On the Importance of Skewness and Asymmetric Dependence in Stock Returns for Asset Allocation

Andrew J. Patton*
University of California, San Diego

PRELIMINARY VERSION

11 November, 2001

Keywords: stock returns, normality, skewness, copulas, asymmetry, predictability, GARCH.
J.E.L. Codes: G11, C32, C51.

*This work is preliminary, and comments or questions are welcome. The author would like to thank Rob Engle, Clive Granger, Bruce Lehmann and Raffaella Giacomini for helpful comments. Special thanks are due to Allan Timmermann for many useful discussions and suggestions on this topic. All remaining deficiencies are the responsibility of the author. Many thanks go to Vince Crawford and the UCSD Economics Experimental and Computational Laboratory for providing the computational resources required for this project. Financial support from the UCSD Project in Econometric Analysis fellowship is gratefully acknowledged. Contact information: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0508, U.S.A. Email: apatton@econ.ucsd.edu.

On the Importance of Skewness and Asymmetric Dependence in Stock Returns for Asset Allocation

Abstract

Two types of asymmetries have been found in the joint distribution of stock returns: skewness in individual stock returns, and the observation that stock returns are more dependent during market downturns than during market upturns. This latter characteristic is referred to as 'asymmetric dependence'. Evidence of skewness has been widely reported, and evidence that stock returns are more dependent during market downturns than during market upturns has been reported by several authors in recent years, see Erb, *et al.*, (1994), Ang and Chen (2001) and Longin and Solnik (2001), for example. We examine in this paper the connection between these two types of asymmetries, and their implications for asset allocation decisions. We provide a link between univariate skewness and asymmetric dependence between assets: asymmetric dependence between assets can lead to skewed portfolios, even if the individual assets are not themselves skewed. Thus a preference for positive skew translates to an aversion to assets that exhibit greater dependence in bear markets than in bull markets. We consider the problem of a CRRA investor allocating wealth between a small-cap and a large-cap portfolio, using monthly data from January 1954 to December 1999. We use models that can capture the empirically observed time-varying means and variances of stock returns, and also the presence of (possibly time-varying) skewness and kurtosis. Further, we employ models of the time-varying dependence structure that allow for, but do not impose, greater dependence during bear markets than bull markets. Our models are developed using copula theory, which enables the construction of flexible multivariate distributions. The importance of skewness and asymmetric dependence for asset allocation is assessed by comparing the risk-adjusted performance of a portfolio based on a bivariate normal distribution model with a portfolio based on a model developed using copula theory. We find that a portfolio based on a copula distribution model significantly outperforms a portfolio based on the bivariate normal distribution for various performance measures and levels of relative risk aversion.

Keywords: stock returns, normality, skewness, copulas, asymmetry, predictability, GARCH.
J.E.L. Codes: G11, C32, C51.

1 Introduction

Two characteristics of stock returns that violate traditional mean-variance analysis are that they exhibit skewness, and that they appear to be more dependent during market downturns than during market upturns. This latter characteristic will be referred to as 'asymmetric dependence'. The presence of either of these asymmetries violates the assumption of elliptically distributed asset returns, which is required for mean-variance analysis, see Ingersoll (1987).

Evidence of skewness in stock returns has been widely reported[1], and is by now generally regarded as a common feature of stock returns. Evidence that stock returns exhibit some form of asymmetric dependence has been reported by several authors in recent years; published work includes Erb, *et al.*, (1994), Longin and Solnik (2001), Ang and Chen (2001) and Ang and Bekaert (2001), who all report that correlations between stock returns are greater during bear markets than during bull markets. Further evidence is reported in numerous unpublished studies[2].

In this paper we show how skewness and asymmetric dependence are related: asymmetric dependence between asset returns can lead to skewed portfolio returns, under some conditions. We show that the distribution of portfolios of individually symmetric assets will be asymmetric if the dependence structure is asymmetric. (The impatient reader may sneak a look at Figure 4, which shows the distribution of an equally weighted portfolio of normally distributed assets that have an asymmetric dependence structure.)

Arrow (1971) suggests that a desirable property of a utility function is that it exhibits non-increasing absolute risk aversion. Under non-increasing absolute risk aversion investors can be shown to have a preference for positively skewed assets, in the same way that positive marginal utility leads to a preference for assets with higher mean returns, and diminishing marginal utility leads to risk aversion. Harvey and Siddique (2000) report evidence that skewness is indeed priced by the market. Utility functions that exhibit non-increasing absolute risk aversion include the constant absolute risk aversion (CARA), or exponential, utility function, and the constant relative risk aversion (CRRA), or 'narrow power', utility function, see Huang and Litzenberger (1988). In this paper we work under the assumption that investors have the CRRA utility function, as the CARA utility function implies that the demand for risky assets is constant for all wealth levels, so the proportion of wealth invested in risky assets decreases as wealth increases; an implication that jars with observed behaviour.

Much of the existing work on asset allocation focussed on special cases where the combination of utility function and distribution model were such than an analytical solution for the optimal portfolio decision exists, see Kandel and Stambaugh (1996) or Campbell and Viceira (1999) amongst

[1] See Kraus and Litzenberger (1976), Friend and Westerfield (1980), Singleton and Wingender (1986), Lim (1989), Richardson and Smith (1993), Harvey and Siddique (1999, 2000), Aït-Sahalia and Brandt (2001) and Chen (2001), amongst others.

[2] See Bae, *et al.*, (2000), Rosenberg (2000) and Campbell, *et al.*, (2001) amongst others.

many others, such as the combination of quadratic or exponential utility with elliptical distributions, or where the utility function was assumed to be a function of a certain number of moments of the distribution of returns. The focus on such analytically tractable special cases was motivated, at least in part, by computational constraints and certainly not by the fact that the utility functions or distributional assumptions were considered realistic.

Recent work by Brandt (1999) and Aït-Sahalia and Brandt (2001) overcome the problem of the appropriate distributional assumption to combine with a given utility function by using the method of moments and the first-order conditions of the investor's optimisation problem to obtain an optimal portfolio decision. Doing so allows them to use whichever utility function they please; like us, they use the CRRA utility function. Theirs is indeed an interesting approach, however it has the drawback that its nonparametric nature imposes restrictions on the number of exogenous regressors that may be included in the model, as in Brandt (1999), or on the way a larger number of regressors may enter into the problem, as in Aït-Sahalia and Brandt (2001). Our framework instead involves a flexible parametric approach to distribution modelling.

For a case study of the importance of skewness and asymmetric dependence for asset allocation we choose to look at the problem of allocating wealth between the CRSP small-cap and large-cap indices[3], using monthly data from January 1954 to December 1999. This problem is representative of that of choosing between a high risk - high return asset and a lower risk - lower return asset. We use distribution models that can capture the empirically observed time-varying means and variances of stock returns, and also the presence of (possibly time-varying) skewness and kurtosis. Further, we employ models of the dependence structure that allow for, but do not impose, greater dependence during bear markets than bull markets, and allow for changes in this dependence structure through time. The empirical section of this paper can be seen as a first step in addressing the suggestions of Harvey and Siddique (2000) and Longin and Solnik (2001), who propose investigating the impact of conditional skewness (Harvey and Siddique) and asymmetric dependence (Longin and Solnik) on portfolio choices.

Our models are developed using copula theory, which enables the construction of very flexible multivariate distributions. In Section 2 we provide a non-technical introduction to copula theory; a more thorough introduction is presented in Appendix 1, and the interested reader is referred to Nelsen (1999), or Schweizer and Sklar (1983) and Joe (1997), for a complete introduction.

We measure the importance of skewness and asymmetric dependence for asset allocation by comparing the risk-adjusted performance of a portfolio based a bivariate normal distribution model with a portfolio based on a model developed using copula theory. The significance of the differences in measures of portfolio performance are tested using bootstrap methods, and we find substantial evidence in most cases that skewness and asymmetric dependence do indeed have important impli-

[3]The small-cap index is comprised of the smallest 10% of U.S. stocks, by market capitalisation and the large-cap index is comprised of the largest 10% of U.S. stocks.

cations for asset allocation. For example[4], while a constant equally weighted portfolio of the two assets generates a Sharpe ratio of 14.2% and the portfolio based on the bivariate normal model generates a Sharpe ratio of 23.0%, the portfolio developed using copula theory attains a Sharpe ratio of 27.8%. Thus the gains to modelling the distribution of returns are increased by over 50% (according to this measure) by capturing and modelling deviations from joint normality.

One of the costs of using more flexible models for the joint distribution of stock returns is that we are forced by computational constraints to be relatively unsophisticated in other aspects of the project. Firstly, we use the entire data set to estimate the models' parameters and to select the models' functional forms. An improvement on this would be to recursively select and estimate models using only data available at each point in time, and then to develop an out-of-sample forecast distribution of returns and compute the corresponding optimal portfolio weights. Secondly, we ignore the effects of parameter estimation uncertainty on the investor's decision problem, though this has been found to be important, see Kandel and Stambaugh (1996) and the references cited therein. Finally, we only consider the investor's problem for the one-period-ahead investment horizon. For one of the utility functions we consider, the log utility function, this approach is correct, however for the remaining utility functions the optimal weights will have both a 'myopic' component and a 'hedging' component, see Merton (1971). The myopic component is the solution we focus on: the investor simply seeks to maximise the next-period expected utility. The hedging component represents the deviation from the myopic optimal weight that occurs when the investor seeks to hedge possible future adverse movements in the investment opportunity set. Ang and Bekaert (2001) and Aït-Sahalia and Brandt (2001) find, however, only weak evidence of hedging demand, though Brandt (1999) reports it to be quite significant. Relaxing these assumptions of the current project design is left for future work[5]. The goal of this project is to determine whether skewness and asymmetric dependence have important implications for asset allocation in our simplified setting.

The remainder of the paper is structured as follows. Immediately below we introduce the notation used in the paper, in Section 2 we provide a non-technical introduction to copula theory and its use in modelling stock returns, and in Section 3 we present some theoretical results on skewness and asymmetric dependence and their impact on portfolios of assets. Some numerical examples illustrating the theory are presented in Section 3.3. In Section 4 we present empirical results on the asset allocation problem for a portfolio of a small-cap index and a large-cap index: Section 4.1 presents the investor's problem in detail, Section 4.3 presents the models employed; Section 4.4

[4]The figures here are taken from Table 11, for the investor with relative risk aversion of seven.

[5]These three simplifications are presented in increasing order of difficulty to relax: estimating recursively the models should not be too daunting a task; dealing with parameter estimation uncertainty (or even 'model uncertainty') will be a tougher computational challenge; computing optimal portfolio weights for forecast horizons greater than one will be a formidable challenge when using flexible distribution models, as such models generally do not lead to closed-form expressions for multi-step ahead density forecasts.

presents the estimation results and some statistical goodness-of-fit tests. The performance of the resulting portfolios are evaluated in 4.5, and the benefits to modelling skewness and asymmetric dependence are compared in Section 4.6. Finally, we conclude in Section 5. In Appendix 1 we present a brief introduction to copula theory, in Appendix 2 we present some useful results on Hansen's skewed t distribution and in Appendix 3 we provide the functional forms of the copulas considered in Section 4. All proofs are contained in Appendix 4.

1.1 Notation

We have two (scalar) random variables of interest, X and Y, and some conditioning variables $\mathbf{W}$. The variables' joint conditional distribution is: $(X_t, Y_t) | \mathcal{F}_{t-1} \sim H_t \equiv C_t (F_t, G_t)$, where H_t is some bivariate distribution function, the marginal distributions of X_t and Y_t are F_t and G_t, and the copula is C_t. (The notation '$H \equiv C(F, G)$' will become clear in the next section.) We will assume that all distributions are continuous, though this assumption may be relaxed at the expense of further complication. The information set is defined as $\mathcal{F}_t \equiv \sigma (X_t, Y_t, \mathbf{W}_{t+1}, X_{t-1}, Y_{t-1}, \mathbf{W}_t, ...)$. As usual, we will denote random variables in upper case, X_t, and realisations of random variables in lower case, x_t. We will often need to refer to the history of the random variables, which will be denoted $Z^t \equiv \left(X_t, Y_t, \mathbf{W}'_{t+1}, X_{t-1}, Y_{t-1}, \mathbf{W}'_t, ...\right)'$. Throughout this paper we will denote the distribution (or *c.d.f.*) of a random variable using an upper case letter, and the corresponding density (or *p.d.f.*) using the lower case letter. We will denote the extended real line as $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{\pm\infty\}$.

2 Flexible multivariate distribution models using copulas

In this paper we use copula theory to develop flexible parametric models of the joint distribution of returns. We allow for time variation in the conditional mean, variance, skewness and kurtosis of the individual returns, and for time-variation in the conditional dependence between the two assets. We then compute the investor's optimum portfolio based on the resulting distribution model. Below we provide a non-technical introduction to copula theory; a more rigorous introduction is presented in Appendix 1.

A copula is a function that links together two (or more) marginal distributions to form a joint distribution. The marginal distributions that it couples can be of any type: a normal and an exponential, or a Student's t and a Uniform, for example. The theory of copulas dates back to Sklar (1959), but it wasn't until Clayton (1978) that copulas were used in the modelling of data. Since then numerous applications have appeared in the applied statistics literature, see Cook and Johnson (1981), Oakes (1989), Genest and Rivest (1993) and Fine and Jiang (2000), amongst others, and more recently in the analysis of economic data, see Rosenberg (1999) and (2000), Li (2000), Scaillet (2000), Embrechts, *et al.,* (2001) and Patton (2001a,b,c). The main theorem in copula theory is that of Sklar (1959), and below we present a modification of it for conditional distributions.

Theorem 1 (Sklar's Theorem for Continuous Conditional Distributions) *Let H_t be a conditional bivariate distribution function with continuous margins F_t and G_t, and let $\mathcal{F}_{t-1}$ be some conditioning set. Then there exists a unique conditional copula $C_t : [0,1] \times [0,1] \to [0,1]$ such that*

$$H_t(x, y|\mathcal{F}_{t-1}) = C_t(F_t(x|\mathcal{F}_{t-1}), G_t(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}), \quad \forall (x, y) \in \bar{\mathbb{R}}^2 \tag{1}$$

Conversely, if C_t is a conditional copula and F_t and G_t are the conditional distribution functions of two random variables X_t and Y_t, then the function H_t defined by equation (1) is a bivariate conditional distribution function with margins F_t and G_t.

Sklar's theorem allows us to decompose a bivariate distribution, H_t, into three components: the two marginal distributions, F_t and G_t, and the copula, C_t. Since all of the univariate information on X and Y is contained in the marginal distributions, what remains is all of the dependence information *between* X and Y, which is captured in the copula.

We may model the individual variables using whatever marginal distribution fits each one best, and then work on modelling the dependence structure via a model for the copula. This is precisely how we approach the problem of modelling the joint distribution of the two assets under analysis in this paper.

It should be pointed out that copula theory extends very naturally to higher dimensions. We focus on the bivariate case in this paper for simplicity.

2.1 Copulas and the existence of moments of portfolios

A possible concern regarding employing flexible models for the dependence structure is that moments of linear combinations of the random variables may not exist, even though they do exist for the individual variables. In this section we show that this is not the case: a sufficient condition for the existence of the k^{th} moment of a linear combination of two random variables is the existence of k moments for each random variable. Thus, for example, the existence of the third moments of X and Y is sufficient for the skewness of $Z \equiv \omega X + (1-\omega) Y$ to exist.

Definition 1 (Central moment) *The k^{th} central moment of X is defined as*

$$M_k[X] \equiv E\left[(X - \mu_x)^k\right], \text{ where } \mu_x \equiv E[X].$$

Definition 2 (Central co-moment) *A k^{th}-order central co-moment of (X, Y) is defined to be*

$$M_{i,k-i}[X, Y] \equiv E\left[(X - \mu_x)^i (Y - \mu_y)^{k-i}\right], \text{ for } 0 \le i \le k.$$

Proposition 1 (Expression for portfolio moments) *The k^{th} central moment of a portfolio $Z \equiv \omega X + (1-\omega) Y$ for $\omega \in \mathbb{R}$ is given by*

$$E\left[(Z-\mu_z)^k\right] = \sum\nolimits_{i=0}^{k} \omega^i (1-\omega)^{k-i} \binom{k}{i} E\left[(X-\mu_x)^i \left(Y-\mu_y\right)^{k-i}\right] \tag{2}$$

where $\binom{k}{i} \equiv \frac{k!}{(k-i)!i!}$ are the binomial coefficients and $\mu_z \equiv \omega\mu_x + (1-\omega)\mu_y$.

Lemma 1 (Existence of co-moments) *If $E\left[|X-\mu_x|^k\right] < \infty$ and $E\left[\left|Y-\mu_y\right|^k\right] < \infty$, then $E\left[|X-\mu_x|^i \cdot \left|Y-\mu_y\right|^{k-i}\right] < \infty$ for $0 \le i \le k$.*

Proposition 2 (Existence of portfolio moments) *Let $Z \equiv \omega X + (1-\omega) Y$, $\omega \in \mathbb{R}$. If $E\left[|X-\mu_x|^k\right] < \infty$ and $E\left[\left|Y-\mu_y\right|^k\right] < \infty$, then $E\left[(Z-\mu_z)^k\right] < \infty$.*

3 Asymmetric dependence, skewness and portfolios

In this section we look at the relationship between the copula and the skewness of a portfolio of assets. It turns out that the copula plays a significant role in determining the skewness of a portfolio. In particular, below we show that the tendency for assets to have greater dependence during downward movements of the market than during upward movements can lead to a negatively skewed portfolio, even if the original assets themselves are not skewed. We will firstly present a few preliminary definitions and results. In Section 3.1 we present the results pertaining to symmetric distributions, in Section 3.2 we move on to the more interesting case of asymmetric distributions, and in Section 3.3 we present a few numerical examples illustrating the results.

3.1 Symmetry

We present below the formal definition of univariate symmetry, and then the most common means of measuring symmetry: skewness.

Definition 3 (Univariate symmetry) *The random variable $X \sim F$ is said to be 'symmetric about μ_x' if $X-\mu_x$ has the same distribution as $\mu_x - X$.*

Definition 4 (Skewness) *The 'skewness' of a random variable X is defined as:*

$$Skew[X] \equiv \sigma_x^{-3} M_3[X] \tag{3}$$

where $\sigma_x^2 \equiv M_2[X]$, is the variance of X.

Definition 5 (Co-skewness) *The 'co-skewness' measures of two random variables are defined as*

$$CoSkew_{12}[X,Y] \equiv \sigma_x^{-1}\sigma_y^{-2} M_{12}[X,Y], \text{ and} \tag{4}$$
$$CoSkew_{21}[X,Y] \equiv \sigma_x^{-2}\sigma_y^{-1} M_{21}[X,Y] \tag{5}$$

The skewness measures are simply third central moments standardised by variance[6]. Note that the definition of 'co-skewness' is slightly different to some previous definitions, as in Friend and Westerfield (1980) and Harvey and Siddique (2000). One first result to note is that if $Skew[X] < 0$, then $Skew[-X] > 0$, and so by short-selling a negatively skewed asset we create a positively skewed asset.

Proposition 3 (Skewness of a portfolio) *Let $(X,Y) \sim H = C(F,G)$. Then the skewness of a random variable $Z \equiv \omega X + (1-\omega) Y$, $\omega \in \mathbb{R}$ is equal to:*

$$\begin{aligned} Skew[Z] &= \sigma_z^{-3} \Big(\omega^3 M_3[X] + (1-\omega)^3 M_3[Y] \\ &\quad + 3\omega (1-w)^2 M_{12}[X,Y] + 3\omega^2 (1-w) M_{21}[X,Y] \Big) \end{aligned} \tag{6}$$

The literature on three moment capital asset pricing models, starting with Kraus and Litzenberger (1976), generally looked only at the marginal contribution of an asset to the skewness of a diversified portfolio. The focus in this paper is instead on a portfolio of just two assets, which will only rarely represent a fully diversified portfolio.

The concepts of skewness and symmetry are related by the fact that a symmetric distribution has zero skew. Thus zero skew is a necessary, though not sufficient condition for symmetry.

There are numerous generalisations of the concept of skewness to the multivariate case, see Mardia (1970), Nelsen (1993) and Avérous and Meste (1997). We will use a definition due to Nelsen.

Definition 6 (Bivariate radial symmetry) *$(X,Y) \sim H = C(F,G)$ are said to be 'radially symmetric about (μ_x, μ_y)' if the joint distribution of $(X - \mu_x, Y - \mu_y)$ is the same as that of $(\mu_x - X, \mu_y - Y)$.*

A further result of interest to us is that the joint distribution of two symmetrically distributed random variables is radially symmetric if and only if their dependence structure is symmetric. This is stated more formally below.

Theorem 2 (Nelsen, 1993) *Let $(X,Y) \sim H = C(F,G)$, and let X and Y be symmetric about μ_x and μ_y respectively. Then (X,Y) are radially symmetric about (μ_x, μ_y) if and only if their copula, C, is radially symmetric.*

We now present a lemma that has some use in the results that follow. The lemma shows that if X and Y have a radially symmetric joint distribution, then the two conditional expectation functions, $E[X|Y=y]$ and $E[Y|X=x]$, also satisfy a symmetry concept. These regression functions will be useful in some results below.

[6] Some authors refer to the third central moment $M_3[X]$ as the skewness of the random variable X. We will instead use the standardised third central moment given in Definition 4.

Lemma 2 (Symmetry of conditional expectation functions) *Let* $(X,Y) \sim H = C(F,G)$ *be radially symmetric about* (μ_x, μ_y). *Then the regression functions,* $E[X|Y=y]$ *and* $E[Y|X=x]$, *satisfy:*

$$\begin{aligned} E\left[X|Y = y + \mu_y\right] - \mu_x &= \mu_x - E\left[X|Y = \mu_y - y\right] \text{ for all } y \in \mathbb{R}, \text{ and} \\ E\left[Y|X = x + \mu_x\right] - \mu_y &= \mu_y - E\left[Y|X = \mu_x - x\right] \text{ for all } x \in \mathbb{R}. \end{aligned}$$

If $\mu_x = \mu_y = 0$, *then the regression functions satisfy:*

$$\begin{aligned} E[X|Y=y] &= -E[X|Y=-y], \text{ and} \\ E[Y|X=x] &= -E[Y|X=-x]. \end{aligned}$$

In the following proposition we present our first result on the implications of bivariate symmetry on the distribution of linear combinations of random variables: that if the joint distribution of two assets is radially symmetric, then any portfolio of these two assets will have a symmetric distribution.

Proposition 4 *Let* $(X,Y) \sim H = C(F,G)$ *be radially symmetric about* (μ_x, μ_y), *for* $(\mu_x, \mu_y) \in \mathbb{R}^2$. *Then* $Z \equiv \omega X + (1-\omega) Y$, $\omega \in \mathbb{R}$, *is symmetric about* $\mu_z \equiv \omega\mu_x + (1-\omega)\mu_y$.

3.2 Asymmetry

In this section we relate the property of skewness to that of asymmetric dependence. Concretely defining what is meant by 'asymmetric dependence' is somewhat difficult: most measures of dependence, such as linear correlation, rank correlation, Kendall's τ, etcetera, summarise the dependence between the variables over their entire support. Some authors, such as Erb, *et al.*, (1994), Ang and Chen (2001) and Longin and Solnik (2001) look at measures originally intended for the entire support of the distribution, such as linear correlation, on just a subset of the support, such as the negative quadrant or the positive quadrant. Ang and Chen (2001) and Longin and Solnik (2001) look at a measure the latter authors call 'exceedence correlation'. The cut-offs for the exceedence correlation may be defined an various ways; we use the quantiles of the variables:

$$\bar{\rho}(q) \equiv \begin{cases} Corr\left[X, Y \mid X \leq Q_x(q) \cap Y \leq Q_y(q)\right], & \text{for } q \leq 0.5 \\ Corr\left[X, Y \mid X \geq Q_x(q) \cap Y \geq Q_y(q)\right], & \text{for } q \geq 0.5 \end{cases}$$

where $Q_x(q)$ and $Q_y(q)$ are the q^{th} quantiles of X and Y respectively.

As Ang and Chen (2001) and Longin and Solnik (2001) point out, how the exceedence correlations behave depends on what underlying distribution is assumed for the data. Even for a standard bivariate normal distribution the exceedence correlation plot is non-linear in the cut-offs. In Figure 1 we plot the empirical exceedence correlations based on the (raw) excess returns on the two assets that are analysed in Section 4, along with the exceedence correlations that would be expected if the data had the bivariate normal distribution. In Figure 2 we plot the empirical exceedence correlations between the transformed standardised residuals of the models for the two excess return series, along with what would be expected it they had the normal copula, and the 'rotated Gumbel' copula, which is described in the next section. These two plots clearly indicate the presence of asymmetric dependence between these assets.

[INSERT FIGURES 1 AND 2 HERE]

We will use an alternative measure of asymmetric dependence that relates directly to skewness. From the definitions of skewness, and the expression given in equation (6), we know that the skewness of a portfolio of two assets may be broken down into three components: the skewness of each of the individual assets and the co-skewness terms. While this might resemble the decomposition of the joint distribution into the marginal distributions and the copula in equation (1), it is not quite as neat as that. The individual skewness coefficients relate only to the marginal distributions, but the co-skewness coefficient cannot be solely defined in terms of the copula. The presence of skewed margins may lead to a non-zero co-skewness coefficient even in the case that the copula is radially symmetric. However, if the marginal distributions are symmetric then the presence of skewness in the portfolio is driven *completely* by asymmetry in the dependence structure. We have the simple proposition:

Proposition 5 *Let* $(X, Y) \sim H = C(F, G)$*, and let* X *and* Y *be either negatively (positively) skewed or symmetric. If the* H *exhibits negative (positive) co-skewness, then* $Z \equiv \omega X + (1 - \omega) Y$ *for* $\omega \in (0, 1)$ *will be negatively (positively) skewed.*

For a portfolio weight, ω, outside the interval $[0, 1]$ the sign of the skewness of the portfolio will depend on the relative magnitudes of the skewness and co-skewness coefficients, and on the particular value for ω.

3.2.1 Co-skewness and asymmetric dependence

The analysis below will make use of the regression functions, $E[X|Y = y]$ and $E[Y|X = x]$, introduced in Section 3.1. We will compare the distance between the conditional mean of $Y|X = \mu_x + x$, for example, with the unconditional mean of Y. This distance provides some measure of the amount of information on Y contained in X when $X = \mu_x + x$. The further the conditional mean is from the unconditional mean, the more informative is the conditioning set, and the greater the dependence between the variables. By comparing the distance between the conditional mean and the unconditional mean for $X = \mu_x + x$ and $X = \mu_x - x$ we may capture asymmetries in the dependence structure: if X and Y are more dependent during bear markets, we would expect $\mu_y - E[Y|X = \mu_x - x] \geq E[Y|X = \mu_x + x] - \mu_y$ for $x \geq 0$.

Proposition 6 *Let* $(X,Y) \sim H = C(F,G)$, *and let* X *and* Y *be symmetric about* μ_x *and* μ_y *respectively. If* $\mu_x - E[X|Y = \mu_y - y] \geq E[X|Y = \mu_y + y] - \mu_x$ *for all* $y \geq 0$, *and* $\mu_y - E[Y|X = \mu_x - x] \geq E[Y|X = \mu_x + x] - \mu_y$ *for all* $x \geq 0$, *with at least one of the weak inequalities holding strictly for some* x *or* y, *then* $Z \equiv \omega X + (1-\omega) Y$ *for* $\omega \in (0,1)$ *will be negatively skewed.*

The above proposition shows that a portfolio (with weight in between 0 and 1) of assets that are individually symmetric, but that exhibit greater dependence in bear markets, will exhibit negative skewness. Of course, if the original variables are also negatively skewed then this will serve to increase the severity of the negative skewness of the convex combination. If one or both of the original variables exhibits positive skewness, then the sign of the skewness of Z may be negative or positive.

It should be pointed out that the conditions in the above proposition are merely sufficient conditions for the skewness of Z to be negative. There are many alternative situations where these conditions do not hold but the skewness of Z is negative; all that is required is that the co-skewness terms are both negative. In some examples that we consider below the condition in Proposition 6 is not met, but we do indeed find negative co-skewness coefficients.

What if we consider linear, non-convex combinations of the variables? That is, what if the investor is able to sell short one asset and invest an amount greater than 100% in the other asset? Below we show that such an investment would lead to positive skewness rather than negative skewness.

Proposition 7 *Let* $(X,Y) \sim H = C(F,G)$, *and let* X *and* Y *be symmetric about* μ_x *and* μ_y *respectively. For simplicity, let* $M_{12}[X,Y] = M_{21}[X,Y]$. *If* $CoSkew_{12}[X,Y] < 0$, *then* $Z \equiv \omega X + (1-\omega) Y$ *for* $\omega < 0$ *or* $\omega > 1$ *will be positively skewed.*

Cases where $M_{12}[X,Y] \neq M_{21}[X,Y]$ can also lead to positively skewed portfolios, though the magnitudes of each and the particular weight chosen would all play a factor. Nevertheless, the point is made that by changing the portfolio weights we may take a negatively skewed portfolio and make it positively skewed.

3.3 A few examples

In this section we illustrate the theoretical results presented in the previous sections with a few numerical examples. We will consider a simple case, where we have two assets, X and Y, which both have the standard normal distribution, $N(0,1)$. Thus these are individually symmetric assets. We will examine four joint distributions, all of which are such that the linear correlation between X and Y is 0.5. These joint distributions are constructed via four different copulas: the normal copula, Clayton's copula, Plackett's copula and the Gumbel copula. Figure 3 below presents the contour plots of the Clayton, Plackett and Gumbel copula distributions (the normal copula distribution's contour plot is the familiar elliptical case) on the left, and the regression functions $E[Y|X=x]$ on the right.

This plot shows the correspondence between asymmetric dependence and the regression functions. The contours of the Clayton copula distribution are more peaked in the lower quadrant, indicating greater dependence, and the conditional mean function is further from the unconditional mean, drawn with a dashed line, for negative values of the conditioning variable than for positive values.

The Plackett copula is radially symmetric, and since the individual assets are symmetric, the resulting joint distribution is also radially symmetric, by Theorem 2. We showed in Lemma 2 that radial symmetry implies a symmetric regression function, and this is confirmed in the second panel of the right column of Figure 3.

The lower left and right panels of Figure 3 show the contour plot and conditional mean function for the Gumbel copula distribution. We can see that it resembles the Clayton distribution, though rotated so that the increased dependence is in the positive quadrant rather than the negative quadrant.

[INSERT FIGURE 3 HERE]

The first illustration of the impact of the copula on the distribution of portfolio returns is given in Figure 4. In this figure we plot the probability density of a portfolio of X and Y, where each asset is individually a standard Normal$(0,1)$ random variable, and where their copula is Clayton's copula, again calibrated to yield a linear correlation coefficient of 0.5. The portfolio we consider here is the equally weighted portfolio, $Z = \frac{1}{2}X + \frac{1}{2}Y$.

[INSERT FIGURE 4 HERE]

The above figure clearly shows the asymmetry of the distribution of portfolio returns. Even though the individual assets have the familiar $N(0,1)$ distribution, a simple average of the two does not. The negatively skewed distribution of the portfolio returns is driven by the fact that Clayton's copula implies increased dependence during bear markets; the co-skewness coefficient for this distribution is -0.30.

We plot in Figure 5 the skewness coefficient of portfolio returns as a function of the weight in the first asset, for the Clayton and Gumbel distributions. (The normal and Plackett copula distributions are radially symmetric, implying that any portfolio will have zero skew.) In terms of portfolio variance, it is easily shown that the optimal portfolio in this case is $\omega = 0.5$ for all four copulas, however this is not necessarily so in terms of portfolio skewness. Figure 5 shows that an even mix of both assets yields the best portfolio skewness *only if* the assets exhibit positive (and equal) co-skewness, as in the Gumbel copula distribution. If the assets exhibit negative co-skewness, as in the Clayton copula distribution, then the portfolio with weight $\omega = 0.5$ is actually the *worst* in terms of portfolio skewness; this portfolio has the largest negative skew of all possible portfolio weights. In this case, the preferences of the investor must be used to trade off increased variance for decreased negative skew.

[INSERT FIGURE 5 HERE]

Finally, we plot below the skewness coefficient of a portfolio of the two assets used in the following section. In Figure 6 we compute the skewness based on a portfolio of the (raw) excess returns for various weights. This figure, however, will be influenced by both the presence of skewness in the marginal distributions and asymmetric dependence. In Figure 7 we 'normalise' the excess returns[7] and plot the skewness coefficients of a portfolio of the normalised assets. The shape of this figure is determined by the shape of the dependence structure. The negative skewness of portfolios with weights in between zero and one confirms the presence of increased dependence between these assets during bear markets.

[INSERT FIGURES 6 AND 7 HERE]

The results presented in this section make it very clear exactly how investors are made worse off by increased dependence between assets during market downturns: such a dependence structure leads to negatively skewed portfolios. All reasonable utility functions have the property that investors prefer positive skew, thus negative skew is an economic 'bad'.

4 A portfolio of small cap and large cap stocks

In this section we consider an investor with constant relative risk aversion facing the problem of allocating wealth between two assets: a portfolio of low market capitalisation stocks ('small caps') and a portfolio of high market capitalisation stocks ('large caps'). These two assets were chosen as being representative of the general problem of balancing a portfolio comprised of a high risk - high return asset and a lower risk - lower return asset. The small cap and large cap portfolios fit this problem: the average annualised return on these indices was 9.95% and 7.97% respectively, and their annualised standard deviations were 21.29% and 14.29%.

The rest of this section is structured as follows: in Section 4.1 we describe in detail the investor's decision problem, in Section 4.2 we provide summary statistics of the data used. In Sections 4.3 and 4.4 we describe the models considered for the joint distribution of the returns on the two indices and in Section 4.5 we analyse the performance of the different models in terms of the risk-adjusted returns on the portfolios generated by the models, both via performance statistics and statistical

[7] Specifically, we use the empirical *cdf* of the excess returns to transform the data to being Uniformly distributed on $[0, 1]$, and then we use the inverse $Normal\,(0,1)$ *cdf* to transform these to having the $N\,(0,1)$ distribution, which has zero skew.

tests for superior performance. Finally in Section 4.6 we compare the benefits to modelling skewness and asymmetric dependence.

4.1 The investor's optimisation problem

The utility functions we assume for our hypothetical investors are from the class of constant relative risk aversion (CRRA) utility functions:

$$\mathcal{U}(\omega) = \begin{cases} (1-\gamma)^{-1} \cdot (W_0 \cdot \exp\{\omega X + (1-\omega) Y\})^{1-\gamma} & \text{if } \gamma \neq 1 \\ \log(W_0 \cdot \exp\{\omega X + (1-\omega) Y\}) & \text{if } \gamma = 1 \end{cases} \tag{7}$$

where W_0 is the initial wealth, X represents the return on the small-cap index and Y represents the return on the large-cap index. The degree of relative risk aversion (RRA) is denoted by γ. For this utility function the initial wealth does not affect the choice of optimal weight and so we will set $W_0 = 1$. We consider five different levels of relative risk aversion: $\gamma = 1, 3, 7, 10$ and 20. This range of risk aversion levels was also considered in Aït-Sahalia and Brandt (2001).

The set-up of the investor's problem is as follows. Let the two assets under consideration be denoted X_t and Y_t. These assets have some joint distribution, H_t, with associated marginal distributions, F_t and G_t, and a copula, C_t. That is, $(X_t, Y_t) | \mathcal{F}_{t-1} \sim H_t \equiv C_t(F_t, G_t)$. We will develop estimates of this joint distribution by modelling the conditional marginal distributions, $\hat{F}_t$ and $\hat{G}_t$, and the conditional copula, $\hat{C}_t$, and use them to compute the optimal weights, ω^*_{t+1}, for the portfolio. The optimal weights are found by maximising the expected utility of the end-of-period wealth under the estimated probability density:

$$\begin{aligned} \omega^*_{t+1} &\equiv \arg\max_{\omega} \hat{E}_t[\mathcal{U}(\exp\{\omega X_{t+1} + (1-\omega) Y_{t+1}\})] \\ &\equiv \arg\max_{\omega} \iint \mathcal{U}(\exp\{\omega x + (1-\omega) y\}) \cdot \hat{h}_{t+1}(x, y) \cdot dx \cdot dy \\ &= \arg\max_{\omega} \iint \mathcal{U}(\exp\{\omega x + (1-\omega) y\}) \cdot \hat{f}_{t+1}(x) \cdot \hat{g}_{t+1}(y) \cdot \hat{c}_{t+1}\left(\hat{F}_{t+1}(x), \hat{G}_{t+1}(y)\right) \cdot dx \cdot dy \end{aligned}$$

We consider in this paper only the problem of maximising one-step-ahead expected utility. As mentioned in the introduction, for the utility functions with $RRA \neq 1$ the optimal weights will have both a 'myopic' component and a 'hedging' component, see Merton (1971). We are forced by computational constraints to ignore the possible hedging demands of the investors.

The double-integral defining the expected utility of wealth does not have a closed-form solution for our case. We use adaptive quadrature to numerically approximate this integral, see Judd (1998) for details[8]. The objective function $\varphi_{t+1}(\omega) \equiv \iint \mathcal{U}(\exp\{\omega x + (1-\omega) y\}) \cdot \hat{h}_{t+1}(x, y) \cdot dx \cdot dy$ was found to be very well-behaved (smooth and having a unique global optimum) for our choices of

[8]We used adaptive Lobatto quadrature for the numerical integrals, which is the function *quadl* in Matlab. Adaptive quadrature methods are useful when the integrals are poorly behaved, or when we know little about the

utility functions and density models and so we employed the BFGS algorithm to locate the optimum, ω^*_{t+1}, at each point in time.

4.2 Description of the data

We use data from the Center for Research in Security Prices (CRSP) on the top 10% and bottom 10% of stocks sorted by market capitalisation to form indices - the 'big cap' and 'small cap' indices. Our data is at the monthly frequency, from January 1954 to December 1999, yielding 552 observations. This data was also analysed in a different context by Perez-Quiros and Timmermann (2001). Descriptive statistics on the two portfolios are presented in Table 1.

[INSERT TABLE 1 HERE]

Table 1 reveals that the small cap index had a higher mean and higher volatility than the large cap index. The small cap index also exhibited slightly positive skewness, while the large cap index exhibited substantial negative skewness. Both indices exhibit excess kurtosis. The Jarque-Bera statistic indicates that neither series is unconditionally normal (suggesting that the assumption conditional normality may be somewhat dubious), and the unconditional correlation coefficient indicates a high degree of linear dependence.

We use three further variables as explanatory variables in our analysis. The first is the one-month treasury bill rate, denoted R_{ft}, which is taken as the risk-free rate. This variable has been used by Fama (1981) and others as a proxy for shocks to expected growth in the real economy. The second variable is the difference between the yield on corporate bonds with Moody's rating Baa versus those with an Aaa rating, denoted SPR_t, which is called the 'default spread'. This variable tracks the cyclical variation in the risk premium on stocks, see Perez-Quiros and Timmermann (2001). Finally, we look at the dividend yield, denoted DIV_t, which is measured as the total dividends paid over the previous 12 months divided by the stock price at the end of the month. This variable acts as a proxy for time-varying expected returns. For a comprehensive review of the variables that have been used in previous studies as predictive variables for stock returns see Aït-Sahalia and Brandt (2001, pp1297-1298).

4.3 Analysis of the different models

We consider a number of different models. The first three are naïve portfolios: 100% weight in the small-cap index, 100% weight in the large-cap index, and an even mix of both indices.

integral (Judd, 1998, p269). The main benefit of adaptive quadrature methods, as opposed to other methods like Gauss-Legendre quadrature, is that the approximation error is reduced by re-computing the integral until the result 'converges'. The main cost of this method is in computational time.

The fourth portfolio is one based solely on the unconditional distribution of returns. For this portfolio we assume that the investor optimises once, and only once, his/her portfolio weights and then maintains them for the entire sample. Rather than propose a functional form for the unconditional distribution of returns we use the empirical joint distribution, H_n, to compute the expected utility:

$$\begin{aligned}\omega^*_{uncond} &\equiv \arg\max_{\omega} \ \hat{E}\left[\mathcal{U}\left(\exp\left\{\omega X + (1-\omega)Y\right\}\right)\right] \\ &= \arg\max_{\omega} \ \iint \mathcal{U}\left(\exp\left\{\omega x + (1-\omega)y\right\}\right)\cdot dH_n(x,y) \\ &= \arg\max_{\omega} \ n^{-1}\sum\nolimits_{t=1}^{n}\mathcal{U}\left(\exp\left\{\omega x_t + (1-\omega)y_t\right\}\right) \qquad (8)\end{aligned}$$

where $\{(x_t, y_t)\}_{t=1}^{n}$ are the observed excess returns on the two assets. This portfolio is 'somewhat naïve', in that the investor does perform some optimisation, but assumes that the joint distribution of these two assets is *i.i.d.* throughout the sample. A comparison of the performance of this portfolio with those constructed using parametric conditional distribution models may then be interpreted as a measure of the benefits to modelling the conditional distribution of these stock returns.

The benchmark parametric model for our study is the bivariate normal distribution, which is compared with another parametric model constructed using copula theory. Both parametric models have the same forms for the conditional means, μ_t^x and μ_t^y, and variances, h_t^x and h_t^y: the conditional mean of the small cap returns was set as a linear function of a constant, the lagged risk-free rate, the lagged default spread, and the first and twelfth lags of the small cap returns. Lags of the dividend yield and the large cap returns were found to be not significant, and so were not included in the model. The conditional mean model for the large cap returns was set as a linear function of a constant, the lagged risk-free rate, the lagged default spread, and the first lag of the large cap returns. Again, lags of the dividend yield and the small cap returns were found to be not significant. For each conditional variance model we employed a TARCH(1,1) specification[9], with the lagged risk-free rate as a regressor. Other variables were tried as volatility regressors, though none were found to be significant.

For the bivariate normal model, all that remains to be specified is a model for the correlation. The conditional correlation was set as a function of the lagged risk-free rate, default spread, dividend yield, and the forecasts of the conditional means of the two variables. The conditional mean forecasts were included in the model to capture any changes in the dependence structure when the variables were both predicted to fall or rise. The bivariate normal model is:

[9]The general TARCH(1,1) specification is: $h_t = \omega + \beta h_{t-1} + \alpha_+ \cdot \varepsilon_{t-1}^2 \cdot \mathbf{1}\{\varepsilon_{t-1} > 0\} + \alpha_- \cdot \varepsilon_{t-1}^2 \cdot \mathbf{1}\{\varepsilon_{t-1} < 0\} + \alpha_z \cdot z_{t-1}$, where ε_t is the residual from the model for the mean, and z_t is an exogenous regressor.

Bivariate normal specification

$$\left(\frac{X_t-\mu_t^x}{\sqrt{h_t^x}},\frac{Y_t-\mu_t^y}{\sqrt{h_t^y}}\right) \sim N\left(\mathbf{0},\left[\begin{array}{cc}1 & \rho_t\\ \rho_t & 1\end{array}\right]\right) \quad (9)$$

$$\rho_t = \Lambda\left(\alpha_0+\alpha_1 R_{ft-1}+\alpha_2 SPR_{t-1}+\alpha_3 DIV_{t-1} +\alpha_4\mu_t^x+\alpha_5\mu_t^y\right) \quad (10)$$

where $\Lambda(x)=\frac{1-e^{-x}}{1+e^{-x}}$ is the modified logistic transformation, designed to keep ρ_t in $(-1,1)$ at all times.

The models constructed using copula theory are called 'copula distribution specifications'. We model separately the marginal densities of each of the variables, and then model their copula. Since both variables, however, are stock returns, it is not unlikely that they will be best fitted by similar models. This is indeed what we find: Hansen's (1994) skewed Student's t distribution was found to provide a good fit for both assets. Some results on this distribution are presented in Appendix 2. In addition to time-varying conditional means and variances, the skewed t can capture time-varying conditional skewness and kurtosis. For these two assets we found that time variation in conditional skewness was significant, as did Harvey and Siddique (1999), but that the conditional kurtosis of both of these variables was constant. We modelled the conditional skewness parameter, λ, as a function of the lagged risk-free rate, default spread and dividend yield, and used the logistic transformation to ensure that the parameter remained within $(-1,1)$ at all times, as suggested by Hansen (1994). The copula distribution model is:

Copula distribution specification

$$\left(\frac{X_t-\mu_t^x}{\sqrt{h_t^x}},\frac{Y_t-\mu_t^y}{\sqrt{h_t^y}}\right) \sim C\left(Skewed\ t\left(\lambda_t^x,\bar{\nu}^x\right), Skewed\ t\left(\lambda_t^y,\bar{\nu}^y\right);\delta_t\right) \quad (11)$$

$$\delta_t = \Gamma\left(\beta_0+\beta_1 R_{ft-1}+\beta_2 SPR_{t-1}+\beta_3 DIV_{t-1} +\beta_4\mu_t^x+\beta_5\mu_t^y\right) \quad (12)$$

where $\Gamma(x)$ is a function designed to keep δ_t in the feasible region for the copula C at all times.

One concern that may arise in this design is the existence of $\hat{E}_t\left[\mathcal{U}\left(\exp\left\{\omega X_{t+1}+(1-\omega)Y_{t+1}\right\}\right)\right]$ for certain density models. Given CRRA utility, any density model that assigns positive probability to the case of bankruptcy would preclude the existence of $\hat{E}_t[\mathcal{U}]$. Both of the above specifications do assign some (extremely small) positive probability to bankruptcy, and so left unmodified $\hat{E}_t[\mathcal{U}]$ will not exist. We get around this by 'squashing' the tails of the distribution: we apply a logistic transformation to the lower tail of the portfolio return distribution so that all probability mass assigned to the region $(-\infty,\varepsilon)$ is re-located to the region $(0,\varepsilon)$, where ε is some extremely small number. We do an equivalent transformation for the upper tail. In this way the density is still continuous and $\hat{E}_t[\mathcal{U}]$ exists.

4.4 Estimation results and goodness-of-fit testing

The parametric models were estimating using the two-stage maximum likelihood estimator of copula models for time series presented in Patton (2001b). The two sets of results for the marginal distributions, those obtained assuming normality and those obtained assuming the skewed t distribution, are presented in Tables 2a and 2b below.

[INSERT TABLES 2a AND 2b HERE]

The above results illustrate some interesting features of the data. The lagged risk-free rate has a negative impact on expected returns, while the lagged default bond spread has a positive impact. The results for the model of conditional variance clearly indicate the asymmetric impacts of positive and negative innovations. The coefficient on positive lagged innovations squared was not significant for the large-cap margin, and so we removed it from the model. The estimated degrees-of-freedom parameters indicate substantial excess kurtosis. The time paths of the conditional skewness parameters are presented in Figure 8. This figure shows that there is some persistence in the conditional skewness of these two assets, contrary to the results of Singleton and Wingender (1985). These authors, however, look at persistence by comparing skewness coefficients in adjacent 5 and 10 year blocks of data, which may be longer period than the persistence we find. Also note that the conditional skewness of both of these assets changes sign numerous times during the sample period.

[INSERT FIGURE 8 HERE]

Since the skewed t distribution nests the normal distribution we can perform a likelihood ratio test that the additional 5 parameters in the skewed t distribution significantly improve the fit. The test statistic[10] (p-value) is 41.2034 (0.0000) for the small-caps and 25.4028 (0.0000), indicating that indeed the skewed t distribution does provide a statistically better fit than the normal distribution for both the small-caps and the large-caps. We can also perform other tests on the skewed t: we test that the skewed t distribution provides a better fit than a standard t distribution by testing that skewness is constant and zero. This yields a p-value for the small-caps (large-caps) of 0.0339 (0.0117), indicating that skewness is significantly different from zero for both margins. We test for the significance of the time variation in the skewness parameter, and find a p-value for the small

[10] All of the test statistics have the χ^2_p distribution under the null hypothesis, where p is the number of restrictions being tested.

caps (large caps) of 0.0156 (0.1335), rejecting the null that the skewness parameter is constant for the small-caps but failing to reject for the large-caps[11].

We now turn to tests of goodness-of-fit of the marginal distribution models. The evaluation of the goodness-of-fit of the models for the marginal distributions is of critical importance: the joint distribution of the transformed variables, $U_t \equiv F_t(X_t;\hat{\varphi}_{n_x})$ and $V_t \equiv G_t(Y_t;\hat{\gamma}_{n_y})$, will be modelled with a copula, which has margins that are Uniform$(0,1)$ by construction. If the marginal distribution models are misspecified then the variables U_t and V_t will not be uniform and the copula will be misspecified. In light of this, we employ a number of tests of the marginal specifications. The first two follow Diebold, *et al.* (1998), who suggested testing that $U_t \sim i.i.d.\ Unif\,(0,1)$ and $V_t \sim i.i.d.\ Unif\,(0,1)$ in two stages: firstly testing that U_t and V_t are *i.i.d.* via LM tests, and then testing that they are Uniform(0,1). We test the *i.i.d.* assumption by regressing $(U_t - \bar{U}_n)^k$ and $(V_t - \bar{V}_n)^k$ on twelve lags of both variables for $k = 1,2,3,4$, where $\bar{U}_n$ and $\bar{V}_n$ are the sample averages of U_t and V_t. We test the $Unif\,(0,1)$ hypothesis via the well-known Kolmogorov-Smirnov test. The results of these tests are presented in Table 3 below. As this table shows, all four marginal distribution models pass both tests.

[INSERT TABLE 3 HERE]

We employ two further tests, suggested in Patton (2001a). These tests jointly test the hypotheses of *i.i.d.* and uniformity via 'hit' tests. The support of the distribution is divided into five regions, R_j, according to quantiles, with boundaries at 0, 0.1, 0.25, 0.75, 0.9 and 1. The regions correspond to the extreme upper and lower tails, the intermediate upper and lower tails, and the centre of the distribution. The hit random variable is defined as taking the value 1 if the transformed variable (U_t or V_t) lies in the region and zero else. That is: $Hit^X_{jt} \equiv \mathbf{1}\{U_t \in R_j\}$ and $Hit^Y_{jt} \equiv \mathbf{1}\{V_t \in R_j\}$, for $j = 1,2,..,5$. Under the null hypothesis that $\hat{U}_t \sim i.i.d.\ Unif\,(0,1)$ we have that $Hit^X_{jt} \sim i.i.d.$ $Bernoulli\left(\Pr_{\hat{F}_t}\left[Hit^X_{jt} = 1\right]\right)$. We may then test this hypothesis for each of the five regions. Testing each region separately enables us to see if deficiencies exist in the model's fit in particular regions (such as in the tails if regions 1 and 5 are misspecified, or in uncaptured skewness, if both upper or both lower regions are misspecified). We may also test the joint hypothesis that all five regions are well specified via a multinomial test, also described in Patton (2001a). We perform these tests, including as regressors the hit indicator for the previous month and the total number of hits in the previous year to test for serial dependence, and a constant to test for misspecification of the conditional density. The results are presented in Table 4 below.

[INSERT TABLE 4 HERE]

[11] Although the LR test for the significance of time-variation in the skewness parameter fails to reject the null of constancy, the specification tests employed suggested that the time variation was important.

The hit tests indicate that the normal distribution models are misspecified: both margins fail the joint test at the 5% alpha level. The small-cap normal model fails in regions 2 and 3, which correspond to the intermediate lower tail and the centre of the distribution. The large-cap normal model fails in regions 2, 3 and 5. The skewed t distribution models appear better specified: both margins pass the joint test, and only one rejection is found, in the central region for the large-cap model. We take this as evidence that the skewed t marginal models are adequately specified, and proceed to modelling the copula.

The copula of the bivariate normal distribution is by definition the normal copula. Thus the transformed residuals from the normal marginal distributions are modelled with a normal copula. A total of nine different copulas were estimated on the transformed residuals from the skewed t models, in the search for the best fitting copula. The copulas considered were the normal, Student's t, Clayton, rotated Clayton, Joe-Clayton, Plackett, Frank, Gumbel, rotated Gumbel and copulas; the functional forms of these copulas are contained in Appendix 3. This list includes almost all of the copulas considered in the various applications of copulas in statistics and economics[12]. We estimated these copulas with conditional dependence modelled as a function of the lagged risk-free rate, default spread and dividend yield, and the forecast conditional mean for both series, as in equation (12). Unfortunately, of the $\binom{9}{2} = 36$ possible pair-wise comparisons, only one involves copulas that are nested (the Student's t copula nests the normal copula: a likelihood ratio test indicates that we may reject the normal copula specification in favour of the Student's t copula, with a p-value of 0.0026.) Thus the problem of comparing the performance of the rest of these copulas involves non-nested hypothesis testing. This problem has been addressed by Rivers and Vuong (1999), who consider pair-wise comparisons of non-nested models. To compare a given model with multiple alternatives, we employ a test for superior performance proposed by White (2000), and modified by Hansen (2001): the bootstrap reality check[13]. In Table 5 below we present the copula likelihoods and the bootstrap reality check p-values (lower bound, consistent and upper bound) for all copula models considered.

[INSERT TABLE 5 HERE]

The results of the reality check test for superior performance indicate that the Clayton, rotated Clayton and Gumbel copulas may all be rejected at the 10% alpha level. This still leaves us with five copulas that are not statistically distinguishable from the best copula considered. To keep

[12]One copula that was consciously omitted from this list is the Farlie-Gumbel-Morgenstern copula. This copula was excluded due to the limited amount of dependence it is able to consider: correlation under this copula is typically bounded at about one-third.

[13]This test was developed for the comparison of out-of-sample forecasts, and its applicability to in-sample problems such as ours has not yet been verified. With the results of Rivers and Vuong (1999) we expect that this test is valid here. We will verify this in future work.

the amount of computation required for this study tractable, we elected to proceed only with the rotated Gumbel copula, which attained the highest copula likelihood. In Table 6 we present the results of the normal copula estimated on the transformed residuals of the normal models for the marginal distributions, and the rotated Gumbel copula estimated on the transformed residuals of the skewed t models for the marginal distributions.

[INSERT TABLE 6 HERE]

The constant correlation bivariate normal model indicates correlation of 0.72, which corresponds with the figure reported on the raw returns in Table 1. The implied correlation from the constant rotated Gumbel copula model[14] is not constant, due to the time-varying skewness in the marginal distributions, but also averaged 0.72. In Figure 9 we present the time paths of the conditional correlations from the time-varying copula models. Both models show substantial time variation in dependence. This is confirmed by likelihood ratio tests of the significance of the time variation, which give p-values of less than 0.001 for both models. This result is in line with that of Harvey and Siddique (2000), who find substantial evidence of time-varying co-skewness. It is interesting to note that the linear correlations implied by the normal and the copula models are quite similar, though their implications for nonlinearities are different.

[INSERT FIGURE 9 HERE]

The dependence regressors that are most significant are the dividend yield and the conditional mean of the small-cap index returns. Dependence is positively related to the dividend yield, which in turn is thought to track time-varying expected returns. Interestingly, dependence appears negatively related to the conditional mean of the small-cap index returns. This means that downward movements in the conditional mean of the small-cap index coincide with increases in the conditional dependence between the small-cap and large-cap index returns. The fact that both significant regressors are related to the expected value of the two asset returns indicates some relationship between the level of returns and the dependence between the returns, as has been proposed previously in the literature.

In Table 7 below we present the results of the logistic hit and multinomial specification tests of the copula models. These tests are the multivariate extensions of the hit tests discussed above, details on which may be found in Patton (2001a). This table reveals that the bivariate normal models fails in regions 2 and 5, corresponding to the extreme joint upper tail and the central regions, but passes the joint test. The rotated Gumbel copula model estimated on the residuals from the skewed t marginal models also fails in region 5, but passes the joint test of correct specification in all regions. That both copula models fail in region 5 may be a result of the fact that both the

[14] We obtain the correlation coefficient implied by the skewed t - rotated Gumbel model via quadrature.

normal and the skewed t marginal model for the large-cap returns fail the hit test of the central region (region 3 in the marginal hit tests, which corresponds to region 5 in the bivariate hit tests).

[INSERT TABLE 7 HERE]

4.5 Performance of the different models

We now analyse the performance of the different asset allocation decisions made using the various models. We consider five levels of relative risk aversion ($\gamma = 1$, 3, 7, 10 and 20), and six models. The six models are:

1. Always hold the small cap index; $\omega = 1$
2. Always hold the large cap index: $\omega = 0$
3. Always hold an even mix of the two indices: $\omega = 0.5$
4. Optimise the portfolio weight only once, and always use that weight: ω^*_{uncond}
5. Find the optimal portfolio weight for each period using the normal distribution model: $\omega^*_{t,NORM}$
6. Find the optimal portfolio weight for each period using the skewed t - rotated Gumbel copula model: $\omega^*_{t,GUMBEL}$.

The first three portfolios are based on naïve rules, in that they are not the result of an optimisation problem. The fourth portfolio is based on the result of an optimisation problem, and so is marked with an asterix, but the weight is constant; based on the unconditional distribution of returns. The last two portfolios are also based on the result of an optimisation problem, and so are denoted with an asterices, and have weights that (potentially) vary over time, and so are marked with a subscript t. The name of the model used to find the weights is also in the subscript.

We also consider the case that the investor is subject to a short sales constraint, so that ω_t must lie in $[0, 1]$ at all times. This leads, effectively, to a further three models to consider, as the optimised portfolio weights were not always inside $[0, 1]$.

4.5.1 Summary statistics

Firstly, let us look at some summary statistics of the sequence of portfolio returns based on the different models. These are presented in Tables 9 and 10. Ignore for now the column labelled $\omega^*_{t,NORMCOP}$; this column will be discussed in Section 4.6. We present five summary statistics on the monthly return series: the mean, standard deviation, skewness, 5% Value-at-Risk (5% VaR) and 1% Value-at-Risk (1% VaR). We repeat the results on the three naïve portfolios for each level of risk aversion, even though they are obviously invariant to risk aversion, so that comparison with the three optimised portfolios' results can be more easily made. One first feature to notice is that the results for ω^*_{uncond} are very similar to the naïve portfolios. The optimal weights for this portfolio were :

Table 8: *Unconditional portfolio weights,* ω^*_{uncond} .

RRA	1	3	7	10	20
ω^*_{uncond}	0.7423	0.4965	−0.0232	−0.0904	−0.2621

Thus the amount of wealth held in the more risky small cap index is a decreasing function of relative risk aversion. That these weights are not too different from the naïve portfolio weights explains the similarity of their results. For the short sales constrained investors the results are even more similar, since the three most risk averse investors always hold all their wealth in the large cap index.

[INSERT TABLES 9 AND 10 HERE]

Another striking feature of the summary statistics is the much greater mean and standard deviation of the portfolio returns based on the distribution models (normal and Gumbel) than the portfolios with constant weights. This is particularly so for the less risk averse investors ($RRA = 1$ and 3) investor. We ignore parameter estimation uncertainty, and so the query may be raised as to whether the investors would so aggressively invest if they knew that they were using parameter estimates rather than the true parameters. Kandel and Stambaugh (1996) and Brandt (1999) both find that even when parameter estimation uncertainty is accounted for a CRRA investor aggressively seeks the best portfolio. The results for the short sales constrained investors reveal a much smaller difference in mean and risk between the distribution portfolios and the constant weight portfolios.

The final summary statistic to note is the skewness coefficient. As noted in previous sections, CRRA investors have a preference for positively skewed assets, *ceteris paribus*. We can see that the distribution portfolios generally obtain positively skewed portfolios. The skewness coefficients are positive for all risk aversion levels, whereas the portfolios with constant weights generally have negative skewness coefficients. We can see that for all but the least risk averse investor the skewness coefficient on the Gumbel portfolio returns is greater than that on the normal portfolio returns, indicating that the Gumbel model was better able to capture positive skewness. This result also holds for short sales constrained investors.

4.5.2 Performance statistics

Tables 11 and 12 contain some risk-adjusted performance measures on the realised portfolio returns. Again, ignore for now the column labelled $\omega^*_{t,NORMCOP}$. These tables present measures in the form of ratios of average return to some measure of risk. In addition to the usual Sharpe ratio (mean to standard deviation) we present two alternative measures: mean to 5% VaR and mean to 1% VaR. The presence of skewness in the distribution of returns, reported in Tables 9 and 10, implies that standard deviation may not be an appropriate measure of risk. Given the invariance to positive skewness, and its popularity among practitioner, we employ 5% Value-at-Risk (VaR) and 1% VaR as alternative measures of risk. In all cases the VaRs are reported as 'losses' and so are positive numbers; a larger (positive) mean/VaR ratio implies a greater return per unit of risk.

In both of these tables we also present bootstrap 90% confidence intervals on the performance statistics. These are constructed as follows. For each level of risk aversion we obtain the six sequences, $\left\{\omega_t^i(\gamma)\right\}_{t=1}^n$ for $i = 1$ to 6, of portfolio weights from the models considered. Using these weights we obtain the matrix of portfolio returns. We generated 1000 bootstrap samples of the portfolio returns, and computed the three performance measures for each bootstrap sample. We employ the stationary bootstrap of Politis and Romano (1994) to deal with our serially dependent returns[15]. Given the bootstrap distribution of the performance measures, we follow standard procedures, see Efron and Tibshirani (1993) for example, to obtain the confidence intervals.

[INSERT TABLES 11 AND 12 HERE]

The performance statistics indicate that substantial gains may be obtained by employing weights obtained from a model of the conditional distribution of stock returns. Of course, it should be re-iterated that these results are not realistic in a number of important ways: the results are in-sample, and the investor faces no transactions costs and is assumed to be able to trade without moving the market (which is a reasonable assumption for all but the largest fund managers). With these *caveats* in mind, we proceed to describe the risk-adjusted return results. Given the similarities between the portfolios with constant weights, we will use the 50:50 portfolio as a representative of this group of portfolios.

For all levels of risk aversion, the portfolios found using the skewed t -rotated Gumbel models generated Sharpe ratios that were almost twice as large as those of the naïve portfolios. So for each unit of standard deviation, the Gumbel portfolios had a return almost double that of the naïve portfolios. The Sharpe ratios from the normal portfolios were around 1.6 times as great as those

[15]The stationary bootstrap is a type of block bootstrap, where the block lengths are distributed as a Geometric(q) random variable. The average block length is $1/q$. The parameter q is assumed to go to zero as $n \to \infty$, and thus the average block length approaches infinity. However, in small samples the choice of q is somewhat arbitrary. We choose q by running univariate regressions of each portfolio's returns on 120 lags, in both levels and squares to capture serial dependence in the conditional mean and variance. We set $1/q$ equal to the largest significant lag in the $6 \times 2 = 12$ regressions. The results suggested an average block length of about 100 observations. We investigated whether the results were sensitive to the choice of average block length, and found that the results were basically unchanged for block lengths between 36 and 240. Some differences were found for block lengths of 12 to 24; attributable to the serial dependence in the data.

of the naïve portfolios. This indicates that although the risk of the distribution model portfolios were much greater, the returns on these portfolios more than compensated for the increase.

When using Value-at-Risk as the measure of risk the returns to actively modelling and trading these stocks are even greater: for both the 5% VaR and 1% VaR the Gumbel portfolio generated risk-adjusted returns between 2.2 and 3.5 times as large as the 50:50 portfolio. The normal portfolio performed slightly worse than the Gumbel portfolio, attaining between 1.75 and 2.6 times as great a risk-adjusted return. The confidence intervals on all of these measures, however, are quite wide. In the next section we conduct tests to determine whether the differences in these measures between the portfolios are significant.

For the short sales constrained portfolios the differences between the distribution model portfolios and the naïve portfolios are reduced. The constrained Gumbel portfolio attains a Sharpe ratio about 1.43 times as large as the 50:50 portfolio, and the constrained normal portfolio about 1.27 times as large. For the mean/VaR measures the constrained Gumbel portfolio risk-adjusted returns are between 1.65 and 1.81 as large as the 50:50 portfolio, and the constrained Gumbel portfolio risk-adjusted returns are between 1.2 and 1.59 times as large. Thus although these models are constrained to take weights inside $[0, 1]$, they still manage to perform better according to all three performance metrics than the naïve portfolios. Further, the Gumbel portfolio outperforms the normal portfolio in all cases.

4.5.3 Tests for superior portfolio performance

In this section we present the results of two statistical tests for superior performance: a bootstrap test of pair-wise comparisons, and the reality check of White (2000), as modified by Hansen (2001).

We conduct pair-wise comparisons by looking at the bootstrap distribution of the difference in the performance measures of two portfolios. Let the performance measure of portfolio i be μ_i. If the lower bound of the bootstrap 90% confidence interval of $\mu_i - \mu_j$ is greater than zero, then we take model i to be significantly better than model j. If the upper bound of the interval is less than zero then we take model j to be significantly better than model i. If the confidence interval includes zero, then the test is inconclusive, and we cannot statistically distinguish models i and j according to that performance measure. This comparison is a bootstrap version of the test proposed by Rivers and Vuong (1999), and is similar to the test of Diebold and Mariano (1995). The results of these tests are presented in Tables 13 and 14 below. In these tables, we include only the 50:50 portfolio of the three naïve portfolios to save space. The results from the pair-wise comparisons involving this portfolio are representative of the results from comparisons involving the other two naïve portfolios. Pair-wise comparisons involving only naïve portfolios were generally not significant.

[INSERT TABLES 13 AND 14 HERE]

For the unconstrained investor, with any level of risk aversion, the results of the pair-wise comparisons clearly indicate that the portfolio based on the Gumbel model is preferred. The Gumbel portfolio wins in pair-wise comparisons for all three performance measures and all five risk aversion levels, against all alternative portfolios, including the normal portfolio. The normal portfolio almost always beats the naïve portfolio, and beats the unconditional portfolio for all three measures for low levels of risk aversion, but fails to beat the unconditional portfolio on the mean-to-standard deviation and mean-to-1% VaR measures for risk aversion greater than three. The unconditional portfolio is generally not significantly better than the naïve portfolio.

The results for the short sales constrained investor are slightly less conclusive. The Gumbel portfolio still always beats the naïve portfolio, and beats the unconditional portfolio everywhere except for the mean-to-1% VaR measure for risk aversion greater than three. The Gumbel and the normal portfolios are not generally distinguishable: the mean/standard deviation and mean/5% VaR measures generally indicate equal performance, however we do find that the Gumbel portfolio performs significantly better than the normal portfolio on the mean-to-1% VaR measure for all levels of risk aversion.

We now present the results of two applications of the reality check: the first takes the 50:50 portfolio as the benchmark portfolio, and the second takes the normal portfolio as the benchmark portfolio. We test whether the benchmark portfolio is significantly beaten by the best alternative portfolio according to the Sharpe ratio, the mean/5% VaR and the mean/1% VaR ratios. The results of the reality check tests are presented in Tables 15 and 16 below.

[INSERT TABLES 15 AND 16 HERE]

Table 15 clearly shows that we may reject the hypothesis that the 50:50 portfolio performs as well as the best alternative portfolio. This is true for all three performance measures, and all five risk aversion levels, both with and without the short sales constraint. In all of these cases the consistent estimate of the p-value of the test is less than 0.01, (and even the upper bounds on the p-values are less than this figure) indicating that we may very safely conclude that this portfolio is significantly worse than the best alternative. That is, there are significant gains, however measured, to modelling the conditional distribution of returns.

The left panel of Table 16 presents the results of the reality check using the normal portfolio as the benchmark, without a short sales constraint. In this case we may reject the normal portfolio as performing as well as the best alternative for all performance measures and all levels of risk aversion, at the 10% level, except two. The upper bounds on the p-values are also quite low, ranging between 0.0110 and 0.2670, confirming the conclusion that we may reject the null that the normal portfolio performs as well as the best alternative.

When the investor is subject to a short sales constraint the normal portfolio is rejected for fewer cases. Under a short sales constraint we may reject the normal portfolio only when using the mean-to-1% VaR performance measure, and only for investors with risk aversion less than 10. When using the mean-to-standard deviation or mean-to-5% VaR performance measures we cannot reject the hypothesis that the normal portfolio subject to a short sales constraint performs as well as the best alternative portfolio.

The overall conclusions that we may draw based on the results presented above are that there exist significant gains to be had by basing portfolio decisions on a model of the conditional distribution of returns. For various risk-adjusted performance measures and risk aversion levels we found substantial evidence that the naïve portfolio rules were significantly worse than the portfolio rules based on models of the time-varying conditional joint density of the returns. Further, we found that the Gumbel model, developed using copula theory, performed significantly better than the model assuming joint normality of returns in the case that the investor was not subject to a short sales constraint. In the constrained case, the evidence was weaker: the normal portfolio never beat the Gumbel portfolio, but only for the mean-to-1% VaR performance measure did we find evidence that the Gumbel portfolio performed significantly better than the normal portfolio.

4.6 Where are the gains? Marginals versus the copula

The two distribution models analysed and compared in the previous sections are different in both their specifications of the marginal distributions and their specifications of the copula: the bivariate normal model uses normal marginals and a normal copula, while the 'Gumbel' model uses Hansen's (1994) skewed t marginals and a rotated Gumbel copula. We found in the previous section that the Gumbel model generally performed significantly better than the normal model, in terms of risk-adjusted returns. In this final section we investigate where the gains are made: in the improved marginal distribution models or in the improved copula model, or both.

We conduct this analysis by constructing a model that has skewed t marginal distributions, and a normal copula. We will call this model 'NormCop'. The portfolio gains to be had by using more flexible models of the marginal distributions may be gauged by comparing the performance of the normal model with the NormCop model. The gains attributable to a better copula model, *conditioning on better marginal distribution models*, may be measured by comparing the performance of the NormCop model to the Gumbel model.

Tables 9 and 10 contain some summary statistics on the realised returns of the NormCop portfolio, and Tables 11 and 12 contain the performance measures discussed in previous sections. Tables 11 and 12 show quite clearly that a lot of the difference in the risk-adjusted portfolio performance statistics is attributable to the more flexible marginal distribution. While the Gumbel portfolio has better performance than the NormCop portfolio in all but one case, the NormCop portfolio is much closer in performance to the Gumbel portfolio than it is to the normal portfolio. This holds true for both the unconstrained and the short sales constrained investors.

The benefits to modelling the margins more flexibly are clearly significant: we find that the NormCop portfolio is significantly better than the normal portfolio for all fifteen risk aversion / performance measure combinations except one.

Although the NormCop portfolio performs almost as well as the Gumbel portfolio, Table 17 shows that in just under half of the pair-wise comparisons of these portfolios the Gumbel portfolio significantly beats the NormCop portfolio, and in no case does the NormCop portfolio beat the Gumbel portfolio. Thus although the magnitude of the differences in performance measures between the Gumbel and NormCop portfolios are relatively small, they are quite often statistically significant. This supports that idea that the benefits to more flexibly modelling the copula are significant.

[INSERT TABLES 17 AND 18 HERE]

For the short sales constrained investor the differences between the portfolios are reduced, as in Table 14. The NormCop portfolio significantly beats the normal portfolio according to the mean-to-1% VaR measure for all risk aversion levels, but does not in general beat the normal portfolio according to any other measure. The Gumbel portfolio significantly beats the NormCop portfolio for the mean-to-1% VaR for investors with risk aversion greater than 7, and for the mean-to-5% VaR measure for investor with risk aversion of 7, but nowhere else. The NormCop portfolio never significantly beats the Gumbel portfolio. Thus we have some evidence that the NormCop portfolio beats the normal portfolio, and some weak evidence that the Gumbel portfolio beats the NormCop portfolio, in the short sales constrained case. These results are not conclusive, however.

Finally, we perform the bootstrap reality check test with the NormCop portfolio as the benchmark. The results are presented in Table 19 below. In the unconstrained case we are able to reject the NormCop portfolio as being as good as the best alternative when using the Sharpe ratio, for risk aversion of 3, 7 and 10. For the mean-to-5% VaR performance measure we can reject the NormCop portfolio only for risk aversion of 7, and for the mean-to-1% VaR performance measure we can reject the NormCop portfolio for risk aversion of 10 and 20. Thus, we have some evidence that the NormCop portfolio is significantly beaten by the best alternative model, in the unconstrained case. For the short sales constrained investor we are unable to reject the NormCop portfolio for any performance measure/risk aversion level combination. For the constrained case we thus conclude that it performs as well as the best alternative portfolio.

[INSERT TABLE 19 HERE]

Overall, the results presented in this section suggest that the most significant benefit of copula theory for asset allocation is that it allows us to form joint distributions with flexibly specified marginal distributions. In our case, allowing for excess kurtosis and time-varying skewness of the individual assets greatly improves the risk-adjusted returns of the resulting optimal portfolios, relative to the bivariate normal distribution. The use of a more carefully selected model for the copula leads to an additional small but generally significant improvement in risk-adjusted portfolio returns. When the investor is constrained to only take portfolio weights between 0 and 1 the significance of the benefits to more flexibly specifying the distribution are reduced, though in general still positive.

5 Conclusions and future work

In this paper we considered the impact that skewness and asymmetric dependence have on the portfolio decisions of a CRRA investor. Evidence of skewness in stock returns has been widely reported over the years, and is generally accepted as a common feature of stock returns. This is of interest as any investor that exhibits non-increasing absolute risk aversion, a very weak requirement, can be shown to exhibit a preference for positively skewed assets, *ceteris paribus*.

Recent work, see Erb, *et al.*, (1994), Ang and Chen (2001) and Longin and Solnik (2001) *inter alia*, has produced evidence that stock returns exhibit greater dependence during bear markets than during bull markets. We showed the link between univariate skewness and asymmetric dependence between assets: asymmetric dependence between assets can lead to skewed portfolios, even if the individual assets are not themselves skewed. In particular, if two assets exhibit greater dependence in bear markets than in bull markets, then portfolios of these assets (when the portfolio weight is in between 0 and 1) will be negatively skewed. Thus the preference for positive skew translates to an aversion to assets that exhibit greater dependence in bear markets than in bull markets.

We considered the problem of allocating wealth between the CRSP small-cap and large-cap indices, using monthly data from January 1954 to December 1999. This problem is representative of that of choosing between a high risk - high return asset and a lower risk - lower return asset. We adopted a parametric approach, using distribution models that are able to capture time-varying means and variances of stock returns, and also (possibly time-varying) skewness and kurtosis. Further, we employed models of the dependence structure of these asset returns that allowed for greater dependence during bear markets than bull markets, and allowed for changes in this dependence structure through time. Our distribution models were constructed using copula theory, a field of multivariate statistics that is gaining increasing interest in economics and finance. Using copula theory we were able to model separately the individual assets' distributions and their dependence structure. The class of possible multivariate distributions using copula theory is much greater than the class of existing multivariate distributions in economics, and the estimation of the model was simplified by employing a two-stage estimator discussed in Patton (2001b).

We measured the importance of skewness and asymmetric dependence for asset allocation by comparing the risk-adjusted performance of a portfolio based on a bivariate normal distribution model with a portfolio based on a model developed using copula theory. The significance of the differences in portfolio performance were tested using bootstrap methods, and substantial evidence was found that skewness and asymmetric dependence do have important implications for asset allocation: we found that portfolio based on the copula distribution model performed significantly better than the portfolio based on the bivariate normal distribution for all performance measures considered and all levels of relative risk aversion. When the investor was subject to a short sales constraint the evidence was reduced, though we still found for many cases that the portfolio based on the normal distribution model could be rejected in favour of that based on the copula distribution model. We conclude that accounting for the observed skewness and asymmetric dependence in stock returns leads to significantly better asset allocation decisions.

The relative importance of marginal distribution skewness and dependence structure asymmetry was also analysed. We introduced a second copula distribution model, that captured the time-varying skewness of the individual assets, but that imposed the same dependence structure as the bivariate normal distribution. The portfolio based on this third distribution model was found to significantly beat the portfolio based on the normal distribution, indicating the importance of skewness. The portfolio based on the distribution model that allowed for asymmetric dependence was found to beat the portfolio that allowed for skewness but no asymmetric dependence in only some cases. We thus concluded that the primary benefit of copula theory for asset allocation problems is that it allows the construction of multivariate distributions with flexible marginal distributions. The benefits to more flexibly specifying the dependence structure, given better specified marginal distribution models, are less significant though still positive.

For every question answered by this paper many new questions arise. The models and optimisation problems in this paper were all in-sample, and the finding that significant benefits existed to modelling asymmetries leads to the question of whether this result would hold in an out-of-sample study. Further, it would be of interest to extend the problem to that of multiple assets - do the benefits to flexibly modelling the joint distribution increase with the dimension of the distribution? In this paper we ignored the impact of parameter estimation uncertainty on the investor's optimisation problem, and it would be interesting to determine how the results would change when this is taken into account. Our copula distribution models are more heavily parameterised than the normal distribution model, and it may be that the benefits derived from the flexible specification are outweighed by the increased parameter estimation uncertainty. Finally, it would be of great interest to compare the results of the methods presented in this paper with other parametric approaches, such as Ang and Bekaert (2001), and with nonparametric approaches, such as those of Brandt (1999) or Aït-Sahalia and Brandt (2001). All of these questions are left for future work.

6 Appendix 1: A brief introduction to copula theory

The introduction presented below follows closely that of Patton (2001a). We will firstly introduce the copula via standard theory on the distribution of transformations of random variables. Following that, the more general theory of conditional copulas is presented. A very readable and thorough introduction to the theory of copulas may be found in Nelsen (1999).

An important transformation in copula theory is the 'probability integral transformation' (PIT). The first analysis of the distribution of the PIT is quite old, dating back to Fisher (1932). For a more recent reference see, for example, Casella and Berger (1990). Let $U_t \equiv F_t(X_t)$ and $V_t \equiv G_t(Y_t)$. We then say that U_t and V_t are the 'PITs of X_t and Y_t'. The distribution of the PIT is given in Theorem 3 below.

Theorem 3 (Fisher, 1932) *Let $X_t \sim F_t$ and let F_t be a continuous distribution function. Then $U_t \equiv F_t(X_t) \sim Unif(0,1)$.*

Proof. *See Theorem 2.1.4 of Casella and Berger (1990).* ■

With this result in hand, we may introduce the copula using basic statistical theory.

6.1 The copula and transformations of random variables

In this section, for the sake of simplicity, we will suppress the dependence of the random variables and their distributions on t. Let $U \equiv F(X)$ and $V \equiv G(Y)$, as above. We will now find the joint density of U and V according to basic results in mathematical statistics on the distribution of transformations of random variables. One standard reference for this is Casella and Berger (1990). We will denote the joint density of U and V as c, which turns out to be the 'copula density'.

Since F and G are strictly increasing and continuous, we have that $X = F^{-1}(U)$ and $Y = G^{-1}(V)$, and $\frac{\partial X}{\partial U} = \left(\frac{\partial U}{\partial X}\right)^{-1} = \left(\frac{\partial F(X)}{\partial X}\right)^{-1} = f(X)^{-1}$ and $\frac{\partial Y}{\partial V} = \left(\frac{\partial V}{\partial Y}\right)^{-1} = \left(\frac{\partial G(Y)}{\partial Y}\right)^{-1} = g(Y)^{-1}$. Note that $\frac{\partial X}{\partial V} = \frac{\partial Y}{\partial U} = 0$. Then,

$$\begin{aligned} c(u,v) &= h(X(u), Y(v)) \cdot \begin{vmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{vmatrix} \\ &= h\left(F^{-1}(u), G^{-1}(v)\right) \cdot \frac{\partial X}{\partial U} \cdot \frac{\partial Y}{\partial V} \\ c(u,v) &= \frac{h\left(F^{-1}(u), G^{-1}(v)\right)}{f\left(F^{-1}(u)\right) \cdot g\left(G^{-1}(v)\right)} \qquad (13) \end{aligned}$$

Equation (13) shows that the copula density of X and Y is equal to the ratio of the joint density, h, to the product of the marginal densities, f and g. From this expression we can obtain a first result on the properties of copulas: if X and Y are independent, then the copula density takes the value 1 everywhere, since in that case the joint density is equal to the product of the marginal densities. Since we know that the marginal densities of U and V are uniform, by Theorem 3 above, we thus have that if X and Y are independent the joint distribution of U and V is the bivariate Uniform$(0,1)$ distribution.

We can also use equation (13) to derive an expression for h as a function of x and y instead:

$$\begin{aligned} h\left(F^{-1}(u), G^{-1}(v)\right) &= f\left(F^{-1}(u)\right) \cdot g\left(G^{-1}(v)\right) \cdot c(u,v) \\ h(x,y) &= f(x) \cdot g(y) \cdot c(F(x), G(y)) \end{aligned} \tag{14}$$

Equation (14) is the 'density version' of Sklar's (1959) theorem: the joint density, h, can be decomposed into product of the marginal densities, f and g, and the copula density, c. Sklar's theorem holds under more general conditions than the ones we imposed for this illustration, and below we discuss the general proof.

6.2 The theory of the conditional copula

For an introduction to the general theory of copulas the reader is referred to Nelsen (1999) or Chapter 6 of Schweizer and Sklar (1983). We will start with a few very basic, but very important, definitions based on those in Nelsen (1999). The second condition below refers to the 'H_t-volume' of a rectangle $[x_1, x_2] \times [y_1, y_2]$ in $\bar{\mathbb{R}}^2$, denoted by V_{H_t}. This is simply the probability of observing a point in the region $[x_1, x_2] \times [y_1, y_2]$. It is expressed in the following way as it generalises more easily to the multivariate case.

Definition 7 *A conditional bivariate distribution function is a right continuous function* $H_t : \bar{R}^2 \to [0,1]$ *with the properties:*

1. $H_t(x, -\infty|\mathcal{F}_{t-1}) = H_t(-\infty, y|\mathcal{F}_{t-1}) = 0,$ *and* $H_t(\infty, \infty|\mathcal{F}_{t-1}) = 1$
2. $V_{H_t}([x_1,x_2] \times [y_1,y_2]) \equiv H_t(x_2, y_2|\mathcal{F}_{t-1}) - H_t(x_1, y_2|\mathcal{F}_{t-1}) - H_t(x_2, y_1|\mathcal{F}_{t-1}) + H_t(x_1, y_1|\mathcal{F}_{t-1}) \geq 0$ *for all* $x_1, x_2, y_1, y_2 \in \bar{\mathbb{R}}$, *and* $x_1 \leq x_2$, $y_1 \leq y_2$.

   where $\mathcal{F}_{t-1}$ *is some conditioning set.*

The first condition simply provides the upper and lower bounds on the distribution function. The second condition ensures that the probability of observing a point in the region $[x_1, x_2] \times [y_1, y_2]$ is non-negative[16]. We now define the conditional copula.

Definition 8 *A two-dimensional conditional copula is a function* $C_t : [0,1] \times [0,1] \to [0,1]$ *with the following properties:*

[16] If we set $x_2 = x_1 + \varepsilon$ and $y_2 = y_1 + \varepsilon$ and let $\varepsilon \to 0^+$, then it becomes clear that this definition is just the generalisation of the condition that if the bivariate density exists, it must be non-negative on the domain of H_t.

1. $C_t(u,0|\mathcal{F}_{t-1}) = C_t(0,v|\mathcal{F}_{t-1}) = 0$, *and* $C_t(u,1|\mathcal{F}_{t-1}) = u$ *and* $C_t(1,v|\mathcal{F}_{t-1}) = v$, *for every* u, v *in* $[0,1]$

2. $V_{C_t}([u_1,u_2] \times [v_1,v_2] |\mathcal{F}_{t-1}) \equiv C_t(u_2,v_2|\mathcal{F}_{t-1}) - C_t(u_1,v_2|\mathcal{F}_{t-1}) - C_t(u_2,v_1|\mathcal{F}_{t-1}) + C_t(u_1,v_1|\mathcal{F}_{t-1}) \geq 0$ *for all* $u_1,u_2,v_1,v_2 \in [0,1]$, *such that* $u_1 \leq u_2$ *and* $v_1 \leq v_2$.

   where $\mathcal{F}_{t-1}$ *is some conditioning set.*

The first condition of Definition 8 provides the lower bound on the distribution function, and ensures that the marginal distributions, $C_t(u,1|\mathcal{F}_{t-1})$ and $C_t(1,v|\mathcal{F}_{t-1})$, are uniform. The condition that V_{C_t} is non-negative has the same interpretation as the second condition of Definition 7: it simply ensures that the probability of observing a point in the region $[u_1,u_2] \times [v_1,v_2]$ is non-negative.

By drawing on the above conditions for the conditional copula, and extending its domain to $\bar{\mathbb{R}}^2$, we may alternatively define a conditional copula as the conditional bivariate distribution of a pair of random variables (U_t, V_t) having margins that are $Unif(0,1)$. The extension of the domain to $\bar{\mathbb{R}}^2$ is accomplished as follows:

$$\text{Let} \quad C_t^*(u,v|\mathcal{F}_{t-1}) = \begin{cases} 0 & \text{for } u < 0 \text{ or } v < 0, \\ C_t(u,v|\mathcal{F}_{t-1}) & \text{for } (u,v) \in [0,1] \times [0,1], \\ u & \text{for } u \in [0,1], v > 1, \\ v & \text{for } u > 1, v \in [0,1], \\ 1 & \text{for } u > 1, v > 1. \end{cases} \tag{15}$$

The link between the probability integral transformation and the theory of copulas now becomes clear: the copula is the joint distribution function of the probability integral transforms of each of the variables X_t and Y_t with respect to their marginal distributions, F_t and G_t. We now move on to an extension of the the key result in the theory of copulas: Sklar's (1959) theorem for conditional distributions:

Theorem 4 (Sklar's Theorem for Continuous Conditional Distributions) *Let* H_t *be a conditional bivariate distribution function with continuous margins* F_t *and* G_t, *and let* $\mathcal{F}_{t-1}$ *be some conditioning set. Then there exists a unique conditional copula* $C_t : [0,1] \times [0,1] \to [0,1]$ *such that*

$$H_t(x,y|\mathcal{F}_{t-1}) = C_t(F_t(x|\mathcal{F}_{t-1}), G_t(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}), \quad \forall x, y \in \bar{\mathbb{R}} \tag{16}$$

Conversely, if C_t *is a conditional copula and* F_t *and* G_t *are the conditional distribution functions of two random variables* X_t *and* Y_t, *then the function* H_t *defined by equation (1) is a bivariate conditional distribution function with margins* F_t *and* G_t.

Proof. See Patton (2001a). ■

The density function equivalent of (1) is useful for maximum likelihood analysis, and is obtained quite easily, provided that F_t and G_t are differentiable, and H_t and C_t are twice differentiable.

$$\begin{aligned}
h_t(x, y|\mathcal{F}_{t-1}) &\equiv \frac{\partial^2 H_t(x, y|\mathcal{F}_{t-1})}{\partial x \partial y} \\
&= \frac{\partial F_t(x|\mathcal{F}_{t-1})}{\partial x} \cdot \frac{\partial G_t(y|\mathcal{F}_{t-1})}{\partial y} \cdot \frac{\partial^2 C_t(F_t(x|\mathcal{F}_{t-1}), G_t(y|\mathcal{F}_{t-1})|\mathcal{F}_{t-1})}{\partial (F_t(x|\mathcal{F}_{t-1}) \partial (G_t(y|\mathcal{F}_{t-1}))} \\
&= f_t(x|\mathcal{F}_{t-1}) \cdot g_t(y|\mathcal{F}_{t-1}) \cdot \frac{\partial^2 C_t(u, v|\mathcal{F}_{t-1})}{\partial u \partial v} \\
&\equiv f_t(x|\mathcal{F}_{t-1}) \cdot g_t(y|\mathcal{F}_{t-1}) \cdot c_t(u, v|\mathcal{F}_{t-1}), \quad \forall (x, y) \in \bar{\mathbb{R}}^2 \qquad (17)
\end{aligned}$$

where $u \equiv F_t(x|\mathcal{F}_{t-1})$, and $v \equiv G_t(y|\mathcal{F}_{t-1})$. The expression in equation (17) is precisely the same as that in equation (14), which we obtained using the theory on the distribution of transformations of random variables. Taking logs of both sides we obtain:

$$\mathcal{L}_{XY} = \mathcal{L}_X + \mathcal{L}_Y + \mathcal{L}_C \qquad (18)$$

and so the joint log-likelihood is equal to the sum of the marginal log-likelihoods and the copula log-likelihood.

We can also obtain a corollary to Theorem 1, analogous to that of Nelson's (1999) corollary to Sklar's Theorem, which enables us to extract the conditional copula from any conditional bivariate distribution function, but first we need the definition of the 'quasi-inverse' of a function.

Definition 9 *The quasi-inverse, $F^{(-1)}$, of a distribution function F is defined as:*

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\}, \text{ for } u \in [0, 1]. \qquad (19)$$

If F is strictly increasing then the above definition returns the usual functional inverse of F, but more importantly it allows us to consider inverses of non-strictly increasing functions.

Corollary 1 *Let H_t be any conditional bivariate distribution with continuous marginal distributions, F_t and G_t, and let $F_t^{(-1)}$ and $G_t^{(-1)}$ denote the (quasi-) inverses of the marginal distributions. Finally, let $\mathcal{F}_{t-1}$ be some conditioning set. Then there exists a unique conditional copula $C_t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that*

$$C_t(u, v|\mathcal{F}_{t-1}) = H_t\left(F_t^{(-1)}(u|\mathcal{F}_{t-1}), G_t^{(-1)}(v|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}\right), \quad \forall u, v \in [0, 1] \qquad (20)$$

Proof. See Patton (2001a). ■

This corollary completes the idea that a bivariate distribution function may be decomposed into three parts. Given any two marginal distributions and any copula we have a joint distribution, and from any given joint distribution we can extract the implied marginal distributions and copula.

7 Appendix 2: Some results on the skewed t distribution.

We provide in this section a few results on Hansen's (1994) skewed t distribution. In the original article the *p.d.f.* of the skewed t random variable was provided; below we provide the *c.d.f.* and inverse *c.d.f.* (useful for random number generation) of the skewed t in terms of the standard Student's t random variable. The motivation for doing this is that most econometric packages (such as Gauss and Matlab) have code available for the standard Students t. With the following results it can be utilised for the skewed t distribution. Matlab code for each of the functions presented below will be available on the author's web site in the near future.

Let Y be a skewed t random variable, with density function $g(\nu, \lambda)$. The variable Y has mean zero and variance one by construction, and so is a suitable model for the standardised residuals of some conditional mean and variance model. The parameters ν and λ control the kurtosis and skewness of the variable.

Skewed t density

$$g(y;\nu,\lambda) = \begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{by+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{for } y \leq -\frac{a}{b} \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{by+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{for } y > -\frac{a}{b} \end{cases}, \text{ where} \tag{21}$$

$$c \equiv \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}} \tag{22}$$

$$b \equiv \sqrt{1+3\lambda^2-a^2} \tag{23}$$

$$a \equiv 4\lambda c\left(\frac{\nu-2}{\nu-1}\right) \tag{24}$$

Let X be a (standard) Student's t_ν random variable, with mean zero and variance $\frac{\nu}{\nu-2}$. Denote the *c.d.f.* of X as $F(\nu)$. Below we derive an expression for the *c.d.f.* of a skewed t random variable in terms of F.

Skewed t cumulative distribution function

$$G(y;\nu,\lambda) = \begin{cases} (1-\lambda)\cdot F\left(\sqrt{\frac{\nu}{\nu-2}}\left(\frac{by+a}{1-\lambda}\right);\nu\right) & \text{for } y \leq -\frac{a}{b} \\ \frac{1-\lambda}{2}+(1+\lambda)\cdot\left[F\left(\sqrt{\frac{\nu}{\nu-2}}\left(\frac{by+a}{1-\lambda}\right);\nu\right)-0.5\right] & \text{for } y > -\frac{a}{b} \end{cases} \tag{25}$$

where a, b and c are as defined for the density function.

Finally, we present the inverse *c.d.f.* of the skewed t distribution, which is denoted $G^{-1}(\nu, \lambda)$. We will express it in terms of the inverse distribution of a Student's t random variable, denoted $F^{-1}(\nu)$.

Inverse Skewed t cumulative distribution function

$$G^{-1}(u;\nu,\lambda)=\begin{cases}\frac{1-\lambda}{b}\sqrt{\frac{\nu-2}{\nu}}\cdot F^{-1}\left(\frac{u}{1-\lambda};\nu\right)-\frac{a}{b} & \text{for } 0<u<\frac{1-\lambda}{2}\\ \frac{1+\lambda}{b}\sqrt{\frac{\nu-2}{\nu}}\cdot F^{-1}\left(0.5+\frac{1}{1+\lambda}\left(u-\frac{1-\lambda}{2}\right)\right)-\frac{a}{b} & \text{for } \frac{1-\lambda}{2}\leq u<1\end{cases}$$

The inverse *c.d.f.* can be used to generate random draws from the skewed t distribution as follows: firstly obtain n draws from the Uniform$(0,1)$ distribution, $\{u_t\}_{t=1}^{n}$. Almost all software packages provide such a feature. Then define $y_t \equiv G^{-1}(u_t;\nu,\lambda)$. The resulting sequence $\{y_t\}_{t=1}^{n}$ are draws from the skewed t distribution. The ability to generate such random variables is useful for Monte Carlo studies involving this distribution, amongst other things.

8 Appendix 3: Copula functional forms

In this appendix we provide the functional forms of the copulas used in this paper. The *c.d.f.* forms will be denoted C, and the *p.d.f.* forms c. For further details on any of these copulas, or for other copulas, the reader is referred to Joe (1997) and Nelsen (1999).

Normal Copula

$$C_N(u,v;\rho) = \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left\{\frac{-(r^2-2\rho rs+s^2)}{2(1-\rho^2)}\right\}dr\,ds$$

$$c_N(u,v;\rho) = \frac{1}{\sqrt{1-\rho^2}}\exp\left\{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)}{2(1-\rho^2)}+\frac{\Phi^{-1}(u)^2\,\Phi^{-1}(v)^2}{2}\right\}$$

$$\rho \in (-1,1)$$

Clayton Copula (Kimeldorf and Sampson Copula in Joe (1997))

$$C_C(u,v;\theta) = \left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$$

$$c_C(u,v;\rho) = (1+\theta)(uv)^{-\theta-1}\left(u^{-\theta}+v^{-\theta}-1\right)^{-2-1/\theta}$$

$$\theta \in [0,\infty)$$

Rotated Clayton Copula

$$C_{RC}(u,v;\theta) = u+v-1+C_C(1-u,1-v;\theta)$$

$$c_{RC} = c_C(1-u,1-v;\theta)$$

$$\theta \in [0,\infty)$$

Joe-Clayton Copula (family BB7 in Joe (1997))

$$
\begin{aligned}
C_{JC}(u,v|\kappa,\gamma) &= 1-\left(\left\{\left[1-(1-u)^{\kappa}\right]^{-\gamma}+\left[1-(1-v)^{\kappa}\right]^{-\gamma}-1\right\}^{-1/\gamma}\right)^{1/\kappa} \\
c_{JC}(u,v|\kappa,\gamma) &= \textit{very long and complicated. Available from author on request.} \\
\kappa &\geq 1,\ \gamma>0
\end{aligned}
$$

Plackett Copula

$$
\begin{aligned}
C_P(u,v;\pi) &= \frac{1}{2(\pi-1)}\left(1+(\pi-1)(u+v)-\sqrt{(1+(\pi-1)(u+v))^2-4\pi(\pi-1)uv}\right) \\
c_P(u,v;\pi) &= \frac{\pi(1+(\pi-1)(u+v-2uv))}{\left((1+(\pi-1)(u+v))^2-4\pi(\pi-1)uv\right)^{3/2}} \\
\pi &\in [0,\infty)
\end{aligned}
$$

Frank Copula

$$
\begin{aligned}
C_F(u,v;\lambda) &= \frac{-1}{\lambda}\log\left(\frac{\left(1-e^{-\lambda}\right)-\left(1-e^{-\lambda u}\right)\left(1-e^{-\lambda v}\right)}{\left(1-e^{-\lambda}\right)}\right) \\
c_F(u,v;\lambda) &= \frac{\lambda\left(1-e^{-\lambda}\right)e^{-\lambda(u+v)}}{\left(\left(1-e^{-\lambda}\right)-\left(1-e^{-\lambda u}\right)\left(1-e^{-\lambda v}\right)\right)^2} \\
\lambda &\in [0,\infty)
\end{aligned}
$$

Gumbel Copula

$$
\begin{aligned}
C_G(u,v;\delta) &= \exp\left\{-\left((-\log u)^{\delta}+(-\log v)^{\delta}\right)^{1/\delta}\right\} \\
c_G(u,v;\delta) &= \frac{C_G(u,v;\delta)\left((-\log u)(-\log v)\right)^{\delta-1}}{uv\left((-\log u)+(-\log v)\right)^{2-1/\delta}}\left(\left((-\log u)^{\delta}+(-\log v)^{\delta}\right)^{1/\delta}+\delta-1\right) \\
\delta &\in [1,\infty)
\end{aligned}
$$

Rotated Gumbel Copula

$$
\begin{aligned}
C_{RG}(u,v;\delta) &= u+v-1+C_G(1-u,1-v;\delta) \\
c_{RG}(u,v;\delta) &= c_G(1-u,1-v;\delta) \\
\delta &\in [1,\infty)
\end{aligned}
$$

9 Appendix 4: Proofs (draft)

Proof of Proposition 1. Straightforward algebraic manipulation. ■

Proof of Lemma 1. For $i = 0$, $E\left[|X-\mu_x|^i \cdot |Y-\mu_y|^{k-i}\right] = E\left[|Y-\mu_y|^k\right] < \infty$ by assumption. Similarly, for $i = k$, $E\left[|X-\mu_x|^i \cdot |Y-\mu_y|^{k-i}\right] = E\left[|X-\mu_x|^k\right] < \infty$ by assumption. For $0 < i < k$ we make use of Hölder's inequality:

$$\begin{aligned} E\left[|X-\mu_x|^i \cdot |Y-\mu_y|^{k-i}\right] &\leq E\left[\left(|X-\mu_x|^i\right)^{\frac{k}{i}}\right]^{\frac{i}{k}} \cdot E\left[\left(|Y-\mu_y|^{k-i}\right)^{\frac{k}{k-i}}\right]^{\frac{k-i}{k}} \\ &= E\left[|X-\mu_x|^k\right]^{\frac{i}{k}} \cdot E\left[|Y-\mu_y|^k\right]^{\frac{k-i}{k}} \\ &< \infty \quad \text{by assumption.} \end{aligned}$$

Thus $E\left[|X-\mu_x|^i \cdot |Y-\mu_y|^{k-i}\right] < \infty$ for $0 \leq i \leq k$. ■

Proof of Proposition 2.

$$\begin{aligned} E\left[(Z-\mu_z)^k\right] &= \sum_{i=0}^{k} \omega^i (1-\omega)^{k-i} \binom{k}{i} E\left[(X-\mu_x)^i \cdot \left(Y-\mu_y\right)^{k-i}\right] \\ &\leq \sum_{i=0}^{k} \omega^i (1-\omega)^{k-i} \binom{k}{i} E\left[|X-\mu_x|^i \cdot |Y-\mu_y|^{k-i}\right] \\ &< \infty \quad \text{by the above Proposition.} \end{aligned}$$

■

Proof of Proposition 3. Follows from the definition of skewness given above, and Proposition 1. ■

Proof of Theorem 2. See the proof of Theorem 3.2 of Nelsen (1993). ■

Proof of Lemma 2. We will show that the statement holds for $E[Y|X=x]$. The same steps can be applied for $E[X|Y=y]$. Notice that

$$\begin{aligned} E[Y|X=x+\mu_x] - \mu_y &= \mu_y - E[Y|X=\mu_x - x] \text{ for all } x \in \mathbb{R} \Leftrightarrow \\ E[Y|X=x] &= 2\mu_y - E[Y|X=2\mu_x - x] \text{ for all } x \in \mathbb{R} \end{aligned}$$

We will use the latter of these two equivalent statements in the proof.

$$\begin{aligned}
E[Y|X=x] &\equiv \int_{-\infty}^{\infty} y \cdot h_{y|x}(y|x) \cdot dx \\
&= \int_{-\infty}^{\infty} y \cdot \frac{h(x,y)}{f(x)} \cdot dx, \text{ by the definition of a conditional density} \\
&= \int_{-\infty}^{\infty} y \cdot \frac{h(x,y)}{f(2\mu_x - x)} \cdot dx, \text{ by the symmetry of } X \\
&= \int_{-\infty}^{\infty} y \cdot \frac{h(2\mu_x - x, 2\mu_y - y)}{f(2\mu_x - x)} \cdot dx, \text{ by the radial symmetry of } (X,Y) \\
&= \int_{-\infty}^{\infty} (2\mu_y - y) \cdot \frac{h(2\mu_x - x, y)}{f(2\mu_x - x)} \cdot dx, \text{ change of variables} \\
&= 2\mu_y \int_{-\infty}^{\infty} \frac{h(2\mu_x - x, y)}{f(2\mu_x - x)} \cdot dx - \int_{-\infty}^{\infty} y \cdot \frac{h(2\mu_x - x, y)}{f(2\mu_x - x)} \cdot dx \\
&= 2\mu_y - E[Y|X = 2\mu_x - x], \text{ integral of conditional density is one,} \\
&\quad \text{and definition of conditional expectation}
\end{aligned}$$

The statement for the case that $\mu_x = \mu_y = 0$ follows trivially. ■

Proof of Proposition 4. Let $Z \sim K$. We need to show that K is symmetric, i.e., that $k(\mu_z + z) = k(\mu_z - z)\ \forall\ z$. This is achieved by considering the joint distribution of (X,Z), which we will define as H^*. We will show that H^* is radially symmetric, and thus that both margins of H^*, F and K, are symmetric. The joint distribution of (X,Z) is found using theory on the distribution of transforms of random variables:

$$h^*(x,z) = \left|\frac{1}{1-\omega}\right| \cdot h\left(x, \frac{1}{1-\omega}(z - \omega x)\right)$$

H^* is radially symmetric if and only if $h^*(x,z) = h^*(2\mu_x - x, 2\mu_z - z)\ \forall\ (x,z) \in \mathbb{R}^2$.

$$\begin{aligned}
h^*(2\mu_x - x, 2\mu_z - z) &= \left|\frac{1}{1-\omega}\right| \cdot h\left(2\mu_x - x, \frac{1}{1-\omega}\left(2\left(\omega\mu_x + (1-\omega)\mu_y\right) - z - \omega(2\mu_x - x)\right)\right) \\
&= \left|\frac{1}{1-\omega}\right| \cdot h\left(2\mu_x - x, 2\mu_y - \frac{1}{1-\omega}(z - \omega x)\right) \\
&= \left|\frac{1}{1-\omega}\right| \cdot h\left(x, \frac{1}{1-\omega}(z - \omega x)\right), \text{ by the radial symmetry of } h \\
&= h^*(x,z)
\end{aligned}$$

Thus h^* is radially symmetric about (μ_x, μ_z), which implies that the distribution of Z is symmetric about μ_z. ■

Proof of Proposition 5. Follows directly from the definition of negative bivariate skewness and the equation for the skewness of a portfolio, given in Proposition 3. ■

Proof of Proposition 6. Since X and Y are symmetric, we need only look at the co-skewness terms. Consider $M_{12}[X,Y]$:

$$\begin{aligned}
M_{12}[X,Y] &\equiv E\left[(X-\mu_x)\left(Y-\mu_y\right)^2\right] \\
&= E\left[\left(Y-\mu_y\right)^2 (E[X|Y=y]-\mu_x)\right] \\
&\equiv \int_{-\infty}^{\infty} \left(y-\mu_y\right)^2 (EE[X|Y=y]-\mu_x)\, g(y)\, dy \\
&= \int_{\mu_y}^{\infty} \left(y-\mu_y\right)^2 (E[X|Y=y]-\mu_x)\, g(y)\, dy \\
&\quad + \int_{-\infty}^{\mu_y} \left(y-\mu_y\right)^2 (E[X|Y=y]-\mu_x)\, g(y)\, dy \\
&= \int_{\mu_y}^{\infty} \left(y-\mu_y\right)^2 (E[X|Y=y]-\mu_x)\, g(y)\, dy \\
&\quad + \int_{\mu_y}^{\infty} \left(\mu_y-y\right)^2 \left(E\left[X|Y=2\mu_y-y\right]-\mu_x\right) g\left(2\mu_y-y\right) dy \\
&= \int_{\mu_y}^{\infty} \left(y-\mu_y\right)^2 \left(E[X|Y=y]-2\mu_x+E\left[X|Y=2\mu_y-y\right]\right) g(y)\, dy \\
&\leq 0
\end{aligned}$$

since $\left(y-\mu_y\right)^2$ and $g(y)$ are positive for all y, and $E[X|Y=y]-2\mu_x+E\left[X|Y=2\mu_y-y\right]$ is (weakly) negative for all $y \geq \mu_y$.

We can similarly show that $M_{21}[X,Y] \leq 0$, and thus that $Skew[Z] < 0$ for $\omega \in (0,1)$. ◾

Proof of Proposition 7.

$$\begin{aligned}
M_3[Z] &= 3\omega(1-\omega)^2 M_{12}[X,Y] + 3\omega^2(1-\omega) M_{21}[X,Y] \\
&= 3\omega(1-\omega) M_{12}[X,Y] \\
&< 0
\end{aligned}$$

since $M_{12} < 0$ and $\omega(1-\omega) < 0$ for $\omega < 0$ and $\omega > 1$. ◾

10 Tables

Table 1: *Descriptive Statistics*		
	Small caps	*Large caps*
Mean*	9.9549	7.9748
Std Dev*	21.2932	14.2888
Skewness	0.0558	−0.3795
5% VaR	8.7973	6.2306
1% VaR	18.9576	9.6657
Kurtosis	7.5647	4.9088
Min	−29.3153	−20.8934
Max	38.3804	16.8145
Jarque-Bera	479.5162	97.0484
p-val	*0.0000*	*0.0000*
Correlation	0.7210	

Note to Table 1: The statistics marked with an asterix were annualised to ease interpretation. 'Jarque-Bera' refers to the test for normality of the unconditional distribution of returns.

Table 3: *LM Tests of independence and Kolmogorov-Smirnov tests of the density*				
	Normal		*Skewed t*	
	Small caps	Large caps	Small caps	Large caps
First moment	83.3762	75.4059	85.0092	76.9122
p-value	*0.4158*	*0.6499*	*0.3706*	*0.6060*
Second moment	79.8539	90.7621	75.7961	87.0697
p-value	*0.5186*	*0.2317*	*0.6386*	*0.3168*
Third moment	81.2354	85.9532	76.9359	83.3897
p-value	*0.4777*	*0.3455*	*0.6053*	*0.4154*
Fourth moment	77.9323	93.5672	74.5048	89.2062
p-value	*0.5758*	*0.1778*	*0.6755*	*0.2657*
K-S stat	0.0419	0.0377	0.0159	0.0157
K-S p-value	*0.2921*	*0.4186*	*0.9991*	*0.9993*

Note to Table 3: This table presents the results of LM tests of the independence of the first four moments of the variables U_t and V_t, described in the text. We regress $(u_t - \bar{u})^k$ and $(v_t - \bar{v})^k$ on twelve lags of both variables, for $k = 1, 2, 3, 4$. The test statistic is $(T - 24) \cdot R^2$ for each regression, and is distributed under the null as χ^2_{24}.

Table 2a: *Results for the small-cap marginal distribution*				
	Normal		*Skewed t*	
	Coeff	Std Error	Coeff	Std Error
Conditional mean				
Constant	0.1530	1.6819	0.1530	1.1544
AR(1)	0.1368	0.0479	0.1368	0.0596
AR(12)	0.1395	0.1378	0.1395	0.0423
R_{ft-1}	−0.5323	1.1410	−0.5323	0.1926
SPR_{t-1}	3.3595	5.1048	3.3595	1.1358
Conditional variance				
Constant	−2.1826	11.7108	−0.8247	1.6754
$\varepsilon_{t-1}^2 \cdot \mathbf{1}\{\varepsilon_{t-1} < 0\}$	0.2669	0.1740	0.2242	0.0992
$\varepsilon_{t-1}^2 \cdot \mathbf{1}\{\varepsilon_{t-1} > 0\}$	0.0052	0.0753	0.0868	0.0900
h_{t-1}^x	0.5326	0.4204	0.6644	0.1178
R_{ft-1}	2.7656	5.4877	1.4565	0.9489
Degrees of freedom and conditional skewness				
Deg. freedom			6.0645	2.0658
Constant			−0.8893	1.1077
R_{ft-1}			−0.0699	0.1633
SPR_{t-1}			1.2130	0.8075
DIV_{t-1}			0.0458	0.1936
$\mathcal{L}_X$	*−1661.4707*		*−1682.0724*	

Note to Tables 2a and 2b: These are the parameter estimates and asymptotic standard errors of the models for the small cap (Table 2a) and large cap (Table 2b) marginal distributions described in Section 4.3. $\mathcal{L}_X$ and $\mathcal{L}_Y$ refer to the value of the log-likelihood function at the optimum.

Table 2b: *Results for the large-cap marginal distribution*				
	Normal		*Skewed t*	
	Coeff	Std Error	Coeff	Std Error
Conditional mean				
Constant	0.8332	0.3859	0.8832	0.3859
AR(1)	−0.0309	0.0470	−0.0309	0.0470
R_{ft-1}	−0.4035	0.0728	−0.4035	0.0728
SPR_{t-1}	2.0567	0.4662	2.0567	0.4662
Conditional variance				
Constant	2.7672	1.5559	2.1358	1.0770
$\varepsilon_{t-1}^2 \cdot \mathbf{1}\{\varepsilon_{t-1} < 0\}$	0.2517	0.1115	0.2217	0.0729
h_{t-1}^y	0.4010	0.2158	0.5431	0.1083
R_{ft-1}	0.9354	0.5021	0.6434	0.2572
Degrees of freedom and conditional skewness				
Deg. freedom			11.2962	5.2663
Constant			−1.2442	0.4553
R_{ft-1}			−0.0198	0.0564
SPR_{t-1}			0.5003	0.4059
DIV_{t-1}			0.1766	0.1288
$\mathcal{L}_Y$	*−1514.4827*		*−1526.7341*	

Table 4: *Logistic hit tests and multinomial tests*				
	Normal		*Skewed t*	
	Small caps	Large caps	Small caps	Large caps
Region 1	7.3493	2.1389	2.1687	2.1261
p-value	*0.1959*	*0.8296*	*0.8253*	*0.8314*
Region 2	13.7482	10.6012	4.5223	8.5520
p-value	*0.0173*	*0.0599*	*0.4769*	*0.1283*
Region 3	20.0499	16.6170	8.6407	15.5125
p-value	*0.0012*	*0.0053*	*0.1243*	*0.0084*
Region 4	4.6298	2.7723	1.2006	1.5536
p-value	*0.4627*	*0.7350*	*0.9448*	*0.9068*
Region 5	4.1818	11.1156	4.7193	7.8785
p-value	*0.5236*	*0.0491*	*0.4511*	*0.1631*
Region ALL	35.1861	29.0053	16.0724	25.1755
p-value	*0.0191*	*0.0491*	*0.7121*	*0.1948*

Note to Table 4: 'Test stat' refers to the likelihood ratio statistic testing the null hypothesis that the model is correctly specified. 'P-value' refers to the area in the right tail of the distribution of the test statistic, a χ_5^2 random variable for the individual region tests and a χ_{20}^2 random variable for the joint test. The numbers 1 through 5 refer to the regions of the marginal distribution support described in the text. 'ALL' refers to the joint test of all regions simultaneously.

Table 5: *Results from the copula specification search*

Model	$\mathcal{L}_C$	Lower	Consistent	Upper
		Reality check p-values		
Normal	*212.7606*	not tested: nested in Student's t copula		
Student's t	*217.3120*	0.0690	0.1590	0.2380
Clayton	*209.3237*	0.0870	0.0870	0.1220
Rotated Clayton	*125.2558*	0.0000	0.0000	0.0000
Joe-Clayton	*216.9659*	0.1260	0.1290	0.2200
Plackett	*219.8094*	0.1830	0.2760	0.3660
Frank	*221.1679*	0.2300	0.3310	0.4510
Gumbel	*175.2964*	0.0000	0.0000	0.0000
Rotated Gumbel	*228.5322*	0.1990	0.8250	0.9620

Note to Table 5: Presented here are the nine copula specifications tried for the copula distribution model. The copula likelihood is denoted $\mathcal{L}_C$. This table presents the results of the reality check of White (2000), as modified by Hansen (2001). 'Lower', 'Consistent' and 'Upper' refer to three estimates of the p-value of the test statistic. A p-value of less than 0.10 indicates that we may reject the hypothesis that the benchmark model performs as well as the best alternative model considered according its log-likelihood value. We do not include the normal copula in the test, as it is nested in the Student's t copula. A standard LR test showed that the normal copula is rejected in favour of the t copula, with a p-value of 0.0026.

Table 6: *Copula model results*

		Coeff	Std Error	$\mathcal{L}_C$
Constant normal	$\bar{\rho}$	0.7230	0.0194	*199.6688*
	Constant	1.3013	0.6333	
	R_{ft-1}	0.1760	0.3275	
Time-varying	SPR_{t-1}	-0.6248	1.6709	*213.0969*
normal	DIV_{t-1}	0.1084	0.0989	
	μ_t^x	-0.1620	0.0700	
	μ_t^y	0.4456	0.7739	
Constant rotated Gumbel	$\bar{\delta}$	2.0914	0.0028	*216.9563*
	Constant	0.7963	0.2917	
	R_{ft-1}	-0.0804	0.1470	
Time-varying	SPR_{t-1}	0.5011	0.7390	*228.5322*
rotated Gumbel	DIV_{t-1}	0.1153	0.0473	
	μ_t^x	-0.0867	0.0322	
	μ_t^y	-0.1600	0.3410	

Note to Table 6: Here we present the copula parameters and standard errors for the two copula models considered: the normal and the rotated Gumbel copula. We present both the constant copula and the time-varying conditional copula results to show the significance of the time variation in conditional dependence over this sample. Standard LR tests yield p-values of less than 0.001 for both copulas, indicating that time variation is very significant.

Table 7: *Hit test results for the copula models*

	Normal	*Rotated Gumbel*
Test stat 1	0.5314	0.9556
p-value 1	*0.9119*	*0.8210*
Test stat 2	7.2028	1.8300
p-value 2	*0.0657*	*0.6084*
Test stat 3	1.0553	0.7386
p-value 3	*0.7879*	*0.8641*
Test stat 4	3.7632	2.6114
p-value 4	*0.2882*	*0.4555*
Test stat 5	16.4333	6.8271
p-value 5	*0.0009*	*0.0776*
Test stat 6	0.2307	1.5575
p-value 6	*0.9725*	*0.6691*
Test stat 7	2.0660	0.5909
p-value 7	*0.5588*	*0.8985*
Test stat ALL	28.8569	12.4958
p-value ALL	*0.1175*	*0.9253*

Note to Table 7: 'Test stat' refers to the likelihood ratio statistic testing the null hypothesis that the model is correctly specified. 'P-value' refers to the area in the right tail of the distribution of the test statistic, a χ^2_3 random variable for the individual region tests and a χ^2_{21} random variable for the joint test. The numbers 1 through 7 refer to the regions of the copula support described in the text. 'ALL' refers to the joint test of all regions simultaneously.

(Table 8 is in the body of the paper.)

Table 9: *Realised portfolio return summary statistics*							
	$\omega = 1$	$\omega = 0$	$\omega = 0.5$	ω^*_{uncond}	$\omega^*_{t,NORM}$	$\omega^*_{t,NORMCOP}$	$\omega^*_{t,GUMBEL}$
				RRA=1			
Mean	0.7528	0.6053	0.6791	0.7148	10.9125	13.9379	14.0138
Std Dev	6.1609	4.1222	4.7835	5.389	48.304	50.1613	49.9096
Skewness	0.0687	−0.3763	−0.3835	−0.1707	1.8756	1.7247	1.8264
5% VaR	8.7973	6.2306	7.2147	7.8653	47.375	48.252	45.3884
1% VaR	18.9576	9.6657	14.3318	16.5739	97.7722	94.585	94.4374
				RRA=3			
Mean	0.7528	0.6053	0.6791	0.6785	4.5339	6.0423	5.9977
Std Dev	6.1609	4.1222	4.7835	4.7758	19.5269	21.9183	21.3493
Skewness	0.0687	−0.3763	−0.3835	−0.386	1.8953	2.9797	2.7509
5% VaR	8.7973	6.2306	7.2147	7.2000	20.0055	19.5651	19.5629
1% VaR	18.9576	9.6657	14.3318	14.2994	37.0224	36.2156	35.6962
				RRA=7			
Mean	0.7528	0.6053	0.6791	0.6019	2.3415	3.0166	2.9998
Std Dev	6.1609	4.1222	4.7835	4.1162	10.1933	11.0499	10.795
Skewness	0.0687	−0.3763	−0.3835	−0.3561	1.3613	2.5956	2.3455
5% VaR	8.7973	6.2306	7.2147	6.1911	11.3981	11.3703	10.7019
1% VaR	18.9576	9.6657	14.3318	9.4895	24.2598	20.381	20.1106
				RRA=10			
Mean	0.7528	0.6053	0.6791	0.592	1.8617	2.3262	2.3158
Std Dev	6.1609	4.1222	4.7835	4.112	8.2497	8.711	8.5493
Skewness	0.0687	−0.3763	−0.3835	−0.2919	1.018	2.1227	1.9089
5% VaR	8.7973	6.2306	7.2147	6.305	9.4905	8.8991	8.9146
1% VaR	18.9576	9.6657	14.3318	9.5607	18.7055	16.0814	15.7535
				RRA=20			
Mean	0.7528	0.6053	0.6791	0.5667	1.2697	1.5025	1.5008
Std Dev	6.1609	4.1222	4.7835	4.1924	6.1943	6.2734	6.2166
Skewness	0.0687	−0.3763	−0.3835	−0.1188	0.3748	1.0287	0.9157
5% VaR	8.7973	6.2306	7.2147	6.3775	7.7317	7.2026	7.1422
1% VaR	18.9576	9.6657	14.3318	9.602	15.4614	13.0429	12.8697

Note to Table 9: 'RRA' refers to the coefficient of relative risk aversion. The first three portfolios are based on naïve rules, the fourth portfolio is based on a weight that is optimised once and used for all periods, the fifth portfolio is based on the normal distribution model, the sixth portfolio on the skewed t - normal copula model and the seventh portfolio is based on the skewed t - rotated Gumbel copula model.

Table 10: *Realised portfolio return summary statistics*							
	$\omega = 1$	$\omega = 0$	$\omega = 0.5$	ω^*_{uncond}	$\omega^*_{t,NORM}$	$\omega^*_{t,NORMCOP}$	$\omega^*_{t,GUMBEL}$
				RRA=1, *subject to short sales constraint*			
Mean	0.7528	0.6053	0.6791	0.7148	1.0981	1.1506	1.1606
Std Dev	6.1609	4.1222	4.7835	5.389	6.0057	5.7247	5.7202
Skewness	0.0687	−0.3763	−0.3835	−0.1707	0.0691	0.2508	0.2491
5% VaR	8.7973	6.2306	7.2147	7.8653	7.3223	7.2881	7.3223
1% VaR	18.9576	9.6657	14.3318	16.5739	18.9576	14.2657	14.2657
				RRA=3, *subject to short sales constraint*			
Mean	0.7528	0.6053	0.6791	0.6785	1.0827	1.1362	1.1417
Std Dev	6.1609	4.1222	4.7835	4.7758	5.9792	5.6888	5.6679
Skewness	0.0687	−0.3763	−0.3835	−0.386	0.0626	0.2262	0.2495
5% VaR	8.7973	6.2306	7.2147	7.2	7.3223	7.2881	7.2881
1% VaR	18.9576	9.6657	14.3318	14.2994	18.9576	14.2657	14.2657
				RRA=7, *subject to short sales constraint*			
Mean	0.7528	0.6053	0.6791	0.6053	1.0556	1.1286	1.1307
Std Dev	6.1609	4.1222	4.7835	4.1222	5.9284	5.5976	5.5773
Skewness	0.0687	−0.3763	−0.3835	−0.3763	0.0669	0.2852	0.3097
5% VaR	8.7973	6.2306	7.2147	6.2306	7.3223	7.1775	7.0758
1% VaR	18.9576	9.6657	14.3318	9.6657	17.2534	14.2657	14.1989
				RRA=10, *subject to short sales constraint*			
Mean	0.7528	0.6053	0.6791	0.6053	1.0525	1.1292	1.1338
Std Dev	6.1609	4.1222	4.7835	4.1222	5.869	5.5742	5.5574
Skewness	0.0687	−0.3763	−0.3835	−0.3763	0.1179	0.3053	0.33
5% VaR	8.7973	6.2306	7.2147	6.2306	7.3223	7.0758	7.0758
1% VaR	18.9576	9.6657	14.3318	9.6657	16.4543	13.9199	13.5453
				RRA=20, *subject to short sales constraint*			
Mean	0.7528	0.6053	0.6791	0.6053	1.0304	1.0933	1.0968
Std Dev	6.1609	4.1222	4.7835	4.1222	5.6709	5.3858	5.3806
Skewness	0.0687	−0.3763	−0.3835	−0.3763	0.1033	0.0576	0.0838
5% VaR	8.7973	6.2306	7.2147	6.2306	7.3223	7.0881	7.0657
1% VaR	18.9576	9.6657	14.3318	9.6657	15.4614	12.9283	12.7622

Note to Table 10: 'RRA' refers to the coefficient of relative risk aversion. The first three portfolios are based on naïve rules, the fourth portfolio is based on a weight that is optimised once and used for all periods, the fifth portfolio is based on the normal distribution model, the sixth portfolio on the skewed t - normal copula model and the seventh portfolio is based on the skewed t - rotated Gumbel copula model.

Table 11: *Realised portfolio return performance statistics, with bootstrap confidence intervals*

	$\omega=1$	$\omega=0$	$\omega=0.5$	ω^*_{uncond}	$\omega^*_{t,NORM}$	$\omega^*_{t,NORMCOP}$	$\omega^*_{t,GUMBEL}$
				RRA=1			
Mean/StdDev	0.1222	0.1468	0.142	0.1326	0.2259	0.2779	0.2808
	[0.06,0.18]	[0.05,0.23]	[0.07,0.2]	[0.07,0.19]	[0.17,0.29]	[0.23,0.32]	[0.24,0.32]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0909	0.2303	0.2889	0.3088
	[0.04,0.12]	[0.02,0.16]	[0.03,0.13]	[0.04,0.13]	[0.15,0.32]	[0.21,0.36]	[0.23,0.39]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0431	0.1116	0.1474	0.1484
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0,0.06]	[0.07,0.15]	[0.11,0.19]	[0.11,0.19]
				RRA=3			
Mean/StdDev	0.1222	0.1468	0.142	0.1421	0.2322	0.2757	0.2809
	[0.06,0.17]	[0.05,0.23]	[0.06,0.2]	[0.06,0.2]	[0.19,0.28]	[0.24,0.31]	[0.24,0.31]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0942	0.2266	0.3088	0.3066
	[0.04,0.12]	[0.02,0.15]	[0.03,0.13]	[0.03,0.13]	[0.16,0.3]	[0.25,0.39]	[0.24,0.37]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0475	0.1225	0.1668	0.168
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0,0.07]	[0.09,0.16]	[0.13,0.21]	[0.13,0.2]
				RRA=7			
Mean/StdDev	0.1222	0.1468	0.142	0.1462	0.2297	0.273	0.2779
	[0.06,0.18]	[0.06,0.23]	[0.07,0.21]	[0.05,0.23]	[0.19,0.27]	[0.23,0.31]	[0.23,0.31]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0972	0.2054	0.2653	0.2803
	[0.04,0.13]	[0.02,0.15]	[0.03,0.14]	[0.02,0.15]	[0.15,0.26]	[0.2,0.32]	[0.23,0.34]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0634	0.0965	0.148	0.1492
	[0,0.06]	[0.03,0.1]	[0,0.07]	[0.03,0.1]	[0.05,0.12]	[0.11,0.18]	[0.11,0.18]
				RRA=10			
Mean/StdDev	0.1222	0.1468	0.142	0.144	0.2257	0.267	0.2709
	[0.06,0.18]	[0.05,0.24]	[0.06,0.21]	[0.05,0.24]	[0.18,0.27]	[0.21,0.3]	[0.22,0.31]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0939	0.1962	0.2614	0.2598
	[0.03,0.12]	[0.02,0.16]	[0.03,0.14]	[0.02,0.16]	[0.14,0.24]	[0.21,0.32]	[0.2,0.31]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0619	0.0995	0.1447	0.147
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0.02,0.1]	[0.06,0.13]	[0.11,0.19]	[0.12,0.19]
				RRA=20			
Mean/StdDev	0.1222	0.1468	0.142	0.1352	0.205	0.2395	0.2414
	[0.06,0.17]	[0.06,0.23]	[0.07,0.2]	[0.04,0.23]	[0.14,0.25]	[0.18,0.28]	[0.18,0.28]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0889	0.1642	0.2086	0.2101
	[0.03,0.12]	[0.02,0.15]	[0.03,0.13]	[0.02,0.15]	[0.11,0.21]	[0.16,0.26]	[0.17,0.26]
Mean/1%VaR	0.0397	0.0626	0.0474	0.059	0.0821	0.1152	0.1166
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0.02,0.1]	[0.04,0.11]	[0.08,0.16]	[0.08,0.17]

Table 12: *Realised portfolio return performance statistics, with bootstrap confidence intervals*

	$\omega = 1$	$\omega = 0$	$\omega = 0.5$	ω^*_{uncond}	$\omega^*_{t,NORM}$	$\omega^*_{t,NORMCOP}$	$\omega^*_{t,GUMBEL}$
			RRA=1, *subject to short sales constraint*				
Mean/StdDev	0.1222	0.1468	0.142	0.1326	0.1828	0.201	0.2029
	[0.05,0.18]	[0.06,0.23]	[0.07,0.2]	[0.06,0.19]	[0.13,0.22]	[0.13,0.25]	[0.13,0.25]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0909	0.15	0.1579	0.1585
	[0.03,0.12]	[0.02,0.15]	[0.03,0.13]	[0.03,0.13]	[0.11,0.19]	[0.11,0.2]	[0.11,0.2]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0431	0.0579	0.0807	0.0814
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0,0.06]	[0.01,0.07]	[0.04,0.11]	[0.03,0.11]
			RRA=3, *subject to short sales constraint*				
Mean/StdDev	0.1222	0.1468	0.142	0.1421	0.1811	0.1997	0.2014
	[0.06,0.18]	[0.05,0.23]	[0.06,0.2]	[0.06,0.2]	[0.13,0.22]	[0.13,0.25]	[0.13,0.25]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0942	0.1479	0.1559	0.1566
	[0.03,0.12]	[0.02,0.15]	[0.03,0.13]	[0.03,0.13]	[0.11,0.19]	[0.1,0.2]	[0.1,0.2]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0475	0.0571	0.0796	0.08
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0,0.07]	[0.01,0.07]	[0.04,0.11]	[0.04,0.11]
			RRA=7, *subject to short sales constraint*				
Mean/StdDev	0.1222	0.1468	0.142	0.1468	0.1781	0.2016	0.2027
	[0.06,0.17]	[0.06,0.23]	[0.07,0.21]	[0.06,0.23]	[0.13,0.22]	[0.14,0.25]	[0.14,0.25]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0972	0.1442	0.1572	0.1598
	[0.03,0.12]	[0.02,0.15]	[0.03,0.14]	[0.02,0.15]	[0.11,0.18]	[0.11,0.2]	[0.11,0.2]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0626	0.0612	0.0791	0.0796
	[0,0.05]	[0.02,0.1]	[0,0.07]	[0.02,0.1]	[0.02,0.08]	[0.04,0.11]	[0.04,0.11]
			RRA=10, *subject to short sales constraint*				
Mean/StdDev	0.1222	0.1468	0.142	0.1468	0.1793	0.2026	0.204
	[0.06,0.17]	[0.06,0.23]	[0.07,0.2]	[0.06,0.23]	[0.12,0.22]	[0.14,0.25]	[0.14,0.26]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0972	0.1437	0.1596	0.1602
	[0.04,0.12]	[0.02,0.15]	[0.03,0.13]	[0.02,0.15]	[0.11,0.18]	[0.11,0.2]	[0.11,0.2]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0626	0.064	0.0811	0.0837
	[0,0.06]	[0.02,0.1]	[0,0.07]	[0.02,0.1]	[0.03,0.09]	[0.05,0.11]	[0.05,0.12]
			RRA=20, *subject to short sales constraint*				
Mean/StdDev	0.1222	0.1468	0.142	0.1468	0.1817	0.203	0.2038
	[0.06,0.17]	[0.05,0.23]	[0.07,0.2]	[0.05,0.23]	[0.12,0.23]	[0.14,0.26]	[0.14,0.26]
Mean/5%VaR	0.0856	0.0972	0.0941	0.0972	0.1407	0.1542	0.1552
	[0.03,0.12]	[0.02,0.16]	[0.03,0.13]	[0.02,0.16]	[0.1,0.18]	[0.1,0.19]	[0.1,0.19]
Mean/1%VaR	0.0397	0.0626	0.0474	0.0626	0.0666	0.0846	0.0859
	[0,0.05]	[0.02,0.1]	[0,0.07]	[0.02,0.1]	[0.03,0.09]	[0.05,0.12]	[0.05,0.12]

Notes to Tables 11 and 12: 'RRA' refers to the coefficient of relative risk aversion. Each of the six columns of figures refer to a particular portfolio: the first three portfolios are based on naïve rules, the fourth portfolio is based on a weight that is optimised once and used for all periods, the fifth portfolio is based on the normal distribution model, the sixth portfolio is based on the skewed t - normal copula model and the seventh portfolio is based on the skewed t - rotated Gumbel copula model. The bootstrap 90% confidence interval is presented in brackets below the actual figure. The details on how the confidence interval was constructed are presented in the text.

Table 13: *Pair-wise comparisons of the models' risk-adjusted performance*

	Mean / StdDev	Mean / 5% VaR	Mean / 1% VaR
		RRA=1	
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	0	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	Normal	Normal	Normal
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	Gumbel	Gumbel	Gumbel
		RRA=3	
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	Normal	Normal	Normal
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	Gumbel	Gumbel	Gumbel
		RRA=7	
Naïve vs. Uncond	0	0	Uncond
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	Normal	0
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	Gumbel	Gumbel	Gumbel
		RRA=10	
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	Normal	0
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	Gumbel	Gumbel	Gumbel
		RRA=20	
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	Normal	0
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	Gumbel	Gumbel	Gumbel

Table 14: *Pair-wise comparisons of the models' risk-adjusted performance*

	Mean / Std Dev	Mean / 5% VaR	Mean / 1% VaR
	RRA=1, *subject to short sales constraint*		
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	Normal	Normal	0
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	Normal	Normal	Normal
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	0	0	Gumbel
	RRA=3, *subject to short sales constraint*		
Naïve vs. Uncond	0	0	0
Naïve vs. Normal	Normal	Normal	0
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	Normal	Normal	0
Uncond vs. Gumbel	Gumbel	Gumbel	Gumbel
Normal vs. Gumbel	0	0	Gumbel
	RRA=7, *subject to short sales constraint*		
Naïve vs. Uncond	0	0	Uncond
Naïve vs. Normal	Normal	Normal	0
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	0	0
Uncond vs. Gumbel	Gumbel	Gumbel	0
Normal vs. Gumbel	0	0	Gumbel
	RRA=10, *subject to short sales constraint*		
Naïve vs. Uncond	0	0	Uncond
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	0	0
Uncond vs. Gumbel	Gumbel	Gumbel	0
Normal vs. Gumbel	0	0	Gumbel
	RRA=20, *subject to short sales constraint*		
Naïve vs. Uncond	0	0	Uncond
Naïve vs. Normal	Normal	Normal	Normal
Naïve vs. Gumbel	Gumbel	Gumbel	Gumbel
Uncond vs. Normal	0	0	0
Uncond vs. Gumbel	Gumbel	Gumbel	0
Normal vs. Gumbel	Gumbel	0	Gumbel

Note to Tables 13 and 14: These tables present the results of pair-wise comparisons of the 50:50 portfolio (denoted 'naïve'), the unconditionally optimal portfolio and the portfolios based on the normal distribution and skewed t - rotated Gumbel copula models. The tests were conducted at the 10% alpha level. A zero was reported if the test was inconclusive, and the name of the model was reported if that model significantly beat the other.

Table 15: *Bootstrap Reality Check p-values,* $\omega = 0.5$ *as benchmark.*

	Unconstrained			**Short sales constrained**		
	Lower	Consistent	Upper	Lower	Consistent	Upper
			RRA=1			
Mean / Std Dev	0.0070	0.0070	0.0070	0.0030	0.0030	0.0030
Mean / 5% VaR	0.0000	0.0000	0.0000	0.0020	0.0020	0.0020
Mean / 1% VaR	0.0020	0.0020	0.0020	0.0010	0.0010	0.0010
			RRA=3			
Mean / Std Dev	0.0000	0.0000	0.0000	0.0060	0.0070	0.0070
Mean / 5% VaR	0.0000	0.0000	0.0000	0.0010	0.0010	0.0010
Mean / 1% VaR	0.0000	0.0000	0.0000	0.0030	0.0030	0.0030
			RRA=7			
Mean / Std Dev	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean / 5% VaR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean / 1% VaR	0.0000	0.0000	0.0000	0.0030	0.0030	0.0030
			RRA=10			
Mean / Std Dev	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean / 5% VaR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean / 1% VaR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
			RRA=20			
Mean / Std Dev	0.0010	0.0020	0.0020	0.0010	0.0010	0.0010
Mean / 5% VaR	0.0000	0.0000	0.0000	0.0010	0.0010	0.0010
Mean / 1% VaR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note to Table 15: This table presents the results of the reality check of White (2000), as modified by Hansen (2001). 'Lower', 'Consistent' and 'Upper' refer to three estimates of the p-value of the test statistic. A p-value of less than 0.10 indicates that we may reject the hypothesis that the benchmark model performs as well as the best alternative model considered according to some performance measure.

Table 16: *Bootstrap Reality Check p-values, $\omega^*_{t,NORM}$ as benchmark.*						
	Unconstrained			**Short sales constrained**		
	Lower	Consistent	Upper	Lower	Consistent	Upper
			RRA=1			
Mean / Std Dev	0.0710	0.0710	0.2650	0.1930	0.3420	0.3740
Mean / 5% VaR	0.0160	0.0160	0.2220	0.3350	0.3350	0.6390
Mean / 1% VaR	0.0120	0.0120	0.1880	0.0670	0.0670	0.0670
			RRA=3			
Mean / Std Dev	0.0390	0.0390	0.2670	0.1930	0.3620	0.3930
Mean / 5% VaR	0.0140	0.0140	0.1690	0.3730	0.3730	0.6270
Mean / 1% VaR	0.0150	0.0150	0.1780	0.0710	0.0710	0.0710
			RRA=7			
Mean / Std Dev	0.0060	0.0060	0.2200	0.1170	0.2630	0.2820
Mean / 5% VaR	0.0000	0.0000	0.1150	0.1630	0.3700	0.4400
Mean / 1% VaR	0.0000	0.0000	0.0110	0.1190	0.1300	0.1300
			RRA=10			
Mean / Std Dev	0.0020	0.0020	0.2200	0.1070	0.2820	0.2960
Mean / 5% VaR	0.0000	0.0000	0.1250	0.1630	0.3670	0.4340
Mean / 1% VaR	0.0000	0.0000	0.0110	0.1060	0.1240	0.1280
			RRA=20			
Mean / Std Dev	0.0190	0.2460	0.2460	0.0580	0.2520	0.2730
Mean / 5% VaR	0.0020	0.0020	0.1640	0.1080	0.3380	0.4060
Mean / 1% VaR	0.0200	0.1040	0.1040	0.0840	0.1100	0.1150

Note to Table 16: This table presents the results of the reality check of White (2000), as modified by Hansen (2001). 'Lower', 'Consistent' and 'Upper' refer to three estimates of the p-value of the test statistic. A p-value of less than 0.10 indicates that we may reject the hypothesis that the benchmark model performs as well as the best alternative model considered according to some performance measure.

Table 17: *Pair-wise comparisons involving the Normal Copula model*			
	Mean / Std Dev	Mean / 5% VaR	Mean / 1% VaR
		RRA=1	
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	NormCop	0	NormCop
Gumbel vs. NormCop	0	Gumbel	0
		RRA=3	
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	NormCop	NormCop	NormCop
Gumbel vs. NormCop	Gumbel	0	0
		RRA=7	
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	NormCop	NormCop	NormCop
Gumbel vs. NormCop	Gumbel	Gumbel	0
		RRA=10	
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	NormCop	NormCop	NormCop
Gumbel vs. NormCop	Gumbel	0	0
		RRA=20	
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	NormCop	NormCop	NormCop
Gumbel vs. NormCop	Gumbel	0	Gumbel

Note to Table 17: This table presents the results of pair-wise comparisons of the 50:50 portfolio (denoted 'naïve'), the unconditionally optimal portfolio and the portfolios based on the normal distribution, the skewed t - rotated Gumbel copula and the skewed t - normal copula models. The tests were conducted at the 10% alpha level. A zero was reported if the test was inconclusive, and the name of the model was reported if that model significantly beat the other.

Table 18: *Pair-wise comparisons involving the Normal Copula model*

	Mean / Std Dev	Mean / 5% VaR	Mean / 1% VaR
	RRA=1, *subject to short sales constraint*		
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	0	0	NormCop
Gumbel vs. NormCop	0	0	0
	RRA=3, *subject to short sales constraint*		
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	NormCop
Normal vs. NormCop	0	0	NormCop
Gumbel vs. NormCop	0	0	0
	RRA=7, *subject to short sales constraint*		
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	0
Normal vs. NormCop	0	0	NormCop
Gumbel vs. NormCop	0	Gumbel	0
	RRA=10, *subject to short sales constraint*		
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	0
Normal vs. NormCop	0	0	NormCop
Gumbel vs. NormCop	0	0	Gumbel
	RRA=20, *subject to short sales constraint*		
Naïve vs. NormCop	NormCop	NormCop	NormCop
Uncond vs. NormCop	NormCop	NormCop	0
Normal vs. NormCop	NormCop	0	NormCop
Gumbel vs. NormCop	0	0	Gumbel

Note to Table 18: This table presents the results of pair-wise comparisons of the 50:50 portfolio (denoted 'naïve'), the unconditionally optimal portfolio and the portfolios based on the normal distribution, the skewed t - rotated Gumbel copula and the skewed t - normal copula models. The tests were conducted at the 10% alpha level. A zero was reported if the test was inconclusive, and the name of the model was reported if that model significantly beat the other.

Table 19: *Bootstrap Reality Check p-values, with $\omega^*_{t,NORMCOP}$ as benchmark*						
	Unconstrained			**Short sales constrained**		
	Lower	Consistent	Upper	Lower	Consistent	Upper
			RRA=1			
Mean / Std Dev	0.1120	0.1120	0.7980	0.2360	0.8470	0.8860
Mean / 5% VaR	0.0140	0.1880	0.5330	0.5200	0.6860	0.9310
Mean / 1% VaR	0.5620	0.5620	0.8810	0.4660	0.8900	0.9080
			RRA=3			
Mean / Std Dev	0.0080	0.0080	0.7150	0.2000	0.8100	0.8650
Mean / 5% VaR	0.7720	0.8300	0.9670	0.5200	0.6410	0.9060
Mean / 1% VaR	0.3480	0.3480	0.7410	0.4870	0.6700	0.9340
			RRA=7			
Mean / Std Dev	0.0030	0.0030	0.6990	0.1900	0.8740	0.9230
Mean / 5% VaR	0.0040	0.0040	0.4970	0.0440	0.3230	0.8450
Mean / 1% VaR	0.4950	0.4950	0.7800	0.4440	0.8500	0.8870
			RRA=10			
Mean / Std Dev	0.0000	0.0000	0.7200	0.0880	0.8560	0.9110
Mean / 5% VaR	0.6150	0.7700	0.9500	0.3120	0.6480	0.9130
Mean / 1% VaR	0.0760	0.0760	0.8020	0.1260	0.7290	0.8000
			RRA=20			
Mean / Std Dev	0.0210	0.3710	0.7610	0.1140	0.8580	0.8990
Mean / 5% VaR	0.1930	0.1930	0.8560	0.1060	0.3940	0.7990
Mean / 1% VaR	0.0970	0.0970	0.8460	0.1360	0.8620	0.9090

Note to Table 19: This table presents the results of the reality check of White (2000), and modified by Hansen (2001). 'Lower', 'Consistent' and 'Upper' refer to three estimates of the p-value of the test statistic. A p-value of less than 0.10 indicates that we may reject the hypothesis that the benchmark model performs as well as the best alternative model considered according to some performance measure.

References

[1] Aït-Sahalia, Yacine, and Brandt, Michael W., 2001, Variable Selection for Portfolio Choice, *Journal of Finance*, 56(4), 1297-1355.

[2] Ang, Andrew, and Bekaert, Geert, 2001, International Asset Allocation with Regime Shifts, forthcoming, *Review of Financial Studies.*

[3] Ang, Andrew, and Chen, Joe, 2001, Asymmetric Correlations of Equity Portfolios, forthcoming, *Journal of Financial Economics.*

[4] Arrow, Kenneth J., 1971, *Essays in the Theory of Risk Bearing*, Markham Publishing Co., Chicago.

[5] Avérous, Jean and Meste, Michel, 1997, Skewness for Multivariate Distributions: Two Approaches, *Annals of Statistics*, 25(5), 1984-1997.

[6] Bae, Kee-Hong, Karolyi, Andrew and Stulz, René M., 2000, A New Approach to Measuring Financial Contagion, Working Paper 2000-13, Fisher College of Business, Ohio State University.

[7] Brandt, Michael W., 1999, Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach, *Journal of Finance*, 54(5), 1609-1645.

[8] Campbell, John Y., and Viceira, Luis M., 1999, Consumption and Portfolio Decisions when Expected Returns are Time Varying, *Quarterly Journal of Economics*, 114, 433-495.

[9] Campbell, Rachel, Koedijk, Kees, and Kofman, Paul, 2000, Increased Correlation in Bear Markets: A Downside Risk Perspective, Working Paper, Faculty of Business Administration, Erasmus University Rotterdam.

[10] Casella, George, and Berger, Roger L., 1990, *Statistical Inference*, Duxbury Press, U.S.A.

[11] Chen, Yi-Ting, 2001, Testing Conditional Symmetry with an Application to Stock Returns, working paper, Institute for Social Science and Philosophy, Academia Sinica.

[12] Clayton, D. G., 1978, A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, 65(1), 141-151.

[13] Cook, R. Dennis, and Johnson, Mark E., 1981, A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data, *Journal of the Royal Statistical Society, Series B,* 43(2), 210-218.

[14] Diebold, Francis X., and Mariano, Roberto S., 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13(3), 253-263.

[15] Diebold, Francis X., Gunther T., and Tay, Anthony S., 1998, Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review*, 39, 863-883.

[16] Efron, B., and Tibshirani, R. J., 1993, *An Introduction to the Bootstrap*, Chapman and Hall, New York.

[17] Embrechts, Paul, Höing, Andrea and Juri, Alessandro, 2001, Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks, working paper, Department of Mathematics, ETHZ, Zurich, Switzerland.

[18] Erb, Claude B., Harvey, Campbell R., and Viskanta, Tadas E., 1994, Forecasting International Equity Correlations, *Financial Analysts Journal*, 50, 32-45.

[19] Fama, Eugene F., 1981, Stock Returns, Real Activity, Inflation, and Money, *American Economic Review*, 71, 545-565.

[20] Fine, J. P., and Jiang, H., 2000, On Association in a Copula with Time Transformations, *Biometrika*, 87(3), 559-571.

[21] Fischer, R. A., 1932, *Statistical Methods for Research Workers.*

[22] Friend, Irwin, and Westerfield, Randolph, 1980, Co-Skewness and Capital Asset Pricing, *Journal of Finance*, 35(4), 897-913.

[23] Genest, C., and Rivest, L.-P., 1993, Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, 88(423), 1034-1043.

[24] Hansen, Bruce E., 1994, Autoregressive Conditional Density Estimation, *International Economic Review*, 35(3), 705-730.

[25] Hansen, Peter R., 2001, An Unbiased and Powerful Test for Superior Predictive Ability, working paper, Department of Economics, Brown University.

[26] Harvey, Campbell R., and Siddique, Akhtar, 1999, Autoregressive Conditional Skewness, *Journal of Financial and Quantitative Analysis*, 34(4), 465-488.

[27] Harvey, Campbell R., and Siddique, Akhtar, 2000, Conditional Skewness in Asset Pricing Tests, *Journal of Finance*, 55(3), 1263-1295.

[28] Huang, Chi-fu, and Litzenberger, Robert H., 1988, *Foundations for Financial Economics*, Prentice-Hall Inc., New Jersey.

[29] Ingersoll, Jonathon E., Jr., 1987, *Theory of Financial Decision Making*, Rowman and Littlefield Publishers Inc., Maryland.

[30] Joe, Harry, 1997, *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability 73, Chapman and Hall, London, U.K.

[31] Judd, Kenneth L., 1998, *Numerical Methods in Economics*, MIT Press, Cambridge, Massachusetts.

[32] Kandel, Shmuel, and Stambaugh, Robert F., 1996, On the Predictability of Stock Returns: An Asset Allocation Perspective, *Journal of Finance*, 51(2), 385-424.

[33] Kraus, Alan, and Litzenberger, Robert H., 1976, Skewness Preference and the Valuation of Risk Assets, *Journal of Finance*, 31(4), 1085-1100.

[34] Li, David X., 2001, On Default Correlation: A Copula Function Approach, Working Paper 99-07, The RiskMetrics Group.

[35] Lim, Kian-Guan, 1989, A New Test of the Three-Moment Capital Asset Pricing Model, *Journal of Financial and Quantitative Analysis*, 24(2), 205-216.

[36] Longin, François, and Solnik, Bruno, 2001, Extreme Correlation of International Equity Markets, *Journal of Finance*, 56(2), 649-676.

[37] Mardia, K. V., 1970, Measures of Multivariate Skewness and Kurtosis, with Applications, *Biometrika*, 57(3), 519-530.

[38] Merton, Robert C., 1971, Optimal Consumption and Portfolio Rules in a Continuous-Time Model, *Journal of Economic Theory*, 3, 373-413.

[39] Nelsen, Roger B., 1993, Some Concepts of Bivariate Symmetry, *Nonparametric Statistics*, 3, 95-101.

[40] Nelsen, Roger B., 1999, *An Introduction to Copulas*, Springer-Verlag, New York.

[41] Oakes, David, 1989, Bivariate Survival Models Induced by Frailties, *Journal of the American Statistical Association*, 84(406), 487-493.

[42] Patton, Andrew J., 2001a, Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula, Working Paper 2001-09, Department of Economics, University of California, San Diego.

[43] Patton, Andrew J., 2001b, Estimation of Copula Models for Time Series of Possibly Different Lengths, working paper, Department of Economics, University of California, San Diego.

[44] Patton, Andrew J., 2001c, Linear Dependence: Defined and Tested, work-in-progress.

[45] Perez-Quiros, Gabriel, and Timmermann, Allan, 2001, Business Cycle Asymmetries in Stock Returns: Evidence from Higher Order Moments and Conditional Densities, *Journal of Econometrics*, 103, 259-306.

[46] Politis, Dimitris N., and Romano, Joseph P., 1994, The Stationary Bootstrap, *Journal of the American Statistical Association*, 89, 1303-1313.

[47] Richardson, Matthew and Smith, Tom, 1993, A Test for Multivariate Normality in Stock Returns, *Journal of Business*, 66(2), 295-321.

[48] Rivers, Douglas, Vuong, Quang, 1999, Model Selection Tests for Nonlinear Dynamic Models, working paper, Department of Economics, University of Southern California.

[49] Rosenberg, Joshua V., 1999, Semiparametric Pricing of Multivariate Contingent Claims, working paper, Department of Finance, Stern School of Business, New York University.

[50] Rosenberg, Joshua V., 2000, Nonparametric Pricing of Multivariate Contingent Claims, working paper, Department of Finance, Stern School of Business, New York University.

[51] Scaillet, O., 2000, Nonparametric Estimation of Copulas for Time Series, working paper, Département des Sciences Economiques, Université Catholique de Louvain, Belgium.

[52] Schweizer, B., and Sklar, A., 1983, *Probabilistic Metric Spaces*, Elsevier Science, New York.

[53] Singleton, J. Clay, and Wingender, John, 1986, Skewness Persistence in Stock Returns, *Journal of Financial and Quantitative Analysis*, 21(3), 335-341.

[54] Sklar, A., 1959, Fonctions de répartition à n dimensionset leurs marges, *Publ. Inst. Statis. Univ. Paris*, 8, 229-231.

[55] White, Halbert, 2000, A Reality Check for Data Snooping, *Econometrica*, 68(5), 1097-1126.

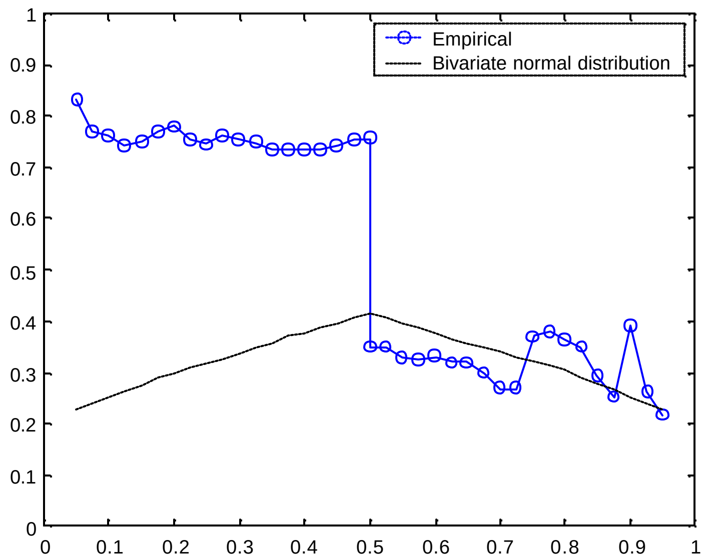

Figure 1: *Exceedence correlations between excess returns (X and Y) on small caps and large caps. The horizontal axis shows the cut-off quantile, and the vertical axis shows the correlation between the two assets given that both exceed that quantile.*

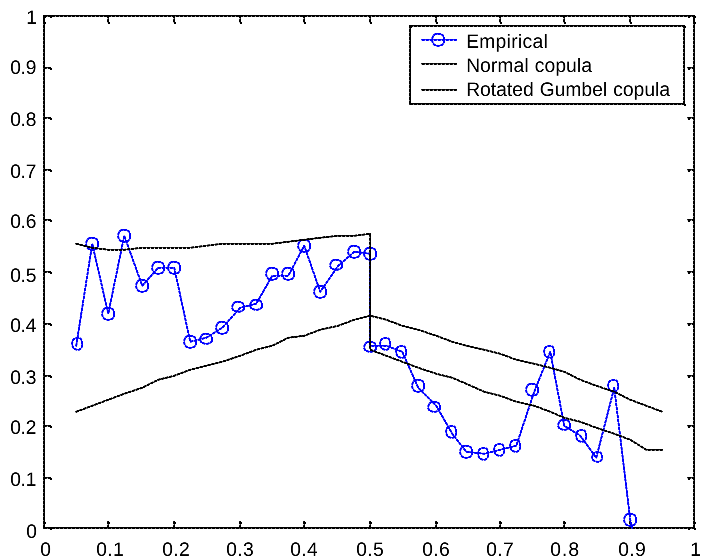

Figure 2: *Exceedence correlations between transformed residuals (U and V) of small caps and large caps. The horizontal axis shows the cut-off quantile, and the vertical axis shows the correlation between the two assets given that both assets exceed that quantile.*

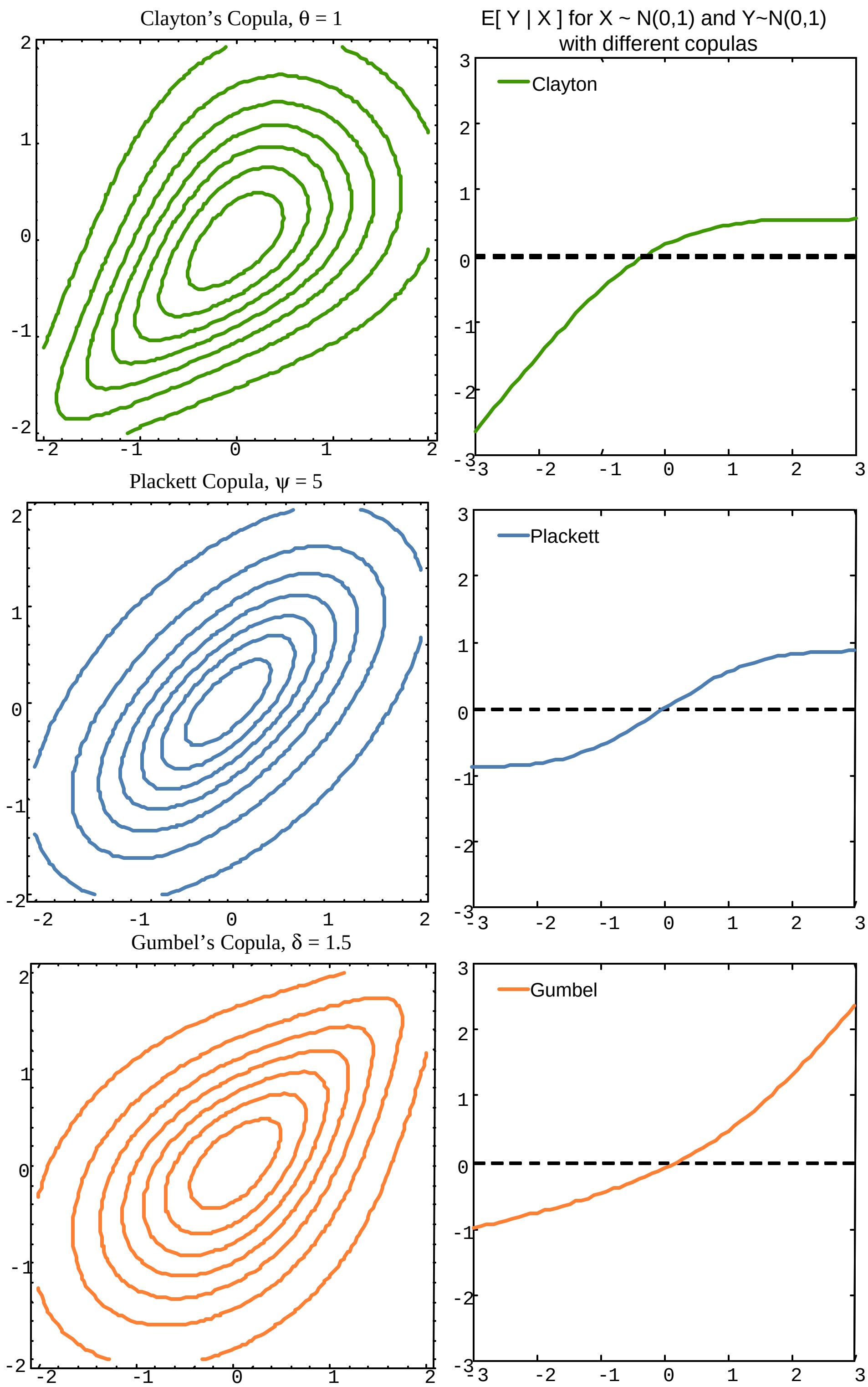

Figure 3: *Aspects of various bivariate distributions, all with standard* N(0,1) *margins and linear correlation coefficients of 0.5. The left three panels present density contour plots, and the right three panels present the regression functions of Y given X.*

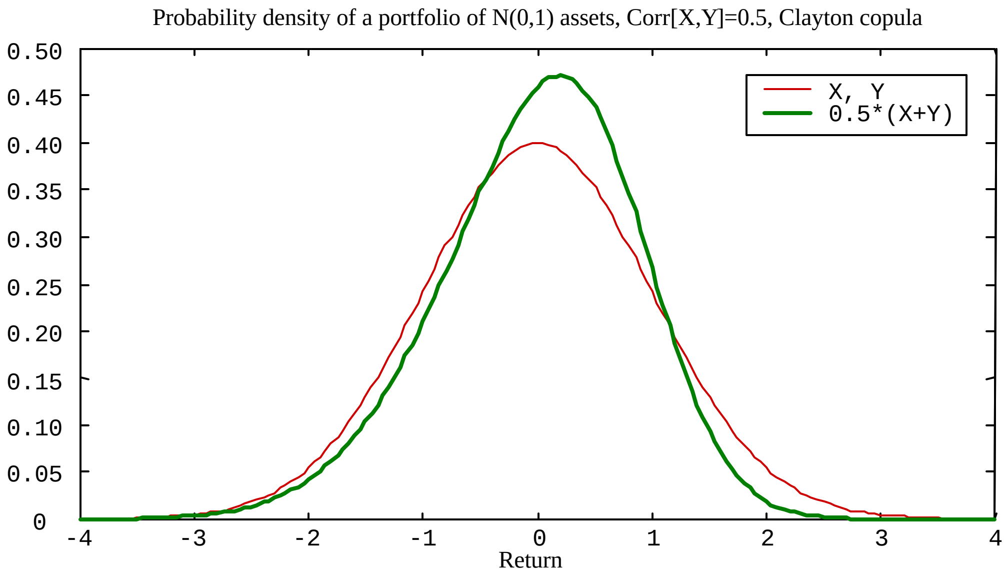

Figure 4: *The distribution of portfolio returns* $Z = \frac{1}{2}X + \frac{1}{2}Y$*, when* $X \sim N(0,1)$*,* $Y \sim N(0,1)$ *and correlation is 0.5, using the Clayton copula.*

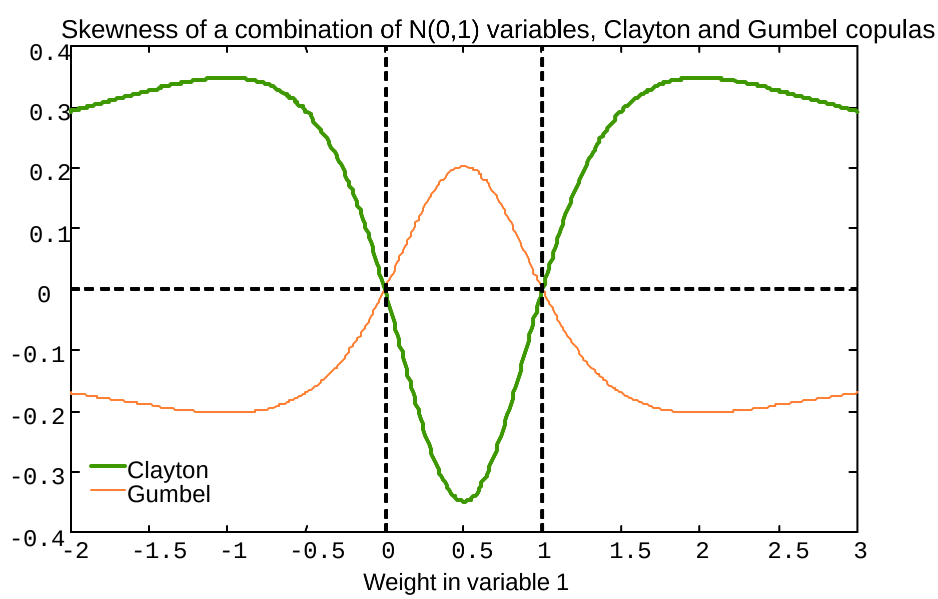

Figure 5: *The skewness of portfolios of N(0,1) assets with linear correlation of 0.5, but with different copulas.*

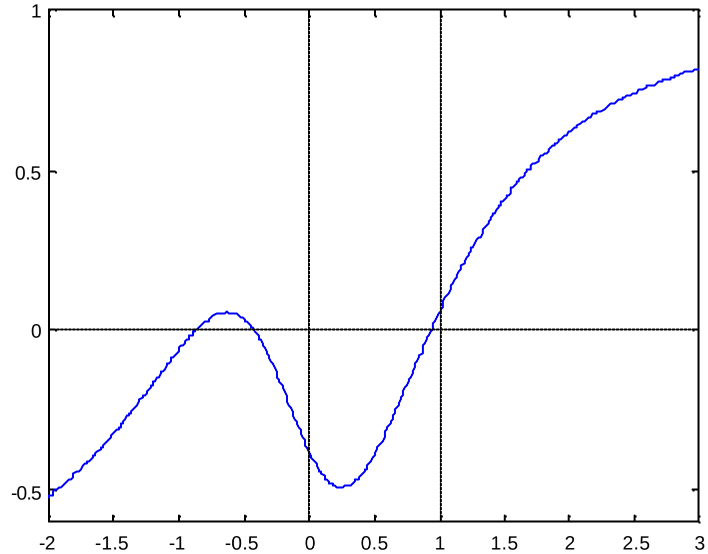

Figure 6: *Skewness of portfolios of the small cap and large cap indices. The horizontal axis shows the weight in the small cap index, and the vertical axis the skewness of the corresponding portfolio.*

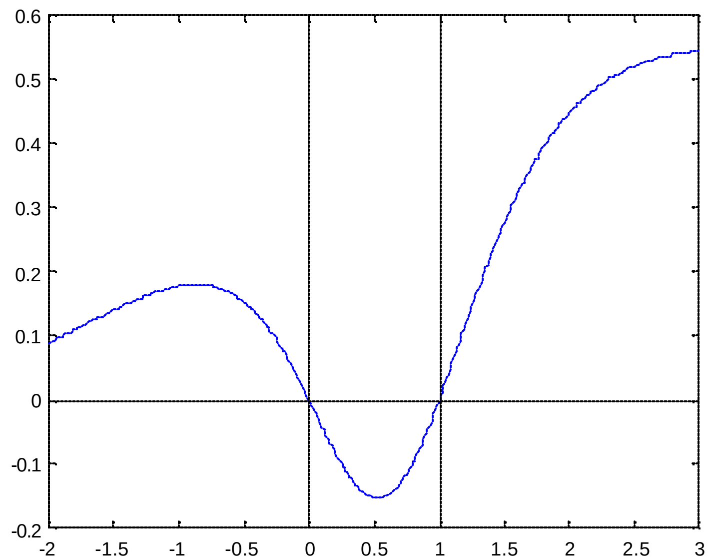

Figure 7: *Skewness of portfolios of the* normalised *small cap and large cap indices. The horizontal axis shows the weight in the normalised small cap index, and the vertical axis the skewness of the corresponding portfolio.*

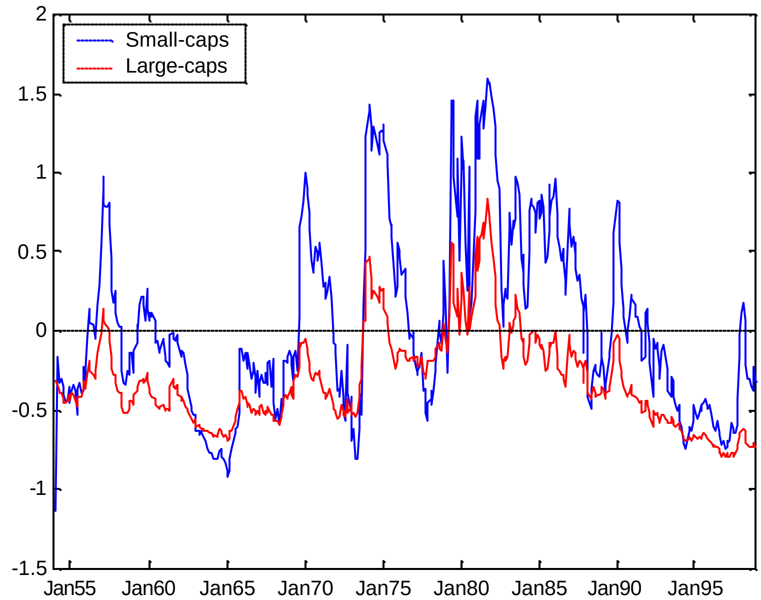

Figure 8: *Conditional skewness of small cap and large cap returns, estimated using the Skewed t marginal distribution model.*

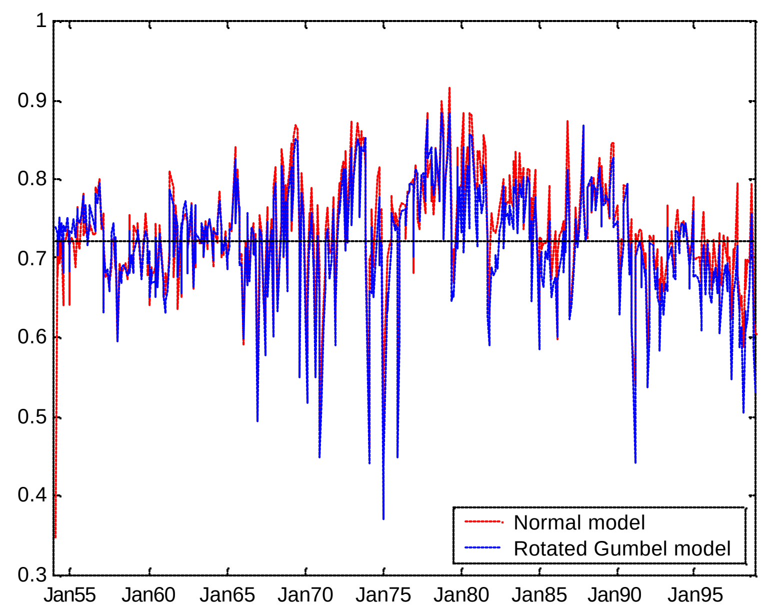

Figure 9: *Conditional correlation estimates from the two distribution models.*