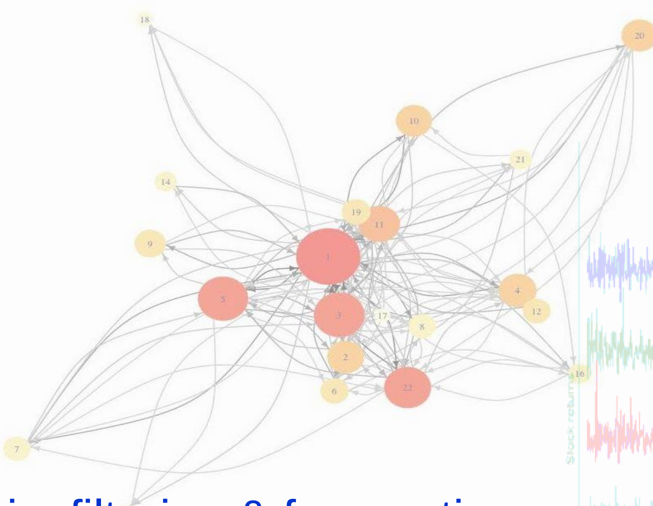
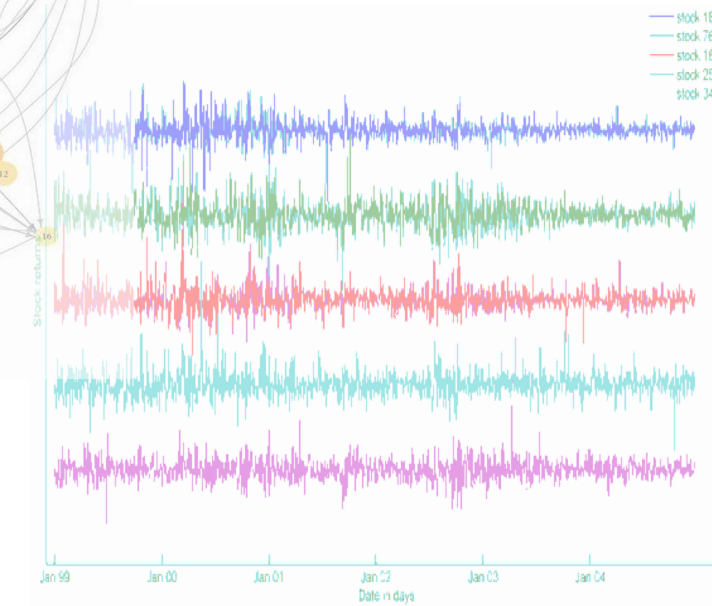
Structured Dynamic Graphical Models & Scaling Multivariate Time Series Methodology

Mike West
Duke University



$$\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$$



Aims:

- Scale up in m
  - sequential analysis, filtering & forecasting

- "Decoupled" univariate series / models
  - Series-specific independent predictors
  - Series-specific states, time evolutions, volatilities, hyperparameters
  - Sequential analysis: Analytic/fast? Parallelisation?

- Critical cross-series / multivariate structure
  - Sensitive co-volatility modelling
  - Coherent joint forecast distributions

Simultaneous models: Coupled set of univariate DLMs

$$
\begin{aligned}
y_{jt} &= \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt} + \nu_{jt} \\
&= \mathbf{x}'_{jt}\phi_{jt} + \mathbf{y}'_{sp(j),t}\gamma_{jt} + \nu_{jt}
\end{aligned}
$$

Multiple univariate models
- *"decoupled"*
- *in parallel*

Independent predictors

Simultaneous "parents"

$$\mathbf{F}_{jt} = \begin{pmatrix} \mathbf{x}_{jt} \\ \mathbf{y}_{sp(j),t} \end{pmatrix} \qquad \boldsymbol{\theta}_{jt} = \begin{pmatrix} \phi_{jt} \\ \gamma_{jt} \end{pmatrix}$$

Simultaneous parents

$$sp(j) \subseteq 1:m \backslash j$$

**Current* values of (some) related series as predictors*

**Independent* across j: residuals & univariate volatilities:* $\nu_{jt} \sim N(0, 1/\lambda_{jt})$

$$\boldsymbol{\Lambda}_t = \mathrm{diag}(\lambda_{1t}, \ldots, \lambda_{mt})$$

Implied coherent multivariate model

$$y_{jt} = \mathbf{x}'_{jt}\phi_{jt} + \mathbf{y}'_{sp(j),t}\gamma_{jt} + \nu_{jt}$$

$$\mu_{jt}$$

$$\boxed{\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Gamma}_t\mathbf{y}_t + \boldsymbol{\nu}_t}$$

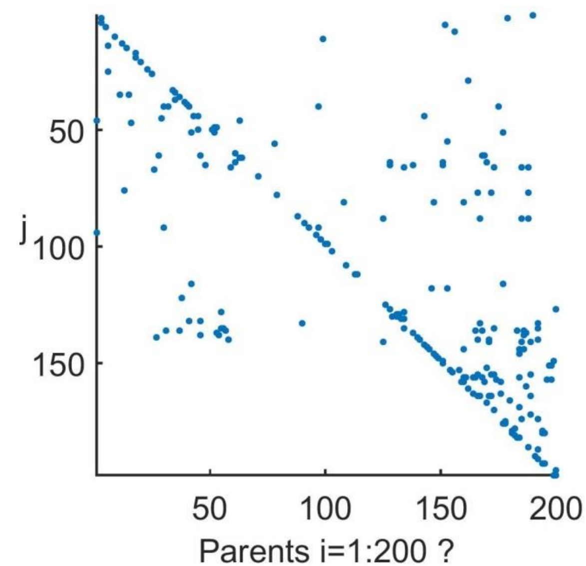
dynamic regressions & precision/volatility

$$\mathbf{y}_t = (\mathbf{I} - \boldsymbol{\Gamma}_t)^{-1}\boldsymbol{\mu}_t + N(\mathbf{0}, \boldsymbol{\Omega}_t^{-1})$$

$$\boldsymbol{\Omega}_t = (\mathbf{I} - \boldsymbol{\Gamma}_t)'\boldsymbol{\Lambda}_t(\mathbf{I} - \boldsymbol{\Gamma}_t)$$

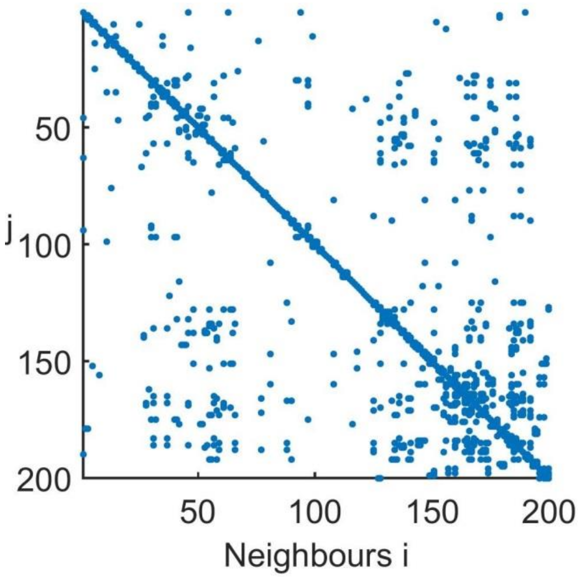
$$\boldsymbol{\Gamma}_t = \begin{pmatrix} 0 & \gamma_{1,2,t} & \gamma_{1,3,t} & \cdots & \gamma_{1,m,t} \\ \gamma_{2,1,t} & 0 & \gamma_{2,3,t} & \cdots & \gamma_{2,m,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{m-1,1,t} & \cdots & \gamma_{m-1,m-2,t} & 0 & \gamma_{m-1,m,t} \\ \gamma_{m,1,t} & \gamma_{m,2,t} & \cdots & \gamma_{m,m-1,t} & 0 \end{pmatrix}$$

Sparse



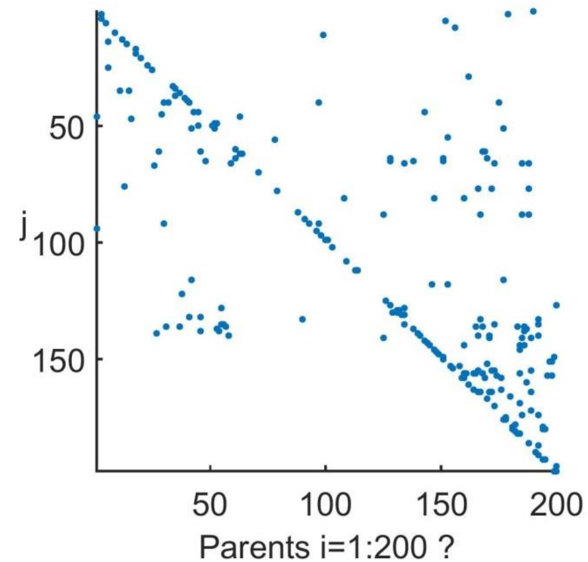
Non-zeros in $\boldsymbol{\Gamma}_t$

Notation: extend γ_{jt} to row *j* and pad with 0 entries



Non-zeros in $\boldsymbol{\Omega}_t$

zero precision:
conditional
independence



Non-zeros in $\boldsymbol{\Gamma}_t$

$$\boldsymbol{\Omega}_t = (\mathbf{I} - \boldsymbol{\Gamma}_t)' \boldsymbol{\Lambda}_t (\mathbf{I} - \boldsymbol{\Gamma}_t)$$

Class of volatility matrix decompositions

Simultaneous models: Coupled set of univariate DLMs

Multiple univariate models
- *"decoupled"*
- *in parallel*

$$y_{jt} = \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt} + \nu_{jt}$$

Parallel states *- independent evolutions*

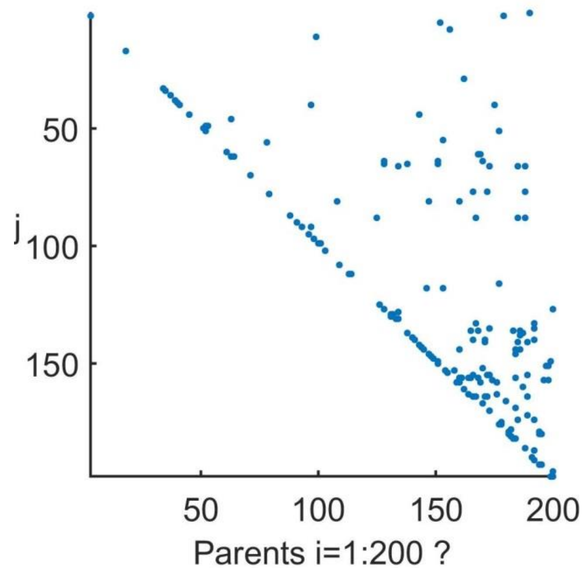
$$\boldsymbol{\theta}_{jt} = \mathbf{G}_{jt}\boldsymbol{\theta}_{j,t-1} + \boldsymbol{\omega}_{jt}, \quad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \mathbf{W}_{jt})$$

Subclass of simultaneous models

$$\mathbf{\Omega}_t = (\mathbf{I} - \mathbf{\Gamma}_t)'\mathbf{\Lambda}_t(\mathbf{I} - \mathbf{\Gamma}_t)$$

- $\mathbf{\Gamma}_t$ *triangular form: by choice or chance*
- *Sparse Cholesky-style precision (… popular)*
- *Directed graphical model*
- *Compositional model form enables computation*

Order of series relevant



Non-zeros in $\mathbf{\Gamma}_t$

[Multiregression DLMs: JQ Smith, C Queen 1990s
DDNMs: Z Zhao & MW 2014,15
Latent threshold versions: J Nakajima & MW 2012-15]

General class of SGDLMs

$$sp(j) \in \{1 : m \backslash j\}$$

$$\begin{aligned} y_{jt} &= \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt} + \nu_{jt}, \quad \nu_{jt} \sim N(0, 1/\lambda_{jt}) \\ &= \mathbf{x}'_{jt}\boldsymbol{\phi}_{jt} + \mathbf{y}'_{sp(j),t}\boldsymbol{\gamma}_{jt} + \nu_{jt} \end{aligned}$$

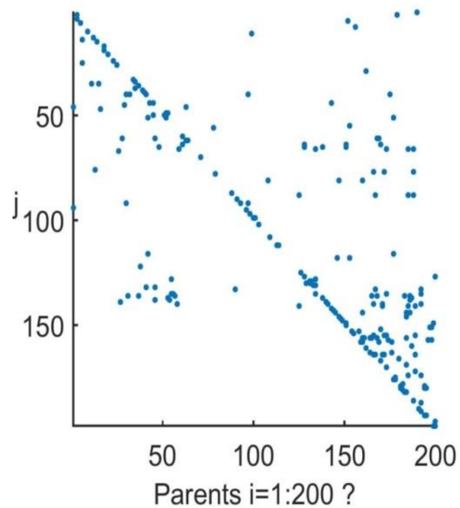
Simultaneous parents of j

$$\begin{aligned} \boldsymbol{\Theta}_t &= \{\boldsymbol{\theta}_{1t}, \ldots, \boldsymbol{\theta}_{mt}\} \\ \boldsymbol{\Lambda}_t &= \text{diag}(\lambda_{1t}, \ldots, \lambda_{mt}) \end{aligned}$$

Joint pdf …
likelihood for states and volatilities:

$$\boxed{p(\mathbf{y}_t|\boldsymbol{\Theta}_t, \boldsymbol{\Lambda}_t) = |\mathbf{I} - \boldsymbol{\Gamma}_t| \ \prod_{j=1:m} N(y_{jt}|\mathbf{F}'_{jt}\boldsymbol{\theta}_{jt}, 1/\lambda_{jt})}$$

- *"almost" decoupled!*
- *- coupling/coherence: "complicating" determinant*

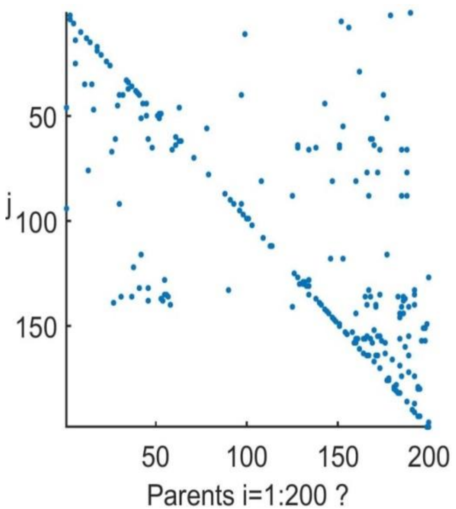


Non-zeros in $\boldsymbol{\Gamma}_t$

Decoupling/Recoupling sequential analysis of SGDLM

Intuition:

- Increasingly sparse $\boldsymbol{\Gamma}_t$ for larger *m*
- $|\mathbf{I} - \boldsymbol{\Gamma}_t|$ "minor" correction to likelihood
    - Triangular: compositional parallel models
    - Partly triangular: no impact on determinant
    - Very sparse: determinant close to 1



Non-zeros in $\boldsymbol{\Gamma}_t$

Sequential analysis strategy - At each *t:*

Decoupled parallel models: Conjugate forms
Forecast: simulate in parallel & recouple

Time *t* updating:
Update decoupled models then recouple for coherent inferences

Evolution to *t+1:* decoupled models

Decoupling/recoupling: Forecasting at time t

$$\mathbf{y}_t = (\mathbf{I}-\mathbf{\Gamma}_t)^{-1}\boldsymbol{\mu}_t + N(\mathbf{0},\mathbf{\Omega}_t^{-1})$$
$$\mathbf{\Omega}_t = (\mathbf{I}-\mathbf{\Gamma}_t)'\mathbf{\Lambda}_t(\mathbf{I}-\mathbf{\Gamma}_t)$$

$$y_{jt} = \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt} + \nu_{jt}, \quad \nu_{jt} \sim N(0, 1/\lambda_{jt})$$
$$= \mathbf{x}'_{jt}\phi_{jt} + \mathbf{y}'_{sp(j),t}\gamma_{jt} + \nu_{jt}$$

Parallel, normal/gamma priors:
$\prod_{j=1:m} p_{jt}(\boldsymbol{\theta}_{jt}, \lambda_{jt}|\mathbf{y}_{1:t-1})$

Directly simulate and combine samples:
$\mathbf{\Theta}_t^r, \mathbf{\Lambda}_t^r$

Transform and recouple for prediction:
$\boldsymbol{\mu}_t^r, \mathbf{\Gamma}_t^r, \mathbf{\Omega}_t^r$

Simulate the future: Monte Carlo sample
$\mathbf{y}_t^r|\boldsymbol{\mu}_t^r, \mathbf{\Gamma}_t^r, \mathbf{\Omega}_t^r \sim p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$

Predict further as needed e.g.

$$\mathbf{\Theta}_t^r, \mathbf{\Lambda}_t^r \rightarrow \mathbf{\Theta}_{t+1}^r, \mathbf{\Lambda}_{t+1}^r$$
$$\mathbf{\Theta}_{t+1}^r, \mathbf{\Lambda}_{t+1}^r, \mathbf{y}_t^r \rightarrow \mathbf{y}_{t+1}^r \sim p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t-1})$$

Exact posterior:

$$p(\boldsymbol{\Theta}_t, \boldsymbol{\Lambda}_t | \mathbf{y}_{1:t}) \propto |\mathbf{I} - \boldsymbol{\Gamma}_t| \; \prod_{j=1:m} \tilde{p}_{jt}(\boldsymbol{\theta}_{jt}, \lambda_{jt} | \mathbf{y}_{1:t})$$

"Complicating" term

Parallel, normal/gamma
"Most" of the posterior

Full posterior inferences: Monte Carlo importance sampling

$$\boldsymbol{\Theta}_t^i, \boldsymbol{\Lambda}_t^i \;\leftarrow\; \{\; \boldsymbol{\theta}_{jt}^i, \lambda_{jt}^i \;\sim\; \tilde{p}_{jt}(\cdot, \cdot | \mathbf{y}_{1:t}), \; j = 1:m \;\}$$

Importance weights: $\alpha_{it} \propto |\mathbf{I} - \boldsymbol{\Gamma}_t^i|$

- inference using weighted samples
- very large samples: trivial computations
  … decoupled conjugate posteriors in parallel

Posterior "almost" decoupled normal-inverse gammas

Decoupling: match product form with full posterior importance sample

$$q(\boldsymbol{\Theta}_t, \boldsymbol{\Lambda}_t | \mathbf{y}_{1:t}) = \prod_{j=1:m} p_{jt}(\boldsymbol{\theta}_{jt}, \lambda_{jt} | \mathbf{y}_{1:t}) \quad \xleftrightarrow{\text{approx}} \quad p(\boldsymbol{\Theta}_t, \boldsymbol{\Lambda}_t | \mathbf{y}_{1:t}) = \{\boldsymbol{\Theta}_t^i, \boldsymbol{\Lambda}_t^i; \alpha_{it}\}$$

normal-gamma

min Kullback-Leibler divergence

Variational Bayes (mean field approximation)

Decoupling/recoupling: Evolution to time *t+1*

Parallel states *- independent evolutions*

$$\boldsymbol{\theta}_{j,t+1} = \mathbf{G}_{j,t+1}\boldsymbol{\theta}_{jt} + \boldsymbol{\omega}_{j,t+1}, \quad \boldsymbol{\omega}_{j,t+1} \sim N(\mathbf{0}, \mathbf{W}_{j,t+1})$$

Kullback-Leibler divergence: Decreases through evolutions

Time *t+1:* Parallel, normal/inverse-gamma priors:

$$\prod_{j=1:m} p_{j,t+1}(\boldsymbol{\theta}_{j,t+1}, \lambda_{j,t+1}|\mathbf{y}_{1:t})$$

Time *t+1:* *Decoupled priors*

Recouple to forecast

Update

Decouple to evolve ….

Completes the *t : t+1* forecast/update/evolve sequence

… Continue ...

Example: SGDLM for 400 S&P stock price series

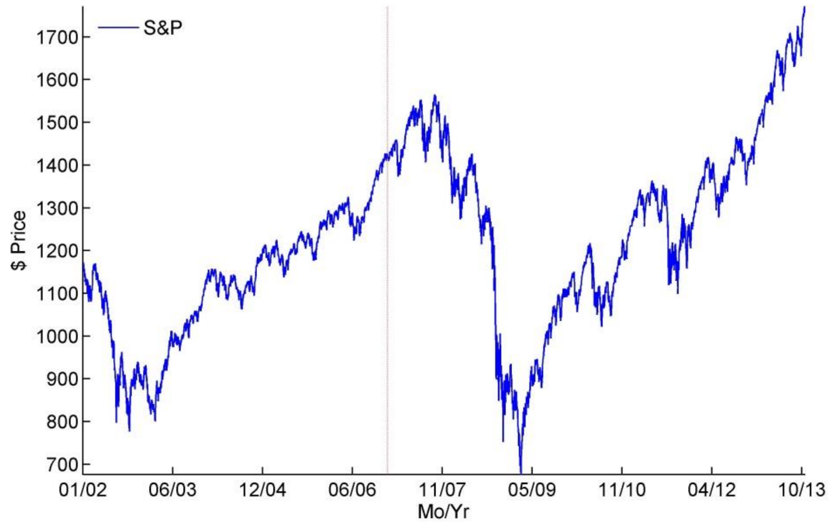
m=400 stocks & S&P index :
daily closing prices

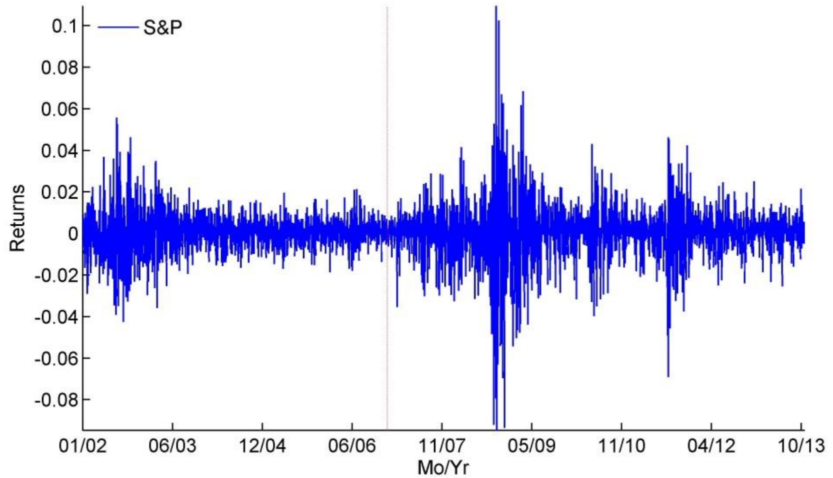
1/2002 – 10/2013

Model (log-difference) returns

Training data : 1/2002 – 12/2006

Test/prediction/portfolio decisions :
1/2007 – 10/2013





S&P example – Models for comparison

Model	Predictors
WDLM 1 SGDLM 1	$\mathbf{F}_{jt} = \mathbf{F}_t = 1$
WDLM 2 SGDLM 2	$\mathbf{F}_{jt} = \mathbf{F}_t = (1, y_{\mathsf{SPX},t-1}, \mathsf{TNX}_{t-1})'$
WDLM 3 SGDLM 3	$\mathbf{F}_{jt} = \mathbf{F}_t = (1, \Delta y_{\mathsf{SPX},t-1}, \Delta\mathsf{TNX}_{t-1}, \Delta^2\mathsf{VIX}_{t-1})'$
WDLM 4 SGDLM 4	$\mathbf{F}_{jt} = \mathbf{F}_t = (1, 0.5\sum_{k=1:2} y_{\mathsf{SPX},t-k} - 0.2\sum_{k=1:5} y_{\mathsf{SPX},t-k}, \Delta^2\mathsf{VIX}_{t-1})'$
SGDLM 5	$\mathbf{F}_{jt} = (1, 0.5\sum_{k=1:2} y_{j,t-k} - 0.2\sum_{k=1:5} y_{j,t-k}, \Delta^2\mathsf{VIX}_{t-1})'$

WDLM	Wishart discount DLM - common predictor model
$y_{\mathsf{SPX},t}$	S&P index series *(j=1 of 401)*
TNX_t	Annualized 10-year T-bill rate
VIX_t	Volatility index (annualized, implied 30-day volatility)

Simultaneous parental set specification?

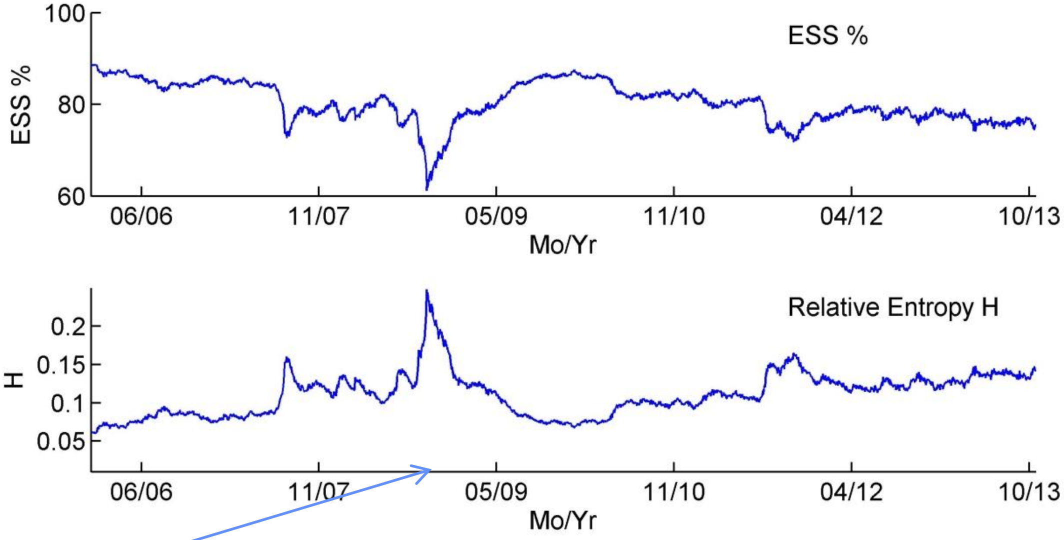
Bayesian model selection / model averaging

Parallel stochastic search over parental sets

“Simpler” specification methods

S&P example – Monitoring decoupling/recoupling

Importance sampling-recoupling:
I=10,000 Monte Carlo sample size



Banking crisis

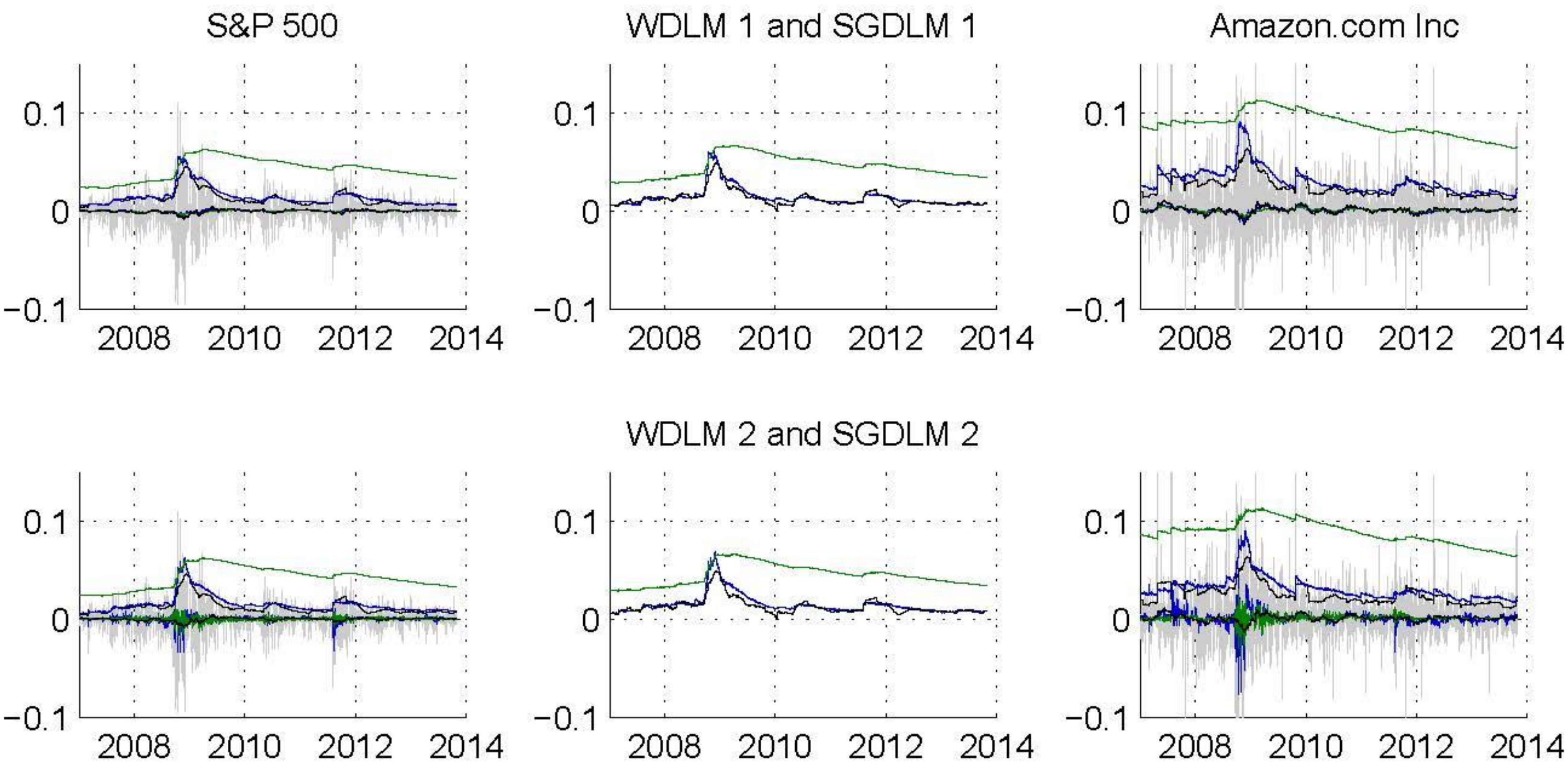
Potential to intervene- "signal" of deterioration of ESS:
Increase sample size
Modify model: parental sets, discounts, etc

S&P SGDLMs : 1-step forecast insights

data
1-step forecasts
volatilities

co-volatilities

data
1-step forecasts
volatilities

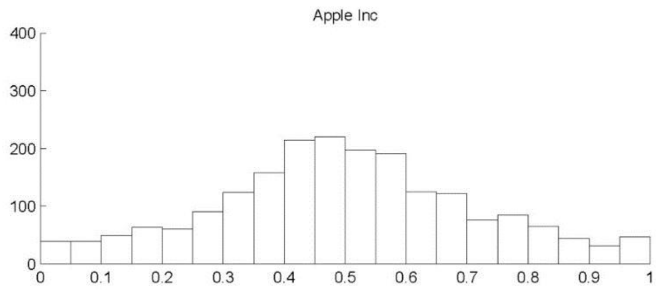


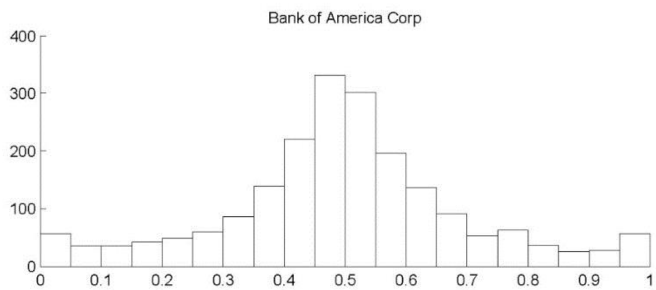
Uni-bivariate/empirical - Wishart DLM - SGDLM

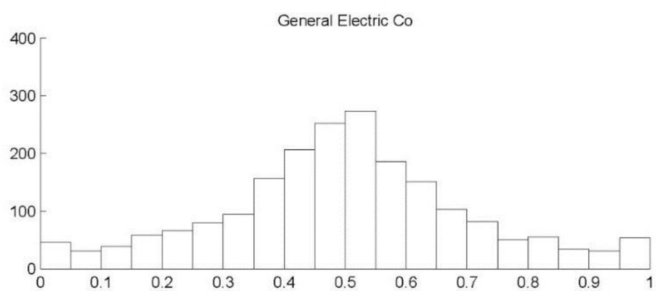
S&P SGDLMs : Can we ignore recoupling?

Realized 1-step forecast CDF transform : U(0,1) is perfect

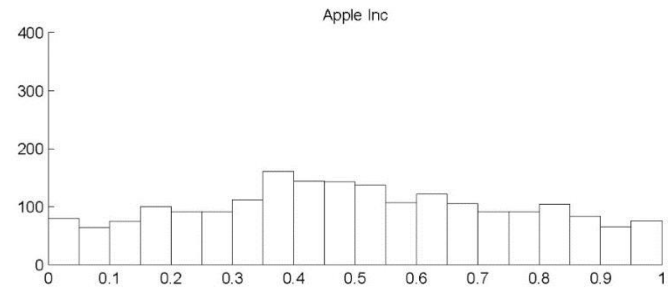
No recoupling

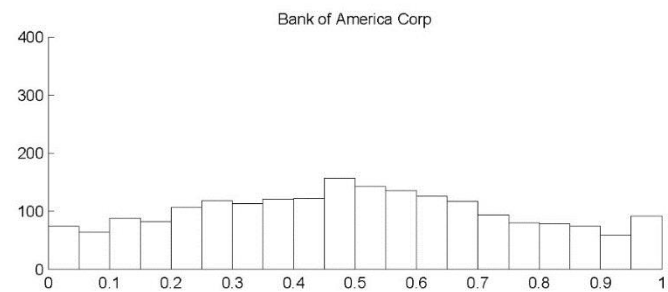


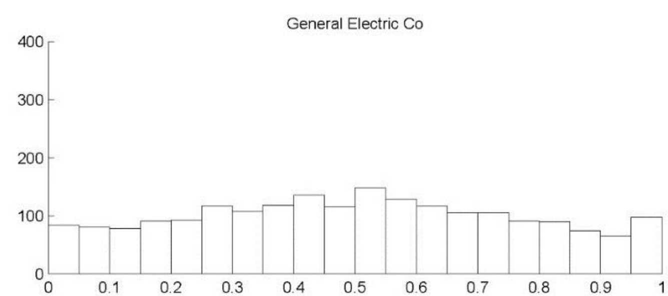




Full recoupling







Traditional Bayesian/Markowitz portfolios:
target return
minimize risk (portfolio SD)
subject to constraints

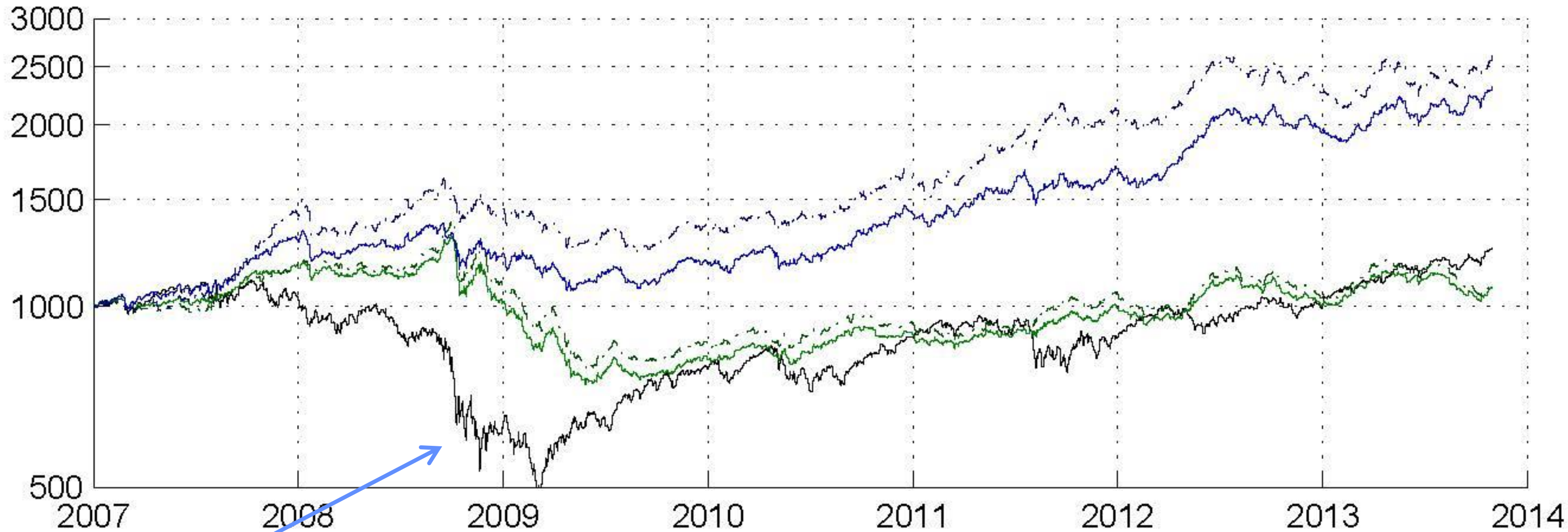
Test/prediction/portfolio decisions period: 1/2007 – 10/13
Hypothetical trading cost: 20bp

Strategy	Constraints
$1	target return $\tau_t = 10\%/252$
$2	target return $\tau_t = 15\%/252$
$3	target return $\tau_t = \hat{\mu}_{\text{SPX},t} + 5\%/252$
$4	SPX neutral, target return $\tau_t = 10\%/252$
$5	SPX neutral, target return $\tau_t = 15\%/252$
$6	SPX neutral, target return $\tau_t = \hat{\mu}_{\text{SPX},t} + 5\%/252$

S&P SGDLMs and portfolios: Cumulative returns

Strategy	\$1	\$2	\$3	\$4	\$5	\$6
WDLM 1	−0.96%	−0.46%	0.05%	−1.13%	−1.04%	−1.28%
SGDLM 1	10.09%	9.94%	9.47%	10.18%	11.65%	13.45%
WDLM 2	−0.34%	−0.15%	0.96%	−0.71%	−0.22%	1.20%
SGDLM 2	8.63%	8.98%	6.74%	9.14%	8.64%	6.60%
WDLM 3	0.43%	0.36%	−2.08%	−0.38%	−1.08%	−3.46%
SGDLM 3	7.11%	8.16%	6.22%	7.83%	7.07%	9.33%
WDLM 4	−0.12%	−0.39%	−0.21%	−0.11%	−0.30%	−1.14%
SGDLM 4	10.91%	11.43%	11.76%	12.44%	12.64%	13.36%
SGDLM 5	11.25%	12.19%	6.31%	12.13%	11.77%	11.26%

Targets undershot or overshot 2007-2013



SPX

Best performing in model+decision rule classes

Portfolios $3 (solid) and $6 (dash-dot)
WDLM 2 SGDLM 4

SGDLM decoupling/recoupling and computations

- *Flexible, customisable univariate models: decoupled, in parallel*

- *Simultaneous parental structure: flexible, sparse - & improved - multivariate stochastic volatility models*

- *No series-order dependence*

- *Scale-up : conceptual & computational*

- *On-line sequential learning: Analytics + simulation + VB*
    - *all parallelisable per step in time*

- *Exploit multi-core, GPU implementation for technical scale-up*

- *Recoupling for coherent posterior and predictive inferences*
    - *importance sampling recouples: return to CPU*

Zoey Zhao PhD (Duke 2015)
Quantitative Researcher in Statistical Arbitrage Trading
Citadel, Chicago

Dynamic dependence networks: Financial time series forecasting & portfolio decisions, *ASMBI, 2015 (to appear)*

Lutz F Gruber PhD (TUM 2015)
Visiting Assistant Professor in Statistical Science
Duke University

GPU-accelerated Bayesian learning in simultaneous graphical dynamic linear models, *Bayesian Analysis, 2015*
Bayesian forecasting and portfolio decisions using simultaneous graphical dynamic linear models, *submitted*

S&P example – Simultaneous parental set specification

Bayesian selection &/or model averaging:

Parallel stochastic model search

Relevant/worthwhile?
- *prediction focus*
- *major collinearities*

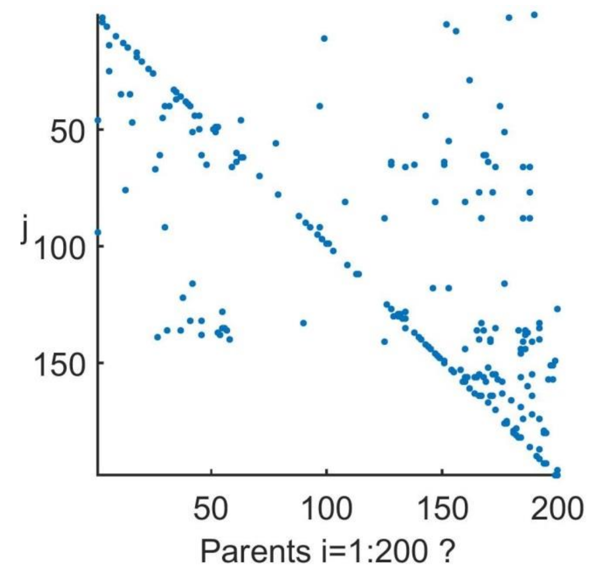
Example here - (ad-hoc) Bayesian "hotspot"

- choose 10-20 parents per series
- Update/refresh periodically (end of each year, 252 days)

How?

Run Wishart discount DLM in parallel to SGDLM

Quick simulation at any time point:
- 20 "most important" parents
- Posterior ranks on abs(precision matrix entries)



S&P example – Simultaneous parental sets

2007	2008	2009	2010	2011	2012	2013
SPX	SPX	SPX	SPX	SPX	SPX	SPX
C	ABT	AAPL	AAPL	AAPL	ABC	BCR
EBAY	AEE	ABT	AEE	BCR	AXP	BF/B
EQR	AFL	APD	BEN	CPB	BCR	BMY
EXC	APD	BEN	BLL	DOV	BK	CPB
HES	BBT	BLL	DOV	EBAY	CPB	EBAY
HRS	BHI	EBAY	EBAY	EIX	D	INTC
IFF	COP	EIX	EIX	HCN	DNB	JNJ
MKC	DOV	HES	ETR	INTC	DOV	MCK
MRO	EBAY	JNJ	HD	JPM	EBAY	MSFT
MTB	GIS	MAS	INTC	KO	INTC	NI
NSC	LMT	MKC	L	L	KO	OMC
PCP	NSC	MTB	MSFT	MSFT	MCK	PCLN
PEP	PX	NEE	NEE	NI	MSFT	PG
PX	SIAL	PG	NSC	PG	NI	RSG
SIAL	SO	SIAL	PKI	PKI	PG	SO
SO	TEG	SO	T	TMO	SO	TROW
STI	TMO	TEG	TMO	TROW	TMO	TRV
T	UPS	UNP	VZ	VTR	XEL	XEL
YHOO	XL	UPS	XEL	XEL	XOM	XOM

sp(Amazon) in SGDLM 1:

Turn-over – red parents switched out at end of year reassessment