Dynamic Covariance Models for Multivariate Financial Time Series

UNIVERSITY OF CAMBRIDGE

Yue Wu, José Miguel Hernández-Lobato, and Zoubin Ghahramani

Department of Engineering, Cambridge University

ICML | Atlanta
International Conference on Machine Learning

Dynamic Covariances for Financial Data

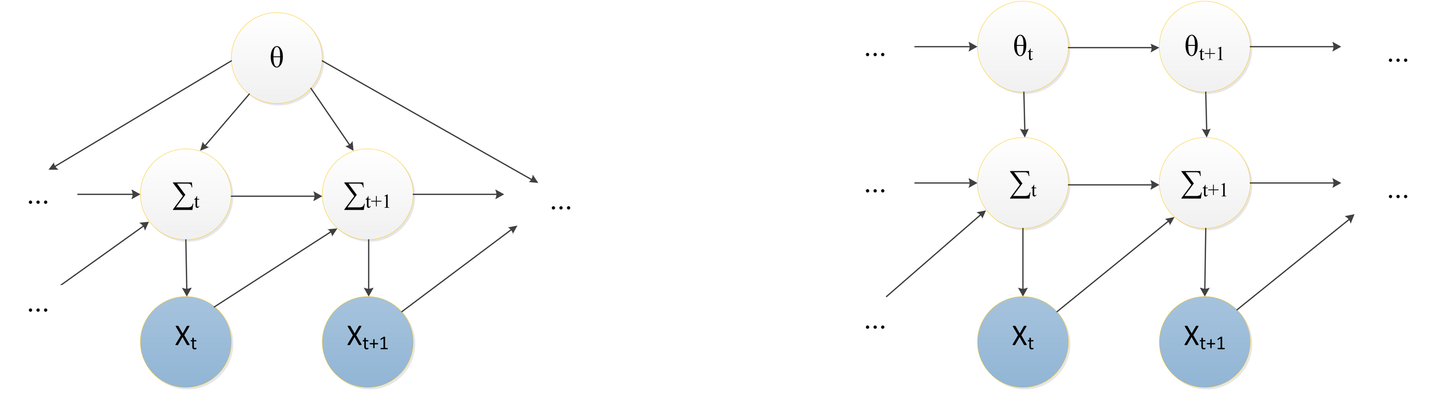
The accurate prediction of time-changing covariances is an important problem in the modeling of multivariate financial data. Standard models such as BEKK [Engle and Kroner 1995] suffer from:

1. Failure to capture shifts in market conditions.
2. Overfitting problems and multiple local optima from taking maximum likelihood estimates.
3. Large computational costs.

We introduce a Bayesian Multivariate Dynamic Covariance model (BMDC) that

1. Captures changes in market conditions with time-varying model parameters.
2. Avoids overfitting by using Bayesian inference.
3. Fast computation with particle filters.

Graphical Models for BMDC and BEKK



Left, graphical model for BEKK. The parameters $\theta = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ are static.
Right, graphical model for BMDC with time-varying parameters $\theta_t = (\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t)$.

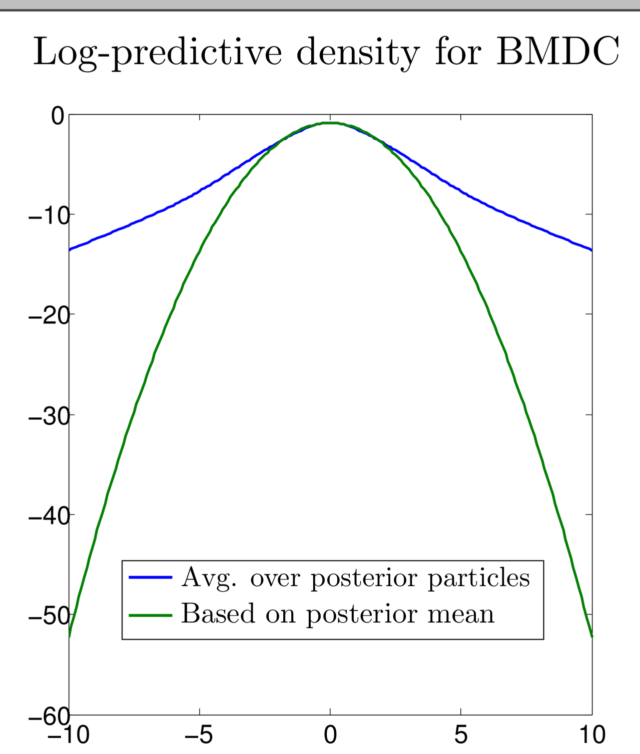
Experiments on Financial Data

Datasets:
We used 66 univariate and multivariate datasets:

1. 21 Daily FX datasets from Jan 2008-Jan 2011. 780 observations each. Dataset dimensions ranged from one to five.
2. 15 Intraday FX datasets from 2008. Five $1D$, five $2D$ and five $3D$ datasets, each with 5000 observations at 5 minute intervals.
3. 30 Daily Equity datasets from Jan 2000-Dec 2011, corresponding to 3000 observations. Dataset dimensions ranged from $1D$ to $3D$.

Experiments:
We compared the average predictive log-likelihood and execution times for BEKK, BMDC and their heavy-tailed variants, BEKK-T and BMDC-T.

BMDC Posterior Predictive is heavy-tailed

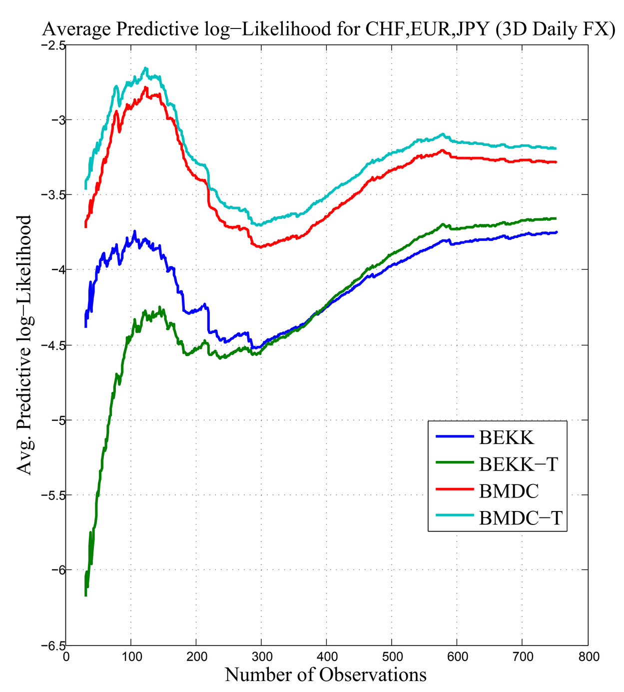


The BMDC posterior predictive distribution is heavy-tailed since the log posterior predictive density is quite flat. The log posterior predictive density is evaluated on the particles that approximate the posterior distribution. In contrast, the log predictive density evaluated on the empirical mean computed across all the particle is not heavy-tailed.

Observations:

1. BEKK is less heavy-tailed, as it yields one maximum likelihood prediction. This prediction will have log predictive density like the BMDC empirical mean.
2. The BMDC-T posterior will be more heavy-tailed than BMDC's, with its heavy-tailed observation model.
3. This confirms that financial data have fat tails and are best modeled with heavy-tailed models.

Average Predictive log-Likelihood



The plot shows typical average predictive log-likelihood for a $3D$ Daily FX time series.

1. BMDC-T is the most predictive method for any number of observations.
2. BEKK-T underperforms BEKK early on, when relatively few observations are available to fit the models.
3. With more data, BEKK-T is less susceptible to overfitting and outperforms BEKK.
4. However, BEKK and BEKK-T still underperform BMDC and BMDC-T with more data.

Particle Filter Sensitivity and Executions Times

Sensitivity in the average predictive log-likelihood of BMDC-T to the number of particles used in the RAPF. Execution times in minutes shown in parentheses. Results for BEKK-T given for comparison.

Data Dim.	BEKK-T	N=1000	N=4000	N=9000	N=25000
1D	−1.33 (71)	−1.32 (8)	−1.32 (20)	−1.32 (52)	−**1.32** (173)
2D	−2.56 (339)	−2.58 (8)	−2.56 (25)	−**2.55** (55)	−2.56 (176)
3D	−3.66 (478)	−3.24 (9)	−3.21 (29)	−3.20 (57)	−**3.18** (180)
4D	−4.52 (1003)	−4.25 (9)	−4.28 (30)	−4.20 (64)	−**4.16** (183)
5D	−5.95 (2971)	−5.60 (9)	−5.55 (32)	−5.53 (68)	−**5.50** (202)
10D	− (−)	−9.01 (9)	−8.93 (45)	−7.96 (75)	−**7.84** (252)
20D	− (−)	−20.04 (20)	−18.10 (37)	−17.27 (83)	−**16.11** (359)

Conclusions

- We have introduced a novel Bayesian Multivariate Dynamic Covariance model with time-varying parameters.
- Our proposed model can adapt its parameters to changing dynamics in financial markets.
- Our proposed model has significantly better predictive performance than standard econometric models, such as BEKK, and more recent methods such as the Generalized Wishart Process.
- We have presented an inference method based on particle filtering that yields substantial savings in computation time, enabling scalable inference to high-dimensional and high-frequency datasets.

BMDC Model

The BMDC probabilistic model is:

$$\mathbf{x}_t \sim \mathcal{N}(0, \Sigma_t)\,, \tag{1}$$

$$\Sigma_t = \mathbf{C}_t^\top \mathbf{C}_t + \mathbf{B}_t \mathbf{x}_{t-1} \mathbf{x}_{t-1}^\top \mathbf{B}_t + \mathbf{A}_t \Sigma_{t-1} \mathbf{A}_t\,. \tag{2}$$

The model parameters are time-varying and follow diffusion processes:

$$\text{vec}(\mathbf{A}_t) \sim \mathcal{N}(\text{vec}(\mathbf{A}_{t-1}), \alpha^2 \mathbf{I})\,, \tag{3}$$

$$\text{vec}(\mathbf{B}_t) \sim \mathcal{N}(\text{vec}(\mathbf{B}_{t-1}), \beta^2 \mathbf{I})\,, \tag{4}$$

$$\text{vec}(\mathbf{C}_t) \sim \mathcal{N}(\text{vec}(\mathbf{C}_{t-1}), \gamma^2 \mathbf{I})\,, \tag{5}$$

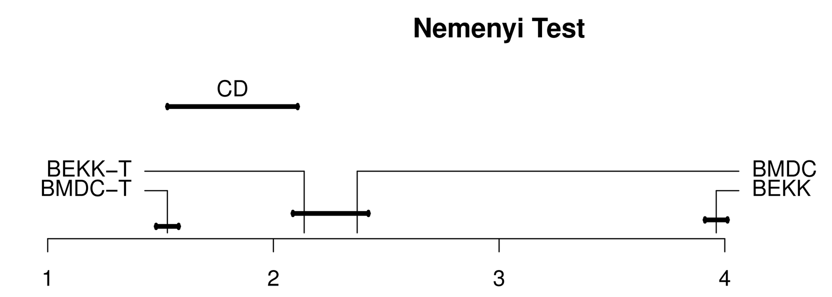
$$\alpha \sim \mathcal{N}(\kappa, \tau)\,, \quad \beta \sim \mathcal{N}(\kappa, \tau)\,, \quad \gamma \sim \mathcal{N}(\kappa, \tau)\,. \tag{6}$$

Bayesian Inference with Particle Filters

BMDC is learnt using the Regularized Auxiliary Particle Filter (RAPF) [Liu and West 1999]. RAPF is a fast online algorithm that:

1. Performs Sequential Monte Carlo [Doucet et al. 2001] on the extended space, $\Omega = (\Sigma_t, \theta_t, \alpha, \beta, \gamma)$.
2. Initially draw hyper-parameters, α, β and γ, from the hyper-priors, Equation 6.
3. Draw initial parameters, $\mathbf{A}_0$, $\mathbf{B}_0$ and $\mathbf{C}_0$, from their priors.
4. Generate new dynamic parameters according to Equations 3, 4 and 5.
5. Construct dynamic covariances using Equation 2.
6. Importance weight the particles and propagate forward.
7. At each sequential step, propose new hyper-parameters using a shrinkage kernel. The shrinkage kernel introduces artificial dynamics for the hyper-parameters, but is useful as it prevents a collapse in the diversity of the hyper-parameter particles.

Results for BMDC vs. BEKK

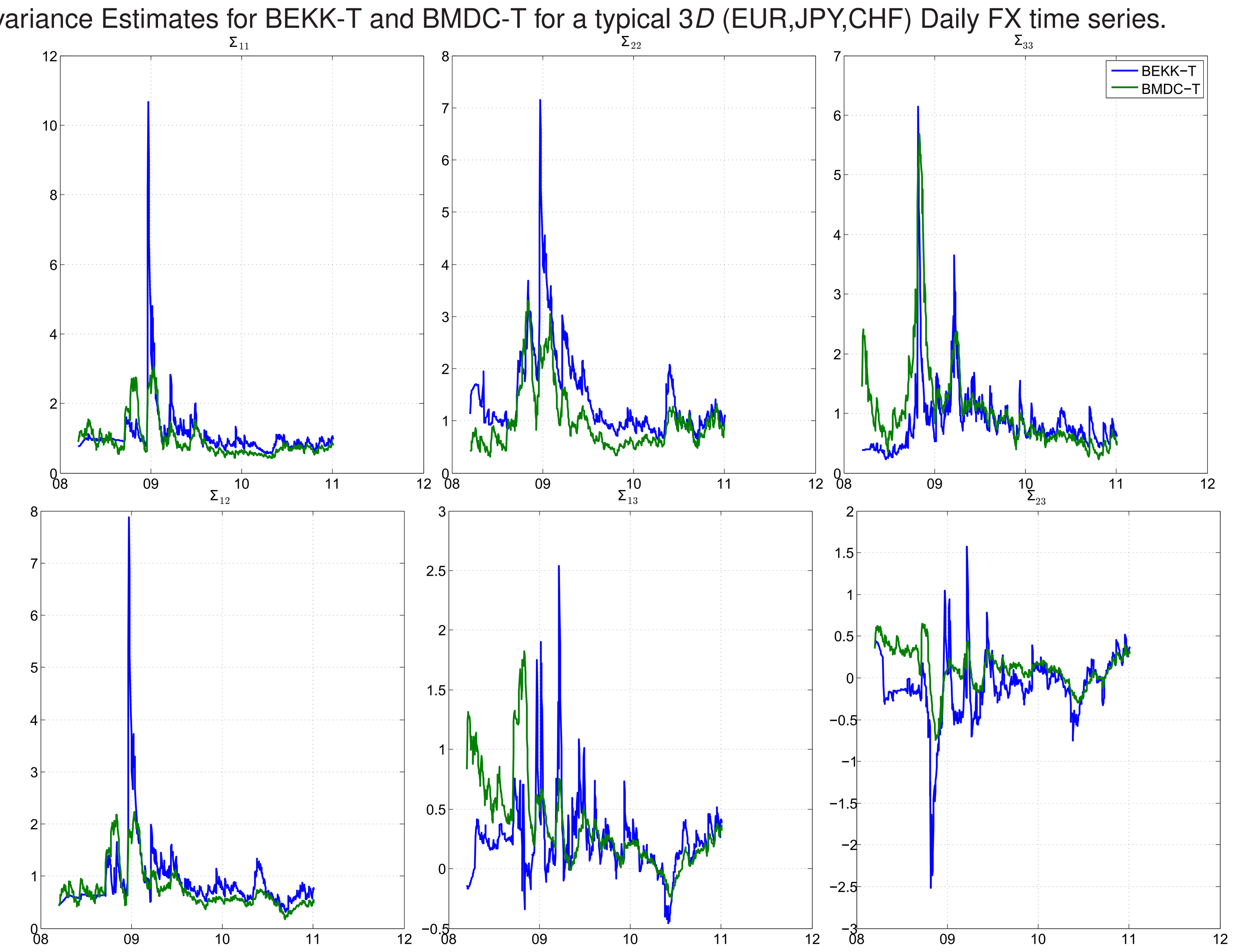


All to all comparison between BMDC-T, BMDC, BEKK-T and BEKK via a Nemenyi test [Demšar 2006] on the 66 datasets. Conclusions:

1. BMDC-T was statistically the most predictive model at $\alpha = 0.05$.
2. Dynamic models, BMDC-T and BMDC, more predictive than static models, BEKK-T and BEKK.
3. Heavy-tailed models, BMDC-T and BEKK-T, beat less heavy-tailed models, BMDC and BEKK.

Visualization of Dynamic Covariances

Covariance Estimates for BEKK-T and BMDC-T for a typical $3D$ (EUR,JPY,CHF) Daily FX time series.



BMDC-T identifies three spikes in the variance terms, Σ_{11}, Σ_{22} and Σ_{33}. The spikes correspond to the collapse of Lehman Brothers in September 2008, concerns over Eurozone indebtedness at the start of 2009, and recurring worries about European debt levels in April 2009. In contrast BEKK-T finds one intense spike. Furthermore, BEKK-T estimates for the covariance terms show large swings, while BMDC-T estimates are smoother. These fluctuations in the BEKK-T estimates are likely to be caused by overfitting. This overfitting leads to lower average predictive log-likelihood.

BMDC vs. Generalized Wishart Process

Recent non-parametric approaches for modeling dynamic covariances include the Generalized Wishart Process (GWP) [Wilson and Ghahramani 2011]:

$$u_{id}(t) \sim \mathcal{GP}(0, k) \tag{7}$$

$$\Sigma(t) = \sum_{i=1}^{\nu} \mathbf{L}\mathbf{u}_i(t)\mathbf{u}_i^\top(t)\mathbf{L}^\top \sim \mathcal{W}_D(\mathbf{L}\mathbf{L}^\top, \nu) \tag{8}$$

$$\mathbf{u}_i(t) = (u_{i1}(t), \ldots, u_{iD}(t))^\top \tag{9}$$

Predictive performance of BMDC versus BEKK and GWP in terms of cumulative predictive log-likelihood:

Dataset	BEKK-Full	BEKK	GWP	BMDC
FX (3D)	2025	2050	2020	**2130**
Equity (5D)	2785	2800	2930	**3090**

References

1. R.F. Engle and K.F. Kroner. Multivariate simultaneous generalized ARCH. Econometric theory, 11(01):122-150, 1995.
2. A. Wilson, Z. Ghahramani. Generalised Wishart Processes. In UAI-11, pp. 736-744, 2011.
3. A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice.* Springer Verlag, 2001.
4. J. Liu, and M. West. Combined parameter and state estimation in simulation-based filtering. Institute of Statistics and Decision Sciences, Duke University, 1999.
5. J. Demšar, Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research, 7:1-30, 2006.