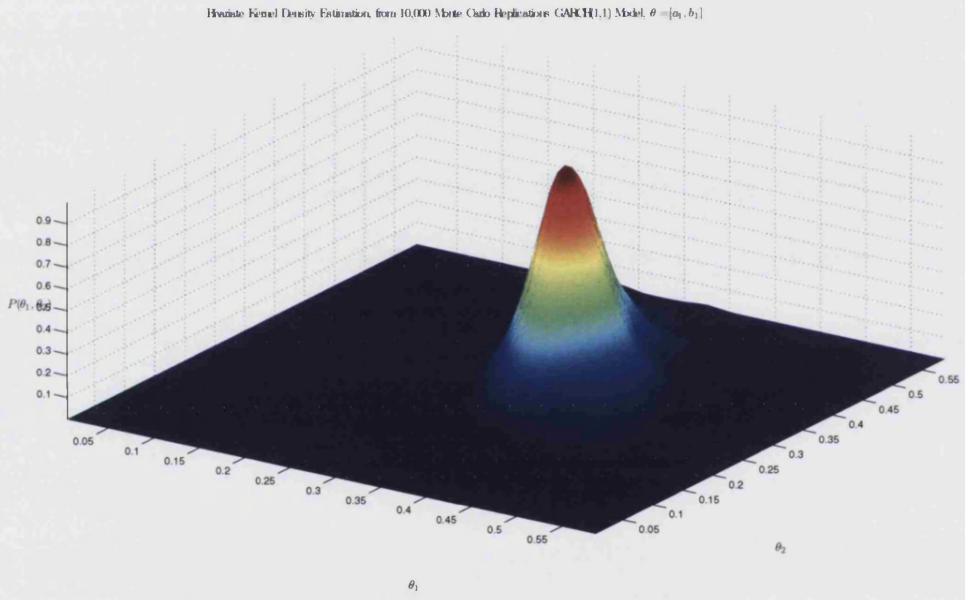

Figure 2.12: The Surface density of the parameter estimates from 10,000 simulations, the true parameter values are labeled. For a simple model such as this the SQP method with BGFS Hessian updating is the industry standard constrained optimization method and produces very consistent estimates.

Chapter 3

Non-Linear Dependency: in time series models: Copula Models

3.1 Chapter Abstract

The copula approach stems from Sklar's theorem, published in (1959) [240], which treats a multivariate distribution as a set of marginal distributions coupled by some joint coupling distribution. This is an extremely intuitive design, by which the dependency is generated by some multi-variate distribution such as a multi-normal or multi-t distribution, but taken in isolation marginal processes are drawn from some univariate distribution, such as an extreme value distribution or a non-central-f-distribution. This allows for a very easy system identification algorithm, identify individually the marginal distributions and parameterize them, then find joint inverse probability density function and fit this to a multivariate distribution.

3.2 A Brief Review of the Copula Literature

As previously stated the basic origin of the copula stems from Sklar (1959), up until 1997 however the term copula was not widely recognized in the statistics literature, something dwelt on at length in the introduction to Nelson's (1998) first real mainstream review of the underlying statistical theory. However the concept of dependence has been a subject of considerable interest in statistics, Jogdeo (1982) [154] notes that dependency is arguably the most studied topic in statistics. As such the practical use of copulas in finance with its great emphasis on dependency was soon forthcoming. A second review by Cherubini *et al* (2004) [51] offer a more finance orientated view to the basic theory, offering a variety of potential uses in the finance domain the most interesting being to the pricing of baskets of credit default obligations. However in the main they restrict themselves to the bivariate case and the main theoretical underpinnings. Moving away from the broad reviews of copula methods the first major practical applications for copula based methodologies were in the actuarial literature. Carriere (1994) [48] presents a compelling model of survival probabilities for coupled lives and uses this data to form an annuity portfolio. This well received work has formed the basis for a wide variety of insurance based applications. However an interesting off shoot in the finance based world is the idea of modelling the survival distributions of defaultable bonds. Which is covered in depth in Hamilton *et al* (2002), who review the likelihood of changes in credit rating and default on a

large database of bonds. The model uses a poisson based jump diffusion method to simulate arrivals in credit rating transitions. The model then utilizes an Ait-Mikhail-Haq copula to model the dependency between two bond processes and then calibrates the model against a real data-set. Monte-Carlo routines were then applied to price derivatives on single and multiple bond portfolios. Embrechts *et al* 2002 review the finance angle of dependency in terms of the value at risk approach discussing strong tail dependency in a variety of contexts. Patton (2005) [215] presents a portfolio model which utilizes skewness to capture some of the properties of dependency, he back tests his model using a bivariate portfolio with a skewed marginal distribution and measuring the conditional skewness using Hansen's (1994) [136] ARCD model, which is effectively a univariate GARCH model with univariate normal objective function replaced by the non-central t-distribution. The deformation parameter is then a function of the third moment of the distribution. Patton extends the work of Hansen by including a variety of bivariate copulae in constructing the dependency between the skewed marginal distributions. The logical extension to the conditional marginal distributions is the conditional copula proposed by Lee and Long (2005) who propose an extension of the dynamic correlation model of Engle (2001) [90] and Engle and Sheppard (2002) [89]. Their bivariate specification models the correlation between two variables, post standardization under the normal two step method. However unlike Patton, they now allow some of the copula parameters to be conditional and

re-specify the likelihood function accordingly.

3.2.1 Notation

This section uses the following notation,

$\mathfrak{F}$	Fraktur Capital F	Non explicit probability density function
$\mathfrak{H}$	Fraktur Capital H	Non explicit cumulative density function
$\mathfrak{C}$	Fraktur Capital C	Non explicit joint density function
$\mathfrak{G}$	Fraktur Capital G	Non explicit multivariate density function
$\mathfrak{L}$	Fraktur Capital L	Log likelihood function
$U(lb, ub)$		Uniform distribution between upper bound ub and 1ower bound lb
$N(\mu, \sigma)$		Univariate Normal distribution mean μ and variance σ^2
$N(\mu, \mathbf{\Sigma})$		Multivariate Normal distribution vector of means μ and covariance matrix $\mathbf{\Sigma}$
$T(v)$		Univariate $t-$ distribution degrees of freedom v
$T(v, \mathbf{R})$		Multivariate $t-$ distribution degrees of freedom v and correlation matrix $\mathbf{R}$
θ		Parameter vector
θ^*		Optimal parameter vector
$\iota', \iota'', \iota'''$		Cohorts of draws from a uniform distribution

3.2.2 The Copula Approach

A generalized copula framework could be conceived as follows, consider a draw from the i^{th} marginal distribution,

$$u_i^* \sim \mathfrak{F}(u_i | \theta_i) \tag{3.2.1}$$

where, u_i is drawn from a vector, where each element has its own marginal distribution,

$$\mathbf{u}^* = [u_1^*, ..., u_n^*]^{\mathrm{T}} \tag{3.2.2}$$

functionally inverting each of the marginal draws into a uniform $[0, 1]$ using the cumulative density function analogue of the marginal probability density function,

$$\begin{aligned} p_i &= \mathfrak{H}_i\left(u_i^* \,|\theta_i\right) && (3.2.3)\\ \mathfrak{H}_i &= \int_a^b \mathfrak{F}_i\left(u_i^* \,|\theta_i\right) dt && (3.2.4)\\ \mathfrak{F}_i &\in \mathbb{F} && (3.2.5)\\ p_i &\sim U\left(0, 1\right) && (3.2.6) \end{aligned}$$

Therefore the n-length vector $\mathbf{p}$ is drawn from an n dimensional unit hypercube, $[0, 1] \times [0, 1] \times ... \times [0, 1]$. Now consider that, each draw of the vector $\mathbf{p}$ is a functional transformation of some other random vector $\mathbf{u}$, where,

$$\begin{aligned} \mathbf{p} &= \int_{-\infty}^{\infty} \mathfrak{F}_{\mathfrak{C}}\left(\mathbf{u} \,|\theta_{\mathfrak{C}}\right) dt && (3.2.7)\\ \mathbf{u} &\sim \mathfrak{G}\left(\mathbf{u} \,|\theta_{\mathfrak{C}}\right) && (3.2.8)\\ \mathfrak{C}\left(\mathbf{p}\right) &= \mathfrak{F}\left(\mathfrak{G}\left(\mathbf{u} \,|\theta_{\mathfrak{C}}\right)\right) && (3.2.9)\\ \mathbf{p} &\sim \mathfrak{C}\left(\mathbf{p}\right) && (3.2.10) \end{aligned}$$

Therefore there is some continuous random vector $\mathbf{u}$ and a transformation,

$$\int_{-\infty}^{\infty} \mathfrak{F}_{\mathfrak{C}(\mathbf{u}|\theta_{\mathfrak{C}})} dt \tag{3.2.11}$$

that results in a random vector with dependent uniform elements. An example could be $\mathbf{u}$ is drawn from a zero centred multivariate Normal/Gaussian with normalized variance/covariance matrix $\mathbf{\Sigma}$, i.e. unit variance and the function $\mathfrak{F}_{\mathfrak{C}}$ is the standard normal cumulative density function. In this case $\mathbf{p}$ is drawn from a Gaussian copula, which is an elliptical copula.

Sklar's Theorem

The underlying concept behind the copula approach is the decomposition of a complex multivariate distribution into n marginal distributions and a single joint distribution.

$$\mathbf{u}^* = [u_1^*, ..., u_n^*]^{\mathrm{T}} \tag{3.2.12}$$

$$\mathfrak{G}(\mathbf{u}^*) = \mathfrak{C}(\mathbb{F}(\mathbf{u}^*)) \tag{3.2.13}$$

Where $\mathfrak{G}$ is some continuous multivariate distribution, $\mathfrak{C}$ is a continuous n-variate joint distribution and $\mathbb{F}$ is a set of n continuous marginal distributions, where the i^{th} marginal, $\mathfrak{H}_i \in \mathbb{F}$.

Elliptical Copulas

Copulas are divided into two basic typologies, those which are based on identified multivariate distributions such as the multi normal/Gaussian or multi students-t, generally referred to as Elliptical copulas, and those based on specific invertible functions, Archimedean copulas.

Example: Generating Random numbers from an n-variate Gaussian Copula

To create a draw of random numbers drawn from a bivariate copula, with univariate normal marginal distribution and a multi-normal joint distribution is simple, first generate a multi-normal distribution, compute the cumulative marginal probabilities and then reconstruct the marginal's from the inverse density function. Therefore to start with a vector of uniformly distributed random numbers $U(0,1)$, a common uniform generator is the linear congruential generator, however there is no such thing as a genuinely random set of numbers, power spectral density analysis on random double precision uniform [0,1] numbers generated by a linear congruential generator will yield distinct peaks from a very large number of draws.

$$\iota_i \sim [0,1] \tag{3.2.14}$$

$$\iota = [\iota_1, ..., \iota_n]^{\mathrm{T}} \tag{3.2.15}$$

$$\varepsilon_i = \mathfrak{F}^{-1}(\iota_i | \mu_i = 0, \sigma_i = 1) \tag{3.2.16}$$

$$= \{\varepsilon_i : \mathfrak{F}(\varepsilon_i | \mu_i = 0, \sigma_i = 1) = \iota_i\} \tag{3.2.17}$$

$$\mathfrak{F}(\varepsilon_i | \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{u_i} \exp\left(\frac{-(t-\mu_i)^2}{2(\sigma_i)^2}\right) dt \tag{3.2.18}$$

$$\varepsilon \sim N(\mathbf{0}, \mathbf{I}) \tag{3.2.19}$$

Next generate the multi-normal draws in the usual way,

$$\mathbf{u} = \boldsymbol{\Sigma}^{*\frac{1}{2}}\varepsilon + \mu^* \tag{3.2.20}$$

$$\mathbf{u} \sim N(\mu^*, \boldsymbol{\Sigma}^*) \tag{3.2.21}$$

$$\mathfrak{C}(\mathbf{u}) = \frac{1}{\sqrt{2\pi^n |\boldsymbol{\Sigma}^*|}} \exp\left(-\tfrac{1}{2}\left((\mathbf{u}-\mu^*)^{\mathrm{T}} (\boldsymbol{\Sigma}^*)^{-1} (\mathbf{u}-\mu^*)\right)\right) \tag{3.2.22}$$

Decomposing the multi-normals by the diagonal of the covariance matrix and utilize the univariate Gaussian/normal cdf to construct a probability vector this may then be transformed by the inverse cdf of the marginal, in this the univariate inverse Gaussain/normal cdf.

$$\left[\sigma^*_{i,j}\right] = \boldsymbol{\Sigma}^* \tag{3.2.23}$$

$$\sigma^*_i = \sqrt{\sigma^*_{i,i}} \tag{3.2.24}$$

$$p_i = \mathfrak{H}_i(u_i | \mu^*_i, \sigma^*_i) = \frac{1}{\sigma^*_i\sqrt{2\pi}} \int_{-\infty}^{u_i} \exp\left(\frac{-(t-\mu^*_i)^2}{2(\sigma^*_i)^2}\right) dt \tag{3.2.25}$$

$$u^*_i = \mathfrak{H}_i^{-1}(p_i | \mu^*_i, \sigma^*_i) = \{\varepsilon_i : \mathfrak{H}(\varepsilon_i | \mu^*_i, \sigma^*_i) = p_i\} \tag{3.2.26}$$

$$\mathbf{u}^* \sim \mathfrak{G}(\mathbf{u}^*) \tag{3.2.27}$$

The vector draw $\mathbf{u}^*$ is now drawn from an n-variate Gaussian copula with univariate Gaussian/normal marginal distributions. There is no restriction to the normal/multinormal distribution and any joint/marginal combination may be used as long as the cumulative density function of the marginal distribution is invertible.

Example: Generating Random numbers from a bivariate copula with distinct marginal distributions

Consider a bivariate case, where the joint distribution is a bivariate students t with 2 degrees of freedom and the marginal distributions are respectively an extreme value distribution and a central f-distribution. The multivariate students-t is simple to generate, first generate a multi-normal distribution and then divide by a random vector drawn from a normalized central χ^2 with appropriate degrees of freedom. The covariation in a multivariate students-t is defined by a correlation matrix, a normalized covariance matrix therefore,

$$\mathbf{R} = [\rho_{i,j}] \tag{3.2.28}$$

$$\rho_{i,j} = \frac{\sigma_{i,j}}{\sqrt{\sigma_{i,i}}\sqrt{\sigma_{j,j}}} \tag{3.2.29}$$

An interesting note on the properties of a correlation matrix for $\mathbf{R}$ to be a valid correlation matrix, the following must hold, $\sum_{i=1}^{n} \lambda_i = n$, where λ_i is the i^{th} eigenvalue of $\mathbf{R}$. The first step is to generate two cohorts,ι' and ι'' of uniform random numbers,

$$\iota'_i \sim U(0,1) \tag{3.2.30}$$

$$\iota''_i \sim U(0,1) \tag{3.2.31}$$

$$\iota' = [\iota'_1, ..., \iota'_n]^{\mathrm{T}} \tag{3.2.32}$$

$$\iota'' = [\iota''_1, ..., \iota''_n]^{\mathrm{T}} \tag{3.2.33}$$

The first cohort, will generate a zero centred n-length multi-normal vector and the second will generate an n-length vector of $\chi^2(v)$ random variables, where v is the number of degrees of freedom of the objective multi students-t distribution.

$$\begin{aligned}
\varepsilon_i' &= \mathfrak{F}'^{-1}(\iota_i' | \mu_i = 0, \sigma_i = 1) && (3.2.34)\\
&= \{\varepsilon_i : \mathfrak{F}'(\varepsilon_i' | \mu_i = 0, \sigma_i = 1) = \iota_i'\} && (3.2.35)\\
\mathfrak{F}'(\varepsilon_i' | \mu_i, \sigma_i) &= \frac{1}{\sigma_i\sqrt{2\pi}} \int_{-\infty}^{\varepsilon_i'} \exp\left(\frac{-(t-\mu_i)^2}{2(\sigma_i)^2}\right) dt && (3.2.36)\\
\varepsilon_i' &\sim N(0,1) && (3.2.37)\\
\varepsilon' &\sim N(\mathbf{0}, \mathbf{I}) && (3.2.38)\\
\mathbf{u}' &= \mathbf{R}^{\frac{1}{2}}\varepsilon' && (3.2.39)\\
\mathbf{u}' &= [u_1', ..., u_n']^{\mathrm{T}} && (3.2.40)\\
\varepsilon_i'' &= \mathfrak{F}''^{-1}(\iota_i'' | v = 2) && (3.2.41)\\
&= \{\varepsilon_i'' : \mathfrak{F}''(\varepsilon_i'' | v = 2) = \iota_i''\} && (3.2.42)\\
\mathfrak{F}''(\varepsilon_i'' | v = 2) &= \int_0^{\varepsilon''} \frac{t^{\frac{1}{2}(v-2)} \exp\left(-\frac{1}{2}t\right)}{2^{\frac{1}{2}v}\Gamma\left(\frac{1}{2}v\right)} && (3.2.43)\\
\varepsilon_i'' &\sim \chi^2(2) && (3.2.44)
\end{aligned}$$

Dividing the two newly constructed random vectors element by element yields a vector drawn from an n-variate multi students-t,

$$u_i = u_i'\left(\varepsilon_i''\right)^{-1} \tag{3.2.45}$$

$$\mathbf{u} = [u_1, ..., u_n]^{\mathrm{T}} \tag{3.2.46}$$

$$\mathbf{u} \sim T(v, \mathbf{R}) \tag{3.2.47}$$

$$\mathfrak{C}(\mathbf{u}\,|v, \mathbf{R}) = T(v, \mathbf{R}) \tag{3.2.48}$$

$$p_i = \mathfrak{F}(u_i\,|v) = \int_{-\infty}^{u_i} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{v\pi}} \frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}} dt \tag{3.2.49}$$

$$\mathbf{p} = [p_1, ..., p_n] \equiv [p_1, p_2] \tag{3.2.50}$$

From the cdf of the univariate students t the [0,1] uniform vector is computed and the inverse cdf's of the marginal distribution is used to reconstruct the variable, in the bivariate case this yields,

$$u_1^* = \mathfrak{H}_1^{-1}(p_1\,|v_1, v_2) = \{u_1 : \mathfrak{H}_1(u_1\,|v_1, v_2) = p_1\} \tag{3.2.51}$$

$$p_1 = \mathfrak{H}_1(u_1\,|\mu, \sigma) \tag{3.2.52}$$

$$= \int_{-\infty}^{u_1} \sigma^{-1} \exp\left(\frac{t-\mu}{\sigma}\right) \exp\left(-\exp\left(\frac{t-\mu}{\sigma}\right)\right) dt$$

$$u_2^* = \mathfrak{H}_2^{-1}(p_2\,|v_1, v_2) = \{u_2 : \mathfrak{H}_2(u_2\,|v_1, v_2) = p_2\} \tag{3.2.53}$$

$$p_2 = \mathfrak{H}_2(u_1\,|v_1, v_2) \tag{3.2.54}$$

$$= \int_0^{u_2} \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_2}{v_2}\right)^{\left(\frac{v_1}{2}\right)} \frac{t^{\left(\frac{v_1+v_2}{2}\right)}}{\left(1+\left(\frac{v_2}{v_1}\right)t\right)^{\frac{v_1+v_2}{2}}} dt$$

Therefore the vector, $\mathbf{u}^* = [u_1^*, u_2^*]$ has two marginal distributions the first drawn from an extreme value distribution u_1^* and the second u_2^* from a central f-distribution.

Copula Estimation Methodologies for Elliptical Copulas

We have shown how it is possible to generate random draws from a variety of Elliptical copulae, the next step is to identify and parameterize Elliptical copulas from vector data. Consider a sample of vector data drawn from from some unknown multivariate distribution, $\mathbf{u} \in \mathbb{R}^n$, which maybe decomposed into n-marginal distributions and one joint distribution, under Sklar's theorem.

$$\mathbf{u}_m = [u_{1,m}, ..., u_{n,m}]^{\mathrm{T}} \tag{3.2.55}$$

$$\mathbf{u}_m \in \mathcal{U} \tag{3.2.56}$$

$$\mathcal{U} = [\mathbf{u}_{m=1}, ..., \mathbf{u}_{m=\tau}] \tag{3.2.57}$$

$$\mathbf{u}_m \sim G(\mathbf{u}_m | \mathbb{Q}) \tag{3.2.58}$$

$$\mathbb{Q} = \{\theta_1, ..., \theta_n, \theta_{\mathfrak{C}}\} \tag{3.2.59}$$

The set $\mathbb{Q}$ contains the parameter vectors for the n marginal distributions θ_i and the joint distribution $\theta_{\mathfrak{C}}$. From the sample $\mathcal{U}$, $\mathbf{u}_m$ is the m^{th} draw from τ draws of the random vector $\mathbf{u}$. Our objective is to form a likelihood function $\mathfrak{L}(\mathbb{Q})$ for which,

$$\max_{\mathbb{Q}} \left(\mathfrak{L}\left(\mathfrak{H}_1\left(\theta_1\right), ..., \mathfrak{H}_n\left(\theta_n\right), \mathfrak{C}\left(\theta_{\mathfrak{C}}\right)\right)\right) \tag{3.2.60}$$

The simplest method is to breakdown the optimization into two distinct steps, first identify the marginal distributions, fit their parameters via maximum likelihood and

convert the marginal distributions into n [0,1] uniform distributions, via the marginal cumulative density function.

$$\mathfrak{F}_i(\theta_i^*) \triangleq \max_{\theta_i}(\log \mathfrak{F}_i(\theta_i)) \tag{3.2.61}$$

$$p_i = \mathfrak{H}_i(u_i, \theta_i^*) \tag{3.2.62}$$

$$p_i \sim U(0,1) \tag{3.2.63}$$

$$\mathbf{p} = [p_1, ..., p_n]^{\mathrm{T}} \tag{3.2.64}$$

In the second step each element of the vector of probabilities is then converted via the joint distributions univariate analogue's inverse cdf and then the joint distribution is again fitted via maximum likelihood.

$$\mathcal{P} = [\mathbf{p}_1, ..., \mathbf{p}_\tau] \tag{3.2.65}$$

$$u_i^* = \mathfrak{H}_{\mathfrak{U}}(p_i | \theta_{\mathfrak{U}}) = \{u_i^* : \mathfrak{H}_{\mathfrak{U}}(u_i^* | \theta_{\mathfrak{U}}) = u_i^*\} \tag{3.2.66}$$

$$\mathcal{U}^* = [\mathbf{u}_1^*, ..., \mathbf{u}_\tau^*] \tag{3.2.67}$$

$$\mathfrak{C}(\theta_{\mathfrak{C}}^*) \triangleq \max_{\theta}(\mathfrak{L}(\mathfrak{C}(\theta))) \tag{3.2.68}$$

The parameter vector of the univariate inverse density function, $\theta_{\mathfrak{U}}$, are free parameters, although they should be uniform for each element of $\mathbf{p}$.

3.2.3 Problems with the two step Approach

The basic problem is system identification, i.e. which marginal/joint combination to use, once the system is identified then parameterization is simple, using the two step

MLE method. The major difficulty when the length of $\mathbf{u}$ exceeds $n = 3$, is identifying the joint distribution. However there are only four viable multivariate distributions to choose from, the multi-normal, multi-t, the multi-gamma and in exceptional cases the Multivariate Dirichlet distribution. A comparison test between the fit of each possible joint distribution maybe undertaken using the stratification techniques, which will be addressed later in this chapter.

3.3 Empirical Copulas

An alternative approach is to evaluate the copulas empirically, this maybe employed at either stage of the two step approach. Again the first step is to construct the marginal cdf and convert each vector innovation of the marginal distribution into probability vectors. This maybe undertaken via a simple histogram approach or via kernel density smoothing. Once the marginal distributions have been converted into [0,1] uniform densities the copula is then modelled directly. First the n dimensional hypercube is constructed, containing the stratification boundaries, second hypercube with elements containing the population abundances for each of the stratification cells. This gives an n dimensional probability array, to construct a Monte-Carlo routine, first generate a vector of independent $U(0, 1)$ draws and then couple those via the probability array. The discrete strata are made continuous by assuming that a draw from anywhere in the strata is equally likely. This approach has none of the drawbacks of the maximum likelihood approach, i.e. distributional assumptions, but is extremely computationally

intensive and requires a large amount of data, for example a Monte-Carlo simulation of a 10-variate system, with 100 stratifications, requires a stratified hypercube with 100,000,000,000,000,000,000 elements, in addition to the 100 stratifications for each of the univariate empirical cdfs. The following example looks at a bivariate example, first consider the two marginal distributions,

$$\mathbf{u} = [u_1, u_1]^{\mathrm{T}} \tag{3.3.1}$$

$$u_1 \sim \mathfrak{F}_1(u_1) \tag{3.3.2}$$

$$u_2 \sim \mathfrak{F}_2(u_2) \tag{3.3.3}$$

$$p_1 = \mathfrak{H}_1(u_1) \tag{3.3.4}$$

$$p_2 = \mathfrak{H}_2(u_2) \tag{3.3.5}$$

$$\mathbf{p} = [p_1, p_1]^{\mathrm{T}} \tag{3.3.6}$$

The probability density array is therefore a matrix $\mathcal{C}$, that operates over an array $\mathcal{S}$, with S stratifications,

$$\mathcal{S} = \begin{bmatrix} \begin{bmatrix} (0,0.2) \\ (0,0.2) \end{bmatrix} & \begin{bmatrix} (0,0.2) \\ (0.2+\delta,0.2) \end{bmatrix} & \cdots & \begin{bmatrix} (0,0.2) \\ (0.8+\delta,1) \end{bmatrix} \\ \begin{bmatrix} (0.2+\delta,0.4) \\ (0,0.2) \end{bmatrix} & \begin{bmatrix} (0.2+\delta,0.4) \\ (0.2+\delta,0.2) \end{bmatrix} & \cdots & \begin{bmatrix} (0.2+\delta,0.4) \\ (0.8+\delta,1) \end{bmatrix} \\ \vdots & \vdots & \ddots & \vdots \\ \begin{bmatrix} (0.8+\delta,1) \\ (0,0.2) \end{bmatrix} & \begin{bmatrix} (0.8+\delta,1) \\ (0.2+\delta,0.2) \end{bmatrix} & \cdots & \begin{bmatrix} (0.8+\delta,1) \\ (0.8+\delta,1) \end{bmatrix} \end{bmatrix} \tag{3.3.7}$$

assuming that,

$$\begin{aligned}\mathcal{S} &= [s_{i,j}] \\ \mathcal{C} &= \begin{bmatrix} \mathbb{P}(s_{1,1}) & \mathbb{P}(s_{1,2}) & \cdots & \mathbb{P}(s_{1,S}) \\ \mathbb{P}(s_{2,1}) & \mathbb{P}(s_{2,2}) & \cdots & \mathbb{P}(s_{2,S}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{P}(s_{S,1}) & \mathbb{P}(s_{S,2}) & \cdots & \mathbb{P}(s_{S,S}) \end{bmatrix}\end{aligned}$$

Where δ is the smallest double precision floating operator, the real minimum. More generally for S stratifications and an n-variate vector process, using the Khatri-Rao product to vectorize the hypercube,

$$\underset{k\times m}{\mathbf{M}} \odot \underset{l\times m}{\mathbf{N}} = [\mathbf{m}_1 \otimes \mathbf{n}_1, ..., \mathbf{m}_m \otimes \mathbf{n}_m] \tag{3.3.8}$$

$$\mathbf{m}_1 = [m_{1,1}, ..., m_{k,1}]^{\mathrm{T}} \tag{3.3.9}$$

$$\mathbf{n}_1 = [n_{1,1}, ..., n_{l,1}]^{\mathrm{T}} \tag{3.3.10}$$

The Khatri-Rao product allows for the creation of an indexing matrix, $\mathbf{A}$ which is then used to systematize the construction of $\mathcal{C}$,

$$\mathbf{s} = [1, ..., S]^{\mathrm{T}} \tag{3.3.11}$$

$$\mathbf{a}_i = \left[\left(\underset{S^i\times 1}{\mathbf{e}} \odot \underset{S\times 1}{\mathbf{s}} \right)_1, ..., \left(\underset{S^i\times 1}{\mathbf{e}} \odot \underset{S\times 1}{\mathbf{s}} \right)_{n-i} \right]^{\mathrm{T}} \tag{3.3.12}$$

$$i \in [1, .., n] \tag{3.3.13}$$

$$\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_n] \tag{3.3.14}$$

The address matrix is characterized as,

$$\underset{S^n \times n}{\mathbf{A}} = \begin{bmatrix} 1 & \cdots & 1 & 1 & 1 \\ 1 & \cdots & 1 & 1 & 2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & 1 & 1 & S \\ 1 & \cdots & 1 & 2 & 1 \\ 1 & \cdots & 1 & 2 & 2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & 1 & S-1 & S \\ 1 & \cdots & 1 & S & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S & \cdots & S & S & S-1 \\ S & \cdots & S & S & S \end{bmatrix} \tag{3.3.15}$$

Therefore each element of $\mathcal{S}$ and $\mathcal{C}$ is indexed in terms of a vector $\dashv_j$, which is a row from $\mathbf{A}$.

$$\mathbf{a}_j = [a_{j,1}, ..., a_{j,n}] \tag{3.3.16}$$

therefore in vector notation,

$$j \in [1, ..., S^n] \tag{3.3.17}$$

$$\mathcal{S} = [s_j] \tag{3.3.18}$$

$$\mathcal{C} = [c_j] \tag{3.3.19}$$

each element of $\mathcal{S}$ and $\mathcal{C}$ is therefore,

$$s_j = \begin{bmatrix} \left(\frac{1}{S}(a_{j,1}-1)+\delta, \frac{1}{S}a_{j,1}\right) \\ \left(\frac{1}{S}(a_{j,2}-1)+\delta, \frac{1}{S}a_{j,2}\right) \\ \vdots \\ \left(\frac{1}{S}(a_{j,n}-1)+\delta, \frac{1}{S}a_{j,n}\right) \end{bmatrix} \tag{3.3.20}$$

$$c_j = \mathbb{P}(s_j) \tag{3.3.21}$$

The following constraint is necessary to ensure $\mathcal{C}$ is a valid density array,

$$\mathbf{e}^{\mathrm{T}} vec\,(\mathcal{C}) = 1 \tag{3.3.22}$$

Therefore if each element in the probability array $\mathcal{C}$ is given equal weight, the sum of those probabilities must be one. The probability density array $\mathcal{C}$ is then vectorized, the empirical cumulative density array, $\mathcal{H}$ is then computed,

$$\mathcal{H} = \sum \downarrow vec\,(\mathcal{C}) \tag{3.3.23}$$

The symbol $\downarrow$ is used to represent the sorting of the vectorized array and $\sum$ to represent the cumulative sum of the sorted array. Once the probability density array is estimated, generating random numbers is very easy, simply generate a 1×1 draw from a uniform $U(0,1)$ density function, the closest element in the cumulative density array $\mathcal{H}$ locates the position of the draw in $\mathcal{S}$, if any of the γ elements are equally probable then there is an $\frac{1}{\gamma}$ chance of drawing them. Once we know which strata in $\mathcal{S}$ we are in then we draw the vector from an n-variate uniform distribution, with limits defined by the strata array, e.g. if in the bivariate example $\mathcal{C}$ is found to be as follows,

$$\mathcal{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.05 & 0.05 & 0.15 \\ 0 & 0.2 & 0.3 & 0.1 & 0 \\ 0 & 0.1 & 0.05 & 0.05 & 0 \\ 0.15 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.3.24}$$

Therefore the empirical cumulative density function is as follows,

$$\mathcal{H}= \begin{bmatrix} 0,0,0,0,0,0,0,0,0,0,0,... \\ 0.05,0.1,0.15,0.2,0.3,... \\ 0.4,0.5,0.7,1 \end{bmatrix} \tag{3.3.25}$$

Therefore a single draw from a $U(0,1)$ of 0.65, will place the draw in the cell, designated $s_{3,3}$,

$$s_{3,3} = \begin{bmatrix} (0.4+\delta,0.6) \\ (0.4+\delta,0.6) \end{bmatrix} \tag{3.3.26}$$

Therefore the resulting random vector will be a draw,

$$\mathbf{u} = \begin{bmatrix} u_1 \sim U\left(0.4+\delta,0.6\right) \\ u_2 \sim U\left(0.4+\delta,0.6\right) \end{bmatrix} \tag{3.3.27}$$

The more stratifications then the less pronounced the natural stepping effect caused by this type of stratification. This methodology is a specific extension of Latin hypercube sampling, which is commonly used in Monte-Carlo analysis to increase the convergence of sample paths of latent stochastic processes.

Empirical Example

A typical empirical copula application could be as follows, suppose a structurer has written a knock-out call option on a basket of bonds, the knockout option includes a complex set of n-clauses on a variety of macroeconomic and financial variables the counter-party wishes to know the probability that any of the knock-out clauses will be enacted over the lifetime of the option. This is a common problem in modern

finance, as derivative instruments become more complex, analyzing the risk of such instruments becomes increasingly complicated. The following is a real trade between a large bank, the issuer, and a smaller bank from an emerging market, the counter-party. The option is a call on a currency swap, in six months, the issuer incorporates a series of knock-out clauses which invalidates the option if at any given time over the period of the option they are activated. Suppose there are n-instruments, the clauses will activate if any of the instruments moves below a certain value, for ease of notation all of the clauses are specified as a floors. In addition to the individual clauses, if the total weighted value of the instruments drops below a certain level then another knock-out clause activates.

$$\mathbf{u}_t = [u_{1,t}, ..., u_{n,t}]^{\mathrm{T}} \tag{3.3.28}$$

$$\omega = [\omega_{1,t}, ..., \omega_{n,t}]^{\mathrm{T}} \tag{3.3.29}$$

$$v_t = \omega^{\mathrm{T}} \mathbf{u}_t \tag{3.3.30}$$

$$\xi = [\xi_1, ..., \xi_n, \xi_v] \tag{3.3.31}$$

The probability measure is therefore,

$$\mathbb{P}\left(\bigcup\left\{\begin{array}{l} u_{1,t} \leq \xi_1 : \{t = 1, ..., t = T\}, ... \\ u_{n,t} \leq \xi_n : \{t = 1, ..., t = T\}, \\ v_t \leq \xi_v : \{t = 1, ..., t = T\} \end{array}\right\}\right) \tag{3.3.32}$$

Taking the sample history of $\mathbf{u}_t$, constructing the hypercube probability measure and then use Monte-Carlo simulation to construct a large number of sample paths, the

proportion of sample paths that activate the clauses, gives the empirical probability of invalidation of the option.

3.3.1 Smoothing the Hypercube

One of the major problems with this methodology is that as the number of dimensions increases, the spacial density decreases, for a set number of observations, this will mean that blank spots in the hypercube are more likely which will result in more pronounced stepping effects in the monte-carlo pathways. When estimating the marginal distributions and compute the cumulative density functions and their quadrature derived inverses, a smoothing kernel is utilized to reduce the stepping effect caused by zero populated elements in the main probability mass. Given a suitable number of observations the mono-dimensional nature makes this smoothing largely redundant, however for the stratified hypercube the number of observations is soon outstripped by the number of elements in the hypercube itself. To this end we turn to a technique known as n-dimensional tensor produce splines. The proposed methodology tackles the smoothing issues in two ways, first by running a simple moving average smoother over the hypercube and re-normalizing the probability densities. And second by looking at clusters of gaps, particularly prevalent in very high dimensional systems. The second method requires a variety of assumptions, first how many high probability mass centers there are in the hypercube, second the rate of functional decay to zero probability mass at the boundaries of these high mass centers. If the data generating

process is from an elliptical copula then this is relatively simple, the decay is exponential and the number of high probability mass is under most circumstances normally one, unless the largest eigenvalue of the non-negative dependency matrix is very low and the probability mass is almost uniformly diffuse across the hypercube.

The Exponential Smoother

Consider a window described by a hypercube $\mathcal{W}_i$ with dimension of length, $\omega = \frac{1}{\alpha}S$, where S is the number of stratifications of $\mathcal{C}$ and i is the i^{th} element of the total elements of $\mathcal{C}$, confining ω to being an odd number then at the center of $\mathcal{W}_i$ is a unique element of $\mathcal{C}$, w_i. We can define the packing of the elements around ω in terms of the address matrix $\mathbf{A}$, described previously. We can now adjust smooth the value of w_i by means of an exponentially weighted average of itself and its nearest neighbours. For an element w_i, with address vector, $\mathbf{a}_i$, the hypercube address matrix will be,

$$\mathbf{A}_{\mathcal{W}_i} = \begin{bmatrix} \mathbf{a}_i \\ \mathbf{a}_i + \delta_1 \\ \mathbf{a}_i + \delta_2 \\ \vdots \\ \mathbf{a}_i + \delta_{\omega^n} \end{bmatrix} \tag{3.3.33}$$

Where the vector δ_j is the j^{th} row of the matrix, $\mathbf{\Delta}$.

$$\underset{\varsigma^n \times n}{\mathbf{\Delta}} = \begin{bmatrix} 0 & \cdots & 0 & -\varsigma \\ 0 & \cdots & -\varsigma & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\varsigma & \cdots & 0 & 0 \\ 0 & \cdots & 0 & -\varsigma+1 \\ \vdots & \vdots & \vdots & \vdots \\ -\varsigma+1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \varsigma & \cdots & 0 & 0 \end{bmatrix} \tag{3.3.34}$$

where,

$$\varsigma = \tfrac{1}{2}(\omega + 1) \tag{3.3.35}$$

For those elements within ς of the boundary aperiodic (torus) boundary conditions for the smoothing are used, (i.e. dampened reflexivity).

$$w_i = \sum_{j=1}^{\omega^n - 1} \alpha \exp\left(\beta w_j^{\gamma}\right) \tag{3.3.36}$$

Chapter Appendix III illustrates the effect of smoothing on the empirical estimation of a bivariate t-copula and a bivariate Gumbel copula. The marginal distributions are both back computed from univariate normals. Its easy to see that the representation of the smoothed copula gives a much more realistic representation of the probability mass density function.

Testing the Robustness of Monte-Carlo Pathways from an Empirical Copula

The reliability of pathways may be tested using a simple monte-carlo procedure, a random sample from an standard elliptical copula is used as the test data set, the empirical copula is then estimated and smoothed, a new large set of random draws is generated from the empirical copula and the original copula specification is parameterized from these draws using maximum likelihood, the velocity of convergence between the original and estimated parameters gives an indication of the power of the technique.

Monte Carlo Testing Methodology

First generate sets of draws $\mathbf{u}_\xi \in \mathcal{U}_\xi$ of varying lengths from a real t-copula with true parameter vector $\hat{\theta}$, second fit the stratified hypercube and generate a new fixed length (very large) set of draws $\mathbf{u}^*_\zeta \in \mathcal{U}_\zeta$ and then from this new set of draws estimate the parameters of the original copula specification, θ^*, then compute the naive quadratic deviation from the true parameter vector, $\theta^*_i - \hat{\theta}_i$ and compare this to a maximum likelihood fit on the original sample, θ'^* and its naive quadratic deviation from the true vector, $\theta'^*_i - \hat{\theta}_i$, where subscript i denotes the i^{th} parameter.

Monte-Carlo Test

The example here, uses a 20-variate t-copula with a variety of marginal distributions, taken from the normal, log-normal, extreme value (basic Paretian) and f-distribution.

The marginal cumulative density functions were estimated using kernel density estimation with two hundred and fifty nodes and Epanechnikov smoothing, see Moral *et al* (2002). The copula was estimated using two hundred and fifty stratifications per variate dimension. This yields, 250^{20} elements within the stratified copula hypercube, (most of the elements are zero or near zero). The experiment was conducted on generated data of lengths between 10,000 and 10,000,000 observations. Estimation took 6 days using a quad processor Xeon workstation, in Linux. Figure 3.1 clearly shows that as expected the MLE estimated parameters of the Monte-Carlo pathways from the stratified copula hypercube converge to the true value as the sample size increases past 10,000 observations. Figure illustrates the performance of the maximum likelihood estimator with varying sample size and demonstrates that even with a known system identification, parametrization converges monotonically to the true estimates only for sample sizes of over $\sim 5,000$.

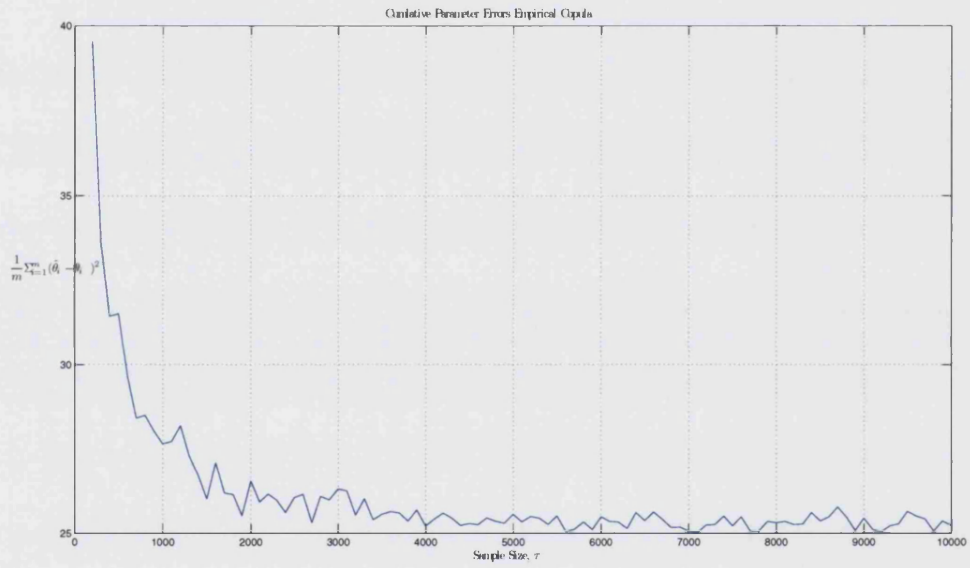

Figure 3.1: The cumulative parameter errors with increasing sample size for an empirical copulas.

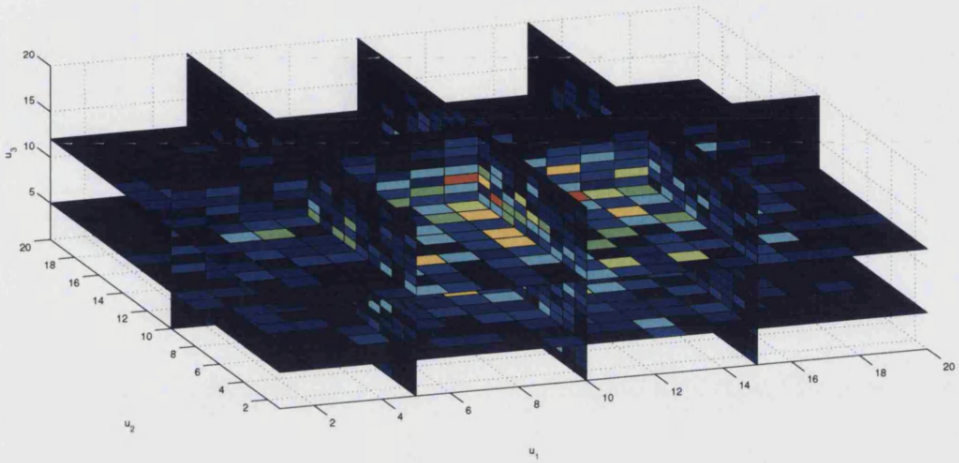

Figure 3.2: 3D Depiction of the Empirical Copula, the color designates the intensity of the density mass in the Unit Cube

3.4 Non-Spherical Copulas, the Archimedean Copula

Archimedean copulas use generator functions to impart dependency between marginal distributions. Consider the Laplace transform,

$$\mathfrak{L}(s) = \int_0^\infty \mathfrak{X}(t)\, e^{-st} dt \tag{3.4.1}$$

$$\mathfrak{P}(v) = \int_0^\infty \mathfrak{C}(u)\, e^{-uv} du \tag{3.4.2}$$

$$\mathfrak{P}(\mathfrak{C}(u)) = \int_0^\infty \mathfrak{C}(u)\, e^{-uv} du \tag{3.4.3}$$

For any given generator function $\mathfrak{F}$, the copula function is defined as follows,

$$\mathfrak{C}(u) = \mathfrak{X}^{-1}\sum_{i=1}^{n}\mathfrak{X}(u_i) \tag{3.4.4}$$

$$\sum_{i=1}^{n}\mathfrak{X}(u_i) \leq \mathfrak{X}(0) \tag{3.4.5}$$

To be a valid copula generating function $\mathfrak{F}$ must be a positive Laplace transform of $\mathfrak{C}(u)$. The following, non-exhaustive, list gives the commonly used generating functions, their inverses and the bivariate form of the Archimedean copula, those listed here have been utilized in the finance literature and include copulas from Hamilton *et al* (1994) [135] and Patton (2002) [215]. For a fuller list of important Archimedean Copulas see Nelson (1991) [205].

The Clayton Copula

Generator function,

$$\mathfrak{X}(u) = (u^{-\theta} - 1) \tag{3.4.6}$$

$$\mathfrak{X}^{-1}(u) = (1+u)^{-\frac{1}{\theta}} \tag{3.4.7}$$

Parameter domain,

$$\theta > 0 \tag{3.4.8}$$

The bivariate version is defined as,

$$\mathfrak{C}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{3.4.9}$$

The Gumbel Copula

Generator function,

$$\mathfrak{X}(u) = (-\ln u)^{\theta} \tag{3.4.10}$$

$$\mathfrak{X}^{-1}(u) = \exp\left(-u^{\frac{1}{\theta}}\right) \tag{3.4.11}$$

Parameter domain,

$$\theta > 1 \tag{3.4.12}$$

The bivariate version is defined as,

$$\mathfrak{C}(u_1, u_2) = \exp\left(-\left((-\ln(u_1))^{\theta} + (-\ln(u_2))^{\theta}\right)^{\frac{1}{\theta}}\right) \tag{3.4.13}$$

The Franks Copula

Generator function,

$$\mathfrak{X}(u) = -\ln\left(\frac{\exp(-\theta u) - 1}{\exp(\theta) - 1}\right) \tag{3.4.14}$$

$$\mathfrak{X}^{-1}(u) = -\tfrac{1}{\theta}\left(\ln 1 - \exp(-u)\left(1 - \exp(-\theta)\right)\right) \tag{3.4.15}$$

Parameter domain,

$$\theta > 1 \tag{3.4.16}$$

The bivariate copula is defined as,

$$\mathfrak{C}(u_1, u_2) = \theta^{-1} \ln\left(1 + \frac{(\exp(\theta u_1) - 1)(\exp(\theta u_2) - 1)}{(\exp(\theta) - 1)}\right) \tag{3.4.17}$$

Ait-Mikhail-Haq Copula

Generator function,

$$\mathfrak{X}(u) = \ln\left(\frac{1-\theta(1-u)}{u}\right) \tag{3.4.18}$$

$$\mathfrak{X}^{-1}(u) = \frac{\exp(u)}{1-\theta(1-u)} \tag{3.4.19}$$

Parameter domain,

$$-1 \leq \theta > 1 \tag{3.4.20}$$

The bivariate copula is defined as,

$$\mathfrak{C}(u_1, u_2) = \frac{u_1 u_2}{1-\theta(1-u_1)(1-u_2)} \tag{3.4.21}$$

The Archimedean copulas can produce very complex tail dependencies in the bivariate case, however in the n-variate form much of the advantages break down. There is no cross dependency and each pair of the n-variate marginal distributions has the identical dependency structure.

3.5 Chapter Appendix: Some Copula Dependency Examples

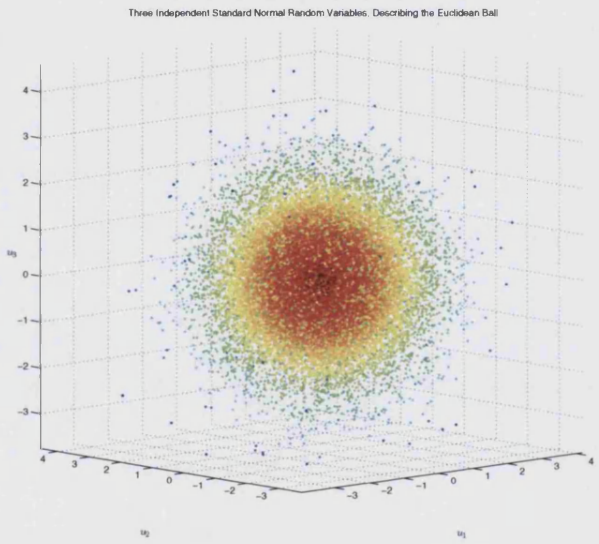

Figure 3.3: IID Tri-variate Normal distributions, i.e. $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I})$, when plotted describe the contours of the Euclidean ball, deformations from this spherical structure are demonstrative of complex multivariate dependence. However extreme values maybe obtained even with no explicit tail correlation.

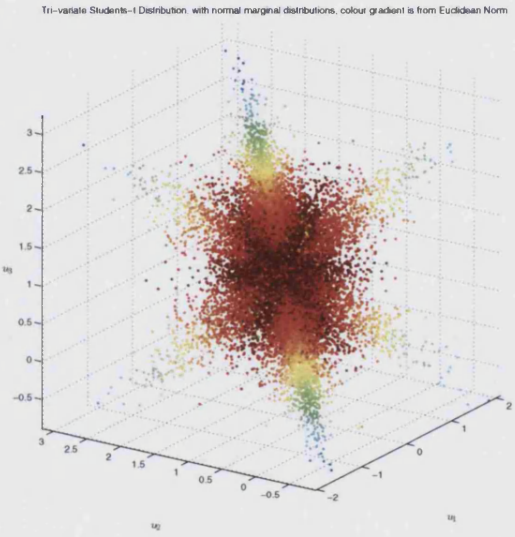

Figure 3.4: In this case the marginal distributions are IID normal and the joint distribution is a tri-variate students-t, however dependency structure is still defined by the identity matrix $\mathbf{I}$, he nature of the H-function, from which the multi-t is derived means that there exists tail domains, where extreme values are far more likely than under the multi-normal

Chapter 4

Co-movement and Contagion: A Case Study of the UK Biotechnology Sector

4.1 Chapter Abstract

By use of the Biotechnology industry as a case study the definitions of co-movement and contagion are strictly rationalized in the context of market categorization. Building upon the ideas developed by Chan (2003) [50] and Barberis *et al* (2003a) [19] in order to investigate the impact of news events on changes in stock co-movement. In this chapter it is suggested that changes in the conditional correlation of assets need not always be thought of in terms contagion effects but as jump changes in the stochastic correlation structure.

4.2 Introduction and Literature

This chapter adds to the existing research on the incorporation of information into asset prices and co-movement between assets. The contribution of this paper is to build on this research by examining the impact of news events specifically relating to just one firm on the remaining firms within a sector. The aim is to increase our understanding of how investors incorporate information into prices by measuring the response of firms to news that does not specifically relate to them. Changes in co-movement may be explained in several ways, first of all by changes in volatility, second changes in the market value of assets, third by changes in the internal structure of the firms. In general finding out which of these processes is really at work when analyzing dynamic correlation structures is difficult. In the biotechnology sector in this sense is there is fairly widely available information as to the potential value of drugs and compounds currently being researched. Likewise there is also a substantial history of information pertaining to success and failure rates of potential drugs and compounds, so to a certain extent there is enough knowledge to assess the fundamental value of firms in this sector, to be able to fully specify and parameterize the processes which underlie the dependency structure of asset price processes.

The literature on the inclusion of news announcements into stock price dynamics is broad and generally focuses on individual stocks as opposed to the general effect on a sector of news announcements to a single element. Work by Bernard and Thomas

(1989) shows that the information contained in earnings announcements takes some time to become fully incorporated into prices. Similar evidence of drift following announcements relating to stock splits and changes in analyst recommendations has been documented by Ikenberry and Ramnath (2002) /citeIkenberry133 and Michaely and Womack (1999) [198] respectively. Pritamani and Singal (2001) [221] suggests there is evidence of a similar lagged adjustment following large price changes, defined as an abnormal return more than three standard deviations from the mean. However, Chan (2003) [50] distinguishes between large price changes associated with news events and those that appear to be unrelated to news, and finds a lagged adjustment, or drift, only following a news event. Large price changes that are not associated with a news event result in reversal. Moreover, the drift is stronger for bad news events, implying that the adjustment to bad news is slower than for good news. Conrad, Cornell and Landsman (2002) [55] document a further asymmetry. Examining the stock price response to earnings announcements, they show that the response to bad news is stronger when the market is rising. On the other hand, the response to good news decreases when stock prices are high. Underpinning this research is the increasingly widespread view that investors under-react to new information. This under-reaction may result from information only gradually becoming incorporated into prices, and is perpetuated by limited arbitrage due to transaction costs, particularly where there is bad news. At the same time, the reversal documented by Chan (2003) [50] following

non-news related price movements suggests a tendency for investors to overreact to such price movements. In a similar approach, Daniel and Titman (2005) [65] distinguishes between tangible and intangible information, where intangible information is 'that part of the stock's past return that cannot be linked directly to accounting numbers, but which presumably reflects changes in expectations about future cash flows.' They argue that investors overreact only to intangible information. Chan (2003) [50] suggests his results are consistent with the model in Hong and Stein (1999). This model is based on two types of trader, one that focuses on news and one that focuses on price movements. Alternatively, the overreaction to intangible information in Daniel and Titman (2005) [65] may be consistent with the overconfidence and self-attribution bias discussed in Daniel, Hirshleifer and Subrahmanyam (1998) [64]. Several studies have also examined the transfer of information between firms within an industry. These studies find that an announcement relating to one firm can have a significant impact on the prices of shares within the industry. Most notably, Lang and Stulz (1992) [175] find that bankruptcy announcements have a negative impact on rival firms within the industry. This effect can however be inverted where the industry is highly concentrated. Szewczyk (1992) [245] shows that rival firms respond negatively to an SEO announcement, whilst Tawatnuntachai and D'Mello (2002) [248] finds that rival firms respond positively to stock split announcements. This research is also related to the empirical studies of market co-movement and

contagion. These studies examine the extent to which markets move together, and in particular how this co-movement is affected by a shock in one market. Whereas the focus of this research is on the transmission of a news shock from one stock to another, studies of market contagion focus on the transmission of shocks between different stock markets. In this context, Forbes and Rigobon (2002) [97] define contagion as 'a significant increase in cross-market linkages after a shock to one country'. Thus contagion is more than simply co-movement in stock returns. A shock in one market that either induces, or occurs at the same time as, a shock in another market, is evidence of co-movement, not of contagion. Contagion requires that the extent of the co-movement between markets be impacted significantly by a shock. An early example of the research supporting the presence of cross-country contagion is that of King and Wadhwani (1990) [169]. They find that the cross-market correlations between the US, UK and Japan rose significantly after the crash in 1987. Lee and Kim (1993) obtain similar results for a larger subset of countries. Most recently, Forbes and Rigobon (2002) [97] concludes that there was no contagion during several major international shocks, including the 1987 crash . Studies focusing on co-movement (and changes in co-movement) among stock returns are also closely related to our study. Barberis et al (2003b) develops a model in which analysts and/or investors assign stocks to categories - what they term style investing. Additionally, they assume that the allocation of funds among different styles is based upon their relative

past performance. As a result, investors switching between styles create patterns in demand that generate return co-movement within a style that is independent of cash-flow co-movement. Consistent with this, Froot and Dabora (1999) [63] shows that twin stocks co-move most closely with the market in which they are traded. A further prediction of Barberis et al (2003b) [17] is that the reclassification of a stock into a new style increases the stock's co-movement with that style. Research in this area has focused on the change in correlation, or co-movement, in returns between a stock and the market in which that particular stock is classified. There has been consistent evidence that a stock's co-movement with an index increases after inclusion in the index, and reduces after deletion from the index (see, for example, Vijh (1994) [258], Greenwood and Sosner (2002) [121] and Barberis et al (2003a) [17]. Greenwod and Sosner (2002) [121] conclude that it is trading that induces the observed co-movement in returns.

4.2.1 Current literature on the Biotechnology Sector

The choice of the biotechnology sector to analyze stock co-movement is driven by the extent of the interdependency between the firms in the sector and a reasonable transparent dependency structure. Most obviously, the firms often engage in the joint licensing of candidate drugs and technologies, and as a result news impacting on one firm's cash flows will have a clear and definite impact on other firms' cash flows. In addition, companies can become inter-linked through product development. Complex

drugs and treatments may use combinations of candidate drugs, often with patents held by a variety of companies, so that news affecting certain companies may have both direct and indirect effects on others within the sector. Perhaps less obviously, a news event that is firm-specific may cause a re-assessment of the prospects for other firms in the sector. For example, an announcement regarding the development of a particular compound by one firm may have important implications for the compounds developed (or in the process of being developed) by the other firms in the sector. Companies often normally hold similar portfolios of licensed candidate drugs/compounds. More generally, a firm-specific news event may have sector-wide implications with regard to the way that the sector as a whole is perceived by investors. Lerner and Merges (1998) [179] identifies the biotechnology sector as one that is particularly prone to information asymmetries, and that as a result news affecting one firm can have a dramatic effect on the other firms in the sector. Lerner, Shane and Tsai (2003) [180] finds that 'unexpected events occurring at a single biotechnology firm e.g., the rejection of a promising drug candidate - had dramatic effects on all firms' abilities to raise equity.' In addition to the interdependencies that are likely to exist between firms in the sector, a further complication that investors face is simply the difficulty in assessing the potential impact of a news event. This is aggravated by the fact that the development of a new drug, from concept to delivery, can take a decade or more. Patent law in the US runs to around 80 pages for headline contents, whilst

drug discovery alone has several hundred pages of law devoted to its complexities. An additional uncertainty has also been introduced regarding firms' ability to patent some types of intellectual assets, particularly those relating to gene sequencing and DNA profiles. Moreover, complications can arise as a result of inconsistencies in legislation between the EU and the US. Finally, in order to maximize profit from their drugs, firms have a strong incentive to attempt to limit the availability of information relating to them, as patent disputes tend to be very costly, therefore protection is not always assumed simply by patent ownership.

4.3 Models of Dynamic Co-dependency With Jump Diffusions

Consider an k length stochastic vector price process $\mathbf{s}_t$, driven by a set of common factors $\mathbf{x}_t$ and k length vector noise process $\mathbf{w}_t$, $\mathbf{J}_t$,

$$d\mathbf{s}_t = \mu\left(\phi\left(\mathbf{x}_t\right)\right) + d\boldsymbol{\Sigma}_t^{\frac{1}{2}} d\mathbf{w}_t + d\mathbf{J}_t \tag{4.3.1}$$

here the final term $\mathbf{J}$ is a jump process, where,

$$\mathbf{J}_t = \begin{bmatrix} \sum_{j^1=1}^{N^1(t)} \left(Y_j^1 - 1\right) \\ \sum_{j^2=1}^{N^2(t)} \left(Y_j^2 - 1\right) \\ \vdots \\ \sum_{j^k=1}^{N^k(t)} \left(Y_j^k - 1\right) \end{bmatrix} \tag{4.3.2}$$

This type of model is commonly used to describe asset price dynamics. Now consider a second state vector of jump processes $\mathbf{x}_t$, where $\mathbf{x}_t \in \mathbb{R}^n$, where $n = n_1 + n_2 + ... + n_k$, where n_i is the number of projects undertaken by company i. Each price process is a transformed Numeriaré, (change if measure) of a value process that is the summation of some set of risky projects, in the biotechnology case the development of various drug projects. Again this state space is defined in terms of some sort of jump diffusion model as set of news events impact on the value.

$$d\mathbf{x}_t = \mu(\mathbf{x}_t)\, dt + d\boldsymbol{\Omega}_t^{\frac{1}{2}} d\mathbf{w}_t^x + d\mathbf{J}_t \tag{4.3.3}$$

The integral price process is now,

$$\mathbf{s}_t = \mathbf{s}_0 + \boldsymbol{\Psi}^{\mathrm{T}} \int_0^t \mathbf{x}_t d\mathbf{x}_t + \boldsymbol{\Sigma}^{\frac{1}{2}} \left(\boldsymbol{\Psi}^{\mathrm{T}} \int_0^t \mathbf{x}_t^s d\mathbf{w}_t^s \right) \tag{4.3.4}$$

Where the matrix, $\boldsymbol{\Psi}$, is partitioned with unit vectors $\underset{n_i \times 1}{\mathbf{e}_i}$ to determine which company owns to which project, $\mathbf{w}_t^x$ and $\mathbf{w}_t^s$ are respectively an n-length and k-length vector Wiener processes.

$$\underset{k \times n}{\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_2^{\mathrm{T}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{e}_k^{\mathrm{T}} \end{bmatrix} \tag{4.3.5}$$

The non-negative matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ determine the quadratic covariation between respectively the diffusion of the projects conducted by the companies in the sector and the actual companies themselves. The unfiltered quadratic covariation of the

company processes is therefore,

$$E\left(\mathbf{s}_t\mathbf{s}_t^{\mathrm{T}}\right) = \mathbf{\Psi}\left(E\left(\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}\right)\right)\mathbf{\Psi}^{\mathrm{T}} \tag{4.3.6}$$

Given a sequence of jumps, caused by news events to individual projects, the expected quadratic covariation will jump according to the change in the state process. Given this structure a series of inferences may be made about the structure of return processes in the biotechnology sector,

- There will be a deterministic shift (jump) in the covariation between assets within the biotechnology sector following a news event.
- This jump in covariation will be in some way proportional to the size of the jump caused by the news event to the original company.
- The return processes of companies not undergoing the news event will have jumps, even if the news event is not directly attributable to them, i.e. transmitted through the state space $\mathbf{x}_t$.

Given that it is difficult to understand the exact properties of the balance sheet items, we assume that the state space is unobserved and as such must be inferred, using probabilistic models.

4.4 An Empirical Study of Co-movement in the UK Biotechnology Sector

We follow Barberis *et al* (2003) [17] and consider the pair wise co-movement between firms in the same sector, based on a two step regression system. Specifically we treat the abnormal returns to a firm undergoing an event as the explanatory variable in simple linear regression. For each asset over the event period the excess returns are computed from a Sharpe-Lintner CAPM model, as follows,

$$\mathbf{r}_{i,\tau} = \mathbf{r}_{m,\tau}\beta_{i,m} + \mathbf{u}_{i,\tau} \tag{4.4.1}$$

$$\mathbf{r}_{j,\tau} = \mathbf{r}_{m,\tau}\beta_{j,m} + \mathbf{u}_{j,\tau} \tag{4.4.2}$$

Where $\mathbf{r}_{i,\tau}$ is the vector of returns over the sample window for the company undergoing an identifiable news event, the integer τ is the length of sample window before the event time $t = 0$, $\mathbf{r}_{j,\tau}$ is another company in the same sector and $\mathbf{r}_{m,\tau}$ is the market return over the equivalent period. The symbol $_\tau \equiv [-\tau, 0, \tau]$, in respect of a column vector over event time, is used to signify that the vector is measured in event time. Now consider a vector $\mathbf{s}_{i,\tau}$ of cumulative abnormal returns where

$$\mathbf{s}_{i,\tau} = [s_{i,-\tau}, ..., s_{i,\tau}]^{\mathrm{T}} \tag{4.4.3}$$

$$s_{i,t} = \sum_{t=-\tau}^{\tau} u_{i,t} \tag{4.4.4}$$

Once the cumulative excess returns are computed the second stage regression is as follows,

$$\mathbf{s}_{j,\tau} = \mathbf{E}_{i,\tau}\mathbf{b}_{j,\tau}^{\mathrm{T}} + \mathbf{u}_{i,j} \tag{4.4.5}$$

Where,

$$\mathbf{E}_{i,\tau} = [\mathbf{e}, \mathbf{s}_{i,\tau}, \boldsymbol{\Delta}\mathbf{s}_{i,\tau}] \tag{4.4.6}$$

$$\underset{2\tau+1\times 2\tau+1}{\boldsymbol{\Delta}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \underset{\tau+1\times\tau+1}{\mathbf{I}} \end{bmatrix} \tag{4.4.7}$$

and the linear coefficients $\mathbf{b}_{i,\tau} = [\beta_{i,\tau,1}, \beta_{i,\tau,2}, \beta_{i,\tau,3}]^{\mathrm{T}}$, are estimated via OLS regression. The magnitude of the jump $J_{i,\tau}$ from the news event is therefore estimated by,

$$J_{i,\tau} = \tfrac{1}{\tau}\sum_{t=-\tau}^{-1} s_{i,t} - \tfrac{1}{\tau}\sum_{t=0}^{\tau} s_{i,t} \tag{4.4.8}$$

For each news event the vector of estimated regression coefficients $\hat{\mathbf{b}}_{j,t}$, is collected for all the other assets in the sector as is the magnitude of the jump. The third element of $\mathbf{b}$, $\beta_{i,\tau,3}$, is used to proxy for the degree of extra co-movement after the event time.

Selected Hypotheses

The following hypotheses are proposed from the time series regressions,

$$H := \begin{cases} H_0 :, & \beta_3 \to 0, \text{ No Change in Co-movement post event;} \\ H_{-1} :, & \beta_3 \leq 0, \text{ Decrease in Co-movement post event;} \\ H_1 :, & \beta_3 \geq 0, \text{ Increase in Co-movement post event.} \end{cases}$$

By collecting a large number of significant events, the relationship between changes in co-movement and news events, maybe measured for an industrial sector.

4.5 Sample Window and News Events

Co-movement and the change in co-movement is measured between 16 companies for the period 1988 - 2003, a daily window of 61 days is used, i.e. $\tau = 30$. This 15-year period yields 251 carefully selected drug discovery events from a potential sample of nearly 2000 events, the events selected specifically contained new information as to the success or failure candidates drugs and offered significant extra information to the market on the potential value of the candidate drug.

Daily closing returns and the risk free rate (the UK 6 month treasury rate, at the time of each event), were obtained from Thompson Financial DataStream^{TM}.

4.5.1 Event Selection Criteria

The firms within the sample made a total of around 2,000 announcements during the sample period. We apply a number of methods to filter the events and produce a sample of significant news events for this study. This selection is critically important to ensure the validity of the findings. A number of criteria were used to choose the news events. The news announcement must relate to products, and contain new information; the firm must hold intellectual rights to the candidate drug in question; subsequent announcements for a particular firm within the event window were excluded. The news announcements were then categorized in general terms of impact, positive or negative. This yielded a total of 251 usable news events. Our selection

criterion for an event was designed to produce a catalogue of events that were primarily significant to the valuation of the candidate drug project. We applied a series of tests to these events to evaluate their significance and filter them appropriately. These included expert evaluation of the content and econometric tests to the return series for the company stock price to analyze excess returns and the impact of the event on the stock price. The press releases were obtained either directly from the companies or from the data vendors used to distribute them. These include RNS [The London Stock Exchange news wire service), AFX [a wholly-owned subsidiary of Agence France-Presse (AFP)] and PR Newswire [United Business Media]. Generally the information flows from the company to the press release medium, often via third party media consultants who design the press releases. Whilst the date of conclusion of a clinical trial may be prior to release, the press release date is the date at which full access to the information is available and is therefore the date used. A test set of events based upon deviations from a standard multifactor linear model was compared to the event dates. All of the significant shocks for each company coincided within +/- 5 days of the announcements within our catalogue.

4.6 Cross Sectional Analysis and Regressions

The selected events were partitioned into two groups, those relating to positive news events and those relating to negative ones. To provide evidence for the underlying structure the following cross sectional regressions are proposed, for a set of m events,

first or the β_2 coefficients,

$$\mathbf{b}_{2,j,\tau} = \left[\beta^1_{2,j,\tau}, \beta^2_{2,j,\tau}, ..., \beta^m_{2,j,\tau}\right]^{\mathrm{T}} \tag{4.6.1}$$

$$\mathbf{j}_i = \left[J^1_i, J^2_i ..., J^m_i\right]^{\mathrm{T}} \tag{4.6.2}$$

$$\underset{m\times 1}{\mathbf{b}_{2,j,\tau}} = \underset{m\times 1}{\mathbf{j}_i}\, \phi_{\beta_2} + \underset{m\times 1}{\mathbf{v}} \tag{4.6.3}$$

and second for the β_3, coefficients,

$$\mathbf{b}_{3,j,\tau} = \left[\beta^1_{3,j,\tau}, \beta^2_{3,j,\tau}, ..., \beta^m_{3,j,\tau}\right]^{\mathrm{T}} \tag{4.6.4}$$

$$\mathbf{j}_i = \left[J^1_i, J^2_i ..., J^m_i\right]^{\mathrm{T}} \tag{4.6.5}$$

$$\underset{m\times 1}{\mathbf{b}_{3,j,\tau}} = \underset{m\times 1}{\mathbf{j}_i}\, \phi_{\beta_3} + \underset{m\times 1}{\mathbf{v}} \tag{4.6.6}$$

Where $\mathbf{v}$ is an m length disturbance vector. The coefficient ϕ is estimated using least squares regression and the standard diagnostic tests were carried out.

Cross Section Hypotheses

The cross sectional regressions attempt to infer the relationship between the magnitude of shocks and the magnitude of the linear dependence between the abnormal returns exhibited by companies stock returns, the following cross sectional hypotheses are proposed,

$$H := \begin{cases} H_0 :, & \phi_{\beta_2} \to 0, \text{ no linear dependence: } \beta_2 \text{ and shock size;} \\ H_1 :, & \phi_{\beta_2} \geq 0, \text{ evidence for linear dependence: } \beta_2 \text{ and shock size.} \end{cases} \tag{4.6.7}$$

and for the β_3 coefficients,

$$H := \begin{cases} H_0 :, & \phi_{\beta_3} \to 0, \text{ no linear dependence } \beta_3 \text{ and shock size;} \\ H_1 :, & \phi_{\beta_3} \geq 0, \text{evidence of linear dependence: } \beta_3 \text{ and shock size.} \end{cases} \tag{4.6.8}$$

Positive Results	Shock	% p.s. β_2	% p.s. β_2	$\beta_2 \bigcup \beta_3$
Mean	0.591084	43.9721641	32.34314133	8.198111412
Standard Deviation	0.76208	23.4712193	21.64117906	11.80533741
Sample Variance	0.145192	550.8981354	468.3406311	139.3659914
Negative Events				
Mean	-0.5566741	40.83353912	37.45934351	7.123101617
Standard Deviation	0.5662177	24.6998481	23.95261722	12.75920723
Sample Variance	0.0801506	610.0824961	573.7278716	162.7973692

Table 4.1: The events are split between positive and negative events, the percentage of positive significant, p.s. $\beta_2's$ illustrates the degree of integration in the sector, on average over the 251 sample windows, at lest 40% of the stocks had significant linear dependency, over the whole sample window. On average the number of p.s. $\beta_3's$ is around 30% for positive events and 23% for negative events from the sample. This resulted in a total of 1098 p.s. β_3 observations from 251 events, out of a total of 4096 estimated.

The table gives the descriptive statistics for the number of positive and significant (p.s.) β_3 coefficients as a result of a shock to one company in the sector. The sample size for the cross sectional regression is 1098 observations, from a possible 251 $\times$ 16, (4,096) observations.

4.6.1 Results for Shock Size Against the Magnitude Coefficients

Table 4.6.1, demonstrates the dependence between shock magnitude and β_3. The regression results suggest that the β_2 coefficient exhibits an insignificant linear relationship to the shock size and as such the null hypothesis from 4.6.7 is accepted. By contrast the results for the β_3 coefficients suggest that there is a strong and significant linear relationship between shock magnitude and coefficient magnitude and as such

Table 4.2: The magnitude of the β_2 and β_3 coefficients regressed against the shock magnitude. Not all of the companies were floated over the whole period, however no news event had less than 12 companies in the cross section

Coefficient	Regression Model 4.6.1	Regression Model 4.6.4
Positive Events	-0.118740191	0.021794969
s.e.	0.097245724	-0.186941943
t-stat	-1.221032518	-0.116586833
Bootstrapped critical boundary	3.043416006	3.386314249
Negative Events	0.469767274	0.605754163
s.e.	0.082018289	0.169269747
t-stat	5.727591754	3.578632174
Bootstrapped critical boundary	3.041747044	3.243782724

the null hypothesis from 4.6.8 is rejected.

4.7 Concluding Remarks

This chapter lends evidence to the conjecture that the dependency between asset returns is dynamic. The case study demonstrates the change in co-movement between assets are a series of observed events which updated the information set. The size of the dependency adjustment is shown to be strongly correlated to the size of the jump that characterizes the event, which is in line with the theoretical model predictions outlined in section 4.3. The biotechnology sector, as a case study, offers a sector with highly interdependent firms that are constantly undergoing significant events which affect their firm specific valuation and the valuation of the many other firms in the sector. However the contribution of these project shocks to changes in dependency is non-uniform, as different projects share common properties between different firms. A model of linkages between firms is proposed, which suggests that such linkages may

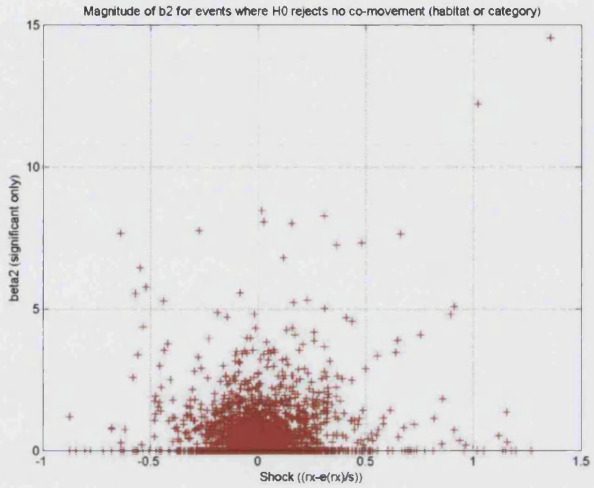

Figure 4.1: The plot points represent the magnitude of the β_2 coefficients against the magnitude of the shock, there is no visible trend and the results appear to be as a perfect half Euclidean ball about the origin, this lack of pattern is in keeping with the regression results

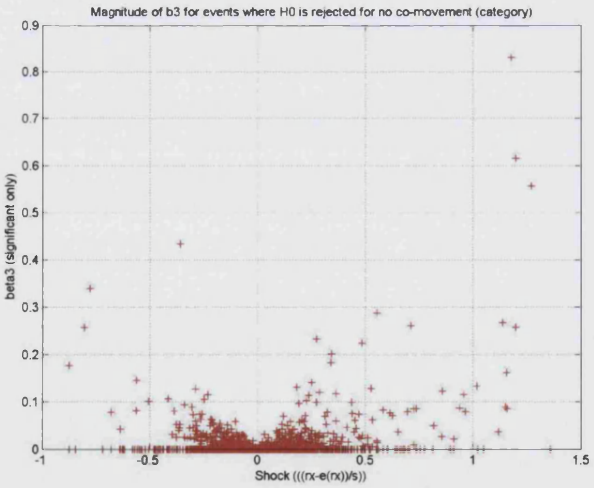

Figure 4.2: The plot points represent the magnitude of the β_3 coefficients against the magnitude of the shock, the trends represented in the regression results are immediately visible, as is a certain degree of asymmetry, which is only marginally significant.

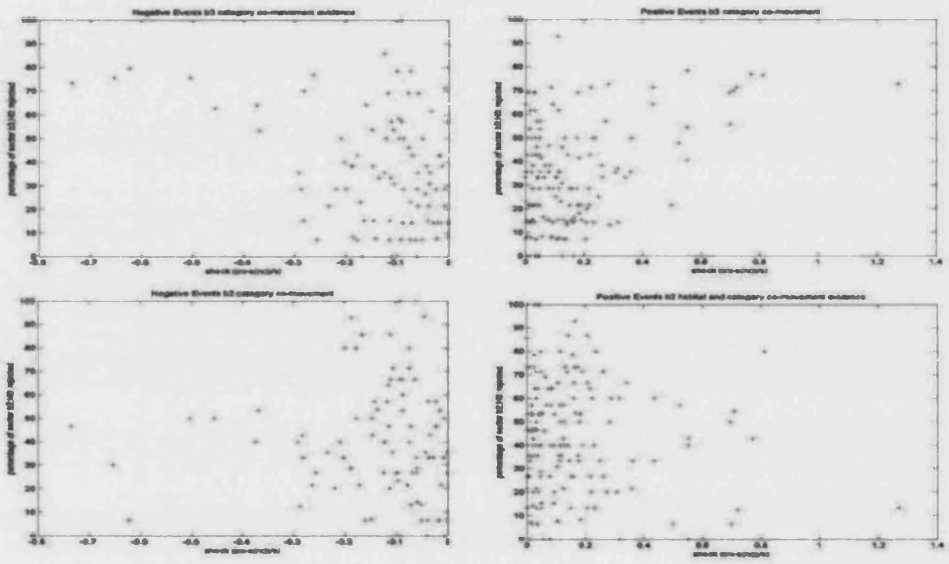

Figure 4.3: The plot points represent percentage of p.s. β_3 coefficients against the magnitude of the shock, the trends represented in the regression are not as immediate visible as those for the magnitude plots as the sample size is cut to 251.

strengthen or weaken depending on the tightness of the correlation of a particular project undergoing an event. The resultant change in the correlation structure is then related to the relative change size of the project in relation the remaining portfolio of drugs, therefore the dependency will jump in proportion to the size of the event.

4.8 Chapter Appendix

4.8.1 Companies and Events

1. Antisoma; Biopharmaceutical Products, Anticancer
2. Biocompaitbles; Biopharmaceutical Products, Polymer Science,
3. British Biotech; Biopharmaceutical Products, Anticancer, Antibiotics
4. CAT; Biopharmaceutical Products, Antibody Technologies
5. CANTAB; Biopharmaceutical Products, Anticancer
6. CellTech; Immunology
7. CENES; Biopharmaceutical Products, Anticancer
8. Amersham; Imaging, Analysis, Genomics,
9. Oxford Biomedica; Biosciences
10. Oxford Glycosciences; Custom drug discovery technologies
11. Powderject; Drug delivery systems

12. PPL; Genomics + Cloning

13. Proteome; Proteomics

14. Provalis; Biopharmaceutical Products, Anticancer

15. Smithkline Beecham; Pharmaceutical

16. Xenova; Diversified Biosciences

Table of Events

1 24/10/2002 Pemtumomab gastric phase II preliminary results,

2 22/08/2002 Antisoma completes patient recruitment for phase I trial of TheraFab,

3 17/06/2002 Phase I data for Antisomas DMXAA presented at International Conference on Vascular Targeting,

4 08/04/2002 Antisoma presents data showing Thioplatin active on cancer cells resistant to cisplatin

5 10/09/2001 New product, Theranase, in-licensed into Antisomas Targeted Apoptosis programme,

6 30/08/2001 Antisoma in-licenses DMXAA, a promising tumour blood vessel targeting agent in Phase I trials,

7 06/06/2001 Recruitment to Phase III study of Antisomas lead product pemtumomab extended after initial target achieved,

8 30/04/2001 Humanised HMFG1, under development by Antisoma as Therex, shown to be effective at harnessing components of the bodys immune system.

9 11/04/2001 Antisoma and BioInvent sign BC-1 monoclonal antibody manufacturing agreement,

10 01/03/2001 Antisoma appoints Medical Oncology Advisor,

11 29/01/2001 TheraFab, Antisomas latest product candidate starts clinical studies,

12 30/10/2000 Antisoma in-licenses Thioplatin, a novel, platinum-based, targeted anti-cancer therapy,

13 11/07/2000 Antisoma resumes enrolment into its SMART Phase III pivotal trial,

14 13/06/2000 Antisoma plc Company share option plan grants,

15 22/05/2000 Theragyn Phase III Study Enrolment Temporarily Delayed,

16 03/04/2000 A gentler way to kill cancer cells,

17 16/03/2000 Antisoma announces adoption of Phase I clinical product Therex,

18 14/02/2000 Antisoma plc reports Q2 results and progress in development pipeline,

19 16/12/1999 Listing on the London Stock Exchange,

20 21/08/2001 Commencement of clinical trial programme for the Batimastat BiodivYsio stent

21 24/05/2001 Biocompatibles Update on New Product Developments

22 03/10/2000 PMA Approval for BiodivYsio Coronary Stent

23 12/09/2000 Approvable Notification Received from FDA for BiodivYsioTM Stent

24 22/05/2000 Positive Results of US Coronary Stent Clinical Trial and Update on US Product Approval Process

25 28/02/2000 BiodivYsio Stent Update

26 26/11/1999 BiodivYsio Drug Delivery Stent receives European Marketing Approval

27 29/06/1999 Placing and Open Offer, Stent Distribution Agreement for Japan, Private Label Contact Lens initiative with Specsavers, CE Mark and Launch for Small Vessel Stent and Trading Update

28 27/07/1998 Biocompatibles Intl. - Re Clinical Results

29 01/10/2002 British Biotech and GeneSoft announce start of human trials for first-in-class antibiotic BB-83698

30 08/08/2002 British Biotech and ImmunoGen start second Phase I study of BB-10901

31 20/05/2002 Data from ongoing Phase I/II clinical study of BB-10901/huN901-DM1 presented at 2002 meeting of American Society of Clinical Oncology

32 26/03/2002 Novel thrombolytic drug cleared by FDA for phase II clinical testing

33 08/03/2002 Suspension of patient recruitment in the BRILLIANT II clinical trial for the Batimastat BiodivYsio stent

34 12/02/2002 British Biotech and MethylGene sign agreement on Phase II antisense drug for cancer

35 26/07/2001 British Biotech to start phase II trial of E21R in acute myeloid leukaemia

36 29/05/2001 British Biotech and ImmunoGen to begin phase I/II trial of BB-10901 in small cell lung cancer

37 02/05/2001 Technical note on long term follow-up of marimastat Study 145

38 06/03/2001 Bresagen Ltd commences patient recruitment for E21R phase II study

39 19/01/2001 British Biotech announces update on marimastat Studies 140 and 117 in patients with small cell lung cancer

40 26/09/2000 Results of marimastat study 186 in patients with advanced overian cancer

41 23/06/2000 Results of marimastat Study 131 in patients with glioblastoma

42 05/05/2000 British Biotech in collaboration with ImmunoGen to develop and commercialise huN901-DM1 for the treatment of small cell lung cancer

43 06/04/2000 British Biotech in collaborative agreement with CareScience, Inc. to develop the CareScript Oncology Extranet

44 23/03/2000 British Biotech plc and DevCo Pharmaceuticals Limited sign agreement on lexipafant

45 25/01/2000 Results of marimastat Study 193 in advanced pancreatic cancer

46 07/09/1999 British Biotech in collaborative agreement with Schering-Plough to develop matrix metalloproteinase inhibitors for cancer

47 26/08/1999 Results of marimastat Study 145 in gastric cancer

48 25/03/1999 British Biotech plc

49 15/02/1999 Results of marimastat Study 128 - pancreatic cancer monotherapy trial

50 14/11/1996 British Biotech and Tanabe in $74 million agreement to develop and market marimastat

51 30/09/2002 Cambridge Antibody Technology Partners with Chugai to Develop Novel Human Monoclonal Antibodies

52 08/05/2002 Cambridge Antibody Technology Announces Positive 12 Month Phase II Results for CAT-152

53 14/11/2001 Cambridge Antibody Technology Announces Six Month Follow-Up Results for CAT-152

54 01/11/2001 Human Genome Sciences Announces Phase I Trial of a Human Monoclonal Antibody Discovered in Collaboration with Cambridge Antibody Technology

55 25/09/2001 CAT Announces Granting of Regulatory Approval to Start UK Patient Trials of CAT-213

56 22/08/2001 Cambridge Antibody Technology Announces Phase II Clinical Trial Results Using CAT-152 in Combined Cataract and Glaucoma Surgery: Findings at Three Months Following Operation and Treatment

57 12/06/2001 Cambridge Antibody Technology Starts Phase I Clinical Trials of CAT-213

58 30/04/2001 CAT Announces Further Information on CAT-152

59 15/01/2001 Cambridge Antibody Technology Granted Further Key Patent in US

60 04/05/2000 One Year Results for CAT-152 in Glaucoma Surgery

61 04/02/2000 D2E7 Becomes the First Fully Human Monoclonal Antibody into Phase III Clinical Trials

62 17/11/1999 New Data from Clinical Trials with D2E7 for Rheumatoid Arthritis

63 06/09/1999 CAT Receives Millennium Product Status for Two Products - CAT Library and ProAb

64 09/06/1999 D2E7 - Further Clinical Data

65 24/05/1999 CAT Acquires Rights to Exclusive Gene Libraries

66 10/05/1999 Impressive Pre-Clinical Study Results with CATs Fully Human Monoclonal Antibodies

67 27/04/1999 Cambridge Antibody Technology holds a granted US patent with broad claims directed to human antibodies to human proteins isolated by phage display and to methods for producing human antibodies.

68 25/03/1999 Cambridge Antibody Technology Strengthens Intellectual Property Position

69 09/11/1998 Impressive Clinical Data with CAT/BASF Human Antibody D2E7 in Rheumatoid Arthritis

70 25/09/1998 Cambridge Antibody Technology Initiates Patent Infringement Action Against MorphoSys

71 16/07/1998 Acquisition of Aptein Inc. Closed & Approval of European Patent for Polysome Display Received

72 26/05/1998 CATs Monoclonal Antibody Approved to Enter Clinical Trials for Glaucoma Surgery

73 02/03/1998 Cambridge Antibody Technology Announces Phage Display Agreement with Dyax Corporation

74 05/11/1997 CATs Collaborative Partner Techniclone Receives FDA Permission to Begin Clinical Trials with TNT for Malignant Glioma

75 08/08/1997 Approval to Start Clinical Trials for Human Anti-TGF-beta2 Monoclonal Antibody

76 08/01/1997 Cambridge Antibody Technology & Eli Lilly Enter into a Collaborative Agreement Involving Antibody-Based Therapeutics

77 12/02/1996 Cambridge Antibody Technology Signs Several Million Dollar Agreement with PFIZER

78 26/01/1996 Cambridge Antibody Technology Limited Licenses Rights to use Anti-TGF-Beta Antibodies

79 18/12/1995 ObeSys and Cambridge Antibody Technology to Collaborate on Discovery and Development of Antibody based Anti-Obesity Product

80 01/02/1994 CAT Improves Binding Characteristics of Mercks HIV Neutralising Antibody

81 05/12/1991 Antibody Breakthrough: Human Antibodies Made in Bacteria without Immunisation

82 06/12/1990 CAT Scientists Develop New Method of Isolating Monoclonal Antibodies

83 18/10/2000 Cantab Pharmaceuticals plc - Genital Warts vaccine clinical trial results

84 15/06/2000 Cantab Pharmaceuticals plc - Cantabs Cervical Dysplasia Vaccine Enters Phase I Clinical Trials

85 15/02/2000 Cantab Pharmaceuticals plc - Anti cancer findings with OX-401 in the journal of immunology

86 03/02/2000 Cantab Pharmaceuticals plc - Received a notification of interference on US patents for its DISC virus vaccine technology.

87 04/11/1999 Cantab Pharmaceutical plc - Phase II trials with DISC HSV genital herpes vaccine begin

88 07/10/1999 Cantab Pharmaceuticals plc - Cantab licenses anti-OX40 antibodies

89 23/09/1999 Cantab Pharmaceuticals plc - Positive phase I clinical trial results for DISC HSV genital herpes vaccines.

90 11/01/1999 Cantab Pharmaceuticals - Phogen - Report on new improved gene therapy method

91 18/05/1998 Cantab Pharmaceuticals plc - Results for first human trials of DISC HSV

92 08/05/1998 Cantab pharmaceuticals plc - Phogen VP22 technology

93 24/04/2002 Celltech And Biogen Announce World-Wide Collaboration, to Develop And Commercialise Phase III Crohns Disease Product

94 28/02/2002 Celltech Outlines CDP 870 Phase II Results in Crohns Disease

95 12/07/2001 Celltech Outlines CDP 870 Phase II Results in Rheumatoid Arthritis

96 05/03/2001 Celltech and Pharmacia Reach Worldwide Agreement on Development and Marketing of CDP 870

97 22/01/2001 Celltech and Johnson & Johnson to Develop New Oral Treatments for Cancer

98 18/05/2000 American Home Products Gains FDA Approval for Mylotarg(tm); First Antibody-Targeted Chemotherapy Provides High Tech Cancer-Fighting Option

99 99 22/03/2000 New Product Pipeline

100 100 10/01/2000 Celltech Announces the Disposal of Rapigene to Qiagen

101 101 11/11/1999 Celltech Chirosciences CDP 571 achieves positive results in CROHNS DISEASE Studies

102 102 18/05/1999 Potential to Develop New Treatments to Reverse Osteoporosis

103 103 03/11/1998 Medeva launches new dry powder inhaler product Asmabec Clickhaler

104 104 27/06/1997 Chiroscience and Schering-Plough to Develop New Oral Treatments For Astma and Other Inflammatory Diseases

105 105 20/05/1997 Bayer Announce Disappointing Results With New Drug For Treatment of Septic Shock

106 106 01/02/1996 Merck/Celltech Collaboration On PDE Type IV Inhibitors

107 107 15/05/2002 CeNeS Announces Further Progress in its Neuropathic Pain Phase II Programme

108 108 06/09/2001 CeNeS launches novel pain drug and establishes specialised hospital sales force

109 109 26/01/2001 Positive Phase II clinical trial results for M6G a morphine alternative

110 08/01/2001 Cambridge NeuroScience Awarded a Phase II SBIR Grant

111 30/06/2000 Moraxen Approved for UK Market

112 18/05/2000 Positive results from its sleep disorders programme

113 10/04/2000 CeNeS Pharmaceuticals announces ion channel technology milestones and aqcuisition of contract molecular biology company

114 18/02/2000 CeNeS Pharmaceuticals announces drug delivery collaboration with Swiss multinational 27/3/2000, Institutional Placing Raising 2.1 million of New Funds

115 31/01/2000 Bioglan and CeNeS Pharmaceuticals - Joint development and licensing agreement in pain control

116 13/01/2000 CeNeS Pharmaceuticals enters its first drug delivery collaboration

118 10/10/2002 Amersham in Metastron promotion deal with UKs Link Pharmaceuticals AFX

119 26/11/2001 Amersham wins US FDA approval for expanded use of MyoviewO coronary diagnosis AFX

120 29/10/2001 Amersham to market Corixas Bexxa drug for non-hodgkins Lymphoma in Europe AFX

121 14/06/1999 Amersham Pharmacia Biotech signs 2 major deals AFX

122 10/12/1998 Amersham Pharmacia collaborate on study AFX

123 11/07/2002 Oxford BioMedicas Trovax Poised to Enter Phase II Trials

124 06/06/2002 Oxford BioMedica: Successful Preclinical Results for New Anaemia Product

125 20/02/2002 Interim Clinical Results - MetXia and TroVax

126 13/12/2001 Progress Made in Nerve Repair Programme - Issues of Shares to Kings College London

127 12/11/2001 Positive Preclinical Results in Gene Therapy for Parkinsons Disease

128 15/10/2001 Successful Interim Phase I/II Trovax Results

129 10/07/2001 GTAC approval for development of METXIA clinical program

130 26/06/2001 Oxford BioMedica Obtains Fundamental Patent for Lentiviral Gene Therapy Technology

131 27/03/2001 Oxford BioMedica Obtains Gene Therapy Patent in the USA

132 05/02/2001 Oxford BioMedica: Acquisition of Novel Technology for Nerve Regeneration and Issue of Shares

133 02/01/2001 Oxford BioMedica Commences Clinical Trial of TroVax

134 06/11/2000 Oxford BioMedica Announces First Clinical Trial Results for its Cancer Therapeutic, MetXia.

135 16/10/2000 Oxford BioMedica Obtains Ethical Approval for TroVax to Enter Trials for Colorectal Cancer.

136 15/08/2000 Oxford BioMedica Presents Discovery of Novel Genes Relevant to Cancer, Heart Disease and Inflammatory Disease at Drug Discovery Conference

137 26/06/2000 Oxford BioMedica Announces Significant Advances in Gene Transfer to the Brain

138 06/06/2000 Oxford BioMedica Announces Collaboration Extension with Aventis Pharmaceuticals Inc. on Gene Therapy for Cardiovascular Disease.

139 18/05/2000 Oxford BioMedica: Acquisition of Therapeutic Genes and Issue of Shares

140 02/05/2000 Oxford BioMedica Acquires Novel Therapeutic Genes for Treatment of Cancer and Other Diseases

141 16/03/2000 Oxford BioMedica Comment on Human Genome/Patenting Issue

142 21/02/2000 Oxford BioMedica Signs Deal on Novel Tumour Vaccine

143 02/02/2000 Oxford BioMedica Annouces Successful Developments in MacroGen Cancer Therapy

144 07/01/2000 Oxford BioMedica Announces New Developments in Cancer Therapy

145 16/11/1999 Oxford BioMedica Develops Vector System for Gene Transfer in Therapy for Neural Diseases

146 06/07/1999 Oxford BioMedica Links Up With Modex Therapeutics To Develop Diabetes Therapy

147 10/06/1999 Oxford BioMedica Releases pre-Clinical Data on MetXia-P450, the Companys First Product to go into Clinical Trial

148 29/01/1999 Oxford BioMedica Announces Positive Preclinical Tumour Vaccine Data

149 16/12/1998 Gene Therapy Advisory Committee Gives Go Ahead for Oxford BioMedica Breast Cancer Trial

150 16/10/1998 Additional Commercial Opportunities For The Hypoxia Response Element Revealed At Gene Therapy Meeting

151 27/07/1998 UKs Leading Gene Therapy Company Awarded DTI Grant To Support Clinical Trial In Breast Cancer

152 20/05/1998 Lentivirus VectorsNew Gene Delivery Technology for Gene Therapy

153 26/01/1998 OXFORD BIOMEDICA ANNOUNCES INTERACTION WITH RHONE-POULENC RORER Agreement is signed with leading pharmaceutical company

154 09/01/1998 MAKING HIV-BASED GENE THERAPY SAFER Oxford BioMedica Develops a New Gene Delivery System to Extend the Applications for Gene Therapy

155 17/11/1997 OXFORD BIOMEDICA ANNOUNCES DETAILS OF ITS FIRST CLINICAL TRIAL IN CANCER Breast Cancer to be Targeted with Novel Gene Therapy.

156 20/10/1997 OXFORD BIOMEDICA ACQUIRES NOVEL ANTI-CANCER TECHNOLOGY Licensing Agreement is Signed with Leading US Researchers

157 12/09/1997 OXFORD BIOMEDICA RECEIVES GOVERNMENT GRANT FOR REVOLUTIONARY NEW TECHNOLOGY IN GENE THERAPY. New Technology Aims to Deliver Therapeutic Genes with the Highest Efficiency to Date.

158 30/04/1997 NOVEL MECHANISM IDENTIFIED FOR ANTI-CANCER GENE THERAPY Scientists Demonstrate the Potential for Delivery of Gene-Based Therapeutics Utilising Tumour-Specific Hypoxia.

159 05/11/2002 OGS announces the start of a Phase I Clinical Study with OGT 923

160 19/09/2002 OGS and FDA to Evaluate Protein Markers Predictive of Drug Toxicity

161 26/07/2002 OGS Announces Positive CPMP Opinion for Zavesca*

162 24/04/2002 OGS temporarily halts Vevesca (OGT 918) treatment in Israel as a precaution pending investigation of an unexplained adverse event

163 21/08/2001 OGS announces completion of New Drug Application submission for Vevesca (OGT 918) with the US Food and Drug Administration. 24 month monotherapy data to be presented at ESGLD Conference

164 12/02/2001 OGS Reports Progress with Vevesca Clinical Trials in type 1 Gaucher disease. Plans to file NDA/MAA on schedule this year

165 25/10/2000 OGT 918 (Vevesca) Designated as Orphan Medicinal Product by EU Commission

166 05/10/2000 Oxford GlycoSciences discovers novel member of key cancer enzyme family

167 08/09/2000 Positive Results on Extended Use Vevesca (OGT 918) Reported at the European Working Group on Gaucher Disease Meeting in Jerusalem

168 19/06/2000 Oxford GlycoSciences Receives Fast Track Designation for OGT 918 in Gaucher Disease from U.S. FDA

169 28/04/2000 First Efficacy Data on OGT 918 in Gaucher Disease Published in The Lancet

170 04/04/2000 OGS Files Patent Application with Pfizer on Alzheimers Disease Markers

171 22/03/2000 OGS proteomics operations passes major protein patent filing milestone

172 09/03/2000 OGS and Bayer partner in proteomics based respiratory disease research

173 29/06/1999 Incyte and Oxford GlycoSciences Announce Launch of Toxicology and Pharmacology Proteomics Database

174 26/04/1999 Incyte and Oxford GlycoSciences Announce Launch of Powerful, Integrated Proteomics Database Products

175 08/01/1999 OGS Completes Patient Accrual in Initial Trial of OGT 918 in Gaucher Disease Phase I/II in Fabry Disease to begin this month

176 10/06/1998 Oxford GlycoSciences Announces Orphan Drug Designation for OGT 918

177 13/01/1998 Incyte and Oxford GlycoSciences to Link Genomics and Proteomics

178 10/09/2002 PowderJect Initiates Clinical Trial of Powder Injection Flu Vaccine

179 09/08/2002 PowderJect Voluntarily Recalls BCG Vaccine

180 24/10/2001 PowderJect Announces Successful Opposition of Vical Incs Core DNA Vaccines Patent

181 08/05/2001 PowderJect Awarded Key DNA and Conventional Vaccine Patents

182 24/04/2001 PowderJect DNA Vaccine First To Achieve Protective Immunity In Non-Responders To Commercial Vaccine

183 20/10/2000 PowderJect Clarifies Position on Oral Polio Vaccine

184 12/06/2000 Serono and PowderJect Announce Achievement of Development Milestone

185 02/05/2000 PowderJect Announces Promising Progress Towards Development of HIV DNA Vaccine

186 21/03/2000 Glaxo Wellcome purchases a further DNA vaccine licence from PowderJect

187 29/11/1999 Positive Clinical Results Confirm Cellular Immunity With PowderJect DNA Vaccine

188 01/09/1999 PowderJect Commences First European Clinical Trial In Healthy Volunteers With a DNA Vaccine

189 30/03/1999 PowderJect Announces Positive Clinical Data Demonstrating Cellular Immunity With a PowderJect DNA Vaccine

190 11/02/1999 PowderJect and Ares-Serono Sign $100 Million Agreement Covering Multiple Proteins In PowderJect System

191 07/12/1998 PowderJects Hepatitis B DNA Vaccine First to Successfully Elicit Protective Immune Response In Humans

192 29/09/1998 PowderJect and Chiroscience Announce Clinical Results

193 16/09/1998 PowderJect Scientific Collaborator Reports Cancer Vaccine Research Results

194 20/04/1998 PowderJect Announces Advances With Needlefree Male Impotence Program

195 04/03/1998 PowderJect & Glaxo Wellcome Sign DNA Vaccine Agreement

196 09/09/1997 PowderJect Vaccines Announces Preliminary Results of Phase I Clinical Trial With Novel Hepatitis B Vaccine

197 18/03/2002 Product Update AAT and Fibrin 1 more

198 05/12/2001 PPL Therapeutics plc (PPL) Announces Status of Phase III AAT Trial more

199 06/09/2001 Positive phase II clinical trial result on tgBSSL in Cystic Fibrosis patients more

200 10/11/2000 PPL Therapeutics technology platform is to be extended into diabetes treatment. Creation of a new subsidiary company with potential to be separately listed. Board changes more

201 20/09/2000 PPL Therapeutics Plc Announces Positive Results from AAT Long Term Safety Study in Cystic Fibrosis Patients more

202 13/06/2000 PPL Therapeutics successfully expresses Novel Antimicrobial Peptide Transgenically more

203 04/04/2000 Positive Clinical Trial Result on BSSL more

204 09/02/2000 PPL Therapeutics Plc PPL announces results from two additional Phase II trials more

205 29/11/1999 Successful Proof of Principle Trial of Aerosolised AAT in the Treatment of Congenital AAT Deficiency. more

206 21/07/1999 Announcement of new groundbreaking technique - birth of Transgenic sheep with Targeted gene more

207 10/06/1999 FDA Orphan Drug Designation for AAT in the treatment of Congenital Deficiency more

208 22/02/1999 PPL announces positive results from its Phase II clinical trial of AAT in cystic fibrosis patients more

209 15/01/2002 Proteome US affiliate successfully corrects gene defect for cystic fibrosis AFX

210 11/04/2001 Proteome Sciences, Israels Mindset receive 450,000 stg Alzheimers grant AFX

211 05/07/2000 Proteome Sciences says well placed to exploit human genome opportunities AFX

212 01/07/1999 Proteome Sciences finds new targets for diabetes, obesity AFX

213 01/03/1999 Gene Therapy advancement

214 18/01/1999 Proteome Sciences develops cystic fibrosis therapy AFX

215 13/12/2001 Granted US Patent

216 12/11/2001 Provalis GlycosalO Test Granted CLIA Waiver Status in the US

217 26/03/2001 Provalis launches Pennsaid(R) Topical Solution in UK

218 23/08/2000 Regulatory Approval

219 17/07/2000 Provalis Concludes Agreement With Bio-Rad Laboratories Inc,

For Worldwide Distribution Of Glycosal(Tm),Files 510K Application For Glycosal(Tm) In The Usa, And Makes Shipments Of First Orders

220 15/03/2000 Launching of Glycosal, etc.

221 25/10/1999 Response to Press Comment, on sale of therapeutics division

222 22/07/1999 Collaboration Agreement Terminated

223 29/06/1999 Cortecs is today holding its scheduled research anddevelopment update, announcing progress in its two lead therapeutic products,MacrulinTM and PseudostatTM, its diagnostic programmes, and outlining the Companys business and scientific strategy.

224 25/05/1999 Phase III Trials

225 24/02/1999 Macritonin TM Trial

226 01/12/1998 Product News, board Changes

227 12/12/2000 (US/UK) SB Terminates Clinical Trial of Lotrafiban

228 14/11/2000 (US) SB Completes Enrollment and Initial Treatment Phase of PRESTO Trial

229 26/10/2000 European Commission Licenses New Hexavalent Paediatric Vaccine From SB For Use In All 15 EU Member States

230 05/10/2000 (US) Coulter Pharmaceutical and SB Announce Priority Review Status for Bexxar

231 18/09/2000 (US) First Study to Confirm That Antibiotic Choice Significantly Impacts Long-Term Patient Outcomes for Acute Infections (ICAAC/Factive)

232 29/08/2000 (US) Groundbreaking Data From Major International Trial Presented Today at European Society of Cardiology Annual Congress

233 18/08/2000 Nice Delivers Positive Guidance For Smithkline Beechams Avandia

234 26/07/2000 (US/UK) Study of fat-reducing protein opens new path toward obesity treatment

235 03/07/2000 European CPMP Adopts Positive Opinion For New Pentavalent And Hexavalent Paediatric Vaccines From SB

236 21/06/2000 (US) Avandia Demonstrates long-term blood sugar control (ENDO)

237 10/06/2000 (US) New evidence suggests that Avandia improves beta cell function in the pancreas (ADA)

238 20/05/2000 (US) New Phase III Study continues positive results seen with Hycamtin in first-relapse ovarian cancer (ASCO)

239 04/04/2000 (US) Study Shows drug combination may slow progression of type 2 Diabetes (JAMA)

240 21/03/2000 (US/UK) Major heart failure trial stopped due to significant survival Benefit

241 12/10/1999 New Study Shows Ariflo Improved Pulmonary Function In COPD

242 08/10/1998 Seroxat Approved In Social Anxiety Disorder/Social Phobia

243 19/05/1998 (UK) New Mmr Vaccine More Comfortable For Infants

244 24/02/1998 SB and SkyePharma Disclose Technology Agreement for Paxil

245 03/10/2002 Xenovas Tariquidar Granted FDA Fast Track Status

246 28/06/2002 Phase III Trials Begin For Tariquidar

247 02/04/2002 Patient Dosing Begins in Phase Ha Dose Escalation Trial for Anti-Cocaine Addiction Vaccine TA-CD

248 05/02/2002 Anti-cancer compound XR11576 enters clinical trials

249 26/10/2001 Results Phase IIa HPV, Phase I CIN and Start Phase II Prime Boost Trials

250 10/10/2001 Results of TA-HSV Phase II Trial for the Treatment of Genital Herpes

251 10/09/2001 Anti-Nicotine Addiction Vaccine TA-NIC Enters Phase I Trials

252 12/05/2001 ASCO Phase II Clinical Results XR9576

253 18/10/2000 Positive Results for XR9576/doxorubicin

254 01/06/2000 XR5000 Phase II Trial Data, Colorectal Study

255 11/05/2000 Q1 and Positive Interim PK data XR9576/doxorubicin

256 07/03/2000 Prelims and XR9576 to Phase III

257 24/11/1999 IND Approval Received for XR9576 US Trials to Begin Immediately

258 11/11/1999 Third Quarter Announcement Interim Phase II PK Data XR9576, Plus selection next generation cytotoxic

259 12/07/1999 Sale of MetaXen to Exelixis

260 18/05/1999 Phase I Results XR9576

261 14/12/1998 European and US Phase II Trials Planned for P-gp Inhibitor XR9576

262 12/11/1998 Third Quarter Results Announcement, Phase II Efficacy Trials Planned for XR5000

263 12/05/1998 Phase I Clinical Trials begin for XR9576

264 18/02/1998 Lilly teams with Xenova Group to develop novel antithrombotic drugs for chronic use (UK + US versions)

265 15/01/1998 Xenova Discovery and Institute of Grassland and Environmental Research Form Innovative Plant Chemistry Alliance

Chapter 5

A new approach to Conditional Multivariate ARCH problems: The Bounded Dynamic Covariance Model

5.1 Chapter Abstract

One of the major goals of modern finance is to accurately model, test and forecast the conditional covariance of very large multivariate data-sets. This is of significant use in modelling forward rate correlations, FX returns, the dynamics of equity return correlations and the evolution of the cross section of macro-economic variables. Traditional methodologies have approached this problem by assuming a simple autoregressive framework, however for very large systems this requires the estimation of a large number of parameters and results in a very flat objective function. Our model strips down the modelling of conditional covariance into a simple two stage

framework. First by filtering the data matrix and attempting to isolate different covariance states and second applying a regime switching model, which then captures the transitions between these states. We demonstrate that this simple but robust approach is applicable to examining the evolution of the covariance of a large cross by applying the model and associated tests to a data matrix created from the filtered returns of the S&P 500.

5.2 Introduction and Literature

General models of multivariate conditional covariance utilize a vec/vech type representation to construct a matrix process that describes the time evolution of the covariance matrix. This chapter introduces a new generalized type of model that describes the time evolution of the conditional covariance matrix as a set of deviations from a set of *a priori* computed state matrices. The model adds to the available tools utilized in analyzing these types of problems by suggesting a model that is analytically tractable, quick to implement and easy to interpret. A great deal of literature has focused on the statistical properties of individual univariate series, however far less focus has been applied to the multivariate domain, an anomaly which is commented on extensively in Chapter 3 of Rachev and Mitnik (2003) [222], regime switching models in variance have been proposed previously see Ding, Granger and Engle (1993) [73], however switching models in covariance have not been proposed in full, however this chapter presents evidence that this type of model maybe the most appropriate in

modelling the probabilistic properties of equity returns. A modern treatment of correlation and dependency is undertaken in Embrechts, McNeil and Straumann (2005) [82] who look at non-linear structures in terms of joint distributions or copulas. Research into copulas has been extensive in the recent literature Patton (2002) [215] constructs a simple two asset scenario and models it utilizing various copulas and a threshold ARCH structure. The findings suggests that the correlation structure is indeed non-linear, however much more work needs to be undertaken in this area, in particular the issue or copula selection needs to be addressed, in a bivariate case simple visual inspection of historical returns will yield the dependency structure, however in larger scale multivariate systems there is no standard methodology to fully evaluate which of the multitude of available generator functions should be used. An additional note should be made in respect of the Archimedean copulas, a specific subset of joint distributions which exhibit many of the most promising properties. These functions are notoriously difficult to specify with regards to multivariate systems with more than two variables, see Williams and Ioannidis (2005b) [262]. Multivariate GARCH models, as previously stated, have not enjoyed the same success of their univariate analogues, in most part due to the prohibitive numbers of parameters required for estimation. This has two major effects first, in the practical sense is the time consuming nature of computation, second in a more abstract form is the difficulty in applying an intuitive rationale to the parameters. The majority literature concentrating on

numerical implementation and asymptotic properties Alexander (2001) [5], Engle (2000) [91], Engle and Sheppard (2002) [92] and on analysis and performance of the implied dynamic covariance matrix and ease of implementation, for example in Pelagatti and Rondena (2004) [216] and Brooks, Burke and Persand (2003) [44] respectively. The sheer technicality of the task involved in correctly specifying and computing the model has now detracted from the real value of the models, which is to better understand the market dynamics. As Green (2000) points out in chapter 18, page 808, with regards to the non-linear architecture of univariate GARCH models, "[at this point the programming] appears fairly complicated,,on the other hand its taken a fair amount of programming to get this far". This is just for a univariate model, therefore in a multivariate context Bollerslev's (1990) most simplistic multivariate GARCH model the constant conditional correlation, CCC, model has the most attractive attributes, in respect of intuition, given that we can apply the following logical hypothesis 'correlation does not change, only variance does, therefore, the off diagonal elements in the covariance matrix are solely affected by changes in the diagonal'. Bollerslev, Engle and Woodridge (1988) [40] had previously specified the vech type model which required a large number of parameters even under basic specification, but this had the advantage of including off diagonal dynamics in the covariance matrix, which were not solely dependent on the diagonal. The dynamic conditional correlation DCC algorithm of Engle and Sheppard (2002) [89]

offers some of the simplicity of the CCC model however offering time varying correlations. Several interesting observations have come about in the last 5 years made about the nature of the dynamics of conditional covariance matrices, Silberberg and Pafka (2001) [238] prove that elemental multiplication of ordered sets of positive semi definite matrices definitely yields positive semi-definite matrices. Jerez, Casals and Sotoca (2001) [153] suggest that the likelihood function specified in MV-GARCH models is ill-conditioned when approaching the optimal parameter structure. They suggest a non-linear transform to the residuals in order to improve the speed of convergence in the optimization process, as an interesting side note they suggest that the type of transformation improves the capturing of very short run correlation dynamics and aids in the system identification. A basic concern with this multivariate model and several others is 'what mechanism exists within asset markets, which creates these time varying correlations and are the dynamics proposed realistic and in line with the general market structure theory?' The first commentary may be made in the context of the elegance of the multivariate form of time varying heteroscedasticity, which has overtaken its intuitive value, in that one would expect under a heterogeneous agent based market that correlation would in some form vary, Cont (2005) [57]suggests that correlation emerges from simple interactions and that correlation and variance and hence variance are not static but dynamic and in some ways stochastic about some

fundamental linear dependency structure. From a more empirical perspective Gallant, Hseih and Tauchen (1991) [102] demonstrate that the conditional innovations of a generalized dependency model will in fact not have a consistent distribution. It is this finding to which we apply the concept which we will call 'categorized state space dependency', which will be discussed in the following section. The theory of market categorization has been put forward strongly by Barberis and Schleifer (2003) [18]and Barberis, Schleifer and Wurgler (2002) [19], who suggest that investors will group assets into categories, before making investment decisions. Merton (1997) [197] addresses this categorization suggesting that it could be the result of transaction costs, international trading restrictions or simply un-informed agents. The interaction of these heterogeneous agents it is argued, creates complex time dependent correlation structures. Barberis and Schleifer, (2003) [18] suggest that there are two possible types of agent acting as investors or facilitators for investment within the market, category and habitat. Their names are somewhat self identifying, category traders tend act over the whole market, dividing it into subsets, habitat, tend to exist in one area of stocks, for example value stocks or biotechnologies. Williams and Ioannidis (2005) [263] suggest that correlation may be a consequence of investor response to past shocks and those positive shocks may actually produce alterations to the correlations, between assets over and above changes in variance, due to realignment of the

weighting applied to various balance sheet components. Finally time varying correlation has been the subject of considerable interest in recent years in the derivatives literature, Joshi (2003) [157], Pelsser (2003) [218] and Rebonato (2004) [223] all address the pricing and hedging of derivative instruments based on implied correlation in the underlying assets, be they FX, equity or interest rates. The literature extends into the idea of dependent and stochastic correlation, addressed in Hull et al (2005) [149] and Collin-Dufresney and Goldsteinz (2001) [53] respectively. This type of auto-correlation is in many respects orthogonal to the cross sectional correlation which we have discussed before in that correlation may be seen amongst forward rates of the same underlying asset(s), however the functional dynamics of this correlation is similarly enigmatic and difficult to isolate. In this article we construct a very simple general model of discrete time dynamic covariation and extend it into three separate but linked models.

5.3 Some Observable Empirical Evidence

The major goal of the multivariate time series literature is to produce a simple and easy to estimate model of time varying covariance which captures the majority of the fluctuations in the conditional covariance matrix. Consider the following simple model, where $\mathbf{y}_t$ is a vector of returns from the S&P 500 cross section, by applying a

simple univariate ARX(p) model for first stage filtration, where,

$$y_{i.t} = f\left(y_{i,t-1}, y_{i,t-2}, ..., y_{i,t-p}\right) + g\left(x_t, z_t\right) + u_{i,t} \tag{5.3.1}$$

and $u_{i,t} \in \mathbf{u}_t$, where $\mathbf{u}_t$ is an n length column vector of residuals and x_t is the market return at time t, z_t an appropriate risk free rate and $t \in [1, ..., \tau]$. The estimated residuals $\hat{\mathbf{u}}_t$ from each model are collected, forming the data matrix $\mathbf{U} = [\mathbf{u}_{t=1}, \mathbf{u}_{t=2}, ..., \mathbf{u}_{t=\tau}]^{\mathrm{T}}$, the estimate of the unconditional covariance matrix is therefore $\hat{\mathbf{\Sigma}} = \frac{1}{\tau}\mathbf{U}^{\mathrm{T}}\mathbf{U}$. Using Cholesky decomposition the vector of eigenvalues ς is computed. Consider the eigenvalues of $\frac{1}{\tau}\mathbf{U}^{\mathrm{T}}\mathbf{U}$, plotted in for a sample period of 20 years, where $\mathbf{u}_t$ are the residuals from the simple filtration, 5.3.1, of daily data. From the 500 firms, 423 have data available over the whole sample period. Now consider the evolution of the largest eigenvalue of the instantaneous matrix $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$, from the S&P 500, The empirical distribution of the largest eigenvalue, φ_{max} of $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$, is illustrated in, 5.3, Compare this distribution to an empirically estimated one generated from using following method,

$$\hat{\mathbf{U}} = \mathbf{E}\left(\hat{\mathbf{\Sigma}}^{\frac{1}{2}}\right)^{\mathrm{T}} \tag{5.3.2}$$

where,

$$\mathbf{E} = \left[\varepsilon_{t=1}^{\mathrm{T}}, \varepsilon_{t=2}^{\mathrm{T}}, ..., \varepsilon_{t=\eta}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.3.3}$$

$$\varepsilon_t \sim N\left(\mathbf{0}, \mathbf{I}\right) \tag{5.3.4}$$

This evidence suggests that either the multivariate characteristic function generating

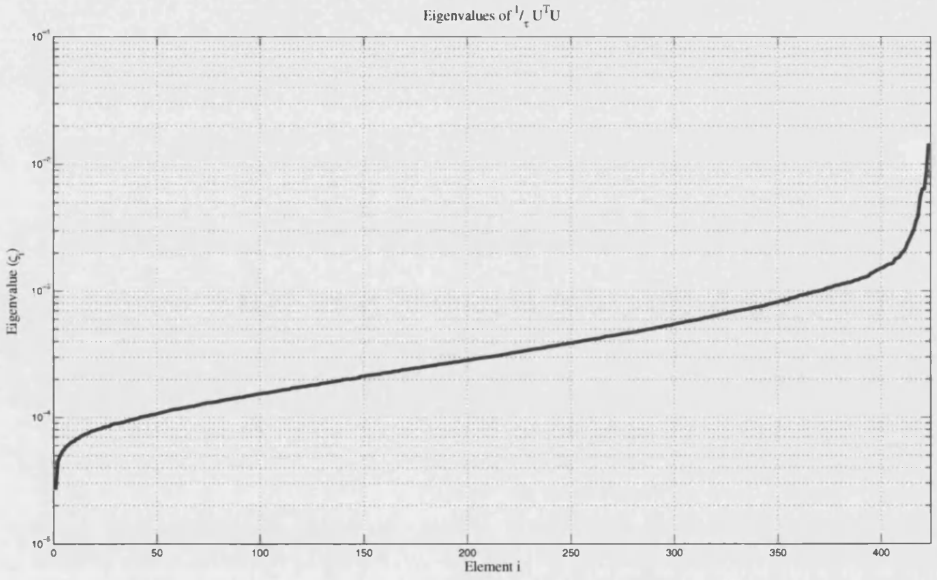

Figure 5.1: A plot of the Eigenvalues of the sample covariance matrix, $\frac{1}{\tau}\mathbf{U}^{\mathrm{T}}\mathbf{U}$, from 423 firms out of the S&P 500, for daily data over 20 years.

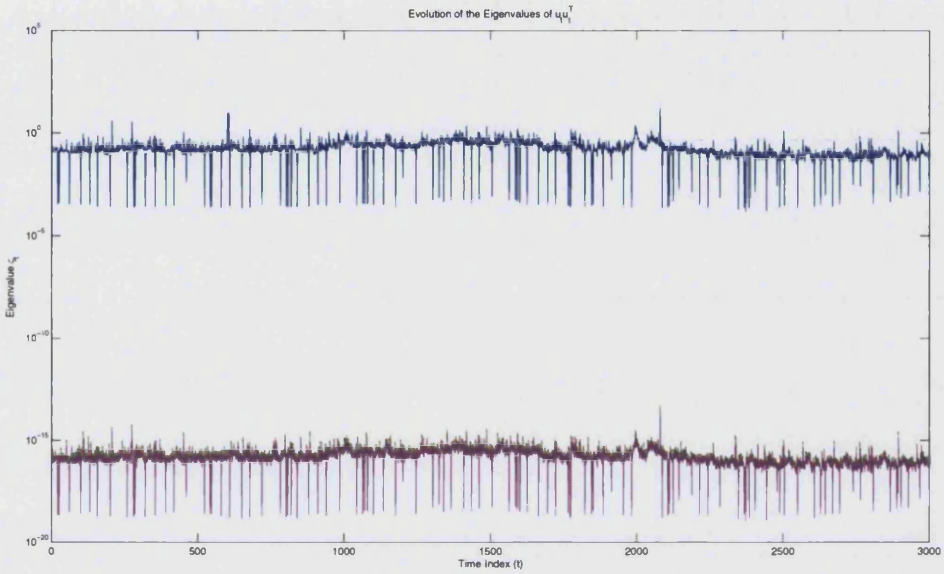

Figure 5.2: Evolution of the Eigenvalues of $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$, over 3000 days, original stock price data source: DataStream$^{\mathrm{TM}}$.

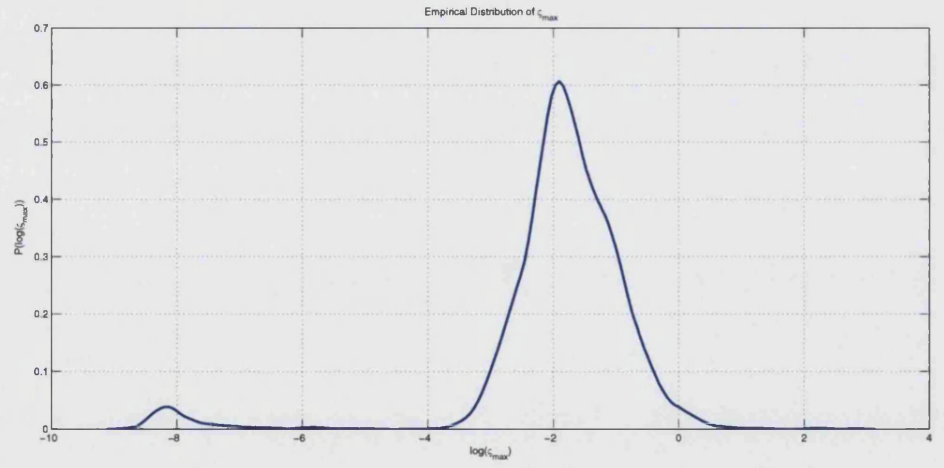

Figure 5.3: Empirical distribution of the largest eigenvalue of $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$

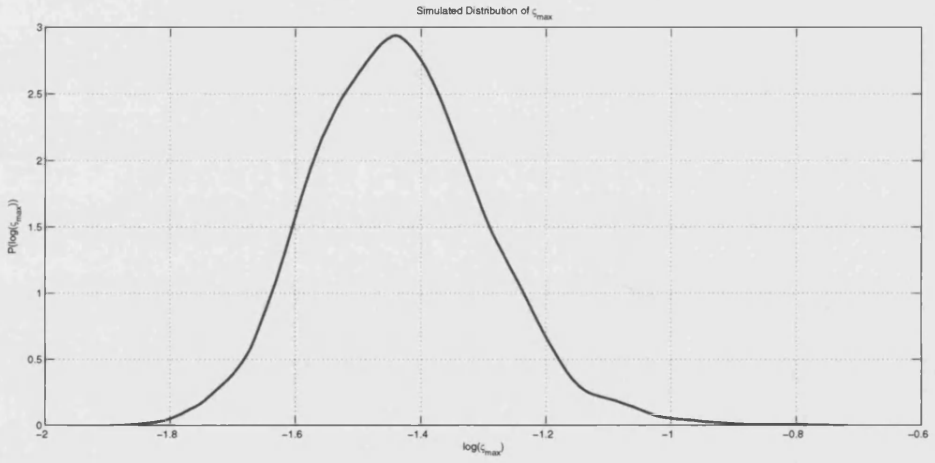

Figure 5.4: Simulated distribution of the largest eigenvalue of $\hat{\mathbf{u}}_t \hat{\mathbf{u}}_t^{\mathrm{T}}$, where $\mathbf{u}_t \sim N\left(\mathbf{0}, \hat{\mathbf{\Sigma}}\right)$. It is immediately apparent that for large covariance matrices the distribution of the largest eigenvalue for draws from a zero entered multi-normal distribution are far more tightly distributed than the distribution observed from 20 years of data from the S&P 500.

the second moments of the distribution of the filtered asset returns is significantly different from a multi-normal distribution or more plausibly we are observing some form of mixture, most likely of multi-normal distributions. From this observed information we can infer that $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t)$, where $\mathbf{\Sigma}_t$, is a matrix process, furthermore that this process is conditioned on some information set Ω_t and as such,

$$\mathbf{\Sigma}_t = \Psi\left(\Omega_{t|t-1,t-2,\ldots,t=1}\right) \tag{5.3.5}$$

Where Ψ is a matrix function operating over Ω_t, and results in the non-negative definite matrix $\mathbf{\Sigma}_t$. In general this is normally a matrix autoregressive process, under some specifications a vector process is used and matricized to form the conditional covariance matrix. The next section reviews two general specifications and outlines their uses and limitations.

5.4 Capturing the Dynamic Covariance Effects

Two common specifications of the covariance process are the MARCH model of Engle and Ding (2002) [72] and the BEKK model of Engle and Kroner (1994) [?], the following section looks at both models in detail.

5.4.1 The MARCH Model

The ethos of the MARCH model was to design a very simple and easy to use approach to modelling high-variate conditional covariance systems. The model offers a relatively simple set of dynamics, but despite this built in simplicity, it is still

computationally very intensive to estimate for these largest systems.

$$\mathbf{\Sigma}_t = \bar{\mathbf{\Sigma}} + \sum_{i=1}^{p} \mathbf{A}_i \times \left(\mathbf{u}_{t-i}\mathbf{u}_{t-i}^{\mathrm{T}}\right) + \sum_{j=1}^{q} \mathbf{B}_j \times \mathbf{\Sigma}_{t-j} \tag{5.4.1}$$

where, $\mathbf{A}_i \in \mathbb{C}^{n\times n}$, $\mathbf{B}_i \in \mathbb{C}^{n\times n}$ and $\bar{\mathbf{\Sigma}} \in \mathbb{C}^{n\times n}$, these parameter matrices may be specified in an unconstrained optimization using the *ivech* transformation,

$$\mathbf{A}_i = (ivech\mathbf{a}_i)\,(ivech\mathbf{a}_i)^{\mathrm{T}} \tag{5.4.2}$$

$$\mathbf{B}_j = (ivech\mathbf{b}_j)\,(ivech\mathbf{b}_j)^{\mathrm{T}} \tag{5.4.3}$$

$$\bar{\mathbf{\Sigma}} = (ivech\mathbf{k})\,(ivech\mathbf{k})^{\mathrm{T}} \tag{5.4.4}$$

And the individual parameter vectors are then $\mathbf{k} \in \mathbb{R}^{n(n+1)\times 1}$, $\mathbf{a}_i \in \mathbb{R}^{n(n+1)\times 1}$ and $\mathbf{b}_j \in \mathbb{R}^{n(n+1)\times 1}$, the full specification parameter θ vector is therefore,

$$\theta = \left[\mathbf{k}^{\mathrm{T}}, \mathbf{a}_1^{\mathrm{T}}, ..., \mathbf{a}_p^{\mathrm{T}}, \mathbf{b}_1^{\mathrm{T}}, ..., \mathbf{b}_q^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.4.5}$$

The stationarity conditions, i.e.

$$\mathbb{P}\left(\mathrm{tr}\left(\mathbf{\Sigma}_{t|t\in[1,...,\tau]}\right) < \infty\right) = 1 \tag{5.4.6}$$

are fairly simple and are defined as follows, using a tensor fibre based approach, where

$$\mathbf{A}_i = [a_{i,j,k}] \tag{5.4.7}$$

$$\mathbf{B}_i = [b_{i,j,k}] \tag{5.4.8}$$

and,

$$\mathcal{A} = [\mathbf{A}_1, ..., \mathbf{A}_p] \tag{5.4.9}$$

$$\mathcal{B} = [\mathbf{B}_1, ..., \mathbf{B}_q] \tag{5.4.10}$$

a simple permutation and fibre extraction yields

$$\underset{p\times 1}{\lambda_{j,k}} = [a_{1,j,k}, a_{2,j,k}, ..., a_{p,j,k}]^{\mathrm{T}} \tag{5.4.11}$$

$$\underset{q\times 1}{\gamma_{j,k}} = [b_{1,j,k}, b_{2,j,k}, ..., b_{q,j,k}]^{\mathrm{T}} \tag{5.4.12}$$

for stationarity the process underlying each fibre must have stationary roots, i.e. the roots of the following sets of polynomial pairs,

$$1 - \left(\lambda_{1,j,k}z + \lambda_{2,j,k}z^2 + ... + \lambda_{p,j,k}z^p\right) = 0 \tag{5.4.13}$$

$$1 - \left(\gamma_{1,j,k}z + \gamma_{2,j,k}z^2 + ... + \gamma_{p,j,k}z^p\right) = 0 \tag{5.4.14}$$

must lie within the unit circle, by setting $a_i \in \mathbb{R}_+$ and $b_i \in \mathbb{R}_+$, this simplifies to, $\sum_{i=1}^{p} \lambda_i = 1$ and $\sum_{i=1}^{q} \gamma_i = 1$. The MARCH specification is typically of use in the construction of very large conditional covariance matrices, however even with the use of the simple element by element approach to the time varying dynamics, maximum likelihood estimation still requires the construction and inversion of a large number of non-negative matrices, coupled with the lack of cross variation in the diagonal and off-diagonal elements the model lacks some of the sophistication of it's peers.

The BEKK Model

An alternative specification is the BEKK model of Engle and Kroner (1994), which whilst pre-dating the MARCH model offers a great deal more flexibility in the dynamics of the conditional covariance matrices. This appears to be an attractive feature

when considering fairly large scale covariance matrices. In particular the models intrinsic structure allows it to capture very localized effects in the off-diagonal elements and still ensure that $\mathbf{\Sigma}_t$ is non-negative.

$$\mathbf{\Sigma}_t = \bar{\mathbf{\Sigma}} + \sum_{i=1}^{p} \mathbf{A}_i^{\mathrm{T}} \left(\mathbf{u}_{t-i} \mathbf{u}_{t-i}^{\mathrm{T}} \right) \mathbf{A}_i + \sum_{j=1}^{q} \mathbf{B}_i^{\mathrm{T}} \mathbf{\Sigma}_{t-j} \mathbf{B}_i \tag{5.4.15}$$

another attractive feature of the BEKK model is the simple partitioned structure of the parameter vector, only the parameters forming the unconditional covariance matrix $\bar{\mathbf{\Sigma}}$, need to be parsed via the *vech* transformation to ensure that it is non-negative definite. As such if $\mathbf{A}_i \in \mathbb{R}_+^{n \times n}$, $\mathbf{B}_i \in \mathbb{R}_+^{n \times n}$ the parameter vector is therefore,

$$\theta = \left[\mathbf{k}^{\mathrm{T}}, vec\mathbf{A}_1^{\mathrm{T}}, ..., vec\mathbf{A}_p^{\mathrm{T}}, vec\mathbf{B}_1^{\mathrm{T}}, ..., vec\mathbf{B}_q^{\mathrm{T}} \right]^{\mathrm{T}} \tag{5.4.16}$$

With $\bar{\mathbf{\Sigma}}$, being the same as in the MARCH specification. Both models treat the recursion of the covariance matrix as a matrix autoregressive moving average process. Assuming that the conditional distribution has a multivariate normal density function, $\mathfrak{F}_t \left(\mathbf{u}_t \right)$, the natural log of which is,

$$\log \mathfrak{F}_t \left(\mathbf{u}_t \left| \theta \right. \right) = -\tfrac{1}{2} n \log 2\pi - \tfrac{1}{2} \log \left| \mathbf{\Sigma}_t \right| - \tfrac{1}{2} \mathbf{u}_t^{\mathrm{T}} \mathbf{\Sigma}_t^{-1} \mathbf{u}_t \tag{5.4.17}$$

the overall objective log-likelihood function $\mathfrak{L} \left(\theta \right)$ is therefore,

$$\mathfrak{L} \left(\theta \right) = -\tfrac{1}{2} \tau n \log 2\pi - \tfrac{1}{2} \sum_{t=1}^{\tau} \log \left| \mathbf{\Sigma}_t \right| + \mathbf{u}_t^{\mathrm{T}} \mathbf{\Sigma}_t^{-1} \mathbf{u}_t \tag{5.4.18}$$

Estimation of the parameters proceeds using a standard non-linear optimization approach. The stationarity constraints on the BEKK model are very simple primarily

due to the quadratic form of the model, if $\mathcal{A} = [\mathbf{A}_1, ..., \mathbf{A}_p]$ and $\mathcal{B} = [\mathbf{B}_1, ..., \mathbf{B}_q]$ are the arrays of parameter matrices then for condition 5.4.6 to hold, additionally the following stationarity condition is required to ensure ergodic variance,

$$vec(\mathcal{A})^{\mathrm{T}} vec(\mathcal{A}) + vec(\mathcal{B})^{\mathrm{T}} vec(\mathcal{B}) \leq 1 \tag{5.4.19}$$

Where *vec* is vectorization operator. The BEKK model obviously offers a more wide ranging variability by utilizing the quadratic form, however under all but the simplest specifications the dimensionality of the parameter vector is very problematic and will lead to a very flat objective function. However both models treat the conditional covariance, $\mathbf{\Sigma}_t$ has some form of functional dependence on the past evolution of the system. The quadratic form of each model and the necessary stationarity conditions mean that longer decays and rapid switches in the covariance structure may not be adequately captured using these models.

5.5 An Alternative Approach to Conditional Covariance: The Boundary Model

As an alternative to the direct matrix autoregressive approach, suggested in the MARCH and BEKK models, we suggest a methodology that simplifies the conditional covariance matrix to a weighted system of static matrices, the simplest specification of which is a regime switching boundary model. Consider the following multivariate linear model with disturbances drawn from a zero-centered conditionally multi-normal

mixture,

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times m}{\mathbf{\Pi}^{\mathrm{T}}}\ \underset{m\times 1}{\mathbf{x}_t} + \underset{n\times 1}{\mathbf{u}_t} \tag{5.5.1}$$

Where the disturbances $\mathbf{u}_t$, are drawn from,

$$\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{5.5.2}$$

$$\varepsilon_t \sim N\left(\mathbf{0}, \underset{n\times n}{\mathbf{I}}\right) \tag{5.5.3}$$

Where $\mathbf{\Sigma}_t$ is a matrix process and $\mathbf{\Sigma}_t^{\frac{1}{2}}$, is the Cholesky or upper triangular factorization, where $\mathbf{\Sigma}_t \in \mathbb{C}^{n\times n}$ and is defined as non-negative definite Hermitian, i.e. symmetric with non-zero eigenvalues. A very simple decomposition of $\mathbf{\Sigma}_t$ suggests a second moment model with two boundary states, designated by the subscripts u and d,

$$\mathbf{\Sigma}_t = \psi_t\mathbf{\Sigma}_u + (1-\psi_t)\,\mathbf{\Sigma}_d \tag{5.5.4}$$

Where the boundary matrices $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_d$ are both non-negative definite Hermitian matrices. $\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_u\mathbf{a} > \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_d\mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^n$. We can treat the scalar process ψ_t as some form of dynamic process constrained to the unit field, now consider a function $\xi(\phi_t;\omega)$ with parameter set ω, where ϕ_t is some d dimensional process, $\phi_t \in \mathbb{R}^d$, to be a valid switching function, the following must hold, $\xi : \mathbb{R}^d \to [0,1]$. For example if $d = 1$ the logistic function is a useful transition function,

$$\xi(\phi_t;\alpha,\beta,\delta) = \left(1+\exp\left(-\alpha\left(\phi_t+\beta\right)^{\delta}\right)\right)^{-1} \tag{5.5.5}$$

therefore,

$$\psi_t = \xi\left(\phi_t; \alpha, \beta, \delta\right) \tag{5.5.6}$$

yields,

$$\xi : \mathbb{R} \rightarrow [0, 1] \tag{5.5.7}$$

And as such constrains, $\psi_t \in [0, 1]$ to the unit field. The scalar process ϕ_t may now be virtually any linear or non-linear dynamic process of our choosing. For this example we choose the following autoregressive quadratic form,

$$\phi_t = \sum_{i=1}^{p} \lambda_i^{\mathrm{T}} \left(\mathbf{u}_{t-i}\mathbf{u}_{t-i}^{\mathrm{T}}\right) \lambda_i + \sum_{j=1}^{q} \gamma_j^{\mathrm{T}} \boldsymbol{\Sigma}_{t-j} \gamma_j \tag{5.5.8}$$

Where, λ_i and γ_i are parameters vectors. The model parameter vector θ is therefore defined as follows,

$$\boldsymbol{\Lambda} = [\lambda_1, ..., \lambda_p] \tag{5.5.9}$$

$$\boldsymbol{\Gamma} = [\gamma_1, ..., \gamma_q] \tag{5.5.10}$$

$$\theta = \left[\alpha, \beta, \delta, \left(vec\boldsymbol{\Lambda}\right)^{\mathrm{T}}, \left(vec\boldsymbol{\Gamma}\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5.5.11}$$

The parameters domains are as follows, $\delta \in \mathbb{N}_+$, $\beta \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $\lambda_i \in \mathbb{R}_+^{(n\times 1)}$, $\gamma_i \in \mathbb{R}_+^{(n\times 1)}$, $\boldsymbol{\Sigma}_u \in \mathbb{C}^{n\times n}$ and $\boldsymbol{\Sigma}_d \in \mathbb{C}^{n\times n}$. Where $\mathbb{C}^{n\times n}$ is the set of all non-negative $n \times n$ definite hermitian matrices and $\mathbb{N}_+$, is the set of positive natural integers. The model offers an extremely simple representation of dynamic covariation, however several attractive properties are immediately apparent. The first interesting aspect is that $\boldsymbol{\Sigma}_t$, will always be non-negative hermitian under most common conditions, if the

boundary matrices $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_d$ are non-negative definite. This is easily demonstrated as follows, consider a subset $\mathcal{M}$ with countable number of elements $k \geq 2$ from $\mathcal{N}$, where $\mathcal{N}$ is the countable set of n elements constituting the vector $\mathbf{y}_t$. If for all possible configurations of $\mathcal{M}$ the matrix $\mathbf{\Omega}_t$ is non-negative definite then by extension any matrix $\mathbf{\Sigma}_t$ from $\mathcal{N}$ must be non-negative definite. Consider the case when $k = 2$,

Proposition

If the boundary matrices $\mathbf{\Omega}_u$ and $\mathbf{\Omega}_d$ are non-negative definite hermitian then $\mathbf{\Omega}_t$ will have a real matrix square root.

Proof

Consider the following decomposition of the boundary matrices, $\mathbf{\Omega}_u = \mathbf{K}_u\mathbf{K}_u^{\mathrm{T}}$ and $\mathbf{\Omega}_d = \mathbf{K}_d\mathbf{K}_d^{\mathrm{T}}$, where,

$$\mathbf{K}_u = \begin{bmatrix} \kappa_{u,1} & \kappa_{u,2} \\ 0 & \kappa_{u,3} \end{bmatrix} \tag{5.5.12}$$

$$\mathbf{K}_d = \begin{bmatrix} \kappa_{d,3} & \kappa_{d,2} \\ 0 & \kappa_{d,3} \end{bmatrix} \tag{5.5.13}$$

$$\kappa_{u,i} \in \mathbb{R} \tag{5.5.14}$$

$$\kappa_{d,j} \in \mathbb{R} \tag{5.5.15}$$

By construction $\mathbf{\Omega}_u$ and $\mathbf{\Omega}_d$ are non-negative definite and maybe factorized by the Cholesky algorithm. Now consider the algebraic construction of $\mathbf{\Omega}_t$,

$$\mathbf{\Omega}_t = \psi_t \begin{bmatrix} \kappa_{u,1}^2 + \kappa_{u,2}^2 & \kappa_{u,2}\kappa_{u,3} \\ \kappa_{u,2}\kappa_{u,3} & \kappa_{u,3}^2 \end{bmatrix} + (1 - \psi_t) \begin{bmatrix} \kappa_{d,1}^2 + \kappa_{d,2}^2 & \kappa_{d,2}\kappa_{d,3} \\ \kappa_{d,2}\kappa_{d,3} & \kappa_{d,3}^2 \end{bmatrix} \tag{5.5.16}$$

It is then relatively simple to show that $\mathbf{\Omega}_t$ has an analytic matrix square root and as such maybe factorized using the Cholesky method, setting $\mathbf{Q}_t = \mathbf{\Omega}_t^{\frac{1}{2}}$, the elements of $\mathbf{Q}_t$ will be,

$$q_{1,1} = \left(\psi_t\left(\kappa_{u,1}^2 + \kappa_{u,2}^2\right) - \left(1-\psi_t\right)\left(\kappa_{d,1}^2 + \kappa_{d,2}^2\right)\right)^{\frac{1}{2}} \tag{5.5.17}$$

$$q_{1,2} = \frac{\psi_t\kappa_{u,2}\kappa_{u,3} + \kappa_{d,2}\left(\kappa_{d,3} - \psi_t\kappa_{d,3}\right)}{\left(-\left(1-\psi_t\right)\left(\kappa_{d,1}^2 + \kappa_{d,2}^2\right) + \psi_t\left(\kappa_{u,1}^2 + \kappa_{u,2}^2\right)\right)^{\frac{1}{2}}} \tag{5.5.18}$$

$$q_{2,1} = 0 \tag{5.5.19}$$

$$q_{2,2} = \left(- \kappa_{d,3}\left(1-\psi_t\right) + \kappa_{u,3}^2\psi_t - \ldots \right. \tag{5.5.20}$$

$$\left. \frac{\left(\left(1-\psi_t\right)\kappa_{d,2}\kappa_{d,3} - \psi_t\kappa_{u,2}\kappa_{u,3}\right)^2\left(\frac{1}{-\left(1-\psi_t\right)\left(\kappa_{d,1}^2+\kappa_{d,2}^2\right)+\psi_t\left(\kappa_{u,1}^2+\kappa_{u,2}^2\right)}\right)^{\frac{1}{2}}}{\left(-\left(1-\psi_t\right)\left(\kappa_{d,1}^2 + \kappa_{d,2}^2\right) + \psi_t\left(\kappa_{u,1}^2 + \kappa_{u,2}^2\right)\right)^{\frac{1}{2}}} \right)^{\frac{1}{2}}$$

By definition, the diagonal elements of $\mathbf{Q}_t$ must be > 0 and if they are then $\mathbf{\Omega}_t$ must also be non-negative definite hermitian, QED. The analytic solutions to the elements of $\mathbf{Q}_t$ maybe computed for any length of $k > 1$. By utilizing the Cholesky method an implicit ordering is introduced and subsequently the elements of $\mathbf{Q}_t$ have some implicit structure based on this ordering, which may often result in order specific results.

Proposition

If the boundary matrices $\mathbf{\Omega}_u$ and $\mathbf{\Omega}_d$ are non-negative definite Hermitian then $\mathbf{\Omega}_t$ will be non-negative definite and by extension the matrix $\mathbf{\Sigma}_t$, will be non-negative definite.

Proof

Given that any $k \times k$ matrix $\mathbf{\Omega} \in \mathbb{C}^{k \times k}$, formed from subsets of $\mathcal{N}$ is non-negative and by definition any sub-matrix from $\mathbf{\Sigma}_t \in \mathbb{C}^{n \times n}$ must be non-negative, if $\mathbf{\Omega}_t$ is non-negative as $k \longrightarrow n$ (see appendix I), all matrices are non-negative then so is $\mathbf{\Sigma}_t$, See Horn and Johnson (1985) [146]. The standard proof of non-negativity, is whether the quadratic form of the matrix for an arbitrary real valued $k = 2$-length vector $\mathbf{a}$, is only valued on $\mathbb{R}_+$. It is relatively simple to demonstrate that if,

$$\psi_t = \left(1 + \exp\left(\alpha\left(\phi_t + \beta\right)^{\delta}\right)\right)^{-1} \tag{5.5.21}$$

$$\mathbf{\Omega}_t = \psi_t \mathbf{\Omega}_d + \left(1 - \psi_t\right) \mathbf{\Omega}_d \tag{5.5.22}$$

Substituting 5.5.21 into 5.5.16 then the quadratic form, $\sigma = \mathbf{a}^{\mathrm{T}} \mathbf{\Omega}_t \mathbf{a}$, may be simplified to,

$$\sigma = \frac{\kappa_{u,1}^2 a_1^2 + \left(\kappa_{u,2}^2 a_1 + \kappa_{u,3}^2 a_2\right)^2 + \exp\left(\alpha\left(\phi_t + \beta\right)^{\delta}\right)\left(\kappa_{d,1}^2 a_1^2 + \left(\kappa_{d,2}^2 a_1 + \kappa_{d,3}^2 a_2\right)^2\right)}{1 + \exp\left(\alpha\left(\phi_t + \beta\right)^{\delta}\right)} \tag{5.5.23}$$

Given that simple inspection demonstrates that both the numerator and denominator of the expression of σ cannot be negative then by extension $\mathbf{\Omega}_t$, must be non-negative,

QED.

5.5.1 Eigenvalues

For completeness the algebraic limit for the smallest eigenvalue of $\mathbf{\Omega}_t$, must always be > 0. Let ς_i, be the smallest eigenvalue of $\mathbf{\Omega}_t$, for $k = 2$ let the transition function be a generalized logistic function, $\xi(\phi_t, \alpha, \beta, \delta)$, substituting for ψ_t, and assuming $\phi_t \in \mathbb{R}_+$, the eigenvalue decomposition of $\mathbf{\Omega}_t$ is therefore, if the diagonal of the matrix $\mathbf{Q}_t$, are guaranteed positive then matrix $\mathbf{\Sigma}_t$ is by definition non-negative, therefore setting,

$$k_u = \kappa_{u,1}^2 + \kappa_{u,2}^2 + \kappa_{u,3}^2 \tag{5.5.24}$$

$$k_d = \kappa_{d,1}^2 + \kappa_{d,2}^2 + \kappa_{d,3}^2 \tag{5.5.25}$$

$$\psi_t = \left(1 + \exp\left(\alpha\left(\phi_t + \beta\right)^{\delta}\right)\right)^{-1} \tag{5.5.26}$$

and using Gaussian the elimination to generate the eigenvalues and taking logs the following results are obtained for the conditional eigenvalues of $\mathbf{\Omega}_t$,

$$\begin{aligned}\log \varsigma_{1,t} &= -\log 2 - \log\left(\psi_t^{-1}\right) + \log\left(k_d + k_d\right) \\ &\quad -\tfrac{1}{2}\log\Big(-4\exp\left(2\alpha\left(\phi_t+\beta\right)^{\delta}\right)\kappa_{d,1}^2\kappa_{d,3}^2 - 4\kappa_{u,1}^2\kappa_{u,3}^2 \\ &\quad +\left(\left(\exp\left(\alpha\left(\phi_t+\beta\right)^{\delta}\right)\right)k_d + k_u\right)^2 \\ &\quad -4\exp\left(\alpha\left(\phi_t+\beta\right)^{\delta}\right)\kappa_{d,3}^2\left(\kappa_{u,1}^2+\kappa_{u,2}^2\right) \\ &\quad -2\kappa_{d,2}^2\kappa_{d,3}^2\kappa_{u,2}^2\kappa_{u,3}^2 + \kappa_{u,3}^2\left(\kappa_{d,1}^2+\kappa_{d,2}^2\right)\Big)\end{aligned} \tag{5.5.27}$$

$$\begin{aligned}\log \varsigma_{2,t} &= -\log 2 - \log\left(\psi_t^{-1}\right) + \log\left(k_d + k_d\right) \\ &\quad +\tfrac{1}{2}\log\Big(-4\exp\left(2\alpha\left(\phi_t+\beta\right)^{\delta}\right)\kappa_{d,1}^2\kappa_{d,3}^2 - 4\kappa_{u,1}^2\kappa_{u,3}^2 \\ &\quad +\left(\left(\exp\left(\alpha\left(\phi_t+\beta\right)^{\delta}\right)\right)k_d + k_u\right)^2 \\ &\quad -4\exp\left(\alpha\left(\phi_t+\beta\right)^{\delta}\right)\kappa_{d,3}^2\left(\kappa_{u,1}^2+\kappa_{u,2}^2\right) \\ &\quad -2\kappa_{d,2}^2\kappa_{d,3}^2\kappa_{u,2}^2\kappa_{u,3}^2 + \kappa_{u,3}^2\left(\kappa_{d,1}^2+\kappa_{d,2}^2\right)\Big)\end{aligned} \tag{5.5.28}$$

The 2×2 example used here maybe scaled to any size $k \times k$ example and will yield the same result.

5.5.2 Model Properties

The model has two main operational modes, first when α is very large, either by parameterization or through *a priori* specification the model has the effect of generating rapid switches between the boundary matrices. The alternative specification restricts the size of α and the conditional covariance is characterized by a continuum of matrices between $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_d$. However the criticality of speed of this adjustment is dependent on all the other parameters, $\theta = \left[\lambda^{\mathrm{T}}, \gamma^{\mathrm{T}}, \alpha, \beta, \delta\right]^{\mathrm{T}}$. The asymptotic behavior of the first model is relatively simple to demonstrate as the extensive literature on finite mixture models is well developed and essentially resolves to the following problem,

$$\mathbb{P}_{\tau\to\infty}\left(\Sigma_t \approx \Sigma_u\right) = f\left(\lambda, \gamma, \alpha, \beta, \delta\right) = \nu_1 \tag{5.5.29}$$

$$\mathbb{P}_{\tau\to\infty}\left(\Sigma_t \approx \Sigma_d\right) = 1 - \mathbb{P}_{\tau\to\infty}\left(\Sigma_t = \Sigma_u\right) \equiv 1 - f\left(\lambda, \gamma, \alpha, \beta, \delta\right) = \nu_2 \tag{5.5.30}$$

Where $\mathbb{P}_{\tau\to\infty}\left(\Sigma_t \approx \Sigma_u\right)$ is the asymptotic probability of the conditional covariance matrix being the upper boundary. The asymptotic density function of the model, $\mathfrak{P}(.)$ will be as follows,

$$\mathfrak{P}\left(\mathbf{u}_t\right) = \nu_1 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{\Sigma}_u\right.\right) + \nu_2 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{\Sigma}_d\right.\right) \tag{5.5.31}$$

where $\mathfrak{F}(.)$ is the multivariate normal density function, the probabilities are conditional on the assumptions underlying the noise process ε_t. More formally the i.i.d. disturbance process is drawn from a matrix normal distribution, $\mathbf{E} = \left[\varepsilon_{t=1}^{\mathrm{T}}, \varepsilon_{t=2}^{\mathrm{T}}, ..., \varepsilon_{t=\tau}^{\mathrm{T}}\right]^{\mathrm{T}}$,

where the distribution of the rows and columns is matrix normal $\mathbf{E} \sim N(\mathbf{0}, \mathbf{I}, \mathbf{I})$. Given this matrix normal assumption and the fact assumption that the distribution of the quadratic form ϕ_t is non-central χ^2 distribution.

5.5.3 Relaxing Some of the Constraints

As previously stated once the assumption of rapid switching is rejected then the asymptotic properties of the model become more complex, however an interesting problem is the following limit,

$$\lim_{\theta \to \theta_0} \left(\mathfrak{P}(\mathbf{u}_t) \to \nu_1 \mathfrak{F}(\mathbf{u}_t | \mathbf{0}, \mathbf{\Sigma}_u) + \nu_2 \mathfrak{F}(\mathbf{u}_t | \mathbf{0}, \mathbf{\Sigma}_d) \right) \tag{5.5.32}$$

At present the following ad-hoc inequality is proposed, $\alpha^{\frac{2}{\delta}} > \left(\lambda^{\mathrm{T}} \gamma\right)^{\frac{1}{\delta}}$, this accounts for the size of the second moment variation in ϕ_t and the speed of the adjustment factor, δ. During the following discussion models with this structure imposed on them will be designated as a boundary switching model BSM. The next problem to address is estimation, as previously illuded to, maximum likelihood methods are only adequate when the size of the system being considered is small, for problems involving large numbers of variables, (in this case large is considered to be > 100), then the objective function is far too flat to give reasonable parameter estimates and the amount of computation time is very prohibitive. One of the attractive properties of the BSM specification is that the boundary matrices, BM, maybe imposed prior to the optimization. Once the, BM, have been estimated, maximum likelihood can

proceed as normal, however only two matrix inversions are necessary during the entire estimation, ensuring a quick convergence. Therefore the total number of parameters needed to estimate an n dimensional model is $2n+3$, i.e a linear increase in parameters by contrast the BEKK and MARCH models have a quadratic increase in parameters.

Note on Stationarity

By construction the resultant vector process $\mathbf{u}_t$ is always stationary if $\psi_t \in [0,1]$. However the only model specifications of interest are when, $\mathbb{P}\left(\sum_{t=1}^{\tau} \phi_t^2 < \infty\right) = 1$, i.e. when the underlying one dimensional driving process is ergodic. The following restrictions are proposed to ensure stationarity in ϕ_t,

$$0 < \lambda^{\mathrm{T}}\lambda \leq 1 \tag{5.5.33}$$

$$0 < \gamma^{\mathrm{T}}\gamma \leq 1 \tag{5.5.34}$$

5.6 Identification of Boundary Matrices

Consider a mixture distribution with the following conditional distribution, $\mathfrak{P}(.)$,

$$\mathfrak{P}_t\left(\mathbf{x}\right) = \sum_{i=1}^{k} \nu_{i,t} \mathfrak{F}\left(\mathbf{x} \,|\mathbf{0}, \mathbf{\Sigma}_i\right) \tag{5.6.1}$$

and the log likelihood is,

$$\mathfrak{L}\left(\theta\right) = \sum_{t=1}^{\tau} \log \mathfrak{P}_t\left(x \,|\theta\right) \tag{5.6.2}$$

where $\sum_{i=1}^{k} \nu_{i,t} = 1$, $\mathfrak{F}(.)$ is the zero centred multivariate normal distribution with covariance matrix $\mathbf{\Sigma}_i$ and $\psi_{i,t}$ is the conditional weighting at time t. In a fully

switching model then if $\psi_{i,t|\{t-1,t-2,...,t-p\}} = 1$ for all other $j \in [1,...,k], j \neq i$, then $\psi_{j,t|\{t-1,t-2,...,t-p\}}$. Thus for every $t \in [1,...,\tau]$, the system is instantaneously drawn from a single distribution characterized by parameters $\mathbf{\Sigma}_i$ drawn from a finite set of distributions. This is the simplest case, which allows for direct inference of the mixture components by some secondary filtration mechanism. A more complex model assumes merely that $\sum_{i=1}^{k} \psi_{i,t} = 1$, and the vector process, ψ_t is some function of its past evolution and some exogenous vector process $\mathbf{z}_t$, i.e. $\psi_t = f(\psi_{t-1}, \psi_{t-1}, ..., \psi_{t-p}, \mathbf{z}_t)$.

5.6.1 Estimating the Switching model in Two Stages

Assuming that $\psi_{i,t|\{t-1,t-2,...,t-p\}} = 1$ for all other $j \in [1,...,k], j \neq i$ and setting the number of distributions in the mixture to two, $k = 2$, by design $\psi_{2,t} = 1 - \psi_{1,t}$, then given that the uniqueness of the two components of the mixture is characterized by the covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, then it is first necessary to construct identify a test boundary of the following hypothesis,

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \equiv \mathbf{\Sigma} \tag{5.6.3}$$

For this purpose the Takemura and Kuriki (2000) [246] test of sample covariance matrix equality, is used. Consider a $\tau \times n$ data matrix $\mathbf{X}$, drawn from some conditional mixture of multivariate normal distributions. Now consider a $\tau_1 \times n$ sub-sample $\mathbf{X}_1$ and a $\tau_2 \times n$ sub-sample $\mathbf{X}_2$, selected through some systematic methodology, for example the von-Mises step method, or via the distribution of the maximum eigenvalue

ς_{max} of the set of matrices $\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}$, were $t \in [1, ..., \tau]$. The following assumptions apply, $\tau_1 + \tau_2 = \tau$ and if $\mathbf{x}_{1,t} \in \mathbf{X}_1$ and $\mathbf{x}_{2,t} \in \mathbf{X}_2$ then $\mathbf{x}_{1,t} \notin \mathbf{X}_2$ and $\mathbf{x}_{2,t} \notin \mathbf{X}_1$ hold. Defining the subset sample covariance matrices as follows, $\hat{\mathbf{\Sigma}}_1 = \frac{1}{\tau_1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1$, $\hat{\mathbf{\Sigma}}_2 = \frac{1}{\tau_2}\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2$ and $\hat{\mathbf{\Sigma}} = \frac{1}{\tau}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, the following test statistic is proposed by Takemura and Kuriki (2000) [246] as an extension to the method set out in Roy (1953),

$$\vartheta = \max_{a,b \in \mathbb{R}^n} \left(\frac{\left(\mathbf{a}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_1\mathbf{b}\right)\left(\mathbf{a}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_2\mathbf{b}\right)}{\left(\left(\tau_1^{-1} + \tau_2^{-1}\right)\left(\mathbf{a}^{\mathrm{T}}\hat{\mathbf{\Sigma}}\mathbf{a}\right)\left(\mathbf{b}^{\mathrm{T}}\hat{\mathbf{\Sigma}}\mathbf{b}\right) + \left(\mathbf{a}^{\mathrm{T}}\hat{\mathbf{\Sigma}}\mathbf{b}\right)^2\right)^{\frac{1}{2}}} \right) \tag{5.6.4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are real value parameter vectors, maximization is undertaken using a standard quadratic programming approach. The under the null the statistic $\vartheta \sim \chi^2(\nu)$, with $\nu = \frac{1}{2}n(n+1)$ degrees of freedom, i.e. one degree of freedom for each unique element in $\mathbf{\Sigma}$. Therefore if $\varphi > z$, where,

$$z = \mathfrak{H}^{-1}(p\,|\nu) = \{z : \mathfrak{H}(z\,|\nu) = p\} \tag{5.6.5}$$

$$p = \mathfrak{H}(z\,|\nu) = \int_0^z \mathfrak{F}(z\,|\nu)\,dx \tag{5.6.6}$$

and p is some chosen probability and $\mathfrak{F}(.)$ is the $\chi^2(\nu)$ density function with degrees of freedom parameter ν, then $H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$ is accepted. If there is stylistic evidence to support the assumption that the number of distributions in the mixture is $k > 2$, then the method is extensible to a k^{th}-order mixture, forming the lower triangular matrix, $\underset{k\times k}{\mathbf{\Phi}} = [\vartheta_{i,j}]$. In practice for $k = 2$ this approach is a reasonably tractable and a computationally less intensive method of tackling conditional covariance problems

than the normal MV-GARCH approaches. Take a large covariance matrix estimation problem, such as estimating the conditional covariance of the S&P 500, the major computational hurdles in this approach are the identification of the mixture component matrices and the subsequent estimation of the parameters of the recursive function $f(.)$. After identifying possible data subsets $\mathbf{X}_1$ and $\mathbf{X}_2$ the major computation difficulty is in the quadratic optimization that underlies the estimation of the test statistic φ. However this is a very conventional quadratic optimization problem and as such very large scale algorithms are available which can robustly approach this problem in a relatively small number of operations, the MIN-PACK (2000) [199] documentation, Fletcher (1994) [96] or Press *et al* (1997) [261]. Once the boundary matrices are identified then only $k = 2$ matrix inversions are required in order to partially parameterize the maximum likelihood function. Restriction testing for the exact identification of $f(.)$, is then relatively straight forward assuming a reasonably simple specification for the evolution of ψ_t.

5.6.2 Empirical Example of the Regime Switching Model

The BSM model is demonstrated using the filtered residuals from the S&P 500 dataset initially analyzed in 5.3, these residuals are filtered using the maximum eigenvalue approach and the boundary covariance matrices are estimated. The Takemura and Kuriki maximum equality test (MEQT) is then applied. Accepting the Boundary hypothesis, the boundary matrices are inverted and the model estimated using the

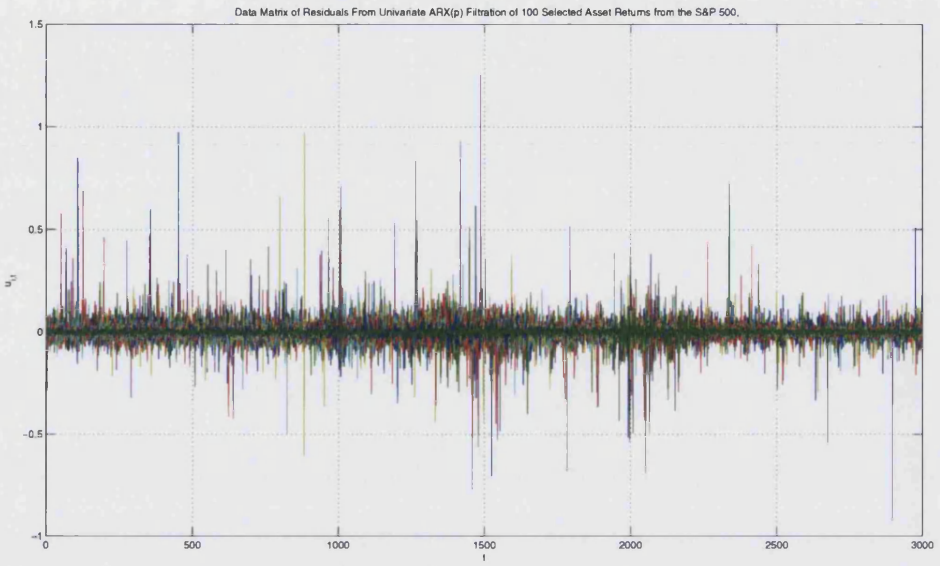

Figure 5.5: 100 Randomly samples stocks from the S&P 500 over 3000 daily observations.

various parameter constraints. Assuming that the parameter estimates are normally distributed the information matrix is estimated and the parameter standard errors computed.

The Data

For computational reasons a randomly sampled 100 asset sub-sample of the master dataset is chosen, figure 5.5 illustrates the residuals from the univariate filtration, Using the empirical distribution of the largest eigenvalues, a high-band filter is used to create two data matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ for MEQT stage, table The High Band filters

Table 5.1: Table of MEQT statistics from a variety of filtration specification.

Filtration	MEQT statisitic	Degrees of Freedom ν	Threshold (95%)
High Band 1	283.5858	100	124.3421
High Band 2	238.3220	100	124.3421
High Band 3	97.3510	100	124.3421

Table 5.2: Estimation Results from the estimation of θ, using the BSM model structure.

Model Specification	Parameters	Log - Likelihood	LR - Ratio versus Mod1	$\chi^2(\nu)$
Mod1 $q = 1$,$p = 1$	303	-437827.9359	N/A	
Mod2 $q = 2$,$p = 1$	403	-437798.3894	-29.5465	124.3421
Mod3 $q = 1$,$p = 2$	403	-437650.2673	-177.6686	124.3421
Mod4 $q = 2$,$p = 2$	503	-437648.9149	-179.0210	233.9943

are illustrated in figure 5.6,

Model Estimation

The next stage in the model estimation is the optimization that yields the estimated parameter vector θ. The model is estimated over a number of lags the basic specification (Mod1) uses one lag with 303 parameters,

Illustrating the Dynamic Correlations

Figure 5.7, illustrates the evolution of the dynamic correlation, $\rho_{i,j,t} = \sigma_{i,j,t}\left(\sigma_{i,t}\sigma_{j,t}\right)^{-1}$, between five randomly chosen stocks from the S&P 500, the first line (blue) represents the BSM and the second (red), the STBM. Whilst the BSM model appear to capture the direction of the dynamics of the conditional correlation the STBM model appears

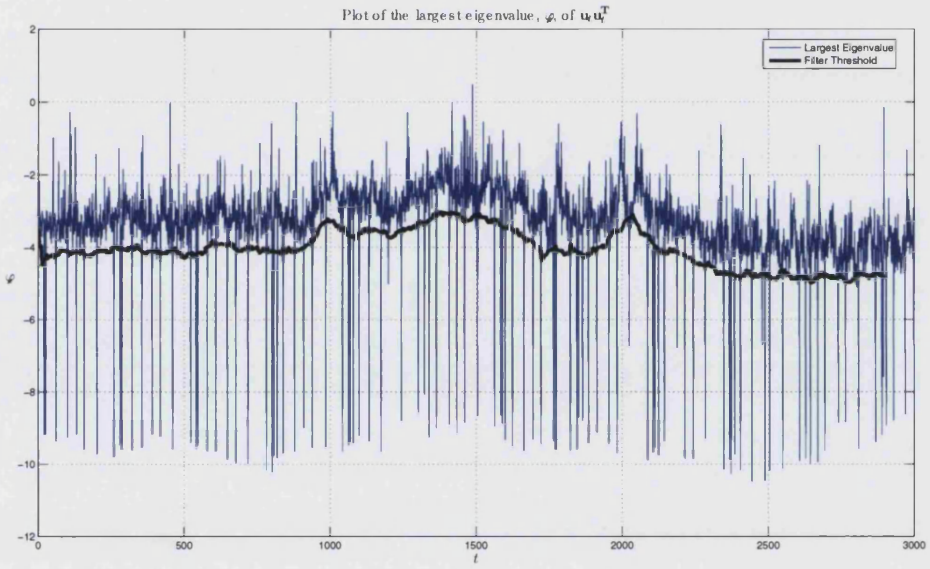

Figure 5.6: The High-Band Filter Schemes are illustrated as a set of arbitrary lines differentiating the proposed residual states.

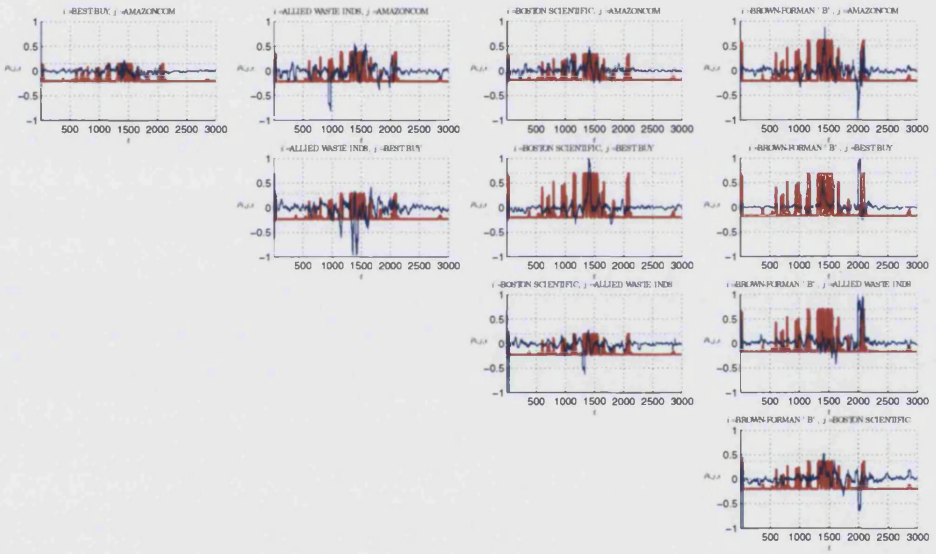

Figure 5.7: The Dynamic correlations of 5 sample assets from the 100 chosen, against the estimated state of the system.

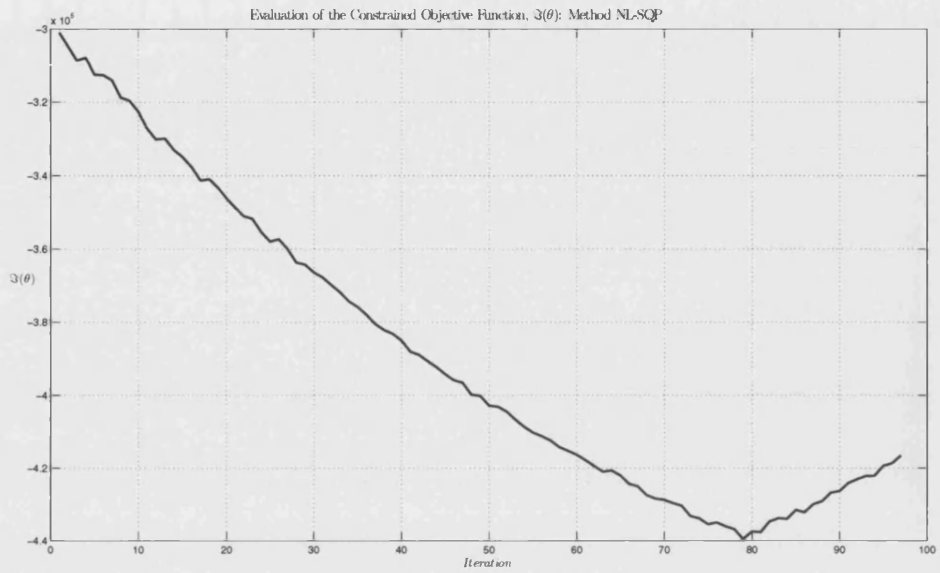

Figure 5.8: The Iterations of the Objective function of the model

to capture the magnitude more effectively. Whilst the model cannot capture the variety of dynamics that the BEKK model, the major innovation is in the ability to tractably estimate a conditional covariance model for very large systems. Figure 5.8, illustrates that even for a 100-variate system the likelihood function still has a reasonable gradient, compared to a 10-variate ($p = 1, q = 1$) BEKK and MARCH model.

5.7 Extensions to the Basic Model

As suggested in section 5.5.2, the specification has several useful properties, in particular the tractability in the modelling of large conditional covariance matrices. However the basic model lacks much of the variability of its major alternatives, namely the BEKK and MARCH models. In this section two alternative specifications are briefly outlined, that extend the original model to allow for a greater freedom in the dynamics of the elements of the conditional covariance matrix.

5.7.1 Boundary Volatility-Correlation Model

The simplest approach is to separate the volatility and correlation dynamics, utilizing the standard decomposition,

$$\mathbf{\Sigma}_t = \mathbf{R}_t \circ \mathbf{H}_t \tag{5.7.1}$$

essentially the correlation and volatility dynamics maybe separated, this gives a freedom in terms of the evolution of the dynamics of $\mathbf{\Sigma}_t$. The boundary covariance matrices maybe estimated in the same manner as previously, the volatility (driven by the diagonal elements) and the correlation (driven by the off-diagonal elements), may then be modelled separately, as follows, first decompose the boundary matrices,

using the standard notation,

$$\begin{array}{ll}
\boldsymbol{\Sigma}_u = \mathbf{R}_u \circ \mathbf{H}_u & \boldsymbol{\Sigma}_d = \mathbf{R}_d \circ \mathbf{H}_d \\
\boldsymbol{\Sigma}_u = \begin{cases} \left[\sigma_{i,u}^2\right]_{i=j} \\ \left[\sigma_{i,j,u}\right]_{i\neq j} \end{cases} & \boldsymbol{\Sigma}_d = \begin{cases} \left[\sigma_{i,d}^2\right]_{i=j} \\ \left[\sigma_{i,j,d}\right]_{i\neq j} \end{cases} \\
\mathbf{R}_u = \begin{cases} \left[\rho_{i,j,u}\right]_{i=j} = 1 \\ \left[\rho_{i,j,u}\right]_{i\neq j} = \frac{\sigma_{i,j,u}}{\sigma_{i,u}\sigma_{j,u}} \end{cases} & \mathbf{R}_d = \begin{cases} \left[\rho_{i,j,d}\right]_{i=j} = 1 \\ \left[\rho_{i,j,d}\right]_{i\neq j} = \frac{\sigma_{i,j,d}}{\sigma_{i,d}\sigma_{j,d}} \end{cases} \\
\mathbf{H}_u = \begin{cases} \left[h_{i,j,u}\right]_{i=j} = \sigma_{i,u}^2 \\ \left[h_{i,j,u}\right]_{i\neq j} = \sigma_{i,u}\sigma_{j,u} \end{cases} & \mathbf{H}_d = \begin{cases} \left[h_{i,j,d}\right]_{i=j} = \sigma_{i,d}^2 \\ \left[h_{i,j,d}\right]_{i\neq j} = \sigma_{i,d}\sigma_{j,d} \end{cases}
\end{array} \tag{5.7.2}$$

Now specify two separate dynamics, $\psi_{\mathbf{R},t}$ and $\psi_{\mathbf{H},t}$, for the correlation and volatility processes respectively, for simplicity only the first order $p = 1, q = 1$ version is specified,

$$\psi_{\mathbf{R},t} = \xi\left(\phi_{\mathbf{R},t} \left| \alpha_{\mathbf{R}}, \beta_{\mathbf{R}}, \delta_{\mathbf{R}}\right.\right) \tag{5.7.3}$$

$$\psi_{\mathbf{H},t} = \xi\left(\phi_{\mathbf{H},t} \left| \alpha_{\mathbf{H}}, \beta_{\mathbf{H}}, \delta_{\mathbf{H}}\right.\right) \tag{5.7.4}$$

$$\phi_{\mathbf{R},t} = \lambda_{\mathbf{R}}^{\mathrm{T}} \epsilon_{t-1} \epsilon_{t-1}^{\mathrm{T}} \lambda_{\mathbf{R}} + \gamma_{\mathbf{R}}^{\mathrm{T}} \mathbf{R}_{t-1} \gamma_{\mathbf{R}} \tag{5.7.5}$$

$$\phi_{\mathbf{H},t} = \lambda_{\mathbf{R}}^{\mathrm{T}} \mathbf{u}_{t-1} \mathbf{u}_{t-1}^{\mathrm{T}} \lambda_{\mathbf{R}} + \gamma_{\mathbf{R}}^{\mathrm{T}} \mathbf{H}_{t-1} \gamma_{\mathbf{R}} \tag{5.7.6}$$

where, $\epsilon_{i,t} \in \epsilon_t$ is the normalized residual, i.e. $\epsilon_{i,t} = u_{i,t}\sigma_{i,t}^{-1}$. As in the previous model there are two main model dynamics depending on the constraints placed upon the parameter vector, again for simplicity, only the case where the ad-hoc constraints $\alpha_{\mathbf{R}}^{\frac{2}{\delta_{\mathbf{R}}}} > \left(\lambda_{\mathbf{R}}^{\mathrm{T}}\gamma_{\mathbf{R}}\right)^{\frac{1}{\delta_{\mathbf{R}}}}$ and $\alpha_{\mathbf{H}}^{\frac{2}{\delta_{\mathbf{H}}}} > \left(\lambda_{\mathbf{H}}^{\mathrm{T}}\gamma_{\mathbf{H}}\right)^{\frac{1}{\delta_{\mathbf{H}}}}$ is considered, asymptotically the process distribution will be,

$$\begin{aligned}
\mathfrak{P}\left(\mathbf{u}_t\right) = {} & \nu_1 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{H}_u \circ \mathbf{R}_u\right.\right) + \nu_2 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{H}_d \circ \mathbf{R}_d\right.\right) \\
& + \nu_3 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{H}_u \circ \mathbf{R}_d\right.\right) + \nu_4 \mathfrak{F}\left(\mathbf{u}_t \left| \mathbf{0}, \mathbf{H}_d \circ \mathbf{R}_u\right.\right)
\end{aligned} \tag{5.7.7}$$

where, given the parameter vector, $\theta = \left[\lambda_{\mathbf{H}}^{\mathrm{T}}, \lambda_{\mathbf{R}}^{\mathrm{T}}, \gamma_{\mathbf{H}}^{\mathrm{T}}, \gamma_{\mathbf{R}}^{\mathrm{T}}, \alpha_{\mathbf{R}}, \beta_{\mathbf{R}}, \delta_{\mathbf{R}}, \alpha_{\mathbf{H}}, \beta_{\mathbf{H}}, \delta_{\mathbf{H}}\right]^{\mathrm{T}}$, the asymptotic weights are,

$$\nu_1 = \mathbb{P}\left(\mathbf{\Sigma}_t \approx \mathbf{H}_u \circ \mathbf{R}_u \,|\theta\right) \tag{5.7.8}$$

$$\nu_2 = \mathbb{P}\left(\mathbf{\Sigma}_t \approx \mathbf{H}_d \circ \mathbf{R}_d \,|\theta\right) \tag{5.7.9}$$

$$\nu_3 = \mathbb{P}\left(\mathbf{\Sigma}_t \approx \mathbf{H}_u \circ \mathbf{R}_d \,|\theta\right) \tag{5.7.10}$$

$$\nu_4 = \mathbb{P}\left(\mathbf{\Sigma}_t \approx \mathbf{H}_d \circ \mathbf{R}_u \,|\theta\right) \tag{5.7.11}$$

Again relaxing the ad-hoc constraints and allowing $\psi_{\mathbf{R},t}$ and $\psi_{\mathbf{H},t}$ to vary in the continuum between $[0, 1]$, produces a far more complex asymptotic distribution.

5.7.2 Multiple Regime Models

The second extension was alluded to in section 5.6.1, where the general finite mixture model distribution is as follows,

$$\mathfrak{P}_t\left(\mathbf{u}_t\right) = \sum_{i=1}^{m} \nu_{i,t} \mathfrak{F}\left(\mathbf{x} \,|\mathbf{0}, \mathbf{\Sigma}_i\right) \tag{5.7.12}$$

The general conditional covariance is therefore decomposed as,

$$\mathbf{\Sigma}_t = \sum_{i=1}^{m} \psi_{i,t} \mathbf{\Sigma}_i \tag{5.7.13}$$

where for the first order case,

$$\psi_{i,t} = \xi\left(\phi_{i,t} \,|\alpha_i, \beta_i, \delta_i\right) \tag{5.7.14}$$

$$\phi_{i,t} = \lambda_i^{\mathrm{T}} \mathbf{u}_{t-1} \mathbf{u}_{t-1}^{\mathrm{T}} \lambda_i + \gamma_i{}^{\mathrm{T}} \mathbf{\Sigma}_{t-1} \gamma_i \tag{5.7.15}$$

the following conditions are required, $\sum_{i=1}^{m} \psi_{i,t} = 1$, and $\psi_{i,t} \geq 0$. This model has the advantage of being able to capture a large degree of variation in the elements of $\mathbf{\Sigma}_t$, however each additional matrix requires a filtration to extract that regime, the estimation of the asymptotic distribution weights becomes increasingly complex and the large number of parameters means that the model could cease to offer significant advantages over the alternative BEKK and MARCH specifications.

5.8 Concluding Remarks

This chapter has presented a new method of estimating evolution of the multivariate second moments for high-variate models. The method has a well developed asymptotic theory and captures some of the covariance properties present in real asset return data. The model in it's basic form is shown to be positive definite under the common parameter constraints and is fairly tractable in it's estimation. Interestingly relaxing the core model constraints yields a very flexible model that can capture a great of variability on the diagonal and off-diagonal elements of the conditional covariance matrix. The innovative two stage structure also yields a very helpful mechanism for interpreting the definition of up and down states and this maybe useful in assigning risk bands to collections of assets.

The basic model is not only applicable to modelling the dynamic interdependencies in the asset market, but could also be used in evaluating forward correlations in interest rates and the dynamic dependencies in factors relating to credit markets. Another

useful approach could be to model FX regimes, particularly when *a priori* information maybe utilized to identify the different boundary conditions, such as during implementation of crawling pegs and other policy driven FX market interventions. The model is parameterized over a very high-variate sample of stocks, i.e. 100 asset returns from the S&P 500, an analysis which would not be tractable using the alternative models, the BEKK model of Engle and Kroner and the MARCH model of Engle and Kim. Further developments of the model relate to a complete derivation of the algebraic relationship between the parameter vectors θ and the asymptotic distribution weights ν and further investigation of the asymptotic properties of the model when the ad-hoc switching assumptions are relaxed, i.e. ψ_t, varies continuously in [0,1]. In conclusion this model appears to offer a solution to the middle ground between the fully functional MV-ARCH models with their associated problems, regarding parameters and topology of the objective function and the totally ad-hoc methods such as the RiskMetrics^{TM} smoother and the exponentially weighted correlation model.

5.8.1 Optimization of the Objective Function

Maximum likelihood estimation requires optimization usually via a generic algorithm, we choose the Davidon-Fletcher-Powell (DFP) method for unconstrained optimization and the Lagrange Newton method or Sequential Quadratic Programming (SQP) method in a constrained optimization framework utilizing a non-linear constraint to restrict the spacing of the boundary matrices via a penalty function acting on the

boundary parameters. For either method the general optimization problem is as follows,

$$\min_{\theta}\left[\mathfrak{L}(\theta)\triangleq-\tfrac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+\log\left(|\mathbf{\Sigma}_t|\right)+\mathbf{u}_t^{\mathrm{T}}\mathbf{\Sigma}_t^{-1}\mathbf{u}_t\right)\right] \tag{5.8.1}$$

The analytical gradient is therefore,

$$\nabla\mathfrak{L}(\theta)=\left[\frac{\partial\mathfrak{L}(\theta)}{\partial\theta_1},\frac{\partial\mathfrak{L}(\theta)}{\partial\theta_1},...,\frac{\partial\mathfrak{L}(\theta)}{\partial\theta_k}\right]^{\mathrm{T}} \tag{5.8.2}$$

5.9 Chapter Appendix I: The DFP Algorithm

Consider a matrix of second order derivatives $\mathbf{H}$ of the likelihood function $\mathfrak{L}(\theta)$, where in general $\mathbf{H}=\nabla\mathfrak{L}(\theta)\nabla^{\mathrm{T}}\mathfrak{L}(\theta)$. The following iterative mechanism is suggested, by utilizing a two stage matrix chain rule, (see Apostal (1967), page 273), the gradients of the state parameters is then found in relation to the transition parameters these matrices are concatenated and the resulting matrix multiplied by its transpose to produce the Hessian, this has proved more effective than simply formulating the finite differences. This is somewhat similar to Engle and Shepherd (2001) [89], however the general ethos of the regime mechanism is maintained. The first step forms the analytical first order derivative matrices via the matrix chain rule, a set of finitely differenced matrices, formed by multiplication of the vector fields by the vector field of the transition parameters, for precise details of the product rule for vectors and its

inclusion in the gradient functional.

$$\begin{aligned}
\nabla\mathfrak{L}(\theta) &= \nabla\mathfrak{L}(\theta_i) + (\theta-\theta_i)\cdot\nabla\mathfrak{L}(\theta_i) \\
&\quad +\tfrac{1}{2}(\theta-\theta_i)^{\mathrm{T}}\cdot\boldsymbol{\Omega}\cdot(\theta-\theta_i) \\
\nabla\mathfrak{L}(\theta) &= \nabla\mathfrak{L}(\theta_i) + \boldsymbol{\Theta}\cdot(\theta-\theta_i) \\
\nabla\mathfrak{L}(\theta_i)\cdot(\theta-\theta_i) &= -(\theta-\theta_i)^{\mathrm{T}}\boldsymbol{\Omega}(\theta-\theta_i) < \mathbf{0} \\
\mathbf{H} &= \boldsymbol{\Omega}^{-1}
\end{aligned} \tag{5.9.1}$$

The Hessian must functionally invert the gradient fields back to the original differencing of the parameters:

$$\theta-\theta_i = \mathbf{H}_{i+1}\cdot\nabla\mathfrak{L}_{i+1} - \nabla\mathfrak{L}_i \tag{5.9.2}$$

The Hessian difference function is now of the following form,

$$\begin{aligned}
\mathbf{H}_{i+1} &= \mathbf{H}_i + \frac{(\theta_{i+1}-\theta_i)\otimes(\theta_{i+1}-\theta_i)}{(\theta_{i+1}-\theta_i)\cdot(\nabla\mathfrak{L}_{i+1}-\nabla\mathfrak{L}_i)} - \\
&\quad \frac{(\mathbf{H}_i(\nabla\mathfrak{L}_{i+1}-\nabla\mathfrak{L}_i))\otimes(\mathbf{H}_i(\nabla\mathfrak{L}_{i+1}-\nabla\mathfrak{L}_i))}{(\nabla\mathfrak{L}_{i+1}-\nabla\mathfrak{L}_i)\cdot\mathbf{H}_i\cdot(\nabla\mathfrak{L}_{i+1}-\nabla\mathfrak{L}_i)}
\end{aligned} \tag{5.9.3}$$

The initial Hessian is formed via finite differencing, using the absolute smallest available floating point number the *real minimum.*

$$\mathbf{H}_i = \begin{bmatrix}
\frac{\partial\mathfrak{L}_i}{\partial^2\theta_1} & \frac{\partial\mathfrak{L}_i}{\partial\theta_2\partial\theta_1} & \cdots & \frac{\partial\mathfrak{L}_i}{\partial\theta_2\partial\theta_N} \\
\frac{\partial\mathfrak{L}_i}{\partial\theta_2\partial\theta_1} & \frac{\partial\mathfrak{L}_i}{\partial^2\theta_2} & \cdots & \frac{\partial\mathfrak{L}_i}{\partial\theta_2\partial\theta_N} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial\mathfrak{L}_i}{\partial\theta_N\partial\theta_1} & \frac{\partial\mathfrak{L}_i}{\partial\theta_N\partial\theta_2} & \cdots & \frac{\partial\mathfrak{L}_i}{\partial^2\theta_N}
\end{bmatrix} \tag{5.9.4}$$

This is a quasi Newton type algorithm and may be poorly specified if the variables are badly scaled, continuous miss specification of the optimal parameters may be

dealt with utilizing the simplex method of Nelder and Mead, which does not require the formal approximation of the analytical gradient matrix, see Press et al (2002) [261], for a detailed discussion on implementing multi-variable constrained functional minimization and Fletcher (1986) [96] for theoretical background and discussion of optimization without gradients. Under fairly common conditions the proposed log likelihood function does exhibit this scaling issue, (particularly prevalent because of the discontinuous form of the vector θ), which may mislead optimization algorithms, we therefore propose the following strategy, compute $\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}$, for each observation and then find the largest Eigenvalue within this matrix, sorting this list of eigenvalues an taking the largest an smallest (upper an lower boundaries) yields an estimate of the boundary matrices. In practice this methodology has two benefits, first it imposes a measurable distance between the boundary matrices and this significantly improves the models ability to capture both short persistence and long memory events. Furthermore if you accept this estimate of the boundary matrices

5.10 Chapter Appendix II: The SQP Algorithm

A significant problem with regards to the unconstrained DFP method is that under fairly simple specifications the the last element of the eigenvector of the boundary matrices either collapses together or explodes apart, which results in unsatisfactory dynamics. We have suggested methods via which the boundary matrices maybe empirically estimated, however this may also be unsatisfactory in terms of computation

time and completeness. Therefore we specify the following general constrained programming problem,

$$\min_{\theta}\left[\mathfrak{L}(\theta)\triangleq-\tfrac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi)+\log\left(|\mathbf{\Sigma}_t|\right)+\mathbf{u}_t^{\mathrm{T}}\mathbf{\Sigma}_t^{-1}\mathbf{u}_t\right)\right] \tag{5.10.1}$$

subject to the following constraints

$$\begin{aligned} G_i(\theta)&=0 \qquad i=1,...,m_e \\ G_i(\theta)&\leq 0 \qquad i=m_e+1,...,m \end{aligned} \tag{5.10.2}$$

The Kuhn-Tucker representation is then as follows

$$\mathfrak{L}(\theta,\gamma)=\nabla\mathfrak{L}(\theta)+\sum_{j=1}^{m}\gamma_j.G_j(\theta) \tag{5.10.3}$$

The quadratic sub problem, is then simply,

$$\min_{\mathbf{d}}\left[\tfrac{1}{2}\mathbf{d}^{\mathrm{T}}\mathbf{H}_i\mathbf{d}+\nabla^{\mathrm{T}}\mathfrak{L}(\theta_i)\,\mathbf{d}\right] \tag{5.10.4}$$

Where $\mathbf{d}\in\mathbb{R}^k$ and is the transient differencing vector, this sub problem is subject to the following linearized constraints,

$$\begin{aligned} \nabla^{\mathrm{T}}G_i(\theta)\,\mathbf{d}+G_i(\theta)&=0 \qquad i=1,..,m_e \\ \nabla^{\mathrm{T}}G_i(\theta)\,\mathbf{d}+G_i(\theta)&\leq 0 \qquad i=m_e+1,...,m \end{aligned} \tag{5.10.5}$$

The updated parameter vector is now simply,

$$\theta_{k+1}=\theta_k+\iota_k\mathbf{d}_k \tag{5.10.6}$$

Where ι is finite step length parameter determined via line search. Utilizing the Broydon-Fletcher-Goldfarb-Shanno, (BGFS) formula, see Fletcher (1994), the Hessian is updated as follows,

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\partial_k \partial_k^{\mathrm{T}}}{\partial_k^{\mathrm{T}} s_k} - \frac{\mathbf{H}_k^{\mathrm{T}} \mathbf{H}_k}{s_k^{\mathrm{T}} \mathbf{H}_k s_k} \tag{5.10.7}$$

Where, $s_k = \theta_{k+1} - \theta_k$ and,

$$\partial_k = \nabla \mathfrak{L}(\theta_{k+1}) + \sum_{j=1}^{n} l_j \nabla G_i(\theta_{k+1}) - \left(\nabla \mathfrak{L}(\theta_k) + \sum_{j=1}^{m} l_j . \nabla G_j(\theta) \right) \tag{5.10.8}$$

the non-linear constraint $G(\theta)$, is in the form of the ratio of the last element of the eigenvectors of the boundary matrices $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_d$, i.e.

$$\begin{aligned} \mathbf{\Sigma}_u \mathbf{x}_u &= y_u \mathbf{x}_u \\ \mathbf{\Sigma}_d \mathbf{x}_d &= y_d \mathbf{x}_d \\ \frac{x_{n,u}}{x_{n,d}} &= z \end{aligned} \tag{5.10.9}$$

Where z is an arbitrarily imposed value and in most cases $z > 1$. This is an extremely useful representation that allows a great deal of control over the optimization, whilst still allowing enough flexibility to reach a realistic estimated global minimum.

Chapter 6

Impulse Response Models in Variance and Covariance

6.1 Chapter Abstract

This chapter presents a unified framework for the computation of impulse response functions for VAR models with MV-GARCH disturbances. By solving from the quadratic form as given by the second moment equations, it is demonstrated that these solutions should be used to adjust the impulse response functions from the mean equations. These adjustments often result in significant different time profiles for the unadjusted impulse responses. This approach provides for a system consistent solution for multivariate linear autoregressive models, with time varying second moments.

6.2 Introduction and Literature

The response of possibly non-linear econometric structures to linear and non-linear shocks has been addressed in a number of papers, most notably Koop et. al (1996) [171] and Gallant et. al (1993) [103]. The first study develops the concept and derive rigorously what they term as 'generalised impulse response' function, They suggest that such a concept can deal adequately with a series of problems that arise when one departs from the calculation of the responses of the simple univariate linear model. One of the most attractive features of this methodology is that the calculation of the impulse response within a multivariate linear system is independent of the order that the variables enter the structure. In a later paper Gourieroux and Jasiak (1999) [114] by extracting the nonlinear residuals as proxies for the non-linear innovations develop a set of test statistics that can be used in testing the specification of the underlying structure.

In an earlier study Gallant et al (1993) provided the 'basic' methodology for computing the impulse response function for non-linear time series by computing the differences between what they term as 'the baseline approach' to the conditional moment profile using a semi-parametric approach. Among the different examples that this methodology can be used is the return volatility-trading volume relationship that it is postulated in term of the second moments of the returns and the first moments in volume.

The growing popularity of the multivariate GARCH models has led to the requirement of calculating the impulse response analysis of the conditional volatility after a shock. Lin (1997) [183] derives and analyses such measure in the context of a multivariate GARCH model and evaluates the small sample properties of the standard errors that surround the IRF by means of a Monte Carlo study for a MV-GARCH(1,1). The purpose of this study is to provide an analytic framework for the computation of impulse response function for a multivariate linear model, that is accompanied by MV-GARCH set of equations that describe the conditional volatility of each/some of its elements. Unlike previous work this chapter does not delve into the computation of impulse response functions of non-linear models, as the systems which this chapter deals with are essentially linear structures that are related in both first (VAR) and second moments (MV-GARCH).

The macroeconomics literature has embraced the general r^{th} order vector auto-regression models as a fundamental tool for the analysis of complex multivariate systems. The use of impulse response functions (IRFs) to graphically illustrate the interactions of sets of dynamic simultaneous equations models is a similarly well developed technique. By contrast MV-GARCH models have not found nearly as much popularity in the literature as their univariate precursors. Impulse response functions and MV-GARCH models have an even more limited history, the VAR-GARCH model of Sin (2005) [239] Polasek and Ren (1998) [219] and (2000) [220] illustrates a mechanism

for implementing an impulse response function in variance for a general VAR model. Elder (2003) offers an extremely limited description of a proposed algorithm for defining impulse responses in variance. Hafner and Herwartz (2001) [128] demonstrate a model of impulse response in variance and apply it to exchange rates utilizing the responses as a guide to system identification. However none of the current derivations of impulse responses in variance incorporate the dynamics of a fully specified MV-GARCH model.

This paper attempts to specify the definitive mechanism that should be used to construct a generalized impulse response function for a given VAR/GARCH combination. We show that the effect of innovations in variance/covariance should be computed in conjunction with the underlying model in means and that by solving a matrix polynomial function the complex interaction between the first and second moments of a dynamic model are understood in a tractable manner. The paper is organized as follows, in the following section develops the mean and volatility models and explore their interactions in the presence of random shocks. The third section provides a simulated bivariate example and demonstrates the difference in the impulse response functions when the interaction between the mean and volatility are taken into account and contrast it to the usual case where despite MV-ARCH effects the impulse response functions ignore it. Our brief conclusions are provided in the final section.

6.3 The General Model

Consider the general linear vector auto-regression model that may include a vector of forcing variables: where $\mathbf{y}_t$ is an n length column vector process in $\mathbb{R}^n$ with discrete time increments t, where $t \in [1, 2, ..., T]$, $\mathbf{x}$ is an m length column vector process of exogenous driving variables, $\mathbf{u}_t$ is a conditionally multi-normal disturbance term, $\mathbf{\Pi}_0$ and $\mathbf{\Pi}_1$ are parameter matrices, of $n \times nr$ and $n \times m$ dimensions respectively.

$$\mathbf{y}_t = \mathbf{\Pi}_0^{\mathrm{T}} vec\left(\mathcal{Y}_t\right) + \mathbf{\Pi}_1^{\mathrm{T}} \mathbf{x}_t + \mathbf{u}_t \tag{6.3.1}$$

The past evolution of $\mathbf{y}_t$ is contained in the array $\mathcal{Y}_t$, thus

$$\mathcal{Y}_t = [\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-r}] \tag{6.3.2}$$

The conditional expectations of the first two moments of the disturbance terms are as follows,

$$E\left(\mathbf{u}_t\right) = \mathbf{0} \tag{6.3.3}$$

$$E\left(\mathbf{u}_t \mathbf{u}_t^{\mathrm{T}}\right) = \mathbf{\Sigma}_t \tag{6.3.4}$$

Given the assumption of conditional normality the distribution of the disturbances is as follows,

$$\mathbf{u}_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_t\right) \tag{6.3.5}$$

The matrix process that describes the evolution of the conditional covariance matrix is described by a modification of the Ding and Engle (2002) MARCH model.

$$\mathbf{\Sigma}_t = \bar{\mathbf{\Sigma}} + \sum_{j=1}^{p} \mathbf{A}_j \times \mathbf{u}_{t-j}\mathbf{u}_{t-j}^{\mathrm{T}} + \sum_{i=1}^{q} \mathbf{B}_i \times \mathbf{\Sigma}_{t-i} \tag{6.3.6}$$

the use of $\times$ in the matrix equation represents the Hadamard product (element by element) between two identically sized arrays. The following stationarity conditions are applied to the VAR model,

$$\mathbf{F} = \begin{bmatrix} \underset{nr\times n}{\mathbf{\Pi}_0^{\mathrm{T}}} \\ \underset{n(r-1)\times n(r-1)}{\mathbf{I}} \quad \underset{n(r-1)\times n}{\mathbf{0}} \end{bmatrix} \tag{6.3.7}$$

For stationarity the eigenvalues of $\mathbf{F}$ must lie within the unit circle. For the MV-GARCH specification, the stationarity conditions are as follows,

$$\mathbf{a} = vec(\mathcal{A}) \tag{6.3.8}$$

$$\mathbf{b} = vec(\mathcal{B}) \tag{6.3.9}$$

$$0 < \sum_{i=1}^{pn^2} a_i + \sum_{j=1}^{qn^2} b_j \leq 1 \tag{6.3.10}$$

Consider the conditional predictor of $\mathbf{y}_t$,

$$\hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}} = \mathbf{\Pi}_0 vec(\mathcal{Y}_t) + \mathbf{\Pi}_1 \mathbf{x}_t \tag{6.3.11}$$

The forecast error is therefore,

$$\mathbf{y}_t - \hat{\mathbf{y}}_t = \mathbf{u}_t \tag{6.3.12}$$

and by assumption,

$$\begin{aligned} E\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}}\right) &= E\left(\mathbf{u}_t\right) \\ &= \mathbf{0} \end{aligned} \tag{6.3.13}$$

and,

$$E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}}\right)^{\mathrm{T}}\right) = \mathbf{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}} \tag{6.3.14}$$

Consider $\mathbf{\Sigma}_t^{\frac{1}{2}}$ as the upper triangular Cholesky factorization, of the non-negative Hermitian matrix $\mathbf{\Sigma}_t$ and a second zero mean n length vector disturbance process, ε consisting of n standard normal i.i.d. elements, then the forecast errors can be expressed as,

$$\mathbf{u}_t = \mathbf{\Sigma}_t^{\frac{1}{2}} \varepsilon_t \tag{6.3.15}$$

and,

$$\varepsilon_t \sim N\left(\mathbf{0}, \mathbf{I}\right) \tag{6.3.16}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix. Substituting (6.3.15) into (6.3.6) obtaining,

$$\mathbf{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}} = \bar{\mathbf{\Sigma}} + \sum_{j=1}^{p} \mathbf{A}_j \times \left(\mathbf{\Sigma}_{t-j}^{\frac{1}{2}} \varepsilon_{t-j}\right)\left(\mathbf{\Sigma}_{t-j}^{\frac{1}{2}} \varepsilon_{t-j}\right)^{\mathrm{T}} + \sum_{i=1}^{q} \mathbf{B}_i \times \mathbf{\Sigma}_{t-i} \tag{6.3.17}$$

Therefore, VAR structure is updated as follows,

$$\mathbf{y}_t = \hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}} + \mathbf{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}}^{\frac{1}{2}} \varepsilon_t \tag{6.3.18}$$

and the forecast errors maybe written as,

$$\begin{aligned} E\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|\{t-1,\ldots,t-r\}}\right) &= \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|\{t-1,\ldots,t-p,t-q\}} E\left(\varepsilon_t\right) \\ &= \mathbf{0} \end{aligned} \tag{6.3.19}$$

From this basic structure the innovation of shocks in mean and variance maybe computed.

6.3.1 Impulse Response Functions

The s step ahead impulse response is given from the following vector of partial derivatives,

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{u}_t} = \boldsymbol{\Psi}^{\mu}_{s} \tag{6.3.20}$$

Substituting the conditional covariance representation yields,

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \left(\boldsymbol{\Sigma}^{\frac{1}{2}}_{t|\{t-1,\ldots,t-p,t-q\}} \varepsilon_t\right)^{\mathrm{T}}} = \boldsymbol{\Psi}^{\mu}_{s} \tag{6.3.21}$$

Using the factorized covariance matrix,

$$\begin{aligned} \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|\{t-1,\ldots,t-p,t-q\}} &\equiv \boldsymbol{\Lambda}_t \\ \boldsymbol{\Lambda}_t &= \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,n} \\ 0 & \lambda_{2,2} & \cdots & \lambda_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_{n,n} \end{bmatrix}_t \\ \lambda_{i,t} &= [\lambda_{i,1}, \ldots, \lambda_{i,n}]^{\mathrm{T}}_t \end{aligned}$$

We can now express the set of partial derivatives as,

$$\Delta \mathbf{y}_{t+s} = \frac{\partial \mathbf{y}_{t+s}}{\partial \lambda_{1,t}^{\mathrm{T}} \varepsilon_t} \delta_1 + \frac{\partial \mathbf{y}_{t+s}}{\partial \lambda_{2,t}^{\mathrm{T}} \varepsilon_t} \delta_2 + ... + \frac{\partial \mathbf{y}_{t+s}}{\partial \lambda_{n,t}^{\mathrm{T}} \varepsilon_t} \delta_n \tag{6.3.22}$$

where the unit shifts are,

$$\delta = [\delta_1, \delta_2 ..., \delta_n]^{\mathrm{T}} \tag{6.3.23}$$

the above specification defines the impact of an innovation in mean, whilst taking into account the covariance structure of the system, however it is not sufficient to integrate both the responses of the mean and volatility equations as it does provide us with an explicit response of the volatility system to an innovation. We now turn our attention in formulating the impulse response functions in variance to a single unit shock in ε.

6.3.2 Impulse Responses in Variance and Covariance

We define the response in variance to unit shock in ε as a matrix process instead of the usual vectorized process.

$$\frac{\partial \left(\left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|\{t-1,...,t-r\}} \right) \left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|\{t-1,...,t-r\}} \right)^{\mathrm{T}} \right)}{\partial \left(\left(\boldsymbol{\Sigma}_{t|\{t-1,...,t-p,t-q\}}^{\frac{1}{2}} \varepsilon_t \right) \left(\boldsymbol{\Sigma}_{t|\{t-1,...,t-p,t-q\}}^{\frac{1}{2}} \varepsilon_t \right)^{\mathrm{T}} \right)} = \boldsymbol{\Psi}_s^{\sigma} \tag{6.3.24}$$

setting,

$$\begin{aligned} \left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|\{t-1,...,t-r\}} \right) \left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s|\{t-1,...,t-r\}} \right)^{\mathrm{T}} &\equiv \mathbf{F} \left(\boldsymbol{\Xi}_t \right)_{t+s} \\ \mathbf{F} \left(\boldsymbol{\Xi}_t \right)_{t+s} &= \left[f_{i,j} \left(\boldsymbol{\Xi}_t \right) \right]_{t+s} \end{aligned} \tag{6.3.25}$$

and

$$\left(\boldsymbol{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}}^{\frac{1}{2}}\varepsilon_t\right)\left(\boldsymbol{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}}^{\frac{1}{2}}\varepsilon_t\right)^{\mathrm{T}} = \boldsymbol{\Xi}_t \tag{6.3.26}$$

$$\boldsymbol{\Xi}_t = \left[\xi_{i,j}\right]_t \tag{6.3.27}$$

To derive the analytical solution for 6.3.24 the derivatives for each of the innovations of the covariance matrix with respect to a matrix of shocks with respect to 6.3.23 needs to be computed.

$$\frac{\partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s}}{\partial \xi_{i,j,t}} = \begin{bmatrix} \partial f_{1,1}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{i,j,t} & \cdots & \partial f_{1,n}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{i,j,t} \\ \vdots & \ddots & \vdots \\ \partial f_{n,1}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{i,j,t} & \cdots & \partial f_{n,n}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{i,j,t} \end{bmatrix} \tag{6.3.28}$$

This matrix forms the individual partitions for a larger matrix of the following set of partial derivatives, for 6.3.29,

$$\frac{\partial\left(\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s|\{t-1,\ldots,t-r\}}\right)\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s|\{t-1,\ldots,t-r\}}\right)^{\mathrm{T}}\right)}{\partial\left(\left(\boldsymbol{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}}^{\frac{1}{2}}\varepsilon_t\right)\left(\boldsymbol{\Sigma}_{t|\{t-1,\ldots,t-p,t-q\}}^{\frac{1}{2}}\varepsilon_t\right)^{\mathrm{T}}\right)} = \frac{\partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s}}{\partial \boldsymbol{\Xi}_t} \tag{6.3.29}$$

and the matrix is therefore,

$$\frac{\partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s}}{\partial \boldsymbol{\Xi}_t} = \begin{bmatrix} \partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{1,1,t} & \cdots & \partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{1,q,t} \\ \vdots & \ddots & \vdots \\ \partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{n,1,t} & \cdots & \partial \mathbf{F}\left(\boldsymbol{\Xi}_t\right)/\partial \xi_{n,n,t} \end{bmatrix} \tag{6.3.30}$$

defining,

$$\mathfrak{D}\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s|\{t-1,\ldots,t-r\}}\right)\left(\mathbf{y}_{t+s}-\hat{\mathbf{y}}_{t+s|\{t-1,\ldots,t-r\}}\right)^{\mathrm{T}} = \mathfrak{D}\mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s} \tag{6.3.31}$$

Where $\mathfrak{D}$ denotes the matrix differential (Jacobian) operator. Which is then expressed more compactly as,

$$\mathfrak{D}\mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s} = \frac{\partial vec\mathbf{F}\left(\boldsymbol{\Xi}_t\right)_{t+s}}{\partial\left(vec\boldsymbol{\Xi}_t\right)^{\mathrm{T}}} = \boldsymbol{\Psi}_s^{\sigma} \tag{6.3.32}$$

For a matrix of unit shocks in the quadratic form,

$$\mathbf{D} = \delta\delta^{\mathrm{T}} \tag{6.3.33}$$

$$\mathbf{D} = [\delta_i \delta_j] \tag{6.3.34}$$

the s step ahead impulse responses are computed as,

$$\begin{aligned}\left[\Delta\left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s}\right)\left(\mathbf{y}_{t+s} - \hat{\mathbf{y}}_{t+s}\right)^{\mathrm{T}}\right]_{i,j,t} = {} & \delta_1^2\left(\frac{\partial \mathbf{F}\left(\Xi_t\right)}{\partial \Xi_t}\right) + \\ & \ldots + \delta_1\delta_n\left(\frac{\partial \mathbf{F}\left(\Xi_t\right)}{\partial \Xi_t}\right) + \\ & \ldots + \delta_n\delta_1\left(\frac{\partial \mathbf{F}\left(\Xi_t\right)}{\partial \Xi_t}\right) + \\ & \ldots + \delta_n^2\left(\frac{\partial \mathbf{F}\left(\Xi_t\right)}{\partial \Xi_t}\right)\end{aligned} \tag{6.3.35}$$

This set of equations fully describe the impact of an innovation on both the mean and volatility systems. More explicitly the eigenvalues of each partition of 6.3.30, correspond to the adjusted elements in 6.3.22. For any given estimated model structure, of particular interest will be the responses to single unit shock in the second moments of model. This maybe of particular interest to models involving the transmission of volatility from shifts in macroeconomic variables to financial markets and the cross contagion transmission of volatility between international financial markets.

6.4 Theoretical Example

The following example is for a bivariate VAR(1), MV-GARCH(1,1) model with 2 driving variables. The generic model is therefore,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \mathbf{\Pi}_0^{\mathrm{T}} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \mathbf{\Pi}_1^{\mathrm{T}} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{6.4.1}$$

expanding the coefficients matrices forms,

$$\begin{aligned} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} &= \begin{bmatrix} \pi_{1,1} y_{1,t-1} + \pi_{1,2} y_{2,t-1} \\ \pi_{2,1} y_{1,t-1} + \pi_{2,2} y_{2,t-1} \end{bmatrix}_0 + \begin{bmatrix} \pi_{1,1} x_{1,t} + \pi_{1,2} x_{2,t} \\ \pi_{2,1} x_{1,t} + \pi_{2,2} x_{2,t} \end{bmatrix}_1 + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \\ &= \begin{bmatrix} \hat{y}_{1,t} \\ \hat{y}_{2,t} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \end{aligned} \tag{6.4.2}$$

this gives the full form of the model,

$$\begin{bmatrix} \hat{y}_{1,t} \\ \hat{y}_{2,t} \end{bmatrix} = \begin{bmatrix} \pi_{1,1} y_{1,t-1} + \pi_{1,2} y_{2,t-1} \\ \pi_{2,1} y_{1,t-1} + \pi_{2,2} y_{2,t-1} \end{bmatrix}_0 + \begin{bmatrix} \pi_{1,1} x_{1,t} + \pi_{1,2} x_{2,t} \\ \pi_{2,1} x_{1,t} + \pi_{2,2} x_{2,t} \end{bmatrix}_1 \tag{6.4.3}$$

setting up the recursive expectations structure,

$$E\left(\begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6.4.4}$$

$$E\left(\begin{bmatrix} u_{1,t}^2 & u_{1,t}u_{2,t} \\ u_{2,t}u_{1,t} & u_{2,t}^2 \end{bmatrix}\right) = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{1,2,t} \\ \sigma_{2,1,t} & \sigma_{2,t}^2 \end{bmatrix} \tag{6.4.5}$$

Defining,

$$\begin{aligned} \mathbf{k} &= [k_1, k_2, k_3]^{\mathrm{T}} \\ \mathbf{K} &= \begin{bmatrix} k_1 & 0 \\ k_2 & k_3 \end{bmatrix} \end{aligned} \tag{6.4.6}$$

The unconditional covariance matrix is thus defined as,

$$\bar{\mathbf{\Sigma}} = \mathbf{K}\mathbf{K}^{\mathrm{T}} \tag{6.4.7}$$

which subsequently yields,

$$\begin{aligned}\mathbf{\Sigma}_t &= \begin{bmatrix} k_1^2 & k_1k_2 \\ k_2k_1 & k_1^2 + k_2^2 \end{bmatrix} + \begin{bmatrix} a_{1,1}u_{1,t-1}^2 & a_{1,2}u_{1,t-1}u_{2,t-1} \\ a_{2,1}u_{2,t-1}u_{1,t-1} & a_{2,2}u_{2,t-1}^2 \end{bmatrix} \\ &+ \begin{bmatrix} b_{1,1}\sigma_{1,t-1}^2 & b_{1,2}\sigma_{1,2,t-1} \\ b_{2,1}\sigma_{2,1,t-1} & b_{2,2}\sigma_{2,t-1}^2 \end{bmatrix} \end{aligned} \tag{6.4.8}$$

$$\tag{6.4.9}$$

therefore the elements of the conditional covariance matrix are as follows,

$$\mathbf{\Sigma}_t = \begin{bmatrix} k_1^2 + a_{1,1}u_{1,t-1}^2 + b_{1,1}\sigma_{1,t-1}^2 & k_1k_2 + a_{1,2}u_{1,t-1}u_{2,t-1} + b_{1,2}\sigma_{1,2,t-1} \\ k_1k_2 + a_{1,2}u_{1,t-1}u_{2,t-1} + b_{1,2}\sigma_{1,2,t-1} & k_1^2 + k_2^2 + a_{2,2}u_{2,t-1}^2 + b_{2,2}\sigma_{2,t-1}^2 \end{bmatrix} \tag{6.4.10}$$

defining that $\mathbf{\Sigma}_t^{\frac{1}{2}} = \mathbf{\Lambda}_t$, using the Cholesky-Banachiewicz algorithm, the elements of $\mathbf{\Lambda}_t$ maybe found in terms of the recursive information of the system. See appendix 1. And similarly analytical expressions for the disturbance terms $\mathbf{u}_t$ as the previous elements of $\mathbf{\Lambda}_t$ and the i.i.d. standard normal vector process ε as in 6.3.26. We now have an algebraic expression for the covariance matrix and its factorized counterpart, therefore computing the Jacobian matrix is a tractable process. For the forward recursion in mean we compute the following,

$$\Delta\mathbf{y}_{t+s} = \frac{\partial\mathbf{y}_{t+s}}{\partial\lambda_{1,t}^{\mathrm{T}}\varepsilon_t}\delta_1 + \frac{\partial\mathbf{y}_{t+s}}{\partial\lambda_{2,t}^{\mathrm{T}}\varepsilon_t}\delta_2 \tag{6.4.11}$$

where the impulses are as follows,

$$\delta = [\delta_1, \delta_2]^{\mathrm{T}} \tag{6.4.12}$$

For the variance process we simply need to partition the Jacobian and sum its components to produce the response structure,

$$\left[\Delta\left(\begin{bmatrix} y_{1,t+s} \\ y_{2,t+s} \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,t+s} \\ \hat{y}_{2,t+s} \end{bmatrix}\right)\left(\begin{bmatrix} y_{1,t+s} \\ y_{2,t+s} \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,t+s} \\ \hat{y}_{2,t+s} \end{bmatrix}\right)^{\mathrm{T}}\right]_{i,j,t} = \delta_1^2\left(\frac{\partial \mathbf{F}(\Xi_t)}{\partial \Xi_t}\right) + \delta_1\delta_2\left(\frac{\partial \mathbf{F}(\Xi_t)}{\partial \Xi_t}\right) + \delta_1\delta_2\left(\frac{\partial \mathbf{F}(\Xi_t)}{\partial \Xi_t}\right) + \delta_2^2\left(\frac{\partial \mathbf{F}(\Xi_t)}{\partial \Xi_t}\right)$$

Each element of the matrix above represents the reaction of individual elements in the covariance matrix given the nature of innovation of shocks into the model. For example given an innovation in the first equation, only the first element of this matrix will be non-zero, the remaining three are therefore set to zero. We now proceed to provide the s-step ahead impulse responses from the above model under two alternative specifications. In this first specification the full interaction between the mean and volatility equations is permitted whilst in the second the interaction is suppressed.

6.5 Numerical Simulation

The bivariate example is parameterized with a set of randomly drawn coefficients and then compute the impulse responses for a given set of exogenously applied shocks. 6.4 and 6.3 present the impulse response functions for both the means and second moments when the MV-ARCH effects have been taken into account. In graph 3 we present the imputed IRF of the correlation coefficient following a unit shock in

the first 'mean' equation. In 6.1 we plot the IRF for both 'mean' equations for the same shock when the MV-ARCH effects are suppressed. The usually reported impulse response function, converge to equilibrium rapidly, given our assumptions of stationarity, within ten periods the shock has been fully dissipated in the structure. However when the MV-GARCH effects are taken into account the persistence of the same shock is far more pronounced. The I.R.F. becomes monotonic after 15 periods at the same time, there are pronounced changes in all the elements of the covariance matrix, that persist. The imputed correlation coefficient increases rapidly after the shock and then declines monotonically. In the context of a two asset portfolio model, the above result shows that following a once and for all shock in the expected returns of one asset the whole portfolio risk changes. If sets of asset returns exhibit MV-GARCH behavior despite the stationarity of the system, our model suggests dynamic re-balancing of the portfolio is required for the appropriate hedging of risk. In the context of macro-economic models that use VAR specifications, if the disturbances of one of the equations exhibits ARCH/GARCH effects, then a once and for all shock in this equation can propagate into the other equations and lengthen the number of forward steps that the model takes to return to the long run equilibrium.

6.6 Concluding Remarks

We have formulated in detail a methodology for constructing discrete impulse response functions in mean, variance and covariance for a multi-equation model exhibiting autoregressive conditional heteroscedasticity in its residuals. This model has many attractive properties and treats the modelling of the first and second moments in a unified and comprehensive manner. When the MV-GARCH effects are not taken into account in the context of a VAR model the usually reported impulse response functions are distortions of the true adjustment path and in the context of our simulations severely under-predict both the magnitude of the disturbance to the system and the number of periods required for return to the long run equilibrium. Our results have relevance for both portfolio construction and macroeconomic policy evaluation. Our methodology suggests that even in a fully stationary environment dynamic portfolio re-balancing is required for controlling for the level of risk. Macroeconomic policy evaluation that ignores possible MV-ARCH effects will underestimate the size and duration of the impact of policy shocks on the economy.

6.7 Chapter Appendix I

The bivariate VAR-GARCH(1,1) conditional covariance matrix, as function of the impulses, ε_{t-1} at the pervious innovation, is as follows, the variance of the disturbance

of equation 1,

$$\sigma_{1,t}^2 = \left({a_{1,1}}^2 \left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2 + 2\,a_{1,1}\,a_{1,2}\,\varepsilon_{2,t-1} \right.$$
$$\sqrt{\sigma_{1,t-1}}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\sqrt{-\left(\frac{{\sigma_{1,2,t-1}}^2}{\sigma_{1,t-1}}\right) + \sigma_{2,t-1}} +$$
$$\sigma_{1,t-1}\,\left({k_1}^2 + 2\,b_{1,1}\,b_{1,2}\,\sigma_{1,2,t-1} + {b_{1,1}}^2\,\sigma_{1,t-1} + {b_{1,2}}^2\,\sigma_{2,t-1}\right) +$$
$$\left. {a_{1,2}}^2\,{\varepsilon_{2,t-1}}^2\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right) \right) \left(\sigma_{1,t-1}\right)^{-1}$$

The covariance of the disturbances is therefore,

$$\sigma_{1,2,t} = \left(\sigma_{1,t-1}\,\left(k_1\,k_2 + b_{1,2}\,b_{2,1}\,\sigma_{1,2,t-1} + b_{1,1}\,b_{2,2}\,\sigma_{1,2,t-1} + b_{1,1}\,b_{2,1}\,\sigma_{1,t-1} + b_{1,2}\,b_{2,2}\,\sigma_{2,t-1}\right) + \right.$$
$$a_{1,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\left(a_{2,1}\,\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + a_{2,1}\,\varepsilon_{1,t-1}\,\sigma_{1,t-1} + \right.$$
$$\left. a_{2,2}\,\varepsilon_{2,t-1}\,\sqrt{\sigma_{1,t-1}}\,\sqrt{-\left(\frac{{\sigma_{1,2,t-1}}^2}{\sigma_{1,t-1}}\right) + \sigma_{2,t-1}}\right) + a_{1,2}\,\varepsilon_{2,t-1}$$
$$\left(a_{2,1}\,\sqrt{\sigma_{1,t-1}}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\sqrt{-\left(\frac{{\sigma_{1,2,t-1}}^2}{\sigma_{1,t-1}}\right) + \sigma_{2,t-1}} + \right.$$
$$\left. \left. a_{2,2}\,\varepsilon_{2,t-1}\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right)\right)\right)\left(\sigma_{1,t-1}\right)^{-1}$$

The variance of the disturbance of the second equation,

$$\sigma_{1,t}^2 = \left({a_{2,1}}^2\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2 + 2\,a_{2,1}\,a_{2,2}\,\varepsilon_{2,t-1}\,\sqrt{\sigma_{1,t-1}} \right.$$
$$\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\sqrt{-\left(\frac{{\sigma_{1,2,t-1}}^2}{\sigma_{1,t-1}}\right)}$$
$$+\sigma_{2,t-1} + \sigma_{1,t-1}\,\left({k_2}^2 + {k_3}^2 + 2\,b_{2,1}\,b_{2,2}\,\sigma_{1,2,t-1} + {b_{2,1}}^2\,\sigma_{1,t-1} + {b_{2,2}}^2\,\sigma_{2,t-1}\right)$$
$$\left. +{a_{2,2}}^2\,{\varepsilon_{2,t-1}}^2\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right) \right) \left(\sigma_{1,t-1}\right)^{-1}$$

Using the Cholesky-Banachiewicz algorithm the covariance matrix is factorized around pivot points along each diagonal and upper diagonal element, the three non-zero elements of the factorized conditional Covariance matrix $\mathbf{\Sigma}_{t-1}^{\frac{1}{2}} = \mathbf{\Lambda}$, consist of the first pivot element,

$$\begin{aligned}\lambda_{1,1,t} \;=\; & \Big({a_{1,1}}^2 \left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2 + \\ & \sigma_{1,t-1}\,\left({k_1}^2 + 2\,b_{1,1}\,b_{1,2}\,\sigma_{1,2,t-1} + {b_{1,1}}^2\,\sigma_{1,t-1} + {b_{1,2}}^2\,\sigma_{2,t-1}\right) + \\ & 2\,a_{1,1}\,a_{1,2}\,\varepsilon_{2,t-1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}} + \\ & {a_{1,2}}^2\,{\varepsilon_{2,t-1}}^2\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right)\Big)\left(\sigma_{1,t-1}\right)^{-1}\end{aligned}$$

The lower diagonal element is zero,

$$\lambda_{2,1,t} = 0$$

The second pivot element factorizes to,

$$
\begin{aligned}
\lambda_{1,2,t} \;=\; & \left(\left(\sigma_{1,t-1}\,\left(k_1\,k_2 + b_{1,2}\,b_{2,1}\,\sigma_{1,2,t-1} + b_{1,1}\,b_{2,2}\,\sigma_{1,2,t-1} + b_{1,1}\,b_{2,1}\,\sigma_{1,t-1} + b_{1,2}\,b_{2,2}\,\sigma_{2,t-1}\right) + \right.\right. \\
& a_{1,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right) \\
& \left(a_{2,1}\,\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + a_{2,1}\,\varepsilon_{1,t-1}\,\sigma_{1,t-1} + a_{2,2}\,\varepsilon_{2,t-1}\,\sqrt{-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}}\right) + \\
& a_{1,2}\,\varepsilon_{2,t-1}\,\left(a_{2,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}} + \right. \\
& \left.\left. a_{2,2}\,\varepsilon_{2,t-1}\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right)\right)\right) \\
& \left(\left(\sigma_{1,t-1}\,\left({a_{1,1}}^2\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2 + \right.\right.\right. \\
& \sigma_{1,t-1}\,\left({k_1}^2 + 2\,b_{1,1}\,b_{1,2}\,\sigma_{1,2,t-1} + {b_{1,1}}^2\,\sigma_{1,t-1} + {b_{1,2}}^2\,\sigma_{2,t-1}\right) + \\
& 2\,a_{1,1}\,a_{1,2}\,\varepsilon_{2,t-1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1} + \varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}} + \\
& \left.\left.\left.{a_{1,2}}^2\,{\varepsilon_{2,t-1}}^2\,\left(-{\sigma_{1,2,t-1}}^2 + \sigma_{1,t-1}\,\sigma_{2,t-1}\right)\right)\right)^{-\frac{1}{2}}\right)^{\frac{1}{2}}
\end{aligned}
$$

The final pivot element of $\mathbf{\Lambda}$, is

$$
\begin{aligned}
\lambda_{2,2,t} \;=\; & \left(a_{2,1}{}^2 \left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2+\right.\\
& \sigma_{1,t-1}\,\left(k_2{}^2+k_3{}^2+2\,b_{2,1}\,b_{2,2}\,\sigma_{1,2,t-1}+b_{2,1}{}^2\,\sigma_{1,t-1}+b_{2,2}{}^2\,\sigma_{2,t-1}\right)+\\
& 2\,a_{2,1}\,a_{2,2}\,\varepsilon_{2,t-1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}}+\\
& a_{2,2}{}^2\,\varepsilon_{2,t-1}{}^2\,\left(-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}\right)-\\
& \left(\left(\sigma_{1,t-1}\,\left(k_1\,k_2+b_{1,2}\,b_{2,1}\,\sigma_{1,2,t-1}+b_{1,1}\,b_{2,2}\,\sigma_{1,2,t-1}+b_{1,1}\,b_{2,1}\,\sigma_{1,t-1}+b_{1,2}\,b_{2,2}\,\sigma_{2,t-1}\right)+\right.\right.\\
& a_{1,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\\
& \left(a_{2,1}\,\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+a_{2,1}\,\varepsilon_{1,t-1}\,\sigma_{1,t-1}+a_{2,2}\,\varepsilon_{2,t-1}\,\sqrt{-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}}\right)+\\
& a_{1,2}\,\varepsilon_{2,t-1}\,\left(a_{2,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}}+\right.\\
& \left.\left.\left.a_{2,2}\,\varepsilon_{2,t-1}\,\left(-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}\right)\right)\right)\,\left(-\left(a_{1,2}\,a_{2,2}\,\varepsilon_{2,t-1}{}^2\,\sigma_{1,2,t-1}{}^2\right)+k_1\,k_2\,\sigma_{1,t-1}+\right.\\
& b_{1,2}\,b_{2,1}\,\sigma_{1,2,t-1}\,\sigma_{1,t-1}+b_{1,1}\,b_{2,2}\,\sigma_{1,2,t-1}\,\sigma_{1,t-1}+b_{1,1}\,b_{2,1}\,\sigma_{1,t-1}{}^2+\\
& a_{1,1}\,a_{2,1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2+b_{1,2}\,b_{2,2}\,\sigma_{1,t-1}\,\sigma_{2,t-1}+\\
& a_{1,2}\,a_{2,2}\,\varepsilon_{2,t-1}{}^2\,\sigma_{1,t-1}\,\sigma_{2,t-1}+\\
& \left(a_{1,2}\,a_{2,1}+a_{1,1}\,a_{2,2}\right)\,\varepsilon_{2,t-1}\,\sqrt{\sigma_{1,t-1}}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-\left(\frac{\sigma_{1,2,t-1}{}^2}{\sigma_{1,t-1}}\right)+\sigma_{2,t-1}}\\
& \left(\left(a_{1,1}{}^2\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2+\right.\right.\\
& \sigma_{1,t-1}\,\left(k_1{}^2+2\,b_{1,1}\,b_{1,2}\,\sigma_{1,2,t-1}+b_{1,1}{}^2\,\sigma_{1,t-1}+b_{1,2}{}^2\,\sigma_{2,t-1}\right)+\\
& 2\,a_{1,1}\,a_{1,2}\,\varepsilon_{2,t-1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}}+\\
& \left.\left.a_{1,2}{}^2\,\varepsilon_{2,t-1}{}^2\,\left(-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}\right)\right)\,\left(\sigma_{1,t-1}\right)^{-1}\right)^{\frac{1}{2}}\Bigg)^{-1}\Bigg)\\
& \left(\sigma_{1,t-1}\,\left(a_{1,1}{}^2\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)^2+\right.\right.\\
& \sigma_{1,t-1}\,\left(k_1{}^2+2\,b_{1,1}\,b_{1,2}\,\sigma_{1,2,t-1}+b_{1,1}{}^2\,\sigma_{1,t-1}+b_{1,2}{}^2\,\sigma_{2,t-1}\right)+\\
& 2\,a_{1,1}\,a_{1,2}\,\varepsilon_{2,t-1}\,\left(\varepsilon_{2,t-1}\,\sigma_{1,2,t-1}+\varepsilon_{1,t-1}\,\sigma_{1,t-1}\right)\,\sqrt{-\sigma_{1,2,t-1}{}^2+\sigma_{1,t-1}\,\sigma_{2,t-1}}+
\end{aligned}
$$

6.8 Appendix II

Practical implementation of a comparison model between impulse responses with MV-ARCH effects and one with a time invariant covariance matrix, in this respect the deviation from equilibrium does not need to be fully explained, only the deviation each model specification results in, given a single unit shock,

$$\mathbf{y}_t = \hat{\mathbf{y}}_t + \mathbf{u}_t \tag{6.8.1}$$

$$(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{u}_t \tag{6.8.2}$$

$$E(\mathbf{y}_t - \hat{\mathbf{y}}_t) = \mathbf{0} \tag{6.8.3}$$

$$E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}}\right) = \mathbf{\Sigma}_t \tag{6.8.4}$$

now construct two equations, one with a dynamic covariance matrix $\mathbf{\Sigma}_t$ and one with a robustly estimated time invariant covariance matrix, $\mathbf{\Sigma}$,

$$E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}}\right) - \mathbf{\Sigma}_t = \mathbf{0} \tag{6.8.5}$$

$$E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}}\right) - \mathbf{\Sigma} = \mathbf{0} \tag{6.8.6}$$

The impulse responses may be characterized as the deviation in response from the equilibrium model, given a vector of shocks δ,

$$\mathbf{y}_t^{\boldsymbol{\Sigma}_t} = \mathbf{y}_t \left| \left\{ E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \right) - \boldsymbol{\Sigma}_t = \mathbf{0} \right\} \right. \tag{6.8.7}$$

$$\mathbf{y}_t^{\boldsymbol{\Sigma}} = \mathbf{y}_t \left| \left\{ E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \right) - \boldsymbol{\Sigma} = \mathbf{0} \right\} \right. \tag{6.8.8}$$

$$\psi_t = \mathbf{y}_t^{\boldsymbol{\Sigma}_t} - \mathbf{y}_t^{\bar{\boldsymbol{\Sigma}}} \tag{6.8.9}$$

$$\boldsymbol{\Psi}_{s|\delta} = \left[\psi_{t+1|\delta}, ..., \psi_{t+s|\delta} \right] \tag{6.8.10}$$

Elimination of $E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \right)$, by subtraction, yields, the differences in the unconditional and conditional covariance matrices,

$$\begin{aligned} \boldsymbol{\Phi}_t &= \left(E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \right) - \boldsymbol{\Sigma}_t \right) - \left(E\left((\mathbf{y}_t - \hat{\mathbf{y}}_t)(\mathbf{y}_t - \hat{\mathbf{y}}_t)^{\mathrm{T}} \right) - \boldsymbol{\Sigma} \right) \\ &= \boldsymbol{\Sigma}_t - \boldsymbol{\Sigma} \end{aligned}$$

(6.8.9)

The eigenvalues of this matrix are anticipated deviations in mean between the MV-ARCH and the time invariant forms of the VAR. For the bivariate example, results in the following recursion,

6.9 Figures 1 - 4

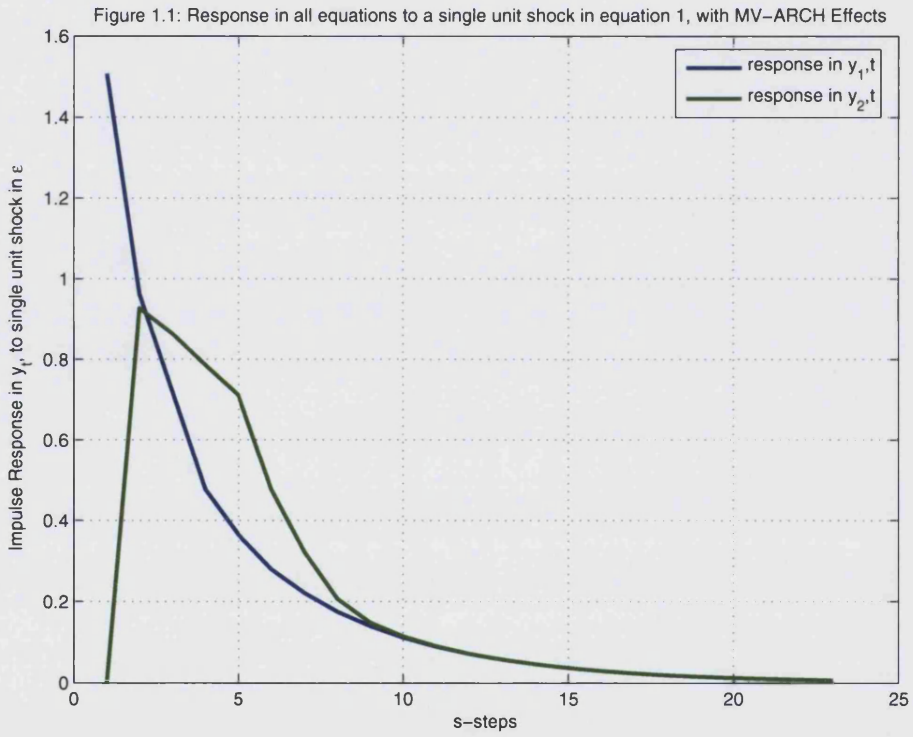

Figure 6.1: The mean equation with ARCH effects suppressed.

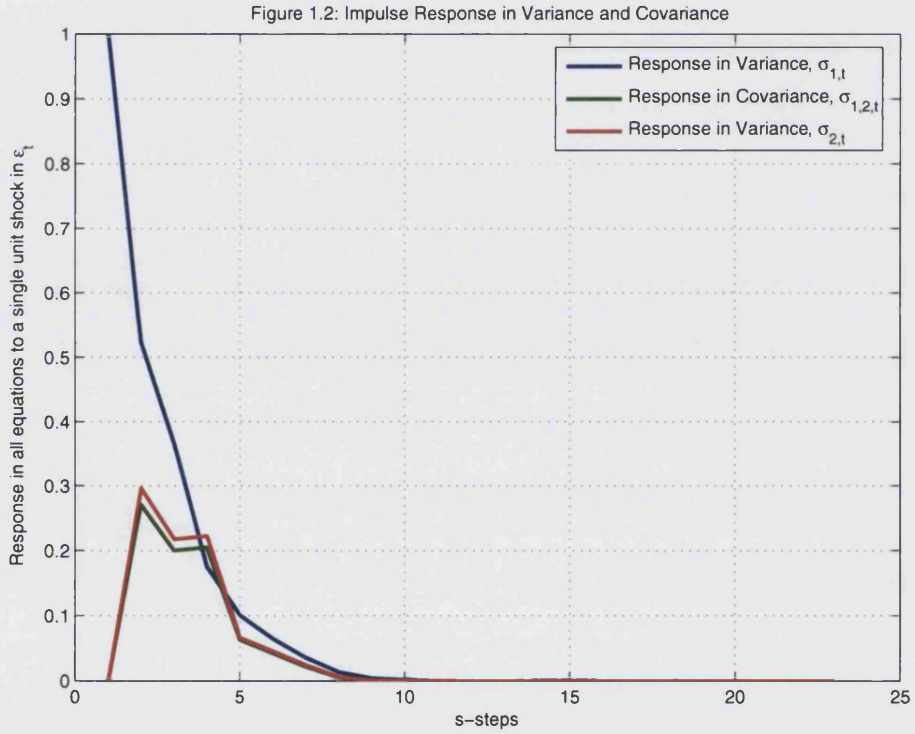

Figure 6.2: The Impulse Response in the Dynamic Correlation Coefficient, $\rho_{1,2,t} = \sigma_{1,2,t} \left(\sigma_{1,t}\sigma_{1,t}\right)^{-1}$

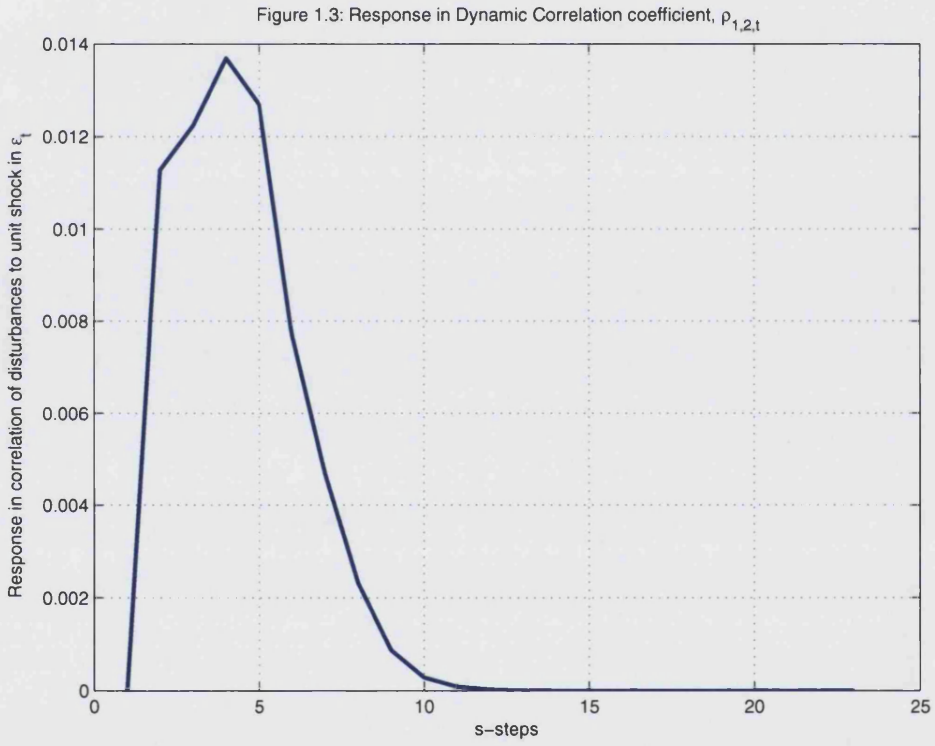

Figure 6.3: The impulse response in variance and covariance from 6.3.24.

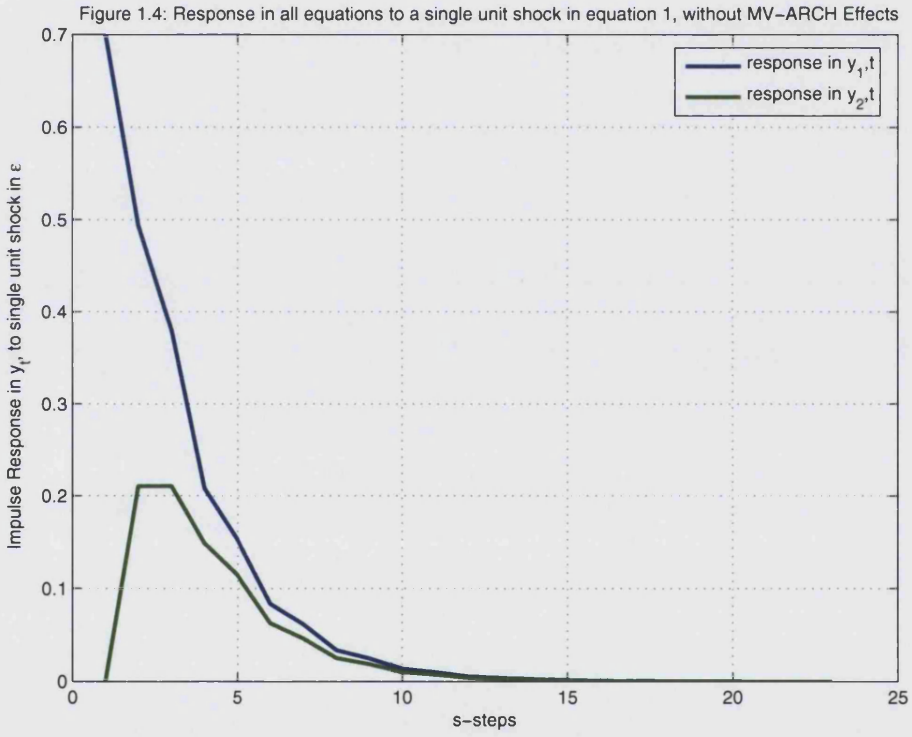

Figure 6.4: The mean equation with ARCH effects.

Chapter 7

Static and Dynamic Semi Parametric Asset Allocation Problems: Asset Allocation Methodologies with Higher Moments

7.1 Chapter Abstract

This Chapter suggests a new, non-parametric approach to static and dynamic asset allocation problems in terms of the moments of a multivariate distribution. By use of a general class of H-distributions, the portfolio density function is constructed from the moment sequence derived from the multivariate co-moments of its components. We use these high order dependencies in order to capture tail dependency, and the associated issues diversification failure and miss-estimation of portfolio risk.

7.2 Introduction and Literature

This chapter contributes to the growing literature on asset allocation where the assumption of unconditional or conditional normality has been relaxed. In order to accomplish this the probabilistic properties of a portfolio given an r length sequence of measured moments, where $r \in \mathbb{N}^+$ is inferred. This chapter reviews both a static and a dynamic representation of these moments and suggest approaches to testing the performance of the methodology.

Asset allocation problems are at the heart of modern financial decision making. The Markowitz mean/variance paradigm, the Sharpe-Lintner CAPM through to APT type models have defined excess returns as a multi-normal/lognormal process, fully characterized by the first two moments, see classic texts by Markowitz (1952,1959) [194], Sharpe (1964) and Ross (1976) [59]. Almost as soon as the mean/variance approach to asset allocation was introduced critiques appeared suggesting that the distributive properties of asset returns had marginal distributions which deviated strongly from normality and joint distributions which exhibited dependencies not full characterized by quadratic covariation, Mandelbrot and Taylor (1967) [193] through to modern treatments by Rachev and Mittnik (2000) [222]. Recent contributions to the literature by Athayde and Flores (2004) [11], Jondeau and Rockinger (2006) [156], Harvey *et al* (2005) [228] and Cvitanic, Polimenis and Zapatero (2005) [61] have looked at the inclusion of the first four moments into the distribution. The third and fourth

moments are generally classified in relation to deviations from those expected under the normal distribution assumptions given the estimates of the first two moments. It should be noted that all distributions characterized by continuously differentiable functions will have existing moments of $r > 2$. Additionally the moment sequence will be characterized by differentiation of the characteristic function derived by the Laplace transform of the density function. The question is less whether higher moments should be included in the asset allocation problem, as to do they actually matter, under some coherent performance measure.

7.3 Random Variables and Generalized Density Functions

Consider a random variable X, with some probability measure on the positive real line $\mathbb{R}^+$, where the probability measure is defined by a sequence of general **H**-function distributions, $\mathfrak{F}_i$,

$$\mathfrak{F}_i(x) = \begin{cases} k\mathbf{H}_{p,q}^{m,n}\left[z\left(\mathbf{A},\mathbf{B}\right)\right], x > 0 \\ 0 \end{cases} \tag{7.3.0}$$

where the **H** is defined as,

$$\mathbf{H}_{p,q}^{m,n}\left[z\left(\mathbf{A},\mathbf{B}\right)\right] = \mathfrak{H}(z) \tag{7.3.1}$$

and,

$$\mathfrak{H}(z) = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{C}} \frac{\prod_{j=1}^{m} \Gamma\left(b_{1,j} - b_{2,j}s\right) \prod_{j=1}^{n} \Gamma\left(1 - a_{1,j} + a_{1,j}s\right)}{\prod_{j=m+1}^{q} \Gamma\left(1 - b_{1,j} + b_{1,j}s\right) \prod_{j=n+1}^{p} \Gamma\left(a_{1,j} - a_{2,j}s\right)} z^s ds \tag{7.3.2}$$

where $\mathsf{i} = \sqrt{-1}$, Γ is the Euler gamma function and the matrices $\mathbf{A}$ and $\mathbf{B}$ are partitioned as follows,

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]^{\mathrm{T}} \tag{7.3.3}$$

$$\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2]^{\mathrm{T}} \tag{7.3.4}$$

the parameters n, m, p, q have the following boundaries and constraints, $0 \leq m \leq q$, $0 \leq n \leq p$, $a_{2,j} > 0$ and $b_{2,j} > 0$, for $j \in [1, 2, ..., p]$. Here $a_{1,j}$ and $b_{1,j}$ are complex numbers such that no pole of $\Gamma(b_{1,j} - b_{2,j}s)$ for $j \in [1, 2, ..., n]$ coincides with any pole of $\Gamma(1 - a_{1,j} + a_{1,j}s)$ for $j \in [1, 2, ..., m]$. Furthermore the set $\mathbb{C}$ is a contour in the complex s-plane from $-\mathsf{i}\infty$ to $+\mathsf{i}\infty$, see Springer (1979) for a more detailed description of **H** functions. Furthermore the parameter matrices are constrained such that,

$$\int_0^{\infty} \mathfrak{F}(x)\, dx = 1 \tag{7.3.4}$$

This looks a little abstract but in reality this functional form offers a very flexible set of probability density functions, which underlines the majority of commonly used probability distributions. However an extremely useful generic property of these distributions is as follows, consider the characteristic function or Fourier transform of 7.3, $\mathfrak{C}(x)$, which is given as,

$$\mathfrak{C}(x) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = \int_0^{\infty} e^{itx} H_{p,q}^{m,n}\left[z(\mathbf{A}, \mathbf{B})\right] dx \tag{7.3.5}$$

$$= \mathbf{L}_{-it}\left\{k\mathfrak{H}_{p,q}^{m,n}\left[z(\mathbf{A}, \mathbf{B})\right]\right. \tag{7.3.6}$$

Where $\mathbf{L}_{-it}$ is the inverse Laplace transform of 7.3.2 and $k \in \mathbb{N}^{+}$, assuming that the integral within the characteristic function converges absolutely to the definition of the 7.3.2, then the characteristic function of the H-distribution can be given as,

$$\mathfrak{C}(x) = \frac{k}{c}\mathbf{H}_{q,p+1}^{m+1,m}\left[z\left(\mathbf{A}^{-\mathbf{L}}, \mathbf{B}^{-\mathbf{L}}\right)\right] \tag{7.3.6}$$

where, the new inverse Laplace parameters are defined as,

$$z = -\frac{i}{c}t \tag{7.3.7}$$

$$\mathbf{a}_1^{-\mathbf{L}} = 1 - \mathbf{b}_1 \tag{7.3.8}$$

$$\mathbf{a}_1^{-\mathbf{L}} = \pm\mathbf{b}_2 \tag{7.3.9}$$

$$\mathbf{b}_1^{-\mathbf{L}} = \left[0, [\mathbf{e} - \mathbf{a}_1]^{\mathrm{T}}\right]^{\mathrm{T}} \tag{7.3.10}$$

$$\mathbf{b}_2^{-\mathbf{L}} = \left[1, [\mathbf{e} - \mathbf{a}_2]^{\mathrm{T}}\right]^{\mathrm{T}} \tag{7.3.11}$$

Here $\mathbf{e}$ represents an appropriately sized unit vector. If a random variable is drawn from a sequence of H-distributions then the r^{th} moment of these distributions is defined by the derivatives of the characteristic function 7.3, therefore if the moments of a distribution are defined as follows,

$$\mu_r = E(x^r) = \int_{-\infty}^{\infty} x^m \mathfrak{F}(x)\, dx \tag{7.3.11}$$

Then if $\mathbf{M}_r\mathfrak{F}(x) = E(x^{r-1})$, where $\mathbf{M}_r$ is the Mellin transform of 7.3.2, see appendix 1. The moments of the H-distribution are therefore,

$$\mu_r = \frac{k}{c^{r+1}} \frac{\prod_{j=1}^{m} \Gamma(b_{1,j} + b_{2,j} + b_{2,j}r) \prod_{j=1}^{n} \Gamma(1 - a_{1,j} - a_{1,j}r)}{\prod_{j=m+1}^{q} \Gamma(1 - b_{1,j} - b_{2,j} - b_{2,j}r) \prod_{j=n+1}^{p} \Gamma(a_{1,j} + a_{2,j} + a_{2,j}r)} \tag{7.3.11}$$

Where the ratio $\frac{k}{c^{r+1}}$ is an identity, such that c and k are either exogenously applied parameters or are free to vary when matching to a moment sequence. Given that a large number of common distributions are derived as special cases of the H-distribution and that by varying c, k, $\mathbf{A}$ and $\mathbf{B}$ powerful mathematical framework will be generated, which maybe utilize to reconstruct an estimate of the characteristic function from sequences of integer moments of random variables and by extension the density function.

7.4 Multivariate Characteristic Functions and Associated Moments

Given an n-variate vector process $\mathbf{x} \in \mathbb{R}^n$, the multivariate characteristic function, (m.c.f),$\mathfrak{H}$, is the Fourier transform of the multivariate probability density function therefore if, $\mathfrak{F}^n$ is the n-variate density function then the m.c.f. is characterized by the following multiple integral,

$$\mathfrak{H}^n(\mathbf{x}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} \exp\left(it_1x_1 + it_2x_2 + \dots + it_nx_n\right) d\mathbf{x} \tag{7.4.1}$$

$$d\mathbf{x} = [dx_1, dx_2, \dots, dx_n]^{\mathrm{T}} \tag{7.4.2}$$

Therefore the sets of co-moments are characterized as the r^{th} derivatives of this function. In this respect co-moments are defined in terms of the expectations of products

of vectors, $\mathbf{z}$ which are subsets of $\mathbf{x}$,

$$\mu_{\mathbf{z}}^{s} = E\left(\prod_{i=1}^{s} z_i\right) \tag{7.4.3}$$

$$z_i \in \mathbf{z} \tag{7.4.4}$$

$$\mathbf{z} \subset \mathbf{x} \tag{7.4.5}$$

The number of co-moment conformations from the characteristic function will therefore be $\imath = n^r$, and as such an array containing the r^{th} co-moments will have $\imath$ elements. Next consider a univariate process which is the weighted summation of the vector process $\mathbf{x}$, therefore if $x = \omega^{\mathrm{T}}\mathbf{x}$, where ω is some set of arbitrary summation weights. The contribution to the r^{th} moment in the univariate moment sequence $\mu = [\mu_1, \mu_2, ..., \mu_r, ..., \mu_\infty]^{\mathrm{T}}$, by the i^{th} co-moment is defined as,

$$\gamma_{\mathbf{z}}^{s} = \prod_{i=1}^{s} w_i \tag{7.4.5}$$

where, $w_i \in \mathbf{w}$ and $\mathbf{w} \subset \omega$. We now have an expression which will yield the parameters of a generalized distribution which is the weighted summation of n marginal distributions, if these marginal distributions are drawn from some set of dependent H-distributions that characterize the distribution of the resultant distribution in terms of moments. Furthermore by assuming that the resultant distribution takes the form of some infinite or finite mixture then this approach offers the tools to deal with this type problem.

7.5 Specifying Higher Order Moments

Consider the following Hankel matrices of moments for a univariate random variable,

$$\boldsymbol{\Delta}_r^{(1)} = \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_{r+1} \\ \mu_2 & \mu_3 & \cdots & \mu_{r+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{r+1} & \mu_{r+2} & \cdots & \mu_{2r+1} \end{bmatrix} \tag{7.5.1}$$

$$\boldsymbol{\Delta}_{r+1} = \begin{bmatrix} 1 & \mu^{\mathrm{T}} \\ \mu & \boldsymbol{\Delta}_r^{(1)} \end{bmatrix} \tag{7.5.2}$$

The Hankel matrix is constructed from a valid set of moments if, $|\boldsymbol{\Delta}_r| \geq 0$ and for any continuation of the moment sequence if $k > r$, then $\{|\boldsymbol{\Delta}_r| = 0 : |\boldsymbol{\Delta}_k| = 0\}$, for $k, r \in \mathbb{N}^+$, Shohat (1943) [236]. For a random variable x where $x = \omega^T \mathbf{x}$, the moment sequence maybe defined as follows, if $\mathcal{M}_r$ is an array of co-moments, where,

$$E\left(\mathbf{x}^{[r]}\right) = \mathcal{M}_r \tag{7.5.2}$$

Specifying $\mathbf{x}^{[r]}$ as the vector permuted outer product of the vector $\mathbf{x}$ with itself, for the second co-moments, this changes to the transpose operator, i.e. $\mathbf{x}^{[2]} \equiv \mathbf{x}\mathbf{x}^{\mathrm{T}}$. The resultant array $\mathcal{M}_r$ is an r dimensional array (classical tensor), with r^{th} order super-symmetry. Matricizing from a multidimensional array to a flat matrix, containing identical elements to $\mathcal{M}_r$, the following definitions apply,

$$vec\left(\mathbf{M}_r\right) = vec\left(\mathcal{M}_r\right) \tag{7.5.2}$$

In this case the *vec* operator stacks the fibre bundles of the array $\mathcal{M}_r$, in such a way that the inverses are homeomorphic, i.e. for a pair of inverse transformations

$ivec_r$,$ivec_2$ of an array, the following definitions hold,

$$ivec_r\left(vec\left(\mathbf{M}_r\right)\right) = \mathcal{M}_r \tag{7.5.3}$$

$$ivec_2\left(vec\left(\mathcal{M}_r\right)\right) = \mathbf{M}_r \tag{7.5.4}$$

Then defining the matrix $\mathbf{M}_r$ in terms of Kronecker powers and first order permutations, (transposes) of $\mathbf{x}$

$$E\left(\left(\mathbf{x}^{[\otimes r-2]}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_r \tag{7.5.4}$$

Imposing $\mathbf{x} = \mathbf{x}^{[\otimes 0]}$. If for a given vector of weights, ω, then the weights array is similarly defined as $\omega^{[r]} = \mathcal{W}_r$, again specifying following matricizing condition, $vec\left(\mathbf{W}_r\right) = vec\left(\mathcal{W}_r\right)$, therefore, $\left(\omega^{[\otimes r-2]}\right)\omega^{\mathrm{T}} = \mathbf{W}_r$. From this definition r^{th} moment μ_r of a random variable $x = \omega^{\mathrm{T}}\mathbf{x}$, is therefore,

$$\mu_r = \mathcal{W}_r \bullet \mathcal{M}_r \tag{7.5.4}$$

Where $\bullet$ is the inner tensor product of two identical arrays, rearranging this and reframing in vector notation as,

$$\mu_r = vec\left(\mathbf{W}_r\right)^{\mathrm{T}} vec\left(\mathbf{M}_r\right) \equiv vec\left(\mathcal{W}_r\right)^{\mathrm{T}} vec\left(\mathcal{M}_r\right) \tag{7.5.4}$$

Therefore the r^{th} moment of x is maybe then defined in matrix notation as,

$$\mu_r = vec\left(\left(\omega^{[\otimes r-2]}\right)\omega^{\mathrm{T}}\right)^{\mathrm{T}} vec\left(E\left(\left(\mathbf{x}^{[\otimes r-2]}\right)\mathbf{x}^{\mathrm{T}}\right)\right) \tag{7.5.4}$$

Table 7.5 illustrates the first 4 co-moments of $\mathbf{x}$,

Moment Array by Vector Permutations	Matrix Notation	Kronecker Power Notation
$E\left(\mathbf{x}^{[1]}\right) = \mathcal{M}_1$	$E\left(\mathbf{x}\right) = \mathbf{M}_1$	$E\left(\mathbf{x}\right) = \mathbf{M}_1$
$E\left(\mathbf{x}^{[2]}\right) = \mathcal{M}_2$	$E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_2$	$E\left(\left(\mathbf{x}^{[\otimes 0]}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_2$
$E\left(\mathbf{x}^{[3]}\right) = \mathcal{M}_3$	$E\left(\left(\mathbf{x}\otimes\mathbf{x}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_3$	$E\left(\left(\mathbf{x}^{[\otimes 1]}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_3$
$E\left(\mathbf{x}^{[4]}\right) = \mathcal{M}_4$	$E\left(\left(\mathbf{x}\otimes\mathbf{x}\otimes\mathbf{x}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_4$	$E\left(\left(\mathbf{x}^{[\otimes 2]}\right)\mathbf{x}^{\mathrm{T}}\right) = \mathbf{M}_4$

Table 7.5

The first 4 moments of $x = \omega^{\mathrm{T}}\mathbf{x}$, are therefore,

$$\begin{aligned}
\mu_0 &= 1 \\
\mu_1 &= \omega^{\mathrm{T}}\mathbf{x} \\
\mu_2 &= vec\left(\omega\omega^{\mathrm{T}}\right)^{\mathrm{T}} vec\left(E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\right) \equiv \omega^{\mathrm{T}}\left(E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\right)\omega \\
\mu_3 &= vec\left(\left(\omega\otimes\omega\right)\omega^{\mathrm{T}}\right)^{\mathrm{T}} vec\left(E\left(\left(\mathbf{x}\otimes\mathbf{x}\right)\mathbf{x}^{\mathrm{T}}\right)\right) \\
\mu_4 &= vec\left(\left(\omega\otimes\omega\otimes\omega\right)\omega^{\mathrm{T}}\right)^{\mathrm{T}} vec\left(E\left(\left(\mathbf{x}\otimes\mathbf{x}\otimes\mathbf{x}\right)\mathbf{x}^{\mathrm{T}}\right)\right)
\end{aligned} \tag{7.5.4}$$

The moment sequence, μ may the be used to reconstruct the density function from the generalized H-distribution density function. The essential task is a maximum likelihood problem, for a given set of sample moments, which H-density function most appropriately fits. There are a number of differing approaches to this problem, the original method proposed by Hill (1969) in Springer (1979) suggests using a Gram-Charlier type A (Hermite polynomial, i.e. mixture of normals) series or Leguerre polynomial series. A second method suggests attacking the problem via estimation of the first four moments and fitting these to a Pearson system of probability density functions.

7.5.1 Transforming the Raw Aggregate Moments into $\mathbb{R}^+$

The H-function class defines a probability measure over $\mathbb{R}^+$, however the raw moment sequences operate over $\mathbb{R}^n$. Therefore a transformation is necessary in order to satisfy, the portfolio distribution constraint. A simple method is to transform the multivariate sequence into the positive domain, for return sequences the lowest value is bounded to -1, therefore adding 1, to $\mathbf{x}$, will yield a multivariate distribution, in $\mathbb{R}^{n,+}$. Then the un-centered moment estimates will yield an aggregate distribution in $\mathbb{R}^+$. Alternatively for unbounded multivariate distributions, using the exponential function results in transformation to $\mathbb{R}^+$.

7.6 Reconstructing the Density Function of x

One of the oldest methods of reconstructing a density function from its moments is the Von Mises (1964) [260] stepwise approach, which proposes matching the moment sequence to the moments of some arbitrary density function. The more moments that are used the more unique the moment sequence and as such the better the identification of the density function. At the time of publication in (1964) nobody could have envisioned that one could realistically optimize a library of general H-function distributions and for example test the fit using quadrature density matching, without an *a priori* imposition of some distribution which is then parameterized via its moments. Given that the H-function serves as the basis for the half/log normal,

Weibull, Gamma, Maxwell, Exponential, Chi-Squared, Rayleigh, Hypergeometric, Cauchy, Student, F distributions and many more then the choice was fairly eclectic. At present with access to more powerful computational resources and a an easier implementation paradigm, such as object orientated programming, it is possible to generalize from the H-distribution, to finite sequence of moments the optimal density function based on some preset criterion. The first step is to deal with the information content that the moment sequence contains, recall the Hankel matrix $\mathbf{\Delta}_r$ from 7.5.1, this matrix maybe partitioned into sequences of moments as follows,

$$\mathbf{\Delta}_{2r} = \left[\mu_{[0,r]}, \mu_{[1,r+1]}, ..., \mu_{[r,2r]}\right]^{\mathrm{T}} \tag{7.6.0}$$

A new partitioned matrix, $\mathbf{\Phi}$ may now be formed,

$$\mathbf{\Phi} = \begin{bmatrix} \mu_{[0,r]} & \mu_{[0,r]} & \cdots & \mu_{[0,r]} & \mu_{[0,r]} \\ \mathbf{0} & \mu_{[1,r+1]} & \cdots & \mu_{[1,r+1]} & \mu_{[1,r+1]} \\ \mathbf{0} & \mathbf{0} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \mu_{[r,2r-1]} & \mu_{[r,2r-1]} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mu_{[r,2r]} \end{bmatrix}^{\mathrm{T}} \tag{7.6.0}$$

where,

$$\mu_{[p,q]} = \left[\mu_p, ..., \mu_q\right]^{\mathrm{T}} \tag{7.6.0}$$

For a vector of coefficients $\mathbf{c} = [c_0, c_1, c_2, ..., c_{2r}]^{\mathrm{T}}$ the following simple linear algebra problem is then solved,

$$\mathbf{\Phi c} + \mu_{[r+1,2r+1]} = \mathbf{0} \tag{7.6.0}$$

To obtain the abscissa values, $\{\nu_1, \nu_2, ..., \nu_{2r-1}\}$, the roots of the $2r - 1^{th}$ polynomial with coefficients described by $\mathbf{c}$ are evaluated. This system maybe rewritten as a

set of Legendre polynomials, see Devroye (1989) [68]. Using the definition in 7.3 a general H-function and generating the abscissa values via backward induction yields a generic representation of the density function.

7.6.1 Application: Generating Multivariate Price Processes With Higher Moment Arrays

Consider the following monte-carlo problem, in order to replicate a sample multivariate sample path a moment matching scheme is produced as follows,

- First Estimate the Number of central co-moments needed to fully describe the characteristic function.
- Measure the central co-moments.
- Match the measured multivariate moments to a weighted finite mixture drawn from a known density function, e.g. multivariate normal.
- Generate the sample paths, using a standard linear congruential quasi-random number generator.

This type of schema is a direct application of the standard sequential monte-carlo approach utilized in a variety of different areas, such as signal processing and tracking and control, See Ristic *et al*, (2005) [224].

7.6.2 Moment Matching

Consider a library of potential density functions, $\mathfrak{F}_i \in \mathfrak{F}$, which are all variants of a multivariate H distribution, the general multivariate density function of the mixture distribution, is given by,

$$\mathfrak{P}(\mathbf{x}) = \sum_{j=1}^{k} \alpha_j \mathfrak{F}_j(\theta_i) \tag{7.6.0}$$

where $\alpha_j \in [0,1]$ and $\sum_{j=1}^{k} \alpha_j = 1$. If the multivariate characteristic function, $\mathfrak{H}(\mathbf{x})$, of the density function $\mathfrak{P}(\mathbf{t})$, is r-differentiable then the k^{th}, multivariate *raw* moment of $\mathbf{x}$, is,

$$\bar{\mathcal{M}}_r(\mathbf{x}) = \frac{1}{i^r} \frac{d^r}{d\mathbf{t}^r} H(\mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} \tag{7.6.0}$$

where, $\mathbf{t} \in \mathbb{R}^n$, where $\mathbf{t}$ is dummy vector the same length as $\mathbf{x}$. The central moment is similarly,

$$\hat{\mathcal{M}}_r(\mathbf{x}) = \bar{\mathcal{M}}_r(\mathbf{x} - E(\mathbf{x})) = \frac{1}{i^r} \frac{d^r}{d\mathbf{t}^r} H(\mathbf{t}) - E(\mathbf{x})^{\mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}} \tag{7.6.0}$$

Now consider a measured set of moments, $r \in [1, ..., \rho]$, from some multivariate dataset,

$$\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_r, ..., \mathcal{M}_\rho\} \tag{7.6.0}$$

matricizing, $\mathbf{M}_r = E(\mathbf{x}^{\otimes r}) = \mathcal{M}_r$, yields a sequence of matrices $\mathbf{M}_r \in \{\mathbf{M}_1, ..., \mathbf{M}_\rho\}$, comparing these to the expected matrices, from the distribution,

$$\mathbf{\Delta}_r = \mathbf{M}_r - \hat{\mathbf{M}}_r \tag{7.6.0}$$

substituting, 7.6.2 into 7.6.2 and rewriting as matricized arrays yields,

$$\Delta_r = mat\left(\mathcal{M}_r - \frac{1}{i^r}\frac{d^r}{d\mathbf{t}^r}H(\mathbf{t}) - E(\mathbf{x})^{\mathbf{t}}\Big|_{\mathbf{t}=\mathbf{0}}\right) \tag{7.6.0}$$

A loss function is now needed in order to optimize the parameters arrays $\{\alpha_j, \theta_i\} \in \theta$, the following basic loss function is proposed,

$$\hat{f}(\theta) = \sum_{r=1}^{\rho} \frac{f^r(\mathbf{t})}{r!}\left(vec(\Delta_r)^{\mathrm{T}}{}_{n^r\times 1}\mathbf{e}\right) \tag{7.6.0}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ and is r differentiable, the simplest case is when, $f(\mathbf{t}) = \exp\left(-\frac{1}{h}|\mathbf{t}|\right)$, where h is a bandwidth parameter used to tune importance weighting of various moments. Minimizing $\left(\hat{\theta}\right)$, subject to the parameter vector θ, yields the parameter estimates of the optimal mixture density function.

$$\hat{f}(\theta^*) \triangleq \min_{\theta}\left(\sum_{r=1}^{\rho} \frac{f^r(\mathbf{t})}{r!}\left(vec(\Delta_r)^{\mathrm{T}}{}_{n^r\times 1}\mathbf{e}\right)\right) \tag{7.6.0}$$

7.6.3 Empirical Example, Fitting a Bivariate Mixture using Moment Matching Libraries

Consider the case of a bivariate portfolio, using the moment matching procedure for the library of distributions in table 7.1 Figures 7.1 and 7.2 demonstrate the higher multivariate moment system for a bivariate system of stock and index.

7.7 Filtration Problems

The previous two sections addressed the identification of a generic density function from its moments. In these sections the evolution of a random variable from a random

Table 7.1: Distribution Library, for Moment Matching Procedure

Distribution	Density Funcion
Bivariate Normal	$F_x(\mathbf{x}_x) = 2\pi^{-\frac{n}{2}} \left\|\boldsymbol{\Sigma}\right\|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - E(\mathbf{x})\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - E(\mathbf{x})\right)\right)$
McKay's Bivariate Gamma	$F_y(\mathbf{x}_y) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)} x_1^{a-1}(x_2 - x_1)^{a-1} \exp(-cx_2)$
General Bivariate Extreme Value Distribution	$F_z(\mathbf{x}_z) = F_{x_1}(x_1) F_{x_2}(x_2) \exp\left(-\alpha\left(\frac{1}{\log F_{x_1}(x_1)} + \frac{1}{\log F_{x_2}(x_2)}\right)^{-1}\right)$ $F_{x_i}(x_i) = \exp(-\exp(-x_i))$

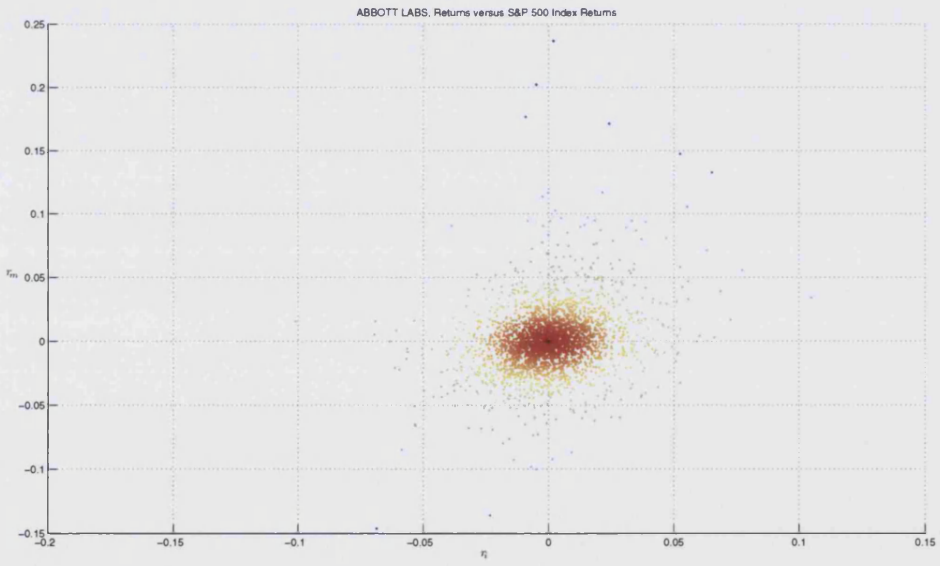

Figure 7.1: Scatter Plot of Abbot Labs Returns versus the S&P 500 Index Returns.

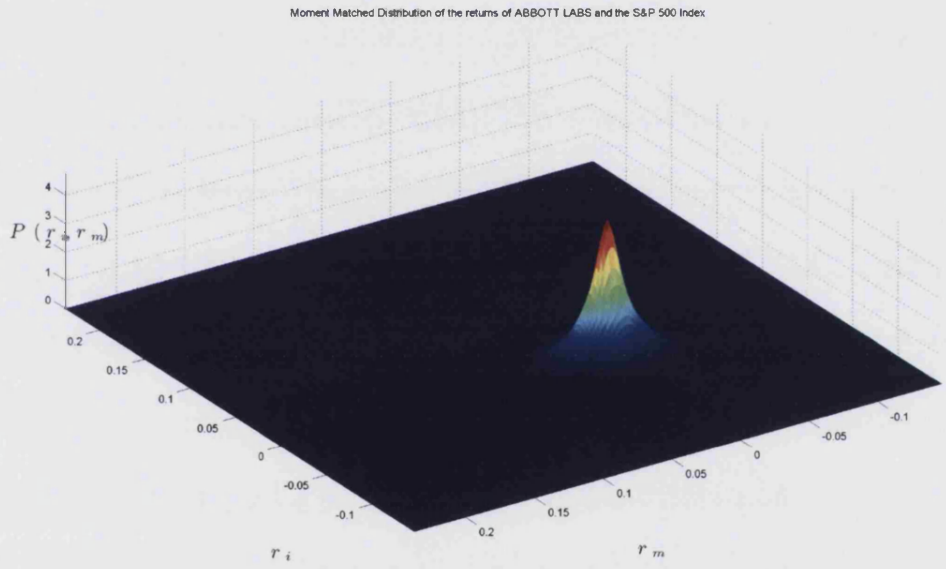

Figure 7.2: The bivariate density function, evaluated from the library.

sequence of density functions is defined, in this following section a more specific notation is used and in particular the following problem of the dynamic evolution of a moment array is considered. If for a given set of polynomials conditioned in the time dimension, the moment sequence at any given innovation is then conditioned on the past evolution of the prior moments. More specifically the non-linear evolution of the moment sequence is considered, in the context of a simple autoregressive mechanism. Therefore the following moment problem for a given conditional moment sequence, $\mu_{[0,r],t}$, is specified in terms of the conditional distribution at time t and will therefore be defined as $\mathfrak{F}_t$.

7.7.1 Dynamic Specification

Consider the following random vector process $\mathbf{u}_t$, whereby $\mathbf{u}_t$, is the zero mean residue from some filtration,

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{m\times n}{\mathbf{\Pi}^{\mathrm{T}}}\ \underset{m\times 1}{\mathbf{x}_t} + \underset{n\times 1}{\mathbf{u}_t} \tag{7.7.0}$$

Where $\mathbf{\Pi}$ is a matrix of coefficients, $\mathbf{y}_t$ is the pre-filtered vector process, $\mathbf{x}_t$ is some deterministic set of explanatory variables. For a given Borel measure of $\mathbf{u}_t \in \mathbb{R}^n$, the conditional density of $\mathbf{u}_t$ is given by the conditional density function $\mathfrak{F}_t$, and the epoch of the density function is described by some Labesgue measure on $\mathbb{R}^{nt}$, therefore there is a bounded functional set $\mathbb{F}$, for which $\mathfrak{F}_t \in \mathbb{F}$. For a conditional moment/co-moment sequence $\mu_{[0,\infty],t}$, which fully describes the density function $\mathfrak{F}_t$ then there exists a set $\mathbb{M}$, for which $\mu_{[0,\infty],t} \in \mathbb{M}$ and as such there is some function g, for which,

$g : \mathbb{M} \rightarrow \mathbb{F}$. In practice this means that if a sequence of random density functions is fully described by their conditional moment sequences then the parameters of the polynomial which underlies this sequence will be fully described the properties of the random variables generated from this distribution sequence.

7.7.2 Autoregressive Conditional Moments

Consider the following general moment innovation, for the i^{th} marginal distribution,

$$\mu_{r,i,t} = E\left(u_{i,t}^{r} \middle| u_{i,t-1}^{r}, u_{i,t-2}^{r}, ..., u_{i,t-p(r)}^{r}, u_{i,t-1}^{r-1}, ..., u_{i,t-p(r-1)}^{r-1}, ..., u_{i,t-p(1)}^{1}\right) \tag{7.7.0}$$

Therefore for the i^{th} variable the infinite moment sequence is described as,

$$\mu_{[0,\infty],i,t} = \left[1, E\left(u_{i,t}^{1}\right), E\left(u_{i,t}^{2}\right), ..., E\left(u_{i,t}^{\infty}\right)\right]^{\mathrm{T}} \tag{7.7.1}$$

$$\mu_{r,t} = \left[\mu_{r,1,t}, \mu_{r,2,t}, ..., \mu_{r,n,t}\right]^{\mathrm{T}} \tag{7.7.2}$$

Now consider the conditional r^{th} order co-moments, $\mathcal{M}_r$, from the multivariate characteristic function 7.4.3. The array maybe decomposed as follows,

$$vec\left(\mathcal{M}_t\right) = vec\left(\mathcal{H}_t \times \mathcal{R}\right) \equiv vec\left(\mathbf{H}_t \times \mathbf{R}_t\right) \tag{7.7.2}$$

Where $\times$ is the element by element multiplication of two identical arrays, when the arrays are matrices then $\times$ is the Hadamard product. The following structure is proposed,

$$\mathbf{h}_{r,t} = \left[\mu_{r,1,t}^{\frac{1}{r}}, \mu_{r,2,t}^{\frac{1}{r}}, ..., \mu_{r,n,t}^{\frac{1}{r}}\right]^{\mathrm{T}} \tag{7.7.3}$$

$$vec\left(\mathbf{H}_{r,t}\right) = vec\left(\left(\mathbf{h}_{r,t}^{[\otimes r-2]}\right)\mathbf{h}_{r,t}^{\mathrm{T}}\right) = vec\left(\mathcal{H}_{r,t}\right) \tag{7.7.4}$$

And the unconditional array is as follows,

$$vec\left(\tilde{\mathbf{M}}_r\right) = vec\left(E\left(\left(\mathbf{u}^{[\otimes r-2]}\right)\mathbf{u}^{\mathrm{T}}\right)\right) = vec\left(\tilde{\mathcal{M}}_r\right) \tag{7.7.5}$$

$$\mathbf{d}_r = sdiag\left(\tilde{\mathcal{M}}_r\right) \tag{7.7.6}$$

$$vec\left(\mathbf{R}_r\right) = vec\left(\left(\mathbf{h}_r^{[\otimes r-2]}\right)\mathbf{h}_r^{\mathrm{T}}\right) = vec\left(\mathcal{R}_r\right) \tag{7.7.7}$$

Where $\tilde{\mathcal{M}}_r = E\left(\mathbf{u}^{[r]}\right)$ is the unconditional expectation of the higher moment array and *sdiag* is the super diagonal of the r-dimensional array $\tilde{\mathcal{M}}_r$.

7.7.3 Univariate Moment Specification

Given the generic derivation suggested in 7.7.2, a more specific recursion assumes that the moments are a linear progression,

$$\mu_{r,i,t} = \begin{cases} E\left(u_{i,t}^r\right) = \sigma_i^r + \sum_{i=1}^r \alpha_i^{\mathrm{T}} vec\left(\mathcal{U}_{t-p(i)}^i\right), r = 2k-1 \\ E\left(u_{i,t}^r\right) = \sigma_i^r + \sum_{i=2}^r \alpha_i^{\mathrm{T}} vec\left(\mathcal{U}_{t-p(i)}^i\right), r = 2k \end{cases} \tag{7.7.8}$$

$$\mathcal{U}_{t-p(i)}^i = \left[u_{i,t-1}^r, u_{i,t-2}^r, ..., u_{i,t-p(i)}^r\right] \tag{7.7.9}$$

Where $k \in \mathbb{N}^+$ and $\alpha_i \in \alpha$ are parameter vectors, here the recursion between the odd and even moments is distinguished, to avoid dealing with possible complex roots in $\mathbf{h}_{r,t}$. A simplification of this specification is already implemented in the Power ARCH model of Ding, Granger and Engle (1993) [88].

$$\mu_{r,i,t} = E\left(u_{i,t}^r\right) = \sigma_i^r + \alpha_r^{\mathrm{T}} vec\left(\mathcal{U}_{t-p(r)}^r\right) \tag{7.7.9}$$

Kristensen and Rahbek (2005) [172] derive the quasi maximum likelihood estimation for this model, with parameter space $\theta = [\sigma, \alpha]$.

7.8 Application to Optimal Portfolio Selection and Utility Theory

One of the basic problems in finance is the optimal allocation of risky assets in a portfolio. Consider a risk averse investor with utility function $u(v) = \{v \,|v \in \mathbb{R} : u(v) \rightarrow \mathbb{R}\}$, where v is the value of wealth, the investor must choose the optimal allocation of a set of n risky assets, with returns $\mathbf{x} = [x_1, x_2, ...x_n]^{\mathrm{T}}$ over two time periods and receives wealth of $v = v_0\left(1 + \omega^{\mathrm{T}}\mathbf{x}\right)$ in the second period, where v_0 is the initial wealth and ω is the portfolio weights. The i^{th} derivative of the investors utility w.r.t. wealth is defined as,

$$u^{(i)}(v) = \frac{d^{(i)}u(v)}{dv} \tag{7.8.0}$$

The expected utility in the second period $E(u)$is defined in terms of the Taylor expansion with remainder about the initial wealth,

$$E(u) = \left(\sum_{i=0}^{m} \frac{u^{(i)}(v_0)}{i!} E(v - v_0)^i\right) + \left(\frac{u^{(i+1)}(v_0 + \theta(v - v_0))}{i+1!} E(v - v_0)^{i+1}\right) \tag{7.8.0}$$

assuming that, the portfolio moments maybe substituted into the expected utility function,

$$E(v - v_0)^i = \left(v_0\left(1 + (\mu_i)^{\frac{1}{i}}\right)\right)^i \tag{7.8.0}$$

rewriting the expected utility as a function of the weight allocation and the portfolio moments,

$$\begin{aligned} E(u|\omega) &= L(\omega) && (7.8.1) \\ &= -\left(\sum_{i=0}^{m} \frac{u^{(i)}(v_0)}{i!} v_0 \left(1 + (f_i(\omega))^{\frac{1}{i}}\right)\right) && (7.8.2) \\ &\quad -\left(\frac{u^{(i+1)}(v_0 + \theta(v - v_0))}{i+1!} v_0 \left(1 + (f_i(\omega))^{\frac{1}{i}}\right)\right) \end{aligned}$$

The optimization problem is therefore defined as follows,

$$L(\omega^*) \triangleq \min_{\omega} (L(\omega) | \omega \in \Omega) \quad (7.8.1)$$

Where Ω is the set of all feasible portfolios. The gradient vector of partial derivatives w.r.t ω is therefore,

$$\nabla L(\omega) = \left[\frac{\partial L(\omega)}{\partial \omega_i}\right]_{i=[1,\ldots,n]} \quad (7.8.1)$$

The second order matrix (Hessian) of partial derivatives are,

$$\nabla^2 L(\omega) = \left[\frac{\partial^2 L(\omega)}{\partial \omega_i \partial \omega_j}\right]_{i,j=[1,\ldots,n]} \quad (7.8.1)$$

The optimal weighting is the global maxima at which,

$$\nabla L(\omega) = \mathbf{0} \quad (7.8.1)$$

7.8.1 Utility Functions and their derivatives

A variety of utility functions are commonly utilized in financial decision making, the effect on the optimal investment allocation are strongly influenced by the nature of

the curvature underlying the utility function. Risk aversion is characterized by convex preferences, however higher moments are dictated by the nature of this convexity. A commonly used utility function is the CRRA or power function,

$$u(v) = \begin{cases} \frac{v^{1-\gamma}}{1-\gamma} & l > 1 \\ \log(v) & \gamma = 1 \end{cases} \tag{7.8.1}$$

A standard HARA type function

$$u(v) = 1 - e^{-\gamma v} \tag{7.8.1}$$

A more exotic hyperbolic function

$$u(v) = 1 - ArcSinh\left(v^{-\gamma}\right) \tag{7.8.1}$$

And finally the Lamberts W function

$$u\left(v\right) = \begin{cases} ue^{u} = x \\ x = v^{-1} \end{cases} \tag{7.8.1}$$

Where γ is the risk aversion parameter.

7.8.2 Empirical Example

Figure 7.3 illustrates the tangency preference set (using a CRRA type utility function) and efficient frontier, solution using a three moment system, the frontier is computed using a standard constrained quadratic optimization algorithm, see Fletcher (2000) [96]. Figure 7.4, illustrates the impact of the inclusion of one extra moment in the decision system.

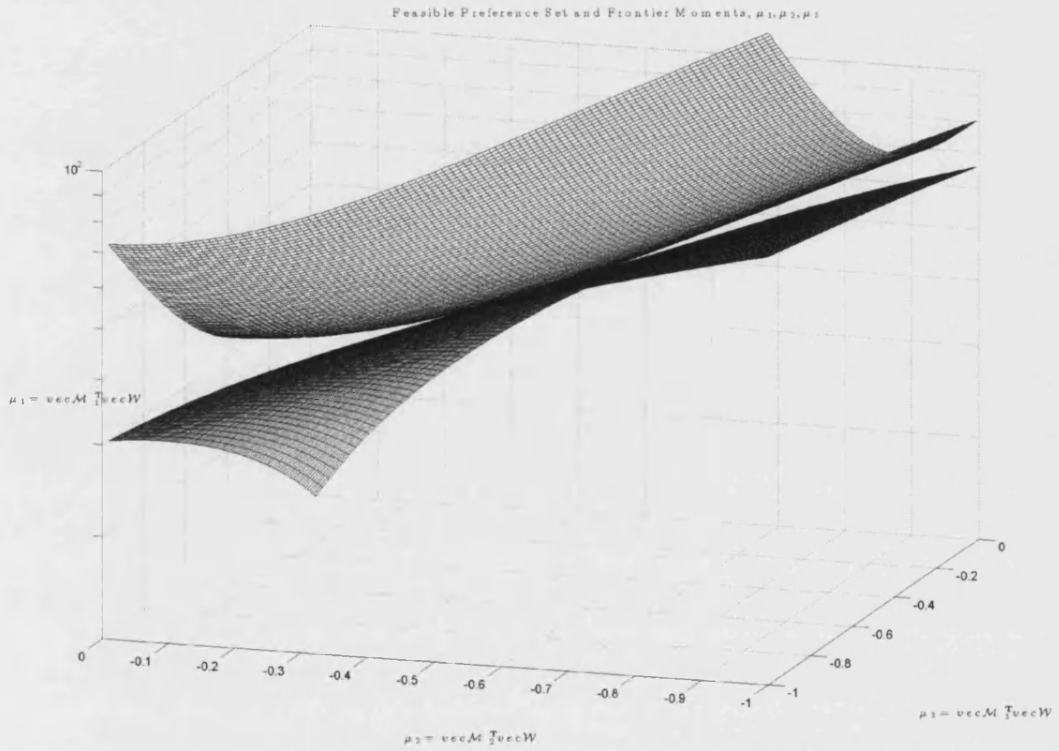

Figure 7.3: Three Moment Preference Set and Interpolated Frontier Moments for a multi-asset portfolio, the data is from a 4-variate Copula.

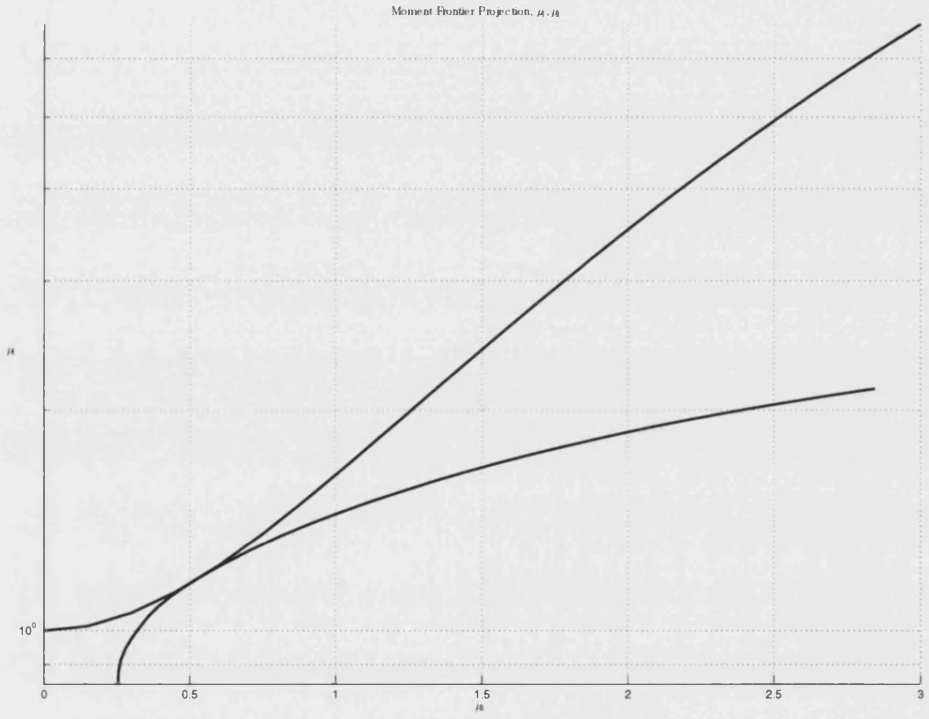

Figure 7.4: This diagram illustrates that the optimal solution for the efficient may be inside the mean-variance frontier when extra moments are included.

7.9 Concluding remarks

This chapter has introduced a series of tools designed to utilize empirically estimated moments in the asset allocation problem. The tools utilize the empirically estimated moments and match those moments to a generalized density function. Furthermore an expected utility maximizing framework is demonstrated that combines the higher moments into a coherent framework, utilizing the Scott and Horvarth preference directions. This approach appears to be a simple method for increasing the information used in asset allocation problems and dealing with observed deviations from multivariate-normality and it's associated moment sequence that appear in many asset return series.

7.10 Chapter Appendix I: Properties Higher dimensional arrays

For any rank-m co-moment tensor there exists an m-1 ranked tensor object which defines the symmetry of them^{th} co-moment array. For any given tensor with dimensions $J \times K \times L$ the Tucker decomposition is for a 3 dimensional array is $\mathcal{M} = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \sum_{k=1}^{d_3} \sigma_{ijk} (\mathbf{u}_i \circ \mathbf{v}_j \circ \mathbf{w}_k)$ where $\mathbf{u}_i \in \mathbb{R}^{d_1}, \mathbf{v}_j \in \mathbb{R}^{d_2}, \mathbf{w}_k \in \mathbb{R}^{d_3}$ and the tensor $\mathcal{S}^c = \sigma_{ijk}$, is called the core tensor. If the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are column vectors from orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$, then the tucker decomposition is called the Higher-Order Single Value Decomposition, HOSVD. The HOSVD can then be defined

in terms of the matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ as $\mathcal{M} = \mathcal{S}\mathbf{UV}$ The rank of the tensor is not revealed unless the tensor has a diagonal core, the diagonal of the tensor is when the contra-variant indices are equal i.e. in a three dimensional array $i = j = k$. The vector of the diagonal in a m^{th} co-moment array is a line on the symmetry array, which is equivalent to Σ. Therefore an m^{th} co-moment array is a tensor object, which exhibits high order super symmetry. Thus for each element $[a_{ijk}]$ in a three dimensional array, the super-symmetry properties hold because $a_{ijk} = a_{jki} = a_{kij} = a_{kji} = a_{ikj} = a_{jik}$ The super diagonal for an appropriate flattening of a co-skewness matrix (which is designated a super cube, due to its symmetry properties). $\mathcal{M} = \sum_{i=1}^{\mathbf{R}} \mathbf{u}_i.\mathbf{v}_i.\mathbf{w}_i$ Re-writing this in terms of flattened matrices, gives $\mathcal{M} = \mathbf{U}\left(\mathbf{V} \odot \mathbf{W}\right)^{\mathrm{T}}$ Where $\odot$ is the Khatri-Rao-Bro product finally we can rewrite this transformation in terms of the more familiar Kronecker product as $\mathcal{M} = \mathbf{UI}\left(\mathbf{V} \otimes \mathbf{W}\right)^{\mathrm{T}}$ where $\mathbf{I}$ is an n way identity matrix with ones on the super diagonal, zeros elsewhere, if the tensor $\mathcal{M}$ is for example a co-skewness matrix with 3 assets, then the identity matrix is of the following form, where the ones specify the points on the super diagonal.

$$\mathbf{I} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{7.10.0}$$

Finally we can describe the decomposition of the tensors in terms of sets of matrices, as well as columns, using the Kronecker (or tensor) product, this is useful as it allows us to create simple functions to construct the flat weights matrix. $\mathcal{W} = \sum_{i=1}^{R} \sigma_i(U_i \otimes W_i)$, where $\bar{R} = rank(\mathcal{M})$, conveniently as the weights matrix is self ascribed, by the

column vector ω then:

$$\mathcal{W} = \omega^{\mathrm{T}} \left(\omega \otimes \omega\right)^{[1,\ldots,m-1]} \tag{7.10.0}$$

So from these results we can flatten any m^{th} order hypercube array and have a function that produces the equivalent weights array, in a flattened form.

Chapter 8

Continuous Time Stochastic Covariance Problems and option pricing models: Multivariate Asset Price Dynamics With Stochastic Covariation

8.1 Chapter Abstract

This chapter illustrates the implications of multivariate stochastic covariation on the pricing of contingent claims on assets. Stochastic volatility models such as those of Heston (1993) [144] and Hull and White (1987) [149] are often used to model volatility risk in the pricing and hedging of contingent claims on risky assets. However some recent empirical evidence has shown that the models under general specifications often do not fully capture the volatility dynamics observed *in situ*. We suggest that some of the fitting issues maybe partially explained by correlation risk, with other assets in the market.

8.2 Introduction and Review

In this chapter a multivariate approach to modelling volatility risk is proposed. Typical examples of the incorporation of stochastic volatility in the pricing of contingent claims on single assets are the Heston (1993) [144]and the more generalized Hull and White (1987) [149]models.

This chapter demonstrates that when the underlying asset exhibits quadratic covariation with a number of other assets and that this covariation is itself a multidimensional diffusion then the estimation of the risk for holding this asset will not be captured by two dimensional brownian motion approaches. In particular much of the criticism presented in Hagan *et al* (2002) [133] on the lack of fit between generalized local volatility models and the observed smile surface can be countered if several other processes were interacting in the volatility equation. Although this is a valid criticism, it does not provide advice on the exact number of such processes.

In modelling the time evolution of the quadratic covariation of asset returns Barberis *et al* (2004) [17]demonstrate that the degree of co-movement observed amongst assets, decomposed as variance and correlation, is a dynamic process affected by changes in the investors information set, chapter 4, similarly demonstrated that the degree of correlation between stocks in a sector. Similarly, in collateralized default obligations (CDOs) the correlation between defaults changes through time and is often driven

by several interacting macro-economic variables. These may not directly vary historically with the asset price, therefore the interaction is 'hidden' in the volatility component. See Berd *et al* (2005) [27], who utilize a threshold approach to modelling the volatility of synthetic CDOs.

Practical approaches to constructing a simple multi-asset multi-volatility are a relatively recent phenomenon, articles by Gourieroux (2005) [113], Gourieroux and Sufana (2003 [117], 2004 [118]) utilize the Wishart distribution to attempt to create a fully featured multivariate volatility model with closed form solutions. The model proposed here does not have a closed form solution, in keeping with many other stochastic volatility models.

The general approach to modelling multivariate risk would be to treat each of the elements of the covariance matrix as unique diffusion process. However the resulting covariance matrix will not be guaranteed to be non-negative definite and as such may have negative eigenvalues. There are well defined approaches to generating guaranteed non-negative definite matrices using a variety of matrix transforms, such as the *ivech* transform utilized in Bollerslev *et al* (1988) [41] and Engle and Kroner (1994) [89]. One of the problems with this approach is that the elements of the factorized covariance matrix are functionally quite complex, particularly as the

number of elements increases. The system presented in this paper is based entirely on functional products and square roots of standard integral stochastic processes and it is is algorithmically scalable, albeit algebraically inelegant.

The article is divided into three sections. The first section presents a review of generalized stochastic volatility models. In the second, the derivation of a matrix stochastic volatility model is presented, first as a basic autoregressive model in discrete time and then as a full continuous time model, for which the univariate form is a generalization of the standard stochastic volatility approaches. The final section addresses a hedging problem in the presence of stochastic covariance and demonstrates why the multivariate approach requires a different hedging strategy to fully encompass each of the diffusions acting on the observed process.

8.2.1 A Representation of Stochastic Volatility

Consider the evolution of a k-dimensional vector stochastic process $\mathbf{x}(t)$, where $\mathbf{x}(t) \in \mathbb{R}^k$, defining the elements of $\mathbf{x}(t)$ as,

$$Y(t) = \log S(t) \tag{8.2.1}$$

$$\mathbf{z}(t) = \left[Z^1(t), Z^2(t), \ldots Z^{k-1}(t)\right]^{\mathrm{T}} \tag{8.2.2}$$

$$\mathbf{x}(t) = \left[Y(t), \mathbf{z}(t)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{8.2.3}$$

Where $S(t)$, is the observed price evolution of a single asset and $\mathbf{z}(t)$ is a $k-1$ vector of volatility processes, where $\mathbf{z}(t) \in \mathbb{R}^{k-1}$. A typical generic stochastic volatility model maybe set out as follows, first the process $\mathbf{x}(t)$ is a multidimensional stochastic differential equation,

$$d\mathbf{x}(t) = \mu(\mathbf{x}(t)|\theta)\,dt + \sigma(\mathbf{x}(t)|\theta)\,d\mathbf{w}(t) \tag{8.2.3}$$

where $\mathbf{w}(t) = \left[dW^1(t), dW^2(t), ..., dW^k(t)\right]^{\mathrm{T}}$ is a vector of independent Brownian motions, where for every $t, h \geq 0, \mathbf{w}_{t+h} - \mathbf{w}_t$, has a Gaussian distribution specified as,

$$\mathbf{w}(t+h) - \mathbf{w}(t) \sim N(\mathbf{0}, \Xi_h) \tag{8.2.4}$$

$$\Xi_h = \begin{cases} [\xi_{i,j}]_{i=j} = h \\ [\xi_{i,j}]_{i\neq j} = 0 \end{cases} \tag{8.2.5}$$

Based on this structure, the covariation between the stochastic processes will be conditional on the structure of $\sigma(\mathbf{x}(t)|\theta)$. Assuming that the functions, $\mu(\mathbf{x}(t)|\theta)$ and $\sigma(\mathbf{x}(t)|\theta) = \Psi(\mathbf{x}(t))$ are simple matrix functions,

$$\mu(\mathbf{x}(t)|\theta) = \Phi(\mathbf{x}(t)) \tag{8.2.6}$$

$$\sigma(\mathbf{x}(t)|\theta) = \Psi(\mathbf{x}(t)) \tag{8.2.7}$$

Then for a standard two dimensional problem, i.e 1 asset price process and 1 volatility process, the vector process may be specified as,

$$d\begin{bmatrix} Y(t) \\ Z(t) \end{bmatrix} = \begin{bmatrix} \phi_S Y(t) \\ \phi_Z Z(t) \end{bmatrix} dt + \begin{bmatrix} \psi_{1,1}(Y(t), Z(t)) & \psi_{1,2}(Y(t), Z(t)) \\ 0 & \psi_{2,2}(Y(t), Z(t)) \end{bmatrix} \begin{bmatrix} dW^1(t) \\ dW^2(t) \end{bmatrix} \tag{8.2.7}$$

where,

$$\mathbf{\Phi}(\mathbf{x}(t)) = \begin{bmatrix} \phi_S & 0 \\ 0 & \phi_Z \end{bmatrix} \begin{bmatrix} Y(t) \\ Z(t) \end{bmatrix} \tag{8.2.8}$$

$$\mathbf{\Psi}(\mathbf{x}(t)) = \begin{bmatrix} \psi_{1,1}(Y(t)) & \psi_{1,2}(Z(t)) \\ 0 & \psi_{2,2}(Z(t)) \end{bmatrix} \tag{8.2.9}$$

Many of continuous time stock price models maybe derived using this approach, such as the Heston (1993) [144] model,

$$\phi_S = (r-d) \tag{8.2.10}$$

$$\phi_Z = \kappa(\nu - Z(t)) \tag{8.2.11}$$

$$\psi_{1,1}(Y(t), Z(t)) = \left((1-\rho)^2 Y(t)\right)^{\frac{1}{2}} \tag{8.2.12}$$

$$\psi_{1,2}(Y(t), Z(t)) = \rho(Y(t))^{\frac{1}{2}} \tag{8.2.13}$$

$$\psi_{2,2}(Y(t), Z(t)) = \sigma(Y(t))^{\frac{1}{2}} \tag{8.2.14}$$

where, $r-d$ is the real risk free rate and model parameterization proceeds via estimation of the parameter vector θ,

$$\theta = [\nu, \rho, \sigma, \kappa]^{\mathrm{T}} \tag{8.2.14}$$

This is a square root diffusion model and operates under the assumption that $Y(t) \in \mathbb{R}^+$. An interesting observation is that if $\sigma = 0$ and $Y_0 = \nu$, then the model reverts to a Black and Scholes (1973) [31] type approach.

8.3 The Multivariate Approach

Consider a stochastic process with k assets, in order to fully describe the instantaneous quadratic covariation of those assets, $\frac{1}{2}k(k+1)$, pieces of information are required, k variances and $\frac{1}{2}k(k-1)$ correlations, to fully describe the volatility of the system. Structurally the instantaneous covariance matrix, $\mathbf{\Sigma}_t$ maybe decomposed as follows,

$$\mathbf{\Sigma}_t = \mathbf{A}_t \mathbf{A}_t^{\mathrm{T}} \tag{8.3.0}$$

Where the $k \times k$ matrix $\mathbf{A}_t$, is composed of $\frac{1}{2}k(k+1)$ unique elements.

$$\mathbf{A}_t = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & a_{k+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_k & a_{k+(k-1)} & \cdots & a_{\frac{1}{2}k(k+1)} \end{bmatrix} \tag{8.3.0}$$

If $\mathbf{A}_t = ivech(\mathbf{a}_t)$ and $\mathbf{a}_t \in \mathbb{R}^{\frac{1}{2}k(k+1)}$, then $\mathbf{\Sigma}_t \in \mathbb{C}^{k \times k}$, where $\mathbb{C}^{k \times k}$ is the set of all non-negative Hermitian matrices. Therefore the Upper Triangular Matrix Square Root, (UTMSR) $\mathbf{\Sigma}_t^{\frac{1}{2}} = chol(\mathbf{\Sigma}_t)$ exists within the kernel of $\mathbf{\Sigma}_t$ and the eigenvalues of $\mathbf{\Sigma}_t$, $\varsigma_i \in \{\varsigma_1, ..., \varsigma_k\}$ are all non-negative and *chol* is the functional result of the Cholesky-Banachiewicz algorithm. Using the *ivech* approach is a useful method of specifying the structure of a conditional covariance matrix and allows an algebraic representation of the elements of the UTMSR of $\mathbf{\Sigma}_t$.

8.3.1 A Discrete Time Autoregressive Model

Consider the discrete time model of the evolution of a k-dimensional process, $\mathbf{u}_t$ where, $\mathbf{u}_t \in \mathbb{R}^k$, $\mathbf{a}_t \in \mathbb{R}^{\frac{1}{2}k(k+1)}$ and ϵ_t and ξ_t are uncorrelated noise drawn from a

zero centered unit variance normal distribution. Using some of the results from the derivation of the classical Bartlett decomposition, see Kollo and von Rosen (2005) [170]a simple model of stochastic covariance is as follows,

$$\mathbf{u}_t = \boldsymbol{\Sigma}_t^{\frac{1}{2}} \varepsilon_t \tag{8.3.1}$$

$$\boldsymbol{\Sigma}_t = \mathbf{A}_t \mathbf{A}_t^{\mathrm{T}} \tag{8.3.2}$$

$$\mathbf{A}_t = ivech\left(\mathbf{a}_t\right) \tag{8.3.3}$$

$$\mathbf{a}_t = \boldsymbol{\Pi} \mathbf{a}_{t-1} + \boldsymbol{\Omega}^{\frac{1}{2}} \xi_t \tag{8.3.4}$$

$$\left[\varepsilon_t^{\mathrm{T}}, \xi_t^{\mathrm{T}}\right]^{\mathrm{T}} \sim N\left(\mathbf{0}, \underset{\frac{1}{2}(k(k+3)) \times \frac{1}{2}(k(k+3))}{\mathbf{I}}\right) \tag{8.3.5}$$

Where, $\boldsymbol{\Pi}$ is a $\frac{1}{2}k\left(k+1\right) \times \frac{1}{2}k\left(k+1\right)$ parameter matrix and the disturbances are independent zero centered Gaussian processes with unit variance, if $\boldsymbol{\Pi} = \mathbf{I}$ then the process underlying $\mathbf{a}_t$ is a multidimensional random walk. The static covariance matrix, $\boldsymbol{\Omega}$ maybe estimated using a standard Wishart approach, the moments of the conditional covariance matrix is therefore,

$$\mathbf{u}_t = \left(ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1} + \boldsymbol{\Omega}^{\frac{1}{2}} diag\left(\xi_t\right)\right) ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1} + \boldsymbol{\Omega}^{\frac{1}{2}} diag\left(\xi_t\right)\right)\right)^{\frac{1}{2}} \varepsilon_t \tag{8.3.6}$$

$$E\left(\boldsymbol{\Sigma}_{t|t-1}\right) = ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}\right) ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}\right)^{\mathrm{T}} \tag{8.3.7}$$

Therefore the quadratic variation of the covariance matrices is,

$$E\left(vec\boldsymbol{\Sigma}_{t|t}vec\boldsymbol{\Sigma}_{t|t}^{\mathrm{T}}\right) = vec\left(ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}+\boldsymbol{\Omega}^{\frac{1}{2}}diag\left(\mathbf{I}\right)\right)ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}+\boldsymbol{\Omega}^{\frac{1}{2}}diag\left(\mathbf{I}\right)\right)^{\mathrm{T}}\right)$$
$$vec\left(ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}+\boldsymbol{\Omega}^{\frac{1}{2}}diag\left(\mathbf{I}\right)\right)ivech\left(\boldsymbol{\Pi}\mathbf{a}_{t-1}+\boldsymbol{\Omega}^{\frac{1}{2}}diag\left(\mathbf{I}\right)\right)^{\mathrm{T}}\right)^{\mathrm{T}} \tag{8.3.6}$$

Specifying the m^{th} moment structure of the observed random vector is,

$$\mathbf{M}_4 = E\left(\mathbf{u}_t\left(\mathbf{u}_t^{\mathrm{T}}\right)^{\otimes m-1}\right) \tag{8.3.6}$$

Therefore the fourth moment of the resulting process is,

$$\underset{k^2\times k^2}{\mathbf{M}_4} = E\left(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\otimes\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\right) \tag{8.3.6}$$

setting,

$$\boldsymbol{\Psi}_t = E\left(vec\boldsymbol{\Sigma}_{t|t-1}vec\boldsymbol{\Sigma}_{t|t-1}^{\mathrm{T}}\right) \tag{8.3.7}$$

$$\mathbf{H}_t = \mathbf{u}_t\mathbf{u}_t^{\mathrm{T}} \tag{8.3.8}$$

$$\mathbf{h}_t = vec\mathbf{H}_t \tag{8.3.9}$$

and expressing the expectation of the fourth moment in terms of the model components yields,

$$E\left(\varepsilon_t\varepsilon_t^{\mathrm{T}}\otimes\varepsilon_t\varepsilon_t^{\mathrm{T}}\right) = \boldsymbol{\Phi} \tag{8.3.10}$$

$$E\left(\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\otimes\mathbf{u}_t\mathbf{u}_t^{\mathrm{T}}\right) = \boldsymbol{\Phi}^{\frac{1}{2}}\boldsymbol{\Psi}_t\left(\boldsymbol{\Phi}^{\frac{1}{2}}\right)^{\mathrm{T}} \tag{8.3.11}$$

The fourth moment of a zero centered normal distribution is defined in Kollo and Von Rosen (2005) [170]. The simplest method is to take the distribution of each element and minimize the following score function over the fourth moment,

$$\mathfrak{F}_{t|t-1}(\theta) = \left|\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}_t\left(\mathbf{\Phi}^{\frac{1}{2}}\right)^{\mathrm{T}}\right|^{-\frac{1}{2}} + \exp\left(-\tfrac{1}{2}\mathbf{h}_t^{\mathrm{T}}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}_t\left(\mathbf{\Phi}^{\frac{1}{2}}\right)^{\mathrm{T}}\mathbf{h}_t\right) \tag{8.3.11}$$

over the sample period $t \in [1, ..., \tau]$, it is simple to show that the maxima of this function attains estimates of $\mathbf{\Pi}$ and $\mathbf{\Omega}$, under the standard asymptotic normality assumptions.

$$\mathfrak{L}(\theta) = -\tfrac{1}{2}\log\left(\prod_{t=1}^{\tau}\left|\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}_t\left(\mathbf{\Phi}^{\frac{1}{2}}\right)^{\mathrm{T}}\right| + \exp\left(-\tfrac{1}{2}\mathbf{h}_t^{\mathrm{T}}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Psi}_t\left(\mathbf{\Phi}^{\frac{1}{2}}\right)^{\mathrm{T}}\mathbf{h}_t\right)\right) \tag{8.3.11}$$

The proposed approach maybe reformulated by including the following decomposition, $\hat{\mathbf{u}}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1,\Omega_t}$. For example a standard linear first order autoregressive model,

$$\mathbf{y}_t = \mathbf{F}\mathbf{y}_{t-1} + \mathbf{\Sigma}_t^{\frac{1}{2}}\varepsilon_t \tag{8.3.12}$$

$$\mathbf{a}_t = \mathbf{\Pi}\mathbf{a}_{t-1} + \mathbf{\Omega}^{\frac{1}{2}}\xi_t \tag{8.3.13}$$

$$\mathbf{\Sigma}_t = \mathbf{A}_t\mathbf{A}_t^{\mathrm{T}} \tag{8.3.14}$$

$$\mathbf{A}_t = ivech(\mathbf{a}_t) \tag{8.3.15}$$

This type of model appears to have a great deal of applications to the evaluation of multivariate risk analysis, in particular for the analysis of large portfolios of heterogenous assets. Interestingly the stationarity conditions for the first order model are extremely simple, i.e.

$$\mathbb{P}\left(\sum_{t=1}^{\tau}\operatorname{tr}\mathbf{\Sigma}_t < \infty\right) = 1 \tag{8.3.15}$$

and,

$$\mathbb{P}\left(\sum_{t=1}^{\tau}\mathrm{tr}\left(\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}\right)<\infty\right)=1 \tag{8.3.15}$$

are simply dependent on the eigenvalues of the square matrices $\mathbf{\Pi}$ and $\mathbf{F}$ lying within the unit circle.

8.4 A Continuous Time Approach

The previous section demonstrated an approach to the specification and estimation of a simple discrete time model, where the degree of quadratic covariation of the disturbance terms follows a simple discrete time multivariate matrix stochastic process. However, generally continuous time stochastic processes are more useful in the financial engineering context. Recalling the stochastic volatility model demonstrated in 8.2.1, a multivariate continuous time analogue along the lines of the model illustrated in 8.3.12, should optimally simplify to 8.2.1 when $k=1$. The following multivariate model is proposed,

$$\mathbf{y}(t) = \left[\log S^1(t), \log S^2(t), ..., \log S^1(t)\right]^{\mathrm{T}} \tag{8.4.1}$$

$$\mathbf{z}(t) = \left[Z^1(t), Z^2(t), ..., Z^1(t)\right]^{\mathrm{T}} \tag{8.4.2}$$

Where $\mathbf{y}(t)$ is a k-dimensional vector of log stock prices $S^i(t)$, whose evolution is defined by a multidimensional stochastic differential equation. Setting, $\mathbf{x}(t) = [\mathbf{y}(t), \mathbf{z}(t)]^{\mathrm{T}}$ and recalling,

$$d\mathbf{x}(t) = \mu(\mathbf{x}(t)|\theta)\,dt + \sigma(\mathbf{x}(t)|\theta)\,d\mathbf{w}(t) \tag{8.4.2}$$

We define the terms as two matrix functions,

$$\mu(\mathbf{x}(t)|\theta)\,dt = \mathbf{\Phi}(\mathbf{x}(t)) \tag{8.4.3}$$

$$\sigma(\mathbf{x}(t)|\theta) = \mathbf{\Omega}(\mathbf{x}(t))\,\mathbf{\Psi}(\mathbf{x}(t)) \tag{8.4.4}$$

For simplicity the following matrix operator is used,

$$\mathbf{\Omega}(\mathbf{x}(t)) = idiag(\mathbf{x}(t)) = \begin{bmatrix} X^1(t) & 0 & \cdots & 0 \\ 0 & X^2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X^2(t) \end{bmatrix} \tag{8.4.4}$$

Again for simplicity it is assumed that the processes underling $\mathbf{x}(t)$ and $\mathbf{z}(t)$ are not linked via the first term, $\mu(\mathbf{x}(t)|\theta)$.

$$\mathbf{\Phi}(\mathbf{x}(t)) = \begin{bmatrix} \mathbf{\Phi}^{\mu} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}^{\sigma} \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix} \tag{8.4.5}$$

$$\mathbf{\Psi}(\mathbf{x}(t)) = \begin{bmatrix} \mathbf{\Omega}\left(d\mathbf{\Sigma}(t)^{\frac{1}{2}}, \mathbf{x}(t)\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}^{\frac{1}{2}} \end{bmatrix} \tag{8.4.6}$$

Again the *ivech* operator is utilized to generate the instantaneous covariance matrix $\mathbf{\Sigma}(t)$,

$$d\mathbf{\Sigma}(t) = ivech(d\mathbf{z}(t))\,ivech(d\mathbf{z}(t))^{\mathrm{T}} \tag{8.4.6}$$

Again simplifying the covariance process,

$$\mathbf{\Omega}\left(d\mathbf{\Sigma}(t)^{\frac{1}{2}}, \mathbf{x}(t)\right) = d\mathbf{\Sigma}(t)^{\frac{1}{2}} \tag{8.4.6}$$

The general sde for the system is as follows,

$$d\begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}^{\mu} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}^{\sigma} \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix} dt + \tag{8.4.7}$$

$$\begin{bmatrix} idiag(\mathbf{y}(t)) & \mathbf{0} \\ \mathbf{0} & idiag(\mathbf{z}(t)) \end{bmatrix} \left(\begin{bmatrix} d\mathbf{\Sigma}(t)^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} d\mathbf{w}^{\mu}(t) \\ d\mathbf{w}^{\sigma}(t) \end{bmatrix} \right)$$

The Brownian motion is partitioned as follows,

$$\mathbf{w}(t) = \left[\mathbf{w}^{\mu}(t)^{\mathrm{T}}, \mathbf{w}^{\mu}(t)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{8.4.6}$$

The system may now be rewritten as a pair of vector processes,

$$d\mathbf{y}(t) = \mathbf{\Phi}^{\mu}\mathbf{y}(t)\,dt + \mathbf{\Omega}(\mathbf{y}(t))\,d\mathbf{\Sigma}(t)^{\frac{1}{2}}\,d\mathbf{w}^{\mu}(t) \tag{8.4.7}$$

$$d\mathbf{z}(t) = \mathbf{\Phi}^{\sigma}\mathbf{z}(t)\,dt + \mathbf{\Omega}(\mathbf{z}(t))\,\mathbf{\Psi}^{\frac{1}{2}}d\mathbf{w}^{\sigma}(t) \tag{8.4.8}$$

Using the Cholesky-Banachiewicz approach to factorizing non-negative definite hermitian matrices, the $k = 2$ model is illustrated,

$$\begin{aligned} d\begin{bmatrix} Y^1(t) \\ Y^2(t) \end{bmatrix} &= \begin{bmatrix} \phi_{1,1}Y^1(t) + \phi_{1,2}Y^2(t) \\ \phi_{2,1}Y^1(t) + \phi_{2,2}Y^2(t) \end{bmatrix} dt \\ &+ \begin{bmatrix} Y^1(t) & 0 \\ 0 & Y^2(t) \end{bmatrix} \\ &\times \left(\begin{bmatrix} \left(\left(Z^1(t)\right)^2 + \left(Z^2(t)\right)^2\right)^{\frac{1}{2}} & \frac{\left(Z^1(t)Z^2(t)\right)}{\left(\left(Z^1(t)\right)^2 + \left(Z^2(t)\right)^2\right)^{\frac{1}{2}}} \\ 0 & \frac{\left(Z^1(t)^2 Z^3(t)^2\right)^{\frac{1}{2}}}{\left(\left(Z^1(t)\right)^2 + \left(Z^2(t)\right)^2\right)^{\frac{1}{2}}} \end{bmatrix} \begin{bmatrix} dW^{\mu,1}(t) \\ dW^{\mu,2}(t) \end{bmatrix} \right) \end{aligned} \tag{8.4.9}$$

The structure of the volatility model is therefore,

$$\begin{aligned} d\begin{bmatrix} Z^1(t) \\ Z^2(t) \\ Z^3(t) \end{bmatrix} &= \begin{bmatrix} \phi_{1,1}Z^1(t) + \phi_{1,2}Z^2(t) + \phi_{1,3}Z^3(t) \\ \phi_{2,1}Z^1(t) + \phi_{2,2}Z^2(t) + \phi_{2,3}Z^3(t) \\ \phi_{3,1}Z^1(t) + \phi_{3,2}Z^2(t) + \phi_{3,3}Z^3(t) \end{bmatrix} dt \\ &+ \begin{bmatrix} Z^1(t) & 0 & 0 \\ 0 & Z^2(t) & 0 \\ 0 & 0 & Z^3(t) \end{bmatrix} \left(\begin{bmatrix} \psi_{1,1} & \psi_{1,2} & \psi_{1,3} \\ 0 & \psi_{2,2} & \psi_{2,3} \\ 0 & 0 & \psi_{3,3} \end{bmatrix} \begin{bmatrix} dW^{\sigma,1}(t) \\ dW^{\sigma,2}(t) \\ dW^{\sigma,2}(t) \end{bmatrix} \right) \end{aligned} \tag{8.4.8}$$

Consider the general problem of pricing a call option in the presence of stochastic volatility, SV, the general approach is to form two separate hedges for each Brownian motion underlying the price of the asset, see Rebonato (2002) [223], Musiela and

Rukowski (2004) [203]for overviews. However the appropriate specification of the volatility model has been the subject of some considerable discussion in the recent technical literature. Most approaches utilize a two dimensional Brownian motion approach. If the stock price dynamics exhibit some form of stochastic quadratic covariation then the the volatility process cannot in itself be explained by some arbitrary volatility process, but the dynamics must include the contribution of all of the other elements in the system. This is in essence the transmission of volatility from one component to another by a relative change in the level of correlation, correlation risk. The type of model suggested in 8.4.8, implicitly includes this type of risk and in effect once the diffusions underlying this extra risk are identified and they can be Delta hedged out in the appropriate manner.

8.5 Application, Portfolio Replication with Stochastic Covariation

Utilizing the bi-variate model a method of formulating a self replicating hedge is demonstrated, utilizing the stochastic covariation method. In this method a replicating hedge is constructed that includes the drift element of the price processes, the stochastic volatility and the off diagonal stochastic covariation. Let the price and covariance processes for three assets be as in 8.4.8, the value of a properly priced call option on the first asset will therefore be,

$$C(t) = C\left(Y^1(t), Z^1(t), Z^2(t), Z^3(t), t\right) \tag{8.5.0}$$

Where $C : \mathbb{R}^d \rightarrow \mathbb{R}$, rewriting as a vector function yields,

$$\mathbf{x}(t) = \left[Y^1(t), Z^1(t), Z^2(t), Z^3(t)\right]^{\mathrm{T}} \tag{8.5.1}$$

$$C(t) = C(\mathbf{x}(t)) \tag{8.5.2}$$

Taking a Taylor expansion around a fixed price, C_0,

$$\begin{aligned} \hat{C}(t, \mathbf{x}(t)) &= C_0 + \frac{\partial C}{\partial t} + \nabla C(\mathbf{x}(t))^{\mathrm{T}} E(\mathbf{x}(t)) \\ &\quad + \tfrac{1}{2}\left(E(\mathbf{x}(t))^{\mathrm{T}} \nabla^2 C(\mathbf{x}(t)) E(\mathbf{x}(t))^{\mathrm{T}}\right) \end{aligned} \tag{8.5.3}$$

In order to hedge out the Brownian motions in the volatility processes, a vector of volatility hedging instruments, $\mathbf{v}(t)$, is required,

$$\mathbf{v}(t) = \left[V^1(t), V^2(t), V^3(t)\right]^{\mathrm{T}} \tag{8.5.2}$$

setting, $V^i(t) = V^i(Z^i(t), dW^i(t), t)$, then the portfolio value of the stock, a bond, the call option and volatility instruments is as follows,

$$\mathbf{s}(t) = \left[Y^1(t), B(t), C(t), \mathbf{v}(t)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{8.5.2}$$

for a given set of weights, ω, the portfolio value is therefore,

$$\pi(t) = \omega^{\mathrm{T}} \mathbf{s}(t) \tag{8.5.3}$$

$$\pi(t) \triangleq \max_{\omega} \left(\omega^{\mathrm{T}} \mathbf{s}(t)\right) \tag{8.5.4}$$

In essence the Brownian motion underlying the stock price and each of the Brownian motions underlying the volatility processes need to be delta hedged out, if only one

extra hedging instrument is included then the replication will not be complete and value may bleed in (or more likely out) of the hedge, see Rebonato (2003) [223].

Applications to Portfolio Hedging

A major advancement of the proposed model is the ability to derive analytical solutions to the weights that should be applied to the volatility instruments for each of the hedges and a schema for assessing how many volatility instruments should be included in hedging model, something which has not been addressed to any significant extent in the option pricing literature, in fact Sahalia and Kimmel (2006) [2], briefly mention this problem and conclude that it should be a system identification issue. However using this approach the number of hedging instruments is immediately apparent and the only system identification issue is in the inclusion issue of how many assets should be included in the system, estimation of the system via 8.3.1, will automatically include a restriction test using a likelihood ratio type test. This methodology would appear to be a significant advancement over the current ad-hoc approaches currently suggested in the literature.

8.6 Numerical Simulations

Figures 8.1 to 8.1 demonstrate a single path from a stochastic covariance simulation for a tri-variate system, the system includes an exogenous driving variable, which is a single interest rate pathway generated by the Williams and Ioannidis (2006) [263]

regime switching rate path method.

8.7 Chapter Concluding Remarks

This chapter has taken much of the ground work carried out in previous chapters and formulated an innovative approach to modelling multivariate stochastic volatility models. The major innovation is way in which the price processes and hedging maybe integrated into a unified replicating framework, which allows for the robust identification of appropriate hedging instruments and their weighting. This is a very useful approach as it implicitly endogenizes volatility hedging in the manner of conventional "out of model" vega hedging approaches. General extensions of this approach are the inclusion of non-linear functions in the asset price models to mimic the smile surfaces observed empirically in option prices, see Musiela and Rutkowski (2005) [203] for discussion.

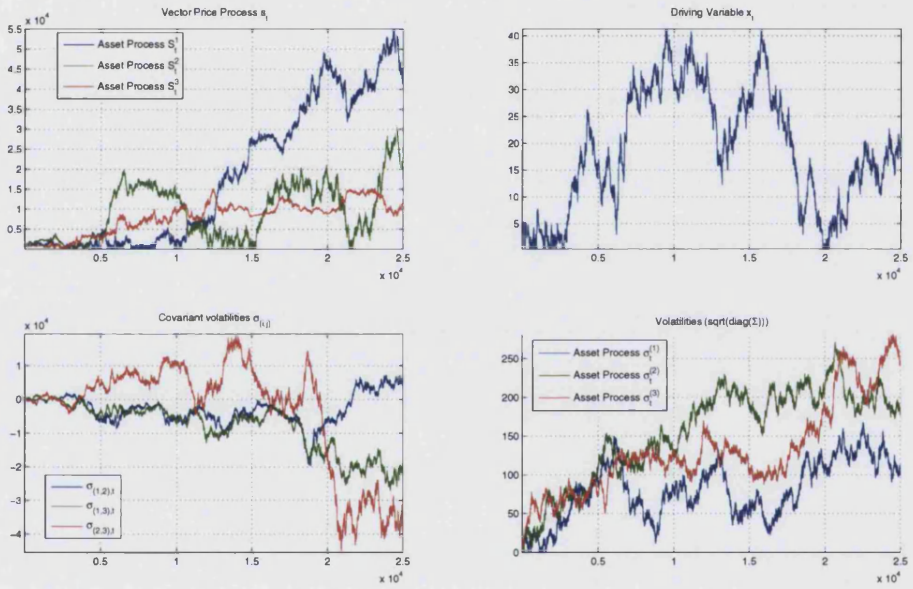

Figure 8.1: A Simulation of Three price processes, with Stochastic Covariance, the simulation demonstrates that the degree of quadratic covariation may even swap signs as the processes evolve.

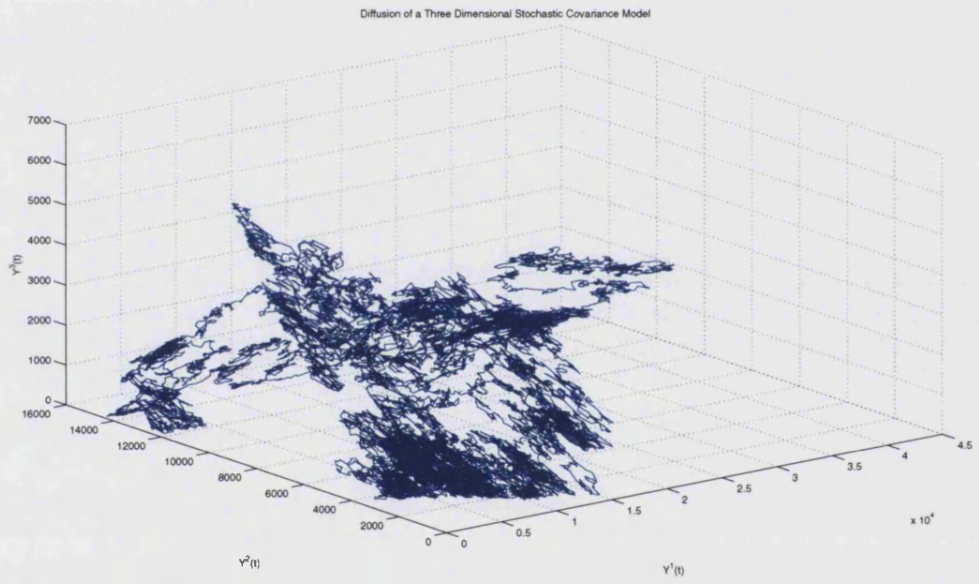

Figure 8.2: The diffusion of the three price process in three dimensional space.

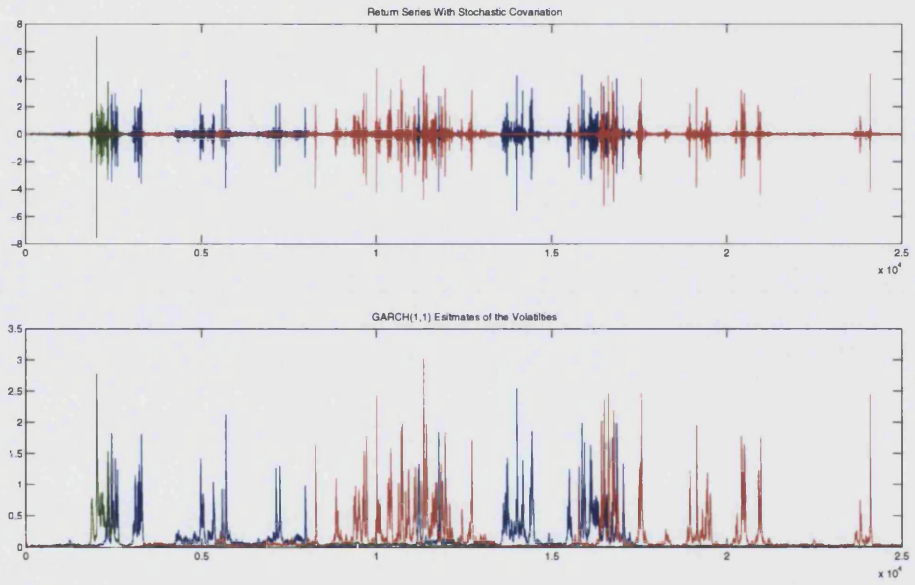

Figure 8.3: The resultant return series of the three processes from 8.1

Concluding Remarks

This final section summarizes the core contribution to the literature of each of the preceding chapters and suggests some avenues for future work in this area.

8.7.1 Notes on Chapter 1

Chapter 1 reviewed the current multivariate time series literature and introduced several new approaches to recursive modelling of dynamic systems. Abadir and Magnus (2002) suggest a standardized notation framework for econometrics and this framework provides the basis for the notation structure of this and subsequent chapters.

8.7.2 Notes on Chapter 2

Chapter 2 reviewed the current MV-GARCH literature and demonstrated some empirical applications for these types of model. The chapter only briefly touched upon the asymptotic theory underlying these models and this is currently the main area of interest in this field. Future work needs to be directed at designing robust asymptotic tests of model fit to increase the robustness of diagnostics over and above the current information matrix, Wald, LM and LR test suite available,

8.7.3 Notes on Chapter 3

Chapter 3 reviewed the copula literature and suggested several new approaches to monte-carlo simulation utilizing this extremely useful toolkit for designing multivariate distributional models. The empirical copula suggested in the latter part of the chapter offers a very powerful and flexible tool for "Black Box" analysis of multivariate data.

8.7.4 Notes on Chapter 4

Chapter 4 is a case study of dynamic dependency in the UK biotechnology sector and lays some of the stylistic evidence of non-static correlation in asset returns. The biotechnology sector is a useful sector to study as the news and information updating process is fairly transparent and allows for the design of very simple event study experiments to test some of the stylized assumptions used in later chapters.

8.7.5 Notes on Chapter 5

Chapter 5 provides a new type of regime switching model of dynamic covariance and demonstrates that this model is one of a class of regime models. The model's major attributes are its ability to comfortably deal with very high-variate systems. This type of model is extremely useful in risk management applications and provides a solid method of analyzing potential multivariate volatility risk. The chapter does not fully develop the asymptotic theory underlying this type of model and this is an area

for future research.

8.7.6 Notes on Chapter 6

Chapter 6 covers the inclusion of multivariate higher moments in the asset allocation problem. The key goal of the chapter was to demonstrate an innovative toolbox of methods which may be used when standard distributional assumptions do not fully encapsulate the properties of asset returns.

8.7.7 Notes on Chapter 7

The final chapter addressed continuous time hedging in presence of multivariate stochastic volatility and adds to the growing literature on stochastic quadratic covariation models. The chapter's main goals are the specification and estimation of a new asset price model which includes stochastic covariation implicitly within it's structure and the impact that this has on portfolio replication strategies. The main are of development in this area is the inclusion of more sophisticated pricing structures and direct development of models of CDO and other credit based contingent claims, for which this model appears ideally suited.

Bibliography

[1] Y. Ait-Sahalia, *Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach*, Econometrica **70** (2002), 223–262.

[2] Y. Ait-Sahalia and R. Kimmel, *Maximum likelihood estimation of stochastic volatility models*, Journal of Financial Economics (2006).

[3] Y. Ait-Sahalia, P. Mykland, and L. Zhang, *How often to sample a continuous-time process in the presence of market microstructure noise*, Review of Financial Studies **18** (2005), 351–416.

[4] C. Alexander, *Orthogonal methods for generating large positive semi-definite covariance matrics*, Mimeo, Reading University **www.ismacentre.rdg.ac.uk/pdf/discussion/DP2000-06.pdf** (1997).

[5] ———, *Market models*, Wiley, London (2000).

[6] C. Alexander and A. Chibumba, *Multivariate orthogonal factor garch*, Mimeo, Reading University (2001).

[7] T. Andersen, T. Bollerslev, F. Diebold, and P.Labys, *Modeling and forecasting realized volatility*, Econometrica (1963).

[8] B.D.O. Anderson and J.B. Moore, *Optimal filtering*, Prentice-Hall (1979).

[9] C. A. Andersson and R. Bro., *The n-way toolbox for matlab.*, Chemom.Intell.Lab.Syst. **52(1)** (2000), 1–5.

[10] G. E. M. Anscombe, *Intention*, Harvard University Press; 2nd edition (1991).

[11] G. Athayde and R. Flores, *Finding a maximum skewness portfolio*, Computing in Economics and FInance (2004).

[12] O. Attanasio, *Risk time-varying second moments and market efficiency*, Review of Economic Studies (1996), 479494.

[13] M. El Babsiri and J.-M. Zakoian, *Contemporaneous asymmetry in garch processes*, Journal of Econometrics (1982), 257294.

[14] R. Baillie, T. Bollerslev, and H. Mikkelsen, *Fractionally integrated generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics (2003), 330.

[15] F. Bandi and P.C.B. Phillips, *Nonstationary continuous-time models*, Handbook of Financial Econometrics (2002).

[16] P. Bank and Hans Fllmer, *American options, multiarmed bandits, and optimal consumption plans: A unifying view*, Springer Berlin / Heidelberg (2003).

[17] N. Barberis, M. Huang, and R. Thaler, *Individual preferences, monetary gambles and the equity premium*, Mimeo, NBER working paper series (2003).

[18] N. Barberis and A. Schleifer, *Style investing*, Journal of Financial Economics (2002), 161–199.

[19] N. Barberis, A. Schleifer, and J. Wurgler, *Comovement*, Mimeo, NBER working paper series (2001), http://www.nber.org/papers/w8895.pdf.

[20] O. Barndorff-Nielsen and N. Shephard, *Normal modified stable processes theory of probability and mathematics statistics*, Mimeo, Nuffield College, Oxford **http://www.nuff.ox.ac.uk/economics/papers/2001/w6/nmsprocnew1.pdf** (2003), 119.

[21] L. Bauwens and S. Laurent, *A new class of multivariate skew densities with application to garch models*, Journal of Business and Economic Statistics, Alexandria, US **23/3** (2005), 346–354.

[22] L. Bauwens, S. Laurent, and J. Rombouts, *Multivariate garch models a survey*, Journal of Applied Econometrics, Wiley **21, Issue 1** (2006), 79 – 109.

[23] L. Bauwens and J. Rombouts, *Bayesian clustering of similar garch models*, Working paper (2002), http://ideas.repec.org/p/ecm/nawm04/370.html.

[24] A. Bera, P. Garcia, and J. Roh, *Estimation of time-varying hedging ratios for corns and soybeans bgarch and random co-efficient approaches*, Sankhya (2003), 346368.

[25] A. Bera and M. Higgins, *Arch models properties: Estimation and testing*, Journal of Economic Surveys (2002), 305362.

[26] A. Bera and S. Kim, *Testing constancy of correlation and other specificationsof the bgarch model with an application to international equity returns*, Journal of Empirical Finance (1997), 171195.

[27] A. Berd, R. Engle, and A. Voronov, *The underlying dynamics of credit correlations*, Available at SSRN: http://ssrn.com/abstract=837824 (2005).

[28] M. Billio, M. Caporin, and M. Gobbo, *Block dynamic conditional correlation multivariate garch models*, University of California at San Diego, Economics Working Paper Series (2001), http://www.greta.it/italiano/pagine/PdfFile/03.03.PDF.

[29] F. Black, *Studies of stock market volatility changes*, Proceedings of the American Statistical Association (1990), 177181.

[30] F. Black, E. Derman, and W. Toy, *A one factor model of interest rates and its application to treasury bond options*, Financial Analysts Journal (1986).

[31] F. Black and M. S. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy **81** (1973), 637–654.

[32] T. Bollerslev, *Generlaized autoregressive conditional heteroskedasticity*, Journal of Econometrics **31** (1986).

[33] ———, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics (1990), 307327.

[34] ———, *Modeling the coherence in short-run nominal exchange rates a multivariate generalized arch model*, Review of Economics and Statistics (1992), 498505.

[35] T. Bollerslev, R. Chou, and K. Kroner, *Arch modeling in finance: A review of the theory and empirical evidence*, Journal of Econometrics (1988), 559.

[36] T. Bollerslev and R. Engle, *Common persistencein conditional variances*, Econometrica (1970), 167186.

[37] T. Bollerslev, R. Engle, and D. Nelson, *Arch models*, in Handbook of Econometrics (1988), 29593038.

[38] T. Bollerslev, R. Engle, and J.M. Woodridge, *A capital-asset pricing model with time-varying coefficients*, Journal of Political Economy (1988).

[39] T. Bollerslev, R. Engle, and J. M. Wooldbridge, *A capital asset pricing model with time varying covariances*, Journal of Political Economy (1994), 116–131.

[40] T. Bollerslev, R. Engle, and J. Wooldridge, *A capital asset pricing mode1 with time varying covariances*, Journal of Political Economy (1993), 116131.

[41] T. Bollerslev and J. Wooldridge, *Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances*, Econometric Reviews (1992), 143 172.

[42] F. Boussama, *Asymptotic normality for the quasi-maximum likelihood estimator of a garch model*, Comptes Rendus de l'Academie des Sciences **331 (Serie I)** (2000), 81–84.

[43] G. Box and G. Jenkins, *Time series analysis*, Forecasting and Control (2003).

[44] C. Brooks, S. Burke, and G. Persand, *Multivariate garch models software choice and estimation issues*, forthcoming in Journal of Applied Econometrics. (2000).

[45] C. Brunetti and C. Gilbert, *Bivariate figarch and fractional cointegration*, Journal of Empirical Finance (2000), 509530.

[46] R. Bucy and P. Joseph, *Filtering for stochastic processes*, Wiley (1968).

[47] H. N. E. Byström, *Orthogonal garch and covariance matrix forecasting: The nordic stock markets during the asian financial crisis 1997-1998*, The European Journal of Finance **10, No 1** (2004), 44 – 67.

[48] J. F. Carriere, *A large sample test for one parameter families of copulas*, Communication Statistics Theory Methods **23** (1994), 1311–1317.

[49] J. Casals, S. Sotoca, and M. Jerez, *The likelihood of multivariate garch models is ill-conditioned*, Mimeo, Universidad Complutense de Madrid (2002), http://www.ucm.es/info/ecocuan/mjm/vgarch.pdf.

[50] W. S. Chan, *Stock price reaction to news and no-news: Drift and reversal after headlines*, Journal of Financial Economics (2003).

[51] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula methods in finance*, John Wiley & Sons Hoboken (2004).

[52] G. Christodoulakis and S. Satchell, *Correlated archmodelling the time-varying correlation between financial asset returns*, European Journal of Operations Research (2003), 351370.

[53] P. Collin-Dufresney and R. S. Goldsteinz, *Stochastic correlation and the relative pricing of caps and swaptions in a generalized-affine framework*, Available at SSRN: http://ssrn.com/abstract=286322 (2001), http://wpweb2.tepper.cmu.edu/wfa/wfapdf/cor72.pdf.

[54] F. Comte and O. Lieberman, *Asymptotic theory for multivariate garch processes*, Journal of Multivariate Analysis (2005), 6184.

[55] J. Conrad and B. Cornell ad W. R. Landsman, *When is bad news really bad news?*, The Journal of Finance **57 Issue 6** (2002), 2507.

[56] R. Cont, *Understanding correlations in returns: an agent-based approach*, Mimeo, Essex University (1985), http://www.essex.ac.uk/wehia05/Abstract/SpecialSession/Cont.doc.

[57] R. Cont and E. Voltchkova, *Integro-differential equations for option prices in exponential lvy models*, Springer Berlin / Heidelberg (2005).

[58] J. C. Cox, J. E. Ingersoll, and S. A. Ross, *A theory of term structure of interest rates*, Econometrica (2000), 129–151.

[59] J. C. Cox and S. A. Ross, *The valuation of options for alternative stochastic processes*, Journal of Financial Economics (1976).

[60] J. S. Cramer, *Econometric applications of maximum likelihood methods*, Cambridge University Press (1986).

[61] J. Cvitanic, V. Poltimenis, and F. Zapatero, *Optimal portfolio allocation with higher moments*, Available at SSRN: http://ssrn.com/abstract=734703 (2006).

[62] S. Cyganowski, P.E. Kloeden, and J. Ombach, *From elementary probability to stochastic differential equations with maple*, Springer New York (2002).

[63] E. Dabora and K. Froot, *How are stock prices affected by the location of trade?*, Journal of Financial Economics **53** (1999), 189–216.

[64] K. Daniel, D. Hirshleifer, and A. Subrahmanyam, *Investor psychology and security market under- and overreactions*, The Journal of Finance **53 Issue 6** (1998), 1839.

[65] K. Daniel and S. Titman, *Testing factor-model explanations of market anomalies*, Mimeo, Robert H. Smith School of Business,University of Maryland, US (2005), http://www.rhsmith.umd.edu/finance/pdfsdocs/seminarfall06/Daniel.pdf.

[66] J. Davidson, *Stochastic limit theory: An introduction for econometricicans*, Oxford University Press (1994), 405.

[67] P. I. Davies and N. J. Higham, *Numerically stable generation of correlation matrices and their factors*, BIT Numerical Mathematics, Springer **40 (4)** (2006), 640–651.

[68] L. Devroye, *Non-uniform random variate generation*, Springer-Verlag (1986).

[69] F. Diebold and M. Nerlove, *The dynamics of exchange rate volatilitya multivariate latent factor arch model*, Journal of Applied Econometrics (2001), 121.

[70] F. X. Diebold, T. A. Gunther, and A. S. Tay, *Evaluating density forecasts with applications to financial risk management*, International Economic Review (1989), 863883.

[71] F. X. Diebold, J. Hahn, and A. S. Tay, *Multivariate density forecast evaluation andcalibrationinfinancialriskmanagementhigh-frequencyreturns onforeignexchange*, Review of Economics and Statistics (1998), 863883.

[72] Z. Ding and R. Engle, *Large scale conditional covariance matrix modeling*, Estimation and Testing (1993).

[73] Z. Ding, C. W. J. Granger, and R. F. Engle, *A longmemory property of stock market returns and a new model*, Journal of Empirical Finance (2003), 83106.

[74] C. Doz and E. Renault, *Conditionally heteroskedastic factor modelsidenti?cation and instrumental variables estimation*, (1993).

[75] C. Drost and C.Klaassen, *Efficient estimation in semiparametric garch models*, Journal of Econometrics (2003), 193221.

[76] C. Drost and T. Nijman, *Temporal aggregation of garch processes*, Econometrica (1997), 909927.

[77] P. Duchesne and S. Lalancette, *On testing for multivariate arch effects in vector time series models*, Forthcoming in La revue canadienne de statistique. (2003).

[78] D. Duffie and P. Glynn, *Estimation of continuous-time markov processes sampled at random time intervals*, Econometrica **72 Issue 6** (2004), 1773.

[79] J.-M. Dufour, L. Khalaf, and M.-C. Beaulieu, *Finite-sample diagnostics for multivariate regression with applications to linear asset pricing models*, Mimeo, Centre interuniversitaire de recherche en conomie quantitative, CIREQ (2000).

[80] M. Dungey, V. Martin, and A. Pagan, *A multivariate latent factor decomposition of international bond yield spreads*, Journal of Applied Econometrics (2001), 697715.

[81] J. Elder, *An impulse response function for a vector autoregression with multivariate garch in mean*, Economics Letters **79** (2003), 79, 21–26.

[82] P. Embrechts, A. McNeil, and D Straumann, *Correlation and dependence in risk management: Properties and pitfalls*, Cambridge University Press (2002).

[83] R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom in?ation*, Econometrica (1982), 9871007.

[84] ———, *Dynamic conditional correlation -a simple class of multivariate garch models*, Journal of Business and Economic Statistics (2002), 339350.

[85] R. F. Engle and T. Bollerslev, *Modelingt he persistence of conditional variances*, Econometric Reviews (1990), 150.

[86] R. F. Engle and G. Gonzalez-Rivera, *Semiparametric arch model*, Journal of Business and Economic Statistics (1996), 345360.

[87] R. F. Engle, T. Ito, and W.-L. Lin, *Meteor showers or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market*, Econometrica (1951), 525542.

[88] R. F. Engle and Granger C. W. J., *Co-integration and error correction: Representation, estimation, and testing*, Econometrica **55, No. 2** (1987), 251–276.

[89] R. F. Engle and F. Kroner, *Multivariate simultaneous generalized arch*, Econometric Theory (1994), 122150.

[90] R. F. Engle and J. Mezrich, *Garch for groups*, RISK **9** (1996), 3640.

[91] R. F. Engle, V. K. Ng, and M. Rothschild, *Asset pricing with a factor-arch covariance structureempirical estimates for treasury bills*, Journal of Econometrics (2002), 213238.

[92] R. F. Engle and K. Sheppard, *Theoretical and empirical properties of dynamic conditional correlation multivariate garch*, Working Paper, National Bureau of Economic Research, Inc (2002).

[93] W. Feller, *Two singular diffusion problems*, Annals of Mathematics (1996), 173:182.

[94] G. Fiorentini, G. Calzolari, and L. Panattoni, *Analytic derivatives and the computation of garch estimates*, Journal of Applied Econometrics (2003), 399417.

[95] J. Fleming, C. Kirby, and B. Ostdiek, *The economic value ofvolatility timing using realized volatility*, Journal of Financial Economics (2000), 473509.

[96] R. Fletcher, *Practical methods of optimization*, John Wiley and Sons Ltd (2000).

[97] K. J. Forbes and Roberto Rigobon, *No contagion, only interdependence: Measuring stock market comovements*, The Journal of Finance **57 Issue 5** (2002), 2223.

[98] R. V. Foutz and R. C. Srivastava, *Statistical inference for markov processes when the model is incorrect*, Advances in Applied Probability **11, No. 4** (1979), 737–749.

[99] P. Franses and D. van Dijk, *Non-linear series models in empirical finance.*, (1987).

[100] K. French, G. Schwert, and R. Stambaugh, *Expected stock returns and volatility*, Journal of Financial Economics (1996), 329.

[101] W.A. Fuller, *Introduction to statistical time series*, Wiley New York (1976).

[102] A.R. Gallant, D. Hsieh, and G. Tauchen, *Estimation of stochastic volatility models with diagnostics*, Journal of Econometrics, (1997), 159–192.

[103] A.R Gallant, Rossi P.E, and G. Tauchen, *Nonlinear dynamic structures*, Econometrica **61, No 4** (1993), 871–907.

[104] R. Gerlach, C. W.S. Chen, D. S. Y. Lin, and H. Ming-Hsiang, *Asymmetric responses of international stock markets to trading volume*, Physica A: Statistical Mechanics and its Applications **360, Issue 2** (2005), 422–444.

[105] R. Geske and H. E. Johnson, *The american put option valued analytically*, Journal of Finance **39, No. 5** (1984), 1511–1524.

[106] R. Geske and R. Roll, *On valuing american call options with the black-scholes european formula*, Journal of Finance **39, No. 2** (1984), 443–455.

[107] E. Ghysels, A. Harvey, and E. Renault, *Stochastic volatility*, in Handbook of Statistics (2001), 119191.

[108] P. Giot and S. Laurent, *Valut-at-risk for longand short positions*, forthcoming in Journal of Applied Econometrics. (2000).

[109] P. Glasserman, *Monte carlo methods in financial engineering*, Springer; 1 edition (2003).

[110] G. Glen, L. M. Leemis, and J. H. Drew, *Computing the distribution of the product of two continuous random variables*, Computational statistics & data analysis, Elsevier Science, Amsterdam **vol. 44, no3** (1993), 451–464.

[111] L. Glosten, R. Jagannathan, and D. Runkle, *On the relation between expected value and the volatility of the nominal excess return on stocks*, Journal of Finance (2004), 17791801.

[112] G. Golub and C. van Loan, *Matrix computations*, The Johns Hopkins University Press (1996).

[113] C. Gourieroux, *Arch models and financial applications*, Springer-Verlag New York Inc. (2000).

[114] C. Gourieroux and J. Jasiak, *Financial econometrics. princeton university press.*, (1999).

[115] ———, *Non-linear innovations and impulse responses*, http://dept.econ.yorku.ca/jasj/papers/impul.ps (2002).

[116] C. Gourieroux, J. Jasiak, and R. Sufana, *The wishart autoregressive process of multivariate stochastic volatility*, CRES and CEPREMAP (2004).

[117] C. Gouriéroux and R. Sufana, *Wishart quadratic term structure models*, Les Cahiers du CREF of HEC Montreal Working Paper No. 03-10 (1997), Available at SSRN: http://ssrn.com/abstract=757307.

[118] ———, *Derivative pricing with wishart multivariate stochastic volatility applications to credit risk*, Les Cahiers du CREF of HEC Montral Working Paper No. CREF 04-09 (2004), Available at SSRN: http://ssrn.com/abstract=757312.

[119] C. W. J. Granger, *Long memory relationships and the aggregation of dynamic models*, Journal of Econometrics **14** (1983), 227–238.

[120] W. H. Greene, *Econometric analysis*, Prentice Hall; 5 edition (2002).

[121] R. Greenwood and N. Sosner, *Trade and the comovement of stock returns: Evidence from japan*, Available at SSRN: http://ssrn.com/abstract=305059 or DOI: 10.2139/ssrn.305059 (2002).

[122] M. S. Grewal and A. P. Andrews, *Kalman filtering: Theory and practice using matlab*, Prentice Hall (1993).

[123] M. S. Grewal and Angus P. Andrews, *Kalman filtering: Theory and practice*, Prentice Hall (1993).

[124] P. A. Grout and A. Zalewska, *The impact of regulation on market risk*, Journal of Financial Economics **80, No.1** (2006), 149–184.

[125] I. Hacking, *The looping effect of human kinds*, in D. Sperber ed. et al. Causal Cognition An Interdisciplinary Approach. (2003), 351–383.

[126] C. Hafner, *Temporal aggregation of multivariate garch processes*, Econometric Institute Report (1998).

[127] ———, *Fourth moment structure of multivariate garch processes*, Journal of Financial Econometrics (2004), 2654.

[128] C. Hafner and H. Herwartz, *Volatility impulse response functions for multivariate garch models*, CORE Discussion Paper 2001/3 (2001).

[129] ———, *Analytical quasi maximum likelihood inference in multivariate volatility models*, Econometric Institute Report 21 (2003).

[130] ———, *Time-varying market price of risk in the capm. approaches, empirical evidence and implications*, Mimeo, Humboldt Universitaet Berlin (2003), 93112.

[131] C. Hafner and J. Rombouts, *Semiparametric multivariate volatilitymodels*, Econometric Institute Report 21 (2003).

[132] ———, *Estimation of temporally aggregated multivariate garch models*, Center for Operations research and econometrics, Universit catholique de Louvain, France (2004), http://neumann.hec.ca/pages/jeroen.rombouts/research/hafnerrombouts2006.pdf.

[133] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward, *Managing smile risk*, Wilmott Magazine (2002).

[134] P. Hall and P. Yao, *Inference in arch and garch models with heavy-tailed errors*, Econometrica (1998), 285317.

[135] J. D. Hamilton, *Time series analysis*, Princeton University Press (1994).

[136] B. E. Hansen, *Autoregressive conditional density estimation*, International Economic Review **35, No. 3** (1994), 705–730.

[137] L. P. Hansen, *Large sample properties of generalized method of moments estimators*, Econometrica **50, No. 4** (1982), 1029–1054.

[138] L. P. Hansen and J. A. Scheinkman, *Back to the future: Generating moment implications for continuous-time markov processes*, Econometrica **63, No. 4** (1995), 767–804.

[139] B. Hansson and P. Hordahl, *Testing the conditional capm using multivariate garch-m*, Applied Financial Economics (1994), 377388.

[140] A. Harvey, E. Ruiz, and N. Shephard, *Multivariate stochastic variance models*, Review of Economic Studies (2002), 247264.

[141] A.C. Harvey, *Forecasting, structural time series models and the kalman filter*, Cambridge University Press (1989).

[142] C. He and Terasvirta, *An extended conditional correlation garch model and its fourth-moment structure*, Working Paper Series in Economics and Finance from Stockholm School of Economics, Sweden (1980).

[143] ———, *An application of the analogy between vector arch and vector random coefficient autoregressive models*, Working Paper Series in Economics and Finance from Stockholm School of Economics, Sweden (2002).

[144] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies **6, Number 2** (1993), 327–343.

[145] A. Holly and J. R. Magnus, *A note on instrumental variables and maximum likelihood estimatoin procedures*, Annales c'Economic et de Statistique **10** (1988), 121–138.

[146] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press (1990).

[147] S. D. Horn and D. B. Duncan, *Linear dynamic recursive estimation from the viewpoint of regression analysis*, Journal of the American Statistical Association (1972).

[148] J. Hosking, *The multivariate portmanteau statistic*, Journal of American Statistical Association (1996), 602608.

[149] J. C. Hull and A. D. White, *The pricing of options on assets with stochastic volatilities*, Journal of Finance **42, issue 2** (1987), 281–300.

[150] P. Dellaportas I. Vrontos and D. Politis, *A full-factor multivariate garch model*, Econometrics Journal (2005), 311333.

[151] D. L. Ikenberry and S. Ramnath, *Underreaction to self-selected news events: The case of stock splits*, Review of Financial Studies **15, No 2** (2002), 489–526.

[152] T. Jeantheau, *Strong consistency of estimators for multivariate arch models*, Econometric Theory (2001), 7086.

[153] M. Jerez, J. Casals, and S. Sotoca, *Exact smoothing for stationary and nonstationary time series*, International Journal of Forecasting **16** (2000), 59–69.

[154] K. Jogdeo, *Dependence, concepts of. in: Samuel kotz, norman l. johnson, and campbell b. read*, Encyclopedia of Statistical Sciences **2** (1982), 324–334.

[155] M. Johannes and N. Polson, *Mcmc methods for continuous-time financial econometrics*, Available at SSRN: http://ssrn.com/abstract=480461 or DOI: 10.2139/ssrn.480461 (2003).

[156] E. Jondeau and M. Rockinger, *The copula-garch model of conditional dependencies an international stock-market application*, Journal of International Money and Finance, Elsevier **25** (2006), 827–853.

[157] M. Joshi, *C++ design patterns and derivatives pricing*, Cambridge University Press (2003).

[158] H.P. McKean K. Ito, *Diffusion processes and their sample paths*, Springer-Verlag Berlin and Heidelberg GmbH & Co. K (1996).

[159] R. E. Kalman, *Contributions to the theory of optimal control*, Bol. Soc. Mat. Mexicana (1960).

[160] G. Kapetanios and M. Massimiliano, *Impulse response functions from structural dynamic factor models: A monte carlo evaluation*, CEPR Discussion Paper No. 5621 (2006), at SSRN: http://ssrn.com/abstract=914165.

[161] ———, *A parametric estimation method for dynamic factor models of large dimensions*, CEPR Discussion Paper No. 5620 (2006), Available at SSRN: http://ssrn.com/abstract=913392.

[162] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus.*

[163] T. Kariya, *Mtv model and its application to the prediction of stock prices*, in Proceedings of the Second International Tampere Conference in Statistics (1995), 161176.

[164] G. Karolyi, *A multivariate garch model of international transmission of stock returns and volatilitythe case of the united states and canada*, Journal of Business and Economic Statistics (2003), 1125.

[165] H. Kawakatsu, *Cholesky factor garch*, Journal of Applied Econometrics **21, Issue 1** (2000), 79 – 109.

[166] C. Kearney and A. Patton, *Multivariate garch modelling of exchange rate volatility transmission in the european monetary system*, Financial Review (2000), 2948.

[167] J. Kim, *The relationship between the monetary regime and output volatilitya multivariate garch-m model of the japanese experience 1919-1996*, Japan and the World Economy (1994), 4969.

[168] M. King, E. Sentana, and S. Wadhwani, *Volatility and links between national stock markets*, Econometrica (1998), 901933.

[169] M. King and S. Wadhwani, *Transmission of volatility between stock markets*, Review of Financial Studies **3** (1990), 5–33.

[170] T. Kollo and D. von Rosen, *Advanced multivariate statistics with matrices*, Kluwer Academic Publishers (2005).

[171] G. Koop, M. N. Pesaran, and S. M. Potter, *Impulse response analysis in non-linear multivariate models*, Journal of Econometrics **74** (1996), 119–147.

[172] D. Kristensen and A. Rahbek, *Asymptotics of the qmle for a class of arch(q) models*, Econometric Theory **21** (2005), 946–961.

[173] F. Kroner and V. Ng, *Modellingasymmetric comovements of asset returns*, The Review of Financial Studies (2003), 817844.

[174] U. Kuchler and M. Srensen, *A note on limit theorems for multivariate martingales*, Research Report no. 350, dept. theor. statist., University of Aarhus (1996), citeseer.ist.psu.edu/kuchler96note.html.

[175] L. Lang and R. Stulz, *Contagion and competitive intra-industry effects of bankruptcy announcements*, Journal of Financial Economics **32** (1992), 45–60.

[176] M. Lanne and P. Saikkonen, *Non-linear garch models for highly persistent volatility*, The Econometrics **8 Issue 2** (2005), 251.

[177] O. Ledoit, P. Santa-Clara, and M. Wolf, *Flexible multivariate garch modeling with an application to international stock markets*, The Review of Economics and Statistics (1994), 735747.

[178] S. Lee and B. Hansen, *Asymptotic properties ofthe maximum likelihood estimator and test of the stability of parameters of the garch and igarch models*, Econometric Theory (2002), 2952.

[179] J. Lerner and R. P. Merges, *The control of technology alliances: An empirical analysis of the biotechnology industry*, The Journal of Industrial Economics (1998).

[180] J. Lernera, H. Shane, and A. Tsai, *Do equity financing cycles matter? evidence from biotechnology alliances*, Journal of Financial Economics **67** (2003), 411–446.

[181] D. Lien and Y. Tse, *Some recent developments in futures hedging*, Journal of Economic Surveys (1992), 357396.

[182] W. Lin, *Alternative estimators for factor garch models -a monte carlo comparison*, Journal of Applied Econometrics (1997), 259279.

[183] ———, *Impulse response function for conditional volatility in garch models*, Journal of Business and Economic Statistics **15, No 1** (1997), 15–25.

[184] S. Ling and W. Li, *Diagnostic checkingof nonlinear multivariate time series with multivariate arch errors*, Journal of Time Series Analysis (1995), 447464.

[185] S. Ling and M. McAleer, *Stationarity and the existence of moments of a family of garch processes*, Journal of Econometrics (2002).

[186] L. Ljungqvist and T. J. Sargent, *Recursive macroeconomic theory,*, The MIT Press (2004).

[187] F. Longin and B. Solnik, *Is the correlation in international equity returns constant1960-1990?*, Journal of International Money and Finance (2002), 326.

[188] R. Lucchetti, *Analytic score for multivariate garch models*, Computational Economics (1996), 133143.

[189] R. Lumsdaine, *Asymptotic properties of the quasi maximum likelihood estimatorin garch*, (1999).

[190] G.S. Maddala and In-Moo Kim, *Unit roots, cointegration and structural change*, Cambridge University Press (1998).

[191] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications to statistics and econometrics*, John Wiley & Sons (1998).

[192] ———, *Elimination matrix: some lemmas and applications*, John Wiley & Sons (1999).

[193] B. Mandelbrot and H. M. Taylor, *On the distribution of stock price differences*, Operations Research **15, No. 6** (1967), 1057–1062.

[194] H Markowitz, *Portfolio selection*, The Journal of Finance (1952).

[195] B. McCullough and H. Vinod, *The numerical reliability of econometric software*, Journal of Economic Literature (2003), 633665.

[196] F. Mencá and E. Sentana, *Estimation and testingof dynamic models with generalised hyperbolic innovations*, (1987).

[197] R. C. Merton, *A simple model of capital market equilibrium with incomplete information*, Journal of Finance, American Finance Association, vol. 42 (1997).

[198] R. Michaely and K. L. Womack, *Conflict of interest and the credibility of underwriter analyst recommendations*, Review of Financial Studies **12, No 4** (1999), 653–686.

[199] J. J. Moore, B. S. Garbow, and K. E. Hillstorm, *User guide for minpack-1*, Argonne National Laboratory Report (1980).

[200] C. D. Moravitz-Martin, *Tensor decompositions workshop discussion notes*, American Insti-tute of Mathematics (AIM) (2004), http://www.aimath.org/WWN/tensordecomp/.

[201] J. P. Morgan, *Riskmetrics technical document*, 4th ed. J.P.Morgan (1998).

[202] R. J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons Inc (2005).

[203] M. Musiela and M. Rutkowski, *Martingale methods in financial modelling*, Springer (1999).

[204] ———, *Martingale methods in financial modelling 2nd edition*, Springer (2002).

[205] R. Nelsen, *An introduction to copulas*, Springer Verlag, New York. (1991).

[206] D. Nelson, *Conditional heteroskedasticity in asset returnsa new approach*, Econometrica (1992), 349370.

[207] D. Nelson and C. Cao, *Inequality constraints in the univariate garch model*, Journal of Business and Economic Statistics (1985), 229235.

[208] H. Neudecker and S. Liu, *Matrix trace inequalities involving simple, kronecker, and hadamard products*, Econometric Theory (1993).

[209] W. Newey, *Maximum likelihood specification testing and conditional moment tests*, Econometrica (1996), 10471070.

[210] T. Nijman and E. Sentana, *Marginalization and contemporaneous aggregation in multivariate garch processes*, Journal of Econometrics (1996), 7187.

[211] R.H. Norden, *A survey of maximum likelihood estimation*, International Statistical Review (1972), 329–354.

[212] B. Øksendale, *Stochastic differential equations*, xyz **abc** (2001), 14:18.

[213] A. Pagan, *The econometrics of financial markets*, Journal of Empirical Finance (1996), 15102.

[214] F. Palm, *Garch models of volatility*, in Handbook of Statistics (2000), 209240.

[215] A. Patton, *Modelling time-varying exchange rate dependence using the conditional copula*, UCSD Discussion Paper No. 01-09 (2005), Available at SSRN: http://ssrn.com/abstract=275591 or DOI: 10.2139/ssrn.275591.

[216] M. M. Pelagatti and S. Rondena, *Dynamic conditional correlation with elliptical distributions*, Available at SSRN: http://ssrn.com/abstract=888732 (2003).

[217] D. Pelletier, *Regime switching for dynamic correlations*, Journal of Econometrics. **131, 1-2** (2000), 445–473.

[218] A. Pelsser, *Efficient methods for valuing interest rate derivatives*, (2003).

[219] W. Polasek and L. Ren, *Model selection for var-garch models using the basel package*, Institute of Statistics and Econometrics, University of Basel (1998).

[220] ———, *Generalised impulse response functions for var-garch-m models*, Institute of Statistics and Econometrics, University of Basel (2000).

[221] M. Pritamani and V. Singal, *Return predictability following large price changes and information releases*, Journal of Banking and Finance **25** (2001), 631–656.

[222] S. Rachev and S. Mittnik, *Stable paretian models in finance*, John Wiley & Sons (2004).

[223] R. Rebonato, *Volatility and correlation the perfect hedger and the fox*, John Wiley & Sons; 2 edition (2003).

[224] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the kalman filter: Particle filters for tracking applications*, Artech House Publishers (2004).

[225] C. Robinson, *Dynamical systems: Stability, symbolic dynamics and chaos*, CRC Press (1999).

[226] L.C.G. Rogers and D. Williams, *Diffusions, markov processes and martingales: Foundations*, Cambridge University Press (2000).

[227] J. Rombouts and M. Verbeek, *Dynamic semiparametric portfolio management in var framework*, forthcoming CORE DP. (2005).

[228] G. De Rossi and A. Harvey, *Time-varying quantiles*, Cambridge Working Papers in Economics (2006).

[229] F. Russo and P. Vallois, *Product of two multiple stochastic integrals with respect to a normal martingale*, Stochastic Processes and their Applications, Elsevier Science **73, Number 1** (1995), 47–68.

[230] B. P. Sack and R. Rigobon, *The impact of monetary policy on asset prices*, FEDS Working Paper No. 2002-04 (2004), Available at SSRN: http://ssrn.com/abstract=298365 or DOI: 10.2139/ssrn.298365.

[231] G. De Santis and B. Gerard, *How big is the premium for currency risk?*, Journal of Financial Economics (1999), 375412.

[232] E. Sentana, *Quadratic archmodels*, Review of Economic Studies (1998), 639661.

[233] ———, *The relation between conditionally heteroskedastic factor models and factor garch models*, Econometrics Journal (2001), 19.

[234] E. Sentana and G. Fiorentini, *Identification estimation and testing of conditionally heteroskedastic factor models*, Journal of Econometrics (1996), 143164.

[235] N. Shephard, *Statistical aspects of arch and stochastic volatility*, Time Series Models: In Econometrics (2001), 167.

[236] J. A. Shohat and J. D. Tamarkin, *The problem of moments*, American Mathematical Society (1943).

[237] S. E. Shreve, *Stochastic calculus for finance ii: Continuous-time models*, Springer-Verlag New York Inc. (2004).

[238] G. Silberberg and S. Pafka, *A sufficient condition for the positive definiteness of the covariance matrix of a multivariate garch model*, CEU Department of Economics Working Paper No. 7/2001 (2001), Available at SSRN: http://ssrn.com/abstract=275136 or DOI: 10.2139/ssrn.275136.

[239] C.Y. Sin, *Impulse response and variance decomposition in the presence of time varying volatility: with applications to price discovery*, Mimeo, Department of Economics, Hong Kong Baptist University (2005).

[240] A. Sklar, *Fonctions de répartition 'a n dimensions et leurs marges*, Inst. Statist. Univ. Paris **8** (1959), 229231.

[241] M. S. Srivastava, *Methods of multivariate statistics*, John Wiley & Sons Inc (2002).

[242] M. S. Srivastava and C. G. Khatri, *An introduction to multivariate statistics*, North-Holland Publ (1979).

[243] J. M. Steele, *Stochastic calculus and financial applications*, Springer; 1 edition (1985).

[244] N. L. Stokey and R. E. Lucas, *Recursive methods in economic dynamics*, Harvard University Press (1989).

[245] S. H. Szewczyk, *The intra-industry transfer of information inferred from announcements of corporate security offerings*, Journal of Finance **47, No. 5** (1992), 1935–1945.

[246] A. Takemura and S. Kuriki, *Maximum covariance difference test for equality of two covariances matrices*, Algebraic Methods in Statistics and Probability, Contemporary Mathematics **287** (2001), 283–301.

[247] G. Tauchen, *Diagnostic testing and evaluation of maximum likelihood models*, Journal of Econometrics (1992), 415443.

[248] O. Tawatnuntachai and R. D'Mello, *Intra-industry reactions to stock split announcements*, The Journal of Financial Research **25 Issue 1** (2002), 39.

[249] J. C. Taylor, *Skew products regular conditional probabilities and stochastic differential equations: a technical remark*, Sminaire de probabilits de Strasbourg **26** (2000), 113–126.

[250] A. Timmermann, *Density forecasting in economics and finance*, Journal of Forecasting (2002), 120123.

[251] R. Tsay, *Analysis of financial time series*, Wiley-Interscience; 1st edition (2000).

[252] Y. Tse, *Residual-based diagnostics for conditional heteroscedasticity models*, Econometrics Journal (1999), 358373.

[253] ———, *A test for constant correlations in a multivariate garch model*, Journal of Econometrics (2002), 107127.

[254] Y. Tse and A. Tsui, *A multivariate garch model with time-varying correlations*, Journal of Business and Economic Statistics (2002), 351362.

[255] ———, *A note on diagnosing multivariate conditional heteroscedasticity models*, Journal of Time Series Anaysis (2002), 679691.

[256] R. van der Weide, *Go-garch: A multivariate generalized orthogonal garch model*, Journal of Applied Econometrics (1993), 549564.

[257] J. H. van Zanten, *A multivariate central limit theorem for continuous local martingales*, Centrum voor Wiskunde en Informatica (CWI) (1998), 10.

[258] A. M. Vijh, *The spinoff and merger ex-date effects*, Journal of Finance **49, No. 2** (1994), 581–609.

[259] P. Vlaar and F. Palm, *The message in weekly exchange rates in the european monetary system mean reversion conditional heteroskedasticity and jumps*, Journal of Business and Economic Statistics (2003), 351360.

[260] R. von Mises, *Mathematical theory of probability and statistics*, New York: Academic Press (1964).

[261] W. Vetterling W. H. Press, S. A. Teukolsky and B. P. Flannery, *Numerical recipes in c++: The art of scientific computing*, Cambridge University Press (2002).

[262] J. M. Williams and C. Ioannidis, *An explanation of interdependence within the uk biotechnology sector*, Mimeo, Working papers, University of Bath (2004).

[263] ———, *Multifactor forward rates models and monetary policy responses*, Mimeo, Working papers, University of Bath (2005).

[264] J. M. Zakoian, *Threshold heteroskedasticity models*, Journal of Economic Dynamics and Control (2005), 931955.