High Dimension Copula-Based Distributions with Mixed Frequency Data

Dong Hwan Oh

Federal Reserve Board

Andrew J. Patton

Duke University

February 2015
Motivation

A model for the distribution of returns on a collection of financial assets is crucial for risk management and asset allocation.

And these collections tend to be large: eg, median number of stocks held by US mutual funds is 94 (25/75 percentiles are 46 and 208).

But there are relatively few dynamic, high-dimension models available.

Many are based on multivariate Normality, despite its limitations.

Almost all use data from a common sampling frequency.

We propose a new approach for constructing and estimating high dimension distribution models, drawing on two areas of recent research:

1. High frequency data is very useful for estimating lower-frequency second moments (eg, correlation).

2. Copula-based distributions are useful for constructing flexible models in high dimensions.
Exploiting high frequency data in lower-frequency copula-based models is not straightforward:

- Unlike covariances, the copula of daily returns is not generally a known function of the copula of high frequency returns.
- So most of the nice theory from high frequency financial econometrics cannot be used directly.

We propose decomposing the dependence structure of daily returns into **linear** and **nonlinear** components:

- **High frequency data** is used to accurately model the linear dependence.
- **Low frequency data** and a new class of copulas is used to capture the remaining nonlinear dependence.
Decomposition of dependence

- **Linear** dependence: Captured by correlation
- **Nonlinear** dependence: Any dependence beyond linear correlation
Standard use of copulas in the literature

- Chasing two rabbits with only one tool
- A heavy burden for the copula model
Our approach: in pictures

- Chasing two rabbits with two tools: high frequency data and copulas
- High frequency data shares the heavy burden with the copula model
Our approach: in equations

- We construct a model for a $N$-vector of daily returns $\mathbf{r}_t$ as follows. Let:

\[
\mathbf{r}_t = \mu_t + \mathbf{H}_t^{1/2} \mathbf{e}_t
\]

where $\mathbb{E}_{t-1} [\mathbf{e}_t] = 0$, $\mathbb{E}_{t-1} [\mathbf{e}_t \mathbf{e}_t'] = \mathbf{I}$

Use standard methods to estimate $\mu_t$ and $\mathbf{H}_t$. Use high frequency data to obtain improved estimates of $\mathbf{H}_t$.

We propose a HAR-type model for $\mathbf{H}_t$ (more details below).

Decompose the distribution of the uncorrelated residuals $\mathbf{e}_t$ iid.

$\mathbb{F}(\mathbf{e}_1; \cdots; \mathbf{F}_N(\mathbf{e}_N; \cdots))$

Can easily choose $\mathbb{F}_i$ to ensure that $\mathbb{E}[\mathbf{e}_t] = 0$ and $\mathbb{E}[\mathbf{e}_t^2] = 1$.

But also need to ensure that $\mathbb{F}$ is such that $\mathbb{E}[\mathbf{e}_t \mathbf{e}_j] = 0$ for $i \neq j$. 

Patton (Duke)
Our approach: in equations

- We construct a model for a $N$-vector of daily returns $r_t$ as follows. Let:
  \[
  r_t = \mu_t + H_t^{1/2} e_t
  \]
  where $E_{t-1}[e_t] = 0$, $E_{t-1}[e_t e'_t] = I$

- Use standard methods to estimate $\mu_t$

- Use high frequency data to obtain improved estimates of $H_t$
  - We propose a HAR-type model for $H_t$ (more details below)
We construct a model for a $N$-vector of daily returns $r_t$ as follows. Let:

$$r_t = \mu_t + H_t^{1/2} e_t$$

where $\mathbb{E}_{t-1}[e_t] = 0$, $\mathbb{E}_{t-1}[e_t e_t'] = I$

Use standard methods to estimate $\mu_t$

Use high frequency data to obtain improved estimates of $H_t$

We propose a HAR-type model for $H_t$ (more details below)

Decompose the distribution of the uncorrelated residuals as

$$e_t \sim iid \ F(\cdot; \eta) = C(F_1(\cdot; \eta), \ldots, F_N(\cdot; \eta); \eta)$$
Our approach: in equations

- We construct a model for a \( N \)-vector of daily returns \( r_t \) as follows. Let:

\[
    r_t = \mu_t + H_t^{1/2} e_t
\]

where \( \mathbb{E}_{t-1}[e_t] = 0 \), \( \mathbb{E}_{t-1}[e_t e'_t] = 1 \)

- Use standard methods to estimate \( \mu_t \)

- Use **high frequency data** to obtain improved estimates of \( H_t \)
  - We propose a HAR-type model for \( H_t \) (more details below)

- Decompose the distribution of the **uncorrelated residuals** as

\[
    e_t \sim iid \ F (\; ; \eta) = C (F_1 (\; ; \eta), \ldots, F_N (\; ; \eta) ; \eta)
\]

- Can easily choose \( F_i \) to ensure that \( \mathbb{E}[e_{it}] = 0 \) and \( \mathbb{E}[e_{it}^2] = 1 \)

- But also need to ensure that \( F \) is such that \( \mathbb{E}[e_{it} e_{jt}] = 0 \) \( \forall i \neq j \)
Contributions of this paper

This paper makes four main contributions. We:

1. Propose a new class of “jointly symmetric” copulas, useful in MV density models that contain a covariance matrix model (e.g., DCC, HAR, SV, etc.)

2. Show that composite likelihood methods can be used to estimate these new models, and verify good finite sample properties via simulations.


4. Apply these new models to a detailed study of 104 US equity returns, and show that they outperform existing approaches both in- and out-of-sample.
1. Introduction

2. Models of linear and nonlinear dependence
   - Jointly symmetric copulas
   - A new covariance matrix model

3. Estimation and comparison via composite likelihood

4. Simulation study

5. Analysis of S&P 100 equity returns
1. Introduction

2. Models of linear and nonlinear dependence
   - Jointly symmetric copulas
   - A new covariance matrix model

3. Estimation and comparison via composite likelihood

4. Simulation study

5. Analysis of S&P 100 equity returns
A model for uncorrelated residuals

- A key building block for our model is an $N$-dim distribution $F$ that guarantees an identity correlation matrix.

- There are very few existing copulas that do this:
  - Normal copula with identity correlation matrix (i.e., independence copula)
  - $t$ copula with identity correlation matrix, when combined with symmetric marginals

- The idea in this paper is to exploit the fact that multivariate distributions that satisfy a certain symmetry condition automatically ensure zero correlation.
Joint symmetry and lack of correlation

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

\[
\tilde{\mathbf{X}}^{(i)} = \begin{bmatrix} \tilde{X}_1^{(i)} & \cdots & \tilde{X}_1^{(N)} \end{bmatrix}, \quad i = 1, 2, \ldots, 2^N
\]

where \( \tilde{X}_j^{(N)} = (X_j - a_j) \) or \( (a_j - X_j) \) for \( j = 1, 2, \ldots, N \).
Joint symmetry and lack of correlation

Definition: Let $\mathbf{X}$ be a vector of $N$ variables and let $\mathbf{a} \in \mathbb{R}^N$. Then $\mathbf{X}$ is **jointly symmetric** about $\mathbf{a}$ if the following $2^N$ vectors of $N$ random variables have the same joint distribution

$$
\tilde{\mathbf{X}}^{(i)} = \begin{bmatrix}
\tilde{X}_1^{(i)}, \ldots, \tilde{X}_1^{(N)}
\end{bmatrix}, \quad i = 1, 2, \ldots, 2^N
$$

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

Lemma 1: If $\mathbf{X}$ is jointly symmetric and has finite second moments, then it has an **identity correlation matrix**.
Joint symmetry and lack of correlation

**Definition:** Let $\mathbf{X}$ be a vector of $N$ variables and let $\mathbf{a} \in \mathbb{R}^N$. Then $\mathbf{X}$ is **jointly symmetric** about $\mathbf{a}$ if the following $2^N$ vectors of $N$ random variables have the same joint distribution

\[
\mathbf{X}^{(i)} = \begin{bmatrix} \tilde{X}_1^{(i)} & \ldots & \tilde{X}_N^{(i)} \end{bmatrix}, \quad i = 1, 2, \ldots, 2^N
\]

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

**Lemma 1:** If $\mathbf{X}$ is jointly symmetric and has finite second moments, then it has an **identity correlation matrix**.

**Lemma 2:** Let $\mathbf{X} \sim \mathbf{F} = \mathbf{C} (F_1, \ldots, F_N)$, where $X_i$ is symmetric about $a_i \ \forall \ i$. Then $\mathbf{X}$ is jointly symmetric iff $\mathbf{C}$ is jointly symmetric.
Joint symmetry and lack of correlation

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

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

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

**Lemma 1:** If \( \mathbf{X} \) is jointly symmetric and has finite second moments, then it has an **identity correlation matrix**.

**Lemma 2:** Let \( \mathbf{X} \sim \mathbf{F} = \mathbf{C} (F_1, \ldots, F_N) \), where \( X_i \) is symmetric about \( a_i \) \( \forall \ i \). Then \( \mathbf{X} \) is jointly symmetric iff \( \mathbf{C} \) is jointly symmetric.

**Result:** Any combination of symmetric marginals and jointly symmetric copula yields a jointly symmetric joint distribution, implying an identity correlation matrix.
There are numerous interesting/useful copula models in the literature, almost none of which are jointly symmetric.

We overcome this lack of choice by proposing a novel way to obtain a jointly symmetric copula: **rotate existing copulas**
Example: rotations of the Clayton copula

Bivariate distributions with rotated Clayton copulas and N(0,1) margins
Example: a jointly symmetric Clayton copula

Bivariate distributions with jointly symmetric Clayton copula and N(0,1) margins
Example: other jointly symmetric distributions

Bivariate distributions with jointly symmetric copulas and N(0,1) margins

- Independence copula
- Jointly sym. Plackett, $\theta = 11$
- Jointly sym. Clayton, $\theta = 2$
- Jointly sym. $t$ copula, $\nu = 3$
- Jointly sym. Frank, $\theta = 6.6$
- Jointly sym. Gumbel, $\theta = 1.9$
N-dimensional jointly symmetric copulas

**Theorem:** Given any \( N \)-dimensional copula \( C \) with density \( c \), then

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

\[
C^{JS}(u_1, \ldots, u_N) = \frac{1}{2^N} \left[ \sum_{k_1=0}^{2} \cdots \sum_{k_N=0}^{2} (-1)^R \cdot C(\tilde{u}_1, \ldots, \tilde{u}_N) \right]
\]

where \( \tilde{u}_i = \begin{cases} 1, & k_i = 0 \\ u_i, & k_i = 1 \\ 1 - u_i, & k_i = 2 \end{cases}, \) and \( R = \sum_{i=1}^{N} 1 \{k_i = 2\} \)

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

\[
c^{JS}(u_1, \ldots, u_N) = \frac{1}{2^N} \left[ \sum_{k_1=1}^{2} \cdots \sum_{k_N=1}^{2} c(\tilde{u}_1, \ldots, \tilde{u}_N) \right]
\]
1. Introduction

2. Models of linear and nonlinear dependence
   - Jointly symmetric copulas
   - A new covariance matrix model

3. Estimation and comparison via composite likelihood

4. Simulation study

5. Analysis of S&P 100 equity returns
Let $\Delta$ be the sampling frequency (e.g., five minutes), yielding $1/\Delta$ observations per trade day, and define the realized covariance matrix as

$$RCov_t^\Delta = \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} r'_{t-1+j\Delta} = \sqrt{RV_t^\Delta} RCorr_t^\Delta \sqrt{RV_t^\Delta}$$

where $RV_t^\Delta = \text{diag} \{ [RV_{1t}^\Delta, \ldots, RV_{Nt}^\Delta] \}$
Let $\Delta$ be the sampling frequency (eg, five minutes), yielding $1/\Delta$ observations per trade day, and define the realized covariance matrix as

$$RCov_t^\Delta = \sum_{j=1}^{1/\Delta} r_{t-1+j\Delta} r'_{t-1+j\Delta} = \sqrt{RV_t^\Delta} RCorr_t^\Delta \sqrt{RV_t^\Delta}$$

where $RV_t^\Delta = \text{diag} \{RV_{1t}^\Delta, \ldots, RV_{Nt}^\Delta\}$

We suggest using a HAR model (Corsi, 2009) for the log realized variances:

$$\log RV_{i,t}^\Delta = \phi_i^{(const)} + \phi_i^{(day)} \log RV_{i,t-1}^\Delta + \phi_i^{(week)} \frac{1}{4} \sum_{j=2}^{5} \log RV_{i,t-j}^\Delta + \phi_i^{(month)} \frac{1}{15} \sum_{j=6}^{20} \log RV_{i,t-j}^\Delta + \xi_{it}$$

estimated via OLS for each variance.
We next propose a HAR-type model for the realized correlation matrix, imposing parameter constraints similar to the DCC model of Engle (2002):

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

where \((a, b, c) \in \mathbb{R}^3\).

This parsimonious model can easily be estimated via OLS, and guarantees positive definiteness if \((a, b, c) > 0\) and \(a + b + c < 1\).
Outline

1. Introduction

2. Models of linear and nonlinear dependence
   - Jointly symmetric copulas
   - A new covariance matrix model

3. Estimation and comparison via composite likelihood

4. Simulation study

5. Analysis of S&P 100 equity returns
Our model for the vector of asset returns is

\[ r_t = \mu_t + H_t^{1/2} e_t \]

where \( e_t \sim iid F(\cdot; \eta) = C (F_1(\cdot; \eta), \ldots, F_N(\cdot; \eta); \eta) \)

and \( F \) is constrained so that \( \mathbb{E}[e_t] = 0 \) and \( \mathbb{E}[e_t e'_t] = \mathbf{I} \).

We will first discuss estimation of \( C \), and then consider estimation of the rest of the model (in stages).

Inference methods will take into account the multi-stage estimation method
Our method for constructing a JS copula requires $2^N$ evaluations of a given original copula density. Even for moderate dimensions this can be very slow.

Eg: computation time for **one evaluation** of density of JS Clayton:

<table>
<thead>
<tr>
<th>$N$</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.23 sec</td>
<td>4 min</td>
<td>70 hours</td>
<td>$10^6$ years</td>
<td>$10^{17}$ years</td>
</tr>
</tbody>
</table>
Our method for constructing a JS copula requires $2^N$ evaluations of a given original copula density. Even for moderate dimensions this can be very slow.

Eg: computation time for one evaluation of density of JS Clayton:

<table>
<thead>
<tr>
<th>$N$</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.23 sec</td>
<td>4 min</td>
<td>70 hours</td>
<td>10^6 years</td>
<td>10^{17} years</td>
</tr>
</tbody>
</table>

We propose overcoming this difficulty by using composite likelihood methods (Lindsay 1988):

- Estimate parameters of the full model by maximizing the likelihoods of submodels
- Consistent if submodels are sufficient to identify parameter of full model
- Less efficient, though loss need not be great
Composite likelihood estimation of the copula II

- Composite likelihood is particularly attractive for jointly symmetric copulas:

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

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

where $c_{ij}$ is the $(i,j)$ marginal copula density of the original $N$-dimensional copula.

- Thus while the full copula model requires $2^N$ rotations of the original density, bivariate marginal copulas only require $2^2$ rotations.
Similar to Engle, et al. (2008), we consider CL based either on all pairs, adjacent pairs, or just one pair of variables:

\[
CL_{all} (u_1, \ldots, u_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \log c_{i,j} (u_i, u_j)
\]

\[
CL_{adj} (u_1, \ldots, u_N) = \sum_{i=1}^{N-1} \log c_{i,i+1} (u_i, u_{i+1})
\]

\[
CL_{first} (u_1, \ldots, u_N) = \log c_{1,2} (u_1, u_2)
\]
Similar to Engle, et al. (2008), we consider CL based either on all pairs, adjacent pairs, or just one pair of variables:

\[
CL_{all} (u_1, \ldots, u_N) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \log c_{i,j} (u_i, u_j)
\]

\[
CL_{adj} (u_1, \ldots, u_N) = \sum_{i=1}^{N-1} \log c_{i,i+1} (u_i, u_{i+1})
\]

\[
CL_{first} (u_1, \ldots, u_N) = \log c_{1,2} (u_1, u_2)
\]

Comparison of computation times for single evaluation of log-likelihood:

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full likelihood</td>
<td>0.23 sec</td>
<td>4 min</td>
<td>70 hours</td>
<td>10^6 years</td>
<td>10^{17} years</td>
</tr>
<tr>
<td>All pairs CL</td>
<td>0.05 sec</td>
<td>0.21 sec</td>
<td>0.45 sec</td>
<td>1.52 sec</td>
<td>5.52 sec</td>
</tr>
<tr>
<td>Adjacent pairs CL</td>
<td>0.01 sec</td>
<td>0.02 sec</td>
<td>0.03 sec</td>
<td>0.06 sec</td>
<td>0.11 sec</td>
</tr>
<tr>
<td>First pair CL</td>
<td>0.001 sec</td>
<td>0.001 sec</td>
<td>0.001 sec</td>
<td>0.001 sec</td>
<td>0.001 sec</td>
</tr>
</tbody>
</table>
The maximum composite likelihood estimator (MCLE) is then obtained as:

$$\hat{\theta}_{MCLE} = \arg \max_{\theta} \sum_{t=1}^{T} CL\left(u_{1t}, \ldots, u_{Nt}; \theta\right)$$

Under standard regularity conditions, Cox and Reid (2004) show that

$$\sqrt{T} \left(\hat{\theta}_{MCLE} - \theta_0\right) \xrightarrow{d} N\left(0, \mathcal{H}_0^{-1} \mathcal{J}_0 \mathcal{H}_0^{-1}\right)$$

A key condition for CL to work is that the submodels used are rich enough to identify the parameters.

This needs to be verified on a case by case basis.

Is easily satisfied for the jointly symmetric copulas we consider: all have just a single unknown parameter, which appears in all bivariate submodels.
We first define the composite Kullback-Leibler information criterion (cKLIC) following Varin and Vidoni (2005).

**Definition (Varin and Vidoni, 2005):** Given an $N$-dimension random variable $Z$ with true density $g$, the composite Kullback-Leibler information criterion (cKLIC) of a density $h$ relative to $g$ is

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

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 adjacent-pair composite likelihoods using the true density $g$ and a competing density $h$.

Above uses CL with adjacent pairs, but other cKLICs can be defined.
Note that the expectation is with respect to the (complete) true density $g$ rather than the CL of the true density, so it possible to interpret cKLIC as a linear combination of the KLICs of submodels:

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

This implies that existing in-sample model selection tests, such as those of Vuong (1989) and Rivers and Vuong (2002) can be applied to model selection using cKLIC.
We may also wish to select the best model in terms of out-of-sample (OOS) forecasting performance measured by some scoring rule, $S$, for the model.

Gneiting and Raftery (2007) define “proper” scoring rules as those which ensure that the true density always receives a higher score than other densities.

- The log density, i.e. $S(h(Z_{t+1})) = \log h(Z_{t+1})$ is proper.

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

  $$S(h(Z_{t+1})) = \sum_{i=1}^{N-1} \log h_i(Z_{i,t+1}, Z_{i+1,t+1})$$

- We show that this scoring rule is also proper.

- Thus OOS tests based on CL are related to the cKLIC, analogous to OOS tests based on the (full) likelihood being related to the KLIC.
Multi-stage estimation of the complete model

- In our empirical work we use an AR(1) for the mean: $\hat{\theta}_i^{\text{mean}} \forall i$
- Estimate the individual variance models using the HAR model: $\hat{\theta}_i^{\text{var}} \forall i$
- Estimate the HAR-correlation model: $\hat{\theta}_i^{\text{corr}}$
- Compute the standardized uncorrelated residuals

$$\hat{e}_t = \hat{A}_t^{-1/2} r_t$$

and estimate their (symmetric) marginal distributions: $\hat{\theta}_i^{\text{mar}} \forall i$
- Estimate the jointly symmetric copula model: $\hat{\theta}_i^{\text{cop}}$
- Define

$$\hat{\theta}_{MSML} = \left[ \hat{\theta}_1^{\text{mean}}, \ldots, \hat{\theta}_N^{\text{mean}}, \hat{\theta}_1^{\text{var}}, \ldots, \hat{\theta}_N^{\text{var}}, \hat{\theta}_i^{\text{corr}}, \hat{\theta}_1^{\text{mar}}, \ldots, \hat{\theta}_N^{\text{mar}}, \hat{\theta}_i^{\text{cop}} \right]$$

- Multi-stage ML estimation (including with a composite likelihood stage) is a form of multi-stage GMM estimation, and under standard regularity conditions it can be shown (see Newey and McFadden, 1994) that

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

2 Models of linear and nonlinear dependence
   ■ Jointly symmetric copulas
   ■ A new covariance matrix model

3 Estimation and comparison via composite likelihood

4 Simulation study

5 Analysis of S&P 100 equity returns
We consider the estimation of jointly symmetric copula parameters via composite likelihood, compared with maximum likelihood (where feasible).

We use the JS Clayton and JS Gumbel copulas

Dimension of problem varies: \( N \in \{2, 3, 5, 10, \ldots, 100\} \).

Sample size is \( T = 1000 \).
Standard deviation of ML and CL as a function of N
ML is best, but infeasible for N>10; CL-adj gets close to CL-all for N>50
Outline

1 Introduction

2 Models of linear and nonlinear dependence
   - Jointly symmetric copulas
   - A new covariance matrix model

3 Estimation and comparison via composite likelihood

4 Simulation study

5 Analysis of S&P 100 equity returns
We study daily returns on all constituents of the S&P 100 index \((N = 104)\) over the period January 2006–December 2012 \((T = 1761)\).

High frequency data is from the NYSE TAQ database, cleaned following Barndorff-Nielsen, Hansen, Lunde and Shephard (2009).

- We adjust for stock splits and dividends using the adjustment factor from CRSP.

- We use 5-minute sampling to compute the realized covariance matrix.
### Summary stats and mean models

**Cross-sectional distribution**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>-0.07</td>
<td>-0.66</td>
<td>-0.32</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.56</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>11.86</td>
<td>6.92</td>
<td>8.47</td>
<td>10.50</td>
<td>13.40</td>
<td>20.02</td>
</tr>
<tr>
<td>Corr</td>
<td>0.47</td>
<td>0.33</td>
<td>0.40</td>
<td>0.46</td>
<td>0.52</td>
<td>0.63</td>
</tr>
</tbody>
</table>

**Summary statistics**

- Skewness: $-0.07$, $-0.66$, $-0.32$, $-0.03$, $0.18$, $0.56$
- Kurtosis: $11.86$, $6.92$, $8.47$, $10.50$, $13.40$, $20.02$
- Corr: $0.47$, $0.33$, $0.40$, $0.46$, $0.52$, $0.63$

**Conditional mean model**

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.00</td>
<td>-0.13</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.05</td>
<td>-0.08</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.13</td>
<td>-0.06</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Tests for skewness, kurtosis, and correlation**

<table>
<thead>
<tr>
<th>Test</th>
<th># of rejections</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0 : Skew [r_{it}] = 0$</td>
<td>3 out of 104</td>
</tr>
<tr>
<td>$H_0 : Kurt [r_{it}] = 3$</td>
<td>104 out of 104</td>
</tr>
<tr>
<td>$H_0 : Corr [r_{it}, r_{jt}] = 0$</td>
<td>5356 out of 5356</td>
</tr>
</tbody>
</table>
Volatility and correlation models
We also consider a GJR-GARCH/DCC model (details in paper)

<table>
<thead>
<tr>
<th>Cross-sectional distribution</th>
<th>Mean</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
</tr>
</thead>
</table>

### Variance model

<table>
<thead>
<tr>
<th></th>
<th>(\phi_i^{(const)})</th>
<th>(\phi_i^{(day)})</th>
<th>(\phi_i^{(week)})</th>
<th>(\phi_i^{(mth)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.00</td>
<td>0.38</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>HAR day</td>
<td>-0.08</td>
<td>0.32</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td>HAR week</td>
<td>-0.04</td>
<td>0.35</td>
<td>0.28</td>
<td>0.20</td>
</tr>
<tr>
<td>HAR month</td>
<td>-0.01</td>
<td>0.38</td>
<td>0.31</td>
<td>0.21</td>
</tr>
</tbody>
</table>

### Correlation model

<table>
<thead>
<tr>
<th></th>
<th>Est</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAR day (a)</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>HAR week (b)</td>
<td>0.32</td>
<td>0.02</td>
</tr>
<tr>
<td>HAR month (c)</td>
<td>0.38</td>
<td>0.03</td>
</tr>
</tbody>
</table>
## Marginal distribution models

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>5%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cross-sectional distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Std dev</td>
<td>1.09</td>
<td>0.96</td>
<td>1.02</td>
<td>1.08</td>
<td>1.14</td>
<td>1.29</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.16</td>
<td>-1.58</td>
<td>-0.47</td>
<td>-0.08</td>
<td>0.34</td>
<td>0.72</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>13.12</td>
<td>5.06</td>
<td>6.84</td>
<td>9.87</td>
<td>16.03</td>
<td>32.72</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.00</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HAR</th>
<th>DCC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marginal t distribution parameter estimates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAR</td>
<td>5.30</td>
<td>4.12</td>
</tr>
</tbody>
</table>

|                      |       |      |      |        |      |      |
| **Tests for skewness, kurtosis, and correlation** |       |      |
| $H_0 : Skew [e_{it}] = 0$ | 4 out of 104 | 6 out of 104 |
| $H_0 : Kurt [e_{it}] = 3$  | 104 out of 104 | 104 out of 104 |
| $H_0 : Corr [e_{it}, e_{jt}] = 0$ | 497 out of 5356 | 1 out of 5356 |
We consider three classes of models for the standardized residuals ($e_t$):

- **Jointly symmetric copula** models (Clayton, Gumbel, Frank and $t$) combined with $N$ Student’s $t$ distributions for the marginals
- The **independence** copula model, with $N$ Student’s $t$ dist’ns for the marginals
- A **jointly symmetric multivariate** $t$ distribution

The first two are copula-based approaches, allowing for separate specification of the marginals and copula

The third corresponds to existing “best practice” for this problem

We do not even bother considering the MV Normal distribution...
<table>
<thead>
<tr>
<th></th>
<th>Jointly symmetric copula models</th>
<th>Benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>Clayton</td>
</tr>
<tr>
<td>HAR</td>
<td>Est.</td>
<td>39.44</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>4.35</td>
</tr>
<tr>
<td>t-test of indep</td>
<td></td>
<td>8.45</td>
</tr>
<tr>
<td>Rank of LogL</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>DCC</td>
<td>Est.</td>
<td>28.21</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>5.50</td>
</tr>
<tr>
<td>t-test of indep</td>
<td></td>
<td>6.13</td>
</tr>
<tr>
<td>Rank of LogL</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>
Linear correlation from the HAR model (Citi-GS)

Correlation varies from 0.25 to 0.75 over the sample period
Quantile dependence (1%) from the JS t model (Citi-GS)

Quant dep(q) = C(q,q)/q. Ranges from 0.03 to 0.35
Model comparison tests

- We use the composite likelihood KLIC (cKLIC) to compare these models:
  
  \[ H_0 : E \left[ CL_t^A - CL_t^B \right] = 0 \]
  
  vs.
  
  \[ H_1 : E \left[ CL_t^A - CL_t^B \right] > 0 \]
  
  \[ H_2 : E \left[ CL_t^A - CL_t^B \right] < 0 \]

- Rivers and Vuong (2002) provide a method for testing this null (in-sample) for the non-nested models.

- We use Giacomini and White (2006) to test the out-of-sample analogue of this null.

  - OOS comparisons involve a penalty for excess parameters.
  
  - We use a rolling window estimation scheme, with the last two years as the OOS period.
In-sample model comparison t-statistics: HAR vs HAR

A positive value indicates the column model beats the row model.

<table>
<thead>
<tr>
<th>t^{JS}</th>
<th>t^{JS}</th>
<th>Clayton^{JS}</th>
<th>Frank^{JS}</th>
<th>Gumbel^{JS}</th>
<th>Indep</th>
<th>MV t</th>
</tr>
</thead>
<tbody>
<tr>
<td>t^{JS}</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Clayton^{JS}</td>
<td>2.92</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Frank^{JS}</td>
<td>2.16</td>
<td>1.21</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Gumbel^{JS}</td>
<td>5.38</td>
<td>6.02</td>
<td>1.75</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Indep*</td>
<td>8.45</td>
<td>10.07</td>
<td>13.43</td>
<td>5.25</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MV t</td>
<td>19.70↑</td>
<td>19.52</td>
<td>19.45</td>
<td>19.23</td>
<td>18.40‡</td>
<td>–</td>
</tr>
</tbody>
</table>

- The jointly symmetric t copula model significantly beats all competitors.
- The multivariate t distribution is beaten by all competitors.
In-sample model comparison t-statistics: DCC vs DCC
A positive value indicates the column model beats the row model

<table>
<thead>
<tr>
<th></th>
<th>$t^{JS}$</th>
<th>Clayton$^{JS}$</th>
<th>Frank$^{JS}$</th>
<th>Gumbel$^{JS}$</th>
<th>Indep</th>
<th>MV $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton$^{JS}$</td>
<td>-</td>
<td>4.48</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank$^{JS}$</td>
<td>2.69</td>
<td>1.27</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gumbel$^{JS}$</td>
<td>6.74</td>
<td>7.47</td>
<td>1.74</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indep*</td>
<td>6.13</td>
<td>7.36</td>
<td>10.36</td>
<td>4.40</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MV $t$</td>
<td>18.50†</td>
<td>18.11</td>
<td>17.94</td>
<td>17.60</td>
<td>15.69‡</td>
<td>-</td>
</tr>
</tbody>
</table>

- The jointly symmetric $t$ copula model significantly beats all competitors
- The multivariate $t$ distribution is beaten by all competitors
In-sample model comparison t-statistics: HAR vs DCC

A positive value indicates the column model beats the row model

<table>
<thead>
<tr>
<th>DCC models</th>
<th>t&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Clayton&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Frank&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Gumbel&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Independ</th>
<th>MV t</th>
</tr>
</thead>
<tbody>
<tr>
<td>t&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>7.86</td>
<td>7.85</td>
<td>7.85</td>
<td>7.84</td>
<td>7.82</td>
<td><strong>6.92</strong></td>
</tr>
<tr>
<td>Clayton&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>7.86</td>
<td><strong>7.86</strong></td>
<td>7.85</td>
<td>7.85</td>
<td>7.83</td>
<td>6.93</td>
</tr>
<tr>
<td>Frank&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>7.85</td>
<td>7.85</td>
<td><strong>7.84</strong></td>
<td>7.83</td>
<td>7.82</td>
<td>6.91</td>
</tr>
<tr>
<td>Gumbel&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>7.88</td>
<td>7.87</td>
<td>7.87</td>
<td><strong>7.86</strong></td>
<td>7.84</td>
<td>6.94</td>
</tr>
<tr>
<td>Independ*</td>
<td>7.90</td>
<td>7.90</td>
<td>7.90</td>
<td>7.89</td>
<td><strong>7.87</strong></td>
<td>6.97</td>
</tr>
<tr>
<td>MV t</td>
<td>8.95</td>
<td>8.95</td>
<td>8.94</td>
<td>8.94</td>
<td>8.92</td>
<td><strong>8.03</strong></td>
</tr>
</tbody>
</table>

- **HAR models beat** DCC equivalents for all choices of copula model
- Even the worst HAR model significantly **beats** the best DCC model
The jointly symmetric $t$, Clayton and Frank copula models are signif **better** than all others, but not signif diff from each other.

The multivariate $t$ distribution is still **beaten** by all competitors.
### Out-of-sample model comparison t-statistics: DCC vs DCC

A positive value indicates the column model beats the row model.

<table>
<thead>
<tr>
<th></th>
<th>$t^{JS}$</th>
<th>Clayton$^{JS}$</th>
<th>Frank$^{JS}$</th>
<th>Gumbel$^{JS}$</th>
<th>Indep</th>
<th>MV $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton$^{JS}$</td>
<td>1.55</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank$^{JS}$</td>
<td>1.79</td>
<td>1.34</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gumbel$^{JS}$</td>
<td>2.96</td>
<td>3.31</td>
<td>0.01</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indep</td>
<td>3.10</td>
<td>3.12</td>
<td>2.38</td>
<td>2.44</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>MV $t$</td>
<td>14.65</td>
<td>14.33</td>
<td>14.56</td>
<td>13.88</td>
<td>12.80</td>
<td>-</td>
</tr>
</tbody>
</table>

- The jointly symmetric $t$, Clayton and Frank copula models are signif **better** than all others, but **not signif diff** from each other.

- The multivariate $t$ distribution is still **beaten** by all competitors.
Out-of-sample model comparison t-statistics: HAR vs DCC
A positive value indicates the column model beats the row model

<table>
<thead>
<tr>
<th>HAR models</th>
<th>t&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Clayton&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Frank&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Gumbel&lt;sup&gt;JS&lt;/sup&gt;</th>
<th>Indep</th>
<th>MV t</th>
</tr>
</thead>
<tbody>
<tr>
<td>t&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>5.23</td>
<td>5.23</td>
<td>5.23</td>
<td>5.23</td>
<td>5.22</td>
<td>4.55</td>
</tr>
<tr>
<td>Clayton&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>5.22</td>
<td>5.23</td>
<td>5.23</td>
<td>5.23</td>
<td>5.22</td>
<td>4.55</td>
</tr>
<tr>
<td>Frank&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>5.23</td>
<td>5.22</td>
<td>5.23</td>
<td>5.22</td>
<td>5.21</td>
<td>4.55</td>
</tr>
<tr>
<td>Gumbel&lt;sup&gt;JS&lt;/sup&gt;</td>
<td>5.24</td>
<td>5.24</td>
<td>5.24</td>
<td>5.23</td>
<td>5.22</td>
<td>4.56</td>
</tr>
<tr>
<td>Indep</td>
<td>5.24</td>
<td>5.24</td>
<td>5.24</td>
<td>5.23</td>
<td>5.22</td>
<td>4.56</td>
</tr>
<tr>
<td>MV t</td>
<td>6.05</td>
<td>6.05</td>
<td>6.05</td>
<td>6.05</td>
<td>6.04</td>
<td>5.41</td>
</tr>
</tbody>
</table>

- **HAR models** beat DCC equivalents for all choices of copula model
- Even the worst HAR model significantly **beats** the best DCC model
Summary and conclusion

- We propose a new class of **dynamic, high-dimensional** distribution models
  - We exploit **high frequency data** to accurately measure and model linear dependence (correlation)
  - We use a new class of **jointly symmetric copulas** to capture any remaining nonlinear dependence
  - We consider **composite likelihood** estimation and model comparison to overcome the computational burden of estimating our JS copulas

- In an application to daily returns on 104 US equities, we find:
  - Significant gains to using high frequency data for estimating linear dependence
  - Significant gains from capturing the remaining nonlinear dependence using a jointly symmetric copula
  - Both of the above conclusions hold both **in- and out-of-sample**
Use copula model to capture entire dependence structure


Model covariance matrix and combine with a Normal or Student’s $t$ distribution

- Jondeau and Rockinger (2012), Hautsch et al. (2013), Jin and Maheu (2013), and others

Combine a covariance matrix model for returns and a copula model for the uncorrelated residuals

- This paper, and Lee and Long (2009)
Comparison with Lee and Long (2009)

- Lee and Long also suggest a linear/nonlinear decomposition:

\[ r_t = \mu_t + H_t^{1/2} \Sigma^{-1/2} w_t \]

where \( w_t \sim iid \ G = C_w(G_1, ..., G_N) \)

\[ \mathbb{E}_{t-1}[w_{it}] = 0, \quad \mathbb{E}_{t-1}[w_{it}^2] = 1 \quad \text{and} \quad \Sigma = \mathbb{E}_{t-1}[w_t w'_t] \]

- Key differences from our approach:
  - LL allow for any model \( G \), and impose the zero correlation constraint by rotating the variables, \( w_t \), by their covariance matrix, \( \Sigma \).
  - This step rules out multistage estimation of \( G \), as all marginals and the coupla are needed to compute \( \Sigma \).
  - The covariance matrix \( \Sigma \) usually requires numerical methods for computation.
  - Smaller: LL use a GARCH model for \( H_t \), while we exploit recent work in high frequency methods to estimate this.
Our model:

\[ r_t = \mu_t + H_t^{1/2} e_t \]

where \( e_t \sim iid \ F = C^{JS} (F_1, \ldots, F_N) \)

All components of this model are parametric: covariance, marginals, copula

- All are thus subject to model misspecification
- In high dimension applications some parametric structure is needed

Residuals \( e_t \) are \( iid \) ⇒ all dynamics in this model come from \( H_t \) (and \( \mu_t \))

- Rules out separate variation in higher-order moments or other dep measures
- Second-moment variation is easily most prominent in financial data

Joint symmetry assumption implies returns are conditionally symmetric

- Will not be plausible in some applications
- Can use Lee-Long method if needed (computationally difficult)