Supplemental Appendix for
"High-Dimensional Copula-Based Distributions with Mixed Frequency Data"

by Dong Hwan Oh and Andrew J. Patton

May 19, 2015

S.A.1: The Dynamic Conditional Correlation (DCC) model

The DCC model by Engle (2002) decomposes the conditional covariance matrix $\mathbf{H}_t$ as:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{2}$$

$$\text{where } \mathbf{D}_t = diag\left(\left\{\sqrt{\sigma_{i,t}^2}\right\}_{i=1}^N\right) \tag{3}$$

and then the conditional correlation matrix is assumed to follow:

$$\mathbf{R}_t = diag\left(\mathbf{Q}_t\right)^{-1/2} \mathbf{Q}_t diag\left(\mathbf{Q}_t\right)^{-1/2} \tag{4}$$

$$\text{where } \mathbf{Q}_t = \left(1-\alpha-\beta\right)\overline{\mathbf{Q}} + \alpha\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\right) + \beta\mathbf{Q}_{t-1} \tag{5}$$

$$\text{and } \boldsymbol{\varepsilon}_t = \mathbf{D}_t^{-1}\left(\mathbf{r}_t - \boldsymbol{\mu}_t\right) \tag{6}$$

and $\overline{\mathbf{Q}}$ is the sample correlation matrix of $\boldsymbol{\varepsilon}_t$.All that remains is to specify models for the individual conditional variances, and for those we assume the GJR-GARCH model of Glosten, *et al.* (1993):

$$\sigma_{i,t}^2 = \psi_i + \kappa_i\left(r_{i,t-1} - \mu_{i,t-1}\right)^2 + \zeta_i\left(r_{i,t-1} - \mu_{i,t-1}\right)^2 1_{\left\{\left(r_{i,t-1}-\mu_{i,t-1}\right)<0\right\}} + \lambda_i\sigma_{i,t-1}^2 \tag{7}$$

The total number of parameters to estimate in this model is $4N + N\left(N-1\right)/2 + 2$.

Engle (2002) suggests estimating the model above using Gaussian quasi-maximum likelihood, and we follow this for the volatility estimation stage. For the DCC estimation stage, Engle, *et al.* (2008) find that when N is large the estimates of α and β may be biased due to the impact of estimation error from estimating $\overline{\mathbf{Q}}$ and they suggest the composite likelihood based estimator based on bivariate likelihoods. We follow their suggestion and use composite likelihood for this stage in Section 4.2 and 5.

S.A.2: Additional material on jointly symmetric copulas

For added intuition, consider the bivariate case. Theorem 1 then shows that the jointly symmetric copula CDF is:

$$\begin{aligned}
\mathbf{C}^{JS}(u_1,u_2) &= \frac{1}{4}\sum_{k_1=0}^{2}\sum_{k_2=0}^{2}(-1)^R\cdot\mathbf{C}(\widetilde{u}_1,\widetilde{u}_2) \\
&= \frac{1}{4}[\mathbf{C}(1,1)+\mathbf{C}(1,u_2)-\mathbf{C}(1,1-u_2) \\
&\quad +\mathbf{C}(u_1,1)+\mathbf{C}(u_1,u_2)-\mathbf{C}(u_1,1-u_2) \\
&\quad -\mathbf{C}(1-u_1,1)-\mathbf{C}(1-u_1,u_2)+\mathbf{C}(1-u_1,1-u_2)] \\
&= \frac{1}{4}[\mathbf{C}(u_1,u_2)-\mathbf{C}(u_1,1-u_2)-\mathbf{C}(1-u_1,u_2)+\mathbf{C}(1-u_1,1-u_2)+2u_1+2u_2-1]
\end{aligned}$$

using the fact that $\mathbf{C}(1,1)=1$ and $\mathbf{C}(1,a)=\mathbf{C}(a,1)=a$. The PDF is simply:

$$\mathbf{c}^{JS}(u_1,u_2)=\frac{1}{4}\left[\mathbf{c}(u_1,u_2)+\mathbf{c}(1-u_1,u_2)+\mathbf{c}(u_1,1-u_2)+\mathbf{c}(1-u_1,1-u_2)\right]$$

The PDF has the nice feature that no copula marginals need to be handled, while the CDF requires keeping track of these, a task that gets more complicated in higher dimensions.

The CDF of a jointly symmetric copula constructed via rotations can also be expressed more compactly using the multinomial formula. (We thank Bruno Rémillard for suggesting the following.)

$$\text{Let}\quad (\mathbf{u}_{\mathcal{A}})_i=\begin{cases} u_i, & i\in\mathcal{A}^c \\ 1-u_i, & i\in\mathcal{A}\end{cases}\quad,\quad i=1,2,..,N$$

so $\mathcal{A}$ is the subset of the N variables that are rotated, and below we sum across all possible subsets of these, of which there are 2^N:

Then

$$\begin{aligned}
\mathbf{C}^{JS}(\mathbf{u}) &= \frac{1}{2^N}\sum_{\mathcal{A}\subseteq\{1,\dots,N\}}\Pr[\mathbf{U}_{\mathcal{A}}\le\mathbf{u}]=\frac{1}{2^N}\sum_{\mathcal{A}}E\left[\prod_{i\in\mathcal{A}^c}\mathbf{1}\{U_i\le u_i\}\prod_{i\in\mathcal{A}}(1-\mathbf{1}\{U_i\le 1-u_i\})\right] \\
&= \frac{1}{2^N}\sum_{\mathcal{A}}\sum_{\mathcal{B}\subseteq\mathcal{A}}(-1)^{|\mathcal{B}|}E\left[\prod_{i\in\mathcal{A}^c}\mathbf{1}\{U_i\le u_i\}\prod_{i\in\mathcal{B}}\mathbf{1}\{U_i\le 1-u_i\}\right] \\
&= \frac{1}{2^N}\sum_{\mathcal{A}}\sum_{\mathcal{B}\subseteq\mathcal{A}}(-1)^{|\mathcal{B}|}\mathbf{C}(\mathbf{u}_{\mathcal{B},\mathcal{A}})
\end{aligned}$$

$$\text{where}\quad(\mathbf{u}_{\mathcal{B},\mathcal{A}})_i = \begin{cases} 1-u_i, & i\in\mathcal{B} \\ u_i, & i\in\mathcal{A}^c \\ 1, & i\in\mathcal{A}\backslash\mathcal{B}\end{cases}\quad,\quad i=1,2,...,N.$$

S.A.3 Additional tables and figures

Table A1: 104 Stocks used in the empirical analysis

Ticker	Name	Ticker	Name	Ticker	Name
AA	Alcoa	EMR	Emerson Elec	NOV	National Oilwell
AAPL	Apple	ETR	Entergy	NSC	Norfolk South
ABT	Abbott Lab.	EXC	Exelon	NWSA	News Corp
AEP	American Elec	F	Ford	ORCL	Oracle
ALL	Allstate Corp	FCX	Freeport	OXY	Occidental Petrol
AMGN	Amgen Inc.	FDX	Fedex	PEP	Pepsi
AMZN	Amazon.com	GD	General Dynam	PFE	Pfizer
AVP	Avon	GE	General Elec	PG	Procter Gamble
APA	Apache	GILD	Gilead Science	QCOM	Qualcomm Inc
AXP	American Ex	GOOG	Google Inc	RF	Regions Fin
BA	Boeing	GS	Goldman 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 property
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	MorganStanley	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 the empirical analysis of this paper.

Table A2: Simulation results for a jointly symmetric copula based on the Gumbel copula

	Bias				Std dev				Average Run Time (in sec)			
N	MLE	$MCLE_{all}$	$MCLE_{adj}$	$MCLE_{first}$	MLE	$MCLE_{all}$	$MCLE_{adj}$	$MCLE_{first}$	MLE	$MCLE_{all}$	$MCLE_{adj}$	$MCLE_{first}$
2	-0.0016	-0.0016	-0.0016	-0.0016	0.0757	0.0757	0.0757	0.0757	0.30	0.13	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	0.13
5	-0.0041	-0.0025	-0.0025	-0.0016	0.0368	0.0409	0.0470	0.0757	3.52	1.31	0.53	0.13
10	-0.0021	-0.0023	-0.0016	-0.0016	0.0245	0.0328	0.0369	0.0757	153	6	1	0.13
20		-0.0019	-0.0021	-0.0016		0.0285	0.0312	0.0757	3×10^{5}	25	2	0.13
30		-0.0019	-0.0022	-0.0016		0.0277	0.0297	0.0757	5×10^{8}	61	4	0.13
40		-0.0019	-0.0022	-0.0016		0.0270	0.0285	0.0757	7×10^{11}	97	5	0.13
50		-0.0024	-0.0027	-0.0016		0.0269	0.0283	0.0757	7×10^{14}	166	7	0.13
60		-0.0021	-0.0023	-0.0016		0.0267	0.0282	0.0757	9×10^{17}	236	8	0.13
70		-0.0022	-0.0024	-0.0016		0.0264	0.0276	0.0757	1×10^{21}	326	9	0.13
80		-0.0022	-0.0023	-0.0016		0.0262	0.0272	0.0757	1×10^{24}	435	11	0.13
90		-0.0021	-0.0022	-0.0016		0.0262	0.0272	0.0757	1×10^{27}	509	11	0.13
100		-0.0020	-0.0021	-0.0016		0.0261	0.0272	0.0757	1×10^{30}	664	13	0.13

Note: This table presents the results from 500 simulations of jointly symmetric copula based on Gumbel copula with true parameter 2. The sample size is $T = 1000$. Four different estimation methods are used: MLE, MCLE with all pairs, MCLE with adjacent pairs, MCLE with the first pair. MLE is infeasible for $N > 10$ and so no results are reported in those cases. The first four columns report the average difference between the estimated parameter and its true value. The next four columns are the standard deviation in the estimated parameters. The last four columns present average run time of each estimation method. The reported run times for MLE for $N > 10$ are based on actual single function evaluation times and on an assumption of 40 function evaluations to reach the optimum.