Copulas for High Dimensions: Models, Estimation, Inference, and Applications

by

Dong Hwan Oh

Department of Economics
Duke University

Date: ______________________

Approved:

____
Andrew J. Patton, Supervisor

____
Tim Bollerslev

____
George Tauchen

____
Shakeeb Khan

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Copulas for High Dimensions: Models, Estimation, Inference, and Applications

by

Dong Hwan Oh

Department of Economics
Duke University

Date: ______________________
Approved:

____
Andrew J. Patton, Supervisor

____
Tim Bollerslev

____
George Tauchen

____
Shakeeb Khan

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2014

Abstract

The dissertation consists of four chapters that concern topics on copulas for high dimensions. Chapter 1 proposes a new general model for high dimension joint distributions of asset returns that utilizes high 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 residuals. Estimation of the new class of copulas is conducted using a composite likelihood, making the model 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.

Chapter 2, which is co-authored with Professor Andrew Patton, presents new models for the dependence structure, or copula, of economic variables based on a factor structure. The proposed models are particularly attractive for high dimensional applications, involving fifty or more variables. This class of models generally lacks a closed-form density, but analytical results for the implied tail dependence can be obtained using extreme value theory, and estimation via a simulation-based method using rank statistics is simple and fast. We study the finite-sample properties of the estimation method for applications involving up to 100 variables, and apply the model to daily returns on all 100 constituents of the S&P 100 index. We find

significant evidence of tail dependence, heterogeneous dependence, and asymmetric dependence, with dependence being stronger in crashes than in booms.

Chapter 3, which is co-authored with Professor Andrew Patton, considers the estimation of the parameters of a copula via a simulated method of moments type approach. This approach is attractive when the likelihood of the copula model is not known in closed form, or when the researcher has a set of dependence measures or other functionals of the copula that are of particular interest. The proposed approach naturally also nests method of moments and generalized method of moments estimators. Drawing on results for simulation based estimation and on recent work in empirical copula process theory, we show the consistency and asymptotic normality of the proposed estimator, and obtain a simple test of over-identifying restrictions as a goodness-of-fit test. The results apply to both *iid* and time series data. We analyze the finite-sample behavior of these estimators in an extensive simulation study.

Chapter 4, which is co-authored with Professor Andrew Patton, proposes a new class of copula-based dynamic models for high dimension conditional distributions, facilitating the estimation of a wide variety of measures of systemic risk. Our proposed models draw on successful ideas from the literature on modelling high dimension covariance matrices and on recent work on models for general time-varying distributions. Our use of copula-based models enable the estimation of the joint model in stages, greatly reducing the computational burden. We use the proposed new models to study a collection of daily credit default swap (CDS) spreads on 100 U.S. firms over the period 2006 to 2012. We find that while the probability of distress for individual firms has greatly reduced since the financial crisis of 2008-09, the joint probability of distress (a measure of systemic risk) is substantially higher now than in the pre-crisis period.

To my parents:

Seung Hyeub Oh and Jung Sook Lee

Contents

List of Tables

List of Figures

Acknowledgements

I am deeply indebted to Andrew Patton, Tim Bollerslev, George Tauchen, Jia Li, Shakeeb Khan, and Federico Bugni for their valuable guidance and continuous encouragement through all the years of my study at Duke. I am particularly grateful to my advisor, Andrew Patton, for generously sharing with me his wisdom, knowledge, and precious time whenever needed. His endless patience and support enriched my Ph.D. life to be a thoughtful and rewarding journey. Getting to the point of writing this acknowledgement for my dissertation would not have been possible without him.

I am also much obliged to my graduate advisor from Seoul, Yoon-Jae Whang, for introducing me to the world of econometrics. His academic erudition and work ethics have always inspired me over the course of my research.

I would like to thank my wife, Yi Jin Yun. I cannot imagine what my whole life would have been like without her. Her sacrifice to support my career and our adorable children, Timothy and Ted, has been a great privilege to me.

Last but not least, I would like to thank my parents for making all this possible. I dedicate this work to them.

1

Modelling High Dimension Distributions with High Frequency Data and Copulas

1.1 Introduction

A multivariate joint distribution for the returns on hundreds of financial assets is one of the most crucial components in modern risk managements and asset allocations. Given that the complete information of dependence and marginal behaviors of assets in portfolios is contained in joint distributions, it is no exaggeration to say that financial decision makings are completely determined by joint distributions of constituents of portfolios. Modelling high dimension distributions, however, is not an easy task mainly due to the "curse of dimensionality," so only a few models are available for high dimensions. This is the reason why high dimension Normal distribution is still widely used in practice and academia in spite of its notorious limits, for example, thin tails and zero tail dependence. This paper proposes a new general model for high dimension joint distributions of asset returns that utilizes high frequency data and copulas. This model is sufficiently general that non-normal features of financial data can easily be incorporated, and novel estimation methods enable us to overcome the "curse of dimensionality" in high dimensions.

Over the last decade, there have been two major findings in financial econometrics. First, high frequency (HF) intraday data has been proven more superior to daily data in measuring and forecasting variances and covariances, see Andersen, *et al.* (2001, 2004) and Barndorff-Nielsen and Shephard (2004). This implies that linear dependence represented by covariances is captured quite well by HF data. Second, copulas can be used to construct high dimension distributions with specified dependence and arbitrary marginal distributions, see Patton (2012) for a comprehensive review. Separately specifying all marginal distributions and dependence makes constructing joint distributions much easier than directly modelling them does. These two findings naturally leads to a question: how could HF data and copulas be used to improve the modelling and forecasting of daily return distributions, especially for high dimensions, say hundreds of random variables?

To address this question is not simple because direct use of HF intraday data to estimate the copula of daily returns is not straightforward. Given the fact that daily returns are the sum of intraday returns, it is not reasonable to assume that the copula of daily returns is a known function of copulas of intraday returns. It is like a claim that a joint distribution of sum of random variables is a known function of joint distributions of each variable, which is not generally true except for special cases such as sum of independent Gaussian random variables. In contrast, Barndorff-Nielsen and Shephard (2004) prove that the realized covariance matrix nonparametrically constructed using intraday data converges in probability to the covariance matrix of daily returns as data frequency goes to infinity. This elegant link between intraday and daily returns for the second moments does not generally hold for copulas for the above reason, so alternative approaches are required in order to exploit information of HF data for modelling dependence.

This paper decomposes the dependence structure into linear and nonlinear components. This decomposition enables the use of HF data to accurately measure and forecast linear dependence and the use of a new class of copulas designed to capture nonlinear dependence between the resulting linearly uncorrelated residuals. This approach is novel as it enables an enhanced estimation of a joint distribution by separately assigning HF data and copula to estimate the linear and nonlinear dependence, respectively. The literature on joint distributions using HF data mostly ignores nonlinear dependence by focusing only on the second moments or linear dependence, for example, Chiriac and Voev (2011), Jondeau and Rockinger (2012), Hautsch, *et al.* (2013), and Jin and Maheu (2013), among others, whereas the literature on copula does not use information of HF data to improve modelling dependence, for example, Chen and Fan (2006), Patton (2006b), and Oh and Patton (2011), among others. The aforementioned decomposition permits the use of both HF data and copulas to improve in modelling dependence without further unrealistic assumptions. To model nonlinear dependence, a new class of copulas designed to capture dependence of linearly uncorrelated random variables is necessary. Motivated by the few parametric copulas available for linearly uncorrelated random variables, for example, t copula with zero correlations, this paper proposes a new method to construct copulas for linearly uncorrelated random variables from any given copula. Since this method is based on simple rotations of a given copula, there is nothing difficult to generate new copulas. This new class of copulas enables us to examine various features of nonlinear dependence which might be generally overlooked in the literature.

The major benefits of the proposed models for joint distributions are threefold. First, the proposed model is sufficiently flexible that almost all kinds of joint models for the second moment and return distributions in the literature are interpretable in this framework. We may also rely on the large literature on multivariate second moments such as BEKK (Engle and Kroner, 1995) and DCC (Engle, 2002) for the separate models for the second moments. Second, the proposed model can be easily extended to high dimensions, say 100, and the estimation is feasible and fast. Computational problems often arise in high dimension models, and the proposed model and estimation methods overcome the "curse of dimensionality" by multi-stage estimation with composite likelihood, see Varin (2008) and Varin, *et al.* (2011) for a comprehensive review. Third, the proposed model allows for the simultaneous use of high frequency data and copulas to completely capture dependence. HF data is assigned to capture linear dependence and jointly symmetric copulas are used to capture nonlinear dependence. Fully exploiting the advantages of HF data and copulas leads to superior performance of the proposed model.

Similar approaches have already appeared in the literature. First of all, Lee and Long (2009) consider the decomposition of linear and nonlinear dependence although their method is more difficult to interpret and cannot be extended to high dimensions.[1] Second, the proposed model is related to the literature on copulas for high dimensions, see Patton (2012) for a review. The standard copula approaches for modelling joint distributions usually do not take into account the decomposition of dependence into linear and nonlinear components because the copula captures all kinds of dependence, including the second moments. Those approaches are not usually formulated to exploit information of HF data for dependence. Another related strand of the literature is about modelling the second moments or joint modelling the second moments and return distributions, see Andersen, *et al.* (2006) for a review, Jin and Maheu (2013) and references therein. Since they usually directly model the joint distributions of standardized residuals rather than relying on copula methods, the multivariate Normal or Student's t distributions are used especially for high dimensions, which implies that nonlinear dependence is mostly ignored.

The remainder of the paper is organized as follows. Section 1.2 provides details of proposed models for high dimension distributions. Section 1.3 and Section 1.4

[1] The details are discussed in Section 1.2

present the estimation procedure and a simulation study for the proposed models, respectively. Section 1.5 applies the proposed model to real data and presents in-sample results and out-of-sample forecasting results for density forecasting and a portfolio choice problems. Section 1.6 concludes.

1.2 Joint models for covariances and returns

We construct the model of N daily return random variables $\mathbf{r}_t$ as follows:

$$\mathbf{r}_t = \mu_t + \mathbf{H}_t^{1/2}\mathbf{e}_t \tag{1.1}$$

$$\mathbf{e}_t|\mathcal{F}_{t-1} \sim \mathbf{F}\left(\cdot;\eta\right) \tag{1.2}$$

where $\mu_t \equiv E\left[\mathbf{r}_t|\mathcal{F}_{t-1}\right]$, $\mathbf{H}_t \equiv Cov\left[\mathbf{r}_t|\mathcal{F}_{t-1}\right]$, $\mathcal{F}_t = \sigma\left(\mathbf{r}_t, \mathbf{r}_{t-1}, \ldots\right)$ and $\mathbf{F}\left(\cdot;\eta\right)$ is a parametric distribution with zero mean and identity covariance matrix. To obtain the square root of a matrix, the spectral decomposition (based on eigenvalues and eigenvectors) is used due to its invariance to the order of the variables. Equation (1.1) implies that returns are specified by conditional mean μ_t, conditional variance-covariance matrix $\mathbf{H}_t$, and standardized uncorrelated residuals $\mathbf{e}_t$ with $E\left[\mathbf{e}_t|\mathcal{F}_{t-1}\right] = \mathbf{0}$ and $E\left[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}\right] = \mathbf{I}$. Equation (1.2) describes that standardized uncorrelated residuals $\mathbf{e}_t$ follow a conditional joint distribution $\mathbf{F}\left(\cdot\right)$ with parameter η. We may further consider the decomposition of conditional joint distribution $\mathbf{F}$ of $\mathbf{e}_t$ into marginal distributions F_i and copula $\mathbf{C}$ by Sklar (1959) and Patton (2006b):

$$\mathbf{e}_t|\mathcal{F}_{t-1} \sim \mathbf{F}\left(\cdot;\eta\right) = \mathbf{C}\left(F_1\left(\cdot;\eta\right), ..., F_N\left(\cdot;\eta\right);\eta\right) \tag{1.3}$$

Note that the uncorrelated $\mathbf{e}_t$ does *not* necessarily mean cross-sectional independence. Except for the independence copula, the elements of $\mathbf{e}_t$ are uncorrelated but have cross-sectional dependence which is completely described by copula $\mathbf{C}$.

This approach naturally reveals two kinds of dependences of $\mathbf{r}_t$: the "linear dependence" captured by conditional variance-covariance matrix $\mathbf{H}_t$ and the "nonlinear dependence" remaining in the uncorrelated residuals $\mathbf{e}_t$ captured by copula $\mathbf{C}$. The main aim of this paper is to construct a high dimension flexible forecasting distribution model which can simultaneously capture both linear and nonlinear dependence of daily returns and to propose an fast and accurate estimation method taking advantage of high frequency intraday data.

There are two important advantages in decomposing the joint distribution of $\mathbf{r}_t$ in equation (1.1), (1.2), and (1.3). First, it allows the researcher to draw on the large literature on measuring, modeling and forecasting conditional variance-covariance matrix $\mathbf{H}_t$ with low and high frequency data. For example, GARCH-type observation driven models such as the multivariate GARCH model by Bollerslev, Engle, Wooldridge (1988), the structural BEKK model by Engle and Kroner (1995), and the dynamic conditional correlation (DCC) model by Engle (2002) naturally fit in equation (1.1) and (1.2), and can be estimated with quasi maximum likelihood methods. The increasing availability of high frequency data also enables us to use more accurate models for conditional variance-covariance matrix, for example, among others, Bauer and Vorkink (2011), Chiriac and Voev (2011), and Noureldin, *et al.* (2011), and those models are also naturally adapted in equation (1.1) and (1.2). Second, the model specified by equation (1.1), (1.2), and (1.3) is easily extended to high dimension applications given that multi-stage separate estimations for conditional mean, conditional variance-covariance and standardized residuals with marginal distributions and copula are allowed. The main difficulty of high dimension problem is the proliferation of parameters and huge computation burden as the dimension increases, known as the curse of dimensionality. The above model, however, overcomes this obstacle not only by separating estimation stages but also by using composite likelihood estimation. The details are provided in Sections 1.2.1 and 1.3.3.

To emphasize the prominent features of our model, we contrast our model with some models in the extant literature. Lee and Long (2009) distinguish and model the linear dependence captured by covariance matrix and the nonlinear dependence remaining in $\boldsymbol{\Sigma}_t^{-1/2}\mathbf{w}_t$ captured by copula of $\mathbf{w}_t$:

$$\mathbf{r}_t = \mu_t + \mathbf{H}_t^{1/2}\boldsymbol{\Sigma}_t^{-1/2}\mathbf{w}_t \tag{1.4}$$
$$\mathbf{w}_t|\mathcal{F}_{t-1} \sim \mathbf{G}(\cdot;\eta) = \mathbf{C}_{\mathbf{w}}\left(G_1(\cdot;\eta),...,G_N(\cdot;\eta);\eta\right)$$

where $\boldsymbol{\Sigma}_t$ is covariance matrix implied by $\mathbf{G}(\cdot;\eta)$. Rather than directly modelling uncorrelated residuals $\mathbf{e}_t$, Lee and Long (2009) use $\mathbf{w}_t$ and its covariance $\boldsymbol{\Sigma}_t$ to obtain uncorrelated residuals $\mathbf{e}_t$. This model is unclear to interpret $\mathbf{w}_t$ and its covariance $\boldsymbol{\Sigma}_t$. More importantly, this approach makes them lose a definite advantage of using copula: multi- stage separating modelling for marginal distributions and dependence. In order to convert $\mathbf{w}_t$ into uncorrelated $\mathbf{e}_t$, they need covariance matrix $\boldsymbol{\Sigma}_t$ determined by both marginal distributions and copula of $\mathbf{w}_t$. Consequently, they have to jointly estimate all parameters η of marginal distributions and copula, which is an inevitable burden especially for high dimensions.

In contrast to their model, our model is easy to interpret and quite flexible and manageable in high dimension because we directly model the standardized uncorrelated residuals $\mathbf{e}_t$ to take advantage of benefits from multi-stage separation. The concern, of course, is that there are a few copulas to ensure zero correlations, for example, Gaussian copula with identity correlation matrix (i.e. the independence copula) and t copula with identity correlation matrix combining with symmetric marginals. We suggest, however, methods to generate various copulas ensuring zero correlations given any copulas by constructing jointly symmetric copula in Section 1.2.1.

Chen and Fan (2006) and Oh and Patton (2011), among others, differ from this paper in that they separate only individual variances rather than the entire variance-covariance matrix and model correlated but de-meaned and de-volitilized returns rather than uncorrelated residuals. Their approach has an advantage that whole dependence information including the second moment such as variance-covariance matrix is all captured by the copula, and the task boils down to how to construct flexible copulas to explain various features of dependence in data. In contrast, our model enables us to use all existing models for the variance-covariance matrix by separating linear and nonlinear dependence. In particular, more accurate modelling and forecasting variance-covariance matrix using high frequency data attained widespread popularity, see Chiriac and Voev (2011), Hansen, *et al.* (2012), and Noureldin, *et al.* (2012) among others. Under our model, we are able to effectively exploit the information of high frequency data for linear dependence, e.g. estimating and forecasting $\mathbf{H}_t$ using high frequency data and to estimate the model for uncorrelated but dependent residuals $\mathbf{e}_t$ using low-frequency data. Although recent research on using high frequency data for copula is growing, most have troubles in appropriately linking high frequency data and copula mainly because a copula of low frequency data is not a known function of copulas of high frequency data, which is contrary to the elegant relationship between high frequency volatility measures and low-frequency counterparts. Our model certainly enables us to exploit all information of both high and low frequency data for linear dependence and nonlinear dependence, respectively.[2]

Another important feature of our model is that it is a joint model for returns and covariances like Jondeau and Rockinger (2012), Hautsch, *et al.* (2013), and Jin and Maheu (2013), among others. While those papers use Normal or Student's t distributions for standardized uncorrelated residuals after modelling covariances, our model for standardized returns is flexible enough to nest not only Normal or Student's t distribution but also various non-standard distributions that ensure zero correlations. Those uncorrelated residuals still have dependence defined as nonlinear

[2] In this paper, we use high-frequency data only for covariance matrix estimation and forecasts. Recently, De Lira Salvatierra and Patton (2013) use dependence information from high-frequency data, e.g. realized correlations, for bivariate dynamic copula models. We leave this possbility for future research.

dependence in this paper, and our model is designed to fully capture various types of nonlinear dependence through a new class of copulas for uncorrelated variables. This makes our model prominent compared to existing models in the literature mostly neglecting nonlinear dependence. Since the new copulas for uncorrelated variables are constructed without difficulty from any given copula by simple rotations, our model can be sufficiently flexible to explain various nonlinear dependence. Through the empirical analysis in Section 1.5 we find that models capable of capturing nonlinear dependence significantly outperform models completely or mostly ignoring it such as Normal or Student's t distribution.

As mentioned above, we are able to separately model conditional mean, conditional covariance, and standardized uncorrelated residuals. We first focus on models for uncorrelated residuals $\mathbf{e}_t$ in Section 1.2.1. Next, forecasting models for variance-covariance matrix $\mathbf{H}_t$ are considered in Section 1.2.2.

1.2.1 A model for uncorrelated standardized residuals

To model uncorrelated residuals $\mathbf{e}_t$, we propose combining a jointly symmetric copula with a set of symmetric marginal distributions, which guarantees zero correlations between variables. We first define the symmetry for an univariate random variable and the *jointly symmetry*[3] for N random variables $\{X_i\}_{i=1}^N$.

Definition 1 (Symmetry)**.** *An univariate random variable X is symmetric about a in $\mathbf{R}$ if the distribution functions of $X - a$ and $a - X$ are the same.*

Definition 2 (Joint symmetry, Definition 2.7.1 in Nelsen 2006)**.** *Let $\{X_i\}_{i=1}^N$ be N random variables and let $\{a_i\}_{i=1}^N$ be a point in $\mathbf{R}^N$. $\{X_i\}_{i=1}^N$ is jointly symmetric about $\{a_i\}_{i=1}^N$ if the following 2^N sets of N random variables have a common joint*

[3] Various concepts of symmetry for multivariate random variables are available, for example, exchangibility and radial symmetry. We refer to Nelsen (2006).

distribution:

$$\left(\widetilde{X}_1, \ldots, \widetilde{X}_i, \ldots, \widetilde{X}_N\right)$$

where $\widetilde{X}_i = (X_i - a_i)$ *or* $(a_i - X_i)$ *for* $i = 1, ..., N$. *For* $N = 2$, *for example,* (X_1, X_2) *is jointly symmetric random variables about* (a_1, a_2) *if* $(X_1 - a_1, X_2 - a_2)$, $(a_1 - X_1, X_2 - a_2), (X_1 - a_1, a_2 - X_2)$, *and* $(a_1 - X_1, a_2 - X_2)$ *have a common joint distribution.*

For example, if N random variables $\{X_i\}_{i=1}^N$ follow the Normal distribution with a mean vector μ and an identity covariance matrix, then $\{X_i\}_{i=1}^N$ is jointly symmetric about μ. If those random variables are continuous, there exists a unique copula by Sklar's theorem, so we naturally think about a copula for those jointly symmetric random variables. We first define the *jointly symmetric copula,* and examine the relationship between jointly symmetric random variables and jointly symmetric copulas in Theorem 1.

Definition 3 (Jointly symmetric copula). *A N dimension copula* $\mathbf{C}(u_1, \ldots, u_N)$ *is jointly symmetric if it satisfies*

$$\forall i,\ \mathbf{C}(u_1, \ldots, u_i, \ldots, u_N) = \mathbf{C}(u_1, \ldots, 1, \ldots, u_N) - \mathbf{C}(u_1, \ldots, 1 - u_i, \ldots, u_N) \quad (1.5)$$

where $u_i \in [0, 1]$. $\mathbf{C}(u_1, \ldots, 1, \ldots, u_N)$ *and* $\mathbf{C}(u_1, \ldots, 1 - u_i, \ldots, u_N)$ *mean that the i-th element is* 1 *and* $1-u_i$, *respectively, and other elements are* $\{u_1, .., u_{i-1}, u_{i+1}, .., u_N\}$.

Theorem 1 (Multivariate analog of Exercise 2.30 in Nelsen 2006). *Let* $\{X_i\}_{i=1}^N$ *be N continuous random variables with joint distribution* $\mathbf{F}$, *marginal distributions* $F_1, .., F_N$ *and copula* $\mathbf{C}$. *Further suppose each of* $\{X_i\}_{i=1}^N$ *is symmetric about each of* $\{a_i\}_{i=1}^N$, *respectively. Then,*

(i) $\{X_i\}_{i=1}^N$ *is jointly symmetric about* $\{a_i\}_{i=1}^N$ *if and only if* $\mathbf{C}$ *satisfies equation (1.5)*

(ii) If $\{X_i\}_{i=1}^N$ *is jointly symmetric about* $\{a_i\}_{i=1}^N$ *, then the correlation of any pair* (X_i, X_j) *is zero for* $i \neq j$*.*

The proof is presented in Appendix A.1. The first result of Theorem 1 states that a jointly symmetric copula satisfying equation (1.5) is the copula for jointly symmetric random variables when marginal distributions are symmetric. The second result implies zero correlations of any pair of jointly symmetric random variables.

While numerous copulas have been proposed to explain various features of dependences in the literature, only a few copulas comply with equation (1.5), for example, the Gaussian and t copulas with the identity correlation matrix. With this limited choice of copulas, we could not fully account for characteristics of nonlinear dependence. Thus, we suggest a novel way to construct jointly symmetric copulas by rotating any given copula.

Theorem 2. *Assume that N dimension copula* $\mathbf{C}$ *with density* $\mathbf{c}$ *is given.*

(i) For any given N dimension copula $\mathbf{C}$*, the following copula* $\mathbf{C}^{JS}$ *is jointly symmetric, i.e. satisfying equation (1.5)*

$$\mathbf{C}^{JS}(u_1, \ldots, u_N) = \frac{1}{2^N}\left[\sum_{j_1=1}^{3} \cdots \sum_{j_N=1}^{3} (-1)^J \cdot \mathbf{C}(\widetilde{u}_1, \ldots, \widetilde{u}_i, \ldots, \widetilde{u}_N)\right] \tag{1.6}$$

where $J = \sum_{i=1}^N 1\{j_i = 2\}$ *and* $\widetilde{u}_i = \begin{cases} u_i & \text{for } j_i = 1 \\ 1 - u_i & \text{for } j_i = 2 \\ 1 & \text{for } j_i = 3 \end{cases}$

(ii) The probability density function $\mathbf{c}^{JS}(u_1, \ldots, u_N)$ *of* $\mathbf{C}^{JS}(u_1, \ldots, u_N)$ *is*

$$\begin{aligned} \mathbf{c}^{JS}(u_1, \ldots, u_N) &= \frac{\partial}{\partial u_1 \cdots \partial u_N}\mathbf{C}^{JS}(u_1, \ldots, u_N) \\ &= \frac{1}{2^N}\left[\sum_{j_1=1}^{2} \cdots \sum_{j_N=1}^{2} \mathbf{c}(\widetilde{u}_1, \ldots, \widetilde{u}_i, \ldots, \widetilde{u}_N)\right] \end{aligned} \tag{1.7}$$

where $\widetilde{u}_i = \begin{cases} u_i & \text{for } j_i = 1 \\ 1 - u_i & \text{for } j_i = 2 \end{cases}$

The proof is presented in Appendix A.1. Theorem 2 proves that the sum of mirror-image symmetrical rotations about every axis turns out to be jointly symmetric copula.[4] This theorem tells that any given non-jointly symmetric copulas can be transformed into jointly symmetric ones by simple rotations and beyond Gaussian or t copulas, any copula can be used for modelling jointly symmetry.

$\mathbf{C}^{JS}(u_1, \ldots, u_N)$ in equation (1.6) involves all marginal copulas of the given copula whereas the density $\mathbf{c}^{JS}$ requires only the densities of the given copula rather than marginal copulas. This makes it easier to visualize how to construct a jointly symmetric copula in terms of the copula density $\mathbf{c}^{JS}$ than $\mathbf{C}^{JS}$. We further move the space from unit simplex $[0,1]^N$ to $\mathbf{R}^N$ using copula with standard Normal marginal distributions. Figure 1.1 shows 90, 180 and 270 degree rotations of Clayton copula density with standard Normal marginal densities, which corresponds to $\mathbf{c}(1-u_1, u_2)$, $\mathbf{c}(1-u_1, 1-u_2)$, and $\mathbf{c}(u_1, 1-u_2)$ with the same marginal distributions. Interpreting equation (1.7) in $\mathbf{R}^N$, we find that $\mathbf{c}^{JS}$ is the copula density of a equal weighted sum of rotations of a given copula about every axis. Figure 1.2 is the density of jointly symmetric copula based on Clayton copula with parameter 1 obtained by equal weighted sum of four densities in Figure 1.1.

In addition, we emphasize that zero correlation does not always imply independence. Figure 1.3 highlights the difference between various jointly symmetric copula constructed by equation (1.7) and the independence copula. The copula for uncorrelated variables can be very different from the independence copula, and this difference means there may exist nonlinear dependence in uncorrelated variables.

[4] Note that while this is not the only way to construct jointly symmetric copulas, it requires the least number of rotations. Figure 1.1 uses 90, 180, 270, and 360 degree rotations to generate a jointly symmetric copula following (1.7), but combining 30, 60, 90, ..., 300, 330, and 360 degree rotations produces another jointly symmetric copula although more rotations are needed.

1.2.2 *Forecasting models for multivariate covariance matrix*

Research on forecasting model for multivariate covariance matrix with low-frequency data is pervasive, see Andersen, *et al.* (2006) for a review, and recently forecasting models using widely available high frequency data are growing, e.g. Chiriac and Voev (2011), Noureldin, *et al.* (2012) among others. There are two major concerns about forecasting models for multivariate covariance matrix: parsimony and positive definiteness. Keeping these two concerns in mind, we suggest a new forecasting model using high frequency data.

We combine the essential ideas of the DCC model by Engle (2002) and the heterogeneous autoregressive (HAR) model by Corsi (2009). Following the DCC model, we use two separate steps for individual variances and covariances in order to have a computation advantage and parsimony. We use the HAR model that is known to successfully explain the long-memory behavior of volatility in a simple AR type 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

$$RVarCov_t^{\Delta} = \sum_{j=1}^{1/\Delta} \mathbf{r}_{t-1+j\cdot\Delta}\mathbf{r}'_{t-1+j\cdot\Delta} \tag{1.8}$$

and is re-written in realized variances and realized correlations by

$$RVarCov_t^{\Delta} = \sqrt{RVar_t^{\Delta}} \cdot RCorr_t^{\Delta} \cdot \sqrt{RVar_t^{\Delta}} \tag{1.9}$$

where $RVar_t^{\Delta} = diag\left[RVarCov_t^{\Delta}\right]$ and $RCorr_{t,ij}^{\Delta} = \left\{ \frac{RVarCov_{t,ij}^{\Delta}}{\sqrt{RVarCov_{t,ii}^{\Delta}}\sqrt{RVarCov_{t,jj}^{\Delta}}} \right\}_{ij}$.

We firstly apply the HAR model to each realized variance by

$$\log RVar^{\Delta}_{ii,t} = \phi_i^{(const)} + \phi_i^{(day)} \log RVar^{\Delta}_{ii,t-1} + \phi_i^{(week)} \frac{1}{4} \sum_{k=2}^{5} \log RVar^{\Delta}_{ii,t-k} \tag{1.10}$$

$$+ \phi_i^{(month)} \frac{1}{15} \sum_{k=6}^{20} \log RVar^{\Delta}_{ii,t-k} + \xi_{it}$$

and those coefficients $\left\{\phi_i^{(const)}, \phi_i^{(day)}, \phi_i^{(week)}, \phi_i^{(month)}\right\}_{i=1}^{N}$ are estimated by OLS for each i variable.

Note that we use logarithm of realized variance rather than realized variance itself for two reasons. First, this model is for forecasts of variance, so positiveness of forecasts should be guaranteed and we can easily achieve positiveness by transforming into logarithms. Second, estimation by OLS can be largely affected by outliers, i.e. some large values of realized variances, and our sample period includes 2008 financial crisis resulting in substantial changes in variance. Log-transforming plays a role in dampening down large changes and reduces the impact of outliers in OLS estimation.

Similarly, we model realized correlations using the vech operator[5]

$$vech\left(RCorr^{\Delta}_{t}\right) = CONST + A \cdot vech\left(RCorr^{\Delta}_{t-1}\right) + B \cdot \frac{1}{4} \sum_{k=2}^{5} vech\left(RCorr^{\Delta}_{t-k}\right)$$

$$+ C \cdot \frac{1}{15} \sum_{k=6}^{20} vech\left(RCorr^{\Delta}_{t-k}\right) + \xi_t$$

where $vech\left(RCorr^{\Delta}_{t}\right)$ and $CONST$ are $\frac{N(N-1)}{2} \times 1$ vector and A, B, and C are $\frac{N(N-1)}{2} \times \frac{N(N-1)}{2}$ matrix. However, the number of parameters to estimate grows at the rate of $O\left(N^2\right)$ which is infeasible to estimate when N is large. We may consider some restrictions as the DCC model does, substituting A, B, and C with constant scalar a, b, and c. Then, the number of parameters to estimate is significantly

[5] The vech operator vertically stacks the upper triangular elements excluding diagonal elements.

reduced to $\frac{N(N-1)}{2}+3$ from $\frac{N(N-1)}{2}+3\left(\frac{N(N-1)}{2}\right)^2$. When N is large, constant terms are still many to estimate, so we lean on the idea of "variance targeting" by Engle and Mezrich (1996):

$$
\begin{aligned}
vech\left(RCorr_t^{\Delta}\right) = (1-a-b-c)\,E\left[vech\left(RCorr_t^{\Delta}\right)\right] + a\cdot vech\left(RCorr_{t-1}^{\Delta}\right) \quad (1.11) \\
+ b\cdot\frac{1}{4}\sum_{k=2}^{5} vech\left(RCorr_{t-k}^{\Delta}\right) + c\cdot\frac{1}{15}\sum_{k=6}^{20} vech\left(RCorr_{t-k}^{\Delta}\right) + \xi_t
\end{aligned}
$$

Substituting $E\left[vech\left(RCorr_t^{\Delta}\right)\right]$ with its sample mean, we can rewrite the equation using demeaned $RCorr_t^{\Delta}$:

$$
\begin{aligned}
vech\left(\widetilde{RCorr_t^{\Delta}}\right) = a\cdot vech\left(\widetilde{RCorr_t^{\Delta}}\right) + b\cdot\frac{1}{4}\sum_{k=2}^{5} vech\left(\widetilde{RCorr_{t-k}^{\Delta}}\right) \\
+ c\cdot\frac{1}{15}\sum_{k=6}^{20} vech\left(\widetilde{RCorr_{t-k}^{\Delta}}\right) + \xi_t
\end{aligned}
$$

where $\widetilde{RCorr_t^{\Delta}} = RCorr_t^{\Delta} - \frac{1}{T}\sum_{t=1}^{T} RCorr_t^{\Delta}$. Now, the coefficients a, b, and c are easily estimated by OLS and constant terms are recovered back by estimates of a, b, and c and sample mean of realized correlation.

In order to ensure positive definite forecasts for $RVarCov_t^{\Delta}$, some conditions given in the following Theorem 3 are necessary.

Theorem 3. *Assume the following three conditions*

1. $\Pr\left[\mathbf{x}'\mathbf{r}_t = 0\right] = 0$ *for any nonzero* $\mathbf{x}\in\mathbf{R}^N$*, i.e.* $\mathbf{r}_t$ *does not have redundant assets.*

2. $a, b,$ *and* $c \geqslant 0$

3. $a+b+c<1$

Then, $E\left[RVarCov_t^{\Delta}|\mathcal{F}_{t-1}\right]$ *is positive definite. In addition, if* $E\left[RCorr_t^{\Delta}\right]$ *is estimated by* $\frac{1}{T}\sum_{t=1}^{T} RCorr_t^{\Delta}$ *and* $T \geqslant N$*, then the sample counterpart to* $E\left[RVarCov_t^{\Delta}|\mathcal{F}_{t-1}\right]$ *is positive definite.*

The proof is given in Appendix A.1. Our forecasting model for realized variance-covariance matrix is simple and fast to estimate and positive definiteness of forecasts is ensured by Theorem 3. We note that the above theorem is robust to the misspecification of return distributions, i.e. Theorem 3 holds regardless of whether or not return distribution follows the proposed model specified by equation (1.1), (1.2), and (1.3).

1.3 Estimation methods and model comparisons

1.3.1 Estimation using composite likelihood estimation

The proposed method to construct jointly symmetric copulas in equation (1.7) requires 2^N calculations of the given original copula density. Even for moderate dimensions, say $N = 20$, the likelihood evaluation could be too cumbersome to calculate, as shown in the first row of Table 1.1. For high dimensions, the ordinary maximum likelihood estimation is not feasible for the jointly symmetric copulas and we suggest an alternative that can overcome this computation issue. We construct the composite likelihood rather than the usual full likelihood, and estimate the parameters of jointly symmetric copulas by maximizing the composite likelihood.

The composite likelihood (CL) (Lindsay, 1988) consists of combinations of the valid likelihoods of submodels or marginal models under the assumption that those submodels are independent. See Varin (2008) and Varin, *et al.* (2011) for overview. The essential intuition behind CL is that since submodels include partial information of full dependence governed by parameters of full likelihood, by properly using that partial information, we can estimate parameters of full likelihood, although with some inevitable efficiency loss.

CL can be defined in various ways, but we consider CL with all pairs, adjacent pairs[6] and the first pair of bivariate marginal copula likelihoods of N-dim copula $\mathbf{c}(u_1,\ldots,u_N;\varphi_0)$

$$CL_{all}(u_1,\ldots,u_N)=\prod_{i=1}^{N-1}\prod_{j=i+1}^{N}\mathbf{c}_{i,j}(u_i,u_j;\varphi) \tag{1.12}$$

$$CL_{adj}(u_1,\ldots,u_N)=\prod_{i=1}^{N-1}\mathbf{c}_{i,i+1}(u_i,u_{i+1};\varphi) \tag{1.13}$$

$$CL_{first}(u_1,\ldots,u_N)=\mathbf{c}_{1,2}(u_1,u_2;\varphi) \tag{1.14}$$

where $\mathbf{c}_{ij}(\cdot,\cdot)$ is a bivariate marginal copula of N-dim copula. While there are many different ways to construct composite likelihoods, those all have some common features. First of all, they are valid likelihoods since the likelihood of the submodels or marginal models are involved. Second, the independence assumption for those submodels causes misspecification and information matrix equality does not hold. Third, the computation of the composite likelihood is substantially faster than that of full likelihood. The computation burden, for example, is reduced from $O\left(2^N\right)$ to $O(N)$ when we use adjacent pairs, and $O(N^2)$ when using all pairs, evaluating the density of the jointly symmetric copula constructed by equation (1.7).

Under mild regularity conditions (see Newey and McFadden, 1994 or White, 1994), Cox and Reid (2004) derives the asymptotic behavior of MCLE. For illustration purposes, only CL with adjacent pairs is described in Theorem 4 below, although other CLs could be used.

Theorem 4 (Cox and Reid, 2004)**.** *Assume* $\mathbf{u}_t$ *is iid over* t *and* N *is a fixed number.*

[6] For a given (arbitrary) order of the variables, the "adjacent pairs" CL uses pairs $(u_{i,t},u_{i+1,t})$ for $i=1,\ldots,N-1$.

Consider MCLE defined as

$$\hat{\varphi}_{MCLE} = \arg\max \sum_{t=1}^{T} \sum_{i=1}^{N-1} \log \mathbf{c}_{i,i+1}\left(u_{i,t}, u_{i+1,t}; \varphi\right) \tag{1.15}$$

Under mild regularity conditions, $\hat{\varphi}_{MCLE} \xrightarrow{p} \varphi_0$ *and*

$$\sqrt{T}\left(\hat{\varphi}_{MCLE} - \varphi_0\right) \xrightarrow{d} N\left(0, \mathcal{H}\left(\varphi_0\right)^{-1} \mathcal{J}\left(\varphi_0\right) \mathcal{H}\left(\varphi_0\right)^{-1}\right)$$

where $\mathcal{H}\left(\varphi_0\right) = -E_{\varphi_0}\left[\frac{\partial}{\partial\varphi\partial\varphi'} \sum_{i=1}^{N-1} \log \mathbf{c}_{i,i+1}\left(\cdot;\varphi\right)\right]$ *and*

$\mathcal{J}\left(\varphi_0\right) = Var_{\varphi_0}\left[\frac{\partial}{\partial\varphi} \sum_{i=1}^{N-1} \log \mathbf{c}_{i,i+1}\left(\cdot;\varphi\right)\right]$

We refer to Cox and Reid (2004) for the proof. The consistency and asymptotic normality of $\hat{\varphi}_{MCLE}$ is easily obtained because of the unbiasedness of the score function of CL, which is a linear combination of valid score functions associated with the marginal copula densities forming the composite likelihood:

$$E_{\varphi_0}\left[\frac{\partial}{\partial\varphi} \sum_{i=1}^{N-1} \log \mathbf{c}_{i,i+1}\left(\cdot;\varphi\right)\right] = 0$$

The asymptotic variance of MCLE is a sandwich form and less efficient than MLE by missepcification caused by the independence assumption.

We also note that to identify the parameters, the components of composite likelihoods must be rich enough to include parameters of full likelihood. Suppose that the composite likelihood uses only the first pair like equation (1.14), but φ does not affect the dependence between the first pair. With this CL, φ 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 (1.12) and (1.13) or higher dimension submodels, e.g. trivariate marginal copulas. Throughout the paper, single parameter copulas where only one parameter determines dependence of any pairs are used as baseline copulas for jointly symmetric copulas. In such cases, we have identification from just a single pair of variables, although it leads to an inefficient estimator. Thus we use CL with all pairs or adjacent pairs as well to address issues on inefficiency.

1.3.2 Model selection tests with composite likelihood

In this section, we consider in-sample and out-of-sample model selection tests when composite likelihood is involved. Tests we discuss here are specialized for our empirical analysis in Section 1.5, 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 4. *Given a N-dimension random variable* $\mathbf{Z} = (Z_1, ..., Z_N)$ *with true density* $g(\mathbf{z})$, *the composite Kullback-Leibler information criterion (cKLIC) of a density* $h(\mathbf{z})$ *relative to* $g(\mathbf{z})$ *is*

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

where $\prod_{i=1}^{N-1} g_i(z_i, z_{i+1})$ *and* $\prod_{i=1}^{N-1} h_i(z_i, z_{i+1})$ *are composite likelihood using adjacent pairs corresponding to true density* $g(\mathbf{z})$ *and a density* $h(\mathbf{z})$, *respectively.*

While we focus on CL using adjacent pairs, other composite likelihood such as CL using all pairs can be defined similarly above. We note that the expectation is with respect to the true density $g(\mathbf{z})$ rather than the CL of true density, which makes it possible to interpret cKLIC as a linear combination of the ordinary KLIC of submodels consisting of CL function:

$$
\begin{aligned}
I_c(g,h) &= E_{g(\mathbf{z})}\left[\log \frac{\prod_{i=1}^{N-1} g_i(z_i, z_{i+1})}{\prod_{i=1}^{N-1} h_i(z_i, z_{i+1})}\right] \\
&= \sum_{i=1}^{N-1} E_{g(\mathbf{z})}\left[\log \frac{g_i(z_i, z_{i+1})}{h_i(z_i, z_{i+1})}\right] \\
&= \sum_{i=1}^{N-1} E_{g_i(z_i, z_{i+1})}\left[\log \frac{g_i(z_i, z_{i+1})}{h_i(z_i, z_{i+1})}\right] \qquad (1.16)
\end{aligned}
$$

The last equation holds since submodels or marginal distributions of the true density $g(\mathbf{z})$ are known given the true density $g(\mathbf{z})$. Since cKLIC can be viewed as a linear combination of the ordinary KLIC of submodels, we may use in-sample model selection tests by Vuong (1989) for iid data and Rivers and Vuong (2002) for time series data. To the best of our knowledge, combining cKLIC with Vuong (1989) or Rivers and Vuong (2002) tests is new to the literature.

Let h^A and h^B be two models to be compared. Then the null hypothesis is

$$
\begin{aligned}
& H_0 : E_{g(\mathbf{z})}\left[CL_t^A(\theta_A^*) - CL_t^B(\theta_B^*)\right] = 0 \qquad (1.17) \\
\text{vs. } & H_1 : E_{g(\mathbf{z})}\left[CL_t^A(\theta_A^*) - CL_t^B(\theta_B^*)\right] > 0 \\
& H_2 : E_{g(\mathbf{z})}\left[CL_t^A(\theta_A^*) - CL_t^B(\theta_B^*)\right] < 0
\end{aligned}
$$

where $CL_t^j(\theta_j^*) \equiv \sum_{i=1}^{N-1} \log h_{i,i+1}^j(z_{i,t}, z_{i+1,t}; \theta_j^*)$ for $j = A, B$. It can be shown 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:

$$
\frac{\sqrt{T}\left\{\overline{CL}_T^A\left(\hat{\theta}_A\right) - \overline{CL}_T^B\left(\hat{\theta}_B\right)\right\}}{\hat{\sigma}_T} \to N(0,1) \text{ under } H_0 \qquad (1.18)
$$

where $\overline{CL}_T^j\left(\hat{\theta}_j\right) \equiv \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N-1} \log h_{i,i+1}^j\left(z_{i,t}, z_{i+1,t}; \hat{\theta}_j\right)$, for $j = A, B$ and $\hat{\sigma}_T$ is some consistent estimator of $V\left[\sqrt{T}\left\{\overline{CL}_T^A\left(\hat{\theta}_A\right) - \overline{CL}_T^B\left(\hat{\theta}_B\right)\right\}\right]$, such as HAC estimator such as Newey-West (1987). We note that there might be cases where a test of the null hypothesis based on the full likelihood could give a different answer to one based on the composite likelihood. We leave the study of this possibility for future research.

We may also select the best model in terms of out-of-sample (OOS) forecasting performance measured by some scoring rules. Gneiting and Raftery (2007) introduce "proper" scoring rules which satisfy the condition that the true density receives higher average scores than other densities:

$$E_{g(\mathbf{z})}\left[S\left(h\left(\mathbf{Z}_{t+1}\right)\right)\right] \leqslant E_{g(\mathbf{z})}\left[S\left(g\left(\mathbf{Z}_{t+1}\right)\right)\right]$$

where S is a scoring rule, $g(\mathbf{z})$ is the true density and $h(\mathbf{z})$ is a competing density. The "natural" scoring rule is the log density evaluated out-of-sample, i.e. $S\left(h\left(\mathbf{Z}_{t+1}\right)\right) = \log h\left(\mathbf{Z}_{t+1}\right)$, and it can be shown that this scoring rule is proper. This proper scoring rule is closely related to the KLIC in that equal average scores of two competing models are equivalent to the equal KLICs of those. Since the KLIC measures how close the density forecasts to the true density, the proper scoring rule can be used as a metric to determine which model is more close to the true density.

We may consider a similar scoring rule based on log composite density:

$$S\left(h\left(\mathbf{Z}_{t+1}\right)\right) = \sum_{i=1}^{N-1} \log h_i\left(Z_{i,t+1}, Z_{i+1,t+1}\right) \tag{1.19}$$

and it can be shown to be proper by the following theorem.

Theorem 5. *Consider the composite likelihood with adjacent pairs. The scoring rule based on log composite density given in equation (1.19) is proper, i.e.*

$$E_{g(\mathbf{z})}\left[\sum_{i=1}^{N-1} \log h_i\left(Z_{i,t+1}, Z_{i+1,t+1}\right)\right] \leqslant E_{g(\mathbf{z})}\left[\sum_{i=1}^{N-1} \log g_i\left(Z_{i,t+1}, Z_{i+1,t+1}\right)\right]$$

where the expectation is with respect to the true density $g(\mathbf{z})$, and g_i and h_i are the composite likelihoods of the true density and the competing density, respectively.

The proof is in Appendix A.1. This theorem allows us to interpret that OOS tests based on CL is related to cKLIC just as OOS tests based on full likelihood to KLIC. We may compare values of log CL evaluated at OOS and test the null hypothesis of equal forecasting performance evaluated by log CL:

$$H_0 : E_{g(\mathbf{z})}\left[CL_t^A\left(\theta_A^*\right) - CL_t^B\left(\theta_B^*\right)\right] = 0$$

Our empirical analysis in Section 1.5 employs Giacomini and White (2006) test that incorporates estimation error in the null hypothesis, which punishes a "good" model that is estimated poorly.

$$H_0 : E_{g(\mathbf{z})}\left[CL_t^A\left(\hat{\theta}_{A,t}^*\right) - CL_t^B\left(\hat{\theta}_{B,t}^*\right)\right] = 0$$

We refer to Patton (2012) for general treatments of OOS model selection tests and comparisons under the set-up of Giacomini and White (2006).

1.3.3 Multistage modelling and estimation

In this section, the multistage models for high dimension distributions of returns $\mathbf{r}_t$ with concrete models for $\mathbf{H}_t$ and copulas are considered. Conditional mean and conditional variance-covariance matrix in equation (1.1) are assumed to be modeled using some parametric specification

$$\mu_t \equiv \mu\left(\mathbf{Y}_{t-1}; \theta^{mean}\right), \ \mathbf{Y}_{t-1} \in \mathcal{F}_{t-1}$$
$$\mathbf{H}_t \equiv \mathbf{H}\left(\mathbf{Y}_{t-1}; \theta^{var}\right)$$

This assumption allows for a variety of models for conditional mean, for example, ARMA, VAR, linear and nonlinear regression, and for conditional variance-covariance matrix, for example, the multivariate GARCH-type models such as DCC, BEKK, and

DECO, see Andersen, *et. al* (2006) for a comprehensive review, and the multivariate stochastic volatility models, see Shephard (2005) for a review, as well as the new model proposed in Section 1.2.2.

The standardized uncorrelated residuals in equation (1.1) are defined as

$$\mathbf{e}_t \equiv \mathbf{H}\left(\mathbf{Y}_{t-1};\theta^{var}\right)^{-1/2}\left(\mathbf{r}_t - \mu\left(\mathbf{Y}_{t-1};\theta^{mean}\right)\right)$$

such that $E\left[\mathbf{e}_t|\mathcal{F}_{t-1}\right] = \mathbf{0}$ and $E\left[\mathbf{e}_t\mathbf{e}_t'|\mathcal{F}_{t-1}\right] = \mathbf{I}$, and those are assumed to follow a parametric distribution:

$$\mathbf{e}_t|\mathcal{F}_{t-1} \sim iid\ \mathbf{F}(\cdot) = \mathbf{C}\left(F_1\left(\cdot;\theta_1^{mar}\right),...,F_N\left(\cdot;\theta_N^{mar}\right);\theta^{copula}\right)$$

where marginal distributions F_i are symmetric about zero and the copula $\mathbf{C}$ is jointly symmetric about zero, which together ensures zero correlations of $\mathbf{e}_t$. For zero mean and unit variance of e_{it}, marginal distributions F_i should be standardized.

The parametric specification of $\mu_t, \mathbf{H}_t, F_i$ and a copula $\mathbf{C}$ enables the use of maximum likelihood (ML) estimation:

$$\hat{\theta} = \arg\max_{\theta} \log L_T(\theta)$$

$$where\ \ \log L_T(\theta) = \sum_{t=1}^{T} \log l_t\left(\mathbf{r}_t|\mathcal{F}_{t-1};\theta\right)$$

$$\log l_t\left(\mathbf{r}_t|\mathcal{F}_{t-1};\theta\right) = \sum_{i=1}^{N} \log f_{it}\left(e_{it}\left(\theta^{mean},\theta^{var}\right);\theta_i^{mar}\right)$$

$$+\log \mathbf{c}\left(F_{1t}\left(e_{1t}\left(\theta^{mean},\theta^{var}\right);\theta_1^{mar}\right),..,F_{Nt}\left(e_{Nt}\left(\theta^{mean},\theta^{var}\right);\theta_N^{mar}\right);\theta^{cop}\right)$$

However, when N is large, this one-stage joint estimation is not feasible, and multi-stage ML estimation could be an alternative.

To save space and focus on the conditional variance and residuals, θ^{mean} is assumed to be known. (For example, a common approach is to assume daily returns have a zero mean.) Since equation (1.9), (1.10), and (1.11) are used for the specification of conditional variance $\mathbf{H}_t$ when high frequency data is available and the DCC model is used for daily data (see Appendix A.2 for details about the DCC model), conditional univariate variances and conditional correlations can be separately modeled. θ^{var} can be written as $[\theta_1^{var}, \ldots, \theta_N^{var}, \theta^{corr}]$ where θ_i^{var} denotes parameters in equation (1.10) or (A.5) and θ^{corr} denotes parameters in (1.11) or (A.8). The parameters to estimate are gathered in θ :

$$\theta \equiv \left[\begin{array}{cccccccc} \theta_1'^{var} & \ldots & \theta_N'^{var} & \theta'^{corr} & \theta_1'^{mar} & \ldots & \theta_N'^{mar} & \theta'^{cop} \end{array} \right]'$$

To allow for estimation in separate stages, those parameters are assumed to appear only in their own stages.

The specific multistage estimation is as follows. The first stage is for individual variances of $\mathbf{r}_t$, and θ_i^{var} is estimated using demeaned $\{r_{it}\}_{t=1}^T$ for each i = 1,...,N. The second stage is for correlations of $\mathbf{r}_t$, and θ^{corr} is estimated using demeaned and de-volatilized $\{\mathbf{r}_t\}_{t=1}^T$. The third stage is for marginal distributions of estimated standardized uncorrelated residuals $\hat{\mathbf{e}}_t \equiv \mathbf{H}\left(\mathbf{Y}_{t-1}; \hat{\theta}^{var}\right)^{-1/2}(\mathbf{r}_t - \mu_t)$, and θ_i^{mar} is estimated using $\{\hat{e}_{it}\}_{t=1}^T$ for each i = 1,...,N. The last stage is for copula estimation using so called "probability integral transforms" of $\hat{\mathbf{e}}_t$, i.e. $\left\{\left[F_1\left(\hat{e}_{1t}; \hat{\theta}_1^{mar}\right), \ldots, F_N\left(\hat{e}_{Nt}; \hat{\theta}_N^{mar}\right)\right]\right\}_{t=1}^T$. The estimates of previous stages are substituted in likelihood functions of next stage and the likelihood function of each stage is denoted by l_t^{stage} :

$$\hat{\theta}_i^{var} \equiv \arg\max_{\theta_i^{var}} \sum_{t=1}^{T} \log l_{it}^{var}\left(\theta_i^{var}\right), \; i = 1, \ldots, N$$

$$\hat{\theta}^{corr} \equiv \arg\max_{\theta^{corr}} \sum_{t=1}^{T} \log l_t^{corr}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \theta^{corr}\right)$$

$$\hat{\theta}_i^{mar} \equiv \arg\max_{\theta_i^{mar}} \sum_{t=1}^{T} \log l_{it}^{mar}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \theta_i^{mar}\right), \;\; i = 1, \ldots, N$$

$$\hat{\theta}^{cop} \equiv \arg\max_{\theta^{cop}} \sum_{t=1}^{T} \log l_t^{cop}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, \ldots, \hat{\theta}_N^{mar}, \theta^{cop}\right)$$

If composite likelihood is used for the last stage as we discussed in Section 1.3.1, then l_t^{cop} is composite likelihood rather than full likelihood, but nothing else changes. The estimation errors of previous stages do not affect consistency of estimators of next stages since the consistency of the previous stages guarantees the consistency of the next stage estimators if the likelihood function of the next stage is smooth enough around true parameters. However, the estimation errors of previous stages are accumulated and affect the asymptotic variance of multistage ML estimators and the following theorem with Appendix A.3 explicitly show how multistage estimations influence asymptotic variance.

Theorem 6. *Assume that conditions of Theorem 6.1 of Newey-McFadden (1994) are all satisfied, and that* $\hat{\theta}_{MSML} \equiv \left[\hat{\theta}_1'^{var}, \ldots, \hat{\theta}_N'^{var}, \hat{\theta}'^{corr}, \hat{\theta}_1'^{mar}, \ldots, \hat{\theta}_N'^{mar}, \hat{\theta}'^{cop}\right]'$ *is consistent. Then*

$$\sqrt{T}\left(\hat{\theta}_{MSML} - \theta^*\right) \xrightarrow{d} N\left(0, V_{MSML}^*\right) \;\; as \; T \to \infty$$

and $\hat{V}_{MSML} \xrightarrow{p} V^*_{MSML}$ *where* $\hat{V}_{MSML} = \hat{A}_T^{-1}\hat{B}_T\left(\hat{A}_T^{-1}\right)'$:

$$\hat{B}_T = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{s}}_t\hat{\mathbf{s}}_t'$$

$$where\ \hat{\mathbf{s}}_t \equiv \left[\begin{array}{cccccccc} \hat{\mathbf{s}}_1'^{var} & \ldots & \hat{\mathbf{s}}_N'^{var} & \hat{\mathbf{s}}'^{corr} & \hat{\mathbf{s}}_1'^{mar} & \ldots & \hat{\mathbf{s}}_N'^{mar} & \hat{\mathbf{s}}'^{copula} \end{array}\right]'$$

$$\hat{\mathbf{s}}_i'^{var} = \frac{\partial}{\partial\theta_i^{var}}\log l_{it}^{var}\left(\hat{\theta}_i^{var}\right), \quad i = 1, 2, \ldots, N$$

$$\hat{\mathbf{s}}'^{corr} = \frac{\partial}{\partial\theta^{corr}}\log l_t^{corr}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}\right)$$

$$\hat{\mathbf{s}}_i'^{mar} = \frac{\partial}{\partial\theta_i^{mar}}\log l_{it}^{mar}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_i^{mar}\right), \quad i = 1, 2, \ldots, N$$

$$\hat{\mathbf{s}}'^{copula} = \frac{\partial}{\partial\theta^{cop}}\log l_t^{cop}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, \ldots, \hat{\theta}_N^{mar}, \hat{\theta}^{cop}\right)$$

and

$$\hat{A}_T = \frac{1}{T}\sum_{t=1}^{T}\hat{P}_t$$

see Appendix A.3 for the specific form of $\hat{P}_t$.

Since multistage ML estimation can be viewed as multistage GMM estimation, we refer to Section 6.1 of Newey and McFadden (1994) for detailed discussion and proofs. For inference, $\hat{V}_{MSML}$ is necessary but calculation is not feasible for high dimensions even if the analytical form is known as in Theorem 6. For example, the proposed model used in Section 1.5 for empirical analysis has more than 5000 parameters to estimate, and $\hat{V}_{MSML}$ is larger than a 5000×5000 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 multistage GMM estimators. Although that bootstrap is not expected to have an asymptotic refinement relative to the standard approach, it allows us to avoid having to compute a large

Hessian matrix. The steps are following: (i) generate bootstrap sample of length T using block bootstrap such as stationary bootstrap, see Politis and Romano (1994) or other methods which can preserve time-series structure, and estimate parameters θ with bootstrap samples. Repeat S times (e.g. $S = 500$) and use the quantiles of $\left\{\hat{\theta}_i\right\}_{i=1}^{S}$ as critical values or use $\alpha/2$ and $(1-\alpha/2)$ quantiles of $\left\{\hat{\theta}_i\right\}_{i=1}^{S}$ to obtain $(1-\alpha)$ confidence intervals for parameters.

1.4 Simulation study

In Section 1.4.1, we study finite sample properties of maximum composite likelihood estimators (MCLEs) defined in equation (1.15) for jointly symmetric copula constructed by equation (1.7) through an extensive Monte Carlo simulations for up to one hundred dimensions. In Section 1.4.2, we illustrate the theoretical results of Section 1.3.3 on multistage estimation through simulations with realistic settings.

1.4.1 Finite sample properties of MCLE for jointly symmetric copulas

In this section, we mainly focus on examining the following. First, how big or small is the efficiency loss of MCLE compared to MLE. Second, which one is best to use among three different MCLE constructed in equation (1.12), (1.13) and (1.14) according to accuracy and computation time. Third, how useful is the cross-sectional information for copula estimations as dimension increases.

The data generating process is as follows. A vector $[u_1, u_2, .., u_N]$ is generated from N dimension copula. To make those data jointly symmetric, choose u_i or $1-u_i$ with 1/2 probability for each i = 1,..., N

$$[\tilde{u}_1, \tilde{u}_2, .., \tilde{u}_N]\,, \text{ where } \tilde{u}_i = \begin{cases} u_i & \text{with prob } 1/2 \\ 1-u_i & \text{with prob } 1/2 \end{cases} \tag{1.20}$$

After repeating T times, T by N data is simulated from jointly symmetric copulas.

We consider two jointly symmetric copulas based on Clayton and Gumbel copulas and time series length $T = 1000$ with dimension $N = 2, 3, 5, 10, 20, ..., 100$. Four different estimation methods are applied to the simulated data: MLE, MCLE with all pairs in equation (1.12), MCLE with adjacent pairs in equation (1.13), and MCLE with the first pair in equation (1.14). We repeat these simulations and estimations five hundred times and report bias and standard deviations of those five hundred estimates with computation times in Table 1.2. While MLE is not feasible for $N \geqslant 20$ due to huge computation burdens, the other MCLEs are feasible and very fast even for $N = 100$, see the last four columns of Table 1.2.

The average biases for all dimensions and for all estimation methods are small relative to the standard deviations except for MCLE with the first pair. The standard deviations play a role in a measure of estimator accuracy and those show that for the low dimension ($N \leqslant 10$) , not surprisingly, MLE has smaller standard deviations than three MCLEs and the relative efficiency of MCLE with all pairs to MLE is 1.05 to 1.37, which is moderate. Among three MCLEs, MCLE with all pairs has the smallest standard deviations whereas 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 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) supporting MCLE with adjacent pairs in the DCC model.

Figure 1.4 indicates biases and standard deviations of four estimations as the dimension N increases. Biases of MCLE with all and adjacent pairs are very similar and standard deviations of those two MCLEs quickly decrease and the difference of those gets smaller as N increases. Compared to the standard deviation of MCLE with the first pair staying flat, the other two MCLEs exploits efficiency gains from cross sectional information, which is intuitive because dependence of any pairs is informative for estimating copula parameters.

In sum, MCLE is less efficient but feasible and very fast for high dimensions, and MCLE gets significant efficiency gains as N increases. While the accuracy of MCLE with adjacent pairs is almost similar to that of MCLE with all pairs, especially for high dimensions, the increase in computation is quite large. For this reason, we use MCLE with adjacent pairs for our empirical analysis in Section 1.5.

1.4.2 Finite sample properties of multistage estimation

In this section, we study the multistage estimation for the proposed model with simulated data from the following set up:

$$\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t \tag{1.21}$$

$$\mathbf{H}_t \equiv Cov\left[\mathbf{r}_t|\mathcal{F}_{t-1}\right]$$

$$\mathbf{e}_t|\mathcal{F}_{t-1} \sim iid\ \mathbf{F}\left(\cdot\right) = \mathbf{C}\left(F_1\left(\cdot;\nu_1\right), ..., F_N\left(\cdot;\nu_N\right);\varphi\right)$$

where the mean part is assumed zero, the variance-covariance part $\mathbf{H}_t$ follows the DCC model with GARCH(1,1), see Appendix A.2 with $\zeta_i = 0$, F_i is standardized Student's t distribution with $\nu_i = 6$ and $\mathbf{C}$ is a jointly symmetric copula constructed by equation (1.6) with Clayton copula with $\varphi = 1$. For realistic set up, we use some estimated parameter values from the results of empirical analysis in Section 1.5. The parameters of equation (A.6) and (A.7) for GARCH and DCC models are set as $[\psi_i, \kappa_i, \lambda_i] = [0.05, 0.1, 0.85]$ and $[\alpha, \beta] = [0.02\ 0.95]$, and $\overline{\mathbf{Q}}$ is set to be the unconditional correlations of 100 stock returns that used in the next section. We first simulate data from the jointly symmetric copula following the way described in the previous section, and then using inverse standardized Student's t distribution, transform those data into uncorrelated $\mathbf{e}_t$. Then we recursively update DCC model by equation (A.7) and (A.8) to generate correlation matrix, and apply GARCH effects by equation (A.6). Then, the simulated $\mathbf{e}_t$ can be easily transformed to $\mathbf{r}_t$ whose conditional covariance matrix is following the DCC model.

We follow the multistage estimation described in Section 1.3.3. 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. From those two stages, the estimated standardized uncorrelated residuals $\hat{\mathbf{e}}_t$ are obtained, and those are used to estimate marginal distributions. At the last stage, the copula parameters are estimated by MCLE with adjacent pairs explained in Section 1.3.1. We repeat this scenario 500 times with time series of length $T = 1000$ and cross sectional dimension $N = 10, 50$, and 100. Table 1.4 reports all parameter estimates except $\overline{\mathbf{Q}}$. 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 set to the same numbers across cross sections.

Table 1.4 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 is 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 multistage estimation is quite fast, for example, it takes five minutes for one hundred dimension model in which the 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 copula estimations in Table 1.4 to the corresponding results in Table 1.2. The standard deviation increases by about 30% for $N = 10$, and by about 19% for $N = 50$ and 100. This loss of accuracy caused by having to estimate parameters of the marginals is considerably small given that more than 5000 parameters are estimated in the former stages. We conclude that multi-stage estimations with composite likelihoods result in a large gain in computation and a small loss in estimation error and efficiency.

1.5 Empirical analysis of S&P 100 equity returns

In this section, the proposed multivariate distribution model is applied to equity returns of constituents of S&P 100. The sample period is January 2006 until December 2012, a total of $T = 1761$ trade days. All companies listed at least once on S&P100 over the sample period are considered, but only 104 are selected after excluding companies that were not traded during the whole sample period. The stocks are listed in Table 1.5 with their 3-digit SIC codes. We obtain high frequency transaction data from NYSE's TAQ database, clean it following Barndorff-Nielsen, *et al.* (2009), see Li (2013) for details, and adjust prices affected by splits and dividends using "adjustment" factors from CRSP. Daily returns are calculated by 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.

Table 1.6 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 thick tails. In the middle 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. The lower panel shows the estimates of the parameters of AR(1) models. Constant terms are estimated around zero and the estimates of AR(1) coefficient are slightly negative, both are consistent with values in other studies.

We estimate two different models for conditional variance-covariance matrix $\mathbf{H}_t$, HAR-type model described in Section 1.2.2 and the DCC model in Appendix A.2. While the latter is estimated with (low frequency) demeaned daily returns, the former is estimated with (high frequency) 5-min returns. Table 1.7 presents the estimates of two models in Panel A and Panel B, respectively. The estimates of variance part for HAR-type models in Panel A are similar to those reported in Corsi (2009): coefficients on day, week, and month being around 1/3 and coefficient on day being the largest. In the correlations part, however, the coefficient on month is the largest followed by week and day. The parameter estimates for the DCC model in Panel B are close to other studies of daily stock returns: indicated volatility clustering, asymmetric volatility dynamics, and highly persistent time-varying correlations. The bootstrap standard errors described in Section 1.3.3 are provided in Table 1.7, and those take into account the estimation errors of former stages.

Two data sets of standardized uncorrelated residuals are constructed[7] and summary statistics are reported in the top panel of Table 1.8:

$$\hat{\mathbf{e}}_{t,HAR} \equiv \hat{\mathbf{H}}_{t,HAR}^{-1/2}\left(\mathbf{r}_t - \hat{\mu}_t\right)$$

$$\hat{\mathbf{e}}_{t,DCC} \equiv \hat{\mathbf{H}}_{t,DCC}^{-1/2}\left(\mathbf{r}_t - \hat{\mu}_t\right)$$

The next stage is for the marginal distributions for those residuals. Before specifying marginal distributions, tests for zero skewness and zero excess kurtosis are conducted and reported in Panel B of Table 1.8. It is found that only 4 (or 6) out of 104 cross sectional residuals of $\hat{\mathbf{e}}_{t,HAR}$ (or $\hat{\mathbf{e}}_{t,DCC}$) are rejected at 5% level for zero skewness test, motivating the use of symmetric marginal distributions. In addition, all of them are rejected under 5% level for zero excess kurtosis test, which suggests to use marginal distributions with thick tails. Thus, standardized Student's t distributions are used for marginal distributions of the estimated standardized uncorrelated residuals.

The top panel of Table 1.9 presents the cross-sectional quantiles of 104 estimated

[7] For square root of matrix, the spectral decomposition rather than Cholesky decomposition is employed due to its invariance to the order of the variables.

degrees of freedom parameters of standardized Student's t distributions, ranging from 4.1 at 5 % quantile to 6.9 at 95% quantile for $\hat{\mathbf{e}}_{t,HAR}$ and from 4.2 at 5% quantile to 8.3 at 95% for $\hat{\mathbf{e}}_{t,DCC}$, indicating excess kurtosis of standardized uncorrelated residuals.

The last stage is the estimation of copula designed to capture nonlinear dependence. Four jointly symmetric copulas based on t, Clayton, Frank, and Gumbel copulas are used. While jointly symmetric copulas based on Clayton, Frank and Gumbel are constructed by equation (1.7), the one based on t copula is simply constructed by substituting correlation matrix of t copula with identity matrix. To see whether those models outperform the existing model in the literature, we use two benchmark models: the independence copula and the multivariate Student's t distribution. The independence copula is a special case of jointly symmetric copula, and there is no parameter to estimate.[8] Since the independence copula completely ignores nonlinear dependence, we can see if there is substantial nonlinear dependence by comparing those four jointly symmetric copulas with the independence copula. In addition, to see whether or not the copula approach outperforms a non-copula approach incapable of separately specifying marginals and dependence, the multivariate Student's t distribution is employed as another benchmark. The bottom panel of Table 1.9 reports the parameter estimates for jointly symmetric copulas and the multivariate (standardized) t distribution with bootstrap standard errors in parenthesis that incorporate accumulated estimation errors from former stages. We follow steps explained in Section 1.3.3 to obtain bootstrap standard errors and the average block length for the stationary bootstrap is set to 100.

To see whether nonlinear dependence exists, we propose formal tests for the null hypothesis that there is *no* nonlinear dependence. Since those four jointly symmetric

[8] The independence copula is a product of its arguments, i.e. $C(u_1, ...u_N) = u_1 \times \cdots \times u_N$, and its density is 1.

copulas and the multivariate t distribution nest the independence copula, the null hypotheses are: $H_0 : \frac{1}{\theta^{JS\_t}} = 0$ for jointly symmetric copulas based on t copula, $H_0 : \theta^{JS\_Clayton} = 0$ for those based on Clayton copula, $H_0 : \theta^{JS\_Frank} = 0$ for those based on Frank copula, $H_0 : \theta^{JS\_Gumbel} = 1$ for those based on Gumbel, and $H_0 : \frac{1}{\theta^{MV\_t}} = 0$ for the multivariate (standardized) t distribution. The t-statistics for those tests are reported in the bottom panel of Table 1.9. We note, however, that the parameters are all on the boundary of the parameter space, 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. All of those null hypotheses above are rejected at 1% level, and we conclude that there is substantial nonlinear cross-sectional dependence in daily returns.

To compare those models, we consider the multivariate log-likelihood of daily returns that can be decomposed into three parts by change of variables: log absolute values of determinant of square root of inverse variance-covariance matrix of daily returns, sum of log likelihoods of marginal distributions for standardized residuals, and log composite likelihood of a copula for standardized residuals:

$$\mathbf{r}_t = \mathbf{H}_t^{1/2}\mathbf{e}_t$$

$$f\left(r_{1t}, ..., r_{Nt}\right) = \left|\det\left(\mathbf{H}_t^{-1/2}\right)\right| g\left(e_{1t}, \ldots, e_{Nt}\right)$$

$$= \left|\det\left(\mathbf{H}_t^{-1/2}\right)\right| \times g_1\left(e_{1t}\right) \times .. \times g_N\left(e_{Nt}\right) \times \mathbf{c}\left(G_1\left(e_{1t}\right), .., G_N\left(e_{Nt}\right)\right)$$

$$\log f\left(r_{1t}, ..., r_{Nt}\right) = \log\left|\det\left(\mathbf{H}_t^{-1/2}\right)\right| + \sum_{i=1}^{N} \log g_i\left(e_{it}\right) + \log \mathbf{c}\left(G_1\left(e_{1t}\right), .., G_N\left(e_{Nt}\right)\right)$$

Table 1.10 reports those three parts of log likelihoods, summing those to obtain the entire log likelihoods for daily return distributions. Comparing the values of the entire log likelihoods, we reasonably expect three findings. First, copula methods seem to outperform the multivariate t distribution that does not explicitly use a copula, which can be confirmed by comparing jointly symmetric copulas with the multivariate t distribution. Second, by comparing first four jointly symmetric copula models with the independence copula model, it can be seen that the models where nonlinear dependence is captured outperform the models that ignore it. Third, high frequency data seem to help improve model fits better than daily data. To formally verify those expectations, we next conduct in-sample and out-of-sample model comparison tests.

1.5.1 In-sample model selection

Table 1.11 presents t-statistics from Rivers and Vuong (2002) (pair-wise) model comparison tests[9] introduced in Section 1.3.2 for composite likelihood. A positive t-statistic indicates that the model above beat 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 or not copula approaches outperform non-copula ones represented by the multivariate t distribution. All t-statistics in that row are positive and larger than 15, which strongly supports that copula approaches significantly outperform non-copula ones. The independence copula as well as the four jointly symmetric copulas can separately specify 104 marginal distributions and dependence, which allows for much more flexibility to the model. In contrast, the multivariate t distribution forces 104 marginal distributions and dependence to be bound to each other,[10] which results in the inferiority of it to the models based on copulas. The multivariate t distribution is widely used as an alternative to Normal distribution not only in the literature but

[9] Those tests can be easily extended for the models combining log likelihoods for marginal distributions and log composite likelihood for copulas by re-defining Kullback-Leibler information criterion.

[10] The multivariate t distribution can be viewed as univariate t distributions coupled by t copula with a constraint that all degrees of freedom parameters for margins and copula should be identical.

also in practice due to its thick tails and non-zero tail dependence. It is, however, seen that the proposed copula-based models significantly beat the multivariate t distribution. This outperformance is true whether the HAR model using 5-min data to forecast $\mathbf{H}_t$ is used (see the bottom row of upper panel) or the DCC model using daily data is used (see the right half of the bottom row of lower panel). Next, to see whether or not nonlinear dependence improves model fits, we compare four jointly symmetric copula models designed to capture nonlinear dependence with the independence copula that completely ignores nonlinear dependence. The second bottom row of the upper panel and the right half of the second bottom row of the bottom panel are for those comparisons, and t-statistics in those rows are all positive and significant under 1% level, which implies that capturing non-linear dependence is quite useful to improve model fits.

Lastly, to see whether forecasts of $\mathbf{H}_t$ using high frequency data results in better model fits than forecasts of $\mathbf{H}_t$ using daily data does, the left half of the bottom panel is explored. All t-statistics are positive and significant at 1% level, and this implies that any model using high frequency data for forecasting $\mathbf{H}_t$ significantly outperforms models using daily data for forecasting $\mathbf{H}_t$. We note that even the multivariate t distribution combined with forecasts of $\mathbf{H}_t$ using high frequency data outperforms any models that use daily data for forecasts of $\mathbf{H}_t$, which means that information of high frequency data substantially improve performance of models.

In this section we verify the following three findings. First, copula methods that allow one to separately specify marginal distributions and dependence significantly outperform non-copula methods. Second, nonlinear dependence is so considerable and important that it plays a critical role in improving model fits. Third, accurately measuring and forecasting linear dependence through high frequency data considerably increase performance of models.

1.5.2 Out-of-sample model selection

The previous section revealed that the proposed models significantly beat benchmark models in in-sample model comparison tests. However, since it is essentially a forecasting model for daily return distributions, it has to be investigated whether it has superior out-of-sample (OOS) forecasting performance. In this section, we consider a multivariate density forecasting based on out-of-sample log (composite) likelihoods to compare 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. Giacomini and White (2006) test described in Section 1.3.2 requires rolling window or fixed window estimation scheme rather than expanding window one. To incorporate structural changes, we employ a rolling window rather than a fixed window. We estimate the whole model using the data in the interval $[t - R + 1, t]$ and evaluate the model using the data at $t + 1$ with those estimates at each time in out-of-sample period. We iterate 502 times for these estimations and evaluations.

Out-of-sample density forecast comparisons

Table 1.12 presents t-statistics from pair-wise OOS model comparison tests. Similar to the results from in-sample tests, we discover three finding again. Copula models are significantly better than the multivariate t distribution, and jointly symmetric copula models significantly outperform the independence copula model. Also models using information of high frequency data significantly beat models using information of daily data.

These results reveal that three major components of the proposed model, separately specifying marginal distribution and dependence, capturing nonlinear dependence and exploiting information of high frequency data lead to improved forecasting performance.

Out-of-sample portfolio decision making

To investigate a economic gain of the proposed model, we consider asset allocation problems in an out-of-sample setting proposed by Patton (2004). The basic idea is simple: a better forecasting model should lead to a better portfolio decision.

We introduce a hypothetical portfolio of 104 stocks listed in Table 1.5 and assume that an investor maximizes his expected utility by choosing optimal portfolio weights on 104 stocks. The utility functions for the investor are the class of CRRA (constant relative risk averse) utility functions:

$$\mathcal{U}(W) = \begin{cases} \frac{W^{1-\rho}}{1-\rho} & if\ \rho \neq 1 \\ \log(W) & if\ \rho = 1 \end{cases}$$

where ρ is a relative risk aversion parameter and W is wealth. Optimal portfolio weights are determined by maximizing the expected utility under the multivariate predictive density for $\mathbf{r}_{t+1}$

$$\omega^*_{t+1} = \arg\max_{\omega \in \mathcal{W}} E_t \left[\mathcal{U}\left(W_0\left(1 + \omega' \mathbf{r}_{t+1}\right)\right)\right] \tag{1.22}$$

where ω is $N \times 1$ portfolio weights, W_0 is initial wealth and $\mathcal{W} = \left\{\omega \in [0,1]^N : \mathbf{1}'\omega \leqslant 1\right\}$. For more realistic settings, we only consider an investor with short-sale constraint. Since the conditional expectation above is taken with respect to conditional distributions of next period returns $\mathbf{r}_{t+1}$, we may expect a better forecasting model (conditional distribution) for $\mathbf{r}_{t+1}$ to give better portfolio weights which generate higher average utilities. By comparing those average utilities, we may pick up better forecasting models. However, utility is not intuitively interpretable, so we convert the average utility to a “management fee”, which is a fixed amount that could be charged (or paid) each period making the investor indifferent between portfolio A and portfolio B. The management fee $\mathcal{C}$ is the solution to the following equation:

$$\frac{1}{P}\sum_{t=R+1}^{R+P}\mathcal{U}\left(1+\omega_{A,t+1}^{*\prime}\mathbf{r}_{t+1}\right)=\frac{1}{P}\sum_{t=R+1}^{R+P}\mathcal{U}\left(1+\omega_{B,t+1}^{*\prime}\mathbf{r}_{t+1}-\mathcal{C}\right)$$

where initial wealth W_0 sets to be 1, R is the length of the in-sample period, and P is the length of the out-of-sample period.

We keep the same R and P as in the previous section, and RRA parameter ρ sets to be 7. We obtain the conditional expectation in equation (1.22) through Monte Carlo integrals using simulated data from estimated models in the previous section.

Table 1.13 presents the estimated management fee $\mathcal{C}$ in annualized percent between any two models of twelve competing models. A positive number indicates that the model above outperforms the model to the left, and a negative one indicates the opposite. We compare copula models and the multivariate t distribution to see whether separately specifying marginals and dependence is influential. Portfolio decisions based on the multivariate t distribution yields smaller economic gains than those from copula based models except one based on Clayton copula. The gains by changing models from non-copula models to copula models are from 0.48% to 2.42%. Second, we find that models that use high frequency data come up with higher economic gains than models that do not use high frequency data. The gains range from 0.4% to 6.3%. This confirms the superiority of models capable of employing high frequency data to models incapable of using high frequency data. Lastly, to see how important nonlinear dependence is, we compare copula models with the independence copula. The copula model based on t copula beats the independence copula whereas the other copula models do not. This suggests properly capturing nonlinear dependence generates higher economic gains. Overall, models based on t copula with high frequency data outperform all other models. As aforementioned, the model based on t copula substantially differs from the benchmark model, the multivariate t distribution in that the latter does not separately specify marginal distributions and dependence whereas the former does.

In sum, we find the strong evidence of usefulness of high frequency data and copula approaches and mild evidence of importance of nonlinear dependence under the portfolio decision problems. Through the realistic portfolio decision problems, the proposed model proves to have an excellent forecasting capability which in turn generates large economic gains.

1.6 Conclusion

This paper proposes a new general model for high dimension distributions of daily asset returns that utilizes high frequency data and copulas. The decomposition of dependence into linear and nonlinear dependence makes it possible to fully exploit advantages of high frequency data and copulas. Linear dependence is accurately measured and forecasted by high frequency data whereas nonlinear dependence can be captured by a new class of copulas for linearly uncorrelated residuals. By assigning two different tasks to high frequency data and copulas, this separation significantly improves the performance of models for joint distributions. In addition, the new class of copulas for uncorrelated variables is proposed which is a rich set of copulas for studying dependence of uncorrelated but dependent variables. Though those copulas can be easily constructed by simple rotations of any given copulas, those rotations may cause serious computation burden in high dimensions. We address computation issues by employing composite likelihoods and multistage estimations. Via an extensive Monte Carlo study, we show that multistage estimation with composite likelihood results in small loss in efficiency and large gain in computation speed especially for high dimensions.

We employ our proposed models to study daily return distributions of 104 constituents of the S&P 100 index over the period 2006 to 2012. We confirm the statistical superiority through in- and out-of-sample tests, and we find large economic gains in asset allocation decisions based on the proposed model in an out-of-sample setting. The excellence of the proposed model can be explained by three keywords: copula approaches, nonlinear dependence, and high frequency data. The multivariate t or Normal distribution generally used in the literature is significantly beaten by the proposed model that utilizes benefits of copulas. Surprisingly, nonlinear dependence mostly ignored in the literature turns out to have fairly valuable information which improves the performance of models. Furthermore, linear dependence accurately measured and forecasted by high frequency data considerably enhances the proposed model.

1.7 Tables and figures

Table 1.1: Computation time of jointly symmetric copula

N	10	20	50	100
Full likelihood	0.23 sec	4 min	10^6 years	10^{17} years
Composite likelihood using all pairs	0.05 sec	0.21 sec	1.52 sec	5.52 sec
Composite likelihood using adj. pairs	0.01 sec	0.02 sec	0.06 sec	0.11 sec

Note: Computation time for one evaluation of the density of jointly symmetric copula based on Clayton copula.

Table 1.2: Simulation results for jointly symmetric copula based on Clayton

	Bias				Std dev				Average Run Time (in sec)		
N	MLE	$\underset{all}{MCLE}$	$\underset{adj}{MCLE}$	$\underset{first}{MCLE}$	MLE	$\underset{all}{MCLE}$	$\underset{adj}{MCLE}$	$\underset{first}{MCLE}$	MLE	$\underset{all}{MCLE}$	$\underset{adj}{MCLE}$
2	-0.0027	-0.0027	-0.0027	-0.0027	0.1176	0.1176	0.1176	0.1176	0.15	0.15	0.15
3	-0.0019	-0.0028	-0.0031	-0.0027	0.0798	0.0839	0.0917	0.1176	0.42	0.50	0.24
5	-0.0014	-0.0022	-0.0016	-0.0027	0.0497	0.0591	0.0713	0.1176	1.96	1.49	0.43
10	-0.0051	-0.0047	-0.0039	-0.0027	0.0293	0.0402	0.0495	0.1176	116	7	1
20		-0.0018	-0.0021	-0.0027		0.0365	0.0405	0.1176		27	2
30		-0.0036	-0.0037	-0.0027		0.0336	0.0379	0.1176		63	3
40		-0.0028	-0.0037	-0.0027		0.0311	0.0341	0.1176		117	5
50		-0.0011	-0.0014	-0.0027		0.0298	0.0329	0.1176		192	6
60		-0.0007	-0.0006	-0.0027		0.0314	0.0332	0.1176		256	7
70		-0.0013	-0.0013	-0.0027		0.0306	0.0324	0.1176		364	8
80		-0.0039	-0.0041	-0.0027		0.0309	0.0332	0.1176		471	9
90		0.0012	0.0013	-0.0027		0.0312	0.0328	0.1176		611	11
100		-0.0006	-0.0003	-0.0027		0.0290	0.0305	0.1176		748	12

Note: This table presents the results from 500 simulations of jointly symmetric copula based on Clayton copula with true parameter 1. Four different estimations are used: MLE, MCLE with all pairs, MCLE with adjacent pairs, MCLE with the first pair. Because MLE takes too much time for $N \geqslant 20$, the results for MLE are reported only for low dimension $(N \leqslant 10)$ Dimensions are from $N = 2$ to $N = 100$ and the sample size is $T = 1000$. 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.

Table 1.3: Simulation results for jointly symmetric copula based on Gumbel

	Bias				Std dev				Average Run Time (in sec)		
N	MLE	$MCLE_{all}$	$MCLE_{adj}$	$MCLE_{first}$	MLE	$MCLE_{all}$	$MCLE_{adj}$	$MCLE_{first}$	MLE	$MCLE_{all}$	$MCLE_{adj}$
2	-0.0016	-0.0016	-0.0016	-0.0016	0.0757	0.0757	0.0757	0.0757	0.30	0.13	0.13
3	-0.0021	-0.0018	-0.0023	-0.0016	0.0484	0.0508	0.0583	0.0757	0.71	0.43	0.29
5	-0.0041	-0.0025	-0.0025	-0.0016	0.0368	0.0409	0.0470	0.0757	3.52	1.31	0.53
10	-0.0021	-0.0023	-0.0016	-0.0016	0.0245	0.0328	0.0369	0.0757	153	6	1
20		-0.0019	-0.0021	-0.0016		0.0285	0.0312	0.0757		25	2
30		-0.0019	-0.0022	-0.0016		0.0277	0.0297	0.0757		61	4
40		-0.0019	-0.0022	-0.0016		0.0270	0.0285	0.0757		97	5
50		-0.0024	-0.0027	-0.0016		0.0269	0.0283	0.0757		166	7
60		-0.0021	-0.0023	-0.0016		0.0267	0.0282	0.0757		236	8
70		-0.0022	-0.0024	-0.0016		0.0264	0.0276	0.0757		326	9
80		-0.0022	-0.0023	-0.0016		0.0262	0.0272	0.0757		435	11
90		-0.0021	-0.0022	-0.0016		0.0262	0.0272	0.0757		509	11
100		-0.0020	-0.0021	-0.0016		0.0261	0.0272	0.0757		664	13

Note: This table presents the results from 500 simulations of jointly symmetric copula based on Gumbel copula with true parameter 2. Four different estimations are used: MLE, MCLE with all pairs, MCLE with adjacent pairs, MCLE with the first pair. Because MLE takes too much time for $N \geqslant 20$, the results for MLE are reported only for low dimension ($N \leqslant 10$) Dimensions are from $N = 2$ to $N = 100$ and the sample size is $T = 1000$. 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.

Table 1.4: Simulation results for multistage estimations

	Variance			Correlation		Marginal	Copula
	Const	ARCH	GARCH	DCC α	DCC β	t dist	JS Clay.
	ψ_i	κ_i	λ_i	α	β	ν_i	φ
True	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
Med	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
Diff	0.0756	0.0898	0.1397	0.0144	0.0435	2.4656	0.1370
Time	1 min						
				$N = 50$			
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
Med	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
Diff	0.0751	0.0931	0.1410	0.0098	0.0267	2.4769	0.0899
Time	2 min						
				$N = 100$			
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
Med	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
Diff	0.0756	0.0928	0.1415	0.0084	0.0229	2.4471	0.0811
Time	5 min						

Note: This table presents the results from 500 simulations of multistage models described in Section 1.3.3. Sample size is $T = 1000$ and cross-sectional dimensions are $N = 10, 50$, 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 50^{th}, 90^{th} and 10^{th} percentiles of the distribution of estimated parameters, and the sixth row presents the difference between the 90^{th} and 10^{th} percentiles. The final row presents estimation time per each simulation

Table 1.5: 104 Stocks used in the empirical analysis

Ticker	Name	Ticker	Name	Ticker	Name
AA	Alcoa	EMR	Emerson Elec	NOV	National Oil.
AAPL	Apple	ETR	Entergy	NSC	Norfolk Sou.
ABT	Abbott Lab.	EXC	Exelon	NWSA	News Corp
AEP	American Elec	F	Ford	ORCL	Oracle
ALL	Allstate Corp	FCX	Freeport	OXY	Occidental Pet.
AMGN	Amgen Inc.	FDX	Fedex	PEP	Pepsi
AMZN	Amazon.com	GD	General Dyna	PFE	Pfizer
AVP	Avon	GE	General Elec	PG	P&G
APA	Apache	GILD	Gilead Science	QCOM	Qualcomm Inc
AXP	American Ex	GOOG	Google Inc	RF	Regions Fin
BA	Boeing	GS	Gold. Sachs	RTN	Raytheon
BAC	Bank of Am	HAL	Halliburton	S	Sprint
BAX	Baxter	HD	Home Depot	SBUX	Starbucks
BHI	Baker Hughes	HNZ	Heinz	SLB	Schlumberger
BK	Bank of NY	HON	Honeywell	SLE	Sara Lee Corp.
BMY	Bristol-Myers	HPQ	HP	SO	Southern Co.
BRKB	Berkshire Hath	IBM	IBM	SPG	Simon pro.
C	Citi Group	INTC	Intel	T	AT&T
CAT	Caterpillar	JNJ	JohnsonJ.	TGT	Target
CL	Colgate	JPM	JP Morgan	TWX	Time Warner
CMCSA	Comcast	KFT	Kraft	TXN	Texas Inst
COF	Capital One	KO	Coca Cola	UNH	UnitedHealth
COP	Conocophillips	LLY	Lilly Eli	UNP	Union Pacific
COST	Costco	LMT	Lock'dMartn	UPS	United Parcel
CPB	Campbell	LOW	Lowe's	USB	US Bancorp
CSCO	Cisco	MCD	MaDonald	UTX	United Tech
CVS	CVS	MDT	Medtronic	VZ	Verizon
CVX	Chevron	MET	Metlife Inc.	WAG	Walgreen
DD	DuPont	MMM	3M	WFC	Wells Fargo
DELL	Dell	MO	Altria Group	WMB	Williams Co
DIS	Walt Disney	MON	Monsanto	WMT	WalMart
DOW	Dow Chem	MRK	Merck	WY	Weyerhauser
DVN	Devon Energy	MS	Morgan Stan.	XOM	Exxon
EBAY	Ebay	MSFT	Microsoft	XRX	Xerox
EMC	EMC	NKE	Nike		

Note: This table presents the ticker symbols and names of the 104 stocks used in Section 1.5.

Table 1.6: Summary statistics and conditional mean estimates

Panel A: Summary statistics

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
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

Panel B: Test for skewness, kurtosis, and correlation

	# of rejections
$H_0 : E\left[r_i^3\right] = 0$	3 out of 104
$H_0 : \frac{E\left[r_i^4\right]}{E\left[r_i^2\right]^2} = 3$	104 out of 104
$H_0 : Corr\left(r_i, r_j\right) = 0$	5356 out of 5356

Panel C: Conditional mean

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
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

Note: Panel A presents summary statistics such as simple unconditional moments, correlations and rank correlations of the daily equity returns used in the empirical analysis. Panel B shows the number of rejections for the test of zero skewness and zero excess kurtosis of 104 stocks under 5% level. Also the number of rejections for the test of zero correlations of all 5356 pairs is in the last line of Panel B. Panel C presents the parameter estimates for AR(1) models of the conditional means of returns. The columns present the mean and quantiles from the cross-sectional distribution of the measures or estimates listed in the rows.

Table 1.7: Conditional variance-covariance estimates

Panel A: HAR-type models for 5-min returns

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
Variance part						
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 part		
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 for daily returns

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
Variance part						
Constant $\psi_i \times 10000$	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 part		
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 1.2.2 using high frequency 5-min returns. Panel B presents summaries of the estimated DCC models described in Appendix A.2 using low frequency daily returns. The estimates for variance parts 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 1.8: Summary statistics of standardized uncorrelated residuals

Panel A: Summary statistics of residuals $\hat{\mathbf{e}}_t$

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
Residuals $\hat{\mathbf{e}}_{t,HAR}$						
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
Residuals $\hat{\mathbf{e}}_{t,DCC}$						
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

Panel B: Test for skewness, kurtosis, and correlation

	# of rejections	
	For $\hat{\mathbf{e}}_{t,HAR}$	For $\hat{\mathbf{e}}_{t,DCC}$
$H_0 : E\left[e_i^3\right] = 0$	4 out of 104	6 out of 104
$H_0 : \frac{E\left[e_i^4\right]}{E\left[e_i^2\right]^2} = 3$	104 out of 104	104 out of 104
$H_0 : Corr\left(e_i, e_j\right) = 0$	497 out of 5356	1 out of 5356

Note: Panel A presents summary statistics of the estimated standardized uncorrelated residuals $\hat{\mathbf{e}}_{t,HAR}$ and $\hat{\mathbf{e}}_{t,DCC}$, and Panel B shows the number of rejections for the test of zero skewness and zero excess kurtosis of $\hat{\mathbf{e}}_{t,HAR}$ and $\hat{\mathbf{e}}_{t,DCC}$ under 5% level. Also the number of rejections for the test of zero correlations of all 5356 pairs is in the last line of Panel B. In Panel C, the parameter estimates for standardized Student's t marginal distributions are summarized in the mean and quantiles from the cross-sectional distributions of the estimates.

Table 1.9: Marginal distribution and copula estimates

	Panel A: Marginal distributions for residuals $\hat{\mathbf{e}}_t$					
			Cross-sectional distribution			
	Mean	5%	25%	Median	75%	95%
Residuals $\hat{\mathbf{e}}_{t,HAR}$						
Student t (ν)	5.3033	4.1233	4.7454	5.1215	5.8684	6.8778
Residuals $\hat{\mathbf{e}}_{t,DCC}$						
Student t (ν)	6.0365	4.2280	5.0314	5.9042	7.0274	8.2823

	Panel B: Copula for residuals $\hat{\mathbf{e}}_t$					
	Jointly symmetric copula based on					
	t	Clayton	Frank	Gumbel	Indep	MV t dist
Residuals $\hat{e}_{t,HAR}$						
Copula est	39.4435	0.0876	1.2652	1.0198	-	6.4326†
(Std error)	(4.3541)	(0.0087)	(0.0942)	(0.0038)		(0.1405)
t-stat	**8.45***	**10.07***	**13.43***	**5.25***	-	**45.72***
Residuals $\hat{e}_{t,DCC}$						
Copula est	28.2068	0.1139	1.5996	1.0312	-	7.0962†
(Std error)	(5.4963)	(0.0155)	(0.1540)	(0.0071)		(0.3586)
t-stat	**6.13***	**7.36***	**10.36***	**4.40***	-	**17.80***

Note: Panel A presents the estimates of the marginal distribution of residuals, (standardized) univariate t distribution, summarized in the mean and quantiles from the cross-sectional distributions of the estimates. Panel B 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 bootstrap standard errors that reflect accumulated estimation errors from former stages are reported in parenthesis. We test the null hypothesis that there is *no* nonlinear dependence and report t-statistics denoted with * if significant at the 1% level. †Note that the parameter of the multivariate t distribution is not a copula parameter, but it is reported in this row for simplicity.

Table 1.10: Log (composite) likelihood for each stage

	Jointly symmetric copula based on					
	t	Clayton	Frank	Gumbel	Indep	MV t dist
5-min data						
$\log\lvert\det(\hat{\mathbf{H}}_{t,HAR}^{-1/2})\rvert$	-21724	-21724	-21724	-21724	-21724	-21724
log L of marginals for $\hat{\mathbf{e}}_{t,HAR}$	-260854	-260854	-260854	-260854	-260854	
log CL of copulas for $\hat{\mathbf{e}}_{t,HAR}$	86.86	78.07	65.60	44.67	N/A	-263129†
Sum	**-282491**	**-282500**	**-282512**	**-282533**	**-282578**	**-284853**
Ranking	1	2	3	4	5	6
Daily data						
$\log\lvert\det(\hat{\mathbf{H}}_{t,DCC}^{-1/2})\rvert$	-11451	-11451	-11451	-11451	-11451	-11451
log L of marginals for $\hat{\mathbf{e}}_{t,DCC}$	-277954	-277954	-277954	-277954	-277954	
log CL of copulas for $\hat{\mathbf{e}}_{t,DCC}$	242.73	214.68	187.69	149.23	N/A	-280157†
Sum	**-289162**	**-289190**	**-289217**	**-289255**	**-289404**	**-291607**
Ranking	7	8	9	10	11	12

Note: This table presents log absolute values of determinant of inverse square root of estimated conditional covariance matrix, sum of log likelihoods of 104 marginal distributions, and log composite likelihoods of jointly symmetric copulas at the estimated parameters. By summing those three values, the entire log likelihood for daily return distributions is obtained . Ranking is offered according to those sums. †Note that the multivariate t distribution dose not use copula approach, so the numbers with † is log composite likelihood of (standardized) multivariate t distributions at parameter estimates, but it is reported in this row for simplicity.

Table 1.11: t-statistics from in-sample model comparison tests

	5-min data						Daily data				
	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep	MV t	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep
5-min data											
t^{JS}	-										
C^{JS}	3.06	-									
F^{JS}	2.13	1.15	-								
G^{JS}	5.50	6.12	1.84	-							
Indep	5.32	5.28	4.62	3.70	-						
MV t	21.02	20.72	20.63	20.26	18.94	-					
Daily data											
t^{JS}	4.29	4.29	4.28	4.26	4.23	2.78	-				
C^{JS}	4.31	4.30	4.29	4.28	4.25	2.80	4.56	-			
F^{JS}	4.30	4.30	4.29	4.27	4.24	2.80	2.71	1.27	-		
G^{JS}	4.35	4.34	4.33	4.32	4.29	2.84	6.79	7.50	1.75	-	
Indep	4.44	4.43	4.42	4.41	4.38	2.93	6.21	6.24	5.56	5.09	-
MV t	5.82	5.82	5.81	5.79	5.76	4.35	20.67	19.99	19.91	19.22	16.13

Note: This table presents t-statistics from Rivers and Vuong (2002) (pair-wise) model comparison tests introduced in Section 1.3.2. 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, a special case of jointly symmetric copulas. MV t is the multivariate Student's t distribution. The upper panel includes results for pair-wise comparison tests for competing models that use 5-min data for forecasts $\mathbf{H}_t$. The first half of the bottom panel is for comparisons of models that use 5-min data with models that use daily data for forecasts $\mathbf{H}_t$. The second half of the bottom panel includes results for pair-wise comparison tests for competing models that use daily data for forecasts $\mathbf{H}_t$.

Table 1.12: t-statistics from out-of-sample model comparison tests

	5-min data						Daily data				
	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep	MV t	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep
5-min data											
t^{JS}	-										
C^{JS}	1.60	-									
F^{JS}	0.83	0.35	-								
G^{JS}	2.94	3.13	1.33	-							
Indep	2.63	2.65	2.46	1.88	-						
MV t	10.68	10.56	10.58	10.41	9.92	-					
Daily data											
t^{JS}	5.21	5.21	5.21	5.20	5.19	4.52	-				
C^{JS}	5.21	5.21	5.21	5.20	5.19	4.53	1.63	-			
F^{JS}	5.20	5.20	5.20	5.19	5.18	4.52	1.77	1.31	-		
G^{JS}	5.21	5.21	5.22	5.21	5.20	4.53	3.02	3.34	0.05	-	
Indep	5.22	5.21	5.22	5.21	5.20	4.54	3.16	3.16	2.47	2.47	-
MV t	6.03	6.03	6.03	6.02	6.01	5.39	14.68	14.37	14.59	13.92	12.83

Note: This table presents t-statistics from pair-wise comparisons of the out-of-sample likelihoods of competing density forecasts. 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, a special case of jointly symmetric copulas. MV t is the multivariate Student's t distribution. The upper panel includes results for pair-wise comparison tests for competing models that use 5-min data for forecasts $\mathbf{H}_t$. The first half of the bottom panel is for comparisons of models that use 5-min data with models that use daily data for forecasts $\mathbf{H}_t$. The second half of the bottom panel includes results for pair-wise comparison tests for competing models that use daily data for forecasts $\mathbf{H}_t$.

Table 1.13: Out-of-sample model comparisons using realized portfolio returns

Risk aversion (ρ) of 7

	5-min data						**Daily data**				
	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep	MV t	t^{JS}	C^{JS}	F^{JS}	G^{JS}	Indep
5-min data											
t^{JS}	-										
C^{JS}	3.35	-									
F^{JS}	1.93	-1.42	-								
G^{JS}	1.94	-1.41	0.01	-							
Indep	1.32	-2.03	-0.61	-0.62	-						
MV t	2.42	-0.93	0.48	0.48	1.10	-					
Daily data											
t^{JS}	1.77	-1.58	-0.17	-0.17	0.45	-0.65	-				
C^{JS}	5.50	2.15	3.57	3.56	4.18	3.08	3.73	-			
F^{JS}	6.36	3.01	4.43	4.42	5.04	3.94	4.59	0.86	-		
G^{JS}	6.32	2.98	4.39	4.39	5.01	3.91	4.56	0.83	-0.03	-	
Indep	3.30	-0.05	1.36	1.36	1.98	0.88	1.53	-2.20	-3.06	-3.03	-
MV t	6.30	2.95	4.36	4.36	4.98	3.88	4.53	0.80	-0.06	-0.03	3.00

Note: This table presents the "management fee," in annualized percent returns, that an investor with risk aversion of 7 would be willing to pay to switch from the model to the left to the model above. A positive management fee indicates that the model above generates more OOS economic gains than the model to the left dose, and a negative one indicates the opposite.

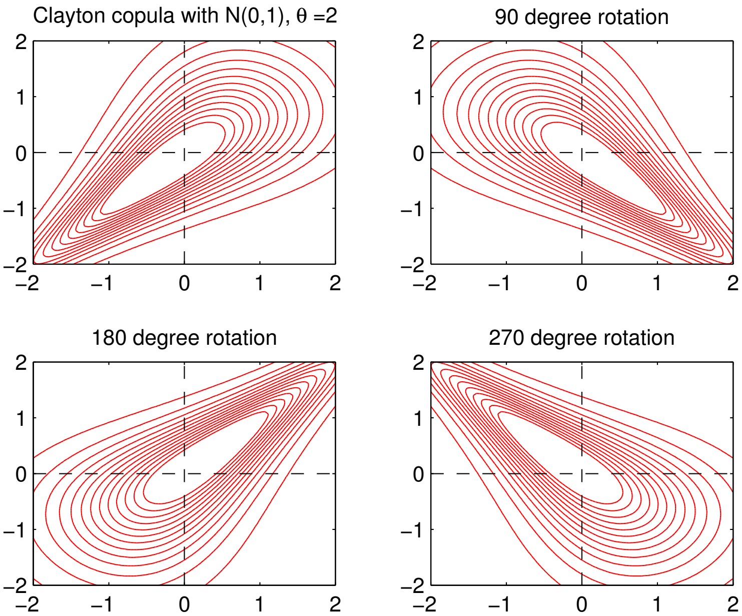


FIGURE 1.1: *90, 180, and 270 degree rotations of the density of Clayton copula ($\theta = 2$) with $N(0, 1)$ margins*

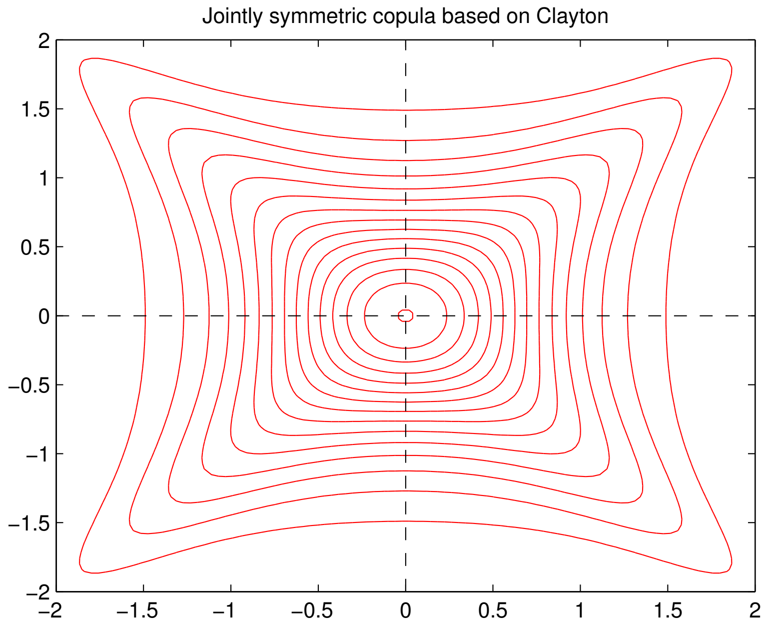


FIGURE 1.2: *Jointly symmetric copula density constructed from Clayton* $(\theta = 2)$ *with* $N(0,1)$ *margins*

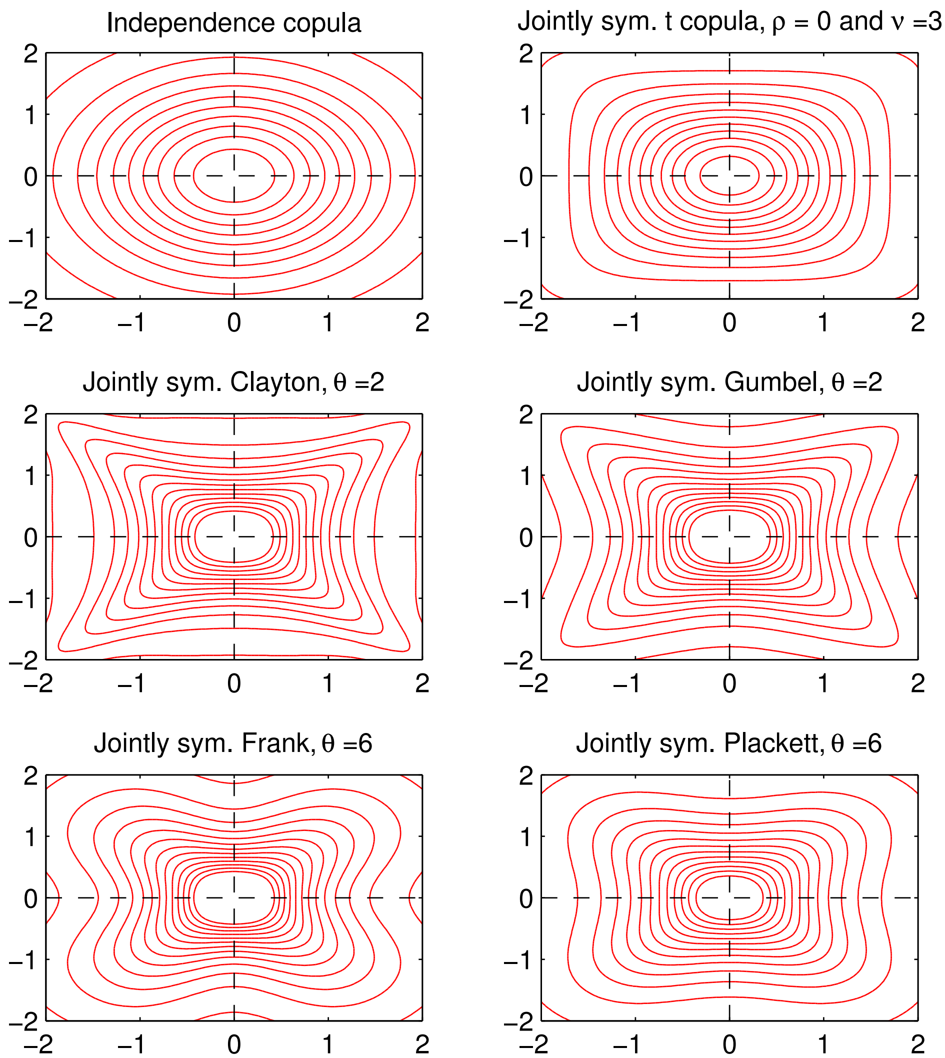


FIGURE 1.3: *Contour plots of densities for independence copula and jointly symmetric copulas based on t, Clayton, Gumbel, Frank, and Plackett copula with* $N(0,1)$ *margins*

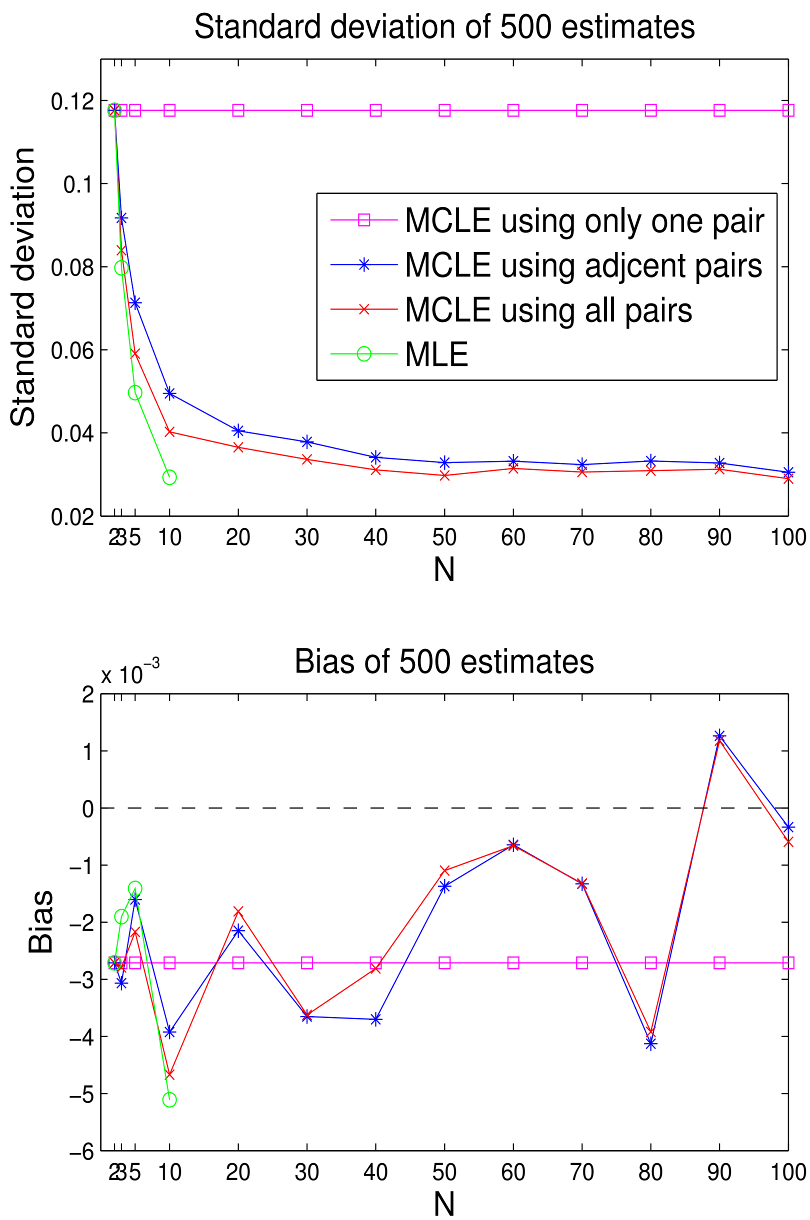


FIGURE 1.4: *Standard deviation and bias of 500 estimators for jointly symmetric copula based on Clayton copula. $T = 1000$ and $N = 2, 3,5,10,\ldots, 100$. Four different esimation methods, MLE, MCLE with all pairs, MCLE with adjacent pairs, and MCLE with the first pair are used*

2

Modelling Dependence in High Dimensions with Factor Copulas (co-authored with Andrew Patton)

2.1 Introduction

One of the many surprises from the financial crisis of late 2007 to 2008 was the extent to which assets that had previously behaved mostly independently suddenly moved together. This was particularly prominent in the financial sector, where poor models of the dependence between certain asset returns (such as those on housing, or those related to mortgage defaults) are thought to be one of the causes of the collapse of the market for CDOs and related securities, see Coval, *et al.* (2009) and Zimmer (2012) for example. Many models that were being used to capture the dependence between a large number of financial assets were revealed as being inadequate during the crisis. However, one of the difficulties in analyzing risks across many variables is the relative paucity of econometric models suitable for the task. Correlation-based models, while useful when risk can be summarized using the second moment, are often built on an assumption of multivariate Gaussianity, and face the risk of neglecting dependence between the variables in the tails, i.e., neglecting the possibility that large crashes may be correlated across assets.

This paper makes two primary contributions. First, we present new models for the dependence structure, or copula, of economic variables. The models are based on a simple factor structure for the copula and are particularly attractive for high dimensional applications, involving fifty or more variables.[1] These copula models may be combined with existing models for univariate distributions to construct flexible, tractable joint distributions for large collections of variables. The proposed copula models permit the researcher to determine the degree of flexibility based on the number of variables and the amount of data available. For example, by allowing for a fat-tailed common factor the model captures the possibility of correlated crashes, and by allowing the common factor to be asymmetrically distributed the model allows for the possibility that the dependence between the variables is stronger during downturns than during upturns. By allowing for multiple common factors, it is possible to capture heterogeneous pair-wise dependence within the overall multivariate copula. High dimension economic applications will often require some strong simplifying assumptions in order to keep the model tractable, and an important feature of the class of proposed models is that such assumptions can be made in an easily understandable manner, and can be tested and relaxed if needed.

Factor copulas do not generally have a closed-form density, but certain properties can nevertheless be obtained analytically. Using extreme value theory we obtain theoretical results on the tail dependence properties for general, multi-factor copulas, and for the specific parametric class of factor copulas that we use in our empirical work.

The second contribution of this paper is a study of the dependence structure of all 100 constituent firms of the Standard and Poor's 100 index, using daily data over the

[1] For related recent work on high dimensional conditional covariance matrix estimation, see Engle and Kelly (2012), Engle, *et al.* (2008), and Hautsch, *et al.* (2010).

period 2008-2010. This is one of the highest dimension applications of copula theory in the econometrics literature. We find significant evidence in favor of a fat-tailed common factor for these stocks (indicative of non-zero tail dependence), implying that the Normal, or Gaussian, copula is not suitable for these assets. Moreover, we find significant evidence that the common factor is asymmetrically distributed, with crashes being more highly correlated than booms. Our empirical results suggest that risk management decisions made using the Normal copula may be based on too benign a view of these assets, and derivative securities based on baskets of these assets, or related securities such as CDOs, may be mispriced if based on a Normal copula. The fact that large negative shocks may originate from a fat-tailed common factor, and thus affect all stocks at once, makes the diversification benefits of investing in these stocks lower than under Normality.

An additional contribution of this paper is a detailed simulation study of the properties of the estimation method for the class of factor copulas we propose. This class does not generally have a closed-form copula likelihood, and we use the SMM estimator proposed in Oh and Patton (2013a). We consider problems of dimension 3, 10 and 100, and confirm that the estimator and associated asymptotic distribution theory have satisfactory finite-sample properties.

Certain types of factor copulas have already appeared in the literature. The models we consider are extensions of Hull and White (2004), in that we retain a simple linear, additive factor structure, but allow for the variables in the structure to have flexibly specified distributions. Other variations on factor copulas are presented in Andersen and Sidenius (2004) and van der Voort (2005), who consider certain non-linear factor structures, and in McNeil *et al.* (2005), who present factor copulas for modelling times-to-default. With the exception of McNeil, *et al.* (2005), the papers to date have not considered estimation of the unknown parameters of these copulas, instead examining calibration and pricing using these copulas. Our formal analysis of the estimation of high dimension copulas via a SMM-type procedure is new to the literature, as is our application of this class of models to a large collection of asset returns.

Some methods for modelling high dimension copulas have previously been proposed in the literature, though few consider dimensions greater than twenty.[2] The Normal copula, see Li (2000) amongst many others, is simple to implement and to understand, but imposes the strong assumption of zero tail dependence, and symmetric dependence between booms and crashes. The (Student's) t copula, and variants of it, are discussed in Demarta and McNeil (2005). An attractive extension of the t copula, the "grouped t" copula, is proposed in Daul *et al.* (2003), who show that this copula can be used in applications of up to 100 variables. This copula allows for heterogeneous tail dependence between pairs of variables, but imposes that upper and lower tail dependence are equal (a finding we strongly reject for equity returns). Smith, *et al.* (2010) extract the copula implied by a multivariate skew t distribution, and Christoffersen, *et al.* (2012) combine a skew t copula with a DCC model for conditional correlations in their study of 33 developed and emerging equity market indices. Archimedean copulas such as the Clayton or Gumbel allow for tail dependence and particular forms of asymmetry, but usually have only a one or two parameters to characterize the dependence between all variables, and are thus quite restrictive when the number of variables is large. Multivariate "vine" copulas are constructed by sequentially applying bivariate copulas to build up a higher dimension copula, see Aas, *et al.* (2009), Heinen and Valdesogo (2009) and Min and Czado (2010) for example, however vine copulas are almost invariably based on an assumption that is hard to interpret and to test, see Acar, *et al.* (2012) for a critique. In our empirical application we compare our proposed factor models with several alternative existing

[2] For general reviews of copulas in economics and finance see Cherubini, *et al.* (2004) and Patton (2012).

models, and show that our model outperforms them all in terms of goodness-of-fit and in an application to measuring systemic risk.

The remainder of the paper is structured as follows. Section 2.2 presents the class of factor copulas, derives their limiting tail properties, and considers some extensions. Section 2.3 considers estimation via a simulation-based method and presents a simulation study of this method. Section 2.4 presents an empirical study of daily returns on individual constituents of the S&P 100 equity index over the period 2008-2010. Appendix B.1 contains all proofs, and Appendix B.2 contains a discussion of the dependence measures used in estimation.

2.2 Factor copulas

For simplicity we will focus on unconditional distributions in the text below, and discuss the extension to conditional distributions in the next section. Consider a vector of N variables, $\mathbf{Y}$, with some joint distribution $\mathbf{F}$, marginal distributions F_i, and copula $\mathbf{C}$:

$$[Y_1, ..., Y_N]' \equiv \mathbf{Y} \sim \mathbf{F} = \mathbf{C}\left(F_1, ..., F_N\right) \tag{2.1}$$

The copula completely describes the dependence between the variables $Y_1, ..., Y_N$. We will use existing models to estimate the marginal distributions F_i, and focus on constructing useful new models for the dependence between these variables, $\mathbf{C}$. Decomposing the joint distribution in this way has two important advantages over considering the joint distribution $\mathbf{F}$ directly: First, it facilitates multi-stage estimation, which is particularly useful in high dimension applications, where the sparseness of the data and the potential proliferation of parameters can cause problems. Second, it allows the researcher to draw on the large literature on models for univariate distributions, leaving "only" the task of constructing a model for the copula, which is a simpler problem.

2.2.1 Description of a simple factor copula model

The class of copulas we consider are those that can be generated by the following simple factor structure, based on a set of $N+1$ latent variables:

$$
\begin{aligned}
X_i &= Z + \varepsilon_i, \;\; i = 1, 2, ..., N \\
Z &\sim F_z(\theta), \;\; \varepsilon_i \sim iid\ F_\varepsilon(\theta), \;\; Z \perp\!\!\!\perp \varepsilon_i \; \forall\ i \\
[X_1, ..., X_N]' \equiv \mathbf{X} &\sim \mathbf{F}_x = \mathbf{C}(G_1(\theta), ..., G_N(\theta); \theta)
\end{aligned} \tag{2.2}
$$

The copula of the latent variables $\mathbf{X}$, $\mathbf{C}(\theta)$, is used as the model for the copula of the observable variables $\mathbf{Y}$.[3] An important point about the above construction is that the marginal distributions of X_i may be different from those of the original variables Y_i, so $F_i \neq G_i$ in general. We use the structure for the vector $\mathbf{X}$ *only* for its copula, and completely discard the resulting marginal distributions. By doing so, we use $\mathbf{C}(\theta)$ from equation (2.2) to construct a model for the *copula* of $\mathbf{Y}$, and leave the marginal distributions F_i to be specified and estimated in a separate step.

The copula implied by the above structure is not generally known in closed form. The leading case where it *is* known is when F_z and F_ε are both Gaussian distributions, in which case the variable $\mathbf{X}$ is multivariate Gaussian, implying a Gaussian copula, and with an equicorrelation dependence structure (with correlation between any pair of variables equal to $\sigma_z^2/(\sigma_z^2 + \sigma_\varepsilon^2)$). For other choices of F_z and F_ε the joint distribution of $\mathbf{X}$, and more importantly the copula of $\mathbf{X}$, is not known in closed form. It is clear from the structure above that the copula will exhibit "equidependence", in that each pair of variables will have the same bivariate copula as any other pair. (This property is known as "exchangeability" in the copula literature.) A similar assumption for correlations is made in Engle and Kelly (2012).

---

[3] This method for constructing a copula model resembles the use of mixture models, e.g. the Normal-inverse Gaussian or generalized hyperbolic distributions, where the distribution of interest is obtained by considering a function of a collection of latent variables, see Barndorff-Nielsen (1978, 1997), Barndorff-Nielsen and Shephard (2009), McNeil, *et al.* (2005).

It is simple to simulate from F_z and F_ε for many classes of distributions, and from simulated data we can extract properties of the copula, such as rank correlation, Kendall's tau, and quantile dependence. These simulated moments can be used in simulated method of moments (SMM) estimation of the unknown parameters, which is studied via simulations in Section 2.3 below.

2.2.2 A multi-factor copula model

The structure of the model in equation (2.2) immediately suggests two directions for extensions. The first is to allow for weights on the common factor that differ across variables. That is, let

$$X_i = \beta_i Z + \varepsilon_i, \;\; i = 1, 2, ..., N \tag{2.3}$$
$$Z \sim F_z, \;\; \varepsilon_i \sim iid\ F_\varepsilon, \;\; Z \perp\!\!\!\perp \varepsilon_i \;\forall\; i$$

with the rest of the model left unchanged. In this "single factor, flexible weights" factor copula, the implied copula is no longer equidependent: a given pair of variables may have weaker or stronger dependence than some other pair. This extension introduces $N - 1$ additional parameters to this model, increasing its flexibility to model heterogeneous pairs of variables, at the cost of a more difficult estimation problem. An intermediate model may be considered, in which sub-sets of variables are assumed to have the same weight on the common factor, which may be reasonable for financial applications with variables grouped ex ante using industry classifications, for example. Such an assumption leads to a "block equidependence" copula, and we will consider this structure in our empirical application.

A second extension to consider is a multi-factor version of the model, where the dependence is assumed to come from a K-factor model:

$$X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \varepsilon_i$$

$$\varepsilon_i \sim iid\ F_\varepsilon, \quad Z_k \perp\!\!\!\perp \varepsilon_i \ \forall\ i, k \tag{2.4}$$

$$[Z_1, ..., Z_K]' \equiv \mathbf{Z} \sim \mathbf{F}_z = \mathbf{C}_{indep}\left(F_{z_1}, ..., F_{z_K}\right)$$

In the most general case one could allow $\mathbf{Z}$ to have a general copula $\mathbf{C}_Z$ that allows dependence between the common factors, however an empirically useful simplification of this model is to impose that the common factors are independent, and thus remove the need to specify and estimate $\mathbf{C}_Z$. A further simplification of this factor model may be to assume that each common factor has a weight equal to one or zero, with the weights specified in advance by grouping variables, for example by grouping stocks by industry.

The above model can be interpreted as a special case of the "conditional independence structure" of McNeil, *et al.* (2005), which is used to describe a set of variables that are independent conditional on some smaller set of variables, $\mathbf{X}$ and $\mathbf{Z}$ in our notation.[4] McNeil, *et al.* (2005) describe using such a structure to generate some factor copulas to model times until default.

2.2.3 Tail dependence properties of factor copulas

Using results from extreme value theory, it is possible to obtain analytically results on the tail dependence implied by a factor copula model despite the fact that we do not have a closed-form expression for the copula. These results are relatively easy to obtain, given the simple linear structure generating the factor copula. Recall the definition of tail dependence for two variables X_i, X_j with marginal distributions

[4] The variables $\mathbf{Z}$ are sometimes known as the "frailty", in the survival analysis and credit default literature, see Duffie, *et al.* (2009) for example.

G_i,G_j :

$$\tau_{ij}^{L} \equiv \lim_{q\to 0} \frac{\Pr\left[X_i \leqslant G_i^{-1}(q), X_j \leqslant G_j^{-1}(q)\right]}{q} \tag{2.5}$$

$$\tau_{ij}^{U} \equiv \lim_{q\to 1} \frac{\Pr\left[X_i > G_i^{-1}(q), X_j > G_j^{-1}(q)\right]}{1-q}$$

That is, lower tail dependence measures the probability of both variables lying below their q quantile, for q limiting to zero, scaled by the probability of one of these variables lying below their q quantile. Upper tail dependence is defined analogously. In Proposition 1 below we present results for a general single factor copula model:

Proposition 1 (Tail dependence for a factor copula)**.** *Consider the factor copula generated by equation (2.3). If F_z and F_ε have regularly varying tails with a common tail index $\alpha > 0$, i.e.*

$$\Pr[Z > s] = A_z^U s^{-\alpha} \quad \text{and} \quad \Pr[\varepsilon_i > s] = A_\varepsilon^U s^{-\alpha}, \quad \text{as } s \to \infty \tag{2.6}$$

$$\Pr[Z < -s] = A_z^L s^{-\alpha} \quad \text{and} \quad \Pr[\varepsilon_i < -s] = A_\varepsilon^L s^{-\alpha} \quad \text{as } s \to \infty$$

where A_Z^L, A_Z^U, A_ε^L and A_ε^U are positive constants. Then, (a) if $\beta_j \geqslant \beta_i > 0$ the lower and upper tail dependence coefficients are:

$$\tau_{ij}^{L} = \frac{\beta_i^\alpha A_z^L}{\beta_i^\alpha A_z^L + A_\varepsilon^L}, \quad \tau_{ij}^{U} = \frac{\beta_i^\alpha A_z^U}{\beta_i^\alpha A_z^U + A_\varepsilon^U} \tag{2.7}$$

(b) if $\beta_j \leqslant \beta_i < 0$ the lower and upper tail dependence coefficients are:

$$\tau_{ij}^{L} = \frac{|\beta_i|^\alpha A_z^U}{|\beta_i|^\alpha A_z^U + A_\varepsilon^L}, \quad \tau_{ij}^{U} = \frac{|\beta_i|^\alpha A_z^L}{|\beta_i|^\alpha A_z^L + A_\varepsilon^U} \tag{2.8}$$

(c) if $\beta_i\beta_j = 0$ or (d) if $\beta_i\beta_j < 0$, the lower and upper tail dependence coefficients are zero.

All proofs are presented in Appendix B.1. This proposition shows that when the coefficients on the common factor have the same sign, and the common factor and idiosyncratic variables have the same tail index, the factor copula generates upper and lower tail dependence. If either Z or ε is asymmetrically distributed, then the upper and lower tail dependence coefficients can differ, which provides this model with the ability to capture differences in the probabilities of joint crashes and joint booms. When either of the coefficients on the common factor are zero, or if they have differing signs, then it is simple to show that the upper and lower tail dependence coefficients are both zero.

The above proposition considers the case that the common factor and idiosyncratic variables have the same tail index; when these indices differ we obtain a boundary result: if the tail index of Z is strictly greater than that of ε and $\beta_i\beta_j > 0$ then tail dependence is one, while if the tail index of Z is strictly less than that of ε then tail dependence is zero.

In our simulation study and empirical work below, we will focus on the skew t distribution of Hansen (1994) as a model for the common factor and the standardized t distribution for the idiosyncratic shocks. Proposition 2 below presents the analytical tail dependence coefficients for a factor copula based on these distributions.

Proposition 2 (Tail dependence for a skew t-t factor copula)**.** *Consider the factor copula generated by equation (2.3). If $F_z = Skew\, t\,(\nu, \lambda)$ and $F_\varepsilon = t\,(\nu)$, then the tail indices of Z and ε_i equal ν, and the constants A_z^L, A_z^U, A_ε^L and A_ε^U from Proposition 1 equal:*

$$A_z^L = \frac{bc}{\nu}\left(\frac{b^2}{(\nu-2)(1-\lambda)^2}\right)^{-(\nu+1)/2}, \quad A_z^U = \frac{bc}{\nu}\left(\frac{b^2}{(\nu-2)(1+\lambda)^2}\right)^{-(\nu+1)/2} \tag{2.9}$$

$$A_\varepsilon^L = A_\varepsilon^U = \frac{c}{\nu}\left(\frac{1}{\nu-2}\right)^{-(\nu+1)/2}$$

where $a = 4\lambda c\,(\nu-2)\,/\,(\nu-1)$, $b = \sqrt{1+3\lambda^2-a^2}$, $c = \Gamma\left(\frac{\nu+1}{2}\right)/\left(\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\,(\nu-2)}\right)$.

Given Proposition 1 and the expressions for A_z^L, A_z^U, A_ε^L and A_ε^U above, we then ob-

tain the tail dependence coefficients for this copula.

In the next proposition we generalize Proposition 1 to allow for a multi-factor model, which will prove useful in our empirical application in Section 2.4.

Proposition 3 (Tail dependence for a multi-factor copula)**.** *Consider the factor copula generated by equation (2.4). Assume F_ε, $F_{z_1}, ..., F_{z_K}$ have regularly varying tails with a common tail index $\alpha > 0$, and upper and lower tail coefficients $A^U_\varepsilon, A^U_1, .., A^U_K$ and $A^L_\varepsilon, A^L_1, .., A^L_K$. Then if $\beta_{ik} \geqslant 0 \ \forall \ i, k$, the lower and upper tail dependence coefficients are:*

$$\tau^L_{ij} = \frac{\sum_{k=1}^K \mathbf{1}\{\beta_{ik}\beta_{jk} > 0\} A^L_k \beta^\alpha_{ik} \delta^\alpha_{L,ijk}}{A^L_\varepsilon + \sum_{k=1}^K A^L_k \beta^\alpha_{ik}} \tag{2.10}$$

$$\tau^U_{ij} = \frac{\sum_{k=1}^K \mathbf{1}\{\beta_{ik}\beta_{jk} > 0\} A^U_k \beta^\alpha_{ik} \delta^\alpha_{U,ijk}}{A^U_\varepsilon + \sum_{k=1}^K A^U_k \beta^\alpha_{ik}}$$

where

$$\delta^{-1}_{L,ijk} \equiv \begin{cases} \max\{1, \gamma_{L,ij}\beta_{ik}/\beta_{jk}\}, & \text{if } \beta_{ik}\beta_{jk} > 0 \\ 1, & \text{if } \beta_{ik}\beta_{jk} = 0 \end{cases} \tag{2.11}$$

$$\delta^{-1}_{U,ijk} \equiv \begin{cases} \max\{1, \gamma_{U,ij}\beta_{ik}/\beta_{jk}\}, & \text{if } \beta_{ik}\beta_{jk} > 0 \\ 1, & \text{if } \beta_{ik}\beta_{jk} = 0 \end{cases}$$

$$\gamma_{L,ij} \equiv \left(\frac{A^L_\varepsilon + \sum_{k=1}^K A^L_k \beta^\alpha_{jk}}{A^L_\varepsilon + \sum_{k=1}^K A^L_k \beta^\alpha_{ik}} \right)^{1/\alpha}, \quad \gamma_{U,ij} \equiv \left(\frac{A^U_\varepsilon + \sum_{k=1}^K A^U_k \beta^\alpha_{jk}}{A^U_\varepsilon + \sum_{k=1}^K A^U_k \beta^\alpha_{ik}} \right)^{1/\alpha} \tag{2.12}$$

The extensions to consider the case that some have opposite signs to the others can be accommodated using the same methods as in the proof of Proposition 1. In the one-factor copula model the variables $\delta_{L,ijk}$ and $\delta_{U,ijk}$ can be obtained directly and are determined by $\min\{\beta_i, \beta_j\}$; in the multi-factor copula model these variables can be determined using equation (2.11) above, but do not generally have a simple expression.

2.2.4 Illustration of some factor copulas

To illustrate the flexibility of this simple class of copulas, Figure 2.1 presents 1000 random draws from bivariate distributions constructed using four different factor copulas. In all cases the marginal distributions, F_i, are set to $N(0,1)$, and the variance of the latent variables in the factor copula are set to $\sigma_z^2 = \sigma_\varepsilon^2 = 1$, so that the common factor (Z) accounts for one-half of the variance of each X_i. The first copula is generated from a factor structure with $F_z = F_\varepsilon = N(0,1)$, implying that the copula is Normal. The second sets $F_z = F_\varepsilon = t(4)$, generating a symmetric copula with positive tail dependence. The third copula sets $F_\varepsilon = N(0,1)$ and $F_z = \mathit{skew}\ t(\infty, -0.25)$, corresponding to a skewed Normal distribution. This copula exhibits asymmetric dependence, with crashes being more correlated than booms, but zero tail dependence. The fourth copula sets $F_\varepsilon = t(4)$ and $F_z = \mathit{skew}\ t(4, -0.25)$, which generates asymmetric dependence and positive tail dependence.

Figure 2.1 shows that when the distributions in the factor structure are Normal or skewed Normal, tail events tend to be uncorrelated across the two variables. When the degrees of freedom is set to 4, on the other hand, we observe several draws in the joint upper and lower tails. When the skewness parameter is negative, as in the lower two panels of Figure 2.1, we observe stronger clustering of observations in the joint negative quadrant compared with the joint positive quadrant.

An alternative way to illustrate the differences in the dependence implied by these four models is to use a measure known as "quantile dependence". This measure captures the probability of observing a draw in the q-tail of one variable given that such an observation has been observed for the other variable. It is defined as:

$$\tau_q \equiv \begin{cases} \frac{1}{q}\Pr\left[U_1 \leqslant q, U_2 \leqslant q\right], & q \in (0, 0.5] \\ \frac{1}{1-q}\Pr\left[U_1 > q, U_2 > q\right], & q \in (0.5, 1) \end{cases} \tag{2.13}$$

where $U_i \equiv G_i(X_i) \sim \mathit{Unif}(0,1)$ are the probability integral transforms of the simulated X_i variables. As $q \to 0$ ($q \to 1$) this measure converges to lower (upper) tail dependence, and for values of q "near" zero or one we obtain an estimate of the dependence "near" the joint tails.

Figure 2.2 presents the quantile dependence functions for these four copulas. For the symmetric copulas (Normal, and t-t factor copula) this function is symmetric about $q = 0.5$, while for the others it is not. The two copulas with a fat-tailed common factor exhibit quantile dependence that increases near the tails: in those cases an extreme observation is more likely to have come from the fat-tailed common factor (Z) than from the thin-tailed idiosyncratic variable (ε_i), and thus an extreme value for one variable makes an extreme value for the other variable more likely. Figure 2.2 also presents the theoretical tail dependence for each of these copulas based on Proposition 2 above using a symbol at $q = 0$ (lower tail dependence) and $q = 1$ (upper tail dependence). The *skew* $t\,(4)$-$t\,(4)$ factor copula illustrates the flexibility of this simple class of models, generating weak upper quantile dependence but strong lower quantile dependence, a feature that may be useful when modelling asset returns.

Figure 2.3 illustrates the differences between these copulas using a truly multivariate approach: Conditional on observing k out of 100 stocks crashing, we present the expected number of the remaining $(100 - k)$ stocks that will crash, a measure based on Geluk, *et al.* (2007):

$$\pi^q\,(j) \equiv \frac{\kappa^q\,(j)}{N - j} \tag{2.14}$$

$$\text{where} \quad \kappa^q\,(j) = E\left[N_q^* | N_q^* \geqslant j\right] - j$$

$$N_q^* \equiv \sum\nolimits_{i=1}^{N} \mathbf{1}\left\{U_i \leqslant q\right\}$$

For this illustration we define a "crash" as a realization in the lower 1/66 tail, corresponding to a once-in-a-quarter event for daily asset returns. The upper panel shows that as we condition on more variables crashing, the expected number of other variables that will crash, $\kappa^q(j)$, initially increases, and peaks at around $j = 30$. At that point, a *skew* $t(4)$-$t(4)$ factor copula predicts that around another 38 variables will crash, while under the Normal copula we expect only around 12 more variables to crash. As we condition on even more variables crashing the plot converges to inevitably zero, since conditioning on having observed more crashes, there are fewer variables left to crash. The lower panel of Figure 2.3 shows that the expected *proportion* of remaining stocks that will crash, $\pi^q(j)$ generally increases all the way to $j = 99$.[5] For comparison, this figure also plots the results for a positively skewed *skew* t factor copula, where booms are more correlated than crashes. This copula also exhibits tail dependence, and so the expected proportion of other stocks that will crash is higher than under Normality, but the positive skew means that crashes are less correlated than booms, and so the expected proportion is less than when the common factor is negatively skewed. This figure illustrates some of the features of dependence that are unique to high dimension applications, and further motivates our proposal for a class of flexible, parsimonious models for such applications.

2.2.5 Non-linear factor copula models

We can generalize the above linear, additive structure to consider more general factor structures. For example, consider the following general one-factor structure:

$$\begin{aligned} X_i &= h(Z, \varepsilon_i), \quad i = 1, 2, ..., N \\ Z &\sim F_z, \;\; \varepsilon_i \sim iid\; F_\varepsilon, \;\; Z \perp\!\!\!\perp \varepsilon_i \;\forall\, i \\ [X_1, ..., X_N]' &\equiv \mathbf{X} \sim \mathbf{F}_x = \mathbf{C}(G_1, ..., G_N) \end{aligned} \tag{2.15}$$

for some function $h : \mathbb{R}^2 \to \mathbb{R}$. Writing the factor model in this general form reveals that this structure nests a variety of well-known copulas in the literature. Some

[5] For the Normal copula this is not the case, however this is perhaps due to simulation error: even with the 10 million simulations used to obtain this figure, joint 1/66 tail crashes are so rare under a Normal copula that there is a fair degree of simulation error in this plot for $j \geqslant 80$.

examples of copula models that fit in this framework are summarized in the table below:

Copula	$h(Z,\varepsilon)$	F_Z	F_ε
Normal	$Z+\varepsilon$	$N(0,\sigma_z^2)$	$N(0,\sigma_\varepsilon^2)$
Student's t	$Z^{1/2}\varepsilon$	$Ig(\nu/2,\nu/2)$	$N(0,\sigma_\varepsilon^2)$
Skew t	$\lambda Z+Z^{1/2}\varepsilon$	$Ig(\nu/2,\nu/2)$	$N(0,\sigma_\varepsilon^2)$
Gen hyperbolic	$\gamma Z+Z^{1/2}\varepsilon$	$GIG(\lambda,\chi,\psi)$	$N(0,\sigma_\varepsilon^2)$
Clayton	$(1+\varepsilon/Z)^{-\alpha}$	$\Gamma(\alpha,1)$	$Exp(1)$
Gumbel	$-(\log Z/\varepsilon)^{\alpha}$	$Stable(1/\alpha,1,1,0)$	$Exp(1)$

where Ig represents the inverse gamma distribution, GIG is the generalized inverse Gaussian distribution, and Γ is the gamma distribution. The skew t and Generalized hyperbolic copulas listed here are from McNeil, *et al.* (2005, Chapter 5), the representation of a Clayton copula in this form is from Cook and Johnson (1981) and the representation of the Gumbel copula is from Marshall and Olkin (1988).

The above copulas all have closed-form densities via judicious combinations of the function h and the distributions F_Z and F_ε. Removing this requirement, and employing simulation-based estimation methods to overcome the lack of closed-form likelihood, one can obtain a much wider variety of models for the dependence structure. In this paper we will focus on linear, additive factor copulas, and generate flexible models by flexibly specifying the distribution of the common factor(s).

2.3 A Monte Carlo study of SMM estimation of factor copulas

As noted above, the class of factor copula models does not generally have a closed-form likelihood, motivating the study of alternative methods for estimation.[6] Oh

[6] In ongoing work, Oh and Patton (2013b), we consider the finite-sample properties of a ML estimator based on a quadrature approximation of the factor copula likelihood. This approach has its own set of numerical and implementation issues, and we leave the consideration of this method aside here.

and Patton (2013a) present a general simulation-based method for the estimation of copula models, which is ideally suited for the estimation of factor copulas, and is described in Section 2.3.2 below. In Section 2.3.3 we present an extensive Monte Carlo study of the finite-sample properties of their SMM estimator in applications involving up to 100 variables (Oh and Patton, 2013a, considers only up to 10 variables in their simulation study).

2.3.1 Description of the model for the conditional joint distribution

We consider the same class of data generating processes (DGPs) as Chen and Fan (2006), Rémillard (2010) and Oh and Patton (2013a). This class allows each variable to have time-varying conditional mean and conditional variance, each governed by parametric models, with some unknown marginal distribution. The marginal distributions are estimated nonparametrically via the empirical distribution function. The conditional copula of the data is assumed to belong to a parametric family, such as a parametric factor copula, and is assumed constant,[7] making the model for the joint distribution semiparametric. The DGP we consider is:

$$\mathbf{Y}_t = \mu_t(\phi_0) + \sigma_t(\phi_0)\,\eta_t \tag{2.16}$$

$$\begin{aligned}
\text{where}\quad \mathbf{Y}_t &\equiv [Y_{1t}, \ldots, Y_{Nt}]' \\
\mu_t(\phi) &\equiv [\mu_{1t}(\phi), \ldots, \mu_{Nt}(\phi)]' \\
\sigma_t(\phi) &\equiv diag\{\sigma_{1t}(\phi), \ldots, \sigma_{Nt}(\phi)\} \\
\eta_t &\equiv [\eta_{1t}, \ldots, \eta_{Nt}]' \sim iid\ \ \mathbf{F}_\eta = \mathbf{C}(F_1, \ldots, F_N; \theta_0)
\end{aligned}$$

where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and independent of η_t. $\mathcal{F}_{t-1}$ is the sigma-field containing information generated by $\{\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots\}$. The $r \times 1$ vector of parameters governing the dynamics of the variables, ϕ_0, is assumed to be $\sqrt{T}$-consistently

[7] The extension to allow for time-varying conditional copulas is relatively simple empirically, but the asymptotic theory for the estiamted parameters needs non-trivial adjustment, and is not considered here.

estimable. If ϕ_0 is known, or if μ_t and σ_t are known constant, then the model becomes one for *iid* data. The copula is parameterized by a $p \times 1$ vector of parameters, $\theta_0 \in \mathbf{\Theta}$, which is estimated using the SMM approach below.

2.3.2 Simulation-based estimation of copula models

The simulation-based estimation method of Oh and Patton (2013a) is closely related to SMM estimation, though is not strictly SMM, as the "moments" that are used in estimation are functions of rank statistics. We will nevertheless refer to the method as SMM estimation. Our task is to estimate the $p \times 1$ vector of copula parameters, $\theta_0 \in \mathbf{\Theta}$, based on the standardized residual $\left\{\hat{\eta}_t \equiv \sigma_t^{-1}\left(\hat{\phi}\right)\left[\mathbf{Y}_t - \mu_t\left(\hat{\phi}\right)\right]\right\}_{t=1}^{T}$ and simulations from the copula model (for example, the factor copula model in equation (2.2). The SMM copula estimator of Oh and Patton (2011) is based on simulation from some parametric joint distribution, $\mathbf{F}_x(\theta)$, with implied copula $\mathbf{C}(\theta)$.

Let $\tilde{\mathbf{m}}_S(\theta)$ be a $(m \times 1)$ vector of dependence measures computed using S simulations from $\mathbf{F}_x(\theta)$, $\{\mathbf{X}_s\}_{s=1}^{S}$, and let $\hat{\mathbf{m}}_T$ be the corresponding vector of dependence measures computed using the standardized residuals $\{\hat{\eta}_t\}_{t=1}^{T}$. (We discuss the empirical choice of which dependence measures to match in Appendix B.2.) The SMM estimator then defined as:

$$\hat{\theta}_{T,S} \equiv \underset{\theta \in \mathbf{\Theta}}{\arg\min}\ Q_{T,S}(\theta) \tag{2.17}$$

$$\text{where}\quad Q_{T,S}(\theta) \equiv \mathbf{g}'_{T,S}(\theta)\,\hat{W}_T\mathbf{g}_{T,S}(\theta)$$

$$\mathbf{g}_{T,S}(\theta) \equiv \hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\theta)$$

and $\hat{W}_T$ is some positive definite weight matrix, which may depend on the data. Under regularity conditions, Oh and Patton (2013a) show that if $S/T \to \infty$ as

$T \to \infty$, the SMM estimator is consistent and asymptotically normal:[8]

$$\sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) \xrightarrow{d} N\left(0, \Omega_0\right) \text{ as } T, S \to \infty \tag{2.18}$$

$$\text{where} \quad \Omega_0 = \left(G_0' W_0 G_0\right)^{-1} G_0' W_0 \Sigma_0 W_0 G_0 \left(G_0' W_0 G_0\right)^{-1}$$

$\Sigma_0 \equiv avar\left[\hat{\mathbf{m}}_T\right]$, $G_0 \equiv \nabla_\theta \mathbf{g}_0\left(\theta_0\right)$, and $\mathbf{g}_0\left(\theta\right) = \text{p-lim}_{T,S\to\infty}\ \mathbf{g}_{T,S}\left(\theta\right)$. Oh and Patton (2011) also present the distribution of a test of the over-identifying restrictions (the "J" test).

The asymptotic variance of the estimator has the same form as in standard GMM applications, however the components Σ_0 and G_0 require different estimation methods than in standard applications. Oh and Patton (2013a) show that a simple *iid* bootstrap can be used to consistently estimate Σ_0, and that a standard numerical derivative of $\mathbf{g}_{T,S}\left(\theta\right)$ at $\hat{\theta}_{T,S}$, denoted $\hat{G}$, will consistently estimate G_0 under the condition that the step size of the numerical derivative goes to zero slower than $T^{-1/2}$. In our simulation study we thoroughly examine the sensitivity of the estimated covariance matrix to the choice of step size.

2.3.3 Finite-sample properties of SMM estimation of factor copulas

In this section we present a study of the finite sample properties of the simulated method of moments (SMM) estimator of the parameters of various factor copulas. In the one case where a likelihood for the copula model is available in closed form we contrast the properties of the SMM estimator with those of the maximum likelihood estimator.

---

[8] Oh and Patton (2013a) also consider the case that $S/T \to 0$ as $S, T \to \infty$, in which case the convergence rate is $\sqrt{S}$ rather than $\sqrt{T}$. In our empirical application we have $S \gg T$, and so we do not present that case here.

Simulation design

We initially consider three different factor copulas, all of them of the form:

$$\begin{aligned} X_i &= Z + \varepsilon_i, \quad i = 1, 2, ..., N \\ Z &\sim Skew\ t\left(\sigma_z^2, \nu, \lambda\right) \\ \varepsilon_i &\sim iid\ t\left(\nu\right), \text{ and } \varepsilon_i \perp\!\!\!\perp Z\ \forall\ i \\ [X_1, ..., X_N]' &\sim \mathbf{F}_x = \mathbf{C}\left(G_x, ..., G_x\right) \end{aligned} \tag{2.19}$$

and we use the skewed t distribution of Hansen (1994) for the common factor. In all cases we set $\sigma_z^2 = 1$, implying that the common factor (Z) accounts for one-half of the variance of each X_i, implying rank correlation of around 0.5. In the first model we set $\nu \to \infty$ and $\lambda = 0$, which implies that the resulting factor copula is simply the Gaussian copula, with equicorrelation parameter $\rho = 0.5$. In this case we can estimate the model by SMM and also by GMM and MLE, and we use this case to study the loss of efficiency in moving from MLE to GMM to SMM. In the second model we set $\nu = 4$ and $\lambda = 0$, yielding a symmetric factor copula that generates tail dependence. In the third case we set $\nu = 4$ and $\lambda = -0.5$ yielding a factor copula that generates tail dependence as well as "asymmetric dependence", in that the lower tails of the copula are more dependent than the upper tails. We estimate the inverse degrees of freedom parameter, ν_z^{-1}, so that its parameter space is $[0, 0.5)$ rather than $(2, \infty]$.

We also consider an extension of the above equidependence model which allow each X_i to have a different coefficient on Z, as in equation (2.3). For identification of this model we set $\sigma_z^2 = 1$. For $N = 3$ we set $[\beta_1, \beta_2, \beta_3] = [0.5, 1, 1.5]$. For $N = 10$ we set $[\beta_1, \beta_2, ..., \beta_{10}] = [0.25, 0.50, ..., 2.5]$, which corresponds to pair-wise rank correlations ranging from approximately 0.1 to 0.8. Motivated by our empirical application below, for the $N = 100$ case we consider a "block equidependence" model, where we assume that the 100 variables can be grouped *ex ante* into 10 groups, and that all variables within each group have the same β_i. We use the same set of values for β_i as in the $N = 10$ case.

We consider two different scenarios for the marginal distributions of the variables of interest. In the first case we assume that the data are *iid* with standard Normal marginal distributions, meaning that the only parameters that need to be estimated are those of the factor copula. This simplified case is contrasted with a second scenario where the marginal distributions of the variables are assumed to follow an AR(1)-GARCH(1,1) process:

$$\begin{aligned} Y_{it} &= \phi_0 + \phi_1 Y_{i,t-1} + \sigma_{it}\eta_{it}, \quad t = 1, 2, ..., T \\ \sigma_{it}^2 &= \omega + \gamma\sigma_{i,t-1}^2 + \alpha\sigma_{i,t-1}^2\eta_{i,t-1}^2 \\ \eta_t &\equiv [\eta_{1t}, ..., \eta_{Nt}] \sim iid \;\; \mathbf{F}_\eta = \mathbf{C}(\Phi, \Phi, ..., \Phi) \end{aligned} \tag{2.20}$$

where Φ is the standard Normal distribution function and $\mathbf{C}$ is the factor copula implied by equation (2.19). We set the parameters of the marginal distributions as $[\phi_0, \phi_1, \omega, \gamma, \alpha] = [0.01, 0.05, 0.05, 0.85, 0.10]$, which broadly matches the values of these parameters when estimated using daily equity return data. In this scenario the parameters of the marginal distribution are estimated in a separate first stage, following which the estimated standardized residuals, $\hat{\eta}_{it}$, are obtained and used in a second stage to estimate the factor copula parameters. In all cases we consider a time series of length $T = 1000$, corresponding to approximately 4 years of daily return data, and we use $S = 25 \times T$ simulations in the computation of the dependence measures to be matched in the SMM optimization. We repeat each scenario 100 times. In all results below we use the identity weight matrix for estimation; corresponding results based on the efficient weight matrix are available in Appendix B.3.[9] In Appendix B.2 we describe the dependence measures we use for the estimation of

[9] The results based on the efficient weight matrix are generally comparable to those based on the identity weight matrix, however the coverage rates are worse than those based on the identity weight matrix.

these models.

Simulation results

Table 2.1 reveals that for all three dimensions ($N = 3$, 10 and 100) and for all three copula models the estimated parameters are centered on the true values, with the average estimated bias being small relative to the standard deviation, and with the median of the simulated distribution centered on the true values. The measures of estimator accuracy (the standard deviation and the 90-10 percentile difference) reveal that adding more parameters to the model, *ceteris paribus*, leads to greater estimation error, as expected; the σ_z^2 parameter, for example, is more accurately estimated when it is the only unknown parameter compared with when it is one of three unknown parameters. Looking across the dimension size, we see that the copula model parameters are almost always more precisely estimated as the dimension grows. This is intuitive, given the equidependence nature of all three models: increasing the dimension of the model does not increase the number of parameters to be estimated but it does increase the amount of information available on the unknown parameters.

Comparing the SMM estimator with the ML estimator, which is only feasible for the Normal copula (as the other two factor copulas do not have a copula likelihood in closed form) we see that the SMM estimator performs quite well. As predicted by theory, the ML estimator is always more efficient than the SMM estimator, however the loss in efficiency is moderate, ranging from around 25% for $N = 3$ to around 10% for $N = 100$. This provides some confidence that our move to SMM, prompted by the lack of a closed-form likelihood, does not come at a cost of a large loss in efficiency. Comparing the SMM estimator to the GMM estimator provides us with a measure of the loss in accuracy from having to estimate the population moment function via simulation. We find that this loss is at most 3% and in some cases ($N = 100$) is slightly negative. Thus little is lost from using SMM rather than GMM.

Table 2.2 shows results for the block equidependence model for the $N = 100$ case with AR-GARCH marginal distributions,[10] which can be compared to the results in the lower panel of Table 2.1. This table shows that the parameters of these models are well estimated using the proposed dependence measures described in Appendix B.2. The accuracy of the "shape" parameters, ν^{-1} and λ, is slightly lower in the more general model, consistent with the estimation error from having to estimate ten factor loadings (β_i) being greater than from having to estimate just a single other parameter (σ_z^2), however this loss is not great.

In Tables 2.3 and 2.4 we present the finite-sample coverage probabilities of 95% confidence intervals based on the estimated asymptotic covariance matrix described in Section 2.3.2. As discussed above, a critical input to the asymptotic covariance matrix estimator is the step size used in computing the numerical derivative matrix $\hat{G}$. This step size, ε_T, must go to zero, but at a slower rate than $T^{-1/2}$. Ignoring constants, our simulation sample size of $T = 1000$ suggests setting $\varepsilon_T > 0.03$, which is much larger than standard step sizes used in computing numerical derivatives.[11] We consider a range of values from 0.0001 to 0.1. Table 2.4 shows that when the step size is set to 0.01, 0.03 or 0.1 the finite-sample coverage rates are close to their nominal levels. However if the step size is chosen too small (0.003 or smaller) then the coverage rates are much lower than nominal levels. For example, setting $\varepsilon_T = 0.0001$ (which is still 16 times larger than the default setting in Matlab) we find coverage rates as low as 38% for a nominal 95% confidence interval. Thus this table shows that the asymptotic theory provides a reliable means for obtaining confidence intervals,

[10] The results for *iid* data, and the results for this model for $N = 3$ and 10, are available in Appendix B.3.

[11] For example, the default in many *Matlab* functions is a step size of $\varepsilon^{1/3} \approx 6 \times 10^{-6} \approx 1/(165,000)$, where $\varepsilon = 2.22 \times 10^{-16}$ is machine epsilon. This choice is optimal in certain applications, see Judd (1998) for example.

so long as care is taken not to set the step size too small.

Finally in Table 2.5 we present the results of a study of the rejection rates for the J test of over-identifying restrictions. Given that we consider $W = I$ in this table, the test statistic has a non-standard distribution (see Proposition 4 of Oh and Patton, 2013a), and we use $10,000$ simulations to obtain critical values. In this case, the limiting distribution also depends on $\hat{G}$, and we present the rejection rates for various choices of step size ε_T. Table 2.5 reveals that the rejection rates are close to their nominal levels, for both the equidependence models and the "different loading" models (which is a block equidependence model for the $N = 100$ case). The J test rejection rates are less sensitive to the choice of step size than the coverage probabilities of confidence intervals, however the best results are again generally obtained when ε_T is 0.01 or greater.

2.4 High-dimension copula models for S&P 100 returns

In this section we apply our proposed factor copulas to a study of the dependence between a large collection of U.S. equity returns. We study all 100 stocks that were constituents of the S&P 100 index as at December 2010. The sample period is April 2008 to December 2010, a total of $T = 696$ trade days. The starting point for our sample period was determined by the date of the latest addition to the S&P 100 index (Philip Morris Inc.), which has had no additions or deletions since April 2008. The stocks in our study are listed in Table 2.6, along with their 3-digit SIC codes, which we will use in part of our analysis below.

Table 2.7 presents some summary statistics of the data used in this analysis. The top panel presents sample moments of the daily returns for each stock. The means and standard deviations are around values observed in other studies. The skewness and kurtosis coefficients reveal a substantial degree of heterogeneity in the shape of the distribution of these asset returns, motivating our use of a nonparametric estimate (the EDF) of this in our analysis.

In the second panel of Table 2.7 we present information on the parameters of the AR(1)–GJR-GARCH models, augmented with lagged market return information, that are used to filter each of the individual return series[12]:

$$r_{it} = \phi_{0i} + \phi_{1i} r_{i,t-1} + \phi_{mi} r_{m,t-1} + \varepsilon_{it} \tag{2.21}$$

$$\begin{aligned} \sigma_{it}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \varepsilon_{i,t-1}^2 + \gamma_i \varepsilon_{i,t-1}^2 \mathbf{1}\{\varepsilon_{i,t-1} \leqslant 0\} \\ + \alpha_{mi} \varepsilon_{m,t-1}^2 + \gamma_{mi} \varepsilon_{m,t-1}^2 \mathbf{1}\{\varepsilon_{m,t-1} \leqslant 0\} \end{aligned} \tag{2.22}$$

Estimates of the parameters of these models are consistent with those reported in numerous other studies, with a small negative AR(1) coefficient found for most though not all stocks, and with the lagged market return entering significantly in 37 out of the 100 stocks. The estimated GJR-GARCH parameters are strongly indicative of persistence in volatility, and the asymmetry parameter, γ, in this model is positive for all but three of the 100 stocks in our sample, supporting the wide-spread finding of a "leverage effect" in the conditional volatility of equity returns. The lagged market residual is also found to be important for volatility in many cases, with the null that $\alpha_{mi} = \gamma_{mi} = 0$ being rejected at the 5% level for 32 stocks.

In the lower panel of Table 2.7 we present summary statistics for four measures of dependence between pairs of standardized residuals: linear correlation, rank correlation, average upper and lower 1% tail dependence (equal to $(\tau_{0.99} + \tau_{0.01})/2$), and the difference in upper and lower 10% tail dependence (equal to $\tau_{0.90} - \tau_{0.10}$). The two correlation statistics measure the sign and strength of dependence, the third and fourth statistics measure the strength and symmetry of dependence in the tails. The two correlation measures are similar, and are 0.42 and 0.44 on average. Across

[12] We considered GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), and GJR-GARCH (Glosten, *et al.*, 1993) models for the conditional variance of these returns, and for almost all stocks the GJR-GARCH model was preferred according to the BIC.

all 4950 pairs of assets the rank correlation varies from 0.37 to 0.50 from the 25^{th} and 75^{th} percentiles of the cross-sectional distribution, indicating the presence of mild heterogeneity in the correlation coefficients. The 1% tail dependence measure is 0.06 on average, and varies from 0.00 to 0.07 across the inter-quartile range. The difference in the 10% tail dependence measures is negative on average, and indeed is negative for over 75% of the pairs of stocks, strongly indicating asymmetric dependence between these stocks.

2.4.1 Results from equidependence copula specifications

We now present our first empirical results on the dependence structure of these 100 stock returns: the estimated parameters of eight different models for the copula. We consider four existing copulas: the Clayton copula, the Normal copula, the Student's t copula, and the skew t copula, with equicorrelation imposed on the latter three models for comparability, and four factor copulas, described by the distributions assumed for the common factor and the idiosyncractic shock: t-Normal, Skew t-Normal, t-t, Skew t-t. All models are estimated using the SMM-type method described in Section 2.3.2. The value of the SMM objective function at the estimated parameters, Q_{SMM}, is presented for each model, along with the p-value from the J-test of the over-identifying restrictions. Standard errors are based on 1000 bootstraps to estimate $\Sigma_{T,S}$, and with a step size $\varepsilon_T = 0.1$ to compute $\hat{G}$.

Table 2.8 reveals that the variance of the common factor, σ_z^2, is estimated by all models to be around 0.9, implying an average correlation coefficient of around 0.47. The estimated inverse degrees of freedom (DoF) parameter in these models is around 1/25, and the standard errors on ν^{-1} reveal that this parameter is significant[13] at

[13] Note that the case of zero tail dependence corresponds to $\nu_z^{-1} = 0$, which is on the boundary of the parameter space for this parameter, implying that a standard t test is strictly not applicable. In such cases the squared t statistic no longer has an asymptotic χ_1^2 distribution under the null, rather it is distributed as an equal-weighted mixture of a χ_1^2 and χ_0^2, see Gourieroux and Monfort (1996, Ch 21). The 90% and 95% critical values for this distribution are 1.64 and 2.71 (compared

the 10% level for the three models that allow for asymmetric dependence, but not significant for the three models that impose symmetric dependence. The asymmetry parameter, λ, is significantly negative in all models in which it is estimated, with t-statistics ranging from -2.1 to -4.4. This implies that the dependence structure between these stock returns is significantly asymmetric, with large crashes being more likely than large booms. Other papers have considered equicorrelation models for the dependence between large collections of stocks, see Engle and Kelly (2012) for example, but empirically showing the importance of allowing the implied common factor to be fat tailed and asymmetric is novel.

Figure 2.4 presents the quantile dependence function from the estimated Normal copula and the estimated skew $t-t$ factor copula, along with the quantile dependence averaged across all pairs of stocks, and pointwise 90% bootstrap confidence intervals for these estimates based on the theory in Rémillard (2010). (The figure zooms in on the left and right 20% tails, removing the middle 60% of the distribution as the estimates and models are all very similar there.) This figure reveals that the Normal copula overestimates the dependence in the upper tail, and underestimates it in the lower tail. This is consistent with the fact that the empirical quantile dependence is asymmetric, while the Normal copula imposes symmetry. The skew $t-t$ factor copula provides a reasonable fit in both tails, though it somewhat overestimates the dependence in the extreme left tail.

Figure 2.5 exploits the high-dimensional nature of our analysis, and plots the expected proportion of "crashes" in the remaining $(100-j)$ stocks, conditional on observing a crash in j stocks. We show this for a "crash" defined as a once-in-a-month (1/22, around 4.6%) event and as a once-in-a-quarter (1/66, around 1.5%) event. For once-in-a-month crashes, the observed proportions track the Skew t-t factor copula well for j up to around 25 crashes, and again for j of around 70. For

with 2.71 and 3.84 for the χ^2_1 distribution), which correspond to t-statistics of 1.28 and 1.65.

j in between 30 and 65 the Normal copula appears to fit quite well. For once-in-a-quarter crashes, displayed in the lower panel of Figure 2.5, the empirical plot tracks that for the Normal copula well for j up to around 30, but for $j = 35$ the empirical plot jumps and follows the skew $t-t$ factor copula. Thus it appears that the Normal copula may be adequate for modeling moderate tail events, but a copula with greater tail dependence (such as the skew $t - t$ factor copula) is needed for more extreme tail events.

The last two columns of Table 2.8 report the value of the objective function (Q_{SMM}) and the p-value from a test of the over-identifying restrictions. The Q_{SMM} values reveal that the three models that allow for asymmetry (skew t copula, and the two skew t factor copulas) out-perform all the other models, and reinforce the above conclusion that allowing for a skewed common factor is important for this collection of assets. The p-values, however, are near zero for all models, indicating that none of them pass this specification test. One likely source of these rejections is the assumption of equidependence, which was shown in the summary statistics in Table 2.7 to be questionable for this large set of stock returns. We relax this in the next section.

2.4.2 Results from block *equidependence copula specifications*

In response to the rejection of the copula models based on equidependence, we now consider a generalization to allow for heterogeneous dependence. We propose a multi-factor model that allows for a common, market-wide, factor, and a set of factors related only to specific industries. We use the first digit of Standard Industrial Classification (SIC) to form seven groups of stocks, see Table 2.6. The model we consider is the copula generated by the following structure:

$$
\begin{aligned}
X_i &= \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \;\; i = 1, 2, ..., 100 \\
Z_0 &\sim Skew\ t\,(\nu, \lambda) \\
Z_S &\sim iid\ t\,(\nu)\,, \;\; S = 1, 2, ..., 7; \;\; Z_S \perp\!\!\!\perp Z_0\ \forall\ S \\
\varepsilon_i &\sim iid\ t\,(\nu)\,, \;\; i = 1, 2, ..., 100; \;\; \varepsilon_i \perp\!\!\!\perp Z_j\ \forall\ i, j
\end{aligned}
\tag{2.23}
$$

where $S\,(i)$ is the SIC group for stock i. There are eight latent factors in total in this model, but any given variable is only affected by two factors, simplifying its structure and reducing the number of free parameters. Note here we impose that the industry factors and the idiosyncratic shocks are symmetric, and only allow asymmetry in the market-wide factor, Z_0. It is feasible to consider allowing the industry factors to have differing levels of asymmetry, but we rule this out in the interests of parsimony. We impose that all stocks in the same SIC group have the same factor loadings, but allow stocks in different groups to have different factor loadings. This generates a "block equidependence" model which greatly increases the flexibility of the model, but without generating too many additional parameters to estimate. In total, this copula model has a total of 16 parameters, providing more flexibility than the 3-parameter equidependence model considered in the previous section, but still more parsimonious (and tractable) than a completely unstructured approach to this 100-dimensional problem.[14]

The results of this model are presented in Table 2.9. The Clayton copula is not presented here as it imposes equidependence by construction, and so is not comparable to the other models. The estimated inverse DoF parameter, ν^{-1}, is around 1/14, somewhat larger and more significant than for the equidependence model, indicating stronger evidence of tail dependence. The asymmetry parameters are also larger (in

[14] We also considered a one-factor model that allowed for different factor loadings, generalizing the equidependence model of the previous section but simpler than this multi-factor copula model. That model provided a significantly better fit than the equidependence model, but was also rejected using the J test of over-identifying restrictions, and so is not presented here to conserve space.

absolute value) and more significantly negative in this more flexible model than in the equidependence model. It appears that when we add variables that control for intra-industry dependence, (i.e., industry-specific factors) we find the market-wide common factor is more fat tailed and left skewed than when we impose a single factor structure.

Focussing on our preferred *skew* $t-t$ factor copula model, the coefficients on the market factor, β_i, range from 0.88 (for SIC group 2, Manufacturing: Food, apparel, etc.) to 1.25 (SIC group 1, Mining and construction), and in all cases significantly different from zero at the 5% level, indicating the varying degrees of inter-industry dependence. The coefficients on the industry factors, γ_i, measure the degree of additional intra-industry dependence, beyond that coming from the market-wide factor. These range from 0.17 to 1.09 for SIC groups 3 and 1 respectively. Even for the smaller estimates, these are significantly different from zero, indicating the presence of industry factors beyond a common market factor. The intra- and inter-industry rank correlations and tail dependence coefficients implied by this model[15] are presented in Table 2.10, and reveal the degree of heterogeneity and asymmetry that this copula captures: rank correlations range from 0.39 (for pairs of stocks in SIC groups 1 and 5) to 0.72 (for stocks within SIC group 1). The upper and lower tail dependence coefficients further reinforce the importance of asymmetry in the dependence structure, with lower tail dependence measures being substantially larger than upper tail measures: lower tail dependence averages 0.82 and ranges from 0.70 to 0.99, while upper tail dependence averages 0.07 and ranges from 0.02 to 0.74.

With this more flexible model we can test restrictions on the factor coefficients, to see whether the additional flexibility is required to fit the data. The p-values from these tests are in the bottom rows of Table 2.9. Firstly, we can test whether all of

[15] Rank correlations from this model are not available in closed form, and we use 50,000 simulations to estimate these. Upper and lower tail dependence coefficients are based on Propositions 2 and 3.

the industry factor coefficients are zero, which reduces this model to a one-factor model with flexible weights. The p-values from these tests are zero to four decimal places for all models, providing strong evidence in favor of including industry factors. We can also test whether the market factor is needed given the inclusion of industry factors by testing whether all betas are equal to zero, and predictably this restriction is strongly rejected by the data. We further can test whether the coefficients on the market and industry factors are common across all industries, reducing this model to an equidependence model, and this too is strongly rejected. Finally, we use the J test of over-identifying restrictions to check the specification of these models. Using this test, we see that the models that impose symmetry are strongly rejected. The skew t copula has a p-value of 0.04, indicating a marginal rejection, and the *skew* $t{-}t$ factor copula performs best, passing this test at the 5% level, with a p-value of 0.07.

Thus it appears that a multi-factor model with heterogeneous weights on the factors, that allows for positive tail dependence and stronger dependence in crashes than booms, is needed to fit the dependence structure of these 100 stock returns.

2.4.3 Measuring systemic risk: Marginal Expected Shortfall

The recent financial crisis has highlighted the need for the management and measurement of systemic risk, see Acharya *et al.* (2010) for discussion. Brownlees and Engle (2011) propose a measure of systemic risk they call "marginal expected shortfall", or MES. It is defined as the expected return on stock i given that the market return is below some (low) threshold:

$$MES_{it} = -E_{t-1}\left[r_{it}|r_{mt} < C\right] \tag{2.24}$$

An appealing feature of this measure of systemic risk is that it can be computed with only a bivariate model for the conditional distribution of (r_{it}, r_{mt}), and Brownlees and Engle (2011) propose a semiparametric model based on a bivariate DCC-GARCH model to estimate it. A corresponding drawback of this measure is that by using the market index to identify periods of crisis, it may overlook periods with crashes in individual firms. With a model for the entire set of constituent stocks, such as the high dimension copula models considered in this paper, combined with standard AR-GARCH type models for the marginal distributions, we can estimate the MES measure proposed in Brownlees and Engle (2011), as well as alternative measures that use crashes in individual stocks as flags for periods of turmoil. For example, one might consider the expected return on stock i conditional on k stocks in the market having returns below some threshold, a "kES":

$$kES_{it} = -E_{t-1}\left[r_{it}\middle|\left(\sum\nolimits_{j=1}^{N}\mathbf{1}\left\{r_{jt} < C\right\}\right) > k\right] \tag{2.25}$$

Brownlees and Engle (2011) propose a simple method for ranking estimates of MES:

$$MSE_i = \frac{1}{T}\sum_{t=1}^{T}\left(r_{it} - MES_{it}\right)^2 \mathbf{1}\left\{r_{mt} < C\right\} \tag{2.26}$$

$$RelMSE_i = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_{it} - MES_{it}}{MES_{it}}\right)^2 \mathbf{1}\left\{r_{mt} < C\right\}$$

Corresponding metrics immediately follow for estimates of "kES".

In Table 2.11 we present the MSE and RelMSE for estimates of MES and kES, for threshold choices of -2% and -4%. We implement the model proposed by Brownlees and Engle (2011), as well as their implementations of a model based on the CAPM, and one based purely on rolling historical information. Along with these, we present results for four copulas: the Normal, Student's t, skew t, and skew $t-t$ factor copula, all with the block equidependence structure from Section 2.4.2 above. In the upper panel of Table 2.11 we see that the Brownlees-Engle model performs the best for both thresholds under the MSE performance metric, with the skew $t-t$ factor copula as the second-best performing model. Under the Relative MSE metric, the factor copula is best performing model, for both thresholds, followed by the skew t copula. Like Brownlees and Engle (2011), we find that the worst-performing methods under both metrics are the Historical and CAPM methods.

The lower panel of Table 2.11 presents the performance of various methods for estimating kES, with k set to 30.[16] This measure requires an estimate of the conditional distribution for the entire set of 100 stocks, and thus the CAPM and Brownlees-Engle methods cannot be applied. We evaluate the remaining five methods, and find that the skew $t-t$ factor copula performs the best for both thresholds, under both metrics. Thus our proposed factor copula model for high dimensional dependence not only allows us to gain some insights into the structure of the dependence between this large collection of assets, but also provides improved estimates of measures of systemic risk.

2.5 Conclusion

This paper presents new models for the dependence structure, or copula, of economic variables based on a simple factor structure for the copula. These models are particularly attractive for high dimensional applications, involving fifty or more variables, as they allow the researcher to increase or decrease the flexibility of the model according to the amount of data available and the dimension of the problem, and, importantly, to do so in a manner that is easily interpreted. The class of factor copulas presented in this paper does not generally have a closed-form likelihood. We use extreme value theory to obtain analytical results on the tail dependence implied by factor copulas, and we consider SMM-type methods for the estimation of factor copulas. Via an extensive Monte Carlo study, we show that SMM estimation has good finite-sample

[16] We choose this value of k so that the number of identified "crisis" days is broadly comparable to the number of such days for MES. Results for alternative values of k are similar.

properties in time series applications involving up to 100 variables.

We employ our proposed factor copulas to study daily returns on all 100 constituents of the S&P 100 index over the period 2008-2010, and find significant evidence of a skewed, fat-tailed common factor, which generates asymmetric dependence and tail dependence. In an extension to a multi-factor copula, we find evidence of the importance of industry factors, leading to heterogeneous dependence. We also consider an application to the estimation of systemic risk, and we show that the proposed factor copula model provides superior estimates of two measures of systemic risk.

2.6 Tables and figures

Table 2.1: Simulation results for factor copula models

	Normal			Factor $t-t$		Factor $skew\ t-t$		
	MLE	GMM	SMM					
	σ_z^2	σ_z^2	σ_z^2	σ_z^2	ν^{-1}	σ_z^2	ν^{-1}	λ
True	1.00	1.00	1.00	1.00	0.25	1.00	0.25	-0.50
				$N=3$				
Bias	0.0141	-0.0143	-0.0164	-0.0016	-0.0185	0.0126	-0.0199	-0.0517
Std	0.0803	0.1014	0.1033	0.1094	0.0960	0.1205	0.1057	0.1477
Med	1.0095	0.9880	0.9949	0.9956	0.2302	1.0050	0.2380	-0.5213
90%	1.1180	1.1103	1.1062	1.1448	0.3699	1.1772	0.3636	-0.3973
10%	0.9172	0.8552	0.8434	0.8721	0.0982	0.8662	0.0670	-0.7538
Diff	0.2008	0.2551	0.2628	0.2727	0.2716	0.3110	0.2966	0.3565
				$N=10$				
Bias	0.0113	-0.0099	-0.0119	-0.0025	-0.0137	-0.0039	-0.0161	-0.0119
Std	0.0559	0.0651	0.0666	0.0724	0.0611	0.0851	0.0790	0.0713
Med	1.0125	0.9874	0.9898	0.9926	0.2360	0.9897	0.2376	-0.5084
90%	1.0789	1.0644	1.0706	1.0967	0.3102	1.1095	0.3420	-0.4318
10%	0.9406	0.9027	0.8946	0.9062	0.1704	0.8996	0.1331	-0.5964
Diff	0.1383	0.1617	0.1761	0.1905	0.1398	0.2100	0.2089	0.1645
				$N=100$				
Bias	0.0167	-0.0068	-0.0080	-0.0011	-0.0138	0.0015	-0.0134	-0.0099
Std	0.0500	0.0554	0.0546	0.0659	0.0549	0.0841	0.0736	0.0493
Med	1.0164	0.9912	0.9956	1.0011	0.2346	0.9943	0.2402	-0.5101
90%	1.0805	1.0625	1.0696	1.0886	0.3127	1.1060	0.3344	-0.4465
10%	0.9534	0.9235	0.9279	0.9112	0.1685	0.8970	0.1482	-0.5734
Diff	0.1270	0.1390	0.1418	0.1773	0.1442	0.2089	0.1861	0.1270

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the $skew\ t-t$ factor copula. The Normal copula is estimated by ML, GMM, and SMM, and the other two copulas are estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 2.3. Problems of dimension $N = 3$, 10 and 100 are considered, the sample size is $T = 1000$ and the number of simulations used is $S = 25 \times T$. 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 50^{th}, 90^{th} and 10^{th} percentiles of the distribution of estimated parameters, and the final row presents the difference between the 90^{th} and 10^{th} percentiles.

Table 2.2: Simulation results for different loadings factor copula model with N=100

	ν^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True	0.25	-0.5	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
	Normal											
Bias	-	-	-0.0010	-0.0038	-0.0040	-0.0072	-0.0071	-0.0140	-0.0178	-0.0119	-0.0194	-0.0208
Std	-	-	0.0128	0.0182	0.0248	0.0322	0.0377	0.0475	0.0651	0.0784	0.1022	0.1291
Med	-	-	0.2489	0.4970	0.7440	0.9942	1.2421	1.4868	1.7279	1.9918	2.2256	2.4832
90%	-	-	0.2645	0.5204	0.7787	1.0291	1.2970	1.5470	1.8226	2.0874	2.3609	2.6458
10%	-	-	0.2304	0.4701	0.7158	0.9502	1.1982	1.4197	1.6526	1.8825	2.0921	2.3090
diff	-	-	0.0341	0.0503	0.0629	0.0788	0.0987	0.1273	0.1700	0.2049	0.2689	0.3368
	Factor $t-t$											
Bias	-0.0120	-	0.0000	0.0009	0.0018	-0.0045	0.0011	-0.0073	-0.0080	-0.0122	-0.0061	-0.0065
Std	0.0574	-	0.0149	0.0236	0.0300	0.0343	0.0443	0.0580	0.0694	0.0867	0.1058	0.1332
Med	0.2384	-	0.2503	0.5056	0.7528	0.9985	1.2550	1.4881	1.7409	1.9820	2.2234	2.4737
90%	0.3056	-	0.2678	0.5255	0.7896	1.0348	1.3052	1.5697	1.8270	2.1012	2.4089	2.6597
10%	0.1683	-	0.2348	0.4689	0.7187	0.9462	1.1965	1.4282	1.6517	1.8744	2.1303	2.3196
diff	0.1373	-	0.0330	0.0566	0.0709	0.0886	0.1086	0.1416	0.1754	0.2268	0.2786	0.3401
	Factor *skew* $t-t$											
Bias	-0.0119	-0.0019	0.0008	0.0001	0.0028	-0.0029	-0.0036	-0.0096	-0.0114	-0.0232	-0.0178	-0.0194
Std	0.0633	0.0451	0.0134	0.0246	0.0320	0.0443	0.0588	0.0806	0.0902	0.1111	0.1373	0.1635
Med	0.2434	-0.5051	0.2477	0.5001	0.7520	0.9986	1.2468	1.4826	1.7417	1.9803	2.2107	2.4786
90%	0.3265	-0.4392	0.2680	0.5309	0.7961	1.0613	1.3028	1.5856	1.8378	2.1094	2.4430	2.7034
10%	0.1550	-0.5527	0.2358	0.4660	0.7155	0.9505	1.1756	1.4042	1.6230	1.8395	2.0494	2.2739
diff	0.1714	0.1134	0.0321	0.0648	0.0807	0.1107	0.1272	0.1814	0.2148	0.2699	0.3936	0.4294

Notes: This table presents the results from 100 simulations of three different factor copulas: the Normal copula, the $t-t$ factor copula and the *skew* $t-t$ factor copula. We divide the $N = 100$ variables into ten groups and assume that all variables in the same group have the same loading on the common factor. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 2.3. The sample size is $T = 1000$ and the number of simulations used is $S = 25\times T$. 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 50^{th}, 90^{th} and 10^{th} percentiles of the distribution of estimated parameters, and the final row presents the difference between the 90^{th} and 10^{th} percentiles.

Table 2.3: Simulation results on coverage rates

	Normal	Factor $t-t$		Factor *skew* $t-t$		
	σ_z^2	σ_z^2	ν^{-1}	σ_z^2	ν^{-1}	λ
			$N=3$			
ε_T						
0.1	89	93	97	99	100	96
0.03	90	94	98	99	98	96
0.01	88	92	98	99	96	95
0.003	85	95	95	96	89	95
0.001	83	89	89	92	84	93
0.0003	58	69	69	74	74	74
0.0001	38	49	53	57	70	61
			$N=10$			
ε_T						
0.1	87	93	99	97	98	99
0.03	87	95	99	97	98	97
0.01	87	94	96	97	98	95
0.003	87	95	95	98	95	96
0.001	87	95	93	96	90	95
0.0003	86	94	87	91	77	93
0.0001	71	87	81	71	81	85
			$N=100$			
ε_T						
0.1	95	93	95	94	95	94
0.03	95	94	94	94	94	94
0.01	95	93	93	94	94	94
0.003	94	95	93	94	94	94
0.001	94	94	92	94	93	95
0.0003	92	94	92	94	92	93
0.0001	84	94	89	94	88	95

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the *skew* $t-t$ factor copula, all estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 2.3. Problems of dimension $N=3$, 10 and 100 are considered, the sample size is $T=1000$ and the number of simulations used is $S=25\times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$. The numbers in the table present the percentage of simulations for which the 95% confidence interval based on the estimated covariance matrix contained the true parameter.

Table 2.4: Coverage rate for different loadings factor copula model with N=100 AR-GARCH data

	ν^{-1}	λ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
	Normal											
ε_T												
0.1	-	-	97	91	92	89	95	93	94	95	95	90
0.03	-	-	97	91	92	90	95	95	94	95	95	90
0.01	-	-	97	91	92	90	95	94	94	96	94	91
0.003	-	-	97	90	93	90	95	94	95	96	95	90
0.001	-	-	97	90	94	93	94	94	94	96	94	92
0.0003	-	-	97	92	93	92	95	94	91	93	92	94
0.0001	-	-	94	94	91	88	90	92	94	91	88	86
	Factor $t-t$											
ε_T												
0.1	95	-	94	93	96	96	98	91	93	92	95	93
0.03	94	-	94	91	96	96	98	92	93	92	97	93
0.01	95	-	94	94	97	96	97	93	93	92	98	93
0.003	94	-	94	94	97	96	97	94	94	95	98	95
0.001	94	-	93	93	97	97	97	92	96	94	100	94
0.0003	90	-	94	95	98	97	99	94	95	95	99	93
0.0001	65	-	95	96	96	98	98	92	96	94	97	91
	Factor *skew* $t-t$											
ε_T												
0.1	93	95	98	95	96	94	94	92	91	91	90	92
0.03	93	95	98	95	95	94	95	92	91	91	89	90
0.01	93	95	97	96	95	94	94	92	92	91	91	91
0.003	93	95	97	96	96	94	95	92	92	92	90	89
0.001	93	94	97	96	95	94	94	91	91	93	89	88
0.0003	84	93	98	95	95	95	95	90	90	88	83	85
0.0001	69	86	98	97	94	91	90	88	87	84	83	80

Notes: This table presents the results from 100 simulations of three different factor copulas: the Normal copula, the $t-t$ factor copula and the *skew* $t-t$ factor copula. We divide the $N = 100$ variables into ten groups and assume that all variables in the same group have the same loading on the common factor. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 2.3. The sample size is $T = 1000$ and the number of simulations used is $S = 25\times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$. The numbers in the table present the percentage of simulations for which the 95% confidence interval based on the estimated covariance matrix contained the true parameter.

Table 2.5: Rejection frequencies for the test of overidentifying restrictions

	Equidependence			Different loadings		
	Normal	Factor $t-t$	Factor *skew* $t-t$	Normal	Factor $t-t$	Factor *skew* $t-t$
			$N=3$			
ε_T						
0.1	97	97	99	95	97	97
0.03	97	98	99	95	95	96
0.01	97	97	100	93	95	95
0.003	97	98	100	92	95	96
0.001	98	96	100	93	93	97
0.0003	99	97	100	91	92	97
0.0001	99	97	99	92	94	98
			$N=10$			
ε_T						
0.1	97	97	98	98	95	98
0.03	98	97	97	98	95	99
0.01	96	97	97	97	94	98
0.003	97	96	97	98	92	99
0.001	98	95	97	96	89	100
0.0003	97	94	97	97	93	100
0.0001	97	94	98	98	95	100
			$N=100$			
ε_T						
0.1	97	95	99	95	95	99
0.03	97	95	98	96	94	99
0.01	97	95	98	96	93	99
0.003	97	95	97	95	94	99
0.001	97	94	99	95	91	100
0.0003	97	94	99	95	89	100
0.0001	98	92	98	93	90	100

Notes: This table presents the results from 100 simulations of three different factor copulas, the Normal copula, the $t-t$ factor copula and the *skew* $t-t$ factor copula, all estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 2.3. Problems of dimension $N = 3$, 10 and 100 are considered, the sample size is $T = 1000$ and the number of simulations used is $S = 25 \times T$. The rows of each panel contain the step size, ε_T, used in computing the matrix of numerical derivatives, $\hat{G}_{T,S}$, needed for the critical value. The confidence level for the test of over-identifying restrictions is 0.95, and the numbers in the table present the percentage of simulations for which the test statistic was greater than its computed critical value.

Table 2.6: Stocks used in the empirical analysis

Ticker	Name	SIC	Ticker	Name	SIC	Ticker	Name	SIC
AA	Alcoa	333	EXC	Exelon	493	NKE	Nike	302
AAPL	Apple	357	F	Ford	371	NOV	National Oilwell	353
ABT	Abbott Lab.	283	FCX	Freeport	104	NSC	Norfolk Sth	671
AEP	American Elec	491	FDX	Fedex	451	NWSA	News Corp	271
ALL	Allstate Corp	633	GD	GeneralDynam	373	NYX	NYSE Euronxt	623
AMGN	Amgen Inc.	283	GE	General Elec	351	ORCL	Oracle	737
AMZN	Amazon.com	737	GILD	GileadScience	283	OXY	OccidentalPetrol	131
AVP	Avon	284	GOOG	Google Inc	737	PEP	Pepsi	208
AXP	American Ex	671	GS	GoldmanSachs	621	PFE	Pfizer	283
BA	Boeing	372	HAL	Halliburton	138	PG	Procter&Gamble	284
BAC	Bank of Am	602	HD	Home Depot	525	QCOM	Qualcomm Inc	366
BAX	Baxter	384	HNZ	Heinz	203	RF	Regions Fin	602
BHI	Baker Hughes	138	HON	Honeywell	372	RTN	Raytheon	381
BK	Bank of NY	602	HPQ	HP	357	S	Sprint	481
BMY	Bristol-Myers	283	IBM	IBM	357	SLB	Schlumberger	138
BRK	Berkshire Hath	633	INTC	Intel	367	SLE	Sara Lee Corp.	203
C	Citi Group	602	JNJ	Johnson&J.	283	SO	Southern Co.	491
CAT	Caterpillar	353	JPM	JP Morgan	672	T	AT&T	481
CL	Colgate	284	KFT	Kraft	209	TGT	Target	533
CMCSA	Comcast	484	KO	Coca Cola	208	TWX	Time Warner	737
COF	Capital One	614	LMT	Lock'dMartn	376	TXN	Texas Inst	367
COP	Conocophillips	291	LOW	Lowe's	521	UNH	UnitedHealth	632
COST	Costco	533	MA	Master card	615	UPS	United Parcel	451
CPB	Campbell	203	MCD	MaDonald	581	USB	US Bancorp	602
CSCO	Cisco	367	MDT	Medtronic	384	UTX	United Tech	372
CVS	CVS	591	MET	Metlife Inc.	671	VZ	Verizon	481
CVX	Chevron	291	MMM	3M	384	WAG	Walgreen	591
DD	DuPont	289	MO	Altria Group	211	WFC	Wells Fargo	602
DELL	Dell	357	PM	Philip Morris	211	WMB	Williams	492
DIS	Walt Disney	799	MON	Monsanto	287	WMT	WalMart	533
DOW	Dow Chem	282	MRK	Merck	283	WY	Weyerhauser	241
DVN	Devon Energy	131	MS	MorganStanley	671	XOM	Exxon	291
EMC	EMC	357	MSFT	Microsoft	737	XRX	Xerox	357
ETR	ENTERGY	491						

Notes: This table presents the ticker symbols, names and 3-digit SIC codes of the 100 stocks used in Section 2.4.

Table 2.7: Summary statistics

	Cross-sectional distribution					
	Mean	5%	25%	Med	75%	95%
Mean	0.0004	-0.0003	0.0001	0.0003	0.0006	0.0013
Std dev	0.0287	0.0153	0.0203	0.0250	0.0341	0.0532
Skewness	0.3458	-0.4496	-0.0206	0.3382	0.6841	1.2389
Kurtosis	11.3839	5.9073	7.5957	9.1653	11.4489	19.5939
ϕ_0	0.0004	-0.0004	0.0001	0.0004	0.0006	0.0013
ϕ_1	-0.0345	-0.2045	-0.0932	-0.0238	0.0364	0.0923
ϕ_m	-0.0572	-0.2476	-0.1468	-0.0719	0.0063	0.1392
$\omega \times 1000$	0.0126	0.0024	0.0050	0.0084	0.0176	0.0409
β	0.8836	0.7983	0.8639	0.8948	0.9180	0.9436
α	0.0240	0.0000	0.0000	0.0096	0.0354	0.0884
γ	0.0593	0.0000	0.0017	0.0396	0.0928	0.1628
α_m	0.0157	0.0000	0.0000	0.0000	0.0015	0.0646
γ_m	0.1350	0.0000	0.0571	0.0975	0.1577	0.3787
ρ	0.4155	0.2643	0.3424	0.4070	0.4749	0.5993
ρ_s	0.4376	0.2907	0.3690	0.4292	0.4975	0.6143
$(\tau_{0.99} + \tau_{0.01})/2$	0.0572	0.0000	0.0000	0.0718	0.0718	0.1437
$(\tau_{0.90} - \tau_{0.10})$	-0.0922	-0.2011	-0.1293	-0.0862	-0.0431	0.0144

Notes: This table presents some summary statistics of the daily equity returns data used in the empirical analysis. The top panel presents simple unconditional moments of the daily return series. The second panel presents summaries of the estimated AR(1)–GJR-GARCH(1,1) models estimated on these returns. The lower panel presents linear correlation, rank correlation, average 1% upper and lower tail dependence, and the difference between the 10% tail dependence measures, computed using the standardized residuals from the estimated AR–GJR-GARCH model. The columns present the mean and quantiles from the cross-sectional distribution of the measures listed in the rows. The top two panels present summaries across the $N = 100$ marginal distributions, while the lower panel presents a summary across the $N(N-1)/2 = 4950$ distinct pairs of stocks.

Table 2.8: Estimation results for daily returns on S&P 100 stocks

	σ_z^2		ν^{-1}		λ		Q_{SMM}	p-val
	Est	Std Err	Est	Std Err	Est	Std Err		
Clayton[†]	0.6017	0.0345	–	–	–	–	0.0449	0.0000
Normal	0.9090	0.0593	–	–	–	–	0.0090	0.0000
Student's t	0.8590	0.0548	0.0272	0.0292	–	–	0.0119	0.0000
Skew t	0.6717	0.0913	0.0532	0.0133	-8.3015	4.0202	0.0010	0.0020
Factor $t-N$	0.8978	0.0555	0.0233	0.0325	–	–	0.0101	0.0000
Factor *skew* $t-N$	0.8954	0.0565	0.0432	0.0339	-0.2452	0.0567	0.0008	0.0002
Factor $t-t$	0.9031	0.0591	0.0142	0.0517	–	–	0.0098	0.0000
Factor *skew* $t-t$	0.8790	0.0589	0.0797	0.0486	-0.2254	0.0515	0.0007	0.0005

Notes: This table presents estimation results for various copula models applied to 100 daily stock returns over the period April 2008 to December 2010. Estimates and asymptotic standard errors for the copula model parameters are presented, as well as the value of the SMM objective function at the estimated parameters and the p-value of the overidentifying restriction test. Note that the parameter λ lies in $(-1, 1)$ for the factor copula models, but in $(-\infty, \infty)$ for the Skew t copula; in all cases the copula is symmetric when $\lambda = 0$. [†]Note that the parameter of the Clayton copula is not σ_z^2 but we report it in this column for simplicity.

Table 2.9: Estimation results for daily returns on S&P 100 stocks for block equidependence copula models

	Normal		Student's t		Skew t		Factor $t-t$		Factor $skew\ t-t$	
	Est	Std Err	Est	Std Err	Est	Std Err	Est	Std Err	Est	Std Err
ν^{-1}	-	-	0.0728	0.0269	0.0488	0.0069	0.0663	0.0472	0.0992	0.0455
λ	-	-	-	-	-9.6597	1.0860	-	-	-0.2223	0.0550
β_1	1.3027	0.0809	1.2773	0.0754	1.1031	0.1140	1.2936	0.0800	1.2457	0.0839
β_2	0.8916	0.0376	0.8305	0.0386	0.7343	0.0733	0.8528	0.0362	0.8847	0.0392
β_3	0.9731	0.0363	0.9839	0.0380	0.9125	0.0638	1.0223	0.0389	1.0320	0.0371
β_4	0.9426	0.0386	0.8751	0.0367	0.7939	0.0715	0.9064	0.0375	0.9063	0.0415
β_5	1.0159	0.0555	0.9176	0.0523	0.8171	0.0816	0.9715	0.0527	0.9419	0.0551
β_6	1.1018	0.0441	1.0573	0.0435	0.9535	0.0733	1.0850	0.0441	1.0655	0.0457
β_7	1.0954	0.0574	1.0912	0.0564	1.0546	0.0839	1.1057	0.0535	1.1208	0.0601
γ_1	1.0339	0.0548	0.9636	0.0603	1.0292	0.0592	1.0566	0.0595	1.0892	0.0582
γ_2	0.4318	0.0144	0.3196	0.0388	0.3474	0.0411	0.3325	0.0215	0.2201	0.0457
γ_3	0.4126	0.0195	0.3323	0.0394	0.2422	0.0458	0.2147	0.0727	0.1701	0.0996
γ_4	0.4077	0.0235	0.3726	0.0328	0.3316	0.0289	0.3470	0.0273	0.2740	0.0495
γ_5	0.4465	0.0403	0.5851	0.0300	0.5160	0.0324	0.5214	0.0312	0.5459	0.0448
γ_6	0.6122	0.0282	0.5852	0.0351	0.5581	0.0286	0.5252	0.0274	0.5686	0.0407
γ_7	0.5656	0.0380	0.5684	0.0464	0.2353	0.0889	0.3676	0.0434	0.3934	0.0548
Q_{SMM}	0.1587		0.1567		0.0266		0.1391		0.0189	
J *p-value*	0.0000		0.0000		0.0437		0.0000		0.0722	
$\gamma_i = 0\ \forall i$	0.0000		0.0000		0.0000		0.0000		0.0000	
$\beta_i = 0\ \forall i$	0.0000		0.0000		0.0000		0.0000		0.0000	
$\beta_i = \beta_j, \gamma_i = \gamma_j\ \forall i,j$	0.0000		0.0000		0.0000		0.0000		0.0000	

Notes: This table presents estimation results for various block equidependence copula models applied to filtered daily returns on collections of 100 stocks over the period April 2008 to December 2010. Note that the parameter λ lies in $(-1,1)$ for the factor copula models, but in $(-\infty,\infty)$ for the Skew t copula; in all cases the copula is symmetric when $\lambda = 0$. The bottom three rows present p-values from tests of constraints on the coefficients on the factors.

Table 2.10: Rank correlation and tail dependence implied by a multi-factor model

	SIC 1	SIC 2	SIC 3	SIC 4	SIC 5	SIC 6	SIC 7
				Rank correlation			
SIC 1	0.72						
SIC 2	0.41	0.44					
SIC 3	0.44	0.45	0.51				
SIC 4	0.41	0.42	0.45	0.46			
SIC 5	0.39	0.40	0.44	0.41	0.53		
SIC 6	0.42	0.43	0.47	0.43	0.42	0.58	
SIC 7	0.45	0.46	0.50	0.46	0.44	0.47	0.57
				Lower \ Upper tail dependence			
SIC 1	0.99 \ 0.74	0.02	0.07	0.02	0.03	0.09	0.13
SIC 2	0.70	0.70 \ 0.02	0.02	0.02	0.02	0.02	0.02
SIC 3	0.92	0.70	0.92 \ 0.07	0.02	0.03	0.07	0.07
SIC 4	0.75	0.70	0.75	0.75 \ 0.02	0.02	0.02	0.02
SIC 5	0.81	0.70	0.81	0.75	0.81 \ 0.03	0.03	0.03
SIC 6	0.94	0.70	0.92	0.75	0.81	0.94 \ 0.09	0.09
SIC 7	0.96	0.70	0.92	0.75	0.81	0.94	0.96\ 0.14

Notes: This table presents the dependence measures implied by the estimated *skew* $t-t$ factor copula model reported in Table 2.9. This model implies a block equidependence structure based on the industry to which a stock belongs, and the results are presented with intra-industry dependence in the diagonal elements, and cross-industry dependence in the off-diagonal elements. The top panel present rank correlation coefficients based on 50,000 simulations from the estimated model. The bottom panel presents the theoretical upper tail dependence coefficients (upper triangle) and lower tail dependence coefficients (lower triangle) based on Propositions 2 and 3.

Table 2.11: Performance of methods for predicting systemic risk

	MSE		**RelMSE**	
Cut-off	*-2%*	*-4%*	*-2%*	*-4%*
	Marginal Expected Shortfall (MES)			
Brownlees-Engle	**0.9961**	**1.2023**	0.7169	0.3521
Historical	1.1479	1.6230	1.0308	0.4897
CAPM	1.1532	1.5547	0.9107	0.4623
Normal copula	1.0096	1.2521	0.6712	0.3420
t copula	1.0118	1.2580	0.6660	0.3325
Skew t copula	1.0051	1.2553	0.6030	0.3040
Skew $t-t$ factor copula	1.0012	1.2445	**0.5885**	**0.2954**
	k-Expected Shortfall (kES)			
Historical	1.1632	1.6258	1.4467	0.7653
Normal copula	1.0885	1.4855	1.3220	0.5994
t copula	1.0956	1.4921	1.4496	0.6372
Skew t copula	1.0898	1.4923	1.3370	0.5706
Skew $t-t$ factor copula	**1.0822**	**1.4850**	**1.1922**	**0.5204**

Notes: This table presents the MSE (left panel) and Relative MSE (right panel) for various methods of estimating measures of systemic risk. The top panel presents results for marginal expected shortfall (MES), defined in equation (2.24), and the lower panel presents results for k-expected shortfall (kES), defined in equation (2.25), with k set to 30. Two thresholds are considered, $C = -2\%$ and $C = -4\%$. There are 70 and 21 "event" days for MES under these two thresholds, and 116 and 36 "event" days for kES. The best-performing model for each threshold and performance metric is highlighted in bold.

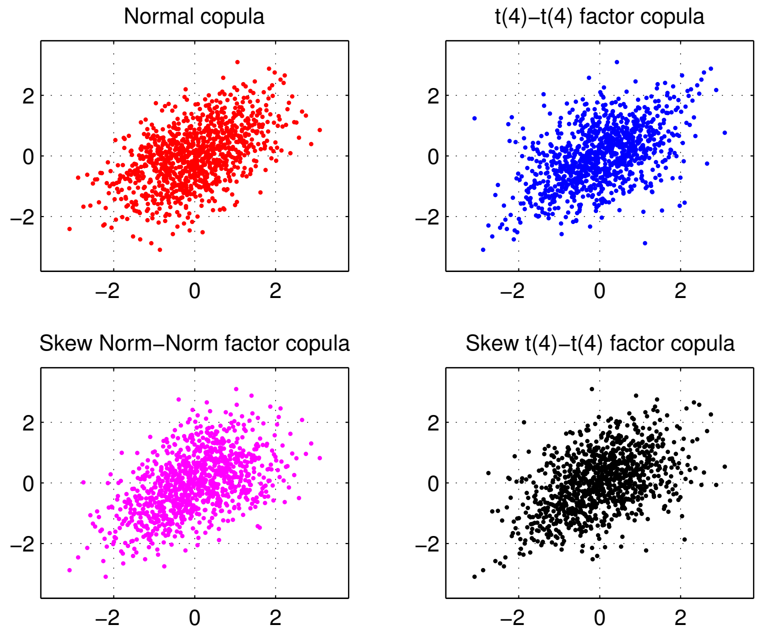


FIGURE 2.1: *Scatter plots from four bivariate distributions, all with N(0,1) margins and linear correlation of 0.5, constructed using four different factor copulas.*

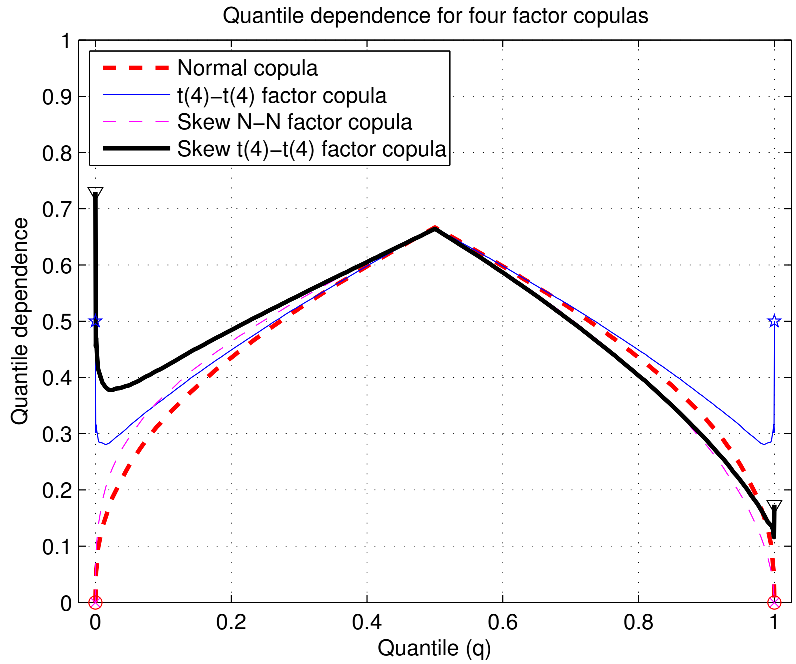


FIGURE 2.2: *Quantile dependence implied by four factor copulas, all with linear correlation of 0.5.*

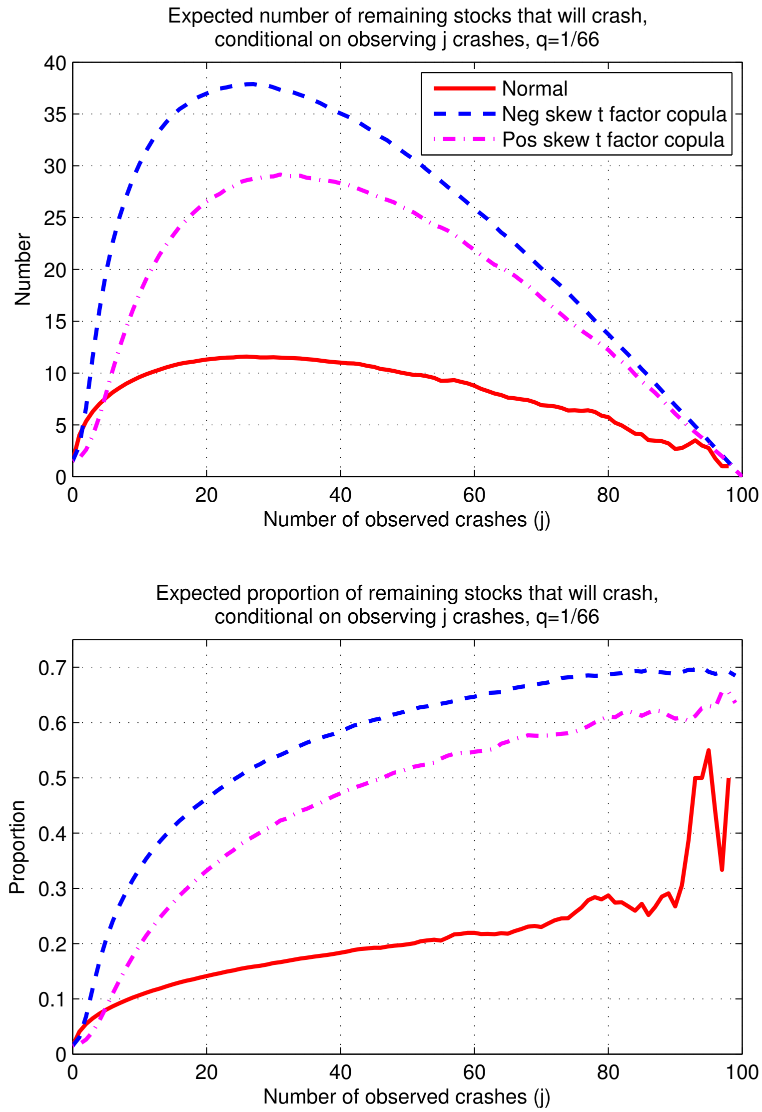


Figure 2.3: *Conditional on observing j out of 100 stocks crashing, this figure presents the expected number (upper panel) and proportion (lower panel) of the remaining (100-j) stocks that will crash. "Crash" events are defined as returns in the lower 1/66 tail.*

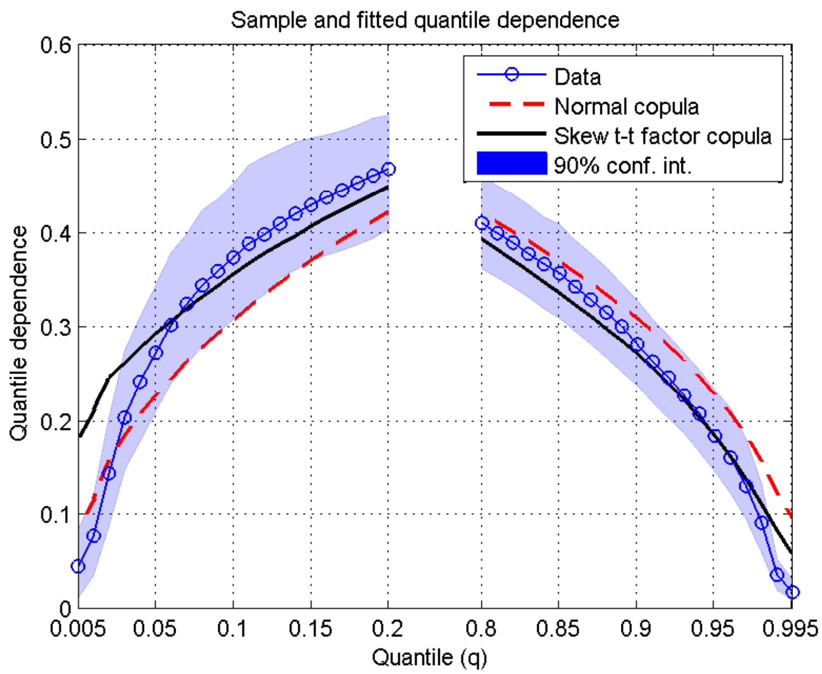
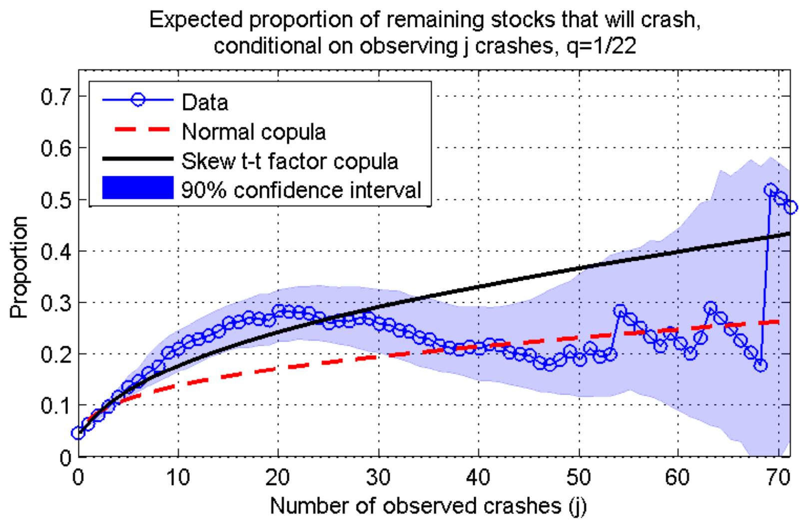


FIGURE 2.4: *Sample quantile dependence for 100 daily stock returns, along with the fitted quantile dependence from a Normal copula and from a Skew t-t factor copula, for the lower and upper tails.*



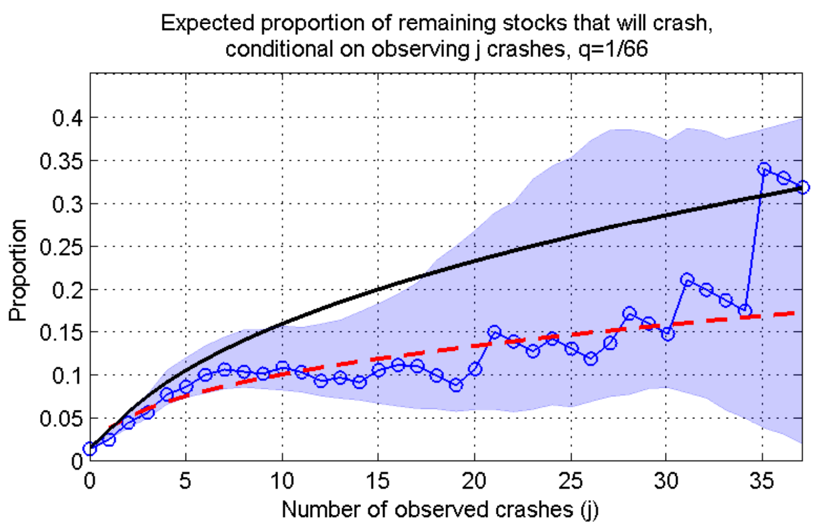


FIGURE 2.5: *Conditional on observing j out of 100 stocks crashing, this figure presents the expected proportion of the remaining (100-j) stocks that will crash. "Crash" events are defined as returns in the lower 1/22 (upper panel) and 1/66 (lower panel) tail. Note that the horizontal axes in these two panels are different, due to limited information in the joint tails.*

3

Simulated Method of Moments Estimation for Copula-Based Multivariate Models (co-authored with Andrew Patton)

3.1 Introduction

Copula-based models for multivariate distributions are widely used in a variety of applications, including actuarial science and insurance (Embrechts, McNeil and Straumann, 2002; Rosenberg and Schuermann 2006), economics (Brendstrup and Paarsch 2007; Bonhomme and Robin 2009), epidemiology (Clayton 1978; Fine and Jiang 2000), finance (Cherubini, Luciano and Vecchiato 2004; Patton 2006a), geology and hydrology (Cook and Johnson 1981; Genest and Favre 2007), among many others. An important benefit they provide is the flexibility to specify the marginal distributions separately from the dependence structure, without imposing that they come from the same family of joint distributions.

While copulas provide a great deal of flexibility in theory, the search for copula models that work well in practice is an ongoing one. This search has spawned a number of new and flexible models, see Demarta and McNeil (2005), McNeil, Frey and Embrechts (2005), Smith, Min, Almeida and Czado (2010), Smith, Gan and Kohn (2012), and Oh and Patton (2011), among others. Some of these models are such that the likelihood of the copula is either not known in closed form, or is complicated to obtain and maximize, motivating the consideration of estimation methods other than MLE. Moreover, in many financial applications, the estimated copula model is used in pricing a derivative security, such as a collateralized debt obligation or a credit default swap (CDO or CDS), and it may be of interest to minimize the pricing error (the observed market price less the model-implied price of the security) in calibrating the parameters of the model. In some cases the mapping from the parameter(s) of the copula to dependence measures (such as Spearman's or Kendall's rank correlation, for example) or to the price of the derivative contract is known in closed form, thus allowing for method of moments or generalized method of moments (GMM) estimation. In general, however, this mapping is unknown, and an alternative estimation method is required. We consider a simple yet widely applicable simulation-based approach to address this problem.

This paper presents the asymptotic properties of a simulation-based estimator of the parameters of a copula model. We consider both *iid* and time series data, and we consider the case that the marginal distributions are estimated using the empirical distribution function (EDF). The estimation method we consider shares features with the simulated method of moments (SMM), see McFadden (1989) and Pakes and Pollard (1989), for example, however the presence of the EDF in the sample "moments" means that existing results on SMM are not directly applicable. We draw on well-known results on SMM estimators, see Newey and McFadden (1994) for example, and recent results from empirical process theory for copulas, see Fermanian, Radulović and Wegkamp (2004), Chen and Fan (2006) and Rémillard (2010), to show the consistency and asymptotic normality of simulation-based estimators of copula models. To the best of our knowledge, simulation-based estimation of copula models has not previously been considered in the literature. An extensive simulation study verifies that the asymptotic results provide a good approximation in finite samples. We illustrate the results with an application to a model of the dependence between the equity returns on seven financial firms during the recent crisis period.

In addition to maximum likelihood, numerous other estimation methods have been considered for copula-based multivariate models. We describe these here and contrast them with the SMM approach proposed in this paper. Multi-stage maximum likelihood, also known as "inference functions for margins" in this literature (see Joe and Xu (1996) and Joe (2005) for *iid* data and Patton (2006b) for time series data) is one of the most widely-used estimation methods. The "maximization by parts" (MBP) algorithm of Song, *et al.* (2005) is an iterative method that improves the efficiency of multi-stage MLE, and attains full efficiency under some conditions. Like MLE, both of these methods only apply when the marginal distributions are parametric. When the marginal distribution models are correctly specified this improves the efficiency of the estimator, relative to the proposed SMM approach using nonparametric margins, however it introduces the possibility of mis-specified marginal distributions, which can have deleterious effects on the copula parameter estimates, see Kim, *et al.* (2007).

Semi-parametric maximum likelihood (see Genest, Ghoudi and Rivest (1995) for *iid* data and Chen and Fan (2006), Chen *et al.* (2009) and Chen, Fan and Tsyrennikov (2006) for time series data) is also a widely-used estimation method and has a number of attractive features. Most importantly, with respect to SMM approach proposed here, it yields fully efficient estimates of the copula parameters, whereas SMM generally does not. Semi-parametric MLE requires, of course, the copula likelihood and for some more complicated models the likelihood can be cumbersome to derive or to compute, e.g. the "stochastic copula" model of Hafner and Manner (2012) or the high dimension factor copula model of Oh and Patton (2011). In such applications it may be desirable to avoid the likelihood and use a simpler SMM approach.

A long-standing estimator of the copula parameter is the method of moments (MM) estimator (see Genest (1987) and Genest and Rivest (1993) for *iid* data and Rémillard (2010) for time series data). This estimator exploits the known one-to-one mapping between the parameters of certain copulas and certain measures of dependence. For example, a Clayton copula with parameter κ implies Kendall's tau of $\kappa/(\kappa+2)$, yielding a simple MM estimator of the parameter of this copula as $\hat{\kappa}=2\hat{\tau}/(1-\hat{\tau})$. This estimator usually has the benefit of being very fast to compute. The SMM estimator proposed in this paper is a direct generalization of MM in two directions. Firstly, it allows the consideration of over-identified models: For some copulas we have more implied dependence measures than unknown parameters (e.g., for the Normal copula we have both Kendall's tau and Spearman's rank correlation in closed form). By treating this as a GMM estimation problem we can draw on the information in all available dependence measures. Secondly, we allow for dependence measures that are not known closed-form functions of the copula parameters. We use simulations to obtain the mapping, making this SMM rather than GMM. In the case that the mapping is known and the number of free parameters equals the number of dependence measures, our SMM approach simplifies to the well-known MM approach.

Other, less-widely used, estimation methods considered in the literature include minimum distance estimation, see Tsukahara (2005), and "expert judgment" estimation, see Britton, Fisher and Whitley (1998). This paper contributes to this literature by considering the properties of a SMM-type estimator, for both *iid* and time series data, nesting GMM and MM estimation of the copula parameter as special cases.

3.2 Simulation-based estimation of copula models

We consider the same class of data generating processes (DGPs) as Chen and Fan (2006), Chen, *et al.* (2009) and Rémillard (2010). This class allows each variable to have time-varying conditional mean and conditional variance, each governed by parametric models, with some unknown marginal distribution. As in those papers, and also earlier papers such as Genest and Rivest (1993) and Genest, Ghoudi and Rivest (1995), we estimate the marginal distributions using the empirical distribution function (EDF). The conditional copula of the data is assumed to belong to a parametric family with unknown parameter θ_0. The DGP we consider is:

$$\begin{aligned} [Y_{1t},\ldots,Y_{Nt}]' &\equiv \mathbf{Y}_t = \mu_t(\phi_0) + \sigma_t(\phi_0)\,\eta_t \qquad (3.1) \\ \text{where} \quad \mu_t(\phi) &\equiv [\mu_{1t}(\phi),\ldots,\mu_{Nt}(\phi)]' \\ \sigma_t(\phi) &\equiv diag\{\sigma_{1t}(\phi),\ldots,\sigma_{Nt}(\phi)\} \\ [\eta_{1t},\ldots,\eta_{Nt}]' &\equiv \eta_t \sim iid \;\; \mathbf{F}_\eta = \mathbf{C}(F_1,\ldots,F_N;\theta_0) \end{aligned}$$

where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and independent of η_t. $\mathcal{F}_{t-1}$ is the sigma-field containing information generated by $\{\mathbf{Y}_{t-1},\mathbf{Y}_{t-2},\ldots\}$. The $r \times 1$ vector of parameters governing the dynamics of the variables, ϕ_0, is assumed to be $\sqrt{T}$-consistently estimable, which holds under mild conditions for many commonly-used models for multivariate time series, such as ARMA models, GARCH models, stochastic volatility models, etc. If ϕ_0 is known, or if μ_t and σ_t are known constant, then the model becomes one for *iid* data. Our task is to estimate the $p \times 1$ vector of copula parameters, $\theta_0 \in \mathbf{\Theta}$, based on the (estimated) standardized residual $\{\hat{\eta}_t \equiv \sigma_t^{-1}(\hat{\phi})[\mathbf{Y}_t - \mu_t(\hat{\phi})\}_{t=1}^T$ and simulations from the copula model, $\mathbf{C}(\cdot;\theta)$.

3.2.1 Definition of the SMM estimator

We will consider simulation from some parametric multivariate distribution, $\mathbf{F}_x(\theta)$, with marginal distributions $G_i(\theta)$, and copula $\mathbf{C}(\theta)$. This allows us to consider cases where it is possible to simulate directly from the copula model $\mathbf{C}(\theta)$ (in which case the G_i are all $Unif(0,1)$) and also cases where the copula model is embedded in some joint distribution with unknown marginal distributions, such as the factor copula models of Oh and Patton (2011).

We use only "pure" dependence measures as moments since those are affected not by changes in the marginal distributions of simulated data ($\mathbf{X}$). For example, moments like means and variances, are functions of the marginal distributions (G_i) and contain no information on the copula. Measures like linear correlation contain information on the copula but are also affected by the marginal distributions. Dependence measures like Spearman's rank correlation and quantile dependence are purely functions of the copula and are unaffected by the marginal distributions, see Nelsen (2006) and Joe (1997) for example. Spearman's rank correlation, quantile dependence, and Kendall's tau for the pair (η_i, η_j) are defined as:

$$\rho^{ij} \equiv 12E\left[F_i(\eta_i) F_j(\eta_j)\right] - 3 = 12\int\int uv dC_{ij}(u,v) - 3 \tag{3.2}$$

$$\lambda_q^{ij} \equiv \begin{cases} P\left[F_i(\eta_i) \leqslant q | F_j(\eta_j) \leqslant q\right] = \frac{C_{ij}(q,q)}{q}, & q \in (0, 0.5] \\ P\left[F_i(\eta_i) > q | F_j(\eta_j) > q\right] = \frac{1-2q+C_{ij}(q,q)}{1-q}, & q \in (0.5, 1) \end{cases} \tag{3.3}$$

$$\tau^{ij} \equiv 4E\left[C_{ij}\left(F_i(\eta_i), F_j(\eta_j)\right)\right] - 1 \tag{3.4}$$

where C_{ij} is the copula of (η_i, η_j). The sample counterparts based on the estimated standardized residuals are defined as:

$$\hat{\rho}^{ij} \equiv \frac{12}{T}\sum_{t=1}^{T} \hat{F}_i(\hat{\eta}_{it}) \hat{F}_j(\hat{\eta}_{jt}) - 3 \tag{3.5}$$

$$\hat{\lambda}_q^{ij} \equiv \begin{cases} \frac{1}{Tq}\sum_{t=1}^{T} 1\{\hat{F}_i(\hat{\eta}_{it}) \leqslant q, \hat{F}_j(\hat{\eta}_{jt}) \leqslant q\}, & q \in (0, 0.5] \\ \frac{1}{T(1-q)}\sum_{t=1}^{T} 1\{\hat{F}_i(\hat{\eta}_{it}) > q, \hat{F}_j(\hat{\eta}_{jt}) > q\}, & q \in (0.5, 1) \end{cases} \tag{3.6}$$

$$\hat{\tau}^{ij} \equiv 4\frac{1}{T}\sum_{t=1}^{T} \hat{C}_{ij}\left(\hat{F}_i(\hat{\eta}_{it}), \hat{F}_j(\hat{\eta}_{jt})\right) - 1 \tag{3.7}$$

where $\hat{F}_i(y) \equiv (T+1)^{-1} \sum_{t=1}^{T} 1\{\hat{\eta}_{it} \leqslant y\}$, and $\hat{C}_{ij}(u,v) \equiv (T+1)^{-1} \sum_{t=1}^{T} 1\{\hat{F}_i(\hat{\eta}_{it}) \leqslant u, \hat{F}_j(\hat{\eta}_{jt}) \leqslant v\}$. We will denote the counterparts based on simulated data as $\tilde{\rho}^{ij}(\theta)$, $\tilde{\lambda}_q^{ij}(\theta)$ and $\tilde{\tau}^{ij}(\theta)$.

Let $\tilde{\mathbf{m}}_S(\theta)$ be a $(m \times 1)$ vector of dependence measures computed using S simulations from $\mathbf{F}_x(\theta)$, $\{\mathbf{X}_s\}_{s=1}^{S}$, and let $\hat{\mathbf{m}}_T$ be the corresponding vector of dependence measures computed using the the standardized residuals $\{\hat{\eta}_t\}_{t=1}^{T}$. These vectors can also contain linear combinations of dependence measures, a feature that is useful when considering estimation of high-dimension models. Define the difference between these as

$$\mathbf{g}_{T,S}(\theta) \equiv \hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\theta) \tag{3.8}$$

Our SMM estimator is based on searching across $\theta \in \boldsymbol{\Theta}$ to make this difference as small as possible. The estimator is defined as:

$$\hat{\theta}_{T,S} \equiv \underset{\theta \in \boldsymbol{\Theta}}{\arg\min}\ Q_{T,S}(\theta) \tag{3.9}$$

$$\text{where} \quad Q_{T,S}(\theta) \equiv \mathbf{g}'_{T,S}(\theta)\, \hat{\mathbf{W}}_T \mathbf{g}_{T,S}(\theta)$$

and $\hat{\mathbf{W}}_T$ is some positive definite weight matrix, which may depend on the data. As usual, for identification we require at least as many moment conditions as there are free parameters (i.e., $m \geqslant p$). In the subsections below we establish the consistency and asymptotic normality of this estimator, provide a consistent estimator of its asymptotic covariance matrix, and obtain a test based on over-identifying restrictions. Appendix C.2 presents details on the computation of the objective function.

3.2.2 Consistency of the SMM estimator

The estimation problem here differs in two important ways from standard GMM or M-estimation: Firstly, the objective function, $Q_{T,S}(\theta)$ is not continuous in θ since $\tilde{\mathbf{m}}_S(\theta)$ will be a number in a set of discrete values as θ varies on Θ, for example, $\left\{0, \frac{1}{Sq}, \frac{2}{Sq}, \ldots, \frac{S}{Sq}\right\}$ for a lower quantile dependence. This problem would vanish if, for the copula model being considered, we knew the mapping $\theta \longmapsto \mathbf{m}_0(\theta) \equiv \lim_{S\to\infty} \tilde{\mathbf{m}}_S(\theta)$ in closed form. The second difference is that a law of large numbers is not available to show the pointwise convergence of $\mathbf{g}_{T,S}(\theta)$, as the functions $\hat{\mathbf{m}}_T$ and $\tilde{\mathbf{m}}_S(\theta)$ both involve empirical distribution functions. We use recent developments in empirical process theory to overcome this difficulty.

We now list some assumptions that are required for our results to hold.

Assumption 1.

(i) The distributions $\mathbf{F}_\eta$ *and* $\mathbf{F}_x$ *are continuous.*

(ii) Every bivariate marginal copula C_{ij} *of* $\mathbf{C}$ *has continuous partial derivatives with respect to* u_i *and* u_j.

If the data $\mathbf{Y}_t$ are *iid*, e.g. if μ_t and σ_t are known constant in equation (3.1), or if ϕ_0 is known, then Assumption 1 is sufficient to prove Proposition 1 below, using the results of Fermanian, *et al.* (2004). If, however, estimated standardized residuals are used in the estimation of the copula then more assumptions are necessary in order to control the estimation error coming from the models for the conditional means and conditional variances. We combine assumptions A1–A6 in Rémillard (2010) in the following assumption. First, define $\gamma_{0t} = \sigma_t^{-1}\left(\hat{\phi}\right) \boldsymbol{\vartheta}_t\left(\hat{\phi}\right)$ and $\gamma_{1kt} = \sigma_t^{-1}\left(\hat{\phi}\right) \boldsymbol{\vartheta}_{kt}\left(\hat{\phi}\right)$ where $\boldsymbol{\vartheta}_t(\phi) = \frac{\partial \mu_t(\phi)}{\partial \phi'}, \boldsymbol{\vartheta}_{kt}(\phi) = \frac{\partial [\sigma_t(\phi)]_{\text{k-th column}}}{\partial \phi'}$, $k = 1, \ldots, N$. Define $\mathbf{d}_t$ as

$$\mathbf{d}_t = \eta_t - \hat{\eta}_t - \left(\gamma_{0t} + \sum_{k=1}^{N} \eta_{kt}\gamma_{1kt}\right)\left(\hat{\phi} - \phi_0\right)$$

where η_{kt} is k-th row of η_t and both γ_{0t} and γ_{1kt} are $\mathcal{F}_{t-1}$-measurable.

Assumption 2.

(i) $\frac{1}{T}\sum_{t=1}^{T}\gamma_{0t} \xrightarrow{p} \mathbf{\Gamma}_0$ *and* $\frac{1}{T}\sum_{t=1}^{T}\gamma_{1kt} \xrightarrow{p} \mathbf{\Gamma}_{1k}$ *where* $\mathbf{\Gamma}_0$ *and* $\mathbf{\Gamma}_{1k}$ *are deterministic for* $k = 1, \ldots, N$.

(ii) $\frac{1}{T}\sum_{t=1}^{T}E\left(\|\gamma_{0t}\|\right), \frac{1}{T}\sum_{t=1}^{T}E\left(\|\gamma_{0t}\|^2\right), \frac{1}{T}\sum_{t=1}^{T}E\left(\|\gamma_{1kt}\|\right)$, *and* $\frac{1}{T}\sum_{t=1}^{T}E\left(\|\gamma_{1kt}\|^2\right)$ *are bounded for* $k = 1, \ldots, N$.

(iii) There exists a sequence of positive terms $r_t > 0$ *so that* $\sum_{t\geqslant 1} r_t < \infty$ *and such that the sequence* $\max_{1\leqslant t\leqslant T}\|\mathbf{d}_t\|/r_t$ *is tight.*

(iv) $\max_{1\leqslant t\leqslant T}\|\gamma_{0t}\|/\sqrt{T} = o_p(1)$ *and* $\max_{1\leqslant t\leqslant T}\eta_{kt}\|\gamma_{1kt}\|/\sqrt{T} = o_p(1)$ *for* $k = 1, \ldots, N$.

(v) $\left(\alpha_T, \sqrt{T}\left(\hat{\phi} - \phi_0\right)\right)$ *weakly converges to a continuous Gaussian process in* $[0,1]^N \times \mathbb{R}^r$, *where* α_T *is the empirical copula process of uniform random variables:*

$$\alpha_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\prod_{k=1}^{N} 1\left(U_{kt} \leqslant u_k\right) - C\left(\mathbf{u}\right)\right\}$$

(vi) $\frac{\partial \mathbf{F}_\eta}{\partial \eta_k}$ *and* $\eta_k\frac{\partial \mathbf{F}_\eta}{\partial \eta_k}$ *are bounded and continuous on* $\bar{\mathbb{R}}^N = [-\infty, +\infty]^N$ *for* $k = 1, \ldots, N$.

With these two assumptions, sample rank correlation and quantile dependence converge in probability to their population counterparts, see Theorems 3 and 6 of Fermanian, Radulović and Wegkamp (2004) for the *iid* case, and combine with Corollary 1 of Rémillard (2010) for the time series case. (See Lemma 1 of Appendix C.1 for details.) When applied to simulated data this convergence holds pointwise for any θ. Thus $\mathbf{g}_{T,S}(\theta)$ converges in probability to the population moment functions defined as follows:

$$\mathbf{g}_{T,S}(\theta) \equiv \hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\theta) \xrightarrow{p} \mathbf{g}_0(\theta) \equiv \mathbf{m}_0(\theta_0) - \mathbf{m}_0(\theta), \text{ for } \forall\theta \in \mathbf{\Theta} \text{ as } T, S \to \infty \tag{3.10}$$

We define the population objective function as

$$Q_0(\theta) = \mathbf{g}_0(\theta)' \mathbf{W}_0 \mathbf{g}_0(\theta) \tag{3.11}$$

where $\mathbf{W}_0$ is the probability limit of $\hat{\mathbf{W}}_T$. The convergence of $\mathbf{g}_{T,S}(\theta)$ and $\hat{\mathbf{W}}_T$ implies that

$$Q_{T,S}(\theta) \xrightarrow{p} Q_0(\theta) \quad \text{for} \quad \forall \theta \in \mathbf{\Theta} \text{ as } T, S \to \infty$$

For consistency of our estimator we need, as usual, *uniform* convergence of $Q_{T,S}(\theta)$, but as this function is not continuous in θ and a law of large numbers is not available, the standard approach based on a uniform law of large numbers is not available. We instead use results on the stochastic equicontinuity of $\mathbf{g}_{T,S}(\theta)$, based on Andrews (1994) and Newey and McFadden (1994).

Assumption 3.

(i) $g_0(\theta) \neq 0$ *for* $\theta \neq \theta_0$

(ii) $\mathbf{\Theta}$ *is compact.*

(iii) Every bivariate marginal copula $C_{ij}(u_i, u_j; \theta)$ *of* $\mathbf{C}(\theta)$ *on* $(u_i, u_j) \in (0,1) \times (0,1)$ *is Lipschitz continuous on* $\mathbf{\Theta}$.

(iv) $\hat{\mathbf{W}}_T$ *is* $O_p(1)$ *and converges in probability to* $\mathbf{W}_0$, *a positive definite matrix.*

Proposition 1. *Suppose that Assumptions 1, 2 and 3 hold. Then* $\hat{\theta}_{T,S} \xrightarrow{p} \theta_0$ *as* $T, S \to \infty$

A sketch of all proofs is presented in Section 3.6, and detailed proofs are in Appendix C.1. Assumption 3(iii) is needed to prove the stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\theta)$, which is a sufficient condition for the stochastic equicontinuity of $\mathbf{g}_{T,S}(\theta)$, and can easily be shown to be satisfied for many bivariate parametric copulas. Assumption 3(ii) requires compactness of the parameter space, a common assumption, and is aided by having outside information (such as constraints from economic arguments) that allow the researcher to bound the set of plausible parameters. While Pakes and Pollard (1989) and McFadden (1989) show the consistency of SMM estimator for T, S diverging at the same rate, Proposition 1 shows that the copula parameter is consistent at any relative rate of T and S as long as both diverge. If we know the function $\mathbf{m}(\theta)$ in closed form, then GMM is feasible and is equivalent to our estimator with $S/T \rightarrow \infty$ as $T, S \rightarrow \infty$.

We focus on weak consistency of our estimator because it suffices for our asymptotic distribution theory, presented below. A corresponding strong consistency result, i.e., $\hat{\theta}_{T,S} \xrightarrow{a.s.} \theta_0$, may be obtained by drawing on recent work by Bouzebda and Zari (2011). The key is to show uniform strong convergence of the sample objective function, from which strong consistency of the estimator easily follows, see Newey and McFadden (1994) for example. Uniform strong consistency of the objective function can be shown by combining minor changes in the above assumptions (eg, $\hat{\mathbf{W}}_T$ must converge *a.s.* to $\mathbf{W}_0$) with pointwise strong convergence of the objective function, which can be obtained using the results of Bouzebda and Zari (2011).

3.2.3 Asymptotic normality of the SMM estimator

As $Q_{T,S}(\theta)$ is non-differentiable the standard approach based on a Taylor expansion is not available, however the asymptotic normality of our estimator can still be established with some further assumptions:

Assumption 4.

(i) θ_0 *is an interior point of* $\mathbf{\Theta}$

(ii) $\mathbf{g}_0(\theta)$ *is differentiable at* θ_0 *with derivative* $\mathbf{G}_0$ *such that* $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ *is nonsingular.*

(iii) $\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)'\hat{\mathbf{W}}_T\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) \leqslant \inf_{\theta\in\boldsymbol{\Theta}} \mathbf{g}_{T,S}(\theta)'\hat{\mathbf{W}}_T\mathbf{g}_{T,S}(\theta) + o_p(1/T + 1/S)$

The first assumption above is standard, and the third assumption is standard in simulation-based estimation problems, see Newey and McFadden (1994) for example. The rate at which the o_p term vanishes in part (iii) turns out to depend on the smaller of T or S, as $o_p(1/T + 1/S) = o_p\left(\min(T,S)^{-1}\right)$, as will become clear from the proposition below. The second assumption requires the population objective function, $\mathbf{g}_0$, to be differentiable even though its finite-sample counterpart is not, which is common in simulation-based estimation. The nonsingularity of $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ is sufficient for local identification of the parameters of this model at θ_0, see Hall (2005) and Rothenberg (1971). With these assumptions in hand we obtain the following result based on three different relative divergence rates of T and S.

Proposition 2. *Suppose that Assumptions 1, 2, 3 and 4 hold. Then*

$$\frac{1}{\sqrt{1/S + 1/T}}\left(\hat{\theta}_{T,S} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{\Omega}_0\right) \text{ as } T, S \to \infty \tag{3.12}$$

where $\boldsymbol{\Omega}_0 = \left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0\mathbf{W}_0\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}$, *and* $\boldsymbol{\Sigma}_0 \equiv avar\left[\hat{\mathbf{m}}_T\right]$.

The rate of convergence is thus shown to equal $\min(S,T)^{1/2}$. In general, one would like to set S very large to minimize the impact of simulation error and obtain a $\sqrt{T}$ convergence rate, however if the model is computationally costly to simulate, then the result for $S \ll T$ may be useful. When S and T diverge at different rates the asymptotic variance of $\min(S,T)^{1/2}\left(\hat{\theta}_{T,S} - \theta_0\right)$ is simply $\boldsymbol{\Omega}_0$. When S and T diverge at the same rate, say $S/T \to k \in (0,\infty)$, the asymptotic variance of $\sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right)$ is $(1 + 1/k)\boldsymbol{\Omega}_0$, which incorporates efficiency loss from simulation error. As usual we find that $\boldsymbol{\Omega}_0 = \left(\mathbf{G}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_0\right)^{-1}$ if $\mathbf{W}_0$ is the efficient weight matrix, $\boldsymbol{\Sigma}_0^{-1}$.

The proof of the above proposition uses recent results for empirical copula processes presented in Fermanian, Radulović and Wegkamp (2004) and Rémillard (2010) to establish the asymptotic normality of the sample dependence measures, $\hat{\mathbf{m}}_T$, and requires us to establish the stochastic equicontinuity of the moment functions, $\mathbf{v}_{T,S}(\theta) = \sqrt{T}\left[\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\right]$. These are shown in Lemmas 6 and 7 in Appendix C.1.

Chen and Fan (2006), Chen, *et al.* (2009) and Rémillard (2010) show that estimation error from $\hat{\phi}$ does not enter the asymptotic distribution of the copula parameter estimator for maximum likelihood or (analytical) moment-based estimators, and the above proposition shows that this surprising result also holds for the SMM-type estimators proposed here. In applications based on *parametric* models for the marginal distributions, the asymptotic covariance matrix of the copula parameter is more complicated. In such cases, the model is fully parametric and the estimation approach here is a form of two-stage GMM (or SMM). In the absence of simulations, this can be treated using existing methods, see White (1994) and Gourieroux, *et al.* (1996) for example. If simulations are used in the copula estimation step, then the lemmas presented in Appendix C.1 can be combined with existing results on two-stage GMM to obtain the limiting distribution. This is not difficult and requires some detailed notation, and so is not presented here.

3.2.4 Consistent estimation of the asymptotic variance

The asymptotic variance of our estimator has the familiar form of standard GMM applications, however the components $\mathbf{\Sigma}_0$ and $\mathbf{G}_0$ require more care in their estimation than in standard applications. We suggest using an *iid* bootstrap to estimate $\mathbf{\Sigma}_0$:

1. Sample with replacement from the standardized residuals $\{\hat{\eta}_t\}_{t=1}^T$ to obtain a bootstrap sample, $\left\{\hat{\eta}_t^{(b)}\right\}_{t=1}^T$. Repeat this step B times.

2. Using $\left\{\hat{\eta}_t^{(b)}\right\}_{t=1}^T$, $b = 1, ..., B$, compute the sample moments and denote as $\hat{\mathbf{m}}_T^{(b)}$,

$b = 1, ..., B$.

3. Calculate the sample covariance matrix of $\hat{\mathbf{m}}_T^{(b)}$ across the bootstrap replications, and scale it by the sample size:

$$\hat{\mathbf{\Sigma}}_{T,B} = \frac{T}{B} \sum_{b=1}^{B} \left(\hat{\mathbf{m}}_T^{(b)} - \hat{\mathbf{m}}_T \right) \left(\hat{\mathbf{m}}_T^{(b)} - \hat{\mathbf{m}}_T \right)' \tag{3.13}$$

For the estimation of $\mathbf{G}_0$, we suggest a numerical derivative of $\mathbf{g}_{T,S}(\theta)$ at $\hat{\theta}_{T,S}$, however the fact that $\mathbf{g}_{T,S}$ is non-differentiable means that care is needed in choosing the step size for the numerical derivative. In particular, Proposition 3 below shows that we need to let the step size go to zero, as usual, but *slower* than the inverse of the rate of convergence of the estimator (i.e., $1/\min\left(\sqrt{T}, \sqrt{S}\right)$). Let $\mathbf{e}_k$ denote the k-th unit vector whose dimension is the same as that of θ, and let $\varepsilon_{T,S}$ denote the step size. A two-sided numerical derivative estimator $\hat{\mathbf{G}}_{T,S}$ of $\mathbf{G}$ has k-th column

$$\hat{\mathbf{G}}_{T,S,k} = \frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S} + \mathbf{e}_k \varepsilon_{T,S}\right) - \mathbf{g}_{T,S}\left(\hat{\theta}_{T,S} - \mathbf{e}_k \varepsilon_{T,S}\right)}{2\varepsilon_{T,S}}, \; k = 1, 2, ..., p \tag{3.14}$$

Combine this estimator with $\hat{\mathbf{W}}_T$ to form:

$$\hat{\mathbf{\Omega}}_{T,S,B} = \left(\hat{\mathbf{G}}'_{T,S} \hat{\mathbf{W}}_T \hat{\mathbf{G}}_{T,S} \right)^{-1} \hat{\mathbf{G}}'_{T,S} \hat{\mathbf{W}}_T \hat{\mathbf{\Sigma}}_{T,B} \hat{\mathbf{W}}_T \hat{\mathbf{G}}_{T,S} \left(\hat{\mathbf{G}}'_{T,S} \hat{\mathbf{W}}_T \hat{\mathbf{G}}_{T,S} \right)^{-1} \tag{3.15}$$

Proposition 3. *Suppose that all assumptions of Proposition 2 are satisfied, and that* $\varepsilon_{T,S} \to 0$, $\varepsilon_{T,S} \times \min\left(\sqrt{T}, \sqrt{S}\right) \to \infty$, $B \to \infty$ *as* T, $S \to \infty$. *Then* $\hat{\mathbf{\Sigma}}_{T,B} \xrightarrow{p} \mathbf{\Sigma}_0$, $\hat{\mathbf{G}}_{T,S} \xrightarrow{p} \mathbf{G}_0$ *and* $\hat{\mathbf{\Omega}}_{T,S,B} \xrightarrow{p} \mathbf{\Omega}_0$ *as* $T, S \to \infty$.

3.2.5 A test of overidentifying restrictions

If the number of moments used in estimation is greater than the number of copula parameters, then it is possible to conduct a simple test of the over-identifying restrictions. When the efficient weight matrix is used in estimation, the asymptotic distribution of this test statistic is the usual chi-squared, however the method of proof is different as we again need to deal with the lack of differentiability of the objective function. We also consider the distribution of this test statistic for general weight matrices, leading to a non-standard limiting distribution.

Proposition 4. *Suppose that all assumptions of Proposition 2 are satisfied and that the number of moments* (m) *is greater than the number of copula parameters* (p). *Then*

$$J_{T,S} \equiv \min(T,S)\, \mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)' \hat{\mathbf{W}}_T \mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) \xrightarrow{d} \mathbf{u}'\mathbf{A}_0'\mathbf{A}_0\mathbf{u} \text{ as } T,S \to \infty$$

where $\mathbf{u} \sim N(0, \mathbf{I})$

and $\mathbf{A}_0 \equiv \mathbf{W}_0^{1/2}\mathbf{\Sigma}_0^{1/2}\mathbf{R}_0,\ \mathbf{R}_0 \equiv \mathbf{I} - \mathbf{\Sigma}_0^{-1/2}\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\mathbf{\Sigma}_0^{1/2}$. *If* $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$, *then* $J_{T,S} \xrightarrow{d} \chi^2_{m-p}$ *as usual.*

As in standard applications, the above test statistic has a chi-squared limiting distribution if the efficient weight matrix ($\hat{\mathbf{\Sigma}}_{T,B}^{-1}$) is used. When any other weight matrix is used, the test statistic has a sample-specific limiting distribution, and critical values in such cases can be obtained via a simple simulation:

1. Compute $\hat{\mathbf{R}}$ using $\hat{\mathbf{G}}_{T,S}$, $\hat{\mathbf{W}}_T$, and $\hat{\mathbf{\Sigma}}_{T,B}$.
2. Simulate $\mathbf{u}^{(k)} \sim iid\ N(0, \mathbf{I})$, for $k = 1, 2, ..., K$, where K is large.
3. For each simulation, compute $J_{T,S}^{(k)} = \mathbf{u}^{(k)\prime}\hat{\mathbf{R}}'\hat{\mathbf{\Sigma}}_{T,B}^{1/2\prime}\hat{\mathbf{W}}_T\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\hat{\mathbf{R}}\mathbf{u}^{(k)}$
4. The sample $(1-\alpha)$ quantile of $\left\{J_{T,S}^{(k)}\right\}_{k=1}^{K}$ is the critical value for this test statistic.

The need for simulations to obtain critical values from the limiting distribution is non-standard but is not uncommon; this arises in many other testing problems, see Wolak (1989), White (2000) and Andrews (2001) for examples. Given that $\mathbf{u}^{(k)}$ is a simple standard Normal, and that no optimization is required in this simulation, and that the matrix $\hat{\mathbf{R}}$ need only be computed once, obtaining critical values for this test is simple and fast.

3.2.6 SMM under model mis-specification

All of the above results hold under the assumption that the copula model is correctly specified. In the event that the specification test proposed in the previous section rejects a model as mis-specified, one is led directly to the question of whether these results, or extensions of them, hold for mis-specified models.

In the literature on GMM, there are two common ways to define mis-specification. Newey (1985) defines a form of "local" mis-specification (where the degree of mis-specification vanishes in the limit), and in that case it is simple to show that the asymptotic behavior of the SMM estimator does not change at all except the mean of limit distribution. Hall and Inoue (2003) consider "non-local" mis-specification. Formally, a model is said to be mis-specified if there is no value of $\theta \in \mathbf{\Theta}$ which satisfies $\mathbf{g}_0(\theta) = \mathbf{0}$. As Hall and Inoue (2003) note, mis-specification is only a concern when the model is over-identified, and so in this section we assume $m > p$. The absence of a parameter that satisfies the population moment conditions means that we must instead consider a "pseudo-true" parameter:

Definition 1. *The pseudo-true parameter is* $\theta_*(W_0) \equiv \arg\min_{\theta\in\mathbf{\Theta}} \mathbf{g}_0'(\theta)\, \mathbf{W}_0 \mathbf{g}_0(\theta)$.

While the true parameter, θ_0, when it exists, is determined only by the population moment condition $\mathbf{g}_0(\theta_0) = \mathbf{0}$, the pseudo-true parameter depends on the moment condition and also on the weight matrix $\mathbf{W}_0$, and thus we denote it as $\theta_*(W_0)$. With the additional assumptions below, the consistency of the SMM estimator under mis-specification can be proven.

Assumption 5. *(i) (Non-local mis-specification)* $\|\mathbf{g}_0(\theta)\| > 0$ *for all* $\theta \in \mathbf{\Theta}$

(ii) (Identification) There exists $\theta_*(\mathbf{W}_0) \in \mathbf{\Theta}$ *such that*
$\mathbf{g}_0(\theta_*(\mathbf{W}_0))' \mathbf{W}_0 \mathbf{g}_0(\theta_*(\mathbf{W}_0)) < \mathbf{g}_0(\theta)' \mathbf{W}_0 \mathbf{g}_0(\theta)$ *for all* $\theta \in \mathbf{\Theta} \backslash \{\theta_*(\mathbf{W}_0)\}$

Proposition 5. *Suppose Assumption 1, 2, 3(ii)-3(iv), and 5 holds. Then* $\hat{\theta}_{T,S} \xrightarrow{p} \theta_*(\mathbf{W}_0)$ *as* $T, S \to \infty$

The above proposition shows that, under mis-specification, the SMM estimator $\hat{\theta}_{T,S}$ converges in probability to the pseudo true parameter $\theta_*(\mathbf{W}_0)$ rather than the true parameter θ_0. This extends existing results for GMM under mis-specification in Hall (2000) and Hall and Inoue (2003), as it is established even under the discontinuity of the moment functions.

While consistency of $\hat{\theta}_{T,S}$ under mis-specification is easily obtained, establishing the limit distribution of $\hat{\theta}_{T,S}$ is not straightforward. A key contribution of Hall and Inoue (2003) was to show that the limit distribution of GMM (with smooth, differentiable moment functions) depends on the limit distribution of the weight matrix, not merely the probability limit of the weight matrix. In SMM applications, it is possible to show that the limit distribution will additionally depend on the limit distribution of the numerical derivative matrix, denoted $\hat{\mathbf{G}}_{T,S}$ above. Some results on the statistical properties of numerical derivatives are presented in Hong, *et al.* (2010), but this remains a relatively unexplored topic. In addition to incorporating the dependence on the distribution of $\hat{\mathbf{G}}_{T,S}$, under mis-specification one needs an alternative approach to establish the stochastic equicontinuity of the objective function, which is required for a Taylor series expansion of the population objective function to be used to obtain the limit distribution of the estimator. We leave the interesting problem of the limit distribution of $\hat{\theta}_{T,S}$ under mis-specification for future research.

3.3 Simulation study

In this section we present a study of the finite sample properties of the simulation-based (SMM) estimator studied in the previous section. We consider two widely-known copula models, the Clayton and the Gaussian (or Normal) copulas, see Nelson (2006) for discussion, and the "factor copula" proposed in Oh and Patton (2011), outlined below. A closed-form likelihood is available for the first two copulas, while the third copula requires a numerical integration step to obtain the likelihood (details on this are presented in Appendix C.3). In all cases we contrast the finite-sample properties of the MLE with the SMM estimator. The first two copulas also have closed-form cumulative distribution functions, and so quantile dependence (defined in equation (3.3) is also known in closed form. For the Clayton copula we have Kendall's tau in closed form ($\tau = \kappa/\left(2+\kappa\right)$) but not Spearman's rank correlation, see Nelsen (2006). For the Normal copula we have both Spearman's rank correlation in closed form ($\rho_S = 6/\pi \arcsin\left(\rho/2\right)$) and Kendall's rank correlation ($\tau = 2/\pi \arcsin\left(\rho\right)$) , see Nelsen (2006) and Demarta and McNeil (2005). This allows us to also compare GMM with SMM for these copulas, to quantify the loss in accuracy from having to resort to simulations.

The factor copula we consider is based on the following structure:

$$\begin{aligned} &\text{Let} \quad X_i = Z + \varepsilon_i, \quad i = 1, 2, ..., N \\ &\text{where} \;\; Z \sim Skew\ t\left(0, \sigma^2, \nu^{-1}, \lambda\right), \;\; \varepsilon_i \sim iid\ t\left(\nu^{-1}\right), \;\; \text{and} \;\; \varepsilon_i \perp\!\!\!\perp Z \;\forall\, i \qquad (3.16) \\ &[X_1, ..., X_N]' \equiv \mathbf{X} \sim \mathbf{F}_x = \mathbf{C}\left(G_x, ..., G_x\right) \end{aligned}$$

where we use the skewed t distribution of Hansen (1994). We use the copula of $\mathbf{X}$ implied by the above structure as our "factor copula" model, and it is parameterized by $\left(\sigma^2, \nu^{-1}, \lambda\right)$. For the factor copula we have neither the likelihood nor any of the above dependence measures in closed form, and so simulation-based methods are required. For the simulation we set the parameters to generate rank correlation of around 1/2, and so set the Clayton copula parameter to 1, the Gaussian copula parameter to 1/2, and the factor copula parameters to $\sigma^2 = 1$, $\nu^{-1} = 1/4$ and $\lambda = -1/2$.

We consider two different scenarios for the marginal distributions of the variables of interest. In the first case we assume that the data are *iid* with standard Normal marginal distributions, meaning that the only parameters that need to be estimated are those of the copula. This simplified case is contrasted with a second scenario where the marginal distributions of the variables are assumed to follow an AR(1)-GARCH(1,1) process, which is widely-used in time series applications:

$$\begin{aligned} Y_{it} &= \phi_0 + \phi_1 Y_{i,t-1} + \sigma_{it}\eta_{it}, \quad t = 1, 2, ..., T \\ \sigma_{it}^2 &= \omega + \beta\sigma_{i,t-1}^2 + \alpha\sigma_{i,t-1}^2\eta_{i,t-1}^2 \\ \eta_t &\equiv [\eta_{1t}, ..., \eta_{Nt}] \sim iid \ \ \mathbf{F}_\eta = \mathbf{C}(\Phi, \Phi, ..., \Phi) \end{aligned} \tag{3.17}$$

where Φ is the standard Normal distribution function and $\mathbf{C}$ can be Clayton, Gaussian, or the factor copula implied by equation (3.16). We set the parameters of the marginal distributions as $[\phi_0, \phi_1, \omega, \beta, \alpha] = [0.01, 0.05, 0.05, 0.85, 0.10]$, which broadly matches the values of these parameters when estimated using daily equity return data. In this scenario the parameters of the models for the conditional mean and variance are estimated, and then the estimated standardized residuals are obtained:

$$\hat{\eta}_{it} = \frac{Y_{it} - \hat{\phi}_0 - \hat{\phi}_1 Y_{i,t-1}}{\hat{\sigma}_{it}}. \tag{3.18}$$

These residuals are used in a second stage to estimate the copula parameters. In all cases we consider a time series of length $T = 1,000$, corresponding to approximately 4 years of daily return data, and we use $S = 25 \times T$ simulations in the computation of the dependence measures to be matched in the SMM optimization. We use five dependence measures in estimation: Spearman's rank correlation, and the

$0.05, 0.10, 0.90, 0.95$ quantile dependence measures, averaged across pairs of assets. We repeat each scenario 100 times, and in the results below we use the identity weight matrix for estimation. (Corresponding results based on the efficient weight matrix are comparable, and available in Appendix C.4.) We also report the computation times (per simulation) for each estimation.

Table 3.1 reveals that for all three dimensions ($N = 2,\ 3$ and 10) and for all three copula models the estimated parameters are centered on the true values, with the average estimated bias being small relative to the standard deviation, and with the median of the simulated distribution centered on the true values. Looking across the dimension size, we see that the copula model parameters are almost always more precisely estimated as the dimension grows. This is intuitive, given the exchangeable nature of all three models.

Comparing the SMM estimator with the ML estimator, we see that the SMM estimators suffer a loss in efficiency of around 50% for $N = 2$ to around 20% for $N = 10$. The loss is greatest for the ν^{-1} parameter of the factor copula, and moderate and similar for the remaining parameters. Some loss is of course expected, and this simulation indicates that the loss is moderate overall. Comparing the SMM estimator to the GMM estimator provides us with a measure of the loss in accuracy from having to estimate the population moment function via simulation. We find that this loss ranges from zero to 3%, and thus little is lost from using SMM rather than GMM. The simulation results in Table 3.2, where the copula parameters are estimated after the estimation of AR(1)-GARCH(1,1) models for the marginal distributions in a separate first stage, are very similar to the case when no marginal distribution parameters are required to be estimated, consistent with Proposition 2. Thus that somewhat surprising asymptotic result is also relevant in finite samples.

In Table 3.3 we present the finite-sample coverage probabilities of 95% confidence intervals based on the asymptotic normality result from Proposition 2 and the asymptotic covariance matrix estimator presented in Proposition 3. As shown in that proposition, a critical input to the asymptotic covariance matrix estimator is the step size used in computing the numerical derivative matrix $\hat{\mathbf{G}}_{T,S}$. This step size, $\varepsilon_{T,S}$, must go to zero, but at a slower rate than $1/\sqrt{T}$. Ignoring constants, our simulation sample size of $T = 1,000$ suggests setting $\varepsilon_{T,S} > 0.001$, which is much larger than standard step sizes used in computing numerical derivatives. (For example, the default in many functions in MATLAB is a step size of around 6×10^{-6}, which is an optimal choice in certain applications, see Judd (1998) for example.) We study the impact of the choice of step size by considering a range of values from 0.0001 to 0.1. Table 3.3 shows that when the step size is set to 0.01 or 0.1 the finite-sample coverage rates are close to their nominal levels. However if the step size is chosen too small (0.001 or smaller) then the coverage rates are much lower than nominal levels. For example, setting $\varepsilon_{T,S} = 0.0001$ (which is still 16 times larger than the default setting in MATLAB) we find coverage rates as low as 2% for a nominal 95% confidence interval. Thus this table shows that the asymptotic theory provides a reliable means for obtaining confidence intervals, so long as care is taken not to set the step size too small.

Table 3.3 also presents the results of a study of the rejection rates for the test of over-identifying restrictions presented in Proposition 4. Given that we consider $W = I$ in this table, the test statistic has a non-standard distribution, and we use $K = 10,000$ simulations to obtain critical values. In this case, the limiting distribution also depends on $\hat{\mathbf{G}}_{T,S}$, and we again compute $\hat{\mathbf{G}}_{T,S}$ using a step size of $\varepsilon_{T,S} = 0.1, 0.01, 0.001$ and 0.0001. The rejection rates are close to their nominal levels 95% for the all three copula models.

We finally consider the properties of the estimator under model mis-specification. In Table 3.4 we consider two scenarios: one where the true copula is Clayton but the model is Normal, and one where the true copula is Normal but the model is

Clayton. The pseudo-true parameters for these two scenarios are not known in closed form, and we use a simulation of length 10 million to estimate it. They vary across the dimension of the problem, and we report them in the top row of each panel of Table 3.4. The remainder of Table 3.4 has the same structure as Tables 3.1 and 3.2. Similar to those tables, in this mis-specified case we see that the estimated parameters are centered on the pseudo-true values, with the average estimated bias being small relative to the standard deviation. Looking across the dimension size, we see that the copula model parameters are almost always more precisely estimated as the dimension grows. These mis-specified scenarios also provide some insight into the power of the specification test based on over-identifying restrictions. We find that for all three dimensions and for both *iid* and AR-GARCH data, the J-test rejected the null of correct specification across all 100 simulations, indicating this test has power to detect model mis-specification.

These simulation results provide support for the proposed estimation method: for empirically realistic parameter values and sample size, the estimator is approximately unbiased, with estimated confidence intervals that have coverage close to their nominal level when the step size for the numerical derivative is chosen in line with our theoretical results, and the test for model mis-specification has finite-sample rejection frequencies that are close to their nominal levels when the model is correctly specified, and has good power to reject mis-specified models.

3.4 Application to the dependence between financial firms

This section considers models for the dependence between seven large financial firms. We use daily stock return data over the period January 2001 to December 2010, a total of $T = 2515$ trade days, on Bank of America, Bank of New York, Citigroup, Goldman Sachs, J.P. Morgan, Morgan Stanley and Wells Fargo. Summary statistics for these returns are presented in Table C.4 of Appendix C.4, and indicate that all series are positively skewed and leptokurtotic, with kurtosis ranging from 16.0 (J.P. Morgan) to 119.8 (Morgan Stanley).

To capture the impact of time-varying conditional means and variances in each of these series, we estimate the following autoregressive, conditionally heteroskedastic models:

$$
\begin{aligned}
r_{it} &= \phi_{0i} + \phi_{1i} r_{i,t-1} + \phi_{2i} r_{m,t-1} + \varepsilon_{it}, \quad \varepsilon_{it} = \sigma_{it}\eta_{it} \\
\text{where} \quad \sigma_{it}^2 &= \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_{1i}\varepsilon_{i,t-1}^2 + \gamma_{1i}\varepsilon_{i,t-1}^2 \cdot \mathbf{1}_{[\varepsilon_{i,t-1} \leqslant 0]} \\
&\quad + \alpha_{2i}\varepsilon_{m,t-1}^2 + \gamma_{2i}\varepsilon_{m,t-1}^2 \cdot \mathbf{1}_{[\varepsilon_{m,t-1} \leqslant 0]}
\end{aligned} \tag{3.19}
$$

where r_{it} is the return on one of these seven firms and r_{mt} is the return on the S&P 500 index. We include the lagged market index return in both the mean and variance specifications to capture any influence of lagged information in the model for a given stock, and in the model for the market index itself we set $\phi_1 = \alpha_1 = \gamma_1 = 0$. Estimated parameters from these models are presented in Table C.5 of Appendix C.4, and are consistent with the values found in the empirical finance literature, see Bollerslev, Engle and Nelson (1994) for example. From these models we obtain the estimated standardized residuals, $\hat{\eta}_{it}$, which are used in the estimation of the dependence structure.

In Table 3.5 we present measures of dependence between these seven firms. The upper panel reveals that rank correlation between their standardized residuals is 0.63 on average, and ranges from 0.55 to 0.76. The lower panel of Table 3.5 presents measures of dependence in the tails between these series. The upper triangle presents the average of the 1% and 99% quantile dependence measures presented in equation (3.6), and we see substantial dependence here, with values ranging between 0.16 and 0.40. The lower triangle presents the difference between the 90% and 10% quantile dependence measures, as a gauge of the degree of asymmetry in the dependence structure. These differences are mostly negative (14 out of 21), indicating greater dependence during crashes than during booms.

Table 3.6 presents the estimation results for three different copula models of these series. The first model is the well-known Clayton copula, the second is the Normal copula and the third is a "factor copula" as proposed by Oh and Patton (2011). The first copula allows for lower tail dependence, but imposes that upper tail dependence is zero. The second copula implies zero tail dependence in both directions. The third copula allows for tail dependence in both tails, and allows the degree of dependence to differ across positive and negative realizations.

For all three copulas we implent the SMM estimator proposed in Section 3.2, with the identity weight matrix and the efficient weight matrix, using five dependence measures: Spearman's rank correlation, and the $0.05, 0.10, 0.90, 0.95$ quantile dependence measures, averaged across pairs of assets. We also implement the MLE for comparison. The value of the SMM objective function at the estimated parameters is presented for each model, along with the p-value from a test of the over-identifying restrictions based on Proposition 4. We use Proposition 3 to compute the standard errors, with $B = 1,000$ bootstraps used to estimate $\Sigma_{T,S}$, and $\varepsilon_{T,S} = 0.1$ used as the step size to compute $\hat{\mathbf{G}}_{T,S}$.

The parameter estimates for the Normal and factor copula models are similar for ML and SMM, while they are quite different for the Clayton copula. This may be explained by the results of the test of over-identifying restrictions: the Clayton copula is strongly rejected (with a p-value of less than 0.001 for both choices of weight matrix), while the Normal is less strongly rejected (p-values of 0.043 and 0.001). The factor copula is not rejected using this test for either choice of weight matrix. The improvement in fit from the factor copula appears to come from its ability to capture tail dependence: the parameter that governs tail dependence (ν^{-1}) is significantly greater than zero, while the parameter that governs asymmetric dependence (λ) is not significantly different from zero.

Given that our sample period spands the financial crisis, one may wonder whether the copula is constant throughout the period. To investigate this, we implement the copula structrual break test proposed by Rémillard (2010). This test uses a Kolmogorov-Smirnov type test statistic to compare the empirical copula before and after a given point in the sample, and then searches across all dates in the sample. We implement this test using 1000 simulations for the "multiplier" method, and find a p-value of 0.001, indicating strong evidence of a change in the copula over this period. Running this test on just the last two years of our sample period (January 2009 to December 2010) results in a p-value of 0.191, indicating no evidence of a change in the copula over this sub-period. We re-estimate our three copula models using data from this sub-period, and present the results in the lower panel of Table 3.6. The estimated parameters all indicate a (slight) increase in dependence in this sub-sample relative to the full sample. Perhaps the largest change is in the ν parameter of the factor copula, which goes from around 8.8 (across the three estimation methods) to around 4.4, indicating a substantial increase in the degree of tail dependence between these firms. The results of the specification tests over this sub-sample are comparable to the full sample results: the Clayton copula is strongly rejected, the Normal copula is rejected but less strongly, and the factor copula is not rejected, using either weight matrix.

Figure 3.1 sheds some further light on the relative performance of these copula models, over the full sample. This figure compares the empirical quantile dependence function with those implied by the three copula models. An *iid* bootstrap with $B = 1,000$ replications is used to construct pointwise confidence intervals for the sample quantile dependence estimates. We see here that the Clayton copula is "too asymmetric" relative to the data, while both the Normal and the factor copula models appear to provide a reasonable fit.

3.5 Conclusion

This paper presents the asymptotic properties of a new simulation-based estimator of the parameters of a copula model, which matches measures of rank dependence implied by the model to those observed in the data. The estimation method shares features with the simulated method of moments (SMM), see McFadden (1989) and Newey and McFadden (1994), for example, however the use of rank dependence measures as "moments" means that existing results on SMM cannot be used. We extend well-known results on SMM estimators using recent work in empirical process theory for copula estimation, see Fermanian, Radulović and Wegkamp (2004), Chen and Fan (2006) and Rémillard (2010), to show the consistency and asymptotic normality of SMM-type estimators of copula models. To the best of our knowledge, simulation-based estimation of copula models has not previously been considered in the literature. We also provide a method for obtaining a consistent estimate of the asymptotic covariance matrix, and a test of the over-identifying restrictions. Our results apply to both *iid* and time series data, and an extensive simulation study verifies that the asymptotic results provide a good approximation in finite samples. We illustrate the results with an application to a model of the dependence between the equity returns on seven financial firms, and find evidence of statistically significant tail dependence, and some evidence that the dependence between these assets is stronger in crashes than booms.

3.6 Sketch of proofs

Detailed proofs are available in Appendix C.1.

Proof of Proposition 1. First note that: (a) $Q_0(\theta)$ is uniquely minimized at θ_0 by Assumption 3(i) and positive definite $\mathbf{W}_0$ of Assumption 3(iv), (b) $\mathbf{\Theta}$ is compact by Assumption 3(ii); (c) $Q_0(\theta)$ consists of linear combinations of rank correlations and quantile dependence measures that are functions of pair-wise copula functions, so $Q_0(\theta)$ is continuous by Assumption 3(iii). The main part of the proof requires establishing that $Q_{T,S}$ uniformly converges in probability to Q_0, which we show using five lemmas in Appendix C.1: Pointwise convergence of $\mathbf{g}_{T,S}(\theta)$ to $\mathbf{g}_0(\theta)$ and stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\theta)$ is shown using results from Fermanian, Wegkamp and Radulović (2004) and Rémillard (2010), combined with Assumption 3(iii). This is sufficient for the stochastic equicontinuity of $\mathbf{g}_{T,S}$ and for the uniform convergence in probability of $\mathbf{g}_{T,S}$ to $\mathbf{g}_0$ by Lemmas 2.8 and 2.9 of Newey and McFadden (1994). Using the triangle and Cauchy-Schwarz inequalities this implies that $Q_{T,S}$ uniformly converges in probability to Q_0. We have thus verified that the conditions of Theorem 2.1 of Newey and McFadden (1994) hold, and we have $\hat{\theta} \xrightarrow{p} \theta_0$ as claimed. □

Proof of Proposition 2. We prove this proposition by verifying the five conditions of Theorem 7.2 of Newey and McFadden (1994) for our problem: (i) $\mathbf{g}_0(\theta_0) = 0$ by construction of $\mathbf{g}_0(\theta) = \mathbf{m}(\theta_0) - \mathbf{m}(\theta)$. (ii) $\mathbf{g}_0(\theta)$ is differentiable at θ_0 with derivative $\mathbf{G}_0$ such that $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ is nonsingular by Assumption 4(ii). (iii) θ_0 is an interior point of $\mathbf{\Theta}$ by Assumption 4(i). (iv) This part requires showing the asymptotic normality of $\sqrt{T}\mathbf{g}_{T,S}(\theta_0)$. We will present the result only for $S/T \to k \in (0,\infty)$. The results for the cases that $S/T \to 0$ or $S/T \to \infty$ are similar. In Lemma 6 of Appendix C.1, we show that $\sqrt{T}(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)) \xrightarrow{d} N(0, \mathbf{\Sigma}_0)$ as $T \to \infty$ and $\sqrt{S}(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)) \xrightarrow{d} N(0, \mathbf{\Sigma}_0)$ as $S \to \infty$ using Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004) and Corollary 1, Proposition 2 and

Proposition 4 of Rémillard (2010). This implies that

$$\sqrt{T}\mathbf{g}_{T,S}(\theta_0) = \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\mathbf{\Sigma}_0)} - \underbrace{\sqrt{\frac{T}{S}}}_{\to 1/\sqrt{k}} \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\mathbf{\Sigma}_0)}$$

and so $\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(\mathbf{0}, (1+1/k)\mathbf{\Sigma}_0)$ as $T, S \to \infty$. (v) This part requires showing that $\sup_{\|\theta-\theta_0\|<\delta} \sqrt{T}\left\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\right\| / \left[1 + \sqrt{T}\|\theta - \theta_0\|\right] \xrightarrow{p} 0$. The main part of this proof involves showing the stochastic equicontinuity of $\mathbf{v}_{T,S}(\theta) = \sqrt{T}\left[\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_0(\theta)\right]$. This is shown in Lemma 7 of Appendix C.1 by showing that $\{\mathbf{g}_{\cdot,\cdot}(\theta) : \theta \in \Theta\}$ is a type II class of functions in Andrews (1994), and then using that paper's Theorem 1. □

Proof of Proposition 3. If μ_t and σ_t are known constant, or if ϕ_0 is known, then the consistency of $\hat{\mathbf{\Sigma}}_{T,B}$ follows from Theorems 5 and 6 of Fermanian, Radulović and Wegkamp (2004). When ϕ_0 is estimated, the result is obtained by combining the results in Fermanian, *et al.* with those of Rémillard (2010): For simplicity, assume that only one dependence measure is used. Let $\hat{\rho}_{ij}$ and $\hat{\rho}_{ij}^{(b)}$ be the sample rank correlations constructed from the standardized residuals $\left\{\hat{\eta}_t^i, \hat{\eta}_t^j\right\}_{t=1}^T$ and from the bootstrap counterpart $\left\{\hat{\eta}_t^{(b)i}, \hat{\eta}_t^{(b)j}\right\}_{t=1}^T$. Also, define the corresponding estimates, $\ddot{\rho}_{ij}$ and $\ddot{\rho}_{ij}^{(b)}$, using the true innovations $\left\{\eta_t^i, \eta_t^j\right\}_{t=1}^T$ and the bootstrapped true innovations $\left\{\eta_t^{(b)i}, \eta_t^{(b)j}\right\}_{t=1}^T$ (where the same bootstrap time indices are used for both $\left\{\hat{\eta}_t^{(b)i}, \hat{\eta}_t^{(b)j}\right\}_{t=1}^T$ and $\left\{\eta_t^{(b)i}, \eta_t^{(b)j}\right\}_{t=1}^T$). Define true ρ as ρ_0. Theorem 5 of Fermanian, Radulović and Wegkamp (2004) shows that

$$\sqrt{T}\left(\ddot{\rho}_{ij} - \rho_0\right) = \sqrt{T}\left(\ddot{\rho}_{ij}^{(b)} - \ddot{\rho}_{ij}\right) + o_p(1)$$

Corollary 1 and Proposition 4 of Rémillard (2010) shows, under Assumption 2, that

$$\sqrt{T}\left(\hat{\rho}_{ij} - \ddot{\rho}_{ij}\right) = o_p\left(1\right) \quad \text{and} \quad \sqrt{T}\left(\hat{\rho}_{ij}^{(b)} - \ddot{\rho}_{ij}^{(b)}\right) = o_p\left(1\right)$$

Combining those three equations, we obtain

$$\sqrt{T}\left(\hat{\rho}_{ij} - \rho_0\right) = \sqrt{T}\left(\hat{\rho}_{ij}^{(b)} - \hat{\rho}_{ij}\right) + o_p\left(1\right), \text{ as } T, B \to \infty$$

and so equation (3.13) is a consistent estimator of $\mathbf{\Sigma}_0$. Consistency of the numerical derivatives $\hat{\mathbf{G}}_{T,S}$ can be established using a similar approach to Theorem 7.4 of Newey and McFadden (1994), and since $\hat{\mathbf{W}}_T \xrightarrow{p} \mathbf{W}_0$ by Assumption 3(iv), we thus have $\hat{\mathbf{\Omega}}_{T,S,B} \xrightarrow{p} \mathbf{\Omega}_0$. □

Proof of Proposition 4. We consider only the case where $S/T \to \infty$ or $S/T \to k > 0$. The case for $k = 0$ is analogous. A Taylor expansion of $\mathbf{g}_0\left(\hat{\theta}_{T,S}\right)$ around θ_0 yields

$$\sqrt{T}\mathbf{g}_0\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_0\left(\theta_0\right) + \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + o\left(\sqrt{T}\left\|\hat{\theta}_{T,S} - \theta_0\right\|\right)$$

and since $\mathbf{g}_0\left(\theta_0\right) = 0$ and $\sqrt{T}\left\|\hat{\theta}_{T,S} - \theta_0\right\| = O_p\left(1\right)$

$$\sqrt{T}\mathbf{g}_0\left(\hat{\theta}_{T,S}\right) = \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + o_p\left(1\right) \tag{3.20}$$

Then consider the following expansion of $\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)$ around θ_0

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}\left(\theta_0\right) + \hat{\mathbf{G}}_{T,S} \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + \mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) \tag{3.21}$$

where the remaining term is captured by $\mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right)$. Combining equations (3.20) and (3.21) we obtain

$$\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) - \mathbf{g}_{T,S}\left(\theta_0\right) - \mathbf{g}_0\left(\hat{\theta}_{T,S}\right)\right] = \left(\hat{\mathbf{G}}_{T,S} - \mathbf{G}_0\right) \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + \mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) + o_p\left(1\right)$$

The stochastic equicontinuity of $\mathbf{v}_{T,S}(\theta) = \sqrt{T}\left[\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\right]$ is established in the proof of Proposition 2, which implies (see proof of Proposition 2) that

$$\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) - \mathbf{g}_{T,S}(\theta_0) - \mathbf{g}_0\left(\hat{\theta}_{T,S}\right)\right] = o_p(1)$$

By Proposition 3, $\hat{\mathbf{G}}_{T,S} - \mathbf{G}_0 = o_p(1)$, which implies $\mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) = o_p(1)$. Thus, we obtain the expansion of $\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)$ around θ_0 :

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}(\theta_0) + \hat{\mathbf{G}}_{T,S} \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + o_p(1) \tag{3.22}$$

The remainder of the proof is the same as in standard GMM applications, see Hall (2005) for example. □

Proof of Proposition 5. Lemma 1, 2, 3 and 4 are not affected by mis-specification. Lemma 5 (i) is replaced by Assumption 5 (ii). Therefore, $\hat{\theta}_{T,S} \xrightarrow{p} \theta_*(\mathbf{W}_0)$. □

3.7 Tables and figures

Table 3.1: Simulation results for iid data

	Clayton				*Normal*			*Factor copula*					
	MLE	GMM	SMM	SMM*	MLE	GMM	SMM	MLE			SMM		
	κ	κ	κ	κ	ρ	ρ	ρ	σ^2	ν^{-1}	λ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	1.00	0.5	0.5	0.5	1.00	0.25	-0.50	1.00	0.25	-0.50
							$N = 2$						
Bias	0.001	-0.014	-0.006	-0.004	0.004	-0.001	-0.001	0.026	-0.002	-0.026	0.016	-0.012	-0.089
St dev	0.085	0.119	0.122	0.110	0.024	0.034	0.034	0.135	0.045	0.144	0.152	0.123	0.199
Median	1.011	0.982	0.991	0.998	0.503	0.497	0.498	1.027	0.251	-0.500	0.985	0.243	-0.548
90-10%	0.216	0.308	0.293	0.294	0.063	0.086	0.087	0.331	0.118	0.355	0.374	0.363	0.540
Time	0.061	0.060	512	49.5	0.021	0.292	0.483		254			103	
							$N = 3$						
Bias	0.015	0.008	0.006	0.008	0.003	-0.004	-0.005	0.014	-0.001	-0.012	0.032	-0.002	-0.057
St dev	0.061	0.090	0.092	0.091	0.020	0.025	0.026	0.120	0.028	0.109	0.124	0.111	0.157
Median	1.013	1.003	0.999	0.998	0.503	0.497	0.499	1.001	0.250	-0.502	1.031	0.256	-0.542
90-10%	0.155	0.226	0.219	0.216	0.049	0.064	0.068	0.297	0.073	0.222	0.297	0.293	0.395
Time	0.113	0.091	1360	56.2	0.023	0.293	0.815		263			136	
							$N = 10$						
Bias	0.008	0.007	0.008	0.004	0.003	-0.002	-0.002	0.011	0.000	0.006	0.026	0.001	-0.011
St dev	0.050	0.068	0.066	0.059	0.014	0.017	0.017	0.092	0.016	0.063	0.093	0.070	0.082
Median	1.005	1.002	1.005	0.999	0.504	0.498	0.499	1.005	0.248	-0.494	1.013	0.255	-0.508
90-10%	0.132	0.198	0.177	0.152	0.035	0.039	0.045	0.240	0.034	0.166	0.248	0.186	0.168
Time	0.409	0.998	22289	170	0.475	0.331	3.140		396			341	

Table 3.2: Simulation results for AR-GARCH data

	Clayton				*Normal*			*Factor copula*					
	MLE	GMM	SMM	SMM*	MLE	GMM	SMM	MLE			SMM		
	κ	κ	κ	κ	ρ	ρ	ρ	σ^2	ν^{-1}	λ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	1.00	0.5	0.5	0.5	1.00	0.25	-0.50	1.00	0.25	-0.50
							$N = 2$						
Bias	-0.005	-0.029	-0.028	-0.020	0.003	-0.001	-0.001	0.021	-0.009	-0.029	0.015	-0.012	-0.073
St dev	0.087	0.124	0.124	0.108	0.024	0.035	0.036	0.137	0.046	0.150	0.155	0.121	0.188
Median	0.998	0.977	0.975	0.982	0.503	0.497	0.499	1.021	0.245	-0.503	0.995	0.235	-0.558
90-10%	0.228	0.327	0.340	0.267	0.061	0.084	0.090	0.343	0.118	0.382	0.411	0.346	0.509
Time	0.026	0.059	525	52	0.030	0.299	0.505		234			95	
							$N = 3$						
Bias	0.006	-0.007	0.002	-0.008	0.003	-0.005	-0.006	0.007	-0.007	-0.011	0.013	-0.020	-0.052
St dev	0.060	0.087	0.088	0.080	0.020	0.026	0.026	0.118	0.028	0.110	0.121	0.106	0.148
Median	1.005	0.991	0.994	0.981	0.502	0.497	0.499	0.997	0.243	-0.502	1.005	0.238	-0.521
90-10%	0.145	0.205	0.213	0.195	0.050	0.065	0.068	0.315	0.074	0.224	0.311	0.297	0.357
Time	0.127	0.108	1577	73	0.022	0.288	1.009		232			119	
							$N = 10$						
Bias	-0.004	-0.002	-0.004	-0.003	0.002	-0.003	-0.004	0.005	-0.006	0.008	-0.004	-0.016	-0.012
St dev	0.049	0.067	0.064	0.059	0.014	0.016	0.017	0.091	0.015	0.063	0.085	0.079	0.071
Median	0.995	0.996	0.987	0.988	0.503	0.497	0.497	1.002	0.243	-0.493	0.990	0.238	-0.508
90-10%	0.134	0.179	0.170	0.152	0.034	0.041	0.045	0.240	0.037	0.169	0.210	0.209	0.165
Time	0.292	1.059	24549	171	1.099	0.392	3.437		430			309	

Notes to Table 3.1: This table presents the results from 100 simulations of Clayton, the Normal copula, and a factor copula. In the SMM and GMM estimation all three copulas use five dependence measures, including four quantile dependence measures ($q = 0.05, 0.10, 0.90, 0, 95$). The Normal and factor copulas also use Spearman's rank correlation, while the Clayton copula uses either Kendall's (GMM and SMM) or Spearman's (SMM*) rank correlation. The marginal distributions of the data are assumed to be *iid* $N(0, 1)$. Problems of dimension $N = 2$, 3 and 10 are considered, the sample size is $T = 1,000$ and the number of simulations used is $S = 25 \times T$. 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 of the estimated parameters. The third and fourth rows present the median and the difference between the 90^{th} and 10^{th} percentiles of the distribution of estimated parameters. The last row in each panel presents the average time in seconds to compute the estimator.

Notes to Table 3.2: This table presents the results from 100 simulations of Clayton, the Normal copula, and a factor copula. In the SMM and GMM estimation all three copulas use five dependence measures, including four quantile dependence measures ($q = 0.05, 0.10, 0.90, 0, 95$). The Normal and factor copulas also use Spearman's rank correlation, while the Clayton copula uses either Kendall's (GMM and SMM) or Spearman's (SMM*) rank correlation. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3.3. Problems of dimension $N = 2$, 3 and 10 are considered, the sample size is $T = 1,000$ and the number of simulations used is $S = 25 \times T$. 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 of the estimated parameters. The third and fourth rows present the median and the difference between the 90^{th} and 10^{th} percentiles of the distribution of estimated parameters. The last row in each panel presents the average time in seconds to compute the estimator.

Table 3.3: Simulation results on coverage rates

	Clayton		Normal		Factor copula			
	κ	J	ρ	J	σ^2	ν^{-1}	λ	J
				$N=2$				
$\varepsilon_{T,S}$								
0.1	91	98	94	98	94	100	95	98
0.01	46	99	92	98	94	99	96	100
0.001	2	99	76	98	76	79	74	99
0.0001	1	99	21	98	54	75	57	97
				$N=3$				
$\varepsilon_{T,S}$								
0.1	97	99	89	97	99	100	96	99
0.01	63	98	88	97	99	96	95	100
0.001	11	98	83	98	92	84	93	100
0.0001	2	100	38	99	57	70	61	99
				$N=10$				
$\varepsilon_{T,S}$								
0.1	96	99	87	97	97	97	95	98
0.01	88	99	87	96	96	97	97	97
0.001	18	100	87	98	97	95	88	97
0.0001	0	98	71	97	73	85	81	98

Notes: This table presents the results from 100 simulations of Clayton copula, the Normal copula, and a factor copula, all estimated by SMM. The marginal distributions of the data are assumed to follow AR(1)-GARCH(1,1) processes, as described in Section 3.3. Problems of dimension $N = 2$, 3 and 10 are considered, the sample size is $T = 1,000$ and the number of simulations used is $S = 25 \times T$. The rows of each panel contain the step size, $\varepsilon_{T,S}$, used in computing the matrix of numerical derivatives, $\hat{\mathbf{G}}_{T,S}$. The numbers in column $\kappa, \rho, \sigma^2, \nu^{-1}$, and λ present the percentage of simulations for which the 95% confidence interval based on the estimated covariance matrix contained the true parameter. The numbers in column J present the percentage of simulations for which the test statistic of over-identifying restrictions test described in Section 3.2 was smaller than its computed critical value under 95% confidence level.

Table 3.4: Simulation results for mis-specified models

	iid		*AR-GARCH*	
True copula	*Normal*	*Clayton*	*Normal*	*Clayton*
Model	Clayton	Normal	Clayton	Normal
		$N = 2$		
Pseudo-true	0.542	0.599	0.543	0.588
Bias	-0.013	0.111	-0.007	0.046
St dev	0.050	0.173	0.035	0.120
Median	0.526	0.659	0.539	0.617
90-10%	0.130	0.433	0.091	0.265
Time	4	72	1	70
J test prob.	0	0	0	0
		$N = 3$		
Pseudo-true	0.543	0.599	0.542	0.607
Bias	0.003	0.077	-0.002	0.006
St dev	0.039	0.164	0.027	0.088
Median	0.544	0.629	0.540	0.609
90-10%	0.107	0.432	0.072	0.198
Time	5	90	1	86
J test prob.	0	0	0	0
		$N = 10$		
Pseudo-true	0.544	0.602	0.544	0.603
Bias	0.001	0.059	-0.001	0.047
St dev	0.033	0.118	0.016	0.116
Median	0.546	0.622	0.540	0.618
90-10%	0.086	0.307	0.043	0.314
Time	20	206	4	207
J test prob.	0	0	0	0

Notes: This table presents the results from 100 simulations when the true copula and the model are different (i.e., the model is mis-specified). The parameters of the copula models are estimated using SMM based on rank correlation and four quantile dependence measures ($q = 0.05, 0.10, 0.90, 0, 95$). The marginal distributions of the data are assumed to be either *iid* $N(0, 1)$ or AR(1)-GARCH(1,1) processes, as described in Section 3.3. Problems of dimension $N = 2$, 3 and 10 are considered, the sample size is $T = 1,000$ and the number of simulations used is $S = 25 \times T$. The pseudo-true parameter is estimated using 10 million observations. The last row in each panel presents the proportion of tests of over-identifying restrictions that are smaller than the 95% critical value.

Table 3.5: Sample dependence statistics

	Bank of America	Bank of N.Y.	Citi Group	Goldman Sachs	JP Morgan	Morgan Stanley	Wells Fargo
BoA		0.586	0.691	0.556	0.705	0.602	0.701
BoNY	0.551		0.574	0.578	0.658	0.592	0.595
Citi	0.685	0.558		0.608	0.684	0.649	0.626
Goldman	0.564	0.565	0.609		0.655	0.759	0.548
JPM	0.713	0.633	0.694	0.666		0.667	0.683
Morgan S	0.604	0.587	0.650	0.774	0.676		0.578
Wells F	0.715	0.593	0.636	0.554	0.704	0.587	
BoA		0.219	0.239	0.219	0.398	0.298	0.358
BoNY	-0.048		0.179	0.199	0.159	0.219	0.199
Citi	-0.045	-0.004		0.199	0.318	0.219	0.199
Goldman	-0.068	0.000	0.032		0.239	0.378	0.199
JPM	-0.024	-0.056	-0.012	0.012		0.239	0.358
Morgan S	-0.060	-0.020	-0.064	-0.036	-0.008		0.219
Wells F	0.020	-0.052	0.044	-0.028	0.024	0.000	

Notes: This table presents measures of dependence between the seven financial firms under analysis. The upper panel presents Spearman's rank correlation (upper triangle) and linear correlation (lower triangle), and the lower panel presents the difference between the 10% tail dependence measures (lower triangle) and average 1% upper and lower tail dependence (upper triangle). All dependence measures are computed using the standardized residuals from the models for the conditional mean and variance.

Table 3.6: Estimation results for daily returns on seven stocks

	Clayton			*Normal*			*Factor*		
	Panel A: Full sample (2001-2010)								
	MLE	SMM	SMM-opt	MLE	SMM	SMM-opt	MLE	SMM	SMM-opt
	κ	κ	κ	ρ	ρ	ρ	$\sigma^2, \nu^{-1}, \lambda$	$\sigma^2, \nu^{-1}, \lambda$	$\sigma^2, \nu^{-1}, \lambda$
Estimate	0.907	1.274	1.346	0.650	0.682	0.659	1.995	2.019	1.955
Std err	0.028	0.048	0.037	0.007	0.010	0.008	0.020	0.077	0.069
Estimate	–	–	–	–	–	–	0.159	0.088	0.115
Std err	–	–	–	–	–	–	0.010	0.034	0.033
Estimate	–	–	–	–	–	–	-0.021	-0.015	-0.013
Std err	–	–	–	–	–	–	0.032	0.035	0.034
$Q_{SMM} \times 100$	–	19.820	19.872	–	0.240	0.719	–	0.040	0.187
J_{pval}	–	0.000	0.000	–	0.043	0.001	–	0.139	0.096
Time	0.7	344	360	0.5	6	6	1734	801	858

Notes: This table presents estimation results for various copula models applied to seven daily stock returns in the financial sector over the period January 2001 to December 2010. Estimates and asymptotic standard errors for the copula model parameters are presented, as well as the value of the SMM objective function at the estimated parameters and the p-value of the overidentifying restriction test. Estimates labeled SMM are estimated using the identity weight matrix; estimates labeled SMM-opt are estimated using the efficient weight matrix.

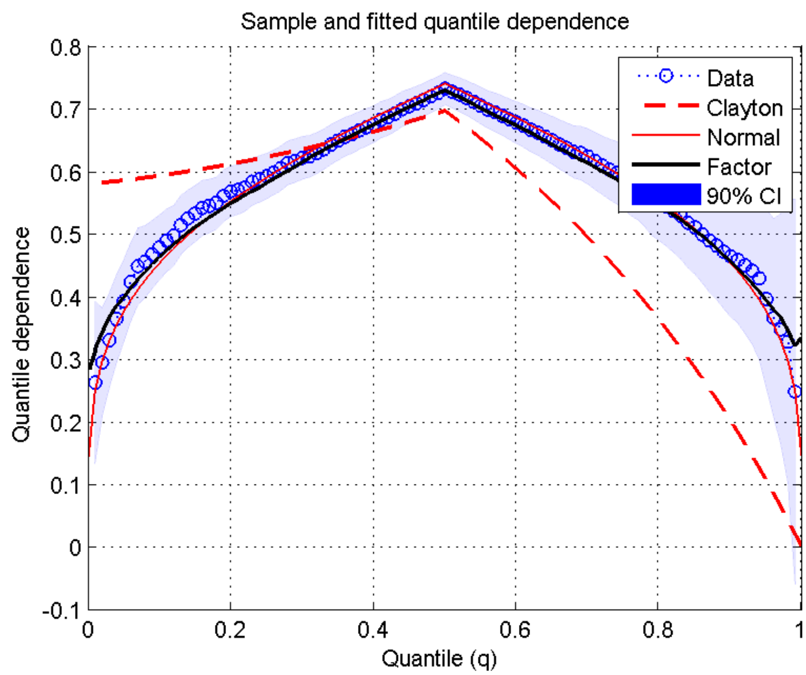


FIGURE 3.1: *This figure plots the probability of both variables being less than their q quantile (q<0.5) or greater than the q quantile (q>0.5). For the data this is averaged across all pairs, and a bootstrap 90% (pointwise) confidence interval is presented.*

4

Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads (co-authored with Andrew Patton)

4.1 Introduction

Systemic risk can be broadly defined as the risk of distress in a large number of firms or institutions. It represents an extreme event in two directions: a large loss (e.g., corresponding to a large left-tail realization for stock returns), across a large proportion of the firms. There are a variety of methods for studying risk and dependence for small collections of assets, see Patton (2012) for a review of copula-based approaches, but a relative paucity of methods for studying dependence between a large collection of assets, which is required for a general analysis of systemic risk.

Some existing methods for estimating systemic risk simplify the task by reducing the dimension of the problem to two: an individual firm and a market index. The "CoVaR" measure of Adrian and Brunnermeier (2009), for example, uses quantile regression to estimate a lower tail quantile (e.g., 0.05) of market returns conditional on a given firm having a returns equal to its lower tail quantile. The "marginal expected shortfall" proposed by Brownlees and Engle (2011) estimates the expected return on a firm conditional on the market return being below some low threshold. These methods have the clear benefit of being parsimonious, but by aggregating the "non firm i" universe to a single market index, useful information about systemic risk may be missed. The objective of this paper is to provide models that can be used to handle large collections of variables, which enables the estimation of a wider variety of systemic risk measures.

We use Sklar's theorem (see Nelsen, 2006), with an extension to conditional distributions from Patton (2006), to decompose the conditional joint distribution of a collection of N variables into their marginal distributions and a conditional copula:

$$\mathbf{Y}_t|\mathcal{F}_{t-1} \sim \mathbf{F}_t = \mathbf{C}_t \left(F_{1t}, ..., F_{Nt}\right) \tag{4.1}$$

We propose new models for the time-varying conditional copula, $\mathbf{C}_t$, that can be used to link models of the conditional marginal distributions (e.g., ARMA-GARCH models) to form a dynamic conditional joint distribution. Of central relevance to this paper are cases where N is relatively large, around 50 to 250. In such cases, models that have been developed for low dimension problems (say, $N < 5$) are often not applicable, either because no generalization beyond the bivariate model exists, or because such generalizations are too restrictive (e.g., Archimedean copulas have just one or two free parameters regardless of N, which is clearly very restrictive in high dimensions), or because the obvious generalization of the bivariate case leads to a proliferation of parameters and unmanageable computational complexity. In high dimension applications, the challenge is to find a balance of flexibility and parsimony.

This paper makes two contributions. First, we propose a flexible and feasible model for capturing time-varying dependence in high dimensions. Our approach draws on successful ideas from the literature on dynamic modeling of high dimension covariance matrices and on recent work on models for general time-varying distributions. In particular, we combine the "GAS" model of Creal, *et al.* (2011, 2013), parameter restrictions and "variance targeting" ideas from Engle (2002) and Engle and Kelly (2012), and the factor copula model of Oh and Patton (2011) to obtain a flexible yet parsimonious dynamic model for high dimension conditional distributions. A realistic simulation study confirms that our proposed models and estimation methods have satisfactory properties for relevant sample sizes.

Our second contribution is a detailed study of a collection of 100 daily credit default swap (CDS) spreads on U.S. firms. The CDS market has expanded enormously over the last decade, growing 40-fold from $0.6 trillion of gross notional principal in 2001 to $25.9 trillion at the end of 2011 according to the International Swaps and Derivatives Association (ISDA), yet it has received relatively little attention in the econometrics literature. (Interest is growing, however, see Conrad, *et al.* (2011), Lucas, *et al.* (2011), Creal, *et al.* (2012) and Christoffersen, *et al.* (2013) for recent work on CDS data.) We use our model of CDS spreads to provide insights into systemic risk, as CDS spreads are tightly linked to the health of the underlying firm. We find that systemic risk rose during the financial crisis, unsurprisingly. More interestingly, we also find that systemic risk remains high relative to the pre-crisis period, even though idiosyncratic risk as fallen.

The remainder of the paper is structured as follows. Section 4.2 presents a dynamic copula model for high dimension applications, and Section 4.3 presents a simulation study for the proposed model. In Section 4.4 we present estimation results for various models of CDS spreads. Section 4.5 presents estimates of time-varying systemic risk, and Section 4.6 concludes. Technical details and some additional results are presented in Appendix D.

4.2 A dynamic copula model for high dimensions

In this section we describe our approach for capturing dynamics in the dependence between a relatively large number of variables. A review of alternative methods from the (small) extant literature is presented in Section 4.2.3. We consider a class of data generating processes (DGPs) that allow for time-varying conditional marginal distributions, e.g., dynamic conditional means and variances, and also possibly time-varying higher-order moments:

$$\begin{aligned} \mathbf{Y}_t &\equiv [Y_{1t}, ..., Y_{Nt}]' \\ \text{where} \quad Y_{it} &= \mu_{it}(\phi_{i,0}) + \sigma_{it}(\phi_{i,0})\eta_{it}, \quad i = 1, 2, ..., N \\ \eta_{it}|\mathcal{F}_{t-1} &\sim F_{it}(\phi_{i,0}) \end{aligned} \tag{4.2}$$

where μ_{it} is the conditional mean of Y_{it}, σ_{it} is the conditional standard deviation, and $F_{it}(\phi_{i,0})$ is a parametric distribution with zero mean and unit variance. We will denote the parameters of the marginal distributions as $\phi \equiv [\phi_1', ..., \phi_N']'$, the parameters of the copula as γ, and the vector of all parameters as $\theta \equiv [\phi', \gamma']'$. We assume that F_{it} is continuous and strictly increasing, which fits our empirical application, though this assumption can be relaxed. The information set is taken to be $\mathcal{F}_t = \sigma(\mathbf{Y}_t, \mathbf{Y}_{t-1}, ...)$. Define the conditional probability integral transforms of the data as:

$$U_{it} \equiv F_{it}\left(\frac{Y_{it} - \mu_{it}(\phi_{i,0})}{\sigma_{it}(\phi_{i,0})}; \phi_{i,0}\right), \quad i = 1, 2, ..., N \tag{4.3}$$

Then the conditional copula of $\mathbf{Y}_t|\mathcal{F}_{t-1}$ is equal to the conditional distribution of $\mathbf{U}_t|\mathcal{F}_{t-1}$:

$$\mathbf{U}_t|\mathcal{F}_{t-1} \sim \mathbf{C}_t(\gamma_0) \tag{4.4}$$

By allowing for a time-varying conditional copula, the class of DGPs characterized by equations (4.2) to (4.4) is a generalization of those considered by Chen and Fan

(2006), for example, however the cost of this flexibility is the need to specify parametric marginal distributions. In contrast, Chen and Fan (2006), Rémillard (2010) and Oh and Patton (2013a) allow for nonparametric estimation of the marginal distributions. The parametric margin requirement arises as the asymptotic distribution theory for a model with nonparametric margins and a time-varying copula is not yet available in the literature. We attempt to mitigate this requirement in our empirical work by using flexible models for the marginal distributions, and conducting goodness-of-fit tests to verify that they provide a satisfactory fit to the data.

4.2.1 Factor copulas

In high dimension applications a critical aspect of any model is imposing some form of dimension reduction. A widely-used method to achieve this in economics and finance is to use some form of factor structure. Oh and Patton (2011) propose using a factor model with flexible distributions to obtain a flexible class of "factor copulas." A one-factor version of their model is the copula for the (latent) vector random variable $\mathbf{X}_t \equiv [X_{1t}, ..., X_{Nt}]'$ implied by the following structure:

$$X_{it} = \lambda_{it}(\gamma_\lambda) Z_t + \varepsilon_{it}, \;\; i = 1, 2, ..., N \tag{4.5}$$

$$\text{where} \quad Z_t \sim F_{zt}(\gamma_z), \;\; \varepsilon_{it} \sim iid \; F_{\varepsilon t}(\gamma_\varepsilon), \;\; Z \perp\!\!\!\perp \varepsilon_i \; \forall \; i$$

where $F_{zt}(\gamma_z)$ and $F_{\varepsilon t}(\gamma_\varepsilon)$ are flexible parametric univariate distributions for the common factor and the idiosyncratic variables respectively, and $\lambda_{it}(\gamma_\lambda)$ is a potentially time-varying weight on the common factor. The conditional joint distribution for $\mathbf{X}_t$ can be decomposed into its conditional marginal distributions and its conditional copula via Sklar's theorem (see Nelsen (2006)) for conditional distributions, see Patton (2006b):

$$\mathbf{X}_t \sim \mathbf{F}_{xt} = \mathbf{C}_t(G_{1t}(\gamma), ..., G_{Nt}(\gamma); \gamma) \tag{4.6}$$

where $\gamma \equiv [\gamma_z', \gamma_\varepsilon', \gamma_\lambda']'$. Note that the marginal distributions of $\mathbf{X}_t$ need not be the same as the marginal distributions of the observed data. Only the *copula* of these variables, denoted $\mathbf{C}_t(\gamma)$, is used as a model for the copula of the observable data $\mathbf{Y}_t$. If we impose that the *marginal* distributions of the observable data are *also* driven by the factor structure in equation (4.5), then this becomes a standard factor model for a vector of variables. However, Oh and Patton (2011) suggest imposing the factor structure only on the component of the multivariate model where dimension reduction is critical, namely the copula, and allow the marginal distributions to be modeled using a potentially different approach. In this case, the factor structure in equation (4.5) is used only for the copula that it implies, and this becomes a "factor copula" model.

The copula implied by equation (4.5) is known in closed form for only a few particular combinations of choices of F_z and F_ε (the most obvious example being where both of these distributions are Gaussian, in which case the implied copula is also Gaussian). For general choices of F_z and F_ε the copula of $\mathbf{X}$ will not be known in closed form, and thus the copula likelihood is not known in closed form. Numerical methods can be used to overcome this problem. Oh and Patton (2013a) propose simulated method of moments-type estimation of the unknown parameters, however their approach is only applicable when the conditional copula is constant. A key objective of this paper is to allow the conditional copula to vary through time and so an alternate estimation approach is required. We use a simple numerical integration method, described in Appendix D.2, to overcome the lack of closed-form likelihood. This numerical integration exploits the fact that although the copula is N-dimensional, we need only integrate out the common factor, which is one-dimensional in the structure above.

Dynamics in the factor copula model in equation (4.5) arise by allowing the loadings on the common factor, λ_{it}, to vary through time, and/or by allowing the distributions of the common factor and the idiosyncratic variables to change through time. For example, holding F_{zt} and $F_{\varepsilon t}$ fixed, an increase in the factor loadings corresponds to an increase in the level of overall dependence (e.g., rank correlation) between the variables. Holding the factor loadings fixed, an increase in the thickness of the tails of the distribution of the common factor increases the degree of tail dependence. In the next section we describe how we model these dynamics.

4.2.2 "GAS" dynamics

An important feature of any dynamic model is the specification for how the parameters evolve through time. Some specifications, such as stochastic volatility models (see Shephard (2005) for example) and related stochastic copula models (see Hafner and Manner (2012) and Manner and Segers (2011)) allow the varying parameters to evolve as a latent time series process. Others, such as ARCH-type models for volatility (see Engle, 1982) and related models for copulas (see Patton (2006b), Jondeau and Rockinger (2006), and Creal, *et al.* (2013) for example) model the varying parameters as some function of lagged observables. An advantage of the latter approach over the former, in particular for high dimension applications, is that it avoids the need to "integrate out" the innovation terms driving the latent time series processes.

Within the class of ARCH-type models ("observation driven", in the terminology of Creal, *et al.* (2013)), the question of *which* function of lagged observables to use as a forcing variable in the evolution equation for the varying parameter arises. For models of the conditional variance, an immediate choice is the lagged squared residual, as in the ARCH model, but for models with parameters that lack an obvious interpretation the choice is less clear. We adopt the generalized autoregressive score (GAS) model of Creal, *et al.* (2013) to overcome this problem. (Harvey (2013) and Harvey and Sucarrat (2012) propose a similar method for modeling time-varying parameters, which they call a "dynamic conditional score," or "DCS," model.) These authors propose using the lagged score of the density model (copula model, in our application) as the forcing variable. Specifically, for a copula with time-varying parameter δ_t, governed by fixed parameter γ, we have:

$$
\begin{aligned}
\text{Let} \quad & \mathbf{U}_t|\mathcal{F}_{t-1} \sim \mathbf{C}(\delta_t(\gamma)) \\
\text{then} \quad \delta_t &= \omega + B\delta_{t-1} + A\mathbf{s}_{t-1} \qquad (4.7) \\
\text{where} \quad \mathbf{s}_{t-1} &= S_{t-1}\nabla_{t-1} \\
\nabla_{t-1} &= \frac{\partial \log \mathbf{c}(\mathbf{u}_{t-1}; \delta_{t-1})}{\partial \delta_{t-1}}
\end{aligned}
$$

and S_t is a scaling matrix (e.g., the inverse Hessian or its square root). While this specification for the evolution of a time-varying parameter is somewhat arbitrary, Creal, *et al.* (2013) provide two motivations for it. Firstly, this model nests a variety of popular and successful existing models: GARCH (Bollerslev (1986)) for conditional variance; ACD (Engle and Russell (1998)) for models of trade durations (the time between consecutive high frequency observations); Davis, *et al.*'s (2003) model for Poisson counts. Secondly, the recursion above can be interpreted as the steepest ascent direction for improving the model's fit, in terms of the likelihood, given the current value of the model parameter δ_t, similar to numerical optimization algorithms such as the Gauss-Newton algorithm. Harvey (2013) further motivates this specification as an approximation to a filter for a model driven by a stochastic latent parameter, or an "unobserved components" model.

GAS dynamics for high dimension factor copulas

We employ the GAS model to allow for time variation in the factor loadings in the factor copula implied by equation (4.5), but to keep the model parsimonious we impose that the parameters governing the "shape" of the common and idiosyncratic variables (γ_z and γ_ε) are constant. We use the skewed t distribution of Hansen (1994) as the model for F_z, and the (symmetric) standardized t distribution as the model for F_ε. The skewed t distribution has two shape parameters, a degrees of freedom parameter ($\nu_z \in (2, \infty]$) and an asymmetry parameter ($\psi_z \in (-1, 1)$). This distribution simplifies to the standardized t distribution when $\psi = 0$. We impose symmetry on the distribution of the idiosyncratic variables for simplicity.

In the general GAS framework in equation (4.7), the N time-varying factor loadings would have $N + 2N^2$ parameters governing their evolution, which represents an infeasibly large number for even moderate values of N. To keep the model parsimonious, we impose that the coefficient matrices (B and A) are diagonal with a common parameter on the diagonal, as in the DCC model of Engle (2002). To avoid the estimation of $N \times N$ scaling matrix we set $S_t = I$. This simplifies our model to be (in logs):

$$\log \lambda_{it} = \omega_i + \beta \log \lambda_{i,t-1} + \alpha s_{i,t-1}, \quad i = 1, 2, ..., N \tag{4.8}$$

where $s_{it} \equiv \partial \log \mathbf{c}(\mathbf{u}_t; \lambda_t, \nu_z, \psi_z, \nu_\varepsilon)/\partial \lambda_{it}$ and $\lambda_t \equiv [\lambda_{1t}, ..., \lambda_{Nt}]'$. The dynamic copula model implied by equations (4.5) and (4.8) thus contains $N + 2$ parameters for the GAS dynamics, 3 parameters for the shape of the common and idiosyncratic variables, for a total of $N + 5$ parameters.

Equidependence vs. heterogeneous dependence

To investigate whether we can further reduce the number of free parameters in this model we consider two restrictions of the model in equation (4.8), motivated by the "dynamic equicorrelation" model of Engle and Kelly (2012). If we impose that $\omega_i = \omega \ \forall \ i$, then the pair-wise dependence between each of the variables will be identical, leading to a "dynamic equidependence" model. (The copula implied by this specification is "exchangeable" in the terminology of the copula literature.) In this case we have only 6 parameters to estimate independent of the number of variables N, vastly reducing the estimation burden, but imposing a lot of homogeneity on the model.

An intermediate step between the fully flexible model in equation (4.8) and the equidependence model is to group the assets using some *ex ante* information (e.g., by industry for stock returns or CDS spreads) and impose homogeneity only within groups. This leads to a "block equidependence" model, with

$$X_{it} = \lambda_{g(i),t} Z_t + \varepsilon_{it}, \quad i = 1, 2, ..., N \tag{4.9}$$
$$\log \lambda_{g,t} = \omega_g + \beta \log \lambda_{g,t-1} + \alpha s_{g,t-1}, \quad g = 1, 2, ..., G$$

where $g(i)$ is the group to which variable i belongs, and G is the number of groups. In this case the number of parameters to estimate in the copula model is $G+2+3$. In our empirical application we have $N = 100$ and we consider grouping variables into $G = 5$ industries, meaning this model has 10 parameters to estimate rather than 105. In our empirical analysis below, we compare these two restricted models ($G = 1$ and $G = 5$) with the "heterogeneous dependence" model which allows a different factor for each variable $(G = N)$.

A "variance targeting" method

Estimating the fully flexible model above involves numerically searching over $N + 5$ parameters, and for $N = 100$ this represents quite a computational challenge. We propose a method to overcome this challenge by adapting an idea from the DCC model of Engle (2002). Specifically, we use a "variance targeting" (Engle and Mezrich (1996)) method to replace the constant ω_i in the GAS equation with a transformation of a sample dependence measure. The nature of our GAS specification means that the variance targeting approach needs to be modified for use here.

The evolution equation for λ_{it} in equation (4.8) can be re-written as

$$\log \lambda_{it} = E[\log \lambda_{it}](1 - \beta) + \beta \log \lambda_{i,t-1} + \alpha s_{i,t-1}$$

using the result from Creal, *et al.* (2013) that $E_{t-1}[s_{it}] = 0$, and so $E[\log \lambda_{it}] = \omega_i / (1 - \beta)$. The proposition below provides a method for using sample rank correlations to obtain an estimate of $E[\log \lambda_{it}]$, thus removing the need to numerically optimize over the intercept parameters, ω_i. The proposition is based on the following assumption.

Assumption 1. *(a) The conditional copula of $\mathbf{Y}_t|\mathcal{F}_{t-1}$ is the time-varying factor copula given in equations (4.5) and (4.8).*

(b) The process $\{\lambda_t\}$ generated by equation (4.8) is strictly stationary.

(c) Let $\rho_{t,X} \equiv vech\left(RankCorr_{t-1}[\mathbf{X}_t]\right)$. Then $\log \lambda_t$ is a linear function of $\rho_{t,X}$.

(d) Let $\rho_{ij,X} \equiv RankCorr[X_i, X_j]$ and $\rho^L_{ij,X} \equiv Corr[X_i, X_j]$. Then, for fixed values of $(\gamma_z, \gamma_\varepsilon)$, the mapping $\rho_{ij} = \varphi(\rho^L_{ij})$ is strictly increasing.

Part (a) of this assumption makes explicit that the copula of the data is the GAS-factor copula model, and so the conditional copula of $\mathbf{Y}_t|\mathcal{F}_{t-1}$ is the same as that of $\mathbf{X}_t|\mathcal{F}_{t-1}$. Blasques, *et al.* (2012) which provide conditions under which univariate GAS models satisfy stationarity conditions; corresponding theoretical results for the multivariate case are not yet available in the literature, and thus in part (b) we simply assume that stationarity holds. Part (c) formalizes the applicability of a Taylor series expansion of the function mapping ρ_t to λ_t. In practice this assumption will hold only approximately, and its applicability needs to be verified via simulation, which we discuss further in Section 4.3. Part (d) enables us to map rank correlations to linear correlations. Note that we can take $(\gamma_z, \gamma_\varepsilon)$ as fixed, as we call this mapping for each evaluation of the log-likelihood, which provides us with a value for $(\gamma_z, \gamma_\varepsilon)$. Importantly, this mapping can be computed *prior* to estimation, and then just called during estimation, rather than re-computed each time the likelihood function is evaluated.

Proposition 1. *Let Assumption 1 hold, and denote the vech of the rank correlation matrix of the standardized residuals as $\bar{\rho}^S_\eta$ and its sample analog as $\hat{\rho}^S_\eta$. Then:*

(i) $E[\log \lambda_t] = H\left(\bar{\rho}^S_\eta\right)$, where H is defined in equation (D.3).

(ii) $\widehat{\log \lambda} = H\left(\hat{\rho}_{\eta}^{S}\right)$ *is a GMM estimator of* $E\left[\log \lambda_t\right]$.

Part (i) of the above proposition provides the mapping from the population rank correlation of the standardized residuals to the mean of the (log) factor loadings, which is the basis for considering a variance-targeting type estimator. Part (ii) shows that the sample analog of this mapping can be interpreted as a standard GMM estimator. This is useful as it enables us to treat the estimation of the entire conditional joint distribution model as multi-stage GMM, and draw on results for such estimators to conduct inference, see White (1994), Engle and Sheppard (2001) and Gonçalves et al. (2013). The latter paper provides conditions under which a block bootstrap may be used to obtain valid standard errors on parameters estimated via multi-stage GMM. The resulting standard errors are not higher-order efficient, like some bootstrap inference methods, but they do enable us to avoid having to handle Hessian matrices of size on the order of $2N \times 2N$. Note that sample rank correlations cannot in general be considered as moment-based estimators, as they depend on the sample ranks of the observed data, and studying their estimation properties requires alternative techniques. However, we exploit the fact that the marginal distributions of the data are known up to an unknown parameter vector, and thus rank correlation can be computed as a sample moment of a nonlinear function of the data.

4.2.3 Other models for dynamic, high dimension copulas

As noted above, the literature contains relatively few models for dynamic, high dimension copulas. Exceptions to this are discussed here. Lucas, *et al.* (2011) combine GAS dynamics with a skewed t copula to model ten sovereign CDS spreads. A similar model, though with an alternative skew t specification and with Engle's (2002) DCC dynamics, is used by Christoffersen, *et al.* (2012, 2013). The former of these two papers analyzes equity returns on up to 33 national stock indices, while the latter studies weekly equity returns and CDS spreads on 233 North American firms (and is the largest time-varying copula model in the extant literature). Almeida *et al.* (2012) use "vine" copulas to model the dependence between 30 German stock return series, with dynamics captured through a stochastic volatility-type equation for the parameters of the copula. Stöber and Czado (2012) also use vine copulas, combined with a regime-switching model for dynamics, to model dependence between ten German stock returns.

4.3 Simulation study

This section presents an analysis of the finite sample properties of maximum likelihood estimation for factor copulas with GAS dynamics. Factor copulas do not have a closed-form likelihood, and we approximate the likelihood using some standard numerical integration methods, details of which can be found in Appendix D.2. Oh and Patton (2013a) propose SMM-type estimation for factor copulas to overcome the lack of a closed-form likelihood, but a likelihood approach allows us to exploit the GAS model of Creal, *et al.* (2013) and so we pursue that here.

We consider three different copula models described for the Monte Carlo simulation: a dynamic equidependence model ($G = 1$), a dynamic block equidependence model ($G = 10$), and a dynamic heterogeneous dependence model ($G = N$), all of them governed by:

$$X_{it} = \lambda_{g(i),t} Z_t + \varepsilon_{it}, \quad i = 1, 2, ..., N \tag{4.10}$$

$$\log \lambda_{g,t} = \omega_g + \beta \log \lambda_{g,t-1} + \alpha s_{g,t-1}, \quad g = 1, 2, ..., G$$

$$Z \sim Skew\ t\,(\nu_z, \psi_z)$$

$$\varepsilon_i \sim iid\ t\,(\nu_\varepsilon)\,, \text{ and } \varepsilon_i \perp\!\!\!\perp Z\ \forall\ i$$

We set $N = 100$ to match the number of series in our empirical application below. For simplicity, we impose that $\nu_z = \nu_\varepsilon$, and we estimate ν^{-1} rather than ν, so that Normality is nested at $\nu^{-1} = 0$ rather than $\nu \to \infty$. Broadly matching the parameter estimates we obtain in our empirical application, we set $\omega = 0$, $\beta = 0.98$, $\alpha = 0.05$, $\nu = 5$, and $\psi_z = -0.1$ for the equidependence model. The block equidependence model uses the same parameters but sets $\omega_1 = -0.03$ and $\omega_{10} = 0.03$, and with ω_2 to ω_9 evenly spaced between these two bounds, and the heterogeneous dependence model similarly uses $\omega_1 = -0.03$ and $\omega_{100} = 0.03$, with ω_2 to ω_{99} evenly spaced between these two bounds. Rank correlations implied by these values range from 0.1 to 0.7. With these choices of parameter values and dependence designs, various dynamic dependence structures are covered, and asymmetric tail dependence, which is a common feature of financial data, is also allowed. We use a sample size of $T = 500$ and we repeat each simulation 100 times.

The results for the equidependence model presented in Panel A of Table 4.1 reveal that the average estimated bias for all parameters is small, and the estimates are centered on true values. The results for the block equidependence model, presented in Panel B, are also satisfactory, and, as expected, the estimation error in the parameters is generally slightly higher for this more complicated model.

The heterogeneous dependence model is estimated using the variance targeting-type approach for the intercepts, $\omega_{i,}$ described in Section 4.2.2, combined with numerical optimization for the remaining parameters. Appendix D.4 contains simulations that verify the applicability of Assumption 1 for this model, and the results presented in Panel C confirm that the approach leads to estimators with satisfactory finite-sample properties. (Panel C reports only every fifth intercept parameter, in the interests of space. The complete set of results is available in Appendix D.3.) The standard errors on the estimated intercept parameters are approximately twice as large, on average, as in the block equidependence case, however this model has seven times as many parameters as the block equidependence (104 vs. 14) and so some loss in accuracy is inevitable. Importantly, all estimated parameters are approximately centered on their true values, confirming that the assumptions underlying

Proposition 1 are applicable for this model.

4.4 Data description and estimation results

4.4.1 CDS spreads

We apply the dynamic copula model described in the previous section to daily credit default swap (CDS) spreads, obtained from Markit. In brief, a CDS is a contract in which the seller provides insurance to the buyer against any losses resulting from a default by the "reference entity" within some horizon. We focus on North American corporate CDS contracts, and the reference entities are thus North American firms. The CDS spread, usually measured in basis points and payable quarterly by the buyer to the seller, is the cost of this insurance. See Duffie and Singleton (2003) and Hull (2012) for more detailed discussions of CDS contracts, and see Barclays "CDS Handbook" (2010) for institutional details.

A key reason for interest in CDS contracts is sensitivity of CDS spreads to changes in market perceptions of the probability of default, see Conrad, *et al.* (2011), Creal, *et al.* (2012) and Christoffersen, *et al.* (2013) for recent empirical studies of implied default probabilities. Under some simplifying assumptions (such as a constant risk free rate and default hazard rate) see Carr and Wu (2011) for example, it can be shown that the CDS spread in basis points is:

$$S_{it} = 100^2 P_{it}^{\mathbb{Q}} L_{it} \tag{4.11}$$

where L_{it} is the loss given default (sometimes shortened to "LGD," and often assumed to equal 0.6 for U.S. firms) and $P_{it}^{\mathbb{Q}}$ is the implied probability of default. The same formula obtains as a first-order approximation at $P_{it}^{\mathbb{Q}} \approx 0$ for other more complicated pricing equations. This expression can be written in terms of the objective probability of default, $P_{it}^{\mathbb{P}}$:

$$S_{it} = 100^2 P_{it}^{\mathbb{P}} \mathcal{M}_{it} L_{it} \tag{4.12}$$

where $\mathcal{M}_{it}$ is the market price of risk (stochastic discount factor). An increase in a CDS spread can be driven by an increase in the LGD, an increase in the market price of default risk for this firm, or an increase in the objective probability of default. Any one of these three effects is indicative of a worsening of the health of the underlying firm.

In the analysis below we work with the log-difference of CDS spreads, to mitigate their autoregressive persistence, and under this transformation we obtain:

$$Y_{it} \equiv \Delta \log S_{it} = \Delta \log P_{it}^{\mathbb{P}} + \Delta \log \mathcal{M}_{it} + \Delta \log L_{it} \tag{4.13}$$

If the loss given default is constant then the third term above vanishes, and if we assume that the market price of risk is constant (as in traditional asset pricing models) or evolves slowly (for example, with a business cycle-type frequency) then daily changes in CDS spreads can be attributed primarily to changes in the objective probability of default. We will use this to guide our interpretation of the empirical results below, but we emphasize here that an increase in any of these three terms represents "bad news" for firm i, and so the isolation of the objective probability of default is not required for our interpretations to follow.

4.4.2 Summary statistics

Our sample period spans January 2006 to April 2012, a total of 1644 days. We study the 5-year CDS contract, which is the most liquid horizon (see Barclays (2010)), and we use "XR" ("no restructuring") CDS contracts, which became the convention for North America following the CDS market standardization in 2009 (the so-called "Big Bang"). To obtain a set of active, economically interesting, CDS data, we took all 125 individual firms in the CDS index covering our sample period (CDX Series 17). Of these, 90 firms had data that covered our entire sample period, and ten firms had no more than three missing observations. We use these 100 firms for our analysis.

(Of the remaining 25 firms, six are not U.S.-based firms and one firm stopped trading because of a firm split. None of the firms defaulted over this sample period.) A plot of these CDS spreads is presented in Figure 4.1, which reveals that the average CDS spread was around 100 basis points (bps), and it varied from a low (averaged across firms) of 24 bps on February 22, 2007, to a high of 304 bps on March 9, 2009.

The levels of our CDS spread data are suggestive of a large autoregressive root, with the median first-order autocorrelation across all 100 series being 0.996 (the minimum is 0.990). Further, augmented Dickey-Fuller tests reject the null hypothesis of a unit root at the 0.05 level for only 12 series. Like interest rate time series, these series are unlikely to literally obey a random walk, as they are bounded below, however we model all series in log differences to avoid the need to consider these series as near unit root processes.

Table 4.2 presents summary statistics on our data. Of particular note is the positive skewness of the log-differences in CDS spreads (average skewness is 1.087, and skewness is positive for 89 out of 100 series) and the excess kurtosis (25.531 on average, and greater than 3 for all 100 firms). Ljung-Box tests for autocorrelation at up to the tenth lag find significant (at the 0.05 level) autocorrelation in 98 out of 100 of the log-differenced CDS spreads, and for 89 series significant autocorrelation is found in the squared log-differences. This motivates specifying models for the conditional mean and variance to capture this predictability.

4.4.3 Conditional mean and variance models

Daily log-differences of CDS spreads have more autocorrelation than is commonly found for daily stock returns (e.g., the average first-order autocorrelation is 0.161) and so the model for the conditional mean of our data needs more structure than the commonly-used constant model for daily stock returns. We use an AR(5) augmented with one lag of the market variable (an equal-weighted average of all 100 series), and we show below that this model passes standard specification tests:

$$Y_{it} = \phi_{0i} + \sum_{j=1}^{5} \phi_{ji} Y_{i,t-j} + \phi_{mi} Y_{m,t-1} + e_{it} \tag{4.14}$$

For the market return we use the same model (omitting, of course, a repeat of the first lag of the market return). We need a model for the market return as we use the residuals from the market return model in our conditional variance specification.

Our model for the conditional variance is the asymmetric volatility model of Glosten, *et al.* (1993), the "GJR-GARCH" model. The motivation for the asymmetry in this model is that "bad news" about a firm increases its future volatility more than good news. For stock returns, bad news comes in the form of a negative residual. For CDS spreads, on the other hand, bad news is a *positive* residual, and so we reverse the direction of the indicator variable in the GJR-GARCH model to reflect this. In addition to the standard GJR-GARCH terms, we also include terms relating to the lagged market residual:

$$\begin{aligned} V_{t-1}\left[e_{it}\right] \equiv \sigma_{it}^2 &= \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i e_{i,t-1}^2 + \delta_i e_{i,t-1}^2 \mathbf{1}\left\{e_{i,t-1} > 0\right\} \\ &+ \alpha_{im} e_{m,t-1}^2 + \delta_{im} e_{m,t-1}^2 \mathbf{1}\left\{e_{m,t-1} > 0\right\} \end{aligned} \tag{4.15}$$

Finally, we specify a model for the marginal distribution of the standardized residuals, η_{it}. We use the skewed t distribution of Hansen (1994), which allows for non-zero skewness and excess kurtosis:

$$\eta_{it} \equiv \frac{e_{it}}{\sigma_{it}} \sim iid\ Skew\ t\left(\nu_i, \psi_i\right) \tag{4.16}$$

Table 4.3 summarizes the results of estimating the above models on the 100 time series. For the conditional mean model, we find strong significance of the first three AR lags, as well as the lagged market return. The conditional variance models reveal only mild statistical evidence of asymmetry in volatility, however the point estimates suggest that "bad news" (a positive residual) increases future volatility about 50% more than good news. The average estimated degrees of freedom parameter is 3.508, suggestive of fat tails, and the estimated skewness parameter is positive for 94 firms, and significantly different from zero for 41 of these, indicating positive skewness.

We now discuss goodness-of-fit tests for the marginal distribution specifications. We firstly use the Ljung-Box test to check the adequacy of these models for the conditional mean and variance, and we are able to reject the null of zero autocorrelation up to the tenth lag for only nine of the residual series, and only two of the squared standardized residual series. We conclude that these models provide a satisfactory fit to the conditional means and variances of these series. Next, we use the Kolmogorov-Smirnov test to investigate the fit of the skewed t distribution for the standardized residuals, using 100 simulations to obtain critical values that capture the parameter estimation error, and we reject the null of correct specification for just eleven of the 100 firms. This is slightly higher than the level of the test (0.05), but we do not pursue the use of a more complicated marginal distribution model for those eleven firms in the interests of parsimony and comparability.

4.4.4 The CDS "Big Bang"

On April 8, 2009, the CDS market underwent changes driven by a move towards more standardized CDS contracts. Details of these changes are described in Markit (2009). It is plausible that the changes to the CDS market around the Big Bang changed the dynamics and distributional features of CDS time series, and we test for that possibility here. We do so by allowing the parameters of the mean, variance, and marginal distribution models to change on the date of the Big Bang, and we test the significance of these changes. We have 591 pre-break observations and 1053 post-break observations.

We find that the conditional mean parameters changed significantly (at the 0.05

level) for 39 firms, and the conditional variance and marginal density shape parameters changed significantly for 66 firms. In what follows, the results we report are based on models that allow for a structural break in the mean, variance and distribution parameters. Given the prevalence of these changes, all of the copula models we consider allow for a break at the date of the Big Bang.

4.4.5 Comparing models for the conditional copula

The class of high dimension dynamic copula models described in Section 4.2 includes a variety of possible specifications: static vs. GAS dynamics; normal vs. skew *t*-*t* factor copulas; equidependence vs. block equidependence vs. heterogeneous dependence.

Table 4.4 presents results for six different dynamic models (a corresponding table for the six static copula models is in Table D.2). Bootstrap standard errors are presented in parentheses below the estimated parameters. (We use the stationary block bootstrap of Politis and Romano (1994) with an average block length of 120 days, applied to the log-difference of the CDS spreads, and we use 100 bootstrap replications.) Similar to other applications of GAS models (see, Creal *et al.* (2011, 2013)) we find strong persistence, with the β parameter ranging from 0.85 to 0.99. (Note that the β parameter in GAS models plays the same role as $\alpha + \beta$ in a GARCH(1,1) model, see Example 1 in Creal, *et al.* (2013)). We also find that the inverse degrees of freedom parameters are greater than zero (i.e., the factor copula is not Normal), which we test formally below. We further find that the asymmetry parameter for the common factor is positive, indicating greater dependence for joint upward moves in CDS spreads. This is consistent with financial variables being more correlated during bad times: for stock returns bad times correspond to joint downward moves, which have been shown in past work to be more correlated than joint upward moves, while for CDS spreads bad times correspond to joint upward moves.

Table 4.4 shows that the estimated degrees of freedom parameter for the common factor is larger than that for the idiosyncratic term. Oh and Patton (2011) show that when these two parameters differ the tail dependence implied by this factor copula is on the boundary: either zero ($\nu_z > \nu_\varepsilon$) or one ($\nu_z < \nu_\varepsilon$); only when these parameters are equal can tail dependence lie inside $(0, 1)$. We test the significance of the difference between these two parameters by estimating a model with them imposed to be equal and then conducting a likelihood ratio test, the log-likelihoods from these two models are reported in Table 4.5. The results strongly suggest that $\nu_z > \nu_\varepsilon$, and thus that extreme movements in CDS spreads are uncorrelated. The average gain in the log likelihood from estimating just this one extra parameter is around 200 points. This does not mean, of course, that "near extreme" movements must be uncorrelated, only that they are uncorrelated in the limit.

Table 4.5 also shows a comparison of the Skew *t*-*t* factor copula with the Normal copula, which is obtained by using a Normal distribution for both the common factor and the idiosyncratic factor. We see very clearly that the Normal copula performs worse than the Skew *t*-*t* factor copula, with the average gain in the log likelihood of the more flexible model being over 2000 points. This represents yet more evidence against the Normal copula model for financial time series; the Normal copula is simply too restrictive.

Finally, Table 4.5 compares the results from models with three different degrees of heterogeneity equidependence vs. block equidependence vs. heterogeneous dependence. We see that the data support the more flexible models, with the block equidependence model improving the equidependence model by around 200 points, and the heterogeneous model improving on the block equidependence model by around 800 points. It should be noted that our use of industry membership to form the "blocks" is just one method, and alternative grouping schemes may lead to better results. We do not pursue this possibility here.

Given the results in Table 4.5, our preferred model for the dependence structure of these 100 CDS spread series is a skew *t*-*t* factor copula, with separate degrees of freedom for the common and idiosyncratic variables, allowing for a separate loading on the common factor for each series (the "heterogeneous dependence" model) and allowing for dynamics using the GAS structure described in the previous section. Figure 4.2 presents the time-varying factor loadings implied by this model, and Figure 4.3 presents time-varying rank correlations. To summarize these results, Figure 4.2 averages the loadings across all firms in the same industry, and Figure 4.3 averages all pair-wise correlations between firms in the same pairs of industries. (Thus the plotted factor loadings and rank correlations are smoother than any individual rank correlation plot.) Also presented in Figure 4.3 are 60-day rolling window rank correlations, again averaged across pairs of the firms in the same pair of industries. This figure reveals a secular increase in the correlation between CDS spreads, rising from around 0.1 in 2006 to around 0.5 in 2013. Interestingly, rank correlations do *not* appear to spike during the financial crisis, unlike individual volatilities and probabilities of default; rather they continue a mostly steady rise through the sample period.

4.5 Time-varying systemic risk

In this section we use the dynamic multivariate model presented above to obtain estimates of measures of systemic risk. A variety of measures of systemic risk have been proposed in the literature to date. One influential measure is "CoVaR," proposed by Adrian and Brunnermeier (2009), which uses quantile regression to estimate the lower tail (e.g., 0.05) quantile of market returns conditional on a given firm having a returns equal to its lower tail quantile. This measure provides an estimate of how firm-level stress spills over to the market index. An alternative measure is "marginal expected shortfall" (MES) proposed by Brownlees and Engle (2011), which estimates the expected return on a firm conditional on the market return being below some low threshold. Segoviano and Goodhart (2009) and Giesecke and Kim (2009) propose measuring systemic risk via the probability that a "large" number of firms are in distress. Lucas, et al. (2011) use the same measure applied to European sovereign debt. Huang, *et al.* (2009) suggest using the price of a hypothetical contract insuring against system-wide distress, valued using a mix of CDS and equity data, as a measure of systemic risk. Schwaab (2010) presents a review of these and other measures of systemic risk.

We consider two different estimates of systemic risk, defined in detail in the following two sub-sections. In all cases we use the dynamic copula model that performed best in the previous section, namely the heterogeneous dependence factor copula model.

4.5.1 Joint probability of distress

The first measure of systemic risk we implement is an estimate of the probability that a large number of firms will be in distress, similar to the measure considered by Segoviano and Goodhart (2009), Giesecke and Kim (2009) and Lucas, et al. (2011). We define distress as a firm's one-year-ahead CDS spread lying above some threshold:

$$D_{i,t+250} \equiv \mathbf{1}\left\{S_{i,t+250} > c^*\right\} \tag{4.17}$$

We choose the threshold as the cross-sectional average of the 99% quantiles of the individual CDS spreads:

$$c^* = \frac{1}{N}\sum_{i=1}^{N} c_i^* \tag{4.18}$$

$$\text{where} \quad \Pr\left[S_{it} \leqslant c_i^*\right] = 0.99$$

In our sample, the 99% threshold corresponds to a CDS spread of 339 basis points. Using equation (4.11) above, this threshold yields an implied probability of default (assuming LGD is 0.6) of 5.7%. (The average CDS spread across all firms is 97 basis points, yielding an implied PD of 1.6%.) We also considered a threshold quantile of 0.95, corresponding to a CDS spread of 245 basis points, and the results are qualitatively similar.

We use the probability of a large proportion of firms being in distress as a measure of systemic risk. Define the "joint probability of distress" as:

$$JPD_{t,k} \equiv \Pr_t\left[\left(\frac{1}{N}\sum_{i=1}^{N} D_{i,t+250}\right) \geqslant \frac{k}{N}\right] \tag{4.19}$$

where k is a user-chosen threshold for what constitutes a "large" proportion of the N firms. We use $k = 30$, and the results corresponding to $k = 20$ and $k = 40$ are qualitatively similar.

With a fixed threshold for distress, such as that in equation (4.18), the average *individual* probability of distress will vary through time. It may thus be of interest, given our focus on *systemic* risk, to consider a scaled version of the JPD, to remove the influence of time variation in individual probabilities of distress. To this end, define:

$$SJPD_{t,k} \equiv \frac{JPD_{t,k}}{AvgIPD_t} \tag{4.20}$$

$$\text{where} \quad AvgIPD_t \equiv \frac{1}{N}\sum_{i=1}^{N} E_t\left[D_{i,t+250}\right] \tag{4.21}$$

The JPD and SJPD estimates must be obtained via simulations from our model, and we obtain these using 10,000 simulations. Given the computational burden, we compute estimates only every 20 trading days (approximately once per month).

The estimated joint probability of distress and scaled joint probability of distress are presented in Figure 4.4. We see from the left panel that the JPD rose dramatically during the financial crisis of late 2008–mid 2009, with the probability of at least 30 firms being in distress reaching around 80%. This panel also reveals that a large part of this increase in JPD is attributable to an increase in the average individual probability of distress, which rose to nearly 50% in the peak of the financial crisis.

In the right panel we report the ratio of these two lines and obtain the scaled probability of distress. This can be thought of as a "multiplier" of individual distress, as it shows the ratio of joint distress to average individual distress. This ratio reached nearly two in the financial crisis. Interestingly, while this ratio fell in late 2009, it rose again in 2010 and in late 2011, indicating that the level of systemic risk implied by observed CDS spreads is substantially higher now than in the pre-crisis period.

4.5.2 Expected proportion in distress

Our second measure of systemic risk more fully exploits the ability of our dynamic copula model to capture heterogeneous dependence between individual CDS spread changes. For each firm i, we compute the expected proportion of stocks in distress conditional on firm i being in distress:

$$EPD_{i,t} \equiv E_t\left[\frac{1}{N}\sum_{j=1}^{N} D_{j,t+250}\middle| D_{i,t+250}=1\right] \tag{4.22}$$

The minimum value this can take is $1/N$, as we include firm i in the sum, and the maximum is one. We use the same indicator for distress as in the previous section (equation (4.17)). This measure of systemic risk is similar in spirit to the CoVaR measure proposed by Adrian and Brunnermeier (2009), in that it looks at distress "spillovers" from a single firm to the market as a whole.

In Figure 4.5 below we summarize the results from the EPD estimates, and present the average, and 10% and 90% quantiles of this measure across the 100 firms in our sample. We observe that the average EPD is around 30% in the pre-crisis period, rising to almost 60% in late 2008, and returning to around 40% in the last year of our sample. Thus this figure, like the JPD and SJPD plot in Figure 4.4, is also suggestive of a large increase in systemic risk around the financial crisis, and higher level of systemic risk in the current period than in the pre-crisis period.

The expected proportion in distress measure enables us to identify firms that are more strongly correlated with market-wide distress than others. When the EPD is low for a given firm, it reveals that distress for that firm is not a signal of widespread distress, i.e., firm i is more idiosyncratic. Conversely, when the EPD is high, it reveals that distress for this firm is a sign of widespread distress, and so this firm is a "bellwether" for systemic risk. To illustrate the information from individual firm EPD estimates, Table 4.6 below presents the top five and bottom five firms according to their EPD on three dates in our sample period, the first day (January 2, 2006), a middle day (January 26, 2009) and the last day (April 17, 2012). We note that SLM Corporation ("Sallie Mae", in the student loan business) appears in the "least systemic" group on all three dates, indicating that periods in which it is in distress are, according to our model, generally unrelated to periods of wider distress. Marsh and McLennan (which owns a collection of risk, insurance and consulting firms) and Baxter International (a bioscience and medical firm) each appear in the "most systemic" group for two out of three dates.

Table 4.6 also provides information on the spread of EPD estimates across firms. At the start of our sample the least systemic firms had EPDs of 2 to 3, indicating that only one to two other firms are expected to be in distress when they are in distress. At the end of our sample the least systemic firms had EPDs of 8 to 12, indicating a wider correlation of distress even among the least correlated. A similar finding is true for the most systemic firms: the EPDs for the most systemic firms rise from 48–53 at the start of the sample to 84–94 at the end. Thus there is a general increase in the correlation between firm distress over this sample period.

4.6 Conclusion

Motivated by the growing interest in measures of the risk of systemic events, this paper proposes new flexible yet parsimonious models for time-varying high dimension distributions. We use copula theory to combine well-known models for conditional means, variances, and marginal distributions with new models of the conditional dependence structure (copula) to obtain dynamic joint distributions. Our proposed new dynamic copula models draw on successful ideas from the literature on dynamic modeling of high dimension covariance matrices, see Engle (2002) and Engle and Kelly (2012) for example, and on recent work on models for general time-varying distributions, see Creal, *et al.* (2011, 2013), along with the "factor copula" of Oh and Patton (2012).

We use the proposed models to undertake a detailed study of a collection of 100 credit default swap (CDS) spreads on U.S. firms, which provide an relatively novel view of the health of these firms. We find, unsurprisingly, that systemic risk was highest during the financial crisis of 2008–09. More interestingly, we also find that systemic risk has remained relatively high, and is substantially higher now than in the pre-crisis period.

4.7 Tables and figures

Table 4.1: Simulation results

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
				Panel A: Equidependence			
ω	0.000	0.005	0.015	0.001	0.027	-0.003	0.030
α	0.050	0.000	0.003	0.050	0.051	0.048	0.003
β	0.980	0.002	0.004	0.980	0.989	0.979	0.010
ν^{-1}	0.200	0.001	0.006	0.200	0.206	0.195	0.010
ψ_z	0.100	0.005	0.017	0.100	0.118	0.097	0.021
				Panel B: Block equidependence			
ω_1	-0.030	0.000	0.005	-0.030	-0.025	-0.035	0.010
ω_2	-0.023	-0.001	0.004	-0.024	-0.020	-0.030	0.010
ω_3	-0.017	0.000	0.005	-0.017	-0.011	-0.023	0.012
ω_4	-0.010	0.000	0.004	-0.011	-0.005	-0.016	0.011
ω_5	-0.003	0.001	0.004	-0.002	0.004	-0.007	0.011
ω_6	0.003	0.001	0.004	0.004	0.009	0.000	0.009
ω_7	0.010	0.002	0.005	0.012	0.018	0.007	0.012
ω_8	0.017	0.001	0.005	0.017	0.025	0.012	0.013
ω_9	0.023	0.001	0.005	0.024	0.030	0.018	0.012
ω_{10}	0.030	0.003	0.006	0.033	0.040	0.024	0.015
α	0.050	0.001	0.005	0.051	0.057	0.045	0.012
β	0.980	-0.001	0.002	0.978	0.981	0.976	0.004
ν^{-1}	0.200	-0.005	0.008	0.196	0.202	0.184	0.018
ψ_z	0.100	0.004	0.025	0.103	0.138	0.071	0.068

Notes: This table presents results from the simulation study described in Section 4.3. Panel A contains results for the equidependence model and Panel B for the "block equidependence" model.

Table 4.1: Simulation results

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
				Panel C: Heterogeneous dependence			
ω_1	-0.030	0.004	0.017	-0.022	-0.005	-0.052	0.047
ω_5	-0.028	0.004	0.016	-0.020	-0.005	-0.046	0.041
ω_{10}	-0.025	0.002	0.016	-0.019	-0.005	-0.041	0.036
ω_{15}	-0.022	0.002	0.013	-0.019	-0.004	-0.043	0.039
ω_{20}	-0.019	0.002	0.011	-0.015	-0.003	-0.033	0.030
ω_{25}	-0.016	0.000	0.010	-0.014	-0.003	-0.030	0.027
ω_{30}	-0.012	0.001	0.008	-0.010	-0.002	-0.022	0.020
ω_{35}	-0.009	0.000	0.008	-0.008	-0.002	-0.020	0.018
ω_{40}	-0.006	-0.001	0.005	-0.006	-0.002	-0.015	0.014
ω_{45}	-0.003	-0.001	0.005	-0.003	0.000	-0.010	0.010
ω_{50}	0.000	-0.002	0.004	-0.002	0.001	-0.007	0.008
ω_{55}	0.003	-0.001	0.004	0.001	0.007	-0.003	0.010
ω_{60}	0.006	-0.002	0.005	0.003	0.010	0.000	0.010
ω_{65}	0.009	-0.002	0.006	0.005	0.013	0.000	0.013
ω_{70}	0.012	-0.004	0.007	0.007	0.017	0.001	0.016
ω_{75}	0.015	-0.004	0.008	0.009	0.019	0.002	0.017
ω_{80}	0.018	-0.004	0.009	0.012	0.026	0.002	0.024
ω_{85}	0.021	-0.006	0.011	0.014	0.032	0.002	0.030
ω_{90}	0.024	-0.006	0.012	0.016	0.036	0.003	0.033
ω_{95}	0.027	-0.006	0.014	0.018	0.040	0.004	0.036
ω_{100}	0.030	-0.007	0.016	0.021	0.040	0.004	0.036
α	0.050	-0.006	0.015	0.045	0.062	0.023	0.039
β	0.980	0.002	0.012	0.983	0.997	0.966	0.031
ν^{-1}	0.200	-0.002	0.009	0.199	0.209	0.186	0.023
ψ_z	0.100	0.008	0.032	0.111	0.152	0.064	0.088

Notes: This table presents results from the simulation study described in Section 4.3. Panel C contains results for the "heterogeneous dependence" model. In the interests of space, Panel C only reports every fifth intercept parameter (ω_i) rather than the complete set of 100 such parameters; the complete table is available in Appendix D.3.

Table 4.2: Summary statistics for daily CDS spreads and log-differences of daily CDS spreads

	Mean	5%	25%	Median	75%	95%
Panel A: Cross-sectional distribution of CDS spreads						
Mean	96.953	37.212	53.561	74.957	123.785	200.346
Std dev	69.950	17.344	27.245	47.508	84.336	180.618
1^{st} autocorr	0.996	0.992	0.995	0.997	0.998	0.998
Skewness	1.203	0.095	0.695	1.280	1.587	2.488
Kurtosis	5.113	2.198	2.943	4.937	6.477	9.486
5%	23.883	9.021	11.741	18.926	29.851	60.538
25%	42.274	20.373	25.212	35.314	47.473	104.704
Median	85.310	35.098	50.105	69.399	113.762	166.208
75%	122.061	46.250	65.862	93.622	154.729	251.112
95%	245.497	72.514	102.554	168.500	313.585	631.924
99%	338.676	80.414	122.885	231.295	435.224	827.098
Panel B: Cross-sectional distribution of log-differences of CDS spreads						
Mean	5.589	-1.634	2.559	5.529	8.521	13.817
Std dev	378.892	308.636	347.627	373.460	400.385	476.533
1^{st} autocorr	0.161	0.030	0.121	0.164	0.217	0.267
Skewness	1.087	-0.285	0.354	0.758	1.488	3.629
Kurtosis	25.531	7.717	10.286	14.557	25.911	74.843
5%	-514.574	-622.282	-551.334	-509.554	-474.027	-415.651
25%	-144.195	-172.319	-155.635	-145.415	-134.820	-111.993
Median	-2.324	-9.045	-3.644	-0.726	0.000	0.000
75%	132.127	95.168	120.514	131.019	144.363	174.645
95%	570.510	457.775	537.093	568.331	612.769	684.984

Panel C: Autocorrelation in CDS spreads

# of reject.	Level	Log-diff	Squared log-diff
ADF test	12	100	–
LB test	–	98	89

Notes: This table presents summary statistics of daily CDS spreads (upper panel) and log-differences of CDS spreads (middle panel), measured in basis points in both cases. The columns present the mean and quantiles from the cross-sectional distribution of the measures listed in the rows. These two panels present summaries across the N = 100 marginal distributions. The bottom panel shows the number of rejections (at the 0.05 level) across the 100 firms for augmented Dickey-Fuller tests of the null of a unit root, as well as Ljung-Box tests for autocorrelation up to 10 lags.

Table 4.3: Marginal distribution parameter estimates

	Cross-sectional distribution					
	Mean	5%	25%	Median	75%	95%
ϕ_0	3.029	-3.760	0.247	3.116	5.861	10.165
ϕ_1	0.005	-0.179	-0.062	0.010	0.082	0.153
ϕ_2	0.025	-0.039	-0.001	0.025	0.050	0.084
ϕ_3	-0.002	-0.058	-0.028	-0.004	0.021	0.064
ϕ_4	0.006	-0.046	-0.014	0.006	0.033	0.054
ϕ_5	0.004	-0.055	-0.022	0.005	0.027	0.060
ϕ_m	0.387	0.163	0.303	0.372	0.480	0.638
$\omega/1000$	5.631	1.401	3.111	5.041	7.260	13.381
β	0.741	0.595	0.699	0.746	0.794	0.845
α	0.114	0.052	0.087	0.106	0.141	0.181
δ	0.022	0.000	0.000	0.000	0.042	0.086
α_m	0.223	0.037	0.137	0.206	0.297	0.494
δ_m	0.072	0.000	0.000	0.059	0.114	0.233
ν	3.620	2.877	3.293	3.571	3.921	4.496
ψ	0.043	-0.003	0.024	0.042	0.062	0.089

	# of rejections
LB test for standardized residuals	9
LB test for squared standardized residuals	2
KS test for skew t dist of std. residuals	11

Notes: The table presents summaries of the estimated AR(5)-GJR-GARCH(1,1)-Skew $t(\nu, \psi)$ models estimated on log-difference of daily CDS spreads. The columns present the mean and quantiles from the cross-sectional distribution of the parameters listed in the rows. The bottom panel shows the number of rejections (at the 0.05 leve) across 100 firms from Ljung-Box tests for serial correlation up to 10 lags. The first row is for standardized residuals of log-difference of daily CDS spreads and the second row for squared standardized residuals. The bottom panel shows the number of rejections across 100 firms from the Kolmogorov–Smirnov test of the Skew $t(\nu, \psi)$ distribution used for the standardized residuals.

Table 4.4: Model estimation results

	Equidep				Block equidep				Heterog			
	Normal		Factor		Normal		Factor		Normal		Factor	
	Pre	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
$\omega_{1\to G}$	-0.013 (0.047)	-0.027 (0.033)	0.005 (0.009)	-0.010 (0.017)	–	–	–	–	–	–	–	–
α_{GAS}	0.024 (0.009)	0.022 (0.003)	0.026 (0.003)	0.026 (0.003)	0.036 (0.015)	0.029 (0.001)	0.033 (0.011)	0.026 (0.010)	0.160 (0.064)	0.144 (0.062)	0.170 (0.022)	0.171 (0.015)
β_{GAS}	0.988 (0.060)	0.847 (0.070)	0.994 (0.026)	0.907 (0.035)	0.996 (0.044)	0.976 (0.042)	0.998 (0.015)	0.992 (0.031)	0.994 (0.047)	0.975 (0.065)	0.994 (0.031)	0.982 (0.040)
ν_z^{-1}	–	–	0.095 (0.025)	0.078 (0.031)	–	–	0.010 (0.029)	0.010 (0.020)	–	–	0.010 (0.027)	0.020 (0.021)
ν_ϵ^{-1}	–	–	0.170 (0.009)	0.178 (0.010)	–	–	0.206 (0.011)	0.190 (0.009)	–	–	0.204 (0.014)	0.179 (0.010)
ψ_z	–	–	-0.018 (0.049)	-0.015 (0.049)	–	–	0.088 (0.093)	0.093 (0.056)	–	–	0.091 (0.074)	0.124 (0.058)
$\log \mathcal{L}$	38,395		40,983		38,518		41,165		39,361		41,913	
AIC	-76,779		-81,943		-77,009		-82,291		-78,314		-83,405	
BIC	-76,747		-81,878		-76,933		-82,182		-77,211		-82,270	

Notes: This table presents parameter estimates for two versions of the factor copula (Normal and Skew t-t), each with one of three degrees of heterogeneity of dependence (equidependence, block equidependence, and heterogeneous dependence). Standard errors based on the stationary bootstrap of Politis and Romano (1994) are presented below the estimated parameters. All models are allowed to have a structural break on April 8, 2009 (see Section 4.4.4), and we denote parameters from the first and second sub-samples as "Pre" and "Post." The log-likelihood at the estimated parameters and the Akaike and Bayesian Information criteria are presented in the bottom three rows. The intercept parameters (ω_i) for the block equidependence and heterogeneous dependence models are not reported to conserve space.

Table 4.5: Model comparison results

		Normal		**Factor** (ν, ψ_z, ν)		**Factor** $(\nu_z, \psi_z, \nu_\varepsilon)$	
		Static	**GAS**	**Static**	**GAS**	**Static**	**GAS**
	# param	2	6	6	10	8	12
Equi-	log $\mathcal{L}$	36,185	38,395	39,223	40,688	39,508	**40,983**
dependence	AIC	-72,369	-76,785	-78,441	-81,367	-79,007	**-81,955**
	BIC	-72,363	-76,769	-78,424	-81,340	-78,986	**-81,922**
Block	# param	10	14	14	18	16	20
equi-	log $\mathcal{L}$	36,477	38,518	39,441	40,799	39,757	**41,165**
dependence	AIC	-72,944	-77,023	-78,867	-81,580	-79,497	**-82,311**
	BIC	-72,917	-76,985	-78,829	-81,532	-79,454	**-82,257**
	# param	200	204	204	208	206	210
Heterogeneous	log $\mathcal{L}$	37,652	39,361	40,357	41,522	40,628	**41,913**
dependence	AIC	-75,105	-78,518	-80,511	-82,835	-81,050	**-83,615**
	BIC	-74,564	-77,966	-79,959	-82,273	-80,494	**-83,048**

Notes: This table presents the log-likelihood at the estimated parameters, as well as the Akaike and Bayesian Information criteria, for a variety of copula models. The preferred model according to each of these criteria is highlighted in bold. Also presented is the number of estimated parameters; note that this accounts for the fact that we allow for a structural break in these parameters, and so the number reported is twice as large as it would be in the absence of a break. We consider models with three degrees of heterogeneity of dependence (equidependence, block equidependence, and heterogeneous dependence); with and without dynamics (static and GAS); and three versions of the factor copula (Normal, Skew *t*-*t* with a common degrees of freedom parameter, and Skew *t*-*t* with separately estimated degrees of freedom parameters).

Table 4.6: Estimates of systemic risk

	2 January 2006		*26 January 2009*		*17 April 2012*	
	EPD	**Firm**	**EPD**	**Firm**	**EPD**	**Firm**
Most systemic	53	Marsh & McLennan	78	Lockheed Martin	94	Wal-Mart
2	50	Hewlett-Packard	77	Campbell Soup	88	Baxter Int'l
3	50	IBM	75	Marsh & McLennan	88	Walt Disney
4	49	Valero Energy	75	Baxter Int'l	87	Home Depot
5	48	Bristol-Myers Squibb	74	Goodrich	84	McDonald's
96	3	Aetna	35	Vornado Realty	12	MetLife
97	3	Xerox	34	Gen Elec Capital	11	The GAP
98	3	Comp Sci Corp	34	Johnson Controls	11	Sallie Mae
99	3	Sallie Mae	34	Alcoa	11	Comp Sci Corp
Least systemic	2	Freeport-McMoRan	33	Sallie Mae	8	Pitney Bowes

Note: This table presents the five firms with the largest estimated expected proportion in distress (EPD), defined in equation (4.22), and the five firms with the smallest EPD, for three dates in our sample period.

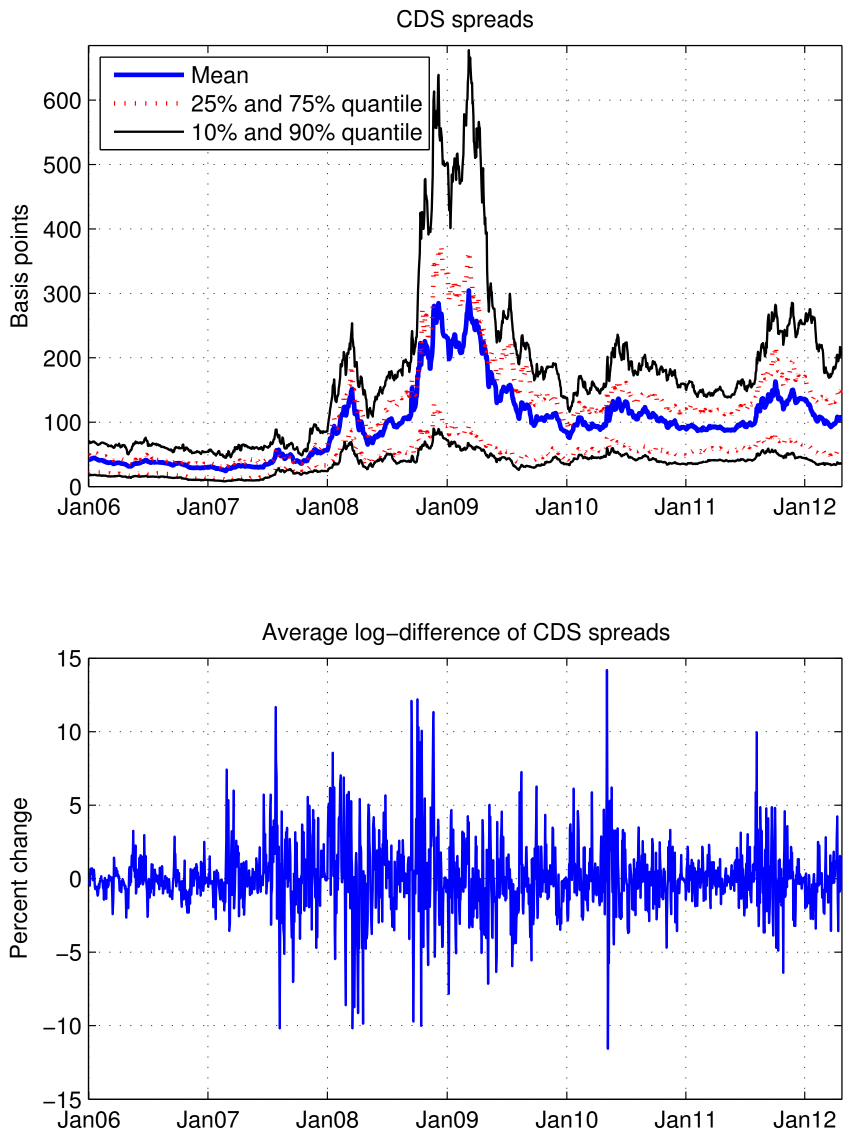


Figure 4.1: *The upper panel plots the mean and 10%, 25%, 75% and 90% quantiles across the CDS spreads for 100 U.S. firms over the period January 2006 to April 2012. The lower panel reports the average (across firms) percent change in CDS spreads for the same time period.*

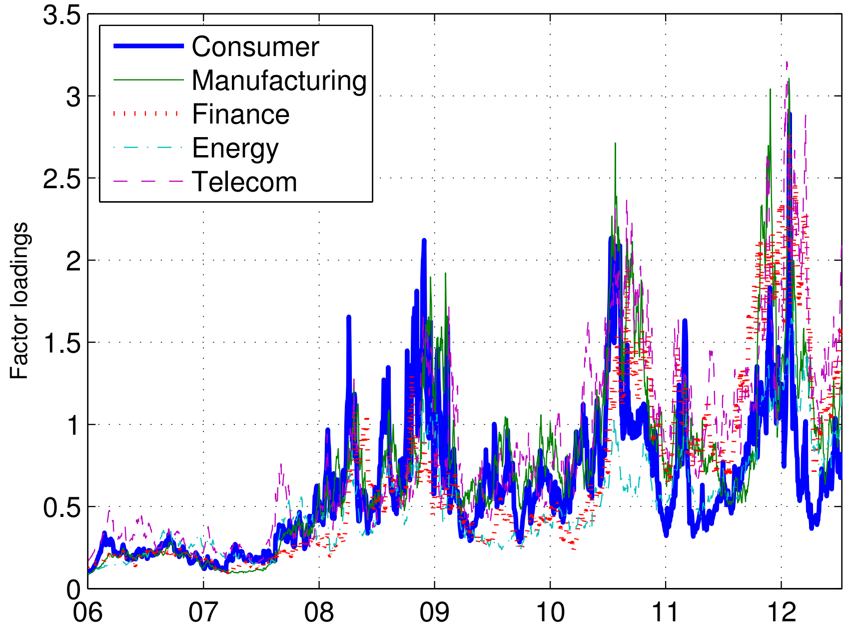


Figure 4.2: *This figure plots the estimated factor loadings* (λ_t) *from the heterogeneous dependence factor copula model, averaged across firms in the same industry.*

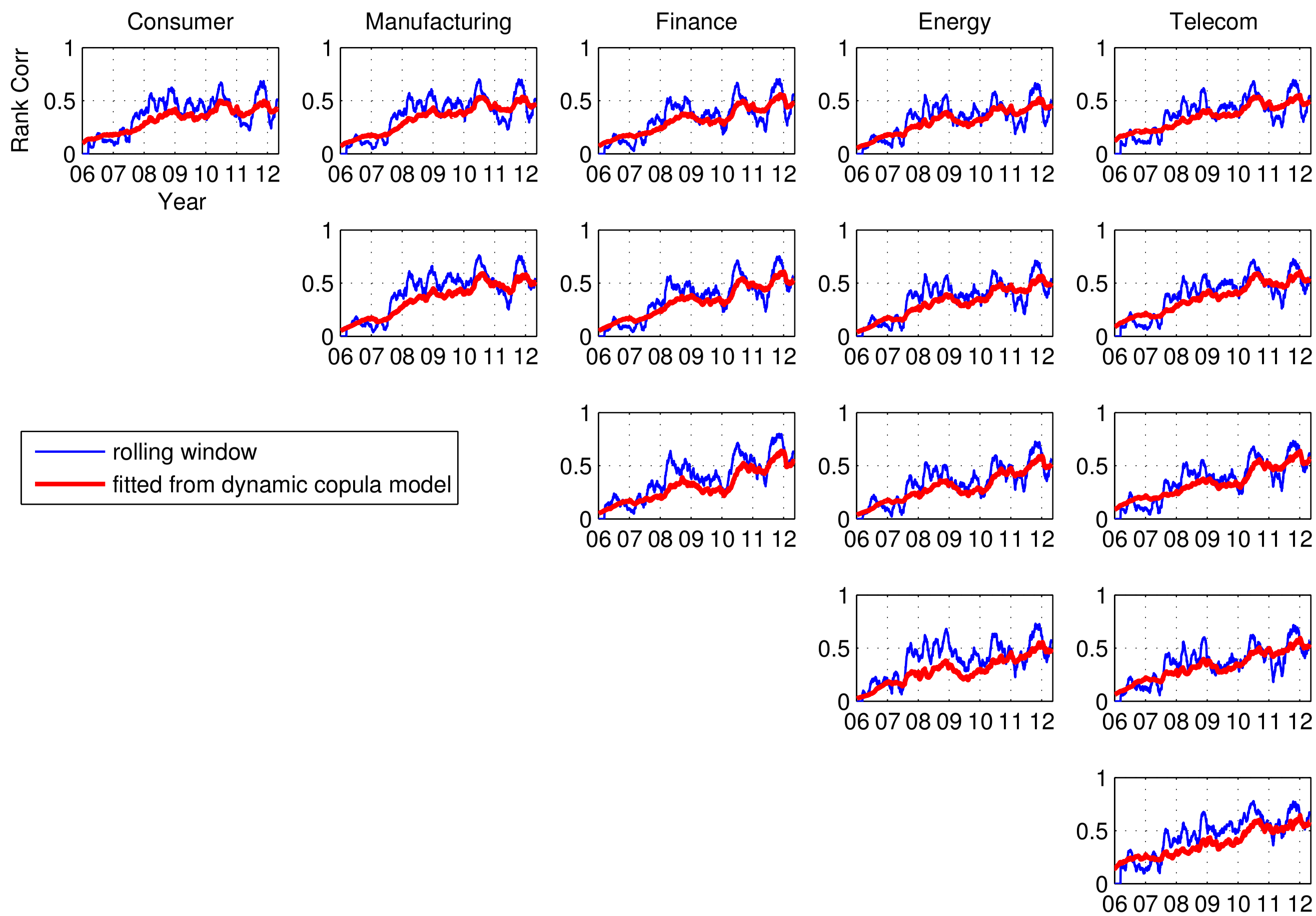


FIGURE 4.3: *This figure plots the time-varying rank correlations implied by the heterogeneous dependence factor copula, as well as 60-day rolling window rank correlations. Each element of this figure is an average of all pair-wise correlations for firms in the same pair of industries.*

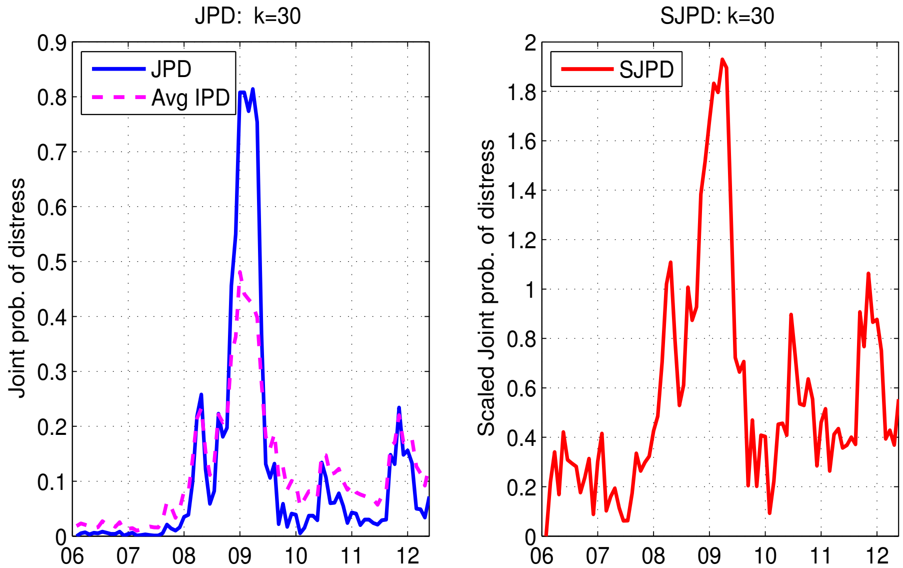


FIGURE 4.4: *The left panel shows the joint probability of distress (JPD) in a solid line and the average individual probability of distress (Avg IPD) in a dashed line. The right panel shows the scaled joint probability of distress (SJPD). Both panels cover the period January 2006 to April 2012.*

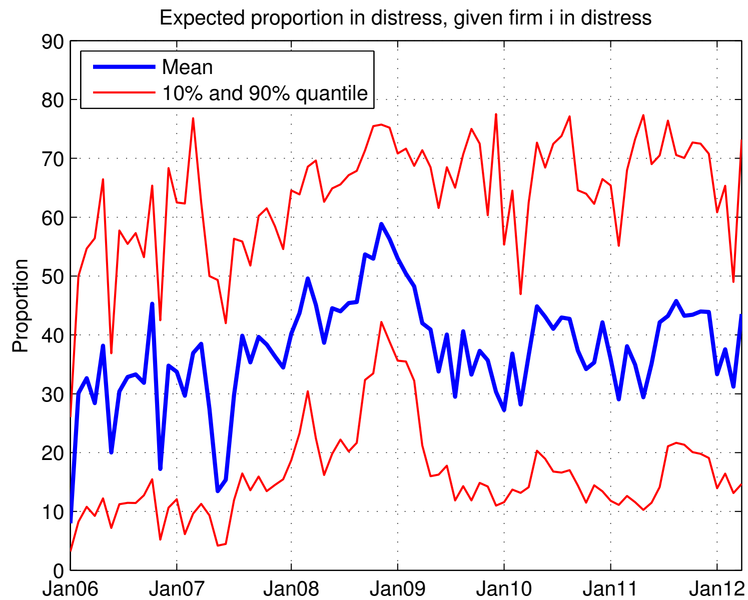


FIGURE 4.5: *This figure shows the expected proportion (in percent) of firms in distress, given firm i in distress, averaged across all 100 firms. The cross-sectional 10% and 90% quantiles are also reported. The sample period is January 2006 to April 2012.*

Appendix A

Appendix to Chapter 1

A.1 Proofs

The following two lemmas are needed to prove Theorem 1.

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

$$\forall i,\ \mathbf{F}\left(a_1 + x_1, \ldots, a_i + x_i, \ldots, a_N + x_N\right) = \mathbf{F}\left(a_1 + x_1, \ldots, \infty, \ldots, a_N + x_N\right) - \mathbf{F}\left(a_1 + x_1, \ldots, a_i - x_i, \ldots, a_N + x_N\right) \tag{A.1}$$

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

Proof. $\Rightarrow$) By Definition 2, the joint symmetry implies that the following holds for any i,

$$\begin{aligned} &\Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i - a_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \\ &= \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, a_i - X_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \end{aligned} \tag{A.2}$$

and with a simple calculation, the right hand side is written as

$$\begin{aligned}
&\Pr\left[X_1 - a_1 \leqslant x_1, \ldots, a_i - X_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \\
&= \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i \leqslant \infty, \ldots, X_N - a_N \leqslant x_N\right] \\
&- \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i \leqslant a_i - x_i, \ldots, X_N - a_N \leqslant x_N\right] \\
&= \mathbf{F}\left(a_1 + x_1, \ldots, \infty, \ldots, a_N + x_N\right) \\
&- \mathbf{F}\left(a_1 + x_1, \ldots, a_i - x_i, \ldots, a_N + x_N\right)
\end{aligned} \tag{A.3}$$

and the left hand side of equation (A.2) is

$$\begin{aligned}
&\Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i - a_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \\
&= \mathbf{F}\left(a_1 + x_1, \ldots, a_i + x_i, \ldots, a_N + x_N\right)
\end{aligned}$$

$\Leftarrow$) Equation (A.1) can be written as

$$\begin{aligned}
&\forall i,\;\; \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i - a_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \\
&= \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i \leqslant \infty, \ldots, X_N - a_N \leqslant x_N\right] \\
&- \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i \leqslant a_i - x_i, \ldots, X_N - a_N \leqslant x_N\right]
\end{aligned}$$

and by equation (A.3), the right hand side becomes

$$\Pr\left[X_1 - a_1 \leqslant x_1, \ldots, a_i - X_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right]$$

Therefore, for any i

$$\begin{aligned}
&\Pr\left[X_1 - a_1 \leqslant x_1, \ldots, X_i - a_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right] \\
&= \Pr\left[X_1 - a_1 \leqslant x_1, \ldots, a_i - X_i \leqslant x_i, \ldots, X_N - a_N \leqslant x_N\right]
\end{aligned}$$

and this satisfies the definition of joint symmetry. □

Lemma 2. *Consider two scalar random variables Z_1 and Z_2, and some constant b_1 in $\mathbf{R}^1$. If $(Z_1 - b_1, Z_2)$ and $(b_1 - Z_1, Z_2)$ have a common joint distribution, then $Cov\left(Z_1, Z_2\right) = 0$*

Proof. $Z_1 - b_1$ and $b_1 - Z_1$ have the same marginal distribution and the same moments, so

$$E[Z_1 - b_1] = E[b_1 - Z_1]$$

$$E[Z_1] = b_1$$

$(Z_1 - b_1, Z_2)$ and $(b_1 - Z_1, Z_2)$ have the same moment, so

$$E[(Z_1 - b_1) Z_2] = E[(b_1 - Z_1) Z_2]$$

$$E[Z_1 Z_2] = b_1 E[Z_2]$$

Covariance of Z_1 and Z_2 is

$$\begin{aligned} Cov(Z_1, Z_2) &= E[Z_1 Z_2] - E[Z_1] E[Z_2] \\ &= b_1 E[Z_2] - b_1 E[Z_2] \\ &= 0 \end{aligned}$$

□

Proof of Theorem 1. (i) ⇒) We follow Lemma 1 and rewrite equation (A.1) as

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

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,

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

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

⇐) Following the reverse way above, equation (1.5) becomes equation (A.1), and the proof is done by Lemma 1.

(ii) This is trivial by Definition 2 for joint symmetry and Lemma 2. □

Proof of Theorem 2. If we prove that $\mathbf{C}^{JS}(u_1, \ldots, u_N)$ in equation (1.6) satisfies equation (1.5), the proof is done:

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

We first prove this for $i = N$. Rewriting equation (1.6) as

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

$$\textit{where } \mathbf{C}_{-N}(u_1, \ldots, u_{N-1}, u_N) = \sum_{j_1=1}^{3} \cdots \sum_{j_{N-1}=1}^{3} (-1)^{J_{-N}} \cdot \mathbf{C}(\tilde{u}_1, \ldots, \tilde{u}_{N-1}, u_N)$$

$$J_{-N} \equiv \sum_{i=1}^{N-1} 1\{j_i = 2\} \text{ and } \tilde{u}_i = \begin{cases} u_i & \text{for } j_i = 1 \\ 1-u_i & \text{for } j_i = 2 \\ 1 & \text{for } j_i = 3 \end{cases}$$

and calculating $\mathbf{C}^{JS}(u_1, \ldots, u_{N-1}, 1)$ and $\mathbf{C}^{JS}(u_1, \ldots, u_{N-1}, 1-u_N)$ result in

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

Similarly, this equation holds for $i = 1, \ldots, N-1$, so the proof is done. □

To prove Theorem 3, we need the following lemma which guarantees that a sample covariance matrix is positive definite if T is large enough relative to dimension N.

Lemma 3. *Consider T vectors $\mathbf{r}_t \in \mathbf{R}^N$ for $t = 1, \ldots, T$. If rank$[\mathbf{r}_1, \ldots, \mathbf{r}_T] \geqslant N$, then $\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}_t'$ is positive definite.*

Proof. Assume that $\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}_t'$ is positive semi-definite. Then there exists a nonzero vector $\mathbf{x} \in \mathbf{R}^{\mathbf{N}}$ such that $\mathbf{x}' \left(\sum_{t=1}^{T} \mathbf{r}_t \mathbf{r}_t' \right) \mathbf{x} = 0$, and this implies $\mathbf{x}' \cdot \mathbf{r}_t = 0$ for any t. On the other hand, if rank$[\mathbf{r}_1, \ldots, \mathbf{r}_T] \geqslant N$, then $[\mathbf{r}_1, \ldots, \mathbf{r}_T]$ span $\mathbf{R}^{\mathbf{N}}$, which implies there exist $\{\alpha_i\}_{i=1}^{T}$ such that

$$\alpha_1 \mathbf{r}_1 + ... + \alpha_T \mathbf{r}_T = \mathbf{x}$$

Premultiplying by $\mathbf{x}'$ gives

$$\alpha_1 \mathbf{x}' \mathbf{r}_1 + ... + \alpha_T \mathbf{x}' \mathbf{r}_T = \mathbf{x}' \mathbf{x}$$

and the left hand side is zero by $\mathbf{x}' \cdot \mathbf{r}_t = 0$ for any t, which contradicts that $\mathbf{x}$ is a nonzero vector. □

Proof of Theorem 3. (i) From equation (1.11), we write the conditional expectation of realized correlation in matrix form:

$$\begin{aligned} E\left[RCorr_t^{\Delta} | \mathcal{F}_{t-1}\right] &= (1 - a - b - c) \cdot E\left[RCorr_t^{\Delta}\right] + a \cdot RCorr_{t-1}^{\Delta} \\ &\quad + b \cdot \frac{1}{4} \sum_{k=2}^{5} RCorr_{t-k}^{\Delta} + c \cdot \frac{1}{15} \sum_{k=6}^{20} RCorr_{t-k}^{\Delta} \end{aligned}$$

$(1 - a - b - c) \cdot E\left[RCorr_t^{\Delta}\right]$ is positive definite by assumptions 1 and 3, and the other three terms are positive semi-definite by assumption 2 and positive semi-definiteness of $RCorr_{t-k}^{\Delta}, k = 1, ..., 20$. Given the fact that the sum of positive definite and positive semi-definite matrices is positive definite, $E\left[RCorr_t^{\Delta} | \mathcal{F}_{t-1}\right]$ is positive definite. Since $E\left[RVar_t^{\Delta} | \mathcal{F}_{t-1}\right]$ is diagonal matrix with positive elements, it is proven that $E\left[RVarCov_t^{\Delta} | \mathcal{F}_{t-1}\right]$ is positive definite.

(ii) The sample counterpart to $E\left[RCorr_t^{\Delta}|\mathcal{F}_{t-1}\right]$ is

$$\hat{E}\left[RCorr_t^{\Delta}|\mathcal{F}_{t-1}\right] = \left(1-\hat{a}-\hat{b}-\hat{c}\right)\cdot\frac{1}{T}\sum_{i=1}^{T}RCorr_i^{\Delta}+\hat{a}\cdot RCorr_{t-1}^{\Delta}$$

$$+\hat{b}\cdot\frac{1}{4}\sum_{k=2}^{5}RCorr_{t-k}^{\Delta}+\hat{c}\cdot\frac{1}{15}\sum_{k=6}^{20}RCorr_{t-k}^{\Delta}$$

where $\hat{a},\hat{b}$, and $\hat{c}$ satisfy assumptions 2 and 3. $\frac{1}{T}\sum_{i=1}^{T}RCorr_i^{\Delta}$ is positive definite by Lemma 3. The remaining proof is similar to the above proof for (i). □

Proof of Theorem 5. By applying $\log(y)\leqslant y-1$ to $\frac{h_i(Z_{i,t+1},Z_{i+1,t+1})}{g_i(Z_{i,t+1},Z_{i+1,t+1})}$, the following is shown

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

where the second line holds since only submodel for $(Z_{i,t+1},Z_{i+1,t+1})$ is needed to evaluate the above expectation, and the third line holds since h_i is a valid density. Thus, we prove that

$$E_{g(\mathbf{z})}\left[\sum_{i=1}^{N-1}\log h_i\left(Z_{i,t+1},Z_{i+1,t+1}\right)\right]\leqslant E_{g(\mathbf{z})}\left[\sum_{i=1}^{N-1}\log g_i\left(Z_{i,t+1},Z_{i+1,t+1}\right)\right]$$

□

A.2 Dynamic conditional correlation (DCC) model

The DCC model by Engle (2002) specifies the conditional covariance matrix $\mathbf{H}_t$ in equation (1.1) by

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{A.4}$$

$$\mathbf{D}_t = diag\left(\left\{\sqrt{\sigma_{i,t}^2}\right\}_{i=1}^{N}\right) \tag{A.5}$$

$$\sigma_{i,t}^2 = \psi_i + \kappa_i \left(r_{i,t-1} - \mu_{i,t-1}\right)^2 + \zeta_i \left(r_{i,t-1} - \mu_{i,t-1}\right)^2 1_{\{(r_{i,t-1}-\mu_{i,t-1})<0\}} + \lambda_i \sigma_{i,t-1}^2 \tag{A.6}$$

$$\varepsilon_t = \mathbf{D}_t^{-1}\left(r_{i,t} - \mu_{i,t}\right)$$

$$\mathbf{Q}_t = (1-\alpha-\beta)\,\overline{\mathbf{Q}} + \alpha\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) + \beta \mathbf{Q}_{t-1} \tag{A.7}$$

$$\mathbf{R}_t = diag\left(\mathbf{Q}_t\right)^{-1/2} \mathbf{Q}_t diag\left(\mathbf{Q}_t\right)^{-1/2} \tag{A.8}$$

The dynamics of each conditional variance is governed by GJR-GARCH, see Glosten, *et al.* (1993) and $\overline{\mathbf{Q}}$ is substituted with sample correlation matrix of ε_t as in Engle (2002). The restrictions $0 \leqslant \alpha, \beta \leqslant 1$ and $\alpha + \beta \leqslant 1$ are imposed for positive definiteness of $\mathbf{Q}_t$ and so $\mathbf{H}_t$. The number of parameters to estimate is $4N + N(N-1)/2 + 2$.

To estimate the DCC model, QMLE based on the quasi-likelihood using normality assumption is feasible as in Engle (2002)

$$\mathbf{r}_t = \mu_t + \mathbf{H}_t^{1/2}\mathbf{e}_t$$

$$\mathbf{e}_t|\mathcal{F}_{t-1} \sim \mathbf{N}\left(\mathbf{0}, \mathbf{I}_{N\times N}\right)$$

and the log likelihood for those estimator can be described as

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim \mathbf{N}\left(\mu_t, \mathbf{H}_t\right)$$

$$\log L = -\frac{1}{2}\sum_{t=1}^{T}\left(N\log(2\pi) + \log|H_t| + \left(\mathbf{r}_t - \mu_t\right)' \mathbf{H}_t^{-1}\left(\mathbf{r}_t - \mu_t\right)\right)$$

Engle, *et al.* (2008), however, indicate that when N is large, bias of estimators for α and β could be substantial due to the impact of estimation error from estimating large matrix $\overline{\mathbf{Q}}$ by sample correlations of ε_t, and they suggest the composite likelihood based estimator. Therefore we follow their estimation method rather than QMLE in Sections 1.4.2 and 1.5.

A.3 Hessian matrix for multistage estimations

The specific form of estimated Hessian matrix in Theorem 6 is following. For illustration purpose, we assume $N = 2$, but it is easy to extend to general N

$$\hat{P}_t = \begin{bmatrix} \bigtriangledown_{var1,var1}^{var} & 0 & 0 & 0 & 0 & 0 \\ 0 & \bigtriangledown_{var2,var2}^{var} & 0 & 0 & 0 & 0 \\ \bigtriangledown_{var1,corr}^{corr} & \bigtriangledown_{var2,corr}^{corr} & \bigtriangledown_{corr,corr}^{corr} & 0 & 0 & 0 \\ \bigtriangledown_{var1,mar1}^{mar} & \bigtriangledown_{var2,mar1}^{mar} & \bigtriangledown_{corr,mar1}^{mar} & \bigtriangledown_{mar1,mar1}^{mar} & 0 & 0 \\ \bigtriangledown_{var1,mar2}^{mar} & \bigtriangledown_{var2,mar2}^{mar} & \bigtriangledown_{corr,mar2}^{mar} & 0 & \bigtriangledown_{mar2,mar2}^{mar} & 0 \\ \bigtriangledown_{var1,cop}^{cop} & \bigtriangledown_{var2,cop}^{cop} & \bigtriangledown_{corr,cop}^{cop} & \bigtriangledown_{mar1,cop}^{cop} & \bigtriangledown_{mar2,cop}^{cop} & \bigtriangledown_{cop,cop}^{cop} \end{bmatrix}$$

$$\bigtriangledown_{var\_i,var\_i}^{var} = \frac{\partial^2}{\partial\theta_i^{var}\partial\theta_i^{var\prime}} \log l_{it}^{var}\left(\hat{\theta}_i^{var}\right), \quad i = 1, 2, \ldots, N,$$

$$\bigtriangledown_{var\_i,corr}^{corr} = \frac{\partial^2}{\partial\theta^{corr}\partial\theta_i^{var\prime}} \log l_t^{corr}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}\right)$$

$$\bigtriangledown_{corr,corr}^{corr} = \frac{\partial^2}{\partial\theta^{corr}\partial\theta^{corr\prime}} \log l_t^{corr}\left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}\right)$$

$$\bigtriangledown^{mar}_{var\_j,mar\_i} = \frac{\partial^2}{\partial \theta_i^{mar} \partial \theta_j^{var\prime}} \log l_{it}^{mar} \left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_i^{mar} \right), \quad i, j = 1, \ldots, N$$

$$\bigtriangledown^{mar}_{corr,mar\_i} = \frac{\partial^2}{\partial \theta_i^{mar} \partial \theta^{corr\prime}} \log l_{it}^{mar} \left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_i^{mar} \right), \quad i = 1, \ldots, N$$

$$\bigtriangledown^{mar}_{mar\_i,mar\_i} = \frac{\partial^2}{\partial \theta_i^{mar} \partial \theta_i^{mar\prime}} \log l_{it}^{mar} \left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_i^{mar} \right), \quad i = 1, \ldots, N$$

$$\bigtriangledown^{cop}_{var\_i,cop} = \frac{\partial^2}{\partial \theta^{cop} \partial \theta_i^{var\prime}} \log l_t^{cop} \left(\hat{\theta}_1^{var}, .., \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, .., \hat{\theta}_N^{mar}, \hat{\theta}^{cop} \right), \quad i = 1, .., N$$

$$\bigtriangledown^{cop}_{corr,cop} = \frac{\partial^2}{\partial \theta^{cop} \partial \theta^{corr\prime}} \log l_t^{cop} \left(\hat{\theta}_1^{var}, \ldots, \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, \ldots, \hat{\theta}_N^{mar}, \hat{\theta}^{cop} \right)$$

$$\bigtriangledown^{cop}_{mar\_i,cop} = \frac{\partial^2}{\partial \theta^{cop} \partial \theta_i^{mar\prime}} \log l_t^{cop} \left(\hat{\theta}_1^{var}, .., \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, .., \hat{\theta}_N^{mar}, \hat{\theta}^{cop} \right), \quad i = 1, .., N$$

$$\bigtriangledown^{cop}_{cop,cop} = \frac{\partial^2}{\partial \theta^{cop} \partial \theta^{cop\prime}} \log l_t^{cop} \left(\hat{\theta}_1^{var}, .., \hat{\theta}_N^{var}, \hat{\theta}^{corr}, \hat{\theta}_1^{mar}, \ldots, \hat{\theta}_N^{mar}, \hat{\theta}^{cop} \right)$$

Appendix B

Appendix to Chapter 2

B.1 Proofs

Proof of Proposition 1. Consider a simple case first: $\beta_1 = \beta_2 = \beta > 0$. This implies that $X_i \sim G$, for $i = 1, 2$, and so we can use the same threshold for both X_1 and X_2. Then the upper tail dependence coefficient is:

$$\tau^U = \lim_{s \to \infty} \frac{\Pr\left[X_1 > s, X_2 > s\right]}{\Pr\left[X_1 > s\right]}$$

From standard extreme value theory, see Hyung and de Vries (2007) for example, we have the probability of an exceedence by the sum as the sum of the probabilities of an exceedence by each component of the sum, as the exceedence threshold diverges:

$$\begin{aligned}
\Pr\left[X_i > s\right] &= \Pr\left[\beta Z + \varepsilon_i > s\right] \\
&= \Pr\left[\beta Z > s\right] + \Pr\left[\varepsilon_i > s\right] + o\left(s^{-\alpha}\right) \quad \text{as } s \to \infty \\
&\approx A_z^U \left(s/\beta\right)^{-\alpha} + A_\varepsilon^U s^{-\alpha} \\
&= s^{-\alpha}\left(A_z^U \beta^\alpha + A_\varepsilon^U\right)
\end{aligned}$$

Further, we have the probability of *two* sums of variables both exceeding some diverging threshold being driven completely be the common component of the sums:

$$\begin{aligned}\Pr\left[X_1 > s, X_2 > s\right] &= \Pr\left[\beta Z + \varepsilon_1 > s, \beta Z + \varepsilon_2 > s\right] \\ &= \Pr\left[\beta Z > s, \beta Z > s\right] + o\left(s^{-\alpha}\right) \quad \text{as } s \to \infty \\ &\approx s^{-\alpha} A_z^U \beta^\alpha\end{aligned}$$

So we have

$$\tau^U = \lim_{s\to\infty} \frac{s^{-\alpha} A_z^U \beta^\alpha}{s^{-\alpha}\left(A_z^U \beta^\alpha + A_\varepsilon^U\right)} = \frac{A_z^U \beta^\alpha}{A_z^U \beta^\alpha + A_\varepsilon^U}$$

(a) Now we consider the case that $\beta_1 \neq \beta_2$, and *wlog* assume $\beta_2 > \beta_1 > 0$. This complicates the problem as the thresholds, s_1 and s_2, must be set such that $G_1(s_1) = G_2(s_2) = q \to 1$, and when $\beta_1 \neq \beta_2$ we have $G_1 \neq G_2$ and so $s_1 \neq s_2$. We can find the link between the thresholds as follows:

$$\Pr\left[X_i > s\right] = \Pr\left[\beta_i Z + \varepsilon_i > s\right] \approx s^{-\alpha}\left(A_z^U \beta_i^\alpha + A_\varepsilon^U\right) \quad \text{for } s \to \infty$$

so find s_1, s_2 such that $s_1^{-\alpha}\left(A_z^U \beta_1^\alpha + A_\varepsilon^U\right) = s_2^{-\alpha}\left(A_z^U \beta_2^\alpha + A_\varepsilon^U\right)$, which implies:

$$s_2 = s_1\left(\frac{A_z^U \beta_2^\alpha + A_\varepsilon^U}{A_z^U \beta_1^\alpha + A_\varepsilon^U}\right)^{1/\alpha}$$

Note that s_1 and s_2 diverge at the same rate. Below we will need to know which of s_1/β_1 and s_2/β_2 is larger. Note that $\beta_2 > \beta_1$, which implies the following:

$$\begin{aligned}&\Rightarrow \beta_2^\alpha > \beta_1^\alpha \quad \text{since } x^\alpha \text{ is increasing for } x, \alpha > 0 \\ &\Rightarrow A_\varepsilon^U \beta_2^\alpha + A_z^U \beta_1^\alpha \beta_2^\alpha > A_\varepsilon^U \beta_1^\alpha + A_z^U \beta_1^\alpha \beta_2^\alpha \\ &\Rightarrow \left(\frac{\beta_2}{\beta_1}\right)^\alpha > \frac{A_\varepsilon^U + A_z^U \beta_2^\alpha}{A_\varepsilon^U + A_z^U \beta_1^\alpha} \Rightarrow \frac{\beta_2}{\beta_1} > \left(\frac{A_\varepsilon^U + A_z^U \beta_2^\alpha}{A_\varepsilon^U + A_z^U \beta_1^\alpha}\right)^{1/\alpha} = \frac{s_2}{s_1} \\ &\Rightarrow \frac{s_1}{\beta_1} > \frac{s_2}{\beta_2}\end{aligned}$$

Then the denominator of the tail dependence coefficient is $\Pr[X_i > s_i] \approx s_i^{-\alpha}\left(A_z^U \beta_i^\alpha + A_\varepsilon^U\right)$, and the numerator becomes:

$$\begin{aligned}
\Pr[X_1 > s_1, X_2 > s_2] &= \Pr[\beta_1 Z + \varepsilon_1 > s_1, \beta_2 Z + \varepsilon_2 > s_2] \\
&\approx \Pr[\beta_1 Z > s_1, \beta_2 Z > s_2] \quad \text{as } s_1, s_2 \to \infty \\
&= \Pr[Z > \max\{s_1/\beta_1, s_2/\beta_2\}] \\
&= \Pr[Z > s_1/\beta_1] = s_1^{-\alpha} A_z^U \beta_1^\alpha
\end{aligned}$$

Finally, using either $\Pr[X_1 > s_1]$ or $\Pr[X_2 > s_2]$ in the denominator we obtain:

$$\tau^U = \frac{s_1^{-\alpha} A_z^U \beta_1^\alpha}{s_1^{-\alpha}\left(A_z^U \beta_1^\alpha + A_\varepsilon^U\right)} = \frac{\beta_1^\alpha A_z^U}{\beta_1^\alpha A_z^U + A_\varepsilon^U}, \text{ as claimed.}$$

(b) Say $\beta_2 < \beta_1 < 0$. Then:

$$\begin{aligned}
\Pr[X_i > s] &= \Pr[\beta_i Z + \varepsilon_i > s] \\
&\approx \Pr[\beta_i Z > s] + \Pr[\varepsilon_i > s] \quad \text{for } s \to \infty \\
&= \Pr[|\beta_i|(-Z) > s] + \Pr[\varepsilon_i > s] \\
&= s^{-\alpha}\left(A_z^L |\beta_i|^\alpha + A_\varepsilon^U\right)
\end{aligned}$$

Next we find the thresholds s_1, s_2 such that $\Pr[X_1 > s_1] = \Pr[X_2 > s_2]$:

$$s_1^{-\alpha}\left(A_z^L |\beta_1|^\alpha + A_\varepsilon^U\right) = s_2^{-\alpha}\left(A_z^L |\beta_2|^\alpha + A_\varepsilon^U\right)$$

$$\text{so} \quad s_2 = s_1 \left(\frac{A_z^L |\beta_2|^\alpha + A_\varepsilon^U}{A_z^L |\beta_1|^\alpha + A_\varepsilon^U}\right)^{1/\alpha}$$

Using the same steps as for part (a), we find that $s_2 > s_1$ but $s_1/|\beta_1| > s_2/|\beta_2|$.

Thus the numerator becomes:

$$\begin{aligned}
\Pr\left[X_1 > s_1, X_2 > s_2\right] &= \Pr\left[\beta_1 Z + \varepsilon_1 > s_1, \beta_2 Z + \varepsilon_2 > s_2\right] \\
&\approx \Pr\left[\beta_1 Z > s_1, \beta_2 Z > s_2\right] \quad \text{for } s_1, s_2 \to \infty \\
&= \Pr\left[\left|\beta_1\right|(-Z) > s_1, \left|\beta_2\right|(-Z) > s_2\right] \\
&= \Pr\left[(-Z) > \max\left\{s_1/\left|\beta_1\right|, s_2/\left|\beta_2\right|\right\}\right] \\
&= \Pr\left[(-Z) > s_1/\left|\beta_1\right|\right] = A_z^L s_1^{-\alpha}\left|\beta_1\right|^{\alpha} \\
\text{so } \tau^U &= \frac{\left|\beta_1\right|^{\alpha} A_z^L}{\left|\beta_1\right|^{\alpha} A_z^L + A_\varepsilon^U}
\end{aligned}$$

(c) If β_1 or β_2 equal zero, then the numerator of the upper tail dependence coefficient limits to zero faster than the denominator. Say $\beta_2 > \beta_1 = 0$:

$$\begin{aligned}
\Pr\left[X_1 > s_1\right] &= s_1^{-\alpha}\left(A_z^U \beta_1^{\alpha} + A_\varepsilon^U\right) = s_1^{-\alpha} A_\varepsilon^U = \mathcal{O}\left(s^{-\alpha}\right) \\
\text{and} \quad \Pr\left[X_2 > s_2\right] &= s_2^{-\alpha}\left(A_z^U \beta_2^{\alpha} + A_\varepsilon^U\right) = \mathcal{O}\left(s^{-\alpha}\right) \\
\text{but} \quad \Pr\left[X_1 > s_1, X_2 > s_2\right] &= \Pr\left[\varepsilon_1 > s_1, \beta_2 Z + \varepsilon_2 > s_2\right] \\
&= \Pr\left[\varepsilon_1 > s_1\right] \Pr\left[\beta_2 Z + \varepsilon_2 > s_2\right] \\
&= A_\varepsilon^U s_1^{-\alpha}\left(A_z^U \beta_2^{\alpha} + A_\varepsilon^U\right) s_2^{-\alpha} \quad \text{as } s \to \infty \\
&= \mathcal{O}\left(s^{-2\alpha}\right) \\
\text{so} \quad \frac{\Pr\left[X_1 > s_1, X_2 > s_2\right]}{\Pr\left[X_1 > s_1\right]} &= \mathcal{O}\left(s^{-\alpha}\right) \to 0 \quad \text{as } s \to \infty.
\end{aligned}$$

(d) Say $\beta_1 < 0 < \beta_2$. Then the denominator will be order $\mathcal{O}\left(s^{-\alpha}\right)$, but the numerator will be of a lower order:

$$\begin{aligned}
\Pr\left[X_1 > s_1, X_2 > s_2\right] &= \Pr\left[\beta_1 Z + \varepsilon_1 > s_1, \beta_2 Z + \varepsilon_2 > s_2\right] \\
&= \Pr\left[\beta_1 Z > s_1, \beta_2 Z > s_2\right] + o\left(s^{-\alpha}\right) \quad \text{as } s \to \infty \\
&= o\left(s^{-\alpha}\right)
\end{aligned}$$

since $\Pr\left[\beta_1 Z > s_1, \beta_2 Z > s_2\right] = 0$ as $s_1, s_2 > 0$ $(\to \infty)$ and $sgn\left(\beta_1 Z\right) = -sgn\left(\beta_2 Z\right)$.

Thus $\tau^U = o\left(s^{-\alpha}\right)/\mathcal{O}\left(s^{-\alpha}\right) = o\left(1\right) \to 0$ as $s \to \infty$. All of the results for parts (a) through (d) apply for lower tail dependence, *mutatis mutandis.* $\square$

Proof of Proposition 2. It is more convenient to work with the density than the distribution function for skew t random variables. Note that if F_z has a regularly varying tails with tail index $\alpha > 0$, then

$$F_z\left(s\right) \equiv \Pr\left[Z \leqslant s\right] = 1 - \Pr\left[Z > s\right] = 1 - A_z^U s^{-\alpha} \quad \text{as} \quad s \to \infty$$

$$f_z\left(s\right) \equiv \frac{\partial F_z\left(s\right)}{\partial s} = -\frac{\partial}{\partial s}\Pr\left[Z > s\right] = \alpha A_z^U s^{-\alpha-1} \quad \text{as} \quad s \to \infty$$

$$\text{so} \quad A_z^U = \lim_{s\to\infty} \frac{f_z\left(s\right)}{\alpha s^{-\alpha-1}}$$

This representation of the extreme tails of a density function is common in EVT, see Embrechts, *et al.* (1997) and Daníelsson, *et al.* (2012) for example. For $\nu \in (2, \infty)$ and $\lambda \in (-1, 1)$, the skew t distribution of Hansen (1994) has density:

$$f_z\left(s;\nu,\lambda\right) = \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2}, & z < -a/b \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2}, & z \geqslant -a/b \end{cases}$$

$$\text{where} \quad a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right),\ b = \sqrt{1 + 3\lambda^2 - a^2},\ c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\left(\nu-2\right)}}$$

and its tail index is equal to the degrees of freedom parameter, so $\alpha = \nu$. Using computational algebra software such as Mathematica, it is possible to show that

$$A_z^U = \lim_{s\to\infty} \frac{f_z\left(s\right)}{\nu s^{-\nu-1}} = \frac{bc}{\nu}\left(\frac{b^2}{\left(\nu-2\right)\left(1+\lambda\right)^2}\right)^{-(\nu+1)/2}$$

For the left tail we have

$$f_z(s) \equiv \frac{\partial F_z(s)}{\partial s} = \frac{\partial}{\partial s} A_z^L (-s)^{-\alpha} \quad \text{as } s \to -\infty$$
$$= \alpha A_z^L (-s)^{-\alpha-1}$$
$$\text{and so } A_z^L = \lim_{s \to -\infty} \frac{f_z(s)}{\nu (-s)^{-\nu-1}}$$

And this can be shown to equal

$$A_z^L = \lim_{s \to -\infty} \frac{f_z(s)}{\alpha (-s)^{-\alpha-1}} = \frac{bc}{\nu} \left(\frac{b^2}{(\nu-2)(1-\lambda)^2} \right)^{-(\nu+1)/2}$$

When $\lambda = 0$ we recover the non-skewed, standardized Student's t distribution. In that case we have $a = 0$, $b = 1$ (and c unchanged), and so we have $A_\varepsilon^U = A_\varepsilon^L = \frac{c}{\nu} \left(\frac{1}{\nu-2} \right)^{-(\nu+1)/2}$. □

Proof of Proposition 3. First consider the denominator of the upper tail dependence coefficient:

$$\Pr[X_i > s_i] = \Pr\left[\sum\nolimits_{k=1}^{K} \beta_{ik} Z_k + \varepsilon_i > s_i\right]$$
$$\approx \Pr[\varepsilon_i > s_i] + \sum\nolimits_{k=1}^{K} \Pr[\beta_{ik} Z_k > s_i] \quad \text{for } s_i \to \infty$$
$$= s_i^{-\alpha} \left(A_\varepsilon^U + \sum\nolimits_{k=1}^{K} A_k^U \beta_{ik}^\alpha \right)$$

We need to choose $s_i, s_j \to \infty$ such that $\Pr[X_i > s_i] = \Pr[X_j > s_j]$, which implies

$$s_j = s_i \left(\frac{A_\varepsilon^U + \sum\nolimits_{k=1}^{K} A_k^U \beta_{jk}^\alpha}{A_\varepsilon^U + \sum\nolimits_{k=1}^{K} A_k^U \beta_{ik}^\alpha} \right)^{1/\alpha} \equiv s_i \gamma_{U,ij}$$

Note again that s_i and s_j diverge at the same rate.

When $\beta_{ik}\beta_{jk} = 0$, the factor Z_k does not contribute to the numerator of the tail dependence coefficient, as it appears in at most one of X_i and X_j. Thus we need only keep track of factors such that $\beta_{ik}\beta_{jk} > 0$. In this case, we again need to determine the larger of s_i/β_{ik} and s_j/β_{jk} for each $k = 1, 2, ..., K$. Unlike the one-factor model, a general ranking cannot be obtained. To keep notation compact we introduce δ_{ijk}. Note

$$\max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_j}{\beta_{jk}}\right\} = \max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_i}{\beta_{jk}}\gamma_{U,ij}\right\} = \frac{s_i}{\beta_{ij}} \max\left\{1, \frac{\beta_{ik}}{\beta_{jk}}\gamma_{U,ij}\right\} \equiv \frac{s_i}{\beta_{ik}\delta_{ijk}}$$

$$\text{where } \ \delta_{ijk}^{-1} \equiv \max\left\{1, \frac{\beta_{ik}}{\beta_{jk}}\gamma_{U,ij}\right\}$$

To cover the case that $\beta_{ik}\beta_{jk} = 0$, we generalize the definition of δ_{ijk} so that it is well defined in that case. The use of any finite number here will work (as it will be multiplied by zero in this case) and we set it to one:

$$\delta_{ijk}^{-1} \equiv \begin{cases} \max\left\{1, \gamma_{U,ij}\beta_{ik}/\beta_{jk}\right\}, & \text{if } \beta_{ik}\beta_{jk} > 0 \\ 1, & \text{if } \beta_{ik}\beta_{jk} = 0 \end{cases}$$

Now we can consider the numerator

$$\Pr\left[X_i > s_i, X_j > s_j\right] = \Pr\left[\sum_{k=1}^{K}\beta_{ik}Z_k + \varepsilon_i > s_i, \sum_{k=1}^{K}\beta_{jk}Z_k + \varepsilon_j > s_j\right]$$

$$\approx \sum_{k=1}^{K} \Pr\left[\beta_{ik}Z_k > s_i, \beta_{jk}Z_k > s_j\right] \quad \text{for } s_i, s_k \to \infty$$

$$= \sum_{k=1}^{K}\mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[\beta_{ik}Z_k > s_i, \beta_{jk}Z_k > s_j\right]$$

$$= \sum_{k=1}^{K}\mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[Z_k > \max\left\{\frac{s_i}{\beta_{ik}}, \frac{s_j}{\beta_{jk}}\right\}\right]$$

$$\equiv \sum_{k=1}^{K}\mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} \Pr\left[Z_k > \frac{s_i}{\beta_{ik}\delta_{ijk}}\right]$$

$$= s_i^{-\alpha}\sum_{k=1}^{K}\mathbf{1}\left\{\beta_{ik}\beta_{jk} > 0\right\} A_k^U \beta_{ik}^{\alpha}\delta_{ijk}^{\alpha}$$

And so we obtain

$$\tau_{ij}^{U}=\lim_{s\rightarrow\infty}\frac{\Pr\left[X_i>s_i,X_j>s_j\right]}{\Pr\left[X_i>s_i\right]}=\frac{\sum_{k=1}^{K}\mathbf{1}\left\{\beta_{ik}\beta_{jk}>0\right\}A_k^U\beta_{ik}^{\alpha}\delta_{ijk}^{\alpha}}{A_{\varepsilon}^{U}+\sum_{k=1}^{K}A_k^U\beta_{ik}^{\alpha}}$$

The results for lower tail dependence can be obtained using similar derivations to those above. □

B.2 Choice of dependence measures for estimation

To implement the SMM estimator of these copula models we must first choose which dependence measures to use in the SMM estimation. We draw on "pure" measures of dependence, in the sense that they are solely affected by changes in the copula, and not by changes in the marginal distributions. For examples of such measures, see Joe (1997, Chapter 2) or Nelsen (2006, Chapter 5). Our preliminary studies of estimation accuracy and identification lead us to use pair-wise rank correlation, and quantile dependence with $q=[0.05,0.10,0.90,0.95]$, giving us five dependence measures for each pair of variables.

Let δ_{ij} denote one of the dependence measures (i.e., rank correlation or quantile dependence at different levels of q) between variables i and j, and define the "pair-wise dependence matrix":

$$D=\left[\begin{array}{cccc}1 & \delta_{12} & \cdots & \delta_{1N}\\ \delta_{12} & 1 & \cdots & \delta_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ \delta_{1N} & \delta_{2N} & \cdots & 1\end{array}\right] \tag{B.1}$$

Where applicable, we exploit the (block) equidependence feature of the models in defining the "moments" to match. For the initial set of simulation results and for the first model in the empirical section, the model implies equidependence, and we use as "moments" the average of these five dependence measures across all pairs, reducing the number of moments to match from $5N(N-1)/2$ to just 5:

$$\bar{\delta} \equiv \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\delta}_{ij} \tag{B.2}$$

For a model with different loadings on the common factor (as in equation 2.3) equidependence does not hold. Yet the common factor aspect of the model implies that there are $\mathcal{O}(N)$, not $\mathcal{O}(N^2)$, parameters driving the pair-wise dependence matrices. In light of this, we use the $N \times 1$ vector $\left[\bar{\delta}_1, ..., \bar{\delta}_N\right]'$, where

$$\bar{\delta}_i \equiv \frac{1}{N} \sum_{j=1}^{N} \hat{\delta}_{ij}$$

and so $\bar{\delta}_i$ is the average of all pair-wise dependence measures that involve variable i. This yields a total of $5N$ moments for estimation.

For the block-equidependence version of this model (used for the $N = 100$ case in the simulation, and in the second set of models for the empirical section), we exploit the fact that (i) all variables in the same group exhibit equidependence, and (ii) any pair of variables (i, j) in groups (r, s) has the same dependence as any other pair (i', j') in the same two groups (r, s). This allows us to average all intra- and inter-group dependence measures. Consider the following general design, where we have N variables, M groups, and k_m variables per group, where $\Sigma_{m=1}^{M} k_m = N$. Then decompose the $(N \times N)$ matrix D into sub-matrices according to the groups:

$$\underset{(N\times N)}{D} = \begin{bmatrix} D_{11} & D_{12}' & \cdots & D_{1M}' \\ D_{12} & D_{22} & \cdots & D_{2M}' \\ \vdots & \vdots & \ddots & \vdots \\ D_{1M} & D_{2M} & \cdots & D_{MM} \end{bmatrix}, \text{ where } D_{ij} \text{ is } (k_i \times k_j) \tag{B.3}$$

Then create a matrix of average values from each of these matrices, taking into account the fact that the diagonal blocks are symmetric:

$$\underset{(M\times M)}{D^*} = \begin{bmatrix} \delta^*_{11} & \delta^*_{12} & \cdots & \delta^*_{1m} \\ \delta^*_{12} & \delta^*_{22} & \cdots & \delta^*_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \delta^*_{1m} & \delta^*_{2m} & \cdots & \delta^*_{mm} \end{bmatrix} \tag{B.4}$$

$$\text{where} \quad \delta^*_{ss} \equiv \frac{2}{k_s\left(k_s-1\right)}\sum\sum\hat{\delta}_{ij}, \;\; \text{avg of } \textit{all upper triangle} \text{ values in } D_{ss}$$

$$\delta^*_{rs} = \frac{1}{k_r k_s}\sum\sum\hat{\delta}_{ij}, \;\; \text{avg of } \textit{all} \text{ elements in matrix } D_{rs}, \quad r \neq s$$

Finally, similar to the previous model, create the vector of average measures $\left[\bar{\delta}^*_1, ..., \bar{\delta}^*_M\right]$, where

$$\bar{\delta}^*_i \equiv \frac{1}{M}\sum_{j=1}^{M}\delta^*_{ij} \tag{B.5}$$

This gives as a total of M moments for each dependence measure, so $5M$ in total.

B.3 Additional tables

Table B.1: Simulation results for different weights factor copula model with N=10, W=I

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True	0.25	-0.5	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
						Normal						
Bias	–	–	-0.0001	-0.0033	-0.0085	0.0004	0.0034	0.0119	0.0075	-0.0031	0.0071	-0.0902
Std	–	–	0.0337	0.0448	0.0535	0.0753	0.0991	0.1089	0.1243	0.1669	0.2061	0.2243
Med	–	–	0.2500	0.4920	0.7466	0.9831	1.2365	1.5021	1.7509	1.9932	2.2281	2.3963
90%	–	–	0.2949	0.5588	0.8032	1.1013	1.3919	1.6698	1.9289	2.1878	2.5385	2.6649
10%	–	–	0.2064	0.4369	0.6727	0.9180	1.1513	1.3938	1.6059	1.7796	1.9910	2.1087
Diff	–	–	0.0886	0.1219	0.1305	0.1833	0.2407	0.2761	0.3229	0.4081	0.5476	0.5562
						t(4)-Normal						
Bias	-0.0032	–	-0.0008	0.0089	0.0022	0.0089	0.0118	0.0133	0.0004	0.0107	0.0195	0.0487
Std	0.0474	–	0.0468	0.0596	0.0798	0.0874	0.1111	0.1482	0.1991	0.2068	0.3063	0.3472
Med	0.2468	–	0.2490	0.5046	0.7344	1.0104	1.2639	1.4930	1.7313	1.9750	2.2259	2.5285
90%	0.2941	–	0.3079	0.5869	0.8475	1.0968	1.4107	1.6629	1.9567	2.2495	2.5667	2.7873
10%	0.1845	–	0.1879	0.4394	0.6692	0.9085	1.1244	1.3755	1.5623	1.7729	1.9852	2.2032
Diff	0.1096	–	0.1200	0.1475	0.1782	0.1883	0.2863	0.2874	0.3943	0.4766	0.5815	0.5841
						Skew t(4,-0.5)-Normal						
Bias	-0.0020	-0.0032	0.0042	0.0034	0.0024	0.0175	0.0293	0.0187	0.0308	0.0355	0.0739	0.1035
Std	0.0481	0.0594	0.0527	0.0593	0.0812	0.1215	0.1270	0.1404	0.1857	0.2292	0.3335	0.3639
Med	0.2474	-0.5017	0.2507	0.4962	0.7430	1.0137	1.2635	1.5152	1.7547	2.0018	2.2757	2.5508
90%	0.3119	-0.4283	0.3226	0.5830	0.8570	1.1679	1.4554	1.7246	2.0268	2.3648	2.7511	3.0743
10%	0.1825	-0.5868	0.1863	0.4340	0.6522	0.8765	1.1469	1.3595	1.5586	1.7609	2.0021	2.2101
Diff	0.1293	0.1586	0.1363	0.1491	0.2048	0.2914	0.3085	0.3651	0.4682	0.6039	0.7490	0.8642

Table B.2: Simulation results for different weights factor copula model with N=10, W=optimal

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True	0.25	-0.5	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
						Normal						
Bias	-	-	-0.0304	-0.0253	-0.0197	-0.0242	-0.0325	-0.0315	-0.0454	-0.0581	-0.0596	-0.0871
Std	-	-	0.0535	0.0410	0.0525	0.0605	0.0641	0.0788	0.0857	0.0978	0.0953	0.1192
Med	-	-	0.2227	0.4794	0.7290	0.9794	1.2114	1.4779	1.7018	1.9403	2.1812	2.4031
90%	-	-	0.2762	0.5185	0.7938	1.0515	1.3109	1.5692	1.8301	2.0783	2.3176	2.5703
10%	-	-	0.1722	0.4209	0.6638	0.9057	1.1395	1.3620	1.5948	1.8123	2.0809	2.2665
Diff	-	-	0.1040	0.0976	0.1300	0.1458	0.1714	0.2072	0.2353	0.2660	0.2367	0.3038
						t(4)-Normal						
Bias	-0.0405	-	-0.0241	-0.0224	-0.0351	-0.0408	-0.0530	-0.0684	-0.0712	-0.0971	-0.1210	-0.1195
Std	0.0607	-	0.0466	0.0601	0.0715	0.0931	0.1193	0.1352	0.1881	0.1756	0.2046	0.2203
Med	0.2086	-	0.2286	0.4727	0.7082	0.9394	1.1816	1.4083	1.6381	1.8670	2.0894	2.3503
90%	0.2629	-	0.2811	0.5515	0.8116	1.0613	1.3281	1.5826	1.9200	2.1332	2.3521	2.6717
10%	0.1438	-	0.1642	0.4111	0.6367	0.8515	1.0602	1.2982	1.4833	1.7163	1.9083	2.1203
Diff	0.1191	-	0.1169	0.1404	0.1749	0.2098	0.2679	0.2844	0.4367	0.4169	0.4438	0.5514
						Skew t(4,-0.5)-Normal						
Bias	-0.0086	-0.0389	-0.0084	-0.0128	-0.0067	0.0075	0.0014	-0.0029	-0.0127	-0.0184	-0.0095	0.0051
Std	0.0496	0.0675	0.0480	0.0551	0.0816	0.1108	0.1209	0.1289	0.1956	0.2079	0.2680	0.2788
Med	0.2482	-0.5326	0.2404	0.4814	0.7390	0.9928	1.2295	1.4721	1.7087	1.9537	2.1688	2.4487
90%	0.2898	-0.4605	0.2993	0.5586	0.8256	1.1433	1.3845	1.6769	1.9767	2.2569	2.5245	2.8424
10%	0.1791	-0.6349	0.1865	0.4268	0.6506	0.8752	1.1275	1.3569	1.5622	1.7658	2.0016	2.2634
Diff	0.1107	0.1744	0.1128	0.1318	0.1750	0.2681	0.2570	0.3199	0.4145	0.4911	0.5229	0.5790

Table B.3: Simulation results for different weights factor copula model with N=100, W=optimal

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
True	0.25	-0.5	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
						Normal						
Bias	-	-	-0.0099	-0.0172	-0.0229	-0.0319	-0.0389	-0.0481	-0.0586	-0.0612	-0.0768	-0.0836
Std	-	-	0.0128	0.0184	0.0242	0.0322	0.0372	0.0454	0.0502	0.0590	0.0622	0.0753
Med	-	-	0.2414	0.4844	0.7259	0.9665	1.2129	1.4507	1.6918	1.9428	2.1678	2.4170
90%	-	-	0.2554	0.5037	0.7574	1.0121	1.2577	1.5105	1.7527	2.0149	2.2589	2.5083
10%	-	-	0.2244	0.4594	0.6970	0.9304	1.1606	1.3983	1.6246	1.8676	2.0960	2.3241
Diff	-	-	0.0311	0.0443	0.0603	0.0818	0.0971	0.1122	0.1281	0.1473	0.1629	0.1842
						t(4)-Normal						
Bias	-0.0683	-	-0.0208	-0.0403	-0.0601	-0.0812	-0.1052	-0.1248	-0.1481	-0.1674	-0.1934	-0.2194
Std	0.0524	-	0.0160	0.0287	0.0402	0.0524	0.0674	0.0802	0.0952	0.1058	0.1220	0.1299
Med	0.1820	-	0.2277	0.4567	0.6842	0.9106	1.1377	1.3694	1.5795	1.8226	2.0378	2.2733
90%	0.2452	-	0.2511	0.4982	0.7402	0.9932	1.2320	1.4817	1.7291	1.9630	2.2169	2.4542
10%	0.1183	-	0.2096	0.4208	0.6476	0.8534	1.0665	1.2843	1.5051	1.7154	1.9190	2.1357
Diff	0.1268	-	0.0415	0.0774	0.0926	0.1398	0.1655	0.1974	0.2240	0.2476	0.2979	0.3185
						Skew t(4,-0.5)-Normal						
Bias	-0.0335	-0.0391	-0.0139	-0.0273	-0.0403	-0.0556	-0.0713	-0.0868	-0.1022	-0.1149	-0.1330	-0.1500
Std	0.0363	0.0482	0.0150	0.0251	0.0370	0.0476	0.0598	0.0729	0.0825	0.0969	0.1099	0.1221
Med	0.2263	-0.5434	0.2367	0.4719	0.7105	0.9381	1.1796	1.4150	1.6414	1.8823	2.1237	2.3574
90%	0.2582	-0.4719	0.2568	0.5059	0.7567	1.0007	1.2524	1.5053	1.7545	2.0035	2.2631	2.5148
10%	0.1693	-0.6029	0.2168	0.4424	0.6611	0.8823	1.1015	1.3199	1.5542	1.7725	1.9882	2.1904
Diff	0.0889	0.1310	0.0400	0.0634	0.0956	0.1185	0.1509	0.1854	0.2003	0.2311	0.2749	0.3245

Table B.4: 95% Coverage rate for different weights factor copula model with N=10 AR-GARCH data W=I

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
	Normal											
ε_T												
0.1	-	-	96	93	91	95	93	96	96	95	97	89
0.03	-	-	96	92	91	94	93	97	96	96	96	89
0.01	-	-	96	92	91	94	92	97	96	96	96	89
0.003	-	-	96	91	90	94	89	94	97	92	95	88
0.001	-	-	96	93	90	93	90	93	92	84	87	79
0.0003	-	-	90	83	85	75	75	76	74	67	74	67
0.0001	-	-	81	70	70	65	65	66	66	57	65	59
	t(4)-Normal											
ε_T												
0.1	96	-	92	97	98	95	94	96	92	94	92	95
0.03	92	-	92	97	97	94	93	97	94	93	94	96
0.01	93	-	92	97	97	93	94	97	92	95	94	96
0.003	91	-	87	94	95	95	91	95	94	90	95	95
0.001	81	-	86	92	91	91	88	89	91	88	84	87
0.0003	76	-	80	82	78	86	76	84	79	71	78	77
0.0001	76	-	81	65	74	77	74	76	64	65	65	79
	Skew t(4,-0.5)-Normal											
ε_T												
0.1	96	94	92	97	95	93	98	98	95	94	97	97
0.03	93	95	91	97	95	93	97	98	95	94	95	98
0.01	91	91	91	96	95	92	96	97	94	93	91	91
0.003	88	91	91	96	94	88	92	96	91	88	87	85
0.001	83	86	86	93	94	85	88	88	82	73	72	81
0.0003	73	70	75	86	81	73	80	73	72	60	65	71
0.0001	68	57	68	77	68	66	69	67	68	61	65	74

Table B.5: 95% Coverage rate for different weights factor copula model with N=10 AR-GARCH data W=optimal

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
	Normal											
ε_T												
0.1	-	-	80	80	81	79	81	80	80	85	91	80
0.03	-	-	80	80	81	79	81	81	80	83	91	80
0.01	-	-	77	80	81	78	80	80	79	82	90	76
0.003	-	-	77	78	80	77	78	78	78	75	90	75
0.001	-	-	73	74	77	70	74	76	69	72	76	62
0.0003	-	-	63	62	67	66	63	70	62	56	71	58
0.0001	-	-	50	53	59	47	57	57	58	57	71	69
	t(4)-Normal											
ε_T												
0.1	74	-	81	80	74	77	69	63	70	71	73	71
0.03	69	-	82	80	72	75	65	62	64	66	69	67
0.01	57	-	78	78	68	70	64	56	56	63	62	62
0.003	46	-	78	76	65	62	59	50	51	52	56	54
0.001	38	-	75	69	55	51	51	50	41	43	43	47
0.0003	35	-	66	55	43	42	39	43	32	35	39	39
0.0001	32	-	62	47	38	44	45	41	36	36	46	49
	Skew t(4,-0.5)-Normal											
ε_T												
0.1	88	81	90	89	89	89	95	93	89	90	94	96
0.03	82	77	88	88	84	83	89	89	81	84	85	89
0.01	80	75	89	88	80	78	82	86	76	79	84	82
0.003	69	68	84	86	74	71	78	79	69	69	68	74
0.001	61	51	78	83	71	62	70	74	64	62	59	63
0.0003	55	36	67	70	62	57	55	65	57	46	54	62
0.0001	50	37	67	64	57	50	54	57	60	52	59	65

Table B.6: 95% Coverage rate for different weights factor copula model with N=100 AR-GARCH data W=optimal

	ν_z^{-1}	λ_z	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
	Normal											
ε_T												
0.1	-	-	78	67	62	55	60	60	57	58	51	54
0.03	-	-	78	67	62	55	60	60	56	58	51	53
0.01	-	-	77	67	62	52	59	61	55	56	49	52
0.003	-	-	77	67	61	51	56	52	52	53	47	50
0.001	-	-	76	63	58	50	56	48	47	50	42	47
0.0003	-	-	71	58	49	37	42	36	37	40	35	33
0.0001	-	-	67	50	39	34	33	27	33	33	26	25
	t(4)-Normal											
ε_T												
0.1	48	-	58	44	37	37	36	36	32	37	34	33
0.03	46	-	54	42	36	35	30	32	29	33	29	29
0.01	44	-	54	40	33	30	27	29	27	32	26	26
0.003	30	-	51	36	29	24	22	20	19	25	22	20
0.001	21	-	47	32	25	21	17	16	14	19	15	14
0.0003	12	-	44	27	22	17	12	10	13	13	11	8
0.0001	6	-	38	21	17	14	11	9	10	8	9	8
	Skew t(4,-0.5)-Normal											
ε_T												
0.1	75	73	79	76	70	69	68	65	67	68	69	65
0.03	75	71	78	72	69	63	62	62	63	62	64	63
0.01	71	68	75	63	66	60	58	60	59	60	58	57
0.003	57	60	70	56	55	46	47	46	45	47	49	43
0.001	51	44	67	49	43	35	33	32	28	30	27	30
0.0003	45	25	63	40	33	27	21	22	14	19	18	21
0.0001	34	15	58	36	27	22	16	16	12	10	13	15

Table B.7: Overidentifying restriction test with W=optimal

	Normal	t(4)-Normal	Skew t(4,-0.5)-Normal	Normal	t(4)-Normal	Skew t(4,–0.5)-Normal
	Equidependce, N=3			Different loading, N=3		
90%	92	92	93	93	96	97
95%	95	95	99	97	98	98
99%	100	98	100	98	100	99
	Equidependce, N=10			Different loading, N=10		
90%	96	92	93	91	95	97
95%	97	94	96	93	98	98
99%	99	96	100	95	99	99
	Equidependce, N=100			Different loading, N=100		
90%	90	89	93	94	90	96
95%	95	92	98	94	95	99
99%	97	96	100	98	99	100

Appendix C

Appendix to Chapter 3

C.1 Proofs

In order to prove Proposition 1, we use the following five lemmas. First, we recall the definition of stochastic equicontinuity.

Definition 1. *(Andrews (1994)) The empirical process* $\{\mathbf{h}_T(\cdot) : T \geqslant 1\}$ *is stochastically equicontinuous if* $\forall\ \varepsilon > 0$ *and* $\eta > 0, \exists\ \delta > 0$ *such that*

$$\limsup_{T\to\infty} P\left[\sup_{\|\theta_1-\theta_2\|<\delta} \|\mathbf{h}_T(\theta_1) - \mathbf{h}_T(\theta_2)\| > \eta\right] < \varepsilon \tag{C.1}$$

Lemma 1. *Under Assumptions 1 and 2,*

(i) $\frac{1}{T}\sum_{t=1}^{T} \hat{F}_i(\hat{\eta}_{it})\hat{F}_j(\hat{\eta}_{jt}) \xrightarrow{p} \iint uvdC_{\eta_i,\eta_j}(u,v;\theta_0)$ *as* $T \to \infty$

(ii) $\frac{1}{T}\sum_{t=1}^{T} 1\left\{\hat{F}_i(\hat{\eta}_{it}) \leqslant q, \hat{F}_j(\hat{\eta}_{jt}) \leqslant q\right\} \xrightarrow{p} C_{\eta_i,\eta_j}(q,q;\theta_0)$ *as* $T \to \infty$

(iii) $\frac{1}{S}\sum_{s=1}^{S} \hat{G}_i(x_{is}(\theta))\hat{G}_j(x_{js}(\theta)) \xrightarrow{p} \iint uvdC_{\eta_i,\eta_j}(u,v;\theta)$ *for* $\forall\ \theta \in \mathbf{\Theta}$ *as* $S \to \infty$

(iv) $\frac{1}{S}\sum_{s=1}^{S} 1\left\{\hat{G}_i\left(x_{is}\left(\theta\right)\right) \leqslant q, \hat{G}_j\left(x_{js}\left(\theta\right)\right) \leqslant q\right\} \xrightarrow{p} C_{\eta_i,\eta_j}\left(q,q;\theta\right)$ *for* $\forall\ \theta \in \boldsymbol{\Theta}$ *as* $S \to \infty$

(v) $\frac{1}{S}\sum_{s=1}^{S} G_i\left(x_{is}\left(\theta\right)\right) G_j\left(x_{js}\left(\theta\right)\right) \xrightarrow{p} \int\int uv dC_{\eta_i,\eta_j}\left(u,v;\theta\right)$ *for* $\forall\ \theta \in \boldsymbol{\Theta}$ *as* $S \to \infty$

(vi) $\frac{1}{S}\sum_{s=1}^{S} 1\left\{G_i\left(x_{is}\left(\theta\right)\right) \leqslant q, G_j\left(x_{js}\left(\theta\right)\right) \leqslant q\right\} \xrightarrow{p} C_{\eta_i,\eta_j}\left(q,q;\theta\right)$ *for* $\forall\ \theta \in \boldsymbol{\Theta}$ *as* $S \to \infty$

Proof of Lemma 1. Under Assumption 1, parts (iii) and (iv) of Lemma 1 can be proven by Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004). Under Assumption 2, Corollary 1 of Rémillard (2010) proves that the empirical copula process constructed by the standardized residuals $\hat{\eta}_t$ weakly converges to the limit of that constructed by the innovations η_t, which combined with Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004) yields parts (i) and (ii) above. In the case where it is possible to simulate directly from the copula rather than the joint distribution, e.g. Clayton/Gaussian copula in Section 3.3 or where we only can simulate from the joint distribution but know the marginal distribution G_i in closed form, it is not necessary to estimate marginal distribution G_i. In this case, instead of (iii) and (iv), (v) and (vi) are used for the later proofs. (v) and (vi) are proven by the standard law of large numbers. □

Lemma 2. *(Lemma 2.8 of Newey and McFadden (1994)) Suppose* $\boldsymbol{\Theta}$ *is compact and* $\mathbf{g}_0\left(\theta\right)$ *is continuous. Then* $\sup_{\theta\in\boldsymbol{\Theta}}\left\|\mathbf{g}_{T,S}\left(\theta\right) - \mathbf{g}_0\left(\theta\right)\right\| \xrightarrow{p} 0$ *as* $T, S \to \infty$ *if and only if* $\mathbf{g}_{T,S}\left(\theta\right) \xrightarrow{p} \mathbf{g}_0\left(\theta\right)$ *for any* $\theta \in \boldsymbol{\Theta}$ *as* $T, S \to \infty$ *and* $\mathbf{g}_{T,S}\left(\theta\right)$ *is stochastically equicontinuous.*

Lemma 2 states that sufficient and necessary conditions for uniform convergence are pointwise convergence and stochastic equicontinuity. The following lemma shows that uniform convergence of the moment functions $\mathbf{g}_{T,S}\left(\theta\right)$ implies uniform convergence of the objective function $Q_{T,S}\left(\theta\right)$.

Lemma 3. *If* $\sup_{\theta \in \boldsymbol{\Theta}} \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\| \xrightarrow{p} 0$ *as* $T, S \to \infty$, *then* $\sup_{\theta \in \boldsymbol{\Theta}} |Q_{T,S}(\theta) - Q_0(\theta)| \xrightarrow{p} 0$ *as* $T, S \to \infty$.

Proof of Lemma 3. By the triangle inequality and Cauchy-Schwarz inequality

$$|Q_{T,S}(\theta) - Q_0(\theta)| \leqslant \left|[\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)]' \hat{\mathbf{W}}_T [\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)]\right| \tag{C.2}$$

$$+ \left|\mathbf{g}_0(\theta)' \left(\hat{\mathbf{W}}_T + \hat{\mathbf{W}}_T'\right) [\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)]\right| \tag{C.3}$$

$$+ \left|\mathbf{g}_0(\theta)' \left(\hat{\mathbf{W}}_T - \mathbf{W}_0\right) \mathbf{g}_0(\theta)\right|$$

$$\leqslant \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\|^2 \left\|\hat{\mathbf{W}}_T\right\| + 2\|\mathbf{g}_0(\theta)\| \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\| \left\|\hat{\mathbf{W}}_T\right\|$$

$$+ \|\mathbf{g}_0(\theta)\|^2 \left\|\hat{\mathbf{W}}_T - \mathbf{W}_0\right\|$$

Then note that $\mathbf{g}_0(\theta)$ is bounded, $\hat{\mathbf{W}}_T$ is $O_p(1)$ and converges to $\mathbf{W}_0$ by Assumption 3(iv), and $\sup_{\theta \in \boldsymbol{\Theta}} \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\| = o_p(1)$ is given. So

$$\sup_{\theta \in \boldsymbol{\Theta}} |Q_{T,S}(\theta) - Q_0(\theta)| \leqslant \left(\sup_{\theta \in \boldsymbol{\Theta}} \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\|\right)^2 O_p(1) \tag{C.4}$$

$$+ 2O(1) \sup_{\theta \in \boldsymbol{\Theta}} \|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\| O_p(1) + o_p(1) = o_p(1)$$

□

Lemma 4. *Under Assumption 1, Assumption 2, and Assumption 3(iii),*

(i) $\mathbf{g}_{T,S}(\theta)$ *is stochastic Lipschitz continuous, i.e.*

$\exists B_{T,S} = O_p(1)$ *such that for all* $\theta_1, \theta_2 \in \boldsymbol{\Theta}, \|\mathbf{g}_{T,S}(\theta_1) - \mathbf{g}_{T,S}(\theta_2)\| \leqslant B_{T,S} \cdot \|\theta_1 - \theta_2\|$

(ii) There exists $\delta > 0$ *such that*

$$\lim \sup_{T,S \to \infty} E\left(B_{T,S}^{2+\delta}\right) < \infty \text{ for some } \delta > 0$$

Proof of Lemma 4. Without loss of generality, assume that $\mathbf{g}_{T,S}(\theta)$ is scalar. By Lemma 1, we know that

$$\tilde{\mathbf{m}}_S(\theta) = \mathbf{m}_0(\theta) + o_p(1) \tag{C.5}$$

Also, by Assumption 3(iii) and the fact that $\mathbf{m}(\theta)$ consists of a function of Lipschitz continuous $C_{ij}(\theta)$, $\mathbf{m}_0(\theta)$ is Lipschitz continuous, i.e. $\exists K$ such that

$$|\mathbf{m}_0(\theta_1) - \mathbf{m}_0(\theta_2)| \leqslant K \|\theta_1 - \theta_2\| \tag{C.6}$$

Then,

$$\begin{aligned} |\mathbf{g}_{T,S}(\theta_1) - \mathbf{g}_{T,S}(\theta_2)| &= |\tilde{\mathbf{m}}_S(\theta_1) - \tilde{\mathbf{m}}_S(\theta_2)| = |\mathbf{m}_0(\theta_1) - \mathbf{m}_0(\theta_2) + o_p(1)| \qquad \text{(C.7)} \\ &\leqslant |\mathbf{m}_0(\theta_1) - \mathbf{m}_0(\theta_2)| + |o_p(1)| \\ &\leqslant K\|\theta_1 - \theta_2\| + |o_p(1)| \\ &= \underbrace{\left(K + \frac{|o_p(1)|}{\|\theta_1 - \theta_2\|}\right)}_{=O_p(1)} \|\theta_1 - \theta_2\| \end{aligned}$$

and let $B_{T,S} = K + \frac{|o_p(1)|}{\|\theta_1 - \theta_2\|}$. Then for some $\delta > 0$

$$\lim \sup_{T,S\to\infty} E\left(B_{T,S}^{2+\delta}\right) = \lim \sup_{T,S\to\infty} E\left(K + \frac{|o_p(1)|}{\|\theta_1 - \theta_2\|}\right)^{2+\delta} < \infty \tag{C.8}$$

□

Lemma 5. *(Theorem 2.1 of Newey and McFadden 1994) Suppose that (i) $Q_0(\theta)$ is uniquely minimized at θ_0; (ii) $\boldsymbol{\Theta}$ is compact; (iii) $Q_0(\theta)$ is continuous; (iv) $\sup_{\theta\in\boldsymbol{\Theta}} \left|\hat{Q}_T(\theta) - Q_0(\theta)\right| \xrightarrow{p} 0$. Then $\hat{\theta} \xrightarrow{p} \theta_0$*

Proof of Proposition 1. We prove this proposition by checking the conditions of Lemma 5.

(i) $Q_0(\theta)$ is uniquely minimized at θ_0 by Assumption 3(i) and Assumption 3(iv).

(ii) $\mathbf{\Theta}$ is compact by Assumption 3(ii).

(iii) $Q_0(\theta)$ consists of linear combinations of rank correlations and quantile dependence measures that are functions of pair-wise copula functions. Therefore, $Q_0(\theta)$ is continuous by Assumption 3(iii).

(iv) The pointwise convergence of $\mathbf{g}_{T,S}(\theta)$ to $\mathbf{g}_0(\theta)$ and the stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\theta)$ are shown by Lemma 1 and by Lemma 4(i), respectively. By Lemma 2.9 of Newey and McFadden (1994), the stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\theta)$ ensures the stochastic equicontinuity of $\mathbf{g}_{T,S}(\theta)$, and under Assumption 3, $\mathbf{\Theta}$ is compact and $\mathbf{g}_0(\theta)$ is continuous in θ. Therefore, $\mathbf{g}_{T,S}$ uniformly converges in probability to $\mathbf{g}_0$ by Lemma 2. This implies that $Q_{T,S}$ uniformly converges in probability to Q_0 by Lemma 3. □

The proof of Proposition 2 uses the following three lemmas.

Lemma 6. *Let the dependence measures of interest include rank correlation and quantile dependence measures, and possibly linear combinations thereof. Then under Assumptions 1 and 2,*

$$\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_0\right) \ \ \text{as } T \to \infty \tag{C.9}$$

$$\sqrt{S}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_0\right) \ \ \text{as } S \to \infty \tag{C.10}$$

Proof of Lemma 6. Follows from Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004) and Corollary 1, Proposition 2 and Proposition 4 of Rémillard (2010). □

We use Theorem 7.2 of Newey & McFadden (1994) to establish the asymptotic normality of our estimator, and this relies on showing the stochastic equicontinuity of $\mathbf{v}_{T,S}(\theta)$ defined below.

Lemma 7. *Suppose that Assumptions 1, 2, and 3(iii) hold. Then when* $S/T \to \infty$ *or* $S/T \to k \in (0,\infty)$, $\mathbf{v}_{T,S}(\theta) = \sqrt{T}\left[\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\right]$ *is stochastically equicontinuous and when* $S/T \to 0$, $\mathbf{v}_{T,S}(\theta) = \sqrt{S}\left[\mathbf{g}_{T,S}(\theta) - \mathbf{g}_0(\theta)\right]$ *is stochastically equicontinuous.*

Proof of Lemma 7. By Lemma 4(i), $\{\mathbf{g}_{\cdot,\cdot}(\theta) : \theta \in \Theta\}$ is a type II class of functions in Andrews (1994). By Theorem 2 of Andrews (1994), $\{\mathbf{g}_{\cdot,\cdot}(\theta) : \theta \in \Theta\}$ satisfies Pollard's entropy condition with envelope $1 \vee \sup_\theta \|\mathbf{g}_{\cdot,\cdot}(\theta)\| \vee B_{\cdot,\cdot}$, so Assumption A of Andrews (1994) is satisfied. Since $\mathbf{g}_{\cdot,\cdot}(\theta)$ is bounded and by the condition of $\limsup_{T,S\to\infty} E\left(B_{T,S}^{2+\delta}\right) < \infty$ for some $\delta > 0$ by Lemma 4(ii), the Assumption B of Andrews (1994) is also satisfied. Therefore, $\mathbf{v}_{T,S}(\theta)$ is stochastically equicontinuous by Theorem 1 of Andrews (1994). $\square$

Lemma 8. *(Theorem 7.2 of Newey & McFadden (1994)) Suppose that* $\mathbf{g}_{T,S}\left(\hat{\theta}\right)' \hat{\mathbf{W}}_T \mathbf{g}_{T,S}\left(\hat{\theta}\right) \leqslant \inf_{\theta\in\boldsymbol{\Theta}} \mathbf{g}_{T,S}(\theta)' \hat{\mathbf{W}}_T \mathbf{g}_{T,S}(\theta) + o_p\left(T^{-1}\right), \hat{\theta} \xrightarrow{p} \theta_0$ *and* $\hat{\mathbf{W}}_T \xrightarrow{p} \mathbf{W}_0, \mathbf{W}_0$ *is positive semi-definite, where there is* $\mathbf{g}_0(\theta)$ *such that (i)* $\mathbf{g}_0(\theta_0) = 0$*,(ii)* $\mathbf{g}_0(\theta)$ *is differentiable at* θ_0 *with derivative* $\mathbf{G}_0$ *such that* $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ *is nonsingular,(iii)* θ_0 *is an interior point of* $\boldsymbol{\Theta}$*,(iv)* $\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_0)$*,(v)* $\exists\delta$ *such that* $\sup_{\|\theta-\theta_0\|\leqslant\delta} \sqrt{T}\left\|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_{T,S}(\theta_0) - \mathbf{g}_0(\theta)\right\| / \left[1 + \sqrt{T}\|\theta - \theta_0\|\right] \xrightarrow{p} 0$*. Then*

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1} \mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0\mathbf{W}_0\mathbf{G}_0 \left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\right).$$

Proof of Proposition 2. We prove this proposition by checking conditions of Lemma 8.

(i) $\mathbf{g}_0(\theta_0) = 0$ by construction of $\mathbf{g}_0(\theta) = \mathbf{m}_0(\theta_0) - \mathbf{m}_0(\theta)$

(ii) $\mathbf{g}_0(\theta)$ is differentiable at θ_0 with derivative $\mathbf{G}_0$ such that $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ is nonsingular by Assumption 4(ii).

(iii) θ_0 is an interior point of $\boldsymbol{\Theta}$ by Assumption 4(i).

(iv) If $S/T \to \infty$ as $T, S \to \infty$,

$$\sqrt{T}\mathbf{g}_{T,S}(\theta_0) = \sqrt{T}\left(\hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\theta_0)\right) \tag{C.11}$$

$$= \sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right) - \sqrt{T}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right)$$

$$= \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\frac{\sqrt{T}}{\sqrt{S}}}_{=o(1)} \times \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}}$$

Therefore,

$$\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty.$$

If $S/T \to k \in (0, \infty)$ as $T, S \to \infty$,

$$\sqrt{T}\mathbf{g}_{T,S}(\theta_0) = \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\frac{\sqrt{T}}{\sqrt{S}}}_{\to 1/\sqrt{k}} \times \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}}$$

Therefore,

$$\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{k}\right)\Sigma_0\right) \text{ as } T, S \to \infty.$$

If $S/T \to 0$ as $T, S \to \infty$,

$$\sqrt{S}\mathbf{g}_{T,S}(\theta_0) = \underbrace{\frac{\sqrt{S}}{\sqrt{T}}}_{=o(1)} \times \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\theta_0) - \mathbf{m}_0(\theta_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}}$$

Therefore,

$$\sqrt{S}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty$$

Consolidating these results across all three combinations of divergence rates for S and T we obtain:

$$\frac{1}{\sqrt{1/S + 1/T}}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty.$$

(v) We established the stochastic equicontinuity of $\mathbf{v}_{T,S}(\theta) = \sqrt{T}\left[\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_0(\theta)\right]$ when $S/T \to \infty$ or $S/T \to k$ by Lemma 7, i.e. for $\forall \varepsilon > 0$, $\eta > 0, \exists \delta$ such that

$$\limsup_{T\to\infty} P\left[\sup_{\|\theta-\theta_0\|<\delta} \|\mathbf{v}_{T,S}(\theta) - \mathbf{v}_{T,S}(\theta_0)\| > \eta\right] \tag{C.12}$$

$$= \limsup_{T\to\infty} P\left[\sup_{\|\theta-\theta_0\|<\delta} \sqrt{T}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\| > \eta\right] < \varepsilon$$

and from the following inequality

$$\frac{\sqrt{T}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\|}{1+\sqrt{T}\,\|\theta-\theta_0\|} \leqslant \sqrt{T}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\| \tag{C.13}$$

we know that

$$\limsup_{T\to\infty} P\left[\sup_{\|\theta-\theta_0\|<\delta} \frac{\sqrt{T}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\|}{1+\sqrt{T}\,\|\theta-\theta_0\|} > \eta\right]$$

$$\leqslant \limsup_{T\to\infty} P\left[\sup_{\|\theta-\theta_0\|<\delta} \sqrt{T}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\| > \eta\right] < \varepsilon \tag{C.14}$$

Similarly, it can be shown that when $S/T \to 0$,

$$\limsup_{S\to\infty} P\left[\sup_{\|\theta-\theta_0\|<\delta} \frac{\sqrt{S}\,\|\boldsymbol{g}_{T,S}(\theta) - \boldsymbol{g}_{T,S}(\theta_0) - \boldsymbol{g}_0(\theta)\|}{1+\sqrt{S}\,\|\theta-\theta_0\|} > \eta\right] < \varepsilon \tag{C.15}$$

□

Proof of Proposition 3. First, we prove the consistency of the numerical derivatives $\hat{\mathbf{G}}_{T,S}$. This part of the proof is similar to that of Theorem 7.4 in Newey and McFadden (1994). We will consider one-sided derivatives first, with the same arguments applying to two-sided derivatives. First we consider the case where $S/T \to \infty$ or $S/T \to k > 0$ as $T, S \to \infty$. We know that $\left\|\hat{\theta}_{T,S} - \theta_0\right\| = O_p\left(T^{-1/2}\right)$ by the conclusion of Proposition 2. Also, by assumption we have $\varepsilon_{T,S} \to 0$ and $\varepsilon_{T,S}\sqrt{T} \to \infty$, so

$$\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S} - \theta_0\right\| \leqslant \left\|\hat{\theta}_{T,S} - \theta_0\right\| + \left\|\mathbf{e}_k\varepsilon_{T,S}\right\| = O_p\left(T^{-1/2}\right) + O\left(\varepsilon_{T,S}\right) = O_p\left(\varepsilon_{T,S}\right)$$

(Recall that $\mathbf{e}_k$ is the k^{th} unit vector.) In the proof of Proposition 2, it is shown that $\exists\delta$ such that

$$\sup_{\|\theta-\theta_0\|\leqslant\delta} \sqrt{T}\left\|\mathbf{g}_{T,S}(\theta) - \mathbf{g}_{T,S}(\theta_0) - \mathbf{g}_0(\theta)\right\| / \left[1 + \sqrt{T}\left\|\theta - \theta_0\right\|\right] = o_p(1)$$

Substituting $\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}$ for θ, then for T, S large, it follows that

$$\frac{\sqrt{T}\left\|\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right) - \mathbf{g}_{T,S}(\theta_0) - \mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)\right\|}{\left[1 + \sqrt{T}\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S} - \theta_0\right\|\right]} \leqslant o_p(1)$$

and

$$\begin{aligned}
&\left\|\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right) - \mathbf{g}_{T,S}(\theta_0) - \mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)\right\| \\
&\leqslant \left[1 + \sqrt{T}\underbrace{\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S} - \theta_0\right\|}_{=O_p\left(\varepsilon_{T,S}\right)}\right] o_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \sqrt{T}O_p\left(\varepsilon_{T,S}\right) o_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\varepsilon_{T,S}\right) o_p(1) \\
&= o_p\left(\varepsilon_{T,S}\right) \qquad \text{(C.16)}
\end{aligned}$$

On the other hand, since $\mathbf{g}_0(\theta)$ is differentiable at θ_0 with derivative $\mathbf{G}_0$ by Assumption 4(ii), a Taylor expansion of $\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)$ around θ_0 is

$$\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right) = \mathbf{g}_0(\theta_0) + \mathbf{G}_0 \cdot \left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S} - \theta_0\right) + o\left(\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S} - \theta_0\right\|\right)$$

with $\mathbf{g}_0(\theta_0) = 0$. Then divide by $\varepsilon_{T,S}$,

$$\begin{aligned}\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S} =&\mathbf{G}_0\cdot\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\theta_0\right)/\varepsilon_{T,S}\\&+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\theta_0\right\|\right)\end{aligned}$$

and

$$\begin{aligned}\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k =&\mathbf{G}_0\cdot\left(\hat{\theta}_{T,S}-\theta_0\right)/\varepsilon_{T,S}\\&+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\theta_0\right\|\right)\end{aligned}$$

The triangle inequality implies that

$$\begin{aligned}\left\|\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k\right\| \leqslant&\left\|\mathbf{G}_0\cdot\left(\hat{\theta}_{T,S}-\theta_0\right)/\varepsilon_{T,S}\right\| && \text{(C.17)}\\&+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\theta_0\right\|\right)\\=&\frac{1}{\sqrt{T}\varepsilon_{T,S}}\left\|\mathbf{G}_0\cdot\sqrt{T}\left(\hat{\theta}_{T,S}-\theta_0\right)\right\| && \text{(C.18)}\\&+\varepsilon_{T,S}^{-1}\left\|\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\theta_0\right\|o(1)\\=&o(1)\,O_p(1)+\varepsilon_{T,S}^{-1}O_p(\varepsilon_{T,S})\,o(1) && \text{(C.19)}\\=&o_p(1)\end{aligned}$$

Combining the inequalities in equations (C.16) and (C.19) gives

$$\begin{aligned}&\left(\frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}(\theta_0)}{\varepsilon_{T,S}}-\mathbf{G}_0\mathbf{e}_k\right)\\=&\left(\frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}(\theta_0)-\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)}{\varepsilon_{T,S}}\right)\\+&\left(\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k\right)\end{aligned}$$

$$
\begin{aligned}
&\left\| \frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\theta_0\right)}{\varepsilon_{T,S}}-\mathbf{G}_0\mathbf{e}_k \right\| \\
&\leqslant \left\| \frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\theta_0\right)-\mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)}{\varepsilon_{T,S}} \right\| \\
&+\left\| \mathbf{g}_0\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k \right\| \\
&\leqslant o_p\left(1\right)
\end{aligned}
$$

Then,

$$
\frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\theta_0\right)}{\varepsilon_{T,S}} \xrightarrow{p} \mathbf{G}_0\mathbf{e}_k
$$

and the same arguments can be applied to the two-sided derivative:

$$
\frac{\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}-\mathbf{e}_k\varepsilon_{T,S}\right)}{2\varepsilon_{T,S}} \xrightarrow{p} \mathbf{G}_0\mathbf{e}_k
$$

This holds for each column $k = 1, 2..., p$. Thus $\hat{\mathbf{G}}_{T,S} \xrightarrow{p} \mathbf{G}_0$.

In the case where $S/T \to 0$ as $T, S \to \infty$, the proof for the consistency of $\hat{\mathbf{G}}_{T,S}$ is done in the similar way using the following facts:

$$
\left\|\hat{\theta}_{T,S}-\theta_0\right\| = O_p\left(S^{-1/2}\right) \tag{C.20}
$$

and $\exists\delta$

$$
\sup_{\|\theta-\theta_0\|\leqslant\delta} \sqrt{S}\left\|\mathbf{g}_{T,S}\left(\theta\right)-\mathbf{g}_{T,S}\left(\theta_0\right)-\mathbf{g}_0\left(\theta\right)\right\|/\left[1+\sqrt{S}\left\|\theta-\theta_0\right\|\right] = o_p\left(1\right) \tag{C.21}
$$

Next, we show the consistency of $\hat{\mathbf{\Sigma}}_{T,B}$. If μ_t and σ_t are known constant, or if ϕ_0 is known, then the result follows from Theorems 5 and 6 of Fermanian, Radulović and Wegkamp (2004). When ϕ_0 is estimated, the result is obtained by combining the results in Fermanian, *et al.* with those of Rémillard (2010), see the Proof of Proposition 3 in the paper for details. $\square$

Proof of Proposition 4. First consider $S/T \to \infty$ or $S/T \to k > 0$. A Taylor expansion of $\mathbf{g}_0\left(\hat{\theta}_{T,S}\right)$ around θ_0 yields

$$\sqrt{T}\mathbf{g}_0\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_0\left(\theta_0\right) + \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + o\left(\sqrt{T}\left\|\hat{\theta}_{T,S} - \theta_0\right\|\right) \tag{C.22}$$

and since $\mathbf{g}_0\left(\theta_0\right) = 0$ and $\sqrt{T}\left\|\hat{\theta}_{T,S} - \theta_0\right\| = O_p\left(1\right)$

$$\sqrt{T}\mathbf{g}_0\left(\hat{\theta}_{T,S}\right) = \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + o_p\left(1\right) \tag{C.23}$$

Then consider the following expansion of $\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)$ around θ_0

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}\left(\theta_0\right) + \hat{\mathbf{G}}_{T,S} \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) + \mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) \tag{C.24}$$

where the remaining term is captured by $\mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right)$. Combining equations (C.23) and (C.24) we obtain

$$\begin{aligned}\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) - \mathbf{g}_{T,S}\left(\theta_0\right) - \mathbf{g}_0\left(\hat{\theta}_{T,S}\right)\right] &= \left(\hat{\mathbf{G}}_{T,S} - \mathbf{G}_0\right) \cdot \sqrt{T}\left(\hat{\theta}_{T,S} - \theta_0\right) \\ &\quad + \mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) + o_p\left(1\right)\end{aligned}$$

Lemma 7 shows the stochastic equicontinuity of $\mathbf{v}_{T,S}\left(\theta\right)$, which implies (see proof of Proposition 2) that

$$\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) - \mathbf{g}_{T,S}\left(\theta_0\right) - \mathbf{g}_0\left(\hat{\theta}_{T,S}\right)\right] = o_p\left(1\right)$$

By Proposition 3, $\hat{\mathbf{G}}_{T,S}-\mathbf{G}_0 = o_p(1)$, which implies $\mathbf{R}_{T,S}\left(\hat{\theta}_{T,S}\right) = o_p(1)$. Thus, we obtain the expansion of $\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)$ around θ_0 :

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}(\theta_0) + \hat{\mathbf{G}}_{T,S}\cdot\sqrt{T}\left(\hat{\theta}_{T,S}-\theta_0\right) + o_p(1) \tag{C.25}$$

The remainder of the proof is the same as in standard GMM applications: From the proof of Proposition 2, we have $\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \xrightarrow{d} N(0,\mathbf{\Sigma}_0)$ and rewrite this as $-\mathbf{\Sigma}_0^{-1/2}\sqrt{T}\mathbf{g}_{T,S}(\theta_0) \equiv \mathbf{u}_{T,S} \xrightarrow{d} \mathbf{u} \sim N(0,\mathbf{I})$, and from Proposition 2, we have $\sqrt{T}\left(\hat{\theta}_{T,S}-\theta_0\right) = \left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\mathbf{\Sigma}_0^{1/2}\mathbf{u}_{T,S} + o_p(1)$. By these two equations and Proposition 3, equation (C.25) becomes

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = -\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\mathbf{u}_{T,S} + \hat{\mathbf{G}}_{T,S}\left(\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,S}\right)^{-1}\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\mathbf{u}_{T,S} + o_p(1) \tag{C.26}$$

$$= -\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\hat{\mathbf{R}}\mathbf{u}_{T,S} + o_p(1)$$

where $\hat{\mathbf{R}} \equiv \left(\mathbf{I} - \hat{\mathbf{\Sigma}}_{T,B}^{-1/2}\hat{\mathbf{G}}_{T,S}\left(\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,S}\right)^{-1}\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\right)$. The test statistic is

$$T\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right)'\hat{\mathbf{W}}_T\mathbf{g}_{T,S}\left(\hat{\theta}_{T,S}\right) = \mathbf{u}_{T,S}'\hat{\mathbf{R}}'\hat{\mathbf{\Sigma}}_{T,B}^{1/2\prime}\hat{\mathbf{W}}_T\hat{\mathbf{\Sigma}}_{T,B}^{1/2}\hat{\mathbf{R}}\mathbf{u}_{T,S} + o_p(1) \tag{C.27}$$

$$= \mathbf{u}'\mathbf{R}_0'\mathbf{\Sigma}_0^{1/2\prime}\mathbf{W}_0\mathbf{\Sigma}_0^{1/2}\mathbf{R}_0\mathbf{u} + o_p(1)$$

where $\mathbf{R}_0 \equiv \left(\mathbf{I} - \mathbf{\Sigma}_0^{-1/2}\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\mathbf{\Sigma}_0^{1/2}\right)$. When $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$, $\hat{\mathbf{R}}$ is symmetric and idempotent with $rank\left(\hat{\mathbf{R}}\right) = tr\left(\hat{\mathbf{R}}\right) = m - p$, and the test statistic converges to a χ^2_{m-p} random variable, as usual. In general, the asymptotic distribution is a sample-dependent combination of m independent standard Normal variables, namely that of $\mathbf{u}'\mathbf{R}_0'\mathbf{\Sigma}_0^{1/2\prime}\mathbf{W}_0\mathbf{\Sigma}_0^{1/2}\mathbf{R}_0\mathbf{u}$ where $\mathbf{u} \sim N(0,\mathbf{I})$.

When $S/T \to 0$, a similar proof can be given using Taylor expansion of $\mathbf{g}_0\left(\hat{\theta}_{T,S}\right)$

$$\sqrt{S}\mathbf{g}_0\left(\hat{\theta}_{T,S}\right) = \sqrt{S}\mathbf{g}_0\left(\theta_0\right) + \mathbf{G}_0 \cdot \sqrt{S}\left(\hat{\theta}_{T,S} - \theta_0\right) + o\left(\sqrt{S}\left\|\hat{\theta}_{T,S} - \theta_0\right\|\right) \tag{C.28}$$

□

C.2 Implementation of the SMM estimator

This section provides further details on the constricution of the SMM objective function and the estimation of the parameter.

Our estimator is based on matching sample dependence measures (rank correlation, quantile dependence, etc) to measures of dependence computed on simulated data from the model evaluated at a given parameter θ. The sample dependence measures are stacked into a vector $\hat{\mathbf{m}}_T$, and the corresponding measures on the simulated data are stacked into a vector $\tilde{\mathbf{m}}_S\left(\theta\right)$. Re-stating equation (3.9) from the paper, our estimator is:

$$\hat{\theta}_{T,S} \equiv \underset{\theta \in \boldsymbol{\Theta}}{\arg\min}\ \mathbf{g}'_{T,S}\left(\theta\right)\hat{\mathbf{W}}_T\mathbf{g}_{T,S}\left(\theta\right) \tag{C.29}$$

where $\mathbf{g}_{T,S}\left(\theta\right) \equiv \hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S\left(\theta\right)$.

We now describe the construction of the SMM objective function. All dependence measures used in this paper are based on the estimated standardized residuals, which are constructed as:

$$\hat{\eta}_t \equiv \frac{\mathbf{Y}_t - \mu_t(\hat{\phi})}{\sigma_t(\hat{\phi})} \tag{C.30}$$

We then compute pair-wise dependence measures such as those in equations (4) and (5) of the paper, eg, $\hat{\rho}^{ij}$ and $\hat{\lambda}_q^{ij}$. For quantile dependence we set $q \in \{0.05, 0.10, 0.90, 0.95\}$.

The copula models we consider all satisfy an "exchangeability" property, and we use that when constructing the moments to use in the estimator. Specifically, we calculate moments $\hat{\mathbf{m}}_T$ as:

$$\hat{\mathbf{m}}_T = \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\begin{array}{ccccc} \hat{\rho}^{ij} & \hat{\lambda}^{ij}_{0.05} & \hat{\lambda}^{ij}_{0.10} & \hat{\lambda}^{ij}_{0.90} & \hat{\lambda}^{ij}_{0.95} \end{array} \right]' \tag{C.31}$$

Next we simulate data $\{\mathbf{X}_s\left(\theta\right)\}_{s=1}^{S}$ from distribution $\mathbf{F}_x\left(\theta\right)$, and compute the vector of dependence measures $\tilde{\mathbf{m}}_S\left(\theta\right)$. It is critically important in this step to keep the random number generator seed *fixed* across simulations, see Gouriéroux and Monfort (1996, *Simulation-Based Econometric Methods*, Oxford University Press). Failing to do so makes the simulated data "jittery" across function evaluations, and the numerical optimization algorithm will fail to converge.

Finally, we specify the weight matrix. In this paper we choose either $\hat{\mathbf{W}}_T = \mathbf{I}$ or $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}^{-1}_{T,B}$. Note that for our estimation problem the estimated efficient weight matrix, $\hat{\mathbf{\Sigma}}^{-1}_{T,B}$, depends on the covariance matrix of the vector of sample dependence measures, and not on the parameters of the model. Thus unlike some GMM or SMM estimation problems, this estimator does not require an initial estimate of the unknown parameter.

We use numerical optimization procedure to find $\hat{\theta}_{T,S}$. As our objective function is not differentiable we cannot use procedures that rely on analytical or numerical derivatives (such as familiar Newton or "quasi-Newton" algorithms). We use "fminsearch" in Matlab, which is a simplex search algorithm that does not require derivatives. As with all numerical optimization procedures, some care is required to ensure that a global optimum has been found. In each estimation, we consider many different starting values for the algorithm, and choose the resulting parameter estimate that leads to the smallest value of the objective function. The models considered here are relatively small, with up to three unknown parameters, but when the number of unknown parameters is large more care is required to ensure that a global optimum has been found, see Judd (1998, *Numerical Methods in Economics*, MIT Press) for more discussion.

C.3 Implementation of MLE for factor copulas

Consider a simple factor model:

$$X_i = Z + \varepsilon_i, \quad i = 1, 2, ..., N$$

$$Z \sim F_Z, \ \varepsilon_i \sim iid\ F_\varepsilon, \ \ \varepsilon_i \perp\!\!\!\perp Z \ \forall \ i$$

$$[X_1, ..., X_N]' \equiv \mathbf{X} \sim \mathbf{F}_x = \mathbf{C}\left(G, ..., G\right)$$

To obtain the copula density $\mathbf{c}$ we must first obtain the joint density, $\mathbf{f}_x$, and the marginal density, g. These can be obtained using numerical integration to "integrate out" the latent common factor, Z. First, note that

$$f_{x_i|z}\left(x_i|z\right) = f_\varepsilon\left(x_i - z\right)$$

$$F_{x_i|z}\left(x_i|z\right) = F_\varepsilon\left(x_i - z\right)$$

$$\text{and} \quad f_{x|z}\left(x_1, \ldots, x_N|z\right) = \prod\nolimits_{i=1}^{N} f_\varepsilon\left(x_i - z\right)$$

Then the marginal density and marginal distribution of X_i are:

$$g\left(x\right) = \int_{-\infty}^{\infty} f_{x,z}\left(x, z\right) dz = \int_{-\infty}^{\infty} f_{x|z}\left(x|z\right) f_z\left(z\right) dz = \int_{-\infty}^{\infty} f_\varepsilon\left(x - z\right) f_z\left(z\right) dz$$

$$G\left(x\right) = \int_{-\infty}^{\infty} \Pr\left[X \leqslant x|Z = z\right] f_z\left(z\right) dz = \int_{-\infty}^{\infty} F_\varepsilon\left(x - z\right) f_z\left(z\right) dz$$

The joint density is similarly obtained:

$$f_x\left(x_1, \ldots, x_N\right) = \int_{-\infty}^{\infty} f_{x|z}\left(x_1, \ldots, x_N|z\right) f_z\left(z\right) dz = \int_{-\infty}^{\infty} \prod\nolimits_{i=1}^{N} f_\varepsilon\left(x_i - z\right) f_z\left(z\right) dz$$

From these, we obtain the copula density:

$$c(u_1, \ldots, u_N) = \frac{f_x\left(G^{-1}(u_1), \ldots, G^{-1}(u_N)\right)}{\prod_{i=1}^{N} g\left(G^{-1}(u_i)\right)}$$

We approximate the above integrals using Gauss-Legendre quadrature, see Judd (1998) for details and discussion. We use the probability integral transformation of Z to convert the above unbounded integals to integrals on $[0,1]$, for example:

$$g(x) = \int_{-\infty}^{\infty} f_\varepsilon(x - z) f_z(z)\, dz = \int_0^1 f_\varepsilon\left(x - F_z^{-1}(u)\right) du$$

A key choice in quadrature methods is the number of "nodes" to use in approximating the integral. We ran simulations using 50, 150, and 250 nodes, and found that the accuracy of the resulting MLE was slightly better for 150 than 50 nodes, and not different for 250 compared with 150 nodes. Thus in the paper we report results for MLE based on quadrature using 150 nodes.

C.4 Additional tables

Table C.1: Simulation results for iid data with optimal weight matrix

	Clayton			*Normal*		*Factor copula*		
	GMM	SMM	SMM*	GMM	SMM	SMM		
	κ	κ	κ	ρ	ρ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	0.5	0.5	1.00	0.25	-0.50
				$N = 2$				
Bias	-0.018	-0.020	-0.018	-0.001	0.000	0.016	-0.026	-0.094
St dev	0.085	0.092	0.091	0.025	0.026	0.144	0.119	0.189
Median	0.984	0.977	0.981	0.497	0.500	0.999	0.200	-0.557
90-10%	0.224	0.247	0.233	0.070	0.069	0.374	0.332	0.447
Time	0.07	515	51	0.41	0.67		112	
				$N = 3$				
Bias	0.008	0.010	0.006	-0.003	-0.003	0.022	-0.009	-0.057
St dev	0.063	0.073	0.068	0.021	0.022	0.110	0.103	0.146
Median	0.996	1.008	1.002	0.495	0.498	1.006	0.238	-0.540
90-10%	0.160	0.172	0.165	0.054	0.061	0.294	0.261	0.366
Time	0.12	1398	59	0.29	1.60		138	
				$N = 10$				
Bias	-0.003	-0.004	-0.005	-0.004	-0.004	0.019	-0.010	-0.023
St dev	0.047	0.049	0.050	0.014	0.015	0.097	0.078	0.085
Median	0.993	0.997	0.997	0.497	0.495	1.006	0.251	-0.514
90-10%	0.121	0.126	0.127	0.036	0.037	0.248	0.189	0.165
Time	1	22521	170	0.34	3		358	

Notes: The simulation design is the same as that of Table 3.1 in Chapter 3 except that we use the optimal weight matrix for W.

Table C.2: Simulation results for AR-GARCH data with optimal weight matrix

	Clayton			*Normal*		*Factor copula*		
	GMM	SMM	SMM*	GMM	SMM	SMM		
	κ	κ	κ	ρ	ρ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	0.5	0.5	1.00	0.25	-0.50
				$N = 2$				
Bias	-0.021	-0.017	-0.014	-0.002	-0.001	0.018	-0.022	-0.083
St dev	0.087	0.097	0.097	0.026	0.026	0.154	0.121	0.188
Median	0.980	0.989	0.987	0.498	0.498	0.997	0.209	-0.553
90-10%	0.225	0.247	0.258	0.070	0.069	0.399	0.346	0.485
Time	0.06	531	60	0.39	0.69		119	
				$N = 3$				
Bias	0.002	-0.004	-0.001	-0.003	-0.003	0.021	-0.009	-0.061
St dev	0.063	0.066	0.068	0.021	0.023	0.114	0.106	0.151
Median	0.995	0.990	0.991	0.495	0.497	1.018	0.243	-0.548
90-10%	0.153	0.166	0.164	0.052	0.058	0.299	0.278	0.336
Time	0.12	1613	76	0.33	1.50		135	
				$N = 10$				
Bias	-0.006	-0.005	-0.007	-0.005	-0.005	0.014	-0.013	-0.027
St dev	0.047	0.051	0.050	0.014	0.015	0.093	0.078	0.097
Median	0.991	0.997	0.993	0.496	0.494	1.000	0.250	-0.513
90-10%	0.120	0.136	0.134	0.037	0.040	0.229	0.193	0.187
Time	2	25492	175	0.41	4		361	

Notes: The simulation design is the same as that of Table 3.2 in Chapter 3 except that we use the optimal weight matrix for W.

Table C.3: Simulation results on coverage rates with optimal weight matrix

	Clayton		Normal		Factor copula			
	κ	J	ρ	J	σ^2	ν^{-1}	λ	J
				$N=2$				
$\varepsilon_{T,S}$								
0.1	89	95	93	99	97	99	96	96
0.01	56		93		95	99	97	
0.001	9		80		77	79	80	
0.0001	1		16		40	54	56	
				$N=3$				
$\varepsilon_{T,S}$								
0.1	91	98	88	95	98	99	97	99
0.01	70		88		98	99	96	
0.001	10		82		88	86	86	
0.0001	0		41		51	59	48	
				$N=10$				
$\varepsilon_{T,S}$								
0.1	93	100	87	97	95	96	94	100
0.01	79		87		94	94	93	
0.001	20		87		89	84	92	
0.0001	5		64		70	70	73	

Notes: The simulation design is the same as that of Table 3.3 in Chapter 3 except that we use the optimal weight matrix for W. The numbers in column J present the percentage of simulations for which the test statistic of over-identifying restrictions test described in Section 3.3 was smaller than its critical value from chi square distribution under 95% confidence level (this test does not require a choice of step size for the numerical derivative, $\varepsilon_{T,S}$, and so we have only one value per model).

Table C.4: Summary statistics on the daily stock returns

	Bank of America	Bank of N.Y.	Citi Group	Goldman Sachs	JP Morgan	Morgan Stanley	Wells Fargo
Mean	0.038	0.015	-0.020	0.052	0.041	0.032	0.047
Std dev	3.461	2.797	3.817	2.638	2.966	3.814	2.965
Skewness	1.048	0.592	1.595	0.984	0.922	4.982	2.012
Kurtosis	28.190	18.721	43.478	18.152	16.006	119.757	30.984

Notes: This table presents some summary statistics of the seven daily equity returns data used in the empirical analysis.

Table C.5: Parameter estimates for the conditional mean and variance models

	BoA	BoNY	Citi	GS	JPM	MS	WF
Constant (ϕ_0)	0.038	0.017	-0.019	0.058	0.043	0.031	0.051
$r_{i,t-1}$	0.020	-0.151	0.053	-0.156	-0.035	0.004	-0.078
$r_{m,t-1}$	-0.053	-0.011	0.029	0.282	-0.141	0.063	-0.099
Constant (ω)	0.009	0.069	0.019	0.034	0.014	0.036	0.008
$\sigma^2_{i,t-1}$	0.931	0.895	0.901	0.953	0.926	0.922	0.926
$\varepsilon^2_{i,t-1}$	0.031	0.017	0.036	0.000	0.025	0.002	0.021
$\varepsilon^2_{i,t-1} \cdot \mathbf{1}_{\{\varepsilon_{i,t-1} \leqslant 0\}}$	0.048	0.079	0.123	0.077	0.082	0.135	0.108
$\varepsilon^2_{m,t-1}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\varepsilon^2_{m,t-1} \cdot \mathbf{1}_{\{\varepsilon_{m,t-1} \leqslant 0\}}$	0.068	0.266	0.046	0.012	0.064	0.077	0.013

Notes: This table presents the estimated models for the conditional mean (top panel) and conditional variance (lower panel).

Appendix D

Appendix to Chapter 4

D.1 Proofs

Proof of Proposition 1. (i) Consider the evolution equation for λ_{it} :

$$\log \lambda_{it} = \omega_i + \beta \log \lambda_{i,t-1} + \alpha s_{i,t-1}, \quad i = 1, 2, ..., N$$

where $s_{i,t-1} \equiv \partial \log \mathbf{c}(\mathbf{u}_{t-1}; \lambda_{t-1}, \nu_z, \psi_z, \nu_\varepsilon)/\partial \lambda_{i,t-1}$. Creal, *et al.* (2013) show $E_{t-1}[s_{i,t}] = 0$, so:

$$E[\log \lambda_{it}] = \omega_i + \beta E[\log \lambda_{i,t-1}] = \frac{\omega_i}{1-\beta}$$

under stationarity of $\{\lambda_t\}$, which holds by assumption 1(b). So we have $\omega_i = E[\log \lambda_{it}](1-\beta)$, and we can re-write our GAS equation in "variance targeting" form:

$$\log \lambda_{it} = E[\log \lambda_{it}](1-\beta) + \beta \log \lambda_{i,t-1} + \alpha s_{i,t-1}$$

The objective of this proposition is to find an estimate of $E[\log \lambda_{it}]$ based on observable data.

Note that the linear correlation between (X_i, X_j) is

$$\rho^L_{ij,X} \equiv Corr\left[X_i, X_j\right] = \frac{\lambda_i \lambda_j}{\sqrt{\left(1+\lambda_i^2\right)\left(1+\lambda_j^2\right)}} \equiv g\left(\lambda_i, \lambda_j\right) \tag{D.1}$$

$$\text{and} \quad \mathbf{R}^L_X \equiv Corr\left[\mathbf{X}\right] = G\left(\lambda\right)$$

By assumption 1(a), this is an exactly- $(N = 3)$ or over- $(N > 3)$ identified system, as we have N parameters $\lambda \equiv [\lambda_1, ..., \lambda_N]'$ and $N(N-1)/2$ correlations. Note that by Assumption 1(d) we have a corresponding exactly- or over-identified system for the rank correlation matrix:

$$\mathbf{R}_X = \varphi\left(\mathbf{R}^L_X\right) = \varphi\left(G\left(\lambda\right)\right) \tag{D.2}$$

(In a slight abuse of notation, we let $\varphi\left(\mathbf{R}^L_X\right)$ map the entire linear correlation matrix to the rank correlation matrix.) Define the (exponential of the) inverse of the function $\varphi \circ G$ as H, so that $\log \lambda = H\left(\rho_X\right)$, where $\rho_X \equiv vech\left(\mathbf{R}_X\right)$. The function H is not known in closed form but it can be obtained by a simple and fast optimization problem:

$$H\left(\rho_X\right) = \arg\min_{\mathbf{a}} \ \left(vech\left\{\varphi\left(G\left(\mathbf{a}\right)\right)\right\} - \rho_X\right)' \left(vech\left\{\varphi\left(G\left(\mathbf{a}\right)\right)\right\} - \rho_X\right) \tag{D.3}$$

This is the GMM analog to the usual method-of-moments estimator used in variance targeting.

Under Assumption 1(c) the function $H\left(\rho_X\right)$ is linear, so

$$E\left[\log \lambda_t\right] = E[H(\rho_{t,X})] = H(\bar{\rho}_X)$$

where $\bar{\rho}_X \equiv E[\rho_{t,X}]$.

Finally, we exploit the fact that $RankCorr\left[\mathbf{X}\right]$ is identical to $RankCorr\left[\eta\right]$ by

Assumption 1(a) and Theorem 5.1.6 of Nelsen (2006). So we obtain:

$$E\left[\log \lambda_t\right] = H(\bar{\rho}_X) = H(\bar{\rho}_\eta)$$

(ii) We use as our "VT estimator" the sample analog of the above expression:

$$\widehat{\overline{\log \lambda}} = H(\hat{\rho}_\eta)$$

First note that, since the marginal distributions of η_t are known, sample rank correlations are a linear functions of a sample moment, see Nelsen (2006, Chapter 5) for example:

$$\hat{\rho}^S_{ij,\eta} = -3 + \frac{12}{T}\sum_{t=1}^{T} F_i\left(\eta_{i,t}\right) F_j\left(\eta_{j,t}\right)$$

Our estimate of $E\left[\log \lambda_{it}\right]$ is obtained in equation (D.3) as:

$$\widehat{\overline{\log \lambda}} = \arg\min_{\mathbf{a}}\ \bar{\mathbf{m}}_T\left(\mathbf{a}\right)' \bar{\mathbf{m}}_T\left(\mathbf{a}\right)$$

$$\text{where} \quad \bar{\mathbf{m}}_T\left(\mathbf{a}\right) \equiv vech\{\varphi\left(G\left(\mathbf{a}\right)\right)\} - \hat{\rho}^S_\eta$$

The element of $\bar{\mathbf{m}}_T$ corresponding to the (i, j) element of the correlation matrix is:

$$\bar{m}^{(i,j)}_T\left(\mathbf{a}\right) = \left[\varphi\left(G\left(\mathbf{a}\right)\right)\right]_{(i,j)} + 3 - \frac{12}{T}\sum_{t=1}^{T} F_i\left(\eta_{i,t}\right) F_j\left(\eta_{j,t}\right)$$

Thus $\widehat{\overline{\log \lambda}}$ is a standard GMM estimator for $N \geqslant 3$. □

D.2 Obtaining the factor copula likelihood

The factor copula introduced in Oh and Patton (2012) does not have a likelihood in closed form, it is relatively simple to obtain the likelihood using numerical integration. Consider the factor structure in equation (4.5) and (4.6). Our objective is to obtain the copula density of $\mathbf{X}_t$.

$$c_t\left(u_1,...,u_N\right)=\frac{f_{xt}\left(G_{1t}^{-1}\left(u_1\right),...,G_{Nt}^{-1}\left(u_N\right)\right)}{g_{1t}\left(G_{1t}^{-1}\left(u_1\right)\right)\cdot\cdots\cdot g_{Nt}\left(G_{Nt}^{-1}\left(u_N\right)\right)} \tag{D.4}$$

where $f_{xt}\left(x_1,...,x_N\right)$ is the joint density of $\mathbf{X}_t$, $g_{it}\left(x_i\right)$ is the marginal density of X_i, and $c_t\left(u_1,...,u_N\right)$ is the copula density. To construct copula density, we need each of the functions that appear on the right-hand side above: $g_{it}\left(x_i\right)$, $G_{it}\left(x_i\right)$, $f_{xt}\left(x_1,...,x_N\right)$ and $G_{it}^{-1}\left(u_i\right)$.

The independence of Z and ε_i implies that:

$$f_{X_i|Z,t}\left(x_i|z\right)=f_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)$$

$$F_{X_i|Z,t}\left(x_i|z\right)=F_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)$$

$$f_{\mathbf{X}|Z,t}\left(x_{1,...,}x_N|z\right)=\prod_{i=1}^{N}f_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)$$

With these conditional distributions, one dimensional integration gives the marginals:

$$\begin{aligned}g_{it}\left(x_i\right)=\int_{-\infty}^{\infty}f_{X_i,Z,t}\left(x_i,z\right)dz&=\int_{-\infty}^{\infty}f_{X_i|Z,t}\left(x_i|z\right)f_{Z,t}\left(z\right)dz\\&=\int_{-\infty}^{\infty}f_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)f_{Z,t}\left(z\right)dz\end{aligned}$$

and similarly

$$G_{it}\left(x_i\right)=\int_{-\infty}^{\infty}F_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)f_{Z,t}\left(z\right)dz$$

$$f_{xt}\left(x_1,...,x_N\right)=\int_{-\infty}^{\infty}\prod_{i=1}^{N}f_{\varepsilon_i}\left(x_i-\lambda_{it}z\right)f_{Z,t}\left(z\right)dz$$

We use a change of variables, $U \equiv F_{Z,t}(z)$, to convert these to bounded integrals:

$$g_{it}(x_i) = \int_0^1 f_{\varepsilon_i}\left(x_i - \lambda_{it} F_{Z,t}^{-1}(u)\right) du$$

$$G_{it}(x_i) = \int_0^1 F_{\varepsilon_i}\left(x_i - \lambda_{it} F_{Z,t}^{-1}(u)\right) du$$

$$f_{xt}(x_1, ..., x_N) = \int_0^1 \prod_{i=1}^{N} f_{\varepsilon_i}\left(x_i - \lambda_{it} F_{Z,t}^{-1}(u)\right) du$$

Thus the factor copula density requires the computation of just one-dimensional integrals. (For a factor copula with J common factors the integral would be J-dimensional.) We use Gauss-Legendre quadrature for the integration, using Q "nodes," (see Judd (1998) for details) and we choose Q on the basis of a small simulation study described below.

Finally, we need a method to invert $G_{it}(x_i)$, and note from above that this is a function of both x and the factor loading λ_{it}, with $G_{it} = G_{js}$ if $\lambda_{it} = \lambda_{js}$. We estimate the inverse of G_{it} by creating a grid of 100 points for x in the interval $[x_{\min}, x_{\max}]$ and 50 points for λ in the interval $[\lambda_{\min}, \lambda_{\max}]$, and then evaluating G at each of those points. We then use two-dimensional linear interpolation to obtain $G^{-1}(u; \lambda)$ given u and λ. This two-dimensional approximation substantially reduces the computational burden, especially when λ is time-varying, as we can evaluate the function G prior to estimation, rather than re-estimating it for each likelihood evaluation.

We conducted a small Monte Carlo simulation to evaluate the accuracy of this numerical approximation. We use quadrature nodes $Q \in \{10, 50, 150\}$ and $[x_{start}, x_{end}] = [-30, 30]$, $[\lambda_{start}, \lambda_{end}] = [0, 6]$ for the numerical inversion. For this simulation, we considered the factor copula implied by the following structure:

$$X_i = \lambda_0 Z_t + \varepsilon_i, \quad i = 1, 2 \tag{D.5}$$

$$\text{where} \quad Z_t \sim Skew\ t\ (\nu_0, \psi_0), \quad \varepsilon_{it} \sim iid\ t(\nu_0), \quad Z \perp\!\!\!\perp \varepsilon_i\ \forall\ i$$

where $\lambda_0 = 1$, $\nu_0^{-1} = 0.25$ and $\psi_0 = -0.5$. At each replication, we simulate $\mathbf{X} = [X_1, X_2]$ 1000 times, and apply empirical distribution functions to transform $\mathbf{X}$ to $\mathbf{U} = [U_1, U_2]$. With this $[U_1, U_2]$ we estimate $[\lambda, \nu^{-1}, \psi]$ by numerically approximated maximum likelihood method.

Table D.3 in Appendix D.3 contains estimation results for 100 replications. We find that estimation with only 10 nodes introduces a relatively large bias, in particular for ν^{-1}, consistent with this low number of nodes providing a poor approximation of the tails of this density. Estimation with 50 nodes gives accurate results, and is comparable to those with 150 nodes in that bias and standard deviation are small. We use 50 nodes throughout the paper.

D.3 Additional tables

Table D.1: Simulation results for the "heterogeneous dependence" model

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
ω_1	-0.030	0.004	0.017	-0.022	-0.005	-0.052	0.047
ω_2	-0.029	0.004	0.018	-0.022	-0.005	-0.048	0.043
ω_3	-0.029	0.004	0.016	-0.021	-0.005	-0.043	0.038
ω_4	-0.028	0.003	0.017	-0.023	-0.005	-0.047	0.042
ω_5	-0.028	0.004	0.016	-0.020	-0.005	-0.046	0.041
ω_6	-0.027	0.004	0.016	-0.020	-0.003	-0.044	0.040
ω_7	-0.026	0.003	0.016	-0.022	-0.004	-0.042	0.038
ω_8	-0.026	0.003	0.016	-0.020	-0.005	-0.043	0.038
ω_9	-0.025	0.003	0.015	-0.019	-0.005	-0.041	0.036
ω_{10}	-0.025	0.002	0.016	-0.019	-0.005	-0.041	0.036
ω_{11}	-0.024	0.002	0.015	-0.018	-0.004	-0.038	0.033
ω_{12}	-0.023	0.003	0.013	-0.018	-0.004	-0.037	0.032
ω_{13}	-0.023	0.003	0.014	-0.018	-0.004	-0.038	0.033
ω_{14}	-0.022	0.003	0.012	-0.018	-0.004	-0.035	0.031
ω_{15}	-0.022	0.002	0.013	-0.019	-0.004	-0.043	0.039
ω_{16}	-0.021	0.002	0.013	-0.016	-0.003	-0.034	0.031
ω_{17}	-0.020	0.003	0.011	-0.015	-0.003	-0.032	0.029
ω_{18}	-0.020	0.002	0.013	-0.015	-0.003	-0.033	0.030
ω_{19}	-0.019	0.002	0.012	-0.016	-0.003	-0.031	0.028
ω_{20}	-0.019	0.002	0.011	-0.015	-0.003	-0.033	0.030
ω_{21}	-0.018	0.003	0.010	-0.013	-0.003	-0.028	0.025
ω_{22}	-0.017	0.002	0.010	-0.013	-0.003	-0.028	0.025
ω_{23}	-0.017	0.003	0.009	-0.013	-0.003	-0.025	0.022
ω_{24}	-0.016	0.002	0.010	-0.013	-0.003	-0.024	0.021
ω_{25}	-0.016	0.000	0.010	-0.014	-0.003	-0.030	0.027

Table D.1: Simulation results for the "heterogeneous dependence" model

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
ω_{26}	-0.015	0.001	0.010	-0.012	-0.003	-0.028	0.025
ω_{27}	-0.014	0.000	0.011	-0.011	-0.003	-0.028	0.025
ω_{28}	-0.014	0.001	0.009	-0.011	-0.002	-0.023	0.021
ω_{29}	-0.013	0.000	0.009	-0.011	-0.002	-0.025	0.023
ω_{30}	-0.012	0.001	0.008	-0.010	-0.002	-0.022	0.020
ω_{31}	-0.012	0.000	0.008	-0.010	-0.002	-0.022	0.020
ω_{32}	-0.011	0.001	0.008	-0.008	-0.002	-0.019	0.017
ω_{33}	-0.011	0.001	0.007	-0.009	-0.002	-0.017	0.015
ω_{34}	-0.010	-0.001	0.008	-0.009	-0.002	-0.021	0.019
ω_{35}	-0.009	0.000	0.008	-0.008	-0.002	-0.020	0.018
ω_{36}	-0.009	0.000	0.007	-0.008	-0.002	-0.018	0.016
ω_{37}	-0.008	0.001	0.005	-0.006	-0.001	-0.014	0.013
ω_{38}	-0.008	0.001	0.006	-0.005	-0.001	-0.016	0.015
ω_{39}	-0.007	0.001	0.005	-0.005	-0.001	-0.013	0.012
ω_{40}	-0.006	-0.001	0.005	-0.006	-0.002	-0.015	0.014
ω_{41}	-0.006	-0.003	0.007	-0.007	-0.002	-0.019	0.017
ω_{42}	-0.005	0.000	0.004	-0.005	-0.001	-0.010	0.009
ω_{43}	-0.005	0.001	0.004	-0.003	0.000	-0.009	0.008
ω_{44}	-0.004	0.000	0.004	-0.003	0.000	-0.010	0.010
ω_{45}	-0.003	-0.001	0.005	-0.003	0.000	-0.010	0.010
ω_{46}	-0.003	0.001	0.003	-0.002	0.002	-0.007	0.008
ω_{47}	-0.002	-0.001	0.003	-0.002	0.000	-0.006	0.006
ω_{48}	-0.002	-0.001	0.003	-0.001	0.001	-0.006	0.008
ω_{49}	-0.001	-0.001	0.004	-0.001	0.002	-0.006	0.008
ω_{50}	0.000	-0.002	0.004	-0.002	0.001	-0.007	0.008
ω_{51}	0.000	-0.002	0.003	-0.001	0.001	-0.006	0.007
ω_{52}	0.001	-0.001	0.004	0.000	0.004	-0.003	0.007
ω_{53}	0.002	0.000	0.004	0.000	0.006	-0.003	0.009
ω_{54}	0.002	-0.003	0.004	-0.001	0.002	-0.005	0.007
ω_{55}	0.003	-0.001	0.004	0.001	0.007	-0.003	0.010
ω_{56}	0.003	-0.002	0.003	0.001	0.007	-0.002	0.009
ω_{57}	0.004	0.000	0.004	0.003	0.010	0.000	0.010
ω_{58}	0.005	-0.002	0.004	0.001	0.008	-0.001	0.009
ω_{59}	0.005	-0.002	0.003	0.003	0.008	0.000	0.008
ω_{60}	0.006	-0.002	0.005	0.003	0.010	0.000	0.010

Table D.1: Simulation results for the "heterogeneous dependence" model

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
ω_{61}	0.006	-0.002	0.005	0.003	0.010	0.000	0.010
ω_{62}	0.007	-0.001	0.005	0.004	0.013	0.001	0.012
ω_{63}	0.008	-0.003	0.005	0.003	0.012	0.000	0.011
ω_{64}	0.008	-0.002	0.005	0.004	0.013	0.001	0.012
ω_{65}	0.009	-0.002	0.006	0.005	0.013	0.000	0.013
ω_{66}	0.009	-0.003	0.005	0.006	0.014	0.001	0.013
ω_{67}	0.010	-0.004	0.006	0.005	0.013	0.000	0.013
ω_{68}	0.011	-0.004	0.006	0.005	0.013	0.001	0.013
ω_{69}	0.011	-0.002	0.007	0.008	0.022	0.002	0.020
ω_{70}	0.012	-0.004	0.007	0.007	0.017	0.001	0.016
ω_{71}	0.012	-0.003	0.007	0.009	0.019	0.001	0.017
ω_{72}	0.013	-0.005	0.007	0.007	0.016	0.001	0.015
ω_{73}	0.014	-0.004	0.008	0.008	0.020	0.001	0.019
ω_{74}	0.014	-0.004	0.009	0.008	0.023	0.002	0.021
ω_{75}	0.015	-0.004	0.008	0.009	0.019	0.002	0.017
ω_{76}	0.016	-0.005	0.009	0.008	0.025	0.002	0.023
ω_{77}	0.016	-0.003	0.009	0.011	0.026	0.002	0.024
ω_{78}	0.017	-0.004	0.009	0.010	0.024	0.002	0.022
ω_{79}	0.017	-0.004	0.010	0.011	0.032	0.002	0.030
ω_{80}	0.018	-0.004	0.009	0.012	0.026	0.002	0.024
ω_{81}	0.019	-0.006	0.009	0.011	0.024	0.002	0.022
ω_{82}	0.019	-0.005	0.010	0.012	0.026	0.003	0.024
ω_{83}	0.020	-0.005	0.010	0.012	0.029	0.002	0.027
ω_{84}	0.020	-0.004	0.012	0.013	0.033	0.004	0.030
ω_{85}	0.021	-0.006	0.011	0.014	0.032	0.002	0.030
ω_{86}	0.022	-0.006	0.011	0.014	0.029	0.003	0.026
ω_{87}	0.022	-0.006	0.013	0.015	0.032	0.003	0.029
ω_{88}	0.023	-0.006	0.011	0.014	0.032	0.004	0.028
ω_{89}	0.023	-0.006	0.012	0.016	0.033	0.003	0.030
ω_{90}	0.024	-0.006	0.012	0.016	0.036	0.003	0.033
ω_{91}	0.025	-0.005	0.014	0.017	0.036	0.004	0.033
ω_{92}	0.025	-0.005	0.014	0.018	0.039	0.003	0.036
ω_{93}	0.026	-0.007	0.012	0.018	0.038	0.003	0.035
ω_{94}	0.026	-0.006	0.015	0.018	0.040	0.004	0.036
ω_{95}	0.027	-0.006	0.014	0.018	0.040	0.004	0.036

Table D.1: Simulation results for the “heterogeneous dependence” model

	True	Bias	Std	Median	90%	10%	Diff (90%-10%)
ω_{96}	0.028	-0.007	0.015	0.019	0.042	0.004	0.038
ω_{97}	0.028	-0.006	0.015	0.019	0.042	0.004	0.038
ω_{98}	0.029	-0.006	0.015	0.020	0.045	0.004	0.041
ω_{99}	0.029	-0.008	0.013	0.020	0.038	0.004	0.034
ω_{100}	0.030	-0.007	0.016	0.021	0.040	0.004	0.036
α	0.050	-0.006	0.015	0.045	0.062	0.023	0.039
β	0.980	0.002	0.012	0.983	0.997	0.966	0.031
ν^{-1}	0.200	-0.002	0.009	0.199	0.209	0.186	0.023
ψ_z	0.100	0.008	0.032	0.111	0.152	0.064	0.088

Notes: This table presents results from the simulation study described in Section 4.3

Table D.2: Static copula model estimation results

	Equidep				Block equidep				Heterog			
	Normal		Factor		Normal		Factor		Normal		Factor	
	Pre	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
$\omega_{1\to G}$	0.326 (0.126)	0.817 (0.099)	0.336 (0.118)	0.802 (0.098)	–	–	–	–	–	–	–	–
α_{GAS}	0	0	0	0	0	0	0	0	0	0	0	0
β_{GAS}	0	0	0	0	0	0	0	0	0	0	0	0
ν_z^{-1}	–	–	0.108 (0.027)	0.009 (0.025)	–	–	0.097 (0.028)	0.011 (0.019)	–	–	0.111 (0.034)	0.008 (0.019)
ν_ϵ^{-1}	–	–	0.214 (0.015)	0.205 (0.010)	–	–	0.198 (0.014)	0.213 (0.009)	–	–	0.191 (0.013)	0.187 (0.009)
ψ_z	–	–	0.055 (0.062)	0.080 (0.056)	–	–	0.085 (0.066)	0.092 (0.060)	–	–	0.105 (0.063)	0.114 (0.061)
$\log \mathcal{L}$	36,185		39,508		36,477		39,757		37,652		40,628	
AIC	-72,367		-78,999		-72,934		-79,481		-74,905		-80,844	
BIC	-72,356		-78,956		-72,880		-79,395		-73,824		-79,731	

Notes: This table presents parameter estimates for two versions of the factor copula (Normal and Skew *t*-*t*), each with one of three degrees of heterogeneity of dependence (equidependence, block equidependence, and heterogeneous dependence). All models are imposed to be constant through time, and so the GAS parameters are fixed at zero. Standard errors based on the stationary bootstrap of Politis and Romano (1994) are presented below the estimated parameters. All models are allowed to have a structural break on April 8, 2009 (see Section 4.4.4 of the paper), and we denote parameters from the first and second sub-samples as "Pre" and "Post." The log-likelihood at the estimated parameters and the Akaike and Bayesian Information criteria are presented in the bottom three rows. The intercept parameters (ω_i) for the block equidependence and heterogeneous dependence models are not reported to conserve space.

Table D.3: Simulation results for MLE with different numbers of quadrature nodes

	10 nodes			50 nodes			150 nodes		
	λ	ν^{-1}	ψ	λ	ν^{-1}	ψ	λ	ν^{-1}	ψ
True	1.000	0.250	-0.500	1.000	0.250	-0.500	1.000	0.250	-0.500
Bias	0.039	-0.062	-0.072	0.023	-0.018	-0.039	0.026	-0.002	-0.026
Std	0.126	0.038	0.203	0.144	0.049	0.168	0.135	0.045	0.144
Median	1.021	0.193	-0.532	1.014	0.236	-0.510	1.027	0.251	-0.500
90% quant	1.204	0.239	-0.341	1.208	0.293	-0.339	1.187	0.305	-0.349
10% quant	0.878	0.136	-0.990	0.846	0.163	-0.751	0.856	0.188	-0.704
diff (90%-10%)	0.326	0.103	0.649	0.362	0.130	0.412	0.331	0.118	0.355

Notes: This table presents the results from the simulation study described in Appendix D.2

D.4 "Variance targeting" assumptions

In Figure D.1 we present simulation evidence supporting the applicability of the assumptions underlying Proposition 1 of the paper. In both panels we use a simulation with 50,000 observations to estimate the true functions. The left panel shows the mapping from rank correlation to linear correlation. This mapping changes slightly with the shape parameters $(\theta_z, \theta_\varepsilon)$, but we see that for all choices presented the function is indeed strictly increasing, supporting assumption (d). We further see that for all three shape parameter choices the function φ is close to being the identity function, and we invoke this approximation in our estimation to increase computational speed. The right panel plots the mapping from rank correlation to log factor loadings, and we see that the true mapping is reasonably approximated by a straight line, particularly for values of rank correlation near the sample average rank correlation in our application, which is around 0.4, supporting assumption (c).

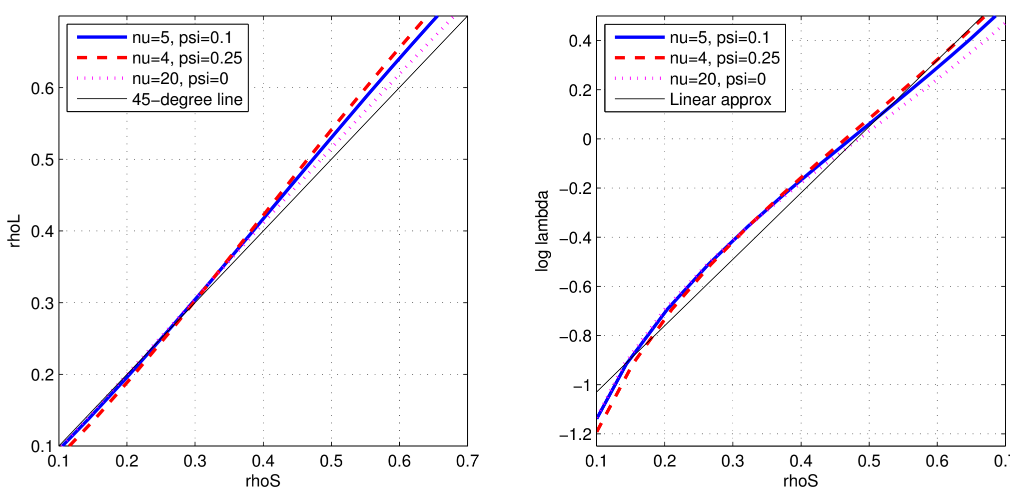


FIGURE D.1: *The left panel plots the mapping from rank correlation to linear correlation for various choices of shape parameters in the factor copula. The right panel compares the true mappings from rank correlation to log-lambda with a linear approximation.*

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44, 182–198.

Acar, E., Genest, C., and Nelehov, J. (2012), "Beyond simplified pair-copula constructions," *Journal of Multivariate Analysis*, 110, 74–90.

Adrian, T. and Brunnermeier, M. (2009), "CoVaR," Staff Report 348, Federal Reserve Bank of New York.

Almeida, C., Czado, C., and Manner, H. (2012), "Modeling high dimensional time-varying dependence using D-vine SCAR models," working paper.

Amisano, G. and Giacomini, R. (2007), "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business and Economic Statistics*, 25, 177–190.

Andersen, L. and Sidenius, J. (2004), "Extensions to the Gaussian copula: random recovery and random factor loadings," *Journal of Credit Risk*, 1, 29–70.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003), "Modeling and forecasting realised volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006), "Volatility and correlation forecasting," in *Handbook of Economic Forecasting*, eds. G. Elliott, C. Granger, and A. Timmermann, vol. 1, chap. 15, pp. 777–878, Elsevier, Oxford.

Andersen, T. G., Bollerslev, T., Frederiksen, P., and Nielsen, M. O. (2010), "Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns," *Journal of Applied Econometrics*, 25, 233–261.

Andrews, D. (1994), "Empirical process methods in econometrics," in *Handbook of Econometrics*, eds. R. Engle and D. McFadden, vol. 4, chap. 37, pp. 2247–2294, Elsevier, Oxford.

Andrews, D. (2001), "Testing when a parameter is on the boundary of the maintained hypothesis," *Econometrica*, 69, 683–734.

Barndorff-Nielsen, O. and Shephard, N. (2004), "Econometric analysis of realised covariation: high frequency based covariance, regression and correlaion in financial economics," *Econometrica*, 72, 885–925.

Barndorff-Nielsen, O. and Shephard, N. (2013), *Financial Volatility in Continuous Time*, Cambridge University Press.

Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. (2009), "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12, 1–32.

Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. (2011), "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *Journal of Econometrics*, 162, 149–169.

Barndorff-Nielsen, O. E. (1978), "Hyperbolic distributions and distributions on hyperbolae," *Scandinavian Journal of Statistics*, 5, 151–157.

Barndorff-Nielsen, O. E. (1997), "Normal inverse Gaussian distributions and stochastic volatility modelling," *Scandinavian Journal of Statistics*, 24, 1–13.

Bauer, G. and Vorkink, K. (2011), "Forecasting multivariate realized stock market volatility," *Journal of Econometrics*, 160, 93–101.

Blasques, F., Koopman, S., and Lucas, A. (2012), "Stationarity and ergodicity of univariate generalized autoregressive score processes," Tinbergen Institute Discussion Paper TI 2012-059/4.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. (1990), "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988), "A capital asset pricing model with time varying covariances," *Journal of Political Economy*, 96, 116–131.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), "ARCH models," in *Handbook of Econometrics*, eds. R. F. Engle and D. McFadden, vol. 4, chap. 49, pp. 2959–3038, Elsevier, Oxford.

Bonhomme, S. and Robin, J.-M. (2009), "Assessing the equalizing force of mobility using short panels: france, 1990-2000," *Review of Economic Studies*, 76, 63–92.

Bouzebda, S. and Zari, T. (2011), "Strong approximation of empirical copula processes by gaussian processes," working paper.

Brendstrup, B. and Paarsch, H. (2007), "Semiparametric identification and estimation in multi-object english auctions," *Journal of Econometrics*, 141, 84–108.

Brownlees, C. and Engle, R. (2011), "Volatility, correlation and tails for systemic risk measurement," working paper, Stern School of Business, New York University.

Capital, B. (2010), *Standard Corporate CDS Handbook*, Barclays Capital.

Carr, P. and Wu, L. (2011), "A simple robust link between american puts and credit protection," *Review of Financial Studies*, 24, 473–505.

Chan, N.-H., Chen, J., Chen, X., Fan, Y., and Peng, L. (2009), "Statistical inference for multivariate residual copula of garch models," *Statistica Sinica*, 19, 53–70.

Chen, X. and Fan, Y. (2006), "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification," *Journal of Econometrics*, 135, 125–154.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006), "Efficient estimation of semiparametric multivariate copula models," *Journal of the American Statistical Association*, 101, 1228–1240.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004), *Copula Methods in Finance*, John Wiley & Sons.

Chiriac, R. and Voev, V. (2011), "Modelling and forecasting multivariate realized volatility," *Journal of Applied Econometrics*, 26, 922–947.

Christoffersen, P., Errunza, V., Jacobs, K., and Langlois, H. (2012), "Is the potential for international diversification disappearing?" *Review of Financial Studies*, 25, 3711–3751.

Christoffersen, P., Jacobs, K., Jin, X., and Langlois, H. (2013), "Dynamic dependence in corporate credit," working paper, Bauer College of Business, University of Houston.

Clayton, D. (1978), "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika*, 65, 141–151.

Conrad, J., Dittmar, R., and Hameed, A. (2011), "Cross-market and cross-firm effects in implied default probabilities and recovery values," working paper, Kenan-Flagler Business School, University of North Carolina.

Cook, R. and Johnson, M. (1981), "A family of distributions for modelling non-elliptically symmetric multivariate data," *Journal of the Royal Statistical Society*, 43, 210–218.

Corsi, F. (2009), "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Coval, J., Jurek, J., and Stafford, E. (2009), "The economics of structured finance," *Journal of Economic Perspectives*, 23, 3–25.

Cox, D. R. and Reid, N. (2004), "A note on pseudolikelihood constructed from marginal densities," *Biometrika*, 91, 729–737.

Creal, D., Koopman, S., and Lucas, A. (2011), "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations," *Journal of Business and Economic Statistics*, 29, 552–563.

Creal, D., Gramercy, R., and Tsay, R. (2012), "Market-based credit ratings," working paper, Booth School of Business, University of Chicago.

Creal, D., Koopman, S., and Lucas, A. (2013), "Generalized autoregressive score models with applications," *Journal of Applied Econometrics*, 28, 777–795.

Danelsson, J., Jorgensen, B., Samorodnitsky, G., Sarma, M., and de Vries, C. (2013), "Fat tails, VaR and subadditivity," *Journal of Econometrics*, 172, 283–291.

Daul, S., Giorgi, E. D., Lindskog, F., and McNeil, A. (2003), "The grouped t-copula with an application to credit risk," *RISK*, 16, 73–76.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. (2003), "Observation driven models for Poisson counts," *Biometrika*, 90, 777–790.

De Lira Salvatierra, I. and Patton, A. J. (2013), "Dynamic copula models and high frequency data," working paper, Duke University.

Demarta, S. and McNeil, A. J. (2005), "The t copula and related copulas," *International Statistical Review*, 73, 111–129.

Diks, C., Panchenko, V., and van Dijk, D. (2010), "Out-of-sample comparison of copula specifications in multivariate density forecasts," *Journal of Economic Dynamics & Control*, 34, 1596–1609.

Diks, C., Panchenko, V., and van Dijk, D. (2011), "Likelihood-based scoring rules for comparing density forecasts in tails," *Journal of Econometrics*, 163, 215–230.

Diks, C., Panchenko, V., Sokolinskiy, O., and van Dijk, D. (2013), "Comparing the accuracy of copula-based multivariate density forecasts in selected regions of support," working paper.

Ding, Z. and Engle, R. (2001), "Large scale conditional covariance matrix modeling, estimation and testing," *Academia Economic Papers*, 29, 157–184.

Duffie, D. and Singleton, K. (2003), *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press, New Jersey.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009), "Frailty correlated default," *Journal of Finance*, 64, 2089–2123.

Embrechts, P., Klppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Springer-Verlag.

Embrechts, P., McNeil, A., and Straumann, D. (2002), "Correlation and dependence properties in risk management: properties and pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. M. Dempster, Cambridge University Press.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of uk inflation," *Econometrica*, 50, 987–1007.

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.

Engle, R. F. and Kelly, B. (2012), "Dynamic equicorrelation," *Journal of Business and Economic Statistics*, 30, 212–228.

Engle, R. F. and Mezrich, J. (1996), "GARCH for groups," *Risk*, 9, 36–40.

Engle, R. F. and Russell, J. R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Engle, R. F. and Sheppard, K. (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate garch," working paper, University of California, San Diego.

Engle, R. F., Shephard, N., and Sheppard, K. (2008), "Fitting and testing vast dimensional time-varying covariance models," working paper, Oxford-Man Institute of Quantitative Finance.

Fan, J., Li, Y., and Ke, Y. (2012), "Vast volatility matrix estimation using high frequency data for portfolio selection," *Journal of the American Statistical Association*, 107, 412–428.

Fermanian, J., Radulovi, D., and Wegkamp, M. (2004), "Weak convergence of empirical copula process," *Bernoulli*, 10, 847–860.

Fine, J. and Jiang, H. (2000), "On association in a copula with time transformations," *Biometrika*, 87, 559–571.

Gao, X. and Song, P. X.-K. (2010), "Composite likelihood Bayesian information criteria for model selection in high-dimensional data," *Journal of the American Statistical Association*, 105, 1531–1540.

Geluk, J., de Haan, L., and de Vries, C. (2007), "Weak and strong financial fragility," Tinbergen Institute Discussion Paper TI 2007-023/2.

Genest, C. (1987), "Frank's family of bivariate distributions," *Biometrika*, 74, 549–555.

Genest, C. and Favre, A.-C. (2007), "Everything you always wanted to know about copula modeling but were afraid to ask," *Journal of Hydrologic Engineering*, 12, 347–368.

Genest, C. and Rivest, L.-P. (1993), "Statistical inference procedures for bivariate archimedean copulas," *Journal of the American Statistical Association*, 88, 1034–1043.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, 82, 543–552.

Giacomini, R. and White, H. (2006), "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

Giesecke, K. and Kim, B. (2011), "Systemic risk: What defaults are telling us," *Management Science*, 57, 1387–1405.

Glosten, R. T., Jagannathan, R., and Runkle, D. (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

Gneiting, T. and Raftery, A. (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378.

Gonalves, S. and White, H. (2002), "The bootstrap of the mean for dependent heterogeneous arrays," *Econometric Theory*, 18, 1367–1384.

Gonalves, S., Hounyo, U., Patton, A., and Sheppard, K. (2013), "Bootstrapping two-stage extremum estimators," working paper, Oxford-Man Institute of Quantitative Finance.

Gouriroux, C. and Monfort, A. (1996), *Statistics and Econometric Models, Volume 2, translated from the French by Q. Vuong*, Cambridge University Press.

Group, M. (2009), "The CDS big bang: understanding the changes to the global CDS contract and North American conventions," research report.

Hafner, C. M. and Manner, H. (2012), "Dynamic stochastic copula models: estimation, inference and applications," *Journal of Applied Econometrics*, 27, 269–295.

Hall, A. (2000), "Covariance matrix estimation and the power of the overidentifying restrictions test," *Econometrica*, 68, 1517–1528.

Hall, A. (2005), *Generalized Method of Moments*, Oxford University Press, Oxford.

Hall, A. and Inoue, A. (2003), "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114, 361–394.

Hansen, B. E. (1994), "Autoregressive conditional density estimation," *International Economic Review*, 35, 705–730.

Hansen, P., Huang, Z., and Shek, H. (2012), "Realized GARCH: a joint model for returns and realized measures of volatility," *Journal of Applied Econometrics*, 27, 877–906.

Harvey, A. (2013), *Dynamic Models for Volatility and Heavy Tails, Econometric Society Monograph 52*, Cambridge University Press, Cambridge.

Harvey, A. and Sucarrat, G. (2012), "EGARCH models with fat tails, skewness and leverage," working paper CWPE 1236, Cambridge University.

Hautsch, N., Kyj, L. M., and Oomen, R. C. A. (2012), "A blocking and regularization approach to high-dimensional realized covariance estimation," *Journal of Applied Econometrics*, 27, 625–645.

Hautsch, N., Kyj, L., and Malec, P. (2013), "Do high-frequency data improve high-dimensional portfolio allocations?" working paper.

Heinen, A. and Valdesogo, A. (2009), "Asymmetric CAPM dependence for large dimensions: the canonical vine autoregressive model," working paper, Univerisidad Carlos III.

Hong, H., Mahajan, A., and Nekipelov, D. (2010), "Extremum estimation and numerical derivatives," working paper, Stanford University.

Huang, X., Zhou, H., and Zhu, H. (2009), "A framework for assessing the systemic risk of major financial institutions," *Journal of Banking and Finance*, 33, 2036–2049.

Hull, J. (2012), *Risk Management and Financial Institutions, Third Edition*, John Wiley & Sons, New Jersey.

Hull, J. and White, A. (2004), "Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation," *Journal of Derivatives*, 12, 8–23.

Hyung, N. and de Vries, C. (2007), "Portfolio selection with heavy tails," *Journal of Empirical Finance*, 14, 383–400.

Jin, X. and Maheu, J. M. (2013), "Modeling realized covariances and returns," *Journal of Financial Econometrics*, 11, 335–369.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chanpman and Hall.

Joe, H. (2005), "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94, 401–419.

Joe, H. and Xu, J. (1996), "The estimation method of inference functions for margins for multivariate models," working paper, Department of Statistics, University of British Columbia.

Jondeau, E. and Rockinger, M. (2006), "The copula-GARCH model of conditional dependencies: An international stock market application," *Journal of International Money and Finance*, 25, 827–853.

Jondeau, E. and Rockinger, M. (2012), "On the importance of time variability in higher moments for asset allocation," *Journal of Financial Econometrics*, 10, 84–123.

Judd, K. (1998), *Numerical Methods in Economics*, MIT Press.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007), "Comparison of semiparametric and parametric methods for estimating copulas," *Computational Statistics & Data Analysis*, 51, 2836–2850.

Lee, T. and Long, X. (2009), "Copula-based multivariate GARCH model with uncorrelated dependent errors," *Journal of Econometrics*, 150, 207–218.

Li, D. X. (2000), "On default correlation: A copula function approach," *Journal of Fixed Income*, 9, 43–54.

Li, S. Z. (2013), "Continuous beta, discontinuous beta and the cross-section of expected stock returns," working paper, Duke University.

Lucas, A., Schwaab, B., and Zhang, X. (2011), "Conditional probabilities for euro area sovereign default risk," Tinbergen Institute discussion paper, 11-176/2/DSF29.

Maheu, J. M. and McCurdy, T. H. (2011), "Do high-frequency measures of volatility improve forecasts of return distributions?" *Journal of Econometrics*, 160, 69–76.

Manner, H. and Segers, J. (2011), "Tails of correlation mixtures of elliptical copulas," *Insurance: Mathematics and Economics*, 48, 153–160.

Marshall, A. W. and Olkin, I. (1988), "Families of multivariate distributions," *Journal of the American Statistical Association*, 83, 834–841.

McFadden, D. (1989), "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 47, 995–1026.

McNeil, A. J., Frey, R., and Embrechts, P. (2005), *Quantitative Risk Management*, Princeton University Press.

Min, A. and Czado, C. (2010), "Bayesian inference for multivariate copulas using pair-copula constructions," *Journal of Financial Econometrics*, 8, 511–546.

Nelsen, R. (2006), *An Introduction to Copulas*, Springer.

Nelson, D. (1991), "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–370.

Newey, W. (1985), "Generalized method of moments specification testing," *Journal of Econometrics*, 29, 229–256.

Newey, W. and McFadden, D. (1994), "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, eds. R. Engle and D. McFadden, vol. 4, chap. 36, pp. 2111–2245, Elsevier, Oxford.

Newey, W. and West, K. D. (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Noureldin, D., Shephard, N., and Sheppard, K. (2012), "Multivariate high-frequency-based volatility (HEAVY) models," *Journal of Applied Econometrics*, 27, 907–933.

Oh, D.-H. and Patton, A. J. (2011), "Modelling dependence in high dimensions with factor copulas," working paper, Duke University.

Oh, D.-H. and Patton, A. J. (2013a), "Simulated method of moments estimation for copula-based multivariate models," *Journal of the American Statistical Association*, 108, 689–700.

Oh, D.-H. and Patton, A. J. (2013b), "Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads," working paper, Duke University.

Pakes, A. and Pollard, D. (1989), "Simulation and the asymptotics of optimization estimators," *Econometrica*, 47, 1027–1057.

Palm, F. and Urbain, J.-P. (2011), "Factor structures for panel and multivariate time series data," *Journal of Econometrics*, 163, 1–3.

Patton, A. J. (2004), "On the out-of-sample importance of skewness and asymmetric dependence for asset allocation," *Journal of Financial Econometrics*, 2, 130–168.

Patton, A. J. (2006a), "Estimation of multivariate models for time series of possibly different lengths," *Journal of Applied Econometrics*, 21, 147–173.

Patton, A. J. (2006b), "Modelling asymmetric exchange rate dependence," *International Economic Review*, 47, 527–556.

Patton, A. J. (2012), "A review of copula models for economic time series," *Journal of Multivariate Analysis*, 110, 4–18.

Patton, A. J. (2013), "Copula methods for forecasting multivariate time series," in *Handbook of Economic Forecasting*, eds. G. Elliott and A. Timmermann, vol. 2, chap. 16, pp. 899–960, Elsevier, Oxford.

Politis, D. and Romano, J. (1994), "The Stationary bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Rivers, D. and Vuong, Q. H. (2002), "Model selection tests for nonlinear dynamic models," *Econometrics Journal*, 5, 1–39.

Rosenberg, J. and Schuermann, T. (2006), "A general approach to integrated risk management with skewed, fat-tailed risks," *Journal of Financial Economics*, 79, 569–614.

Rothenberg, T. (1971), "Identification of parametric models," *Econometrica*, 39, 577–591.

Rmillard, B. (2010), "Goodness-of-fit tests for copulas of multivariate time series," working paper, McGill University.

Schwaab, B. (2010), "New quantitative measures of systemic risk," Financial Stability Review Special Feature E, European Central Bank.

Segioviano, M. and Goodhart, C. (2009), "Banking stability measures," IMF Working Paper WP/09/4.

Shephard, N. (2005), *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.

Sklar, A. (1959), "Fonctions de repartition a n dimensions et leurs marges," *Publications de Institut Statistique de Universite de Paris 8*, pp. 229–231.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010), "Modeling longitudinal data using a pair-copula decomposition of serial dependence," *Journal of the American Statistical Association*, 105, 1467–1479.

Smith, M., Gan, Q., and Kohn, R. (2012), "Modeling dependence using skew t copulas: Bayesian inference and applications," *Journal of Applied Econometrics*, 27, 500–522.

Song, P., Fan, Y., and Kalbfleisch, J. (2005), "Maximization by parts in likelihood inference," *Journal of the American Statistical Association*, 100, 1145–1158.

Stber, J. and Czado, C. (2012), "Detecting regime switches in the dependence structure of high dimensional financial data," working paper, Technische Universitt Mnchen.

van der Voort, M. (2005), "Factor copulas: totally external defaults," working paper, Erasmus University Rotterdam.

Varin, C. (2008), "On composite marginal likelihoods," *Advances in Statistical Analysis*, 92, 1–28.

Varin, C. and Vidoni, P. (2005), "A note on composite likelihood inference and model selection," *Biometrika*, 92, 519–528.

Varin, C., Reid, N., and Firth, D. (2011), "An overview of composite likelihood methods," *Statistica Sinica*, 21, 5–42.

Vuong, Q. H. (1989), "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, 57, 307–333.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press.

White, H. (2000), "A reality check for data snooping," *Econometrica*, 68, 1097–1126.

Wolak, F. (1989), "Testing inequality contraints in linear econometic models," *Journal of Econometrics*, 41, 205–235.

Zimmer, D. (2012), "The role of copulas in the housing crisis," *Review of Economics and Statistics*, 94, 607–620.

Biography

Dong Hwan Oh was born in Republic of Korea on April 16, 1980. He earned his B.A. in economics with Summa Cum Laude in February 2003, and M.A. in economics from Seoul National University, in September 2008. He began his graduate studies at Duke University in August 2008. He plans on graduating from Duke University with a doctorate in economics in the spring of 2014.