Dynamic Factor Copula Models with Estimated Cluster Assignments

Dong Hwan Oh
Federal Reserve Board

Andrew Patton
Duke University

Summer 2021

Motivation

- A small but growing literature considers methods for studying risk and dependence for large collections $(N \geq 100)$ of assets
  - **Covariances**: Fan *et al.* (2008, *JoE*), Tao *et al.* (2011, *JASA*), Hautsch *et al.* (2012, *JAE*), Engle *et al.* (2019, *JBES*) and many more.
  - **Copulas**: Creal and Tsay (2015, *JoE*), Christoffersen *et al.* (2018, *RoF*), Oh and Patton (2018, JBES), Opschoor *et al.* (2020, *JBES*).
- Applications in risk management, asset allocation, stress testing.
- Such models inevitably have to impose some structure to overcome the curse of dimensionality
  - A common assumption is some sort of factor or group structure.

★ How should we group the assets?
  - SIC codes are commonly used, but are they optimal?

What we do

- We propose a dynamic factor copula for high-dimensional applications, where group assignments are estimated from the data.
  - **Dynamics** captured via a GAS model (Creal *et al.*, 2013 *JAE*; Harvey, 2013, *book*).
  - **Shape** modeled using skew t copula, nesting *t* and Gaussian.
  - ★ **Clustering** done via an EM-type algorithm.
- We show that group assignments can be consistently estimated using a simpler *misspecified model*, reducing the computational burden.
- We find empirically that a model with estimated group assignments significantly outperforms one using SIC-based groups.
  - Opschoor *et al.* (2020, *JBES*) show that SIC-based groups are better than groups based on size, value, etc.

Outline

1. Motivation
2. A dynamic skewed t factor model
3. Estimating cluster assignments
4. Simulation study
5. Application to 110 US equities

Outline

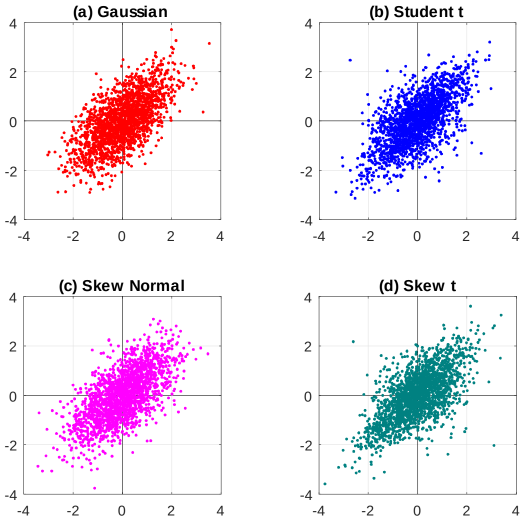
1. Motivation
2. **A dynamic skewed t factor model**
3. Estimating cluster assignments
4. Simulation study
5. Application to 110 US equities

A skewed t factor copula model

- We consider an extension of the model considered in Creal and Tsay (2015, *JoE*) and Opschoor *et al.* (2020, *JBES*) to allow for asymmetric dependence, namely a **skewed t factor copula**.
  - As expected, it nests the Student's t and Gaussian copulas.
- Unlike the model of Oh and Patton (2018, *JBES*) this copula has a (quasi) closed-form log-likelihood and score function.
  - It contains a univariate CDF that must be computed numerically. This CDF only has two free parameters, and so it is not a burden.

Draws from bivariate distributions with different copulas

All with correlation of 0.5 and all with N(0,1) marginal distributions



Factor structure implies a block correlation matrix

- The copula is obtained from the joint distribution of **X**, where

$$X_i = \left(\sum_{g=0}^{G} \lambda_{i,g} Z_g + \sigma_i \epsilon_i \right) \sqrt{W} + \zeta W, \quad i = 1, ..., N$$

where Z and ϵ are standard Normal, and W is inverse Gamma.

- We assume each variable X_i loads on two factors:
a common market factor, Z_0, and a group-specific factor, Z_g.

- As usual for a factor model, the structure of this model implies that a key correlation matrix, $\mathbf{R}_t$, in the log-likelihood satisfies:

$$\mathbf{R}_t = \mathbf{L}_t' \mathbf{L}_t + \mathbf{D}_t$$

  - Creal and Tsay (2015, *JoE*) discuss how to exploit this structure.

GAS dynamics for the factor loadings

- We capture time-varying dependence by allowing the factor loadings to evolve according to a "**generalized autoregressive score**" (GAS) model (Creal *et al*., 2013, *JAE*; Harvey, 2013, book).

- We adopt the most flexible specification considered in Opschoor *et al.* (2020, *JBES*). For each $g = 1, ..., G$:

$$\lambda_{M,g,t+1} = \omega_g^M + \beta^M \lambda_{M,g,t} + \alpha^M \frac{\partial \log \mathbf{c}_t \left(\mathbf{x}_t ; \boldsymbol{\theta} \right)}{\partial \lambda_{M,g}}$$

$$\lambda_{g,g,t+1} = \omega_g^C + \beta^C \lambda_{g,g,t} + \alpha^C \frac{\partial \log \mathbf{c}_t \left(\mathbf{x}_t ; \boldsymbol{\theta} \right)}{\partial \lambda_{g,g}}$$

Outline

1. Motivation
2. A dynamic skewed t factor model
3. **Estimating cluster assignments**
4. Simulation study
5. Application to 110 US equities

Clustering on a misspecified model I

- Our target model is a skew t factor copula with GAS dynamics and estimated cluster assignments.
  - That is a complicated model, and is hard to estimate.
- We exploit two key features of this model to dramatically speed up estimation:

Clustering on a misspecified model II

1. The block structure of the correlation matrix $\mathbf{R}$ in the factor copula likelihood holds regardless of the skew t shape parameters

   - Can cluster using a Gaussian copula rather than a skew t copula.

Clustering on a misspecified model II

1. The block structure of the correlation matrix $\mathbf{R}$ in the factor copula likelihood holds regardless of the skew t shape parameters
   - Can cluster using a Gaussian copula rather than a skew t copula.
2. The block structure of the *conditional* correlation matrix $\mathbf{R}_t$ also, generally, manifests in the unconditional correlation matrix $\overline{\mathbf{R}}$
   - Can cluster using a static copula rather than a dynamic copula
   - The time series averaging in $\overline{\mathbf{R}}$ may make the clusters harder to detect than in $\mathbf{R}_t$

Clustering on a misspecified model II

1. The block structure of the correlation matrix $\mathbf{R}$ in the factor copula likelihood holds regardless of the skew t shape parameters
   - Can cluster using a Gaussian copula rather than a skew t copula.
2. The block structure of the *conditional* correlation matrix $\mathbf{R}_t$ also, generally, manifests in the unconditional correlation matrix $\overline{\mathbf{R}}$
   - Can cluster using a static copula rather than a dynamic copula
   - The time series averaging in $\overline{\mathbf{R}}$ may make the clusters harder to detect than in $\mathbf{R}_t$

- Given the clusters obtained using the static Gaussian copula, we estimate the more flexible dynamic skew t copula.

Estimation of cluster assignments I

- Let $\Gamma = [\gamma_1, ..., \gamma_N]$ where $\gamma_i \in \{1, ..., G\}$ for $i = 1, ..., N$, denote the vector of cluster assignments
- Let $\boldsymbol{\theta}$ be the vector of static Gaussian factor copula parameters.

1. Given an estimate of the cluster assignment vector, $\hat{\Gamma}^{(s)}$ the copula log-likelihood is maximized over $\boldsymbol{\theta}$ to yield:

$$\hat{\boldsymbol{\theta}}^{(s+1)} = \arg\max_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} \log \mathbf{c}\left(\mathbf{u}_t; \boldsymbol{\theta}, \hat{\Gamma}^{(s)}\right)$$

   - This is a standard numerical optimization step.

Estimation of cluster assignments II

2. Given copula parameter $\hat{\boldsymbol{\theta}}^{(s+1)}$, the log-likelihood is maximized over cluster assignments γ_i for $i = 1, ..., N$:

$$\hat{\gamma}_i^{(s+1)} = \arg \max_{g \in \{1,...,G\}} \frac{1}{T} \sum_{t=1}^{T} \log \mathbf{c} \left(\mathbf{u}_t ; \hat{\boldsymbol{\theta}}^{(s+1)}, \tilde{\Gamma}_i^{(s)} (g) \right)$$

where $\tilde{\Gamma}_i^{(s)}(g)$ equals $\hat{\Gamma}^{(s)}$ except that the i^{th} element is set to g.

- This step is very fast, only requiring $G \times N$ likelihood evaluations.

- The iteration between steps (1)–(2) continues until convergence.
- Convergence to a local optimum is guaranteed, and we use 100 randomly-chosen starting values to improve accuracy.

Consistency of estimated cluster assignments

- Under standard regularity conditions
  1. Stationarity
  2. LLNs holding for log-likelihoods and scores
  3. Clusters being "well separated"
- We obtain the following quasi MLE-type result:

$$\hat{\Gamma}_T \xrightarrow{p} \Gamma_0 \quad \text{as} \quad T \rightarrow \infty$$

Super-consistency of estimated cluster assignments?

- Results from related contexts suggest that if our GAS process satisfies certain mixing properties, a *large deviations principle* may be applied
    - Hahn and Moon (2010, *ET*), Bonhomme and Manresa (2015, ECMA).

- This would enable obtaining a rate result, usually of the form:

$$\Pr\left[\hat{\Gamma}_T \neq \Gamma_0\right] \leq C_1 \exp\left\{-C_2 T^{\kappa}\right\}$$

for some constants $C_1, C_2, \kappa > 0$.

    - Estimation error in estimated cluster assignments vanishes much faster than the usual $\sqrt{T}$ rate.

- Unfortunately, general results on the mixing properties of GAS processes are not yet available, and we do not pursue this result.

Outline

1. Motivation
2. A dynamic skewed t factor model
3. Estimating cluster assignments
4. **Simulation study**
5. Application to 110 US equities

Simulation design

- We consider a variety of factor copulas (Gaussian, t and skew t), with or without GAS dynamics.
    - Parameters chosen to approximately match what we find empirically.
- We set $N = 100$, $T \in \{1000, 2000\}$, $G \in \{10, 20\}$.
- We use 100 random starting values for the clustering step, and repeat each simulation (only) 100 times.

Simulation results: Cluster assignment accuracy

What's the best way to show 100% accuracy in a table??

$T = 1000,\ G = 10$	**Gaussian**	**t**	**skew t**
Group assignment estimation accuracy			
Number incorrect			
0	100	100	100
1	0	0	0
≥ 2	0	0	0
Estimation details			
Time (hours)	0.56	0.84	0.84
EM iterations	3.45	4.31	4.23

Simulation results: Cluster assignment accuracy

What's the best way to show 100% accuracy in a table??

$T = 2000,\ G = 20$	**Gaussian**	**t**	**skew t**
Group assignment estimation accuracy			
Number incorrect			
0	100	100	100
1	0	0	0
$\geq$2	0	0	0
Estimation details			
Time (hours)	1.44	2.10	2.78
EM iterations	6.94	4.49	3.90

Simulation results: Cluster assignment accuracy

Accuracy is no longer perfect when G is increased

$T = 1000,\ G = 20$	**Gaussian**	**t**	**skew t**
Group assignment estimation accuracy			
Number incorrect			
0	100	95	98
1	0	5	2
≥ 2	0	0	0
Estimation details			
Time (hours)	0.72	0.95	1.43
EM iterations	3.45	4.31	4.23

Outline

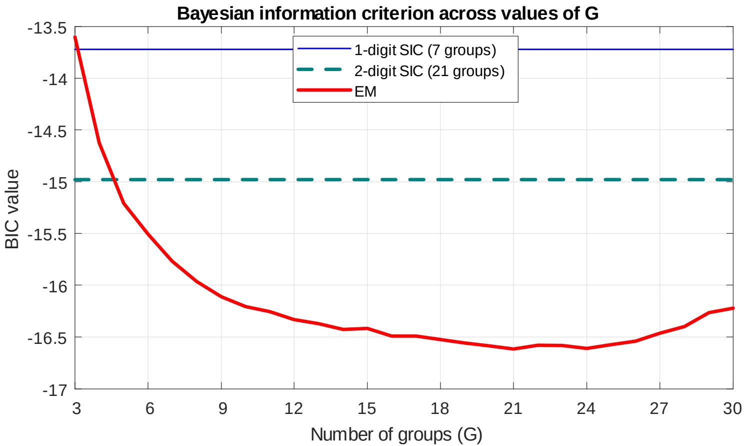
1. Motivation
2. A dynamic skewed t factor model
3. Estimating cluster assignments
4. Simulation study
5. **Application to 110 US equities**

Data and marginal models

- Every stock that was ever a constituent of the S&P 100 index during this sample, and which traded for the full sample period.
  - $N = 110$
- Daily returns from January 4, 2010 to December 31, 2019.
  - $T = 2516$
- We use an AR(1)–GJR-GARCH(1,1) model with skew t innovations to model the marginal distributions.
  - DoF parameter is around 5, skewness is slightly negative

Finding the optimal number of groups

BIC is minimized at G=21. Beats SIC for G>=5.



Estimated cluster assignments: Largest 2 groups

Many estimated clusters appear to be SIC groups with a twist

Ticker	Name	SIC	Ticker	Name	SIC
ABT	Abbott Lab.	28	BAC	Bank of America	60
AGN	Actavis	28	BK	Bank Of NY	60
AMGN	Amgen	28	C	Citigroup Inc	60
BAX	Baxter	**38**	COF	Capital One	60
BIIB	Biogen	28	GS	Goldman Sachs	**62**
BMY	Bristol-Myers	28	JPM	JP Morgan	60
GILD	Gilead	28	MET	Metlife	**63**
JNJ	Johnson & Johnson	28	MS	Morgan Stanley	60
LLY	Lilly Eli	28	RF	Regions Fin	60
MDT	Medtronic	**38**	USB	U S Bancorp	60
MRK	Merck	28	WFC	Wells Fargo	60
PFE	Pfizer	28			
UNH	Unitedhealth	**63**			

Estimated cluster assignments: Small groups

Some smaller groups reveal some obvious pairs

Ticker	Name	SIC	Ticker	Name	SIC
HD	Home Depot	52	T	AT&T	48
LOW	Lowes	52	VZ	Verizon	48
MCD	Mcdonalds	58	MA	Mastercard	**73**
SBUX	Starbucks	58	V	Visa	**61**

Estimated cluster assignments: Tech clusters

Estimated assignments correct (?) some out-dated SIC codes

Ticker	Name	SIC	Ticker	Name	SIC
AAPL	Apple	**35**	CSCO	Cisco Sys	36
ADBE	Adobe	73	HPQ	Hewlett Pac	**35**
AMZN	Amazon	73	INTC	Intel	36
CRM	Salesforce	73	MSFT	Microsoft	**73**
EBAY	Ebay	73	NVDA	Nvidia	36
GOOGL	Google	73	QCOM	Qualcomm	36
NFLX	Netflix	**78**	TXN	Texas Instru	36
PCLN	Priceline	73			

Out-of-sample model comparisons

- To formally compare the various models we consider, we turn to out-of-sample forecast performance.
  - Economically interesting in its own right
  - Econometrically helpful, because estimated group assignments have estimation error that is hard to handle theoretically
- We split our sample in half:
  - First half: Cluster assignments and copula parameters estimated
  - Second half: Models evaluated using out-of-sample log-likelihood and Diebold-Mariano (1995, *JBES*) tests.

Dynamic vs. Static

Positive t-stat indicates GAS is preferred to Static: GAS preferred everywhere

t-statistics	Static vs. GAS		
	Gaussian	*t*	skew *t*
SIC 1 digit	7.86	12.07	11.55
SIC 2 digit	9.89	15.73	16.50
3 groups	6.53	6.64	6.76
5 groups	7.55	10.80	10.91
20 groups	10.91	16.19	14.74
25 groups	11.82	19.00	19.14
30 groups	10.92	17.17	15.92

Gaussian vs. t vs. skew t

Positive t-stat indicates latter model preferred to former: Symmetric t is preferred

t-statistics	Copula shape		
	G vs. *t*	G vs. skew *t*	*t* vs. skew *t*
SIC 1 digit	9.29	8.60	-2.98
SIC 2 digit	8.94	8.47	-2.85
3 groups	8.34	7.71	-2.26
5 groups	9.24	9.09	-1.72
20 groups	9.43	8.00	-3.80
25 groups	9.48	8.57	-2.36
30 groups	9.70	8.72	-2.96

Choice of the number of clusters

Positive t-statistic indicates column model beats row model

t-statistics	SIC-1	SIC-2	3	5	20	30
SIC-1		26.17	-4.78	21.90	33.13	29.98
SIC-2	-26.17		-21.58	2.53	23.29	17.08
3 groups	4.78	21.580		24.18	30.93	28.72
5 groups	-21.90	-2.53	-24.18		22.54	15.19
18 groups	-33.50	-23.19	-31.40	-23.18	2.38	-12.35
19 groups	-32.07	-21.20	-30.85	-20.56	7.60	-9.83
20 groups	-33.13	-23.29	-30.93	-22.54		-13.06
21 groups	-33.29	-23.36	-31.25	-22.69	2.92	-12.80
22 groups	-32.97	-21.26	-31.53	-20.07	7.89	-7.74
30 groups	-29.98	-17.08	-28.72	-15.19	13.06	
$\log \mathcal{L}$	34,074	37,887	33,175	38,304	42,146	40,559

Choice of the number of clusters

20 estimated groups significantly beats all alternatives

t-statistics	SIC-1	SIC-2	3	5	20	30
SIC-1		26.17	-4.78	21.90	33.13	29.98
SIC-2	-26.17		-21.58	2.53	23.29	17.08
3 groups	4.78	21.58		24.18	30.93	28.72
5 groups	-21.90	-2.53	-24.18		22.54	15.19
18 groups	-33.50	-23.19	-31.40	-23.18	2.38	-12.35
19 groups	-32.07	-21.20	-30.85	-20.56	7.60	-9.83
20 groups	-33.13	-23.29	-30.93	-22.54		-13.06
21 groups	-33.29	-23.36	-31.25	-22.69	2.92	-12.80
22 groups	-32.97	-21.26	-31.53	-20.07	7.89	-7.74
30 groups	-29.98	-17.08	-28.72	-15.19	13.06	
$\log \mathcal{L}$	34,074	37,887	33,175	38,304	**42,146**	40,559

Choice of the number of clusters

5 estimated groups signif beats SIC-based models with 7 or 21 groups.

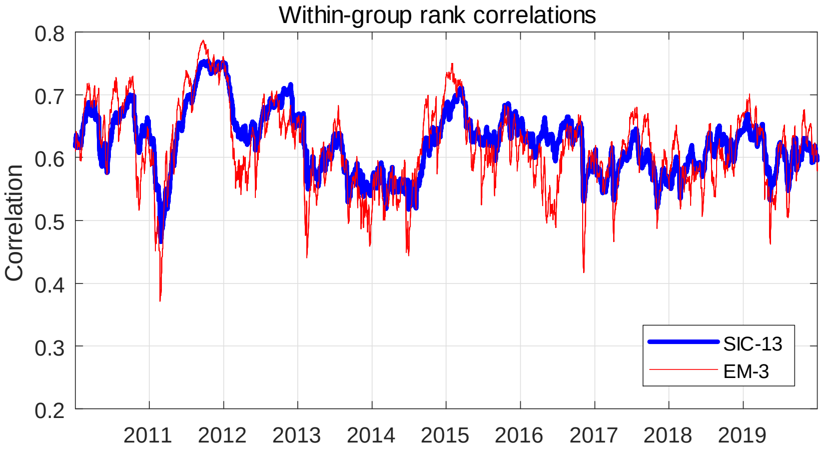
t-statistics	SIC-1	SIC-2	3	**5**	20	30
SIC-1		26.17	-4.78	**21.90**	33.13	29.98
SIC-2	-26.17		-21.58	**2.53**	23.29	17.08
3 groups	4.78	21.58		24.18	30.93	28.72
5 groups	-21.90	-2.53	-24.18		22.54	15.19
18 groups	-33.50	-23.19	-31.40	-23.18	2.38	-12.35
19 groups	-32.07	-21.20	-30.85	-20.56	7.60	-9.83
20 groups	-33.13	-23.29	-30.93	-22.54		-13.06
21 groups	-33.29	-23.36	-31.25	-22.69	2.92	-12.80
22 groups	-32.97	-21.26	-31.53	-20.07	7.89	-7.74
30 groups	-29.98	-17.08	-28.72	-15.19	13.06	
$\log \mathcal{L}$	34,074	37,887	33,175	**38,304**	42,146	40,559

Summary of empirical results

- The above results reveal that a model with as few as 5 estimated groups significantly out-performs an **otherwise identical** model using 21 groups based on SIC codes.
  - A model with 20 (or 21) estimated groups out-performs the SIC-based model by even more.
- We now try to shed more light on *where* the improvement in forecast performance is coming from.

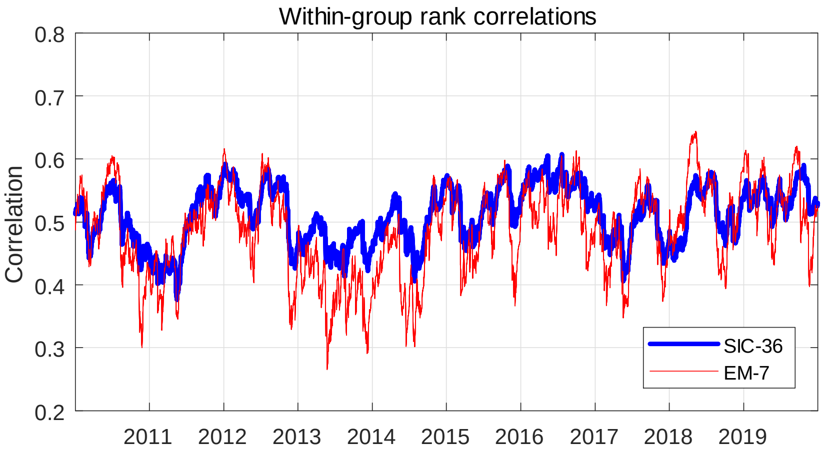
Within-group rank correlations: EM 3 & SIC 13

Correlations appear more "dynamic" for the EM model



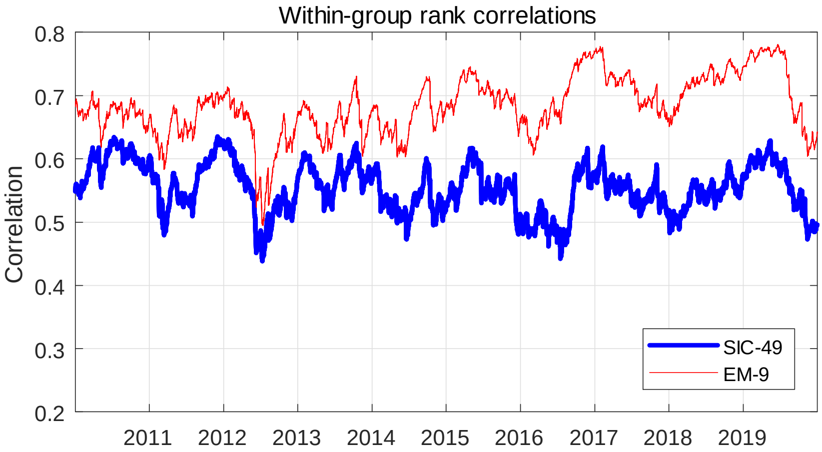
Within-group rank correlations: EM 7 & SIC 36

Correlations appear more "dynamic" for the EM model



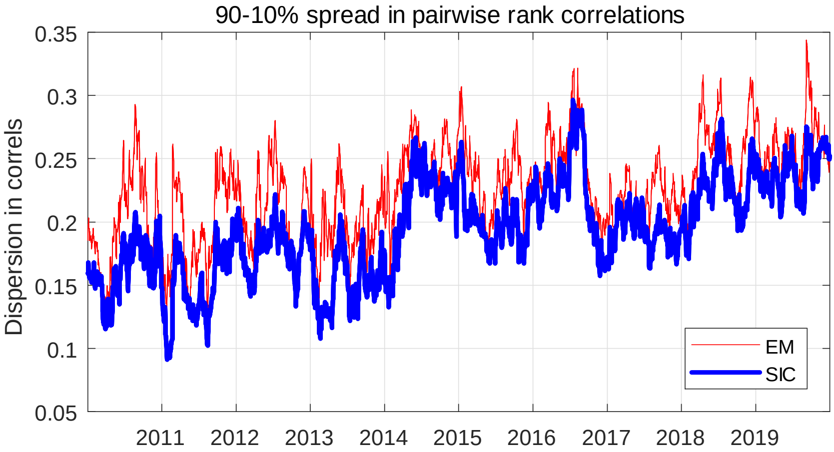
Within-group rank correlations: EM 9 & SIC 49

"EM 9" contains all SIC-49 firms but one; moving it increases within-group correlation



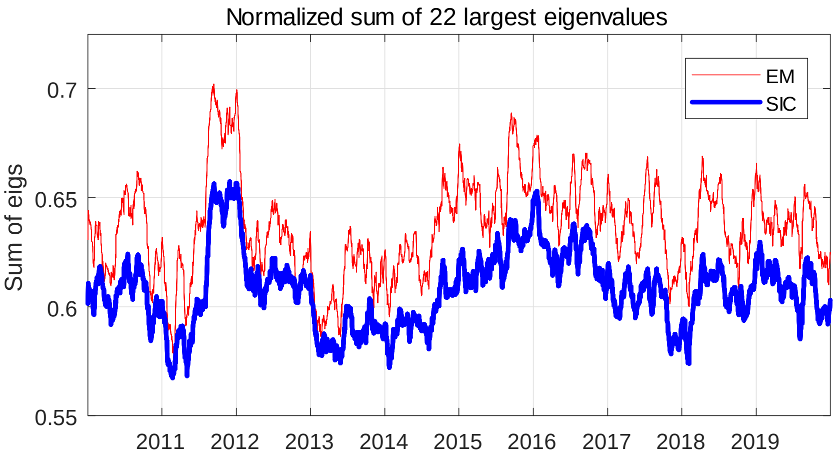
Spread in pairwise rank correlations

Greater cross-sectional spread in pairwise correlations from the EM model



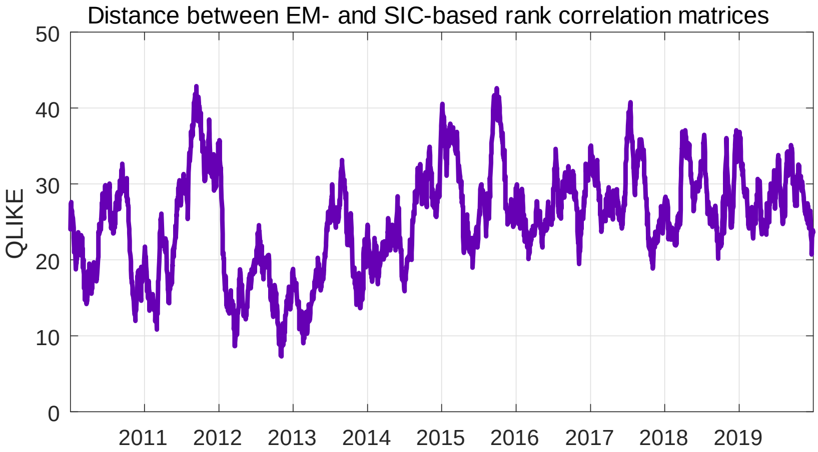
Normalized sum of first 22 eigenvalues

Explanatory power of factors is greater for model with estimated group assignments



Distance between model-implied rank correlation matrices

Distance is related to explanatory power of the factors in the model (see eig. plot)



Summary

- We propose a high-dimensional dynamic copula model based on a cluster structure, where **cluster assignments are estimated**.
  - Estimated assignments are close to SIC groupings, but with meaningful differences. Eg: Apple *used* to be in manufacturing, now it's tech.
- We show that these assignments can be estimated from a simpler **misspecified model** making the method computationally feasible.
  - Clustering on a misspecified model does *not* generally work; but we show that it does here.
- We show that estimating cluster assignments leads to dramatically **better forecast performance** than using SIC-based groups.
  - A model with as few as 5 estimated clusters outperforms a model with 21 clusters formed using two-digit SIC codes.

Appendix

Simulation results: Parameter estimation accuracy

$T = 1000,\ G = 10$		Gaussian		skew t	
	True	Mean	Std Dev	Mean	Std Dev
ω_1^M	0.04	0.042	0.007	0.044	0.008
ω_5^M	0.04	0.042	0.007	0.043	0.007
ω_{10}^M	0.04	0.042	0.007	0.044	0.008
ω_1^C	0.04	0.043	0.007	0.042	0.007
ω_5^C	0.04	0.043	0.007	0.041	0.007
ω_{10}^C	0.04	0.043	0.007	0.041	0.007
α^M	0.02	0.020	0.002	0.020	0.002
β^M	0.90	0.894	0.015	0.893	0.017
α^C	0.02	0.020	0.002	0.020	0.002
β^C	0.90	0.896	0.016	0.898	0.014
ν	5.00			5.016	0.108
ζ	-0.10			-0.100	0.007