Fitting Vast Dimensional Time-Varying Covariance Models

Kevin Sheppard

Department of Economics
University of Oxford

Joint with
Cavit Pakel, Bilkent
Robert F. Engle, NYU
Neil Shephard, Harvard

October 4, 2017

UNIVERSITY OF
OXFORD

The Problem

- Most efforts in covariance modeling modeling suffers from an extreme curse of dimensionality
  - L assets
  - $O\left(L^2\right)$ parameters
  - $O\left(L^3\right)$ single likelihood computational complexity
  - Up $O\left(L^5\right)$ cost to perform a single Netwon-Raphson update
- Most models are feasible for between 5 and 50 assets
- Substantial interest in estimating dynamic covariance models in realistic dimensions: 100 – 1000 assets
  - Multivariate ARCH
    - Engle (2002), Tse & Tsui (2002), Ledoit, Santa Clara & Wolf (2003), Cappiello, Engle & Sheppard (2006), Engle & Kelly (2007), Engle, Ledoit & Wolf (2016), *others*
  - Realized Covariance
    - Hautsch, Kyj & Oomen (2010), Lunde, Sheppard, Shephard (2015)

Contributions

- Develops a general estimation strategy applicable to many dynamic covariance specifications
  - Well suited to models where the dimension of the parameters governing the dynamics is small relative to the number of assets
- Identifies a nuisance parameter problem relevant for estimating high-dimensional models
  - Affects a wide range of dynamic models, including high-dimensional dynamic copulas
- Introduce a new estimator based on composite likelihoods
- Provide asymptotic theory for the two cases:
  - Classic: Fixed N, large T
  - Diagonal: Large N, large T
- Establish economic gains to using new estimator

Notation

UNIVERSITY OF OXFORD

- T time periods
- L assets
  - Interested in the case where $L \approx T$ or even $L > T$
- N submodels – typically $N = O\left(L^2\right)$ buy may be smaller
- r_t is a L by 1 vector of financial asset returns
  - For simplicity assume conditional mean 0
- H_t is an $L \times L$ positive definite *conditional* covariance matrix

$$H_t = \text{Cov}\left[r_t | \mathcal{F}_{t-1}\right] = \text{E}\left[r_t r_t' | \mathcal{F}_{t-1}\right]$$

- $\theta = (\gamma, \lambda)$ are model parameters
  - "Static" parameters:$|\gamma| = O\left(L^2\right)$
  - "Dynamic" parameters:$|\lambda| = O\left(1\right)$

Nuisance Parameter Problem

UNIVERSITY OF OXFORD

- Scalar BEKK using covariance targeting

$$H_t = (1 - \alpha - \beta)\,\Sigma + \alpha r_{t-1} r'_{t-1} + \beta H_{t-1}$$

- Σ is estimated using a moment-based estimator

$$\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} r_t r'_t, \quad \hat{\lambda} = \mathrm{vech}\left(\hat{\Sigma}\right)$$

- α, β estimated conditioning on the moment estimator

$$\max_{\alpha,\beta} \log L\left(\alpha, \beta; \hat{\lambda}, r\right)$$

  - We call this 2-step MLE (2MLE)
  - Returns assumed to be conditionally normally distributed

Estimates based on the S&P Indices

- Application to S&P 100 and S&P 500
- Used all members continuously available from 1997–2006 (96/480)
- Parameter estimates follow a distinct pattern

	S&P 100			S&P 500	
L	α	β	L	α	β
5	.0189	.9794	5	.0261	.9715
10	.0125	.9865	10	.0080	.9909
25	.0081	.9909	25	.0055	.9932
50	.0056	.9926	50	.0034	.9934
96	.0041	.9932	100	.0015	.9842
			480	.0032	.5630

- Dynamics become less pronounced as L grows

General problem representation

- Transform r_t into $X_t = \{X_{1t}, \ldots, X_{Nt}\}$
  - X_{jt} is a vector containing a subset of r_t
  - Dimensions need not be common
- We focus on unique pairing

$$X_{1t} = (r_{1t}, r_{2t})'$$
$$X_{2t} = (r_{1t}, r_{3t})'$$
$$\vdots$$
$$X_{\frac{(L(L-1))}{2}} = (r_{L-1t}, r_{Lt})$$

- Here $N = L(L-1)/2$, but we consider smaller N as well
- Important: subsets need not be exhaustive

Bivariate Covariance Model

- Define

$$X_{jt} = \left(r_{k_{j1}t}, r_{k_{je}t}\right)', \ k_{j1}, k_{j2} \in \{1, 2, \ldots, K\}$$

so that $\text{Cov}\,(r_{mt}, r_{nt} | \mathcal{F}_{t-1}) = h_{mnt}$

- The conditional covariance is then

$$\text{Cov}\left[X_{jt} | \mathcal{F}_{t-1}\right] = \left[\begin{array}{cc} h_{k_{j1}tk_{j1}t} & h_{k_{j1}tk_{j2}t} \\ h_{k_{j1}tk_{j2}t} & h_{k_{j2}tk_{j2}t} \end{array}\right] = H^{\star}_{jt}$$

- A valid estimator can be constructed using only this pair using

$$\log L_j\left(\psi_j\right) = \sum_{t=1}^{t} l_{jt}\left(\psi_j\right), \ l_{jt} = \log f\left(X_{jt}; \psi_j\right)$$

where $\psi_j = \left(\gamma_j', \lambda'\right)'$

Composite Likelihood

- The pair-wise loglikelihood is

$$l_{jT}\left(\gamma_j,\lambda\right)=\frac{1}{T}\sum_{t=1}^{T}\left[\log\left|H_{jt}\left(\gamma_j,\lambda\right)\right|+X_{jt}'H_{jt}^{-1}\left(\gamma_j,\lambda\right)X_{jt}\right],$$

- The composite loglikelihood is defined by averaging across all pairwise loglikelihoods

$$l_{NT}\left(\gamma_1,...,\gamma_N,\lambda\right)=\frac{1}{N}\sum_{j=1}^{N}l_{jT}\left(\gamma_j,\lambda\right)$$

  - Would be log-likelihood if $\{X_{1t},\ldots X_{Nt}\}$ where independent
  - Composite naming follows Lindsay (1988)
- The parameters of interest, λ, are estimated

$$\hat{\lambda}=\arg\min_{\lambda}l_{NT}\left(\hat{\gamma}_1,...,\hat{\gamma}_N,\lambda\right)$$

Computational Aspects

- Likelihood evaluation of standard estimators is $O\left(L^3\right)$
- Using all pairs composite likelihood (2MCL) is $O\left(L^2\right)$
- Using only contiguous pair composite likelihood (2SMCL) is $O\left(L\right)$
  - Random selection is strictly exogenous, so no consequences for inference
- Tempting to use $o\left(L\right)$ pairs in very large panels
  - Simplifies asymptotic theory
  - Little efficiency loss for $L \geq 20$
  - Asymptotic efficiency is bounded by the common factor structure in financial returns

Estimation Time

- Estimation using randomly ordered members of S&P 500

L	2MLE	2MCLE	2MSCLE
5	24s	0.1s	0.2s
25	46s	2.1s	0.2s
50	2m 10s	10s	0.5s
100	1h 50m	39s	0.8s
250	15h 11m	4m 7s	1.6s
480	85h 33m	18m 6s	4.5s

- Further gains possible exploiting variation-free likelihood for efficient parallel computation

Asymptotic Theory

UNIVERSITY OF OXFORD

- Starting from pair-wise BEKK likelihood

$$H_{jt}(\gamma_j, \lambda) = \Gamma_j (1 - \alpha - \beta) + \alpha X_{j,t-1} X'_{j,t-1} + \beta H_{j,t-1}(\gamma_j, \lambda)$$

- Model parameters are $\theta = \left(\gamma'_1, ..., \gamma'_N, \lambda'\right)'$
- Asymptotic theory focuses on inference for λ
- Estimator behavior determined by

$$Z_{t,T,N} = \frac{1}{N} \sum_{j=1}^{N} \frac{\partial l_{jt}\left(\gamma_{j0}, \lambda_0\right)}{\partial \lambda} + \frac{1}{N} \sum_{j=1}^{N} \frac{\partial^2 l_{jT}\left(\gamma_{j0}, \lambda_0\right)}{\partial \lambda \partial \gamma'_j} \left(\hat{\gamma}_j - \gamma_{j0}\right)$$

- Focus is on case where $N = O\left(T\right)$
  - Asymptotic theory for case where $N = O\left(1\right)$ (L finite) is standard, aside from composite-likelihood structure

Key Assumptions

Omitting common assumptions (compactness...)

Assumption

$\lim_{L,T\to\infty} L/T = c$ *where* $0 < c < \infty$.

- Number of asset is growing, but not too quickly
- Number of pairs is growing fast

Assumption

$\max_{1\leq j\leq N,\, 1\leq t\leq T} E[||X_{jt}||^{10}] < \infty$, *as* $N, T \to \infty$.

- Returns have at least 10 moments. L finite requires at least 6.

Key Assumptions

Assumption

$\mathcal{I} = \lim_{N,T\to\infty} Var\left(T^{-1/2}\sum_{t=1}^{T} Z_{t,T,N}\right)$ *exists and is positive definite.*

Also, $\mathcal{D}^{-1} = \lim_{N\to\infty}\left\{-E\left[\partial^2 l_{NT}\left(\gamma_{10},\ldots,\gamma_{N0},\lambda_0\right)/\partial\lambda\partial\lambda'\right]\right\}^{-1}$ *exists.*

$T^{-1/2}\sum_{t=1}^{T} Z_{t,T,N} \xrightarrow{d} N\left(0,\mathcal{I}\right)$, *as* $N,T\to\infty$.

- Key assumption: Cross-sectional average is $O_p\left(1\right)$ – requires strong dependence
- Problems with weaker dependence usually converge faster but often have asymptotic biases

Asymptotic Behavior

Key Results

Theorem (Consistency)

Suppose Assumptions hold for $N = O(L)$. Then, as $N, T \to \infty$, we have $\max_{1 \le j \le N} \left|\left|\hat{\gamma}_j - \gamma_{j0}\right|\right| \xrightarrow{p} 0$. If $N = O(L)$, then $\lambda \xrightarrow{p} \lambda_0$ as $N, T \to \infty$.

Theorem (Asymptotic Normality)

Suppose Assumptions hold for $N = O(L)$. Then,

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \xrightarrow{d} N\left(0, \mathcal{D}^{-1}\mathcal{I}\mathcal{D}^{-1}\right).$$

In-sample estimates

Scalar BEKK for S&P 100

- CL parameter estimates are stable as cross-section dimension grows
- Standard errors decline, but only slowly in L

	2MLE		2MCLE		2SMCLE	
L	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
5	.0189	.9794	.0287 (.0081)	.9692 (.0092)	.0284 (.0083)	.9696 (.0094)
10	.0125	.9865	.0281 (.0055)	.9699 (.0063)	.0272 (.0054)	.9709 (.0062)
25	.0081	.9909	.0308 (.0047)	.9667 (.0055)	.0307 (.0049)	.9668 (.0056)
50	.0056	.9926	.0319 (.0046)	.9645 (.0056)	.0316 (.0047)	.9647 (.0057)
96	.0041	.9932	.0334 (.0041)	.9636 (.0049)	.0335 (.0043)	.9634 (.0051)

- Inference is *very slow* in 2MLE

Efficiency

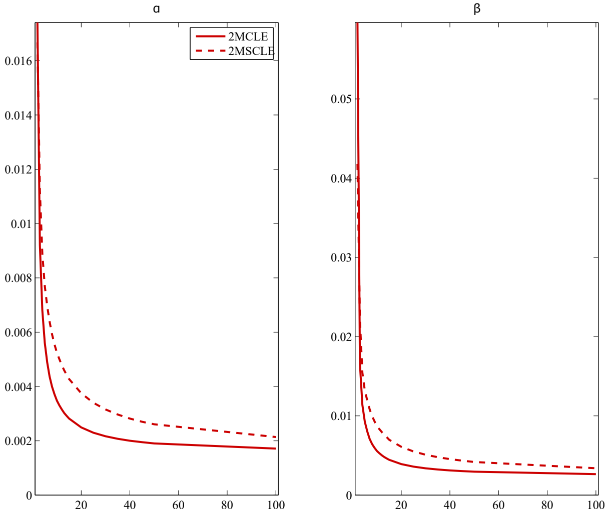
- Asymptotic theory requires that cross-sectional average is still $O_p(1)$
- However, does not mean there are no gains to using larger cross-sections
- Conceptually

$$Z_{t,T,j} \approx f_t + \epsilon_{T,j}$$

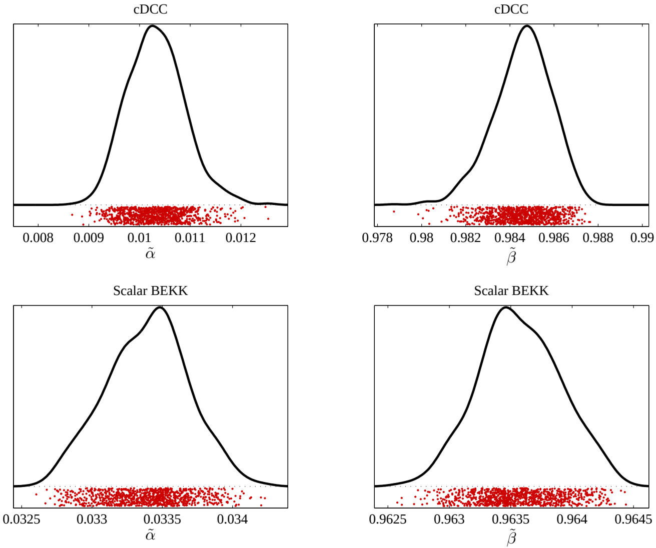
- Cross-sectional averaging will reduce impact of $\epsilon_{T,j}$ terms, but will not eliminate f_t term
- Compare all pair estimator with contiguous pair estimator

Monte Carlo Evidence: Efficiency

- Scalar BEKK, $\alpha = .05, \beta = .93$



Random Subset Estimator



Empirical Application

- Empirical application to S&P 100 components
- Model specification is an EGARCH-cDCC
- Marginal variances specified as EGARCH (Nelson (1991)) processes
- The market is a standard EGARCH model

$$\ln h_{\bullet,t} = \omega_\bullet + \alpha_\bullet|\epsilon_{\bullet,t-1} - \sqrt{2/\pi}| + \kappa_\bullet\epsilon_{\bullet,t-1} + \beta_\bullet \ln h_{\bullet,t-1}, \quad \epsilon_{\bullet,t} = \bar{r}_t h_{\bullet,t}^{-1/2}.$$

- Individual assets use an extended form

$$\ln \tilde{h}_{j,t} = \omega_j + \alpha_j|\epsilon_{j,t-1} - \sqrt{2/\pi}| + \kappa_j\epsilon_{j,t-1} + \beta_j \ln h_{j,t-1}, \quad h_{j,t} = h_{\bullet,t}\tilde{h}_{j,t}, \quad \epsilon_{j,t} = r_{j,t}h_{j,t}^{-1/2}$$

- Conditional correlation follows cDCC (Aielli (2013))

$$R_{jt} = P_{jt}^{-1/2} Q_{jt} P_{jt}^{-1/2}, \quad P_{jt} = \begin{pmatrix} Q_{11jt} & 0 \\ 0 & Q_{22jt} \end{pmatrix},$$

$$Q_{jt} = \Psi_j (1 - \alpha - \beta) + \alpha P_{jt-1}^{1/2} \left(S_{jt-1}S'_{jt-1} - R_{jt-1} \right) P_{jt-1}^{1/2} + (\alpha + \beta) Q_{jt-1}, \quad \Psi_j = \begin{pmatrix} 1 & \varphi_j \\ \varphi_j & 1 \end{pmatrix}$$

In-sample Parameter Estimates

- Composite cDCC parameter estimates are very stable across dimension
- MLE estimated parameter collapse in even moderate dimensions

L		**2MLE**	**2MCLE**		**2MSCLE**	
5	.0101	.9823	.0143 (.0487)	.9829 (.0846)	.0099 (.0033)	.9885 (.0045)
10	.0030	.9908	.0107 (.0012)	.9881 (.0016)	.0093 (.0016)	.9886 (.0018)
25	.0018	.9882	.0100 (.0009)	.9871 (.0017)	.0089 (.0011)	.9889 (.0012)
50	.0015	.9524	.0101 (.0008)	.9856 (.0018)	.0092 (.0010)	.9869 (.0019)
96	.0020	.5561	.0103 (.0009)	.9846 (.0019)	.0094 (.0009)	.9860 (.0014)

Comparing Estimation Methods

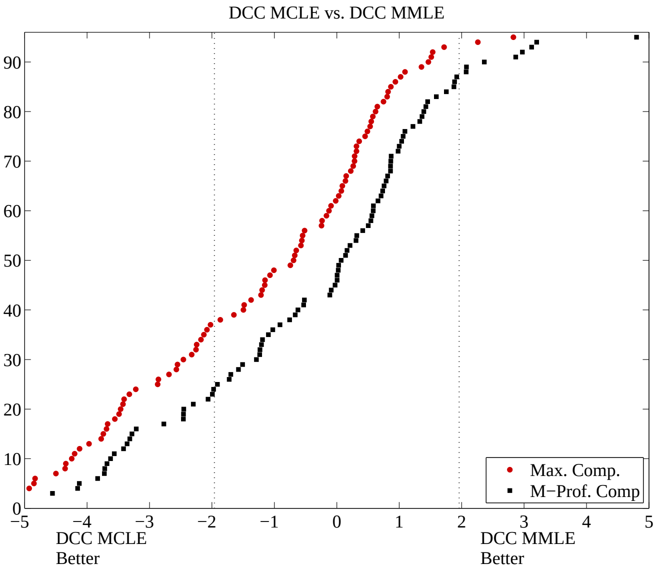
- Models can be estimated using alternative methods
- Natural to compare models using Giacomini-White framework
  - Forecast error assumed to be asymptotically non-negligible
- Performance based on variance of out-of-sample market-hedged residual

$$\hat{\epsilon}_{jt} = r_{jt} - \hat{\beta}_t r_{mt}$$

  - $\hat{\beta}_{jt}$ depends on estimated model variances and correlation

Comparing Estimation Methods

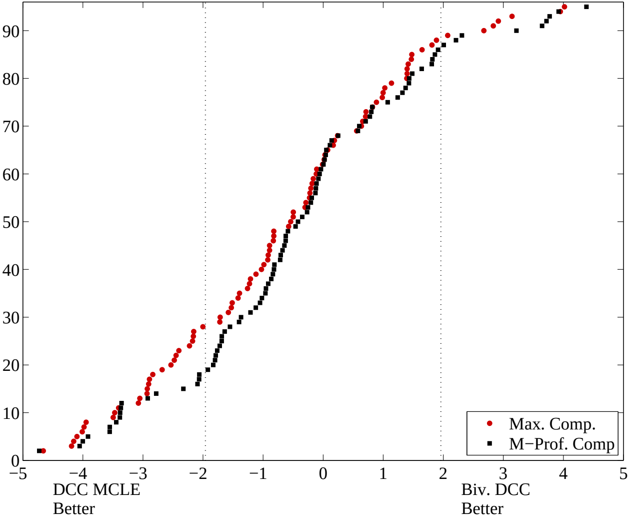
UNIVERSITY OF OXFORD



Comparing against Bivariate Models

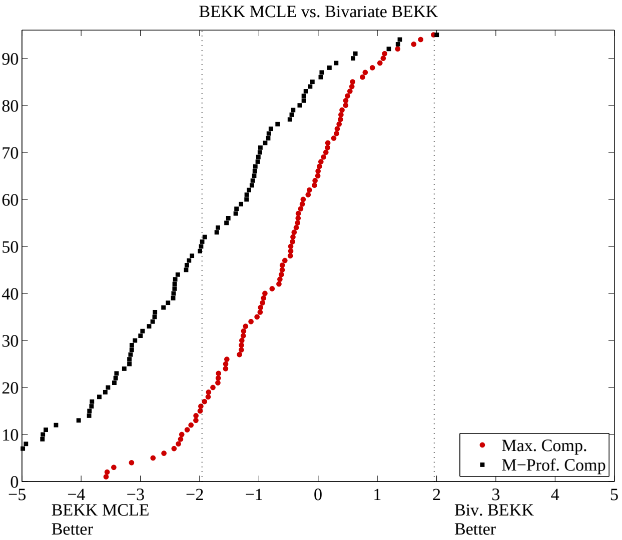
cDCC

DCC MCLE vs. Bivariate DCC

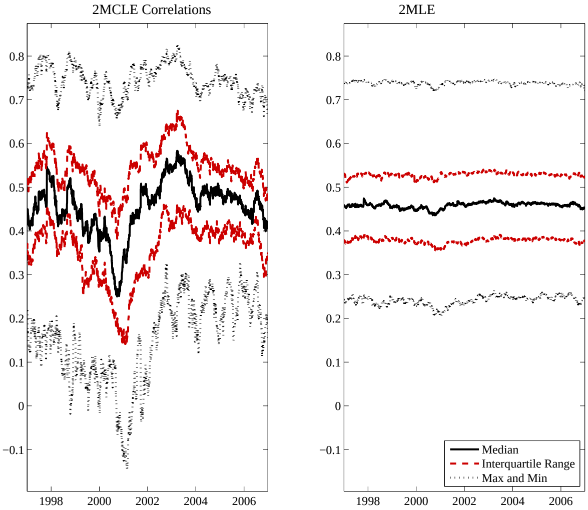


Comparing against Bivariate Models

Scalar BEKK



Forecasting Differences



Understanding the Source of the Problem

- Entire difference results in estimation error in long-run target
  - $\hat{\Sigma}$ in a scalar BEKK
- Examined the estimated average condition numbers of $\left\{\hat{\Sigma}_t\right\}$ as a function of L and T for MLE and MCLE
  - Ratio of maximum to minimum eigenvalue
- Standardized these by true condition number since depends on L

$$\underbrace{\hat{u}_{Lt} = \frac{\hat{\lambda}_{t,\max}/\hat{\lambda}_{t,\min}}{\lambda_{t,\max}/\lambda_{t,\min}}}_{\text{MLE}}, \quad \hat{u}_{jt} = \underbrace{\frac{\hat{\lambda}_{jt,\max}/\hat{\lambda}_{jt,\min}}{\lambda_{jt,\max}/\lambda_{jt,\min}}}_{\text{MCLE}}$$

- Interest in $\bar{u}_{c,LT} = (NT)^{-1} \sum_{j=1}^{N} \sum_{t=1}^{T} \hat{u}_{jt}$ or $\bar{u}_{LT} = T^{-1} \sum_{t=1}^{T} \hat{u}_{Lt}$

Understanding the Source of the Problem

UNIVERSITY OF OXFORD

- Regressed these on L and T to approximate rate

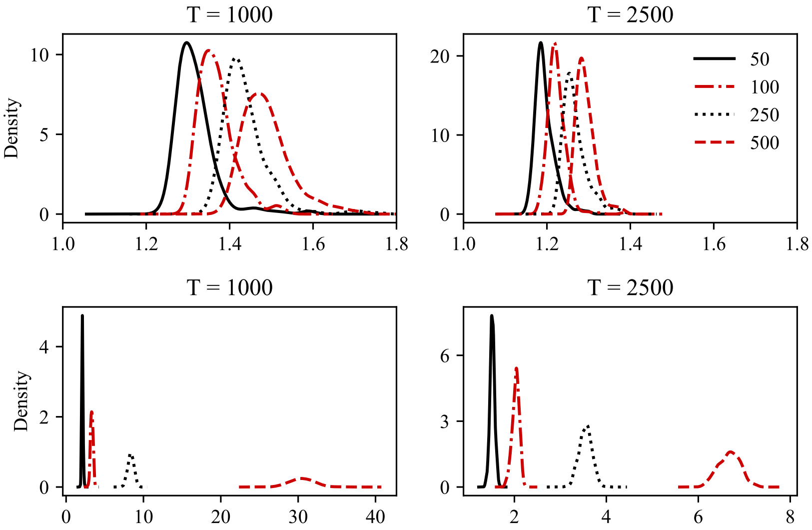
$$y_{L,T,r} = \beta_0 + \beta_1 \ln L + \beta_2 \ln T + \eta_{L,T,r},$$

- $y_{LT,r}$ is one of the two measures, $\bar{u}_{LT,r}$ or $\bar{u}_{c,LT,r}$ as a function of L or T
- Measure convergence rates $y_{L,T,r} = O(\ln L^{\beta_1} T^{\beta_2})$

	L	T
2MLE	1.0046	$-.8383$
2MCLE	$.0487$	$-.1448$

- Increasing L hurts, increasing T helps.
- Sum of coefficients confirms that diagonal is much worse for 2MLE

Understanding the Source of the Problem



Performance in Monte Carlo

- Monte Carlo varying L and T
- Here $T = 2,000$
- Focus on Scalar BEKK $\alpha = .05, \beta = .93$

Bias

	2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β
3	-.000	-.008	-.000	-.009	.000	-.010
10	-.001	-.005	-.000	-.007	-.000	-.007
50	-.006	-.003	-.000	-.006	-.000	-.006
100	-.012	-.004	-.000	-.006	-.000	-.006

- Bias in α is nearly 25% when $L = 100$

Monte Carlo Continued

RMSE

	2MLE		2MCLE		2MSCLE	
L	α	β	α	β	α	β
3	.008	.023	.009	.025	.010	.029
10	.003	.009	.005	.014	.006	.015
50	.006	.004	.003	.009	.003	.009
100	.012	.004	.003	.009	.003	.009

- RMSE cost in small model where $L = 3$ is negligible
- For moderately large L bias dominates RMSE for 2MLE

Conclusions

UNIVERSITY OF OXFORD

- Introduce composite likelihood for estimation of vast dimensional models
  - Faster then MLE
  - Less biased than MLE for most interesting problem sizes
  - Feasible when $L > T$
  - Asymptotically valid when L grows with T
  - Only a small price in extra PEE even when MLE would be better
- Shows clear gains in empirically relevant situations hedging
- Additional application areas include dynamic hedge fund betas and macroeconomic risk measurement
- Well suited to unbalanced panels since easy to include a pair when available