

High-Dimensional Copula-Based Distributions with Mixed Frequency Data*

Dong Hwan Oh[†] Andrew J. Patton[‡]
Federal Reserve Board Duke University

This version: January 23, 2015.

Abstract

This paper proposes a new general model for high-dimensional distributions of asset returns that utilizes mixed frequency data and copulas. The dependence between returns is decomposed into linear and nonlinear components, which enables the use of high frequency data to accurately measure and forecast linear dependence, and the use of a new class of copulas designed to capture nonlinear dependence among the resulting linearly uncorrelated, low frequency, residuals. Estimation of the new class of copulas is conducted using composite likelihood, making this approach feasible even for hundreds of variables. A realistic simulation study verifies that multistage estimation with composite likelihood results in small loss in efficiency and large gain in computation speed. In- and out-of-sample tests confirm the statistical superiority of the proposed models applied to daily returns on all constituents of the S&P 100 index.

Keywords: high frequency data, forecasting, composite likelihood, nonlinear dependence.

J.E.L. codes: C32, C51, C58.

*We thank Tim Bollerslev, Federico Bugni, Jia Li, Oliver Linton, Bruno Rémillard, Enrique Sentana, Neil Shephard, and George Tauchen as well as seminar participants at the Federal Reserve Board, Rutgers University, SUNY-Stony Brook, Toulouse School of Economics, University of Cambridge, and University of Montreal for their insightful comments. We also benefitted from data mainly constructed by Sophia Li and Ben Zhao. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Board.

[†]Quantitative Risk Analysis Section, Federal Reserve Board, Washington DC 20551. Email: donghwan.oh@frb.gov

[‡]Department of Economics, Duke University, Box 90097, Durham NC 27708. Email: andrew.patton@duke.edu

1 Introduction

A model for the multivariate distribution of the returns on large collections of financial assets is a crucial component in modern risk management and asset allocation. Modelling high-dimensional distributions, however, is not an easy task and only a few models are typically used in high dimensions, most notably the Normal distribution, which is still widely used in practice and academia despite its notorious limits, for example, thin tails and zero tail dependence.

This paper provides a new approach for constructing and estimating high-dimensional distribution models. Our approach builds on two active areas of recent research in financial econometrics. First, high frequency data has been shown to be superior to daily data for measuring and forecasting variances and covariances, see Andersen, *et al.* (2006) for a survey of this very active area of research. This implies that there are gains to be had by modelling linear dependence, as captured by covariances, using high frequency data. Second, copula methods have been shown to be useful for constructing flexible distribution models in high dimensions, see Christoffersen, *et al.* (2013), Oh and Patton (2013) and Creal and Tsay (2014). These two findings naturally lead to the question of whether high frequency data and copula methods can be combined to improve the modelling and forecasting of high-dimensional return distributions.

Exploiting high frequency data in a lower frequency copula-based model is not straightforward as, unlike variances and covariances, the copula of low frequency (say daily) returns is not generally a known function of the copula of high frequency returns. Thus the link between high frequency volatility measures (e.g., realized variance and covariance) and their low frequency counterparts cannot generally be exploited when considering dependence via the copula function. We overcome this hurdle by decomposing the dependence structure of low frequency asset returns into linear and nonlinear components. We then use high frequency data to accurately model the linear dependence, as measured by covariances, and a new class of copulas to capture the remaining dependence in the low frequency standardized residuals.

The difficulty in specifying a copula-based model for standardized, uncorrelated, residuals, is that the distribution of the residuals must imply an identity correlation matrix. Independence is only sufficient for uncorrelatedness, and we wish to allow for possible nonlinear dependence between these linearly unrelated variables. Among existing work, only the multivariate Student's t distribution has been used for this purpose, as an identity correlation matrix can be directly

imposed on this distribution. We dramatically increase the set of possible models for uncorrelated residuals by proposing methods for generating “jointly symmetric” copulas. These copulas can be constructed from any given (possibly asymmetric) copula, and when combined with any collection of (possibly heterogeneous) symmetric marginal distributions they guarantee an identity correlation matrix. Evaluation of the density of our jointly symmetric copulas turns out to be computationally difficult in high dimensions, but we show that composite likelihood methods (see Varin, *et al.* 2011 for a review) may be used to estimate the model parameters and undertake model selection tests.

This paper makes four main contributions. Firstly, we propose a new class of “jointly symmetric” copulas, which are useful in multivariate density models that contain a covariance matrix model (e.g., GARCH-DCC, HAR, stochastic volatility, etc.) as a component. Second, we show that composite likelihood methods may be used to estimate the parameters of these new copulas, and in an extensive simulation study we verify that these methods have good finite-sample properties. Third, we propose a new and simple model for high-dimensional covariance matrices drawing on ideas from the HAR model of Corsi (2009) and the DCC model of Engle (2002), and we show that this model outperforms the familiar DCC model empirically. Finally, we present a detailed empirical application of our model to 104 individual U.S. equity returns, showing that our proposed approach significantly outperforms existing approaches both in-sample and out-of-sample.

Our methods and application are related to several existing papers. Most closely related is the work of Lee and Long (2009), who also consider the decomposition into linear and nonlinear dependence, and use copula-based models for the nonlinear component. However, Lee and Long (2009) focus only on bivariate applications, and their approach, which we describe in more detail in Section 2, is computationally infeasible in high dimensions. Our methods are also clearly related to copula-based density models, some examples of which are cited above, however in those approaches only the variances are modelled prior to the copula stage, meaning that the copula model must capture both the linear and nonlinear components of dependence. This makes it difficult to incorporate high frequency data into the dependence model. Papers that employ models for the joint distribution of returns that include a covariance modelling step include Chiriac and Voev (2011), Jondeau and Rockinger (2012), Hautsch, *et al.* (2013), and Jin and Maheu (2013). As models for the standardized residuals, those papers use the Normal or Student’s t distributions, both of which are nested in our class of jointly symmetric models, and which we show are significantly beaten in our application to U.S. equity returns.

The paper is organized as follows. Section 2 presents our approach for modelling high-dimensional distributions. Section 3 presents multi-stage, composite likelihood methods for model estimation and comparison, which are studied via simulations in Section 4. Section 5 applies our model to daily equity returns and compares it with existing approaches. Section 6 concludes. An appendix contains all proofs, and a web appendix contains additional details, tables and figures.

2 Models of linear and nonlinear dependence

We construct a model for the conditional distribution of the N -vector \mathbf{r}_t as follows:

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{H}_t^{1/2} \mathbf{e}_t \tag{1}$$

$$\text{where } \mathbf{e}_t \sim iid \mathbf{F}(\cdot; \boldsymbol{\eta}) \tag{2}$$

where $\mathbf{F}(\cdot; \boldsymbol{\eta})$ is a joint distribution with zero mean, identity covariance matrix and “shape” parameter $\boldsymbol{\eta}$, and $\boldsymbol{\mu}_t = E[\mathbf{r}_t | \mathcal{F}_{t-1}]$, $\mathbf{H}_t = V[\mathbf{r}_t | \mathcal{F}_{t-1}]$, $\mathcal{F}_t = \sigma(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots)$, and \mathbf{Y}_t includes \mathbf{r}_t and possibly other time t observables, such as realized variances and covariances. To obtain $\mathbf{H}_t^{1/2}$, we suggest using the spectral decomposition due to its invariance to the order of the variables. Note that by assuming that \mathbf{e}_t is *iid*, we impose that all dynamics in the conditional joint distribution of \mathbf{r}_t are driven by the conditional mean and (co)variance. This common, and clearly strong, assumption goes some way towards addressing the curse of dimensionality faced when N is large.

In existing approaches, see Chiriac and Voev (2011), Jondeau and Rockinger (2012), Hautsch, *et al.* (2013), and Jin and Maheu (2013) for example, \mathbf{F} would be assumed multivariate Normal (which reduces to independence, given that \mathbf{e}_t has identity covariance matrix) or Student’s t , and the model would be complete. Instead, we consider the decomposition of the joint distribution \mathbf{F} into marginal distributions F_i and copula \mathbf{C} using Sklar’s (1959) theorem:

$$\mathbf{e}_t \sim \mathbf{F}(\cdot; \boldsymbol{\eta}) = \mathbf{C}(F_1(\cdot; \boldsymbol{\eta}), \dots, F_N(\cdot; \boldsymbol{\eta}); \boldsymbol{\eta}) \tag{3}$$

Note that the elements of \mathbf{e}_t are uncorrelated but may still exhibit cross-sectional dependence, which is completely captured by the copula \mathbf{C} . Combining equations (1)–(3) we obtain the following

density for the distribution of returns:

$$\mathbf{f}_t(\mathbf{r}_t) = \det(\mathbf{H}_t^{-1/2}) \times \mathbf{c}(F_1(e_{1t}), \dots, F_N(e_{Nt})) \times \prod_{i=1}^N f_i(e_{it}) \quad (4)$$

Thus this approach naturally reveals two kinds of dependence between returns: “linear dependence,” captured by conditional covariance matrix \mathbf{H}_t , and any “nonlinear dependence” remaining in the uncorrelated residuals \mathbf{e}_t , captured by the copula \mathbf{C} . There are two important advantages in decomposing a joint distribution of returns in this way. First, it allows the researcher to draw on the large literature on measuring, modeling and forecasting conditional covariance matrix \mathbf{H}_t with low and high frequency data. For example, GARCH-type models such as the multivariate GARCH model of Bollerslev, *et al.* (1988), the BEKK model of Engle and Kroner (1995), and the dynamic conditional correlation (DCC) model of Engle (2002) naturally fit in equations (1) and (2). The increasing availability of high frequency data also enables us to use more accurate models for conditional covariance matrix, see, for example, Bauer and Vorkink (2011), Chiriac and Voev (2011), and Noureldin, *et al.* (2012), and those models are also naturally accommodated by equations (1)–(2).¹ Second, the model specified by equations (1)–(3) is easily extended to high-dimensional applications given that multi-stage separate estimation of the conditional mean of the returns, the conditional covariance matrix of the returns, the marginal distributions of the standardized residuals, and finally the copula of the standardized residuals is possible. Of course, multi-stage estimation is less efficient than one-stage estimation, however the main difficulty in high-dimensional applications is the proliferation of parameters and the growing computational burden as the dimension increases. By allowing for multi-stage estimation we overcome this obstacle.

Lee and Long (2009) were the first to propose decomposing dependence into linear and nonlinear components, and we now discuss their approach in more detail. They proposed the following model:

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \mathbf{H}_t^{1/2} \boldsymbol{\Sigma}^{-1/2} \mathbf{w}_t \quad (5)$$

where $\mathbf{w}_t \sim iid \mathbf{G}(\cdot; \boldsymbol{\eta}) = \mathbf{C}_w(G_1(\cdot; \boldsymbol{\eta}), \dots, G_N(\cdot; \boldsymbol{\eta}); \boldsymbol{\eta})$

and $\boldsymbol{\Sigma} \equiv Cov[\mathbf{w}_t]$

¹In the part of our empirical work that uses realized covariance matrices, we take these as given, and do not take a stand on the specific continuous-time process that generates the returns and realized covariances. This means that, unlike a DCC-type model, for example, which only considers daily returns, or a case where the continuous-time process was specified, we cannot simulate or generate multi-step ahead predictions from these models.

Thus rather than directly modelling uncorrelated residuals \mathbf{e}_t as we do, Lee and Long (2009) use \mathbf{w}_t and its covariance matrix $\mathbf{\Sigma}$ to obtain uncorrelated residuals $\mathbf{e}_t = \mathbf{\Sigma}^{-1/2}\mathbf{w}_t$. In this model it is generally hard to interpret \mathbf{w}_t , and thus to motivate or explain choices of models for its marginal distribution or copula. Most importantly, this approach has two aspects that make it unamenable to high-dimensional applications. Firstly, the structure of this model is such that multi-stage estimation of the joint distribution of the standardized residuals is not possible, as these residuals are linear combinations of the latent variables \mathbf{w}_t . Thus the entire N -dimensional distribution \mathbf{G} must be estimated in a single step. Lee and Long (2009) focus on bivariate applications where this is not difficult, but in applications involving 10, 50 or 100 variables this quickly becomes infeasible. Secondly, the matrix $\mathbf{\Sigma}$ implied by $\mathbf{G}(\cdot; \boldsymbol{\eta})$ can usually only be obtained by numerical methods, and as this matrix grows quadratically with N , this is a computational burden even for relatively low dimension problems. In contrast, we directly model the standardized uncorrelated residuals \mathbf{e}_t to take advantage of benefits from multi-stage separation and to avoid having to compute $\mathbf{\Sigma}$. In addition to proposing methods that work in high dimensions, our approach extends that of Lee and Long (2009) to exploit recent developments in the use of high frequency data to estimate lower frequency covariance matrices.

We next describe how we propose modelling the uncorrelated residuals, \mathbf{e}_t , and then we turn to models for the covariance matrix \mathbf{H}_t .

2.1 A density model for uncorrelated standardized residuals

A building block for our model is an N -dimensional distribution, $\mathbf{F}(\cdot; \boldsymbol{\eta})$, that guarantees an identity correlation matrix. The concern is that there are only a few copulas that ensure zero correlations, for example, the Gaussian copula with identity correlation matrix (i.e. the independence copula) and the t copula with identity correlation matrix, when combined with symmetric marginals. To overcome this lack of choice, we now propose methods to generate many copulas that ensure zero correlations by constructing “jointly symmetric” copulas.

We exploit the result that multivariate distributions that satisfy a particular type of symmetry condition are guaranteed to yield an identity correlation matrix, which is required by the model specified in equations (1)-(2). Recall that a scalar random variable X is symmetric about the point a if the distribution functions of $(X - a)$ and $(a - X)$ are the same. For vector random variables there are two main types of symmetry: in the bivariate case, the first (“radial symmetry”) requires

only that $(X_1 - a_1, X_2 - a_2)$ and $(a_1 - X_1, a_2 - X_2)$ have a common distribution function, while the second (“joint symmetry”) further requires that $(X_1 - a_1, a_2 - X_2)$ and $(a_1 - X_1, X_2 - a_2)$ also have that common distribution function. The latter type of symmetry is what we require for our model. The definition below for the N -variate case is adapted from Nelsen (2006).

Definition 1 (Joint symmetry) *Let \mathbf{X} be a vector of N random variables and let \mathbf{a} be a point in \mathbb{R}^N . Then \mathbf{X} is jointly symmetric about \mathbf{a} if the following 2^N sets of N random variables have a common joint distribution:*

$$\tilde{\mathbf{X}}^{(i)} = \left[\tilde{X}_1^{(i)}, \dots, \tilde{X}_N^{(i)} \right]', \quad i = 1, 2, \dots, 2^N$$

where $\tilde{X}_j^{(i)} = (X_j - a_j)$ or $(a_j - X_j)$, for $j = 1, 2, \dots, N$.

From the following simple lemma we know that all jointly symmetric distributions guarantee an identity correlation matrix, and are thus easily used in a joint distribution model with a covariance modelling step.

Lemma 1 *Let \mathbf{X} be a vector of N jointly symmetric random variables with finite second moments. Then \mathbf{X} has an identity correlation matrix.*

If the variable \mathbf{X} in Definition 1 has *Unif*(0, 1) marginal distributions, then its distribution is a jointly symmetric copula. It is possible to show that, given symmetry of the marginal distributions,² joint symmetry of the copula is necessary and sufficient for joint symmetry of the joint distribution, via the N -dimensional analog of Exercise 2.30 in Nelsen (2006):

Lemma 2 *Let \mathbf{X} be a vector of N continuous random variables with joint distribution \mathbf{F} , marginal distributions F_1, \dots, F_N and copula \mathbf{C} . Further suppose X_i is symmetric about $a_i \forall i$. Then \mathbf{X} is jointly symmetric about $\mathbf{a} \equiv [a_1, \dots, a_N]$ if and only if \mathbf{C} is jointly symmetric.*

Lemma 2 implies that any combination of symmetric marginal distributions, of possibly different forms (e.g., Normal, Student’s t with different degrees of freedom, double-exponential, etc.) with any jointly symmetric copula yields a jointly symmetric joint distribution, and by Lemma 1, all

²We empirically test the assumption of marginal symmetry in our application in Section 5, and there we also describe methods to overcome this assumption if needed.

such distributions have an identity correlation matrix, and can thus be used in a model such as the one proposed in equations (1)-(2).

While numerous copulas have been proposed to explain various features of dependences in the literature, only a few existing copulas are jointly symmetric, for example, the Gaussian and t copulas with a identity correlation matrix. To overcome this limited choice of copulas, we next propose a novel way to construct jointly symmetric copulas by “rotating” any given copula, thus vastly increasing the set of possible copulas to use in applications.

Theorem 1 *Assume that N dimensional copula \mathbf{C} , with density \mathbf{c} , is given.*

(i) *The following copula \mathbf{C}^{JS} is jointly symmetric:*

$$\mathbf{C}^{JS}(u_1, \dots, u_N) = \frac{1}{2^N} \sum_{k_1=0}^2 \cdots \sum_{k_N=0}^2 (-1)^R \cdot \mathbf{C}(\tilde{u}_1, \dots, \tilde{u}_i, \dots, \tilde{u}_N) \quad (6)$$

$$\text{where } R = \sum_{i=1}^N \mathbf{1}\{k_i = 2\}, \quad \text{and } \tilde{u}_i = \begin{cases} 1, & \text{if } k_i = 0 \\ u_i, & \text{if } k_i = 1 \\ 1 - u_i, & \text{if } k_i = 2 \end{cases}$$

(ii) *The probability density function \mathbf{c}^{JS} implied by \mathbf{C}^{JS} is*

$$\mathbf{c}^{JS}(u_1, \dots, u_N) = \frac{\partial^N \mathbf{C}^{JS}(u_1, \dots, u_N)}{\partial u_1 \cdots \partial u_N} = \frac{1}{2^N} \sum_{k_1=1}^2 \cdots \sum_{k_N=1}^2 \mathbf{c}(\tilde{u}_1, \dots, \tilde{u}_i, \dots, \tilde{u}_N) \quad (7)$$

Theorem 1 shows that the average of mirror-image rotations of a potentially asymmetric copula about every axis generates a jointly symmetric copula. Note that the copula cdf, \mathbf{C}^{JS} , in equation (6) involves all marginal copulas (i.e., copulas of dimension 2 to $N - 1$) of the original copula, whereas the density \mathbf{c}^{JS} requires only the densities of the (entire) original copula, with no need for marginal copula densities. Also notice that \mathbf{c}^{JS} requires the evaluation of a very large number of densities even when N is only moderately large, which may be slow even when a single evaluation is quite fast. In Section 3 below we show how composite likelihood methods may be employed to overcome this computational problem.

We next show that we can further increase the set of jointly symmetric copulas for use in applications by considering convex combinations of jointly symmetric copulas.

Proposition 1 *Any convex combination of N -dimensional distributions that are jointly symmetric around a common point $a \in \mathbb{R}^1$ is jointly symmetric around a . This implies that (i) any convex combination of univariate distributions symmetric around a common point a is symmetric around a , and (ii) any convex combination of jointly symmetric copulas is a jointly symmetric copula.*

It is simple to visualize how to construct a jointly symmetric copula in terms of the copula density: the upper panels of Figure 1 show density contour plots for zero, 90-, 180- and 270-degree rotations of the Clayton copula, when combined with standard Normal marginal densities. The “jointly symmetric Clayton” copula is obtained by taking an equal-weighted average of these four densities, and is presented in the lower panel of Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

Figure 2 presents density contour plots for six jointly symmetric distributions that differ only in their jointly symmetric copula. The upper left panel is the case of independence, and the top right panel presents the jointly symmetric t copula, which is obtained when the correlation parameter of that copula is set to zero. The remaining four panels illustrate the flexibility of the models that can be generated using Theorem 1. To aid interpretability, the lower four copulas have parameters chosen so that they are each approximately equally distant from the independence copula based on the Kullback-Leibler information criterion (KLIC). Figure 2 highlights the fact that the copula for uncorrelated random variables can be very different from the independence copula, capturing different types of “nonlinear” dependence.

[INSERT FIGURE 2 ABOUT HERE]

2.2 Forecasting models for multivariate covariance matrix

Research on forecasting models for multivariate covariance matrices with low-frequency data is pervasive, see Andersen, *et al.* (2006) for a review, and research on forecasting models using high frequency data is growing, e.g. Chiriac and Voev (2011), Noureldin, *et al.* (2012) among others. There are two major concerns about forecasting models for multivariate covariance matrices: parsimony and positive definiteness. Keeping these two concerns in mind, we combine the essential ideas of the DCC model of Engle (2002) and the heterogeneous autoregressive (HAR) model of Corsi (2009) to obtain a simple and flexible new forecasting model for covariance matrices. Following the

DCC model, we estimate the variances and correlations separately, to reduce the computational burden. We use the HAR model structure, which is known to successfully capture the long-memory behavior of volatility in a simple autoregressive way.

Let Δ be the sampling frequency (e.g., 5 minutes), which yields $1/\Delta$ observations per trade day. The $N \times N$ realized covariance matrix for the interval $[t-1, t]$ is defined by

$$RVarCov_t^\Delta = \sum_{j=1}^{1/\Delta} \mathbf{r}_{t-1+j\cdot\Delta} \mathbf{r}'_{t-1+j\cdot\Delta} \quad (8)$$

and is re-written in terms of realized variances and realized correlations as:

$$RVarCov_t^\Delta = \sqrt{RVar_t^\Delta} \cdot RCorr_t^\Delta \cdot \sqrt{RVar_t^\Delta} \quad (9)$$

where $RVar_t^\Delta = \text{diag} \{RVarCov_t^\Delta\}$ is a diagonal matrix with the realized variances on the diagonal, and $RCorr_t^\Delta = (RVar_t^\Delta)^{-1/2} \cdot RVarCov_t^\Delta \cdot (RVar_t^\Delta)^{-1/2}$.

We propose to first apply the HAR model to each (log) realized variance:

$$\begin{aligned} \log RVar_{ii,t}^\Delta &= \phi_i^{(const)} + \phi_i^{(day)} \log RVar_{ii,t-1}^\Delta + \phi_i^{(week)} \frac{1}{4} \sum_{k=2}^5 \log RVar_{ii,t-k}^\Delta \\ &\quad + \phi_i^{(month)} \frac{1}{15} \sum_{k=6}^{20} \log RVar_{ii,t-k}^\Delta + \xi_{it}, \quad i = 1, 2, \dots, N. \end{aligned} \quad (10)$$

and the coefficients $\{\phi_i^{(const)}, \phi_i^{(day)}, \phi_i^{(week)}, \phi_i^{(month)}\}_{i=1}^N$ are estimated by OLS for each variance. We use the logarithm of the realized variance to ensure that all variance forecasts are positive, and also to reduce the influence of large observations, which is important as the sample period in our empirical analysis includes the 2008 financial crisis.

Next, we propose a model for realized correlations, using the *vech* operator. Consider the following HAR-type model for correlations:

$$\begin{aligned} vech(RCorr_t^\Delta) &= vech\left(\overline{RCorr_t^\Delta}\right) (1 - a - b - c) + a \cdot vech(RCorr_t^\Delta) \\ &\quad + b \cdot \frac{1}{4} \sum_{k=2}^5 vech(RCorr_{t-k}^\Delta) + c \cdot \frac{1}{15} \sum_{k=6}^{20} vech(RCorr_{t-k}^\Delta) + \boldsymbol{\xi}_t \end{aligned} \quad (11)$$

where $\overline{RCorr_t^\Delta} = \frac{1}{T} \sum_{t=1}^T RCorr_t^\Delta$ and $(a, b, c) \in \mathbb{R}^3$. A more flexible version of this model would allow (a, b, c) to be replaced with $N(N-1)/2 \times N(N-1)/2$ matrices (A, B, C) , however the

number of free parameters in such a specification would be $\mathcal{O}(N^2)$, and is not feasible for high-dimensional applications. In this parsimonious specification, the coefficients a, b , and c are easily estimated by OLS regardless of the dimension. Note that the form of the model in equation (11) is such that the predicted value will indeed be a correlation matrix (when the *vech* operation is undone), and so the residual in this specification, $\boldsymbol{\xi}_t$, is one that lives in the space of differences of correlation matrices. As we employ OLS for estimation, we are able to avoid having to specify a distribution for this variable.

Let $\widehat{RVarCov}_t^\Delta$ denote a forecast of the covariance matrix based on equations (10) and (11) and estimated parameters. The theorem below provides conditions under which $\widehat{RVarCov}_t^\Delta$ is guaranteed to be positive definite.

Theorem 2 *Assume that (i) $\Pr[\mathbf{x}'\mathbf{r}_t = 0] = 0$ for any nonzero $\mathbf{x} \in \mathbb{R}^N$ (i.e. \mathbf{r}_t does not have redundant assets), (ii) $[\hat{a}, \hat{b}, \hat{c}] \geq 0$, and (iii) $\hat{a} + \hat{b} + \hat{c} < 1$. Then, $\widehat{RVarCov}_t^\Delta$ is positive definite.*

Our forecasting model for the realized covariance matrix is simple and fast to estimate and positive definiteness is ensured by Theorem 2. We note that the above theorem is robust to the misspecification of return distributions, i.e. Theorem 2 holds regardless of whether or not return distribution follows the proposed model specified by equations (1)–(2).

3 Estimation methods and model comparisons

This section proposes a composite likelihood approach to estimate models from the class of jointly symmetric copulas proposed in Theorem 1, and then describes corresponding methods for model comparison tests of copula models specified and estimated in this way. Finally, we present results on how to handle the estimation error for the complete model, taking into account the multi-stage nature of the proposed estimation methods.

3.1 Estimation using composite likelihood

The proposed method to construct jointly symmetric copulas in Theorem 1 requires 2^N evaluations of the given original copula density. Even for moderate dimensions, say $N = 20$, the likelihood evaluation may be too slow to calculate. We illustrate this using a jointly symmetric copula based on the Clayton copula, which has a simple closed-form density and requires just a fraction of a

second for a single evaluation.³ The first row of Table 1 shows that as the dimension, and thus the number of rotations, increases, the computation time for a single evaluation of the jointly symmetric Clayton copula grows from less than a second to several minutes to many years.⁴

[INSERT TABLE 1 ABOUT HERE]

For high dimensions, ordinary maximum likelihood estimation is not feasible for our jointly symmetric copulas. A composite likelihood (Lindsay, 1988) consists of combinations of the likelihoods of submodels or marginal models of the full model, and under certain conditions maximizing the composite likelihood (CL) can be shown to generate parameter estimates that are consistent for the true parameters of the model.⁵ The essential intuition behind CL is that since submodels include partial information on the parameters of the full model, by properly using that partial information we can estimate the parameters of full model, although of course subject to some efficiency loss.

The composite likelihood can be defined in various ways, depending on which sub-models of the full model are employed. In our case, the use of bivariate sub-models is particularly attractive, as a bivariate sub-model of the jointly symmetric copula generated using equation (6) requires only four rotations. This is easily shown using some copula manipulations, and we summarize this result in the proposition below.

Proposition 2 *For N -dimensional jointly symmetric copulas generated using Theorem 1, the (i, j) bivariate marginal copula density is obtained as*

$$\mathbf{c}_{ij}^{JS}(u_i, u_j) = \frac{1}{4} \{ \mathbf{c}_{ij}(u_i, u_j) + \mathbf{c}_{ij}(1 - u_i, u_j) + \mathbf{c}_{ij}(u_i, 1 - u_j) + \mathbf{c}_{ij}(1 - u_i, 1 - u_j) \}$$

where \mathbf{c}_{ij} is the (i, j) marginal copula density of the original N -dimensional copula.

Thus while the full model requires 2^N rotations of the original density, bivariate marginal models only require 2^2 rotations. Similar to Engle, *et al.* (2008), we consider CL based either on all pairs

³All computational times reported are based on using Matlab R2014a on a 3.4GHz Intel PC with Windows 7.

⁴While evaluation of the likelihood is slow, simulating from this model is simple and fast (see Section 4.1 for details). This suggests that simulation-based alternatives to maximum likelihood might be feasible for these models. We leave the consideration of this interesting possibility for future research.

⁵See Varin, *et al.* (2011) for an overview of this method, and see Engle, *et al.* (2008) for an application of this method in financial econometrics.

of variables, only adjacent pairs of variables,⁶ and only the first pair of variables:

$$CL_{all}(u_1, \dots, u_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \log \mathbf{c}_{i,j}(u_i, u_j) \quad (12)$$

$$CL_{adj}(u_1, \dots, u_N) = \sum_{i=1}^{N-1} \log \mathbf{c}_{i,i+1}(u_i, u_{i+1}) \quad (13)$$

$$CL_{first}(u_1, \dots, u_N) = \log \mathbf{c}_{1,2}(u_1, u_2) \quad (14)$$

As one might expect, estimators based on these three different CLs will have different degrees of efficiency, and we study this in detail in our simulation study in the next section.

While there are many different ways to construct composite likelihoods, they all have some common features. First of all, they are valid likelihoods since the likelihood of the sub-models are themselves valid likelihoods. Second, the joint model implied by taking products of densities of sub-models (i.e., imposing an incorrect independence assumption) causes misspecification and the information matrix equality will not hold. Third, the computation of the composite likelihood is substantially faster than that of full likelihood. In our application the computational burden is reduced from $\mathcal{O}(2^N)$ to $\mathcal{O}(N^2)$, $\mathcal{O}(N)$ or $\mathcal{O}(1)$ when we use all pairs, only adjacent pairs, or only the first pair of variables. The bottom three rows in Table 1 show the computation gains from using a composite likelihood based on one of the three combinations in equations (12)–(14) compared with using the full likelihood.

Let us define a maximum composite likelihood estimator (MCLE) as:

$$\hat{\boldsymbol{\theta}}_{MCLE} = \arg \max_{\boldsymbol{\theta}} \sum_{t=1}^T CL(u_{1t}, \dots, u_{Nt}; \boldsymbol{\theta}) \quad (15)$$

where CL is a composite log-likelihood, such as one of those in equations (12)–(14). Under mild regularity conditions (see Newey and McFadden, 1994 or White, 1994), and an identification condition we discuss in the next paragraph, Cox and Reid (2004) show that

$$\sqrt{T} \left(\hat{\boldsymbol{\theta}}_{MCLE} - \boldsymbol{\theta}_0 \right) \xrightarrow{d} N \left(0, \mathcal{H}_0^{-1} \mathcal{J}_0 \mathcal{H}_0^{-1} \right) \quad (16)$$

where $\mathcal{H}_0 = -E \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} CL(u_{1t}, \dots, u_{Nt}; \boldsymbol{\theta}_0) \right]$ and $\mathcal{J}_0 = V \left[\frac{\partial}{\partial \boldsymbol{\theta}} CL(u_{1t}, \dots, u_{Nt}; \boldsymbol{\theta}_0) \right]$. We refer the

⁶For a given (arbitrary) order of the variables, the “adjacent pairs” CL uses pairs (u_i, u_{i+1}) for $i = 1, \dots, N - 1$. Similarly, the “first” pair is simply whichever series were arbitrarily labelled as the first two.

reader to Cox and Reid (2004) for the proof. The asymptotic variance of MCLE takes a “sandwich” form, and is of course weakly greater than that of the MLE. We investigate the extent of the efficiency loss of MCLE relative to MLE in the simulation study in the next section.

The identification condition required for CL estimation comes from the first-order condition implied by the optimization problem. Specifically, it is required that

$$E \left[\frac{\partial}{\partial \boldsymbol{\theta}} CL(u_{1t}, \dots, u_{Nt}; \boldsymbol{\theta}) \right] \begin{cases} = \mathbf{0} & \text{for } \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \neq \mathbf{0} & \text{for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases} \quad (17)$$

That is, the components of the composite likelihood must be rich enough to identify the parameters of full likelihood. As a problematic example, consider a composite likelihood that uses only the first pair of variables (as in equation 14), but some elements of $\boldsymbol{\theta}$ do not affect the dependence between the first pair. With such a CL, $\boldsymbol{\theta}$ would not be identified, and one would need to look for a richer set of submodels to identify the parameters, for example using more pairs, as in equation (12) and (13), or using higher dimension submodels, e.g. trivariate marginal copulas. In our applications, we consider as “generating” copulas only those with a single unknown parameter that affects all bivariate copulas, and thus all of the CLs in equations (12)–(14) are rich enough to identify the unknown parameter.

3.2 Model selection tests with composite likelihood

We next consider in-sample and out-of-sample model selection tests when composite likelihood is involved. The tests we discuss here are guided by our empirical analysis in Section 5, so we only consider the case where composite likelihoods with adjacent pairs are used. We first define the composite Kullback-Leibler information criterion (cKLIC) following Varin and Vidoni (2005).

Definition 2 *Given an N -dimensional random variable $\mathbf{Z} = (Z_1, \dots, Z_N)$ with true density \mathbf{g} , the composite Kullback-Leibler information criterion (cKLIC) of a density \mathbf{h} relative to \mathbf{g} is*

$$I_c(\mathbf{g}, \mathbf{h}) = E_{\mathbf{g}(\mathbf{z})} \left[\log \prod_{i=1}^{N-1} \mathbf{g}_i(z_i, z_{i+1}) - \log \prod_{i=1}^{N-1} \mathbf{h}_i(z_i, z_{i+1}) \right]$$

where $\prod_{i=1}^{N-1} \mathbf{g}_i(z_i, z_{i+1})$ and $\prod_{i=1}^{N-1} \mathbf{h}_i(z_i, z_{i+1})$ are adjacent-pair composite likelihoods using the true density \mathbf{g} and a competing density \mathbf{h} .

We focus on CL using adjacent pairs, but other cKLICs can be defined similarly. Note that the composite log-likelihood for the joint distribution can be decomposed using Sklar’s theorem (equation 3) into the marginal log-likelihoods and the copula composite log-likelihood. We use this expression when comparing our joint density models in our empirical work below.⁷

$$\begin{aligned}
CL_h &\equiv \sum_{i=1}^{N-1} \log \mathbf{h}(z_i, z_{i+1}) \\
&= \log h_1(z_1) + \log h_N(z_N) + 2 \sum_{i=1}^{N-1} \log h_i(z_i) + \sum_{i=1}^{N-1} \log \mathbf{c}(H_i(z_i), H_{i+1}(z_{i+1}))
\end{aligned} \tag{18}$$

Secondly, notice that the expectation in the definition of cKLIC is with respect to the (complete) true density \mathbf{g} rather than the CL of the true density, which makes it possible to interpret cKLIC as a linear combination of the ordinary KLIC of the submodels used in the CL:

$$I_c(\mathbf{g}, \mathbf{h}) = \sum_{i=1}^{N-1} E_{\mathbf{g}(\mathbf{z})} \left[\log \frac{\mathbf{g}_i(z_i, z_{i+1})}{\mathbf{h}_i(z_i, z_{i+1})} \right] = \sum_{i=1}^{N-1} E_{\mathbf{g}_i(z_i, z_{i+1})} \left[\log \frac{\mathbf{g}_i(z_i, z_{i+1})}{\mathbf{h}_i(z_i, z_{i+1})} \right] \tag{19}$$

The second equality holds since the expectation of a function of (Z_i, Z_{i+1}) only depends on the bivariate distribution of those two variables, not the entire joint distribution. The above equation shows that cKLIC can be viewed as a linear combination of the ordinary KLICs of submodels, which implies that existing in-sample model selection tests, such as those of Vuong (1989) for *iid* data and Rivers and Vuong (2002) for time series, can be straightforwardly applied to model selection using cKLIC.⁸ To the best of our knowledge, combining cKLIC with Vuong (1989) or Rivers and Vuong (2002) tests is new to the literature.

We may also wish to select the best model in terms of out-of-sample (OOS) forecasting performance measured by some scoring rule, \mathcal{S} , for the model. Gneiting and Raftery (2007) define “proper” scoring rules as those which satisfy the condition that the true density always receives a higher score, in expectation, than other densities. Gneiting and Raftery (2007) suggest that the “natural” scoring rule is the log density, i.e. $\mathcal{S}(\mathbf{h}(\mathbf{Z})) = \log \mathbf{h}(\mathbf{Z})$, and it can be shown that this

⁷In our empirical work, we also include the pair (z_1, z_N) in the “adjacent” composite likelihood so that all marginals enter into the joint composite likelihood twice.

⁸We note that a model selection test based on the full likelihood could give a different answer to one based on a composite likelihood. We leave the consideration of this possibility for future research.

scoring rule is proper.⁹ We may consider a similar scoring rule based on log composite density:

$$\mathcal{S}(\mathbf{h}(\mathbf{Z})) = \sum_{i=1}^{N-1} \log \mathbf{h}_i(Z_i, Z_{i+1}) \quad (20)$$

This scoring rule is shown to be proper in the following theorem.

Theorem 3 *The scoring rule based on log composite density given in equation (20) is proper, i.e.*

$$E \left[\sum_{i=1}^{N-1} \log \mathbf{h}_i(Z_i, Z_{i+1}) \right] \leq E \left[\sum_{i=1}^{N-1} \log \mathbf{g}_i(Z_i, Z_{i+1}) \right] \quad (21)$$

where the expectation is with respect to the true density \mathbf{g} , and \mathbf{g}_i and \mathbf{h}_i are the composite likelihoods of the true density and the competing density respectively.

This theorem allows us to interpret OOS tests based on CL as being related to the cKLIC, analogous to OOS tests based on the full likelihood being related to the KLIC. In our empirical analysis below we employ a Giacomini and White (2006) test based on an OOS CL scoring rule.

3.3 Multi-stage estimation and inference

We next consider multi-stage estimation of models such as those defined by equations (1)-(3). We consider general parametric models for the conditional mean and covariance matrix:

$$\begin{aligned} \boldsymbol{\mu}_t &\equiv \boldsymbol{\mu}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}^{mean}), \quad \mathbf{Y}_{t-1} \in \mathcal{F}_{t-1} \\ \mathbf{H}_t &\equiv \mathbf{H}(\mathbf{Y}_{t-1}; \boldsymbol{\theta}^{var}) \end{aligned} \quad (22)$$

This assumption allows for a variety of models for conditional mean, for example, ARMA, VAR, linear and nonlinear regressions for the mean, and various conditional covariance models, such as DCC, BEKK, and DECO, and stochastic volatility models (see Andersen, *et. al* (2006) and Shephard (2005) for reviews) as well as the new model proposed in Section 2.2.

The standardized uncorrelated residuals in equation (3) follow a parametric distribution:

$$\mathbf{e}_t \sim iid \mathbf{F} = \mathbf{C} \left(F_1(\cdot; \boldsymbol{\theta}_1^{mar}), \dots, F_N(\cdot; \boldsymbol{\theta}_N^{mar}); \boldsymbol{\theta}^{copula} \right) \quad (23)$$

⁹The expectation of the log scoring rule is equal to the KLIC up to an additive constant. Since the KLIC measures how close the density forecasts to the true density, the log scoring rule can be used as a metric to determine which model is closer to the true density.

where the marginal distributions F_i have zero mean, unit variance, and are symmetric about zero and the copula \mathbf{C} is jointly symmetric, which together ensures an identity correlation matrix for \mathbf{e}_t . The parametric specification of $\boldsymbol{\mu}_t$, \mathbf{H}_t , F_i and \mathbf{C} theoretically enables the use of (one-stage) maximum likelihood estimation, however, when N is large, this estimation strategy is not feasible, and multi-stage ML (MSML) estimation is a practical alternative. We describe MSML estimation in detail below. To save space $\boldsymbol{\theta}^{mean}$ is assumed to be known in this section. (For example, it is common to assume that daily returns are mean zero.)

The covariance model proposed in Section 2.2 allows for the separate estimation of the conditional variances and the conditional correlation matrix, similar to the DCC model of Engle (2002) which we also consider in our empirical application below. Thus we can decompose the parameter $\boldsymbol{\theta}^{var}$ into $[\boldsymbol{\theta}_1^{var}, \dots, \boldsymbol{\theta}_N^{var}, \boldsymbol{\theta}^{corr}]$, and then represent the complete set of unknown parameters as

$$\boldsymbol{\theta} \equiv \begin{bmatrix} \boldsymbol{\theta}_1^{var} & \dots & \boldsymbol{\theta}_N^{var} & \boldsymbol{\theta}^{corr} & \boldsymbol{\theta}_1^{mar} & \dots & \boldsymbol{\theta}_N^{mar} & \boldsymbol{\theta}^{cop} \end{bmatrix}. \quad (24)$$

As usual for multi-stage estimation, we assume that each sub-vector of parameters is estimable in just a single stage of the analysis, and we estimate the elements of $\boldsymbol{\theta}$ as follows:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_i^{var} &\equiv \arg \max_{\boldsymbol{\theta}_i^{var}} \sum_{t=1}^T \log l_{it}^{var}(\boldsymbol{\theta}_i^{var}), \quad i = 1, \dots, N \\ \hat{\boldsymbol{\theta}}^{corr} &\equiv \arg \max_{\boldsymbol{\theta}^{corr}} \sum_{t=1}^T \log l_t^{corr}(\hat{\boldsymbol{\theta}}_1^{var}, \dots, \hat{\boldsymbol{\theta}}_N^{var}, \boldsymbol{\theta}^{corr}) \\ \hat{\boldsymbol{\theta}}_i^{mar} &\equiv \arg \max_{\boldsymbol{\theta}_i^{mar}} \sum_{t=1}^T \log l_{it}^{mar}(\hat{\boldsymbol{\theta}}_1^{var}, \dots, \hat{\boldsymbol{\theta}}_N^{var}, \hat{\boldsymbol{\theta}}^{corr}, \boldsymbol{\theta}_i^{mar}), \quad i = 1, \dots, N \\ \hat{\boldsymbol{\theta}}^{cop} &\equiv \arg \max_{\boldsymbol{\theta}^{cop}} \sum_{t=1}^T \log l_t^{cop}(\hat{\boldsymbol{\theta}}_1^{var}, \dots, \hat{\boldsymbol{\theta}}_N^{var}, \hat{\boldsymbol{\theta}}^{corr}, \hat{\boldsymbol{\theta}}_1^{mar}, \dots, \hat{\boldsymbol{\theta}}_N^{mar}, \boldsymbol{\theta}^{cop}) \end{aligned} \quad (25)$$

In words, the first stage estimates the N individual variance models based on QMLE; the next stage uses the standardized returns to estimate the correlation model, using QMLE or a composite likelihood method (as in Engle, *et al.*, 2008); the third stage estimates the N marginal distributions of the estimated standardized uncorrelated residuals; and the final stage estimates the copula of the standardized residuals based on the estimated “probability integral transforms.” This final stage may be maximum likelihood (if the copula is such that this is feasible) or composite likelihood, as described in Section 3.1. We denote the complete vector of estimated parameters obtained from these four stages as $\hat{\boldsymbol{\theta}}_{MSML}$.

As is clear from the above, later estimation stages depend on previously estimated parameters, and the accumulation of estimation error must be properly incorporated into standard error calculations for $\hat{\boldsymbol{\theta}}_{MSML}$. Multi-stage ML estimation (and, in particular, multi-stage ML with a composite likelihood stage) can be viewed as a form of multi-stage GMM estimation, and under standard regularity conditions, it can be shown (see Newey and McFadden, 1994, Theorem 6.1) that

$$\sqrt{T} \left(\hat{\boldsymbol{\theta}}_{MSML} - \boldsymbol{\theta}^* \right) \xrightarrow{d} N(0, V_{MSML}^*) \text{ as } T \rightarrow \infty \quad (26)$$

Consistent estimation of V_{MSML}^* is theoretically possible, however in high dimensions it is not computationally feasible. For example, the proposed model used in Section 5 for empirical analysis has more than 1000 parameters, making V_{MSML}^* a very large matrix. An alternative is a bootstrap inference method, see Gonçalves, *et al.* (2013) for conditions under which block bootstrap may be used to obtain valid standard errors for multi-stage GMM estimators. Although this bootstrap approach is not expected to yield any asymptotic refinements, it allows us to avoid having to compute a large Hessian matrix. The bootstrap procedure is as follows: (i) generate a bootstrap sample of length T using a block bootstrap, such as the stationary bootstrap of Politis and Romano (1994), to preserve time series dependence in the data; (ii) obtain $\hat{\boldsymbol{\theta}}_{MSML}^{(b)}$ from the bootstrap sample, (iii) repeat steps (i)-(ii) B times and use the quantiles of $\left\{ \hat{\boldsymbol{\theta}}_{MSML}^{(b)} \right\}_{b=1}^B$ as critical values, or use $\alpha/2$ and $(1 - \alpha/2)$ quantiles of $\left\{ \hat{\boldsymbol{\theta}}_{MSML}^{(b)} \right\}_{b=1}^B$ to obtain $(1 - \alpha)$ confidence intervals for parameters.

4 Simulation study

4.1 Finite sample properties of MCLE for jointly symmetric copulas

In this section we use simulations to study the efficiency loss from maximum composite likelihood estimation (MCLE) relative to MLE, and we compare the efficiency of the three composite likelihoods presented in equations (12)–(14), namely “all pairs,” “adjacent pairs,” and “first pair.”

We specify the data generating process as follows, based on some copula \mathbf{C} and a set of independent Bernoulli random variables:

$$\tilde{u}_{it} = Z_{it}u_{it} + (1 - Z_{it})(1 - u_{it}), \quad t = 1, 2, \dots, T \quad (27)$$

$$\text{where } [u_{1t}, \dots, u_{Nt}] \equiv \mathbf{u}_t \sim iid \mathbf{C}(\theta)$$

$$\text{and } Z_{it} \sim iid \text{Bernoulli}(1/2), \text{ and } Z_{it} \perp Z_{jt} \forall i \neq j$$

We consider two choices for \mathbf{C} , the Clayton copula and with parameter equal to one and the Gumbel copula with parameter equal to two. We set $T = 1000$ and we consider dimensions $N = 2, 3, 5, 10, 20, \dots, 100$. We repeat all simulations 500 times.

We consider four different estimation methods: MLE, MCLE with all pairs (equation 12), MCLE with adjacent pairs (equation 13), and MCLE with the first pair (equation 14). MLE is not computationally feasible for $N > 10$, but the MCLEs are feasible for all dimensions considered.¹⁰ We report estimated run times for MLE for $N \geq 20$ to provide an indication of how long MLE would take to complete in those dimensions.

Table 2 presents the simulation results for the Clayton copula, and the web appendix presents corresponding results for the Gumbel copula. The average biases for all dimensions and for all estimation methods are small relative to the standard deviations. The standard deviations show, unsurprisingly, that MLE is more accurate than the three MCLEs; the efficiency loss of MCLE with “all pairs” to MLE is ranges from 5% to 37%. Among the three MCLEs, MCLE with all pairs has the smallest standard deviations and MCLE with the first pair has the largest, as expected. Comparing MCLE with adjacent pairs to MCLE with all pairs, we find that loss in efficiency is 23% for $N = 10$, and 5% for $N = 100$, and computation speed is two times faster for $N = 10$ and

¹⁰Note that the four estimation methods are equivalent when $N = 2$, and so the results are identical in the top row. Also note that the “first pair” MCLE results are identical across values of N , but we repeat the results down the rows for ease of comparison with the other estimation methods.

70 times faster for $N = 100$. For high dimensions, it is confirmed that MCLE with adjacent pairs performs quite well compared to MCLE with all pairs according to accuracy and computation time, which is similar to results in Engle, *et al.* (2008) on the use of adjacent pairs in the estimation of the DCC model.

In sum, MCLE is less efficient than MLE but still approximately unbiased and very fast for high dimensions. The accuracy of MCLE based only on adjacent pairs is similar to that of MCLE with all pairs, especially for high dimensions, and the gains in computation time are large. For this reason, we use MCLE with adjacent pairs for our empirical analysis in Section 5.

[INSERT TABLE 2 ABOUT HERE]

4.2 Finite sample properties of multi-stage estimation

Next we study multi-stage estimation for a representative model for daily asset returns. We assume:

$$\begin{aligned} \mathbf{r}_t &= \mathbf{H}_t^{1/2} \mathbf{e}_t & (28) \\ \mathbf{H}_t &\equiv \text{Cov}[\mathbf{r}_t | \mathcal{F}_{t-1}] \\ \mathbf{e}_t &\sim \text{iid } \mathbf{F} = \mathbf{C}(F_1(\cdot; \nu_1), \dots, F_N(\cdot; \nu_N); \boldsymbol{\varphi}) \end{aligned}$$

We set the mean return to zero, and we assume that the conditional covariance matrix, \mathbf{H}_t , follows a GARCH(1,1)-DCC model (see the web appendix for details of this specification). We use parameter values for these models based approximately on our empirical analysis in Section 5: we set the GARCH parameters as $[\psi_i, \kappa_i, \lambda_i] = [0.05, 0.1, 0.85] \forall i$, the DCC parameters as $[\alpha, \beta] = [0.02 \ 0.95]$, and we set the unconditional correlation matrix to equal the sample correlations of the first N stock returns used in our empirical analysis. We use a standardized Student's t distribution for the marginal distributions of the standardized residuals, F_i , and set the degrees of freedom parameter to six. We specify \mathbf{C} as a jointly symmetric copula constructed via Theorem 1, using the Clayton copula with parameter equal to one.

We estimate the model using the multi-stage estimation described in Section 3.3. The parameters of GARCH for each variables are estimated via QML at the first stage, and the parameters of the DCC model are estimated via variance targeting and composite likelihood with adjacent pairs, see Engle, *et al.* (2008) for details. We use ML to estimate the marginal distributions of the

standardized residuals, and finally we estimate the copula parameters using MCLE with adjacent pairs as explained in Section 3.1. We repeat this scenario 500 times with time series of length $T = 1000$ and cross-sectional dimensions of $N = 10, 50$, and 100. Table 3 reports all parameter estimates except $\bar{\mathbf{Q}}$. The columns for $\psi_i, \kappa_i, \lambda_i$ and ν_i report the summary statistics obtained from $500 \times N$ estimates since those parameters are the same across all variables.

[INSERT TABLE 3 ABOUT HERE]

Table 3 reveals that the estimated parameters are centered on the true values with the average estimated bias being small relative to the standard deviation. As the dimension size increases, the copula model parameters are more accurately estimated, which was also found in the previous section. Since this copula model keeps the dependence between any two variables identical, the amount of information on the unknown copula parameter increases as the dimension grows. The average computation time is reported in the bottom row of each panel, and it indicates that multi-stage estimation is quite fast: for example, it takes five minutes for the one hundred dimension model, in which the total number of parameters to estimate is more than 5000.

To see the impact of estimation errors from the former stages to copula estimation, we compare the standard deviations of the estimated copula parameters in Table 3 with the corresponding results in Table 2. The standard deviation increases by about 30% for $N = 10$, and by about 19% for $N = 50$ and 100. The loss of accuracy caused by having to estimate the parameters of the marginals is relatively small, given that more than 5000 parameters are estimated in the former stages. We conclude that multi-stage estimation with composite likelihood results in a large reduction in the computational burden (indeed, they make this estimation problem feasible using current computing power) and yields reliable parameter estimates.

5 Empirical analysis of S&P 100 equity returns

In this section we apply our proposed multivariate distribution model to equity returns over the period January 2006 to December 2012, a total of $T = 1761$ trade days. We study every stock that was ever a constituent of the S&P 100 equity index during this sample, and which traded for the full sample period, yielding a total of $N = 104$ assets. The web appendix contains a table with the names of these 104 stocks. We obtain high frequency transaction data on these stocks

from the NYSE TAQ database, and clean these data following Barndorff-Nielsen, *et al.* (2009), see Bollerslev, *et al.* (2014) for details. We adjust prices affected by splits and dividends using “adjustment” factors from CRSP. Daily returns are calculated using the log-difference of the close prices from high frequency data. For high frequency returns, log-differences of five minute prices are used and overnight returns are treated as the first return in a day.

5.1 Volatility models and marginal distributions

Table 4 presents the summary statistics of the data and the estimates of conditional mean model. The top panel presents unconditional sample moments of the daily returns for each stock. Those numbers broadly match values reported in other studies, for example, strong evidence for fat tails. In the lower panel, the formal tests for zero skewness and zero excess kurtosis are conducted. The tests show that only 3 stocks out of 104 have a significant skewness, and all stocks have a significant excess kurtosis. For reference, we also test for zero pair-wise correlations, and we reject the null for all pairs of asset returns. The middle panel shows the estimates of the parameters of AR(1) models. Constant terms are estimated to be around zero and the estimates of AR(1) coefficient are slightly negative, both are consistent with values in other studies.

[INSERT TABLE 4 ABOUT HERE]

We estimate two different models for conditional covariance matrix: the HAR-type model described in Section 2.2 and a GJR-GARCH–DCC model.¹¹ The latter model uses daily returns, and the former exploits 5-minute intra-daily returns,¹² both models are estimated using quasi-maximum likelihood. The estimates of HAR variance models are presented in Panel A of Table 5, and are similar to those reported in Corsi (2009): coefficients on past daily, weekly, and monthly realized variances are around 0.38, 0.31 and 0.22. For the HAR-type correlation model, however, the coefficient on past monthly correlations is the largest followed by weekly and daily. The parameter estimates for the DCC model presented in Panel B are close to other studies of daily stock returns, indicating volatility clustering, asymmetric volatility dynamics, and highly persistent time-varying correlations. The bootstrap standard errors described in Section 3.3 are provided for the correlation models, and they take into account the estimation errors of former stages.

¹¹In the interests of space, we report the details of this familiar specification in the web appendix to this paper.

¹²We use the standard realized covariance matrix, see Barndorff-Nielsen and Shephard (2004), in the HAR models, and we do not try to correct for the (weak) AR dynamics captured in the conditional mean model.

[INSERT TABLE 5 ABOUT HERE]

The standardized residuals are constructed as $\hat{\mathbf{e}}_{t,M} \equiv \hat{\mathbf{H}}_{t,M}^{-1/2} (\mathbf{r}_t - \hat{\boldsymbol{\mu}}_t)$ where $M \in \{HAR, DCC\}$. We use the spectral decomposition rather than the Cholesky decomposition to compute the square-root matrix due to the former's invariance to the order of the variables. Summary statistics on the standardized residuals are presented in Panels A and B of Table 6.

Our proposed approach for modelling the joint distribution of the standardized residuals is based on a jointly symmetric distribution, and thus a critical first step is to test for *univariate* symmetry of these residuals. We do so in Panel D of Table 6. We find that we can reject the null of zero skewness for only 4/104 and 6/104 series based on the HAR and DCC models. Thus the assumption of symmetry appears reasonable for this data set.¹³ We also test for zero excess kurtosis and we reject it for all 104 series for both volatility models. These two test results motivate our choice of a standardized Student's t distribution for the marginal distributions of the residuals. Finally, as a check of our conditional covariance models, we also test for zero correlations between the residuals. We find that we can reject this null for 9.2% and 0.0% of the 5356 pairs of residuals, using the HAR and DCC models. Thus both models provide a reasonable estimate of the time-varying conditional covariance matrix, although by this metric the DCC model would be preferred over the HAR model.

Panel C of Table 6 presents the cross-sectional quantiles of 104 estimated degrees of freedom parameters of standardized Student's t distributions. These estimates range from 4.1 (4.2) at the 5% quantile to 6.9 (8.3) at the 95% quantile for the HAR (DCC) model. Thus both sets of standardized residuals imply substantial kurtosis, and, interestingly for the methods proposed in this paper, substantial *heterogeneity* in kurtosis. A simple multivariate t distribution could capture the fat tails exhibited by our data, but it imposes the same degrees of freedom parameter on all 104 series. Panel C suggests that this restriction is not supported by the data, and we show in formal model selection tests below that this assumption is indeed strongly rejected.

[INSERT TABLE 6 ABOUT HERE]

¹³If these tests indicated the presence of significant asymmetry, then an alternative approach based on a combination of the one presented here and that of Lee and Long (2009) might be employed: First, use the current approach for the joint distribution of the variables for which symmetry is not rejected. Then use Lee and Long's approach for the joint distribution of the asymmetric variables. Finally combine the two sets of variables invoking the assumption that the entire (N -dimensional) copula is jointly symmetric. As discussed in Section 2, such an approach will be computationally demanding if the number of asymmetric variables is large, but this hybrid approach offers a substantial reduction in the computational burden if a subset of the variables are symmetrically distributed.

5.2 Specifications for the copula

We next present the most novel aspect of this empirical analysis: the estimation results for a selection of jointly symmetric copula models. Parameter estimates and standard errors for these models are presented in Table 7. We consider four jointly symmetric copulas based on the t , Clayton, Frank, and Gumbel copulas. The jointly symmetric copulas based on Clayton, Frank and Gumbel are constructed using Theorem 1, and the jointly symmetric t copula is obtained simply by imposing an identity correlation matrix for that copula.¹⁴ We compare our jointly symmetric specifications with two well-known benchmark models: the independence copula and the multivariate Student's t distribution. The independence copula is a special case of a jointly symmetric copula, and there is no parameter to estimate. The multivariate t distribution is what would be obtained if our jointly symmetric t copula and all 104 univariate t distributions had the same degrees of freedom parameter, and in this case there would be no gains to using Sklar's theorem to decompose the joint distribution of the residuals into marginal distributions and the copula. Note that while the independence copula imposes a stronger condition on the copula specification than the multivariate t distribution, it does allow each of the marginal distributions to be possibly heterogeneous Student's t distributions, and so the ordering of these two specifications is not clear *ex ante*. This table also reports bootstrap standard errors which incorporate accumulated estimation errors from former stages. We follow steps explained in Section 3.3 to obtain these standard errors. The average block length for the stationary bootstrap is set to 100.

[INSERT TABLE 7 ABOUT HERE]

The log-likelihoods of the complete model for all 104 daily returns are reported for each of the models in Table 7, along with the rank of each model according to its log-likelihood, out of the twelve competing specifications presented here. Comparing the values of the log-likelihoods, we draw two initial conclusions. First, copula methods (even the independence copula) outperform the multivariate t distribution, which imposes strong homogeneity on the marginal distributions and the copula. Second, high frequency data improves the fit of all models relative to the use of daily data: the best six performing models are those based on the HAR specification.

¹⁴It is important to note that the combination of a jointly symmetric t copula with the 104 univariate t marginal distributions does *not* yield a multivariate t distribution, except in the special case that all 105 degrees of freedom parameters are identical. We test that restriction below and find that it is strongly rejected.

We next study the importance of allowing for nonlinear dependence. The independence copula assumes no nonlinear dependence, and we can test for the presence of nonlinear dependence by comparing the remaining specifications with the independence copula. Since the four jointly symmetric copulas and the multivariate t distribution all nest the independence copula,¹⁵ we can implement this test as a simple restriction on an estimated parameter. The t -statistics for those tests are reported in the bottom row of each panel of Table 7. Independence is strongly rejected in all cases, and we thus conclude that there is substantial nonlinear cross-sectional dependence in daily returns. While linear correlation and covariances are important for describing this vector of asset returns, these results reveal that these measures are *not* sufficient to completely describe their dependence.

Our model for the joint distribution of returns invokes an assumption that while linear dependence, captured via the correlation matrix, is time-varying, nonlinear dependence, captured through the distribution of the standardized residuals, is constant. We test this assumption by estimating the parameters of this distribution (the copula parameter, and the parameters of the 104 univariate Student's t marginal distributions) separately for the first and second half of our sample period, and then test whether they are significantly different. We find that 16 (19) of the HAR (DCC) marginal distribution parameters are significantly different at the 5% level, but none of the copula parameters are significantly different. Importantly, when we implement a joint test for a change in the entire parameter vector, we find no significant evidence (the p -values are both 0.99), and thus overall we conclude that this assumption is consistent with the data.¹⁶

We now turn to formal tests to compare the remaining, mostly non-nested, models. We consider both in-sample and out-of-sample tests.

¹⁵The t copula and the multivariate t distribution nest independence at $\theta^{-1} = 0$; the Clayton and Frank jointly symmetric copulas nest independence at $\theta = 0$; the Gumbel jointly symmetric copula nests independence at $\theta = 1$. We note, however, that independence is nested on the boundary of the parameter space in all cases, which requires a non-standard t test. The asymptotic distribution of the squared t -statistic no longer has χ_1^2 distribution under the null, rather it follows an equal-weighted mixture of a χ_1^2 and χ_0^2 , see Gouriéroux and Monfort (1996, Ch 21). The 90%, 95%, and 99% critical values for this distribution are 1.28, 1.64, and 2.33 which correspond to t -statistics of 1.64, 1.96, and 2.58.

¹⁶An alternative approach to capturing time-varying nonlinear dependence could be to specify a generalized autoregressive score (GAS) model (Creal, et al., 2013) for these parameters. GAS models have been shown to work well in high dimensions, see Oh and Patton (2013). We leave this interesting extension for future research.

5.3 Model selection tests

5.3.1 In-sample tests

As discussed in Section 3.2, the composite likelihood KLIC, (cKLIC) is a proper scoring rule, and can be represented as a linear combination of bivariate KLICs, allowing us to use existing in-sample model selection tests, such as those of Rivers and Vuong (2002). In a Rivers and Vuong test comparing two models, A and B , the null and alternative hypotheses are:

$$\begin{aligned}
 H_0 & : E [CL_t^A (\theta_A^*) - CL_t^B (\theta_B^*)] = 0 \\
 \text{vs. } H_1 & : E [CL_t^A (\theta_A^*) - CL_t^B (\theta_B^*)] > 0 \\
 H_2 & : E [CL_t^A (\theta_A^*) - CL_t^B (\theta_B^*)] < 0
 \end{aligned} \tag{29}$$

where $CL_t^M (\theta_M^*)$ is the day t composite likelihood for the joint distribution from model $M \in \{A, B\}$, and the expectation is taken with respect to the true, unknown, joint distribution. Rivers and Vuong (2002) show that a simple t -statistic on the difference between the sample averages of the log-composite likelihood has the standard Normal distribution under the null hypothesis:

$$\frac{\sqrt{T} \left\{ \overline{CL}_T^A (\hat{\theta}_A) - \overline{CL}_T^B (\hat{\theta}_B) \right\}}{\hat{\sigma}_T} \rightarrow N(0, 1) \text{ under } H_0 \tag{30}$$

where $\overline{CL}_T^M (\hat{\theta}_M) \equiv \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^{N-1} \log h_{i,i+1}^M (z_{i,t}, z_{i+1,t}; \hat{\theta}_M)$, for $M \in \{A, B\}$ and $\hat{\sigma}_T$ is some consistent estimator of $V \left[\sqrt{T} \left\{ \overline{CL}_T^A (\hat{\theta}_A) - \overline{CL}_T^B (\hat{\theta}_B) \right\} \right]$, such as the HAC estimator of Newey and West (1987).

Table 8 presents t -statistics from Rivers and Vuong (2002) model comparison tests. A positive t -statistic indicates that the model above beats the model to the left, and a negative one indicates the opposite. We first examine the bottom row of the upper panel to see whether the copula-based models outperform the multivariate t distribution. The multivariate t distribution is widely used as an alternative to Normal distribution not only in the literature but also in practice due to its thick tails and non-zero tail dependence. We observe that all t -statistics in that row are positive and larger than 18, indicating strong support in favor of the copula-based models. This outperformance is also achieved when the GARCH–DCC model using daily data is used (see the right half of the bottom row of the lower panel).

[INSERT TABLE 8 ABOUT HERE]

Next we consider model comparisons for the volatility models, to see whether a covariance matrix model that exploits high frequency data provides a better fit than one based only on daily data. The diagonal elements of the left half of the lower panel present these results, and in all cases we find that the model based on high frequency data significantly out-performs the corresponding model based on lower-frequency data. In fact, *all* t -statistics in the left half of the lower panel are positive and significant, indicating that the worst high frequency model is better than the best daily model. This is strong evidence of the gains from using high frequency data for capturing dynamics in conditional covariances.

Finally, we identify the best-fitting model of all twelve models considered here. The fact that all t -statistics in Table 8 are positive indicates that the first model listed in the top row is the best, and that is the model based on the jointly symmetric t copula. This model significantly beats all alternative models. (The second-best model is based on the jointly symmetric Clayton copula.) In Figure 3 we present the model-implied conditional correlation and the 1% quantile dependence, a measure of lower-tail dependence,¹⁷ for one pair of assets in our sample, Citi Group and Goldman Sachs, using the best model. The plot shows that the correlation between this pair ranges from 0.25 to around 0.75 over this sample period. The lower tail dependence implied by the jointly symmetric t copula ranges from 0.02 to 0.34, with the latter indicating very strong lower-tail dependence.

[INSERT FIGURE 3 ABOUT HERE]

5.3.2 Out of sample tests

We next investigate the out-of-sample (OOS) forecasting performance of the competing models. We use the period from January 2006 to December 2010 ($R = 1259$) as the in-sample period, and January 2011 to December 2012 ($P = 502$) as the out-of-sample period. We employ a rolling window estimation scheme, re-estimating the model each day in the OOS period. We use the Giacomini and White (2006) test to compare models based on their OOS composite likelihood. The implementation of these tests is analogous to the Rivers and Vuong test described above. We note here that the Giacomini and White test punishes complicated models that provide a good (in-sample) fit but are

¹⁷For two variables with a copula \mathbf{C} , the q -quantile dependence measure is obtained as $\tau^q = \mathbf{C}(q, q)/q$, and is interpretable as the probability that one of the variables will lie in the lower q tail of its distribution, conditional on the other variable lying in its lower q tail.

subject to a lot of estimation error. This feature is particularly relevant for comparisons of our copula-based approaches, which have 104 extra parameters for the marginal distribution models, with the multivariate t distribution, which imposes that all marginal distributions and the copula have the same degrees of freedom parameter.¹⁸

Table 9 presents t -statistics from these pair-wise OOS model comparison tests, with the same format as Table 8. The OOS results are broadly similar to the in-sample results, though with somewhat lower power. We again find that the multivariate t distribution is significantly beaten by *all* competing copula-based approaches, providing further support for the models proposed in this paper. We also again find strong support for the use of high frequency data for the covariance matrix model, with the HAR-type models outperforming the daily GARCH-DCC models.

Comparing the independence copula with the jointly symmetric copulas we again find that the independence copula is significantly beaten, providing evidence for the out-of-sample importance of modeling dependence beyond linear correlation. One difference in Table 9 relative to Table 8 is in the significance of the difference in performance between the four jointly symmetric copulas: we find that the jointly symmetric Gumbel copula is significantly beaten by the t and the Clayton, but neither of these latter two significantly beats the other, nor the Frank copula. The jointly symmetric t remains the model with the best performance, but it is not significantly better than the jointly symmetric Clayton or Frank models out of sample.

[INSERT TABLE 9 ABOUT HERE]

6 Conclusion

This paper proposes a new general model for high-dimensional distributions of asset returns that utilizes mixed frequency data and copulas. We decompose dependence into linear and nonlinear components, and exploit recent advances in the analysis of high frequency data to obtain more accurate models for linear dependence, as measured by the covariance matrix, and propose a new class of copulas to capture the remaining dependence in the low frequency standardized residuals. By assigning two different tasks to high frequency data and copulas, this separation significantly improves the performance of models for joint distributions. Our approach for obtaining jointly

¹⁸Also note that the Giacomini and White (2006) test can be applied to nested and non-nested models, and so all elements of Table 9 are computed in the same way. See Patton (2012) for more details on implementing in-sample and out-of-sample tests for copula-based models.

symmetric copulas generates a rich set of models for studying dependence of uncorrelated but dependent variables. The evaluation of the density of our jointly symmetric copulas turns out to be computationally difficult in high dimensions, but we show that composite likelihood methods may be used to estimate the parameters of the model and undertake model selection tests.

We employ our proposed models to study daily return distributions of 104 U.S. equities over the period 2006 to 2012. We find that our proposed models significantly outperform existing alternatives both in-sample and out-of-sample. The improvement in performance can be attributed to three main sources. Firstly, the use of a copula-based approach allows for the use of heterogeneous marginal distributions, relaxing a constraint of the familiar multivariate t distribution. Secondly, the use of copula models that allow for dependence beyond linear correlation, which relaxes a constraint of the Normal copula, leads to significant gains in fit. Finally, consistent with a large extant literature, we find that linear dependence, as measured by the covariance matrix, can be more accurately modelled by using high frequency data than using daily data alone.

Appendix: Proofs

The following two lemmas are needed to prove Lemma 2.

Lemma 3 *Let $\{X_i\}_{i=1}^N$ be N continuous random variables with joint distribution \mathbf{F} , marginal distributions F_1, \dots, F_N . Then $\{X_i\}_{i=1}^N$ is jointly symmetric about $\{a_i\}_{i=1}^N$ if and only if*

$$\begin{aligned} \mathbf{F}(a_1 + x_1, \dots, a_i + x_i, \dots, a_N + x_N) &= \mathbf{F}(a_1 + x_1, \dots, \infty, \dots, a_N + x_N) \\ &\quad - \mathbf{F}(a_1 + x_1, \dots, a_i - x_i, \dots, a_N + x_N) \quad \forall i \end{aligned} \tag{31}$$

$\mathbf{F}(a_1 + x_1, \dots, \infty, \dots, a_N + x_N)$ and $\mathbf{F}(a_1 + x_1, \dots, a_i - x_i, \dots, a_N + x_N)$ mean that only the i^{th} element is ∞ and $a_i - x_i$, respectively, and other elements are $\{a_1 + x_1, \dots, a_{i-1} + x_{i-1}, a_{i+1} + x_{i+1}, \dots, a_N + x_N\}$.

Proof. (\Rightarrow) By Definition 1, the joint symmetry implies that the following holds for any i ,

$$\Pr[X_1 - a_1 \leq x_1, \dots, X_i - a_i \leq x_i, \dots, X_N - a_N \leq x_N] = \Pr[X_1 - a_1 \leq x_1, \dots, a_i - X_i \leq x_i, \dots, X_N - a_N \leq x_N] \tag{32}$$

and with a simple calculation, the right hand side of equation (32) is written as

$$\begin{aligned}
& \Pr [X_1 - a_1 \leq x_1, \dots, a_i - X_i \leq x_i, \dots, X_N - a_N \leq x_N] \tag{33} \\
&= \Pr [X_1 - a_1 \leq x_1, \dots, X_i \leq \infty, \dots, X_N - a_N \leq x_N] - \Pr [X_1 - a_1 \leq x_1, \dots, X_i \leq a_i - x_i, \dots, X_N - a_N \leq x_N] \\
&= \mathbf{F}(a_1 + x_1, \dots, \infty, \dots, a_N + x_N) - \mathbf{F}(a_1 + x_1, \dots, a_i - x_i, \dots, a_N + x_N)
\end{aligned}$$

and the left hand side of equation (32) is

$$\Pr [X_1 - a_1 \leq x_1, \dots, X_i - a_i \leq x_i, \dots, X_N - a_N \leq x_N] = \mathbf{F}(a_1 + x_1, \dots, a_i + x_i, \dots, a_N + x_N)$$

(\Leftrightarrow) Equation (31) can be written as

$$\begin{aligned}
& \Pr [X_1 - a_1 \leq x_1, \dots, X_i - a_i \leq x_i, \dots, X_N - a_N \leq x_N] \\
&= \Pr [X_1 - a_1 \leq x_1, \dots, X_i \leq \infty, \dots, X_N - a_N \leq x_N] - \Pr [X_1 - a_1 \leq x_1, \dots, X_i \leq a_i - x_i, \dots, X_N - a_N \leq x_N] \quad \forall i
\end{aligned}$$

and by equation (33), the right hand side becomes $\Pr [X_1 - a_1 \leq x_1, \dots, a_i - X_i \leq x_i, \dots, X_N - a_N \leq x_N]$.

Therefore

$$\Pr [X_1 - a_1 \leq x_1, \dots, X_i - a_i \leq x_i, \dots, X_N - a_N \leq x_N] = \Pr [X_1 - a_1 \leq x_1, \dots, a_i - X_i \leq x_i, \dots, X_N - a_N \leq x_N] \quad \forall i$$

and this satisfies the definition of joint symmetry. ■

Equation (31) provides a definition of joint symmetry for general CDFs. The corresponding definition for copulas is given below.

Definition 3 (Jointly symmetric copula) *A N -dimensional copula \mathbf{C} is jointly symmetric if it satisfies*

$$\mathbf{C}(u_1, \dots, u_i, \dots, u_N) = \mathbf{C}(u_1, \dots, 1, \dots, u_N) - \mathbf{C}(u_1, \dots, 1 - u_i, \dots, u_N) \quad \forall i \tag{34}$$

where $u_i \in [0, 1] \quad \forall i$. “ $\mathbf{C}(u_1, \dots, 1, \dots, u_N)$ ” and “ $\mathbf{C}(u_1, \dots, 1 - u_i, \dots, u_N)$ ” are taken to mean that the i^{th} element is 1 and $1 - u_i$, respectively, and other elements are $\{u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N\}$.

Lemma 4 *Consider two scalar random variables X_1 and X_2 , and some constant b_1 in \mathbb{R}^1 . If $(X_1 - b_1, X_2)$ and $(b_1 - X_1, X_2)$ have a common joint distribution, then $\text{Cov}[X_1, X_2] = 0$.*

Proof. $X_1 - b_1$ and $b_1 - X_1$ have the same marginal distribution and the same moments, so $E[X_1 - b_1] = E[b_1 - X_1] \Rightarrow E[X_1] = b_1$. The variables $(X_1 - b_1, X_2)$ and $(b_1 - X_1, X_2)$ also have the same moments,

so $E[(X_1 - b_1)X_2] = E[(b_1 - X_1)X_2] \Rightarrow E[X_1X_2] = b_1E[X_2]$. Thus the covariance of X_1 and X_2 is $Cov[X_1, X_2] = E[X_1X_2] - E[X_1]E[X_2] = 0$. ■

Proof of Lemma 1. Joint symmetry of (X_i, X_j) around (b_i, b_j) , for $i \neq j$, is sufficient for Lemma 4 to hold. This is true for all pairs (i, j) of elements of the vector \mathbf{X} , and so $Corr[\mathbf{X}] = I$. ■

Proof of Lemma 2. (\Rightarrow) We follow Lemma 3 and rewrite equation (31) as

$$\begin{aligned} & \mathbf{C}(F_1(a_1 + x_1), \dots, F_i(a_i + x_i), \dots, F_N(a_N + x_N)) \\ &= \mathbf{C}(F_1(a_1 + x_1), \dots, 1, \dots, F_N(a_N + x_N)) - \mathbf{C}(F_1(a_1 + x_1), \dots, F_i(a_i - x_i), \dots, F_N(a_N + x_N)) \quad \forall i \end{aligned}$$

and we know $F_i(a_i + x_i) = 1 - F_i(a_i - x_i)$ due to the assumption of the symmetry of each X_i . Therefore,

$$\mathbf{C}(u_1, \dots, u_i, \dots, u_N) = \mathbf{C}(u_1, \dots, 1, \dots, u_N) - \mathbf{C}(u_1, \dots, 1 - u_i, \dots, u_N) \quad \forall i$$

where $u_i \equiv F_i(a_i + x_i)$.

(\Leftarrow) Following the reverse steps to above, equation (34) becomes equation (31), and the proof is done by Lemma 3. ■

Proof of Theorem 1. We seek to show that \mathbf{C}^{JS} in equation (6) satisfies equation (34), i.e.:

$$\mathbf{C}^{JS}(u_1, \dots, u_i, \dots, u_N) = \mathbf{C}^{JS}(u_1, \dots, 1, \dots, u_N) - \mathbf{C}^{JS}(u_1, \dots, 1 - u_i, \dots, u_N) \quad \forall i$$

We first show this equality for $i = N$. Re-write equation (6) as

$$\mathbf{C}^{JS}(u_1, \dots, u_N) = \frac{1}{2^N} [\mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, u_N) - \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1 - u_N) + \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1)]$$

$$\begin{aligned} \text{where } \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, u_N) &= \sum_{k_1=0}^2 \cdots \sum_{k_{N-1}=0}^2 (-1)^{R_{(-N)}} \cdot \mathbf{C}(\tilde{u}_1, \dots, \tilde{u}_{N-1}, u_N) \\ R_{(-N)} &\equiv \sum_{i=1}^{N-1} 1\{k_i = 2\} \text{ and } \tilde{u}_i = \begin{cases} 1 & \text{for } k_i = 0 \\ u_i & \text{for } k_i = 1 \\ 1 - u_i & \text{for } k_i = 2 \end{cases} \end{aligned}$$

Then similarly re-write $\mathbf{C}^{JS}(u_1, \dots, u_{N-1}, 1)$ and $\mathbf{C}^{JS}(u_1, \dots, u_{N-1}, 1 - u_N)$ to obtain:

$$\begin{aligned}
& \mathbf{C}^{JS}(u_1, \dots, u_{N-1}, 1) - \mathbf{C}^{JS}(u_1, \dots, u_{N-1}, 1 - u_N) \\
= & \frac{1}{2^N} \left[\mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1) - \underbrace{\mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 0)}_{=0} + \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1) \right] \\
& - \frac{1}{2^N} [\mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1 - u_N) - \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, u_N) + \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1)] \\
= & \frac{1}{2^N} [\mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, u_N) - \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1 - u_N) + \mathbf{C}_{(-N)}(u_1, \dots, u_{N-1}, 1)] \\
= & \mathbf{C}^{JS}(u_1, \dots, u_N)
\end{aligned}$$

This equation holds similarly for all $i = 1, \dots, N - 1$, so the proof is done. ■

Proof of Proposition 1. Let $\{\mathbf{F}_s\}_{s=1}^S$ be a collection of distributions jointly symmetric around a . By Lemma 2 this implies

$$\mathbf{F}(a + x_1, \dots, a + x_i, \dots, a + x_N) = \mathbf{F}(a + x_1, \dots, \infty, \dots, a + x_N) - \mathbf{F}(a + x_1, \dots, a - x_i, \dots, a + x_N) \quad \forall i$$

Next let $\mathbf{G}(\mathbf{x}) \equiv \sum_{s=1}^S \omega_s \mathbf{F}_s(\mathbf{x})$. Then

$$\begin{aligned}
& \mathbf{G}(a + x_1, \dots, a + x_i, \dots, a + x_N) \\
= & \sum_{s=1}^S \omega_s \mathbf{F}_s(a + x_1, \dots, a + x_i, \dots, a + x_N) \\
= & \sum_{s=1}^S \omega_s \mathbf{F}(a + x_1, \dots, \infty, \dots, a + x_N) - \sum_{s=1}^S \omega_s \mathbf{F}(a + x_1, \dots, a - x_i, \dots, a + x_N) \quad \forall i \\
\equiv & \mathbf{G}(a + x_1, \dots, \infty, \dots, a + x_N) - \mathbf{G}(a + x_1, \dots, a - x_i, \dots, a + x_N) \quad \forall i
\end{aligned}$$

and thus \mathbf{G} is jointly symmetric around a by Lemma 2. Claim (i) of the proposition is proved by noting that joint symmetry reduces to (univariate) symmetry when $N = 1$. Claim (ii) is proven by noting that if $\{\mathbf{F}_s\}_{s=1}^S$ all have $Unif(0, 1)$ marginal distributions then they are all jointly symmetric copulas, and as convex combinations of copulas are copulas, \mathbf{G} is then a jointly symmetric copula. ■

To prove Theorem 2, we need the following lemma. Below we use M to denote the number of daily observations for the DCC model, and the total number of intra-daily observations for the HAR-type model.

Lemma 5 *If $\text{rank}[\mathbf{y}_1, \dots, \mathbf{y}_M] \geq N$, where $\mathbf{y}_m \in \mathbb{R}^N$, then $\sum_{m=1}^M \mathbf{y}_m \mathbf{y}_m'$ is positive definite.*

Proof. Assume that $\sum_{m=1}^M \mathbf{y}_m \mathbf{y}_m'$ is only positive semi-definite. Then there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x}' \left(\sum_{m=1}^M \mathbf{y}_m \mathbf{y}_m' \right) \mathbf{x} = 0$, and this implies $\mathbf{x}' \cdot \mathbf{y}_m = 0$ for any m . On the other hand, if

$\text{rank}[\mathbf{y}_1, \dots, \mathbf{y}_M] \geq N$, then $[\mathbf{y}_1, \dots, \mathbf{y}_M]$ span \mathbb{R}^N , which implies there exist $\{\alpha_i\}_{i=1}^M$ such that

$$\alpha_1 \mathbf{y}_1 + \dots + \alpha_M \mathbf{y}_M = \mathbf{x}.$$

$$\text{Premultiplying by } \mathbf{x}' \text{ gives } \alpha_1 \mathbf{x}' \mathbf{y}_1 + \dots + \alpha_M \mathbf{x}' \mathbf{y}_M = \mathbf{x}' \mathbf{x}.$$

The left hand side is zero since $\mathbf{x}' \cdot \mathbf{y}_m = 0$ for any m , which contradicts that \mathbf{x} is a nonzero vector. Therefore, $\sum_{m=1}^M \mathbf{y}_m \mathbf{y}_m'$ is positive definite. ■

Proof of Theorem 2. First we note that equation (10) is guaranteed to yield a positive variance forecast, and so the diagonal matrix of variance forecasts is positive definite. Given the fact that if matrices U and V are positive definite, then UVU is also positive definite, we just need to establish positive definiteness of the correlation matrix forecast \widehat{RCorr}_t^Δ below. Substituting $(\hat{a}, \hat{b}, \hat{c})$ for (a, b, c) in equation (11), and undoing the “vech” operation, we obtain \widehat{RCorr}_t^Δ :

$$\widehat{RCorr}_t^\Delta = \overline{RCorr}_T^\Delta (1 - \hat{a} - \hat{b} - \hat{c}) + \hat{a} \cdot RCorr_{t-1}^\Delta + \hat{b} \cdot \frac{1}{4} \sum_{k=2}^5 RCorr_{t-k}^\Delta + \hat{c} \cdot \frac{1}{15} \sum_{k=6}^{20} RCorr_{t-k}^\Delta$$

The first term is positive definite since $\overline{RCorr}_T^\Delta$ is positive definite by Lemma 5 if the number of days multiplied by the intra-day frequency is greater than N and $(1 - \hat{a} - \hat{b} - \hat{c})$ is greater than zero by assumption. The other three terms are positive semi-definite by the positive semi-definiteness of realized correlation and the assumption that $[\hat{a}, \hat{b}, \hat{c}] \geq 0$. Thus \widehat{RCorr}_t^Δ is positive definite. ■

Proof of Proposition 2. We obtain the marginal copula density by first obtaining the marginal copula CDF. Recall that a the (i, j) bivariate copula CDF implied by an N -dimensional copula CDF is obtained by setting all arguments of the original copula to 1, except i and j :

$$\mathbf{C}_{ij}(u_i, u_j) = \mathbf{C}(1, \dots, 1, u_i, u_j, 1, \dots, 1) \tag{35}$$

For jointly symmetric copulas generated using equation (6) this implies

$$\begin{aligned} \mathbf{C}_{ij}^{JS}(u_i, u_j) &= \mathbf{C}^{JS}(1, \dots, 1, u_i, u_j, 1, \dots, 1) \\ &= \frac{1}{2^N} \sum_{j_1=0}^2 \dots \sum_{j_N=0}^2 (-1)^R \cdot \mathbf{C}(\tilde{u}_1, \dots, \tilde{u}_N) \\ &= \frac{1}{2^N} \sum_{k_1=0}^1 \dots \sum_{k_i=0}^2 \sum_{k_j=0}^2 \dots \sum_{k_N=0}^1 (-1)^R \cdot \mathbf{C}(\tilde{u}_1, \dots, \tilde{u}_N) \end{aligned}$$

since $u_m = 1 \forall m \notin \{i, j\}$, and so $\tilde{u}_m = 0$ whenever $k_m = 2$, and $\mathbf{C} = 0$ whenever any one of its arguments

is equal to zero. Then

$$\mathbf{C}_{ij}^{JS}(u_i, u_j) = \frac{1}{2^N} 2^{N-2} \sum_{k_i=0}^2 \sum_{k_j=0}^2 (-1)^R \cdot \mathbf{C}_{ij}(\tilde{u}_i, \tilde{u}_j) = \frac{1}{4} \sum_{k_i=0}^2 \sum_{k_j=0}^2 (-1)^R \cdot \mathbf{C}_{ij}(\tilde{u}_i, \tilde{u}_j)$$

since $\tilde{u}_m = 1$ for $k_m = 0$ or $k_m = 1$, for all $m \notin \{i, j\}$. Expanding above we obtain

$$\mathbf{C}_{ij}^{JS}(u_i, u_j) = \frac{1}{4} \{2u_i + 2u_j - 1 + \mathbf{C}_{ij}(u_i, u_j) - \mathbf{C}_{ij}(u_i, 1 - u_j) - \mathbf{C}_{ij}(1 - u_i, u_j) + \mathbf{C}_{ij}(1 - u_i, 1 - u_j)\},$$

and then taking the second cross partial derivative we find

$$\mathbf{c}_{ij}^{JS}(u_i, u_j) \equiv \frac{\partial^2 \mathbf{C}_{ij}^{JS}(u_i, u_j)}{\partial u_i \partial u_j} = \frac{1}{4} \{\mathbf{c}_{ij}(u_i, u_j) + \mathbf{c}_{ij}(u_i, 1 - u_j) + \mathbf{c}_{ij}(1 - u_i, u_j) + \mathbf{c}_{ij}(1 - u_i, 1 - u_j)\}$$

as claimed. ■

Proof of Theorem 3. By applying $\log(y) \leq y - 1$ to $\frac{\mathbf{h}_i(Z_i, Z_{i+1})}{\mathbf{g}_i(Z_i, Z_{i+1})}$, the following is shown:

$$\begin{aligned} \sum_{i=1}^{N-1} E_{g(\mathbf{z})} \left[\log \frac{\mathbf{h}_i(Z_i, Z_{i+1})}{\mathbf{g}_i(Z_i, Z_{i+1})} \right] &\leq \sum_{i=1}^{N-1} \left(E_{\mathbf{g}(\mathbf{z})} \left[\frac{\mathbf{h}_i(Z_i, Z_{i+1})}{\mathbf{g}_i(Z_i, Z_{i+1})} \right] - 1 \right) \\ &= \sum_{i=1}^{N-1} \left(E_{\mathbf{g}_i(z_i, z_{i+1})} \left[\frac{\mathbf{h}_i(Z_i, Z_{i+1})}{\mathbf{g}_i(Z_i, Z_{i+1})} \right] - 1 \right) \\ &\equiv \sum_{i=1}^{N-1} \left(\int \mathbf{g}_i(z_i, z_{i+1}) \frac{\mathbf{h}_i(z_i, z_{i+1})}{\mathbf{g}_i(z_i, z_{i+1})} dz_i dz_{i+1} - 1 \right) = 0 \end{aligned}$$

where the second line holds since only submodel for (Z_i, Z_{i+1}) is needed to evaluate the above expectation, and the third line holds since \mathbf{h}_i is a valid density. Thus $E \left[\sum_{i=1}^{N-1} \log \mathbf{h}_i(Z_i, Z_{i+1}) \right] \leq E \left[\sum_{i=1}^{N-1} \log \mathbf{g}_i(Z_i, Z_{i+1}) \right]$ as claimed. ■

References

- [1] Andersen, T. G., T. Bollerslev, P. F. Christoffersen, F. X. Diebold, 2006, Volatility and Correlation Forecasting, in G. Elliott, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 1, 779-879.
- [2] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, Modeling and forecasting realised volatility, *Econometrica*, 71, 579-625.
- [3] Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard, 2009, Realized kernels in practice: trades and quotes, *Econometrics Journal*, 12, 1-32.
- [4] Barndorff-Nielsen, O. and N. Shephard, 2004, Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics, *Econometrica*, 72, 885-925.
- [5] Bauer, G. H. and K. Vorkink, 2011, Forecasting multivariate realized stock market volatility, *Journal of Econometrics*, 160, 93-101.
- [6] Bollerslev, T., R. F. Engle, and J. M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy*, 96, 116-131.
- [7] Bollerslev, T., S. Z. Li, and V. Todorov, 2014, Roughing up Beta: Continuous vs Discontinuous Betas, and the Cross-Section of Expected Stock Returns, working paper, Department of Economics, Duke University.
- [8] Chiriac, R. and V. Voev, 2011, Modelling and forecasting multivariate realized volatility, *Journal of Applied Econometrics*, 26, 922-947.
- [9] Christoffersen, P., K. Jacobs, X. Jin and H. Langlois, 2013, Dynamic dependence in corporate credit, working paper, Bauer College of Business, University of Houston.
- [10] Corsi, F., 2009, A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, 7, 174-196.
- [11] Cox, D. R. and N. Reid, 2004, A note on pseudolikelihood constructed from marginal densities, *Biometrika*, 91, 729-737.
- [12] Creal, D., S. J. Koopman, and A. Lucas, 2013, Generalized autoregressive score models with applications, *Journal of Applied Econometrics*, 28, 777-795.
- [13] Creal, D. and R. Tsay, 2014, High-dimensional dynamic stochastic copula models, Working paper, University of Chicago.
- [14] Engle, R. F., 2002, Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, 20, 339-351.
- [15] Engle, R. F. and K. F. Kroner, 1995, Multivariate Simultaneous Generalized ARCH, *Econometric Theory*, 11(1), 122-150.

- [16] Engle, R. F., N. Shephard, and K. Sheppard, 2008, Fitting and testing vast dimensional time-varying covariance models, Working paper, Oxford-Man Institute, University of Oxford.
- [17] Giacomini, R. and H. White, 2006, Tests of conditional predictive ability, *Econometrica*, 74, 1545-1578.
- [18] Glosten, R. T., R. Jagannathan, and D. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1801.
- [19] Gneiting, T. and A. Raftery, 2007, Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102, 359-378.
- [20] Gonçalves, S., U. Hounyo, A.J. Patton and K. Sheppard, 2013, Bootstrapping two-stage extremum estimators, working paper, Oxford-Man Institute of Quantitative Finance.
- [21] Gouriéroux, C. and A. Monfort, 1996, *Statistics and Econometric Models, Volume 2*, translated from the French by Q. Vuong, Cambridge University Press, Cambridge.
- [22] Hautsch, N., L. M. Kyj, and P. Malec, 2013, Do high-frequency data improve high-dimensional portfolio allocations?, *Journal of Applied Econometrics*, forthcoming.
- [23] Jin, X. and J. M. Maheu, 2013, Modeling realized covariances and returns, *Journal of Financial Econometrics*, 11, 335-369.
- [24] Jondeau, E. and M. Rockinger, 2012, On the importance of time variability in higher moments for asset allocation, *Journal of Financial Econometrics*, 10, 84-123.
- [25] Lee, T. and X. Long, 2009, Copula-based multivariate GARCH model with uncorrelated dependent errors, *Journal of Econometrics*, 150, 207-218.
- [26] Lindsay, B., 1988, Composite likelihood methods, *Statistical Inference from Stochastic Processes*, American Mathematical Society, 221-239
- [27] Maheu, J. M. and T. H. McCurdy, 2011, Do high-frequency measures of volatility improve forecasts of return distributions?, *Journal of Econometrics*, 160, 69-76.
- [28] Nelsen, R. B., 2006, *An Introduction to Copulas*, Second Edition, Springer, U.S.A.
- [29] Newey, W. K. and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, in R.Engle and D. McFadden (eds.), *Handbook of Econometrics*, 4, 2111-2245.
- [30] Newey, W. K. and K. D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.
- [31] Noureldin, D., N. Shephard, and K. Sheppard, 2012, Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics*, 27, 907-933.
- [32] Oh, D.-H. and A. J. Patton, 2013, Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads, Working paper, Duke University.

- [33] Patton, A. J., 2012, Copula methods for forecasting multivariate time series, in G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2, ch16.
- [34] Politis, D. N., and J. P. Romano, 1994, The Stationary bootstrap, *Journal of the American Statistical Association*, 89, 1303-1313.
- [35] Rivers, D. and Q. H. Vuong, 2002, Model selection tests for nonlinear dynamic models, *Econometrics Journal*, 5, 1-39.
- [36] Shephard, N., 2005, *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.
- [37] Sklar, A., 1959, Fonctions de repartition a n dimensions et leurs marges. Publications de Institut Statistique de Universite de Paris 8, 229-231.
- [38] Varin, C., N. Reid, and D. Firth, 2011, An overview of composite likelihood methods, *Statistica Sinica*, 21, 5-42.
- [39] Varin, C. and P. Vidoni, 2005, A note on composite likelihood inference and model selection, *Biometrika*, 92, 519-528.
- [40] Vuong, Q. H., 1989, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57, 307-333.
- [41] White, H., 1994, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs No. 22, Cambridge University Press, Cambridge, U.K.

Table 1: Computation times for jointly symmetric copulas

| N | 10 | 20 | 30 | 50 | 100 |
|---|-----------|-----------|-----------|--------------|-----------------|
| Full likelihood | 0.23 sec | 4 min | 70 hours | 10^6 years | 10^{17} years |
| Composite likelihood using all pairs | 0.05 sec | 0.21 sec | 0.45 sec | 1.52 sec | 5.52 sec |
| Composite likelihood using adjacent pairs | 0.01 sec | 0.02 sec | 0.03 sec | 0.06 sec | 0.11 sec |
| Composite likelihood using first pair | 0.001 sec | 0.001 sec | 0.001 sec | 0.001 sec | 0.001 sec |

Note: Computation times for one evaluation of the density of jointly symmetric copula based on the Clayton copula. These times are based actual computation times for a single evaluation of an N -dimension Clayton copula, multiplied by the number of rotations required to obtain the jointly symmetric copula likelihood (2^N) or composite likelihood based on all pairs ($2N(N-1)$), adjacent pairs ($4(N-1)$), or a single pair (4).

Table 2: Simulation results for a jointly symmetric copula based on the Clayton copula

| N | Bias | | | | Std dev | | | | Average Run Time (in sec) | | | |
|-----|---------|--------------|--------------|----------------|---------|--------------|--------------|----------------|---------------------------|--------------|--------------|----------------|
| | MLE | $MCLE_{all}$ | $MCLE_{adj}$ | $MCLE_{first}$ | MLE | $MCLE_{all}$ | $MCLE_{adj}$ | $MCLE_{first}$ | MLE | $MCLE_{all}$ | $MCLE_{adj}$ | $MCLE_{first}$ |
| 2 | -0.0027 | -0.0027 | -0.0027 | -0.0027 | 0.1176 | 0.1176 | 0.1176 | 0.1176 | 0.12 | 0.12 | 0.12 | 0.12 |
| 3 | -0.0019 | -0.0028 | -0.0031 | -0.0027 | 0.0798 | 0.0839 | 0.0917 | 0.1176 | 0.42 | 0.50 | 0.24 | 0.12 |
| 5 | -0.0014 | -0.0022 | -0.0016 | -0.0027 | 0.0497 | 0.0591 | 0.0713 | 0.1176 | 1.96 | 1.49 | 0.43 | 0.12 |
| 10 | -0.0051 | -0.0047 | -0.0039 | -0.0027 | 0.0293 | 0.0402 | 0.0495 | 0.1176 | 116 | 7 | 1 | 0.12 |
| 20 | -0.0018 | -0.0018 | -0.0021 | -0.0027 | 0.0365 | 0.0365 | 0.0405 | 0.1176 | 2×10^5 | 27 | 2 | 0.12 |
| 30 | -0.0036 | -0.0036 | -0.0037 | -0.0027 | 0.0336 | 0.0336 | 0.0379 | 0.1176 | 3×10^8 | 63 | 3 | 0.12 |
| 40 | -0.0028 | -0.0028 | -0.0037 | -0.0027 | 0.0311 | 0.0311 | 0.0341 | 0.1176 | 4×10^{11} | 117 | 5 | 0.12 |
| 50 | -0.0011 | -0.0011 | -0.0014 | -0.0027 | 0.0298 | 0.0298 | 0.0329 | 0.1176 | 5×10^{14} | 192 | 6 | 0.12 |
| 60 | -0.0007 | -0.0007 | -0.0006 | -0.0027 | 0.0314 | 0.0314 | 0.0332 | 0.1176 | 7×10^{17} | 256 | 7 | 0.12 |
| 70 | -0.0013 | -0.0013 | -0.0013 | -0.0027 | 0.0306 | 0.0306 | 0.0324 | 0.1176 | 8×10^{20} | 364 | 8 | 0.12 |
| 80 | -0.0039 | -0.0039 | -0.0041 | -0.0027 | 0.0309 | 0.0309 | 0.0332 | 0.1176 | 9×10^{23} | 471 | 9 | 0.12 |
| 90 | 0.0012 | 0.0012 | 0.0013 | -0.0027 | 0.0312 | 0.0312 | 0.0328 | 0.1176 | 9×10^{26} | 611 | 11 | 0.12 |
| 100 | -0.0006 | -0.0006 | -0.0003 | -0.0027 | 0.0290 | 0.0290 | 0.0305 | 0.1176 | 1×10^{30} | 748 | 12 | 0.12 |

Note: This table presents the results from 500 simulations of jointly symmetric copula based on Clayton copula with true parameter 1. The sample size is $T = 1000$. Four different estimation methods are used: MLE, MCLE with all pairs, MCLE with adjacent pairs, MCLE with the first pair. MLE is infeasible for $N > 10$ and so no results are reported in those cases. The first four columns report the average difference between the estimated parameter and its true value. The next four columns are the standard deviation in the estimated parameters. The last four columns present average run time of each estimation method. The reported run times for MLE for $N > 10$ are based on actual single function evaluation times and on an assumption of 40 function evaluations to reach the optimum.

Table 3: Simulation results for multi-stage estimation

| | Variance | | | Correlation | | Marginal | Copula |
|------------|----------|------------|-------------|--------------|-------------|----------|------------|
| | Const | ARCH | GARCH | DCC α | DCC β | t dist | JS Clayton |
| | ψ_i | κ_i | λ_i | α | β | ν_i | φ |
| True value | 0.05 | 0.10 | 0.85 | 0.02 | 0.95 | 6.00 | 1.00 |
| $N = 10$ | | | | | | | |
| Bias | 0.0123 | 0.0007 | -0.0162 | -0.0012 | -0.0081 | 0.1926 | -0.0122 |
| Std | 0.0442 | 0.0387 | 0.0717 | 0.0060 | 0.0277 | 1.1023 | 0.0650 |
| Median | 0.0536 | 0.0959 | 0.8448 | 0.0184 | 0.9459 | 5.9837 | 0.9920 |
| 90% | 0.1027 | 0.1478 | 0.9015 | 0.0263 | 0.9631 | 7.5215 | 1.0535 |
| 10% | 0.0271 | 0.0580 | 0.7619 | 0.0119 | 0.9196 | 5.0559 | 0.9165 |
| 90-10 Diff | 0.0756 | 0.0898 | 0.1397 | 0.0144 | 0.0435 | 2.4656 | 0.1370 |
| Time/rep. | 54 sec | | | | | | |
| $N = 50$ | | | | | | | |
| Bias | 0.0114 | 0.0012 | -0.0149 | -0.0018 | -0.0051 | 0.1880 | -0.0136 |
| Std | 0.0411 | 0.0412 | 0.0687 | 0.0040 | 0.0111 | 1.0936 | 0.0390 |
| Median | 0.0529 | 0.0958 | 0.8454 | 0.0179 | 0.9458 | 6.0000 | 0.9880 |
| 90% | 0.1019 | 0.1499 | 0.9025 | 0.0234 | 0.9580 | 7.5223 | 1.0312 |
| 10% | 0.0268 | 0.0567 | 0.7615 | 0.0135 | 0.9313 | 5.0454 | 0.9413 |
| 90-10 Diff | 0.0751 | 0.0931 | 0.1410 | 0.0098 | 0.0267 | 2.4769 | 0.0899 |
| Time/rep. | 138 sec | | | | | | |
| $N = 100$ | | | | | | | |
| Bias | 0.0119 | 0.0017 | -0.0158 | -0.0020 | -0.0041 | 0.1813 | -0.0133 |
| Std | 0.0419 | 0.0404 | 0.0691 | 0.0034 | 0.0094 | 1.0748 | 0.0362 |
| Median | 0.0533 | 0.0966 | 0.8440 | 0.0177 | 0.9467 | 6.0002 | 0.9886 |
| 90% | 0.1025 | 0.1504 | 0.9022 | 0.0223 | 0.9566 | 7.4963 | 1.0244 |
| 10% | 0.0270 | 0.0576 | 0.7607 | 0.0139 | 0.9337 | 5.0492 | 0.9432 |
| 90-10 Diff | 0.0756 | 0.0928 | 0.1415 | 0.0084 | 0.0229 | 2.4471 | 0.0811 |
| Time/rep. | 329 sec | | | | | | |

Note: This table presents the results from 500 simulations of multi-stage estimation of the model described in Section 3.3. The sample size is $T = 1000$ and cross-sectional dimensions are $N = 10, 50, \text{ and } 100$. The first row of each panel presents the average difference between the estimated parameter and its true value. The second row presents the standard deviation in the estimated parameters. The third, fourth and fifth rows present the 50th, 90th and 10th percentiles of the distribution of estimated parameters, and the final row presents the difference between the 90th and 10th percentiles.

Table 4: Summary statistics and conditional mean estimates

| | <i>Cross-sectional distribution</i> | | | | | |
|--|-------------------------------------|---------|---------|---------|---------|------------------------|
| | Mean | 5% | 25% | Median | 75% | 95% |
| <i>Panel A: Summary statistics</i> | | | | | | |
| Mean | 0.0002 | -0.0006 | 0.0001 | 0.0002 | 0.0004 | 0.0006 |
| Std dev | 0.0219 | 0.0120 | 0.0159 | 0.0207 | 0.0257 | 0.0378 |
| Skewness | -0.0693 | -0.6594 | -0.3167 | -0.0318 | 0.1823 | 0.5642 |
| Kurtosis | 11.8559 | 6.9198 | 8.4657 | 10.4976 | 13.3951 | 20.0200 |
| Corr | 0.4666 | 0.3294 | 0.4005 | 0.4580 | 0.5230 | 0.6335 |
| <i>Panel B: Conditional mean</i> | | | | | | |
| Constant | 0.0002 | -0.0006 | 0.0000 | 0.0002 | 0.0004 | 0.0006 |
| AR(1) | -0.0535 | -0.1331 | -0.0794 | -0.0553 | -0.0250 | 0.0105 |
| <i>Panel C: Test for skewness, kurtosis, and correlation</i> | | | | | | |
| | | | | | | <i># of rejections</i> |
| $H_0 : Skew [r_{it}] = 0$ | | | | | | 3 out of 104 |
| $H_0 : Kurt [r_{it}] = 3$ | | | | | | 104 out of 104 |
| $H_0 : Corr [r_{it}, r_{jt}] = 0$ | | | | | | 5356 out of 5356 |

Note: Panel A presents summary statistics on the daily equity returns used in the empirical analysis. The columns present the mean and quantiles from the cross-sectional distribution of the measures listed in the rows. Panel B presents the parameter estimates for AR(1) models of the conditional means of returns. Panel C shows the number of rejections at the 5% level for tests of zero skewness, zero excess kurtosis, and zero cross-correlation for the 104 stocks under 5% level. (The total number of pairs of stocks is 5356.)

Table 5: Conditional covariance model parameter estimates

Panel A: HAR-type models based on 5-min returns

| | <i>Cross-sectional distribution</i> | | | | | |
|------------------------------|-------------------------------------|---------|---------|---------|--------|--------|
| | Mean | 5% | 25% | Median | 75% | 95% |
| Variance models | | | | | | |
| Constant $\phi_i^{(const)}$ | -0.0019 | -0.0795 | -0.0375 | -0.0092 | 0.0207 | 0.1016 |
| HAR day $\phi_i^{(day)}$ | 0.3767 | 0.3196 | 0.3513 | 0.3766 | 0.3980 | 0.4414 |
| HAR week $\phi_i^{(week)}$ | 0.3105 | 0.2296 | 0.2766 | 0.3075 | 0.3473 | 0.3896 |
| HAR month $\phi_i^{(month)}$ | 0.2190 | 0.1611 | 0.1959 | 0.2146 | 0.2376 | 0.2962 |
| | Est | Std Err | | | | |
| Correlation model | | | | | | |
| HAR day (a) | 0.1224 | 0.0079 | | | | |
| HAR week (b) | 0.3156 | 0.0199 | | | | |
| HAR month (c) | 0.3778 | 0.0326 | | | | |

Panel B: DCC models based on daily returns

| | | | | | | |
|-------------------------------|--------|---------|--------|--------|--------|--------|
| Variance models | | | | | | |
| Constant $\psi_i \times 10^4$ | 0.0864 | 0.0190 | 0.0346 | 0.0522 | 0.0811 | 0.2781 |
| ARCH κ_i | 0.0252 | 0.0000 | 0.0079 | 0.0196 | 0.0302 | 0.0738 |
| Asym ARCH ζ_i | 0.0840 | 0.0298 | 0.0570 | 0.0770 | 0.1015 | 0.1535 |
| GARCH λ_i | 0.9113 | 0.8399 | 0.9013 | 0.9228 | 0.9363 | 0.9573 |
| | Est | Std Err | | | | |
| Correlation model | | | | | | |
| DCC ARCH (α) | 0.0245 | 0.0055 | | | | |
| DCC GARCH (β) | 0.9541 | 0.0119 | | | | |

Note: Panel A presents summaries of the estimated HAR-type models described in Section 2.2 using 5-minute returns. Panel B presents summaries of the estimated GJR-GARCH–DCC models using daily returns. The parameter estimates for variance models are summarized in the mean and quantiles from the cross-sectional distributions of the estimates. The estimates for correlation parts are reported with bootstrap standard errors which reflect accumulated estimation errors from former stages.

Table 6: Summary statistics and marginal distributions for the standardized residuals

| | <i>Cross-sectional distribution</i> | | | | | |
|--|-------------------------------------|----------------|---------|---------|---------|---------|
| | Mean | 5% | 25% | Median | 75% | 95% |
| <i>Panel A: HAR standardized residuals</i> | | | | | | |
| Mean | 0.0023 | -0.0122 | -0.0042 | 0.0016 | 0.0076 | 0.0214 |
| Std dev | 1.0921 | 0.9647 | 1.0205 | 1.0822 | 1.1423 | 1.2944 |
| Skewness | -0.1613 | -1.5828 | -0.4682 | -0.0837 | 0.3420 | 0.7245 |
| Kurtosis | 13.1220 | 5.0578 | 6.8422 | 9.8681 | 16.0303 | 32.7210 |
| Correlation | 0.0026 | -0.0445 | -0.0167 | 0.0020 | 0.0209 | 0.0502 |
| <i>Panel B: GARCH-DCC standardized residuals</i> | | | | | | |
| Mean | 0.0007 | -0.0155 | -0.0071 | 0.0004 | 0.0083 | 0.0208 |
| Std dev | 1.1871 | 1.1560 | 1.1737 | 1.1859 | 1.2002 | 1.2240 |
| Skewness | -0.1737 | -1.4344 | -0.5293 | -0.0307 | 0.2628 | 0.7920 |
| Kurtosis | 12.6920 | 5.0815 | 6.7514 | 10.1619 | 15.9325 | 28.8275 |
| Correlation | -0.0011 | -0.0172 | -0.0073 | -0.0008 | 0.0053 | 0.0145 |
| <i>Panel C: Marginal t distribution parameter estimates</i> | | | | | | |
| HAR | 5.3033 | 4.1233 | 4.7454 | 5.1215 | 5.8684 | 6.8778 |
| DCC | 6.0365 | 4.2280 | 5.0314 | 5.9042 | 7.0274 | 8.2823 |
| <i>Panel D: Test for skewness, kurtosis, and correlation</i> | | | | | | |
| | # of rejections | | | | | |
| | HAR | DCC | | | | |
| $H_0 : Skew [e_{it}] = 0$ | 4 out of 104 | 6 out of 104 | | | | |
| $H_0 : Kurt [e_{it}] = 3$ | 104 out of 104 | 104 out of 104 | | | | |
| $H_0 : Corr [e_{it}, e_{jt}] = 0$ | 497 out of 5356 | 1 out of 5356 | | | | |

Note: Panel A presents summary statistics of the uncorrelated standardized residuals obtained from the HAR-type model, and Panel B presents corresponding results based on the GARCH-DCC model. Panel C presents the estimates of the parameters for the marginal distribution of standardized residuals, obtained from the two volatility models. Panel D reports the number of rejections, at the 5% level, for tests of zero skewness, zero excess kurtosis, and zero cross-correlation.

Table 7: Estimation results for the copula models

| | | Jointly symmetric copula models | | | | Benchmarks | |
|-----|-----------------|---------------------------------|--------------------|--------------------|--------------------|------------|---------------------------------|
| | | t | Clayton | Frank | Gumbel | Indep | MV t dist |
| HAR | Est. (s.e.) | 39.4435 (4.3541) | 0.0876 (0.0087) | 1.2652 (0.0942) | 1.0198 (0.0038) | - | 6.4326 [†] (0.1405) |
| | log L | -282491 | -282500 | -282512 | -282533 | -282578 | -284853 |
| | Rank | 1 | 2 | 3 | 4 | 5 | 6 |
| | t-test of indep | 8.45 | 10.07 | 13.43 | 5.25 | - | 45.72 |
| DCC | Est. (s.e.) | 28.2068 (5.4963) | 0.1139 (0.0155) | 1.5996 (0.1540) | 1.0312 (0.0071) | - | 7.0962 [†] (0.3586) |
| | log L | -289162 | -289190 | -289217 | -289255 | -289404 | -291607 |
| | Rank | 7 | 8 | 9 | 10 | 11 | 12 |
| | t-test of indep | 6.13 | 7.36 | 10.36 | 4.40 | - | 17.80 |

Note: This table presents the estimated parameters of four different jointly symmetric copula models based on t , Clayton, Frank, and Gumbel copulas, as well as the estimated parameter of the (standardized) multivariate t distribution as a benchmark model. The independence copula model has no parameter to estimate. Bootstrap standard errors are reported in parentheses. Also reported is the log-likelihood from the complete distribution model formed by combining the copula model with the HAR or DCC volatility model. (The MV t distribution is not based on a copula decomposition, but its joint likelihood may be compared with those from copula-based models.) The bottom row of each panel reports t -statistics for a test of no nonlinear dependence. [†]The parameter of the multivariate t distribution is not a copula parameter, but it is reported in this row for simplicity.

Table 8: t -statistics from in-sample model comparison tests

| | HAR models | | | | | | GARCH-DCC models | | | | |
|-------------------------|--------------------|----------|----------|----------|--------------------|--------|--------------------|----------|----------|----------|--------------------|
| | t^{JS} | C^{JS} | F^{JS} | G^{JS} | Indep | MV t | t^{JS} | C^{JS} | F^{JS} | G^{JS} | Indep |
| HAR models | | | | | | | | | | | |
| t^{JS} | – | | | | | | | | | | |
| C^{JS} | 2.92 | – | | | | | | | | | |
| F^{JS} | 2.16 | 1.21 | – | | | | | | | | |
| G^{JS} | 5.38 | 6.02 | 1.75 | – | | | | | | | |
| Indep* | 8.45 | 10.07 | 13.43 | 5.25 | – | | | | | | |
| MV t | 19.70 [†] | 19.52 | 19.45 | 19.23 | 18.40 [‡] | – | | | | | |
| GARCH-DCC models | | | | | | | | | | | |
| t^{JS} | 7.86 | 7.85 | 7.85 | 7.84 | 7.82 | 6.92 | – | | | | |
| C^{JS} | 7.86 | 7.86 | 7.85 | 7.85 | 7.83 | 6.93 | 4.48 | – | | | |
| F^{JS} | 7.85 | 7.85 | 7.84 | 7.83 | 7.82 | 6.91 | 2.69 | 1.27 | – | | |
| G^{JS} | 7.88 | 7.87 | 7.87 | 7.86 | 7.84 | 6.94 | 6.74 | 7.47 | 1.74 | – | |
| Indep* | 7.90 | 7.90 | 7.90 | 7.89 | 7.87 | 6.97 | 6.13 | 7.36 | 10.36 | 4.40 | – |
| MV t | 8.95 | 8.95 | 8.94 | 8.94 | 8.92 | 8.03 | 18.50 [†] | 18.11 | 17.94 | 17.60 | 15.69 [‡] |

Note: This table presents t -statistics from pair-wise Rivers and Vuong (2002) model comparison tests introduced in Section 3.2. A positive t -statistic indicates that the model above beat the model to the left, and a negative one indicates the opposite. t^{JS} , C^{JS} , F^{JS} , and G^{JS} stand for jointly symmetric copulas based on t , Clayton, Frank, and Gumbel copulas respectively. “Indep” is the independence copula. MV t is the multivariate Student’s t distribution. The upper panel includes results for models that use 5-min data and the HAR-type covariance model introduced in Section 2.2, the lower panel includes results for models based on a GARCH-DCC covariance model. *The comparisons of jointly symmetric copula-based models with the independence copula, reported in the penultimate row of the top panel, and the right half of the penultimate row of the lower panel, are nested comparisons and the Rivers and Vuong (2002) test does not apply. The t -statistics here are the same as those in Table 7. [†]The MV t density is nested in the density based on the jointly symmetric t copula, and so strictly the Rivers and Vuong (2002) test does not apply, however it is computationally infeasible to implement the formal nested test; we report the Rivers and Vuong t -statistic here for ease of reference. [‡]The MV t density and the density based on the independence copula are nested only at a single point, and we apply the Rivers and Vuong (2002) test here.

Table 9: t -statistics from out-of-sample model comparison tests

| | HAR models | | | | | | GARCH-DCC models | | | | | |
|-------------------------|------------|----------|----------|----------|-------|--------|------------------|----------|----------|----------|-------|--|
| | t^{JS} | C^{JS} | F^{JS} | G^{JS} | Indep | MV t | t^{JS} | C^{JS} | F^{JS} | G^{JS} | Indep | |
| HAR models | | | | | | | | | | | | |
| t^{JS} | – | | | | | | | | | | | |
| C^{JS} | 1.50 | – | | | | | | | | | | |
| F^{JS} | 0.89 | 0.44 | – | | | | | | | | | |
| G^{JS} | 2.88 | 3.09 | 1.21 | – | | | | | | | | |
| Indep | 2.57 | 2.60 | 2.34 | 1.84 | – | | | | | | | |
| MV t | 10.75 | 10.63 | 10.65 | 10.48 | 10.00 | – | | | | | | |
| GARCH-DCC models | | | | | | | | | | | | |
| t^{JS} | 5.23 | 5.23 | 5.23 | 5.23 | 5.22 | 4.55 | – | | | | | |
| C^{JS} | 5.23 | 5.23 | 5.23 | 5.23 | 5.22 | 4.55 | 1.55 | – | | | | |
| F^{JS} | 5.23 | 5.22 | 5.23 | 5.22 | 5.21 | 4.55 | 1.79 | 1.34 | – | | | |
| G^{JS} | 5.24 | 5.24 | 5.24 | 5.23 | 5.22 | 4.56 | 2.96 | 3.31 | 0.01 | – | | |
| Indep | 5.24 | 5.24 | 5.24 | 5.23 | 5.22 | 4.56 | 3.10 | 3.12 | 2.38 | 2.44 | – | |
| MV t | 6.05 | 6.05 | 6.05 | 6.05 | 6.04 | 5.41 | 14.65 | 14.33 | 14.56 | 13.88 | 12.80 | |

Note: This table presents t -statistics from pair-wise comparisons of the out-of-sample likelihoods of competing density forecasts based on the test of Giacomini and White (2006). A positive t -statistic indicates that the model above beat the model to the left, and a negative one indicates the opposite. t^{JS} , C^{JS} , F^{JS} , and G^{JS} stand for jointly symmetric copulas based on t , Clayton, Frank, and Gumbel copulas respectively. “Indep” is the independence copula. MV t is the multivariate Student’s t distribution. The upper panel includes results for models that use 5-min data and the HAR-type covariance model introduced in Section 2.2, the lower panel includes results for models based on a GARCH-DCC covariance model.

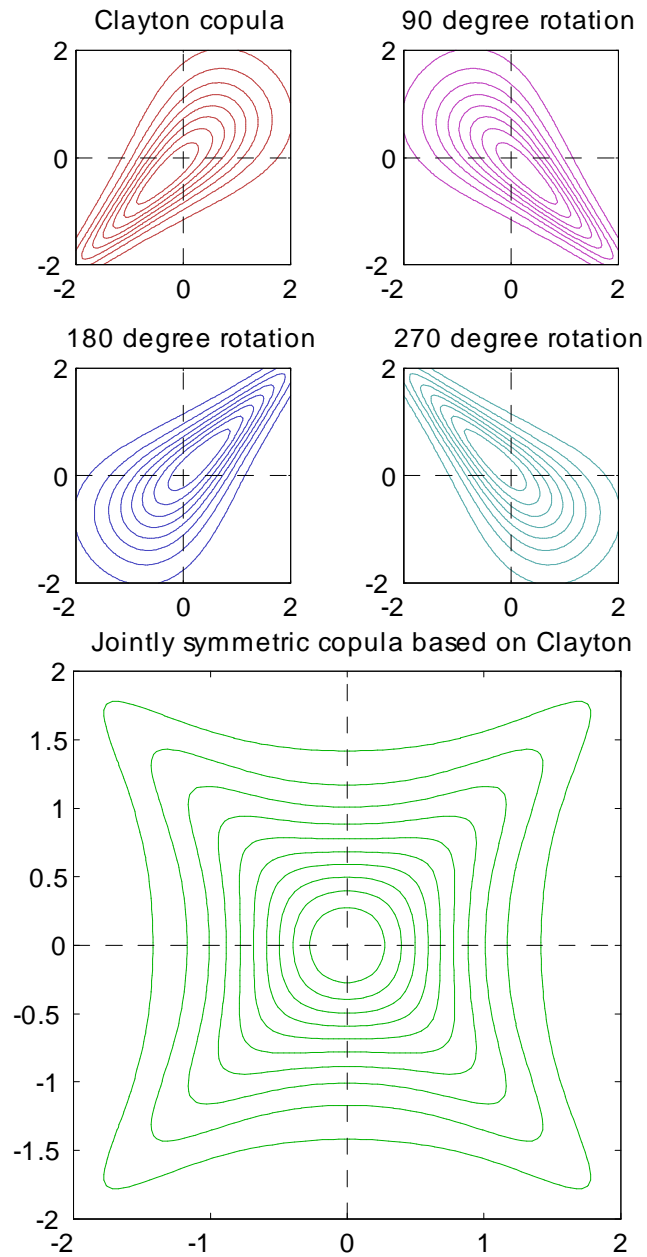


Figure 1: *Iso-probability contour plots of joint distributions with standard Normal margins and various copulas: the Clayton copula ($\theta = 2$), and its 90-, 180-, and 270-degree rotations (upper panel), and an equal-weighted average of the four Clayton copulas (lower panel).*

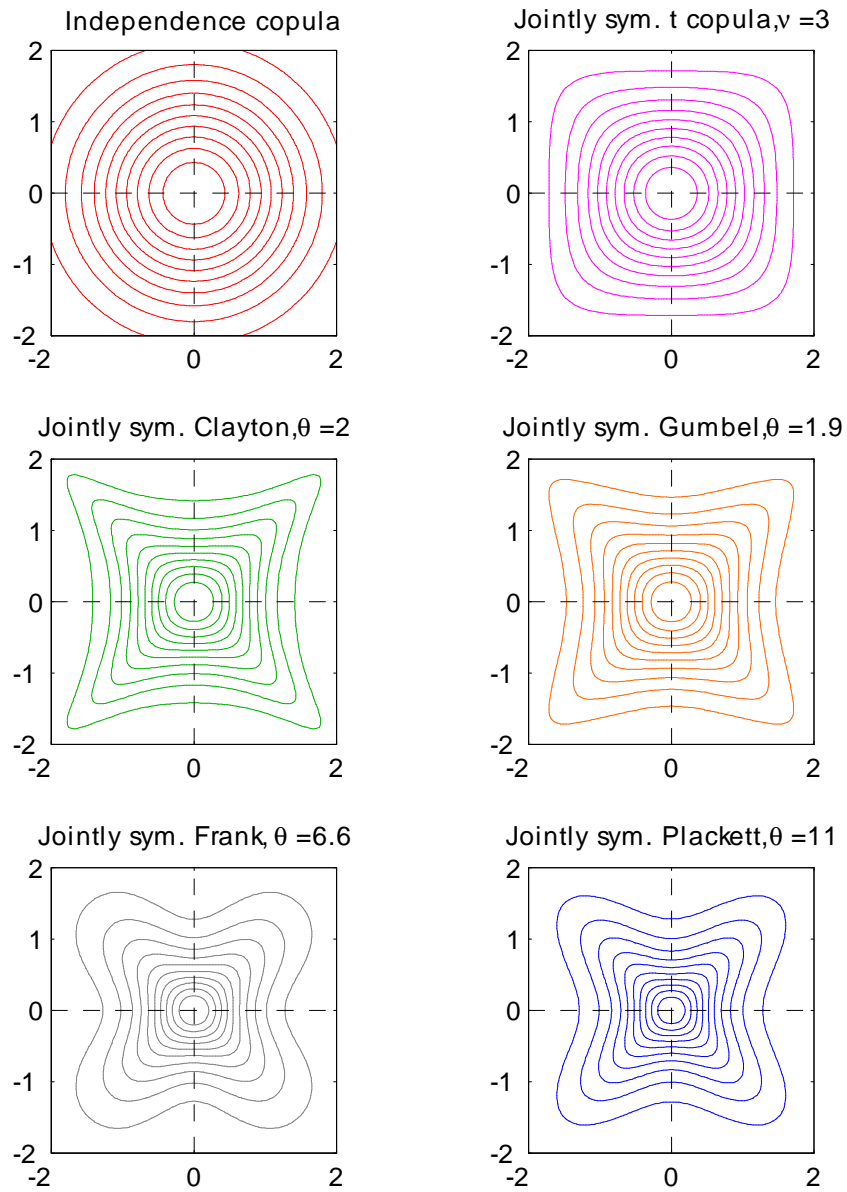


Figure 2: *Iso-probability contour plots of joint distributions with standard Normal margins and various jointly symmetric copulas.*

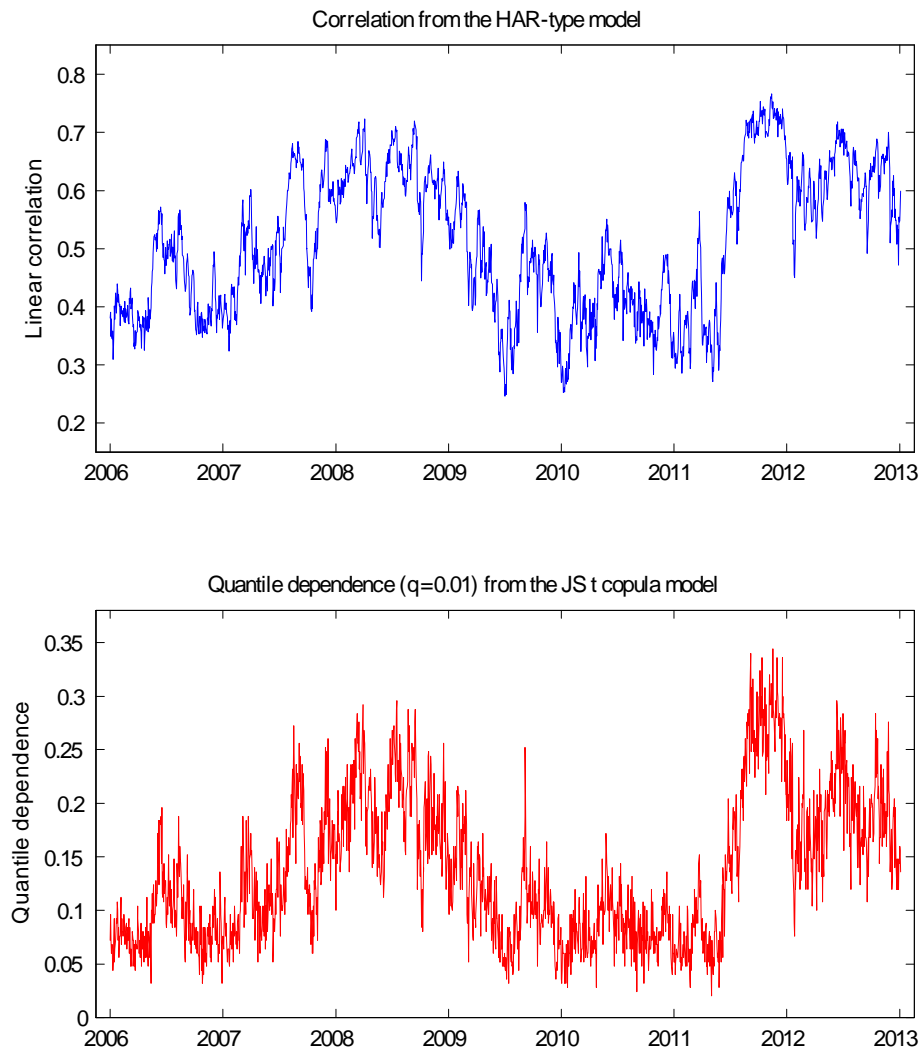


Figure 3: *Model-implied linear correlation (upper panel) and 1% quantile dependence (lower panel) for daily returns on Citi Group and Goldman Sachs, based on the HAR-type model for the conditional covariance matrix, and the jointly symmetric t copula model.*