Modelling Dependence in High Dimensions with Factor Copulas

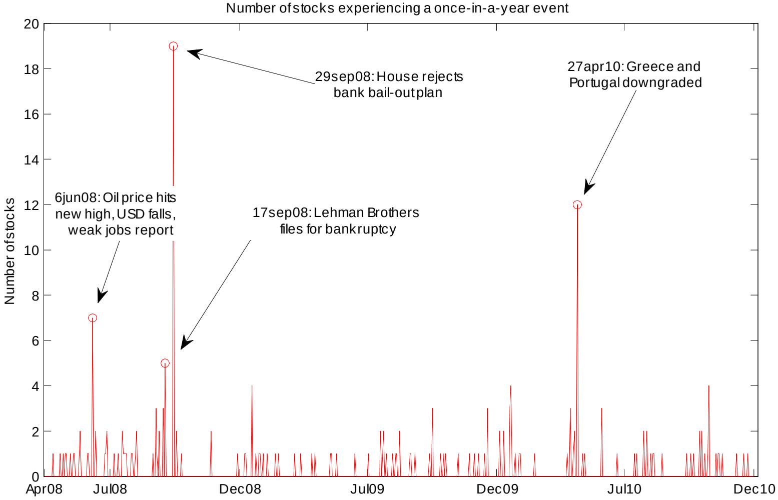
Dong Hwan Oh and Andrew Patton

Department of Economics
Duke University

September 2012

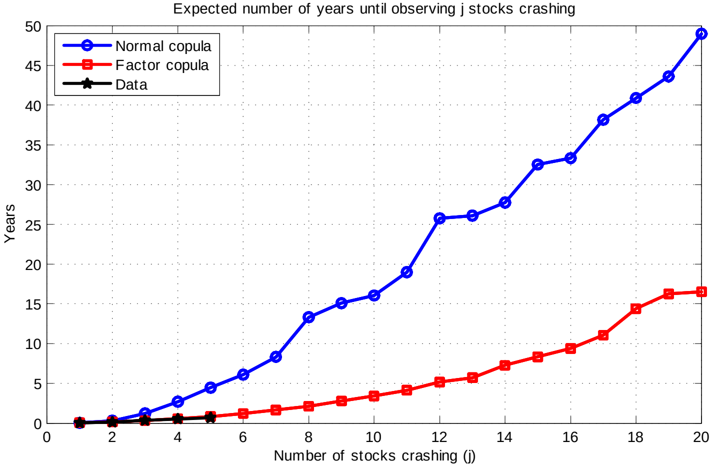
Joint crashes, S&P 100 stocks, 2008-2010

Number of stocks experiencing a once-in-a-year crash



Expected wait times to observe j once-in-a-year crashes

Observed frequency of joint crashes is *much* higher than as implied by a Normal copula



Motivation

- Crashes are not usually contained to pairs or small collections of assets; they can (and do) sometimes affect **large collections** of assets (eg, returns on financial companies, defaults on credit, etc. during the financial crisis)

- One of the difficulties with studying the dependence between large numbers of assets is the relative paucity of models available
  - There is a growing literature on models for **large covariance matrices** (eg, Engle and Kelly, 2008, Engle, Shephard and Sheppard, 2008, Hautsch, Kyj and Oomen, 2010)
  - However there is a relative **lack of models for dependence** beyond linear correlation

Joint distributions and copulas

- From Sklar (1959), we can decompose a N-dim joint distribution into its N univariate marginal distributions and a N-dim copula:

$$\mathbf{F}(x_1, x_2, \cdots, x_N) = \mathbf{C}(F_1(x_1), F_2(x_2), \cdots, F_N(x_N)) \quad \forall\, \mathbf{x} \in \mathbb{R}^N$$

$$\text{where } \mathbf{C} : [0,1]^N \rightarrow [0,1]$$

- Existing work in econometrics provides us with good, flexible models for univariate (conditional) distributions, F_i.
- The goal of this paper is to provide new models for C**,** with particular attention paid to the case that N is large (between, say 20 and 100).

Main contributions of this paper

1. A flexible, easily interpreted, class of **copula models** that may be applied in high dimensional problems.
   - Closed-form expression for these models not generally available, but:
     - Simulation-based estimation is simple and fast
     - Analytical results on tail dependence available using EVT
2. In our empirical **application** we study S&P 100 daily equity returns
   - We find significant evidence of tail dependence, asymmetric dependence, and heterogeneous dependence.
   - Our new copula models provide improved estimates of systematic risk.

Outline

Characteristics of a high dimension copula model

1. Allows for **tail dependence**?

   → non-zero probability of joint extreme events

2. Allows for **asymmetric dependence**?

   → crashes and booms are not assumed identical

3. Allows for **heterogeneous dependence**?

   → dependence between (Y_i, Y_j) not assumed identical for all pairs

4. Allows for **interpretable constaints on dependence?**

   → parameter restrictions can be explained, understood, and tested

Existing models for high dimension copulas

- **Normal:** Li (2000, *J.Fix Inc*), and many others
- **Student's t:** Embrechts *et al.* (2002), Fang *et al.* (2002 *JMVA*)
- **Archimedean** (eg, Clayton, Gumbel, Frank)

★ **Nested Archimedean copulas:** Hofert and Scherer (2011 *Quant Fin*)**,** Joe (1997 *book*), McNeil et al. (2005 *book*)

★ **Grouped t-copulas:** Daul et al. (2003 *RISK*)

★ **Pair-copula constructions (vines):** Aas, *et al.* (2007 *IME*), Heinen and Valdesogo (2009 wp), Acar, *et al.* (2012 *JMVA*)

★ **Skew t copulas:** Smith, *et al.* (2010 *JAE*), Christoffersen, *et al.* (2011 wp), Demarta and McNeil (2004 *ISR*)

“Explicit” and “Implicit” copulas

- McNeil, Embrechts and Frey (2005) characterize copula models as:

1. “**Explicit**”: those with a known (simple) functional form:

$$\text{Clayton:} \quad \mathbf{C}(u_1, u_2) = \left(u_1^{-\gamma} + u_2^{-\gamma} - 1\right)^{-1/\gamma}$$

"Explicit" and "Implicit" copulas

- McNeil, Embrechts and Frey (2005) characterize copula models as:

1. "**Explicit**": those with a known (simple) functional form:

$$\text{Clayton:} \quad \mathbf{C}(u_1, u_2) = \left(u_1^{-\gamma} + u_2^{-\gamma} - 1\right)^{-1/\gamma}$$

2. "**Implicit**": those obtained from a known multivariate distribution:

$$\text{Normal:} \quad \mathbf{C}(u_1, u_2) = \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$$

"Explicit" and "Implicit" copulas

- McNeil, Embrechts and Frey (2005) characterize copula models as:

1. "**Explicit**": those with a known (simple) functional form:

$$\text{Clayton:} \quad \mathbf{C}(u_1, u_2) = \left(u_1^{-\gamma} + u_2^{-\gamma} - 1\right)^{-1/\gamma}$$

2. "**Implicit**": those obtained from a known multivariate distribution:

$$\text{Normal:} \quad \mathbf{C}(u_1, u_2) = \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$$

★ "**Really implicit**": *those obtained from known multivariate models*

A simple factor copula model

- Consider a vector of n variables, $\mathbf{Y}$, with some joint distribution $\mathbf{F}^*$, marginal distributions F_i^*, and copula $\mathbf{C}^*$

$$[Y_1, ..., Y_n]' \equiv \mathbf{Y} \sim \mathbf{F}^* = \mathbf{C}^* \left(F_1^*, ..., F_n^*\right)$$

A simple factor copula model

- Consider a vector of n variables, $\mathbf{Y}$, with some joint distribution $\mathbf{F}^*$, marginal distributions F_i^*, and copula $\mathbf{C}^*$

$$\left[Y_1, ..., Y_n\right]' \equiv \mathbf{Y} \sim \mathbf{F}^* = \mathbf{C}^* \left(F_1^*, ..., F_n^*\right)$$

- Our model for $\mathbf{C}^*$ is the copula $\mathbf{C}\left(\boldsymbol{\theta}\right)$ implied by the following model:

$$\begin{aligned} \text{Let} \quad X_i &= Z + \varepsilon_i, \quad i = 1, 2, ..., n \\ Z &\sim F_z\left(\boldsymbol{\theta}\right), \quad \varepsilon_i \sim iid\ F_\varepsilon\left(\boldsymbol{\theta}\right), \quad Z \perp\!\!\!\perp \varepsilon_i \ \forall\ i \\ \text{So} \quad \left[X_1, ..., X_n\right]' &\equiv \mathbf{X} \sim \mathbf{F}_x\left(\boldsymbol{\theta}\right) = \mathbf{C}\left(G_1\left(\boldsymbol{\theta}\right), ..., G_n\left(\boldsymbol{\theta}\right); \boldsymbol{\theta}\right) \end{aligned}$$

A simple factor copula model

- Consider a vector of n variables, $\mathbf{Y}$, with some joint distribution $\mathbf{F}^*$, marginal distributions F_i^*, and copula $\mathbf{C}^*$

$$[Y_1, ..., Y_n]' \equiv \mathbf{Y} \sim \mathbf{F}^* = \mathbf{C}^* \left(F_1^*, ..., F_n^*\right)$$

- Our model for $\mathbf{C}^*$ is the copula $\mathbf{C}(\boldsymbol{\theta})$ implied by the following model:

$$\begin{aligned} \text{Let} \quad X_i &= Z + \varepsilon_i, \quad i = 1, 2, ..., n \\ Z &\sim F_z(\boldsymbol{\theta}), \quad \varepsilon_i \sim iid\ F_\varepsilon(\boldsymbol{\theta}), \quad Z \perp\!\!\!\perp \varepsilon_i\ \forall\ i \\ \text{So} \quad [X_1, ..., X_n]' &\equiv \mathbf{X} \sim \mathbf{F}_x(\boldsymbol{\theta}) = \mathbf{C}\left(G_1(\boldsymbol{\theta}), ..., G_n(\boldsymbol{\theta}); \boldsymbol{\theta}\right) \end{aligned}$$

- In general we won't know $\mathbf{C}(\boldsymbol{\theta})$ in closed form, but we can nevertheless use it as a model for the true copula $\mathbf{C}^*$.

A simple factor copula model, discussion I

$$\begin{aligned} [Y_1, ..., Y_n]' &\equiv \mathbf{Y} \sim \mathbf{F}^* &= \mathbf{C}(F_1^*, \quad ... \quad, F_n^* \quad ; \boldsymbol{\theta}) \\ [X_1, ..., X_n]' &\equiv \mathbf{X} \sim \mathbf{F}_x(\boldsymbol{\theta}) &= \mathbf{C}(G_1(\boldsymbol{\theta}), ..., G_n(\boldsymbol{\theta}); \boldsymbol{\theta}) \end{aligned}$$

1. The structure for $[X_1, ..., X_n]' \equiv \mathbf{X}$ also provides marginal distributions for X_i, but we discard these and only "extract" the copula from this structure. In estimation we will only use "copula information" from this structure
   - This is what makes this a model for the *copula* of the data and not for the joint distribution. (Marginal distributions are modelled in a separate stage.)
   - This also means that we must be careful that all parameters in $\boldsymbol{\theta}$ are identified given only the copula.

A simple factor copula model, discussion II

2. When Z and ε_i are Normal, the resulting factor copula is Normal. For almost all other combinations, the distribution of $\mathbf{X}$ is not previously known and **not known in closed form**.

   - Thus it is simple to nest the important (but restrictive) Normal copula. Useful for model comparisons
   - When distribution of $\mathbf{X}$ is not known in closed form, we propose using simulation-based methods to estimate $\boldsymbol{\theta}$
   - For many types of factor structures simulation is **fast and simple**.

A simple factor copula model, discussion III

3. Even without a closed-form expression for $\mathbf{C}(\boldsymbol{\theta})$, we can still obtain properties of that copula. Eg:

   - If Z is at least as fat-tailed as ε_i, then this copula will exhibit **tail dependence**
   - If Z is skewed and ε_i is symmetric then this copula will exhibit **asymmetric dependence**
   - With a trivial modification of above structure, can acommodate both **positive and negative dependence**.
   - Can impose **equi-dependence,** block equi-dependence, or allow for **heterogeneous dependence**

Tail dependence properties of a factor copula

- Despite the lack of closed-form expression, we can still obtain tail dependence properties of (linear) factor copulas using EVT:

Proposition (Tail dependence for a linear factor copula): If F_z and F_ε have regularly varying tails with a common tail index $\alpha > 0$ and constants A_z^L, A_z^U, A_ε^L, A_ε^U, eg:

$$\Pr[Z < -s] = A_z^L s^{-\alpha}, \text{ as } s \to \infty$$

then the lower tail dependence coefficient of the implied factor copula is:

$$\lambda^L \equiv \lim_{s \to \infty} \frac{\Pr[X_1 < -s \cap X_2 < -s]}{\Pr[X_1 < -s]} = \frac{A_z^L}{A_z^L + A_\varepsilon^L}$$

A corresponding result holds for the upper tail dependence coefficient.

Extensions of the simple factor copula

- **Heterogeneous dependence between pairs of assets:**

$$
\begin{aligned}
X_i &= \beta_i Z + \varepsilon_i, \quad i = 1, 2, ..., n \\
Z &\sim F_z(\boldsymbol{\theta}), \quad \varepsilon_i \sim iid \ F_\varepsilon(\boldsymbol{\theta}), \quad Z \perp\!\!\!\perp \varepsilon_i \ \forall \ i
\end{aligned}
$$

Extensions of the simple factor copula

- **Heterogeneous dependence between pairs of assets:**

$$\begin{aligned} X_i &= \beta_i Z + \varepsilon_i, \quad i = 1, 2, ..., n \\ Z &\sim F_z(\boldsymbol{\theta}), \quad \varepsilon_i \sim iid \ F_\varepsilon(\boldsymbol{\theta}), \quad Z \perp\!\!\!\perp \varepsilon_i \ \forall \ i \end{aligned}$$

- **Multiple common factors:**

$$\begin{aligned} X_i &= \sum_{j=1}^{J} \beta_{ij} Z_j + \varepsilon_i, \quad i = 1, 2, ..., n \\ \varepsilon_i &\sim iid \ F_\varepsilon, \quad Z_j \perp\!\!\!\perp \varepsilon_i \ \forall \ i, j \\ [Z_1, ..., Z_J]' &\equiv \mathbf{Z} \sim \mathbf{F}_z = \mathbf{C}_z\left(G_{z_1}, ..., G_{z_J}\right) \end{aligned}$$

Extensions of the simple factor copula

- **Heterogeneous dependence between pairs of assets:**

$$\begin{aligned} X_i &= \beta_i Z + \varepsilon_i, \quad i = 1, 2, ..., n \\ Z &\sim F_z(\boldsymbol{\theta}), \quad \varepsilon_i \sim iid \ F_\varepsilon(\boldsymbol{\theta}), \quad Z \perp\!\!\!\perp \varepsilon_i \ \forall \ i \end{aligned}$$

- **Multiple common factors:**

$$\begin{aligned} X_i &= \sum_{j=1}^{J} \beta_{ij} Z_j + \varepsilon_i, \quad i = 1, 2, ..., n \\ \varepsilon_i &\sim iid \ F_\varepsilon, \quad Z_j \perp\!\!\!\perp \varepsilon_i \ \forall \ i, j \\ [Z_1, ..., Z_J]' &\equiv \mathbf{Z} \sim \mathbf{F}_z = \mathbf{C}_z\left(G_{z_1}, ..., G_{z_J}\right) \end{aligned}$$

- **Time-varying dependence:** let $\sigma^2_{z,t} \equiv V_{t-1}[Z_t]$, where

$$\sigma^2_{z,t} = \omega + \beta \sigma^2_{z,t-1} + \alpha \cdot h\left(Y_{1,t-1}, ..., Y_{N,t-1}\right)$$

Extensions to non-linear factor copulas

- It is simple to generalize the class of factor copulas to consider other ways of combining Z and ε_i :

$$X_i = h(Z, \varepsilon_i), i = 1, 2, ..., n$$

$$\begin{aligned} Z &\sim F_z(\boldsymbol{\theta}) \\ \varepsilon_i &\sim iid\ F_\varepsilon(\boldsymbol{\theta}), \quad Z \perp\!\!\!\perp \varepsilon_i \ \forall\ i \end{aligned}$$

$$[X_1, ..., X_n]' \equiv \mathbf{X} \sim \mathbf{F}_x(\boldsymbol{\theta}) = \mathbf{C}(G_1, ..., G_n; \boldsymbol{\theta})$$

- By considering different functions h and distributions (F_Z, F_ε) we can obtain existing copulas and generate new models:

Existing non-linear factor copulas

- $X_i = h(Z, \varepsilon_i) \Rightarrow [X_1, ..., X_n]' \sim \mathbf{F}_X = \mathbf{C}(G_1, ..., G_n; \boldsymbol{\theta})$

$\mathbf{C}(\boldsymbol{\theta})$	$h(z, \varepsilon)$	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	$-(\log z - \log \varepsilon)^{\alpha}$	$Stable(1/\alpha, 1, 1, 0)$	$Exp(1)$

- By abandoning the requirement for a closed-form expression for the copula, much flexibility is gained

Illustration of some factor copulas

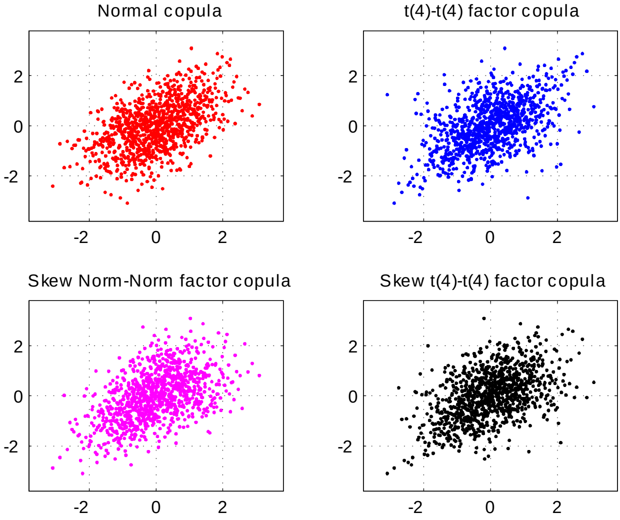
- Consider the following factor structure:

$$\begin{aligned} \text{Let} \quad X_i &= Z + \varepsilon_i, \quad i = 1, 2, \ldots, n \\ \varepsilon_i &\sim iid \;\; t(\nu), \quad Z \perp\!\!\!\perp \varepsilon_i \; \forall \; i \\ Z &\sim Skew \;\; t\left(\sigma_z^2, \nu, \lambda\right) \\ \nu &\in [2, \infty], \quad \lambda \in [-0.99, 0] \end{aligned}$$

- We set σ_z^2 so that the factor copula implied by this structure generates linear correlation of 0.5. (Roughly, $\sigma_z^2 = 1$.)

Scatterplots of joint dist'ns with factor copulas

Marginal dist'ns are N(0,1), linear correlation = 0.5.



Quantile dependence and "crash" dependence

1. **Quantile dependence**: conditional on one variable being in its q tail, what is prob that other variable is in its q tail?

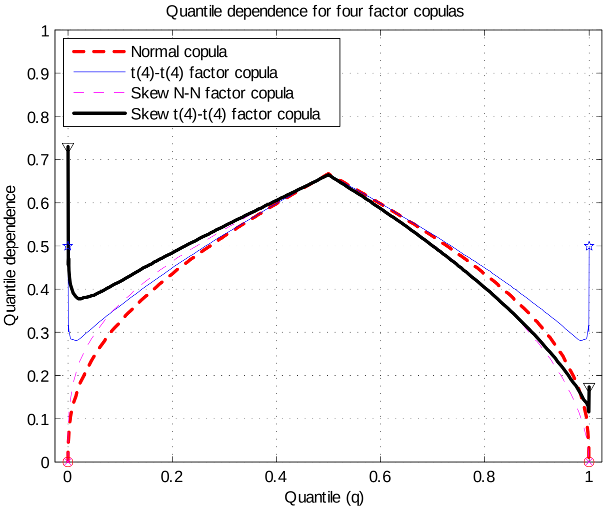
$$\begin{aligned}\lambda_q^L &\equiv \Pr\left[U_1 \leq q | U_2 \leq q\right] = \frac{\mathbf{C}\left(q, q; \boldsymbol{\theta}\right)}{q} \\ \lambda_q^U &\equiv \Pr\left[U_1 > q | U_2 > q\right] = \frac{1 - 2q + \mathbf{C}\left(q, q; \boldsymbol{\theta}\right)}{1 - q}\end{aligned}$$

2. **Crash dependence** (related to a measure in Embrechts, *et al.*, 2000): Conditional on j variables being in their q tails, what is expected proportion of remaining variables that are in their q tails?

$$\begin{aligned}\pi_j^q &\equiv \frac{\kappa_j^q}{n - j} \\ \text{where} \quad \kappa_j^q &= E\left[N_q^* | N_q^* \geq j\right] - j \\ N_q^* &\equiv \textstyle\sum_{i=1}^{n} \mathbf{1}\left\{U_i \leq q\right\}\end{aligned}$$

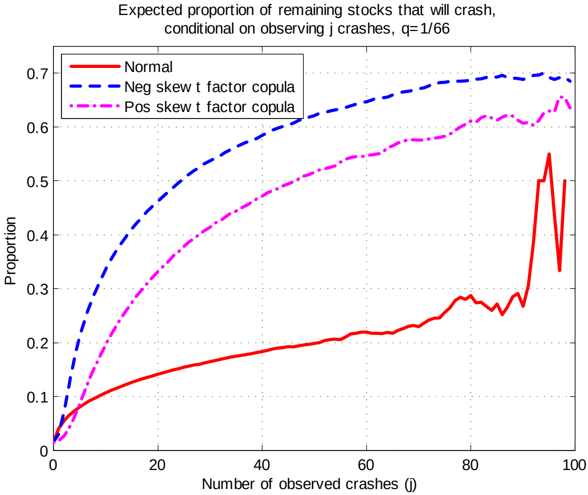
Quantile dependence for factor copulas

Probabilities of joint crashes and joint booms of varying severity



Proportion of remaining stocks that will crash

"Crash" defined as a 1/66 event = once in a quarter for daily asset returns



Outline

Estimation via SMM-type method

- Estimation of the proposed factor copulas is complicated by the fact that their likelihood is not known in closed form
- In a companion paper, we propose estimation via a **simulated method of moments-type** estimator
  - We match measures of dependence computed from the data with those measured from simulations of the model
  - Our estimator is *not* strictly SMM, as our "moments" are functions of sample ranks and not sample means

Definition of the estimator I

- Consider the following measures of dependence

$$\hat{\rho}_T \equiv 12 \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \left(Y_{it} \right) \hat{F}_j \left(Y_{jt} \right) - 3$$

$$\hat{\lambda}_T^q \equiv \frac{1}{q} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1} \left\{ \hat{F}_i \left(Y_{it} \right) < q \cap \hat{F}_j \left(Y_{jt} \right) < q \right\}$$

where $\hat{F}_i \left(y \right) \equiv \left(T + 1 \right)^{-1} \sum_{t=1}^{T} \mathbf{1} \left\{ Y_{it} < y \right\}$.

- Could consider other dependence measures: Kendall's tau, Blomqvist's beta, etc.

Definition of the estimator II

- Let $\hat{\mathbf{m}}_T$ denote a vector of sample measures of dependence

- Let $\tilde{\mathbf{m}}_S(\boldsymbol{\theta})$ denote the corresponding vector computed using S simulations of the factor copula using parameter vector $\boldsymbol{\theta}$

- Our estimator is defined as

$$\hat{\boldsymbol{\theta}}_{T,S} \equiv \arg\min_{\boldsymbol{\theta}} \mathbf{g}'_{T,S}(\boldsymbol{\theta}) W_T \mathbf{g}_{T,S}(\boldsymbol{\theta})$$

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

- This is implemented in the same way as SMM, but the asymptotic properties of $\hat{\boldsymbol{\theta}}_{T,S}$ require different methods of proof.

Main findings from the simulation

- The asymptotic results for our simulation-based estimator provide a **good approximation in finite samples**:

1. Parameter estimates are approximately unbiased
2. Confidence intervals have satisfactory coverage rates, when the step size is not too small
3. Tests of over-identifying restrictions have satisfactory finite-sample size
4. Estimation error from marginal distribution dynamics (AR-GARCH) does *not* affect accuracy of copula parameter estimates
5. Loss in efficiency (under Normality) is low :
    - From ML to SMM is moderate: around $10 - 25\%$
    - From GMM to SMM is negligible: around $0 - 3\%$

Outline

1. Introduction
2. Factor copula models
3. Simulation-based estimation and a Monte Carlo study
4. **Application to 100 U.S. stock returns**
5. Conclusion

Modelling equity return dependence I

- Daily US equity return data from CRSP:
  - $N = 100$, the individual stocks in the S&P 100 index
  - $T = 696$, sample period is Apr 08–Dec 10
- Our model for the joint distribution of these returns is based on:

$$\mathbf{Y}_t | \mathcal{F}_{t-1} \sim \mathbf{F}_t = \mathbf{C}\left(F_{1t}, F_{2t}, \cdots, F_{Nt}\right)$$

  - Similar to the CCC model of Bollerslev (1990)

Modelling equity return dependence II

- **Marginal distributions:**
  - AR-GARCH models to construct standardized residuals:

$$
\begin{aligned}
r_{it} &= \phi_0 + \phi_1 r_{i,t-1} + \phi_m r_{m,t-1} + \varepsilon_{it} \\
\sigma_{it}^2 &= \omega + \beta\sigma_{i,t-1}^2 + \alpha\varepsilon_{i,t-1}^2 + \gamma\varepsilon_{i,t-1}^2 \mathbf{1}\{\varepsilon_{i,t-1} < 0\} \\
&\quad + \alpha_m \varepsilon_{m,t-1}^2 + \gamma_m \varepsilon_{m,t-1}^2 \mathbf{1}\{\varepsilon_{m,t-1} < 0\} \\
\eta_{it} &\equiv \varepsilon_{it}/\sigma_{it} \sim iid\ F_{\eta,i}
\end{aligned}
$$

  - Empirical distribution functions for $F_{\eta,i}$:

$$
\hat{F}_{\eta,iT}(\eta) \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{\eta_{it} \leq \eta\}
$$

Modelling equity return dependence III

- **Copula:**
  - We consider a variety of new and existing copula models:

Copula
Clayton
Normal
Student's t
Skew t
Factor (t-t)
Factor (Skew t-t)

Heterogeneity
Equi-dependence
Block equi-dependence
Multi-factor

Summary statistics for pair-wise dependence

- Dependence measures using standardized residuals (4950 pairs)

	Mean	**5%**	**25%**	**Median**	**75%**	**95%**
ρ (linear corr)	0.42	0.28	0.35	0.42	0.48	0.60
ρ_s (rank corr)	0.44	0.30	0.37	0.43	0.50	0.62
$\frac{1}{2}(\tau_{0.99} + \tau_{0.01})$	0.07	0.00	0.00	0.07	0.07	0.22
$(\tau_{0.90} - \tau_{0.10})$	-0.08	-0.19	-0.13	-0.09	-0.04	0.03

Parameter estimates for copula models

Equi-dependence models

	σ_z^2		ν_z^{-1}		λ_z		Q_{SMM}	p-val
	Est	s.e.	Est	s.e.	Est	s.e.		
Clayton*	0.60	0.03	-	-	-	-	4.49	0.000
Normal	0.91	0.06	-	-	-	-	0.90	0.000
Student's t	0.86	0.05	0.03	0.03	-	-	1.19	0.000
Skew t	0.67	0.09	0.05	0.01	-8.30	4.02	0.10	0.002
$t - t$	0.90	0.06	0.01	0.05	-	-	0.98	0.000
Skew $t - t$	0.88	0.06	0.08	0.05	-0.23	0.05	0.07	0.001

- $\hat{\rho} = \frac{\hat{\sigma}_z^2}{\hat{\sigma}_z^2+1} \approx 0.46, \quad \hat{v}_z \approx 25$
- All models are rejected using the J-test

Block equi-dependence copula, by industry group

- We next divide the stocks into 7 groups using their 1-digit SIC code:

SIC	Industry	Number of Stocks
1	Mining & construction	6
2	Manufacturing: Food, apparel, furniture, etc	26
3	Manufacturing: Electronics, machinery, etc	25
4	Transportation, communications, utilities	11
5	Wholesale and retail trade	8
6	Finance, insurance, real estate	18
7	Services	6
	Total	**100**

A multi-factor copula model

- Consider a more flexible two-factor copula model:

$$X_i = \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \quad i = 1, 2, ..., n$$

$$\begin{aligned} Z_0 &\sim Skew\ t(\nu, \lambda), \quad \varepsilon_i \sim iid\ t(\nu), \quad i = 1, 2, ..., n \\ Z_S &\sim t(\nu), \quad S = 1, 2, ..., 7 \\ Z_i &\perp Z_j \perp \varepsilon_k \ \forall\ i, j, k \end{aligned}$$

- We impose that the parameters β_i and γ_i are the same for stocks in the same industry ($\Rightarrow$*block equi-dependence*)
- This model allows for more flexibility in modelling intra- and inter-industry correlations

Parameter estimates for copula models

Multi-factor block equi-dependence models - shape parameters

Copula	ν_z^{-1}		λ_z		$Q_T\left(\hat{\theta}\right)$	p-val
	Est	SE	Est	SE		
Normal	–	–	–	–	0.159	0.000
Student's t	0.073	0.027	–	–	0.157	0.000
Skew t	0.049	0.007	-9.660	1.086	0.027	0.044
$t-t$	0.066	0.047	–	–	0.139	0.000
Skew $t-t$	0.099	0.046	-0.222	0.055	0.019	0.072

- $\hat{\nu}_z \approx 14$

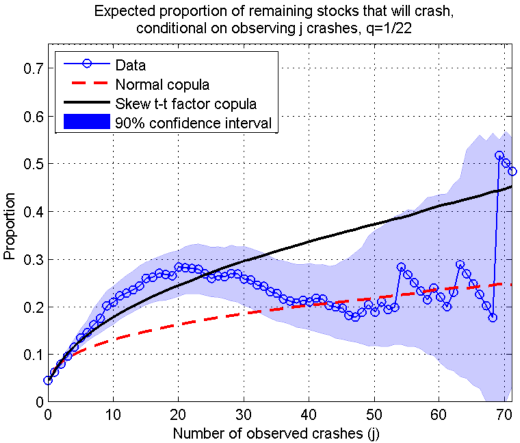
Parameter estimates for copula models

Block equi-dependence models - block dependence parameters

	Skew $t-t$					
SIC industry		Est	Std Err		Est	Std Err
Food, Apparel	β_2	0.865	0.040	γ_2	0.266	0.040
Transportation	β_4	0.916	0.041	γ_4	0.262	0.059
Trade	β_5	0.926	0.056	γ_5	0.554	0.047
Elec, machinery	β_3	1.017	0.037	γ_3	0.229	0.058
Finance, Ins.	β_6	1.067	0.046	γ_6	0.567	0.037
Services	β_7	1.129	0.063	γ_7	0.341	0.088
Mining	β_1	1.290	0.085	γ_1	1.022	0.064

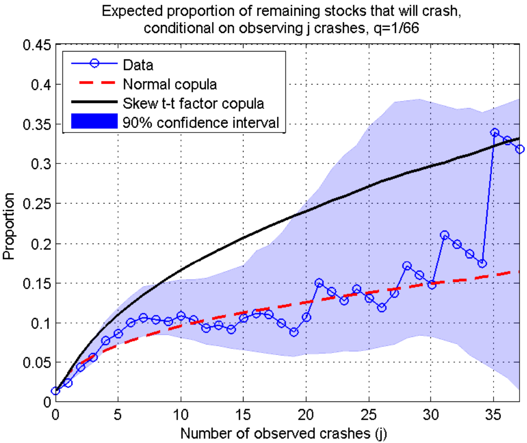
Sample and fitted multiple-stock crash probabilities

Crash quantile = 1/22



Sample and fitted multiple-stock crash probabilities

Crash quantile = 1/66



Rank correlation matrix implied by the two-factor structure

Correlation between stocks within / across industries

SIC industry	Mining	Food	Elec.	Trnsprt	Trade	Fin.	Serv.
Mining	0.72						
Food, Apparel	0.41	0.44					
Elec, Mach.	0.44	0.45	0.51				
Transport	0.41	0.42	0.45	0.46			
Trade	0.39	0.40	0.44	0.41	0.53		
Finance, Ins.	0.42	0.43	0.47	0.43	0.42	0.58	
Servies	0.45	0.46	0.50	0.46	0.44	0.47	0.57

- Pair-wise rank correlation ranges from 0.39 to 0.72.

Lower tail dependence matrix

Lower tail dep between stocks within / across industries

SIC industry	Mining	Food	Elec.	Trnsprt	Trade	Fin.	Serv.
Mining	0.99						
Food, Apparel	0.70	0.70					
Elec, Mach.	0.92	0.70	0.92				
Transport	0.75	0.70	0.75	0.75			
Trade	0.81	0.70	0.81	0.75	0.81		
Finance, Ins.	0.94	0.70	0.92	0.75	0.81	0.94	
Servies	0.96	0.70	0.92	0.75	0.81	0.94	0.96

- Lower tail dependence ranges from 0.70 to 0.99.

Upper tail dependence matrix

Upper tail dep between stocks within / across industries

SIC industry	Mining	Food	Elec.	Trnsprt	Trade	Fin.	Serv.
Mining	0.74	0.02	0.07	0.02	0.03	0.09	0.13
Food, Apparel		0.02	0.02	0.02	0.02	0.02	0.02
Elec, Mach.			0.07	0.02	0.03	0.07	0.07
Transport				0.02	0.02	0.02	0.02
Trade					0.03	0.03	0.03
Finance, Ins.						0.09	0.09
Servies							0.14

- Upper tail dependence ranges from 0.02 to 0.74.

Testing restrictions on the two-factor copula model

- The two factor model:

$$X_i = \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \quad i = 1, 2, ..., n$$

Testing restrictions on the two-factor copula model

- The two factor model:

$$X_i = \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \quad i = 1, 2, ..., n$$

1. Test that only common factor is needed $\Rightarrow \gamma_i = 0 \; \forall \; i$
   p-value=0.000

Testing restrictions on the two-factor copula model

- The two factor model:

$$X_i = \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

1. Test that only common factor is needed $\Rightarrow \gamma_i = 0 \ \forall \ i$
   p-value=0.000

2. Test that only industry factors are needed $\Rightarrow \beta_i = 0 \ \forall \ i$
   p-value=0.000

Testing restrictions on the two-factor copula model

- The two factor model:

$$X_i = \beta_i Z_0 + \gamma_i Z_{S(i)} + \varepsilon_i, \quad i = 1, 2, ..., n$$

1. Test that only common factor is needed $\Rightarrow \gamma_i = 0 \; \forall \; i$
   p-value=0.000
2. Test that only industry factors are needed $\Rightarrow \beta_i = 0 \; \forall \; i$
   p-value=0.000
3. Test of equidependence assumption $\Rightarrow \beta_i = \beta_j \cap \gamma_i = \gamma_j \; \forall i, j$
   p-value=0.000

Measuring systemic risk va "expected shortfall"

- Brownlees and Engle (2011, wp) propose a measure of systemic risk they call "marginal expected shortfall" (MES):

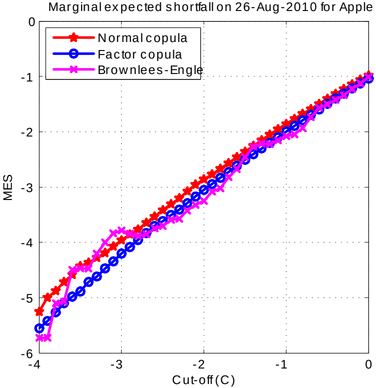
$$MES_{it} \equiv E_{t-1}\left[r_{it} | r_{mt} < C\right]$$

- That is, it is the expected return on stock i, conditional on the market return being below some threshold C (eg, $C = -2\%$).
  - Brownless and Engle also provide a way to estimate this using a bivariate GARCH model and nonparametric estimation of the tail.
- With a high dimension copula model one can recover the MES measure, and also alternative measures, such as:

$$kES_{it} \equiv E_{t-1}\left[r_{it} \,\middle|\, \left(\sum_{j=1}^{n} \mathbf{1}\left\{r_{jt} < C\right\}\right) > k\right]$$

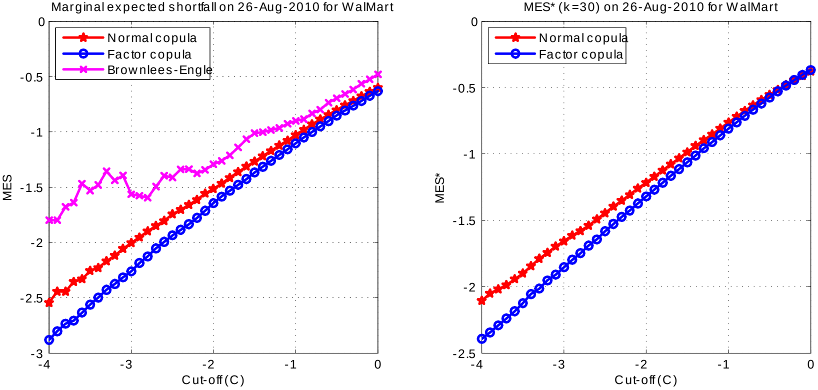
Marginal expected shortfall as a function of the threshold

MES estimates for Apple are similar across all models



Marginal expected shortfall as a function of the threshold

MES estimates for Walmart have sizeable differences



Evaluating MES forecasts

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

$$
\begin{aligned}
MSE_i &= \frac{1}{T}\sum_{t=1}^{T}\left(r_{it} - \widehat{MES}_{it}\right)^2 \mathbf{1}\left\{r_{mt} < C\right\} \\
RelativeMSE_i &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_{it} - \widehat{MES}_{it}}{\widehat{MES}_{it}}\right)^2 \mathbf{1}\left\{r_{mt} < C\right\}
\end{aligned}
$$

- Corresponding metrics can also be constructed for kES

Performance of methods for predicting MES

Brownlees-Engle does best under MSE; Factor copula best under RelMSE

	MSE		**RelMSE**	
Cut-off	*-2%*	*-4%*	*-2%*	*-4%*
Gaussian copula	1.0096	1.2521	0.6712	0.3420
Factor copula	1.0012	1.2445	0.5885	0.2954
t copula	1.0118	1.2580	0.6660	0.3325
skew t copula	1.0051	1.2553	0.6030	0.3040
BE	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

Performance of methods for predicting kES (k=30)

MES estimates from the proposed factor copula perform best

	MSE		RelMSE	
Cut-off	*-2%*	*-4%*	*-2%*	*-4%*
Gaussian copula	1.0885	1.4855	1.3220	0.5994
Factor copula	1.0822	1.4850	1.1922	0.5204
t copula	1.0956	1.4921	1.4496	0.6372
skew t copula	1.0898	1.4923	1.3370	0.5706
Historical	1.1632	1.6258	1.4467	0.7653

Summary and conclusion

- We present a simple and flexible class of **factor copula models** that may be applied in high dimensions.
  - Analytical results on tail dependence available using EVT
  - Estimation using SMM has good properties in finite samples
- We applied the new copulas to a collection of 100 daily equity returns
  - Among the **highest dimension** copula application to date
  - Significant evidence of **tail dependence**, **asymmetric dependence**, and **heterogeneous dependence**
  - Improved estimates of measures of **systematic risk**

Simulation study

- We use simulations to study the finite-sample properties of our estimator. We consider a variety of scenarios:

1. **Factor copula**
    1. Equi-dependence or heterogeneous dependence
    2. Normal, Student's $t\,(4)$, or skewed $t\,(4, -0.5)$ common factor

    - All copulas generate correlation of 0.5
2. **Dimension:** $N = 3$, 10 and 100
3. **Marginal distributions:** *iid* or AR(1)-GARCH(1,1) for conditional mean, conditional variance

- **Moments:** rank correlation, and quantile dependence for $q \in \{0.05, 0.10, 0.90, 0.95\}$
- Sample size: $T = 1000$, $S = 25 \times T$, replications$= 100$.

Factor copula with Normal common factor

SMM is only slightly less efficient than MLE, about same as GMM

$$
\begin{aligned}
X_i &= Z + \varepsilon_i \\
Z &\sim N(0,1) \\
\varepsilon_i &\sim iid\ \ N(0,1)\ \text{ and } Z \perp \varepsilon_i
\end{aligned}
$$

	Bias			**Std dev**		
# of variables	3	10	100	3	10	100
MLE $\hat{\sigma}_z^2$	0.017	0.015	0.018	0.081	0.057	0.051
GMM $\hat{\sigma}_z^2$	-0.012	-0.004	-0.003	0.100	0.067	0.057
SMM $\hat{\sigma}_z^2$	-0.013	-0.006	-0.005	0.103	0.069	0.056

iid vs AR-GARCH marginal dynamics: Factor copula with Normal common factor

AR-GARCH estimation error does not affect copula parameter estimate

	Bias			**Std dev**		
# of variables	3	10	100	3	10	100
MLE *iid*	0.017	0.015	0.018	0.081	0.057	0.051
MLE *GARCH*	0.014	0.011	0.017	0.080	0.056	0.050
SMM *iid*	-0.013	-0.006	-0.005	0.103	0.069	0.056
SMM *GARCH*	-0.016	-0.012	-0.008	0.103	0.067	0.055

iid vs AR-GARCH marginal dynamics: Factor copula with skewed t common factor

AR-GARCH estimation error does not affect copula parameter estimate

		Bias			Std dev		
# of variables		3	10	100	3	10	100
$\hat{\sigma}_z^2$	*iid*	0.079	0.056	0.042	0.317	0.197	0.160
$\hat{\sigma}_z^2$	*GARCH*	0.061	0.040	0.019	0.307	0.180	0.141
$\hat{\nu}^{-1}$	*iid*	-0.004	0.001	-0.001	0.068	0.049	0.041
$\hat{\nu}^{-1}$	*GARCH*	-0.008	-0.003	-0.005	0.068	0.049	0.038
$\hat{\lambda}$	*iid*	-0.019	-0.005	-0.003	0.122	0.066	0.053
$\hat{\lambda}$	*GARCH*	-0.020	-0.006	-0.000	0.121	0.066	0.054

Coverage probabilities: N=3

95% confidence intervals, AR-GARCH data, for different step sizes

	Normal	Factor $t - t$		Factor *Skew* $t - t$		
ε_T	σ_z^2	σ_z^2	ν_z^{-1}	σ_z^2	ν_z^{-1}	λ_z
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

Coverage probabilities: N=10

95% confidence intervals, AR-GARCH data, for different step sizes

	Normal	Factor $t-t$		Factor *Skew* $t-t$		
ε_T	σ_z^2	σ_z^2	ν_z^{-1}	σ_z^2	ν_z^{-1}	λ_z
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

Coverage probabilities: N=100

95% confidence intervals, AR-GARCH data, for different step sizes

	Normal	Factor $t-t$		Factor *Skew* $t-t$		
ε_T	σ_z^2	σ_z^2	ν_z^{-1}	σ_z^2	ν_z^{-1}	λ_z
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

J-test rejection frequencies

AR-GARCH data, step size of 0.1, 0.05 level

	iid data			AR-GARCH data		
	Norm	*t-t*	*Skew t-t*	Norm	*t-t*	*Skew t-t*
N=3	0.03	0.03	0.01	0.05	0.03	0.03
N=10	0.03	0.03	0.02	0.02	0.05	0.02
N=100	0.03	0.05	0.01	0.05	0.05	0.01