CITY UNIVERSITY

Higher Moment Models for Risk and Portfolio Management

by

Alexios Ghalanos

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Faculty of Finance
Cass Business School

December 2012

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my academic advisors Steve Thomas and Andrew Clare. I am especially thankful to Eduardo Rossi and Giovanni Urga for their collaboration on a number of research projects.

Dedicated to Anna and Liana.

Thesis Deposit Agreement

COVERED WORK

I, Alexios Ghalanos, of 8A Drayton Court, London SW10 9RQ, "the Depositor",would like to deposit "Higher Moment Models for Risk and Portfolio Management", hereafter referred to as the "Work", in the City University Institutional Repository and agree that rights granted to the City University Institutional Repository through this agreement are entirely non-exclusive and royalty free. I am free to publish the Work in its present version or future versions elsewhere. I agree that the City University Institutional Repository administrators or any third party with whom the City University Institutional Repository has an agreement to do so may, without changing content, translate the Work to any medium or format for the purpose of future preservation and accessibility.

DEPOSIT IN THE CITY UNIVERSITY INSTITUTIONAL REPOSITORY

I understand that Work deposited in the City University Institutional Repository will be accessible to a wide variety of people and institutions - including automated agents - via the World Wide Web. I also agree to an electronic copy of my thesis being included in the British Library Electronic Theses On-line System (EThOS). I understand that once the Work is deposited, a citation to the Work will always remain visible. Removal of the Work can be made after discussion with the City University Institutional Repository, who shall make best efforts to ensure removal of the Work from any third party with whom the City University Institutional Repository has an agreement.

I AGREE AS FOLLOWS:

- That I am the author of the Work and have the authority to make this agreement and to hereby give the City University Institutional Repository administrators the right to make available the Work in the way described above.
- That I have exercised reasonable care to ensure that the Work is original, and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other Intellectual Property Right. Where I have included third party copyright material, I have fully acknowledged its source.

- The administrators of the City University Institutional Repository do not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of breach of intellectual property rights, or any other right, in the material deposited.

CITY UNIVERSITY

Abstract

Faculty of Finance
Cass Business School

This thesis considers specific topics related to the dynamic modelling and management of risk, with a particular emphasis on the generation of asymmetric and fat tailed behavior observed in practise. Specifically, extensions to the dynamics of the popular GARCH model, to capture time variation in higher moments, are considered in the univariate and multivariate context, with a special focus on the Generalized Hyperbolic distribution. In Chapter 1, I consider the extension of univariate GARCH processes with higher moment dynamics based on the Autoregressive Conditional Density model of Hansen (1994), with conditional distribution the Generalized Hyperbolic. The value of such dynamics are analyzed in the context of risk management, and the question of ignoring them discussed. In Chapter 2, I review some popular multivariate GARCH models with a particular emphasis on the dynamic correlation model of Engle (2002), and alternative distributions such those from the Generalized Asymmetric Laplace of Kotz, Kozubowski, and Podgorski (2001). In Chapter 3, I propose a multivariate extension to the Autoregressive Conditional Density model via the independence framework of the Generalized Orthogonal GARCH models, providing the first feasible model for large dimensional multivariate modelling of time varying higher moments. A comprehensive out-of-sample risk and portfolio management application provides strong evidence of the improvement over non time varying higher moments. Finally, in Chapter 4, I consider the benefits of active investing when the benchmark index is not optimally weighted. I investigate advances in the definition and use of risk measures in portfolio allocation, and propose certain simple solutions to challenges arising in the optimization of these measures. Combining the models discussed in the previous chapters, within a fractional programming optimization framework and using a range of popular risk measures, a large scale out-of-sample portfolio application on the point in time constituents of the Dow Jones Industrial Average is presented and discussed, with clear implications for active investing and benchmark policy choice.

Introduction

The modelling of security returns is of vital importance in forecasting and risk management, with implications on the methods for allocating capital among competing prospects. It forms a key part of the value added process of the investment allocation life cycle, which is the dynamic process of model building, forecasting and the generation of optimal allocations of capital based on some risk-reward trade-off. Traditionally, this process has been completely dominated by the classical model of Bachelier (1964), developed further, most notably, by Fama (1970) in the Efficient Market Hypothesis, and the modern portfolio theory of Markowitz (1952). Under this framework, security returns are seen as being random, absent of any distorting 'effects' and normally distributed. While the randomness concept is not generally disputed, the absence of distorting effects and the normality assumption, while offering tractability and elegance to the mathematical modelling setup, are not borne out in reality. The distorting effects, also called stylized facts, lead to behavior which deviates from the i.i.d. assumption, with obvious consequences with respect to the summation of random sequences and any inferences or forecasts based on this. There is an abundance of empirical evidence documenting such effects as fat tails and skewness - a clear departure from normality, autocorrelation in returns and volatility clustering. This has given rise to a new set of models, methods and distributions aimed at providing a more realistic model of security dynamics, supported by more positive theories of investor behavior such as in Kahneman and Tversky (1979). In what Rom and Ferguson (1994) called the 'Post-Modern Portfolio' age, the recognition of fat tails and asymmetry has given rise to new models and measures of risk which more adequately capture these effects. Foremost among these new models has been the ARMA model popularized by Box and Cox (1964) and the GARCH model of Bollerslev (1986). These models have sought to provide simple dynamics for the conditional first and second moments of the distribution of security returns. Numerous extensions such as the incorporation of the asymmetric effect of negative and positive returns (TAR and TGARCH), long memory (ARFIMA and FIGARCH), the Taylor effect[1] (APARCH), and omnibus distributions such as the Generalized Hyperbolic of Barndorff-Nielsen (1977) have been used in an attempt to capture increasing degrees of complexity in the observed dynamics of the underlying securities.

The implications for risk management are clear, with a large number of studies showing that GARCH models provide superior estimates of conditional density based tail measures such as VaR. In industry, J.P. Morgan's Riskmetrics methodology (see Morgan (1994)), being a restricted GARCH model with fixed parameters, has proved simple and popular among practitioners leading to its wide adoption. Nevertheless, despite the popularity of GARCH models and their applicability across a wide range of applications, they cannot account for the large security price fluctuations observed in practice, even when accounting for fat tails and skewness with the use of highly parameterized distributions. To correct this limitation, while staying within the tractable GARCH framework, Hansen (1994) introduced dynamics to the parameters which control the shape and skew of the distribution, thus potentially allowing for extreme events to be modelled. However, this Autoregressive Conditional Density (ACD) model has not been as popular as the simpler GARCH model, and many open questions remain such as the generalization to the multivariate domain, parameter consistency, and the overall net benefit for the added complexity.

The objectives of this thesis, are twofold. First, to address the issue of time varying higher moments in the joint modelling of security returns with the objective of providing for a feasible and value added input to the portfolio allocation and risk management setup. Second, to test this and related models using a new set of risk-reward models in a large empirical setup and at the same time consider the question of the optimality of the weighting schemes used by benchmark indices in the presence of such dynamics.

In Chapter 1, I review the literature on the ACD models, much of which attempts to establish their value using a variety of mainly in-sample empirical applications covering securities from real estate to foreign exchange, daily to monthly frequencies. The chapter tries to answer a more general question, namely what is the cost of using GARCH when the dynamics include time varying higher moments. Using a variety of tests, both in- and out-of-sample, based on simulated and real data, I show that there is a real cost to using GARCH in these circumstances but little cost to using ACD dynamics to capture the infrequent extremes seen in practice. The chapter introduces the ACD-GH

[1]Named after Taylor (1986) who observed that the sample autocorrelation of absolute returns was usually larger than that of squared returns

model, a far more flexible representation than previously used in the literature, and discusses its properties and the modelling challenges arising from the highly nonlinear dynamics and presence of 2 shape parameters. The question of parameter consistency is also partially addressed within a simulation exercise, something very rarely seen in the literature.

Chapter 2 provides a general review of some popular multivariate GARCH models and discusses the tradeoff between feasibility and complexity in the modelling process with some interesting insights. In particular, the value of the DCC model versus more established models like the BEKK, a question posed by Caporin and McAleer (2012), is discussed and some evidence with general observations provided for why there is value in a first stage univariate filter. Some recent enhancements to the DCC model are reviewed with a particular emphasis on the problem of estimation in the presence of covariance targeting. More positive enhancements in terms of the Copula DCC model are summarized and form part of the large empirical application of Chapter 4.

In Chapter 3, the ACD model of the first chapter is extended to the multivariate domain via the GO-GARCH framework of van der Weide (2002), providing the first feasible and tractable approach for the joint modelling of time varying higher moments. The affine representation of the model which gives rise to closed form expressions of higher co-moment matrices and a semi-analytic form for the weighted portfolio density has clear applications in portfolio and risk management. An extensive set of empirical applications establishes the value of the model and its features using two different datasets and frequencies.

In Chapter 4, the models reviewed and introduced in the thesis form the data generating processes from which a large scale scenario based portfolio allocation exercise is undertaken on the weekly point in time constituents of the Dow 30 index. Using both long-only and long-short portfolios, by deriving new smooth NLP representations of some popular risk measures such as CVaR and LPM, the application considers the benefits of applying these new tools to beating the benchmark indices. In the presence of the observed dynamics in the moments and co-moments of securities, it is argued that static or simplistic weighting schemes used by the benchmark indices make them suboptimal with clear implications for both active and passive investment.

Chapter 1

GARCH Dynamics and Time Varying Higher Moments

Introduction

Since Mandelbrot (1963), researchers have discovered numerous statistical properties in real market time series that contradict the theoretical results of their models. These so called stylized facts, together with the paradigm shift away from the completely rational, representative agent to a boundedly rational, heterogeneous agent, has motivated researchers to model financial markets with a new set of tools, distributions and models. Among these, the pioneering work of Box and Cox (1964) in the area of autoregressive moving average models paved the way for related work in the area of volatility modelling with the introduction of ARCH and then GARCH models by Engle (1982) and Bollerslev (1986), respectively. In terms of the statistical framework, these models provide motion dynamics for the dependency in the conditional time variation of the mean and variance, in an attempt to capture such phenomena as autocorrelation in returns and squared returns. Extensions to these models have included more sophisticated dynamics such as threshold models to capture the asymmetry in the news impact, as well as distributions other than the normal to account for the skewness and excess kurtosis observed in practise. With the introduction of flexible distributions, such as those from the normal mixture family, which possess many desirable properties, the importance of conditional variations in moments other than the mean and variance has been researched. Capturing the asymmetries and thick tails that are typically observed in the distribution of financial returns is particularly important in the context of risk management and portfolio theory, with a number of authors, including

Lai (1991), Prakash, Chang, and Pactwa (2003), and Jondeau and Rockinger (2006b), providing evidence suggesting that the incorporation of higher moments in portfolio allocation leads to superior approximations of expected utility.

Although models from the GARCH family are able under certain assumptions and parameterizations to produce thick-tailed and skewed unconditional distributions, they typically assume that the shape and skewness parameters are time invariant. This also leads to the assumption that the conditional distribution of the standardized innovations is independent of the conditioning information, for which there is no good reason to believe so a-priori. Different models have been developed in the literature to capture dependencies in higher moments, starting with Hansen (1994) who considered the problem of modelling the full parameters of a generalized skew-student distribution by imposing a quadratic law of motion on the conditioning information. With few exceptions, the research on time varying higher moments has mostly explored different parameterizations, in terms of dynamics and distributions, with little attention to the performance of these models out-of-sample or in their ability to outperform GARCH models with respect to such measures as Value at Risk (VaR) which are important in risk management. The question of how these models perform out-of-sample with respect to a range of measures is addressed in this chapter through an empirical application on 14 international equity indices and a range of popular distributions. More generally, using a Monte Carlo experiment, the more important question of the cost of ignoring time varying higher moment dynamics is addressed with clear implications for risk management. The chapter is organized as follows: Section 1.1 provides an introduction to the ACD models dynamics, followed by a literature review of these models in Section 1.2. Section 1.4 introduces a new model based on the Generalized Hyperbolic distribution which is an omnibus distribution with many well researched sub-families, and discusses its properties, estimation challenges and standardization. Section 1.5 discusses more general features of the ACD models including their estimation, forecasting, simulation and inference methods, followed by an out-of-sample empirical application comparing the performance of 14 international equity indices using both ACD and GARCH models, and 5 very flexible and feature-rich distributions. The more general question of ignoring ACD dynamics is addressed in Section 1.6 using a variety of tests, and Section 1.7 concludes.

1.1 The Autoregressive Conditional Density Model

The Autoregressive Conditional Density (ACD) model[1], formally introduced by Hansen (1994), generalizes GARCH type dynamics to time varying conditional higher moments and as such subsumes them. In GARCH models, the density function is usually written in terms of the location and scale parameters, normalized to give zero mean and unit variance. The same arguments follow in ACD models, and I follow the exposition of Hansen (1994) and consider the density function $f(y|\alpha)$, partitioned so that

$$\alpha_t = (\mu_t, \sigma_t, \omega_t)\,, \tag{1.1}$$

where the conditional mean is given by

$$\mu_t = \mu\left(\theta, x_t\right) = E\left(y_t|x_t\right), \tag{1.2}$$

and the conditional variance is,

$$\sigma_t^2 = \sigma^2\left(\theta, x_t\right) = E\left(\left(y_t - {\mu_t}^2\right)|x_t\right), \tag{1.3}$$

with $\omega_t = \omega(\theta, x_t)$ denoting the remaining parameters of the distribution, such as a shape and skew parameter. The conditional mean and variance are used to scale the innovations,

$$z_t\left(\theta\right) = \frac{y_t - \mu\left(\theta, x_t\right)}{\sigma\left(\theta, x_t\right)}, \tag{1.4}$$

having conditional density which may be written as,

$$g\left(z|\omega_t\right) = \frac{d}{dz}P\left(z_t < z|\omega_t\right), \tag{1.5}$$

and related to $f(y|\alpha)$ by,

$$f\left(y_t|\mu_t, \sigma_t^2, \omega_t\right) = \frac{1}{\sigma_t}g\left(z_t|\omega_t\right). \tag{1.6}$$

The difference between ACD and GARCH models is that in the latter case, ω_t is time invariant. A first order constant-GARCH(1,1) model with general ACD dynamics can

[1] Also abbreviated as $ARCD$ in the literature, whilst ACD has also been used for the Autoregressive Conditional Duration model of Engle and Russell (1998).

thus be represented as,

$$\begin{aligned}
x_t &= \mu_t + \varepsilon_t, \\
\varepsilon_t &= \sigma_t z_t, \\
z_t &\sim \Delta\left(0, 1, \omega_{jt}\right), \quad j = 1, \ldots, l \\
\sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\
\omega_{jt} &= \Phi\left(\bar{\omega}_{jt}\right),
\end{aligned} \tag{1.7}$$

where Δ is some appropriately scaled distribution with $j = 1, \ldots, l$ higher order time varying dynamics denoted by parameters ω_{jt}, and $\Phi(.)$ represents some appropriate transformation, to the unconstrained motion dynamics $\bar{\omega}_{jt}$, constraining them within their distribution specific bounds. Contrary to variance, which is directly modelled and constrained to be positive, the modelling of higher order moments is done indirectly via the some distributional parameters which usually have to be constrained within a specific range. The shape parameter(s) controls the tail thickness while the skew parameter(s) the asymmetry, and both may be needed in the calculation of the higher order central moments such as skewness and kurtosis. A key requirement for any autoregressive type process is the self-decomposability of the conditional distribution, while possessing the linear transformation property is required to center $(x_t - \mu_t)$ and scale (ϵ_t/σ_t) the innovations, after which the modelling is carried out directly using the zero-mean, unit variance, distribution of the standardized variable z (which is a scaled version of the same conditional distribution of x_t). The ACD model can be modelled with any type of dynamics for the variance, and higher moment parameters, but because of the varying nature of the latter only certain types of dynamics will immediately lead to closed form solutions for persistence in the conditional variance equation, while for the higher moments only simulation methods are available to evaluate the unconditional moments. For the simple model given above, the persistence and unconditional value of the variance are easily derived from the literature on ARMA type processes, and useful for imposing some stationarity conditions during estimation and for n-ahead forecasting,

$$\begin{aligned}
E\left(\sigma^2\right) &= \frac{\omega}{1 - (\alpha_1 + \beta_1)}, \\
E\left(\omega_j\right) &\equiv E\left(\Phi\left(\bar{\omega}_j\right)\right),
\end{aligned} \tag{1.8}$$

which because of the nonlinear transformation, $E\left(\Phi\left(\bar{\omega}_{jt}\right)\right) \neq \Phi\left(E\left(\bar{\omega}_{jt}\right)\right)$. As a result, the unconditional value of the higher moment parameters, and hence the higher central moments, must be estimated via simulation. A popular choice for the transformation

$\Phi(.)$ is the logistic cumulative distribution function (CDF),[2] in which case the expectation reduces to the following form for the higher moment parameters,

$$\begin{aligned} E(\omega_j) &= E\left(L_{\omega_j} + \frac{(U_{\omega_j} - L_{\omega_j})}{1+e^{-\bar{\omega}}}\right) \\ &= L_{\omega_j} + (U_{\omega_j} - L_{\omega_j})E\left(\frac{1}{1+e^{-\bar{\omega}}}\right), \end{aligned} \tag{1.9}$$

where U_ω and L_ω represent the upper and lower distributional bounds of the higher moment parameters. Unlike some other nonlinear GARCH models, where the logistic transformation is used[3], there does not appear to be any trick which can be used to simplify this (such as an antisymmetric relationship around the expected value) as a result of transforming the whole rather than parts of the function. Therefore, the transition from simple GARCH to ACD models takes one well into the domain of non-linear modelling. The original model of Hansen (1994) used quadratic type dynamics for the higher order parameters:

$$\bar{\omega}_{1t} = \zeta_0 + \zeta_1 z_{t-1} + \zeta_2 z_{t-1}^2 + \xi\bar{\omega}_{1t-1}, \tag{1.10}$$

while Jondeau and Rockinger (2003) used piecewise linear dynamics:

$$\bar{\omega}_{1t} = \zeta_0 + \zeta_1 z_{t-1} I_{z_{t-1}\leqslant y} + \zeta_2 z_{t-1} I_{z_{t-1}>y} + \xi\bar{\omega}_{1t-1}, \tag{1.11}$$

where I is the indicator function taking on the value 1 if true and 0 otherwise, y the threshold value (normally set to 0) and z_t the standardized residuals, though the residuals have also been used by some researchers instead. I have found that using instead a simple threshold value for y of 1 (i.e. one standard deviation either side of zero) reduces some of the noise inherent in the higher moment dynamics, and hence all subsequent reference to the piecewise linear dynamics will use this setup. The question of parameter consistency of these different model dynamics is addressed in Section 1.5.1.

[2] Another choice would be the truncated Normal distribution which provides for a somewhat finer control on the shape of the CDF.

[3] Such as the logistic Smooth Transition Models.

1.2 Empirical Studies

The previous section set out a rather general setup, but other parameterizations have been considered in the literature with regards to the proxy for the motion dynamics of the higher moment parameters, and the underlying conditional distribution. In Harvey and Siddique (1999), Brooks, Burke, Heravi, and Persand (2005), Premaratne and Bera (2000), Rockinger and Jondeau (2002) and Jondeau and Rockinger (2003) for example, different laws of motion and distributions for modelling the full time varying conditional density parameters were considered, on instruments varying from real estate to foreign exchange returns. The results were mixed, with Harvey and Siddique (1999) finding significant evidence of time varying skewness, Jondeau and Rockinger (2003) finding both time varying skewness and kurtosis significant, while Premaratne and Bera (2000), Brooks, Burke, Heravi, and Persand (2005) and Rockinger and Jondeau (2002) found little evidence of either. With regards to the frequency of observation, Jondeau and Rockinger (2003) found the presence of time varying skewness and kurtosis in daily but not weekly data, partly consistent with the observation that excess kurtosis diminishes with temporal aggregation, while others including Hansen (1994), Bond and Patel (2003) and Harvey and Siddique (1999) did find evidence of time varying skewness and kurtosis in weekly and even monthly data. Other researchers including Bond and Patel (2003) found the whole premise algorithmically unstable leading to convergence problems, arising because of the constraints required to limit the distribution parameters within certain bounds and the fact that both skewness and kurtosis are driven by extreme events making their identification with a particular law of motion and (standardized) residuals very hard. Brannas and Nordman (2003) found that depending on the type of distribution, the results will vary, with more richly parameterized distributions leading to a better overall fit and inference about time variation in the parameters. Moreover, as argued by Jondeau and Rockinger (2003), the method used to estimate the model may lead to very different results, demonstrated by their use of an algorithm using a Sequential Quadratic Programming (SQP) solver which constrains the parameters directly at every point in time. However, they also pointed out that constraining the parameters at every time point does not solve the problem of forecasting as the elimination of the nonlinear bounding transformation leaves the dynamics absent of any rule to constrain the forecast from violating the parameter bounds. Additionally, it is useful to be able to check and impose stationarity on the variance through the calculation of the persistence, also necessary for n-step ahead forecasting. In this setup, where the shape and skew parameters are allowed to vary, this creates a constraint on the type of variance dynamics which may be used with such a constraint, effectively limiting the motion dynamics to symmetric type processes. Consider for example the model of Jondeau and Rockinger (2003), where a piecewise linear based law of motion is used for the variance,

$$\sigma_t^2 = \omega + \alpha_1 \max\left(z_{t-1}, 0\right) + \alpha_2 \min\left(z_{t-1}, 0\right) + \beta_1 \sigma_{t-1}^2 \ . \tag{1.12}$$

The problem with this setup is that in order to calculate persistence and hence impose stationarity rules during estimation, the expectation of the min and max of the distribution is required, by evaluating for instance the following integral for the max case,

$$E\left(max\left(z, 0\right)\right) = \int_{-\infty}^{\infty} \max\left(z, 0\right) f\left(z; 0, 1, \omega_j\right) dz, \tag{1.13}$$

which proves difficult and impractical when the higher order parameters ω_j are time varying. Therefore, laws of motion such as those considering asymmetric effects pose certain challenges, unless one is prepared to assume that obtaining these values by simulation after the fact is reasonable, which is usually the case in most nonlinear models, where stationarity conditions may be deemed sufficient but not necessary[4]. This should not be seen as a severe limitation of ACD models, since by allowing the skew and shape parameters to vary, there is less of a need to capture asymmetries from the law of motion of the variance. Evidence of this was partly presented in Harvey and Siddique (1999) where the inclusion of time varying skewness affected the persistence of the conditional variance and caused some of the asymmetries in the variance to disappear (through a reduced asymmetry coefficient in the variance dynamics).

With the exception of Wilhelmsson (2009), no other paper to my knowledge, has investigated the out-of-sample performance of the ACD models with respect to applied measures such as VaR. In fact, the vast majority of papers reviewed simply investigated the in-sample dynamics of the models, inferring from the estimated parameters the presence or absence of time variation in the higher moments without attempting to address the issue of how this translates, out-of-sample, to a value added input in an applied setting. Using the NIG distribution, Wilhelmsson (2009) does provide a large out-of-sample application, using a range of measures such as the test of VaR exceedances of Christoffersen (1998) and misspecification test of Hong and Li (2005), to compare the ACD-NIG model against 4 other models including the model of Hansen

[4] In any case, simulation is required to evaluate the unconditional expectation of the higher order parameters.

(1994), and 3 GARCH models. Unfortunately, the application is based entirely on the S&P 500 index, though it does go as far back as 1962, and uses a rolling scheme of re-estimating the model, rolling the 1-ahead forecast every single day! In section 1.4, I generalize the dynamics to the Generalized Hyperbolic distribution, which embeds the NIG, and discuss its properties and estimation challenges.

1.3 Dataset Choice and Motivation

The empirical application considered in this chapter, while using daily data going back only as far as 1996, includes 14 international equity indices, and thus a wide sample from which to draw conclusions. I use a dataset comprised of indices simply because of the continuity they offer for research purposes which is not always possible when choosing individual equities. The fact that they represent weighted aggregates of their respective constituents also alleviates any possible dataset bias from choosing securities with highly pronounced idiosyncratic properties. Finally, the international nature of this dataset provides diversity in the characteristics of the returns as well as comparison between and across regional groupings. The period under consideration is also one of the most feature rich in financial history presenting a unique testing ground for risk and portfolio models.[5] The importance of these exchange traded international equity indices in portfolio allocation, and a more thorough description of their characteristics can be found, among others, in Miffre (2007), Amenc and Goltz (2007) and Chen and Huang (2010). A comprehensive survey of exchange traded funds (ETF) and the ways they are used in asset management is published by the Edhec-Risk Institute.[6].

In Table 1.1 I provide a summary of some key features of this dataset, made up of 14 MSCI iShares[7] representing a country sampling from 3 key geographic regions, namely Americas, Asia and Europe, for the period 12/08/1996 to 02/03/2011, where the starting date was chosen as the earliest available date common to all indices, after truncating a few periods for the presence of an excessive amount of non-trading (zero

[5]The period includes the Asian Financial Crisis (1997), the Russian Financial Crisis (1998), the Dot-Com bubble (2000) and subsequent economic downturn (2002), the Chinese Stock Bubble (2007), the US Bear Market of 2007-2009, the European sovereign debt crisis (2010) as well as the flash crash of May 2010.

[6]http://www.edhec-risk.com

[7]The data, comprised of adjusted log returns, was downloaded from Yahoo Finance with symbols SPY (USA), EWC (Canada), EWW (Mexico), EWA (Australia), EWH (Hong Kong), EWJ (Japan), EWS (Singapore), EWG (Germany), EWQ (France), EWP (Spain), EWI (Italy), EWU (UK), EWL (Switzerland), EWD (Sweden).

Table 1.1: Summary statistics for 14 MSCI World iShares

	mean	sd	skewness	kurtosis	min	ARCH-LM(1)	Ljung-Box(1)	Ljung-Box(2)	JB
USA	2.49E-04	0.013	-0.03	11.8	-0.104	0.00	0.00	0.00	0
Canada	4.58E-04	0.016	-0.48	8.1	-0.116	0.00	0.02	0.03	0
Mexico	5.43E-04	0.022	-0.01	11.2	-0.186	0.00	0.71	0.00	0
Australia	4.05E-04	0.019	-0.17	11.4	-0.132	0.00	0.00	0.00	0
Hong.Kong	2.05E-04	0.021	0.30	10.1	-0.133	0.00	0.00	0.00	0
Japan	-2.81E-05	0.017	0.36	8.8	-0.109	0.00	0.00	0.00	0
Singapore	1.40E-04	0.021	0.20	8.4	-0.122	0.00	0.00	0.00	0
Germany	2.45E-04	0.018	0.08	10.1	-0.120	0.00	0.00	0.00	0
France	2.81E-04	0.017	-0.09	8.3	-0.116	0.00	0.00	0.00	0
Spain	3.95E-04	0.018	-0.05	8.7	-0.117	0.00	0.00	0.00	0
Italy	2.45E-04	0.018	-0.06	8.7	-0.112	0.00	0.00	0.00	0
UK	2.07E-04	0.016	-0.12	11.5	-0.128	0.00	0.00	0.00	0
Switzerland	2.60E-04	0.015	-0.20	6.9	-0.086	0.00	0.00	0.00	0
Sweden	3.71E-04	0.022	-0.14	7.5	-0.147	0.00	0.00	0.00	0

Notes to table 1.1: The Table presents summary statistic for the daily returns of 14 MSCI World iShares for the period 12/08/1996 to 02/03/2011, including the mean (mean), minimum (min), standard deviation (sd), skewness (skewness), kurtosis (kurtosis), the p-value of the ARCH-LM test of Engle (1982) with 1 lag, the p-value of the test of Ljung and Box (1978) for independence using 1 and 2-lags, and the p-value of the normality test of Jarque and Bera (1987).

returns). Notable among the country groups is the mostly positive skewness of Asian indices (excluding Australia), where mostly negative skewness is observed in all other indices as would be expected (see for example French, Schwert, and Stambaugh (1987) and Hong and Stein (2003)). Various theories have been put forward as to the reason for the positive skewness, such as short sale bans (see for example Bris, Goetzmann, and Zhu (2007)) and poor corporate governance (see for example Bae, Lim, and Wei (2006)). Kurtosis is evenly high across all indices, beyond what one would expect if the returns were normally distributed, and confirmed by almost zero p-values from the test of Jarque and Bera (1987) (column 'JB') which overwhelmingly rejects the null hypothesis of normally distributed returns. Another effect present in the dataset is that of non constant volatility, as evidenced by the rejection of the ARCH-LM test of Engle (1982) under the null hypothesis of no heteroscedasticity. In terms of autocorrelation in the returns, with the possible exception of Mexico, this is quite pronounced using both 1 and 2 lags, as the low p-values from the test of Ljung and Box (1978) show, under the null hypothesis of serial independence. Notably absent from the table is a test of time varying higher moments. One could consider extending the concept of autocorrelation to measures of autocoskewness and autocokurtosis[8], but it is not clear what the distribution of these measures are and hence how to make any meaningful inference. Instead, the presence or absence of ACD dynamics must be made post-estimation, for which a number of misspecification tests exist and are extensively used in the sections that follow.

1.4 The ACD-GH Model

Motion dynamics for the parameters form half of the modelling setup, with the choice of distribution forming the other half. The type of distribution chosen depends on it possessing certain desirable properties as mentioned in Section 1.1. Beyond the basic requirements, a general distribution which might contain as special cases other distributions, and also possess a multivariate counterpart is preferred. The Generalized Hyperbolic distribution (*GH*), introduced by Barndorff-Nielsen, Blæsild, Jensen, and Bagnold (1985) in the context of a sand project, is a variance-mean mixture of the normal and Generalized Inverse Gaussian (GIG) distributions. It is an extremely flexible distribution, allowing for skewness and fat tails, and nesting a large number of other distributions which have proved popular in the empirical modelling of financial asset returns, such as the Hyperbolic (*HYP*), Normal Inverse Gaussian (*NIG*), Variance Gamma (*VG*), (skew) Laplace and as limiting cases, the Normal and (skew) Student distributions. Tail flexibility is one particularly attractive feature of the GH model, which allows for modelling asymmetrically the upper and lowers tails. The Generalized Hyperbolic Skew Student (*GHST*) distribution for example, analyzed in Aas and Haff (2006), allows for the modelling of one heavy (with polynomial behavior) and one

[8]The $lag-l$ sample autocoskewness of x_t may be defined as:

$$\hat{\rho}_{l,2-1} = \frac{\sum_{t=l+1}^{T}\left[\left(x_t^2 - \hat{\mu}_{x^2}\right)\left(x_{t-l} - \hat{\mu}_x\right)\right]}{\sqrt{\sum_{t=1}^{T}\left(x_t^2 - \bar{\mu}_{x^2}\right)^2}\sqrt{\sum_{t=1}^{T}\left(x_t - \bar{\mu}_x\right)^2}} \tag{1.14}$$

The $lag-l$ sample autocokurtosis of x_t may be defined as:

$$\hat{\rho}_{l,3-1} = \frac{\sum_{t=l+1}^{T}\left[\left(x_t^3 - \hat{\mu}_{x^2}\right)\left(x_{t-l} - \hat{\mu}_x\right)\right]}{\sqrt{\sum_{t=1}^{T}\left(x_t^3 - \bar{\mu}_{x^2}\right)^2}\sqrt{\sum_{t=1}^{T}\left(x_t - \bar{\mu}_x\right)^2}} \tag{1.15}$$

semi-heavy (with exponential behavior) tail.[9] The GH distribution is part of an even larger family of distributions called the Normal Mean-Variance Mixture distributions, discussed in Barndorff-Nielsen, Kent, and Sörensen (1982).

Definition 1. The n-dimensional random variable $\boldsymbol{X}$ is said to have a normal mean-variance mixture distribution of the following form:

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\boldsymbol{A}\boldsymbol{Z}, \tag{1.16}$$

where $\boldsymbol{Z} \sim N_q(0, I_q)$, $W \in \mathbb{R}^1_+$, $\boldsymbol{A} \in \mathbb{R}^{n\times q}$, and $\boldsymbol{\mu}$, $\boldsymbol{\gamma} \in \mathbb{R}^n$. From the definition it follows that,

$$\begin{aligned} &\boldsymbol{X}\,|W \sim N_q\left(\boldsymbol{\mu} + W\boldsymbol{\gamma}, W\boldsymbol{\Sigma}\right), \\ &E\left(\boldsymbol{X}\right) = \boldsymbol{\mu} + E\left(W\right)\boldsymbol{\gamma}, \\ &Cov\left(\boldsymbol{X}\right) = E\left(W\right)\boldsymbol{\Sigma} + Var\left(W\right)\boldsymbol{\gamma}\boldsymbol{\gamma}', \end{aligned} \tag{1.17}$$

where $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}'$, and the mixing variable W is positive and has finite variance. A very useful property is that if the distribution of W is infinitely divisible, then the distribution of $\boldsymbol{X}$ is also infinitely divisible. This implies that there exists a Lèvy process with support over the entire real line, which is distributed at time $t = 1$ according to the law of X. Since the theoretical properties of Lèvy processes are well established, this translates into the possibility of formulating financial models directly in terms of such processes. A very popular choice for the mixing variable is the Generalized Inverse Gaussian (*GIG*) distribution, so that $W \sim GIG(\lambda, \chi, \psi)$[10], in which case the multivariate GH distribution is obtained, which depends on the three real parameters of the GIG distribution, the location ($\boldsymbol{\mu}$) and skewness ($\boldsymbol{\gamma}$) vectors in $\mathbb{R}^n$, and a positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$. The kurtosis (tail behavior), described by the λ and χ parameters, is driven by the univariate GIG mixing distribution and is therefore similar in all dimensions. I leave the definition and discussion of the n-dimensional case for Chapter 3, and follow Prause (1999) in defining the GH distribution in the 1-dimensional case.

Definition 2. The 1-dimensional Generalized Hyperbolic distribution, represents the mixture of $X\,|W$ with respect to W and given by:

[9] It is in fact the only distribution in the GH family to allow for one polynomial and one exponential tail.

[10] The χ and ψ parameters have also been represented as δ^2 and $\alpha^2 - \beta^2$ respectively in the literature.

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|W}(x|w) f_W(w)\, dw \\ &= \int_0^\infty N_{x|w}(\mu + \beta w, w)\, GIG_w\left(\lambda, \delta^2, \alpha^2 - \beta^2\right) dw \\ &= c(\lambda, \alpha, \beta, \delta)\left(\delta^2 + (x-\mu)^2\right)^{(\lambda - 1/2)/2} \times \mathbf{K}_{\lambda - 1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) e^{\beta(x-\mu)} \end{aligned} \tag{1.18}$$

$$c(\lambda, \alpha, \beta, \delta) = \frac{\left(\alpha^2 - \beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda - 1/2}\delta^\lambda \boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}$$

where N and GIG are the Normal and Generalized Inverse Gaussian distribution density functions respectively, and parameter domain of variation $0 \leqslant |\beta| < \alpha$, $\mu, \lambda \in \mathbb{R}$, $\delta > 0$ and $\mathbf{K}_\lambda$ is the modified bessel function of the third kind. The asymmetry of the GH distribution is purely down to the term $e^{\beta(x-\mu)}$ in the above definition. Special cases of the distribution are obtained by varying λ. For example, the NIG distribution has proved a very popular choice in the modelling of skewed and fat tailed financial asset returns and is obtained by setting λ to $-\frac{1}{2}$, while the HYP introduced in Eberlein and Keller (1995) with applications in option pricing is obtained by setting $\lambda = 1$. The GHST mentioned earlier is obtained by setting λ to $-\frac{\nu}{2}$ (with ν representing the degrees of freedom), and $\alpha \to |\beta|$, while the symmetric student distribution as $\beta \to 0$. Of particular note is that the Normal and (skew) Laplace can be represented as a limiting cases, and using the (χ, ξ) parametrization given below, when $\chi \to 1$ and $\chi \to 0$ respectively. The parameters of the distribution may be interpreted as location (μ), scale (δ), skewness (β) and shape (α and λ), thus allowing a richly parameterized setup for modelling the observed financial market features of asymmetry and likelihood of extreme events. A number of location and scale invariant parameterizations of the GH have been proposed in the literature,

$$\begin{aligned} &\zeta = \delta\sqrt{\alpha^2 - \beta^2}, \qquad \rho = \frac{\beta}{\alpha}, \\ &\xi = (1 - \zeta)^{-\frac{1}{2}}, \qquad \chi = \xi\rho, \\ &\bar{\alpha} = \alpha\delta, \qquad \bar{\beta} = \beta\delta. \end{aligned} \tag{1.19}$$

Bläsild (1981) proved that a linear transformation of the form $aX + b$ of a variable X distributed according to a GH distribution would again lead to a variable distributed with the same distribution and parameters $\lambda^* = \lambda$, $\alpha^* = \alpha / |a|$, $\beta^* = \beta / |a|$, $\delta^* = \delta\, |a|$, and $\mu^* = a\mu + b$. Therefore, for the modelling of (0,1) processes such as we find in models which are centered and scaled by their mean and standard deviation, one can use any of these location and scale invariant parametrization plus the following theoretical moment formulae for the Generalized Hyperbolic (needed to apply the centering and scaling):

$$\begin{aligned} E(\mathbf{X}) &= \mu + \frac{\beta\delta^2}{\sqrt{\alpha^2-\beta^2}} \frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}, \\ Var(\mathbf{X}) &= \delta^2 \left(\frac{\mathbf{K}_{\lambda+1}(\zeta)}{\zeta \boldsymbol{K}_{\lambda}(\zeta)} + \frac{\beta^2}{\alpha^2-\beta^2} \left[\frac{\mathbf{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} - \left(\frac{\mathbf{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} \right)^2 \right] \right). \end{aligned} \tag{1.20}$$

Prause (1999) suggests the use of the $(\bar{\alpha}, \bar{\beta})$ parametrization, which is adopted by Jensen and Lunde (2001) as well as Wilhelmsson (2009) in their GARCH and ACD - NIG models respectively. However, using either the (ζ, ρ) or (ξ, χ) parameterizations seems more natural as the two parameter representation is more directly linked with skewness and kurtosis[11]. In any case, moving between any of these parameterizations is a simple matter of applying the appropriate transformation. In Appendix A I provide the necessary formulae for scaling and centering the GH density in the (ζ, ρ) and (ξ, χ) parameterizations for use in GARCH type processes and the more simplified NIG standardization. In the ACD-GH model, the centered and scaled random variable z_t is conditionally distributed as a standardized GH, i.e., $GH(z_t; \lambda, \rho_t, \zeta_t)$, with the dynamics for the skew and shape parameters ρ_t and ζ_t defined as:

$$\begin{aligned} \rho_t &= -0.99 + \frac{1.98}{1+e^{-\bar{\rho}_t}} \\ \zeta_t &= 0.1 + \frac{24.9}{1+e^{-\bar{\zeta}_t}} \end{aligned} \tag{1.21}$$

where the effective bounds of the distributional parameters are $[-0.99, 0.99]$ and $[0.1, 25]$ for ρ and ζ respectively, while the GIG shape parameter λ is allowed to vary between $[-5, 5]$. Because most of the variation in kurtosis with respect to the shape parameter is obtained near the lower limit of ζ, care should be taken when choosing the upper limit in the presence of the transformation function lest too narrow a CDF range results when the bounds are too wide. The actual dynamics for the unconstrained parameters $\bar{\rho}_t$ and $\bar{\zeta}_t$ may be quadratic or piece-wise linear (or any other suitable representation for that matter), and I selectively consider both in the sections that follow. In order to transform the higher order parameters into higher order moments, I make use of the

[11] In fact, for the HYP distribution Barndorff-Nielsen and Bläsild (1981) show that skewness and kurtosis are approximately equal to 3χ and $3\xi^2$ respectively

moment generating function of the GH distribution,

$$\begin{aligned} M_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) &= e^{\mu u} M_{GIG\left(\lambda,\delta\sqrt{\alpha^2-\beta^2}\right)}\left(\frac{u^2}{2}+\beta u\right) \\ &= e^{\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda/2} \frac{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-(\beta+u)^2}\right)}{\boldsymbol{K}_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \end{aligned} \tag{1.22}$$

which for the NIG distribution we can greatly simplify because $\lambda = -0.5$, and the fact that $\boldsymbol{K}_\lambda(x) = \boldsymbol{K}_{-\lambda}(x)$ to obtain the skewness (S) and kurtosis (K),

$$\begin{aligned} K_{NIG} &= 3 + \frac{3\left(1+4\beta^2/\alpha^2\right)}{\delta\gamma} \\ S_{NIG} &= \frac{3\beta}{\alpha\sqrt{\delta\gamma}} \end{aligned} \tag{1.23}$$

where $\gamma = \sqrt{\alpha^2-\beta^2}$. Scott, Würtz, Dong, and Tran (2011) provide some new results and a quick recursion algorithm to calculate moments of any positive integer order for the GH.

The choice of λ in the GH Distribution

The *GIG* mixing distribution shape parameter λ is responsible for defining the effective variation in the shape and skew parameters of the GH. When the higher moments are not time varying, estimating all 3 parameters in a GARCH setup, while challenging, is quite feasible. However, when these parameters are time varying, there is a problem of identification and uniqueness since some combinations of λ with ζ_t or ρ_t will yield the same or very close likelihood. Further, since the likelihood surface becomes quite flat, particularly for values of ζ beyond a narrow range, this poses substantial estimation problems. It is therefore advisable to pre-specify λ. To illustrate, I consider, in Figure 1.1, contour plots of the log excess kurtosis [12] and skewness for 5 different values of λ and combinations of ρ and ζ. It is quite clear that when $\lambda = -0.5$, i.e. the NIG distribution, both kurtosis and skewness have a much larger region of variation, and similarly for $\lambda = -2$ and $\lambda = -4$. The excess kurtosis contour plot also reveals two additional insights. First, that the maximum kurtosis occurs as $\zeta \to 0$, for each value of

[12] The Log transformation is used to aid visual interpretation since differences in the admissible regions of kurtosis for the different λ values are very large.

λ. However, a more thorough investigation reveals that for values of λ below -2, we no longer have the nice stepped contour but instead it looks more like a plateau with steep peaks on either side of the extreme limits of ρ with the maximum kurtosis then being slightly above the minimum limit of ζ. Secondly, while for larger values of ζ the shape of the excess kurtosis contours looks similar for all values of λ, this rapidly changes as $\zeta \to 0$ and $\rho \to |1|$. For instance, when $\lambda = -2$, the contour near the lower limit of ζ is quite flattened meaning that a large change towards the extremes of ρ is required to move to a higher contour (i.e higher excess kurtosis) than would be required for the case when $\lambda = -0.5$. Similar arguments apply for the skewness contour plot. While I have considered here only a small range of values for λ, it is clear that in a higher moment time varying context, it may be preferable to fix this parameter to some value which offers a good range of values for skewness and kurtosis rather than estimating it, both because of the non-uniqueness of those moments for different combinations of the distributional parameters in a certain range, and the nonlinearity of the model. Alternatively, controlling the domain of variation of λ in the estimation procedure to a limited range is also a viable strategy, though care must be taken to ensure that the likelihood is not lower than any of the known subclasses which may be checked post-estimation (which is the strategy followed in this chapter).

1.5 Estimation

Following Hansen (1994), the log-likelihood function of ACD models can be written as,

$$\ln L\left(\theta | x_1, x_2, ., x_n\right) = \sum_{t=1}^{n} l_t\left(\theta\right), \tag{1.24}$$

where

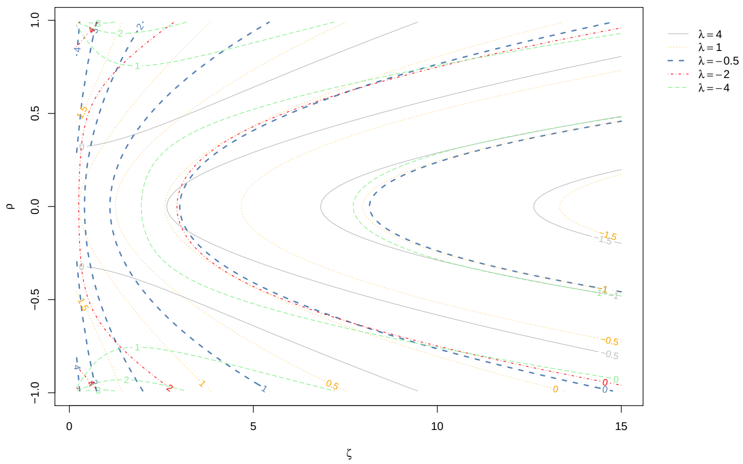
$$l_t\left(\theta\right) = \ln g\left(z_t\left(\theta\right) | \rho_t\left(\theta\right), \zeta_t\left(\theta\right)\right) - \ln \sigma\left(\theta, x_t\right). \tag{1.25}$$

The maximum likelihood estimate (MLE) of the model, $\hat{\theta}$, is obtained by maximizing the conditional log-likelihood (1.24). Assuming a correct specification, the likelihood scores,

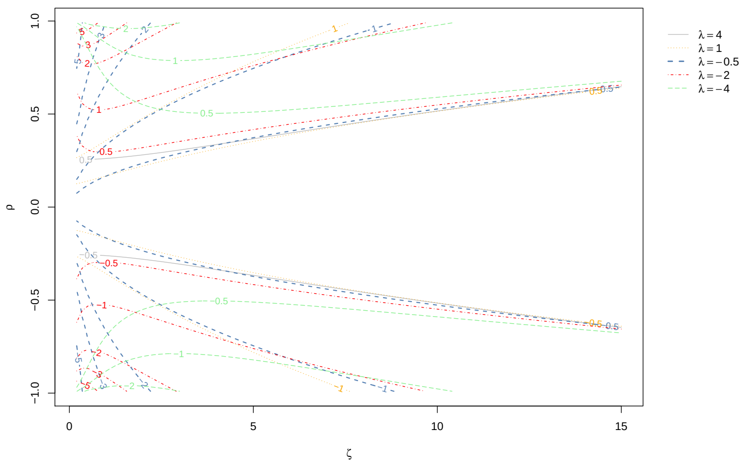
$$\frac{\partial}{\partial\theta} l_t\left(\theta\right) = \frac{\partial}{\partial\theta} \ln g\left(z_t\left(\theta\right) | \omega_t\left(\theta\right)\right) - \frac{\partial}{\partial\theta} \ln \sigma\left(\theta, x_t\right), \tag{1.26}$$

are martingale differences and have variance, V,

$$V = V\left(\theta_0\right), \qquad V\left(\theta\right) = -E\left(\frac{\partial}{\partial\theta\partial\theta'} l_t\left(\theta\right)\right) \tag{1.27}$$



(A) Excess Kurtosis



(B) Skewness

FIGURE 1.1: GH Skewness and Excess Kurtosis Contour Plots

where θ_0 denotes the true parameter value. Consistency of the MLE is obtained if $E(l_t(\theta)) < \infty$ and $E(\partial/\partial\theta)l_t(\theta) < \infty$, uniformly in θ, while asymptotic normality is obtained if $V < \infty$ and the likelihood is sufficiently well behaved in the neighborhood of θ_0. As noted by Hansen (1994), a proof of this in such a general setting will be hard to provide. However, it is instructive to obtain some insight into the distribution of the parameters, given different motion dynamics, since this will provide guidance, among other things, on the length of data required for modelling such processes with a certain degree of confidence. This issue is also considered in Chapter 4 (see Section 4.4.2) when generating simulated density forecasts from ACD type models. To assess this question of parameter consistency, I employ in the next subsection some standard simulation methods to investigate the properties of the parameter distributions and behavior as the length of the series, T, increases.

1.5.1 Simulated Parameter Consistency

Table 1.2 presents the average higher moment parameter estimates from data simulated with quadratic and piece-wise linear dynamics, for data sizes of 1000, 2000, 3000, 4000 and 6000. For each data size, 500 independent simulated paths were created and fitted under the different assumptions on the dynamics of the higher moments. The true parameters from which the data were simulated are given in the first line of each subtable (θ), while numbers in parenthesis under each estimates ($\hat{\theta}$) are the Root Mean Squared Error ($RMSE$). The parameters chosen were based on different ACD models fitted to the S&P 500 log returns for the period 10/03/1987 to 13/01/2003.

TABLE 1.2: Simulated parameter density and RMSE of ACD higher moment dynamics

PANEL A

	$NIG_{quad}(\rho_t,\zeta)$					$NIG_{quad}(\rho,\zeta_t)$				
	ζ	χ_0	χ_1	χ_2	ξ_1	ρ	κ_0	κ_1	κ_2	ψ_1
θ	**1.0000**	**-0.0320**	**0.4364**	**0.0774**	**0.5831**	**-0.6000**	**-0.3098**	**-0.0516**	**-0.0029**	**0.7806**
$\hat{\theta}_{1000}$	1.0288	-0.0356	0.4659	0.0852	0.5468	-0.6071	-0.3709	-0.0583	-0.0031	0.7369
	[0.24244]	[0.09434]	[0.14088]	[0.06652]	[0.17753]	[0.05093]	[0.44520]	[0.04655]	[0.00326]	[0.16281]
$\hat{\theta}_{2000}$	1.0248	-0.0318	0.4558	0.0801	0.5693	-0.6032	-0.3180	-0.0543	-0.0031	0.7693
	[0.16848]	[0.05866]	[0.09200]	[0.04023]	[0.10314]	[0.03176]	[0.18229]	[0.01238]	[0.00068]	[0.06645]
$\hat{\theta}_{3000}$	1.0138	-0.0313	0.4493	0.0778	0.5754	-0.6015	-0.3144	-0.0527	-0.0030	0.7745
	[0.13175]	[0.04651]	[0.07504]	[0.03056]	[0.08137]	[0.02602]	[0.15495]	[0.00970]	[0.00053]	[0.05219]
$\hat{\theta}_{4000}$	1.0116	-0.0308	0.4452	0.0775	0.5786	-0.6009	-0.3085	-0.0523	-0.0030	0.7784
	[0.11281]	[0.03782]	[0.06456]	[0.02562]	[0.06604]	[0.02281]	[0.09028]	[0.00808]	[0.00045]	[0.03256]
$\hat{\theta}_{6000}$	1.0056	-0.0306	0.4421	0.0771	0.5824	-0.6000	-0.3124	-0.0519	-0.0029	0.7786
	[0.08796]	[0.02983]	[0.04971]	[0.01980]	[0.05182]	[0.01860]	[0.13005]	[0.00673]	[0.00038]	[0.04413]

PANEL B

	$NIG_{pwl}(\rho_t,\zeta)$					$NIG_{pwl}(\rho,\zeta_t)$				
	ζ	χ_0	χ_1	χ_2	ξ_1	ρ	κ_0	κ_1	κ_2	ψ_1
θ	**1.0000**	**-0.0762**	**0.1426**	**0.4683**	**0.6191**	**-0.1086**	**-0.3678**	**0.1272**	**-1.0906**	**0.7642**
$\hat{\theta}_{1000}$	1.0613	-0.0766	0.1485	0.4560	0.6017	-0.1078	-0.7234	0.0135	-1.1506	0.6223
	[0.26179]	[0.08574]	[0.18083]	[0.20547]	[0.21352]	[0.06741]	[0.80938]	[0.54059]	[0.37507]	[0.27997]
$\hat{\theta}_{2000}$	1.0525	-0.0760	0.1482	0.4559	0.6120	-0.1050	-0.6115	0.0602	-1.1162	0.6647
	[0.18164]	[0.05704]	[0.11913]	[0.13461]	[0.15035]	[0.04672]	[0.63329]	[0.32997]	[0.25700]	[0.23000]
$\hat{\theta}_{3000}$	1.0356	-0.0763	0.1497	0.4566	0.6165	-0.1071	-0.5837	0.0543	-1.0970	0.6810
	[0.14129]	[0.04367]	[0.09329]	[0.11004]	[0.10970]	[0.03868]	[0.55987]	[0.28811]	[0.19110]	[0.19702]
$\hat{\theta}_{4000}$	1.0327	-0.0753	0.1481	0.4552	0.6209	-0.1047	-0.5731	0.0749	-1.0894	0.6846
	[0.12370]	[0.03693]	[0.08017]	[0.09528]	[0.09124]	[0.03283]	[0.54302]	[0.22428]	[0.16618]	[0.19570]
$\hat{\theta}_{6000}$	1.0245	-0.0742	0.1470	0.4503	0.6294	-0.1044	-0.5213	0.1005	-1.0916	0.7038
	[0.09739]	[0.02762]	[0.06502]	[0.07981]	[0.07692]	[0.02652]	[0.44998]	[0.12714]	[0.14317]	[0.16529]

Notes to table 1.2: The Table presents the average higher moment parameter estimates from data simulated with dynamics given by the subtable headings, for data lengths of 1000, 2000, 3000, 4000 and 6000. For each data size, 500 independent simulated paths were created and fitted under the different assumptions on the dynamics of the higher moments. The true parameters from which the data were simulated are given in the first line of each subtable (θ), while numbers in parenthesis under each estimates ($\hat{\theta}$) are the RMSE values.

Assuming $\sqrt{N}$ consistency, the rate of change of the RMSE as the length increases from T_N to T_{N+1} should be approximately equal to $\sqrt{T_N/T_{N+1}}$. Taking for instance the RMSE for $T = 6000$ and comparing it with $T = 2000$, the average decrease in the RMSE for the parameters is 0.56 which is very close to the expected RMSE decrease of $0.58(\sqrt{2000/6000})$. The notable exceptions are the autoregressive parameters in the case of the time varying shape dynamics ζ_t in the piecewise linear models which show closer to cubic root consistency. Dark (2006) considered higher moment time variation using the Generalized Skewed Student density of Hansen (1994) and a range of GARCH models including the symmetric GARCH, asymmetric power ARCH ($APARCH$) and a Hyperbolic APARCH model for long memory processes. In a Monte Carlo study of the parameter behavior and simulated distribution, he found similar results in that the skew parameter was well behaved while the shape parameter was not in terms of RMSE, using both quadratic and restricted non quadratic dynamics. Because this exercise was carried out for only one set of parameters per model, it is hard to generalize to all cases particularly because of the nonlinear transformation which affects quite strongly parameters close to the distribution bounds. In addition, it is rare to see any papers on ACD models publishing results of such simulations and as such this is one area which could certainly benefit from more research. The implications would be that for some combination of model dynamics and parameters a lot more data is required in order to obtain the same degree of confidence as in GARCH models.

1.5.2 Inference and Goodness of Fit

Even though the GH is an extremely flexible distribution, the are many different types of dynamics and alternative models which might fit the underlying data and belong to the domain of the 'correct' model. It is therefore recommended to report and use the robust standard errors of White (1982) which produce asymptotically valid confidence intervals by calculating the covariance of the parameters V as:

$$\hat{V} = -(A)^{-1}B(-A)^{-1}, \tag{1.28}$$

where

$$\begin{aligned} A &= L''\left(\hat{\theta}\right), \\ B &= \sum_{i=1}^{n} g_i\left(x_i\left|\hat{\theta}\right.\right)^T g_i\left(x_i\left|\hat{\theta}\right.\right), \end{aligned} \tag{1.29}$$

which is the Hessian and covariance of the scores at the optimum. The robust standard errors are the square roots of the diagonal of V.

To investigate how well the time varying higher moments fit real data, I perform an LR test on the log returns of the 14 MSCI index iShares, introduced in Section 1.3, for the full dataset period 12/08/1996 to 02/03/2011. The null is the restricted model whereby the skew and shape parameters ($[\chi_1, \chi_2, \xi_1]$ and $[\kappa_1, \kappa_2, \psi_1]$ respectively) excluding the intercepts are zero, which is effectively the GARCH model without time varying higher moment dynamics. Table 1.3 reports the results of the test together with all parameters and their respective robust p-values. Because of the presence of autocorrelation in the return series, an AR(2) model was used to filter the conditional mean, the estimation of which was performed in a joint step.[13] Starting with the conditional mean, with the exception of the USA and Canada, the intercept and autoregressive parameters are all significant at the 10% level. The 3 GARCH parameters (ω, α_1, β_1) are significant for all the securities with persistence hitting the estimation program's constrained upper bound of 0.99.[14] This is likely the result of some structural break over this long and specific time period considered rather than an indication of the presence of integrated GARCH dynamics. Finally, with the exception of Canada[15], all indices display some degree of time varying higher moment dynamics as evidenced by their p-values.

While the likelihood of the ACD model will always be higher than that from a GARCH model, given that the latter is a restricted version of the former, it is important to consider the marginal value in the fit. In this particular case, and for the time periods considered, the p-values from the LR test indicate a clear rejection of the restricted GARCH model at the 10% significance level, in all cases, in favor of the ACD dynamics. However, this does not immediately translate into out-of-sample out-performance vis-a-vis a GARCH model with respect to some operational measure such as VaR exceedances or Expected Shortfall (ES), which is why I consider a comparative empirical application in Section 1.5.6, unlike the vast majority of the literature reviewed which has been restricted to some in-sample inference procedures which even then are not always very informative. For example, Hansen (1994) suggested the use of the parameter constancy test of Nyblom (1989) as an additional diagnostic test. This is a Lagrange Multiplier (LM) test of the null hypothesis that the parameters are constant against the alternative that the parameters follow a martingale process, with a joint parameter test looking at the stability of the whole parameter vector. The problem with such a test, is that

[13] While it is also possible to perform the estimation in 2 steps, the 1 step approach provides more efficient parameter estimates, particularly for not very large datasets.

[14] Persistence in the simple GARCH dynamics used is the sum of α_1 and β_1, and constrained to be less than or equal to 0.99 in the estimation procedure used.

[15] It cannot be discounted that the lack of significance may also be a result of a local solution being reached by the optimizer, and this is taken up in more detail in the next subsection on optimization.

the null hypothesis will often be rejected since it is well known that structural breaks, particularly when including long periods of data for estimation, are likely to create instability in the intercepts of all the dynamics (including the mean, variance and higher moment parameters). These structural breaks do not in themselves denote time varying dynamics and hence such a test may prove misleading. Instead, misspecification tests such as those used in Section 1.6 are more likely to be relevant and informative.

Table 1.3: ACD parameter estimates for 14 MSCI World iShares

	USA	**Canada**	**Mexico**	**Australia**	**Hong Kong**	**Japan**	**Singapore**	**Germany**	**France**	**Spain**	**Italy**	**UK**	**Switzerland**	**Sweden**
μ	7.24E-05	9.06E-04	1.05E-03	5.47E-04	7.51E-04	7.22E-05	6.22E-04	6.54E-04	5.17E-04	5.84E-04	3.99E-04	4.62E-04	4.65E-04	8.32E-04
	[0.73]	*[0.44]*	*[0.00]*	*[0.01]*	*[0.00]*	*[0.71]*	*[0.01]*	*[0.00]*	*[0.02]*	*[0.02]*	*[0.10]*	*[0.02]*	*[0.01]*	*[0.00]*
ar_1	-0.06	-0.05	0.00	-0.14	-0.10	-0.09	-0.11	-0.07	-0.07	-0.07	-0.09	-0.13	-0.14	-0.04
	[0.13]	*[0.85]*	*[0.83]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.01]*	*[0.09]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.03]*
ar_2	-0.07	-0.02	-0.05	-0.05	-0.04	-0.02	-0.07	-0.01	-0.03	-0.02	-0.02	-0.07	-0.04	-0.03
	[0.29]	*[0.97]*	*[0.00]*	*[0.01]*	*[0.02]*	*[0.10]*	*[0.00]*	*[0.27]*	*[0.64]*	*[0.48]*	*[0.19]*	*[0.01]*	*[0.01]*	*[0.11]*
ω	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.02]	*[0.60]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.01]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.01]*	*[0.00]*	*[0.01]*
α_1	0.09	0.06	0.10	0.07	0.09	0.07	0.10	0.08	0.07	0.07	0.09	0.07	0.08	0.07
	[0.00]	*[0.07]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*
β_1	0.90	0.93	0.89	0.91	0.90	0.92	0.89	0.91	0.92	0.92	0.90	0.91	0.91	0.92
	[0.00]	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*
χ_0	-0.02	-0.11	-0.16	-0.01	-0.08	-0.01	-0.27	-0.08	-0.03	-0.11	-0.03	-0.09	-0.25	-0.08
	[0.36]	*[0.95]*	*[0.00]*	*[0.35]*	*[0.01]*	*[0.42]*	*[0.00]*	*[0.11]*	*[0.76]*	*[0.45]*	*[0.83]*	*[0.23]*	*[0.02]*	*[0.09]*
χ_1	-0.01	0.16	0.13	-0.02	0.12	0.00	0.21	0.08	0.02	-0.11	-0.06	0.07	0.18	0.08
	[0.83]	*[0.86]*	*[0.04]*	*[0.04]*	*[0.02]*	*[0.80]*	*[0.03]*	*[0.08]*	*[0.81]*	*[0.31]*	*[0.40]*	*[0.49]*	*[0.04]*	*[0.29]*
χ_2	0.01	0.06	0.02	0.00	0.06	0.01	0.00	0.01	-0.04	0.00	-0.05	0.05	-0.16	0.03
	[0.13]	*[0.87]*	*[0.26]*	*[0.52]*	*[0.00]*	*[0.43]*	*[0.68]*	*[0.55]*	*[0.55]*	*[0.90]*	*[0.19]*	*[0.25]*	*[0.48]*	*[0.16]*
ξ_1	0.99	0.83	0.53	1.00	0.83	0.99	0.00	0.76	0.78	0.55	0.76	0.84	0.00	0.78
	[0.00]	*[0.79]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.91]*	*[0.00]*	*[0.00]*	*[0.13]*	*[0.05]*	*[0.00]*	*[0.99]*	*[0.00]*
κ_0	-0.29	-0.90	-0.31	-0.81	-2.37	-1.87	-1.04	-0.70	-0.92	-0.90	-2.12	-0.20	-1.10	-0.37
	[0.18]	*[0.98]*	*[0.00]*	*[0.34]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.17]*	*[0.09]*	*[0.00]*	*[0.24]*	*[0.04]*	*[0.42]*
κ_1	-1.00	-0.71	-0.35	-0.34	-0.57	-0.37	-0.57	-1.00	-0.85	-0.30	-0.54	-0.34	-0.72	-0.40
	[0.02]	*[0.95]*	*[0.00]*	*[0.16]*	*[0.02]*	*[0.34]*	*[0.00]*	*[0.00]*	*[0.16]*	*[0.18]*	*[0.06]*	*[0.12]*	*[0.16]*	*[0.04]*
κ_2	-0.15	0.01	-0.08	-0.09	0.21	-0.07	0.04	0.08	0.11	0.04	-0.02	-0.06	0.23	-0.05
	[0.28]	*[0.99]*	*[0.00]*	*[0.16]*	*[0.01]*	*[0.60]*	*[0.13]*	*[0.10]*	*[0.69]*	*[0.87]*	*[0.73]*	*[0.45]*	*[0.42]*	*[0.71]*
ψ_1	0.80	0.61	0.85	0.61	0.00	0.00	0.58	0.64	0.59	0.66	0.00	0.88	0.52	0.81
	[0.00]	*[0.96]*	*[0.00]*	*[0.09]*	*[0.99]*	*[1.00]*	*[0.00]*	*[0.00]*	*[0.04]*	*[0.00]*	*[1.00]*	*[0.00]*	*[0.02]*	*[0.00]*
LL (GARCH)	11401.96	10551.45	9490.18	10184.74	9718.38	10298.49	9635.51	10214.47	10356.70	10176.02	10289.88	10629.21	10627.19	9499.30
LL (ACD)	11458.70	10569.09	9501.89	10195.22	9727.88	10307.76	9645.63	10229.41	10370.00	10181.68	10300.53	10638.04	10635.19	9506.57
LR_{stat}	113.48	35.28	23.43	20.97	18.99	18.54	20.24	29.89	26.60	11.33	21.29	17.65	16.00	14.54
p-value	***0.000***	***0.000***	***0.001***	***0.002***	***0.004***	***0.005***	***0.003***	***0.000***	***0.000***	*0.079*	***0.002***	***0.007***	***0.014***	***0.024***

Notes to table 1.3: The Table presents parameter estimates of an AR(2)-GARCH(1,1)-NIG(ρ_t, ζ_t) for the daily log returns of 14 MSCI World iShares for the period 12/08/1996 to 02/03/2011. The ACD NIG dynamics (ρ_t, ζ_t) were estimated using a quadratic model as in Equation (1.10). Values in square brackets represent the p-values from the estimated robust standard errors. The Log-Likelihood of each model ($LL(ACD)$) is reported as well as the Log-Likelihood of the restricted GARCH model ($LL(GARCH)$) where the restriction is of constant skew and shape. The Likelihood Ratio statistic (LR_{stat}) under the null of the restricted model is distributed χ^2_6, with the 6 restrictions representing the dynamic model skew and shape parameters excluding their intercepts. Tested at the 5% level of significance, the GARCH model is rejected in 13 of the 14 securities tested.

1.5.3 Optimization Strategy

The nonlinear transformation required to constrain the higher moment parameters within their distribution specific bounds creates certain challenges in the estimation process. The likelihood surface is no longer smooth necessitating a global optimization approach to solving such problems, and is quite typical in the nonlinear dynamics literature. A number of authors, including Hansen (1994) and Jondeau and Rockinger (2003), have cited the use of a hierarchical type strategy to obtain estimates of the non time varying skew and shape parameters as starting values to a second stage, where the skewness and shape may be estimated incrementally, assessing their significance separately and then jointly. I follow a similar strategy, with certain additional enhancements which I have found to provide some more confidence of optimality. Namely, the parameters from a GARCH model are estimated and used as starting parameters in the second stage ACD model, with the non time-varying skew and shape parameters serving the role of the higher moment dynamics recursion starting values (after transforming from the constrained to the unconstrained domain by inverting the logistic transformation function). Because of the sensitivity of the solution to the starting parameters, I have found that a random search multi-start optimization algorithm offers the best outcome amongst competing methods in getting close to a viable global optimum within an acceptable time limit. I make use of results from Hu, Shonkwiler, and Spruill (1994) who provide for strong arguments in favor of sampling the parameter space from the uniform distribution based on each parameter's upper and lower bounds, evaluating the likelihood at these randomly sampled points, ranking the results and then starting the solver from these different starting points. For this purpose, I use an augmented Lagrange based solver with an SQP interior step method, described in Ye (1997)[16].

1.5.4 ACD Forecasting and Simulation

While there is a common source for the shocks in the ACD models, that arising from the innovations process and affecting the variance, skewness and shape parameters, it is nevertheless time-varying, unlike in GARCH models. This means that the process of simulation requires as many evaluations to the random number generator of the underlying conditional distribution as there are simulated samples. Compare this with

[16]The solver is implemented in the R package *Rsolnp* of Ghalanos and Theussl (2012) and available on the Comprehensive R Archive Network (CRAN). See Appendix F for details.

GARCH models where one call to the generator is made at the beginning of the simulation to generate the N samples required. That is, in order to generate the next sample z_{t+n}, we must know the value of z_{t+n-1}, which means that the process of sampling from ACD models is more time consuming[17] and involved than models with time invariant higher moment parameters.

The 1-step ahead forecast for ACD models is given by the same filtering mechanism as that used in the maximum likelihood fitting phase, since GARCH type models automatically generate 1-step ahead forecasts. For n-step ahead forecasts, the variance forecast should converge to its long run value given that $E(z_t) = 0$, $E(z_t^2) = 1$ and therefore $E(\varepsilon_t^2 | \Omega_{t-1}) = E(\sigma_t^2 z_t^2) = \sigma_t^2$ giving $\sigma_{t+n}^2 = \omega + (\alpha_1 + \beta_1)\sigma_t^2$. For the higher moment n-ahead forecasts these must be derived iteratively since there is no closed form solution in the presence of the nonlinear transformation.

1.5.5 Higher Moment News Impact Curves

The concept of a news impact curve was introduced by Engle and Ng (1993), and provides a visual representation of the impact of shocks on the time varying variance. More specifically, it has been used to compare the role of asymmetric response of variance to positive and negative shocks from which the sign bias tests were developed by the same authors. While ACD models considered here do not provide for any variation in terms of asymmetry in variance or the higher moments, I do extend the concept of the role of standardized shocks on the higher time varying moments as a stepping stone to a surface function in a multivariate extension in Chapter 3. It also serves as a useful diagnostic tool in deciding on the types of dynamics to choose. Figure 1.2 shows the higher moment news impact curves for an ACD-NIG model, with both quadratic and piecewise linear dynamics for the shape and skew parameters. The first curve is of the unconstrained dynamics, before the logistic transformation, the next 2 represent the constrained dynamics and distributional moments, calculated taking into account all the higher moment distributional parameters. Using this visual diagnostic, it appears to make more sense to use quadratic dynamics for the skew parameter, providing for a greater response with respect to shocks, and is the strategy followed in the empirical application of this chapter. Because the news impact curve depends on certain long run relationships to calculate, simulation methods are necessary in the presence of the

[17] This is particularly true when sampling from distributions which have expensive random number generators.

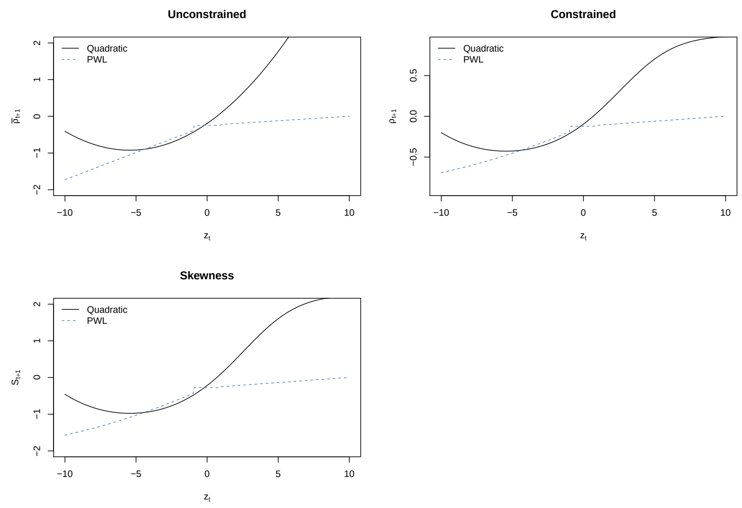
nonlinear transformation. Specifically, I simulate from the model to obtain the long run value of the unconstrained higher moment parameters which I then use to evaluate the news impact, which for the quadratic case is:

$$\begin{aligned}\zeta_{t+1} &= \Phi(\bar{\zeta}_{t+1}) = \Phi\left(\chi_0 + \chi_1 z_t + \chi_2 z_t^2 + \xi_1 \bar{\zeta}^*\right),\\ \rho_{t+1} &= \Phi(\bar{\rho}_{t+1}) = \Phi\left(\kappa_0 + \kappa_1 z_t + \kappa_2 z_t^2 + \psi_1 \bar{\rho}^*\right),\end{aligned} \tag{1.30}$$

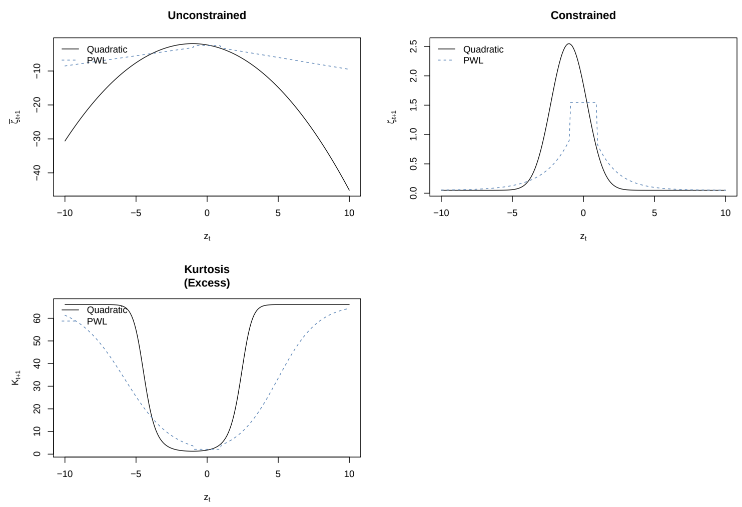
where $\bar{\zeta}^*$ and $\bar{\rho}^*$ are the simulated long run unconstrained shape and skew parameters respectively, and $\Phi(.)$ the CDF function for bounding transformation. The process of translating the resulting curves into standardized skewness and kurtosis is then a simple matter of applying the appropriate formulae from the GH moment generating function.

1.5.6 Competing Distributions

The distributions used in the ACD literature have mostly been limited to some variation of the Student distribution. For example, Hansen (1994) and Jondeau and Rockinger (2003) have used the Generalized Student distribution, Harvey and Siddique (1999) the noncentral Student distribution, Lambert and Laurent (2001a) a skewed Student distribution, while Brooks, Burke, Heravi, and Persand (2005) a standard Student distribution to model only kurtosis. Departures from the Student variations have included a Pearson Type IV distribution in Brannas and Nordman (2003), an Entropy distribution in Rockinger and Jondeau (2002) and a Gram-Charlier expansion of the Normal in León, Rubio, and Serna (2005). More recently, Wilhelmsson (2009) has used the NIG distribution with ACD dynamics in a risk management exercise and shown that compared to a number of competing models without higher moment dynamics, the ACD-NIG model was the only one which could not be rejected as capturing the correct number of VaR exceedances for the S&P500 in a test spanning a long time period. In this section, I consider a number of interesting distributions with ACD dynamics and compare their out-of-sample performance with equivalent GARCH models, using operational measures relevant to risk management. The distributions used are the NIG, HYP and omnibus GH, the Skew Student, and Johnson's SU (see Johnson (1949)) distribution (*JSU*) reparametrized and described in Rigby and Stasinopoulos (2005). The Skew Student distribution (*SSTD*) is based on the inverse scale factor transformations of Fernandez and Steel (1998), with details on its standardization for use in GARCH processes given in Appendix B. To obtain some insight into the type and magnitude of the skewness and kurtosis generated by these distribution, Figure 1.3 illustrates the



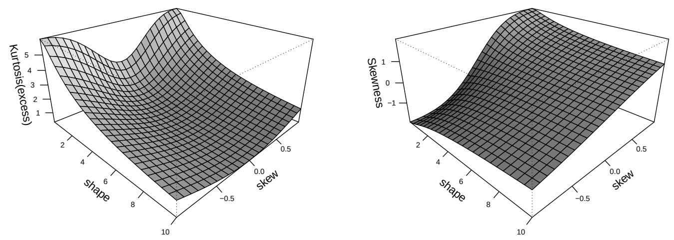
(A) Skewness News Impact



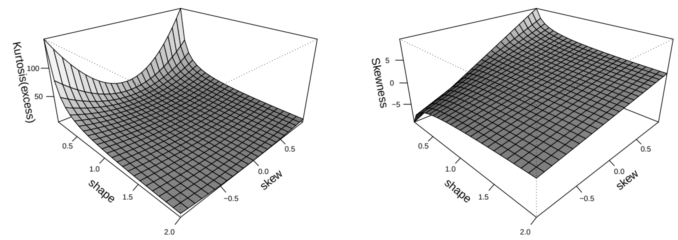
(B) Excess Kurtosis News Impact

Figure 1.2: Higher Moment News Impact Curves

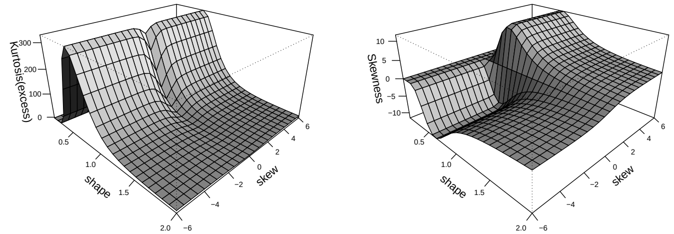
skewness and kurtosis surfaces for different combinations of these distributions' higher moment parameters. The lower bound on the shape parameter is controlled in this setup to enable better illustration.



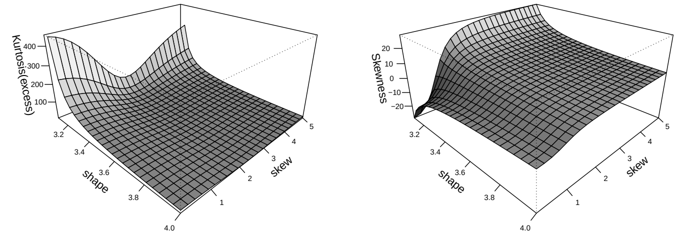
(A) HYP



(B) NIG



(C) JSU



(D) SSTD

FIGURE 1.3: Skewness and Kurtosis Surfaces

In all cases, as the shape approaches a certain bound, kurtosis increases rapidly. For the skewed Student distribution, a skew parameter of 1 translates to no asymmetry while in the JSU case as the shape approaches zero and skew= 0 the distribution approaches the Normal. The GH distribution is not shown since the additional GIG mixing parameter λ makes any such representation difficult in 3 dimensions. Instead, the HYP distribution is shown as a popular modelling choice within the GH family. While the HYP distribution shows the least variation in skewness and kurtosis, the other 3 distributions allow for a very wide range in both, and well beyond what is likely to be observed in practice. While skewness and kurtosis in the HYP, NIG and SSTD distributions achieve their maximum variation as the skew and shape parameters approach their limits, the JSU distribution achieves maximum variation in the region of 0.5. In all cases, the maximum variation is achieved in a very narrow range of values for the skew and shape parameters which should therefore act as a guideline for optimization algorithms seeking to impose some bounds.
The out-of-sample application is based on the 14 MSCI iShares described earlier, spanning the period 12/08/1996 to 02/03/2011. Starting on 17/03/2000, each model was estimated using all available data up-to that point and 1-ahead rolling forecasts for the next 250 days generated. The process was repeated by increasing the data window[18] up to the last period, resulting in 3000 out-of-sample rolling conditional density forecasts. For the GH distribution, the λ shape parameter for the ACD model was fixed to the value obtained from the estimated GARCH-GH model so that the comparison would be based on the same sub-distribution. An AR(2) filter was used for the conditional mean equation, and quadratic dynamics for the skew and shape parameters of the ACD model. To compare the models I made use of some popular tail related tests including the conditional coverage VaR exceedances test of Christoffersen (1998) at the 1% quantile and the ES test of McNeil and Frey (2000) at the 5% quantile, described in more detail in Appendix C. The results are presented in Tables 1.4 and 1.5. It is difficult to conclude that the ACD and GARCH with conditional distribution the NIG provide for substantial differences with respect to VaR exceedances. For none of the indices is the conditional coverage test rejected, with the possible exception of Japan for the GARCH model, and the same applies to the ES test as evidenced by the high p-values. In the case of the HYP distribution, there are marginally too many exceedances in the

[18]In this application I have used an expanding rather than moving window, unlike the multivariate application of Chapter 3 where a moving window is important to capture changes in the co-movements of the factors.

case of the ACD model for Australia and marginally too few in the case of the GARCH model for Japan. Too many exceedances would translate to too little risk capital allocated, which is usually penalized by regulators, whilst too few exceedances means an inefficient use of capital, which over the long run will be punished by shareholders. The results for the GH, JSU and SSTD distributions are similar to the NIG distribution, with the GARCH model for Japan once again being marginally rejected in terms of VaR exceedances. The underestimation of VaR exceedances for the Japan index is the same for all GARCH models, irrespective of the distribution, and this is likely related to the 'whipsaw' pattern of the index for this period for which the strategy of re-estimating the model every 250 days is too long and would probably require a more frequent window size. This is not a problem for the ACD models which can accommodate any such shortcomings as a result of the added flexibility of tail dynamics from the shape parameter. Finally, for the ACD with conditional distribution SSTD, for both Australia and Spain, there are too many exceedances generated indicating perhaps spurious higher moment dynamics. As can be concluded from this application, it is not immediately obvious, either from the preliminary summary statistics nor the in-sample fit (even when covering the whole period) whether an ACD model will outperform out-of-sample the equivalent GARCH model, a point obviously missed by most of the papers on covering such models.

Table 1.4: Out-of-sample VaR and density forecast tests for 14 MSCI World iShares (NIG, HYP and GH)

	USA	**Canada**	**Mexico**	**Australia**	**Hong Kong**	**Japan**	**Singapore**	**Germany**	**France**	**Spain**	**Italy**	**UK**	**Switzerland**	**Sweden**
ACD (NIG)														
$VaRExceed_{1\%}$	22	26	24	34	21	21	27	34	26	35	32	34	29	26
$VaR(cc)p-value$	0.26	0.60	0.43	0.52	0.19	0.19	0.67	0.52	0.60	0.13	0.66	0.54	0.74	0.60
$ES_{5\%}$ p-value	0.96	0.94	0.69	0.33	0.71	0.88	0.42	0.64	0.76	1.00	0.95	0.65	0.37	0.81
GARCH (NIG)														
$VaRExceed_{1\%}$	23	27	21	29	22	18	23	34	25	32	31	33	29	29
$VaR(cc)p-value$	0.34	0.67	0.19	0.56	0.26	0.05	0.34	0.52	0.52	0.66	0.71	0.60	0.74	0.74
$ES_{5\%}$ p-value	0.98	0.97	0.67	0.40	0.55	0.90	0.52	0.56	0.77	0.32	0.52	0.65	0.39	0.65
ACD (HYP)														
$VaRExceed_{1\%}$	32	29	27	41	22	24	31	32	26	41	32	36	25	30
$VaR(cc)p-value$	0.61	0.74	0.67	0.09	0.26	0.43	0.71	0.66	0.60	0.05	0.66	0.43	0.52	0.74
$ES_{5\%}$ p-value	0.77	0.83	0.40	0.08	0.60	0.77	0.32	0.46	0.41	0.13	0.43	0.42	0.65	0.53
GARCH (HYP)														
$VaRExceed_{1\%}$	23	28	23	33	22	19	23	34	25	33	31	33	28	29
$VaR(cc)p-value$	0.34	0.72	0.34	0.59	0.26	0.09	0.34	0.52	0.52	0.59	0.71	0.60	0.72	0.74
$ES_{5\%}$ p-value	0.96	0.90	0.56	0.31	0.50	0.87	0.46	0.45	0.74	0.32	0.48	0.62	0.34	0.63
ACD (GH)														
$VaRExceed_{1\%}$	31	28	29	33	26	22	28	39	26	36	31	36	31	32
$VaR(cc)p-value$	0.61	0.72	0.74	0.60	0.60	0.26	0.72	0.17	0.60	0.12	0.71	0.43	0.71	0.66
$ES_{5\%}$ p-value	0.98	0.93	0.62	0.39	0.47	0.94	0.58	0.38	0.48	0.28	0.62	0.50	0.34	0.44
GARCH (GH)														
$VaRExceed_{1\%}$	22	28	22	28	22	18	23	33	25	33	31	33	28	29
$VaR(cc)p-value$	0.26	0.72	0.26	0.50	0.26	0.05	0.34	0.60	0.52	0.59	0.71	0.60	0.72	0.74
$ES_{5\%}$ p-value	0.99	0.95	0.84	0.49	0.58	0.92	0.65	0.64	0.75	0.45	0.57	0.69	0.40	0.55

Notes to table 1.4: The Table presents comparative out-of-sample tail based forecast tests of 14 MSCI iShares for the period 12/08/1996 to 02/03/2011 based on AR(2) first order GARCH and ACD (with quadratic dynamics for the skew and shape) models with conditional distributions the NIG, HYP and GH. Starting on 17/03/2000, each model was estimated using all available data up-to that point and 1-ahead rolling forecasts for the next 250 days generated. The process was repeated up to the last period resulting in 3000 out-of-sample rolling forecasts from which were calculated the VaR exceedances at the 1% coverage, together with their respective p-values based on the conditional coverage (cc) test of Christoffersen (1998), the ES test of McNeil and Frey (2000) at the 5% coverage (p-values shown).

TABLE 1.5: Out-of-sample VaR and density forecast tests for 14 MSCI World iShares (SSTD and JSU)

	USA	**Canada**	**Mexico**	**Australia**	**Hong Kong**	**Japan**	**Singapore**	**Germany**	**France**	**Spain**	**Italy**	**UK**	**Switzerland**	**Sweden**
ACD (JSU)														
$VaRExceed_{1\%}$	35	40	27	38	24	25	31	33	31	37	38	35	33	31
$VaR(cc)p-value$	0.49	0.13	0.67	0.23	0.43	0.52	0.71	0.60	0.71	0.11	0.23	0.13	0.60	0.71
$ES_{5\%}$ p-value	0.78	0.66	0.71	0.33	0.68	0.72	0.31	0.70	0.60	0.35	0.25	0.44	0.42	0.66
GARCH (JSU)														
$VaRExceed_{1\%}$	23	27	23	32	22	18	23	34	25	32	31	33	29	29
$VaR(cc)p-value$	0.34	0.67	0.34	0.61	0.26	0.05	0.34	0.52	0.52	0.66	0.71	0.60	0.74	0.74
$ES_{5\%}$ p-value	0.98	0.97	0.73	0.39	0.53	0.93	0.59	0.59	0.81	0.41	0.59	0.66	0.37	0.62
ACD (SSTD)	USA	Canada	Mexico	Australia	Hong Kong	Japan	Singapore	Germany	France	Spain	Italy	UK	Switzerland	Sweden
$VaRExceed_{1\%}$	30	26	22	43	22	21	28	32	28	40	32	35	27	32
$VaR(cc)p-value$	0.59	0.60	0.26	0.04	0.26	0.19	0.72	0.66	0.72	0.06	0.66	0.49	0.67	0.66
$ES_{5\%}$ p-value	0.93	0.97	0.81	0.12	0.81	0.94	0.52	0.69	0.62	0.35	0.50	0.62	0.73	0.80
GARCH (SSTD)														
$VaRExceed_{1\%}$	23	28	25	34	24	18	25	36	25	33	31	34	32	30
$VaR(cc)p-value$	0.34	0.72	0.52	0.54	0.43	0.05	0.52	0.37	0.52	0.59	0.71	0.52	0.61	0.74
$ES_{5\%}$ p-value	0.98	0.97	0.73	0.29	0.52	0.94	0.61	0.55	0.74	0.46	0.46	0.67	0.34	0.50

Notes to table 1.5: The Table presents comparative out-of-sample tail based forecast tests of 14 MSCI iShares for the period 12/08/1996 to 02/03/2011 based on AR(2) first order GARCH and ACD (with quadratic dynamics for the skew and shape) models with conditional distributions the Skew-Student of Fernandez and Steel (1998) (SSTD) and the Johnson's SU (JSU) distribution (see Johnson (1949)). Starting on 17/03/2000, each model was estimated using all available data up-to that point and 1-ahead rolling forecasts for the next 250 days generated. The process was repeated up to the last period resulting in 3000 out-of-sample rolling forecasts from which were calculated the VaR exceedances at the 1% coverage, together with their respective p-values based on the conditional coverage (cc) test of Christoffersen (1998), the ES test of McNeil and Frey (2000) at the 5% coverage (p-values shown).

The VaR exceedances test may be considered a rather crude method to capturing the tail risk differences in the models, as it distinguishes only on the basis of integer exceedances and requires a large amount of out-of-sample data to avoid the possibility of data-snooping bias. A more informative way to compare the models is by using the model confidence set (MCS) approach of Hansen, Lunde, and Nason (2011), described in Appendix C.3, and using a tail based loss function which is able to distinguish the average magnitude of the exceedances thus providing a more complete picture of comparative model performance. I follow Gonzalez-Rivera, Lee, and Mishra (2004) and define a statistical loss function used in quantile estimation, which for a given coverage level α is defined as,

$$Q_{loss} \equiv N^{-1} \sum_{t=R}^{T} \left(\alpha - 1\left(r_{t+1} < VaR_{t+1}^{\alpha}\right)\right)\left(r_{t+1} - VaR_{t+1}^{\alpha}\right), \tag{1.31}$$

where $P = T - R$ is the out-of-sample forecast horizon, T the total horizon to include in estimation, and R the start of the out-of-sample forecast. This is an asymmetric loss function, linearly penalizing exceedances more heavily by $(1-\alpha)$. Because of the non-differentiable nature of the indicator function **1**, I adopt the recommendation of Gonzalez-Rivera, Lee, and Mishra (2004) and replace it with the approximation:

$$1\left(r_{t+1} < VaR_{t+1}^{\alpha}\right) \approx \left[1 + \exp\left\{\delta\left(r_{t+1} - VaR_{t+1}^{\alpha}\right)\right\}\right]^{-1} \tag{1.32}$$

which is found to very closely match the indicator function for values of δ equal to 25.[19] Table 1.6 reports the probability of the 10 different models being in the MCS for each of the 14 country indices using a coverage level of 5%. The results are much more informative and provide for a much clearer picture of where the ACD models provide for superior performance. There is a clear rejection of the GARCH models for Canada, Mexico, Hong Kong, Singapore and Germany. Interestingly, the ACD model with conditional distributions the NIG, GH and JSU are also rejected for the Hong Kong index, where the ACD-HYP is almost certainly the superior model, whilst the ACD-SSTD cannot be rejected as belonging to the MCS. Considering that the ACD-NIG and ACD-HYP models had almost the same number of exceedances in Table 1.4, and the ACD-GH had an exceedance value closer to the expected one of 30 than either of the others, this is quite a surprising result and may be due to calibration issues with respect to the distribution bounds. In the case of Spain and the ACD-NIG model,

[19]That is, when dealing with percentages, otherwise for decimals use 2500.

considering the fact that all other ACD models belong to the MCS, this is clearly a case of a badly-fitting distribution rather than the absence of time varying higher moments.

Table 1.6: VaR model comparison on 14 MSCI World iShares

VaR Loss (5%)	ACD (NIG)	ACD (HYP)	ACD (GH)	ACD (SSTD)	ACD (JSU)	GARCH (NIG)	GARCH (HYP)	GARCH (GH)	GARCH (SSTD)	GARCH (JSU)
USA	0.65	0.65	1.00	0.65	0.06	0.42	0.42	0.40	0.40	0.42
Canada	0.82	1.00	0.82	0.82	0.03	0.03	0.01	0.01	0.01	0.01
Mexico	1.00	0.41	0.41	0.41	0.41	0.00	0.00	0.00	0.00	0.00
Australia	1.00	0.71	0.71	0.71	0.80	0.71	0.80	0.71	0.71	0.71
Hong Kong	0.04	1.00	0.07	0.33	0.07	0.03	0.03	0.03	0.04	0.04
Japan	0.95	0.84	0.95	1.00	0.95	0.32	0.32	0.84	0.32	0.32
Singapore	0.85	0.85	0.85	1.00	0.15	0.03	0.03	0.03	0.04	0.04
Germany	0.47	0.45	0.45	1.00	0.77	0.05	0.05	0.05	0.05	0.05
France	0.48	0.13	0.13	1.00	0.48	0.39	0.39	0.39	0.13	0.13
Spain	0.03	0.97	0.92	1.00	0.84	0.97	0.97	0.97	0.92	0.95
Italy	0.48	0.98	1.00	0.98	0.98	0.91	0.95	0.88	0.88	0.88
UK	0.40	1.00	0.18	0.19	0.40	0.40	0.71	0.19	0.38	0.40
Switzerland	0.38	1.00	0.38	0.38	0.38	0.38	0.38	0.35	0.38	0.38
Sweden	1.00	0.44	0.47	0.44	0.44	0.44	0.44	0.44	0.44	0.44

Notes to table 1.6: The Table reports the probability of being in the model confidence set of Hansen, Lunde, and Nason (2011) for each of the 14 MSCI World iShares based on the VaR loss function, at the 5% quantile, defined in Gonzalez-Rivera, Lee, and Mishra (2004). Density forecast for each of the 14 MSCI indices for the period 12/08/1996 to 02/03/2011 were calculated based on AR(2) first order GARCH and ACD (with quadratic dynamics for the skew and shape) model with conditional distributions the NIG, HYP, GH, the Skew-Student (SSTD) of Fernandez and Steel (1998) and Johnson's SU (JSU) distribution (see Johnson (1949)). Starting on 17/03/2000, each model was estimated using all available data up to that point and 1-ahead rolling forecasts for the next 250 days generated. The process was repeated up to the last period resulting in 3000 out-of-sample rolling forecasts from which was calculated the VaR at the 5% quantile for each model and from which the relevant loss function was extracted and compared per country index.

1.6 The Cost of GARCH

Evidence for the presence of higher moment dynamics was presented in the previous section through an empirical application where the benefits where assessed through certain operational measures. It was shown that for some of the indices, there was a clear benefit to using such dynamics over GARCH models and, perhaps just as important, little penalization in using them even when there was no clear out-performance. In this section, the more general question of the cost of using GARCH dynamics in the presence of time varying higher moments is considered. Since ACD models generalize GARCH dynamics to higher moment distributional parameters, choosing which parameters to keep, and hence the dynamics describing the model, is equivalent to evaluating the significance of those fitted parameters, using an information criterion such as the Akaike Information Criterion (AIC) or undertaking an LR based test as in Section 1.5.2. If on the other hand one were to ignore possible higher moment dynamics and simply fit a GARCH model, then a key question would be to decide, in the ACD framework, whether the model was misspecified, and what the cost would be in terms of some operational tail based distribution measure. I investigate in this section, through a small Monte Carlo study, various tests of misspecification relevant to these types of models, as well as the possible cost of assuming constant higher moments when the underlying dynamics are clearly not.

1.6.1 The BDS test of i.i.d.

A key assumption and requirement of most econometric models is that the standardized innovations, that is the standardized noise left after filtering out the underlying dynamics driving the observed process, are i.i.d. Violation of this assumption usually implies a misspecified model which has not adequately captured the underlying dynamics, making it difficult to make correct inferences from the unconditional distribution of the resulting model. The BDS test of Brock, Dechert, and Scheinkman (1993) examines the spatial dependence of a series by embedding that series into an m-dimensional space and counting the near points (defined as those points for which the distance is less than some user defined value ϵ). Formally, the spatial correlation is computed using the correlation integral as:

$$\begin{aligned} C_{\epsilon,m} &= N_m^{-1}\left(N_m^{-1}-1\right)\sum_{i\neq j} I_{i,j}, \\ I_{i,j} &= \begin{cases} 1 & \text{if } \left\| x_i^m - x_j^m \right\| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad . \end{aligned} \tag{1.33}$$

A series is then i.i.d if $C_{\epsilon,m} \approx [C_{\epsilon,1}]^m$, with their difference defined as the BDS statistic and distributed asymptotically standard Normal. In one of the first papers to introduce chaotic dynamics to the finance community, Hsieh (1991) investigated the power of the BDS test to detect different types of dynamics giving rise to non i.i.d. behavior. It was found that the test had good power to detect linear dependence, non-stationarity, chaotic dynamics and non-linear stochastic processes. Using a set of stock indices and CRSP value weighted decile portfolios, Hsieh found that most of the non-i.i.d. behavior was driven by the non-linear dynamics induced by conditional heteroscedasticity[20]. In a further paper, running a large number of power tests, Brock and Dechert (1988) showed the ability of the BDS test to detect nonlinearities in GARCH, NLMA, and TAR models, and given further support in Barnett, Gallant, Hinich, Jungeilges, Kaplan, and Jensen (1997). Nevertheless, the test has two major problems in implementation. One is the choice of the embedding dimension m and the proximity parameter ϵ, with various pairs of the parameters giving rise to different values. Hence without a general guideline as to which combination of parameters to use, it is sometimes difficult to draw clear conclusions. The other problem, examined in Brock and Dechert (1988) and De Lima (1996) is the conditions under which the BDS statistic is nuisance-parameter free. Effectively, what this implies for GARCH models is that in the case of the standard model of Bollerslev (1986), a simple log transformation of the squared standardized residuals from the fitted model allows to place that model in the class of linear additive models for which the BDS statistic is nuisance-parameter free. For all other models not fitting this category, the distribution of the test statistic must be simulated for each set of model parameters estimated.

In Table 1.7 I present a simulation study where an AR(2)-GARCH(1,1) model with conditional distribution NIG(ρ, ζ) is fitted to series simulated from 4 different models. As a test case, the first model is simply an AR(2)-GARCH(1,1) with conditional distribution NIG(ρ, ζ) (i.e. the fitted model). Under the assumption that the test correctly identifies the model as the correct one, then the test should reject the model 5% of the time using a 95% confidence level[21]. With the exception of very low proximity parameters and high embedding dimensions, the test is able to correctly identify the test case with an average percent rejection rate of 5%. For the other models, the situation is not as clear cut. When the dynamics include a time varying shape parameter, the

[20] Though to fully explain the dynamics, a stochastic volatility type representation was used

[21] Under a correctly identified model, the p-values of the test in the simulation should also be uniformly distributed.

Table 1.7: GARCH BDS test under alternative dynamics

	AR(2)-GARCH(1,1)-NIG(ρ, ζ)					AR(2)-GARCH(1,1)-NIG(ρ, $\zeta_t[1,1,1]$)				
	ϵ/σ					ϵ/σ				
m	0.5	1	1.5	2	2.5	0.5	1	1.5	2	2.5
2	4.8	4.3	4.4	4.7	4.9	41.0	38.6	34.8	30.8	25.3
3	5.6	4.7	4.6	4.5	4.3	49.9	47.6	43.4	37.9	31.1
4	8.2	5.0	4.6	4.4	4.4	52.2	48.9	45.1	39.8	32.9
5	16.8	5.2	4.8	5.1	4.5	52.8	49.2	45.2	40.3	33.1
6	38.2	5.2	5.0	5.2	4.5	55.3	47.6	43.3	38.4	32.1
7	63.4	6.8	4.9	4.6	4.7	69.9	45.5	42.0	36.8	31.1
8	81.3	10.3	4.9	4.6	4.6	82.5	43.6	39.5	34.9	29.8
9	98.0	17.7	5.4	4.7	4.8	100.0	45.0	37.3	32.9	28.5
10	100.0	30.6	6.4	4.7	4.7	100.0	49.6	35.9	32.2	27.4

	AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, ζ)					AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, $\zeta_t[1,1,1]$)				
	ϵ/σ					ϵ/σ				
m	0.5	1	1.5	2	2.5	0.5	1	1.5	2	2.5
2	6.1	5.8	6.0	5.9	6.0	48.6	46.3	41.4	36.3	29.8
3	6.9	6.3	5.9	5.6	5.8	52.7	50.7	47.0	42.3	35.3
4	9.3	6.1	5.2	5.3	5.8	52.6	49.4	46.5	41.9	35.2
5	18.0	5.8	5.3	5.3	5.6	52.1	48.3	45.4	40.9	34.3
6	37.0	5.6	5.4	5.2	5.3	53.5	45.9	43.5	39.4	32.5
7	64.0	7.5	5.4	4.9	4.7	67.3	43.6	40.8	37.1	31.6
8	81.5	11.2	5.8	5.0	4.7	82.8	44.0	38.5	35.3	30.1
9	97.7	18.5	6.5	5.0	4.7	100.0	44.1	37.2	33.4	28.9
10	100.0	33.6	7.1	4.7	4.7	100.0	48.8	35.5	31.6	27.0

Note: The table reports the percentage of rejections for the BDS test of i.i.d. (with embedding dimensions m 2 to 10 and ϵ representing the range of standard deviations of the data) at the 95% confidence level, when applied to the log of the squared standardized residuals of the GARCH-NIG model from simulated data under alternative data generating processes. The Monte Carlo experiment used 2000 simulations with 8000 observations each from the AR(2)-GARCH(1,1) model with conditional densities given by NIG(ρ,ζ), NIG(ρ,$\zeta_t[1,1,1]$), NIG($\rho_t[1,1,1]$,ζ) and NIG($\rho[1,1,1]$,$\zeta_t[1,1,1]$).

rejection rate is certainly higher but nowhere near conclusive. In the case of time varying skew parameter, there is almost no difference from the test case implying that such dynamics do not give rise to nonlinearity affecting the i.i.d. of the standardized residuals. As a result, we may conclude that for the case of both time varying skew and shape parameters, the degree of rejection is almost purely the result of the latter dynamics. In the NIG distribution, both skew and shape parameters jointly determine higher moments, with the shape parameter having a more direct impact on higher even moments, and the skew on higher odd moments. It is clearly the case that in the presence of time variation in the higher parameters the standardized residuals are no longer identically distributed. However, whether it is a failing of the BDS methodology in general to pick this up wholly or whether the marginal contribution of the time varying parameters to the generation of a non-i.i.d. process is too small to be picked up, under this distribution, is an open question. In the other tests that follow, the time variation in the higher moments is therefore directly considered as an alternative to a general test like the BDS which may not be sensitive enough to capture this type of misspecification.

1.6.2 GMM Orthogonality Test

The GMM type moment (orthogonality) tests of Hansen (1982) have been applied to test the adequacy of ACD models in Harvey and Siddique (1999) and Jondeau and Rockinger (2003). Under a correctly specified model, certain population moment conditions should be satisfied and hold in the sample using the standardized residuals. The moment conditions can be tested both individually using a t-test or jointly using a Wald test. Formally, the following moment conditions can be tested:

$$\begin{array}{lcl} M_1 & E\left[z_t\right] & =0, \\ M_2 & E\left[z_t^2-1\right] & =0, \\ M_3 & E\left[z_t^3\right] & =0, \\ M_4 & E\left[z_t^4-3\right] & =0, \\ Q_2 & E\left[\left(z_t^2-1\right)\left(z_{t-j}^2-1\right)\right] & =0, \\ Q_3 & E\left[\left(z_t^3\right)\left(z_{t-j}^3\right)\right] & =0, \\ Q_4 & E\left[\left(z_t^4-3\right)\left(z_{t-j}^4-3\right)\right] & =0, \end{array} \tag{1.34}$$

where $j = 1 \ldots, p$ is the lag and usually p is set to 4. The last 3 conditions test the conditional variance, skewness and kurtosis using p lags and may be tested using a Wald test distributed χ^2 with p d.o.f. It is also possible to test all the conditions jointly using a Wald test distributed χ^2 with $4 + 3p$ d.o.f. Table 1.8 reports the results of applying the orthogonality test to standardized residuals of the fitted AR(2)-GARCH(1,1) model with conditional distribution NIG(ρ, ζ) under 4 different DGPs. In a reverse situation to the BDS test discussed previously, the orthogonality test appears to have good power when the skew parameter is time varying, but very poor power when the shape parameter is time varying. Strangely enough, the test also rejects the joint case in the base model which should not happen since it represents the correct model. Ergun and Jun (2010) find that the test does have some problem detecting misspecification, particularly in the dynamics of the fourth moment, under the skew Student distribution.

1.6.3 Non-Parametric Transition Density Test

The non parametric density test of Hong and Li (2005) provides for a powerful misspecification test, making few assumptions about the underlying dynamics. Ergun and Jun (2010) used this to test ACD model misspecification using Hansen's Generalized-Student distribution and compared the size and power of this test with that of the GMM type test described in the previous section. They found that it had significantly

Table 1.8: GARCH orthogonality tests under alternative dynamics

AR(2)-GARCH(1,1)-NIG(ρ, ζ)

	$E[z]$	$E[z^2]-1$	$E[z^3]$	$E[z^4]-3$	Q_2	Q_3	Q_4	**Joint**
Mean	0.000	-0.001	0.001	-0.014				
s.e.	0.000	0.001	0.009	0.240				
t-value	0.00	-0.03	0.01	-0.03	4.13	3.89	9.65	44.00
%Rejections: H0					7	3	44	79

AR(2)-GARCH(1,1)-NIG(ρ, $\zeta_t[1,1,1]$)

	$E[z]$	$E[z^2]-1$	$E[z^3]$	$E[z^4]-3$	Q_2	Q_3	Q_4	**Joint**
Mean	-0.005	0.006	-0.055	1.052				
s.e.	0.000	0.001	0.033	6.899				
t-value	-0.41	0.23	-0.30	0.40	4.10	3.97	5.76	24.78
%Rejections: H0					4	2	16	37

AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, ζ)

	$E[z]$	$E[z^2]-1$	$E[z^3]$	$E[z^4]-3$	Q_2	Q_3	Q_4	**Joint**
Mean	-0.002	0.003	-0.009	0.269				
s.e.	0.000	0.001	0.011	0.374				
t-value	-0.19	0.14	-0.09	0.44	4.24	15.29	9.32	96.22
%Rejections: H0					8	87	41	99

AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, $\zeta_t[1,1,1]$)

	$E[z]$	$E[z^2]-1$	$E[z^3]$	$E[z^4]-3$	Q_2	Q_3	Q_4	**Joint**
Mean	-0.006	0.009	0.180	2.914				
s.e.	0.000	0.001	0.200	261.213				
t-value	-0.56	0.30	0.40	0.18	4.44	10.96	6.23	65.33
%Rejections: H0					8	58	19	96

Note: The table reports average mean, standard error and t-values of the orthogonality moment conditions described in Hansen (1982) of the fitted standardized residuals from the GARCH-NIG model under alternative Data Generating Processes (*DGP*). The Monte Carlo experiment uses 2000 simulations with 8000 observations each from the AR(2)-GARCH(1,1) model with conditional densities given by NIG(ρ,ζ), NIG(ρ,$\zeta_t[1,1,1]$), NIG($\rho_t[1,1,1]$,ζ) and NIG($\rho[1,1,1]$,$\zeta_t[1,1,1]$). Q_2, Q_3, Q_4 are the Wald tests for the joint significance of $E[(z_t^2-1)(z_{t-j}^2-1)]$, $E[(z_t^3)(z_{t-j}^3)]$ and $E[(z_t^4-3)(z_{t-j}^4-3)]$, $j=1,\ldots,4$, respectively and the t-value included is distributed χ^2 with 4 (no. lags) d.o.f. The final column in each table ('Joint') is the Wald test of joint nullness of all the conditions with t-value distributed again as χ^2 with 16 d.o.f. The percentage of rejections of the null hypothesis is also reported for the 2000 simulations from each DGP.

more power than the latter, and since parameter uncertainty has no impact on the asymptotic distribution of the test statistic, it is quite robust to model misspecification under quite general conditions. The details of the test appear in Appendix C and I focus instead on the results in this section. Table 1.9 reports the average statistics for $M(m,l)$ under the first four moments and using 4 lags, and $\hat{W}(4)$, from the Monte Carlo simulation fitting an AR(2)-GARCH(1,1) model with conditional distribution NIG(ρ, ζ) under 4 different DGPs. The test correctly captures the base model with almost no rejections for the moments based tests, and 8% total rejections under the portmanteau test statistic which is close to the 95% confidence level tested. When the shape parameter is time varying, the average value for the third and fourth moment tests, as well as the portmanteau W test are well above the critical value of 1.645 representing the 95% quantile of the standard Normal density. In the case of only time varying skew parameter, the same cannot be said, with the statistics falling well within a correctly specified model. That is, when GARCH dynamics with constant skew and shape parameter from the NIG conditional density are used to model dynamics generated from the same model but with a time varying conditional skew parameter, the test would appear to suggest that there is no explicit cost for this misspecification. This is in direct contrast to the GMM orthogonality test discussed previously, but in line with the results of the BDS test. Finally, when both higher moment parameters are time varying, the test statistics overwhelmingly reject the null hypothesis of correct specification. Interestingly, the third moment test now also rejects the null which is not surprising since the NIG shape and skew parameters jointly determine the higher moments, with variation in both creating a scenario where it is more likely that time varying skewness is more pronounced when both parameters are time varying or both at the limits.

Table 1.9: GARCH Hong-Li tests under alternative dynamics

AR(2)-GARCH(1,1)-NIG(ρ, ζ)					
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
Mean	1.904	-1.701	-1.418	-1.181	-0.926
%Rejections: H0	0	0	1	2	8
AR(2)-GARCH(1,1)-NIG(ρ, $\zeta_t[1,1,1]$)					
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
Mean	-1.799	-0.760	2.431	5.556	2.156
%Rejections: H0	0	3	55	84	57
AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, ζ)					
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
Mean	-1.236	-1.447	-1.127	-0.704	1.061
%Rejections: H0	0	1	2	7	33
AR(2)-GARCH(1,1)-NIG($\rho_t[1,1,1]$, $\zeta_t[1,1,1]$)					
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
Mean	-0.828	-0.964	1.689	3.536	4.895
%Rejections: H0	1	2	44	68	94

Note: The table reports the average value of the Hong and Li (2005) statistic from the Monte Carlo experiment using 2000 simulations with 8000 observations each from the AR(2)-GARCH(1,1) model with conditional densities given by NIG(ρ,ζ), NIG(ρ,$\zeta_t[1,1,1]$), NIG($\rho_t[1,1,1]$,ζ) and NIG($\rho[1,1,1]$,$\zeta_t[1,1,1]$). $M(j,j), j = 1,..,4$, represents the nonparametric test for misspecification in the conditional moments of the standardized residuals from the fitted AR(2)-GARCH-NIG model, and distributed as $N(0,1)$ under the null of a correctly specified model. The statistic W in column 5 of the table is the Portmanteau type test statistic for general misspecification (using 4 lags) and distributed as $N(0,1)$ under the null of a correctly specified model.

1.6.4 Value at Risk and Tail Events

In practice, what the inclusion of time variation in higher moments achieves is to marginally increase the flexibility of capturing extreme tail events and time varying asymmetry. While the tests considered thus far are in-sample misspecification tests, and may be used to test a model prior to its overall usage, it is also important to consider the implication in a more operational setting. Specifically, I consider the out-of-sample fit of GARCH forecasts when the underlying higher moment dynamics are time varying. While the general density test of Berkowitz (2001) provides for a good measure of the overall fit of the model, it is important to consider a test of the tail fit which is where the cost of ignoring time variation in higher moments, particularly the shape parameter, is most likely to manifest in practise. For this purpose, I choose the tail test of Berkowitz (2001) since it provides for a reasonable indication of tail fit and is less likely to depend on the size of the dataset as in the case of the VaR tests discussed in Section 1.5.6. I simulate 1000 independent runs of size 4000 each, from an AR(2)-GARCH(1,1) model with conditional distribution NIG(ρ_t, ζ) , NIG(ρ, ζ_t) and NIG(ρ_t, ζ_t), for a total of 3000 simulated series representing varying degrees of higher moment dynamics, where the same ACD estimates were used as in the previous sub-sections. For each of the series I fit an AR(2)-GARCH(1,1) model with conditional distribution NIG(ρ, ζ) in an out-of-sample rolling forecast application using a starting window size of 2000 and moving the window every 25 days[22] for a total of 2000 out-of-sample density forecasts. Table 1.10 presents the results of this study where the median p-value from the full density and tail tests (at the 5% coverage) of Berkowitz (2001) are given together with the number of rejections of the null under the respective tests with 95% confidence. For the full density test, time variation in the shape

Table 1.10: GARCH Berkowitz density tests under alternative dynamics

$ACD[\rho,\zeta_t]$	$Berkowitz_1$	$Berkowitz_{0.05}$
Median p-value	0.071	0.014
% Rejections	59	81
$ACD[\rho_t,\zeta]$	$Berkowitz_1$	$Berkowitz_{0.05}$
Median p-value	0.288	0.026
% Rejections	21	73
$ACD[\rho_t,\zeta_t]$	$Berkowitz_1$	$Berkowitz_{0.05}$
Median p-value	0.077	0.020
% Rejections	56	75

Note: The table reports the median p-value of the Berkowitz (2001) full density and tail tests (at the 5% quantile) under the null that an AR(1)-GARCH(1,1)-NIG(ρ,ζ) model correctly fits the data out of sample. Using 1000 randomly generated scenarios of size 4000, the last 2000 points were left for out of sample 1-ahead rolling forecasts, using a moving window of size 2000 and re-estimating the GARCH model every 25 periods. The simulated data were generated from an AR(1)-GARCH(1,1) model with higher moment conditional densities given by NIG(ρ,$\zeta_t[1,1,1]$), NIG($\rho_t[1,1,1]$,ζ) and NIG($\rho[1,1,1]$,$\zeta_t[1,1,1]$), using quadratic dynamics (see Equation (1.10)). The table also reports the percent rejection of the null at the 95% confidence level, across all scenarios.

[22] That means that the model is refitted every 25 days and for every fit, 25 rolling 1-ahead forecasts are generated.

parameter appears to be more important, similarly to the results in Section 1.6.3, and with only about 50% rejection. However, when it comes to the tail test, the presence of either shape or skew dynamics leads to a strong rejection of the GARCH model with values well above 70%. This means that the presence of time varying higher moments is unlikely to be well modelled by GARCH dynamics alone, particularly with respect to measures depending on the tail of the distribution.

1.7 Conclusion

The ARMA-GARCH framework is one of the most flexible and popular methods for the parametric modelling of the conditional density of portfolio returns. However, its inability to accommodate extreme swings in returns and changes over time to the asymmetry of the conditional distribution has important consequences for applications depending on forecasts from these models, such as typically found in risk management and portfolio applications . The majority of the literature on the use of ACD models has been mainly restricted to in-sample model evaluation on a small cross section of returns. In this chapter I have provided an out-of-sample application on a larger cross section of international equity indices using a range of relevant conditional distributions. The evidence points to visible benefits in the inclusion of time varying higher dynamics within a GARCH framework, using a range of operational measures, and importantly does not penalize for their inclusion in the absence, or lack of significant prevalence, of such dynamics. Using a Monte Carlo experiment, I have also provided evidence that when time varying higher moments are actually present, there is a high cost to ignoring them using a range of misspecification tests in-sample as well as tail based tests out-of-sample. The computational challenges arising from the presence of highly nonlinear bounding transformations should no longer hinder the use of ACD models since recent advances to global optimization strategies have taken advantage of parallel processing making such problems much more approachable and feasible. Some open questions remain in this field, such as tests for the detection of such dynamics, the asymptotic properties of the parameters under different motion dynamics and the generalization to a feasible multivariate model. I take up the challenge of providing a feasible multivariate extension in Chapter 3.

Chapter 2

Multivariate GARCH Dynamics and Dependence

Portfolio allocation and risk management require the modelling of a diverse and possibly large universe of assets. The goal is usually to obtain some measure of a linear combination of these assets, either in-sample or forecasted. Modelling the linear combination directly as a univariate problem is certainly one viable option. However, this approach throws away information, which may be contained in the multivariate structure of the assets, that may aid in inference and understanding. More importantly, when considering portfolio allocation decisions where weights are not predetermined, this approach is rendered effectively worthless. Instead, one is interested in expressing the multivariate dependence in terms of either a set of simulated scenarios, which approximate a point in time multivariate density, else the conditional moment and co-moment matrices, such as the mean and covariance. Given these measures of the multivariate distribution, it should then be possible, subject to the properties of that distribution, to obtain the weighted marginal density, or an approximation to that, from which any type of measure can then be extracted.

In the classical model of market behavior, models of risk and portfolio allocation are usually based on a static Capital Asset Pricing Model ($CAPM$) or multifactor based regressions. The popularity of these models, particularly the multifactor extensions of the CAPM (see for example Ross (1976)), is evidenced by the large number of investment professionals still using them and supported by numerous software and research companies.[1] While these models may be adequate when using low frequency data such as monthly and quarterly returns, the evidence on the weekly and higher frequency scale suggests that there exist dynamic interactions within and between securities which cannot be modelled by traditional approaches (see for example Bollerslev, Engle, and Wooldridge (1988) for a dynamic CAPM based model). It is therefore important to have a model which captures as much of these interactions as possible, and from which meaningful forecasts can be generated to aid in decision making. One of the first models for capturing dynamic interactions emerged from industry in the form of J.P.Morgan's Riskmetrics methodology (see Morgan (1994)), which was effectively a restricted multivariate GARCH ($MGARCH$) model with fixed parameters. While GARCH models have proved very successful in capturing some of the most salient features of the observed market phenomena, their extension to the multivariate domain has not enjoyed the same level of success, largely due to dimensionality issues and feasibility of estimation. The direct multivariate extensions have sought to generalize the univariate dynamics into matrix form to capture the complex dependencies in the conditional covariance. Unfortunately, this has come at the cost of introducing a very large number of parameters making their estimation infeasible and impractical for even a small set of securities. Reduction in the problem dimension has been achieved through the indirect and factor models by taking advantage of relationships inherent in elliptical distributions or compromises in the complexity of relationships modelled. These more feasible models, coupled with flexible and feature rich multivariate distributions, have provided portfolio and risk managers with an invaluable framework for the modelling of risk. This chapter provides a review of MGARCH models, with a particular focus on the DCC model originally proposed by Engle (2002) and its extension in terms of dynamics and distributions. I try to address the question of dimensionality versus feasibility and the cost, if any, of certain compromises used to address the former. Surprisingly, I find that the 2-stage approach typically used in DCC modelling offers certain advantages beyond estimation ease, contrary to the arguments of Caporin and McAleer (2012), and discuss possible reasons. In a departure from the more general review setup, I also undertake an empirical investigation using the MSCI dataset of Chapter 1 to consider the relative in-sample goodness of fit of the diagonal BEKK and diagonal AGDCC models with a range of different distributions, and attempt to form some general conclusions on their relative merits. Finally, a small Monte Carlo exercise is used

[1] Such as MSCI Barra and Northfield Information Services.

to compare the DCC and BEKK models, and consider, if any, the cost of switching one for the other. This chapter is structured as follows: Section 2.1 considers the extension of the univariate GARCH dynamics to the multivariate domain with a focus on the Asymmetric Generalized Dynamic Conditional Correlation (*AGDCC*) model, and in Section 2.2 the DCC-copula model is considered. The challenge of finding feasible distributions with desirable properties is considered in Section 2.3 with a focus on the normal mean-variance mixture family and the Generalized Asymmetric Laplace (*GAL*) in particular. Section 2.4 examines more closely the diagonal AGDCC model within an empirical application and contrasts it with a diagonal BEKK model under alternative distributions. Section 2.5 concludes.

2.1 Multivariate GARCH

The generalization of univariate GARCH models to the multivariate domain is conceptually simple. Consider the stochastic vector process, $\boldsymbol{x}_t$ $\{t = 1, 2, \ldots, T\}$ of financial returns with dimension $N \times 1$ and mean vector $\boldsymbol{\mu}_t$[2], given the information set $\boldsymbol{I}_{t-1}$:

$$\boldsymbol{x}_t \,|\boldsymbol{I}_{t-1} = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{2.1}$$

where the residuals of the process are modelled as:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{z}_t, \tag{2.2}$$

and $\boldsymbol{H}_t^{1/2}$ is an $N \times N$ positive definite matrix such that $\boldsymbol{H}_t$ is the conditional covariance matrix of $\boldsymbol{x}_t$[3], and $\boldsymbol{z}_t$ an $N \times 1$ i.i.d. random vector, with centered and scaled first 2 moments:

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{z}_{\mathrm{t}}\right] &= 0, \\ \mathrm{Var}[\boldsymbol{z}_{\mathrm{t}}] &= \boldsymbol{I}_{\mathrm{N}}, \end{aligned} \tag{2.3}$$

[2]The mean vector may for example be derived from a VARMA model or may simply represent the unconditional means of the financial returns.

[3]One way to obtain the square root matrix is through the singular value decomposition of $\boldsymbol{H}_t$.

with $\boldsymbol{I}_N$ denoting the identity matrix of order N. The conditional covariance matrix $\boldsymbol{H}_t$ of $\boldsymbol{x}_t$ may be defined as:

$$\begin{aligned}\mathrm{Var}\left(\boldsymbol{x}_{\mathrm{t}}\left|\boldsymbol{I}_{\mathrm{t}-1}\right.\right) = \mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{x}_{\mathrm{t}}) &= \mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{\varepsilon}_{\mathrm{t}}) \\ &= \boldsymbol{H}_t^{1/2}\mathrm{Var}_{\mathrm{t}-1}(\boldsymbol{z}_{\mathrm{t}})(\boldsymbol{H}_{\mathrm{t}}^{1/2})' \\ &= \boldsymbol{H}_t. \end{aligned} \tag{2.4}$$

The literature on the different specifications of $\boldsymbol{H}_t$ may be broadly divided into direct multivariate extensions, factor models and the conditional correlation models. The usual trade-off of model parametrization and dimensionality clearly applies here, with the fully parameterized models offering the richest dynamics at the cost of increasing parameter size, making it unfeasible for modelling anything beyond a couple of assets. There is, also, a not so evident tradeoff between those models which allow flexible univariate dynamics to enter the equation at the cost of some multivariate dynamics, and this is discussed more fully in Section 2.4. The next sections will review these models and the tradeoffs they present for the decision maker. A more complete review of multivariate GARCH (*MGARCH*) models is provided in Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009b).

2.1.1 Direct Multivariate Extension Models

A direct extension of univariate GARCH dynamics to the multivariate domain was proposed by Bollerslev, Engle, and Wooldridge (1988), where each element of the conditional covariance matrix $\boldsymbol{H}_t$ is composed of linear combinations of the lagged errors and cross product errors and lagged values of $\boldsymbol{H}_t$.

Definition 3. The VEC(p,q) Model

$$vech\left(\boldsymbol{H}_t\right) = \boldsymbol{c} + \sum_{j=1}^{p}\boldsymbol{A}_j vech\left(\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}'_{t-j}\right) + \sum_{j=1}^{q}\boldsymbol{B}_j vech\left(\boldsymbol{H}_{t-j}\right), \tag{2.5}$$

where $\boldsymbol{c}$ is the $N(N+1)/2 \times 1$ intercept, *vech* is the operator that stacks the lower triangular portion of the $N \times N$ symmetric matrix as an $N(N+1)/2$ vector, and matrices $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ are square of order $N(N+1)/2$, giving a total of $\frac{1}{2}n^4 + n^3 + n^2 + \frac{1}{2}n$ variables! Because $\boldsymbol{H}_t$ is symmetric, $vech(\boldsymbol{H}_t)$ contains all the unique elements in $\boldsymbol{H}_t$. The richness of the model is immediately visible, as the variance of each individual asset is a function of its own squared errors and variances, all other squared errors and variances and all other cross lagged errors and covariances, and similarly modelled for the off diagonal elements (covariances). There is obviously a high cost to modelling the full interaction of lags and cross lags and hence the contagion-effect, where the (co)variance of an asset may be influenced by the lagged (co)variance of other assets. However, the requirement that $\boldsymbol{H}_t$ be positive definite for all values of ε_t in the sample space is difficult to impose during estimation. The Diagonal VEC (*DVEC*) model was suggested by the same authors to partly alleviate the dimensionality problem[4], by foregoing the effect of cross lags on individual variances and covariances, modelling A_t and B_t as diagonal matrices. Additionally, the diagonal representation, usually expressed in terms of Hadamard products, also benefits from simpler conditions for imposing positive definiteness of $\boldsymbol{H}_t$, derived in Attanasio (1991), which is a drawback of the full VEC model for which such conditions are hard to arrive at.

To overcome the difficulties of imposing positive definiteness in the VEC model and the high dimensionality, while not giving up as much as the DVEC, the BEKK model of Engle and Kroner (1995) was proposed on the premise that co-movements of financial assets are driven by a set of underlying factors. In terms of MGARCH categories, it lies somewhere between the direct extension VEC model, for which it is a special case[5], and a class of factors models most of which can be expressed as special cases of the BEKK model.

Definition 4. The BEKK(p,q,K) Model. The conditional covariance matrix $\boldsymbol{H}_t$ is governed by the following motion dynamics,

$$\boldsymbol{H}_t = \boldsymbol{C}'\boldsymbol{C} + \sum_{k=1}^{K}\sum_{j=1}^{q} \boldsymbol{A}'_{jk}\boldsymbol{\varepsilon}_{t-j}\boldsymbol{\varepsilon}'_{t-j}\boldsymbol{A}_{jk} + \sum_{k=1}^{K}\sum_{j=1}^{p} \boldsymbol{B}'_{jk}\boldsymbol{H}_{t-j}\boldsymbol{B}_{jk}, \tag{2.6}$$

where $\boldsymbol{C}$, $\boldsymbol{A}_{jk}$ and $\boldsymbol{B}_{jk}$ are $N \times N$ matrices, with $\boldsymbol{C}$ being upper triangular. A direct advantage of the BEKK model over the VEC model of Bollerslev, Engle, and Wooldridge (1988), is that positivity of the covariance matrix $\boldsymbol{H}_t$ is easy to impose. The number of parameters is significantly less in the full BEKK model, being $\frac{5}{2}n^2 + \frac{1}{2}n$, and only about $\frac{5}{3}$ times bigger than the DVEC model. Unlike the DVEC model, the BEKK specification does model the dependence of conditional variances (covariances) subject to the lagged values of all other variances (covariances), hence capturing the spillover effect. The quadratic form of the model, means that certain sign restrictions are necessary to ensure identifiability, which for simple models such as when $K = 1$ and $p = q = 1$ is

[4]The DVEC requires $\frac{3}{2}\left(n^2 + n\right)$ parameters.

[5]In fact, for each BEKK model there is an equivalent VEC representation.

a simple matter of ensuring the positivity of the upper diagonal elements of $\boldsymbol{A}_{11}$ and $\boldsymbol{B}_{11}$. Consistency of the BEKK model was proved by Jeantheau (1998) under the log-moment condition, which requires the existence of sixth-order moments, making this untestable. Asymptotic normality of the Quasi Maximum Likelihood (*QML*) Estimates of the BEKK model was established by Comte and Lieberman (2003), this time under the existence of eighth order moments which again are untestable, while Hafner and Preminger (2009) established the asymptotic normality of the VEC model (in which the BEKK is nested) under the existence of sixth order moments. The dynamics of both the VEC and BEKK models can be reduced to achieve dimensionality reduction gains, leading to several variants such as diagonal and scalar models, as well as the use of covariance targeting to reduce the number of parameters in the estimation of the intercept. In the latter case, this is achieved by setting:

$$\boldsymbol{C}'\boldsymbol{C} = \bar{\boldsymbol{\Sigma}} - \boldsymbol{A}'\bar{\boldsymbol{\Sigma}}\boldsymbol{A} - \boldsymbol{B}'\bar{\boldsymbol{\Sigma}}\boldsymbol{B} \tag{2.7}$$

where $\bar{\boldsymbol{\Sigma}}$ is the unconditional covariance matrix of $\boldsymbol{\varepsilon}$ which may be consistently estimated by $N^{-1}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$. In order for $\boldsymbol{H}_t$ to be positive definite in the presence of covariance targeting, the eigenvalues of the intercept must be positive and checked during estimation. This is a highly nonlinear constraint for which I find that standard solvers will fail to converge to a global optimum most of the time, thus necessitating a global optimization approach. Rotating between local and global based solvers (such as simulated annealing or pattern search) will provide for a much more robust solution than either individually[6]. Finally, covariance stationarity in the diagonal BEKK models is simply a vectorized form of the scalar case so that the element-wise sum of the squared diagonal parameters is less than unity:

$$\sum_{i=1}^{p} a_{nn,i}^2 + \sum_{j=1}^{q} b_{nn,j}^2 < 1. \tag{2.8}$$

It would appear that covariance targeting for large dimensional systems eliminates $N(N+1)/2$ parameters from the estimation thus making it more feasible. However,

[6]The strategy is quite simple: an SQP based solver which is guaranteed to find a local minimum initiates the first run, and then a simulated annealing solver runs for N iterations to find a new direction, after which the more efficient SQP solver takes over to find that direction's local minimum. The strategy iterates between the two solvers until the global based solver cannot find a better solution after N iterations. As an added safety step, an alternative global based solver such as that of pattern search, is then run N iterations at which point either the solution does not change else if a better point is found the strategy restarts.

this is only partly true. In the absence of covariance targeting, we can guarantee positive definiteness of the intercept, by construction, through $C'C$. With covariance targeting, the added constraint of positive definiteness provides for 2 possible avenues. The first one, imposes a proper constraint by adding $N(N+1)/2$ parameters to the estimation so that the intercept, $\boldsymbol{\Omega}$, calculated through targeting, is constrained to be positive definite. Formally:

Property 1. A Matrix M is positive definite if and only if there is a positive definite matrix $\boldsymbol{B} > 0$ with $\boldsymbol{B}^2 = \boldsymbol{M}$.[7]

The matrix $\boldsymbol{B}$ is called the 'square root' of $\boldsymbol{M}$. This matrix $\boldsymbol{B}$ is unique, but only under the assumption $\boldsymbol{B} > 0$. In terms of the optimization problem, we can include the following constraint to ensure the positive definiteness of the intercept $\boldsymbol{\Omega}$: $\boldsymbol{B}^2 - \boldsymbol{\Omega} = 0$. If $\boldsymbol{\Omega}$ has a 'square root' then it is positive semi-definite. One therefore models the lower triangular part of $\boldsymbol{B}$ which creates an added $N(N+1)/2$ parameters in the optimization problem. Thus in this case, the full constraint reintroduces back into the model the same number of parameters eliminated because of covariance targeting in the first place, which is possibly the reason it has not been considered in the literature thus far. The second approach, which does not introduce this constraint involves the use of a global optimization approach since checking for positive and real eigenvalues as an 'arbitrary' constraint, introduces non-smoothness and discontinuity in the likelihood, and is likely to lead to many local minima.

An alternative avenue for dimensionality reduction has been explored through the factor and conditional correlation models which exploit certain distributional properties to simplify or avoid the full multivariate density evaluation. The conditional correlation model is discussed in the next section, while a statistical factor MGARCH framework is discussed in Chapter 3.

2.1.2 Conditional Correlation Models

Conditional correlation models are founded on a decomposition of the conditional covariance matrix into conditional standard deviations and correlations, so that they may be expressed in such a way that the univariate and multivariate dynamics may be separated, thus easing the estimation process. This decomposition comes at a cost of some

[7] Notation greater than in this context means positive definite

dynamic structure as well as severe restriction on the type of multivariate distribution which can usually be decomposed in such a way. As a result, the models have been extended to allow for more flexible dynamic structure which unfortunately has led to significant loss in the ease of estimation.

In the constant conditional correlation model (CCC) of Bollerslev (1990), the covariance matrix can be decomposed into

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t = \rho_{ij}\sqrt{h_{iit}h_{jjt}}, \tag{2.9}$$

where $\boldsymbol{D}_t = diag(\sqrt{h_{11,t}}, ..., \sqrt{h_{nn,t}})$, and $\boldsymbol{R}$ is the positive definite constant conditional correlation matrix. The conditional variances, and $h_{ii,t}$, which can be estimated separately, can be written in vector form based on GARCH(p,q) models[8]

$$h_t = \omega + \sum_{i=1}^{p} \boldsymbol{A}_i \varepsilon_{t-i} \odot \varepsilon_{t-i} + \sum_{i=1}^{q} \boldsymbol{B}_i h_{t-i} \tag{2.10}$$

where $\omega \in \mathbb{R}^n$, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are $N \times N$ diagonal matrices, and $\odot$ denotes the Hadamard operator. The conditions for the positivity of the covariance matrix $\boldsymbol{H}_t$ are that $\boldsymbol{R}$ is positive definite, and the elements of ω and the diagonal elements of the matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are positive. In the extended CCC model (E-CCC) of Jeantheau (1998), the assumption of diagonal elements on $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ was relaxed, allowing the past squared errors and variances of the series to affect the dynamics of the individual conditional variances, and hence providing for a much richer structure, albeit at the cost of a lot more parameters. The decomposition in (2.9), allows the log-likelihood at each point in time (LL_t), in the multivariate normal case, to be expressed as

$$\begin{aligned} LL_t &= \frac{1}{2}\left(\log(2\pi) + \log|\boldsymbol{H}_t| + \boldsymbol{\varepsilon'}_t \boldsymbol{H}_t^{-1} \boldsymbol{\varepsilon}_t\right) \\ &= \frac{1}{2}\left(\log(2\pi) + \log|\boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t| + \boldsymbol{\varepsilon'}_t \boldsymbol{D}_t^{-1} \boldsymbol{R}^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{\varepsilon}_t\right) \\ &= \frac{1}{2}\left(\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log|\boldsymbol{R}| + \boldsymbol{z'}_t \boldsymbol{R}^{-1} \boldsymbol{z'}_t\right) \end{aligned} \tag{2.11}$$

where $\boldsymbol{z}_t = \boldsymbol{D}_t^{-1}\boldsymbol{\varepsilon}_t$. This can be described as a term ($\boldsymbol{D}_t$) for the sum of the univariate GARCH model likelihoods, a term for the correlation ($\boldsymbol{R}$) and a term for the covariance which arises from the decomposition.

Because the restriction of constant conditional correlation may be unrealistic in practice,

[8] The GARCH models are not restricted to be of one particular 'flavor', allowing to mix for example APARCH with EGARCH models in the univariate stage.

a class of models termed Dynamic Conditional Correlation (DCC), due to Engle (2002) and Tse and Tsui (2002), allow for the correlation matrix to be time varying with motion dynamics, such that

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t. \tag{2.12}$$

In these models, apart from the fact that the time varying correlation matrix, $\boldsymbol{R}_t$, must be inverted at every point in time (making the calculation that much slower), it is also important to constrain it to be positive definite. The most popular of these DCC models, due to Engle (2002), achieves this constraint by modelling a proxy process, $\boldsymbol{Q}_t$ as:

$$\begin{aligned} \boldsymbol{Q}_t &= \bar{\boldsymbol{Q}} + a\left(\boldsymbol{z}_{t-1}\boldsymbol{z'}_{t-1} - \bar{\boldsymbol{Q}}\right) + b\left(\boldsymbol{Q}_{t-1} - \bar{\boldsymbol{Q}}\right) \\ &= (1-a-b)\bar{\boldsymbol{Q}} + a\boldsymbol{z}_{t-1}\boldsymbol{z'}_{t-1} + b\boldsymbol{Q}_{t-1} \end{aligned} \tag{2.13}$$

where a and b are non negative scalars, with the condition that $a+b<1$ imposed to ensure stationarity and positive definiteness of $\boldsymbol{Q}_t$. $\bar{Q}$ is the unconditional matrix of the standardized errors $\boldsymbol{z}_t$ which enters the equation via the covariance targeting part $(1-a-b)\bar{\boldsymbol{Q}}$, and $\boldsymbol{Q}_0$ is positive definite. The correlation matrix $\boldsymbol{R}$ is then obtained by rescaling $\boldsymbol{Q}_t$ such that,

$$\boldsymbol{R}_t = diag(\boldsymbol{Q}_t)^{-1/2}\boldsymbol{Q}_t diag(\boldsymbol{Q}_t)^{-1/2}. \tag{2.14}$$

The log-likelihood function in Equation (2.10) can be decomposed more clearly into a volatility and correlation component by adding and subtracting $\boldsymbol{\varepsilon'}_t\boldsymbol{D}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{z'}_t\boldsymbol{z}_t$,

$$\begin{aligned} LL &= \frac{1}{2}\sum_{i=1}^{T}\left(N\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log|\boldsymbol{R}_t| + \boldsymbol{z'}_t\boldsymbol{R}_t^{-1}\boldsymbol{z'}_t\right) \\ &= \frac{1}{2}\sum_{i=1}^{T}\left(N\log(2\pi) + 2\log|\boldsymbol{D}_t| + \boldsymbol{\varepsilon'}_t\boldsymbol{D}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{\varepsilon}_t\right) - \frac{1}{2}\sum_{i=1}^{T}\left(\boldsymbol{z'}_t\boldsymbol{z}_t + \log|\boldsymbol{R}_t| + \boldsymbol{z'}_t\boldsymbol{R}_t^{-1}\boldsymbol{z'}_t\right) \\ &= LL_V(\theta_1) + LL_R(\theta_1,\theta_2) \end{aligned} \tag{2.15}$$

where $LL_V(\theta_1)$ is the volatility component with parameters θ_1, and $LL_R(\theta_1,\theta_2)$ the correlation component with parameters θ_1 and θ_2. In the Multivariate Normal case, where no shape or skew parameters enter the density, the volatility component is the sum of the individual GARCH likelihoods which can be jointly maximized by separately maximizing each univariate model. In other distributions, such as the multivariate Student, the existence of a shape parameter means that the estimation must be performed in one step so that the shape parameter is jointly estimated for all models[9]. Separation of the likelihood into 2 parts provides for feasible large scale estimation. Together with the use of covariance targeting, very large scale systems may be estimated in a matter of seconds with the use of parallel and grid computing. Yet as the system becomes larger and larger, it becomes questionable whether the scalar parameters can adequately capture the dynamics of the underlying process. As such, Cappiello, Engle, and Sheppard (2006) generalize the DCC model with the introduction of the Asymmetric Generalized DCC ($AGDCC$) where the dynamics of $\boldsymbol{Q}_t$ are:

$$\boldsymbol{Q}_t = \left(\bar{\boldsymbol{Q}} - \boldsymbol{A}'\bar{Q}\boldsymbol{A} - \boldsymbol{B}'\bar{\boldsymbol{Q}}\boldsymbol{B} - \boldsymbol{G}'\bar{\boldsymbol{Q}}^{-}\boldsymbol{G}\right) + \boldsymbol{A}'\boldsymbol{z}_{t-1}\boldsymbol{z}'_{t-1}\boldsymbol{A} + \boldsymbol{B}'\boldsymbol{Q}_{t-1}\boldsymbol{B} + \boldsymbol{G}'\boldsymbol{z}_t^{-}\boldsymbol{z}'^{-}_t\boldsymbol{G} \tag{2.16}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{G}$ are the $N \times N$ parameter matrices, $\boldsymbol{z}_t^{-}$ are the zero-threshold standardized errors which are equal to $\boldsymbol{z}_t$ when less than zero and zero otherwise, $\bar{\boldsymbol{Q}}$ and $\bar{\boldsymbol{Q}}^{-}$ the unconditional matrices of $\boldsymbol{z}_t$ and $\boldsymbol{z}_t^{-}$ respectively. Because of its high dimensionality, restricted models have been used including the scalar, diagonal and symmetric versions with the specifications nested being

- DCC : $\boldsymbol{G} = [0]\,, \boldsymbol{A} = \sqrt{a}, \boldsymbol{B} = \sqrt{b}$
- ADCC : $\boldsymbol{G} = \sqrt{g}, \boldsymbol{A} = \sqrt{a}, \boldsymbol{B} = \sqrt{b}$
- GDCC : $\boldsymbol{G} = [0]$.

[9] In Section 2.2 I discuss the case of the DCC-Student copula where this may be relaxed

Table 2.1: AGDCC comparative estimates (global equity and bond indices)

AGDCC (MVN) Model	Original Estimates			Revised Estimates		
	a_i^2	g_i^2	b_i^2	a_i^2	g_i^2	b_i^2
Australia stocks	0.006	0.008	0.790	0.002*	0.005*	0.928
Austria stocks	0.003	0.004	0.961	0.007	0.016*	0.890
Belgium stocks	0.010	0.008	0.950	0.010*	0.029*	0.910
Canada stocks	0.005	0.002	0.952	0.002*	0.027*	0.876*
Denmark stocks	0.003	0.005	0.965	0.005*	0.014*	0.933
France stocks	0.009	0.003	0.945	0.017*	0.042	0.853
Germany stocks	0.007	0.007	0.957	0.005*	0.031*	0.909
Hong Kong stocks	0.000*	0.002	0.956	0.000*	0.010*	0.937
Ireland stocks	0.000*	0.007	0.968	0.011	0.002*	0.786
Italy stocks	0.007	0.012	0.957	0.009*	0.025*	0.905
Japan stocks	0.002	0.003	0.953	0.004*	0.003*	0.947*
Mexico stocks	0.001*	0.019	0.938	0.000*	0.011*	0.900
Netherlands stocks	0.006	0.009	0.959	0.008*	0.016*	0.919*
New Zealand stocks	0.001*	0.001*	0.922	0.001*	0.004*	0.956*
Norway stocks	0.002	0.006	0.929	0.001*	0.002*	0.889
Singapore stocks	0.001*	0.002	0.976	0.000*	0.010*	0.925*
Spain stocks	0.006	0.007	0.954	0.016	0.028*	0.857
Sweden stocks	0.005	0.006	0.965	0.003*	0.026*	0.869*
Switzerland stocks	0.015	0.009	0.943	0.011*	0.028	0.900
U.K. stocks	0.007	0.006	0.955	0.006*	0.033*	0.875*
U.S. stocks	0.002	0.004	0.951	0.023	0.009*	0.801
Austria bonds	0.010	0.009	0.976	0.035*	0.003*	0.938
Belgium bonds	0.011	0.009	0.975	0.038*	0.004*	0.934
Canada bonds	0.005	0.006	0.859	0.006*	0.017*	0.859*
Denmark bonds	0.011	0.009	0.973	0.039*	0.008*	0.926
France bonds	0.011	0.008	0.973	0.035	0.007*	0.932
Germany bonds	0.013	0.009	0.972	0.036*	0.005*	0.934
Ireland bonds	0.014	0.007	0.968	0.042	0.006*	0.923
Japan bonds	0.005	0.006	0.964	0.020*	0.005*	0.932*
Netherlands bonds	0.013	0.008	0.972	0.035*	0.004*	0.936
Sweden bonds	0.008	0.012	0.962	0.021*	0.019*	0.939
Switzerland bonds	0.012	0.007	0.974	0.035*	0.002*	0.936
U.K. bonds	0.006	0.004	0.972	0.032	0.000*	0.923
U.S. bonds	0.006	0.003	0.936	0.008*	0.020*	0.898
LL		**78,022**			**78,120**	

Notes to table 2.1: The Table presents the original parameter estimates and Log-Likelihood (LL) of the 34 global equity and bond indices from the ADGCC (MVN) model in the paper by Cappiello, Engle, and Sheppard (2006) (Table 6a) and the revised estimates after re-running the estimation using a global optimization strategy. * denotes insignificance at the 5% level.

As discussed in the previous section, covariance targeting in such high dimensional models where the parameters are no longer scalars, creates difficulties in imposing positive definiteness during estimation while at the same time guaranteeing a global optimum solution. In fact, investigating the model of Cappiello, Engle, and Sheppard (2006), using the same dataset[10], I find rather different results. Starting with their parameters for the diagonal AGDCC model, displayed in Table 6a of their paper, I find a log-likelihood of $78,022$[11], and continue the optimization using a global optimization approach. This leads to a slightly higher likelihood of $78,120$, and more importantly I find no significant asymmetry in the correlation dynamics (except for French and Swiss Equity), unlike the authors who show in Table 6a of their paper that with the exception of New Zealand Equity, all asymmetry parameters of the AGDCC model are significant at the 5% level. Table 2.1 provides a side by side comparison of the original parameters from the paper by Cappiello, Engle, and Sheppard (2006) and the revised estimates. Besides the differences in the significance of the asymmetry parameter (parameters in the original paper were reported as the squares of their values and I follow the same format for this table), there are also significant differences in the shock and persistence parameters a_i^2 and b_i^2 respectively, with the revised estimates exhibiting much less significance at the 5% level. This certainly highlights the practical problems in estimating this model. More substantially, Aielli (2009) points out that the estimation of $\bar{\boldsymbol{Q}}_t$ as the empirical counterpart of the correlation matrix of $\boldsymbol{z}_t$ in the DCC model is inconsistent since $\mathrm{E}[\boldsymbol{z}_\mathrm{t}\boldsymbol{z}_\mathrm{t}] = \mathrm{E}[\boldsymbol{R}_\mathrm{t}] \neq \boldsymbol{E}[\bar{\boldsymbol{Q}}_\mathrm{t}]$. He proposes instead the *cDCC* model which includes a corrective step which eliminates this inconsistency, albeit at the cost of targeting which is not allowed. Whether the identified inconsistency is significant enough to merit widespread adoption is still an open question, since the elimination of the 2 step approach also eliminates most of the advantages of using a DCC type model over the BEKK, a point forcefully taken up by Caporin and McAleer (2012) who question the merits of the DCC model over the BEKK model with covariance targeting which has more consistent properties. I investigate this point further in Section 2.4 and in fact find that there is value in the 2-stage approach.

[10]The dataset was kindly provided by Kevin Sheppard, and consists of FTSE All-World Indices for 21 countries and DataStream-constructed five-year average maturity government bond indices for 13 countries. The sample covers the period from January 8, 1987 to February 7, 2002 (785 observations), of weekly (Thursday-to-Thursday close) continuously compounded returns. I would also note here that this dataset, as provided, appears to be missing the whole month of February 1995, which equates to 4 missing data points (since it is weekly).

[11]Which is also different from the one the authors report in Table 6b of their paper (−25485.1), probably since they exclude some constant or just report the second stage likelihood rather than the overall one.

Other notable DCC extensions have included the Smooth and double Smooth Transition Conditional Correlation models of Silvennoinen and Teräsvirta (2009a) and the Regime Switching Dynamic Correlation of Pelletier (2006). An interesting compromise in the modelling of the dynamics in the AGDCC context was proposed by Billio, Caporin, and Gobbo (2006) in terms of a block-diagonal structure so that the dynamics among groups of highly correlated securities is the same. In a different direction, Pelagatti and Rondena (2006) provide a rather obvious argument that any elliptical distribution could be decomposed and used in DCC modelling, including the Student[12] and Laplace, the latter discussed in more detail in Section 2.3.1. Finally, Kroner and Ng (1998) introduce an omnibus model, the Generalized Dynamic Covariance (*GDC*) Model which nests the VEC, BEKK, FGARCH, CCC and DCC models as special cases and written as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t + \boldsymbol{\Phi} \odot \Theta_t, \tag{2.17}$$

where $\boldsymbol{D}_t = d_{ij,t}$, $d_{ii,t} = \sqrt{\theta_{ii,t}} \forall i$ and $d_{ij,t} = 0 \forall i \neq j$, $\odot$ is the Hadamard operator, $\boldsymbol{R}_t = \rho_{ij,t}$, and $\Theta = \theta_{ij,t}$ following BEKK dynamics as in Equation (2.6). Depending on the parameter restrictions, various models arise such as the BEKK model when $\boldsymbol{R}$ is diagonal and $\boldsymbol{\Phi}$ with off-diagonal values of 1. Other restrictions, leading to other models, are given in proposition 1 of Kroner and Ng (1998). The authors also describe in the same paper an asymmetric version of this model by adjusting the BEKK dynamics in $\theta_{ij,t}$ to incorporate an asymmetry term for the zero-threshold shocks, which is a natural generalization from such univariate models as the GJR-GARCH and T-GARCH of Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994) respectively. Like in the case of the family GARCH model of Hentschel (1995) where comparison of nested models was made via the news impact curve of Engle and Ng (1993), the authors generalize the curve to a surface function providing for some revealing visual insights into the different multivariate dynamics.

The main advantage of DCC models is the 2-step estimation process which effectively allows any type of univariate model for the first stage dynamics. However, there is also a cost to this method when it comes to calculating the standard errors, since any potential speed gains in the 2 step method are almost eliminated when calculating the Hessian which is partitioned[13]. For the scalar DCC model, this is not an issue, but quickly

[12] Technically, the Student cannot be used in a 2-stage estimation because of the presence of a shape parameter in the density, as discussed previously, though in practice a first stage QML estimation has been employed.

[13] Because of the two-stage estimation process, the standard error matrix in these DCC models is a partitioned matrix with the first stage univariate GARCH parameters partitioned off from the second

becomes computationally expensive in the diagonal and full Generalized DCC model as the number of assets increases. In addition, for the Generalized DCC model, aside from the problem of imposing stationarity and positive definiteness in the model, which involves some very nonlinear constraints with the possibility of a non global solution, the Hessian might be ill conditioned for parameters at the intercept's limit of positive definiteness or when the solution is only a local one. In such situations, one must ask whether a BEKK model with covariance targeting is not better suited to the task since there is no 2-stage estimation, and hence the full Hessian is more likely to depend on the immediate optimization result rather than an afterthought in the calculation of the partitioned Hessian. Nevertheless, the DCC model's 2 stage estimation has proved quite popular in practice, and the next section reviews an even more flexible representation of the DCC model when the distribution is fitted using a copula.

2.2 MGARCH with Flexible Margins: The DCC Copula model

Copula functions were introduced by Sklar (1959) as a tool to connect disparate marginal distribution together to form a joint multivariate distribution. They were extensively used in survival analysis and the actuarial sciences for many years before being introduced in the finance literature more than a decade ago by Frey and McNeil (2000) and Li (2000). They have since been very popular in investigating the dependence of financial time series of various assets classes and frequencies. Breymann, Dias, and Embrechts (2003) investigate bivariate hourly forex spot returns finding that the Student copula best fit the data at all horizons (with the shape parameter increasing with the time horizon), while Malevergne and Sornette (2003) find that the Normal copula fits pairs of currencies and equities well on the whole but unsurprisingly fails to capture tail events where the Student copula does best.[14] Junker and May (2005) use a Frank copula with a transformation generator and GARCH dynamics for the margins using the empirical distribution, to analyze the bivariate dependency of the daily returns of 6 stocks and 3 Euro swap rates (with horizons 2,5, and 10 Years). The comparison with a range of popular copulas including the Normal and Student, in a risk exercise

stage DCC parameters. While it is also possible to estimate the system in one stage, that would defeat the main purpose of the model which is ease and speed of estimation.

[14]Interestingly the authors argue that since such events are rare, the goodness of fit test they use cannot always reject the Normal copula.

shows that asymmetric tail dependency is important and usually not accommodated by the Student distribution[15] While most studies are predominantly focused on bivariate copulas, the extension to n-variate models is not overtly challenging particularly for elliptical distributions, or the use of the more recent Vine pair copulas (see for example Joe, Li, and Nikoloulopoulos (2010)).

2.2.1 Copulas

An n-dimensional copula $C(u_1, \ldots, u_n)$ is a distribution in the unit hypercube $[0,1]^n$ with uniform margins. Sklar (1959) showed that every joint distribution F of the random vector $\boldsymbol{X} = (x_1, \ldots, x_n)$ with margins $F_1(x_1), \ldots, F_n(x_n)$, can be represented as:

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{2.18}$$

for a copula C, which is uniquely determined in $[0,1]^n$ for distributions F under absolutely continuous margins and obtained as:

$$C(u_1, \ldots, u_n) = F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right) \tag{2.19}$$

The density function may conversely be obtained as :

$$f(x_1, \ldots, x_n) = c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} f_i(x_i) \tag{2.20}$$

where f_i are the marginal densities and c is the density function of the copula given by:

$$c(u_1, \ldots, u_n) = \frac{f\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right)}{\prod_{i=1}^{n} f_i\left(F_i^{-1}(u_i)\right)}. \tag{2.21}$$

with F_i^{-1} being the quantile function of the margins. A key property of copulas is their invariance under strictly increasing transformation of the components of the $\boldsymbol{X}$, so that for example the copula of the multivariate Normal distribution $F_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the same as that of $F_n(0, \boldsymbol{R})$ where $\boldsymbol{R}$ is the correlation matrix implied by the covariance matrix, and the same for the copula of the multivariate Student distribution reviewed in detail

[15] An alternative would be to use the skew Generalized Hyperbolic Student distribution analyzed in Aas and Haff (2006) which allows for the modelling of one heavy (with polynomial behavior) and one semi-heavy (with exponential behavior) tail.

in Demarta and McNeil (2005). The density of the Normal copula, of the n-dimensional random vector $\boldsymbol{X}$ in terms of the correlation matrix $\boldsymbol{R}$, is then:

$$c\left(u; \boldsymbol{R}\right) = \frac{1}{\left|\boldsymbol{R}\right|^{1/2}} e^{-\frac{1}{2}u'\left(\boldsymbol{R}' - \boldsymbol{I}\right)u} \tag{2.22}$$

where $\boldsymbol{u}_i = \boldsymbol{\Phi}^{-1}\left(F_i\left(\boldsymbol{x}_i\right)\right)$ for $i = 1, \ldots, n$, representing the quantile of the Probability Integral Transformed (PIT) values of $\boldsymbol{X}$, and $\boldsymbol{I}$ is the identity matrix. Because the Normal copula cannot account for tail dependence, the Student copula has been more widely used for modelling of financial assets. The density of the Student copula, of the n-dimensional random vector $\boldsymbol{X}$ in terms of the correlation matrix $\boldsymbol{R}$ and shape parameter ν, can be written as:

$$c\left(\boldsymbol{u}; \boldsymbol{R}, \nu\right) = \frac{\boldsymbol{\Gamma}\left(\frac{\nu+n}{2}\right)\left(\boldsymbol{\Gamma}\left(\frac{\nu}{2}\right)\right)^n \left(1 + \nu^{-1}\boldsymbol{u}'\boldsymbol{R}^{-1}\boldsymbol{u}\right)^{-(\nu+n)/2}}{\left|\boldsymbol{R}\right|^{1/2}\left(\boldsymbol{\Gamma}\left(\frac{\nu+n}{2}\right)\right)^n \boldsymbol{\Gamma}\left(\frac{\nu}{2}\right) \prod_{i=1}^{n}\left(1 + \frac{\boldsymbol{u}_i^2}{\nu}\right)^{-(\nu+1)/2}} \tag{2.23}$$

where $\boldsymbol{u}_i = {t_\nu}^{-1}\left(F\left(x_i; \nu\right)\right)$, where t_ν^{-1} is the quantile function of the student distribution with shape parameter ν.

2.2.2 Correlation and Kendall's $\boldsymbol{\tau}$

Pearson's product moment correlation $\boldsymbol{R}$ totally characterizes the dependence structure in the multivariate Normal case, where zero correlation also implies independence, but can only characterize the ellipses of equal density when the distribution belongs to the elliptical class. In the latter case for instance, with a distribution such as the multivariate Student, the correlation cannot capture tail dependence determined by the shape parameter. Furthermore, it is not invariant under monotone transformations of original variables making it inadequate in many cases. An alternative measure which does not suffer from this is Kendall's $\boldsymbol{\tau}$ (see Kruskal (1958)) based on rank correlations which makes no assumption about the marginal distributions but depends only on the copula C. It is a pairwise measure of concordance calculated as:

$$\boldsymbol{\tau}\left(x_i, x_j\right) = 4\int_0^1 \int_0^1 C\left(u_i, u_j\right) dC\left(u_i, u_j\right) - 1. \tag{2.24}$$

For elliptical distributions, Lindskog, Mcneil, and Schmock (2003) proved that there is a one-to-one relationship between this measure and Pearson's correlation coefficient

ρ given by:

$$\boldsymbol{\tau}\left(x_i, x_j\right) = \left(1 - \sum_{x\in\mathbb{R}} \left(\mathbb{P}\{X_i = x\}^2\right)\right) \frac{2}{\pi} \arcsin \rho_{ij} \tag{2.25}$$

which under certain assumptions (such as in the case of the multivariate Normal) simplifies to $\frac{2}{\pi}\arcsin\rho_{ij}$.[16] Kendall's $\boldsymbol{\tau}$ is also invariant under monotone transformations making it rather more suitable when working with non-elliptical distributions. A useful application arises in the case of the multivariate Student distribution, where a maximum likelihood approach for the estimation of the correlation matrix $\boldsymbol{R}$ becomes unfeasible for large dimensions. In this case, an alternative approach is to estimate the sample counterpart of Kendall's $\boldsymbol{\tau}$[17] from the transformed margins and then translate that into the correlation matrix as detailed in (2.25), providing for a method of moments type estimator.[18] The shape parameter ν may then be estimated keeping the correlation matrix constant, with little loss in efficiency vis-a-vis the full maximum likelihood method.[19]

2.2.3 Transformations and Consistency

The estimation and PIT transformation of the margins provides for a great deal of flexibility, with the possibility of adopting a parametric, semi-parametric or empirical approach. The first method, whereby the margins and transformation are performed using a parametric density, was termed the Inference-Functions-for-Margins (*IFM*) by Joe (1997) who also established the asymptotic theory for it. The semi-parametric method (*SPD*) uses a distribution which couples together tails fitted by the generalized Pareto distribution (*GPD*)[20] with a kernel based interior and described in Davison and Smith (1990), and offers a rather flexible method for capturing fat tails observed in practice. Finally, the empirical approach, also called pseudo-likelihood, was investigated by Genest, Ghoudi, and Rivest (1995) and asymptotic properties established under the assumption that the sequence of $\boldsymbol{X}$ is i.i.d.(see Durrleman, Nikeghbali, and Roncalli (2000) for an excellent summary of the different methods and their properties.)

[16]Another popular measure is Spearman's correlation coefficient ρ_s which under Normality equates to $\frac{6}{\pi}\arcsin\frac{\rho_{ij}}{2}$, and it is usually very close in result to Kendall's measure.

[17]The matrix is build up from the pairwise estimates.

[18]It may be the case that the resultant matrix is not positive definite, in which case a variety of methods exist to tweak it into one.

[19] According to at least one study of Zeevi and Mashal (2002).

[20]For which a Probability Weighted Moment (*PWM*) approach exists which is quite robust.

2.2.4 The DCC Student Copula

The extension of the static copula approach to dynamic models, and in particular GARCH, was investigated by Patton (2006) who extended and proved the validity of Sklarś theorem for the conditional case. Jondeau and Rockinger (2006a) combined the ACD model of Hansen (1994) with a skewed Student distribution to model time-varying or regime switching Student copula for the dependence between pairs of countries, while Chollete, Heinen, and Valdesogo (2009) used a GARCH with skewed Student distribution in the first stage and a regime switching model with a Canonical vine copula for the high dependence regime and a Normal copula for the low dependence regime. The use of the skewed Student distribution in such models, beyond its tractability and desirable features, according to Chollete, Heinen, and Valdesogo (2009) is so as to ensure that the asymmetry in the dependence structure is purely the result of multivariate asymmetry and not an artifact of poor modelling of the margins. Demarta and McNeil (2005) described a skewed Student copula derived from the Normal Mean Variance Mixture distribution (reviewed in the next section), with common shape (ν) univariate skewed student margins and separate skewness (γ) parameters.[21]
In an elliptical distribution setting, adding dynamics to the correlation matrix of the copula seems a natural extension of the 2-stage DCC model, and allows the estimation of a Student copula with disparate shape parameters for the first stage, where this was not possible using the standard DCC model (unless estimated jointly). Let the n-dimensional random vector of asset returns $\boldsymbol{r}_t = r_{it}, \ldots, r_{nt}$ follow a copula GARCH model with joint distribution given by:

$$F\left(\boldsymbol{r}_t|\boldsymbol{\mu}_t, \boldsymbol{h}_t\right) = C\left(F_1\left(r_{1t}|\mu_{1t}, h_{1t}\right), \ldots, F_n\left(r_{nt}|\mu_{nt}, h_{nt}\right)\right) \tag{2.26}$$

where F_i, $i = 1, \ldots, n$ is the conditional distribution of the i^{th} marginal series density, C is the n-dimensional copula. The conditional mean $\mathrm{E}\left[\mathrm{r_{it}}\left|\Im_{t-1}\right.\right] = \mu_{\mathrm{it}}$, where $\Im_{t-1}$ is the σ-field generated by the past realization of r_t, and the conditional variance h_{it}

[21]The reason for the common shape parameter is that the mixing variable $\boldsymbol{W}$ in the Normal Mean Variance mixture is tge Inverse Gaussian distribution, $W \sim Ig\left(\nu/2, \nu/2\right)$. A grouped type copula whereby the shape parameter is also allowed to vary is also described by Demarta and McNeil (2005), in which case each variable has a different value for the mixing variable $\boldsymbol{W}$, so that $\boldsymbol{W}_j \sim Ig\left(\nu_j/2, \nu_j/2\right)$, for $j = 1, \ldots, n$, and the $\boldsymbol{W}_j$ are now perfectly correlated.

follows a GARCH(1,1) process[22]:

$$r_{it} = \mu_{it} + \varepsilon_{it}, \varepsilon_{it} = \sqrt{h_{it}} z_{it}, \tag{2.27}$$

$$h_{it} = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{it-1} \tag{2.28}$$

where z_{it} are i.i.d. random variables which conditionally follow some distribution with the requisite properties. For the purpose of this exposition, and because it is used in an empirical application extensively in Chapter 4, I consider the standardized skew Student distribution, so that $z_{it} \sim f_i(0, 1, \xi_i, \nu_i)$, of Fernandez and Steel (1998) with skew and shape parameters ξ and ν respectively and derived in Appendix B.[23] The conditional GARCH dynamics are such as to ensure positivity and stationarity, namely $(\omega, \alpha_1, \beta_1) > 0$ and $(\alpha_1 + \beta_1) < 1$. The dependence structure of the margins is then assumed to follow a Student copula with conditional correlation $\boldsymbol{R}_t$ and constant shape parameter η. The conditional density at time t is given by:

$$c_t\left(u_{it}, \ldots, u_{nt} \left| \boldsymbol{R}_t, \eta \right.\right) = \frac{f_t\left(F_i^{-1}\left(u_{it} \left| \eta \right.\right), \ldots, F_i^{-1}\left(u_{nt} \left| \eta \right.\right) \left| \boldsymbol{R}_t, \eta \right.\right)}{\prod\limits_{i=1}^{n} f_i\left(F_i^{-1}\left(u_{it} \left| \eta \right.\right) \left| \eta \right.\right)} \tag{2.29}$$

where $u_{it} = F_{it}\left(r_i t \left| \mu_{it}, h_{it}, \xi_i, \nu_i \right.\right)$ is the PIT transformation of each series by its conditional distribution F_{it} estimated via the first stage GARCH process, $F_i^{-1}\left(u_{it} \left| \eta \right.\right)$ represents the quantile transformation of the uniform margins subject to the common shape parameter of the multivariate density, $f_t\left(. \left| \boldsymbol{R}_t, \eta \right.\right)$ is the multivariate density of the Student distribution with conditional correlation R_t and shape parameter η and $f_i\left(. \left| \eta \right.\right)$ is the univariate margins of the multivariate Student distribution with common shape parameter η. The dynamics of R_t are assumed to follow an AGDCC model as in Section 2.1.2, though it is more common to use a restricted scalar DCC model for not too large a number of series. Finally, the joint density of the 2-stage estimation is written as:

$$f\left(\boldsymbol{r}_t \left| \boldsymbol{\mu}_t, \boldsymbol{h}_t, \boldsymbol{R}_t, \eta \right.\right) = c_t\left(u_{it}, \ldots, u_{nt} \left| \boldsymbol{R}_t, \eta \right.\right) \prod_{i=1}^{n} \frac{1}{\sqrt{h_{it}}} f_{it}\left(z_{it} \left| \nu_i, \xi_i \right.\right) \tag{2.30}$$

[22]For simplicity of exposition, a simple GARCH model is chosen, but in fact any combination of GARCH models may be used.

[23]As mentioned in the previous section, one could adopt instead an empirical or semi-parametric estimation and transformation approach.

where it is clear that the likelihood is composed of a part due to the joint DCC copula dynamics and a part due to the first stage univariate GARCH dynamics.
A similar model, with Student margins, was estimated by Ausin and Lopes (2010) using a Bayesian setup, and an empirical risk management application, albeit once again using only a bivariate series (DAX and Dow Jones indices), used to illustrate its applicability and appropriateness. In Chapter 4, it is shown that this model performs particularly well against a range of related models in the feasible MGARCH universe in a large scale out of sample study on 30 Assets, verifying the benefits of flexible margins combined with a copula DCC model.

2.3 Multivariate Distributions and Normal-Mean Variance Mixtures

While univariate dynamics may easily be extended to the multivariate domain, by simple equation manipulation, the concept of a multivariate distribution is far more complicated and forms a constraining element in the data modelling process. Within financial applications the emphasis has mostly been on either the elliptical methodology, whereby the transition from the univariate to the multivariate domain has been achieved through the construction of densities that are quadratic form functions of the margins, or through copulas, where the dependency structure is separate from marginal dynamics. A key step in the maximization of the likelihood function of a multivariate density with GARCH dynamics, is to appropriately scale the data so that they are i.i.d.[24] This implies that a multivariate density with conditional mean $\mu_t(\theta)$ and conditional variance $\boldsymbol{H}_t(\theta)$, can be scaled so that:

$$f\left(\boldsymbol{y}_t\left|(\theta,\eta)\right.,\boldsymbol{\Omega}_{t-1}\right)=\left|\boldsymbol{H}_t\right|^{-0.5}g\left(\boldsymbol{H}_t^{-0.5}\left(\boldsymbol{y}_t-\boldsymbol{\mu}_t\right)|\eta\right) \tag{2.31}$$

where $g(.)$ is the conditional density of the standardized residuals $\boldsymbol{z}_t$ with possible additional nuisance parameters η. Traditionally, because of its tractability and desirable features, the multivariate normal distribution, uniquely determined by its mean

[24]The weaker assumption that they are a martingale difference sequence with respect to the conditioning information leads to a quasi-likelihood approach.

and covariance, has dominated financial modelling. It possesses many desirable features such as invariance under affine linear transformation, infinite divisibility, self-decomposability and formation of subsequences, making it ideal for the regressive and autoregressive modelling as well as portfolio modelling. It also forms a sufficient condition for the use of mean variance analysis developed by Markowitz (1952) and used extensively in industry to this date. Even when the underlying data generating process is not conditionally Multivariate Normal, it will still yield consistent estimates of θ, as shown by Bollerslev and Wooldridge (1992) (see also Gourieroux (1997) for its asymptotic properties in the context of MGARCH) making it a rather 'forgiving' distribution in terms of consistency in the presence of misspecification. However, it has long been established that the returns of financial assets exhibit characteristics such as fat tails and skewness not captured by the normal distribution and for which such characteristics matter in conditional density forecasting. The class of elliptical distributions, introduced by Kelker (1970), may be considered as generalizations of the multivariate normal distribution and therefore share many of its desirable properties, while also allowing for some tail dependence. Very generally, an elliptical distribution can be considered as an affine transformation of a spherical distribution, the latter being a distribution which is invariant under rotations and reflections.

Definition 5. Elliptical Distributions. A random vector $\boldsymbol{X}$ has a multivariate elliptical distribution, denoted as $\boldsymbol{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, if its characteristic function may be expressed as:

$$\boldsymbol{\varphi}_{\boldsymbol{X}}\left(t\right) = e^{(it'\boldsymbol{\mu})\psi\left(\frac{1}{2}t'\boldsymbol{\Sigma}t\right)}, \tag{2.32}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ and is positive definite, and $\psi(t)$ is some characteristic generator function. While it does not follow that $\boldsymbol{X}$ will necessarily have a density, if it does it will be of the following form:

$$f_{\boldsymbol{X}}\left(\boldsymbol{x}\right) = \frac{c_n}{\sqrt{|\boldsymbol{\Sigma}|}} g_n\left(\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right), \tag{2.33}$$

where g_n is the density generator function and c_n is some normalizing constant. The generator may or may not depend on n, as is the case, for example, in the multivariate normal case where the density generator is given by $g(u) = e^{-u}$. If $\boldsymbol{X}$ belongs to a multivariate elliptical distribution, then assuming that the mean exists, it will be $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ and the covariance, again assuming that it exists, will be $\mathrm{Var}(\boldsymbol{X}) = \frac{\partial\psi(0)}{\partial \mathrm{t}}\boldsymbol{\Sigma}$, coinciding with the standard covariance matrix up to a constant. Examples of elliptical distributions without either a mean or variance are the symmetric stable distributions with index of stability (α) less than 1.[25]

A very desirable feature of elliptical distributions is that of invariance under affine linear transformation. If $\boldsymbol{A}$ is some $q \times n$ matrix of rank $q \leqslant n$ and $b \in \mathbb{R}^q$, then

$$\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \sim E_q\left(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}', g_q\right). \tag{2.34}$$

Hence marginal distributions of elliptical distributions are also elliptical distributions. An additional property, particularly important in portfolio allocation, is that of invariance under convolution of distributions, meaning that the sum of elliptical distributions is also an elliptical distribution.

The ability to model tail dependence with elliptical distribution is rather limited because of their radial symmetry, as upper and lower tail dependence is the same. However, as McNeil, Frey, and Embrechts (2005) have observed, lower tail dependence is often much stronger than upper tail dependence while Rachev and Mittnik (2000) find that there is significant variation in the tail index between different assets. Kring, Rachev, Höchstötter, Fabozzi, and Bianchi (2009) proposed a multi tail generalized elliptical distribution which allows for just such asymmetry and variation between assets' tail index, and derive various properties of this class of distributions. In a slightly different direction, a class of distributions called skew-elliptical, is reviewed in Genton (2004), and includes among others various flavors of the skew-student such as those introduced by Azzalini and Capitanio (2003), and the class of skew distributions of Fernandez and Steel (1998) which additionally include the skew-Normal and skew-GED distributions. However, these distributions are generally unable to capture the variation in tail index between assets. As in the univariate case, it is important to be able to express the moments of the distribution in such a way as to make modelling in a GARCH setup possible. For 2-stage modelling, such as is typically used with the DCC model, the distribution must be elliptical. For a 1-stage modelling, there is much more flexibility in the class of distributions chosen, though more richly parameterized distributions usually have a complicated form for the covariance which may lead to problems in imposing positive definiteness at each iteration of the conditional likelihood. In Chapter 1, the Normal Mean-Variance Mixture distribution of Barndorff-Nielsen, Kent, and Sörensen (1982) was referenced in a univariate context for ACD modelling. In the remainder of this section I review its properties in a multivariate setting and 2 very special members

[25]The normal can be considered a special case of the symmetric stable class with index of stability equal to 2, and is itself a spherical distribution.

of its family derived under different mixing variables.
The n-dimensional random variable $\boldsymbol{X}$ is said to have a normal mean-variance mixture distribution of the following form:

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\boldsymbol{A}\boldsymbol{Z}, \tag{2.35}$$

where $\boldsymbol{Z} \sim N_q(0, I_q)$, $W \in \mathbb{R}^1_+$, $\boldsymbol{A} \in \mathbb{R}^{n\times q}$, and $\boldsymbol{\mu}$, $\boldsymbol{\gamma} \in \mathbb{R}^n$. The basic premise behind this distribution is to introduce noise in the covariance matrix and mean vector of a multivariate Normal distribution through the mixing variable W. The vector-valued variable $\boldsymbol{\gamma}$ introduces skewness, and when it is equal to zero, $\boldsymbol{X}$ is distributed as a scale mixture of Normal distributions. Different mixing distributions for W give rise to different families of distributions. A key distribution arising from this representation, already discussed in Chapter 1 in the univariate context, is the GH distribution. When the mixing variable W is GIG, the n-dimensional GH distribution of Barndorff-Nielsen (1977) arises, which allows for the modelling of multivariate data with some very desirable features such as the ability to model skewness individually for each dimension. Additionally, the distribution has the property of infinite divisibility (inherited from the GIG mixing distribution), and is closed under margining, conditioning and linear affine transformations, and in the case of the NIG distribution is also closed under convolution when the margins have the same skew and shape parameters.

Definition 6. The n-dimensional Generalized Hyperbolic distribution of the random vector $\boldsymbol{X} \in \mathbb{R}^n$

$$\begin{aligned} GH_n\left(\boldsymbol{x};\alpha,\boldsymbol{\beta},\delta,\boldsymbol{\mu},\boldsymbol{\Sigma}\right) &= c_n \frac{K_{\lambda-n/2}\left(\alpha\sqrt{\delta^2+\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)}\right)}{\left(\frac{1}{\alpha}\sqrt{\delta^2+\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)}\right)^{n/2-\lambda}} e^{\boldsymbol{\beta}'\left(\boldsymbol{x}-\boldsymbol{\mu}\right)}. \\ c_n &= \frac{\left(\sqrt{\alpha^2-\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}/\delta\right)^{\lambda}}{\left(2\pi\right)^{n/2} K_\lambda\left(\delta\sqrt{\alpha^2-\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}\right)}. \end{aligned} \tag{2.36}$$

with parameter domain of variation, $\lambda \in \mathbb{R}$, $\boldsymbol{\beta}, \boldsymbol{\mu} \in \mathbb{R}^n$, $\delta > 0$, $\alpha^2 > \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ with determinant 1. The fact that the domain of α is 1-dimensional, is due to the univariate GIG mixing distribution, and means that kurtosis is the same for all dimensions. That is, there is one joint representation of extreme events, which may not be an adequate reflection of the multivariate data, especially when they come from not very highly correlated (at least in the tail sense) sets. Special subfamilies of the distribution can be derived, as in the univariate case discussed in Chapter 1, including the multivariate NIG, VG and (skew) Student, while the Laplace and Normal are limiting cases under the restrictions $\alpha \to \infty, \delta \to \infty, \frac{\delta}{\alpha} \to \sigma^2 < \infty$ respectively, which means that they are directly comparable under a range of measures. A nice property of GH distributions, inherited from their mean-variance mixture representation, is that they are closed under margining, conditioning and regular affine transformations[26] as discussed in Bläsild (1981), hence satisfying the requirements of most financial applications. This also implies the ability to create location and scale invariant parameterizations, desirable for the modelling of MGARCH processes. An affine representation of this distribution is considered in Chapter 3 where a new MGARCH model is presented, while I consider another member of the normal mean-variance mixture distribution in the next section which arises when W is Gamma ($\mathbf{\Gamma}$) distributed.

2.3.1 The Multivariate Laplace Model and its Extensions

The Laplace, with its towering peak and heavy tails, has a special place alongside the Normal distribution, being stable under geometric rather than ordinary summation, thus making it suitable for stochastic modelling. In regression modelling, when the errors have a Laplace distribution, then the least absolute deviation is also the maximum likelihood estimate, equivalent to the least squared deviation estimate when the errors have a Normal distribution. This can be easily inferred from the density function of the Laplace which differs mainly from the Normal by including a term for the mean absolute rather squared deviation of a random variable x. It also arises as a special case in the Generalized Error distribution with shape parameters = 1, and the Geometric Stable distribution with stability parameter = 2, and zero skewness and location (also called the Linnik distribution with stability parameter = 2). Because it has tails heavier than the Normal distribution it is more suitable for the modelling of financial returns. In the multivariate setting, the multivariate Laplace has been analyzed, among others, by Anderson (1992) as part of the multivariate Linnik distribution, Marshall and Olkin (1993) and Kalashnikov (1997) as a multivariate distribution generated by i.i.d. univariate Laplace margins, Fernandez, Osiewalski, and Steel (1995) as a natural generalization of the univariate model to d-dimensions in the framework of the multivariate exponential power distribution. Recently, the distribution has been extended by Kotz, Kozubowski, and Podgorski (2001) to the Generalized Asymmetric Laplace

[26] Assuming certain restrictions on the parameters.

(GAL) distribution which is a member of the Normal Mean-Variance Mixture distribution given in (2.35) with mixing variable W distributed $\mathbf{\Gamma}(\theta, 1)$. The density is given by:

$$f_{GAL}\left(\boldsymbol{x};\boldsymbol{m},\boldsymbol{\Sigma},\boldsymbol{\gamma}\right) = \frac{2|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{n/2}\mathbf{\Gamma}\left(\theta\right)} e^{(\boldsymbol{x}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}} \left(\frac{\left(\boldsymbol{x}-\boldsymbol{m}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{m}\right)}{2+\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}\right)^{\frac{2\theta-n}{4}}$$
$$\times K_{\frac{2\theta-n}{2}}\left(\sqrt{\left(2+\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right)\left(\boldsymbol{x}-\boldsymbol{m}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{m}\right)}\right) \tag{2.37}$$

where $\boldsymbol{m} \in \mathbb{R}^n$ is the location vector, $\boldsymbol{\gamma} \in \mathbb{R}^n$ the skewness vector and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ a positive definite scaling matrix. The parameter θ inherited from the $\mathbf{\Gamma}$ mixing distribution determines various subfamilies of this distribution such as the widely cited Asymmetric Multivariate Laplace (AML) described in Kotz, Kozubowski, and Podgorski (2001), Kozubowski and Podgórski (2001) and Kotz, Kozubowski, and Podgórski (2002). This particular subfamily is derived by setting $\theta = 1$, but it is also and possible to derive it as a special case of the GH distribution by setting the restrictions $\lambda = 1$, $\boldsymbol{\mu} = 0$, $\delta^2 = 0$, $\alpha^2 - \boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta} = 2$ and therefore that $\alpha = \sqrt{2 + \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}$, in Equation (2.36). Using an asymptotic relation discussed in Kotz, Kozubowski, and Podgorski (2001) when $\delta^2 = 0$, the 2 parameter n-dimensional AML density ($AML_n(\boldsymbol{m}, \boldsymbol{\Sigma})$) may be written as:

$$AML_n\left(\boldsymbol{x};\boldsymbol{m},\boldsymbol{\Sigma}\right) = \frac{2e^{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \left(\frac{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{2+\boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}\right)^{\nu/2} \mathrm{K}_{\nu}\left(\sqrt{\left(2+\boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right)\left(\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}\right)}\right) \tag{2.38}$$

where $\nu = \left(2 \text{ - n}\right)/2$. The mixing distribution is therefore $GIG(1, 0, 2)$ which is the standard exponential. The moments of the distribution follow from its mean-variance mixture form in Equation (2.35) and are:

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{X}\right) &= \boldsymbol{m} \\ \mathrm{Var}\left(\boldsymbol{X}\right) &= \boldsymbol{\Sigma} + \boldsymbol{m}\boldsymbol{m}'. \end{aligned} \tag{2.39}$$

The distribution is described as closed under a scaling transformation, but not as it is appears under a location shift. Formally, let $\boldsymbol{X} = (X_1, \ldots, X_n)' \sim AML_n\left(\boldsymbol{m}, \boldsymbol{\Sigma}\right)$ and $\boldsymbol{A}$ be a real $q \times n$ matrix, then the random vector $\boldsymbol{AX}$ is $AML_q\left(\boldsymbol{Am}, \boldsymbol{A\Sigma A}'\right)$. While this property is fine for portfolio applications where a linear combination of assets results in a univariate AL distribution, for estimation purposes, it is not possible to obtain a location invariant parametrization of the distribution since the moment we subtract the mean vector we are effectively left with a non-skewed version of the distribution. Consider the distribution of the zero mean residual $\boldsymbol{\varepsilon}_t = \boldsymbol{x}_t - \boldsymbol{m}_t$, where $\boldsymbol{x}_t \sim AML_n\left(\boldsymbol{m}_t, \boldsymbol{\Sigma}_t\right)$, which under a scaling transformation by the matrix $\boldsymbol{A}$, is equal to

$\boldsymbol{A}(\boldsymbol{x}_t - \boldsymbol{m}_t) \sim AML_q(0, \boldsymbol{A\Sigma}_t\boldsymbol{A}') \equiv ML_q(0, \boldsymbol{A\Sigma}_t\boldsymbol{A}')$. Since the estimation of 2-stage DCC models is done on the zero mean residuals, the conditional distribution of $\boldsymbol{z}_t$ (see Equation (2.12)) is that of the symmetric Laplace, while the conditional distribution of the returns is Asymmetric. That is, the conditional mean $\boldsymbol{m}_t$ from a possible filtration process imparts the asymmetry to the conditional distribution of the returns[27]. Cajigas and Urga (2006) discuss the properties of the AML distribution in the context of the AGDCC model and proof of consistency established for the 2-stage estimation under the multivariate symmetric Laplace, despite referencing the AML as the main distribution of the paper. In this case, when $\boldsymbol{m} = 0$, the distribution is Pareto stable, just as in the Normal case, and in contrast to most of the other GH sub-family of distributions. This condition implies an important property necessary for the modelling of financial portfolios known as the additivity property, which is basically the concept that a linear combination of independent random variables with stability index α is also stable with the same parameter α (see Khindanova, Rachev, and Schwartz (2001)). Formally, the random variable $\mathbf{X}$ is said to be Pareto stable if for any $a_i > 0$, $i = 1, ..., n$ there exist a constant $c = d^{1/\alpha}$ and $\mathbf{u}_d \in \mathbb{R}^n$ for any $n \geq 2$ such that,

$$a_1\mathbf{X}^{(1)}+...+a_d\mathbf{X}^{(d)} \stackrel{D}{=} c\mathbf{X} + \mathbf{u}_d \tag{2.40}$$

where $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(d)}$ are independent copies of $\mathbf{X}$. In an alike way Laplace laws are stable, but under geometric summation instead of random summation[28]. To be able

[27] This rather strange situation is further discussed in Lindsey and Lindsey (2006)

[28] In the geometric stable model, the return $r_{f(p)}$ is considered to be the sum of smaller returns $r^{(i)}$ over the period of time $f(p)$ which is a stopping time random variable with geometric probability function $P(f(p) = j) = p(1-p)^{j-1}$, $j = 1, 2, \ldots$.
The geometric stable distribution can be approximated to a normalized geometric stable model sum when the p parameter of the stopping time function $f(p)$ approaches zero. More formally, the random array $\mathbf{X}$ has a geometric stable distribution in $\mathbb{R}^n$ if and only if:

$$a(p)\sum_{i=1}^{f(p)}\left(\kappa(p) + \mathbf{r}^{(i)}\right) \xrightarrow{d} \mathbf{X}, \qquad \text{as } p \to 0 \tag{2.41}$$

where $\left\{\mathbf{r}^{(d)} = \left(r_1^{(d)}, \ldots, r_n^{(d)}\right), d \geq 1\right\}$ is a sequence of i.i.d. random vectors in $\Re^n$ independent of $f(p)$, $a(p) > 0$, $\kappa(p) \in \mathbb{R}^n$, and $\xrightarrow{d}$ denotes convergence in distribution. Kozubowski and Podgórski (2001) show that when each vector in $\mathbf{r}$ has mean m_i, $i = 1, ...n$, a variance σ_{ij}, $i = 1, ...n$, $j = 1, ...n$, and for $a(p) = \sqrt{p}$ and $\kappa(p) = \mathbf{m}\left(\sqrt{p} - 1\right)$, the random variable $\mathbf{X}$ defined by the convergence in distribution property in Equation (2.41) has an AML distribution with the characteristic function:

$$\Psi(\mathbf{t}) = \frac{1}{1 + \frac{1}{2}\mathbf{t}'\mathbf{H}\mathbf{t} - i\mathbf{t}'\mathbf{m}} \tag{2.42}$$

where $\mathbf{t} \in \mathbb{R}^n$, and $\mathbf{H} \in \mathbb{R}^{n\times n}$ is a positive-definite matrix.

to preserve stability we have to constrain the normalising constants $a(p)$ and $\kappa(p)$ in Equation (2.41) to:

$$a(p) = \sqrt[\alpha]{p}, \quad \kappa(p) = \mathbf{0} \tag{2.43}$$

The first condition implies that for the case of the AML distribution $\alpha = 2$. This is the same α value of the normal distribution which is the only Pareto-stable distribution with a finite second moment. The second condition $\kappa(p) = \mathbf{0}$ implies $\mathbf{m} = \mathbf{0}$, limiting one to the use of the symmetric version of the distribution. A final point of interest arises from the mixture representation of the AML in that $m = \Sigma\beta$, where β is the d-dimensional asymmetry vector of the GH distribution. When Σ is time varying, as in the MGARCH models, this induces time variation in β and hence skewness.
In subsequent papers, Kotz, Kozubowski, and Podgórski (2002) and Kozubowski and Podgórski (2001) describe further applications of the AML distribution, albeit in a non dynamic setting. In an attempt to alleviate these problems, Kollo and Srivastava (2005) reparameterized the AML to have covariance which does not depend on the mean vector albeit present only the characteristic function of the distribution, while Arslan (2010) derives an alternative multivariate skew Laplace distribution from the GAL distribution where the mean now depends on location $\boldsymbol{m}$ and asymmetry $\boldsymbol{\gamma}$, $\in R^n$, while the covariance matrix now depends only on the asymmetry vector $\boldsymbol{\gamma}$. More formally, setting θ in Equation (2.37) to $2n-1$ yields the following distribution:

$$f_{MSL}\left(\boldsymbol{x};\boldsymbol{m},\boldsymbol{\Sigma},\boldsymbol{\gamma}\right) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{2^n \pi^{(n-1)/2} \alpha \boldsymbol{\Gamma}\left(\frac{n+1}{2}\right)} e^{-\alpha\sqrt{(\boldsymbol{x}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m})}+(\boldsymbol{x}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}} \tag{2.44}$$

where $(\boldsymbol{x},\boldsymbol{m},\boldsymbol{\gamma}) \in \mathbb{R}^n$, and $\boldsymbol{m}$ now represent the location parameter vector and $\boldsymbol{\gamma}$ the skew parameter vector, $\boldsymbol{\Sigma}$ is an $\mathbb{R}^{n\times n}$ positive definite scaling matrix, and $\alpha = \sqrt{1+\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}$. The first two moments of the distribution are now, $\mathrm{E}\left(\boldsymbol{X}\right) = \boldsymbol{m} + (\mathrm{n}+1)\,\boldsymbol{\gamma}$ and $\mathrm{Var}\left(\boldsymbol{X}\right) = \boldsymbol{H} = (\mathrm{n}+1)\left(\boldsymbol{\Sigma}+2\boldsymbol{\gamma}\boldsymbol{\gamma}'\right)$. Standardization of the distribution proceeds as detailed in Equation (2.31), by making use of the moment conditions and the fact that the MSL distribution is closed under location and scaling transformations[29]. Formally, let:

- $\boldsymbol{H} = \mathrm{Var}\left(\boldsymbol{X}\right) = (\mathrm{n}+1)\left(\boldsymbol{\Sigma}+2\boldsymbol{\gamma}\boldsymbol{\gamma}'\right)$,
- $\boldsymbol{b} = -\mathrm{E}\left(\boldsymbol{X}\right) = -\boldsymbol{m} - (\mathrm{n}+1)\,\boldsymbol{\gamma}$,

[29] If $\boldsymbol{X} \sim MSL_n(\boldsymbol{m},\boldsymbol{\Sigma},\boldsymbol{\gamma})$, and $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}$, where $\boldsymbol{A}^{q\times n}$ and $\boldsymbol{b}\in\mathbb{R}^q$. then $\boldsymbol{Y} \sim MSL_q(\boldsymbol{A}\boldsymbol{m}+\boldsymbol{b},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}',\boldsymbol{A}\boldsymbol{\gamma})$.

- $\boldsymbol{Z} = \boldsymbol{H}^{-0.5}\left(\boldsymbol{X} + \boldsymbol{b}\right)$,

and define the transformed parameters

- $\bar{\boldsymbol{m}} = \boldsymbol{H}^{-0.5}\boldsymbol{m} + \boldsymbol{H}^{-0.5}\boldsymbol{b}$,
- $\bar{\boldsymbol{\Sigma}} = \boldsymbol{H}^{-0.5}\boldsymbol{\Sigma}\boldsymbol{H}'^{-0.5}$,
- $\bar{\gamma} = \boldsymbol{H}^{-0.5}\boldsymbol{\gamma}$,

thus $\boldsymbol{Z} \sim \left(\bar{\boldsymbol{m}}, \bar{\boldsymbol{\Sigma}}, \bar{\gamma}\right)$. The density of $\boldsymbol{X}$ can be represented in terms of the density of the transformed i.i.d variable $\boldsymbol{Z}$ as:

$$\begin{aligned} f\left(\mathbf{H}^{-0.5}\left(\mathbf{X} - E\left(\mathbf{X}\right)\right)\right) &= \left|\mathbf{H}\right|^{-0.5} g\left(\mathbf{Z}\right) \\ &= \left|\mathbf{H}\right|^{-0.5}\left(\frac{\left|\bar{\boldsymbol{\Sigma}}\right|^{-1/2}}{2^{n}\pi^{(n-1)/2}\bar{\alpha}\boldsymbol{\Gamma}\left(\frac{n+1}{2}\right)} e^{-\bar{\alpha}\sqrt{(\mathbf{z}-\bar{\mathbf{m}})'\bar{\boldsymbol{\Sigma}}^{-1}(\mathbf{z}-\bar{\mathbf{m}})} + (\mathbf{z}-\bar{\mathbf{m}})'\bar{\boldsymbol{\Sigma}}^{-1}\bar{\gamma}}\right) \end{aligned} \tag{2.45}$$

In terms of a BEKK-MSL model, the conditional log-likelihood at time t to be maximized is then given by:

$$\begin{aligned} LL_t\left(\boldsymbol{z}_t | \bar{\boldsymbol{m}}_t, \bar{\boldsymbol{\Sigma}}_t, \bar{\gamma}\right) &= -0.5\log\left(\left|\mathbf{H}_t\right|\right) - 0.5\log\left(\left|\bar{\boldsymbol{\Sigma}}_t\right|\right) - n\log\left(2\right) - 0.5\left(n-1\right)\log\left(\pi\right) \\ &- \log\left(\bar{\alpha}\right) - \log\boldsymbol{\Gamma}\left(\frac{n+1}{2}\right) - \bar{\alpha}\sqrt{\left(\mathbf{z}_t - \bar{\mathbf{m}}_t\right)'\bar{\boldsymbol{\Sigma}}_t^{-1}\left(\mathbf{z}_t - \bar{\mathbf{m}}_t\right)} \\ &+ \left(\mathbf{z}_t - \bar{\mathbf{m}}_t\right)'\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\gamma} \end{aligned} \tag{2.46}$$

where $\boldsymbol{H}_t$ is derived from the BEKK dynamics. The parameters of the MSL distribution must be estimated jointly, therefore a 2-stage DCC does not appear feasible. For the BEKK model on the other hand, this seems a perfectly adequate distribution adding only $2 \times n$ parameters beyond the multivariate Normal. It is even possible to reduce this to just n parameters if we make some sacrifices in the efficiency of the location estimator by conditioning it on the unconditional mean of the data $\mathrm{E}\left(\boldsymbol{X}\right) = \boldsymbol{\mu}$ and then only estimating the skewness vector $\boldsymbol{\gamma}$, so that the location $\boldsymbol{m} = \boldsymbol{\mu} - (n+1)\boldsymbol{\gamma}$. Finally, the conditional covariance from the model $\boldsymbol{H}_t$ is used to derive the conditional scaling matrix $\boldsymbol{\Sigma}_t$. The next section considers an in-sample empirical exercise using the BEKK and AGDCC models for a range of distributions.

2.4 Empirical Application

In the BEKK model, the diagonal and off diagonal covariance dynamics are jointly estimated in one single step. In the AGDCC model, the correlation, not covariance dynamics are estimated, given the diagonal volatility estimated from a first stage where there is flexibility as to the model used, so that the standardized residuals may better be filtered for asymmetry and other related effects. The advantage of this two stage approach with flexible marginal dynamic models must be weighed against drawbacks discussed in the previous sections such as the limited choice in admissible multivariate distributions and certain inconsistencies in the DCC correlation setup. In their paper, aptly titled *'Do we really need both BEKK and DCC?'*, Caporin and McAleer (2012) argue that the scalar BEKK, regardless of whether targeting was used or not, is the optimal model to use, though this is not followed by any empirical investigation and is based on derived arguments. In this section, I consider this question by undertaking an in-sample comparison[30] of the diagonal AGDCC and BEKK models, with covariance targeting, using the same dataset presented in Chapter 1 of the 14 MSCI world equity indices for the period 12/08/1996 to 02/03/2011. Admissible multivariate distributions used for both models were the Multivariate Normal (MVN) and Multivariate Laplace (MVL) discussed in the previous section, whilst for diagonal BEKK the Multivariate Student (MVT) and Multivariate Skew Laplace (MSL) were also used, in order to investigate the degree of kurtosis[31] and asymmetry in the conditional distribution. For the diagonal AGDCC model, the first stage estimation consisted of a joint AR(2)-GARCH(1,1) model, whilst for the diagonal BEKK, the dataset was filtered by a first stage AR(2) model.

[30]Because of the dimensionality of the models, an out of sample application was beyond the computational resources available to this researcher.

[31]Since the excess kurtosis in the MVL is fixed at 3, the use of the MVT distribution provides for a more flexible calibration of the actual kurtosis which might be present

Table 2.2: Diagonal BEKK Model under 4 conditional distributions (14 MSCI iShares)

	MVN		**MVT**			**MVL**		**MSL**		
	A	**B**	**A**	**B**	**shape**	**A**	**B**	**A**	**B**	**skew**
USA	0.133***	0.990***	0.117***	0.992***		0.109***	0.993***	0.124***	0.992***	-3.0E-4***
Canada	0.122***	0.991***	0.107***	0.993***		0.098***	0.993***	0.112***	0.992***	-3.8E-4***
Mexico	0.129***	0.990***	0.108***	0.993***		0.100***	0.993***	0.115***	0.992***	-4.7E-4***
Australia	0.117***	0.992***	0.102***	0.994***		0.094***	0.994***	0.109***	0.993***	-5.0E-4***
Hong Kong	0.131***	0.990***	0.115***	0.992***		0.109***	0.992***	0.122***	0.992***	-3.9E-4***
Japan	0.131***	0.990***	0.113***	0.992***		0.105***	0.992***	0.120***	0.991***	-2.6E-4***
Singapore	0.129***	0.991***	0.117***	0.992***		0.111***	0.992***	0.122***	0.992***	-4.2E-4***
Germany	0.127***	0.991***	0.110***	0.993***		0.101***	0.994***	0.116***	0.992***	-4.0E-4***
France	0.114***	0.993***	0.102***	0.994***		0.093***	0.995***	0.107***	0.994***	-3.6E-4***
Spain	0.115***	0.993***	0.103***	0.994***		0.092***	0.995***	0.106***	0.994***	-3.0E-4***
Italy	0.114***	0.993***	0.102***	0.994***		0.092***	0.995***	0.106***	0.994***	-3.2E-4***
UK	0.120***	0.992***	0.104***	0.994***		0.096***	0.994***	0.110***	0.993***	-3.5E-4***
Switzerland	0.116***	0.992***	0.099***	0.994***		0.089***	0.995***	0.105***	0.994***	-3.2E-4***
Sweden	0.121***	0.992***	0.110***	0.993***		0.103***	0.993***	0.114***	0.992***	-4.2E-4***
					10.018***					
LL	**161,696**		**163,255**			**162,082**		**163,118**		

Notes to table 2.2: The Table presents the diagonal BEKK parameter estimates and Log-Likelihood (LL) under 4 different conditional distributions, for the daily log returns of 14 MSCI World iShares for the period 12/08/1996 to 02/03/2011. The data was first filtered by an estimated AR(2) model prior to the MGARCH estimation. The *, ** and *** next to the parameters denote significance at the 10%, 5% and 1% levels respectively.

TABLE 2.3: Diagonal AGDCC model under 2 conditional distributions (14 MSCI iShares)

	MVN			**MVL**		
	A	**B**	**G**	**A**	**B**	**G**
USA	0.152***	0.986***	0.043	0.090***	0.993***	0.071***
Canada	0.126***	0.990***	0.017	0.080**	0.994***	0.062***
Mexico	0.129***	0.988***	0.029	0.068	0.994***	0.073*
Australia	0.129	0.987***	0.056	0.068***	0.995***	0.066**
Hong Kong	0.120***	0.986***	0.076	0.071*	0.993***	0.089**
Japan	0.108**	0.989***	0.044	0.066***	0.995***	0.053**
Singapore	0.122***	0.986***	0.086	0.058	0.994***	0.098*
Germany	0.164***	0.982***	-0.004	0.094***	0.993***	0.040*
France	0.179**	0.980***	-0.013	0.100***	0.993***	0.043*
Spain	0.160***	0.984***	-0.001	0.106***	0.992***	0.048
Italy	0.164**	0.983***	-0.011	0.100***	0.993***	0.047
UK	0.146***	0.986***	0.006	0.082***	0.994***	0.052***
Switzerland	0.125**	0.990***	-0.032	0.078**	0.996***	0.011
Sweden	0.141***	0.986***	-0.003	0.077***	0.994***	0.051***
LL		**162,185**			**162,230**	

Notes to table 2.3: The Table presents the diagonal AGDCC parameter estimates and Log-Likelihood (LL) under 2 different conditional distributions, for the daily log returns of 14 MSCI World iShares for the period 12/08/1996 to 02/03/2011. The data was first filtered by an estimated AR(2)-GARCH(1,1) model prior to the second stage estimation. The *, ** and *** next to the parameters denote significance at the 10%, 5% and 1% levels respectively

The estimated parameters from the diagonal BEKK and AGDCC models are presented in Tables 2.2 and 2.3, respectively. For the BEKK model, all parameter estimates were found to be highly significant. Judging from the value of the shape parameter of the MVT distribution, this implies an excess kurtosis of 1 which is significantly lower than that of the MVL and MSL distributions which is fixed at 3. The MVT was also found to have the highest likelihood among 4 distributions employed, a comparison of which is possible as a result of the distributions being nested in the Multivariate Mean Variance Mixture family. The skew parameter for the MSL distribution was also found to be highly significant and negative for all countries as one should expect, and without any surprises for the Asian indices as found in the unconditional statistics in Table 1.1. For the AGDCC model, the estimates of the shock and persistence parameters (A and B) were mostly significant with the exception of the shock parameters in the case of Australia under the MVN distribution, and Mexico and Singapore under the MVL distribution. The conditional correlation asymmetry parameter (G) was insignificant under the MVN distribution but mostly significant under the MVL distribution. The reason for this is not immediately clear, though the literature on the Laplace indicates that it is more robust in the presence of outliers which, under the Normal, are penalized at a squared rate. Thus if outliers are given more weight under the Normal, then it is likely that dynamics which are more sensitive to the center of the mass, such as those which measure separate reactions to positive and negative shocks, become crowded out by extreme observations which place the focus on the tails. Therefore, in addition to the accommodation of higher excess kurtosis, the MVL likelihood is somewhat higher than that of the MVN.

In order to obtain a more complete picture of the goodness of fit of each model and distribution combination with respect to this dataset, I make use of the GMM and non-parametric density misspecification tests of Hansen (1982) and Hong and Li (2005), respectively, introduced in Sections 1.6.2 and 1.6.3. These are univariate tests for which the comparison is carried out on the weighted conditional density of each model.[32] To alleviate any bias from using any particular set of weights, 5000 randomly sampled set of weights were generated from the exponential distribution and standardized to sum to 1. Table 2.4 reports the average value and t-statistic from the GMM test on the standardized weighted residuals of each model, where the first four columns denote the unconditional population moment conditions ($E[z]$, $E[z^2]-1$,$E[z^3]$, $E[z^4]-3$) that

[32]There are no equivalent feasible or practical tests available in the multivariate domain and this is therefore considered a next best alternative method.

should be satisfied, column Q_2 the test for the conditional variance using 4 lags, and column J the joint moment conditions.[33] It is immediately clear that in terms

TABLE 2.4: MGARCH models: GMM misspecification test (14 MSCI iShares)

	$E[z]$	$E[z^2]-1$	$E[z^3]$	$E[z^4]-3$	Q_2	J
BEKK(D)-MVN	0.0004	-0.0543	-0.4684	2.5419		
	[0.02]	[-1.53]	[-2.72]	[2.58]	[34.60]	[105.07]
BEKK(D)-MVT	-0.0001	-0.0600	-0.4478	1.6785		
	[-0.01]	[-1.66]	[-2.54]	[1.63]	[38.39]	[111.86]
BEKK(D)-MVL	-0.0006	-0.0867	-0.4074	-0.5473		
	[-0.04]	[-2.45]	[-2.40]	[-0.67]	[40.48]	[140.87]
BEKK(D)-MSL	0.0001	-0.0547	-0.2171	-0.3352		
	[0.01]	[-1.52]	[-1.22]	[-0.45]	[37.02]	[89.80]
AGDCC(D)-MVN	-0.0385	-0.0174	-0.5903	2.0646		
	[-2.35]	[-0.52]	[-3.72]	[2.16]	[9.38]	[41.93]
AGDCC(D)-MVL	-0.0380	-0.0205	-0.5769	-1.0127		
	[-2.32]	[-0.62]	[-3.72]	[-1.19]	[10.48]	[139.78]

Note: The Table reports the average statistic and t-values from the GMM moment based test of Hansen (1982) applied to the weighted standardized residuals of the in-sample fit of 14 MSCI iShares for the period 12/08/1996 to 02/03/2011, from the diagonal BEKK and diagonal AGDCC models under alternative conditional distributions. Using a set of 5000 randomly generated weights with full budget constraint, the weighted margins were used to test for misspecification in the individual moments ($E[z]$, $E[z^2]-1$, $E[z^3]$, $E[z^4]-3$), the conditional variance (Q_2) under 4 lags and all the conditions jointly (J). Values in square brackets represent the t-values, which for the individual moment conditions may be tested using a t-test under the null of not being significantly different from zero, while the conditional variance and joint cases may be tested using a Wald test with 4 and 8 d.o.f. respectively (with critical values of 9.49 and 15.5).

of overall fit, none of the models provide an adequate representation of the underlying dynamics with joint statistic t-values well outside the critical value. A closer look reveals why. Considering the small t-value of the first moment condition under the BEKK model, it is clear that the mean of the series has been adequately removed (i.e. the mean is not significantly different from zero). This is in direct contrast to the AGDCC model where the t-values are quite high indicating just the opposite. This is somewhat surprising, since both models were first filtered by estimating an AR(2) model. However, in the case of the AGDCC model, this was based on a first stage AR(2)-GARCH(1,1) joint estimation whilst in the BEKK case an AR(2) model with constant volatility was used for demeaning the series. Estimating the AR(2)-GARCH(1,1) model jointly is the more efficient strategy[34], but it would seem that the DCC decomposition introduces some inefficiency back into the unconditional mean of the standardized residuals. This cost for using the AGDCC model is completely reversed in the case of the unconditional and conditional variance conditions ($E[z^2]-1$ and Q_2), where the BEKK model

[33]For the last 2 columns, only the t-value is reported since the actual value is a vector and cannot be represented compactly in a table. The critical values for the Q_2 and *Joint* conditions at the 95% confidence level, distributed χ^2 with p d.o.f. and $4+p$ d.o.f, are 9.488 and 15.507, respectively.

[34]Two stage estimation is consistent but not as efficient, see Engle (1982).

now seems to fare rather badly, as evidence by the high t-values versus the AGDCC. The first stage univariate GARCH filter appears to provide for some extra efficiency in this sample size. Taken together, these two conditions (making up 5 of the 8 d.o.f. in the test) lead to a much better fit overall. With respect to the higher moment conditions and the distributions, there are no surprises with the MVN failing to capture the skewness and kurtosis in the dataset, the MVT and MVL failing to capture skewness but capturing kurtosis, and the MSL being the only distribution to adequately capture all the unconditional moment conditions. However, the high t-value for the Q_2 moment condition means that the diagonal BEKK model does not adequately capture the structure of the conditional variance. Table 2.5 reports the average statistic

TABLE 2.5: MGARCH models: Hong-Li misspecification test (14 MSCI iShares)

Hong-Li Non-Parametric Test					
	$M(1,1)$	$M(2,2)$	$M(3,3)$	$M(4,4)$	W
BEKK-MVN	43.7	15.9	5.8	1.8	59.8
	[100]	*[99.8]*	*[90.3]*	*[54.7]*	*[100]*
BEKK-MVT	43.1	16.3	6.5	2.5	40.3
	[100]	*[99.9]*	*[92.6]*	*[63.6]*	*[100]*
BEKK-MVL	44.3	19.2	9.1	4.4	40.1
	[100]	*[99.9]*	*[96.4]*	*[80.2]*	*[100]*
BEKK-MSL	43.1	18.8	9.2	4.7	32.4
	[100]	*[99.9]*	*[96.4]*	*[82]*	*[100]*
AGDCC-MVN	28.7	9.9	3.9	2.8	29.1
	[100]	*[99.1]*	*[93]*	*[91.6]*	*[100]*
AGDCC-MVL	28.2	11.8	5.6	3.5	48.6
	[100]	*[99.1]*	*[95.3]*	*[92.5]*	*[100]*

Note: The Table reports the average statistic and percent rejections from the non parametric density test of Hong and Li (2005) applied to the probability integral transformed weighted margins of the in-sample fit of 14 MSCI iShares for the period 12/08/1996 to 02/03/2011, from the diagonal BEKK and diagonal AGDCC models under alternative conditional distributions. $M(j,j), j = 1,..,4$, represent the nonparametric tests for misspecification in the conditional moments, and distributed as $N(0,1)$ under the null of a correctly specified model. The statistic W in column 5 of the table is the Portmanteau type test statistic for general misspecification (using 4 lags) and distributed as $N(0,1)$ under the null of a correctly specified model. Values in square brackets are the percent rejections under the null, with 95% confidence, for the 5000 randomly weighted margins.

and percent rejections from applying the test of Hong and Li (2005) on the randomly weighted margins of each model. The first four columns are tests of the cross-correlation of each moment based on four lags and described in Appendix C, whilst column W is the nonparametric Portmanteau type test of the overall goodness of fit. Considering the first column ($M(1,1)$), this tells a rather different story to the GMM test previously considered and indicates that the AGDCC models, which filter the dataset with a first stage joint AR(2)-GARCH(1,1) model do a better job at capturing the conditional mean than a simple AR(2) filter with constant variance as in the case of the BEKK models. This is not surprising since the GMM test considered the unconditional mean, whereas this test considers the conditional mean. It is difficult to draw substantial conclusions from the other cross-correlation based tests, other than to note that for the conditional fourth moment ($M(4,4)$) the MVL distribution appears to badly fit the tails since it imposes a fixed excess kurtosis of 3 whereas the observed excess kurtosis from the MVT distribution appears to be only marginally higher than what would be expected from the Normal case. Looking at the main goodness of fit statistic in column 5, it would appear that again the AGDCC dynamics provide for a much better relative overall fit, possibly arising from the lower statistic in the first and second conditional moments, closely followed by the BEKK-MSL for which it is not immediately clear why, looking at the conditional moment tests but likely related to the accommodation of both asymmetry and fat tails in the conditional distribution.

The empirical application considered thus far has indicated that the diagonal

Table 2.6: Scalar BEKK vs DCC: cost of misspecification

			DGP: AR(2)-GARCH(SSTD)-DCC-Copula(MVT)		
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
DCC(MVN)					
statistic	[4.405 ; -1.672]	[2.770 ; -1.545]	[2.283 ; -1.393]	[2.708 ; -1.214]	[7.468 ; 0.583]
%Rejections	10.5	11.0	13.0	14.8	38.1
sBEKK(MVN)					
statistic	[27.828 ; -1.550]	[25.920 ; -1.353]	[20.311 ; -1.102]	[15.292 ; -0.848]	[90.130 ; 1.308]
%Rejections	16.7	18.1	21.3	24.1	50.0
			DGP: AR(2)-sBEKK(MVN)		
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
DCC(MVN)					
statistic	[-1.614 ; -1.794]	[-1.459 ; -1.673]	[-1.513 ; -1.247]	[-1.064 ; -1.414]	[-0.706 ; -0.930]
%Rejections	0.2	0.8	2.2	3.6	7.6
			DGP: AR(2)-GARCH(N)-DCC(MVN)		
	M(1,1)	**M(2,2)**	**M(3,3)**	**M(4,4)**	**W**
sBEKK(MVN)					
statistic	[-1.388 ; -1.65]	[-1.189 ; -1.504]	[-0.918 ; -1.308]	[-0.674 ; -1.158]	[-0.348 ; -0.573]
%Rejections	1.2	2.3	4.8	7.8	11.9

Note: The Table reports the mean and median (square brackets) statistic and percent rejections under the test of Hong and Li (2005) for the randomly weighted margins of the DCC and scalar BEKK (sBEKK) models fitted to data sampled under 3 Data Generating Processes (*DGP*) identified in the heading of each subpanel. For each DGP, 1000 scenarios of size 1000 were generated and fitted using the scalar versions of the DCC and BEKK models with multivariate Normal distribution (*MVN*). The probability integral transformation was then applied to the 100 randomly weighted margins of each scenario under each model and the percent of rejections under the non parametric test evaluated. $M(j,j), j = 1,..,4$, represents the nonparametric test for misspecification in the conditional moments, and distributed as $N(0,1)$ under the null of a correctly specified model, while W in column 5 of the table is the Portmanteau type test statistic for general misspecification (using 4 lags) and distributed as $N(0,1)$ under the null of a correctly specified model.

AGDCC somewhat outperforms the diagonal BEKK model with conditional distribution the MVN. The assumption made, given the evidence presented, was that the first stage conditional mean and variance filter provided for greater robustness in the presence of misspecification, and in particular in the presence of non-normally distributed returns. To provide further evidence of this, I perform a Monte Carlo experiment as follows. First, I estimate an AR(2)-GARCH(SSTD)-DCC copula (MVT) model on the 14 MSCI indices for the same period considered previously. Using a DCC-copula model was deemed a very flexible method for capturing the observed underlying features of the dataset, if not the full covariance dynamics since the scalar DCC was used.[35] The first stage Skew Student distribution ($SSTD$) of Fernandez and Steel (1998) was already discussed in Section 1.5.6 and allows the filtering of asymmetric and fat tailed behavior. Across the 14 equity indices, the skew and shape parameters were significant and close to 0.9 and 7 respectively, indicating skewness of around -0.3 and excess kurtosis of approximately 2. The second stage copula MVT had a significant shape parameter of 15 indicating that the majority of the excess kurtosis had been captured in the first stage leaving an excess of around 0.5 for the joint tails. From this model, 1000 scenarios of size 1000×14 were simulated, and estimated using an AR(2)-scalar BEKK ($sBEKK$) and an AR(2)-GARCH(1,1)-DCC model both with MVN conditional distribution. The use of a diagonal MGARCH model was deemed impractical in this application.[36] For each of the estimated scenarios, the weighted standardized residuals were estimated making use of the residuals and conditional covariance matrix of the MGARCH models, using 100 randomly generated sets of weights. These model based weighted standardized residual series were then transformed into U(0,1) series by applying the PIT transformation for use in the test of Hong and Li (2005). Table 2.6 reports the mean and median statistic for the test as well as the percent rejections under the null of a correctly specified model at the 95% confidence level. To complete the exercise, the table also reports the results with DGP an AR(2)-sBEKK model and AR(2)-GARCH(1,1)-DCC model with MVN conditional distribution. The reason I report both the mean and median is because in the case of the DCC-copula model, the distribution of the statistic across the different scenarios and randomly generated weights is heavily skewed, particularly in the case of the sBEKK estimated model. From the percent rejection results it is quite evident that the DCC model is more robust to misspecification, and this is consistently shown in all the moment conditions in columns 1-4 in the table. In particular, the $M(4,4)$ column, which reports the misspecification for the conditional fourth moment, seems to fare considerably better for the DCC model. Since it is known that univariate GARCH models generate excess kurtosis[37], a first stage GARCH filter of the data may therefore provide for more robustness since it allows the filtering of each individual series which

[35] This was not included in the misspecification tests because of the nonlinear copula transformation involved which makes the determination of the analytical weighted moments, required by the tests, impossible.

[36] At a calculated time of 2 hours per estimated scenario for the diagonal model, 1000 estimations for 2 models would have required a total of 167 days under the available resources.

[37] The kurtosis (κ) implied by the GARCH(1,1) process with parameters α and β is:

$$\kappa = \frac{3\left(1 - P^2\right)}{\left(1 - P^2 - 2\alpha^2\right)} > 3 \tag{2.47}$$

is not possible with the BEKK dynamics (at least in the scalar case). Thus, in the presence of skewed and fat-tailed data, even though the DCC model is still rejected 38% of the time, it is still much lower than the 50% rejection of the sBEKK model, in this application. When it comes to using the sBEKK as DGP, the DCC model is well within the expected 5% rejection zone for the 95% confidence level used, with an overall rejection of 7.6% and even lower for the individual conditional moment conditions. However, when it comes to using a DCC model as DGP, the sBEKK model seems to have a higher rejection rate of about 12%. While more research is needed in this area, particularly with respect to more highly parameterized models such as the diagonal variants, it would certainly appear that some of the criticism targeted at the DCC model is not particularly justified.

2.5 Conclusion

The importance of capturing the joint dynamics and features of securities is of paramount importance in portfolio construction, risk management and in trading strategies such as pairs and correlation trading. Multivariate dependence as captured by MGARCH models is severely constrained by dimensionality issues and the availability of flexible distributions with the desired features. Complex dynamics are difficult to model for anything more than a few assets, and tradeoffs such covariance targeting introduce a different set of challenges in the estimation. This chapter reviewed the 2 most popular MGARCH models, the BEKK and DCC, with an in-sample relative comparison using a range of popular distributions. While the BEKK model can be used with a wide choice of conditional distributions, the 2-stage modelling process used by the DCC model has the advantage of a first pass univariate GARCH filter which, from the evidence presented, is quite robust to misspecification. A Monte Carlo application shows that even in the presence of non-normally distributed returns, the cost of misspecification is less with the DCC than the BEKK model. Criticisms leveled against the DCC models, such as lack of published theoretical proofs and inconsistency in model specification do not appear to be backed by the empirical evidence. This chapter also considered the extension of the DCC model to include more complex correlation dynamics in the form of the AGDCC model. Unfortunately, this extension has come at the cost of increasing

where P is the persistence of the process and equal to $\alpha + \beta$, under the assumption of the existence of the stationary fourth moment i.e. $\left(\beta^2 + 2\alpha\beta + 3\alpha^2\right) < 1$.

the dimensionality of the problem and, more importantly, the use of covariance targeting in the presence of such complex nonlinear dynamics introduces non-smoothness to the typical optimization setup increasing the complexity in solving such problems with confidence. A more promising extension has come in the form of the dynamic copula model which increases the flexibility of the types of distributions used in the first stage GARCH estimation, and the possibility of using a richer multivariate distribution in the second stage. The drawback of the use of this model is the absence of certain closed form solutions for the weighted margins, but out-of-sample this can be ameliorated by the use of simulation methods. Nevertheless, security dynamics contain much more complex features than can currently be accommodated in a feasible setup by typical MGARCH models, evidence of which was presented in the empirical application where none of the models appeared to adequately capture the dynamics of a typical international equity index dataset. With the advent of high speed trading and advances in computational power, the importance of fast and realistic estimation of the dynamics of large dimensional systems has become a key requirement in modern applied finance. The only framework in which parallel computation can be fully realized for large dimensional systems is that of independence. In the next chapter, I show how the independent factor framework can be used to jointly model time varying higher moments, capturing a much richer set of features in the underlying data and allowing for very fast estimation of large dimensional systems.

Chapter 3

Multivariate ACD Dynamics and Independence

The multivariate GARCH models covered in the previous chapter quickly become computationally infeasible for large dimensional problems. In addition, modelling of higher moments in a time varying context is almost impossible or at least completely impractical in a multivariate dependence setting mainly because of the difficulty to parameterize marginal and joint distributional parameters. Attempts to capture such higher moment dynamics can be found for instance in Jondeau and Rockinger (2009) who provide for an asymmetric DCC[1] Skew-Student model with time varying higher moment dynamics. Because of the presence of the skew and shape nuisance parameters in the conditional likelihood, the estimation has to be carried out jointly which makes this model infeasible for anything but a few assets.[2] Yet the importance of including time varying higher moments within the GARCH framework, particularly for risk management, is paramount and was covered extensively in Chapter 1, albeit in a univariate context. A feasible multivariate extension has neither been attempted nor is there any comparative empirical evidence as to its performance in a risk or portfolio management application. As in the univariate ACD literature, the few authors which have attempted to model time variation in higher moment in the multivariate domain have not provided for any out-of-sample empirical applications, a sure indication of the dimensionality problem

[1] The authors also argue that an asymmetric BEKK is just as feasible.

[2] The added complexity of the nonlinear transformations required in constraining the higher moment dynamics within their specific bounds does also makes this that much harder in a joint dependence framework.

faced in the typical joint dynamics setup. While an out-of-sample application on a specific period and dataset might not provide for conclusive evidence in favor of a model, when it can be feasibly and confidently estimated, then its uptake by practitioners and researchers will generate enough of a body of empirical evidence over time to gauge its overall value. The typical MGARCH models discussed in Chapter 2, or extensions to those such as proposed by Jondeau and Rockinger (2009) in terms of higher moment dynamics, suffer from dimensionality issues and are therefore unlikely to be very useful in practical risk and portfolio management applications.

In this chapter, I make use of a statistical Independent Factor GARCH framework to create the only truly feasible multivariate time varying higher moment model. Building on the Generalized Orthogonal GARCH (*GO-GARCH*) framework of van der Weide (2002), the Independent Components Analysis methodology of Hyvärinen and Oja (2000), and making use of an affine representation of the multivariate Generalized Hyperbolic distribution proposed by Schmidt, Hrycej, and Stützle (2006), it is possible to reduce the problem of joint time varying higher moment dynamics to one of effectively univariate estimation. Unlike other models, independence offers a greater deal of flexibility in modelling the full marginal dynamics within a multivariate affine factor framework, providing closed form higher co-moments and a semi-analytic representation for the weighted conditional portfolio density, invaluable in risk and portfolio management applications. I provide a comprehensive empirical application using two different datasets, at different frequencies, and covering a very large out of sample period, to show the value of such dynamics in a multivariate setting. The chapter is organized as follows: Section 3.1 introduces the Independent Factor ACD (*IFACD*) model and its motivation. Key features such as the conditional higher co-moment tensors are presented in Section 3.2, portfolio conditional density representation is detailed in Section 3.3 and estimation of the model and the ICA algorithm are discussed in Section 3.4. To evaluate the performance of the proposed model, I carry out an empirical study of VaR performance and dynamic moment based portfolio allocation in Section 3.5, using the log returns on 14 MSCI tradeable country indices used in the previous chapters, and also investigate a dynamic scenario based portfolio application using an alternate dataset of the weekly log total returns of the Dow Jones Industrial Average (*DJIA*) index constituents. Section 3.6 concludes.

3.1 The Independent Factor Model

Factor ARCH models, originally introduced by Engle, Ng, and Rothschild (1990) and with foundations in the Arbitrage Pricing Theory of Ross (1976), are based on the assumption that returns are generated by a set of unobserved underlying factors that are conditionally heteroscedastic, while the dependence framework is non-dynamic as a consequence of large scale estimation in a multivariate setting. The dependence structure of the unobserved factors then determines the type of factor model it belongs to, with correlated factors making up the F-ARCH type models, while uncorrelated and independent factors comprise the Orthogonal and Generalized Orthogonal Models respectively.[3] Because one can always re-discover uncorrelated or independent sources by certain statistical transformations, the correlated factor assumption of F-ARCH models does appear to be restrictive. GO-GARCH models on the other hand make use of those transformations to place the factors in an independence framework with unique benefits such as separability and the fast convolution of the weighted density giving rise to truly large scale, real-time and feasible estimation. Consider a set of N assets whose returns $\boldsymbol{r}_t$ are observed for T periods, with conditional mean $E[\boldsymbol{r}_t|\mathfrak{F}_{t-1}] = \boldsymbol{m}_t$, where $\mathfrak{F}_{t-1}$ is the σ-field generated by the past realizations of $\boldsymbol{r}_t$, i.e. $\mathfrak{F}_{t-1} = \sigma(\boldsymbol{r}_{t-1}, \boldsymbol{r}_{t-2}, \ldots)$. The GO-GARCH Model of van der Weide (2002) maps $\boldsymbol{r}_t - \boldsymbol{m}_t$ onto a set of unobserved independent factors $\boldsymbol{f}_t$ (or "structural errors"),

$$\boldsymbol{r}_t = \boldsymbol{m}_t + \boldsymbol{\epsilon}_t \quad t = 1, \ldots, T \tag{3.2}$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{A}\boldsymbol{f}_t, \tag{3.3}$$

where $\boldsymbol{A}$ is invertible and constant over time and may be decomposed into the de-whitening matrix $\boldsymbol{\Sigma}^{1/2}$, representing the square root of the unconditional covariance, and an orthogonal matrix, $\boldsymbol{U}$, so that:

$$\boldsymbol{A} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{U}, \tag{3.4}$$

and $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{Nt})'$. The rows of the mixing matrix $\boldsymbol{A}$ therefore represent the independent source factor weights assigned to each asset (i.e. rows are the assets and

[3] It should be noted, that most of these factor models may be seen as special cases of the BEKK model. The GO-GARCH model has the following restricted BEKK representation:

$$H_t = C + \sum_{i=1}^{m} A_i x_{t-1} {x'}_{t-1} {A'}_i + B H_{t-1} B'. \tag{3.1}$$

Under the assumption that all A_i and B are restricted to have the same eigenvector Z, with the eigenvalues of A being all zero except the i^{th} one, and the C can be decomposed into ZDZ' where D is some positive definite diagonal matrix, then this is a GO-GARCH (with GARCH(1,1) univariate dynamics) model where Z is the linear ICA map. However, GO-GARCH model is not limited to GARCH(1,1) or any particular process for the factors.

columns the factors). In the IFACD model I assume that the factors have the following specification:

$$\boldsymbol{f}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \tag{3.5}$$

where $\boldsymbol{H}_t = E[\boldsymbol{f}_t\boldsymbol{f}_t'|\mathfrak{F}_{t-1}]$ is a diagonal matrix with elements $(h_{1t}, \ldots, h_{Nt})$ which are the conditional variances of the factors, and $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{Nt})'$. The random variable z_{it} is independent of z_{jt-s} $\forall j \neq i$ and $\forall s$, with $E[z_{it}|\mathfrak{F}_{t-1}] = 0$ and $E[z_{it}^2] = 1$, this implies that $E[\boldsymbol{f}_t|\mathfrak{F}_{t-1}] = \boldsymbol{0}$ and $E[\boldsymbol{\epsilon}_t|\mathfrak{F}_{t-1}] = \boldsymbol{0}$. The factor conditional variances, $h_{i,t}$, can be modelled as a GARCH-type process. The unconditional distribution of the factors is characterized by:

$$E[\boldsymbol{f}_t] = \boldsymbol{0} \quad E[\boldsymbol{f}_t\boldsymbol{f}_t'] = \boldsymbol{I}_N \tag{3.6}$$

which, in turn, implies that:

$$E[\boldsymbol{\epsilon}_t] = \boldsymbol{0} \quad E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \boldsymbol{A}\boldsymbol{A}'. \tag{3.7}$$

It follows that the returns can be expressed as:

$$\boldsymbol{r}_t = \boldsymbol{m}_t + \boldsymbol{A}\boldsymbol{H}_t^{1/2}\boldsymbol{z}_t. \tag{3.8}$$

The conditional covariance matrix, $\boldsymbol{\Sigma}_t \equiv E[(\boldsymbol{r}_t - \boldsymbol{m}_t)(\boldsymbol{r}_t - \boldsymbol{m}_t)'|\mathfrak{F}_{t-1}]$ of the returns is given by:

$$\boldsymbol{\Sigma}_t = \boldsymbol{A}\boldsymbol{H}_t\boldsymbol{A}'. \tag{3.9}$$

The Orthogonal Factor model of Alexander (2001)[4] which uses only information in the covariance matrix, leads to uncorrelated components but not necessarily independent unless assuming a multivariate normal distribution. However, while whitening is not sufficient for independence, it is nevertheless an important step in the preprocessing of the data in the search for independent factors, since by exhausting the second order information contained in the covariance matrix it makes it easier to infer higher order information, reducing the problem to one of rotation (orthogonalization). The original procedure of van der Weide (2002) used a 1-step maximum likelihood approach to jointly estimate the rotation matrix and dynamics making the procedure infeasible for anything other than a few assets. Alternative approaches such as nonlinear least squares and method of moments for the estimation of $\boldsymbol{U}$ have been proposed in van der

[4] When $\boldsymbol{U}$ is restricted to be an identity matrix, the model reduces to the Orthogonal Factor model.

Weide (2004) and Boswijk and van der Weide (2011), respectively. In this model, I estimate the matrix $\boldsymbol{U}$ by ICA as in Broda and Paolella (2009) and Zhang and Chan (2009). One of the computational advantages offered by the Generalized Orthogonal approach is that following the estimation of the independent factors, the dynamics of the marginal density parameters of those factors may be estimated separately. In this context, I propose to extend the dynamics to the full conditional density parameters to model in a multivariate setting time varying higher moments. This builds on the CHICAGO[5] model of Broda and Paolella (2009) where the factor dynamics have non-time varying higher moments within the GH distribution, where the latter only differs from the GHICA[6] model of Chen, Härdle, and Spokoiny (2010) by using GARCH type dynamics for the conditional variance rather than a local exponential smoothing technique. While any multivariate distribution, admitting an affine representation may be used in this setup, the GH distribution, introduced by Barndorff-Nielsen (1977) and representing the case in the mean-variance mixture family given in Equation (2.35) where $W \sim GIG(\lambda, \delta^2, \alpha^2 - \beta^2)$, was chosen for its flexibility and rich parametrization, and already discussed in Chapters 1 and 2.

3.1.1 Conditional factor dynamics

Although models from the GARCH family are able under certain assumptions and parameterizations to produce thick-tailed and skewed unconditional distributions they typically assume that the shape and skewness parameters are time invariant. This also leads to the assumption that the conditional distribution of the standardized innovations ($\boldsymbol{z}_t$) is independent of the conditioning information, for which there is no good reason to believe so a-priori. There is an extensive empirical literature which has investigated time variation in the full conditional density parameters, reviewed and discussed in Chapter 1. In the IFACD model I assume that the centered and scaled random variables z_{it} are conditionally distributed as standardized GH, i.e., $GH(z_{it}; \lambda_i, \mu_i, \delta_i, \alpha_i, \beta_i)$. As in the ACD models presented in Chapter 1, I assume separate dynamics for the skew and shape parameters of the standardized GH distribution in the (ρ, ζ) parametrization. Similar to Jondeau and Rockinger (2003), alternative specifications were explored for the ACD parameter dynamics. Based on the bounds and hence resulting shape of the moment dynamics, I selected piecewise linear dynamics for the shape parameter $(\breve{\zeta}_{i,t})$

[5] Conditionally Heteroscedastic Independent Components Analysis GO-GARCH model.
[6] Generalized Hyperbolic Independent Components Analysis.

and quadratic dynamics for the skew parameter ($\breve{\rho}_{i,t}$)

$$\begin{aligned} \breve{\rho}_{it} &= \chi_{0i} + \chi_{1i} z_{it-1} + \chi_{2i} z_{it-1}^2 + \xi_{1i} \breve{\rho}_{it-1} \\ \breve{\zeta}_{it} &= \kappa_{0i} + \kappa_{1i} z_{it-1} \mathbf{1}_{[z_{it-1} < -1]} + \kappa_{2i} z_{t-1} \mathbf{1}_{[z_{it-1} > 1]} + \psi_{1i} \breve{\zeta}_{it-1}, \end{aligned} \tag{3.10}$$

where $\mathbf{1}$ is the indicator function such that positive (negative) standardized innovations, larger (smaller) than one standard deviation, have a different impact on the shape dynamics. This is in the spirit of the Threshold Autoregressive Model of Tong and Lim (1980), albeit I do not estimate the threshold but impose it a-priori at 1 standard deviation. The intuition from experiments with these models is that most of the variation is already captured by the conditional standard deviation, up to at least 1 standard deviation events, so that shocks beyond this are more likely to be relevant lest one introduces too much noise in the presence of over-parameterized dynamics. Thus the threshold acts as a sensitivity barrier to noise which is already well modelled by variation in the second moment GARCH dynamics. The logistic transform is then used to map the unconstrained processes $\breve{\rho}_t$ and $\breve{\zeta}_t$ into $\rho_{i,t}$ and $\zeta_{i,t}$:

$$\rho_{it} = -0.99 + \frac{1.98}{1 + e^{-\breve{\rho}_{it}}} \tag{3.11}$$

$$\zeta_{it} = 0.1 + \frac{24.9}{1 + e^{-\breve{\zeta}_{it}}} \tag{3.12}$$

where the bounds of the distributional parameters are $[-0.99, 0.99]$ and $[0.1, 25]$ for ρ and ζ, respectively. I limit the upper bound of ζ to 25 for estimation ease, since values beyond this point lead to very little change in the skewness and kurtosis, with the range 0.1 to 25 representing most of the distribution. In theory, the GIG shape parameter λ_i is allowed to vary for each factor, but as argued in Chapter 1, this introduces an added layer of complexity because of certain identification issues. The NIG distribution, with a value of λ equal to -0.5, which results in a very tractable sub-family of the GH, provides for a very rich modelling environment for financial time series, and without much loss of generality I adopt this as the main distribution in the empirical exercise.[7] Eventually, the single factors, $f_{it}, i = 1, \ldots, N$, are conditionally distributed as a $GH(f_{it}; \lambda_i, \mu_{it}\sqrt{h_{it}}, \delta_{it}\sqrt{h_{it}}, \alpha_{it}/\sqrt{h_{it}}, \beta_{it}/\sqrt{h_{it}})$. The relation between the parameters in the $(\alpha, \beta, \delta, \mu)$ parametrization was already covered in Equation (1.19). Finally,

[7]Though I do present results for standardization and portfolio density in both the NIG and more general GH cases.

the vector of returns $\boldsymbol{r}_t$, which can be expressed as a linear transformation of independent factors $\boldsymbol{f}_t \in \mathbb{R}^N$, turns out to be conditionally distributed according to the maGH distribution:

$$\boldsymbol{r}_t|\mathfrak{F}_{t-1} \sim maGH_N(\boldsymbol{m}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\omega}_t), \tag{3.13}$$

where $\boldsymbol{\omega}_t = (\omega_{1t}, \ldots, \omega_{Nt})'$ and $\omega_{it} = (\lambda_i, \alpha_{it}, \beta_{it})'$, representing the conditional shape and skew parameter vectors. The extension of dynamics to all the parameters of the distribution presents the opportunity to go beyond the conditional time-varying covariance matrix, to higher co-moment tensors, the details and importance of which are discussed in the next section.

3.2 Conditional Co-Moments

It seems to be a well-established stylized fact that the unconditional security return distribution is not normal and the mean and variance of returns alone are insufficient to characterize the return distribution completely. This has led researchers to pay attention to the third moment - skewness - and the fourth moment - kurtosis. The validity of the CAPM in the presence of higher-order co-moments and their effects on asset prices has been investigated. The simple, single-factor, CAPM only holds under very specific conditions. When asset prices are non-normal and investors have non-quadratic preferences, then they will care about all return moments and not only mean and variance, as in the standard CAPM. There are a number of extensions to the basic two-moments CAPM which predict a linear relationship in which terms like co-skewness and co-kurtosis are priced. For example, Kraus and Litzenberger (1976), Sears and Wei (1985) extended the CAPM to incorporate skewness in asset valuation models and provided mixed results. A few studies have shown that non-diversified skewness and kurtosis play an important role in determining security valuations. Fang and Tsong-Yue (1997), derived a four-moment CAPM where it was shown that systematic variance, systematic skewness and systematic kurtosis contribute to the risk premium of an asset. Harvey and Siddique (2000) examined an extended CAPM, including systematic co-skewness, reporting that conditional skewness explains the cross-sectional variation of expected returns across assets and is significant even when factors based on size and book-to-market are included. As skewness of a portfolio matters to investors, an asset's contribution to the skewness of a broadly diversified portfolio, referred to as "co-skewness" with the portfolio, may also be rewarded. Skewness preference further suggests that the representative investor may adjust his diversified portfolio such that an individual security's contribution to the skewness of the market portfolio may become a component of the security's expected returns. Mathematically, as demonstrated in Conine and Tamarkin (1981), both individual assets' skewness and co-skewness between assets contribute to the skewness of the portfolio which is composed of these assets. Intuitively, as positive (negative) skewness implies a probability of obtaining a large positive (negative) return (relative to a benchmark such as the normal distribution), a positive co-skewness of an asset with another asset means that, when the price volatility goes up the return of this asset also goes up. The general acceptance that the conditional density of asset returns is not completely and adequately characterized by the first two moments, implies that the derivation of any measure of risk from that density requires estimates for the higher order co-moments of the return distribution if one is work within a multivariate setting. The linear affine representation of the IFACD model allows to identify closed-form expression for the conditional co-skewness and co-kurtosis of asset returns[8], as described in de Athayde and Flôres Jr (2000). The novelty of the IFACD model is that the third and fourth factor co-moment matrices are now time-varying, as a consequence of the ACD specification of the conditional density of the standardized innovations, e.g., z_{it}. The conditional co-moments of $\boldsymbol{r}_t$ of order 3 and 4 are represented as tensor matrices,

$$\begin{aligned} \boldsymbol{M}_t^3 &= \boldsymbol{A}\boldsymbol{M}_{f,t}^3(\boldsymbol{A}' \otimes \boldsymbol{A}'), \\ \boldsymbol{M}_t^4 &= \boldsymbol{A}\boldsymbol{M}_{f,t}^4(\boldsymbol{A}' \otimes \boldsymbol{A}' \otimes \boldsymbol{A}'), \end{aligned} \tag{3.14}$$

where $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$ are the $(N \times N^2)$ conditional third co-moment matrix and the $(N \times N^3)$ conditional fourth co-moment matrix of the factors, respectively. $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$, defined as are given by

$$\boldsymbol{M}_{f,t}^3 = \left[\boldsymbol{M}_{1,f,t}^3, \boldsymbol{M}_{2,f,t}^3, \ldots, \boldsymbol{M}_{N,f,t}^3\right] \tag{3.15}$$

$$\boldsymbol{M}_{f,t}^4 = \left[\boldsymbol{M}_{11,f,t}^4, \boldsymbol{M}_{12,f,t}^4, \ldots, \boldsymbol{M}_{1N,f,t}^4 | \ldots | \boldsymbol{M}_{N1,f,t}^4, \boldsymbol{M}_{N2,f,t}^4, \ldots, \boldsymbol{M}_{NN,f,t}^4\right] \tag{3.16}$$

where $\boldsymbol{M}_{k,f,t}^3, k = 1, \ldots, N$ and $\boldsymbol{M}_{kl,f,t}^4, k, l = 1, \ldots, N$ are the $(N \times N)$ submatrices of $\boldsymbol{M}_{f,t}^3$ and $\boldsymbol{M}_{f,t}^4$, respectively, with elements

$$\begin{aligned} m_{ijk,f,t}^3 &= E[f_{i,t}f_{j,t}f_{k,t}|\mathfrak{F}_{t-1}] \\ m_{ijkl,f,t}^4 &= E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}]. \end{aligned}$$

[8] It is possible to go beyond these moments but the notation becomes cumbersome and the benefits likely to be marginal.

Since the factors f_{it} can be decomposed as $z_{it}\sqrt{h_{it}}$, and given the assumptions on z_{it}, then $E[f_{i,t}f_{j,t}f_{k,t}|\mathfrak{F}_{t-1}] = 0$. It is also true that for $i \neq j \neq k \neq l E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}] = 0$ and when $i = j$ and $k = l$,

$$E[f_{i,t}f_{j,t}f_{k,t}f_{l,t}|\mathfrak{F}_{t-1}] = h_{it}^2 h_{kt}^2.$$

Thus, under the assumption of mutual independence, all elements in the conditional co-moments matrices with at least 3 different indices are zero. Finally, standardizing the conditional co-moments one obtains conditional co-skewness and co-kurtosis of $\boldsymbol{r}_t$,

$$\begin{aligned} \boldsymbol{S}_{ijk,t} &= \frac{m^3_{ijk,t}}{(\sigma_{i,t}\sigma_{j,t}\sigma_{k,t})}, \\ \boldsymbol{K}_{ijkl,t} &= \frac{m^4_{ijkl,t}}{(\sigma_{i,t}\sigma_{j,t}\sigma_{k,t}\sigma_{l,t})}, \end{aligned} \tag{3.17}$$

where $\boldsymbol{S}_{ijk,t}$ represents the asset co-skewness between elements i, j, k of $\boldsymbol{r}_t$, $\sigma_{i,t}$ the standard deviation of $\boldsymbol{r}_{i,t}$, and in the case of $i = j = k$ represents the skewness of asset i at time t, and similarly for the co-kurtosis tensor $\boldsymbol{K}_{ijkl,t}$. Two natural applications of return co-moments matrices are Taylor type utility expansions in portfolio allocation and higher moment news impact surfaces, applications of which are featured in Sections 3.5.4.1 and 3.5.2 respectively.

3.3 The Portfolio Conditional Density

An important question that can be addressed in this framework is the determination of the portfolio conditional density, an issue of vital importance in risk management applications. The N-dimensional NIG distribution, closed under convolution, is suited to problems in portfolio and risk management where a weighted sum of assets is considered. However, when the distributional parameters α and β, representing skew and shape, are allowed to vary, as in the IFACD case, this property no longer holds and numerical methods such as that of the Fast Fourier Transform (FFT) are needed to derive the weighted density by inversion of the characteristic function of the scaled parameters[9]. In the case of the NIG distribution, this is greatly simplified because of the representation of the modified Bessel function for the GIG shape index (λ) with

[9]This effectively means that the weighted density is not necessarily NIG distributed.

value -0.5 which was derived in Barndorff-Nielsen and Bläsild (1981), otherwise the characteristic function of the GH involves the evaluation of the modified Bessel function with complex arguments, which is considerably slower. Appendix A derives the characteristic functions used in the case of independent margins for both the NIG and full GH distributions. Let R_t be the portfolio return:

$$R_t = \boldsymbol{w}_t' \boldsymbol{r}_t = \boldsymbol{w}_t' \boldsymbol{m}_t + (\boldsymbol{w}_t' \boldsymbol{A} \boldsymbol{H}_t^{1/2}) \boldsymbol{z}_t \tag{3.18}$$

where $\boldsymbol{H}_t^{1/2}$ is estimated from the ACD dynamics of $\boldsymbol{y}_t$. The IFACD model allows to express the portfolio variance, skewness and kurtosis in closed form,

$$\begin{aligned} \sigma_{p,t}^2 &= \boldsymbol{w}_t' \boldsymbol{\Sigma}_t \boldsymbol{w}_t, \\ s_{p,t} &= \frac{\boldsymbol{w}_t' \boldsymbol{M}_t^3 (\boldsymbol{w}_t \otimes \boldsymbol{w}_t)}{(\boldsymbol{w}_t' \boldsymbol{\Sigma}_t \boldsymbol{w}_t)^{3/2}}, \\ k_{p,t} &= \frac{\boldsymbol{w}_t' \boldsymbol{M}_t^4 (\boldsymbol{w}_t \otimes \boldsymbol{w}_t \otimes \boldsymbol{w}_t)}{(\boldsymbol{w}_t' \boldsymbol{\Sigma}_t \boldsymbol{w}_t)^2}, \end{aligned} \tag{3.19}$$

where $\boldsymbol{\Sigma}_t$, $\boldsymbol{M}_t^3$ and $\boldsymbol{M}_t^4$ are derived in (3.14). The portfolio conditional density may be obtained via the inversion of the characteristic function through the FFT method as in Chen, Härdle, and Spokoiny (2007) (see Appendix A for details) or by simulation. I choose the former for its accuracy and speed. Provided that $\boldsymbol{z}_t$ is a N-dimensional vector of innovations, marginally distributed as 1-dimensional standardized GH, the density of the weighted asset return, $w_{it} r_{it}$, is

$$w_{i,t} r_{i,t} = (w_{i,t} m_{i,t} + \overline{w}_{i,t} z_{i,t}) \sim GH_{\lambda_i} \left(\overline{w}_{i,t} \mu_{i,t} + w_{i,t} m_{i,t}, |\overline{w}_{i,t}| \, \delta_{i,t}, \frac{\alpha_{i,t}}{|\overline{w}_{i,t}|}, \frac{\beta_{i,t}}{|\overline{w}_{i,t}|} \right) \tag{3.20}$$

where $\overline{\boldsymbol{w}}_t'$ is equal to $\boldsymbol{w}_t' \boldsymbol{A} \boldsymbol{H}_t^{1/2}$, and $\overline{w}_{i,t}$ is the i-th element of $\overline{\boldsymbol{w}}_t$, $m_{i,t}$ the conditional mean of the i-th underlying asset. In order to obtain the density of the portfolio, we must sum the individual weighted densities of $z_{i,t}$. The characteristic function of the portfolio return R_t is

$$\varphi_R(u) = \prod_{i=1}^{n} \varphi_{\bar{w} Z_i}(u) = \exp\left(\mathrm{i}u \sum_{j=1}^{d} \bar{\mu}_j + \sum_{j=1}^{d} \left(\frac{\lambda_j}{2} \log\left(\frac{\gamma}{\upsilon}\right) + \log\left(\frac{\boldsymbol{K}_{\lambda_j}(\bar{\delta}_j \sqrt{\upsilon})}{\boldsymbol{K}_{\lambda_j}(\bar{\delta}_j \sqrt{\gamma})} \right) \right) \right) \tag{3.21}$$

where, $\gamma = \bar{\alpha}_j^2 - \bar{\beta}_j^2$, $\upsilon = \bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2$, and $(\bar{\alpha}_j, \bar{\beta}_j, \bar{\delta}_j, \bar{\mu}_j)$ are the scaled versions of the parameters $(\alpha, \beta_i, \delta_i, \mu_i)$ as shown in (3.20). The density may be accurately approximated by FFT as follows,

$$f_R(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{(-\mathrm{i}ur)} \varphi_R(u) \mathrm{d}u \approx \frac{1}{2\pi} \int_{-s}^{s} e^{(-\mathrm{i}ur)} \varphi_R(u) \mathrm{d}u. \tag{3.22}$$

Expression (3.22) is the base for the calculation of VaR in the empirical application reported in Section 3.5.

3.4 Estimation

The estimation of the factor loading matrix $\boldsymbol{A}$ exploits the decomposition in (3.4). The estimation of $\boldsymbol{\Sigma}^{1/2}$, representing the square root of the unconditional covariance matrix, is usually obtained from the OLS residuals $\widehat{\boldsymbol{\epsilon}}_t = \boldsymbol{r}_t - \widehat{\boldsymbol{m}}_t$, while the orthogonal matrix $\boldsymbol{U}$ can be estimated using ICA (see Broda and Paolella (2009), Zhang and Chan (2009)). ICA is a computational method for separating multivariate mixed signals, $\boldsymbol{x} = [x_1, ..., x_n]'$, into additive statistically independent and non-Gaussian components, $\boldsymbol{s} = [s_1, ..., s_n]'$, such that $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{s}$. The objective is to decompose the observed $\boldsymbol{x} = [x_1, ..., x_n]'$, into independent factors $\boldsymbol{s} = [s_1, ..., s_n]'$ and a linear matrix $\boldsymbol{B}$, such that $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{s}$. The independent source vector $\boldsymbol{s} \in \mathbb{R}^n$, is assumed to be sampled from a joint distribution $f(\boldsymbol{s})$,

$$f(s_1, ..., s_n) = f(s_1)f(s_2)...f(s_n), \tag{3.23}$$

where $\boldsymbol{s}$ is not directly observable, nor is the particular form of the individual distributions, $f(s_i)$, usually known.[10] This forms the key property of independence, namely that the joint density of independent signals is simply the product of their marginals. The estimate of the linear mixing matrix $\boldsymbol{B}$ can be obtained via estimation methods based on a choice of criteria for measuring independence which include the maximization of non-Gaussianity through measures such as kurtosis and negentropy, minimization of mutual information, likelihood and infomax. This follows from the Central Limit Theorem which states that mixtures of independent variables tend to become more Gaussian in distribution when they are mixed linearly, hence maximizing non-Gaussianity leads to independent components (see Hyvärinen and Oja (2000) for more details).[11] Entropy may be thought of as the amount of information inherent within a random variable, being an increasing function of the amount of randomness in that variable. For a discrete

[10] If the distributions are known the problem reduces to a classical maximum likelihood parametric estimation.

[11] Estimation by minimization of the mutual information was first proposed by Comon (1994) who derived a fundamental connection between cumulants, negentropy and mutual information. The approximation of negentropy by cumulants was originally considered much earlier in Jones and Sibson (1987), while the connection between infomax and likelihood was shown in Pearlmutter and Parra (1997), and the connection between mutual information and likelihood was explicitly discussed in Cardoso (2000)

random variable X it is defined as,

$$H(X) = -\sum_i P(X = b_i) \log P(X = b_i), \tag{3.24}$$

with b_i denoting the possible values of X. In the continuous case, for a continuous random variable X with density $f_X(x)$, the entropy[12] H is defined as,

$$H(X) = -\int f_X(x) \log f_X(x) \mathrm{d}x. \tag{3.25}$$

A key result from information theory states that among all random variables of equal variance, a Gaussian variable has the largest entropy. Hence entropy could be used as a measure of non-Gaussianity. A related measure of non-Gaussianity is the negentropy which is always non-negative and zero for a Gaussian variable. It is defined as,

$$J(X) = H(X_{gauss}) - H(X), \tag{3.26}$$

where $H(X_{gauss})$ is the entropy of a Gaussian random variable having the same covariance matrix as X. As shown by Comon (1994), negentropy is invariant for invertible linear transformations and is an optimal estimator of non-Gaussianity with regards to its statistical properties (i.e. consistency, asymptotic variance and robustness). In practice, because we do not know the density, approximations of negentropy are used such as the one by Hyvärinen and Oja (2000),

$$J(X) \approx \sum_{i=1}^{p} k_i [E(G_i(X)) - E(G_i(V))]^2, \tag{3.27}$$

where k_i are positive constants, V is a standardized Gaussian variable and G_i are non-quadratic functions. The choice of the non-quadratic function has an impact on the robustness of the estimators of negentropy. with $G(x) = x^4$ (kurtosis based) being the least robust while more robust measures would include,

$$g_1(u) = \frac{1}{a_1} \log\cosh a_1 u, \qquad g_2(u) = -\exp(-0.5u^2). \tag{3.28}$$

Because these non-quadratic functions present a complex nonlinear optimization problem, sophisticated numerical algorithms are usually necessary. Two main algorithms are

[12]In the continuous case this is usually called *differential entropy.*

used, the online and batch methods, with the former based on stochastic gradient methods while in the latter case a popular choice is the natural gradient ascend of likelihood. The FastICA of Hyvärinen and Oja (2000) is a very efficient batch algorithm with a range of options for the non-quadratic functions. It can be used to estimate the components either one at a time by finding maximally non-Gaussian directions or in parallel by maximizing non-Gaussianity or the likelihood. The estimation procedure of the IFACD model can be summarized as follows. First, the FastICA is applied to the whitened data $\boldsymbol{z}_t = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\epsilon}}_t$, where $\widehat{\boldsymbol{\Sigma}}^{1/2}$ is obtained from the eigenvalue decomposition of the OLS residual covariance matrix, returning an estimate of $\boldsymbol{f}_t$, i.e., $\boldsymbol{y}_t = \boldsymbol{W}\boldsymbol{z}_t$. Second, because of the assumption of independence, the likelihood function of the IFACD model is greatly simplified so that the conditional log-likelihood function is expressed as the sum of the individual conditional log-likelihoods, derived from the conditional marginal densities of the factors, i.e., $GH_{\lambda_i}(y_{it}) \equiv GH(y_{it}; \lambda_i, \mu_{it}\sqrt{h_{it}}, \delta_{it}\sqrt{h_{it}}, \alpha_{it}/\sqrt{h_{it}}, \beta_{it}/\sqrt{h_{it}})$, plus a term for the mixing matrix $\boldsymbol{A}$, estimated in the first step by FastICA:

$$L(\widehat{\boldsymbol{\epsilon}}_t \,|\boldsymbol{\theta}, \boldsymbol{A}) = T \log \left|\boldsymbol{A}^{-1}\right| + \sum_{t=1}^{T}\sum_{i=1}^{N} \log\left(GH_{\lambda_i}(y_{it}|\theta_i)\right) \tag{3.29}$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters in the marginal densities. Because ICA is a linear noiseless model,[13] the implication for this 2 stage estimation in the IFACD model is that uncertainty plays no part in the derivation of the mixing matrix $\boldsymbol{A}$ and hence does not affect the standard errors of the independent factors.

The possibility of modelling the independent factors separately not only increases the flexibility of the model but also its computational feasibility, since the multivariate estimation reduces to N univariate optimization steps plus a term which depends on the factor loading matrix. Thus the independence property of the model allows the estimation of very large scale systems on modern computational grids[14] with the time required to calculate any n-dimensional model equivalent to the time it takes to estimate one single factor in the ACD framework.

[13]According to Hyvärinen and Oja (2000), this can be partially justified by the fact that most of the research on ICA has also concentrated on the noise free model and it has been shown with overwhelming empirical support across a number of different disciplines to be a very good approximation to a more complex model with noise added. Because the estimation of the noise-free model has proved to be a very difficult task in itself, the noise-free model may also be considered a tractable approximation of the more realistic noisy model.

[14]For the large scale out-of-sample backtesting carried out in this application, the model was estimated on the Amazon Elastic Cloud at a fraction of the time it would take to estimate either a DCC model or any other multivariate GARCH type model.

3.5 Empirical application

While the contribution of time varying higher moment dynamics is simple enough to observe and evaluate in univariate models, in the IFACD model each underlying asset is a weighted combination of the independent factors and the diverse dynamics they possess. Assessing the relative goodness of fit of such a multivariate model is a daunting task, not least because of the absence of relevant and feasible measures in this area. Therefore, and in keeping with the applied aspect of the IFACD model in risk and portfolio management, the evaluation of the model with respect to the risk application was performed on a linearly weighted forecast density representing a typical portfolio approach. To avoid any bias from using a particular weighting, a large number of randomly weighted portfolios were formed in each risk management application and the average statistic and percent rejection of each test reported. I continue to make use of the MSCI dataset described in Section 1.3 and used in the empirical exercises of Chapters 1 and 2. Additionally, a different dataset based on the weekly total log returns of the point in time constituents of the DJIA index, was also used in order to gauge the model's performance on a larger cross-section and a lower frequency (weekly). This is described and analyzed in more detail in Section 4.4.1. The next sections present the detailed results of the relative performance of the IFACD model with respect to the CHICAGO and other relevant MGARCH models, in both risk and portfolio management applications.[15]

3.5.1 Model Estimation and In-Sample Fit

The MSCI dataset of 14 international equity indices was found to be well fitted in a univariate context both in and out-of-sample by ACD models and ill fitted by the multivariate GARCH models in-sample. In order to obtain an overall, full sample, goodness of fit of the model, I repeat the empirical exercise of Section 2.4, using the GO-GARCH (MVN), CHICAGO and IFACD models with conditional distribution the maNIG. Table 3.1 reports parameter estimates and summary fit statistics of the IFACD and CHICAGO models for the period 12/08/1996 to 02/03/2011. As in the empirical application of Section 2.4, an AR(2) model was used to filter the data prior to applying the FastICA algorithm using the hyperbolic tangent (tanh) contrast function to separate

[15] See Appendix F for details on the software used to estimate these models.

the signals.[16] In both models, the factor variance dynamics were assumed to follow a GARCH(1,1) model with parameters $(\omega, \alpha_1, \beta_1)$, and for the skew and shape dynamics of the IFACD model, a first order quadratic and piece-wise linear model with parameters $(\chi_0, \chi_1, \chi_2, \xi_1)$ and $(\kappa_0, \kappa_1, \kappa_2, \psi_1)$ respectively was used, as in Equation (3.10).

[16] The nonlinearity contrast function tanh is optimal for a wide range of source distributions including supergaussian and subgaussian.

Table 3.1: IFACD vs CHICAGO: Parameter estimates and in-sample fit (14 MSCI iShares)

IFACD	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
ω	0.0095***	0.0089***	0.0037***	0.0055**	0.0069***	0.0028***	0.0028***	0.0043***	0.0029**	0.0043***	0.0012***	0.0020***	0.0026***	0.0021**
α_1	0.0842***	0.0642***	0.0404***	0.0499***	0.0820***	0.0618***	0.0498***	0.0349***	0.0535***	0.0740***	0.0289***	0.0357***	0.0441***	0.0479***
β_1	0.9094***	0.9277***	0.9554***	0.9456***	0.9138***	0.9363***	0.9479***	0.9608***	0.9441***	0.9252***	0.9697***	0.9623***	0.9539***	0.9520***
χ_0	0.4160***	0.1176***	-0.0793	0.0855	0.1035	-0.0278	-0.0796	0.1336	0.0264	0.0700*	0.0278	0.0123	-0.0430	-0.0026
χ_1	-0.1513***	0.2342***	0.0892	0.1810***	0.0745	-0.0148	0.0558	0.0217	-0.0142	0.0043	-0.0310	0.0016	0.1680*	0.1624***
χ_2	-0.0080	-0.0598***	0.0083	-0.0467	-0.0532*	0.0470*	0.0295	-0.0672	-0.0253	-0.0462*	0.0598**	-0.0512	0.0696**	-0.0046
ξ_1	0.2589	0.4874***	0.2685	0.0069	0.0407	0.6591	0.2727	0.0115	0.8872***	0.8603***	0.0231	0.0007	0.1402	0.8033***
κ_0	-0.3655	-1.5930***	-1.1093***	0.0160	-0.4839	-0.2515	-1.1488***	-0.7475	-1.1168***	-1.5398***	-0.4182	-1.6048***	0.0566	-0.9210***
κ_1	0.7783***	0.2958*	0.5646**	0.0410	0.0980	0.0119	-0.0264	-0.3710	-0.8221	-0.0002	0.2139**	-0.9412	-0.1734	0.2361
κ_2	0.1622	-0.1324	-0.3907	-0.1902	-0.3369***	-0.5411***	0.9861	0.3700	-0.2340	0.2290	0.9651	0.0177	-0.4026***	0.4300*
ψ_1	0.7042**	0.0064	0.0056	0.9707***	0.6292*	0.6401***	0.1185	0.4766	0.0002	0.0056	0.6826***	0.0057	0.9740***	0.1963
CHICAGO	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
ω	0.0082***	0.0073***	0.0037**	0.0055**	0.0065***	0.0029**	0.0031**	0.0041**	0.0030**	0.0041**	0.0015	0.0020**	0.0030**	0.0020**
α_1	0.0762***	0.0580***	0.0404***	0.0454***	0.0771***	0.0623***	0.0519***	0.0346***	0.0546***	0.0724***	0.0299***	0.0347***	0.0433***	0.0441***
β_1	0.9144***	0.9350***	0.9562***	0.9492***	0.9185***	0.9361***	0.9454***	0.9612***	0.9432***	0.9266***	0.9687***	0.9634***	0.9540***	0.9544***
ρ	0.2603***	0.0733**	-0.0515	0.0242	0.0203	0.0394	-0.0314	0.0356	-0.0049	0.0683*	0.0514	-0.0217	0.0141	-0.0078
ζ	2.9990***	2.2366***	2.6669***	3.2576***	2.4429***	3.3794***	3.8397***	3.7550***	3.6924***	2.7143***	3.7625***	2.8129***	6.2489***	4.2903***
$Persistence_{IFACD}[F_i]$	0.994	0.992	0.996	0.995	0.996	0.998	0.998	0.996	0.998	0.999	0.999	0.998	0.998	1.000
$Persistence_{CHICAGO}[F_i]$	0.991	0.993	0.997	0.995	0.996	0.998	0.997	0.996	0.998	0.999	0.999	0.998	0.997	0.999
$LL_{IFACD}[F_i]$	-4240.9	-4706.5	-4788.8	-4816.5	-4602.0	-4514.6	-4580.8	-4872.4	-4608.4	-4662.4	-4803.4	-4700.7	-4816.9	-4680.6
$LL_{CHICAGO}[F_i]$	-4255.5	-4716.8	-4793.7	-4822.4	-4607.9	-4518.4	-4583.9	-4874.4	-4611.3	-4665.6	-4809.1	-4704.6	-4826.8	-4686.5
$LR(stat)$	-29.4	-20.7	-9.8	-11.9	-11.7	-7.6	-6.1	-3.8	-5.7	-6.3	-11.5	-7.8	-19.7	-11.8
$LR(p-value)$	0.00	0.00	0.13	0.07	0.07	0.27	0.41	0.70	0.46	0.39	0.08	0.25	0.00	0.07
$LL_{IFACD}[Model]$	163119.6													
$LL_{CHICAGO}[Model]$	163037.7													

Note: The Table reports the parameter estimates under the IFACD and CHICAGO models under the NIG distribution, for the log returns of 14 MSCI indices from 12/08/1996 to 02/03/2011. The estimates are for the independent factors (F) arising from the ICA transformation of the data, which was first demeaned and filtered for first and second order autocorrelation using an AR(2) model. The *, ** and *** next to the parameters denote significance at the 10%, 5% and 1% levels respectively. The individual factor volatility persistence, and Log-Likelihood are reported as is the overall model Log-Likelihood for comparison between the two models. The Likelihood ratio (LR) statistic reports the difference between the IFACD and CHICAGO models under the assumption that the latter is a restricted version of the former, and distributed as χ^2 with 6 degrees of freedom representing the number of restrictions (time varying higher moment parameters excluding the intercept).

The conditional variance dynamics of the factors follow a GARCH(1,1) model: $h^2_{i,t} = \omega_j + \alpha_j F^2_{i,t-1} + \beta_j h^2_{i,t-1}$, $i = 1, \ldots, 14$. The conditional skew dynamics of the factors are bounded through a logistic transformation such that $\rho_{it} = -0.99 + \frac{1.98}{1+\exp -\tilde{\rho}_{it}}$, where the unconstrained parameters follow a first order quadratic model: $\tilde{\rho}_{it} = \chi_{0i} + \chi_{1i} z_{it-1} + \chi_{2i} z^2_{it-1} + \xi_{1i}\tilde{\rho}_{it-1}$. The conditional shape dynamics of the factors are bounded through a logistic transformation such that $\zeta_{it} = 0.1 + \frac{24.9}{1+\exp -\tilde{\zeta}_{it}}$, where the unconstrained parameters follow first order piecewise linear model: $\tilde{\zeta}_{it} = \kappa_{0i} + \kappa_{1i} z_{it-1} \mathbf{1}_{[z_{it-1}<-1]} + \kappa_{2i} z_{t-1} \mathbf{1}_{[z_{it-1}>1]} + \psi_{1i}\tilde{\zeta}_{it-1}$.

The results of this whole sample estimation suggest that there is some time variation in the shape parameter in about half of the independent factors, and somewhat less in the skew parameter. Compared to the CHICAGO model, and using an LR test under the alternative hypothesis that a restricted model (with the restriction being that of no time variation in the skew and shape parameters i.e. the CHICAGO model), it can be concluded that in half the factors the restricted model can be rejected at the 10% level of significance in favor of the IFACD model. In both models, one can also observe a very high persistence in the variance dynamics which likely indicates some structural break or shift which cannot be accounted for by the simple GARCH(1,1) model. This is less likely to be a problem in the out-of-sample application considered later where a smaller window size is considered in a rolling estimation and forecast setting. Across the 14 factors (F_1 to F_{14}), the average estimated excess kurtosis from the CHICAGO model is about 0.89, very close to what was observed in Section 2.4 for the BEKK (MVT) model, and the average skewness close to zero.[17] To compare the GO-GARCH (MVN), CHICAGO and IFACD models, representing models within the independence framework, with the MGARCH models of Chapter 2, I repeat the misspecification exercise of Section 2.4 using the test of Hong and Li (2005). Table 3.2 displays the average statistic and percent rejections of the test for the 3 models and for ease of comparison includes the results already presented in Table 2.5 for the various MGARCH models used in the previous chapter. While the GO-GARCH (MVN) model does no better than the BEKK (MVN), it is the CHICAGO and IFACD models which have the lowest overall, among all models, Portmanteau (W) statistic indicating a better fit to the data. Despite the AGDCC model having a lower conditional mean and variance ($M(1,1)$ and $M(2,2)$) statistic, for reasons already discussed, the flexibility of the maNIG distribution appear to provide for a lower overall cost of misspecification. However, within this large in-sample application it is not clear whether there are substantial marginal benefits to including time variation in higher moments, despite the slightly lower value for the Portmanteau statistic in the case of the IFACD model. This question is more readily addressed in the out-of-sample application in Sections 3.5.3 and 3.5.4.

[17] The skewness(S) and excess kurtosis(K_{ex}) of the NIG distribution using the $(\alpha, \beta, \delta, \mu)$ parametrization are:

$$\begin{aligned} S &= \frac{3\beta}{\alpha\sqrt{\delta\gamma}}, \\ K_{ex} &= \frac{3}{\delta\sqrt{\alpha^2-\beta^2}}\left(1+4\frac{\beta^2}{\alpha^2}\right). \end{aligned} \tag{3.30}$$

TABLE 3.2: Independence vs Dependence: Hong-Li misspecification test (14 MSCI iShares)

Hong-Li Non-Parametric Test					
	$M(1,1)$	$M(2,2)$	$M(3,3)$	$M(4,4)$	W
BEKK-MVN	43.7	15.9	5.8	1.8	59.8
	[100]	*[99.8]*	*[90.3]*	*[54.7]*	*[100]*
BEKK-MVT	43.1	16.3	6.5	2.5	40.3
	[100]	*[99.9]*	*[92.6]*	*[63.6]*	*[100]*
BEKK-MVL	44.3	19.2	9.1	4.4	40.1
	[100]	*[99.9]*	*[96.4]*	*[80.2]*	*[100]*
BEKK-MSL	43.1	18.8	9.2	4.7	32.4
	[100]	*[99.9]*	*[96.4]*	*[82]*	*[100]*
AGDCC-MVN	28.7	9.9	3.9	2.8	29.1
	[100]	*[99.1]*	*[93]*	*[91.6]*	*[100]*
AGDCC-MVL	28.2	11.8	5.6	3.5	48.6
	[100]	*[99.1]*	*[95.3]*	*[92.5]*	*[100]*
GOGARCH-MVN	43.3	17.4	6.1	1.6	44.6
	[100]	*[99.9]*	*[90.3]*	*[43.6]*	*[100]*
GOGARCH-NIG	39.7	15.0	5.1	1.3	23.8
	[100]	*[99.8]*	*[88.8]*	*[43.7]*	*[100]*
IFACD-NIG	38.8	15.2	5.6	1.7	20.3
	[100]	*[99.8]*	*[90.6]*	*[51]*	*[100]*

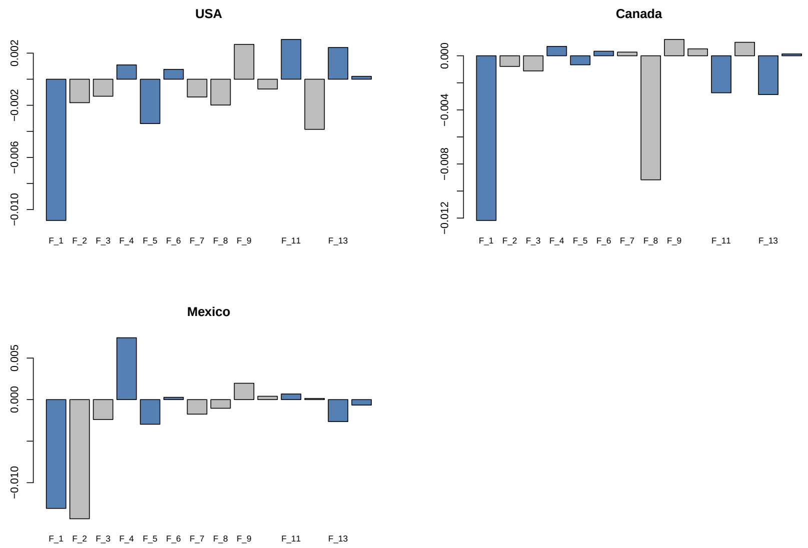
Note: The Table reports the average statistic and percent rejections from the non parametric density test of Hong and Li (2005) applied to the probability integral transformed weighted margins of the in-sample fit of 14 MSCI iShares for the period 12/08/1996 to 02/03/2011, from the diagonal BEKK and diagonal AGDCC under alternative conditional distributions representing models from a dependence based framework, and the GO-GARCH (MVN), CHICAGO (NIG) and IFACD (NIG) representing models from the statistical independence based framework. $M(j,j), j = 1,..,4$, represent the nonparametric tests for misspecification in the conditional moments, and distributed as $N(0,1)$ under the null of a correctly specified model. The statistic W in column 5 of the table is the Portmanteau type test statistic for general misspecification (using 4 lags) and distributed as $N(0,1)$ under the null of a correctly specified model. Values in square brackets are the percent rejections under the null, with 95% confidence, for the 5000 randomly weighted margins.

3.5.2 Co-Moment News Impact Surface

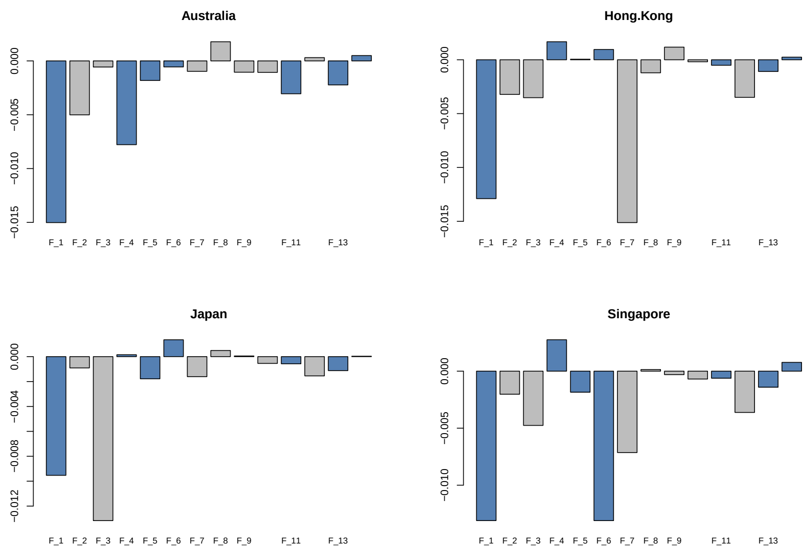
Because the structural errors in the IFACD model are modelled as a linear combination of independent factors, it is revealing to inspect the types of interactions created from the factor dynamics. One way of understanding the impact of the factors on the underlying assets is by simply inspecting the factor loadings in the mixing matrix where common sources of risk may be identified.[18] Back and Weigend (1997) for instance, showed that a dominant set of factors obtained from ICA can reveal more of the underlying structure of the time series than PCA, while Xu (1999) provides for a heuristic criterion for choosing such dominant factors. In Figures 3.1 and 3.2, the factor loadings for the countries in the Americas, Europe and Asia are shown, where the blue colored loadings represent those factors which had some degree of significant time variation in the shape parameter in Table 3.1. While it is not always easy to make inference from statistical factors, one can immediately observe the common to all indices, large and

[18]The columns in the mixing matrix $\boldsymbol{A}$ represent the asset loadings.

negative loading on Factor 1 which may represent, for instance, some common risk to the global equity markets such as an oil shock. Beyond this, it is difficult to infer with confidence anything more about the factors without some substantial analysis and mapping of those factors to some fundamental combination of risk factors, and even then this is a speculative exercise at best. A more revealing method for visualizing the



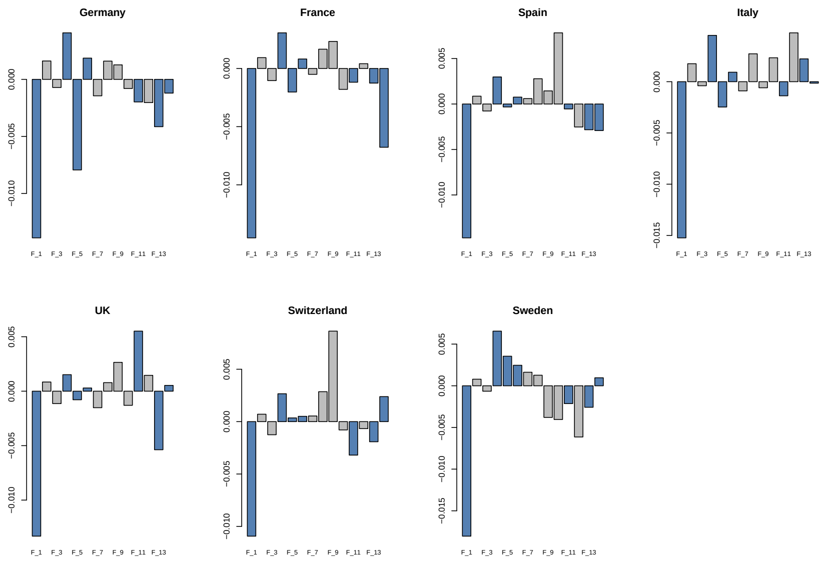
(A) Americas Country Loadings



(B) Asia Country Loadings

Figure 3.1: Factor Loadings

multivariate dynamics in GARCH systems is through the news impact function. This was originally suggested in the univariate literature by Engle and Ng (1993), providing



(A) European Country Loadings

FIGURE 3.2: Factor Loadings

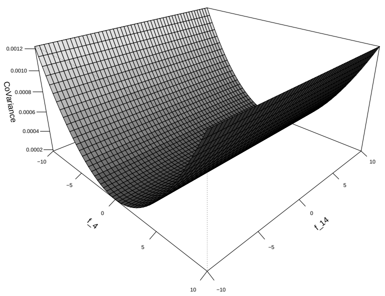
a visual representation of the impact of shocks on the time varying variance. It was extended to a surface function by Kroner and Ng (1998) who compared a number of multivariate GARCH models and the type of surfaces they generate. This was further extended in a natural direction by Jondeau and Rockinger (2009) to include the impact of higher moment co-dependence. While the IFACD model is mainly one of univariate independent dynamics, I investigate the type of interactions generated by the model by constructing news impact surfaces for the covariance and co-skewness. Since shocks impact the factors independently, the news impact surface is a combination of the independent news impact curves of the factors which, when combined via the mixing matrix $\boldsymbol{A}$, create the dynamics for the underlying asset-factor surface function. To achieve this, a set of common shocks values is first passed to the individual factors to obtain the univariate news impact curves of the variance, skew and shape, for which simulation is used to obtain the unconditional long run values required in this setup. The co-moment news impact surface is then obtained by evaluating the contribution of the shocks on the factors under consideration while maintaining all other factors at their no-shock values, and transforming the values using $\boldsymbol{A}$ into weighted shocks on the underlying assets. Formally, let the vector $\boldsymbol{z}_{t-1}$ denote the conditioning variables known at time $t-1$ for the determination of the $h_{i,t}$, $m^3_{i,t}$ and $m^4_{i,t}$, and let Z be the unconditional value of the shocks assumed to be constant, beyond i and j. Let $h_{ab,t}$, $m^3_{abc,t}$ and $m^4_{abcd,t}$ be the conditional covariance, third and fourth co-moments of the assets a,b, and c. The news impact surfaces, with respect to shocks from factors i and

j, are the three dimensional graphs of the functions:

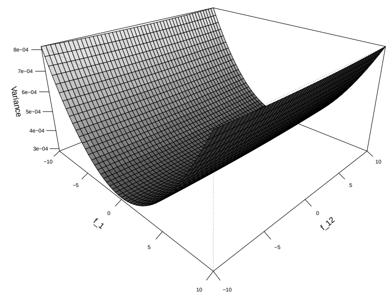
$$\begin{aligned} \sigma_{ab,t} &= f(\boldsymbol{A}, \boldsymbol{H}_{f,t-1} | (\hat{z}_{i,t-1}, \hat{z}_{j,t-1}, Z)), \\ m^3_{abc,t} &= f(\boldsymbol{A}, \boldsymbol{M}^3_{f,t-1} | (\hat{z}_{i,t-1}, \hat{z}_{j,t-1}, Z)), \\ m^4_{abcd,t} &= f(\boldsymbol{A}, \boldsymbol{M}^4_{f,t-1} | (\hat{z}_{j,t-1}, \hat{z}_{j,t-1}, Z)), \end{aligned} \tag{3.31}$$

where, $\boldsymbol{H}_{f,t-1}$, $\boldsymbol{M}^3_{f,t-1}$ and $\boldsymbol{M}^4_{f,t-1}$ represent the factor covariance matrix, the third and fourth co-moment tensor matrices, respectively, as defined in Equations (3.15) and (3.16).

Figures 3.3 and 3.4 display the covariance and co-skewness news impact surfaces



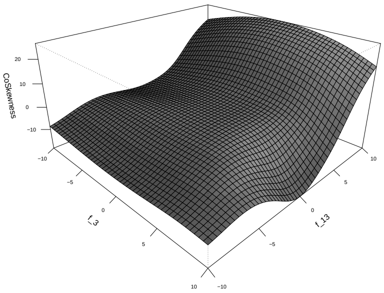
(A) (i,j)(Germany,France)



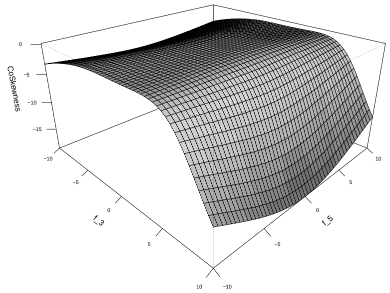
(B) (i,i)(Australia)

FIGURE 3.3: Covariance News Impact Surface

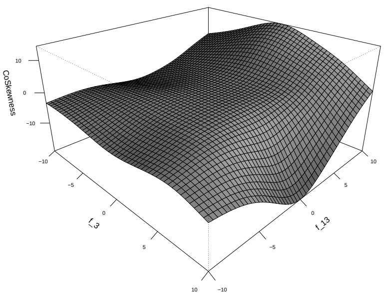
for selected asset-factor combinations. The covariance news impact surfaces shown are mostly 'U' shaped indicating the dominance of one factor over another in addition to the fact that they are independent. For example, the covariance between France and Germany, when all other factors except 3 and 14 provide shocks, is dominated by Factor 3, which a visual inspection of the factor loading also confirms. The co-skewness news impact surface figures, of the shocks from a set of factors to assets iij provide for more interesting insights, since they show how good a hedge one asset (i) is in terms of volatility changes in another (j), with a negative value indicating that asset j's return goes down with a positive increase in the volatility in country i, hence providing for a poor hedge. For example, any shock from Factor 3, but mostly a positive one, leads to a fall in the coskewness iij between Germany and France while a positive (negative) shock from Factor 13 leads to an increase (decrease) in coskewness. Sub-figure 3.4d is perhaps more revealing, showing the UK as a good hedge to increases in the volatility of the US following a shock from Factor 1, as revealed by the increase in the coskewness in reaction to shocks of any sign from that factor.



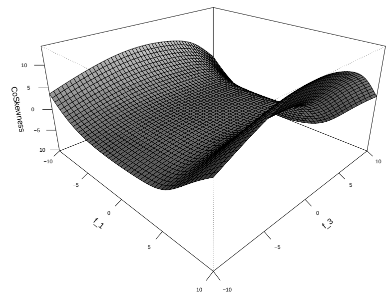
(A) (i,i,i)(Germany)



(B) (i,i,i)(Germany)



(C) (i,i,j)(Germany,France)



(D) (i,i,j)(UK,USA)

FIGURE 3.4: CoSkewness News Impact Surface

While these types of visual diagnostics may prove useful, they should usually be supplemented with more concrete analysis, as suggested by Jondeau and Rockinger (2009), in terms of simulations to determine the finite sample distribution of the reactions to shocks as well as Impulse Response functions to track the decay of the reactions to the shocks over time.

3.5.3 Model Risk Forecast Comparison

The out-of-sample empirical application on the 14 MSCI index log returns is based on a 5 day rolling forecast and re-estimation scheme. Starting from 10/08/2000, the last 4 years of daily log returns are used to estimate the models (IFACD and CHICAGO), from which the next 5 days of 1-ahead rolling forecasts are created. The models are then re-estimated moving the data window 5 days ahead and a new set of 5 day rolling 1-ahead forecasts created for a total of 522 re-estimations resulting in 2610 forecasts. For simplicity, an AR(1) models was used to capture the autocorrelation in the underlying dataset, while consistency in comparison between the 2 models was guaranteed by using the same mixing matrix $\boldsymbol{A}$ across the 2 models for each estimation window so that the only difference would be purely in the dynamics of the factors. Because of the non-dynamic nature of the independence matrix in the IFACD model, the rolling window re-estimation scheme with a fixed window size of 4 years enables the capturing of any changes to the loadings, though tracking such changes is not trivial since the independent components are identified only up to a permutation and scaling of the sources.[19] As an additional benchmark, a DCC-Normal model with AR(1) conditional mean dynamics was also estimated so as to gauge the cost, if any, of non-dynamic (in)dependence. To evaluate the performance of the models out-of-sample, a weighted linear combination of the forecast density was used in order to form portfolios from which measures could easily be calculated. To avoid bias from any particular weighting scheme, 1000 + 1[20] random portfolios were generated by sampling weights from the exponential distribution and dividing by the sum of the randomly generated deviates to create the full investment constraint. The weighted densities of the IFACD and CHICAGO models were estimated using the FFT method described in Section 3.3 from which quantile and distribution functions were then formed for use in the VaR and PIT calculations respectively. To assess the adequacy of the risk forecasts a number of tests were used, namely Berkowitz (2001) for testing the predictive density, Kupiec (1995) and Christoffersen (1998) for VaR exceedances and Christoffersen and Pelletier (2004) for VaR Durations, and described in Appendix C.

Table 3.3 reports the result under the different tests for the equally weighted (EW) and average of the randomly weighted ($RAND$) portfolios. For the Berkowitz test, there does not appear to be a significant difference between the 2 Factor models, with a rejection rate among the randomly weighted portfolios of about 16% and 14% for the IFACD and CHICAGO models respectively, indicating that overall both models fit the out-of-sample forecast density well on average. The DCC model on the other hand appears to fit the conditional weighted forecast densities very badly with an almost 100% rejection rate. For the VaR tests, both at the 1% and 5% coverage rates, the IFACD model does substantially better than CHICAGO as evidenced by the large difference in rejection rates between the two models. The DCC model does very badly

[19] For a small number of factors this is not too challenging as it is possible to use some type of pattern matching on overlapping factors to identify the factors over the rolling window and their position in the mixing matrix $\boldsymbol{A}$. Nevertheless, this requires testing each factor from window i with time index $(t_i - s + r) : t_i$ with every other factor (including the rotated version of the factor) from the next window $i+1$ with time index $t_{i+1} : (t_{i+1} - r)$, where r is the rolling period and s the window size, using for instance a distance minimization criterion.

[20] The 1001^{th} was the equally weighted portfolio.

in capturing tail events as evidenced by the high rejection rates of this test, which is not surprising given the inadequacy of the Normal distribution in capturing the observed behavior of markets. In the VaR duration test, with 1% coverage the IFACD performs somewhat better than the CHICAGO model, but at the 5% it would appear that both Factor models perform equally well. However, the DCC outperforms the other models here which is perhaps indicative of the value of dynamic dependence since the duration indirectly tests for clustering of tail events which is not likely to be fully filtered out in a static independence framework.

Table 3.3: IFACD vs CHICAGO: Forecast density and tail tests

	Berkowitz	$VaR_{1\%}$	$VaR_{5\%}$	$VaR_{Dur_{1\%}}$	$VaR_{Dur_{5\%}}$
IFACD					
EW					
$p-value$	0.097	0.359	0.055	0.085	0.005
RAND					
$p-value$	0.160	0.170	0.100	0.070	0.021
$\%Reject$	15.6	22.8	37.9	61.9	88.7
CHICAGO					
EW					
$p-value$	0.14	0.04	0.05	0.009	0.009
RAND					
$p-value$	0.220	0.070	0.100	0.040	0.020
$\%Reject$	13.7	60.8	53.3	74.7	87.3
DCC(N)					
EW					
$p-value$	0.001	0.000	0.006	0.085	0.136
RAND					
$p-value$	0.010	0.000	0.030	0.090	0.140
$\%Reject$	93.6	100.0	84.2	51.0	33.1

Note: The Table reports the out of sample performance of the IFACD and CHICAGO (maNIG), and DCC (N) models for 14 MSCI indices for the period 11/08/2000 to 28/12/2010 (2610 days) based on the density test of Berkowitz (2001), the conditional coverage test for VaR exceedances of Christoffersen (1998), the ES test of McNeil and Frey (2000) and the duration of VaR exceedances test of Christoffersen and Pelletier (2004). Starting on 10/08/2000 ($T = 1$), the last 4 years of data were used to estimate the 3 models, after which the estimates were used to produce rolling forecasts for the next 5 days. The model parameters were re-estimated taking into account new data every 5 days for a total of 522 re-estimations and 2610 out of sample forecasts. The null hypothesis in all tests is equivalent to a correctly specified model for which the table reports the p-values of the test under an equally weighted (EW) portfolio, and the average p-value of 1000 random weighted (RAND) long only portfolios with full investment budget constraint. For the RAND portfolio, the number of rejections of the null hypothesis at the 5% level of significance is also reported.

3.5.4 Model Optimal Portfolio Forecast Comparison

When moving away from the standard and nonrepresentative quadratic type utility maximization of Markowitz (1952), portfolio allocation usually takes the form of either some other type of utility maximization or minimization of some measure of risk such as those from the family of spectral risk measures defined by Acerbi (2002). In either case, it is usual to use simulated forecast scenarios to approximate the density, from which any measure may then be computed and minimized using either LP or NLP based methods. An alternative approach, seeks to approximate a utility function via a Taylor series expansion as in Jondeau and Rockinger (2006b) where the Constant Absolute Risk Aversion ($CARA$) utility function was approximated using the first 4 moments of the forecast density. This provides an interesting exposition of the value of the IFACD model which generates time varying higher co-moments and as such is considered in the next section. As an alternative, a scenario based optimization approach is also considered based on the MiniMax criterion of Young (1998) which seeks to minimize the regret from obtaining a very large loss, and is in fact the limit of the Conditional Value at Risk (CVaR) as the quantile approaches zero. Since part of the value of ACD based dynamics is in capturing the very extreme type movements which the non time varying GARCH dynamics cannot accommodate, this type of risk measure is believed to be ideally suited for this purpose.

3.5.4.1 Taylor Series Utility Expansion and Higher Moments

The approximation of expected utility based on the first four moments has been covered among others in Jurczenko and Maillet (2006) and Jondeau and Rockinger (2006b), with the higher co-moments exposition which follows taken from the latter. Consider an investor who allocates capital in order to maximize the expected utility (UW) over his end of period wealth W, and with initial wealth set at 1. The optimal allocation problem may be formulated as:

$$\begin{aligned} &\max_{w} E\left[U\left(W\right)\right] \\ &s.t. \sum_{i=1}^{n} w_i = 1 \quad w \geqslant 0, \text{ for i} = 1, \ldots, \text{n}. \end{aligned} \tag{3.32}$$

where we assume the absence of a riskless asset so that the sum of the weights w, representing the fraction of wealth allocated to risky assets, sums to one, and that we forbid short-selling. Given a vector of returns $R = (R_1, \ldots, R_n)$, the end of period wealth $W_T = (1 + \bar{r})$, where $\bar{r} = w'R$. While it is possible to calculate any utility function using the semi-analytic approach described in Section 3.3, this becomes computationally infeasible in an optimization setting. Instead, one can use a Taylor series expansion to approximate the utility using only the moments, such that the expected utility for k moments is given by:

$$E\left[U\left(W\right)\right] = \sum_{k=0}^{\infty} \frac{U^{(k)}\left(\bar{W}\right)}{k!} E\left[\left(W - \bar{W}\right)^{k}\right], \tag{3.33}$$

where $\bar{W} = w'E[R]$, is the weighted mean expected return. Loistl (1976) explored the necessary conditions for this infinite series to converge to the expected utility, strongly depending on the type of utility function used. For the purpose of this exercise we truncate the order to the first four moments for feasibility in the co-moment representation (the co-kurtosis tensor is already a daunting $n \times n^3$), being two orders more than the Mean-Variance criterion, with skewness and kurtosis directly related to investor preference (dislike) for odd (even) moments under certain mild assumptions given in Scott and Horvath (1980). The expected Utility under the Taylor series expansion is:

$$\begin{aligned} E[U(W)] = U\left(\bar{W}\right) + U^{(1)}\left(\bar{W}\right) E\left[W - \bar{W}\right] + \frac{1}{2} U^{(2)}\left(\bar{W}\right) E\left[\left(W - \bar{W}\right)^2\right] \\ + \frac{1}{3} U^{(3)}\left(\bar{W}\right) E\left[\left(W - \bar{W}\right)^3\right] + \frac{1}{4} U^{(4)}\left(\bar{W}\right) E\left[\left(W - \bar{W}\right)^4\right] + O\left(W^4\right) \end{aligned} \tag{3.34}$$

where $O(W^4)$ represents the remainder of the Taylor series due to the truncation. Define the expected return, variance skewness and kurtosis[21] as follows:

$$\begin{aligned} \bar{W} = E[r_p] \quad & = m_p = w'\boldsymbol{m} \\ E\left[\left(W - \bar{W}\right)^2\right] = E\left[(r_p - m_p)^2\right] \quad & = \sigma_p^2 = w'(AHA')w \\ E\left[\left(W - \bar{W}\right)^3\right] = E\left[(r_p - m_p)^3\right] \quad & = s_p^2 = w'\left(AM^3(A' \otimes A')\right)(w \otimes w) \qquad (3.35) \\ E\left[\left(W - \bar{W}\right)^4\right] = E\left[(r_p - m_p)^4\right] \quad & = k_p^4 = w'\left(AM^4(A' \otimes A' \otimes A')\right)(w \otimes w \otimes w) \end{aligned} \tag{3.36}$$

where A is the mixing matrix from the ICA decomposition, H the conditional factor covariance (diagonal) matrix given in (3.9), M^3 and M^4 the conditional higher co-moment tensors given in (3.15) and (3.16) respectively, and $\boldsymbol{m}$ the conditional mean vector. The expected Utility is then approximated by:

$$E[U(W)] \approx U\left(\bar{W}\right) + \frac{1}{2} U^{(2)}\left(\bar{W}\right)\sigma_p^2 + \frac{1}{3!} U^{(3)}\left(\bar{W}\right) s_p^3 + \frac{1}{4!} U\left(\bar{W}\right) k_p^4. \tag{3.37}$$

When the CARA utility function is used, then the approximation resolves to:

$$E[U(W)] = -\exp(-\lambda W) \approx -\exp(-\lambda m_p)\left[1 + \frac{\lambda}{2}\sigma_p^2 + \frac{\lambda^3}{3!} s_p^3 + \frac{\lambda^4}{4!} k_p^4\right] \tag{3.38}$$

[21] All time indices are suppressed in the representations for clarity.

where λ represents the investor's constant absolute risk aversion, with higher (lower) values representing higher (lower) aversion. I maximize this function for all rolling forecasts using an SQP based solver and making use of the first order derivatives given in Jondeau and Rockinger (2006b). A budget constraint is also imposed as are positivity and upper bounds on the assets weights of 50%. Table 3.4 shows the results for the IFACD, CHICAGO and DCC models of CARA utility portfolios maximized under 4 different risk aversion coefficients, λ, representing the mildly risk averse to extremely risk averse investor. The picture which emerges from this exercise is strikingly clear. As the risk aversion coefficient increases, and more weight is given to the higher moments, the IFACD model progressively outperforms the CHICAGO model. This is immediately evident from the p-value of the model confidence set (MCS)[22] procedure of Hansen, Lunde, and Nason (2011), using a simple loss function of the negative of the portfolio returns which overwhelmingly rejects the CHICAGO model from $\lambda = 5$ onwards. A slower progression is seen when it comes to the significance of Risk-Reward (RR) differences given by the test of Ledoit and Wolf (2008), with the IFACD model RR ratio starting to look significantly better at $\lambda = 25$ with 90% confidence. Not surprisingly, the DCC model based on the CARA utility expansion with only the first 2 moments fares worse at all levels of risk aversion, indicating that trading away some dynamics in terms of conditional dependence for the flexibility of dynamic higher moments pays off. Figure 3.5 shows the cumulative wealth (W) of the IFACD and CHICAGO based CARA ($\lambda = 25$) portfolio and the relative difference in the aggregate weights of the 3 regions represented by the 14 MSCI indices during 2 period. Since both models share the exact same conditional mean dynamics and hence ICA mixing matrix, the differences are purely the result of the conditional factor dynamics, captured in the relative weight distribution of the two optimized models in the two sub figures. The first, showing a relatively mild period in 2004, indicates relative differences in the weights of between -20% and 20% (i.e. a -20% on the aggregate Europe region indicates that the CHICAGO model was 20% underweight that region relative to the IFACD model). If that serves as a baseline for relatively mild periods, then surely the next figure which captures the run-up to the 2008 crisis, and showing relative differences of between -60% and 60%, clearly indicates where and how the two models diverged during that extremely turbulent period. Thus, despite any noise present from possibly over-parameterized dynamics, there can be no doubt that without allowing the higher moments to vary, extreme market movements are unlikely to be accommodated

[22]See Appendix C.3 for more details.

TABLE 3.4: Time varying higher co-moments portfolio with CARA utility

	IFACD	CHICAGO	DCC(N)
$\boldsymbol{\lambda = 1}$			
$\bar{W}_T$	75.47	79.72	57.33
$\hat{\mu}$	0.0018	0.0018	0.0017
$\hat{\sigma}$	0.0169	0.0170	0.0168
$\sqrt{(252)}(\frac{\hat{\mu}}{\hat{\sigma}})$	1.69	1.71	1.60
LW [stat ; p-value] vs IFACD		[1.248 ; 0.224]	[1.106 ; 0.278]
MCS [p-value]	[0.357]	[1.000]	[0.357]
Log Relative Wealth		0.05	-0.27
$\boldsymbol{\lambda = 5}$			
$\bar{W}_T$	94.07	58.60	43.89
$\hat{\mu}$	0.0019	0.0017	0.0016
$\hat{\sigma}$	0.0175	0.0161	0.0159
$\sqrt{(252)}\frac{\hat{\mu}}{\hat{\sigma}}$	1.72	1.66	1.58
LW [stat ; p-value] vs IFACD		[0.559 ; 0.564]	[1.186 ; 0.226]
MCS [p-value]	[1.000]	[0.054]	[0.044]
Log Relative Wealth		-0.47	-0.76
$\boldsymbol{\lambda = 10}$			
$\bar{W}_T$	64.20	39.84	26.02
$\hat{\mu}$	0.0017	0.0015	0.0014
$\hat{\sigma}$	0.0163	0.0154	0.0151
$\sqrt{(252)}\frac{\hat{\mu}}{\hat{\sigma}}$	1.69	1.58	1.43
LW [stat ; p-value] vs IFACD		[1.301 ; 0.198]	[2.316 ; 0.020]
MCS [p-value]	[1.000]	[0.021]	[0.003]
Log Relative Wealth		-0.48	-0.90
$\boldsymbol{\lambda = 25}$			
$\bar{W}_T$	23.83	17.82	13.24
$\hat{\mu}$	0.0013	0.0012	0.0011
$\hat{\sigma}$	0.0145	0.0142	0.0140
$\sqrt{(252)}\frac{\hat{\mu}}{\hat{\sigma}}$	1.44	1.34	1.23
LW [stat ; p-value] vs IFACD		[1.636 ; 0.107]	[2.306 ; 0.024]
MCS [p-value]	[1.000]	[0.039]	[0.012]
Log Relative Wealth		-0.29	-0.59

Note: The Table reports the out of sample performance of the IFACD and CHICAGO (maNIG), and DCC (N) models from the optimization of the CARA utility approximation using only the first 4 co-moment matrices, for 14 MSCI indices for the period 11/08/2000 to 28/12/2010 (2610 days). Starting on 10/08/2000 ($T = 1$), the last 4 years of data were used to estimate the 3 models, after which the estimates were used to produce rolling forecasts for the next 5 days. The model parameters were re-estimated taking into account new data every 5 days for a total of 522 re-estimations and 2610 out of sample forecasts. The performance statistics reported are $\bar{W}_T$ representing terminal wealth of a portfolio of starting value of 1, the average return ($\hat{\mu}$), average volatility ($\hat{\sigma}$), the annualized risk-return ($\sqrt{(252)}\,(\hat{\mu}/\hat{\sigma})$), the statistic and p-value of the Ledoit and Wolf (2008) test for the difference in the RR ratio between the IFACD and other models, the p-value of the MCS procedure of Hansen, Lunde, and Nason (2011) using 10000 bootstrap replications under the range statistic and the relative log difference in Terminal Wealth between the IFACD and other models. The CARA utility was optimized under 4 different risk aversion levels (λ) representing the mild to very risk averse investor.

by purely GARCH dynamics even when using a conditional distribution with fat tails and skewness such as the NIG.

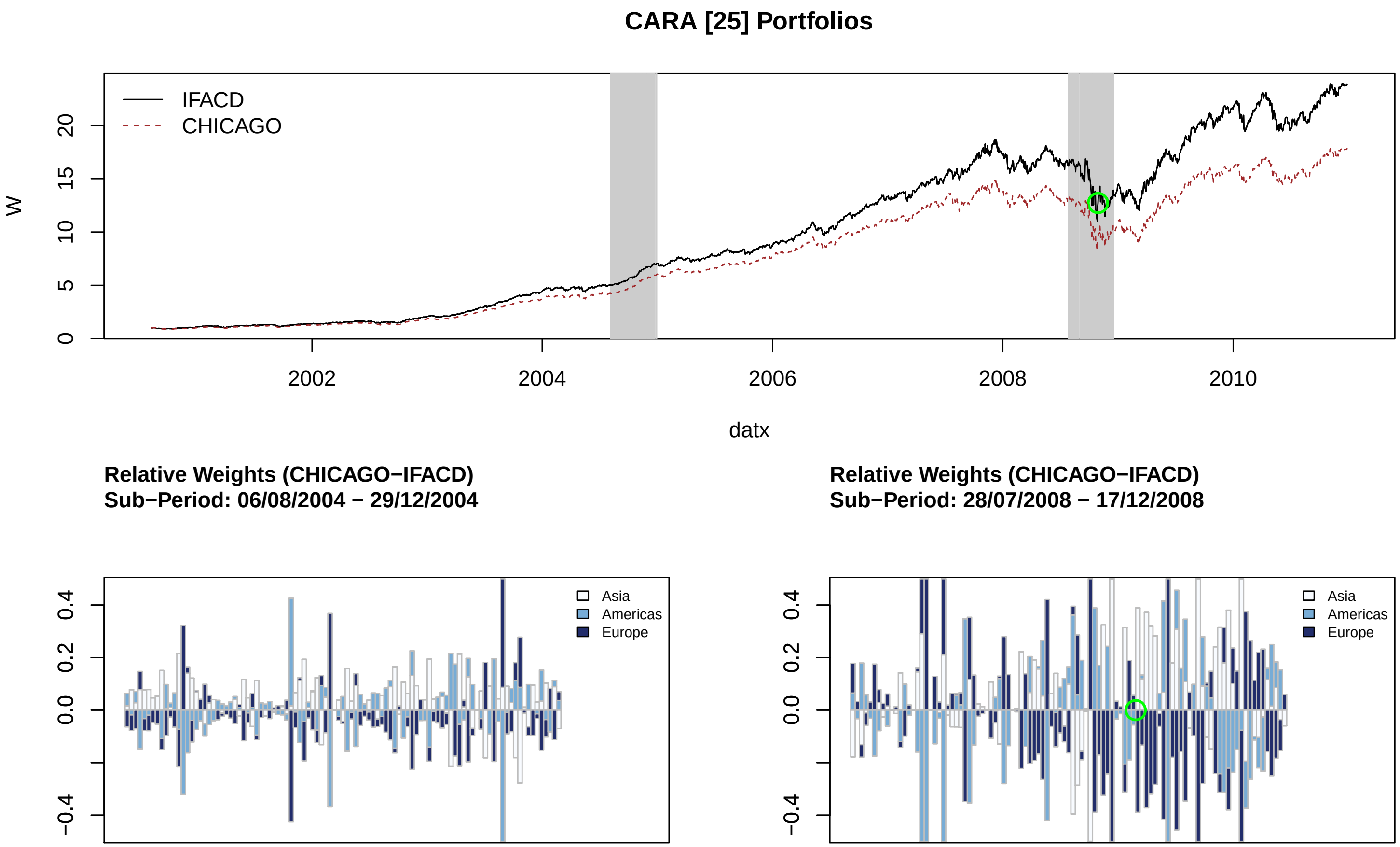


FIGURE 3.5: CARA Based Portfolio ($\lambda = 25$) and Differential Weight Allocation

3.5.4.2 Extreme Loss Aversion via MiniMax Optimization

Since part of the value of the IFACD model dynamics lies in capturing the extremes of the tails, it seems natural to investigate a portfolio application which targets minimization of the maximum regret. The MiniMax[23] model of Young (1998) has game theoretic origins, where agents aim to minimize their expected maximum losses. Formally, when applied to portfolio selection, given N assets and T periods, the model may be represented by as a Linear Programming (LP) problem:

$$\begin{aligned}
&\min_{M_p,w} M_p \\
&s.t. \\
&M_p - \sum_{j=1}^{m} w_j r_{i,j} \leq 0, \forall i = \{1,\dots,n\} \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j
\end{aligned} \tag{3.39}$$

where M_p is the objective minimization value representing the maximum loss of the portfolio[24] given a vector of weights w, C some minimum level of return and μ the forecast return vector on the m securities. While the problem was originally considered on historical data so that n represented the number of periods in the dataset, it is also possible to consider n to be the size (rows) of a 1-ahead forecast scenario, in which case the problem is equivalent to the maximization of the Conditional Value at Risk for a quantile approaching zero. Because of this discretization of the forecast density, the actual minimum quantile achieved will be loosely related to the number of scenarios T generated.[25] The first constraint in equation (3.39) guarantees that M_p is bounded from above by the maximum portfolio loss. Tracing out the set of portfolios for different

[23] When dealing with a loss distribution, MiniMax is the technically correct term, whilst when working with returns, maximizing the minimum negative return is equivalent to the MaxiMin problem, but in keeping with the literature I will only refer to this problem as Minimax.

[24] To express this as a minimization problem I work with losses so that M_p represents the maximum loss rather than the minimum return.

[25] This effectively means that there will always be some uncertainty when comparing models using the same T sized scenarios, but this will always be the case for scenario based portfolio optimization models with almost any measure.

levels of G (using an equality rather than inequality), will generate the portfolio frontier from which the optimal risk to return portfolio may for instance be chosen. Yet there is no reason why we cannot directly estimate this optimal ratio of risk to return making use of fractional programming as described in Charnes and Cooper (1962) and more recently in Stoyanov, Rachev, and Fabozzi (2007). The Minimax LP problem can therefore be reformulated as follows:

$$
\begin{aligned}
&\min_{M_p, w, b} M_p \\
&M_p - \sum_{j=1}^{m} w_j r_{i,j} \leq 0, \forall i = \{1, \ldots, n\} \\
&\sum_{j=1}^{m} w_j \mu_j = 1 \\
&\sum_{j=1}^{m} w_j = b \\
&b \geq 0
\end{aligned}
\tag{3.40}
$$

where b is the multiplier coefficient added to the optimization problem as a result of transforming the fractional risk/reward problem. Further details can be found in Charnes and Cooper (1962) for LP and Dinkelbach (1967) for NLP type problems.

For the empirical application using the MiniMax measure, I have chosen an alternative dataset of weekly log total returns for the point in time constituents of the DJIA index taken from the CRSP database.[26] This dataset is analyzed in more detail in Section 4.4.1 where it is used in a large out-of-sample comparative model and risk measure application. The dataset was chosen because of the continuity offered by the underlying index and the liquidity of the stocks covered, which together with a choice of lower frequency should hopefully provide some diversity in the evidence presented on the value of the IFACD model.

The out-of-sample weekly forecasts, based on weekly re-estimation of the models using all available data going as far back as 1965, covers the period 13/01/1975 to 03/01/2011 for a total of 1878 weeks. For each weekly forecast, a scenario of size 7000×30 was created and optimized using the method in Equation (3.40) subject to positivity constraints on the weights, maximum allocation of 20% per security and a full investment constraint. The conditional mean forecasts where generated from an AR(1) model,

[26] Whenever Dow Jones added, dropped or changed companies in the index, this was immediately reflected in the application.

and the model dynamics used were the same as in the previous application in Section 3.5.4.1. Additionally, an AR(1)-DCC model with multivariate Laplace innovations was also fitted in order the gauge the cost of the non-dynamic independence while providing for thicker tails than the Normal,[27] thus not penalizing the model disproportionately versus the IFACD and CHICAGO models which use the very flexible NIG distribution. Table 3.5 reports the results of the application based on the terminal wealth, return to risk (RR) ratio and VaR at the 1% quantile. I also report the standardized statistic of the difference in the RR ratios between the IFACD and the other two models, together with the p-value under the null that the difference is zero based on the test of Ledoit and Wolf (2008). It would appear that the difference between the IFACD and CHICAGO models based on the RR ratio is only just marginally insignificant at the 10% level, despite almost a doubling of the terminal wealth without adding more risk, whilst in the case of the DCC model the p-value is significantly higher indicating a strong rejection of the alternative hypothesis that the RR ratios differ. Using the MCS procedure as in Section 3.5.4.1, under a simple loss function of the negative portfolio returns, all 3 models belong to the optimal set with 90% confidence, though it is clear from the relative values that the IFACD dominates the CHICAGO with 70% confidence.

TABLE 3.5: IFACD, CHICAGO and DCC based portfolios under MiniMax criterion

MiniMax Optimization	$IFACD(NIG)$	$CHICAGO$	$DCC(MVL)$
$\bar{W}_T$	892.8	486.2	577.7
$\hat{\mu}$	0.003972	0.003650	0.003712
$\hat{\sigma}$	0.026270	0.026425	0.025241
$\sqrt{(52)}\frac{\hat{\mu}}{\hat{\sigma}}$	1.090	0.996	1.060
LW [stat ; p-value]		[1.53 ; 0.12]	[0.48 ; 0.64]
MCS [p-value]	[1.00]	[0.31]	[0.35]
$VaR_{1\%}$	0.0623	0.0637	0.0640

Note: The Table reports the out of sample performance of the IFACD, CHICAGO and DCC (Laplace) models optimized under the mean-Minimax (LP) measure using scenario methods, for the weekly log returns of the 30 point in time constituents of the DJIA index for the period 13/01/1975 to 03/01/2011 (1878 weeks). The models were re-estimated every week, and a scenario forecast for the following week of size 7000×30 generated and optimized under a mean-Minimax fractional linear programming model. The performance statistics reported are $\bar{W}_T$ representing terminal wealth of a portfolio of starting value of \$1, the average mean ($\hat{\mu}$), average volatility ($\hat{\sigma}$), the annualized return to risk (RR) ratio ($\sqrt{(252)}(\frac{\hat{\mu}}{\hat{\sigma}})$), the statistic and p-value of the Ledoit and Wolf (2008) test for the difference in the RR ratio between the IFACD and every other model, the p-value of the MCS procedure of Hansen, Lunde, and Nason (2011) using 10000 bootstrap replications under the range statistic, and the VaR at the 1% coverage rate.

[27]The weighted Laplace portfolio has an excess kurtosis of 3.

3.6 Conclusion

In portfolio and risk management analysis it is important to consider not only asymmetric and heavy tailed multivariate distributions but also time variation in skewness and kurtosis. This is particularly true within a GARCH framework where the variance dynamics are unable to accommodate the extreme swings in security prices. Modelling the conditional density dynamics in a multivariate GARCH setup has proved unfeasible mainly because of the difficulty in dealing with tractable representations of multivariate distributions. The IFACD model presents the opportunity to use time varying dynamics for all distributional parameters in a multivariate setting, without incurring the usually penalty of increasing problem dimensionality, by making use of the statistical independence factor framework. It provides for a truly large scale and very fast computation of systems which can be estimated making use of parallel resources, a feature only available within this framework. Some of the unique features of this model such as closed form higher moments and the semi-analytic expression for the full weighted portfolio density have clear applications in portfolio and risk management as shown in this chapter. Modelling of higher moment dynamics appears to provide the most benefit during periods of market stress when GARCH volatility with static higher moments cannot adjust to extreme expansions in the conditional density representative of large negative returns. Evidence of this was presented in the empirical section of this chapter with a risk and portfolio application using two different datasets. In the VaR application, the IFACD model had the lowest number of rejections for correctly capturing the exceedances at the 1% coverage level, surpassing both the CHICAGO and DCC models. When using only the conditional co-moment forecasts, the IFACD model again outperformed the CHICAGO and DCC models, using a variety of performance measures, as more weight was assigned to the higher moments, commensurate with increasing risk aversion. With the weekly DJIA index dataset, using a very large out-of-sample rolling portfolio application and an extreme aversion risk measures, the IFACD model again provided the superior performance based on the MCS procedure. The tradeoff for working with such flexible dynamics in a multivariate setting is in the use of static independence. However, when compared with a DCC model which benefits from dynamics in multivariate dependence, any drawbacks were not apparent under a frequent re-estimation and rolling update scheme in the empirical applications. Possible avenues for further research with these models would include possible time variation in the ICA mixing matrix, though the research thus far from the ICA community suggests this to be a rather hard problem. Dimensionality reduction via the PCA whitening stage is also a possibility as long as the independent components are contained in the reduced subspace which is difficult to say a-priori. The most valuable research would most likely focus on the univariate ACD estimation, with exploration of alternative types of dynamics, possible use of estimation procedures such as GMM and more robust optimization algorithms to deal with the nonlinear bounding transformation.

Chapter 4

Active Weights for Bad Benchmarks

Since their introduction in the early 90's, Exchange Traded Funds (ETF) have exploded in popularity with assets growing to approximately \$1.35 trillion by Q3:2010,[1] with over 3,011 such funds globally, and accounting for about 12% of all mutual fund assets. Not surprisingly, the top 16 ETF[2] by Assets under Management (AuM) are populated with US based index trackers such as the 'SPY' (State Street S&P 500 Index Fund), 'QQQ' (Invesco PowerShares NASDAQ-100 Index) and 'DIA' (State Street Dow Jones Industrial Average Fund), giving investors a relatively low cost way of passively tracking these market indices. At the heart of this index fund growth phenomenon is a long standing and ongoing debate on the merits of passive versus active investing. One of the first studies to argue in favor of passive investing was Jensen (1968) who studied the performance of 115 mutual fund managers for the period 1955-1964. He reported that only 48 out of 115 mutual funds outperformed the market excluding management fees, but after the fees had been subtracted, this number dropped to 39. Many studies following Jensen's appear to support his findings. Fama (1991) provided for a summary of subsequent studies showing in particular that pension funds under-performed passive benchmarks on a risk adjusted basis. Malkiel (2003) argued on the absence of any recognizable anomalies or irrationalities which would lead to superior returns, hence advocating a passive investment strategy. French (2008) investigated the overall

[1] Source: BlackRock ETP Landscape Industry Highlights, Bloomberg, Year End 2011
[2] Source: http://etfdb.com/compare/market-cap/

cost of active investing, concluding that over the period 1980-2006, the typical investor would have increased his average annual return by 67 basis by switching to a passive market portfolio. Yet despite the pedigree and force of arguments of these authors, active investing continues to thrive, as witnessed by the equally strong growth of hedge funds, which have grown from a 'paltry' \$2.8 billion in 1995 to over \$2 trillion[3] by 2012, with the top 225 managers (approximately less than 2%) holding more than 60% of the AuM, the majority of which came from institutional sources, representing a sophisticated class of investor. Arguments in favor of active investment have looked at certain market anomalies as the reasons for observed out-performance, such as momentum (see for example Grinblatt and Titman (1992) and Jegadeesh and Titman (1993)), and more generally the types of behavioral biases documented in Tversky and Kahneman (1974) and Kahneman and Tversky (1979) which could give rise to non rational investment decisions which can persist. There is little doubt that security returns do not conform to the classical view of markets put forward by Bachelier (1964), as the evidence for the stylized facts is now overwhelming. Specific evidence of this was presented in previous chapters and a feasible solution to an important effect, the time variation in higher moments, was proposed in a multivariate context. When the dynamics of securities are characterized by time varying moments and co-moments, it is not very likely that a simple index weighting scheme such as equal weighting or capitalization weighting will be optimal. Even when using robust methods to calculate the covariance matrix for mean variance optimization as in Scherer (2007) or Bayesian methods as in Pastor and Stambaugh (2000) or Black and Litterman (1992), using the unconditional historical data without accounting for conditional variation in the moments and co-moments is not going to provide for a good forecast.

The majority of empirical studies have used a variety of index benchmarks to gauge the performance of fund managers and the value of active investing. Worst still, these indices form the basis for rewarding a large number of investment managers whose sole aim is to outperform their benchmark. Grinold (1992) asked whether index benchmarks are efficient, using the test of Gibbons, Ross, and Shanken (1989) on the country indices of Australia, Germany, Japan, U.K. and U.S. for the period ending 1991 (with start dates as early as 1973 for the U.S.). He found that 4 out of the 5 indices were not efficient ex-post. Demey, Maillard, and Roncalli (2010) have also argued that neither capitalization nor price weighted indices are efficient, exposing investors to what should be diversifiable risk. DeMiguel, Garlappi, and Uppal (2009) on the other hand

[3] Source: HFMWeek.

argued that it is difficult to beat the equal weighting strategy, considering a number of mostly mean-variance related models on limited history monthly data, but again made no formal statistical comparison of the Sharpe ratio differences. In defense of optimization, and a direct reply to the previous study which was criticized for using very short histories for the estimation of the covariance, Kritzman, Page, and Turkington (2010) provide a very thorough empirical application across a number of portfolios using the mean-variance criterion, including one comprising daily returns of the point in time constituents of the S&P500 from 1998 to 2008, and show substantially large differences (though again not statistically tested) in Sharpe ratios against the equal weighting strategy. Similarly, Martellini (2008) used total volatility as a model-free estimate of a stock's excess expected return in order to design improved equity benchmarks, finding that the maximum Sharpe ratio portfolios significantly outperformed capitalization weighted schemes on a risk adjusted basis. Unfortunately, the differences in Sharpe ratios presented were again evaluated ad-hoc and not using any test of significant statistical difference. Nevertheless, there is little doubt that a well thought out approach to the investment allocation life cycle process can provide significant value added versus either a naive 1/N strategy or many of the benchmark market indices. The growth in active investment products means that investors face a daunting task of sifting through relative risk and performance histories in order to rank and choose a suitable investment. When the comparison is made against one of the typical market indices, as is usually the practice, investors are setting the benchmark too low and rewarding managers too high.

In this chapter I present evidence, through a large out of sample application, that it is possible to outperform the benchmark index on which the performance of so many fund managers is gauged and rewarded. Making use of a variety of models and measures, representing varying degrees of modelling sophistication, and without any superior conditional mean forecast model, I argue that the weighting of one of the most popular indices, the Dow Jones Industrial Average (*DJIA*), is neither efficient nor does it represent a good benchmark. Outperforming this and related indices such as the S&P 500 is both possible and feasible. Using this unique[4] point in time deep historical dataset of the DJIA index members, representing what is likely the most liquid stocks in the US market, it is possible to invest within a compact set which is easy to track and replicate

[4]To the best of my knowledge, there has not been another study of this point in time constituents of the Dow 30 used in an out-of-sample portfolio allocation application.

over time. Unique features of this application also include the use of fractional programming and the derivation and use of certain smooth Nonlinear Programming (*NLP*) formulations allowing the inclusion of leverage in the optimization process, with confidence, and without resorting to solutions requiring strong assumptions of the portfolio weights' trimability (see Section 4.3.2) or global optimization approaches.[5] The models from the previous chapters are used in this application as the data generating processes for the modelling and simulation of scenario based density forecasts. Using a range of models and risk measures it is possible to compare the different portfolios formed from these, and unlike some previous studies in the literature on applied portfolio allocation, the comparison is not limited to terminal wealth but makes use of tests such a the Model Confidence Set of Hansen, Lunde, and Nason (2011) and the test of Ledoit and Wolf (2008) for proper statistical evaluation of Sharpe ratio differences. Finally, the application has a weekly holding/modelling cycle, which is a compromise between the data length requirements of the econometric models which exclude the use of monthly data and the noise in higher frequency datasets as well as the cost of more frequent re-balancing which excludes the use of daily data.

The chapter is organized as follows. Section 4.1 briefly reviews stochastic programming models and scenario optimization. Section 4.2 reviews the measures of risk and deviation used in the empirical application while Section 4.3 discusses methods for optimizing such measures using fractional linear and nonlinear programming methods, whilst also proposing certain smooth approximations to the optimization of some discontinuous functions. The results of the empirical application are presented and discussed in Section 4.4, and Section 4.5 concludes.

4.1 Stochastic Programming Models

> "...there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know." – Donald Rumsfeld

[5] Some evidence is also presented to show that these smooth NLP approximations are very accurate and provide for substantial speed increases versus equivalent LP formulations.

Randomness in the underlying environment leads to uncertainty, which can be characterized, albeit approximately, by a model with a probability distribution. The uncertainty is by no means resolved by the modelling process, but simply structured under a set of assumptions for enabling decision-making. In terms of the quotation, it allows the assignment of probabilities to some unknowns so that they become 'known unknowns'.[6] This structured uncertainty, usually arises in the following areas:

- Model Selection Uncertainty.
  The underlying true Data Generating Process (*DGP*) is never known, but some of its properties inferred from observing some limited history. The set of models and distributions used in this chapter, while quite general are by no means exhaustive and in some cases, particularly with omnibus distributions like the GH will overlap over a certain space.[7] In Bayesian modelling, the problem of model selection uncertainty has been approached by using ensemble learning techniques such as model averaging (*BMA*), bagging and boosting (see for example Raftery, Madigan, and Hoeting (1997) for BMA in linear regression and Polikar (2006) for a general overview). Most of these methods however have mostly been used in univariate linear models and it is not clear how to feasibly extend this to multivariate GARCH type models let alone in a non Bayesian setting.
- Distributional Uncertainty.
  In simulating forecasts from an estimated model, the explicit assumption is that the forecast distribution is the same as the one used for the in-sample estimation, adding another layer of uncertainty. For ARMA and GARCH based models, Pascual, Romo, and Ruiz (2004, 2006) suggest to sample from the empirical distribution of the standardized residuals. However, depending on the actual size of the in-sample data, this may be inadequate given the requirements of scenario optimization using a large discrete set to approximate the continuous distribution. This is likely to be important when the measure applied to the scenario is tail dependent.

[6]The 'unknown unknowns' still remain as a result of imperfect models for capturing the underlying dynamics, or what Nasim Taleb calls 'Black Swans' events.

[7]For example, the DCC Student will overlap with the DCC Normal in cases here the data is multivariate Normal and the shape parameter > 30. However, there are efficiency gains from estimating certain simpler models individually which is why I do not restrict myself to just 2 omnibus models for estimation.

- Parameter Uncertainty.

  In a non Bayesian setting, explicitly accounting for parameter uncertainty is extremely difficult particularly in the multivariate case. In portfolio optimization, this has usually been limited to the Mean-Variance (EV) model using historical data and choices on the prior mean vector and covariance matrix. For example, Garlappi, Uppal, and Wang (2007) tackle both parameter and model uncertainty in an EV setup and conclude that it leads to higher out-of-sample Sharpe ratios for their set of 8 international equity indices.[8]

The purpose of stochastic programming (SP) is to incorporate such uncertainty into the objective or constraint functions with a view to obtaining an optimal set of decisions. This is done by constructing a scenario, or set of scenarios, representing the possible future path or paths of the underlying process (as a discrete time approximation to the continuous case) incorporating the uncertainty with respect to the model and future, and from which decisions can be based. These types of models were originally proposed and analyzed among others by Dantzig (1956, 1992), Beale (1955), Dantzig and Infanger (1993), Madansky (1962) and Charnes and Cooper (1959). An excellent exposition of SP models in asset and liability management can be found in Kouwenberg and Zenios (2006), from which the notation in the remainder of this Section is based on. The basic SP problem may be expressed as follows:

$$\begin{aligned} &\min_{x} \mathbf{E}\left[f_0\left(x,\omega\right)\right] \\ &s.t. \\ &f_i\left(x,\omega\right) \leq 0, i = 1,\ldots,n \end{aligned} \tag{4.1}$$

where x is an m-dimensional vector of decision variables, ω represents the random vector, and the set of objective and constraint functions $f_i : \mathbb{R}^m \times \Omega \to \mathbb{R}$. When the random vector ω can be represented by a discrete and finite distribution with support the set $\Omega = \{\omega_1, \ldots, \omega_N\}$, this is called the scenario set. The 2 basic cases, representing extremes in SP are the anticipative and adaptive models. In the anticipative model, decisions are made before any uncertainty is resolved by conditioning the objective and constraint functions on a random vector representing the anticipated realization of future outcomes given an underlying DGP. In the adaptive model, decisions are made after uncertainty in some of the variables is resolved. This does not provide

[8]Unfortunately, the comparison with other models stops at a simple comparative difference of Sharpe ratios without mention of the significance of those differences.

for a full resolution of uncertainty otherwise this would result in a fully deterministic model. Formally, let $\mathcal{A}$ represent the set of all relevant information that could become available by realizing an observation, being a subfield of the $\sigma-$field of all the outcomes generated from the set Ω of the random vector ω. The decision x on the random vector ω is termed $\mathcal{A}-$measurable, and the adaptive SP can be represented as:

$$\begin{array}{lll} \min\limits_{x(\omega)} & E\left[f_0\left(x\left(\omega\right),\omega\right)|\mathcal{A}\right] & \\ s.t. & E\left[f_i\left(x\left(\omega\right),\omega\right)|\mathcal{A}\right] \leqslant 0 & i=1,\ldots,n \\ & x\left(\omega\right)\in X & \text{almost surely} \end{array} \tag{4.2}$$

The mapping $x:\Omega\rightarrow X$ is such that $x\left(\omega\right)$ is $\mathcal{A}-$measurable, and the problem can be handled by solving a set of deterministic programs for every ω. Between the two extremes of no information (anticipative model) and complete information (distribution model), is the partial information model which allows for intermediate actions or recourse. These multi-stage SP with recourse problems, given a discrete and finite distribution from which a scenario set can be extracted, may be formulated into deterministic large scale Linear (*LP*), Quadratic (*QP*) or Nonlinear Programming (*NLP*) problems. Notable early contributions are Bradley and Crane (1972) who developed a multi-stage SP model for bond portfolio management, where an LP decomposition algorithm was presented allowing for the efficient and recursive solution of sub problems in the general portfolio model. More recently, applications in asset and liability management can be found for instance in Dantzig and Infanger (1993), Kouwenberg (2001), Mulvey and Shetty (2004), Herzog, Dondi, Keel, Schumani, and Geering (2007) and Huang (2010). The evidence from the research on multistage SP models is that they do add significant value to the modelling process. However, they incur a high computational burden for the added complex modelling they enable and as such do not readily lend themselves to the type of large scale empirical back-testing application undertaken in this chapter, which is restricted to a single stage anticipative model, more typically observed in practice.

Irrespective of the type of SP model used, two issues deserve particular attention. First, it is important to check the consistency of the measures chosen against the model generated scenarios so as to gauge a reasonable scenario size and obtain some insight into the degree of error arising from the discretization. When comparing multiple models, it is likely that one size will not fit all. Ideally one would want to choose, for each model, a size which will give the same type of consistency across all models. I take this issue up in more detail in the empirical section. Secondly, it is important to ensure that none of the scenario sets include risk free arbitrage. This might happen if the objective function is unbounded from below (first order arbitrage) or a state exists for which it is zero (second order arbitrage) or alternatively that all states yield positive payoffs. In practice, this can and should be checked post optimization, though is not very common[9] with the types of models and frequency of data used in the empirical application.

4.2 Risk and Deviation: Models and Properties

In portfolio and resource allocation, characterization of the future uncertainty by a scenario of possible outcomes does not in itself provide value to the decision maker unless he is able to rank, choose and allocate among competing alternatives based on a set of preferences. Historically, theories of such preferences have been normative, describing a certain set of principles or axioms for rational behavior. The expected utility theory, first proposed by Bernoulli (1954) as a solution to the St.Petersburg Paradox[10], and formalized by Von Neumann and Morgenstern (1944) into 4 key axioms (Completeness, Transitivity, Independence, Continuity), provides the most popular approach[11] to rational decision making. Risk attitudes in expected utility theory are usually measured by the Arrow-Pratt (see Arrow (1963)) definitions of absolute and relative risk aversion (ARA and RRA respectively) which are standardized measures of the degree of curvature in the utility functions[12] Utility functions of the form $U(W) = -\exp(-\lambda W)$, for instance, have constant absolute risk aversion, an application of which was used in 3.5.4.1. Unfortunately, by solving 1 paradox, the theory introduced 2 others; the Allais Paradox provides a challenge to the Independence axiom and relates to issues of bounded rationality already discussed by Simon (1955) who also argued against the homogeneous decision maker so popular in classical finance theory; the Ellsberg Paradox provides a challenge to the Completeness axiom leading to inconsistent choice and giving rise to ambiguity aversion. In terms of the quotation in Section 4.1, this is related to preference for 'known' versus 'unknown knowns'. Empirical studies have also found inconsistencies between the prescribed behavior of expected utility and the observed behavior of individuals. Lichtenstein and Slovic (1973) found that subjects sometimes

[9] In fact, no occurrence of arbitrage was found in any scenario in the empirical application.

[10] Where the distinction between expected utility and expected return was made.

[11] Though by no means the only approach. See for example Savage (1962) for subjective expected utility, Quiggin (1982) and Schmeidler (1989) for rank dependent utility and Zadeh (1965) for Fuzzy Logic.

[12] Formally, $ARA(W) = -\frac{U''(W)}{U'(W)}$ and $RRA(W) = -\frac{WU''(W)}{U'(W)}$.

exhibit signs of preference reversals with regards to their certainty equivalents of different lotteries which may arise as a result of the way the decision problem is framed. In fact many of the paradoxes investigated, and most importantly the actual behavior observed in experiments is well captured by the behavioral approach to decision making introduced by Kahneman and Tversky (1979) and formalized in their cumulative prospect theory. Underreaction, overreaction and related irrational processing of information (see for example De Bondt and Thaler (1987) and Frazzini (2006)) are well captured by a range of cognitive biases described in the behavioral finance literature. More importantly, cumulative prospect theory is framed in terms of gains and losses rather than Terminal Wealth (TW) which may lead to a more practical approach to defining disutility and risk for the average investor rather than the rigid axioms of expected utility theory.

The rather arbitrary nature of utility functions, and difficulty in pragmatically having a one size fits all approach, has led to a parallel strand of research in an attempt to depart from the utility framework altogether and to make use of criteria based on more objective and concrete concepts, mainly related to loss aversion. A first attempt at quantifying risk as the loss beyond a certain threshold was the Safety-First criterion of Roy (1952) which aimed at minimizing the probability of being below an investor's minimum acceptable return (MAR). Later concepts looked at improving on this measure by penalizing losses below the threshold at different rates,[13] representing different preferences for risk. Irrespective of the type of measure, the more general reward-risk approach has proved very popular both academically and in practise since it enables preferences to be summarized in a few scalar parameters such as the mean and variance. However, it was not until recently that formal qualifications of the properties of such measures were defined in seminal papers by Artzner, Delbaen, Eber, and Heath (1999) and Acerbi (2002) on risk and Rockafellar, Uryasev, and Zabarankin (2006) on deviation, with the latter establishing the connection between the two. Continuing with the notation from Section 4.1, consider the probability space $\{\Omega, \mathcal{A}, P\}$ where P is the probability on the $\mathcal{A}$ measurable subsets of Ω. Rockafellar, Uryasev, and Zabarankin (2006) defined a set of axioms which functions in the linear space $\mathcal{L}^2$ (which includes the mean and variance) should satisfy. Formally, the deviation measure functionals $\mathcal{D} : \mathcal{L}^2(\Omega) \rightarrow [0, \infty]$ should satisfy the following axioms:

- (D1) $\mathcal{D}(C) = 0 \; \forall$ constants C,

[13] The probability of being below a threshold $Pr(R_i < \theta)$ is equivalent to to $E\left[(R_i - \theta)^a\right]$, with $a = 0$.

- (D2) $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ $\forall$ X and $\lambda > 0$,
- (D3) $\mathcal{D}(X + X') \leq \mathcal{D}(X) + \mathcal{D}(X')$ $\forall$ X and X',
- (D4) $\mathcal{D}(X) \geq 0$ $\forall$ X and $\mathcal{D}(X) > 0$ $\forall$ nonconstant X,

where (D1) is the translation invariance property under the special condition given for constants (i.e. insensitivity to constant shifts), (D2) represents the positive homogeneity property, (D3) the subadditivity property , while (D4) is the lower bound implied by the domain of $\mathcal{D}$. Artzner, Delbaen, Eber, and Heath (1999) provide the equivalent 'coherent' risk measure functionals $\mathcal{R} : \mathcal{L}^2(\Omega) \rightarrow (-\infty, \infty]$ which should satisfy the following axioms:

- (R1) $\mathcal{R}(C) = -C$ $\forall$ constants C,
- (R2) $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ $\forall$ X and $\lambda > 0$,
- (R3) $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ $\forall$ X and X',
- (R4) $\mathcal{R}(X) \leq \mathcal{R}(X')$ whenever $X \geq X'$,

where (R1) is the translation invariance property, (R2) is positive homogeneity, (R3) subadditivity property and (R4) the monotonicity property. More plainly, (R1) implies that adding a constant to a set of losses does not change the risk,[14] (R2) that holdings and risk scale by the same linear factor, (R3) that portfolio risk cannot be more than the combined risks of the individual positions, and (R4) that larger losses equate to larger risks. Acerbi (2002) defined the family of spectral risk measures as those with weighted[15] quantiles, possessing the properties of coherent risk measures and additionally:

- (R5) If $\mathcal{F}(X) = \mathcal{F}(Y)$, then $\mathcal{R}(X) = \mathcal{R}(Y)$,

which essentially implies that portfolios with equal cumulative distribution functions ($\mathcal{F}$) should have the same risk. Rockafellar, Uryasev, and Zabarankin (2006) defined a one-to-one relationship between deviation and risk measures[16] which satisfy properties (R1), (R2), (R3) and are strictly expectation bounded ($\mathcal{R}(X) > E[-X]$) as:

[14] A point taken up forcefully by Glyn Holton on his critique of these properties in his blog: http://glynholton.com/2008/09/the-case-for-incoherence/

[15] With positive weights which are normalized to sum to 1.

[16] In the rest of chapter, I will refer to 'risk' to mean both risk and deviation measures.

- $\mathcal{D}(X) = \mathcal{R}(X - E[X])$,
- $\mathcal{R}(X) = E[-X] + \mathcal{D}(X)$.

An alternative approach to the ranking of risky alternatives is based on stochastic dominance theory developed by Quirk and Saposnik (1962), where pointwise comparison between such alternatives is undertaken based on functions constructed from the complete set of possible outcomes (or distribution). Define $\mathcal{F}(X)$ as the distribution function of X, then:

$$F_x^k(r) = \begin{cases} F_x & k = 1 \\ \int\limits_{-\infty}^{r} F_x^{k-1}(t)\,dt & k \geq 2 \end{cases} \tag{4.3}$$

$\forall r \in \mathbb{R}$. Ranking of alternatives X and Y with distribution functions $\mathcal{F}(X)$ and $\mathcal{F}(Y)$ is such that X is preferred to Y with respect to the k^{th} order stochastic dominance[17] if and only if $\mathcal{F}_X^k(r) \leqslant \mathcal{F}_Y^k(r)$, with at least one strict inequality. There are important implications arising from this type of ranking. For example, $X \succ_1 Y \iff E[U(X)] \geq E[U(Y)]$ for every non-decreasing utility function U, which is the choice of rational investors. More importantly however, $X \succ_2 Y \iff E[U(X)] \geqslant E[U(Y)]$ for every non-decreasing and concave utility function U, which is the choice of every rational AND risk averse investor. While very appealing as a theory with sound choice criteria for making investment decisions, stochastic dominance relations are very difficult to apply in practice for more than a couple of outcomes. In addition, the theory requires the complete enumeration of the outcome space by the decision maker which may be infeasible in practice and may leave some prospects unranked. Nevertheless, a certain family of risk measures discussed in this section has a strong link with stochastic dominance which provides for an interesting and feasible alternative.

Whether a measure is coherent, spectral or meets certain stochastic dominance criteria does not establish it as better or worse as a tool for decision making. The output domain of these measures will certainly overlap at times depending for example on the underlying dynamics and state of the market. In the following subsections, I consider the properties and representations of 5 interesting and popular measures which are used in the empirical application of Section 4.4. The first 3 loosely belong to the general L^p

[17] Denoted as $X \succ_k Y$.

function space[18] and include the Absolute Deviation, Variance and Minimax measures, while the other 2 are the threshold based measures of Lower Partial Moments (LPM) and Conditional Value at Risk ($CVaR$).

4.2.1 Mean Variance ($\boldsymbol{EV}$)

Markowitz (1952) ushered in the era of modern portfolio management with the introduction of the Mean-Variance model of risk-return. Variance is a valid measure of risk for ranking preferences if either the investor exhibits quadratic utility (in which case it does not matter whether the underlying data is multivariate normal), or the underlying data is multivariate normal (in which case the utility of the investor is irrelevant since variance is the optimal choice). The optimization problem may be posed as the following NLP problem:

$$
\begin{aligned}
&\min_{w} \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} w_j \left(r_{i,j} - \mu_j \right) \right)^2 \\
&s.t. \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j \in \{1, \dots, m\}
\end{aligned}
\tag{4.5}
$$

where w represent the weights of the $j = 1, \dots, m$ assets, $i = 1, \dots, n$ are the number of periods or scenario points for the returns r and μ_j the forecast return. The problem effectively minimizes portfolio variance subject to the portfolio forecast return being equal to C, a full investment constraint and positivity constraints on the weights. While it is simple to express the problem in its quadratic form such that variance

[18] The L^p function space is defined as:

$$
\|e\|_p = \left(\sum_{j=1}^{m} |e_j|^p \right)^{1/p} \tag{4.4}
$$

with $p = 1$ representing the absolute (or Manhattan distance) measure, $p = 2$ the standard deviation (or Euclidean distance) where we can make use of variance instead because of the monotone transformation property, and $p = \infty$ represents the largest absolute value where we can represent the losses for Minimax optimization.

is equal to $w'\Sigma w$, I leave the problem in its more general NLP form which admits nonlinear constraints which would for example include long-short optimization with a leverage constraint.[19] Criticism of variance as a valid method for ranking portfolios is mainly aimed at the quadratic utility assumption which seems nothing more than a mathematical convenience rather than a reflection of reality, leading to the strange case of investors desiring less to more after a certain point on the utility curve, whilst the multivariate normality assumption is not usually borne out by the data. The symmetric nature of variance, penalizing both up and down deviations at the same rate was criticized quite early by Hanoch and Levy (1969)[20], while its lack of consistency with stochastic dominance relations should have effectively buried it as a method for portfolio allocation. However, its tractability and ease of use has made it a very popular choice, particularly for the modelling of monthly returns, with numerous extensions to provide for robustness and uncertainty mainly in the derivation of the covariance matrix. For example James and Stein (1956) provide for a shrinkage estimator, Black and Litterman (1992) a semi-Bayesian approach while Michaud (1989) a general criticism of the approach with a now patented alternative based on resampling methods.

4.2.2 Mean Absolute Deviation (*MAD*)

In the early days of computer programming, large scale quadratic problems were computationally more demanding to solve than linear problems. In light of this, Konno and Yamazaki (1991) introduced a piece-wise linear formulation of the absolute deviation function as an alternative to the Markowitz (1952) method. The standard NLP objective function may be formulated as:

$$\frac{1}{n}\sum_{i=1}^{n}\left|\sum_{j=1}^{m} w_j\left(r_{i,j}-\mu_j\right)\right| \tag{4.6}$$

[19] In the case of quadratic based constraints, the problem can also be posed as a second order cone (*SOCP*) problem.

[20] The criticism was in fact also aimed at any symmetric dispersion measure, not just variance.

which Konno and Yamazaki (1991) reduced to the following piece-wise linear problem:

$$\begin{aligned}
&\min_{w,d} \frac{1}{n} \sum_{i=1}^{n} d_i \\
&s.t. \\
&\sum_{j=1}^{m} (r_{i,j} - \mu_j)\, w_j \leq y_i, \forall i \in \{1, \dots, n\} \\
&\sum_{j=1}^{m} (r_{i,j} - \mu_j)\, w_j \geq -y_i, \forall i \in \{1, \dots, n\} \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j \in \{1, \dots, m\}
\end{aligned} \tag{4.7}$$

where d represent the absolute deviations of the portfolio from its forecast mean, forming a vector of variables of size n (length of the scenario) to be optimized. However, the constraints imposed to create the piece-wise linear function for the absolute deviation requires two $n \times n$ diagonal matrices stacked together[21] which may lead to computer memory problems for very large scenarios. This is in direct contrast to the EV model which only depends on the number of assets. Furthermore, while in the EV model deviations from the mean are penalized at an increasing rate arising from the square function, in the MAD model deviations are penalized at a linear rate which may not realistically represent the average investor. However, by not giving undue weight to the extreme observations, the MAD model may be more robust to possible misspecification in the dynamics from which the scenario was generated. Extensions to the model have included the addition of skewness in Konno, Shirakawa, and Yamazaki (1993), and semi-absolute deviation first suggested by Speranza (1993) who showed that the mean semi-deviation is a half of the mean absolute deviation from the mean. Similar to the EV model, the MAD model lacks consistency with stochastic dominance relations.

[21]Feinstein and Thapa (1993) provide for a reformulated representation with only one n diagonal matrix.

4.2.3 MiniMax

The MiniMax model of Young (1998), already introduced in an application in Section 3.5.4.2, aims to minimize the maximum loss, $\max\left(\sum_{j=1}^{m} -r_{i,j}w_j, \forall i = 1,\ldots,n\right)$ and as such is a very conservative criterion. It has a very simple LP formulation:

$$
\begin{aligned}
&\min_{M_p,w} M_p \\
&s.t. \\
&M_p - \sum_{j=1}^{m} w_j r_{i,j} \leq 0, \forall i = \{1,\ldots,n\} \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j \in \{1,\ldots,m\}
\end{aligned}
\tag{4.8}
$$

where M_p is the objective minimization value representing the maximum loss of the portfolio and guaranteed to be bounded from above by the maximum portfolio loss as a result of the first constraint. While Young (1998) only considered the problem in light of historical scenarios, there is no reason why r in the formulation may not represent a future simulated forecast scenario. Contrary to the MAD model, it only requires 1 additional variable and an $n \times 1$ additional constraint vector in the LP formulation, and as such does not pose any computational challenges even for very large problems. The Minimax principle is also consistent with expected utility theory at the limit based on a very risk averse decision maker, and a good approximation to the EV model when returns are multivariate Normal. Interestingly, the model is also a limiting case of the Conditional Value at Risk spectral risk measure described in the next sections.

4.2.4 Lower Partial Moments

The concept of penalizing deviations below a certain threshold at a different rate is at the heart of modern risk management and was already hinted at by Markowitz (1952) in a reference to semi-standard deviation. This was later formalized into a very general class of measures by Stone (1973), and the Lower Partial Moment (LPM) framework of Fishburn (1977) which, in the continuous case, may be defined as:

$$LPM_{a,\tau}(f) = \int_{-\infty}^{\tau} (\tau - x)^a f(x)\, dx \tag{4.9}$$

where a is some positive number which represents the rate at which deviations below the threshold τ are penalized and f some density function. In the discrete case, the function may be represented as:

$$LPM_{a,\tau}(x) = E\left[\max(\tau - x, 0)^a\right]. \tag{4.10}$$

Upper Partial Moments (UPM) are defined similarly. Usually, in the portfolio optimization context, the measure is standardized by raising it to the power of $\frac{1}{a}$. Fishburn (1977) derived a utility representation for this measure consistent with the von Neumann-Morgenstern axioms, and represented as:

$$U(x) = \begin{matrix} x - k(\tau - x)^{a_l} & x < \tau \\ x & x \geq \tau \end{matrix} \tag{4.11}$$

where k is a positive constant. Harlow and Rao (1989) describe an asset pricing model in the mean-lower partial moment framework (MLPM) and show that an MLPM framework is consistent with a very general set of utility functions. For example, the hyperbolic absolute risk aversion ($HARA$) class of utility functions is consistent with 1^{st}-degree LPM, whereas any risk averse utility function displaying skewness preference with positive first and third derivatives and negative second derivatives are consistent with 2^{nd}-degree LPM. In addition to this strong link with expected utility theory, Bawa and Lindenberg (1977), Bawa (1978) and Fishburn (1977) showed that stochastic dominance is equivalent to all degrees of n-degree LPM.
The portfolio optimization problem can be posed as follows:

$$\begin{aligned} &\min \left(\frac{1}{n} \sum_{i=1}^{n} \max \left(0, \tau - \left(\sum_{j=1}^{m} w_j r_{j,i} \right) \right)^a \right)^{1/a} \\ &s.t. \\ &\sum_{j=1}^{m} w_j \mu_j = C \\ &\sum_{j=1}^{m} w_j = 1 \\ &w_j \geq 0, \forall j \in \{1, \ldots, m\} \end{aligned} \tag{4.12}$$

Special cases are $a = 0$ representing the shortfall probability or Safety-First model of Roy (1952), $a = 1$ the below target shortfall and $a = 2$ the shortfall variance which is equivalent to the central semi-variance when $\tau = E(x)$. When $a = 1$, an LP formulation exists and is given by:

$$\begin{aligned}
&\min_{w} \frac{1}{n}\sum_{i=1}^{n} d_i \\
&s.t. \\
&\tau - \sum_{j=1}^{m} w_j r_{j,i} \leq d_i, \forall i \in \{1,\ldots,n\} \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j \in \{1,\ldots,m\} \\
&d_i \geq 0, \forall i \in \{1\ldots,n\}
\end{aligned} \tag{4.13}$$

For positive values of a other than 1, the discontinuous max function appears to pose some problems in the optimization strategy. Nawrocki and Staples (1989) devised a heuristic measure which approximates the function using only quadratic programming methods. Instead, I replace the max function with a smoothed approximation for which derivatives exist and discussed further in Section 4.3.2. With regards to the choice of threshold variable τ, the choice may be motivated by the investor's minimum acceptable return, some benchmark rate[22] or any other reasonable choice. A simple choice which makes use of the properties of this deviation measure is to use the mean of the portfolio which is equivalently equal to using a threshold of zero and passing a demeaned scenario matrix.[23]

[22]However, Brogan and Stidham(2005, 2008) have shown that for the linear separation property to hold, which assumes convexity of the mean-LPM space, the threshold must either be equal to the risk free rate or the mean of the portfolio.

[23]This is because the following relationship holds for LPM measures:

$$LPM_{\tau,a}(X) = LPM_{t+C,a}(X + C). \tag{4.14}$$

which is equivalent to property (D1) in Section 4.2 when accounting for the threshold parameter's shift by the constant C. Additionally, and with important implications in fractional programming, the LPM measure also has the scaling property so that:

$$LPM_{\tau,a}(X) = \frac{1}{b} LPM_{bt,a}(bX), \tag{4.15}$$

Because the linear reward function may be too restrictive in practise, Holthausen (1981) extended the LPM model to include a non-linear reward measure so that for $x \geq \tau$ in (4.11) the utility is then $U(x) = x + (x - \tau)^{a_u}$, where a_u is the power exponent for the upper partial moment, thus effectively capturing a range of linear and nonlinear utility curves (such as S-shaped and inverse S-shaped) with reference to gains and losses as illustrated in the example in Figure 4.1. Farinelli and Tibiletti (2008) also proposed

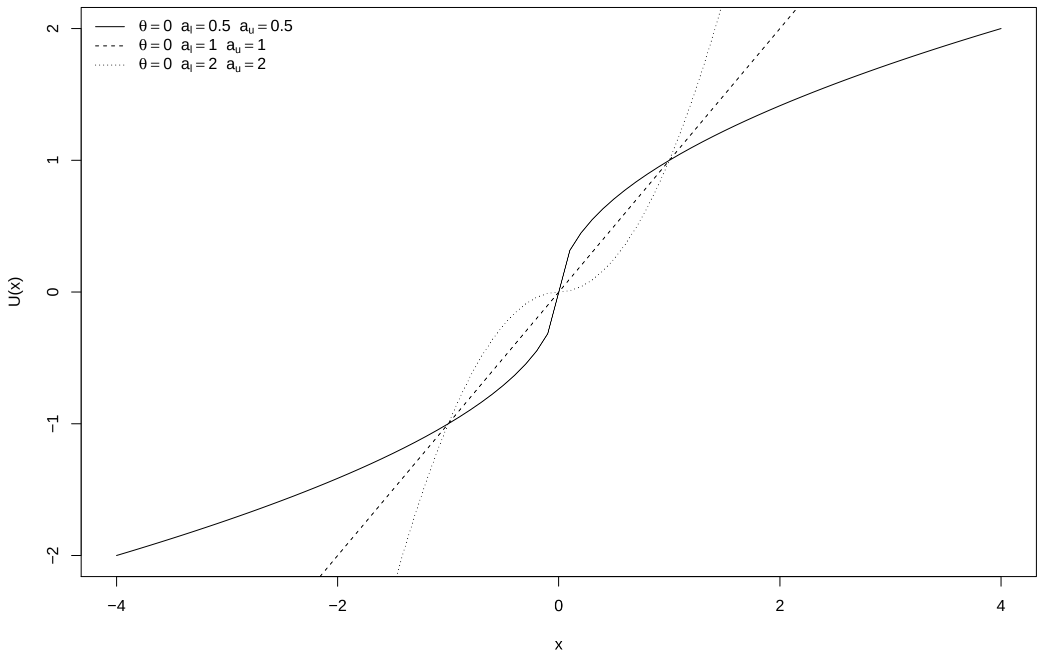


FIGURE 4.1: Upper to Lower Partial Moment Utility

an Upper to Lower partial moment portfolio optimization approach, but surprisingly failed to reference the original extension of Holthausen (1981)[24], nor did they mention

where it is understood that for the non-standardized version of the measure, i.e. when not raised to the power of $\frac{1}{a}$, the measure is multiplied not by $\frac{1}{b}$ but $\frac{1}{b^a}$ instead.

[24] In addition to this glaring omission, they then proceeded in subsequent papers to name the measure after themselves as the Tibilleti-Farinelli ratio. A fashion for naming ratios derived from the LPM framework can be seen in the Sortino ratio which is the return below target divided by the target semi-deviation described in Sortino and Price (1994), the Sortino-Satchell ratio for the return below target divided by the n-degree LPM described in Sortino and Satchell (2001) and the Omega Ratio which is the Upper to Lower Partial Moment of degree 1 described in Keating and Shadwick (2002), and though not named after the authors uses the last letter of the Greek alphabet in an obvious statement on the finality of the measure for performance measurement after which no other measures need be defined!

that such a measure is non-convex meaning that one has to resort to global optimization approaches to obtain a solution.[25]

4.2.5 Conditional Value at Risk and Spectral Measures

Since the report by the Group of Thirty G30 (1993), the use of Value at Risk (*VaR*) is almost universal among banks, trading desks and other financial entities as a key measure for measuring and managing risk. Despite its popularity, it has come under growing pressure as a non coherent (it lacks the subadditivity property) and inadequate measure, or an imperfect measure which has been incorrectly used, overused, abused and over-relied upon. An alternative measure, based on the average loss conditional on the VaR being violated is called Conditional Value at Risk (*CVaR*)[26] which is a coherent and convex risk measure belonging to the class of spectral risk measures of Acerbi and Tasche (2002). Formally, a spectral risk measure M_ψ is a weighted average of the loss distribution quantile q evaluated at p, such that:

$$M_\psi = \int_0^1 \psi(p)\, q_p dp \tag{4.16}$$

where $\psi(p)$ is a weighting function defined over the full range of probabilities $p \in [0, 1]$ and restricted to be non-negative, normalized to sum to 1, and increasing or constant in p (such that higher losses have equal or higher weights to lower losses). VaR is clearly a spectral risk measure with weighting the dirac delta function which is degenerate, while CVaR is based on a step function (constant weight for losses greater than VaR). Cotter and Dowd (2006) investigated alternative weighting functions to account for truly risk averse behavior by considering strictly increasing weight functions in an application for establishing futures clearinghouse margin requirements. While they found that such weighting schemes were superior to the standard CVaR, Dowd, Cotter, and Sorwar (2008) also found some problems in their implementation both in the choice of functions as well as the mixing properties of these measures with respect to nonlinear weighting functions. In a different direction Rockafellar, Uryasev, and Zabarankin (2006) considered the so called mixed-CVaR problem whereby it is possible to mix together CVaR

[25] The paper also alludes to the ease of estimation of some other measures of risk-reward such as the upper to lower CVaR which were shown in other papers to require a Mixed Integer approach which is by no means simple nor computationally feasible for a large number of assets.

[26] Also called Expected Tail Loss with distinctions in the names sometimes denoting differences for the continuous and sample cases, with the latter requiring a specialized representation in order to be deemed convex according to Rockafellar and Uryasev (2000).

at different coverage rates using a weighting function, and established the relationship between this and the spectral risk representation. In terms of the general optimization problem, CVaR may be represented as an NLP minimization problem with objective function given by:

$$\min_{w,v} \frac{1}{na} \sum_{i=1}^{n} \left[\max \left(0, v - \sum_{j=1}^{m} w_j r_{i,j} \right) \right] - v \tag{4.17}$$

where v is the a-quantile of the distribution. For a discrete scenario, this can be represented using auxiliary variables as the following LP problem (due to Rockafellar and Uryasev (2000)):

$$\begin{aligned}
&\min_{w,d,v} \frac{1}{na} \sum_{i=1}^{n} d_i + v \\
&s.t. \\
&\sum_{j=1}^{m} w_j r_{i,j} + v \geq -d_i, \forall i \in \{1, ..., n\} \\
&\sum_{j=1}^{m} w_j \mu_j = C \\
&\sum_{j=1}^{m} w_j = 1 \\
&w_j \geq 0, \forall j \in \{1, \ldots, m\} \\
&d_i \geq 0, \forall i \in \{1, \ldots, n\}
\end{aligned} \tag{4.18}$$

where v represents the VaR at the a-coverage rate and d_i the deviations below the VaR. The formulation presented here is in such a way as to represent the asset returns scenario matrix rather than the more typical loss representation in the literature.

Direct extensions have followed in the same vein as the LPM measures with Biglova, Ortobelli, Rachev, and Stoyanov (2004) proposing the Rachev Ratio as the upper to lower CVaR for which a mixed integer representation is provided in Stoyanov, Rachev, and Fabozzi (2007) (modified by Konno, Tanaka, and Yamamoto (2011) for cases when the returns are completely distributed on the positive side), and also proposed in the same paper the Generalized Rachev Ratio which is the Rachev Ratio but with the numerator and denominator raised to different powers representing different penalization to gains and losses beyond some upper and lower quantiles. Unfortunately, this generalization, like the upper to lower LPM has both a convex numerator and denominator[27]

[27] Technically, both risk and reward CVaR functions are convex for values of the power ≥ 1.

making it non quasi-convex and hence necessitating a global optimization approach.[28] The empirical application in Biglova, Ortobelli, Rachev, and Stoyanov (2004) compared a range of measures and concluded that based on the terminal wealth reached by the Rachev and Generalized Rachev Ratios using a set of 9 German stocks optimized during the period 1999-2003, that they are "...good criteria for portfolio optimization because they yield the best performance for investors in the period examined". It is quite surprising how many authors use terminal wealth to make inference about the performance of their measures given that is it sensitive to the starting period and does not really take into account path riskiness which is the whole point of these measures.

4.3 Applied Optimization

The previous section outlined a number of popular risk-return measures for portfolio allocation, setup up as risk minimization problems subject to some specified reward constraint. Traditionally, at least in the finance literature, the method for choosing the optimal risk-reward portfolio was to trace out the efficient frontier and then search for the highest return-risk combination along that frontier. Additionally, assuming the presence of a risk-free asset, the portfolio optimization was usually performed on the excess forecast returns so that the tangency portfolio would intersect the 45° line running from the origin, with clear links to the classic CAPM framework and two fund separation theorem. The following section discusses the reformulation of the optimal risk-reward ratio programming based on the well established literature on fractional programming. As regards the risk free rate, at no point in this exercise is this used ex-ante to generate excess forecast returns. The reason is quite simply that the AR(1) forecast, used as the model for the conditional mean, is not likely to capture very well the size of the ex-post return, but it is expected that at least it will have some better properties with respect to the sign of the forecast if not the relative size of the forecast across assets. Subtracting the risk free rate, particularly at times like the early 1980's when the annualized 1 Month Treasury Bill rate was almost 20% would have likely changed the sign of the weekly forecast and excluded it from consideration in the optimal set. Furthermore, modelling the excess returns directly when the risk free is not constant but somewhat stochastic creates a host of other model consistency

[28] The mixed-integer approach for the Rachev Ratio is just as difficult to estimate as it is limited by the size of the scenario which determines the number of binary variables required.

problems. Therefore, all the modelling, forecasting and optimization is carried out without reference to any risk free rate, though the ex-post performance measures do account for it as does any portfolio accounting requirements described in more detail in Section 4.4.4.

4.3.1 Fractional Programming and Optimal Risk-Reward Portfolios

Consider the general nonlinear problem of minimizing a risk to reward problem represented as a fraction:

$$
\begin{aligned}
&\min_{\mathbf{w}} \frac{\rho_{risk}\left(\mathbf{Rw}\right)}{\rho_{reward}\left(\mathbf{Rw}\right)} \\
&\mathbf{w}'\mathbf{1} = 1 \\
&\mathbf{L} \leq \mathbf{Aw} \leq \mathbf{U} \\
&\mathbf{w} \geq 0
\end{aligned} \tag{4.19}
$$

where $\mathbf{w}$ is an $m \times 1$ vector of weights, $\mathbf{R}$ the $n \times m$ Scenario matrix of returns so that the risk (ρ_{risk}) and reward (ρ_{reward}) functions are applied on the weighted scenario returns, $\mathbf{1}$ an $m \times 1$ vector of ones and $\mathbf{A}$ a $q \times m$ matrix of linear constraints with lower and upper bounds given by $\mathbf{L}$ and $\mathbf{U}$ respectively. The key developments in the theory of fractional programming were provided in the linear case by Charnes and Cooper (1962), while for nonlinear cases the main contributions can be traced to Dinkelbach (1967) and Schaible(1976a, 1976b). More recently, Stoyanov, Rachev, and Fabozzi (2007) provided a more focused review of fractional programming with reference to portfolio optimization. Under the assumption that both numerator and denominator are positive homogeneous, the problem in (4.19) can be transformed into the following simpler nonlinear fractional programming ($NLFP$) problem:

$$
\begin{aligned}
&\min_{\hat{\mathbf{w}},\upsilon} \rho_{risk}\left(\mathbf{R}\hat{\mathbf{w}}\right) \\
&\rho_{reward}\left(\mathbf{R}\hat{\mathbf{w}}\right) \geq 1 \\
&\hat{\mathbf{w}}'\mathbf{1} = \upsilon \\
&\upsilon\mathbf{L} \leq \mathbf{A}\hat{\mathbf{w}} \leq \upsilon\mathbf{U} \\
&\upsilon \geq 0
\end{aligned} \tag{4.20}
$$

where υ represents a scalar auxilliary scaling variable and $\hat{\mathbf{w}}$ the unconstrained optimal weight vector such that the optimal weight vector $\mathbf{w} = \frac{\hat{\mathbf{w}}}{\upsilon}$. In order for this problem to be convex, the reward function must be concave and the risk function convex, with strict positivity required for both functions.[29] Different relaxations of these basic conditions lead to different classes of problems in the literature, some with unique solutions and others requiring global search methods. Finally, these simple conditions admit both convex risk and deviation measures as defined in Section 4.2. For the purpose of the empirical exercise, the reward measure considered is linear, representing the weighted expected forecast return, whilst the risk/deviation measure for the long only portfolios is either linear or quadratic. The NLFP formulation of all the models used and their derivatives is given in Appendix D.

4.3.2 Smooth Approximations to Discontinuous Functions

While it is preferable to work with an LP formulation of a decision problem, there are certain situations where this poses certain challenges. First, for some LP problems, the dimension of the dataset and constraints may tax the limits of computer memory. Consider for example the MAD model presented in Section 4.2.2 which requires a constraint matrix of size $2n \times m$ in order to create the piecewise linear representation for the absolute value, where for large scenarios (n) and assets (m) memory considerations become important. Second, in practice, many problems and/or constraints simply cannot be expressed in LP form necessitating the use of either QP or NLP based methods. In that case, it is always preferable to have analytic derivatives of the function and constraints, for speed and accuracy versus numerical evaluation methods. Interestingly, some problems, while convex are discontinuous because of the presence of such functions as the min and abs. For these problems, an approximation may be obtained by considering smooth and continuous versions of these functions. Consider for example the CVaR and LPM measures, both of which depend on the max function, for which the following smooth approximation, s_{max} may be used:

$$\max(x,0) \approx s_{\max}(x,0) = \frac{\left(\sqrt{x^2+\varepsilon}+x\right)}{2} \tag{4.21}$$

[29]For the reward function the requirement is a little more relaxed in that there must be at least some combination of the weights and returns for which the reward is positive. Additionally, for a linear reward function the constraint becomes an equality.

where ε is some very small positive number controlling the degree of approximation error. The absolute value may also be approximated with the following function s_{abs}:

$$abs\left(x\right) \approx s_{abs}\left(x\right) = \sqrt{\left(x+\varepsilon\right)^2} \tag{4.22}$$

although alternatives also exist[30]. Apart from allowing the MAD problem to be represented in NLP form with a smooth function, it also allows for the use of short positions, replacing the full investment constraint with a leverage constraint (the absolute sum of positions)[31], without resorting to such methods as described in Jacobs, Levy, and Markowitz (2006) which double the size of the problem and require certain very specific assumptions about the 'trimability' of the portfolio. Finally, for the case of the minimax problem, I make use of the generalized mean function $M_p\left(x_1,\ldots,x_n\right) = \left(\frac{1}{n}\sum\limits_{i=1}^{n} x_i^p\right)^{1/p}$, which approximates the maximum of a set of positive values as $p \to \infty$. In order to obtain the maximum loss for use in the NLP minimax optimization function, I combine this function with the s_{max} function defined in (4.21) applied to the negative of the scenario returns:

$$\left(\frac{1}{n}\sum_{i=1}^{n} s_{\max}\left(-\mathbf{w}'\mathbf{r}_i,0\right)^p\right)^{1/p}. \tag{4.23}$$

In practise, the optimization problem needs to be calibrated for p since very large values will exceed the limits of even 64-bit architectures.[32] To gauge how well these smooth approximations work in practise, Tables 4.1 and 4.2 report the optimal weights and timings of a number of models optimized under standard LP risk measures and their smooth NLP counterparts.[33] In terms of numerical accuracy, the average mean squared error (*MSE*) across all models is 1.4E-4, and when excluding the MiniMax model the average falls to 5.0E-6. The numerical accuracy can be controlled even more tightly by changing ε in Equations 4.21 and 4.22 and the termination criteria of the NLP solver. As was expected, the highest loss in numerical accuracy is with the MiniMax model with an average MSE of 5.0E-5. The timings are also of particular interest with the smooth NLP MAD models being on average 11.5× faster than their LP counterparts,

[30] One such alternative is: $\left(2x/\pi\right)\left(\tan^{-1}\left(ox\right)\right)$, where o is some very large positive number.

[31] A common mistake is to keep the full investment constraint instead of replacing it with the leverage constraint, which makes no sense even when controlling for individual position limits.

[32] The largest number which can be represented is about $1.7976e+308$.

[33] The optimization methods and models used in this chapter are available to use in the Portfolio Allocation and Risk Management Applications (*parma*) package of Ghalanos (2012a). See Appendix F for details. The dataset used here is one comprised of 15 ETFs and taken from the *parma* package from which this example is reproduced.

and between 3.4× and 1.4× for the rest of the models. Not shown is the relative memory usage where the MAD model for instance required 10× more random access memory in LP form than their smooth NLP counterpart. This is likely to be quite important for models with larger dimensions.

TABLE 4.1: LP vs smooth approximations to NLP: (CVaR and LPM)

Panel A: CVaR								
	GO-GARCH (maNIG)		DCC (MVN)		DCC-Copula(MVT)		VAR	
	LP	NLP	LP	NLP	LP	NLP	LP	NLP
IWD	0.0415	0.0416						
TLT	0.3678	0.3678	0.3119	0.3119	0.3145	0.3145	0.3036	0.3025
EWC	0.0907	0.0906	0.1881	0.1881	0.1855	0.1855		
EWL	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
EWJ							0.0757	0.0764
EWQ							0.0119	0.0128
EZA							0.1088	0.1083
Risk/Reward	6.0682	6.0682	6.6554	6.6554	6.4094	6.4094	3.8895	3.8895
Elapsed(sec)	2.9932	2.8799	2.5986	1.0049	2.7549	0.6299	3.1299	2.6006
Panel B: LPM								
	GO-GARCH (maNIG)		DCC (MVN)		DCC-Copula(MVT)		VAR	
	LP	NLP	LP	NLP	LP	NLP	LP	NLP
IWD	0.0661	0.054	0.3531	0.3531	0.343	0.343		
TLT	0.3748	0.3802	0.1469	0.1469	0.1570	0.1570	0.2816	0.2816
EWC	0.0591	0.0657	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
EWL	0.5000	0.5000					0.1061	0.1061
EWJ								
EWQ								
EZA							0.1124	0.1124
EPP								
Risk/Reward	1.2888	1.2889	1.4908	1.4908	1.4080	1.4080	0.9310	0.9310
Elapsed(sec)	2.4893	0.9111	2.3486	0.5049	2.4110	0.5830	2.2705	0.9893

Note: The Table reports the weights, risk to return ratio and timing (in seconds) of portfolios optimized under the optimal fractional programming model using for the CVaR and LPM risk measures using either an LP or smooth approximation NLP, for 1-ahead scenarios generated under 4 different models. The dataset used was comprised of the daily log returns of 15 Exchange Traded Funds for the period 28/05/2003 to 01/06/2012 and available in the *parma* package on R-Forge (see Appendix E).

4.4 Empirical Application

The DJIA index is perhaps the oldest and most cited benchmark index in the US and around the world. Its 30 blue chip constituents are among the most recognized names in the stock market and despite its price weighting methodology, it has mostly matched the performance of the value weighted S&P500. Obviously, as a quoted benchmark which is meant to be followed by both insiders and the general public, it is unlikely to be constructed by complex methods which are difficult to comprehend. Such broad market popular indices are likely to be based on simplified, non-optimal methods as the companies creating them need to consider the factors which will lead to a general acceptance of their product. Unfortunately, once a benchmark is created, accepted and

Table 4.2: LP vs smooth approximations to NLP: (MAD and Minimax)

Panel A: MAD

	GO-GARCH (maNIG)		DCC (MVN)		DCC-Copula(MVT)		VAR	
	LP	**NLP**	**LP**	**NLP**	**LP**	**NLP**	**LP**	**NLP**
IWD	0.0693	0.0692						
TLT	0.3785	0.3785	0.3528	0.3528	0.3331	0.3331	0.2609	0.2609
EWC	0.0522	0.0523	0.1472	0.1472	0.1669	0.1669		
EWL	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
EWJ							0.1155	0.1155
EWQ								
EZA							0.1236	0.1236
Risk/Reward	2.6823	2.6823	3.0807	3.0807	2.9403	2.9403	2.0405	2.0405
Elapsed(sec)	3.8643	0.3643	3.6299	0.3486	2.8955	0.2080	2.8330	0.2549

Panel B: MiniMax

	GO-GARCH (maNIG)		DCC (MVN)		DCC-Copula(MVT)		VAR	
	LP	**NLP**	**LP**	**NLP**	**LP**	**NLP**	**LP**	**NLP**
IWD			0.3428	0.3697	0.2627	0.2075		
TLT	0.5000	0.5000	0.0126	0.0177	0.2548	0.3064	0.3587	0.3576
EWC			0.0444	0.0340	0.0315	0.0085		
EWL	0.5000	0.5000	0.5000	0.5000	0.4510	0.4775	0.5000	0.5000
EWJ			0.0250	0.0199				
EWQ							0.1164	0.1138
EZA							0.0249	0.0287
EPP			0.0752	0.0587				
Risk/Reward	12.8921	12.8921	11.5635	11.6046	10.4623	10.4884	6.5808	6.5898
Elapsed(sec)	0.5488	0.2549	0.5361	0.7236	0.5361	0.5830	0.5674	0.3018

Note: The Table reports the weights, risk to return ratio and timing (in seconds) of portfolios optimized under the optimal fractional programming model using for the MAD and Minimax risk measures using either an LP or smooth approximation NLP, for 1-ahead scenarios generated under 4 different models. The dataset used was comprised of the daily log returns of 15 Exchange Traded Funds for the period 28/05/2003 to 01/06/2012 and available in the *parma* package on R-Forge (see Appendix E).

popularized by the media, it is difficult to change despite any advances in knowledge or computational power. So engrained in the popular culture and psyche can such benchmarks become that they monopolize the space, making it very difficult for any alternative benchmarks to gain a foothold. The application in this section aims to show that the current bar on beating the DJIA or for that matter the S&P500 is set quite low as evidenced by the performance of a number of models and measures which have been considered throughout this thesis. Given the failure to reject tests of the presence of conditional correlation, GARCH effects and fat tails, it is quite possible to create a dynamic allocation strategy which takes these into account and provide superior out of sample performance. There are multiple implications. First, there is certainly value in active investing when the benchmark is sub-optimally constructed. Second, investors should reconsider the reward of managers who benchmark against such indices. Third, investors should reconsider the passive tracking of such indices. This implies that there could be value in constructing more optimally weighted indices.

4.4.1 Data Description and Characteristics

Weekly total returns data, based on the Friday close price, of the point in time constituents of the DJIA was collected from CRSP, for the period 1960 to 2010. The decision to use weekly data was determined by the need to have an adequate historical length for the modelling process, and at the same time a reasonable forecast window. It would have made little sense to use daily data within a static portfolio optimization setting[34], despite the extra window length it would have provided for the modelling, as this would involve costly daily re-balancing, and in any case, the noise in daily data is likely to negatively impact forecasts. Monthly data, while preferable from a re-balancing point of view, does not provide enough historical information to be of much use in the modelling process, and given the type of security dynamics present, would likely be sub-optimal as large swings in prices would invalidate any average monthly density forecast.

Starting on 06/01/1975 and using all available data since 1960[35], a number of models were fitted and the 1-week ahead forecast scenario of size 7000×30 generated which was used in the portfolio optimization. The process was repeated by moving the data window 1 week ahead and taking into account any constituent changes in the Dow membership, until the last forecast scenario, for 03/01/2011, was created for a total of 1878 forecasts. While the estimation and allocation of the models was always done on a Friday, the actual formation of the portfolios was carried out on the following Monday or nearest trading day after that, so as to realistically allow for a lag between the estimation of the models, requiring the closing price of Friday, and the formation which was executed at the closing price of Monday or nearest trading day after that. Table 4.3 shows the Dow constituents and their CRSP PERMNO[36] as of 01/06/1959, and the changes to the index following the original set which was used in this application. The start of 1976 is significant as it marks the end of a painful energy crisis of 1973 and the 1973-1975 recession[37], and the start of a series of more frequent changes to the Dow, which until that date was quite stable in that no changes had occurred since 1959. In fact, this marks a shift from an Industrials based index to one more reflective

[34] As opposed to a dynamic stochastic optimization method discussed in Section 4.1

[35] The data was truncated to start on the first commonly available observation of the entire dataset of 30 members.

[36] This is a unique permanent security identification number assigned by CRSP to each security, which does not change following name or capital structure changes, nor delisting, thus making is possible to track a company's complete trading history.

[37] Source: NBER

of the changes in the US economic landscape with the rise in importance of Technology, Pharmaceutical, Banking/Credit and Entertainment industries. Thus while the calculation of the Dow may be completely inefficient, there is certainly value in the selection of the constituents, representing a small, manageable cross section of large blue-chip companies of the US economy.

TABLE 4.3: Dow Jones Industrial constituent changes since 1959

01/06/1959	PERMNO
Allied Chemical	10145
American Tobacco	10225
American Can	10241
American Telephone & Telegraph	10401
Anaconda Copper	10495
Bethlehem Steel	10786
Chrysler	11260
DuPont	11703
Eastman Kodak Company	11754
Standard Oil (NJ)	11850
General Electric Company	12060
General Motors Corporation	12079
International Harvester	12503
International Nickel	12546
Owens-Illinois Glass	13661
Sears Roebuck & Company	14322
Standard Oil of California	14541
Texaco Incorporated	14736
U.S. Steel	15069
Westinghouse Electric	15368
Woolworth	15456
Union Carbide	15659
General Foods	16109
Goodyear	16432
Johns-Manville	16707
United Aircraft	17830
Procter & Gamble Company	18163
Swift & Company	19713
International Paper Company	21573
Aluminum Company of America	24643

09/08/1976	PERMNO
(-) Anaconda Copper	10495
(-) Standard Oil (NJ)	11850
(-) United Aircraft	17830
(-) International Nickel	12546
(+) Inco	12546
(+) Esmark	19713
(+) Minnesota Mining & Manufacturing	22592
(+) Exxon Corporation	11850
29/06/1979	
(-) Chrysler	11260
(-) Esmark	19713
(+) International Business Machines	12490
(+) Merck & Company, Inc.	22752
30/08/1982	
(-) Johns-Manville	16707
(+) American Express Company	59176
30/10/1985	
(-) Allied Chemical	10145
(-) American Tobacco	10225
(-) Standard Oil of California	14541
(-) General Foods	16109
(+) AlliedSignal Incorporated	10145
(+) Philip Morris Companies Inc.	13901
(+) Chevron	14541
(+) McDonaldŠs Corporation	43449
12/03/1987	
(-) Owens-Illinois Glass	13661
(-) U.S. Steel	15069
(-) Inco	12546
(-) International Harvester	12503
(+) Boeing Company	19561
(+) USX Corporation (formerly U.S. Steel)	15069
(+) Coca-Cola Company	11308
(+) Navistar International Corporation	12503

06/05/1991	PERMNO
(-) American Can	10241
(-) Navistar International Corporation	12503
(-) USX Corporation	15069
(+) Caterpillar Incorporated	18542
(+) Walt Disney Company	26403
(+) J.P. Morgan & Company	48071
17/03/1997	
(-) American Telephone & Telegraph	10401
(-) Bethlehem Steel	10786
(-) Texaco Incorporated	14736
(-) Westinghouse Electric	15368
(-) Woolworth	15456
(+) AT&T Incorporated	10401
(+) Johnson & Johnson	22111
(+) Hewlett-Packard Company	27828
(+) Wal-Mart Stores Incorporated	55976
(+) Travelers Group	70519
01/11/1999	
(-) Sears Roebuck & Company	14322
(-) Chevron	14541
(-) Union Carbide	15659
(-) Goodyear	16432
(-) Travelers Group	70519
(-) Aluminum Company of America	24643
(+) Microsoft Corporation	10107
(+) Alcoa Incorporated	24643
(+) Intel Corporation	59328
(+) SBC Communications Incorporated	66093
(+) Home Depot Incorporated	66181
(+) Citigroup Incorporated	70519

27/01/2003	PERMNO
(-) AlliedSignal Incorporated	10145
(-) Philip Morris Companies Inc.	13901
(-) Minnesota Mining & Manufacturing	22592
(-) J.P. Morgan & Company	48071
(+) Altria Group Incorporated	13901
(+) Honeywell International	18374
(+) 3M Company	22592
(+) J.P. Morgan Chase	47896
08/04/2004	
(-) AT&T Incorporated	10401
(-) Eastman Kodak Company	11754
(-) International Paper Company	21573
(+) Pfizer Incorporated	21936
(+) American International Group Inc.	66800
(+) Verizon Communications Inc.	65875
21/11/2005	
(-) SBC Communications Incorporated	66093
(+) AT&T Incorporated	10401
19/02/2008	
(-) Altria Group Incorporated	13901
(-) Honeywell International	18374
(+) Chevron	14541
(+) Bank of America Corporation	59408
22/09/2008	
(-) American International Group Inc.	66800
(+) Kraft Foods Inc.	89006
08/06/2009	
(-) General Motors Corporation	12079
(-) Citigroup Incorporated	70519
(+) Travelers Companies	59459
(+) Cisco Systems, Inc.	76076

Note: The Table lists the 30 constituent companies of the DJIA as at 06/01/1959 together with their CRSP PERMNO's, and changes to the index since. A (-) before a name indicates a removal while a (+) indicates an addition to the index. These are not necessarily additions or deletions, but may represent mergers, change of name or other corporate actions which necessitated some type of change, which may be inferred from the PERMNO.

In order to check the accuracy of the collected data and constituent changes, and making use of the changes to the Dow Divisor[38], Figure 4.2 shows the overlapping plots of the actual Dow and the Dow recreated using this dataset, constituent change tables and divisor changes.

To gauge the performance of the various models and risk measures against the benchmark indices, I consider in this application the total return DJIA index ($DJIA_{PW}$) for which data only exists back to 1987, an equally weighted total return version of the DJIA ($DJIA_{EW}$) which starts in 1975 (the same date as the empirical application), and the total return S&P500 ($S\&P500_{VW}$) with dates starting in both 1975 and 1987 to allow comparison with the 2 DJIA indices. For the risk free rate, I use a proxy based on the 1 Month Treasury Bill (*TB1M*) where the daily return on this instrument is a simple daily rate that compounds, over the number of trading days in a month, to the TB1M rate.

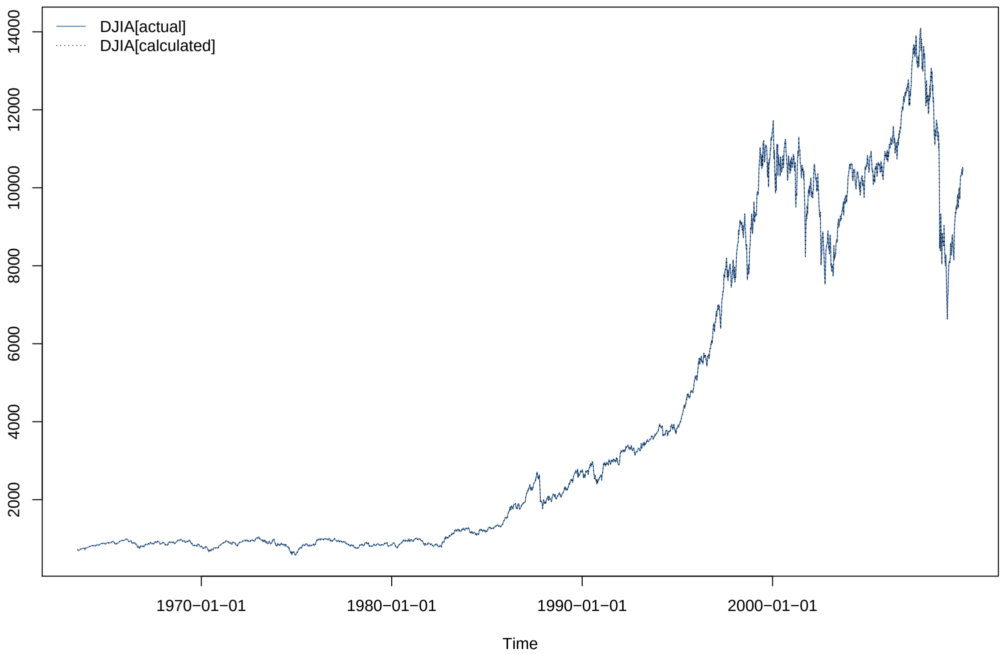


FIGURE 4.2: Actual vs Calculated DJIA

[38]The Dow Divisor adjusts the index for stock splits and other related changes in the index, with a current value as of 02/07/2010 of 0.132129493.

4.4.1.1 ARCH Effects

Table 4.4 reports the min, median and max p-values of the ARCH LM test of Engle (1982) under the null of no ARCH effects using 1 lag, for each constituent which was part of the Dow during the 1878 moving window estimations covering the period 1975-2010. The test was carried out on the standardized squared residuals from a constant-AR(1) model. With the exception of Kraft Food (CRSP PERMNO 89006), the null of no ARCH can safely be rejected for all securities across all estimation windows.

TABLE 4.4: Dow Jones Industrial constituents ARCH effects

ARCH LM Test [p-values]							
PERMNO	**Min**	**Median**	**Max**	**PERMNO**	**Min**	**Median**	**Max**
10107	0.0001	0.0030	0.1142	**17830**	0.0000	0.0387	0.9995
10145	0.0000	0.0005	0.0393	**18163**	0.0000	0.0015	0.5048
10225	0.0002	0.0011	0.0314	**18542**	0.0000	0.3817	0.5552
10241	0.0000	0.0000	0.0001	**19561**	0.0025	0.0260	0.7334
10401	0.0000	0.0192	0.9694	**19713**	0.0001	0.0004	0.0005
10495	0.0000	0.0048	0.0188	**21573**	0.0069	0.1118	0.7644
10786	0.0000	0.1332	0.9785	**21936**	0.0000	0.0001	0.0124
11260	0.0000	0.0000	0.0002	**22111**	0.0000	0.0000	0.0009
11308	0.0000	0.0013	0.0643	**22592**	0.0000	0.0000	0.3609
11703	0.0000	0.0000	0.0989	**22752**	0.0000	0.0000	0.3157
11754	0.0000	0.0001	0.6782	**24643**	0.0000	0.0000	0.5767
11850	0.0000	0.0003	0.0151	**26403**	0.0000	0.0000	0.6884
12060	0.0000	0.0000	0.0002	**27828**	0.0000	0.0002	0.0458
12079	0.0000	0.0000	0.7233	**43449**	0.0000	0.0051	0.9993
12490	0.0000	0.0000	0.9398	**47896**	0.0000	0.0001	0.0125
12503	0.0000	0.0000	0.0000	**48071**	0.0000	0.0007	0.0317
12546	0.0002	0.0354	0.9571	**55976**	0.0000	0.0000	0.1949
13661	0.0000	0.0004	0.3942	**59176**	0.0000	0.0000	0.0027
13901	0.0000	0.0007	0.0190	**59328**	0.0000	0.0001	0.9984
14322	0.0000	0.0000	0.0069	**59408**	0.0000	0.0321	0.1123
14541	0.0000	0.0014	0.1865	**59459**	0.0000	0.0000	0.0000
14736	0.0000	0.0000	0.0000	**65875**	0.0000	0.0000	0.0515
15069	0.0000	0.0590	0.9999	**66093**	0.0278	0.0581	0.5478
15368	0.0000	0.0000	0.0008	**66181**	0.0004	0.0039	0.1088
15456	0.0000	0.0102	0.0554	**66800**	0.0000	0.0000	0.0028
15659	0.0000	0.1397	0.8278	**70519**	0.0000	0.0002	0.0704
16109	0.0036	0.0164	0.9904	**76076**	0.0009	0.0014	0.0029
16432	0.0018	0.0228	0.9996	**89006**	0.4196	0.4872	0.8809
16707	0.0000	0.0000	0.0000				

Note: The Table reports the min, median and max p-value from the ARCH LM test of Engle (1982) under the null of no ARCH, of the constituents of the DJIA during the period 1975 to 2010 based on their CRSP PERMNO. The test was applied to the estimates of the squared standardized residuals from a constant-AR(1) model, at each period and for each constituent of the index under the moving estimation scheme described in Section 4.4.1.

4.4.1.2 Multivariate Normality

To assess the degree of departure from multivariate Normality in the weekly dataset I consider the test of Mardia (1980) which considers measures of multivariate skewness and kurtosis based on the Mahalanobis distance metric. Formally, consider the random sample $Y = (y_1, y_2, \ldots, y_n)$ of size N from a d-variate distribution. The multivariate skewness (b_{1d}) and kurtosis (b_{2d}) are defined in Mardia (1970) as:

$$b_{1d} = N^{-2} \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}^3$$
$$b_{2d} = N^{-1} \sum_{i=1}^{N} r_i^4 \tag{4.24}$$

where r_{ij} is the Mahalanobis angle between vectors y_i and y_j and defined as:

$$(y_i - E(Y))' \Sigma^{-1} (y_j - E(Y)), \tag{4.25}$$

with Σ being the covariance matrix of Y. The limiting distribution of $Nb_{id}/6$ is χ^2 with $d(d+1)(d+2)/6$ d.o.f, while for $\sqrt{N}(b_{2d} - d(d+2))(8d(d+2))^{-1/2}$ the limiting distribution is Normal. Under the assumption of multivariate Normality, the data should not have significant skewness or kurtosis. The first part of Table 4.5 reports the average test statistics for Mardia's kurtosis and skewness across the 1878 rolling estimation windows, encompassing the different point in time Dow 30 constituents, with values of 13550 and 112 and p-values of 0 and 0 respectively, indicating a clear departure from multivariate normality.

TABLE 4.5: Dow Jones Industrial dataset multivariate characteristics

Multivariate Normality Test (Mardia)			
	Min	**Median**	**Max**
b_{1d}	8523	13190	28710
b_{2d}	41.23	110.9	211.8

DCC Test (Engle)			
$1975-2010$	**Min**	**Median**	**Max**
Stat	0.69	8.08	26.50
$[p-value]$	[0.706]	[0.0176]	[0.000]
$1982-1985$			
Stat	0.89	1.20	3.04
$[p-value]$	[0.640]	[0.549]	[0.218]
$1997-2000$			
Stat	0.69	1.51	3.58
$[p-value]$	[0.708]	[0.470]	[0.167]

Note: The Table show average statistics for the point in time DJIA dataset for the period 1975 to 2010. The first table shows the multivariate normality test of Mardia (1970) based on multivariate skewness (b_{1d}) and kurtosis (b_{2d}) both of which are well above their critical values for each period tested. The second table shows the test of constant correlation of Engle and Sheppard (2001) which was based on estimating for each period a constant correlation model (CCC) and testing this under the null of no dynamic correlation. With the exception of the 2 periods given ($1982-1985$ and $1997-2000$), the test rejected the null at the 10% confidence level for the remainder of the periods tested.

4.4.1.3 Constant Correlation

Engle and Sheppard (2001) devised a method to test the assumption of constant correlation, with the null hypothesis $H_0 : R_t = \bar{R} \forall t \in T$ against the alternative H_1 of dynamics in R_t given by:

$$H_1 : vech(R_t) = vech(\bar{R}) + \beta_1 vech(R_{t-1}) + ... + \beta_p vech(R_{t-p}). \tag{4.26}$$

The test effectively estimates equates to estimating a multivariate dataset using the Constant Conditional Correlation (CCC) model of Bollerslev (1990) and after which the standardized residuals[39] should be i.i.d. with covariance the identity matrix. Testing for this can be done using a series of artificial regressions on the outer and lagged product of these residuals and a constant. The second part of Table 4.5 displays the min, median and max of this statistic and its equivalent p-value (the statistic is distributed χ^2 with lags+1 d.o.f.) for the whole period 1975-2010, and 2 subperiods under the rolling estimation scheme. That is, the CCC model was estimated, and the test calculated for each of the 1878 moving windows. With the exception of the 2 subperiods displayed, for most of the period under study the null of constant correlation can be rejected with a high degree of confidence.

4.4.2 Data Generating Models

Having established the presence of such stylized facts as the presence of ARCH effects, departure from multivariate normality and non-constant correlation, I have chosen to use 8 different models, already discussed in Chapters 2 and 3. As a benchmark, and because it is mistakenly used in many cases, I have also included the static Historical (M_1) approach whereby the recent history is used as a proxy for the forecast scenario. That is, the unconditional multivariate density serves as a proxy for the conditional 1 step ahead forecast. Obviously, this method is only going to be valid, irrespective of the use of robust or Baysian methods, in the absence of conditional dynamics in the moments and co-moments. Apart from this, the models used briefly belong to 3 more general processes. The dynamic correlation models (DCC) were chosen with multivariate Normal, Laplace and Copula-Student distributions (M_2, M_3, M_4), the Generalized

[39]Standardized by the symmetric square root decomposition of the estimated constant correlation matrix

Orthogonal models (GO-GARCH and CHICAGO) with multivariate Normal and affine NIG distributions (M_5, M_6), and the Independent Factor ACD models (IFACD) with multivariate affine NIG and GH distributions (M_7, M_8). As argued in Section 4.1, it is important to have some idea of the consistency of measures derived from the scenario forecasts from each model. Using 3 of the models (M_3, M_6, M_8), one from each of the more general classes of models discussed,[40] Table 4.6 reports the average RMSE of 3 forecast measures (mean, standard deviation and CVaR at the 5% coverage) for which analytical (or semi-analytical) solutions exist against their equivalent scenario generated average measure (for the generated scenarios of size 1000, 2000, 4000, and 7000). The comparison is undertaken for the 1878 forecast periods of the empirical application and assuming an equally weighted portfolio of the point in time constituents of the DJIA. Values in square brackets represent the ratio of the RMSE with scenario size 1000 (S_{1000}) against the RMSE with scenario size T (T>1000). Under $\sqrt{N}$ consistency, this number should be close to $\sqrt{T/1000}$. The forecast mean, generated from AR(1) dynamics, is well within the expected consistency for all 3 models. The forecast standard deviation and CVaR are also within the expected consistency for the DCC and GO-GARCH based models, but not so for the IFACD model which appears to have cubic or quartic consistency. This was already discussed in Section 1.5.1, and means that we require a larger number of points per scenario for these types of models to have the equivalent consistency with the other models considered. Taking into account the actual level of the RMSE and computational resource constraints, I opt for a scenario of size 7000 for each of the models except the IFACD based models for which I use a scenario of size 10000 for each 1-ahead forecast.

4.4.3 Risk Models

For each forecast from the 8 models presented in the previous section, portfolios were formed from the optimization of 6 different measures of risk-reward using the NLFP method discussed in Section 4.3.1. These risk measures, already covered in Section 4.2 were CVaR at the 5% coverage rate (R_1), MAD (R_2), EV (R_3), MiniMax (R_4) and LPM of orders 1 and 4 (R_5 and R_6, respectively) with threshold the portfolio mean. Additionally, 2 sets of portfolios were considered, one with the standard long only constraint (denoted [L]) for which an LP or QP formulation was possible, and one

[40] This excludes the DCC-Copula model for which there is usually no closed form analytic expression for most of the measures in order to compare against the scenario average used.

TABLE 4.6: Forecast scenario measures consistency

Forecast Mean (RMSE)	S_{1000}	S_{2000}	S_{4000}	S_{7000}
M_3	0.00057	0.00041	0.00029	0.00022
		[1.3905]	[1.9545]	[2.6132]
M_6	0.00061	0.00041	0.00031	0.00023
		[1.4735]	[1.9567]	[2.6235]
M_8	0.00060	0.00043	0.00031	0.00023
		[1.3772]	[1.9524]	[2.5486]
Forecast SD (RMSE)	S_{1000}	S_{2000}	S_{4000}	S_{7000}
M_3	0.00043	0.00030	0.00023	0.00018
		[1.3995]	[1.8855]	[2.3905]
M_6	0.00045	0.00033	0.00023	0.00017
		[1.376]	[2.0047]	[2.6305]
M_8	0.00069	0.00056	0.00047	0.00045
		[1.2259]	[1.4557]	[1.5415]
Forecast $CVaR_{5\%}$ (RMSE)	S_{1000}	S_{2000}	S_{4000}	S_{7000}
M_3	0.00146	0.00104	0.00075	0.00057
		[1.4026]	[1.9509]	[2.5521]
M_6	0.00168	0.00118	0.00080	0.00062
		[1.4204]	[2.0888]	[2.7019]
M_8	0.00215	0.00169	0.00138	0.00127
		[1.2698]	[1.5553]	[1.6889]

Note: The Table reports the average RMSE of 3 forecast measures (mean, standard deviation and CVaR at the 5% coverage) for which analytical (or semi-analytical) solutions exist against their equivalent scenario generated average measure (for the generated scenarios of size 1000, 2000, 4000, and 7000) under 3 models (M_3, M_6, M_8). The comparison is undertaken for the 1878 forecast periods of the empirical application and assuming an equally weighted portfolio of the point in time constituents of the DJIA index. Values in square brackets represent the ratio of the RMSE with scenario size 1000 (S_{1000}) against the RMSE with scenario size T (T>1000). Under $\sqrt{(N)}$ consistency, this number should be close to $\sqrt{(T/1000)}$.

with no short sale constraints but a leverage constraint of 2 (denoted [LS]) for which an NLP formulation was used, details of which are given in Appendix D. For the [L] portfolios, upper bounds of 20% were used to avoid excess concentration, whilst for the [LS] portfolios, lower and upper bounds of -30% and 30% were used respectively.

4.4.4 Transaction Costs and Long-Short Margin Accounting

The explicit modelling of transaction costs, both fixed and proportional, during the optimization process may be important in light of costly re-balancing of positions vis-a-vis the marginal expected return. Based on a review of current commercial broker rates[41], and for the purpose of the empirical application undertaken, a fixed cost of $10 per transaction, was deemed realistic and adequate. The TB1M was used for the margin

[41] See for example http://www.interactivebrokers.com/en/p.php?f=interest&ib_entity=uk and http://www.schwab.com/public/schwab/investing/pricing_services/fees_minimums

accounting of the [LS] portfolios, for which TB1M+50bp were charged for margins loans, and TB1M-50bp paid for excess cash, where the initial and maintenance margin was set at 30%. Costs arising as a result of differences in the closing price (which was used as the price at which transactions were made), the amount of volume transacted (market impact) and the bid-ask spread, were not explicitly modelled. The use of modern algorithmic trading models has made possible the reduction if not elimination of such costs or at least their meaning in the traditional sense presented in the literature. For example, early studies on momentum trading strategies such as Jegadeesh and Titman (1993) and Moskowitz and Grinblatt (1999) assume round-trip transaction costs of up to 2%, while more recent studies such as in Kritzman, Page, and Turkington (2010) assume round-trip transaction costs of 0.4%. However, such strategies include a large cross-section of the US equity market, including small cap and non-liquid assets, and where market impact was likely to be a significant factor in the pricing. Because the application presented here is based on what is probably the 30 most liquid stocks in the US market, costs relating to the sourcing of stocks for the purpose of shorting, market impact and bid-ask spread costs are assumed to be small. This is not meant to imply that such costs do not exist, but that certain simplifying assumptions had to be made for the purpose of the backtest application, which given the type of securities used, was not believed to stray too far from reality.

The formation and tracking of the optimal portfolios was based on the weights obtained every Friday after the close of trading, based on which an integer number of shares was bought on the following Monday close price, and any required re-balancing of the portfolios undertaken. By forming the weekly portfolios in such a way, using shares and tracking them daily, it was possible to obtain a much more detailed picture on the drawdowns and daily risk as well as allow the realistic tracking of the margin positions of the [LS] portfolios.

4.4.5 Results

To gauge the performance of the portfolios, the DJIA and S&P500 total return indices were used as benchmarks. Because the total return DJIA index only started to be calculated in 1987, whilst the portfolio backtest started in 1975, comparison of the portfolios was carried out using 2 different starting dates. For the starting year 1975, I used an equally weighted total return index calculated from the point in time constituents of the DJIA, and supplemented this with the S&P500 value weighted index. For the starting year 1987, I used the actual total return price weighted DJIA index and again supplemented the comparison by including the S&P500 value weighted index, so that the empirical application was not limited to only one benchmark.

Table 4.7 displays some key performance statistics of these benchmarks, where I have also included an equally weighted version of the total return S&P500 index and the TB1M which acted as a proxy for the risk-free rate. It is interesting to note that the equally weighted indices, in both benchmarks, are clearly superior in terms of Sharpe Ratio to their actual weighting schemes (price and value weighted), a point already discussed in DeMiguel, Garlappi, and Uppal (2009) who argued that on the basis of model and parameter uncertainty, nothing beats the 1/N rule. Not discussed in that paper, was that an equally weighted index will also have the highest drawdown, since by ignoring higher moments a disproportionate weight is placed on securities which may experience large drops in value. This is usually not picked up by the Sharpe ratio alone which is why other measures or ratios should be consulted in the presence of non-normally distributed returns.

In Table 4.8, the alpha and beta from a CAPM regression on the monthly excess

Table 4.7: Benchmark indices

Benchmarks	$TW[1975]$	$TW[1987]$	$CVaR_5\%$	**MaxDD**	SD_A	$Mean_A$	$Sharpe_A$
$DJIA_{PW}$	*	8.0	0.027	0.519	0.175	0.102	0.35
$DJIA_{EW}$	133.1	8.2	0.026	0.628	0.176	0.142	0.43
$S\&P500_{VW}$	55.0	6.8	0.025	0.547	0.171	0.132	0.41
$S\&P500_{EW}$	186.1	12.0	0.026	0.590	0.183	0.172	0.59
$TB(1M)$	9.7	2.7	0.000	0.000	0.041	0.066	-

Note: The Table reports the performance of the S&P500 and Dow Jones Industrial Total Return benchmark indices under alternative weighting schemes (PW=Price Weighted, EW=Equally Weighted, VW=Value Weighted) and the 1 Month Treasury Bill (*TM(1M)*) as a proxy for the risk free rate for the period 13/01/1975 to 03/01/2011 (the DJIA Total Return index starts in 1987). The measures of performance used were Terminal Wealth (TW) of $1 starting from 2 different years, conditional Value at Risk (*CVaR*) at the 5% coverage rate, maximum drawdown (*MaxDD*), standard deviation (SD_A) formed from annual returns, holding period mean return ($Mean_A$) formed from annual returns and the annual Sharpe ratio ($Sharpe_A$).

portfolio returns, using the benchmark indices with different starting dates, is reported for the [L] and [LS] portfolios. To account for the presence of heteroscedasticity and autocorrelation in the portfolios, standard errors were calculated using the covariance estimator of West and Newey (1987), with a lag length of 4, and significance at the 1%, 5% and 10% levels is denoted by ***, ** and * respectively. The first thing that is clear from the table, and quite consistent throughout this empirical analysis, is the significantly bad performance (negative and significant alpha) of M_1 (the Historical Scenario) as a model for stochastic programming. In dynamic markets with the types of stylized facts observed and discussed in Section 4.4.1, a static approach, where the unconditional history is used as a proxy for the forecast and as represented by M_1, is completely inadequate. Further, for the [L] portfolios, alpha is only significant in very few cases. Specifically, when the risk measure is one which involves extreme tail risk penalization, as in the Minimax (R_4) and LPM of order 4 (LPM_4), the IFACD models are the only models with a significant alpha representing an annualized excess return of between 4% and 5% against the DJIA and S&P500 indices, respectively. Additionally, with the exception of the EV (R_3) criterion of Markowitz (1952), the IFACD models display significant alpha in all other risk measures with respect to the S&P500 index, in both the equal and value weighted versions. This is a rather surprising result. While the stocks used, based on the DJIA universe, are also members of the S&P500, it is likely that the latter, which includes another 470 securities, is adversely impacted from a non-optimized allocation scheme of a larger number of securities. This might therefore imply that the S&P500 is even less efficient that the DJIA. Overall, beta is everywhere significant and usually less than 1 which lends support to the notion that these portfolios are not just loading up on beta risk. The fact that the IFACD model was the only one to generate significant alpha in these portfolios, again lends support to the value of accounting for time varying higher moment dynamics, even on weekly data with window length not exceeding 1000 points.

When it comes to the [LS] portfolios the results are quite different with a significant alpha, except for the M_1 model, almost everywhere and representing an average excess annualized return of 8% against both the DJIA and S&P500 indices. Beta here is quite low on average with the lowest value found among the 2 IFACD models. Nevertheless, it is quite difficult to distinguish significant differences between models and risk measures for the [LS] portfolios in these regressions.

TABLE 4.8: Optimal portfolios: Benchmark regressions

	$DJIA_{PW}$ (1987)		$DJIA_{EW}$ (1975)		$S\&P500_{VW}$ (1987)		$S\&P500_{VW}$ (1975)			$DJIA_{PW}$ (1987)		$DJIA_{EW}$ (1975)		$S\&P500_{VW}$ (1987)		$S\&P500_{VW}$ (1975)	
	a	b	a	b	a	b	a	b		a	b	a	b	a	b	a	b
$R[L]_1M_1$	-0.005***	0.86***	-0.005***	0.785***	-0.005***	0.868***	-0.005***	0.834***	$R[LS]_1M_1$	-0.004**	0.22***	-0.004***	0.153***	-0.004 **	0.247***	-0.004***	0.226***
$R[L]_1M_2$	0.002	0.87***	0.001	0.844***	0.003	0.835***	0.002	0.843***	$R[LS]_1M_2$	0.008***	0.18***	0.008***	0.181***	0.008***	0.167***	0.008***	0.206***
$R[L]_1M_3$	0.000	0.90***	0.001	0.850***	0.001	0.855***	0.001	0.863***	$R[LS]_1M_3$	0.008***	0.16***	0.007***	0.131***	0.008***	0.132 **	0.007***	0.153***
$R[L]_1M_4$	0.001	0.89***	0.001	0.854***	0.002	0.841***	0.001	0.851***	$R[LS]_1M_4$	0.008***	0.17***	0.008***	0.185***	0.009***	0.154***	0.008***	0.212***
$R[L]_1M_5$	0.002	0.92***	0.001	0.837***	0.003	0.867***	0.002	0.839***	$R[LS]_1M_5$	0.008***	0.13***	0.007***	0.067 *	0.008***	0.125 **	0.007***	0.102***
$R[L]_1M_6$	0.002	0.93***	0.002	0.859***	0.003	0.887***	0.002	0.861***	$R[LS]_1M_6$	0.008***	0.14***	0.007***	0.087 **	0.008***	0.137***	0.007***	0.120***
$R[L]_1M_7$	0.002	0.92***	0.002	0.843***	0.003 *	0.869***	0.002 *	0.840***	$R[LS]_1M_7$	0.007***	0.14**	0.007***	0.093 **	0.007***	0.131 **	0.007***	0.119***
$R[L]_1M_8$	0.002	0.90***	0.002	0.844***	0.003 *	0.861***	0.002	0.848***	$R[LS]_1M_8$	0.008***	0.11*	0.008***	0.066	0.008***	0.111 *	0.008***	0.102 **
$R[L]_2M_1$	-0.004***	0.83***	-0.003***	0.762***	-0.003 *	0.826***	-0.003***	0.807***	$R[LS]_2M_1$	-0.001	0.60***	-0.002	0.540***	-0.001	0.620***	-0.002	0.576***
$R[L]_2M_2$	0.002	0.87***	0.002	0.840***	0.003	0.835***	0.002	0.843***	$R[LS]_2M_2$	0.008***	0.20***	0.007***	0.198***	0.008***	0.184***	0.007***	0.222***
$R[L]_2M_3$	0.000	0.90***	0.001	0.851***	0.001	0.853***	0.001	0.861***	$R[LS]_2M_3$	0.008***	0.14***	0.007***	0.123***	0.009***	0.114 **	0.007***	0.142***
$R[L]_2M_4$	0.001	0.89***	0.001	0.856***	0.002	0.840***	0.001	0.853***	$R[LS]_2M_4$	0.009***	0.19***	0.008***	0.179***	0.009***	0.164***	0.008***	0.207***
$R[L]_2M_5$	0.002	0.92***	0.002	0.838***	0.003	0.875***	0.002	0.841***	$R[LS]_2M_5$	0.008***	0.13**	0.007***	0.074 *	0.008***	0.123 **	0.007***	0.105 **
$R[L]_2M_6$	0.002	0.94***	0.002	0.861***	0.003	0.894***	0.002	0.864***	$R[LS]_2M_6$	0.007***	0.14***	0.008***	0.084 **	0.008***	0.138***	0.007***	0.117***
$R[L]_2M_7$	0.002	0.92***	0.002	0.847***	0.003 *	0.873***	0.003 *	0.843***	$R[LS]_2M_7$	0.008***	0.14**	0.007***	0.087 **	0.008***	0.134 **	0.007***	0.121***
$R[L]_2M_8$	0.002	0.92***	0.001	0.853***	0.003	0.874***	0.002	0.855***	$R[LS]_2M_8$	0.008***	0.12**	0.008***	0.069 *	0.008***	0.113 *	0.008***	0.104 **
$R[L]_3M_1$	-0.006***	1.04***	-0.007***	0.955***	-0.005***	1.054***	-0.007***	1.031***	$R[LS]_3M_1$	-0.004**	0.37***	-0.004 **	0.274***	-0.004 *	0.385***	-0.005***	0.364***
$R[L]_3M_2$	0.000	1.03***	0.000	0.981***	0.001	1.010***	0.001	0.991***	$R[LS]_3M_2$	0.008***	0.19***	0.008***	0.188***	0.009***	0.174***	0.008***	0.213***
$R[L]_3M_3$	0.000	1.01***	0.000	0.957***	0.001	0.988***	0.001	0.986***	$R[LS]_3M_3$	0.008***	0.14***	0.007***	0.123***	0.009***	0.118 **	0.007***	0.144***
$R[L]_3M_4$	0.002	1.06***	0.001	0.998***	0.003	1.035***	0.001	1.013***	$R[LS]_3M_4$	0.009***	0.17***	0.008***	0.171***	0.009***	0.152***	0.008***	0.199***
$R[L]_3M_5$	0.000	1.10***	0.000	0.999***	0.001	1.057***	0.000	1.007***	$R[LS]_3M_5$	0.007***	0.13**	0.007***	0.074 *	0.008***	0.126 **	0.007***	0.106 **
$R[L]_3M_6$	0.002	1.02***	0.001	0.943***	0.003	0.977***	0.001	0.953***	$R[LS]_3M_6$	0.007***	0.14***	0.007***	0.088 **	0.007***	0.141***	0.007***	0.120***
$R[L]_3M_7$	0.001	1.01***	0.000	0.940***	0.002	0.964***	0.001	0.949***	$R[LS]_3M_7$	0.008***	0.13**	0.008***	0.081 *	0.008***	0.124 **	0.007***	0.114 **
$R[L]_3M_8$	0.002	1.06***	0.001	0.994***	0.003	1.028***	0.002	0.999***	$R[LS]_3M_8$	0.008***	0.11**	0.008***	0.065	0.008***	0.111 *	0.008***	0.102 **
$R[L]_4M_1$	-0.004**	0.88***	-0.004***	0.824***	-0.003 *	0.893***	-0.004***	0.871***	$R[LS]_4M_1$	-0.007***	0.16***	-0.007***	0.145***	-0.008***	0.190***	-0.007***	0.217***
$R[L]_4M_2$	0.001	0.88***	0.000	0.844***	0.002	0.836***	0.001	0.850***	$R[LS]_4M_2$	0.006***	0.17***	0.006***	0.172***	0.006***	0.161***	0.006***	0.194***
$R[L]_4M_3$	0.000	0.90***	0.000	0.855***	0.001	0.836***	0.001	0.855***	$R[LS]_4M_3$	0.008***	0.11**	0.007***	0.108***	0.008***	0.093 *	0.007***	0.127***
$R[L]_4M_4$	0.001	0.89***	0.001	0.850***	0.002	0.827***	0.002	0.845***	$R[LS]_4M_4$	0.007***	0.17***	0.007***	0.171***	0.007***	0.159***	0.007***	0.191***
$R[L]_4M_5$	0.003**	0.88***	0.003 **	0.829***	0.004 **	0.833***	0.003 **	0.836***	$R[LS]_4M_5$	0.007***	0.15***	0.006***	0.083 **	0.007***	0.135 **	0.006***	0.103 **
$R[L]_4M_6$	-0.001	0.94***	0.000	0.861***	0.000	0.894***	0.001	0.863***	$R[LS]_4M_6$	0.005***	0.13***	0.005***	0.088 **	0.005***	0.124 **	0.005***	0.111***
$R[L]_4M_7$	0.003*	0.89***	0.002	0.812***	0.004 **	0.842***	0.002 *	0.822***	$R[LS]_4M_7$	0.006***	0.10**	0.006***	0.063 *	0.007***	0.086 *	0.006***	0.082 **
$R[L]_4M_8$	0.002	0.88***	0.002 *	0.816***	0.003 *	0.832***	0.003 *	0.830***	$R[LS]_4M_8$	0.007***	0.11*	0.006***	0.070 *	0.007***	0.106 *	0.006***	0.095 **
$R[L]_5M_1$	-0.003**	0.81***	-0.003***	0.752***	-0.003 *	0.806***	-0.003***	0.791***	$R[LS]_5M_1$	-0.003	0.53***	-0.004 **	0.442***	-0.003	0.525***	-0.004 **	0.529***
$R[L]_5M_2$	0.002	0.87***	0.001	0.840***	0.003	0.834***	0.002	0.842***	$R[LS]_5M_2$	0.008***	0.19***	0.008***	0.192***	0.009***	0.178***	0.007***	0.214***
$R[L]_5M_3$	0.000	0.90***	0.001	0.852***	0.001	0.853***	0.001	0.862***	$R[LS]_5M_3$	0.008***	0.14***	0.007***	0.118***	0.009***	0.110 **	0.007***	0.138***
$R[L]_5M_4$	0.001	0.89***	0.001	0.857***	0.002	0.841***	0.001	0.853***	$R[LS]_5M_4$	0.009***	0.19***	0.008***	0.181***	0.009***	0.167***	0.008***	0.211***
$R[L]_5M_5$	0.002	0.92***	0.002	0.839***	0.003	0.875***	0.002	0.842***	$R[LS]_5M_5$	0.007***	0.13**	0.007***	0.072 *	0.007***	0.125 **	0.007***	0.105 **
$R[L]_5M_6$	0.002	0.94***	0.002	0.862***	0.003	0.896***	0.002	0.865***	$R[LS]_5M_6$	0.007***	0.14***	0.007***	0.085 **	0.007***	0.137***	0.007***	0.117***
$R[L]_5M_7$	0.002	0.92***	0.002	0.847***	0.003 *	0.872***	0.003 *	0.843***	$R[LS]_5M_7$	0.008***	0.14**	0.007***	0.087 **	0.008***	0.134 **	0.007***	0.121***
$R[L]_5M_8$	0.002	0.92***	0.001	0.854***	0.003	0.875***	0.002	0.856***	$R[LS]_5M_8$	0.008***	0.12**	0.008***	0.075 *	0.008***	0.113 *	0.008***	0.111 **
$R[L]_6M_1$	-0.005***	0.84***	-0.005***	0.777***	-0.004 **	0.825***	-0.004***	0.807***	$R[LS]_6M_1$	-0.006***	0.23***	-0.005***	0.188***	-0.006***	0.256***	-0.006***	0.269***
$R[L]_6M_2$	0.002	0.87***	0.001	0.842***	0.003	0.833***	0.002	0.842***	$R[LS]_6M_2$	0.008***	0.20***	0.007***	0.192***	0.008***	0.185***	0.007***	0.219***
$R[L]_6M_3$	0.000	0.90***	0.001	0.848***	0.001	0.852***	0.001	0.861***	$R[LS]_6M_3$	0.008***	0.15***	0.007***	0.128***	0.008***	0.126 **	0.007***	0.150***
$R[L]_6M_4$	0.001	0.89***	0.001	0.855***	0.002	0.844***	0.002	0.853***	$R[LS]_6M_4$	0.008***	0.18***	0.008***	0.185***	0.009***	0.154***	0.008***	0.210***
$R[L]_6M_5$	0.002	0.92***	0.002	0.838***	0.003	0.871***	0.002	0.841***	$R[LS]_6M_5$	0.007***	0.13***	0.007***	0.075 **	0.008***	0.129 **	0.007***	0.106***
$R[L]_6M_6$	0.002	0.93***	0.002	0.854***	0.003 *	0.883***	0.002 *	0.857***	$R[LS]_6M_6$	0.007***	0.14***	0.007***	0.086 **	0.007***	0.133 **	0.007***	0.118***
$R[L]_6M_7$	0.003	0.92***	0.002 *	0.842***	0.004 **	0.863***	0.003 *	0.837***	$R[LS]_6M_7$	0.007***	0.13**	0.007***	0.087 **	0.008***	0.120 **	0.007***	0.112 **
$R[L]_6M_8$	0.003*	0.88***	0.002 *	0.826***	0.004 **	0.837***	0.003 *	0.830***	$R[LS]_6M_8$	0.008***	0.11*	0.008***	0.064	0.008***	0.106 *	0.007***	0.098 **

Note: The Table reports the results of regressions of the excess monthly returns of the optimal long [L] and long-short [LS] portfolios formed under 6 different measures of risk and 8 models, on the total excess returns of the $DJIA_{PW}$, $DJIA_{EW}$ and $S\&P500_{VW}$ indices. The first column, a, is the CAPM alpha and the second, b, the CAPM beta. Significance at the 1%, 5% and 10% levels is denoted by ***, ** and * respectively, where standard errors were calculated using a heteroscedasticity adjusted covariance matrix of West and Newey (1987) with 4 lags. The starting years used for the portfolio regressions are in parenthesis.

A more revealing pattern is reported in Table 4.9 which shows the difference in Sharpe Ratios (SR)[42] between the portfolios and benchmarks. Among the risk measures for the [L] portfolios benchmarked against the price weighted DJIA and S&P500 indices with start year 1987, the Minimax (R_4) and LPM of order 4 (R_6) measures based on GOGARCH (M_5 and M_6) and IFACD (M_7 and M_8) models have the most significant positive differences, confirming the analysis in Table 4.8. It is interesting to note that based on the ranking of portfolios using the SR difference, the two benchmarks appear interchangeable. Using the equally weighted DJIA and S&P500 benchmarks with start year 1975, the SR differences appear to be positive and significant for the majority of portfolios formed under different measures of risk and models. Also, the SR differences of the equal weighted indices are almost everywhere higher than their value or price weighted equivalents, suggesting that, unlike the conclusions of DeMiguel, Garlappi, and Uppal (2009), the 1/N strategy is less efficient and hence has more room for improvement (as evidenced by the higher SR differences) than value or price weighting schemes.

In the long-short case, the majority of portfolios have positive and significant SR differences with respect to all the benchmarks and starting years, with more than a doubling of the SR difference in some cases. Here, the higher significant SR differences are to be found among the DCC type models (M_2, M_3 and M_4), and this may be related to the short history used and AR(1) model which under the DCC is jointly estimated with the GARCH dynamics so that the parameter estimates are more efficient. Since at least one measure of the second conditional moment, the beta, was lowest among the IFACD models in the CAPM type regressions in Table 4.8, it follows that the SR differences must be in the first conditional moment for which the differences can only be accounted for by the reason stated, and already discussed in Chapter 2. While there is clearly value in removing the short-sale constraint allowing for a greater degree of diversification particularly in periods of falling prices, the simple AR(1) model used for conditional mean in all the models does well in capturing positive trends but less well in capturing negative ones. Table 4.10 shows the average directional accuracy (DA) of each security which was included in the backtest and the breakdown of DA by positive (DA^+) and negative (DA^-) returns. The DA is very high for positive returns which is not surprising since most securities, over relatively sized time windows, have a positive

[42]While there are strong arguments against using the SR for performance measurement, it is still the most widely cited ratio and the only one to my knowledge for which adequate tests for establishing the significance of ratio differences exist.

mean. More sophisticated models are certainly possible such as the factor based models (see for example Grinold and Kahn (2000) and Chen, Roll, and Ross (1986)), but even with this simple model it is possible to significantly outperform the benchmark using weekly returns. As a further illustration of the degree to which the optimal portfolios outperform the benchmarks, Figure 4.3 shows the log TW of selected [L] and [LS] portfolios against the benchmark portfolios with starting years 1975 and 1987. Even though TW is not a good guide to portfolio performance, since it is strongly biased by the starting period and hence is strongly path dependent, it is still a high impact visual method for assessing relative performance. The 2 chosen portfolios, based on the IFACD (NIG) and DCC-Copula(T) models, with risk measure the Conditional Value at Risk, clearly show that an investor choosing to follow an active portfolio strategy using a universe comprised of liquid DJIA constituents members would have more than trebled his terminal wealth versus the S&P500 and DJIA (equal weighted) indices starting in 1975. Indirectly shown here, replicating the S&P500 index could be done more compactly by simply following an equal weighted DJIA constituent strategy, with important implications for managers tracking the former with a limited number of stocks.

TABLE 4.9: Optimal portfolios: Sharpe ratio benchmark difference

	$DJIA_{PW}$[1987]	$DJIA_{EW}$[1975]	$S\&P500_{VW}$[1987]	$S\&P500_{VW}$[1975]		$DJIA_{PW}$[1987]	$DJIA_{EW}$[1975]	$S\&P500_{VW}$[1987]	$S\&P500_{VW}$[1975]
$R[L]_1M_1$	-0.18 *	-0.13	-0.14	-0.10	$R[LS]_1M_1$	-0.51 *	-0.60***	-0.47	-0.57***
$R[L]_1M_2$	0.21	0.28 *	0.25	0.31 *	$R[LS]_1M_2$	0.66***	0.61***	0.70***	0.64***
$R[L]_1M_3$	0.16	0.27***	0.20	0.30 **	$R[LS]_1M_3$	0.60***	0.51 **	0.64***	0.54***
$R[L]_1M_4$	0.17	0.26 **	0.21	0.29 **	$R[LS]_1M_4$	0.67***	0.62***	0.71***	0.65***
$R[L]_1M_5$	0.24	0.28***	0.28 *	0.31***	$R[LS]_1M_5$	0.48 **	0.40 *	0.52 **	0.43 **
$R[L]_1M_6$	0.25 *	0.30***	0.29 *	0.33***	$R[LS]_1M_6$	0.49 **	0.41 *	0.53 **	0.45 **
$R[L]_1M_7$	0.27 *	0.30***	0.31 **	0.33***	$R[LS]_1M_7$	0.45 **	0.37 *	0.49 **	0.40 *
$R[L]_1M_8$	0.25	0.29***	0.29 *	0.32***	$R[LS]_1M_8$	0.51 **	0.41 *	0.55 **	0.44 **
$R[L]_2M_1$	-0.09	-0.03	-0.05	0.00	$R[LS]_2M_1$	-0.13	-0.20	-0.09	-0.17
$R[L]_2M_2$	0.23	0.29 **	0.27	0.32 **	$R[LS]_2M_2$	0.64***	0.59***	0.68***	0.62***
$R[L]_2M_3$	0.16	0.26***	0.20	0.30***	$R[LS]_2M_3$	0.66***	0.53***	0.70***	0.57***
$R[L]_2M_4$	0.18	0.25 **	0.21	0.28 **	$R[LS]_2M_4$	0.69***	0.61***	0.73***	0.64***
$R[L]_2M_5$	0.25 *	0.29***	0.29 *	0.32***	$R[LS]_2M_5$	0.49 **	0.41 *	0.52 **	0.44 **
$R[L]_2M_6$	0.25	0.29***	0.29 *	0.33***	$R[LS]_2M_6$	0.48 **	0.43 **	0.52 **	0.46 **
$R[L]_2M_7$	0.26 *	0.30***	0.30 *	0.33***	$R[LS]_2M_7$	0.49 **	0.41 *	0.53 **	0.44 **
$R[L]_2M_8$	0.23	0.28***	0.27 *	0.31***	$R[LS]_2M_8$	0.46 *	0.41 *	0.50 **	0.44 *
$R[L]_3M_1$	-0.16 *	-0.18 **	-0.12	-0.15 *	$R[LS]_3M_1$	-0.40	-0.46 **	-0.36	-0.43***
$R[L]_3M_2$	0.15	0.20 *	0.19	0.23 *	$R[LS]_3M_2$	0.68***	0.62***	0.72***	0.66***
$R[L]_3M_3$	0.15	0.22 **	0.18	0.25 **	$R[LS]_3M_3$	0.64***	0.53***	0.68***	0.56***
$R[L]_3M_4$	0.24	0.25***	0.28 *	0.28***	$R[LS]_3M_4$	0.68***	0.61***	0.72***	0.64***
$R[L]_3M_5$	0.18	0.21 **	0.22	0.24 **	$R[LS]_3M_5$	0.47 **	0.39 *	0.51 **	0.43 **
$R[L]_3M_6$	0.22	0.24 **	0.26	0.28 **	$R[LS]_3M_6$	0.47 **	0.42 **	0.51 **	0.45 **
$R[L]_3M_7$	0.22	0.22 **	0.26 *	0.25 **	$R[LS]_3M_7$	0.50 **	0.43 *	0.54 **	0.46 **
$R[L]_3M_8$	0.24	0.27***	0.28 *	0.30***	$R[LS]_3M_8$	0.47 *	0.41 *	0.51 **	0.44 **
$R[L]_4M_1$	-0.10	-0.06	-0.06	-0.03	$R[LS]_4M_1$	-0.92***	-0.94***	-0.88***	-0.91***
$R[L]_4M_2$	0.17	0.23 **	0.21	0.26 **	$R[LS]_4M_2$	0.45 **	0.44 **	0.49***	0.47 **
$R[L]_4M_3$	0.16	0.24 **	0.20	0.27 **	$R[LS]_4M_3$	0.56***	0.46 **	0.60***	0.49 **
$R[L]_4M_4$	0.23 *	0.29***	0.27 *	0.32***	$R[LS]_4M_4$	0.60***	0.53***	0.63***	0.56***
$R[L]_4M_5$	0.32 **	0.35***	0.36***	0.38***	$R[LS]_4M_5$	0.45 **	0.29	0.48 **	0.32
$R[L]_4M_6$	0.10	0.21 **	0.14	0.24 **	$R[LS]_4M_6$	0.27	0.24	0.31 *	0.27
$R[L]_4M_7$	0.29 **	0.29***	0.33 **	0.32***	$R[LS]_4M_7$	0.38 *	0.27	0.42 **	0.31
$R[L]_4M_8$	0.25 **	0.32***	0.29 **	0.36***	$R[LS]_4M_8$	0.38 *	0.27	0.42 **	0.30
$R[L]_5M_1$	-0.09	-0.03	-0.05	0.00	$R[LS]_5M_1$	-0.23	-0.30 *	-0.19	-0.26 *
$R[L]_5M_2$	0.22	0.28 *	0.26	0.31 *	$R[LS]_5M_2$	0.68***	0.61***	0.72***	0.64***
$R[L]_5M_3$	0.16	0.26 **	0.19	0.29***	$R[LS]_5M_3$	0.65***	0.52***	0.69***	0.56***
$R[L]_5M_4$	0.18	0.25 **	0.22	0.28 **	$R[LS]_5M_4$	0.69***	0.61***	0.73***	0.64***
$R[L]_5M_5$	0.25 *	0.29***	0.29 *	0.32***	$R[LS]_5M_5$	0.46 **	0.41 *	0.50 **	0.44 **
$R[L]_5M_6$	0.25 *	0.30***	0.29 *	0.33***	$R[LS]_5M_6$	0.46 **	0.41 *	0.50 **	0.44 **
$R[L]_5M_7$	0.27 *	0.30***	0.31 *	0.34***	$R[LS]_5M_7$	0.50 **	0.41 *	0.54 **	0.45 **
$R[L]_5M_8$	0.23	0.27***	0.27 *	0.31***	$R[LS]_5M_8$	0.45 *	0.39 *	0.49 *	0.43 *
$R[L]_6M_1$	-0.18	-0.12	-0.15	-0.09	$R[LS]_6M_1$	-0.65 **	-0.66***	-0.61 **	-0.63***
$R[L]_6M_2$	0.21	0.28 *	0.25	0.31 *	$R[LS]_6M_2$	0.65***	0.60***	0.69***	0.64***
$R[L]_6M_3$	0.16	0.27***	0.20	0.30***	$R[LS]_6M_3$	0.61***	0.52***	0.65***	0.55***
$R[L]_6M_4$	0.19	0.27 **	0.23	0.30 **	$R[LS]_6M_4$	0.69***	0.64***	0.73***	0.67***
$R[L]_6M_5$	0.25 *	0.29***	0.29 *	0.32***	$R[LS]_6M_5$	0.48 **	0.39 *	0.52 **	0.42 **
$R[L]_6M_6$	0.26 *	0.31***	0.30 **	0.34***	$R[LS]_6M_6$	0.46 **	0.40 *	0.50 **	0.43 **
$R[L]_6M_7$	0.29 **	0.32***	0.33 **	0.35***	$R[LS]_6M_7$	0.48 **	0.39 *	0.52 **	0.43 **
$R[L]_6M_8$	0.28 *	0.31***	0.32 **	0.34***	$R[LS]_6M_8$	0.50 **	0.41 *	0.54 **	0.44 **

Note: The Table reports the differences in the annualized SR of the excess returns of the long ([L]) and long-short ([LS]) portfolios formed under 6 different measures of risk and 8 DGPs, against the total excess returns of the $DJIA_{PW}$, $DJIA_{EW}$ and $S\&P500_{VW}$ indices. Significance at the 1%, 5% and 10% levels is denoted by ***, ** and * respectively based on the test of Ledoit and Wolf (2008). The starting years used for the calculations are in parenthesis.

Table 4.10: Directional accuracy (conditional mean dynamics)

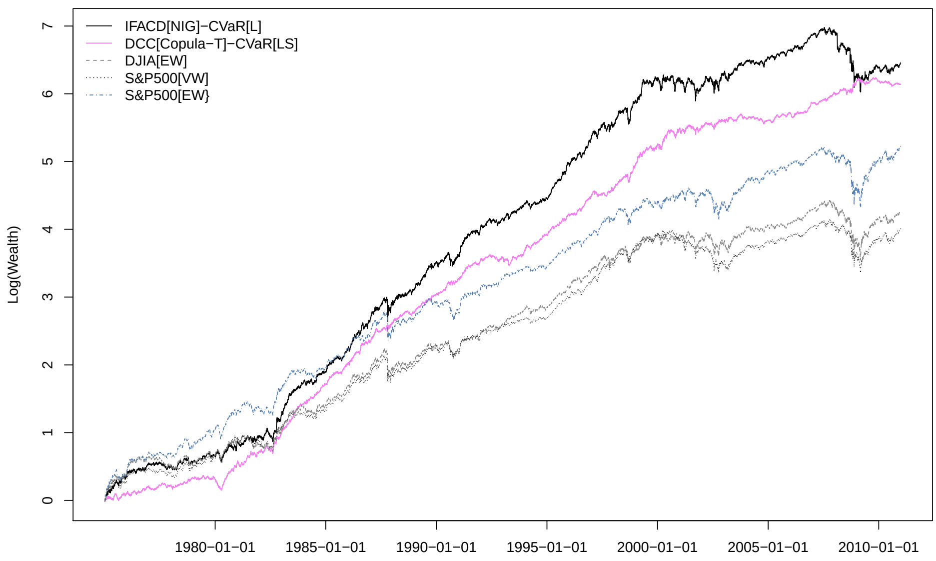
PERMNO	10107	10145	10225	10241	10401	10495	10786	11260	11308	11703	11754	11850	12060	12079	12490
DA	48	53	55	52	50	55	49	50	53	52	50	55	54	52	51
DA^+	89	87	97	94	83	88	90	97	89	87	81	83	80	88	96
DA^-	7	14	2	6	16	23	11	3	12	15	20	21	24	15	2
% Up Weeks	50	53	56	52	51	49	48	50	54	52	50	55	53	51	52
% Down Weeks	50	47	44	48	49	51	52	50	46	48	50	45	47	49	48
No.Weeks	582	1729	565	853	1793	84	1159	234	1242	1878	1527	1878	1878	1797	1644

PERMNO	12503	12546	13661	13901	14322	14541	14736	15069	15368	15456	15659	16109	16432	16707	17830
DA	51	51	52	56	51	54	57	51	54	51	53	57	53	52	55
DA^+	74	79	91	92	78	85	85	85	94	90	96	99	99	89	93
DA^-	34	28	7	12	23	19	24	17	8	11	5	1	1	12	9
% Up Weeks	45	46	53	55	50	53	54	50	54	51	53	57	53	51	55
% Down Weeks	55	54	47	45	50	47	46	50	46	49	47	43	47	49	45
No.Weeks	853	636	636	1164	1296	1445	1159	853	1159	1159	1296	564	1296	400	1878

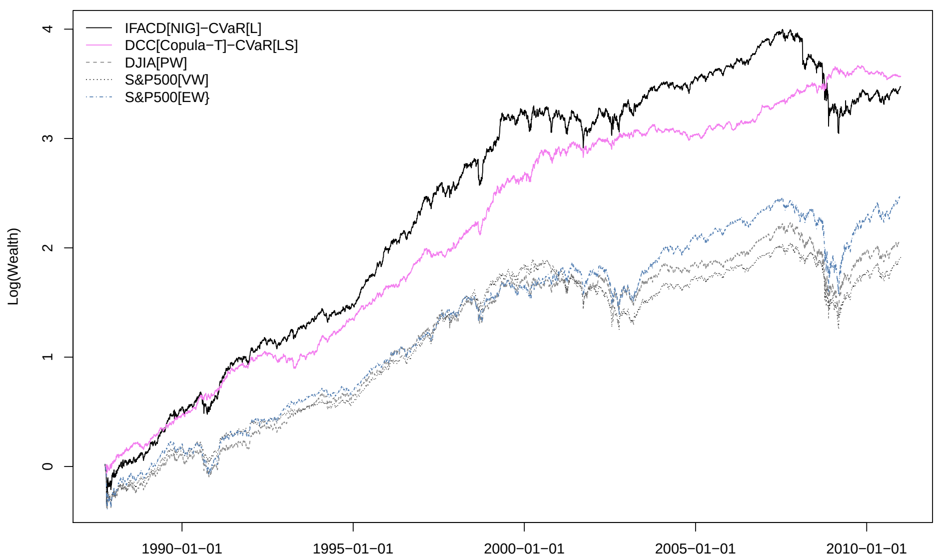
PERMNO	18163	18542	19561	19713	21573	21936	22111	22592	22752	24643	26403	27828	43449	47896	48071
DA	54	52	52	51	52	53	53	54	53	53	52	53	54	52	50
DA^+	85	91	94	89	94	81	92	81	80	88	97	90	95	89	93
DA^-	19	9	6	9	9	26	11	25	21	15	3	12	7	9	5
% Up Weeks	53	53	53	52	51	49	51	53	54	52	52	53	53	53	50
% Down Weeks	47	47	47	48	49	51	49	47	46	48	48	47	47	47	50
No.Weeks	1878	1025	1242	234	1527	351	719	1794	1644	1878	1025	719	1313	413	612

PERMNO	55976	59176	59328	59408	59459	65875	66093	66181	66800	70519	76076	89006
DA	56	53	53	47	54	53	55	47	51	51	54	57
DA^+	89	82	93	68	95	78	83	81	88	88	100	69
DA^-	18	20	8	26	5	27	28	14	16	13	0	43
% Up Weeks	54	53	53	50	54	51	49	50	48	51	54	53
% Down Weeks	46	47	47	50	46	49	51	50	52	49	46	47
No.Weeks	719	1478	582	149	81	351	316	582	233	638	81	118

Note: The Table reports the average percent Directional Accuracy (DA), based on the constant-AR(1) conditional mean dynamics, of the securities (using their CRSP PERMNO) which were included in the backtest and representing the point in time constituents of the DJIA. DA^+ and DA^- denotes the directional accuracy for positive and negative returns respectively.



(A) Start:1975



(B) Start:1987

Figure 4.3: Selected Portfolios vs Benchmarks

TABLE 4.11: Optimal Long-only portfolios under 6 different measures of risk

R_1	TW	$CVaR_{5\%}$	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	32.9	0.024	0.578	0.184	0.119	0.292
M_2	589.2	0.025	0.607	0.221	0.216	0.700
M_3	511.2	0.025	0.538	0.217	0.209	0.687
M_4	477.4	0.025	0.570	0.216	0.208	0.680
M_5	631.5	0.026	0.635	0.223	0.219	0.702
M_6	763.4	0.026	0.638	0.227	0.226	0.719
M_7	727.7	0.026	0.641	0.226	0.224	0.712
M_8	679.3	0.026	0.647	0.229	0.223	0.702

R_2	TW	MAD	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	58.2	0.007	0.492	0.183	0.135	0.387
M_2	595.8	0.008	0.597	0.217	0.216	0.711
M_3	490.6	0.008	0.537	0.217	0.208	0.680
M_4	462.2	0.008	0.595	0.215	0.207	0.679
M_5	677.6	0.008	0.634	0.225	0.222	0.707
M_6	733.2	0.008	0.639	0.229	0.225	0.708
M_7	736.5	0.008	0.654	0.230	0.225	0.706
M_8	638.5	0.008	0.658	0.234	0.222	0.681

R_3	TW	SD	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	24.3	0.013	0.649	0.214	0.115	0.231
M_2	540.6	0.013	0.725	0.255	0.222	0.626
M_3	627.6	0.013	0.617	0.244	0.222	0.657
M_4	747.6	0.013	0.558	0.226	0.223	0.704
M_5	578.5	0.013	0.620	0.250	0.220	0.627
M_6	718.0	0.013	0.579	0.251	0.227	0.652
M_7	550.3	0.013	0.602	0.233	0.215	0.650
M_8	829.2	0.013	0.560	0.237	0.228	0.699

R_4	TW	Min	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	52.1	-0.210	0.685	0.192	0.136	0.368
M_2	392.4	-0.241	0.534	0.211	0.199	0.650
M_3	422.4	-0.200	0.537	0.215	0.202	0.661
M_4	492.8	-0.179	0.517	0.195	0.204	0.741
M_5	923.7	-0.190	0.603	0.217	0.230	0.777
M_6	371.8	-0.200	0.608	0.208	0.200	0.664
M_7	622.8	-0.201	0.588	0.213	0.216	0.717
M_8	763.0	-0.210	0.471	0.218	0.222	0.732

R_5	TW	LPM_1	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	57.2	0.004	0.496	0.175	0.134	0.393
M_2	576.9	0.004	0.600	0.217	0.215	0.706
M_3	481.0	0.004	0.542	0.218	0.207	0.674
M_4	465.6	0.004	0.594	0.216	0.207	0.678
M_5	685.3	0.004	0.633	0.225	0.222	0.710
M_6	739.7	0.004	0.639	0.229	0.226	0.708
M_7	752.6	0.004	0.652	0.230	0.226	0.710
M_8	630.6	0.004	0.656	0.234	0.221	0.678

R_6	TW	LPM_4	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	31.9	0.027	0.618	0.185	0.118	0.287
M_2	576.5	0.042	0.603	0.218	0.215	0.702
M_3	495.9	0.027	0.535	0.218	0.208	0.679
M_4	494.8	0.032	0.570	0.215	0.209	0.687
M_5	665.5	0.026	0.632	0.223	0.221	0.710
M_6	782.8	0.026	0.612	0.224	0.226	0.728
M_7	809.4	0.026	0.622	0.226	0.228	0.728
M_8	779.1	0.029	0.581	0.225	0.226	0.727

Note: The Table reports the out of sample performance of 8 models ($M_1 - M_8$) optimized under 6 different risk measures, using the weekly log total returns of the point in time 30 constituents of the DJIA index for the period 13/01/1975 to 03/01/2011 (1878 weeks). The models were re-estimated every week (Friday), and a scenario forecast for the following week of size 7000×30. The set of weights for each model and measure was obtained by optimizing under optimal risk-reward fractional programming model with short sale constraints and maximum bounds of 20%. Portfolios were formed/rebalanced on the Monday following the Friday's estimation, by calculating the number of shares to allocate/reallocate to each position after subtracting a flat commission of $10 per trade, and tracked daily. The measures of performance used, based on these daily portfolios, were terminal wealth (TW) of $1, maximum drawdown ($MaxDD$), standard deviation ($SD_A$) formed from annual returns, holding period mean return ($Mean_A$) formed from annual returns and the annualized Sharpe ratio ($Sharpe_A$). The statistic for the risk measure optimized was also reported for each model based on the daily portfolio returns.

TABLE 4.12: Optimal Long-Short portfolios under 6 different measures of risk

R_1	TW	$CVaR_{5\%}$	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	3.1	0.017	0.472	0.123	0.039	-0.228
M_2	580.8	0.014	0.198	0.148	0.202	1.000
M_3	349.3	0.014	0.181	0.151	0.186	0.849
M_4	596.7	0.015	0.201	0.153	0.204	0.981
M_5	279.9	0.016	0.202	0.161	0.180	0.779
M_6	311.7	0.016	0.213	0.166	0.184	0.776
M_7	241.1	0.016	0.294	0.160	0.175	0.755
M_8	329.3	0.016	0.245	0.166	0.186	0.784

R_2	TW	MAD	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	29.1	0.007	0.400	0.202	0.114	0.234
M_2	541.9	0.005	0.191	0.150	0.200	0.975
M_3	396.6	0.005	0.162	0.153	0.190	0.868
M_4	610.6	0.005	0.187	0.153	0.205	0.986
M_5	299.7	0.005	0.276	0.163	0.182	0.787
M_6	329.1	0.005	0.294	0.171	0.186	0.773
M_7	306.2	0.005	0.316	0.168	0.184	0.769
M_8	358.3	0.006	0.304	0.166	0.189	0.807

R_3	TW	SD	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	5.1	0.009	0.492	0.142	0.056	-0.070
M_2	613.0	0.007	0.198	0.151	0.204	0.999
M_3	393.8	0.007	0.171	0.149	0.190	0.889
M_4	575.6	0.007	0.198	0.150	0.202	0.986
M_5	277.2	0.008	0.280	0.163	0.180	0.768
M_6	312.9	0.008	0.291	0.168	0.184	0.770
M_7	327.2	0.008	0.294	0.168	0.186	0.785
M_8	347.8	0.008	0.298	0.167	0.188	0.798

R_4	TW	Min	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	0.8	-0.056	0.691	0.121	0.001	-0.592
M_2	251.3	-0.041	0.166	0.127	0.173	0.949
M_3	280.2	-0.053	0.179	0.164	0.180	0.737
M_4	345.0	-0.069	0.233	0.144	0.185	0.882
M_5	151.1	-0.087	0.310	0.161	0.160	0.630
M_6	118.2	-0.077	0.282	0.156	0.152	0.607
M_7	138.8	-0.069	0.276	0.160	0.157	0.631
M_8	148.7	-0.099	0.283	0.159	0.159	0.629

R_5	TW	LPM_1	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	11.6	0.004	0.465	0.148	0.081	0.104
M_2	576.5	0.003	0.194	0.151	0.202	0.981
M_3	388.2	0.003	0.171	0.150	0.189	0.879
M_4	609.6	0.003	0.195	0.150	0.204	0.999
M_5	297.0	0.003	0.275	0.165	0.182	0.784
M_6	291.2	0.003	0.290	0.168	0.182	0.753
M_7	318.7	0.003	0.317	0.168	0.185	0.780
M_8	339.6	0.003	0.304	0.166	0.187	0.798

LPM_4	TW	R_6	$MaxDD$	SD_A	$Mean_A$	$Sharpe_A$
M_1	2.2	0.010	0.570	0.129	0.031	-0.293
M_2	567.1	0.009	0.190	0.150	0.202	0.987
M_3	374.3	0.009	0.181	0.150	0.188	0.874
M_4	634.3	0.010	0.214	0.154	0.206	0.983
M_5	272.0	0.012	0.270	0.164	0.179	0.760
M_6	275.6	0.013	0.271	0.167	0.180	0.748
M_7	270.4	0.012	0.285	0.165	0.179	0.757
M_8	317.5	0.015	0.278	0.167	0.185	0.775

Note: The Table reports the out of sample performance of 8 models ($M_1 - M_8$) optimized under 6 different risk measures, using the weekly log total returns of the point in time 30 constituents of the DJIA index for the period 13/01/1975 to 03/01/2011 (1878 weeks). The models were re-estimated every week (Friday), and a scenario forecast for the following week of size 7000 × 30. The set of weights for each model and measure was obtained by optimizing under optimal risk-reward fractional programming model with maximum absolute bounds of 30% and 2× leverage. Portfolios were formed and rebalanced on the Monday following the Friday's estimation, by calculating the number of shares to allocate/reallocate to each position after subtracting a flat commission of $10 per trade, and tracked daily. Margin for the long and short positions was maintained at 30% of gross value and debit interest on loans was charged at 50*bp* above the 1 Month T-Bill rate while credit interest on cash was set at 50*bp* below the same rate. The measures of performance used, based on these daily portfolios, were terminal wealth (TW) of $1, maximum drawdown ($MaxDD$), standard deviation ($SD_A$) formed from annual returns, holding period mean return ($Mean_A$) formed from annual returns and the annualized Sharpe ratio ($Sharpe_A$). The statistic for the risk measure optimized was also reported for each model based on the daily portfolio returns.

Having presented evidence that there is value in an active portfolio strategy, even when considering alternative weighting schemes for the benchmark, it is interesting to investigate whether there is a particular set of measures or models which performs better among those used to form optimal portfolios. Tables 4.11 and 4.12 present performance statistics for the [L] and [LS] portfolios respectively, under the 6 risk measures and 8 models for the dynamics. Immediately obvious, and already stated in the benchmark comparison, is the poor relative performance of model M_1. Within the [L] portfolios, it is difficult to distinguish any particular model as being better based on the SR, although models based on the Minimax and LPM of order 4 measures (R_4 and R_6 models respectively) appear to provide for marginally better relative performance. Within the Minimax portfolios the model based on the IFACD-GH dynamics (M_8) has the lowest maximum drawdown ($MaxDD$) among all [L] portfolios which highlights the value of flexible dynamic higher moments for extreme losses, a good example of which was already presented in Section 3.5.4.2. Within the [LS] portfolios it would appear that the DCC models provide superior performance, confirming the results in Table 4.8, for the results already stated. Compared with the [L] portfolios, [LS] portfolios have a significantly higher SR and less than half the drawdowns. However, because of the lack of high DA in the negative returns of the [LS] portfolios, even with 2× leverage the mean is still not as high as that of the [L] portfolios. As comparison of SR differences has been notoriously absent in many papers, with most tables simply presenting the SR without stating any formal test of their significance, I provide in Appendix E.1 a detailed set of supplemental tables of the pairwise p-values for the SR differences of all portfolios using the test of Ledoit and Wolf (2008). The tables confirm the overall picture presented thus far, but also provide a more detailed view of the relative differences. For example, the difference between the SR of model M_8 against the DCC models M_2 and M_4 which appear to have the higher SR in the [LS] portfolios, are only on average marginally significant with p-values near 10%. Whilst the differences in many of the portfolios appear marginal, it is important to investigate whether any of them offer any potential advantage in terms of lower turnover. To that end, Table 4.13 reports the average weekly percent turnover rates for the intersection of all models and measures, in the [L] and [LS] portfolios, where turnover is defined as in Barber, Lehavy, McNichols, and Trueman (2001) in a three step procedure. First, for each stock j in the portfolio at time t_0, the percent holding in that portfolio at time t_1 is calculated based assuming no re-balancing, and denoting this as $G_{j,t}$:

$$G_{j,t} = \frac{S_{j,0}P_{j,1}}{\sum\limits_{\forall j} S_{j,0}P_{j,1}}, \tag{4.27}$$

where S_j is the number of shares in stock j and P_j the price of stock j. This is then compared to actual percent holding of firm j in the portfolio at the end of trading at time t taking into account any portfolio re-balancing undertaken, and denoting this as $F_{j,t}$:

$$F_{j,t} = \frac{S_{j,1}P_{j,1}}{\sum\limits_{\forall j} S_{j,1}P_{j,1}}. \tag{4.28}$$

Turnover at time t, T_t, is then defined as

$$T_t = \sum_{\forall j} \max\{G_{j,t} - F_{j,t}, 0\}. \tag{4.29}$$

For the [LS] portfolios the calculations have to be adjusted so that the absolute (gross) value is used to calculate percent holdings, and sign reversals are properly accounted for. Specifically:

- when $G_{j,t} \geq 0$ && $F_{j,t} \geq 0$, then $T_{j,t} = max\,(G_{j,t} - F_{j,t}, 0)$.
- when $G_{j,t} \geq 0$ && $F_{j,t} < 0$, then $T_{j,t} = G_{j,t}$.
- when $G_{j,t} < 0$ && $F_{j,t} \geq 0$, then $T_{j,t} = abs\,(G_{j,t})$.
- when $G_{j,t} < 0$ && $F_{j,t} < 0$, then $T_{j,t} = max\,(F_{j,t} - G_{j,t}, 0)$.

TABLE 4.13: Weekly portfolio turnover

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[L]_1$	4	60	62	61	69	69	69	69
$R[L]_2$	3	60	62	61	69	69	69	69
$R[L]_3$	3	66	66	67	72	72	72	73
$R[L]_4$	4	71	70	71	73	73	73	73
$R[L]_5$	11	60	62	61	69	69	69	69
$R[L]_6$	15	60	62	61	69	69	69	69
$R[LS]_1$	10	61	61	62	67	67	67	67
$R[LS]_2$	16	61	61	62	67	67	67	68
$R[LS]_3$	6	63	63	64	68	68	68	68
$R[LS]_4$	3	61	61	62	67	67	67	67
$R[LS]_5$	18	61	61	62	67	67	68	68
$R[LS]_6$	7	61	61	62	67	67	67	67

Note: The Table reports the average weekly percent turnover for portfolio optimized under 8 different models (M_1 to M_8) and optimized under 6 different measures of risk with no short sale constraint and 2× leverage ($R[LS]_1$ to $R[LS]_6$). Details of the turnover calculations can be found in Equations 4.27, 4.28 and 4.29.

The first thing that is immediately obvious is the overall high percent turnover of these models. Given that this is a weekly model, based on an AR(1) conditional mean forecast, this is not very surprising considering the arguments made with regards to the observed sign changes in the actual underlying returns and captured in the model DA which was already discussed and summarized in Table 4.10. Secondly, there is little variation in the overall percent weekly turnover between risk measures but there is some variation between models. The M_1 model, which has the worst overall performance, is unable to capture the conditional dynamics since using the unconditional history and rolling the window by 1 week at a time is likely to lead to a very slow change in the underlying weight distribution and hence very small turnover. Of more interest is the 7 to 10 percentage point difference between the DCC and the IFACD models. This is surely related to the fact that the time variation in the higher moments modelled by the IFACD model provides for additional variability at the margins of the risk return space.

With the exception of the SR, for which a robust test exists for the statistical comparison of pairwise differences, it is hard to draw strong conclusions from the other average point measures. However, since the SR may not be an optimal measure for comparison and inference in the presence of non-normally distributed returns, I use the MCS method of Hansen, Lunde, and Nason (2011) separately and jointly on the [L] and [LS] portfolios in Tables 4.14, 4.15 and 4.16. As it is not clear a-priori what loss function to use, I present results using both a linear loss function of the negative of the excess (over the TB1M rate) portfolio returns, and one based on the upper to lower Partial Moment utility measure of Holthausen (1981)[43] with threshold the TB1M, and parameters such that an S-shaped utility curve results, based on the theory of Kahneman and Tversky (1979) about investors' reactions to gains and losses. The procedure is run on all the portfolios in the [L], [LS] and the joint set of [L] and [LS] portfolios, and it therefore identifies the intersection of models and measures which belong to the optimal set. Neither the linear nor the partial moment loss functions suggest that any portfolio is superior to any other, with the exception of model M_1 which, as expected, does not belong to the optimal set with at least 90% confidence. To investigate whether the test was particulary sensitive to any other loss function, the MCS procedure was also tested on losses based on a quadratic function, and an inverse S-shaped utility curve which penalized (rewards) losses (gains) at an increasing rate. Neither returned results which were substantially different.

The MCS procedure on the portfolios, using a number of different

[43] The upper to lower partial moment loss function used, was defined as:

$$U(x_t) = x_t - k_l(\theta_t - x_t)^{a_l}, \qquad x_t < \theta_t$$
$$U(x_t) = x_t + k_u(x_t - \theta_t)^{a_u}, \qquad x_t \geq \theta_t$$

were the 1 Month T-Bill rate was used as the threshold(θ), $a_l = 0.3$, $a_u = 0.3$, $k_l = 10$ and $k_u = 10$, so as to represent an S-Shaped utility curve.

Table 4.14: Model Confidence Set portfolios I (Long only)

	Partial Moments Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[L]_1$	0.00	0.68	0.68	0.68	0.68	0.68	0.68	0.68
$R[L]_2$	0.09	0.68	0.68	0.68	0.68	0.68	0.68	0.68
$R[L]_3$	0.00	0.68	0.68	0.68	0.41	0.68	0.48	0.48
$R[L]_4$	0.04	0.68	0.68	0.68	0.68	0.48	0.68	0.68
$R[L]_5$	0.03	0.68	0.68	0.68	0.68	0.92	0.68	0.68
$R[L]_6$	0.00	0.68	0.68	0.68	0.68	1.00	0.68	0.68
	Linear Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[L]_1$	0.00	0.87	0.87	0.87	0.87	0.87	0.87	0.87
$R[L]_2$	0.00	0.87	0.87	0.87	0.87	0.87	0.87	0.87
$R[L]_3$	0.00	0.87	0.87	1.00	0.87	1.00	0.87	1.00
$R[L]_4$	0.01	0.87	0.87	0.87	1.00	0.84	0.87	1.00
$R[L]_5$	0.00	0.87	0.87	0.87	0.87	0.87	1.00	0.87
$R[L]_6$	0.00	0.87	0.87	0.87	0.87	0.87	1.00	1.00

Note: The Table reports the Model Confidence Set (MCS) p-values of the portfolios estimated under 8 different models (M_1 to M_8) and optimized under 6 different measures of risk with short sale constraints ($R[L]_1$ to $R[L]_6$). The partial moment loss function used was based on a modified partial moment utility of Holthausen (1981) and defined as:

$$U(x_t) = x_t - k_l(\theta_t - x_t)^{a_l}, \quad x_t < \theta_t$$
$$U(x_t) = x_t + k_u(x_t - \theta_t)^{a_u}, \quad x_t \geq \theta_t$$

were the 1 month T-Bill rate was used as the threshold (θ), $a_l = 0.3$, $a_u = 0.3$, $k_l = 10$ and $k_u = 10$, so as to represent an S-shaped utility curve. The linear loss function was based on the excess over the 1 month T-Bill rate.

Table 4.15: Model Confidence Set portfolios I (Long-Short)

	Partial Moments Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[LS]_1$	0.00	1.00	0.64	1.00	1.00	1.00	0.82	1.00
$R[LS]_2$	0.01	1.00	0.99	1.00	1.00	1.00	1.00	1.00
$R[LS]_3$	0.00	1.00	0.95	1.00	1.00	1.00	1.00	1.00
$R[LS]_4$	0.00	0.64	0.48	1.00	1.00	0.64	0.48	0.64
$R[LS]_5$	0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$R[LS]_6$	0.00	1.00	0.64	1.00	1.00	1.00	0.98	1.00
	Linear Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[LS]_1$	0.00	0.85	0.85	0.85	0.85	0.85	0.85	0.85
$R[LS]_2$	0.00	0.85	0.85	0.85	0.85	0.85	0.85	0.85
$R[LS]_3$	0.00	0.85	0.85	1.00	0.85	1.00	0.85	1.00
$R[LS]_4$	0.01	0.85	0.85	0.85	1.00	0.84	0.85	1.00
$R[LS]_5$	0.00	0.85	0.85	0.85	0.85	0.85	1.00	0.85
$R[LS]_6$	0.00	0.85	0.85	0.85	0.85	0.85	1.00	1.00

Note: The Table reports the Model Confidence Set (MCS) p-values of the portfolios estimated under 8 different models (M_1 to M_8) and optimized under 6 different measures of risk with no short sale constraint and 2× leverage ($R[LS]_1$ to $R[LS]_6$). The partial moment loss function used was based on a modified partial moment utility of Holthausen (1981) and defined as:

$$U(x_t) = x_t - k_l(\theta_t - x_t)^{a_l}, \quad x_t < \theta_t$$
$$U(x_t) = x_t + k_u(x_t - \theta_t)^{a_u}, \quad x_t \geq \theta_t$$

were the 1 month T-Bill rate was used as the threshold(θ), $a_l = 0.3$, $a_u = 0.3$, $k_l = 10$ and $k_u = 10$, so as to represent an S-shaped utility curve. The linear loss function was based on the excess over the 1 month T-Bill rate.

loss functions, did not provide a compact superior set of models or measures. It is not clear why, but it appears that with the portfolios tested, each with a size of more than 9000 points (daily returns), any marginal benefits from differences in the model dynamics contribute very little to the overall losses based on the type of loss functions used. While it is difficult to highlight one particular measure or model which does

TABLE 4.16: Model Confidence Set portfolios II

	Partial Moments Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[L]_1$	0.01	0.88	0.88	0.88	0.88	0.88	0.88	0.88
$R[L]_2$	0.07	0.88	0.88	0.88	0.88	0.88	0.88	0.88
$R[L]_3$	0.00	0.88	0.88	0.88	0.68	0.88	0.74	0.69
$R[L]_4$	0.05	0.88	0.88	0.88	0.88	0.69	0.88	0.88
$R[L]_5$	0.05	0.88	0.88	0.88	0.88	0.92	0.88	0.88
$R[L]_6$	0.01	0.88	0.88	0.88	0.88	1.00	0.88	0.88
$R[LS]_1$	0.00	0.88	0.68	0.88	0.88	0.88	0.68	0.88
$R[LS]_2$	0.00	0.88	0.68	0.88	0.88	0.88	0.88	0.88
$R[LS]_3$	0.00	0.88	0.68	0.88	0.88	0.88	0.88	0.88
$R[LS]_4$	0.00	0.68	0.17	0.88	0.69	0.68	0.68	0.68
$R[LS]_5$	0.00	0.88	0.69	0.88	0.88	0.88	0.88	0.88
$R[LS]_6$	0.00	0.88	0.68	0.88	0.88	0.88	0.69	0.88
	Linear Loss Function							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
$R[L]_1$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[L]_2$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[L]_3$	0.00	0.95	0.95	1.00	0.95	1.00	0.95	1.00
$R[L]_4$	0.02	0.95	0.95	0.95	1.00	0.95	0.95	1.00
$R[L]_5$	0.00	0.95	0.95	0.95	0.95	0.95	1.00	0.95
$R[L]_6$	0.00	0.95	0.95	0.95	0.95	0.95	1.00	1.00
$R[LS]_1$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[LS]_2$	0.22	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[LS]_3$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[LS]_4$	0.00	0.22	0.95	0.95	0.22	0.15	0.22	0.22
$R[LS]_5$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95
$R[LS]_6$	0.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95

Note: The Table reports the Model Confidence Set (MCS) p-values of the portfolios estimated under 8 different models (M_1 to M_8) and optimized under 6 different measures of risk with short sale constraints ($R[L]_1$ to $R[L]_6$) and no short sale constraint with 2× leverage ($R[LS]_1$ to $R[LS]_6$) evaluated jointly. The partial moment loss function used was based on a modified partial moment utility of Holthausen (1981) and defined as:

$$U(x_t) = x_t - k_l(\theta_t - x_t)^{a_l}, \qquad x_t < \theta_t$$
$$U(x_t) = x_t + k_u(x_t - \theta_t)^{a_u}, \qquad x_t \geq \theta_t$$

were the 1 month T-Bill rate was used as the threshold(θ), $a_l = 0.3$, $a_u = 0.3$, $k_l = 10$ and $k_u = 10$, so as to represent an S-shaped utility curve. The linear loss function was based on the excess over the 1 month T-Bill rate.

better all the time, for the [L] portfolios the LPM measures with the IFACD models appears to be marginally better, when considering the results of Tables 4.8 and 4.9 and some of the point estimates in Table 4.11. For the [LS] portfolios the DCC based models across most measures seems to provide marginal out-performance, and it was hypothesized that this may be related to the joint AR-GARCH estimation under these models, a point also highlighted in Section 2.4. Tests to evaluate significant differences in measures other than the Sharpe Ratio, such as those from the LPM family discussed in Section 4.2.4, would be particularly useful in relative performance evaluation.

4.5 Conclusion

Advances in knowledge and computational power are opening up new frontiers for active investing and portfolio management. No longer is it necessary to make the usual simplifying assumptions of multivariate normality, nor be restricted to the EV model. New models and measures can capture the observed market phenomena, in large dimensional systems with reasonable speed. Using fractional programming with smooth approximations to discontinuous functions, it is possible to optimize with confidence and speed these risk measures, something which previously proved challenging in an NLP setup. Benchmarks for gauging the performance of these new models need to reflect their sophistication and adapt accordingly. In this chapter I have used and extended some of these new measures and models, and provided evidence via a large scale realistic portfolio application of their ability to substantially outperform a popular set of benchmarks. The [L] portfolios based on the IFACD model and optimized under the LPM risk measure provided the most significant alpha in Table 4.8 with an expected excess return of between 4% and 5% against the DJIA and S&P500 indices, and betas below 1. For the [LS] portfolios all models and measures provided a significant alpha with very low betas. The portfolios based on the IFACD model and optimized under the LPM measure had the lowest betas in this group. The SR differences confirmed the findings from the CAPM regressions, with the DCC models having the largest significant differences in the [LS] portfolios. It was hypothesized that, given similar conditional mean forecast dynamics, the reason for this difference was the better consistency of the joint AR-GARCH first stage estimation process of the DCC models. Among the benchmarks, the equal weighted indices were the easiest to outperform with significant SR differences in the [L] portfolios for all models and measures. This is a clear indication that they are sub-optimally weighted against even the value or price weighted indices on which they were calculated, and contrary to the findings of DeMiguel, Garlappi, and Uppal (2009). The terminal wealth and drawdown statistics in Tables 4.11 and 4.12 again confirmed the previous findings with the IFACD models providing the best performance in the [L] portfolios and the DCC based models in the [LS] portfolios. Unfortunately, the MCS procedure under a number of different loss functions did not reveal any significant differences amongst the portfolios tested, neither separately among the [L] and [LS] portfolios nor jointly, in Tables 4.14, 4.15 and 4.16. However, and consistent among all tests undertaken was the dismal performance of all portfolios formed under the M_1 model which completely ignores security dynamics and goes to the very basis of the importance of modelling them.

At the weekly frequency tested, the security dynamics observed make a statically based weighting scheme inefficient, with immediate value to models which capture these dynamics, and asymmetric risk measures which reflect realistic investor preferences for gains and losses. More importantly, active investors should consider more carefully the rewards to actively managed funds which benchmark against these indices, and passive investors should reconsider the optimality of an index tracking strategy. While it is unlikely that the weighting of popularly tracked indices will ever be anything but the most simple of schemes, there is certainly value in the creation of optimally redesigned tradeable funds based on these indices. Because of the high turnover required by such active models which trade on a weekly cycle, using a subset of very liquid and high capitalization stocks is likely to provide for the lowest cost approach. Future research might also consider a multi period dynamic programming model with recourse, opening up the possibility of using daily data for n-ahead optimization and the consideration of path dependent measures such as conditional drawdown at risk of Chekhlov, Uryasev, and Zabarankin (2005). Additionally, the use of more sophisticated conditional mean forecasts would certainly add value beyond the simple AR model tested.

Conclusion

When security returns are characterized by dynamics such as time varying moments and co-moments, and heterogeneous investors trading at different frequencies with varying degrees of rationality, the investment allocation life cycle of modelling, allocation and risk management must be approached with a new set of tools and mindset. In this thesis I have sought to provide one avenue for rationally approaching this process which is consistent with the observed phenomena of modern markets. To this end, a new model for jointly estimating time varying higher moment dynamics was introduced and shown to provide substantial relative value. New approaches for modelling and optimizing portfolios, including smooth NLP based risk measures and fractional programming, were shown to outperform established market aggregates over a long period, and using weekly returns.

In Chapter 1, substantial evidence was presented of the presence and importance of time varying higher moment dynamics for portfolio and risk management in a univariate context. In the majority of the literature on time varying higher moments, only inference on the in-sample estimation of these models has been supplied with little in the way of either the out-of-sample performance nor the value of such models for risk management. Making use of a feature rich distribution, the GH, the cost of ignoring such dynamics using a number of misspecification tests in-sample as well as tail based tests out-of-sample was shown to be high. The application examined the higher moment dynamics present in a set of 14 international equity indices for a 15 year period which included a range of financial crises. Despite the widely held view that indices, representing market aggregates, provide for an efficient weighting with less risk than individual securities, the presence of higher moment dynamics was found to have a very negative impact on the accurate estimation of risk measures such as VaR when using GARCH models. Instead, accounting for such dynamics with ACD models, using a variety of distributions, the evidence clearly showed that better estimates of VaR were obtained and a simulation study confirmed this. The chapter also addressed some more specific issues of ACD models, often overlooked in the literature, such as the consistency of the parameters under different dynamics and the estimation using a global optimization approach, both of which are important in confidently using and applying these models.

Chapter 2 provided a review of multivariate GARCH models, their feasibility and value in an applied setting. More specifically, an alternative answer was put forward to the that provided by Caporin and McAleer (2012) regarding the value of the 2-stage DCC model versus the more established BEKK. Using an in-sample empirical application with the same dataset of Chapter 1, some evidence for the superior performance of DCC models as a result of the 2 stage estimation process was provided and a reasonable explanation put forward as to why this might be the case in general. A small Monte Carlo application provided further support to this. The chapter also discussed in more detail the problem of imposing covariance targeting for the extensions to the DCC model proposed by Cappiello, Engle, and Sheppard (2006), something which the majority of the literature has simply ignored, despite the fact that it may lead to local rather than global optima. A more promising extension to the model in the form of the DCC Student Copula was analyzed, and shown in Chapter 4 to provide for a very flexible and feasible model for security dynamics.

In Chapter 3, the first feasible multivariate higher moment dynamics model, making use of the independent factor framework, was presented and its unique properties such as closed form higher co-moments and semi-analytic weighted density discussed. Continuing with the same dataset from Chapters 1 and 2, the empirical application both highlighted the practical applications of these properties and the superior performance of this model out-of-sample in both risk and portfolio applications in a crisis rich historical subperiod. Additional evidence was provided by using a weekly dataset, comprising the point in time constituents of the DJIA index in a portfolio application using an extreme loss aversion metric. Beyond the unique properties and dynamics of this model, in a multivariate setting, the additional feature of separability means that this model may be estimated in real time on any number of securities by making use of parallel computational resources, making it ideal for mission critical risk management applications.

In chapter 4, the research focus was shifted from the modelling to the allocation stage in the investment life-cycle. The question, given the dynamics established in previous chapters, of outperforming statically or sub-optimally weighted indices was posed and answered in a large scale portfolio application making use of many new innovative methods. Using a weekly dataset of the point in time constituents of the DJIA index going back to 1970, and already used in an application in Chapter 3, a rolling portfolio estimation and optimization backtest provided strong evidence of the statistically significant superior performance of a range of models against both the Dow and S&P500 indices. The proposition of DeMiguel, Garlappi, and Uppal (2009) of the superiority of the 1/N rule was squarely rejected on both theoretical and empirical grounds, as the rich dynamics observed in the underlying securities can simply not be handled by static weighting schemes. A comparison between models, based on some measures, highlighted the superior performance of the IFACD and Copula DCC models. Asymmetric and tail based measures of risk such as the LPM and MiniMax were also shown to provide some marginal improvement over other measures. Conclusions were drawn from statistically robust measures of comparison such as the test of Sharpe ratio differences of Ledoit and Wolf (2008) and the Model Confidence Set of Hansen, Lunde, and Nason (2011), departing from the usual practise of either simply looking at terminal wealth or reporting the Sharpe ratios without testing their relative significance. The inefficiency of these equity indices, already brought to light by Grinold (1992) and Demey, Maillard, and Roncalli (2010), has serious implications for both active and passive investors who rely on them for different purposes. Whether these models and methods can be leveraged to create optimized index products remains an open question.

There are certainly other avenues and models with which to approach the investment allocation lifestyle. Recent advances in machine learning, have opened up a host of new options for identifying hidden trends and information in large datasets. With regards to the models already used and for very large datasets, dimensionality reduction can be used with the IFACD model via the PCA whitening stage. However, the main bottleneck remains the estimation of the ACD dynamics, and hence alternative approaches to their estimation would provide for the most useful contribution to the model. The combination of traditional fundamental or statistical return factor models could certainly be combined with any of the feasible multivariate GARCH extensions, to create even better models for portfolio allocation. Multi-step dynamic programming with recourse is also an avenue worth exploring, particularly with daily data, opening up the possibility of trading at different frequencies and using path dependent measures such as maximum drawdown, among others. Internalizing transaction and turnover costs in the portfolio process is also a possibility, though this would require a mixed integer approach for the traditional minimum risk problem but a global optimization approach when using fractional programming. Finally, the clear evidence of dynamics not fully explained by traditional models, leading to the increased popularity of new risk measures in risk and portfolio management, necessitates a set of robust tests to evaluate such measures, as exist for testing the differences in Sharpe ratios.

Appendix A

The Generalized Hyperbolic Distribution

A.1 The Standardized GH Density

In order to model zero-mean, unit variance processes, the distribution, which must posses the scaling property, needs to be properly standardized. In the case of the GH distribution, because of the existence of location and scale invariant parameterizations and the possibility of expressing the mean and the variance in terms of one of those parametrizations, namely the (ζ, ρ), the task of standardizing the density can be broken down to one of estimating those 2 parameters, representing a combination of shape and skewness, followed by a series of transformation steps to translate the parameters into the $(\alpha, \beta, \delta, \mu)$ parametrization for which standard formulae exist for the likelihood function. The (ξ, χ) parametrization, which is a simple transformation of the (ζ, ρ), could also be used in the first step and then transformed into the latter before proceeding further. The steps to transforming from the (ζ, ρ) to the $(\alpha, \beta, \delta, \mu)$ parametrization, while at the same time standardizing for zero mean and unit variance are given henceforth. Let X be a random variable distributed as a $GH(\zeta, \rho)$, where

$$\zeta = \delta\sqrt{\alpha^2 - \beta^2}. \qquad \rho = \frac{\beta}{\alpha}, \tag{A.1}$$

Inverting (A.1) we can express α and β in terms of ζ, ρ and δ,

$$\alpha = \frac{\zeta}{\delta\sqrt{1-\rho^2}}, \tag{A.2}$$

$$\beta = \alpha\rho. \tag{A.3}$$

For standardization we require that,

$$E(X) = \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} = \mu + \frac{\beta\delta^2}{\zeta}\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} = 0$$
$$Var(X) = \delta^2\left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\zeta\boldsymbol{K}_{\lambda}(\zeta)} + \frac{\beta^2}{\alpha^2-\beta^2}\left(\frac{\boldsymbol{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} - \left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}\right)^2\right)\right). \tag{A.4}$$

It follows that

$$\mu = -\frac{\beta\delta^2}{\zeta}\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} \tag{A.5}$$
$$\delta = \left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\zeta\boldsymbol{K}_{\lambda}(\zeta)} + \frac{\beta^2}{\alpha^2-\beta^2}\left(\frac{\boldsymbol{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} - \left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}\right)^2\right)\right)^{-0.5}. \tag{A.6}$$

Since we can express, $\beta^2/\left(\alpha^2-\beta^2\right)$ as,

$$\frac{\beta^2}{\alpha^2-\beta^2} = \frac{\rho^2}{(1-\rho^2)}, \tag{A.7}$$

then we can re-write the formula for δ in terms of the parameters ζ and ρ as,

$$\delta = \left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\zeta\boldsymbol{K}_{\lambda}(\zeta)} + \frac{\rho^2}{(1-\rho^2)}\left(\frac{\boldsymbol{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} - \left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}\right)^2\right)\right)^{-0.5}. \tag{A.8}$$

Transforming into the $(\alpha, \beta, \delta, \mu)$ parametrization proceeds by first substituting (A.8) into (A.2) and simplifying,

$$\alpha = \frac{\zeta\left(\frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\zeta\boldsymbol{K}_{\lambda}(\zeta)} + \frac{\rho^2\left(\frac{\boldsymbol{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)} - \frac{(\boldsymbol{K}_{\lambda+1}(\zeta))^2}{(\boldsymbol{K}_{\lambda}(\zeta))^2}\right)}{(1-\rho^2)}\right)^{0.5}}{\sqrt{(1-\rho^2)}},$$
$$= \left(\frac{\frac{\zeta\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}}{(1-\rho^2)}\left(1 + \frac{\zeta\rho^2\left(\frac{\boldsymbol{K}_{\lambda+2}(\zeta)}{\boldsymbol{K}_{\lambda+1}(\zeta)} - \frac{\boldsymbol{K}_{\lambda+1}(\zeta)}{\boldsymbol{K}_{\lambda}(\zeta)}\right)}{(1-\rho^2)}\right)\right)^{0.5}. \tag{A.9}$$

Finally, the rest of the parameters are derived recursively from α and the previous results,

$$\beta = \alpha\rho, \tag{A.10}$$

$$\delta = \frac{\zeta}{\alpha\sqrt{1-\rho^2}}, \tag{A.11}$$

$$\mu = \frac{-\beta\delta^2\boldsymbol{K}_{\lambda+1}(\zeta)}{\zeta\boldsymbol{K}_{\lambda}(\zeta)}. \tag{A.12}$$

For the use of the (ξ,χ) parametrization in estimation, the additional preliminary steps of converting to the (ζ,ρ) are,

$$\zeta = \frac{1}{\xi^2} - 1, \tag{A.13}$$

$$\rho = \frac{\chi}{\xi}. \tag{A.14}$$

A.2 The GH characteristic function

The moment generating function (MGF) of the GH distribution is,

$$\begin{aligned} M_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) &= e^{\mu u} M_{GIG\left(\lambda,\delta\sqrt{\alpha^2-\beta^2}\right)}\left(\frac{u^2}{2}+\beta u\right) \\ &= e^{\mu u}\left(\frac{\alpha^2-\beta^2}{\alpha^2-(\beta+u)^2}\right)^{\lambda/2}\frac{\boldsymbol{K}_{\lambda}\left(\delta\sqrt{\alpha^2-(\beta+u)^2}\right)}{\boldsymbol{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \end{aligned} \tag{A.15}$$

where M_{GIG} represents the moment generating function of the Generalized Inverse Gaussian which forms the mixing distribution in this variance-mean mixture subclass. Powers of the MGF, $M_{GH}(u)^p$, only have the representation in (A.15) for $p = 1$, which means that GH distributions are not closed under convolution with the exception of the NIG, and only in the case when the shape and skew parameters are the same. The MGF of the NIG is,

$$M_{NIG(\alpha,\beta,\delta,\mu)}(u) = e^{\mu u}\frac{e^{\delta\sqrt{\alpha^2-\beta^2}}}{e^{\delta\sqrt{\alpha^2-(\beta+u)^2}}}. \tag{A.16}$$

Powers of p are equivalent in this case to multiplication by p of δ and μ, so that,

$$NIG(\alpha,\beta,\delta_1,\mu_1)\times\ldots\times NIG(\alpha,\beta,\delta_n,\mu_n) = NIG(\alpha,\beta,\delta_1+\ldots+\delta_n,\mu_1+\ldots+\mu_n). \tag{A.17}$$

When the distribution is not closed under convolution, numerical methods are required such as the inversion of the characteristic function by FFT. Because the MGF is a holomorphic function for complex z, with $|z| < \alpha - \beta$, we can obtain the characteristic function of the GH distribution, using the following representation,

$$\phi_{GH}(u) = M_{GH}(\mathrm{i}u), \tag{A.18}$$

so that the characteristic function may be written as,

$$\phi_{GH(\lambda,\alpha,\beta,\delta,\mu)}(u) = e^{\mu \mathrm{i}u} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + \mathrm{i}u)^2} \right)^{\lambda/2} \frac{\boldsymbol{K}_\lambda \left(\delta \sqrt{\alpha^2 - (\beta + \mathrm{i}u)^2} \right)}{\boldsymbol{K}_\lambda \left(\delta \sqrt{\alpha^2 - \beta^2} \right)}. \tag{A.19}$$

and for the NIG this is simplified to,

$$\phi_{NIG(\alpha,\beta,\delta,\mu)}(u) = e^{\mu \mathrm{i}u} \frac{e^{\delta\sqrt{\alpha^2-\beta^2}}}{e^{\delta\sqrt{\alpha^2-(\beta+\mathrm{i}u)^2}}}. \tag{A.20}$$

In order to find the portfolio density in the case of the IFACD model, the characteristic function required for the inversion of the NIG density was already used in Chen, Härdle, and Spokoiny (2010) and given below,

$$\phi_{port}(u) = \exp\left\{ \mathrm{i}u \sum_{j=1}^{d} \bar{\mu}_j + \sum_{j=1}^{d} \bar{\delta}_j \left(\sqrt{\bar{\alpha}_j^2 - \bar{\beta}_j^2} - \sqrt{\bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2} \right) \right\} \tag{A.21}$$

where $\bar{\alpha}_j$, $\bar{\beta}_j$, $\bar{\delta}_j$ and $\bar{\mu}_j$ represent the parameters scaled as described in the main text of the paper. In the case of the GH characteristic function, this is a little more complicated as it involves the evaluation of modified Bessel function of the third kind with complex arguments.[1] Taking logs and summing,

$$\phi_{port}(u) = \exp\Bigg\{ \mathrm{i}u \sum_{j=1}^{d} \Big(\bar{\mu}_j + \frac{\lambda_j}{2} \log\left(\bar{\alpha}_j^2 - \bar{\beta}_j^2\right) - \frac{\lambda_j}{2} \log\left(\bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2\right) + \\ \log\left(\boldsymbol{K}_{\lambda_j}\left(\bar{\delta}_j \sqrt{\bar{\alpha}_j^2 - (\bar{\beta}_j + \mathrm{i}u)^2}\right)\right) - \log\left(\boldsymbol{K}_{\lambda_j}\left(\bar{\delta}_j \sqrt{\bar{\alpha}_j^2 - \bar{\beta}_j^2}\right)\right)\Big)\Bigg\} \tag{A.22}$$

which is more than 30 times slower to evaluate than the equivalent NIG function because of the Bessel function evaluations.

[1]Routines for this exist for example on Netlib, see http://www.netlib.org/amos/zbesk.f

Appendix B

The Student and Skewed Student Distributions

B.1 The Standardized Student Density

The GARCH-Student model was first used described in Bollerslev (1987) as an alternative to the Normal distribution for fitting the standardized innovations. It is described completely by a shape parameter ν, but for standardization we proceed by using its 3 parameter representation as follows:

$$f\left(x\right)=\frac{\mathbf{\Gamma}\left(\frac{\nu+1}{2}\right)}{\sqrt{\beta\nu\pi}\mathbf{\Gamma}\left(\frac{\nu}{2}\right)}\left(1+\frac{\left(x-\alpha\right)^{2}}{\beta\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{B.1}$$

where α, β, and ν are the location, scale[1] and shape parameters respectively, and $\mathbf{\Gamma}$ is the Gamma function. Similar to the Normal and Generalized Error distributions, this is a unimodal and symmetric distribution where the location parameter α is the mean (and mode) of the distribution while the variance is:

$$Var\left(x\right)=\frac{\beta\nu}{\left(\nu-2\right)}. \tag{B.2}$$

[1]In some representations, mostly Bayesian, this is represented in its inverse form to denote the precision.

For the purposes of standardization we require that:

$$\begin{aligned} \mathrm{Var(x)} &= \frac{\beta\nu}{(\nu-2)} = 1 \\ \therefore \beta &= \frac{\nu-2}{\nu} \end{aligned} \tag{B.3}$$

Substituting $\frac{(\nu-2)}{\nu}$ into B.1 we obtain the standardized Student's distribution:

$$f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} f(z) = \frac{1}{\sigma}\frac{\mathbf{\Gamma}\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)\,\pi}\mathbf{\Gamma}\left(\frac{\nu}{2}\right)}\left(1+\frac{z^2}{(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)}. \tag{B.4}$$

The Student distribution has zero skewness and excess kurtosis equal to $6/(\nu-4)$ for $\nu > 4$.

B.2 The Standardized Skewed Student Density

Fernandez and Steel (1998) induced skewness into unimodal and symmetric distributions by introducing inverse scale factors in the positive and negative real half lines. Let z be a zero mean, unit variance i.i.d. random variable with unimodal and symmetric density $g(.)$, $z \sim g(0,1)$, and ϵ the mixture of $|z|$ based on a Bernoulli process w with probability of success $\frac{\xi^2}{1+\xi^2}$[2]:

$$\epsilon = w\xi\,|z| - (1-w)\,\frac{1}{\xi}\,|z|\,. \tag{B.5}$$

The density of $f(\epsilon|\xi)$ is then:

$$f(\epsilon;\xi) = \frac{2}{\xi+\xi^{-1}}\Big(g\left(\epsilon\xi^{-1}\right)H(\epsilon) + g(\epsilon\xi)\,H(-\epsilon)\Big) \tag{B.6}$$

where $\xi \in \mathbb{R}^{+}$[3] and $H(.)$ is the Heaviside step function.[4] Assuming that the r^{th} moment of $g(0,1)$ exists, then its skewed version also has a finite r^{th} moment given by

$$\mathrm{E}\left(\epsilon^{\mathrm{r}}\,|\xi\right) = \mathrm{M_r}\frac{\xi^{\mathrm{r}+1} + \frac{(-1)^{\mathrm{r}}}{\xi^{\mathrm{r}+1}}}{\xi+\xi^{-1}} \tag{B.7}$$

[2]This exposition is based on Lambert and Laurent (2001b).
[3]When $\xi = 1$, the distribution is symmetric.
[4]This is equal to $(1+sgn(\epsilon))/2$, for which a number of smooth approximations exist, such as $\frac{1}{1+e^{-2k\epsilon}}$ for large k

where M_r denotes the absolute moment function of $g(0,1)$, and given by:

$$M_r = 2\int_0^\infty \epsilon^r f(\epsilon;\nu)\,d\epsilon \tag{B.8}$$

so that when $g(0,1)$ is the Student distribution, this simplifies to:

$$M_{r|\nu} = \frac{\mathbf{\Gamma}\left(\frac{\nu-r}{2}\right)\mathbf{\Gamma}\left(\frac{1+r}{2}\right)(\nu-2)^{\frac{1+r}{2}}}{\sqrt{\pi(\nu-2)}\mathbf{\Gamma}\left(\frac{\nu}{2}\right)} \tag{B.9}$$

The mean and variance are then defined as:

$$\begin{aligned} \mathrm{E}(\epsilon\,|\xi) &= \mathrm{M}_1\left(\xi-\xi^{-1}\right) \\ \mathrm{Var}(\epsilon\,|\xi) &= \left(\mathrm{M}_2-\mathrm{M}_1^2\right)\left(\xi^2+\xi^{-2}\right)+2\mathrm{M}_1^2-\mathrm{M}_2 \end{aligned} \tag{B.10}$$

which for the Skew Student distribution, and provided that $\nu > 2$, simplifies to:

$$\begin{aligned} \mathrm{E}(\epsilon\,|\xi\,,\nu) \equiv \bar{\mu} &= \frac{\mathbf{\Gamma}\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\mathbf{\Gamma}\left(\frac{\nu}{2}\right)}\left(\xi-\xi^{-1}\right) \\ \mathrm{Var}(\epsilon\,|\xi\,,\nu) \equiv \bar{\sigma}^2 &= \left(\xi^2+\xi^{-2}-1\right)-\bar{\mu}^2 \end{aligned} \tag{B.11}$$

Finally, the density of the standardized variable, z, is given by:

$$\begin{aligned} &f\left(\frac{(x-\mu)}{\sigma}\,|\xi,\nu\right) = \frac{1}{\sigma}f(z\,|\xi,\nu) \\ &= \frac{2\bar{\sigma}}{\sigma\left(\xi+\xi^{-1}\right)}\left[g\left(\xi\left(\bar{\sigma}z+\bar{\mu}\right)|\nu\right)H\left(-\left(z+\bar{\mu}/\bar{\sigma}\right)\right)+g\left(\xi^{-1}\left(\bar{\sigma}z+\bar{\mu}\right)|\nu\right)H\left(z+\bar{\mu}/\bar{\sigma}\right)\right] \end{aligned} \tag{B.12}$$

When $f(.)$ is the skewed Student distribution, $g(.)$ is the standardized Student distribution given in (B.4). Similar arguments can be used to derive the standardized skew Normal and Generalized Error distributions.

Appendix C

Goodness of Fit and Operational Risk Tests

C.1 Parametric and Non Parametric Density Tests

Consider a random variable $(r.v.)$ r_t such that $r_t\,|\Im_{t-1} \sim f\,(\mu_t, \sigma_t, \omega_t)$, with $f(.)$ being the density, μ_t the conditional mean, σ_t the conditional standard deviation and ω_t any additional distributional parameters, given the information set $\Im_{t-1}$ at time $t-1$. A novel method to analyze how well a conditional density fits the underlying data is through the probability integral transformation (PIT) discussed in Rosenblatt (1952) and defined as:

$$x_t = \int_{-\infty}^{r_t} \hat{f}\,(u)\,du = \hat{F}\,(r_t; \mu_t, \sigma_t, \omega_t) \tag{C.1}$$

which transforms the data r_t, using the estimated distribution $\hat{F}$ conditional parameter into i.i.d. $U(0,1)$ under the correctly specified model. The visual test of Tay, Diebold, and Gunther (1998) and formal test of Berkowitz (2001) is based on this transformation. Because of the difficulty in testing for i.i.d. under $U(0,1)$, Berkowitz transforms the uniform data into conditionally standard normal $N(0,1)$ by applying the quantile normal transformation, Φ^{-1}. A likelihood ratio test of the i.i.d assumption can the be formulated as:

$$LR = -2\left(L\,(0,1,0) - L\left(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}\right)\right) \tag{C.2}$$

where the restriction is of zero mean, unit variance and no autocorrelation, and distributed $\chi^2(3)$. It is also possible to extend the methodology to test for tail fit

More recently, Hong and Li (2005) introduced a nonparametric portmanteau test, building on the work of Ait-Sahalia (1996), which tests the joint hypothesis of i.i.d and $U(0,1)$ for the sequence x_t. As noted by the authors, testing x_t using for instance the Kolmogorov-Smirnov test would only check the U(0,1) assumption under i.i.d. and not the joint assumption of $U(0,1)$ and i.i.d. Their approach is to compare a kernel estimator $\hat{g}_j(x_1, x_2)$ for the joint density $g_j(x_1, x_2)$ of the pair $\{x_t, x_{t-j}\}$ (where j is the lag order) with unity, the product of two $U(0,1)$ densities. Given a sample size n and lag order $j > 0$, their joint density estimator is:

$$\hat{g}_j(x_1, x_2) \equiv (n-j)^{-1} \sum_{t=j+1}^{n} K_h\left(x_1, \hat{X}_t\right) K_h\left(x_2, \hat{X}_{t-j}\right) \tag{C.3}$$

where $\hat{X}_t = X_t\left(\hat{\theta}\right)$, and $\hat{\theta}$ is a $\sqrt{n}$ consistent estimator of θ_0. The function K_h is a boundary modified kernel[1] defined as:

$$K_h(x,y) \equiv \begin{cases} h^{-1}k\left(\frac{x-y}{h}\right)\Big/\int_{-(x/h)}^{1} k(u)\,du, & \text{if} x \in [0,h), \\ h^{-1}k\left(\frac{x-y}{h}\right), & \text{if} x \in [h, 1-h], \\ h^{-1}k\left(\frac{x-y}{h}\right)\Big/\int_{-1}^{(1-x)/h} k(u)\,du, & \text{if} x \in (1-h, 1], \end{cases} \tag{C.4}$$

where $h \equiv h(n)$ is a bandwidth such that $h \to 0$ as $n \to \infty$, and the kernel $k(.)$ is a pre-specified symmetric probability density, which is implemented as suggested by the authors using a quartic kernel,

$$k(u) = \frac{15}{16}\left(1-u^2\right)^2 \mathbf{1}(|u| \leq 1), \tag{C.5}$$

where $\mathbf{1}(.)$ is the indicator function. Their portmanteau test statistic is defined as:

$$\hat{W}(p) \equiv p^{-1/2} \sum_{j=1}^{p} \hat{Q}(j), \tag{C.6}$$

where

$$\hat{Q}(j) \equiv \left[(n-j)\,h\hat{M}(j) - A_h^0\right]\Big/V_0^{1/2}, \tag{C.7}$$

and

$$\hat{M}(j) \equiv \int_0^1 \int_0^1 \left[\hat{g}_j(x_1, x_2) - 1\right]^2 dx_1 dx_2. \tag{C.8}$$

[1] This is a key advancement over the test Ait-Sahalia (1996) which has been shown to produce biased estimates near the boundaries of the data as discussed in Chapman and Pearson (2001).

The centering and scaling factors A_h^0 and V_0 are defined as:

$$\begin{aligned} A_h^0 &\equiv \left[\left(h^{-1}-2\right)\int_{-1}^{1} k^2(u)\,du + 2\int_0^1\int_{-1}^{b} k_b^2(u)\,dudb\right]^2 - 1 \\ V_0 &\equiv 2\left[\int_{-1}^{1}\left[\int_{-1}^{1} k(u+v)\,k(v)\,dv\right]^2 du\right]^2 \end{aligned} \tag{C.9}$$

where,

$$k_b(.) \equiv k(.) \Big/ \int_{-1}^{b} k(v)\,dv. \tag{C.10}$$

Under the correct model specification, the authors show that $\hat{W}(p) \rightarrow N(0,1)$ in distribution. Because negative values of the test statistic only occur under the null hypothesis of a correctly specified model, the authors indicate that only upper tail critical values need be considered. The test is quite robust to model misspecification as parameter uncertainty has no impact on the asymptotic distribution of the test statistic as long as the parameters are $\sqrt{n}$ consistent. Finally, in order to explore possible causes of misspecification when the statistic rejects a model, the authors develop the following test statistic:

$$M(m,l) \equiv \left[\sum_{j=1}^{n-1} w^2(j/p)(n-j)\,\hat{\rho}_{ml}^2(j) - \sum_{j=1}^{n-1} w^2(j/p)\right] \Big/ \left[2\sum_{j=1}^{n-2} w^4(j/p)\right]^{1/2} \tag{C.11}$$

where $\hat{\rho}_{ml}(j)$ is the sample cross-correlation between $\hat{X}_t^m$ and $\hat{X}_{t-|j|}^l$, and $w(.)$ is a weighting function of lag order j, and as suggested by the authors implemented as the Bartlett kernel. As in the $\hat{W}(p)$ statistic, the asymptotic distribution of $M(m,l)$ is $N(0,1)$ and upper critical values should be considered.

C.2 Value at Risk Tests

Consider a random variable $(r.v.)$ r_t such that $r_t\,|\Im_{t-1} \sim f(\mu_t, \sigma_t, \omega_t)$, with $f(.)$ being the density, μ_t the conditional mean, σ_t the conditional standard deviation and ω_t any additional distributional parameters, given the information set $\Im_{t-1}$ at time $t-1$. The Value at Risk measure with coverage probability p is then defined as:

$$\Pr\left(r_t \leqslant q_{t|t-1}(p;\mu_t,\sigma_t,\omega_t)\,|\Im_{t-1}\right) = p \tag{C.12}$$

where $q_{t|t-1}$ is the quantile function of the density. If the density is also location and scale invariant, then the quantile function can also be written as:

$$q_{t|t-1}(p) = \mu_t + q_{t|t-1}(p;0,1,\omega_t)\,\sigma_t. \tag{C.13}$$

Define the sequence of VaR exceedances (or hits) as:

$$HIT_t = \begin{cases} 1, & \text{if } \mathrm{r}_t < -VaR_t\,(p) \\ 0, & \text{otherwise} \end{cases} \tag{C.14}$$

where $t = 1, \ldots, N$. The unconditional coverage (uc), or proportion of failures, test of Kupiec (1995) tests whether the observed frequency of VaR exceedances is consistent with the expected exceedances, defined as $p \times N$, given the chosen quantile and a confidence level. Under the null hypothesis of a correctly specified model, the number of exceedances, $X = \sum HIT$, follows a binomial distribution. A probability below a given significance level leads to a rejection of the null hypothesis. The test is usually conducted as a likelihood ratio test, with the statistic taking the form,

$$LR_{uc} = -2\ln\left(\frac{(1-p)^{N-X}p^X}{\left(1-\frac{X}{N}\right)^{N-X}\left(\frac{X}{N}\right)^X}\right) \tag{C.15}$$

Under the null the test statistic is asymptotically distributed as a χ^2 with 1 degree of freedom. The test does not consider any potential violation of the assumption of the independence of the number of exceedances. The conditional coverage (cc) test of Christoffersen, Hahn, and Inoue (2001) corrects this by jointly testing the frequency as well as the independence of exceedances, assuming that the VaR violation is modelled with a first order Markov chain with switching probability matrix given by:

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{C.16}$$

where π_{ij} is the probability of hit-no hit sequence (i.e. 01 represents no-hit followed by hit on 2 consecutive days). The test is again a likelihood ratio, with the statistic taking the form:

$$LR_{cc} = -2\log\left[\left(\frac{(1-p)^{N-X}p^X}{\left(1-\frac{X}{N}\right)^{N-X}\left(\frac{X}{N}\right)^X}\right) + \left(\frac{(1-\pi)^{\pi_{00}+\pi_{10}}\pi^{\pi_{01}+\pi_{11}}}{(1-\pi_0)^{\pi_{00}}\pi_0{}^{\pi_{01}}(1-\pi_1)^{\pi_{10}}\pi_1{}^{\pi_{11}}}\right)\right] \tag{C.17}$$

where $\pi_0 = \frac{\pi_{01}}{\pi_{00}+\pi_{01}}$, $\pi_1 = \frac{\pi_{11}}{\pi_{10}+\pi_{11}}$ and $\pi = \frac{\pi_{01}+\pi_{11}}{\pi_{00}+\pi_{01}+\pi_{10}+\pi_{11}}$. The null is that the conditional and unconditional coverage are equal to p, and is asymptotically distributed as $\chi^2(2)$.

In a further paper, Christoffersen and Pelletier (2004) consider the duration between VaR violations as a stronger test of the adequacy of a risk model. The duration of time between VaR violations (no-hits) should ideally be independent and not cluster.

Under the null hypothesis of a correctly specified risk model, the no-hit duration should have no memory. Since the only continuous distribution which is memory free is the exponential, the test can conducted on any distribution which embeds the exponential as a restricted case, and a likelihood ratio test then conducted to see whether the restriction holds. Following Christoffersen and Pelletier (2004), the Weibull distribution is used with parameter $b = 1$ representing the case of the exponential, by maximizing the following likelihood:

$$\begin{aligned} LL\left(D;\Theta\right) = C_1 \log S\left(D_1\right) + \left(1 - C_1\right) \log f\left(D_1\right) + \sum_{i=1}^{N_T - 1} \log\left(f\left(D_i\right)\right) \\ + C_{N_T} \log S\left(D_{N_T}\right) + \left(1 - C_{N_T}\right) \log f\left(D_{N_T}\right) \end{aligned} \tag{C.18}$$

where $S\left(D_i\right) = 1 - F\left(D_i\right)$ is the survival function, D the duration between hits, and C_i a series used to denote whether a duration is censored or not. While the Weibull density $f_W\left(D;a,b\right) = a^b b D^{b-1} e^{-(aD)^b}$, has 2 parameters, the parameter a can be calculated given b[2] making the problem quite fast and feasible. The likelihood ratio statistic under a restricted model with $b = 1$ (the exponential distribution) is then distributed $\chi^2(1)$. Because VaR tests deal with the occurrences of hits, they are by definition rather crude measures to compare how well one model has done versus another, particularly with short data sets. The expected shortfall test of McNeil and Frey (2000) measures the mean of the excess shortfall given the VaR violations which should be zero under the null of a correctly specified risk model. Formally, define the 1-step ahead Expected Shortfall as:

$$S_{p,t} = \mu_{t+1} + \frac{\sigma}{p} \int_0^p q\left(x;0,1,\omega_{t+1}\right) dx \tag{C.20}$$

where p is the coverage rate, and q is the quantile function with possible higher moment dynamic forecasts ω_{t+1} and forecast conditional mean and standard deviation μ_{t+1} and σ_{t+1} respectively. Equivalently, Equation C.20 can be represented as $S_{p,t} = \mu_{t+1} + \sigma_{t+1} E\left[Z \left| Z > z_p\right.\right]$, where z_p is the upper $p^t h$ quantile of the marginal distribution of Z_t which will only depend on t if ω_t is time varying (as in ACD models). The shortfall

[2]From Equation (29), p. 92 of their paper this is:

$$\hat{a} = \left(\frac{N_T - C_1 - C_{N_T}}{\sum_{i=1}^{N_T} D_i^b} \right)^{1/b} \tag{C.19}$$

residuals S^*_{t+1} are then defined as:

$$S^*_{t+1} = \frac{(S_{p,t} - r_{t+1})\, I_{[r_{t+1} < VaR(r_{t+1})]}}{\sigma_{t+1}} \tag{C.21}$$

where r_{t+1} is the realized return and I the indicator function representing here the cases where the forecast VaR underestimated the loss (negative return). The null is that these shortfall residuals are i.i.d with mean equal to zero, which can be tested usually by a one sided t-test where the alternative is that the mean is greater than zero. McNeil and Frey (2000) suggest bootstrapping the p-value so as to avoid unnecessary assumptions about the distribution of the shortfall residuals, and all tests using this measure in the thesis have reported the bootstrapped p-value.

C.3 The Model Confidence Set

The Model Confidence Set (*MCS*) procedure of Hansen, Lunde, and Nason (2011) (henceforth *HLN*) provides for a ranking of models given some penalized measure on their relative loss function difference. Define a set M^0 as the original model comparison set with i models and t the time index, and let $L_{i,t}(.)$ be some user specified loss function. The ranking of the models is then based on the relative difference of the pairwise loss function, $d_{ij,t}$:

$$d_{ij,t} = L_{i,t} - L_{j,t}\ \forall i, j \in M^0, \tag{C.22}$$

where it is assumed that $\mu_{ij} \equiv E[d_{ij,t}]$ is finite and does not depend on t, and that i is preferred to j if $\mu_{ij} \leq 0$. The set of models which can then be described as superior is defined as:

$$M^* \equiv \left\{ i \in M^0 : \mu_{ij} \leqslant 0\ \forall j \in M^0 \right\}. \tag{C.23}$$

The determination of M^* is done through a sequence of significance tests with models found to be significantly inferior eliminated from the set. The null hypothesis takes the form:

$$H_{0,M} : \mu_{ij} = 0\ \forall i, j \in M \tag{C.24}$$

with $M \in M^0$, and tested using an equivalence test δ_M. In case of rejection of the null, an elimination rule e_M is then used to identify the model to be removed from the set and the procedure repeated until all inferior models are eliminated. Given a significance level a, the models which are not eliminated are deemed the model confidence set $\hat{M}^*_{1-a}$ with the key theorem of the test, given a set of assumptions on δ_M and e_M, being that

$\lim_{n\to+\infty} P\left(M^* \subset \hat{M}^*_{1-a}\right) \geqslant 1-a$. The actual studentized measure used to compare models is defined as:

$$\frac{\hat{d}_i}{\sqrt{\mathrm{var}\left(\hat{d}_i\right)}} \tag{C.25}$$

with $\hat{d}_i$ derived as:

$$\begin{aligned} \hat{d}_{ij} &\equiv \frac{1}{N}\sum_{t=1}^{N} d_{ij,t} \\ \hat{d}_i &\equiv \frac{1}{m-1}\sum_{j\in M} \hat{d}_{ij} \end{aligned} \tag{C.26}$$

where $\hat{d}_{ij}$ measures the relative performance between models, and $\hat{d}_i$ the measures the relative performance of model i to the average of all the models in M, and the variance of $\hat{d}_i$, $var(\hat{d}_i)$ may be derived by use of the bootstrap. The statistic then used to eliminate inferior models is the range statistic[3] and defined as:

$$T_R = \max_{i,j\in M} \frac{\left|\hat{d}_i\right|}{\sqrt{\mathrm{var}\left(\hat{d}_i\right)}}. \tag{C.27}$$

The asymptotic distribution of T_R, and hence the p-values reported, is obtained via the bootstrap procedure, the validity of which is established in HLN.

[3]Other options are available such as the semi-quadratic statistic, but only the range statistic was used in this Thesis for all tests.

Appendix D

Nonlinear Fractional Programming Portfolios

D.1 General Constraints and Derivatives

The long-short nonlinear fractional programming (*NLFP*) portfolios all share the same constraints (**C**) and are summarized in this section. Vector notation is used were possible, and $\hat{\mathbf{w}}$ represents the $m \times 1$ vector of unconditional weights which when scaled by the auxiliary variable υ from the optimization problem will yield the optimal vector of weights. The upper and lower bounds are imposed via the linear inequality constraint and given below.

D.1.1 Linear Reward Fractional Constraint

$$\mathbf{C}_1 : \hat{\mathbf{w}}'\boldsymbol{\mu} - 1 = 0 \tag{D.1}$$

$$\frac{d\mathbf{C}_1}{d\hat{\mathbf{w}}} = \hat{\mathbf{w}}, \tag{D.2}$$

where $\boldsymbol{\mu}$ is the $m \times 1$ vector of forecast returns.

D.1.2 Leverage Constraint

$$\mathbf{C}_2 : \left[\sqrt{(\hat{\mathbf{w}} + \varepsilon)^2}\right]' \mathbf{1} - l\upsilon = 0 \tag{D.3}$$

$$\frac{d\mathbf{C}_2}{d\hat{\mathbf{w}}} = \frac{\hat{\mathbf{w}}}{\sqrt{(\hat{\mathbf{w}} + \varepsilon)^2}} \tag{D.4}$$

$$\frac{d\mathbf{C}_2}{d\upsilon} = l, \tag{D.5}$$

where l is the leverage, $\mathbf{1}$ an $m \times 1$ vector of ones and ε some very small positive number controlling the degree of error in the absolute value smooth function approximation.

D.1.3 Linear Bounds and Inequalities

In the NLFP setup, the upper and lower bounds on the weights ($\hat{\mathbf{w}}$) are unconstrained in the main routine, so that the limits on the final optimal weights are instead imposed as inequality constraints:

$$C_3 : \mathbf{I}\hat{\mathbf{w}} - \upsilon\mathbf{U} \leq 0 \tag{D.6}$$

$$\frac{d\mathbf{C}_3}{d\hat{\mathbf{w}}} = \mathbf{I} \tag{D.7}$$

$$\frac{d\mathbf{C}_3}{d\upsilon} = -\mathbf{U}, \tag{D.8}$$

where $\mathbf{I}$ is an $m \times m$ identity matrix, and $\mathbf{U}$ an $m \times 1$ vector of the upper bounds on $\mathbf{w}$. Similarly, for the lower bounds given by the $m \times 1$ vector $\mathbf{L}$ the constraint and its derivative are given by:

$$C_4 : \upsilon\mathbf{L} - \mathbf{I}\hat{\mathbf{w}} \leq 0 \tag{D.9}$$

$$\frac{d\mathbf{C}_4}{d\hat{\mathbf{w}}} = -\mathbf{I} \tag{D.10}$$

$$\frac{d\mathbf{C}_4}{d\upsilon} = \mathbf{L}. \tag{D.11}$$

D.2 Objective Functions and Derivatives

For ease of exposition, define the portfolio returns $r_{i,p} = \sum_{j=1}^{m} (r_{i,j}\hat{w}_j)$.

D.2.1 Fractional Mean-Variance Objective

$$P_1 : \min_{\hat{\mathbf{w}},\upsilon} \frac{1}{n-1} \sum_{i=1}^{n} r_{i,p}^2 \tag{D.12}$$

$$\frac{dP_1}{d\hat{w}_j} = \frac{1}{n-1} \sum_{i=1}^{n} 2 r_{i,j} r_{i,p}, \quad \forall j = \{1, \ldots, m\} \tag{D.13}$$

$$\frac{dP_1}{d\upsilon} = 0. \tag{D.14}$$

D.2.2 Fractional Mean-Minimax

$$P_2 : \min_{\hat{\mathbf{w}},\upsilon} \left(\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sqrt{r_{i,p}^2 + \varepsilon} - r_{i,p}}{2} \right)^p \right)^{1/p} \tag{D.15}$$

$$\frac{dP_2}{d\hat{w}_j} = \frac{\sum_{i=1}^{n} \left(p \left(-r_{i,j} + \frac{(r_{i,j} r_{i,p})}{\sqrt{r_{i,p}^2 + \varepsilon}} \right) c_i{}^{p-1} \right) \left(\frac{1}{2^p n} \sum_{i=1}^{n} (c_i{}^p) \right)^{\frac{1}{p-1}}}{2^p n p}, \quad \forall j = \{1, \ldots, m\} \tag{D.16}$$

$$\frac{dP_2}{d\upsilon} = 0, \tag{D.17}$$

where $c_i = \sqrt{r_{i,p}^2 + \varepsilon} - r_{i,p}$.

D.2.3 Fractional Mean-MAD

$$P_3 : \min_{\hat{\mathbf{w}},\upsilon} \frac{1}{n-1} \sum_{i=1}^{n} \sqrt{(r_{i,p} + \varepsilon)^2} \tag{D.18}$$

$$\frac{dP_3}{d\hat{w}_j} = \frac{1}{n-1} \sum_{i=1}^{n} \frac{r_{i,j} (\varepsilon + r_{i,p})}{\left(\sqrt{(r_{i,p} + \varepsilon)^2} \right)}, \quad \forall j = \{1, \ldots, m\} \tag{D.19}$$

$$\frac{dP_3}{d\upsilon} = 0. \tag{D.20}$$

D.2.4 Fractional Mean-LPM

$$P_4 : \min_{\hat{\mathbf{w}},\upsilon} \left(\frac{1}{n2^a} \sum_{i=1}^{n} \left(\sqrt{(\tau - r_{i,p})^2 + \varepsilon} + (\tau - r_{i,p}) \right)^a \right)^{1/a} \tag{D.21}$$

$$\frac{dP_4}{d\hat{w}_j} = -\frac{1}{2^a an} \left(\sum_{i=1}^{n} \frac{\left(d_i + \sqrt{\varepsilon + d_i^2}\right)^a}{2^a n} \right)^{\frac{1}{a}-1} \tag{D.22}$$

$$\times \left(\frac{1}{2^a an} \sum_{i=1}^{n} a \left(r_{i,j} + \frac{(r_{i,j} d_i)}{\sqrt{\varepsilon + d_i^2}} \right) \left(d_i + \sqrt{\varepsilon + d_i^2} \right)^{a-1} \right), \quad \forall j = \{1, \ldots, m\}$$

$$\frac{dP_4}{d\upsilon} = 0, \tag{D.23}$$

where $d_i = \tau - r_{i,p}$. A particular note is merited with regards to the LPM measure and fractional programming. Because of the presence of the threshold parameter in the optimization, this must be scaled by the fractional parameter υ which is possible since the measure is both location and scale invariant. In the optimization exercise undertaken the demeaned scenario was used and hence the threshold was zero which why is the formula in this case is greatly simplified, but in all other cases this needs to be addressed using the relationship in Equation (4.15).

D.2.5 Fractional Mean-CVaR

$$P_5 : \min_{\hat{\mathbf{w}},\upsilon,v} -v + \frac{1}{2na} \sum_{i=1}^{n} \left(\sqrt{d_i^2 + \varepsilon} + d \right) \tag{D.24}$$

$$\frac{dP_5}{d\hat{w}_j} = \frac{1}{2na} \sum_{i=1}^{n} \left(\frac{r_{i,j} \left(r_{i,p} - v \right)}{\sqrt{\varepsilon + \left(r_{i,p} - v \right)^2}} - r_{i,j} \right), \quad \forall j = \{1, \ldots, m\} \tag{D.25}$$

$$\frac{dP_5}{dv} = -\frac{1}{a} \left(\sum_{i=1}^{n} \frac{r_{i,p} - v}{2n\sqrt{\varepsilon + \left(r_{i,p} - v \right)^2}} - 0.5 \right) - 1 \tag{D.26}$$

$$\frac{dP_5}{d\upsilon} = 0, \tag{D.27}$$

where v is the VaR estimated at the a-quantile, and $d_i = v - r_{i,p}$.

Appendix E

Supplemental Tables

E.1 Pairwise P-values of Portfolio SR Differences

Table E.1, E.2 and E.3 provide the pairwise p-values of the Ledoit and Wolf (2008) test under the null $H_0 : SR_i - SR_j = 0$, of the [L] and [LS] portfolio SR differences. There are 6 (risk measures) $\times$ 8 (models) $\times$ 2 ([L] and [LS]) portfolios in total, and split among the 3 tables. Table E.1 displays the comparison of the [L] against [L] portfolios, Table E.2 [LS] against [L] portfolios, and Table E.3 [LS] against [LS] portfolios. For visual clarity, a red colored number denotes failure to reject H_0 at the 10% level.

TABLE E.1: Long-only portfolios: Significance of pairwise Sharpe ratio differences

$R[L]_1 M_2$	0.01																																														
$R[L]_1 M_3$	0.00	0.98																																													
$R[L]_1 M_4$	0.00	0.82	0.84																																												
$R[L]_1 M_5$	0.00	0.94	0.95	0.96																																											
$R[L]_1 M_6$	0.00	0.90	0.80	0.73	0.44																																										
$R[L]_1 M_7$	0.00	0.90	0.82	0.74	0.36	0.97																																									
$R[L]_1 M_8$	0.00	0.97	0.98	0.91	0.92	0.58	0.60																																								
$R[L]_2 M_1$	0.00	0.06	0.01	0.02	0.01	0.00	0.01	0.01																																							
$R[L]_2 M_2$	0.01	0.58	0.91	0.70	0.89	0.94	0.94	0.90	0.04																																						
$R[L]_2 M_3$	0.00	0.93	0.62	0.92	0.99	0.75	0.77	0.97	0.01	0.86																																					
$R[L]_2 M_4$	0.00	0.53	0.69	0.56	0.85	0.66	0.67	0.77	0.04	0.39	0.76																																				
$R[L]_2 M_5$	0.00	1.00	0.97	0.88	0.42	0.60	0.56	0.91	0.01	0.95	0.92	0.78																																			
$R[L]_2 M_6$	0.00	0.96	0.90	0.81	0.57	0.57	0.73	0.75	0.00	1.00	0.85	0.74	0.79																																		
$R[L]_2 M_7$	0.00	0.91	0.84	0.75	0.45	0.95	0.86	0.56	0.01	0.95	0.78	0.66	0.62	0.81																																	
$R[L]_2 M_8$	0.00	0.89	0.87	0.98	0.85	0.33	0.35	0.43	0.01	0.82	0.92	0.89	0.65	0.46	0.28																																
$R[L]_3 M_1$	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00																															
$R[L]_3 M_2$	0.03	0.10	0.14	0.12	0.23	0.12	0.14	0.13	0.21	0.07	0.16	0.16	0.18	0.15	0.12	0.18	0.01																														
$R[L]_3 M_3$	0.01	0.42	0.20	0.35	0.28	0.15	0.16	0.24	0.10	0.37	0.22	0.49	0.24	0.18	0.17	0.28	0.00	0.71																													
$R[L]_3 M_4$	0.01	0.67	0.48	0.60	0.47	0.32	0.33	0.51	0.06	0.61	0.52	0.75	0.42	0.38	0.36	0.55	0.00	0.58	0.65																												
$R[L]_3 M_5$	0.02	0.51	0.29	0.43	0.16	0.11	0.10	0.26	0.16	0.44	0.30	0.56	0.12	0.12	0.13	0.27	0.00	0.93	0.84	0.49																											
$R[L]_3 M_6$	0.00	0.56	0.38	0.53	0.26	0.11	0.11	0.26	0.06	0.51	0.42	0.66	0.22	0.15	0.14	0.34	0.00	0.58	0.75	0.91	0.53																										
$R[L]_3 M_7$	0.01	0.55	0.33	0.48	0.17	0.09	0.07	0.26	0.09	0.50	0.35	0.62	0.13	0.11	0.10	0.29	0.00	0.78	0.96	0.69	0.65	0.72																									
$R[L]_3 M_8$	0.00	0.76	0.66	0.77	0.61	0.44	0.43	0.63	0.04	0.72	0.67	0.87	0.56	0.49	0.47	0.68	0.00	0.55	0.58	0.78	0.18	0.69	0.39																								
$R[L]_4 M_1$	0.36	0.02	0.00	0.01	0.00	0.00	0.00	0.01	0.44	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.12	0.11	0.04	0.03	0.10	0.03	0.06	0.02																							
$R[L]_4 M_2$	0.00	0.65	0.53	0.66	0.69	0.48	0.51	0.63	0.02	0.58	0.57	0.85	0.63	0.56	0.51	0.74	0.00	0.34	0.53	0.79	0.52	0.71	0.61	0.94	0.00																						
$R[L]_4 M_3$	0.00	0.78	0.61	0.83	0.80	0.59	0.60	0.77	0.01	0.72	0.68	0.99	0.75	0.67	0.61	0.87	0.00	0.30	0.38	0.67	0.40	0.59	0.48	0.83	0.00	0.85																					
$R[L]_4 M_4$	0.00	0.78	0.60	0.54	0.62	0.82	0.84	0.65	0.00	0.82	0.53	0.50	0.70	0.77	0.82	0.57	0.00	0.12	0.14	0.28	0.14	0.22	0.17	0.41	0.00	0.31	0.39																				
$R[L]_4 M_5$	0.00	0.60	0.38	0.40	0.12	0.30	0.28	0.27	0.00	0.61	0.34	0.40	0.15	0.23	0.31	0.17	0.00	0.09	0.06	0.09	0.01	0.04	0.01	0.11	0.00	0.21	0.25	0.61																			
$R[L]_4 M_6$	0.00	0.54	0.36	0.47	0.29	0.13	0.16	0.29	0.04	0.48	0.37	0.62	0.25	0.18	0.17	0.35	0.00	0.59	0.79	0.94	0.64	0.98	0.82	0.79	0.02	0.67	0.57	0.17	0.04																		
$R[L]_4 M_7$	0.00	0.95	0.89	0.84	0.78	0.95	0.95	0.86	0.00	0.98	0.84	0.77	0.87	0.95	0.98	0.75	0.00	0.28	0.28	0.39	0.10	0.28	0.14	0.45	0.00	0.61	0.66	0.81	0.34	0.27																	
$R[L]_4 M_8$	0.00	0.76	0.64	0.64	0.49	0.70	0.69	0.59	0.00	0.79	0.58	0.61	0.55	0.63	0.69	0.50	0.00	0.22	0.17	0.22	0.04	0.15	0.06	0.20	0.00	0.40	0.44	0.86	0.72	0.16	0.62																
$R[L]_5 M_1$	0.01	0.04	0.00	0.01	0.01	0.00	0.00	0.01	0.77	0.03	0.00	0.03	0.01	0.00	0.00	0.01	0.01	0.18	0.08	0.06	0.16	0.05	0.08	0.04	0.53	0.01	0.01	0.00	0.00	0.03	0.00	0.00															
$R[L]_5 M_2$	0.01	1.00	0.97	0.81	0.94	0.91	0.90	0.96	0.06	0.11	0.93	0.49	1.00	0.95	0.90	0.89	0.00	0.09	0.41	0.64	0.49	0.55	0.54	0.76	0.02	0.64	0.78	0.78	0.59	0.51	0.94	0.76	0.04														
$R[L]_5 M_3$	0.00	0.89	0.33	0.99	0.96	0.71	0.72	0.91	0.01	0.81	0.33	0.81	0.88	0.81	0.73	0.97	0.00	0.17	0.25	0.56	0.33	0.45	0.38	0.71	0.00	0.62	0.76	0.49	0.33	0.41	0.80	0.56	0.01	0.89													
$R[L]_5 M_4$	0.00	0.55	0.72	0.58	0.86	0.67	0.67	0.79	0.05	0.40	0.78	0.49	0.80	0.73	0.67	0.90	0.00	0.16	0.49	0.74	0.55	0.65	0.60	0.86	0.02	0.84	0.98	0.53	0.42	0.61	0.78	0.61	0.03	0.52	0.82												
$R[L]_5 M_5$	0.00	0.99	0.95	0.87	0.36	0.66	0.63	0.86	0.01	0.97	0.90	0.76	0.71	0.87	0.70	0.60	0.00	0.18	0.21	0.42	0.12	0.20	0.12	0.52	0.00	0.60	0.71	0.71	0.16	0.23	0.90	0.57	0.00	0.99	0.85	0.77											
$R[L]_5 M_6$	0.00	0.95	0.89	0.81	0.55	0.63	0.77	0.72	0.01	0.99	0.83	0.71	0.78	0.54	0.85	0.43	0.00	0.16	0.18	0.37	0.14	0.15	0.10	0.47	0.00	0.56	0.64	0.77	0.24	0.17	0.97	0.64	0.00	0.94	0.79	0.73	0.84										
$R[L]_5 M_7$	0.00	0.89	0.80	0.72	0.36	0.95	0.96	0.51	0.00	0.93	0.75	0.64	0.56	0.70	0.05	0.23	0.00	0.11	0.15	0.33	0.12	0.13	0.09	0.45	0.00	0.48	0.58	0.84	0.31	0.16	0.95	0.71	0.00	0.88	0.71	0.65	0.61	0.75									
$R[L]_5 M_8$	0.00	0.88	0.84	0.96	0.79	0.30	0.33	0.36	0.01	0.81	0.90	0.90	0.62	0.42	0.27	0.22	0.00	0.18	0.31	0.58	0.27	0.35	0.30	0.71	0.01	0.75	0.89	0.55	0.16	0.38	0.73	0.50	0.01	0.87	0.95	0.91	0.56	0.39	0.21								
$R[L]_6 M_1$	0.89	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.28	0.05	0.02	0.01	0.04	0.01	0.01	0.01	0.34	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.11	0.01	0.00	0.00	0.00	0.00	0.00	0.00							
$R[L]_6 M_2$	0.01	0.69	0.99	0.85	0.96	0.88	0.88	0.98	0.07	0.42	0.95	0.56	0.98	0.94	0.90	0.90	0.00	0.10	0.44	0.67	0.52	0.57	0.57	0.77	0.02	0.67	0.80	0.76	0.60	0.54	0.94	0.76	0.04	0.82	0.91	0.58	0.97	0.94	0.87	0.88	0.01						
$R[L]_6 M_3$	0.00	0.95	0.70	0.88	0.97	0.78	0.78	0.99	0.00	0.89	0.77	0.75	0.95	0.87	0.80	0.89	0.00	0.16	0.21	0.49	0.29	0.41	0.34	0.66	0.00	0.55	0.62	0.55	0.35	0.35	0.86	0.59	0.00	0.95	0.45	0.76	0.92	0.86	0.78	0.87	0.00	0.98					
$R[L]_6 M_4$	0.00	0.96	0.99	0.22	0.95	0.82	0.83	0.99	0.01	0.86	0.93	0.42	0.97	0.90	0.85	0.88	0.00	0.08	0.29	0.57	0.38	0.44	0.42	0.71	0.00	0.54	0.71	0.61	0.45	0.39	0.90	0.67	0.01	0.95	0.85	0.44	0.95	0.89	0.81	0.86	0.00	0.98	0.97				
$R[L]_6 M_5$	0.00	1.00	0.98	0.89	0.33	0.58	0.53	0.93	0.01	0.94	0.92	0.79	0.95	0.78	0.60	0.66	0.00	0.19	0.23	0.42	0.13	0.21	0.13	0.55	0.00	0.61	0.73	0.68	0.14	0.25	0.87	0.56	0.01	1.00	0.88	0.80	0.73	0.74	0.52	0.64	0.00	0.99	0.95	0.98			
$R[L]_6 M_6$	0.00	0.83	0.69	0.65	0.30	0.44	0.77	0.43	0.00	0.87	0.64	0.57	0.43	0.34	0.69	0.25	0.00	0.10	0.11	0.26	0.08	0.08	0.06	0.35	0.00	0.40	0.48	0.93	0.37	0.05	0.85	0.77	0.00	0.83	0.58	0.58	0.48	0.37	0.77	0.22	0.00	0.82	0.66	0.73	0.41		
$R[L]_6 M_7$	0.00	0.76	0.59	0.57	0.16	0.49	0.26	0.29	0.00	0.78	0.55	0.50	0.23	0.32	0.29	0.12	0.00	0.08	0.10	0.25	0.07	0.07	0.04	0.32	0.00	0.33	0.41	0.95	0.52	0.07	0.69	0.88	0.00	0.75	0.51	0.53	0.26	0.35	0.34	0.12	0.00	0.75	0.56	0.63	0.23	0.70	
$R[L]_6 M_8$	0.00	0.77	0.65	0.57	0.42	0.72	0.76	0.26	0.00	0.82	0.59	0.47	0.52	0.60	0.67	0.15	0.00	0.05	0.14	0.33	0.14	0.12	0.13	0.42	0.00	0.34	0.47	0.97	0.50	0.09	0.82	0.82	0.00	0.77	0.54	0.49	0.56	0.62	0.76	0.14	0.00	0.75	0.61	0.64	0.51	0.92	0.82

Note: The table displays the Ledoit and Wolf (2008) p-values of the differences in Sharpe ratios of the long-only [L] portfolios formed under 6 different measures of risk and deviation ($R[L]_1$ to $R[L]_6$) and 8 models for the conditional dynamics (M_1 to M_8), under the null that the differences are equal to zero. Red colored cells denote significance at at the 10% level.

TABLE E.2: Long&Long/Short portfolios: Significance of pairwise Sharpe ratio differences

$R[LS]_1M_1$	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_1M_2$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_1M_3$	0.00	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.01	0.02	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01
$R[LS]_1M_4$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_1M_5$	0.00	0.17	0.12	0.12	0.08	0.11	0.12	0.12	0.01	0.18	0.12	0.12	0.10	0.11	0.11	0.10	0.00	0.04	0.04	0.05	0.03	0.04	0.03	0.06	0.00	0.09	0.09	0.16	0.20	0.04	0.12	0.16	0.01	0.17	0.11	0.12	0.10	0.11	0.13	0.09	0.00	0.18	0.12	0.14	0.09	0.12	0.15	0.14
$R[LS]_1M_6$	0.00	0.17	0.10	0.11	0.07	0.10	0.09	0.09	0.01	0.17	0.09	0.10	0.08	0.10	0.09	0.08	0.00	0.04	0.03	0.04	0.02	0.04	0.02	0.05	0.00	0.07	0.07	0.15	0.18	0.04	0.10	0.15	0.00	0.17	0.09	0.11	0.08	0.09	0.10	0.09	0.00	0.15	0.10	0.12	0.08	0.11	0.12	0.13
$R[LS]_1M_7$	0.00	0.24	0.16	0.16	0.13	0.15	0.15	0.14	0.01	0.24	0.16	0.16	0.14	0.14	0.14	0.13	0.00	0.07	0.06	0.08	0.04	0.06	0.05	0.08	0.01	0.11	0.12	0.21	0.26	0.06	0.16	0.22	0.01	0.22	0.14	0.17	0.14	0.15	0.16	0.13	0.00	0.23	0.15	0.18	0.13	0.17	0.20	0.19
$R[LS]_1M_8$	0.00	0.19	0.13	0.14	0.11	0.13	0.13	0.12	0.01	0.20	0.12	0.14	0.12	0.13	0.13	0.12	0.00	0.05	0.05	0.07	0.03	0.05	0.05	0.07	0.01	0.11	0.10	0.18	0.23	0.06	0.14	0.20	0.01	0.19	0.12	0.14	0.12	0.12	0.14	0.11	0.00	0.20	0.13	0.14	0.10	0.15	0.18	0.16
$R[LS]_2M_1$	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.58	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_2M_2$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_2M_3$	0.00	0.02	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.01
$R[LS]_2M_4$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_2M_5$	0.00	0.17	0.10	0.11	0.07	0.09	0.10	0.09	0.00	0.16	0.09	0.11	0.08	0.09	0.10	0.08	0.00	0.04	0.04	0.04	0.02	0.03	0.03	0.05	0.00	0.08	0.08	0.15	0.18	0.04	0.10	0.14	0.00	0.15	0.08	0.11	0.08	0.09	0.11	0.07	0.00	0.17	0.09	0.11	0.08	0.11	0.13	0.12
$R[LS]_2M_6$	0.00	0.14	0.08	0.09	0.07	0.08	0.08	0.08	0.00	0.14	0.08	0.08	0.06	0.07	0.08	0.06	0.00	0.04	0.03	0.04	0.02	0.03	0.02	0.03	0.00	0.06	0.06	0.12	0.15	0.03	0.09	0.12	0.00	0.14	0.08	0.09	0.08	0.07	0.08	0.06	0.00	0.14	0.08	0.09	0.06	0.08	0.10	0.10
$R[LS]_2M_7$	0.00	0.17	0.10	0.13	0.08	0.11	0.10	0.10	0.01	0.17	0.10	0.12	0.09	0.10	0.11	0.09	0.00	0.04	0.04	0.05	0.02	0.04	0.03	0.06	0.00	0.08	0.09	0.16	0.20	0.04	0.11	0.15	0.01	0.15	0.10	0.12	0.11	0.10	0.10	0.09	0.00	0.16	0.10	0.13	0.09	0.12	0.14	0.14
$R[LS]_2M_8$	0.01	0.25	0.18	0.19	0.15	0.18	0.18	0.16	0.02	0.25	0.17	0.18	0.16	0.18	0.19	0.16	0.00	0.07	0.08	0.09	0.06	0.08	0.07	0.10	0.01	0.14	0.14	0.26	0.30	0.08	0.19	0.24	0.02	0.24	0.17	0.19	0.16	0.18	0.19	0.14	0.00	0.23	0.16	0.19	0.16	0.21	0.22	0.21
$R[LS]_3M_1$	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_3M_2$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_3M_3$	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01
$R[LS]_3M_4$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_3M_5$	0.00	0.19	0.12	0.13	0.09	0.12	0.12	0.10	0.00	0.18	0.12	0.12	0.09	0.10	0.12	0.09	0.00	0.04	0.05	0.05	0.03	0.05	0.03	0.06	0.01	0.09	0.09	0.18	0.20	0.05	0.12	0.17	0.01	0.18	0.11	0.12	0.11	0.11	0.12	0.09	0.00	0.18	0.12	0.14	0.10	0.13	0.14	0.14
$R[LS]_3M_6$	0.00	0.16	0.09	0.10	0.06	0.08	0.09	0.08	0.01	0.15	0.09	0.09	0.08	0.08	0.09	0.07	0.00	0.03	0.03	0.04	0.02	0.03	0.02	0.04	0.00	0.08	0.07	0.14	0.16	0.03	0.09	0.13	0.00	0.14	0.08	0.11	0.07	0.08	0.09	0.07	0.00	0.15	0.09	0.10	0.07	0.10	0.12	0.11
$R[LS]_3M_7$	0.00	0.15	0.09	0.10	0.07	0.09	0.09	0.08	0.01	0.15	0.09	0.11	0.07	0.08	0.09	0.06	0.00	0.03	0.03	0.04	0.02	0.03	0.02	0.04	0.00	0.07	0.07	0.14	0.16	0.03	0.09	0.12	0.00	0.14	0.09	0.10	0.08	0.08	0.08	0.07	0.00	0.16	0.09	0.11	0.07	0.10	0.11	0.10
$R[LS]_3M_8$	0.00	0.22	0.16	0.16	0.14	0.17	0.17	0.16	0.02	0.22	0.16	0.17	0.15	0.15	0.16	0.14	0.00	0.07	0.07	0.09	0.05	0.07	0.05	0.09	0.01	0.14	0.13	0.24	0.29	0.07	0.18	0.23	0.01	0.22	0.15	0.17	0.16	0.15	0.17	0.14	0.00	0.22	0.16	0.19	0.14	0.17	0.21	0.19
$R[LS]_4M_1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_4M_2$	0.00	0.06	0.02	0.02	0.02	0.03	0.03	0.02	0.00	0.06	0.02	0.03	0.03	0.03	0.03	0.02	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.03	0.06	0.01	0.02	0.04	0.00	0.06	0.02	0.02	0.03	0.03	0.03	0.02	0.00	0.06	0.02	0.02	0.03	0.03	0.04	0.03
$R[LS]_4M_3$	0.00	0.05	0.02	0.02	0.02	0.03	0.03	0.02	0.00	0.05	0.01	0.02	0.02	0.03	0.03	0.02	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.03	0.06	0.01	0.02	0.04	0.00	0.04	0.01	0.02	0.03	0.02	0.03	0.02	0.00	0.05	0.01	0.02	0.02	0.03	0.04	0.03
$R[LS]_4M_4$	0.00	0.02	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.02	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.01	0.01
$R[LS]_4M_5$	0.01	0.35	0.26	0.27	0.22	0.28	0.28	0.26	0.02	0.34	0.26	0.25	0.24	0.26	0.27	0.23	0.00	0.11	0.12	0.14	0.08	0.12	0.11	0.15	0.01	0.22	0.21	0.35	0.43	0.12	0.28	0.35	0.02	0.34	0.24	0.26	0.25	0.25	0.28	0.22	0.01	0.34	0.26	0.29	0.25	0.30	0.33	0.31
$R[LS]_4M_6$	0.01	0.47	0.41	0.39	0.34	0.42	0.41	0.38	0.04	0.49	0.38	0.37	0.38	0.39	0.41	0.35	0.00	0.18	0.20	0.23	0.13	0.20	0.17	0.24	0.02	0.32	0.34	0.50	0.63	0.19	0.41	0.54	0.03	0.47	0.37	0.37	0.38	0.40	0.42	0.35	0.01	0.46	0.39	0.42	0.38	0.45	0.50	0.47
$R[LS]_4M_7$	0.01	0.37	0.31	0.30	0.26	0.32	0.31	0.29	0.02	0.38	0.28	0.28	0.28	0.29	0.31	0.27	0.00	0.12	0.14	0.16	0.09	0.14	0.12	0.18	0.01	0.24	0.25	0.38	0.49	0.14	0.31	0.41	0.02	0.36	0.27	0.28	0.27	0.28	0.31	0.24	0.01	0.38	0.29	0.30	0.28	0.32	0.38	0.36
$R[LS]_4M_8$	0.01	0.44	0.36	0.35	0.32	0.37	0.38	0.34	0.03	0.45	0.35	0.35	0.35	0.35	0.37	0.31	0.00	0.16	0.17	0.21	0.13	0.19	0.15	0.22	0.02	0.28	0.28	0.44	0.56	0.18	0.37	0.48	0.03	0.44	0.33	0.35	0.34	0.38	0.38	0.32	0.01	0.43	0.35	0.37	0.34	0.41	0.45	0.42
$R[LS]_5M_1$	0.32	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.04	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.87	0.04	0.02	0.01	0.03	0.01	0.01	0.01	0.18	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.06	0.02	0.00	0.01	0.00	0.00	0.00	0.01	0.31	0.02	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_5M_2$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_5M_3$	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.02	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01
$R[LS]_5M_4$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_5M_5$	0.00	0.17	0.10	0.11	0.08	0.10	0.10	0.10	0.01	0.17	0.10	0.11	0.09	0.10	0.11	0.09	0.00	0.04	0.04	0.05	0.02	0.04	0.02	0.05	0.00	0.08	0.08	0.15	0.19	0.04	0.10	0.16	0.00	0.16	0.09	0.11	0.09	0.10	0.10	0.08	0.00	0.18	0.11	0.12	0.09	0.11	0.13	0.13
$R[LS]_5M_6$	0.00	0.18	0.11	0.11	0.08	0.11	0.10	0.10	0.01	0.18	0.09	0.11	0.09	0.10	0.11	0.08	0.00	0.04	0.04	0.05	0.03	0.04	0.03	0.06	0.00	0.09	0.08	0.15	0.20	0.04	0.12	0.15	0.00	0.17	0.10	0.11	0.08	0.09	0.11	0.08	0.00	0.16	0.10	0.12	0.09	0.12	0.13	0.13
$R[LS]_5M_7$	0.00	0.18	0.11	0.11	0.08	0.10	0.10	0.09	0.01	0.16	0.09	0.11	0.09	0.09	0.10	0.08	0.00	0.03	0.04	0.04	0.02	0.04	0.03	0.05	0.00	0.08	0.08	0.15	0.18	0.04	0.11	0.15	0.00	0.16	0.09	0.12	0.08	0.09	0.10	0.09	0.00	0.17	0.11	0.12	0.09	0.11	0.14	0.13
$R[LS]_5M_8$	0.01	0.27	0.20	0.20	0.17	0.21	0.20	0.18	0.02	0.27	0.19	0.20	0.18	0.18	0.20	0.18	0.00	0.08	0.09	0.12	0.06	0.09	0.07	0.12	0.01	0.17	0.16	0.28	0.33	0.10	0.21	0.27	0.02	0.26	0.19	0.20	0.18	0.19	0.21	0.17	0.01	0.27	0.19	0.22	0.18	0.22	0.26	0.24
$R[LS]_6M_1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_6M_2$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_6M_3$	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.01	0.01	0.01
$R[LS]_6M_4$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$R[LS]_6M_5$	0.00	0.19	0.12	0.13	0.10	0.12	0.11	0.11	0.01	0.18	0.11	0.13	0.11	0.12	0.12	0.10	0.00	0.04	0.04	0.05	0.03	0.05	0.03	0.06	0.00	0.09	0.09	0.17	0.21	0.04	0.12	0.18	0.00	0.19	0.11	0.12	0.10	0.12	0.13	0.09	0.00	0.18	0.11	0.14	0.09	0.13	0.16	0.15
$R[LS]_6M_6$	0.00	0.20	0.12	0.12	0.09	0.12	0.12	0.10	0.01	0.19	0.11	0.12	0.10	0.09	0.11	0.10	0.00	0.04	0.04	0.05	0.02	0.05	0.03	0.06	0.00	0.09	0.08	0.16	0.20	0.04	0.12	0.17	0.01	0.19	0.10	0.13	0.10	0.11	0.12	0.09	0.00	0.18	0.11	0.13	0.10	0.12	0.15	0.14
$R[LS]_6M_7$	0.00	0.19	0.12	0.13	0.09	0.12	0.12	0.12	0.01	0.19	0.11	0.13	0.11	0.12	0.11	0.10	0.00	0.05	0.05	0.05	0.03	0.04	0.03	0.06	0.00	0.09	0.09	0.17	0.21	0.04	0.12	0.16	0.01	0.18	0.10	0.12	0.10	0.10	0.12	0.09	0.00	0.19	0.11	0.13	0.10	0.12	0.15	0.15
$R[LS]_6M_8$	0.00	0.19	0.13	0.14	0.10	0.14	0.13	0.13	0.01	0.19	0.12	0.14	0.11	0.12	0.14	0.11	0.00	0.05	0.05	0.06	0.03	0.05	0.03	0.07	0.01	0.10	0.10	0.19	0.22	0.06	0.13	0.18	0.01	0.19	0.11	0.14	0.12	0.12	0.14	0.10	0.00	0.20	0.12	0.14	0.11	0.14	0.18	0.15

Note: The table displays the Ledoit and Wolf (2008) p-values of the differences in Sharpe ratios of the long-short [LS] with 2× leverage versus long-only [L] portfolios formed under 6 different measures of risk and deviation ($R[LS]_1$ to $R[LS]_6$ and $R[L]_1$ to $R[L]_6$ respectively) and 8 models for the conditional dynamics (M_1 to M_8), under the null that the differences are equal to zero. Red colored cells denote significance at at the 10% level. The column headings for the pairwise comparison, not displayed here, are equivalent to the row headings of Table E.1.

Table E.3: Long-Short portfolios: Significance of pairwise Sharpe ratio differences

$R[LS]_1M_2$	0.00																																														
$R[LS]_1M_3$	0.00	0.13																																													
$R[LS]_1M_4$	0.00	0.90	0.10																																												
$R[LS]_1M_5$	0.00	0.04	0.24	0.05																																											
$R[LS]_1M_6$	0.00	0.05	0.29	0.06	0.79																																										
$R[LS]_1M_7$	0.00	0.03	0.17	0.04	0.56	0.41																																									
$R[LS]_1M_8$	0.00	0.08	0.31	0.09	0.99	0.80	0.59																																								
$R[LS]_2M_1$	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00																																							
$R[LS]_2M_2$	0.00	0.39	0.25	0.62	0.07	0.07	0.04	0.11	0.00																																						
$R[LS]_2M_3$	0.00	0.26	0.35	0.21	0.18	0.20	0.12	0.23	0.00	0.43																																					
$R[LS]_2M_4$	0.00	0.91	0.13	0.59	0.06	0.08	0.04	0.10	0.00	0.77	0.26																																				
$R[LS]_2M_5$	0.00	0.06	0.29	0.06	0.54	0.95	0.35	0.77	0.00	0.08	0.22	0.08																																			
$R[LS]_2M_6$	0.00	0.09	0.38	0.10	0.42	0.42	0.17	0.47	0.00	0.12	0.28	0.11	0.62																																		
$R[LS]_2M_7$	0.00	0.06	0.27	0.07	0.86	0.91	0.28	0.86	0.00	0.08	0.20	0.08	0.86	0.49																																	
$R[LS]_2M_8$	0.00	0.08	0.27	0.09	0.75	0.60	0.93	0.53	0.00	0.11	0.22	0.11	0.59	0.39	0.61																																
$R[LS]_3M_1$	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00																															
$R[LS]_3M_2$	0.00	0.58	0.09	0.97	0.04	0.04	0.02	0.07	0.00	0.05	0.18	0.76	0.04	0.07	0.05	0.07	0.00																														
$R[LS]_3M_3$	0.00	0.24	0.19	0.19	0.16	0.20	0.12	0.23	0.00	0.40	0.97	0.24	0.21	0.28	0.21	0.21	0.00	0.15																													
$R[LS]_3M_4$	0.00	0.91	0.12	0.50	0.06	0.08	0.04	0.10	0.00	0.78	0.26	0.97	0.07	0.10	0.07	0.10	0.00	0.75	0.23																												
$R[LS]_3M_5$	0.00	0.04	0.23	0.06	0.97	0.76	0.54	0.99	0.00	0.07	0.17	0.06	0.15	0.36	0.83	0.75	0.00	0.04	0.17	0.06																											
$R[LS]_3M_6$	0.00	0.07	0.33	0.08	0.51	0.54	0.22	0.59	0.00	0.10	0.24	0.10	0.78	0.54	0.62	0.45	0.00	0.06	0.25	0.09	0.45																										
$R[LS]_3M_7$	0.00	0.08	0.34	0.10	0.52	0.72	0.06	0.54	0.00	0.12	0.26	0.11	0.73	0.85	0.10	0.41	0.00	0.07	0.26	0.11	0.45	0.96																									
$R[LS]_3M_8$	0.00	0.08	0.27	0.09	0.79	0.64	0.85	0.57	0.00	0.10	0.23	0.11	0.60	0.37	0.64	0.67	0.00	0.06	0.22	0.12	0.81	0.45	0.41																								
$R[LS]_4M_1$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																							
$R[LS]_4M_2$	0.00	0.01	0.51	0.06	0.48	0.54	0.38	0.54	0.00	0.04	0.36	0.07	0.57	0.65	0.54	0.48	0.00	0.01	0.33	0.07	0.49	0.62	0.65	0.50	0.00																						
$R[LS]_4M_3$	0.00	0.12	0.57	0.11	0.42	0.46	0.32	0.46	0.00	0.19	0.32	0.13	0.49	0.56	0.46	0.42	0.00	0.09	0.31	0.12	0.43	0.53	0.56	0.42	0.00	0.82																					
$R[LS]_4M_4$	0.00	0.50	0.66	0.34	0.18	0.21	0.14	0.24	0.00	0.68	0.86	0.47	0.21	0.27	0.22	0.22	0.00	0.43	0.85	0.46	0.17	0.23	0.26	0.21	0.00	0.32	0.48																				
$R[LS]_4M_5$	0.00	0.01	0.08	0.01	0.17	0.19	0.47	0.30	0.00	0.02	0.06	0.01	0.12	0.10	0.20	0.51	0.00	0.01	0.06	0.02	0.19	0.13	0.13	0.45	0.00	0.23	0.18	0.06																			
$R[LS]_4M_6$	0.00	0.00	0.03	0.00	0.07	0.03	0.18	0.12	0.00	0.00	0.02	0.00	0.05	0.02	0.07	0.25	0.00	0.00	0.02	0.00	0.07	0.02	0.03	0.22	0.00	0.09	0.08	0.02	0.57																		
$R[LS]_4M_7$	0.00	0.01	0.07	0.01	0.24	0.17	0.29	0.29	0.00	0.01	0.04	0.01	0.15	0.10	0.17	0.50	0.00	0.01	0.04	0.01	0.21	0.11	0.09	0.44	0.00	0.17	0.14	0.05	0.86	0.69																	
$R[LS]_4M_8$	0.00	0.01	0.06	0.01	0.11	0.07	0.23	0.10	0.00	0.01	0.04	0.01	0.07	0.04	0.08	0.27	0.00	0.01	0.04	0.01	0.10	0.04	0.05	0.23	0.00	0.15	0.12	0.04	0.70	0.84	0.84																
$R[LS]_5M_1$	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00															
$R[LS]_5M_2$	0.00	0.78	0.16	0.79	0.05	0.06	0.03	0.09	0.00	0.38	0.31	0.99	0.06	0.09	0.07	0.09	0.00	0.23	0.28	1.00	0.04	0.08	0.09	0.08	0.00	0.02	0.14	0.55	0.01	0.00	0.01	0.01	0.00														
$R[LS]_5M_3$	0.00	0.19	0.59	0.15	0.22	0.22	0.14	0.26	0.00	0.34	0.58	0.18	0.25	0.31	0.24	0.25	0.00	0.12	0.48	0.18	0.19	0.28	0.31	0.24	0.00	0.43	0.42	0.77	0.07	0.03	0.05	0.04	0.00	0.23													
$R[LS]_5M_4$	0.00	0.90	0.14	0.55	0.07	0.07	0.04	0.10	0.00	0.81	0.29	0.89	0.08	0.11	0.09	0.11	0.00	0.73	0.26	0.92	0.07	0.10	0.11	0.11	0.00	0.07	0.13	0.49	0.02	0.00	0.01	0.01	0.00	0.98	0.22												
$R[LS]_5M_5$	0.00	0.06	0.27	0.07	0.73	0.95	0.43	0.85	0.00	0.08	0.19	0.08	0.76	0.55	0.96	0.65	0.00	0.04	0.19	0.07	0.53	0.70	0.67	0.67	0.00	0.53	0.47	0.20	0.14	0.06	0.16	0.09	0.00	0.06	0.23	0.07											
$R[LS]_5M_6$	0.00	0.05	0.27	0.06	0.83	0.91	0.41	0.85	0.00	0.07	0.20	0.07	0.88	0.11	0.97	0.62	0.00	0.04	0.20	0.07	0.80	0.27	0.62	0.66	0.00	0.54	0.46	0.21	0.20	0.04	0.17	0.08	0.00	0.06	0.23	0.08	0.98										
$R[LS]_5M_7$	0.00	0.06	0.29	0.08	0.81	0.99	0.26	0.80	0.00	0.09	0.21	0.09	0.94	0.58	0.77	0.57	0.00	0.05	0.22	0.08	0.76	0.71	0.39	0.60	0.00	0.56	0.48	0.23	0.19	0.05	0.16	0.07	0.00	0.07	0.25	0.10	0.97	0.94									
$R[LS]_5M_8$	0.00	0.06	0.24	0.07	0.59	0.47	0.90	0.34	0.00	0.09	0.19	0.08	0.45	0.28	0.44	0.47	0.00	0.06	0.18	0.08	0.60	0.33	0.30	0.35	0.00	0.42	0.37	0.19	0.61	0.33	0.57	0.36	0.00	0.07	0.21	0.08	0.51	0.47	0.41								
$R[LS]_6M_1$	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00							
$R[LS]_6M_2$	0.00	0.55	0.19	0.76	0.05	0.06	0.03	0.09	0.00	0.60	0.34	0.95	0.07	0.10	0.06	0.09	0.00	0.24	0.31	0.95	0.05	0.08	0.10	0.09	0.00	0.01	0.13	0.58	0.01	0.00	0.01	0.01	0.00	0.88	0.24	0.97	0.07	0.06	0.08	0.07	0.00						
$R[LS]_6M_3$	0.00	0.18	0.34	0.14	0.21	0.24	0.15	0.26	0.00	0.33	0.61	0.19	0.25	0.32	0.25	0.26	0.00	0.13	0.44	0.17	0.20	0.28	0.30	0.24	0.00	0.41	0.38	0.76	0.07	0.03	0.05	0.04	0.00	0.23	0.97	0.20	0.24	0.24	0.25	0.20	0.00	0.25					
$R[LS]_6M_4$	0.00	0.68	0.05	0.32	0.04	0.05	0.03	0.06	0.00	0.43	0.12	0.20	0.05	0.07	0.05	0.08	0.00	0.80	0.12	0.10	0.04	0.07	0.06	0.07	0.00	0.03	0.06	0.18	0.01	0.00	0.01	0.01	0.00	0.60	0.09	0.19	0.05	0.05	0.06	0.05	0.00	0.55	0.08				
$R[LS]_6M_5$	0.00	0.04	0.22	0.04	0.80	0.71	0.61	0.95	0.00	0.06	0.16	0.06	0.33	0.35	0.78	0.79	0.00	0.03	0.16	0.06	0.83	0.44	0.42	0.84	0.00	0.48	0.40	0.17	0.15	0.07	0.22	0.12	0.00	0.05	0.20	0.06	0.58	0.75	0.71	0.64	0.00	0.05	0.19	0.04			
$R[LS]_6M_6$	0.00	0.04	0.24	0.05	1.00	0.46	0.53	1.00	0.00	0.07	0.17	0.06	0.72	0.16	0.84	0.74	0.00	0.03	0.18	0.05	0.99	0.16	0.50	0.79	0.00	0.49	0.40	0.18	0.24	0.02	0.21	0.11	0.00	0.05	0.20	0.06	0.81	0.70	0.78	0.58	0.00	0.05	0.20	0.04	0.93		
$R[LS]_6M_7$	0.00	0.04	0.24	0.05	0.96	0.75	0.10	0.98	0.00	0.07	0.17	0.06	0.69	0.41	0.75	0.76	0.00	0.04	0.18	0.06	0.97	0.51	0.25	0.82	0.00	0.48	0.40	0.18	0.28	0.08	0.11	0.12	0.00	0.05	0.20	0.07	0.80	0.80	0.69	0.61	0.00	0.05	0.20	0.04	0.97	0.96	
$R[LS]_6M_8$	0.00	0.06	0.28	0.07	0.98	0.77	0.58	0.95	0.00	0.10	0.22	0.09	0.73	0.43	0.82	0.59	0.00	0.05	0.21	0.09	0.99	0.53	0.49	0.61	0.00	0.52	0.45	0.21	0.30	0.11	0.28	0.06	0.00	0.08	0.24	0.09	0.83	0.82	0.76	0.39	0.00	0.08	0.25	0.06	0.96	0.98	0.99

Note: The table displays the Ledoit and Wolf (2008) p-values of the differences in Sharpe ratios of the long-short [LS] portfolios with 2× leverage formed under 6 different measures of risk and deviation ($R[LS]_1$ to $R[LS]_6$) and 8 models for the conditional dynamics (M_1 to M_8), under the null that the differences are equal to zero. Red colored cells denote significance at at the 10% level.

Appendix F

Software

F.1 Univariate GARCH and ACD Models

The ACD models were estimated by modifying the **rugarch** package of Ghalanos (2012c) and available on the Comprehensive R Archive Network (*CRAN*):
http://cran.r-project.org/web/packages/rugarch/index.html. A vignette of the package, *Introduction to the rugarch package*, provides comprehensive details of the models and methods used. For global optimization problems, the multistart *gosolnp* solver found in the **solnp** package of Ghalanos and Theussl (2012) is already linked with the **rugarch** package and includes parallel estimation for use in multicore computer systems.

F.2 Multivariate GARCH and ACD models

The diagonal BEKK and GDCC models, together with the different conditional distributions used, were estimated in Matlab using a combination of the nonlinear solver *fmincon*, the simulated annealing solver*simulannealbnd*, and the direct search algorithm found in *patternsearch*. Scalar DCC and the Copula DCC model described in Chapter 2 were estimated using the multivariate GARCH package **rmgarch** of Ghalanos (2012b) which is available on the R-Forge development repository:
http://rgarch.r-forge.r-project.org/. This package also includes the GO-GARCH models with MVN, maNIG and maGH distributions, along with a complete set of methods and functions such as the FFT based approach to obtain weighted margins, the conditional co-moments tensors etc. In addition, 2 ICA algorithms are included,

the FASTICA of Hyvärinen and Oja (2000) and the RADICAL of Learned-Miller and Fisher III (2003), with a range of locally implemented options such as covariance shrinkage using the method of Ledoit and Wolf (2003), and dimensionality reduction in the PCA stage. To estimate and work with the IFACD model, the GO-GARCH models in the **rmgarch** package were modified to account for the use of time varying higher moments and linked to the modifications made in the **rugarch** package for the estimation of ACD models.

F.3 Portfolio Optimization

All measures and methods described in Chapter 4 are included in the **parma** package of Ghalanos (2012a) and available to download from the R-Forge development repository:
`http://r-forge.r-project.org/projects/parma/`. A vignette, *Portfolio Optimization in parma*, provides comprehensive details of the package with examples.

Bibliography

Aas, K., and I. Haff (2006): “The generalized hyperbolic skew Student’s t-distribution,” *Journal of Financial Econometrics*, 4(2), 275–309.

Acerbi, C. (2002): “Spectral measures of risk: A coherent representation of subjective risk aversion,” *Journal of Banking and Finance*, 26(7), 1505–1518.

Acerbi, C., and D. Tasche (2002): “Expected shortfall: A natural coherent alternative to value at risk,” *Economic Notes*, 31(2), 379–388.

Aielli, G. (2009): “Dynamic conditional correlations: On properties and estimation,” *Available from SSRN: http://ssrn.com/abstract=1507743*.

Ait-Sahalia, Y. (1996): “Testing continuous-time models of the spot interest rate,” *Review of Financial Studies*, 9(2), 385–426.

Alexander, C. (2001): “Orthogonal GARCH,” in *Mastering Risk*, vol. 2, pp. 21–38. Financial Times-Prentice Hall.

Amenc, N., and F. Goltz (2007): “ETFs in core-satellite portfolio management,” *ETF and Indexing*, 2007(1), 75–88.

Anderson, D. (1992): “A multivariate Linnik distribution,” *Statistics and Probability Letters*, 14(4), 333–336.

Arrow, K. (1963): “Uncertainty and the welfare economics of medical care,” *American Economic Review*, 53(5), 941–973.

Arslan, O. (2010): “An alternative multivariate skew Laplace distribution: Properties and estimation,” *Statistical Papers*, 51(4), 865–887.

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999): “Coherent measures of risk,” *Mathematical Finance*, 9(3), 203–228.

Attanasio, O. (1991): "Risk, time-varying second moments and market efficiency," *Review of Economic Studies*, 58(3), 479–494.

Ausin, M., and H. Lopes (2010): "Time-varying joint distribution through copulas," *Computational Statistics and Data Analysis*, 54(11), 2383–2399.

Azzalini, A., and A. Capitanio (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society: Series B (Methodological)*, 65(2), 367–389.

Bachelier, L. (1964): "Theory of speculation (originally published in 1900)," in *The random character of stock market prices*, ed. by P. Cootner. MIT Press, Cambridge, Mass.

Back, A., and A. Weigend (1997): "A first application of independent component analysis to extracting structure from stock returns," *International Journal of Neural Systems*, 8(4), 473–484.

Bae, K., C. Lim, and K. Wei (2006): "Corporate governance and conditional skewness in the world's stock markets," *Journal of Business*, 79(6), 2999–3028.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman (2001): "Can investors profit from the prophets? Security analyst recommendations and stock returns," *Journal of Finance*, 56(2), 531–563.

Barndorff-Nielsen, O. (1977): "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674), 401–419.

Barndorff-Nielsen, O., and P. Bläsild (1981): "Hyperbolic distributions and ramifications: Contributions to theory and application," in *Statistical distributions in scientific work*, ed. by C. Taillie, G. Patil, and B. Baldessari, vol. 4, pp. 19–44. Reidel.

Barndorff-Nielsen, O., P. Blæsild, J. Jensen, and R. Bagnold (1985): "The fascination of sand," in *A celebration of statistics : the ISI centenary volume*, ed. by A. Atkinson, and S. Fienberg, pp. 57–87. Springer.

Barndorff-Nielsen, O., J. Kent, and M. Sörensen (1982): "Normal variance-mean mixtures and z distributions," *International Statistical Review*, 50(2), 145–159.

Barnett, W., A. Gallant, M. Hinich, J. Jungeilges, D. Kaplan, and M. Jensen (1997): "A single-blind controlled competition among tests for nonlinearity and chaos," *Journal of Econometrics*, 82(1), 157–192.

Bauwens, L., S. Laurent, and J. Rombouts (2006): “Multivariate GARCH models: A survey,” *Journal of Applied Econometrics*, 21(1), 79–109.

Bawa, V. (1978): “Safety-first, stochastic dominance, and optimal portfolio choice,” *Journal of Financial and Quantitative Analysis*, 13(2), 255–271.

Bawa, V., and E. Lindenberg (1977): “Capital market equilibrium in a mean-lower partial moment framework,” *Journal of Financial Economics*, 5(2), 189–200.

Beale, E. (1955): “On minimizing a convex function subject to linear inequalities,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 17(2), 173–184.

Berkowitz, J. (2001): “Testing density forecasts, with applications to risk management,” *Journal of Business and Economic Statistics*, 19(4), 465–474.

Bernoulli, D. (1954): “Exposition of a new theory on the measurement of risk (originally published 1738 St. Petersburg, Russia),” *Econometrica*, 22(1), 23–36.

Biglova, A., S. Ortobelli, S. Rachev, and S. Stoyanov (2004): “Different approaches to risk estimation in portfolio theory,” *Journal of Portfolio Management*, 31(1), 103–112.

Billio, M., M. Caporin, and M. Gobbo (2006): “Flexible dynamic conditional correlation multivariate GARCH models for asset allocation,” *Applied Financial Economics Letters*, 2(2), 123–130.

Black, F., and R. Litterman (1992): “Global portfolio optimization,” *Financial Analysts Journal*, 48(5), 28–43.

Bläsild, P. (1981): “The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen’s bean data,” *Biometrika*, 68(1), 251–263.

Bollerslev, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31(3), 307–327.

——— (1987): “A conditionally heteroskedastic time series model for speculative prices and rates of return,” *Review of Economics and Statistics*, 69(3), 542–547.

——— (1990): “Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model,” *Review of Economics and Statistics*, 72(3), 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988): “A capital asset pricing model with time-varying covariances,” *Journal of Political Economy*, 96(1), 116.

Bollerslev, T., and J. Wooldridge (1992): “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances,” *Econometric Reviews*, 11(2), 143–172.

Bond, S., and K. Patel (2003): “The conditional distribution of real estate returns: Are higher moments time varying?,” *Journal of Real Estate Finance and Economics*, 26(2), 319–339.

Boswijk, P., and R. van der Weide (2011): “Method of moments estimation of GO-GARCH models,” *Journal of Econometrics*, 163(1), 118–126.

Box, G., and D. Cox (1964): “An analysis of transformations,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–252.

Bradley, S., and D. Crane (1972): “A dynamic model for bond portfolio management,” *Management Science*, 19(2), 139–151.

Brannas, K., and N. Nordman (2003): “Conditional skewness modelling for stock returns,” *Applied Economics Letters*, 10(11), 725–728.

Breymann, W., A. Dias, and P. Embrechts (2003): “Dependence structures for multivariate high-frequency data in finance,” *Quantitative Finance*, 3(1), 1–14.

Bris, A., W. Goetzmann, and N. Zhu (2007): “Efficiency and the bear: Short sales and markets around the world,” *Journal of Finance*, 62(3), 1029–1079.

Brock, W., and W. Dechert (1988): “A general class of specification tests: The scalar case,” in *Business and Economics Statistics Section of the Proceedings of the American Statistical Society*, pp. 70–79.

Brock, W., W. Dechert, and J. Scheinkman (1993): “A test for independence based on the correlation dimension,” *Econometric Reviews*, 15(3), 197–235.

Broda, S., and M. Paolella (2009): “CHICAGO: A fast and accurate method for portfolio risk calculation,” *Journal of Financial Econometrics*, 7(4), 412.

Brogan, A., and S. Stidham (2005): “A note on separation in mean-lower-partial-moment portfolio optimization with fixed and moving targets,” *IIE Transactions*, 37(10), 901–906.

——— (2008): “Non-separation in the mean–lower-partial-moment portfolio optimization problem,” *European Journal of Operational Research*, 184(2), 701–710.

Brooks, C., S. Burke, S. Heravi, and G. Persand (2005): “Autoregressive conditional kurtosis,” *Journal of Financial Econometrics*, 3(3), 399–421.

Cajigas, J., and G. Urga (2006): "Dynamic conditional correlation models with asymmetric multivariate Laplace innovations," *Centre for Econometric Analysis, Cass Business School*, working paper.

Caporin, M., and M. McAleer (2012): "Do we really need both BEKK and DCC? A tale of two multivariate GARCH models," *Journal of Economic Surveys*, 26(4), 736–751.

Cappiello, L., R. Engle, and K. Sheppard (2006): "Asymmetric correlations in the dynamics of global equity and bond returns," *Journal of Financial Econometrics*, 4(4), 537–572.

Cardoso, J. (2000): "Entropic contrasts for source separation: geometry and stability," in *Unsupervised adaptive filters*, ed. by S. Haykin, pp. 139–190. John Wiley & sons.

Chapman, D., and N. Pearson (2001): "Recent advances in estimating term-structure models," *Financial Analysts Journal*, 57(4), 77–95.

Charnes, A., and W. Cooper (1959): "Chance-constrained programming," *Management Science*, 6(1), 73–79.

——— (1962): "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, 9(3-4), 181–186.

Chekhlov, A., S. Uryasev, and M. Zabarankin (2005): "Drawdown measure in portfolio optimization," *International Journal of Theoretical and Applied Finance*, 8(1), 13–58.

Chen, J., and C. Huang (2010): "An analysis of the spillover effects of exchange-traded funds," *Applied Economics*, 42(9), 1155–1168.

Chen, N., R. Roll, and S. Ross (1986): "Economic forces and the stock market," *Journal of Business*, 59(3), 383–403.

Chen, Y., W. Härdle, and V. Spokoiny (2007): "Portfolio value at risk based on independent component analysis," *Journal of Computational and Applied Mathematics*, 205(1), 594–607.

Chen, Y., W. Härdle, and V. Spokoiny (2010): "GHICA–Risk analysis with GH distributions and independent components," *Journal of Empirical Finance*, 17(2), 255–269.

Chollete, L., A. Heinen, and A. Valdesogo (2009): “Modeling international financial returns with a multivariate regime-switching copula,” *Journal of Financial Econometrics*, 7(4), 437–480.

Christoffersen, P. (1998): “Evaluating interval forecasts,” *International Economic Review*, 39(4), 841–862.

Christoffersen, P., J. Hahn, and A. Inoue (2001): “Testing and comparing value-at-risk measures,” *Journal of Empirical Finance*, 8(3), 325–342.

Christoffersen, P., and D. Pelletier (2004): “Backtesting value-at-risk: A duration-based approach,” *Journal of Financial Econometrics*, 2(1), 84–108.

Comon, P. (1994): “Independent component analysis, a new concept?,” *Signal Processing*, 36(3), 287–314.

Comte, F., and O. Lieberman (2003): “Asymptotic theory for multivariate GARCH processes,” *Journal of Multivariate Analysis*, 84(1), 61–84.

Conine, T. E., and M. J. Tamarkin (1981): “On diversification given asymmetry in returns,” *Journal of Finance*, 36(5), 1143–1155.

Cotter, J., and K. Dowd (2006): “Extreme spectral risk measures: An application to futures clearinghouse margin requirements,” *Journal of Banking and Finance*, 30(12), 3469–3485.

Dantzig, G. (1956): “Recent advances in linear programming,” *Management Science*, 2(2), 131–144.

Dantzig, G., and G. Infanger (1992): “Large-scale stochastic linear programs: importance sampling and Benders decomposition,” in *Computational and applied mathematics, I: algorithms and theory: selected and revised papers from the IMACS 13th World Congress, Dublin, Ireland, July 1991*, vol. 1991, p. 111. North Holland.

——— (1993): “Multi-stage stochastic linear programs for portfolio optimization,” *Annals of Operations Research*, 45(1), 59–76.

Dark, J. (2006): “Time varying skewness and kurtosis and a new model,” *Monash University, Australia*, working paper.

Davison, A., and R. Smith (1990): “Models for exceedances over high thresholds,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3), 393–442.

de Athayde, G., and R. Flôres Jr (2000): "Introducing higher moments in the CAPM: Some basic ideas," in *Advances in Quantitative Asset Management*, ed. by C. Dunis, pp. 3–15. Springer.

De Bondt, W., and R. Thaler (1987): "Further evidence on investor overreaction and stock market seasonality," *Journal of Finance*, 42(3), 557–581.

De Lima, P. (1996): "Nuisance parameter free properties of correlation integral based statistics," *Econometric Reviews*, 15(3), 237–259.

Demarta, S., and A. McNeil (2005): "The t copula and related copulas," *International Statistical Review*, 73(1), 111–129.

Demey, P., S. Maillard, and T. Roncalli (2010): "Risk-Based Indexation," *Available from SSRN: http://ssrn.com/abstract=1582998*.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?," *Review of Financial Studies*, 22(5), 1915–1953.

Dinkelbach, W. (1967): "On nonlinear fractional programming," *Management Science*, 13(7), 492–498.

Dowd, K., J. Cotter, and G. Sorwar (2008): "Spectral risk measures: Properties and limitations," *Journal of Financial Services Research*, 34(1), 61–75.

Durrleman, V., A. Nikeghbali, and T. Roncalli (2000): "Which copula is the right one," *Available from SSRN: http://ssrn.com/abstract=1032545*.

Eberlein, E., and U. Keller (1995): "Hyperbolic distributions in finance," *Bernoulli*, 1(3), 281–299.

Engle, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4), 987–1007.

——— (2002): "Dynamic conditional correlation," *Journal of Business and Economic Statistics*, 20(3), 339–350.

Engle, R., and K. Kroner (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11(1), 122–150.

Engle, R., and V. Ng (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48(5), 1749–1778.

ENGLE, R., V. NG, AND M. ROTHSCHILD (1990): "Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills," *Journal of Econometrics*, 45(1), 213–237.

ENGLE, R., AND J. RUSSELL (1998): "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, 66(5), 1127–1162.

ENGLE, R., AND K. SHEPPARD (2001): "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," *NBER Working Paper*, http://www.nber.org/papers/w8554.pdf.

ERGUN, A., AND J. JUN (2010): "Conditional skewness, kurtosis, and density specification testing: Moment-based versus nonparametric tests," *Studies in Nonlinear Dynamics and Econometrics*, 14(3), 5.

FAMA, E. (1970): "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, 25(2), 383–417.

——— (1991): "Efficient capital markets: II," *Journal of Finance*, 46(5), 1575–1617.

FANG, H., AND L. TSONG-YUE (1997): "Co-kurtosis and capital asset pricing," *Financial Review*, 32(2), 293–307.

FARINELLI, S., AND L. TIBILETTI (2008): "Sharpe thinking in asset ranking with one-sided measures," *European Journal of Operational Research*, 185(3), 1542–1547.

FEINSTEIN, C. D., AND M. N. THAPA (1993): "A reformulation of a mean-absolute deviation portfolio optimization model," *Management Science*, 39(12), 1552–1553.

FERNANDEZ, C., J. OSIEWALSKI, AND M. STEEL (1995): "Modeling and inference with v-spherical distributions," *Journal of the American Statistical Association*, 90(432), 1331–1340.

FERNANDEZ, C., AND M. STEEL (1998): "On Bayesian modeling of fat tails and skewness," *Journal of the American Statistical Association*, 93(441), 359–371.

FISHBURN, P. (1977): "Mean-risk analysis with risk associated with below-target returns," *American Economic Review*, 67(2), 116–126.

FRAZZINI, A. (2006): "The disposition effect and underreaction to news," *Journal of Finance*, 61(4), 2017–2046.

FRENCH, K. (2008): "Presidential address: The cost of active investing," *Journal of Finance*, 63(4), 1537–1573.

FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): “Expected stock returns and volatility,” *Journal of Financial Economics*, 19(1), 3–29.

FREY, R., AND A. MCNEIL (2000): “Modelling Dependent Defaults,” *Available from ETHZ: ftp://magellan-03.math.ethz.ch/hg/users/mcneil/defaults.pdf.*

G30 (1993): *Derivatives: Practices and principles*, vol. 5. Group of Thirty.

GARLAPPI, L., R. UPPAL, AND T. WANG (2007): “Portfolio selection with parameter and model uncertainty: A multi-prior approach,” *Review of Financial Studies*, 20(1), 41–81.

GENEST, C., K. GHOUDI, AND L. RIVEST (1995): “A semiparametric estimation procedure of dependence parameters in multivariate families of distributions,” *Biometrika*, 82(3), 543–552.

GENTON, M. (2004): *Skew-elliptical distributions and their applications: A journey beyond normality.* Chapman & Hall/CRC.

GHALANOS, A. (2012a): *parma: Portfolio Allocation and Risk Management Applications.*Available from R-Forge: https://r-forge.r-project.org/projects/parma/.

——— (2012b): *rmgarch: Multivariate GARCH models.*Available from R-Forge: http://r-forge.r-project.org/projects/rgarch/.

——— (2012c): *rugarch: Univariate GARCH models.*Available from CRAN: http://cran.r-project.org/web/packages/rugarch/index.html.

GHALANOS, A., AND S. THEUSSL (2012): *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method.*Available from CRAN: http://cran.r-project.org/web/packages/Rsolnp/index.html.

GIBBONS, M., S. ROSS, AND J. SHANKEN (1989): “A test of the efficiency of a given portfolio,” *Econometrica*, 57(5), 1121–1152.

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *Journal of Finance*, 48(5), 1779–1801.

GONZALEZ-RIVERA, G., T. LEE, AND S. MISHRA (2004): “Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood,” *International Journal of Forecasting*, 20(4), 629–645.

GOURIEROUX, C. (1997): *ARCH models and financial applications.* Springer.

Grinblatt, M., and S. Titman (1992): “The persistence of mutual fund performance,” *Journal of Finance*, 47(5), 1977–1984.

Grinold, R. (1992): “Are benchmark portfolios efficient?,” *Journal of Portfolio Management*, 19(1), 34–40.

Grinold, R., and R. Kahn (2000): *Active portfolio management: A quantitative approach for providing superior returns and controlling risk*. McGraw-Hill.

Hafner, C., and A. Preminger (2009): “Asymptotic theory for a factor GARCH model,” *Econometric Theory*, 25(2), 336–363.

Hanoch, G., and H. Levy (1969): “The efficiency analysis of choices involving risk,” *Review of Economic Studies*, 36(3), 335–346.

Hansen, B. (1994): “Autoregressive conditional density estimation,” *International Economic Review*, 35(3), 705–730.

Hansen, L. (1982): “Large sample properties of generalized method of moments estimators,” *Econometrica*, 50(4), 1029–1054.

Hansen, P., A. Lunde, and J. Nason (2011): “The model confidence set,” *Econometrica*, 79(2), 453–497.

Harlow, W., and R. Rao (1989): “Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence,” *Journal of Financial and Quantitative Analysis*, 24(3), 285–311.

Harvey, C., and A. Siddique (1999): “Autoregressive conditional skewness,” *Journal of Financial and Quantitative Analysis*, 34(4), 465–487.

Harvey, C., and A. Siddique (2000): “Conditional skewness in asset pricing tests,” *Journal of Finance*, 55(3), 1263–1295.

Hentschel, L. (1995): “All in the family nesting symmetric and asymmetric GARCH models,” *Journal of Financial Economics*, 39(1), 71–104.

Herzog, F., G. Dondi, S. Keel, L. Schumani, and H. Geering (2007): “Solving ALM problems via sequential stochastic programming,” *Quantitative Finance*, 7(2), 231–244.

Holthausen, D. (1981): “A risk-return model with risk and return measured as deviations from a target return,” *American Economic Review*, 71(1), 182–188.

Hong, H., and J. Stein (2003): "Differences of opinion, short-sales constraints, and market crashes," *Review of Financial Studies*, 16(2), 487–525.

Hong, Y., and H. Li (2005): "Nonparametric specification testing for continuous-time models with applications to term structure of interest rates," *Review of Financial Studies*, 18(1), 37–84.

Hsieh, D. (1991): "Chaos and nonlinear dynamics: application to financial markets," *Journal of Finance*, 46(5), 1839–1877.

Hu, X., R. Shonkwiler, and M. Spruill (1994): "Random restarts in global optimization," *Available from Georgia Institute of Technology: http://smartech.gatech.edu/xmlui/handle/1853/31310*.

Huang, H. (2010): "Optimal multiperiod asset allocation: Matching assets to liabilities in a discrete model," *Journal of Risk and Insurance*, 77(2), 451–472.

Hyvärinen, A., and E. Oja (2000): "Independent component analysis: Algorithms and applications," *Neural Networks*, 13(4-5), 411–430.

Jacobs, B. I., K. N. Levy, and H. M. Markowitz (2006): "Trimability and fast optimization of long-short portfolios," *Financial Analysts Journal*, 62(2), 36–46.

James, W., and C. Stein (1956): "Estimation with quadratic loss," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 1–379. University of California Press.

Jarque, C., and A. Bera (1987): "A test for normality of observations and regression residuals," *International Statistical Review*, 55(2), 163–172.

Jeantheau, T. (1998): "Strong consistency of estimators for multivariate ARCH models," *Econometric Theory*, 14(1), 70–86.

Jegadeesh, N., and S. Titman (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 48(1), 65–91.

Jensen, M. (1968): "The performance of mutual funds in the period 1945-1964," *Journal of Finance*, 23(2), 389–416.

Jensen, M., and A. Lunde (2001): "The NIG-S model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model," *Econometrics Journal*, 4(2), 319–342.

Joe, H. (1997): *Multivariate models and dependence concepts*, vol. 73. Chapman & Hall/CRC.

Joe, H., H. Li, and A. Nikoloulopoulos (2010): “Tail dependence functions and vine copulas,” *Journal of Multivariate Analysis*, 101(1), 252–270.

Johnson, N. (1949): “Systems of frequency curves generated by methods of translation,” *Biometrika*, 36(1–2), 149–176.

Jondeau, E., and M. Rockinger (2003): “Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements,” *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.

——— (2006a): “The copula-garch model of conditional dependencies: An international stock market application,” *Journal of International Money and Finance*, 25(5), 827–853.

——— (2006b): “Optimal portfolio allocation under higher moments,” *European Financial Management*, 12(1), 29–55.

——— (2009): “The impact of shocks on higher moments,” *Journal of Financial Econometrics*, 7(2), 77.

Jones, M., and R. Sibson (1987): “What is projection pursuit?,” *Journal of the Royal Statistical Society: Series A (General)*, 150(1), 1–37.

Junker, M., and A. May (2005): “Measurement of aggregate risk with copulas,” *Econometrics Journal*, 8(3), 428–454.

Jurczenko, E., and B. Maillet (2006): “Theoretical Foundations of Asset Allocation and Pricing Models with Higher-order Moments,” in *Multi-moment asset allocation and pricing models*, pp. 1–36. John Wiley & sons.

Kahneman, D., and A. Tversky (1979): “Prospect theory: An analysis of decision under risk,” *Econometrica: Journal of the Econometric Society*, 47(2), 263–291.

Kalashnikov, V. (1997): *Geometric sums: bounds for rare events with applications: risk analysis, reliability, queueing*, vol. 413. Springer.

Keating, C., and W. Shadwick (2002): “A universal performance measure,” *Journal of performance measurement*, 6(3), 59–84.

Kelker, D. (1970): “Distribution theory of spherical distributions and a location-scale parameter generalization,” *Sankhya: The Indian Journal of Statistics*, 32(4), 419–430.

Khindanova, I., S. Rachev, and E. Schwartz (2001): “Stable modeling of value at risk,” *Mathematical and Computer Modelling*, 34(9), 1223–1259.

Kollo, T., and M. Srivastava (2005): "Estimation and testing of parameters in multivariate Laplace distribution," *Communications in Statistics-Theory and Methods*, 33(10), 2363–2387.

Konno, H., H. Shirakawa, and H. Yamazaki (1993): "A mean-absolute deviation-skewness portfolio optimization model," *Annals of Operations Research*, 45(1), 205–220.

Konno, H., K. Tanaka, and R. Yamamoto (2011): "Construction of a portfolio with shorter downside tail and longer upside tail," *Computational Optimization and Applications*, 48(2), 199–212.

Konno, H., and H. Yamazaki (1991): "Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market," *Management Science*, 37(5), 519–531.

Kotz, S., T. Kozubowski, and K. Podgorski (2001): *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*, no. 183. Birkhauser.

Kotz, S., T. Kozubowski, and K. Podgórski (2002): "Maximum likelihood estimation of asymmetric Laplace parameters," *Annals of the Institute of Statistical Mathematics*, 54(4), 816–826.

Kouwenberg, R. (2001): "Scenario generation and stochastic programming models for asset liability management," *European Journal of Operational Research*, 134(2), 279–292.

Kouwenberg, R., and S. Zenios (2006): "Stochastic programming models for asset liability management," *Handbook of asset and liability management*, 1, 253–303.

Kozubowski, T., and K. Podgórski (2001): "Asymmetric Laplace laws and modeling financial data," *Mathematical and Computer Modelling*, 34(9), 1003–1021.

Kraus, A., and R. Litzenberger (1976): "Skewness preference and the valuation of risk assets," *Journal of Finance*, 31(4), 1085 – 1100.

Kring, S., S. Rachev, M. Höchstötter, F. Fabozzi, and S. Bianchi, M.L. (2009): "Multi-tail generalized elliptical distributions for asset returns," *Econometrics Journal*, 12(2), 272–291.

Kritzman, M., S. Page, and D. Turkington (2010): "In defense of optimization: The fallacy of 1/N," *Financial Analysts Journal*, 66(2), 31.

Kroner, K., and V. Ng (1998): "Modeling asymmetric comovements of asset returns," *Review of Financial Studies*, 11(4), 817.

Kruskal, W. (1958): "Ordinal measures of association," *Journal of the American Statistical Association*, 53(284), 814–861.

Kupiec, P. (1995): "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, 3(2), 73–84.

Lai, T. (1991): "Portfolio selection with skewness: A multiple-objective approach," *Review of Quantitative Finance and Accounting*, 1(3), 293–305.

Lambert, P., and S. Laurent (2001a): "Modeling skewness dynamics in series of financial data using skewed location-scale distributions," *Institut de Statistique, Louvain-la-Neuve Discussion Paper*, 119.

——— (2001b): "Modelling financial time series using GARCH-type models and a skewed Student density," *Universite de Liege*.

Learned-Miller, E. G., and J. W. Fisher III (2003): "ICA using spacings estimates of entropy," *Journal of Machine Learning Research*, 4, 1271–1295.

Ledoit, O., and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 10(5), 603–621.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15(5), 850–859.

León, Á., G. Rubio, and G. Serna (2005): "Autoregresive conditional volatility, skewness and kurtosis," *Quarterly Review of Economics and Finance*, 45(4), 599–618.

Li, D. (2000): "On default correlation: A copula function approach," *Journal of Fixed Income*, 9(4), 43–54.

Lichtenstein, S., and P. Slovic (1973): "Response-induced reversals of preference in gambling: An extended replication in Las Vegas.," *Journal of Experimental Psychology*, 101(1), 16–20.

Lindsey, J., and P. Lindsey (2006): "Multivariate distributions with correlation matrices for nonlinear repeated measurements," *Computational Statistics and Data Analysis*, 50(3), 720–732.

LINDSKOG, F., A. MCNEIL, AND U. SCHMOCK (2003): "Kendall's tau for elliptical distributions," in *Credit Risk: Measurement, Evaluation and Management*, pp. 149–156. Physica-Verlag.

LJUNG, G., AND G. BOX (1978): "On a measure of lack of fit in time series models," *Biometrika*, 65(2), 297–303.

LOISTL, O. (1976): "The erroneous approximation of expected utility by means of a Taylor's series expansion: Analytic and computational results," *American Economic Review*, 66(5), 904–910.

MADANSKY, A. (1962): "Methods of solution of linear programs under uncertainty," *Operations Research*, 10(4), 463–471.

MALEVERGNE, Y., AND D. SORNETTE (2003): "Testing the Gaussian copula hypothesis for financial assets dependences," *Quantitative Finance*, 3(4), 231–250.

MALKIEL, B. (2003): "Passive investment strategies and efficient markets," *European Financial Management*, 9(1), 1–10.

MANDELBROT, B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36(4), 394–419.

MARDIA, K. (1970): "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, 57(3), 519–530.

——— (1980): "Tests of unvariate and multivariate normality," *Handbook of Statistics*, 1, 279–320.

MARKOWITZ, H. (1952): "Portfolio selection," *Journal of Finance*, 7(1), 77–91.

MARSHALL, A., AND I. OLKIN (1993): "Maximum likelihood characterizations of distributions," *Statistica Sinica*, 3, 157–171.

MARTELLINI, L. (2008): "Toward the design of better equity benchmarks," *Journal of Portfolio Management*, 34(4), 34–41.

MCNEIL, A., AND R. FREY (2000): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, 7(3), 271–300.

MCNEIL, A., R. FREY, AND P. EMBRECHTS (2005): *Quantitative risk management. Concepts, techniques, and tools.* Princeton University Press.

Michaud, R. (1989): "The Markowitz optimization enigma: Is optimized optimal?," *Financial Analysts Journal*, 45(1), 31–42.

Miffre, J. (2007): "Country-specific ETFs: An efficient approach to global asset allocation," *Journal of Asset Management*, 8(2), 112–122.

Morgan, J. (1994): "Introduction to Riskmetrics," Discussion paper, J.P. Morgan.

Moskowitz, T., and M. Grinblatt (1999): "Do industries explain momentum?," *Journal of Finance*, 54(4), 1249–1290.

Mulvey, J., and B. Shetty (2004): "Financial planning via multi-stage stochastic optimization," *Computers and Operations Research*, 31(1), 1–20.

Nawrocki, D., and K. Staples (1989): "A customized LPM risk measure for portfolio analysis," *Applied Economics*, 21(2), 205–218.

Nyblom, J. (1989): "Testing for the constancy of parameters over time," *Journal of the American Statistical Association*, 84(405), 223–230.

Pascual, L., J. Romo, and E. Ruiz (2004): "Bootstrap predictive inference for ARIMA processes," *Journal of Time Series Analysis*, 25(4), 449–465.

——— (2006): "Bootstrap prediction for returns and volatilities in GARCH models," *Computational Statistics and Data Analysis*, 50(9), 2293–2312.

Pastor, L., and R. Stambaugh (2000): "Comparing asset pricing models: An investment perspective," *Journal of Financial Economics*, 56(3), 335–381.

Patton, A. (2006): "Modelling asymmetric exchange rate dependence," *International Economic Review*, 47(2), 527–556.

Pearlmutter, B. A., and L. C. Parra (1997): "Maximum likelihood blind source separation a context-sensitive generalization of ICA," *Advances in Neural Information Processing Systems*, 9, 613–619.

Pelagatti, M., and S. Rondena (2006): "Dynamic conditional correlation with elliptical distributions," *mimeo*.

Pelletier, D. (2006): "Regime switching for dynamic correlations," *Journal of Econometrics*, 131(1-2), 445–473.

Polikar, R. (2006): "Ensemble based systems in decision making," *Circuits and Systems Magazine, IEEE*, 6(3), 21–45.

Prakash, A., C. Chang, and T. Pactwa (2003): “Selecting a portfolio with skewness: Recent evidence from US, European, and Latin American equity markets,” *Journal of Banking and Finance*, 27(7), 1375–1390.

Prause, K. (1999): “The generalized hyperbolic model: Estimation, financial derivatives, and risk measures,” Ph.D. thesis, University of Freiburg.

Premaratne, G., and A. Bera (2000): “Modeling asymmetry and excess kurtosis in stock return data,” *Illinois Research & Reference Working Paper No. 00-123.*

Quiggin, J. (1982): “A theory of anticipated utility,” *Journal of Economic Behavior and Organization*, 3(4), 323–343.

Quirk, J., and R. Saposnik (1962): “Admissibility and measurable utility functions,” *Review of Economic Studies*, 29(2), 140–146.

Rachev, S., and S. Mittnik (2000): *Stable Paretian models in finance.* John Wiley & sons.

Raftery, A., D. Madigan, and J. Hoeting (1997): “Bayesian model averaging for linear regression models,” *Journal of the American Statistical Association*, 92(437), 179–191.

Rigby, R., and D. Stasinopoulos (2005): “Generalized additive models for location, scale and shape,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554.

Rockafellar, R., and S. Uryasev (2000): “Optimization of conditional value-at-risk,” *Journal of Risk*, 2, 21–42.

Rockafellar, R., S. Uryasev, and M. Zabarankin (2006): “Generalized deviations in risk analysis,” *Finance and Stochastics*, 10(1), 51–74.

Rockinger, M., and E. Jondeau (2002): “Entropy densities with an application to autoregressive conditional skewness and kurtosis,” *Journal of Econometrics*, 106(1), 119–142.

Rom, B., and K. Ferguson (1994): “Post-modern portfolio theory comes of age,” *Journal of Investing*, 3(3), 11–17.

Rosenblatt, M. (1952): “Remarks on a multivariate transformation,” *Annals of Mathematical Statistics*, 23(3), 470–472.

Ross, S. (1976): “The arbitrage theory of capital asset pricing,” *Journal of Economic Theory*, 13(3), 341–360.

Roy, A. (1952): "Safety first and the holding of assets," *Econometrica: Journal of the Econometric Society*, 20, 431–449.

Savage, L. (1962): *Subjective probability and statistical practice*. Methuen.

Schaible, S. (1976a): "Fractional programming. I, Duality," *Management Science*, 22(8), 858–867.

——— (1976b): "Fractional programming. II, On Dinkelbach's algorithm," *Management Science*, 22(8), 868–873.

Scherer, B. (2007): "Can robust portfolio optimisation help to build better portfolios?," *Journal of Asset Management*, 7(6), 374–387.

Schmeidler, D. (1989): "Subjective probability and expected utility without additivity," *Econometrica*, 57(3), 571–87.

Schmidt, R., T. Hrycej, and E. Stützle (2006): "Multivariate distribution models with generalized hyperbolic margins," *Computational Statistics and Data Analysis*, 50(8), 2065–2096.

Scott, D., D. Würtz, C. Dong, and T. Tran (2011): "Moments of the generalized hyperbolic distribution," *Computational Statistics*, 26(3), 459–476.

Scott, R., and P. Horvath (1980): "On the direction of preference for moments of higher order than the variance," *Journal of Finance*, 35(4), 915–919.

Sears, R., and K. Wei (1985): "Asset pricing, higher moments, and the market risk premium: A note," *Journal of Finance*, 40, 1251–1253.

Silvennoinen, A., and T. Teräsvirta (2009a): "Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model," *Journal of Financial Econometrics*, 7(4), 373–411.

——— (2009b): "Multivariate GARCH models," in *Handbook of Financial Time Series*, pp. 201–229. Springer.

Simon, H. (1955): "A behavioral model of rational choice," *Quarterly Journal of Economics*, 69(1), 99–118.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8(1), 11.

Sortino, F., and L. Price (1994): "Performance measurement in a downside risk framework," *Journal of Investing*, 3(3), 59–64.

SORTINO, F., AND S. SATCHELL (2001): *Managing downside risk in financial markets: Theory, practice and implementation*, vol. 1. Butterworth-Heinemann.

SPERANZA, M. (1993): "Linear programming models for portfolio optimization," *Finance*, 14(1), 107–123.

STONE, B. (1973): "A general class of three-parameter risk measures," *Journal of Finance*, 28(3), 675–685.

STOYANOV, S., S. RACHEV, AND F. FABOZZI (2007): "Optimal financial portfolios," *Applied Mathematical Finance*, 14(5), 401–436.

TAY, A., F. DIEBOLD, AND T. GUNTHER (1998): "Evaluating density forecasts, with applications to financial risk management," *International Economic Review*, 39(4), 863–883.

TAYLOR, S. (1986): *Modelling Financial Time Series*. John Wiley & sons.

TONG, H., AND K. LIM (1980): "Threshold autoregression, limit cycles and cyclical data-with discussion," *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3), 245–292.

TSE, Y., AND A. TSUI (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business and Economic Statistics*, 20(3), 351–362.

TVERSKY, A., AND D. KAHNEMAN (1974): "Judgment under uncertainty: Heuristics and biases," *Science*, 185(4157), 1124–1131.

VAN DER WEIDE, R. (2002): "GO-GARCH: a multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17(5), 549–564.

——— (2004): "Wake me up before you GO-GARCH," in *Computing in Economics and Finance*. Society for Computational Economics.

VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of games and economic behavior*. Princeton University Press.

WEST, K., AND W. NEWEY (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55(3), 703–708.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society*, 50(1), 1–25.

WILHELMSSON, A. (2009): "Value at Risk with time varying variance, skewness and kurtosis–the NIG-ACD model," *Econometrics Journal*, 12(1), 82–104.

XU, L. (1999): "An Empirical Method to Select Dominant Independent Components in ICA for Time Series Analysis," in *Proceedings of the 1999 International Joint Conference on Neural Networks (IJCNN'99)*, vol. 6, pp. 3883–3887.

YE, Y. (1997): *Interior point algorithms: theory and analysis*, vol. 44. Wiley-Interscience.

YOUNG, M. (1998): "A minimax portfolio selection rule with linear programming solution," *Management Science*, 44(5), 673–683.

ZADEH, L. (1965): "Fuzzy sets," *Information and control*, 8(3), 338–353.

ZAKOIAN, J. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18(5), 931–955.

ZEEVI, A., AND R. MASHAL (2002): "Beyond correlation: Extreme co-movements between financial assets," *Available from SSRN: http://ssrn.com/abstract=317122*.

ZHANG, K., AND L. CHAN (2009): "Efficient factor GARCH models and factor-DCC models," *Quantitative Finance*, 9(1), 71–91.