Multivariate Time Series:
Forecasting, Decisions, Structure & Scalability

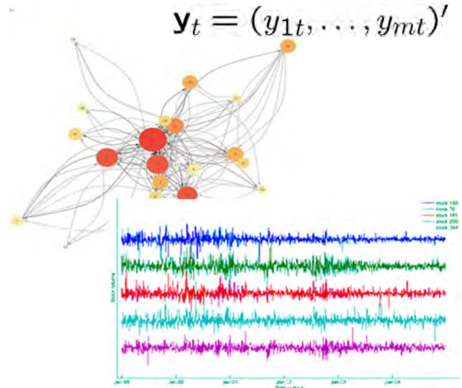
Mike West
Duke University

Time series/dynamic data modelling: Contexts

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

- Challenges -

Flexible multivariate models
Computationally feasible
Scale up in m



- Increasingly large-scale:
  - High-dimensional time series
  - Dynamic networks
  - Large-scale hierarchical systems

- Sequential analysis, forecasting, decisions:
  - Financial portfolios
  - Multi-step macroeconomics
  - Monitoring networks- change/anomaly detection
  - Large-scale commercial sales forecasting
  - Business/corporate financial flows, investment decisions

Bayesian forecasting R&D: Finance & economics

CITADEL
Bank of America
citi
charles SCHWAB
Goldman Sachs
Deutsche Bank
UBS
nikko am
Nikko Asset Management
GROUPE Caisse des Dépôts

Multinationals & boutiques–
- Management Investment
- Global portfolios
- Bayesian econometrics
- Risk management

Bayesian forecasting
& dynamic modelling

Decision analysis
Portfolios
Risk management
Policy/Advisory

Morgan Stanley
HIGHBRIDGE
Capital One
BEST FOUNDATION

Hedge funds ($bb+)

Merrill Lynch
CNA Financial
Corporation
Bank of Japan
The Federal Reserve Board
Point72 Asset Management
financial engines
QMS Capital Management LP
Tudor Capital

Dynamic modelling and forecasting R&D: Industry, Labs

Bayesian forecasting
& dynamic modelling

Commerce
Energy
IT/internet
Economics/Markets

Causality

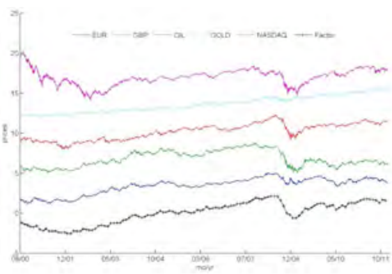
at&t
Google research
GE Global Research
facebook
IBM
LinkedIn
amazon
84.51°
@WalmartLabs
Valassis
RAND CORPORATION
MAXPOINT

State-space models: Structure, interpretation, time

parameters and parameter processes & precision/volatility

vector time series

$$\mathbf{y}_t = \mathbf{F}_t(\mathbf{X}_t, \mathbf{\Theta}_t) + N(\mathbf{0}, \mathbf{\Omega}_t^{-1})$$



State evolution:

$$p(\mathbf{\Theta}_t, \mathbf{\Omega}_t | \mathbf{\Theta}_{t-1}, \mathbf{\Omega}_{t-1})$$

Sequential analysis: step-ahead predictions:

$$p(\mathbf{y}_{t+1:t+k} | \mathbf{y}_{1:t})$$

- enabled by dynamic models/learning: $p(\mathbf{\Theta}_*, \mathbf{\Omega}_* | \mathbf{y}_{1:t})$
- tracking & short-term prediction of volatility dynamics
- relevant partitioning/attributing variation
- open to intervention

Sequential dynamic modelling context

Sequential : model specification
: forecasting
: control/intervention
: learning/updating

Dynamic (linear) state space models

Hidden Markov models

$$\begin{array}{ccccccc} & \mathbf{y}_{t-1} & & \mathbf{y}_t & & \mathbf{y}_{t+1} & \\ & \uparrow & & \uparrow & & \uparrow & \\ \longrightarrow & \mathbf{\Theta}_{t-1} & \longrightarrow & \mathbf{\Theta}_t & \longrightarrow & \mathbf{\Theta}_{t+1} & \longrightarrow \end{array}$$

Closed form analytics:

DLMs: Multi- (or matrix-) normal states/inverse Wishart volatility matrices

Sequential learning & forecasting

$\mathcal{D}_t = \{\mathbf{y}_{1:t}, \oplus_t\}, \ \forall t$

Information sets

$\mathbf{y}_t$

$p(\mathbf{y}_{t:t+k}|\mathcal{D}_{t-1})$ Forecast

$p(\mathbf{y}_t|\Theta_t)$

$\Theta_{t-1} \longrightarrow \Theta_t$

$p(\Theta_t|\Theta_{t-1})$

$\mathcal{D}_{t-1} \rightarrow \mathcal{D}_t$

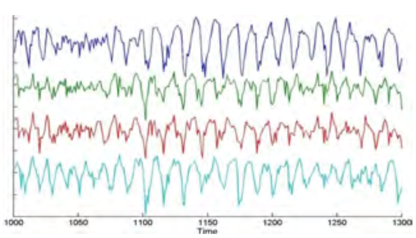
$p(\Theta_{t-1}|\mathcal{D}_{t-1})$ Yesterday's posterior $\longrightarrow$ $p(\Theta_t|\mathcal{D}_{t-1})$ Today's prior $\longrightarrow$ $p(\Theta_t|\mathcal{D}_t)$ Today's posterior $\longrightarrow$

Multivariate time series: TV-VAR models

$$\mathbf{y}_t = \sum_{r=1:d} \mathbf{\Phi}_{rt}\mathbf{y}_{t-r} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{\Omega}_t^{-1})$$

Time-varying lagged connections

Time-varying volatilities and "co-volatilities" in contemporaneous connections

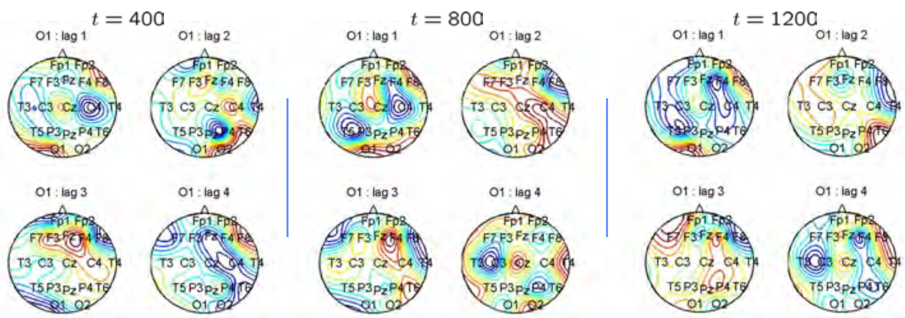


"Workhorse" of much empirical time series analysis

- Macroeconomics (central banks +)
- Natural/engineering sciences (climatology, neuroscience, …)
- Extensible in many directions

Neuroscience: Multiple EEG time series

Coefficients of all series on one chosen series
Dynamic network interconnectivities



TV-VAR(4)

State space models: Exchangeable time series models

$$\mathbf{y}_t = \Theta_t' \mathbf{F}_t + \nu_t, \qquad \nu_t \sim N(\mathbf{0}, \Omega_t^{-1})$$

State matrix

Volatility matrix

Dynamic regression &/or autoregression:

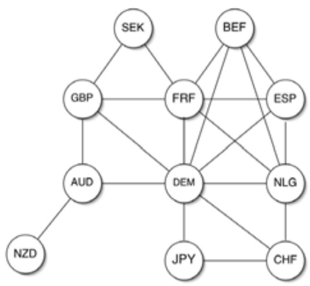
- lagged connections
- exogenous predictors
- dummies for trend, seasonal
- etc

Parameter dimension – BIG with moderate *m, #lags, etc*: scalability?

- Constraints
- Priors
- Parsimony: sparse lagged and cross-sectional dependencies?

A start on sparsity: Graphical volatility models

$$\mathbf{y}_t = \Theta_t' \mathbf{F}_t + \nu_t, \qquad \nu_t \sim N(\mathbf{0}, \Omega_t^{-1})$$

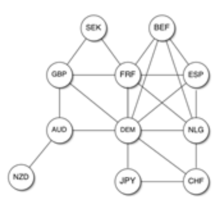


Parsimony:
Conditional independencies
Zeros in precision ~ missing edges

Extend sequential learning & forecasting analytics

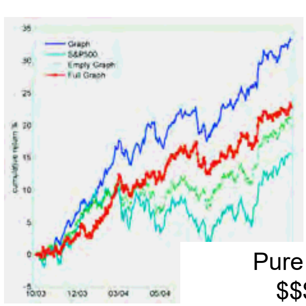
- Hyper-inverse Wishart theory
- Markov volatility on a graph
- Model uncertainty: which graph or graphs?
- Intense computation: model search

Sparsity for structure in dynamic portfolio decisions



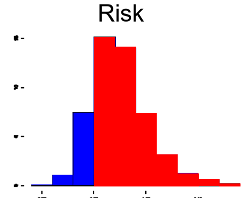
FX rates
Large-scale global equities
Mutual funds

Data "likes" sparse models : Parsimony
- ✓ Improves forecasts
- ✓ Higher realized cumulative returns
- ✓ Lower risk portfolios
- ✓ Lower costs, more stability

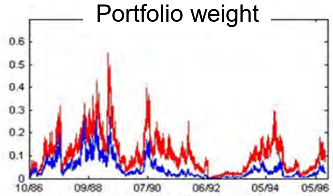


Pure prediction:
$$$ returns

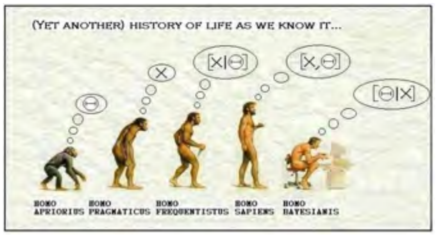
Risk



Portfolio weight



Sparse model
versus
Standard model



Belief analysis★
Modelling, inference, prediction

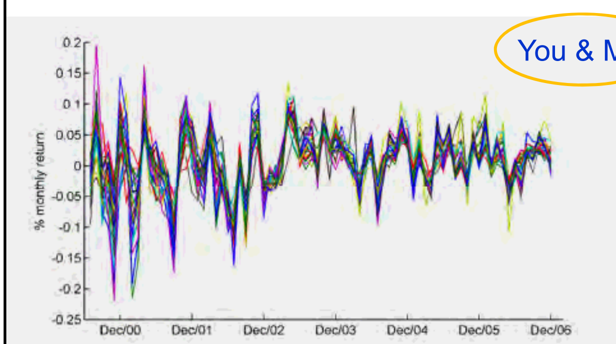
Decision analysis
Utilities, actions, inferential goals

[★ M. Goldstein, discussing D.V. Lindley, Valencia IV, 1992]

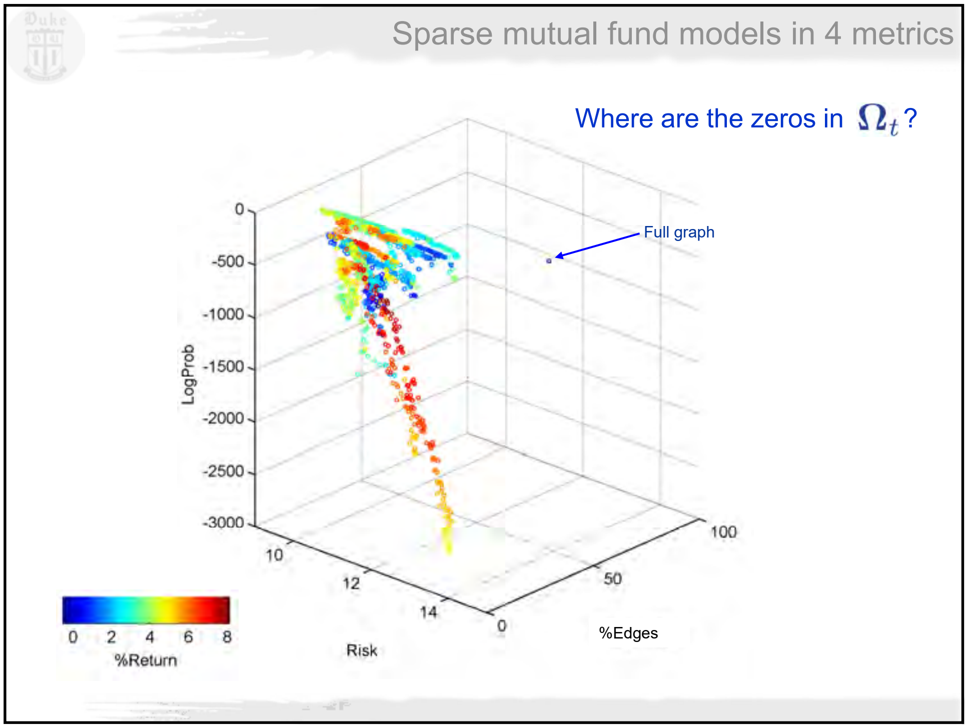
$$\mathbf{y}_t = \mathbf{f}_t + N(\mathbf{0}, \mathbf{\Omega}_t^{-1})$$

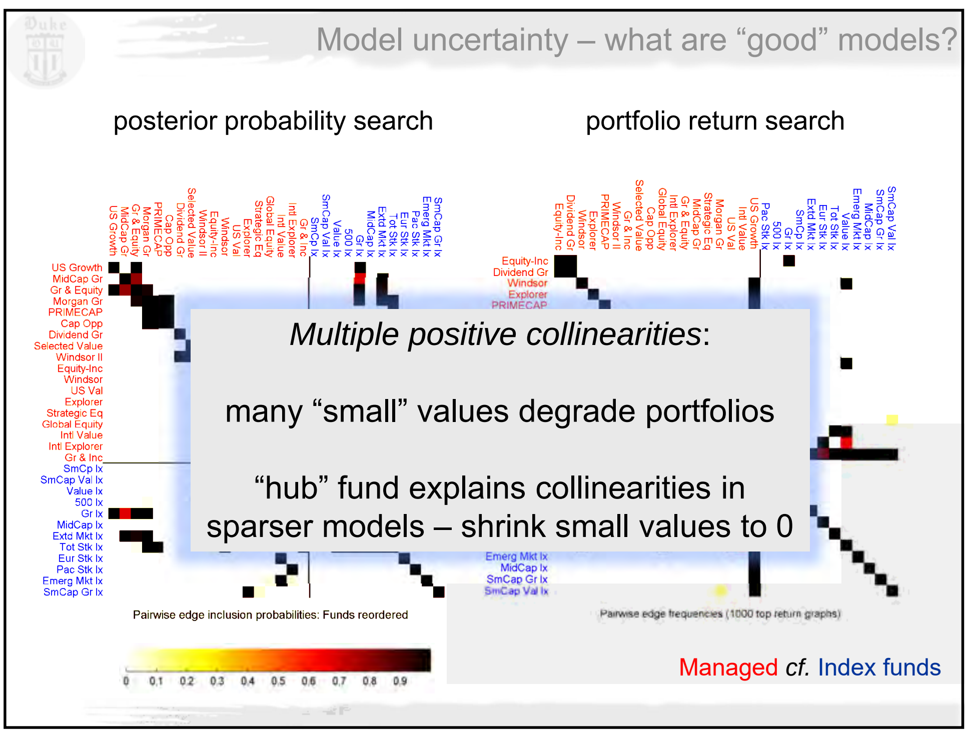
30 Vanguard Mutual Funds
Monthly returns example

$2^{\binom{30}{2}} \approx 10^{131}$ models



You & Me investors: Individual retirement portfolios


Sparse mutual fund models in 4 metrics
Where are the zeros in Ω_t?
Full graph
LogProb
0
-500
-1000
-1500
-2000
-2500
-3000
10
12
14
Risk
0
50
100
%Edges
0 2 4 6 8
%Return


Model uncertainty – what are "good" models?
posterior probability search
portfolio return search
US Growth
MidCap Gr
Gr & Equity
Morgan Gr
PRIMECAP
Cap Opp
Dividend Gr
Selected Value
Windsor II
Equity-Inc
Windsor
US Val
Explorer
Strategic Eq
Global Equity
Intl Value
Intl Explorer
Gr & Inc
SmCp Ix
SmCap Val Ix
Value Ix
500 Ix
Gr Ix
MidCap Ix
Extd Mkt Ix
Tot Stk Ix
Eur Stk Ix
Pac Stk Ix
Emerg Mkt Ix
SmCap Gr Ix
Multiple positive collinearities:
many "small" values degrade portfolios
"hub" fund explains collinearities in sparser models – shrink small values to 0
Pairwise edge inclusion probabilities: Funds reordered
Pairwise edge frequencies (1000 top return graphs)
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Managed *cf.* Index funds

Model comparison, evaluation, selection, combination

GOALS!

More general globally sparse models?

Sparse parameters and parameter processes & precision/volatility

$$\mathbf{y}_t = \mathbf{F}_t(\mathbf{X}_t, \mathbf{\Theta}_t) + N(\mathbf{0}, \mathbf{\Omega}_t^{-1})$$

Bayesian variable selection methods – statistical &/or decision-guided

Sparsity/data-informed parsimony

- *noise control, robust inferences*
- *improved model fit, predictions & decisions (finance/portfolios)*

More general classes of models:

- Customized univariate models?
- Dynamic latent factor models?

Dynamic regression, autoregression & latent factors

$$\mathbf{y}_t = \mathbf{F}_t(\mathbf{X}_t, \boldsymbol{\Theta}_t) + N(\mathbf{0}, \boldsymbol{\Omega}_t^{-1})$$

$$\begin{aligned}
\mathbf{F}_t(\mathbf{X}_t, \boldsymbol{\Theta}_t) &= \mathbf{c}_t && \text{local trends, seasonals} \\
&\pm \mathbf{A}_t \mathbf{z}_t && \text{dynamic regression} \\
&\pm \textstyle\sum_{j=1:d} \mathbf{B}_{jt} \mathbf{y}_{t-j} && \text{TV-VAR(d)} \\
&\pm \mathbf{E}_t \boldsymbol{\phi}_t && \text{dynamic latent factors}
\end{aligned}$$

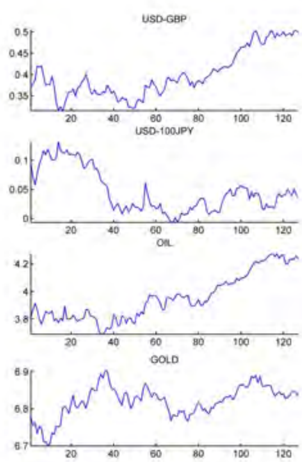
$$\phi_t \sim \text{TV-VAR}(\bullet)$$

e.g. Macro-economics: Factor-augmented TV-VAR models

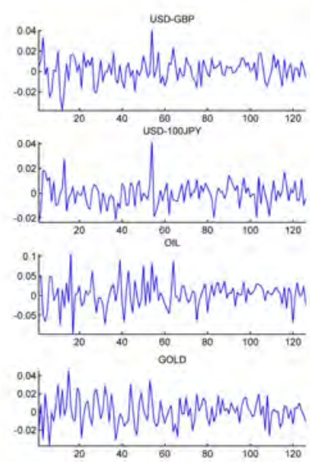
Potential sparsity in $\mathbf{A}_t$, $\mathbf{B}_{(1:d)t}$, $\mathbf{E}_t$, $\boldsymbol{\Omega}_t$?

Example: Commodities & FX volatility modelling

Daily $ US FX and commodities: Jan-Jun 2009



GBP	British Pound Sterling
EUR	Euro
JPY	Japanese Yen
CAD	Canadian Dollar
AUD	Australian Dollar
CHF	Swiss Franc
NOK	Norwegian Krone
NZD	New Zealand Dollar
ZAR	South African Rand
OIL	Brent Crude future
GOLD	Comex Gold future



Introduction to decouple/recouple in example

Dynamic regression, latent factors & volatility
- possible model structure?

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{A}_t\mathbf{z}_t + \mathbf{E}_t\phi_t + N(\mathbf{0}, \mathbf{\Lambda}_t^{-1})$$

local level

dynamic regression
Oil, Gold

VAR(1) dynamic latent factor
- independent components -

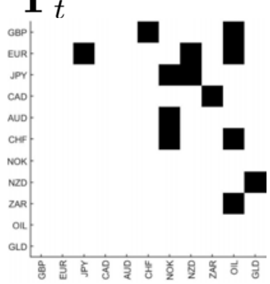
diagonal:
univariate SV models

Introduction to decouple/recouple in example

Dynamic regression, latent factors & volatility
- with contemporaneous dependencies

$$(\mathbf{I} - \mathbf{\Gamma}_t)\mathbf{y}_t = \mathbf{c}_t + \mathbf{A}_t\mathbf{z}_t + \mathbf{E}_t\phi_t + N(\mathbf{0}, \mathbf{\Lambda}_t^{-1})$$

$\mathbf{\Gamma}_t$



Structural model form:
- partly decouples univariate models
- series-specific customization
- flexibility and dynamics
- partial decoupling aids model fitting, Bayesian computation

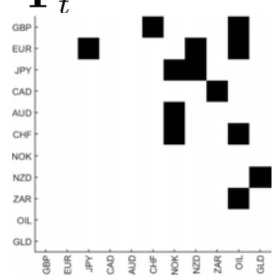
Introduction to decouple/recouple in example

Dynamic regression, latent factors & volatility
- with contemporaneous dependencies

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

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

$\boldsymbol{\Gamma}_t$

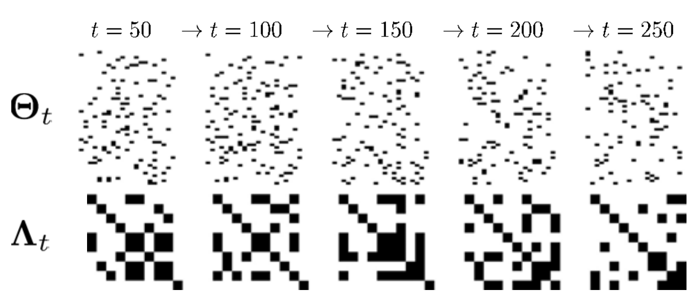


Reduced model form:
- recouples to multivariate model
- cross-talk across series predictors
- Cholesky-style matrix volatility

Identifying structure: Global vs dynamic sparsity

$$\mathbf{y}_t = \mathbf{F}_t(\mathbf{X}_t, \boldsymbol{\Theta}_t) + N(\mathbf{0}, \boldsymbol{\Omega}_t^{-1})$$

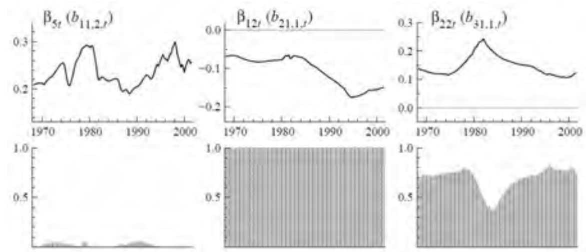
Time-varying SPARSITY PATTERNS?



Dynamic sparsity: A general modelling approach

Latent threshold models:

~ Extended Markov models for states ~
dynamic "variable selection"

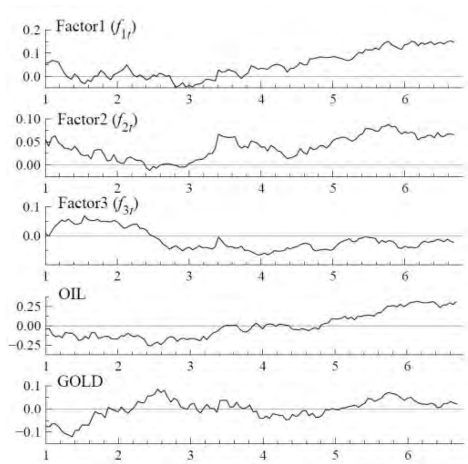


Posterior mean

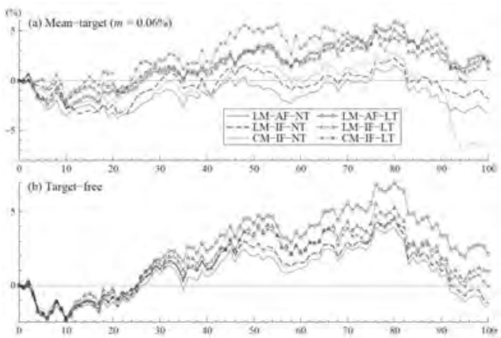
Posterior probability of zero

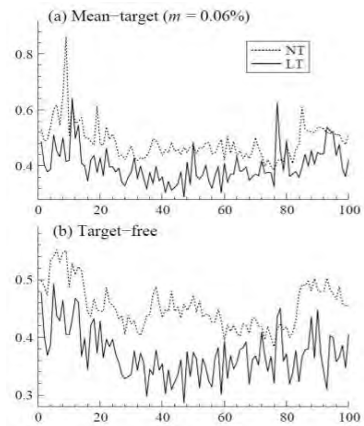
Commodity & FX example: Dynamic factors

$$\cdots + \mathbf{E}_t \phi_t + \cdots$$



Commodity & FX example: Portfolio decisions-returns





Structured sparsity/parsimony improves:

- *forecasts & characterisation of multivariate volatility*
- *portfolio decisions fed by forecasts*
  : realized returns & risk profiles

Bayesian computation: Filtering, forecasting, retrospection

BIG MCMC!

: forward filtering, backward sampling-
layered conditional linear, Gaussian state space models
: coupled sampling of states, latent factors, parameters, ….

MCMC & REPEAT!

: redo each time step
: possible resampling, importance sampling?

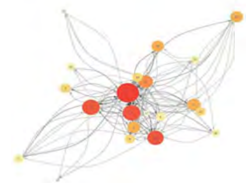
SM

le filtering/learning ?

ABC

Perspective

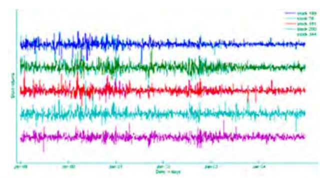
$$\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$$



- o Scale up modelling: Structure, sparsity?
- o Flexible, customizable univariate models?
- o Scale up forward filtering & forecasting ?

- o Revert R&D focus: 95% thinking, 5% computation

- o Maximally exploit analytics ?
- o Partial parallelization ?



Decouple/recouple

1. Macro-economics and finance: Dynamics, volatility, forecasting
   - *m ~ 10-100s+*
2. Network flows: Dynamics, characterizing variation, monitoring
   - *$m=n^2$, n ~ 100s-1,000s*
3. Commercial forecasting: Dynamics, sparsity, hierarchies
   - *m ~ 1,000s-10,000s+*

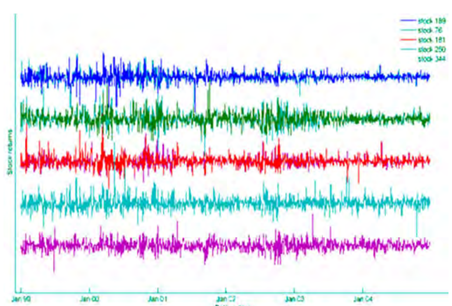
Dynamic Dependency Network Models

Mike West
Duke University

Scale-up Bayesian forecasting

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

Aims:



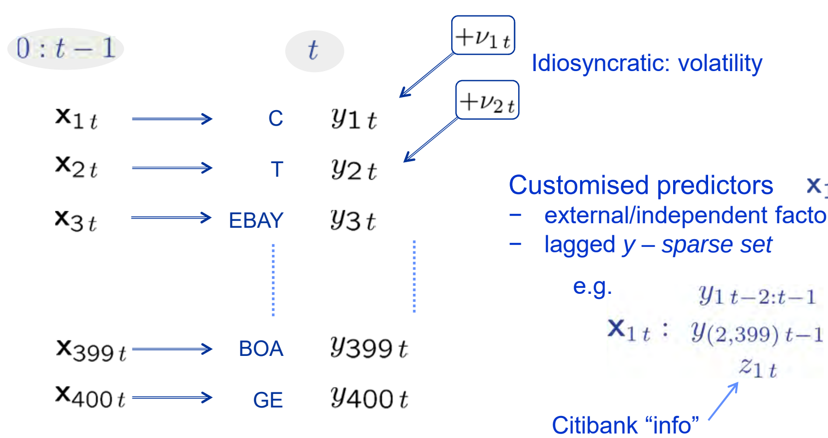
- Scale up in m
  - sequential analysis & forecasting
- "Decoupled" univariate time series
  - o Series-specific models
  - o Sequential analysis: Fast? Parallelisation?
- "Recoupled" multivariate dynamic model - critical cross-series structure
  - o Sensitive co-volatility modelling
  - o Coherent joint forecast distributions

Multiregressions, time-varying VAR models & DDNMs

- Structured subclass of TV-VAR models +
- Econometrics: structural form "recursive" systems
- Extensions & generalizations of "multiregression DLMs"
  - C.M. Queen & J.Q. Smith (RSS B, 1993; and later)
- Cholesky-style multivariate stochastic volatility (MSV)
  - G. Primiceri (2005)
  - H. Lopes, R. McCulloch & R.Tsay (2010 original)
  - J. Nakajima & MW (2013[2],14, 15,17)
    - * *TV-VAR, latent thresholds, factor models:*
    - * *BIG MCMC analyses!*

Decoupled, parallel univariate models

Cross-series: directed & sparse graphical structure – *lagged*



Contemporaneous coupling

Cross-series: *contemporaneous* graphical structure

$0:t-1$ $\quad$ t

$\mathbf{x}_{1\,t} \rightarrow$ C $\quad y_{1\,t}$?

$\mathbf{x}_{2\,t} \rightarrow$ T $\quad y_{2\,t}$

$\mathbf{x}_{3\,t} \rightarrow$ EBAY $\quad y_{3\,t}$

$\vdots$

$\mathbf{x}_{399\,t} \rightarrow$ BOA $\quad y_{399\,t}$

$\mathbf{x}_{400\,t} \rightarrow$ GE $\quad y_{400\,t}$

e.g.

$\leftarrow y_{(2,399)\,t}$

$\leftarrow y_{(7,29,131)\,t}$

$\vdots$

$$y_{jt} \leftarrow \mathbf{y}_{pa(j),t}$$

$$pa(j) \in \{(j+1):m\}$$

Sparse sets: parental predictors

Univariate dynamic linear models

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

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

Known predictors $\quad$ *"Parents"*

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

Parental sets:

$$pa(j) \in \{j+1, j+2, \ldots, m\}$$

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

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

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

Dynamic dependency network models

$$y_{jt} = \mathbf{F}'_{jt}\boldsymbol{\theta}_{jt} + \nu_{jt}$$
$$= \mathbf{x}'_{jt}\boldsymbol{\phi}_{jt} + \mathbf{y}'_{pa(j),t}\boldsymbol{\gamma}_{jt} + \nu_{jt}$$

$pa(j) \in \{j+1, j+2, \ldots, m\}$

Joint pdf: $p(\mathbf{y}_t) = \prod_{j=1:m} q_j(y_{jt}|\mathbf{y}_{pa(j),t})$

Normal pdf in conditional model j

Compositional form of joint density
- directed graphical model -

- *Known* $\mathbf{x}_{jt}$ *& variances:* *Multiregression DLM*
- *TV-VAR with lagged* $\mathbf{y}_{\cdot,t-r}$ *in* $\mathbf{x}_{jt}$*, & unknown volatility processes:*

Dynamic dependency networks

Recoupling: Coherent multivariate model

$$y_{jt} = \mathbf{x}'_{jt}\boldsymbol{\phi}_{jt} + \mathbf{y}'_{pa(j),t}\boldsymbol{\gamma}_{jt} + \nu_{jt}$$

μ_{jt}

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

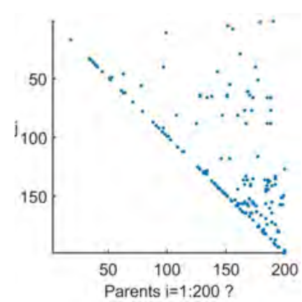
dynamic regressions & precision/volatility

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

- *Multivariate volatility (precision) matrix :*
  - Cholesky-style: upper triangular form

$$\boldsymbol{\Gamma}_t = \begin{pmatrix} 0 & \gamma_{12t} & \gamma_{13t} & \cdots & \gamma_{1mt} \\ 0 & 0 & \gamma_{23t} & \cdots & \gamma_{2mt} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \gamma_{m-1,mt} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

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



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

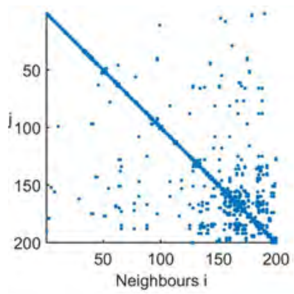
Sparse

Dynamic graphical model structure: DAG to graph

precision/volatility

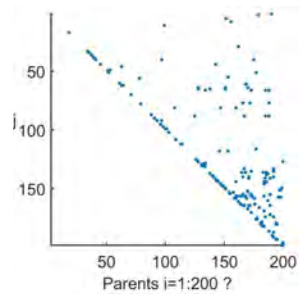
$$\mathbf{y}_t = (\mathbf{I}-\boldsymbol{\Gamma}_t)^{-1}\boldsymbol{\mu}_t + N(\mathbf{0}, \boldsymbol{\Omega}_t^{-1})$$

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



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

zero precision:
conditional
independence



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

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda}_t - \boldsymbol{\Gamma}_t'\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t\boldsymbol{\Gamma}_t + \boldsymbol{\Gamma}_t'\boldsymbol{\Lambda}_t\boldsymbol{\Gamma}_t$$

Links to & from parents

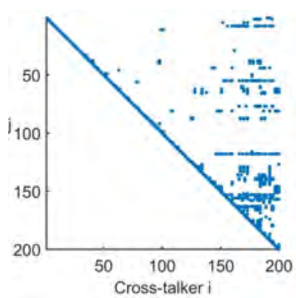
Moralization

DAG to graph

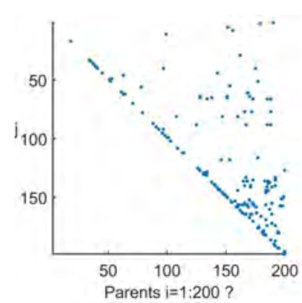
Recoupling: Parental-induced cross-talk in regressions

dynamic regressions

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



Non-zeros in $(\mathbf{I}-\boldsymbol{\Gamma}_t)^{-1}$



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

$$(\mathbf{I}-\boldsymbol{\Gamma}_t)^{-1} = \mathbf{I} + \boldsymbol{\Gamma}_t + \boldsymbol{\Gamma}_t^2 + \boldsymbol{\Gamma}_t^3 + \cdots$$

parents
grandparents
great-grandparents
…….

"Triangular" system of DLMs

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

$$y_{jt} = \mathbf{F}_{jt}'\boldsymbol{\theta}_{jt} + \nu_{jt} = \mathbf{x}_{jt}'\boldsymbol{\phi}_{jt} + \mathbf{y}_{pa(j),t}'\boldsymbol{\gamma}_{jt} + \nu_{jt}$$

$$\mathbf{F}_{jt} = \begin{pmatrix} \mathbf{x}_{jt} \\ \mathbf{y}_{pa(j),t} \end{pmatrix} \qquad \boldsymbol{\theta}_{jt} = \begin{pmatrix} \boldsymbol{\phi}_{jt} \\ \boldsymbol{\gamma}_{jt} \end{pmatrix}$$

Parallel states - *independent evolutions*

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

e.g. $\mathbf{G}_{jt} = \mathbf{I}, \mathbf{W}_{jt} = \text{diag}(\mathbf{W}_{\phi,jt}, \mathbf{W}_{\gamma,jt})$

2 discount factors, series specific

DDNMs: Filtering and forecasting

Parallel, conditionally independent analyses:

$$y_{jt} = \mathbf{F}_{jt}'\boldsymbol{\theta}_{jt} + \nu_{jt}$$
$$\boldsymbol{\theta}_{jt} = \boldsymbol{\theta}_{j,t-1} + \boldsymbol{\omega}_{jt}$$

$$\boldsymbol{\theta}_{jt}|\lambda_{jt}, \mathcal{D}_t \sim N(\mathbf{m}_{jt}, \mathbf{C}_{jt}/(s_{jt}\lambda_{jt}))$$
$$\lambda_{jt}|\mathcal{D}_t \sim G(n_{jt}/2, n_{jt}s_{jt}/2)$$

Example:
– independent, steady state evolutions
– discounting for state and volatility

Forecasting:

- 1-step: multivariate/joint forecast mean and variance matrix
  - analytic: recursively computed
- 1-step:- marginal likelihoods for parameter assessment?
  - require parental predictor values -
  - series *j* specific (discount factors, TVAR model order, etc …)
- And forecasting more steps ahead?

DDNMs: Recoupling in multi-step forecasting

Forecasting:

- TVAR models: Simulate for more steps ahead
- DDNM: Simulate!
  - For steps ahead *k=1:K …*

$$\mathbf{y}_{t+1:t+k} \sim p(\mathbf{y}_{t+1:t+k}|\mathcal{D}_t)$$

$$y_{1,t+k} \leftarrow p(y_{1,t+k}|\mathbf{y}_{pa(1),t+k}, \mathcal{D}_{t+k-1})$$

$$y_{j,t+k} \leftarrow p(y_{j,t+k}|\mathbf{y}_{pa(j),t+k}, \mathcal{D}_{t+k-1})$$

$$y_{m-1,t+k} \leftarrow p(y_{m-1,t+k}|\mathbf{y}_{pa(m-1),t+k}, \mathcal{D}_{t+k-1})$$

$$y_{m,t+k} \leftarrow p(y_{m,t+k}|\mathcal{D}_{t+k-1})$$

DDNMs: Summary

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

- Multiple univariate models: decoupled, in parallel

- Conducive to on-line sequential learning: Analytic, fast, parallel

*- Simultaneous parental sets define *sparse* multivariate stochastic volatility matrix*

- New dynamic graphical models for MSV: evolutions of $\boldsymbol{\Gamma}_t, \boldsymbol{\Lambda}_t$

- Simulation for forecasting: exploit recursive recoupling

*- * Dependent on choice of order of named series j=1:m **

ASMBI example: FX, commodities, stock indices

Daily closing prices in $US
Model log(price)

j	Name	Asset
1	CHF	Swiss Franc
2	EUR	Euro
3	NSD	NASDAQ Composite Index
4	S&P	S&P 500 Index
5	NOK	Norwegian Krone
6	GBP	British Pound

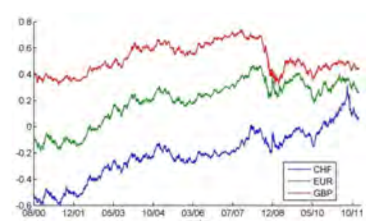
j	Name	Asset
7	AUD	Australian Dollar
8	NZD	New Zealand Dollar
9	ZAR	South African Rand
10	GOL	Gold
11	CAD	Canadian Dollar
12	JPY	Japanese Yen
13	OIL	Crude Oil

ASMBI example: Models, analyses, decisions

50:50 training/test data:

- Train (to mid 4/2006):
  evaluate all $\mathcal{M}_j,\ j = 1{:}m$

- Test (from mid 4/2006):
  use reduced subsets of models



Forecasting and decisions: 5-day look-ahead portfolio

- Commodity Trading Advisors' (CTA) benchmark ("market index")
  Model *m+1* series with CTA included
- Optimization of portfolio: risk/return utility function

ASMBI example: Model structure and hyper-parameters

$$y_{jt} = \mu_{jt} + \sum_{r=1:p_j} \phi_{jrt} y_{j,t-r} + \sum_{k \in pa(j)} \gamma_{jkt} y_{kt} + \nu_{jt}$$

$\mathcal{M}_j = \{ \; p_j,$ TVAR order

$pa(j),$ parental set

$* \; \}$ hyperparameters

Discrete model space:

$\mathcal{M}_j \in \{\mathcal{M}_j^1, \mathcal{M}_j^2, \cdots\}$

- all possible parental sets
- discrete ranges of TVAR order and parameters

Sequential learning over time: adaptive discount learning model probabilities

$$Pr(\mathcal{M}_j | \mathcal{D}_t)$$

Parallel, analytic computation

DDNM Decoupled model uncertainty analysis

Sequential learning over time:

$$Pr(\mathcal{M}_j|\mathcal{D}_t) \propto Pr(\mathcal{M}_j|\mathcal{D}_{t-1}) \; p(y_{jt}|\mathbf{y}_{pa(j),1:t-1}, \mathcal{D}_{t-1}, \mathcal{M}_j)$$

$\mathcal{M}_j \in \{\mathcal{M}_j^1, \mathcal{M}_j^2, \cdots\}$

$j = 1:m$

Parallel!

- across all models for each series j
- independently/parallel across series $j = 1:m$

- computationally accessible
- model space dimension reduction:

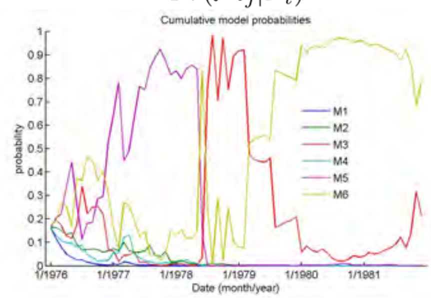
$$\prod_{j=1:m} \#\{\mathcal{M}_j^r\} \quad \longrightarrow \quad \sum_{j=1:m} \#\{\mathcal{M}_j^r\}$$

But first: More time-adaptive structure learning

Model probabilities converge/degenerate

- Sometimes fast
- Always to "wrong" model
- Lack of adaptability over time

$Pr(\mathcal{M}_j|\mathcal{D}_t)$



"Reality" may be outside the "span" of model set?

- all are "poor models"

Power discounting of model uncertainties

$$Pr(\mathcal{M}_j|\mathcal{D}_t) \propto Pr(\mathcal{M}_j|\mathcal{D}_{t-1})^{\alpha}\ p(y_{jt}|\mathbf{y}_{pa(j),1:t-1}, \mathcal{D}_{t-1}, \mathcal{M}_j)$$

Discount ("forgetting") factor

$0 << \alpha \leq 1$

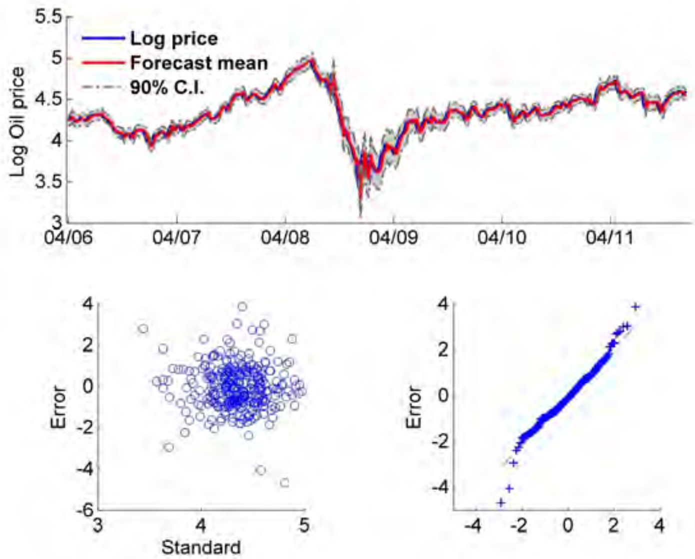
- Exponentially discount past information across models
- "Synthetic" dynamics: Time-varying model structure
- Intervention: Change model probabilities each time step

Extend model uncertainty analysis: grid of α
+ product of model probabilities over *j*

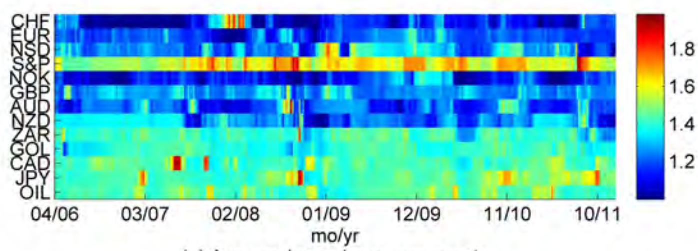
[Raftery et al, 2010
Koop & Korobilis 2012]

[P.J. Harrison, 1960s; West & Harrison, 1987 1st edn]

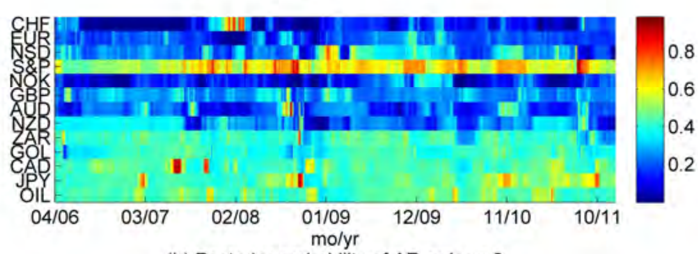
ASMBI example: One series



ASMBI example: TVAR model order by series

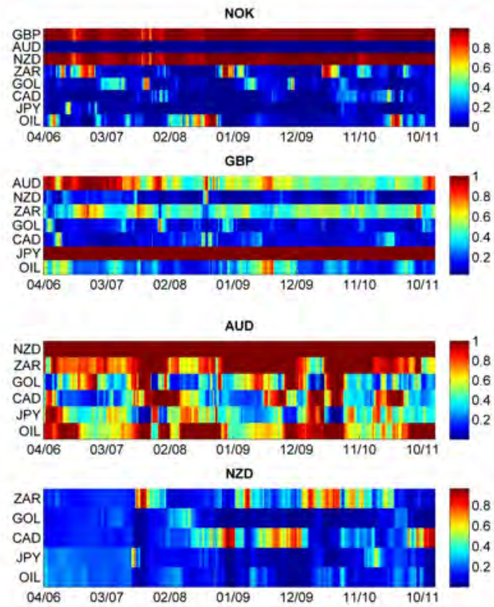


(a) Average lag order across assets

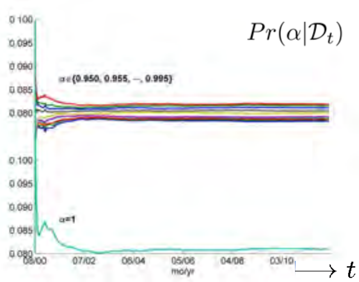


(b) Posterior probability of AR order = 2

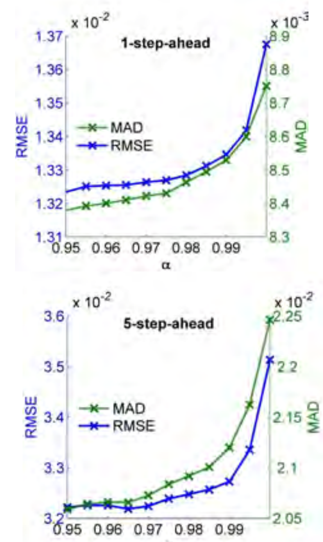
ASMBI example: Some parental inclusion probabilities



ASMBI example: Scoring power discounting



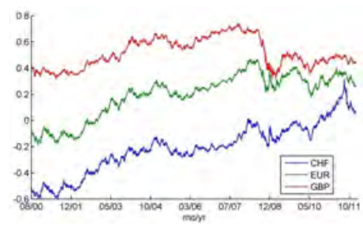
Forecast RMSE & MAD



Improves:
1,5-step forecasts
- adaptability -

Portfolio returns?

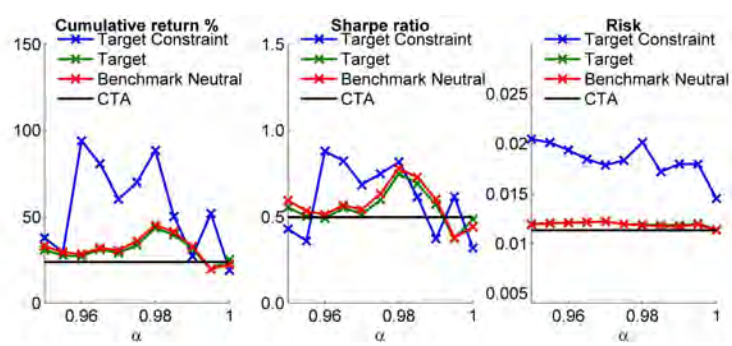
ASMBI example: Portfolio decisions



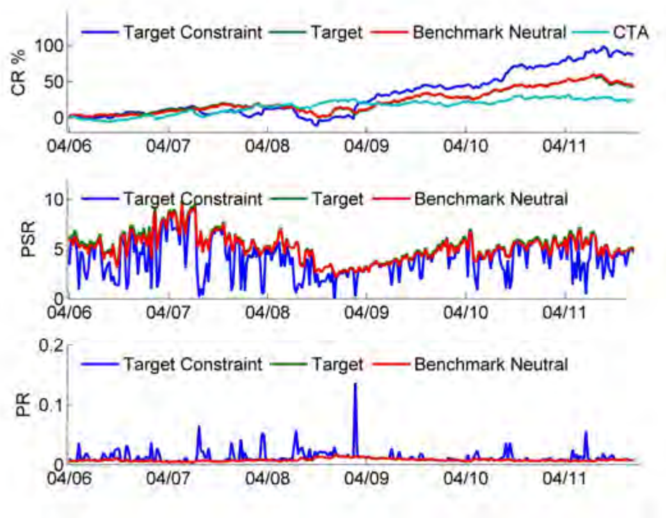
Forecasting and decisions: 5-day look-ahead portfolio

- Commodity Trading Advisors' (CTA) benchmark ("market index")
  - Model *m+1* series with CTA included
- Optimization of portfolio to:
  - minimize predicted portfolio risk (=predictive portfolio return variance)
  - subject to specified %target exceedance over expected/forecast CTA benchmark
  - and decorrelated- in expectation- with CTA benchmark

ASMBI example: 5-day ahead portfolios



ASMBI example: 5-day ahead portfolios



DDNMs

- *Flexible, customisable univariate models: decoupled, in parallel*
- *Parental structure: Dynamic graphical models*
- *On-line sequential learning: Analytic, fast, parallel*
- *Recoupling for coherent posterior and predictive inferences*
- *Simulation-based forecasting*

Dependent on choice of order of named series j=1:m

Choice/specification of order? Parental sets?

Scale-up in dimension? 400 S&P stock price series?

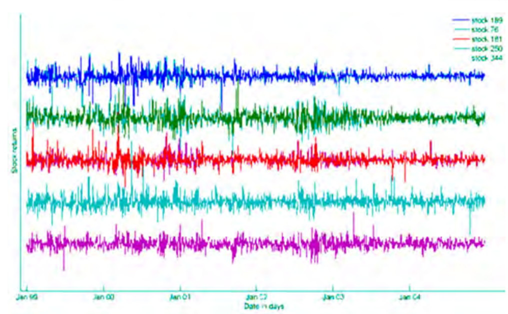
Simultaneous Graphical Dynamic Models

Mike West
Duke University

Concept: Decouple/recouple

- Challenges -

Flexible multivariate models
Computationally feasible
Scale up in m



DECOUPLE - univariate time series

- Customisable series-specific models
- State-space models: analytic tractability
- Sequential analysis: fast? parallelisation?

RECOUPLE - Critical cross-series / multivariate structure

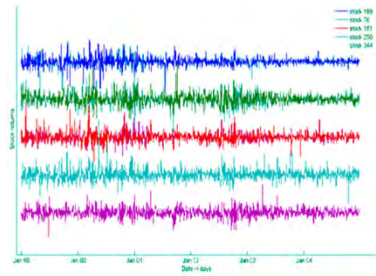
- Sensitive dependence / “co-volatility” modelling
- Sensitive online monitoring, adaptation to change
- Coherent joint forecast distributions, resulting decisions

Multivariate volatility

Financial time series: FX, commodities, stocks ..

Multi-step forecast and portfolio goals

- daily/weekly: sequential, repeat



$$\mathbf{y}_t = \mathbf{f}_t + N(\mathbf{0}, \mathbf{\Sigma}_t)$$

Critical cross-series / multivariate structure

- o Sensitive dependence / "co-volatility" modelling ?

Existing models

- o Wishart or Inverse Wishart processes, MGARCH, …. (In)flexibility?
- o DDNMs …. ordering of series? Parental set selection?
- o Scalable – theoretically? computationally?

Decoupled, parallel univariate dynamic models

FX – daily price $US

Cross-series: directed & sparse *graphical* structure – *lagged*

$0:t-1$ $\quad$ t

$+\nu_{1\,t}$ $\quad$ $+\nu_{2\,t}$ $\quad$ Idiosyncratic: volatility

$\mathbf{x}_{1\,t}$ → ¥ $y_{1\,t}$

$\mathbf{x}_{2\,t}$ → $Au $y_{2\,t}$

$\mathbf{x}_{3\,t}$ → £ $y_{3\,t}$

⋮

$\mathbf{x}_{m-1,t}$ → € $y_{m-1,t}$

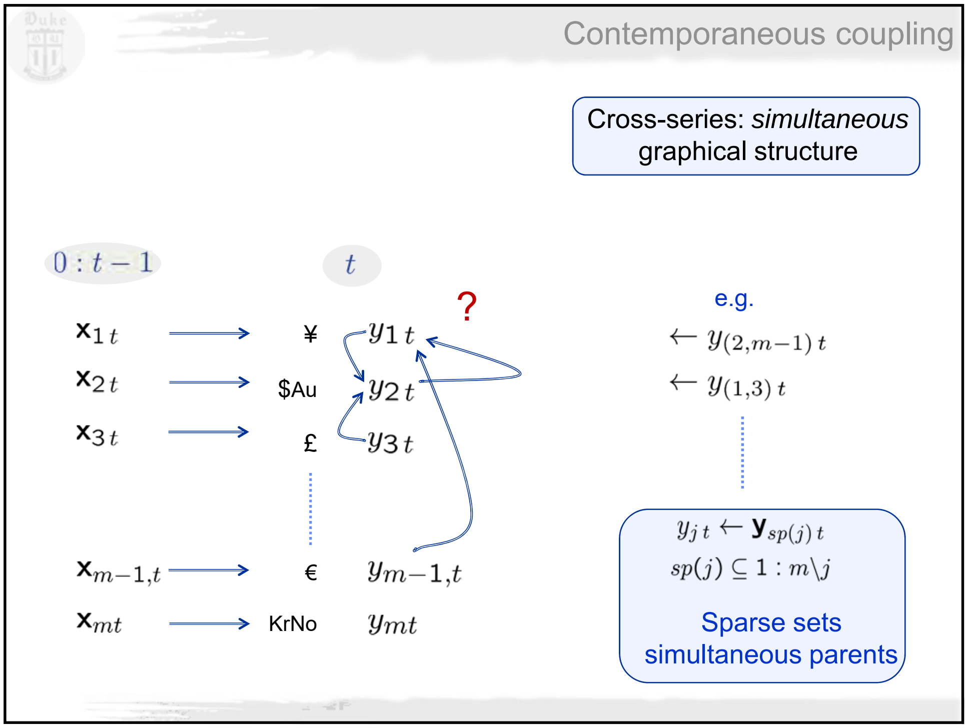
$\mathbf{x}_{mt}$ → KrNo y_{mt}

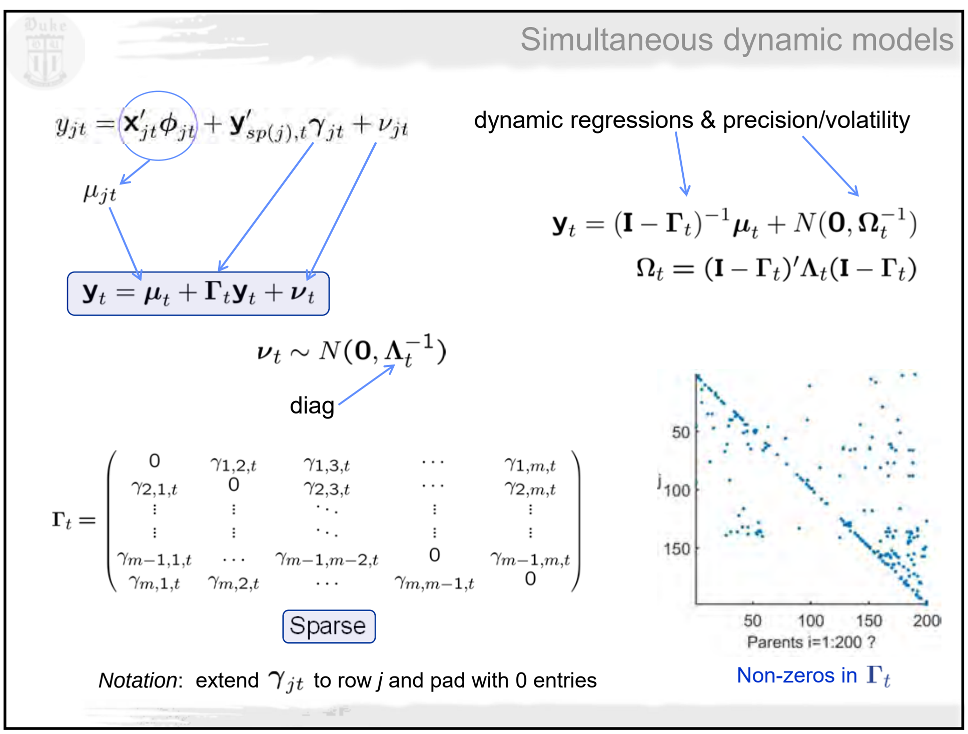
Customised predictors $\mathbf{x}_{1\,t}$

- external/independent factors
- lagged *y* – *sparse set*

e.g. $\mathbf{x}_{1\,t}:$ $y_{1\,t-2:t-1}$, $y_{(2,m-1,m)\,t-1}$, $z_{1\,t}$

¥ "info"


Contemporaneous coupling
Cross-series: *simultaneous* graphical structure
$0 : t-1$
t
?
$\mathbf{x}_{1\,t}$ ¥ $y_{1\,t}$
$\mathbf{x}_{2\,t}$ \$Au $y_{2\,t}$
$\mathbf{x}_{3\,t}$ £ $y_{3\,t}$
$\mathbf{x}_{m-1,t}$ € $y_{m-1,t}$
$\mathbf{x}_{mt}$ KrNo y_{mt}
e.g.
$\leftarrow y_{(2,m-1)\,t}$
$\leftarrow y_{(1,3)\,t}$
$y_{j\,t} \leftarrow \mathbf{y}_{sp(j)\,t}$
$sp(j) \subseteq 1:m \backslash j$
Sparse sets
simultaneous parents

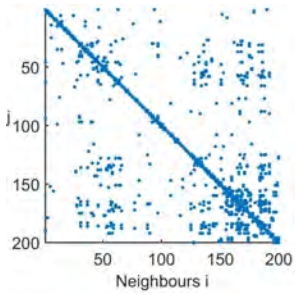

Simultaneous dynamic models
$y_{jt} = \mathbf{x}'_{jt}\boldsymbol{\phi}_{jt} + \mathbf{y}'_{sp(j),t}\boldsymbol{\gamma}_{jt} + \nu_{jt}$
μ_{jt}
dynamic regressions & precision/volatility
$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Gamma}_t\mathbf{y}_t + \boldsymbol{\nu}_t$
$\mathbf{y}_t = (\mathbf{I}-\boldsymbol{\Gamma}_t)^{-1}\boldsymbol{\mu}_t + N(\mathbf{0}, \boldsymbol{\Omega}_t^{-1})$
$\boldsymbol{\Omega}_t = (\mathbf{I}-\boldsymbol{\Gamma}_t)'\boldsymbol{\Lambda}_t(\mathbf{I}-\boldsymbol{\Gamma}_t)$
$\boldsymbol{\nu}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}_t^{-1})$
diag
$\boldsymbol{\Gamma}_t = \begin{pmatrix} 0 & \gamma_{1,2,t} & \gamma_{1,3,t} & \cdots & \gamma_{1,m,t} \\ \gamma_{2,1,t} & 0 & \gamma_{2,3,t} & \cdots & \gamma_{2,m,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{m-1,1,t} & \cdots & \gamma_{m-1,m-2,t} & 0 & \gamma_{m-1,m,t} \\ \gamma_{m,1,t} & \gamma_{m,2,t} & \cdots & \gamma_{m,m-1,t} & 0 \end{pmatrix}$
Sparse
50
j 100
150
50 100 150 200
Parents i=1:200 ?
Non-zeros in $\boldsymbol{\Gamma}_t$
Notation: extend $\boldsymbol{\gamma}_{jt}$ to row *j* and pad with 0 entries

Dynamic graphical model structure induced

precision/volatility

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

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



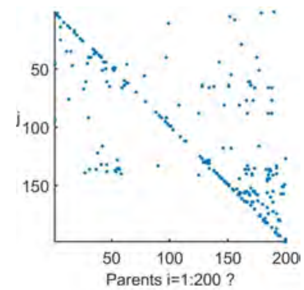
Non-zeros in $\boldsymbol{\Omega}_t$

zero precision: conditional independence

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda}_t - \boldsymbol{\Gamma}_t'\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t\boldsymbol{\Gamma}_t + \boldsymbol{\Gamma}_t'\boldsymbol{\Lambda}_t\boldsymbol{\Gamma}_t$$

Links to & from parents

Moralization



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

Side notes: Modelling interpretation

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

Simultaneous specification: "Structural form" modelling

- Direct, series specific, customized, parallel, efficient

Multivariate model implied: "Reduced form" model

"Over-parametrized"?

- Prediction arbitrates, not parameter interpretation
- In any case, sparsity obviates:

$$\#\{\boldsymbol{\Omega}_t\} > \#\{\boldsymbol{\Gamma}_t, \boldsymbol{\Lambda}_t\} \iff \text{ave}|sp(j)| < (m-1)/2$$

Practical models: small *sp(j)*

Simultaneous dynamic linear/state-space model

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

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

Residual volatilities:
Tractable Markov dynamics

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

$$\boldsymbol{\theta}_{jt} = \begin{pmatrix} \boldsymbol{\phi}_{jt} \\ \boldsymbol{\gamma}_{jt} \end{pmatrix}$$

Parallel states:
Linear, Gaussian, Markov
state evolution models

Special examples: Dynamic dependency network models

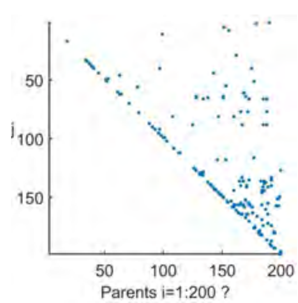
- $\boldsymbol{\Gamma}_t$ *triangular form: by choice or chance*
- *Sparse Cholesky-style precision (… popular)*
  - upper or lower triangular- same story: reorder

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

Order of series relevant

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

Sparse, upper triangular



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

SGDLMs: Simultaneous graphical dynamic models

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

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

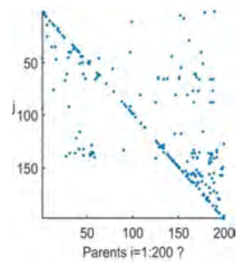
Simultaneous parents of j

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

Joint pdf …
likelihood for states and volatilities:

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

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



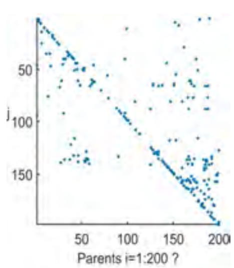
Non-zeros in Γ_t

Decouple/Recouple: Sequential analysis of SGDLM

Basis:

- o Increasingly sparse Γ_t for larger *m*
- o $|\mathbf{I} - \boldsymbol{\Gamma}_t|$ "minor" correction to likelihood

- Triangular: compositional models: $= 1$
- Partly triangular, sparse: ≈ 1

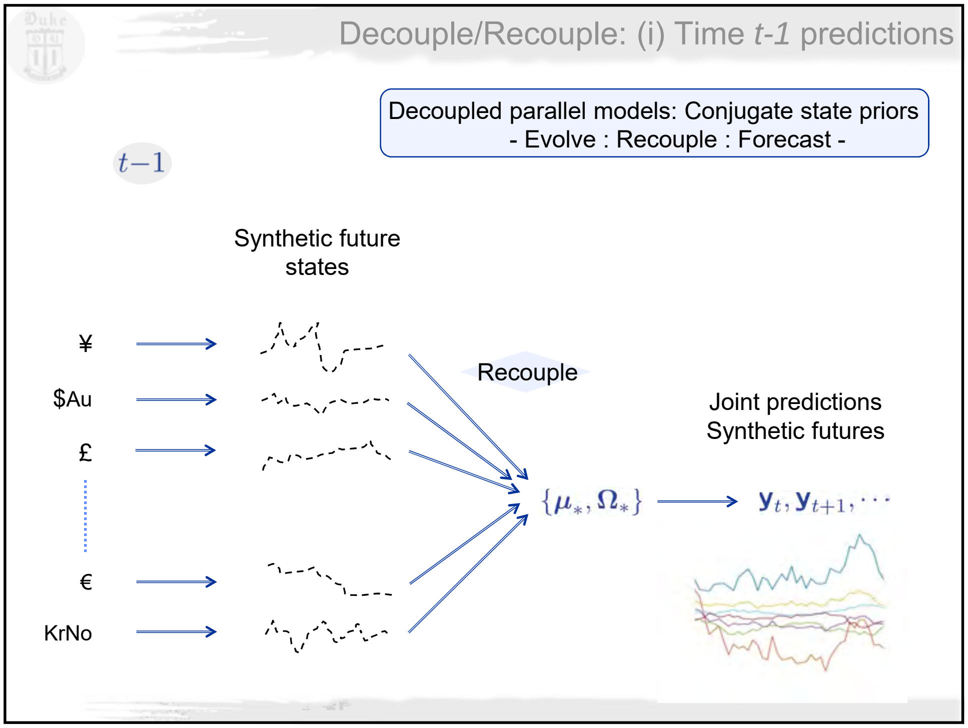


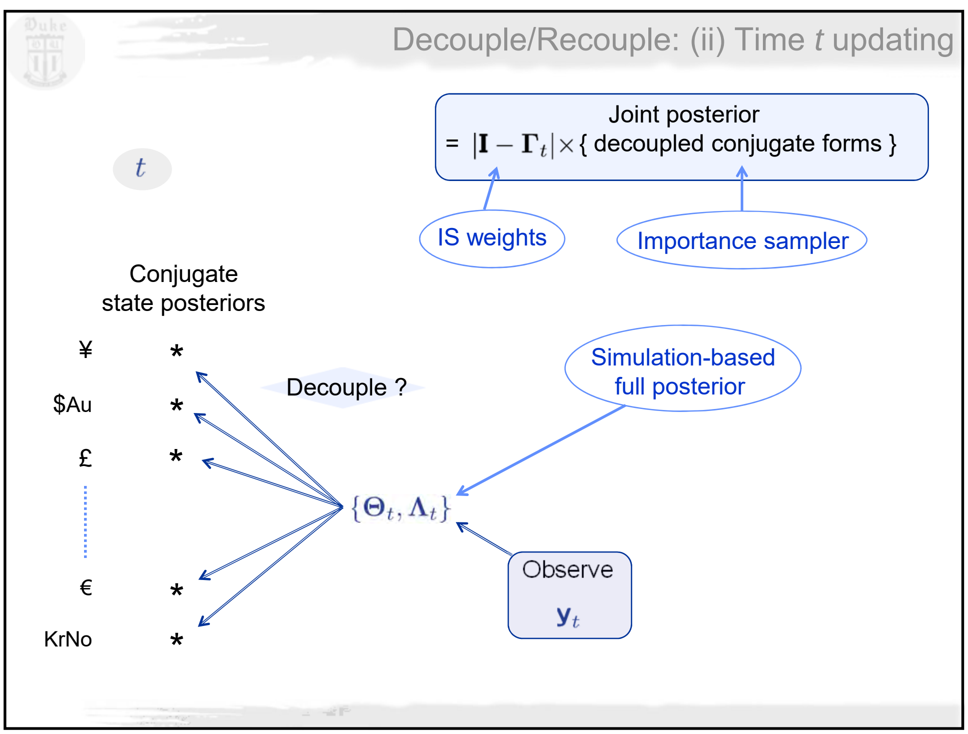
Non-zeros in Γ_t

Strategy:
- exploit DLM analytics in decoupled models
- simulation for forecasting & state inference

Recouple: importance sampling

Decouple: variational Bayes


Decouple/Recouple: (i) Time t-1 predictions
Decoupled parallel models: Conjugate state priors
- Evolve : Recouple : Forecast -
$t-1$
Synthetic future states
¥
$Au
£
€
KrNo
Recouple
Joint predictions Synthetic futures
$\{\mu_*, \Omega_*\}$
$\mathbf{y}_t, \mathbf{y}_{t+1}, \cdots$


Decouple/Recouple: (ii) Time t updating
Joint posterior
= $|\mathbf{I} - \mathbf{\Gamma}_t| \times$ { decoupled conjugate forms }
IS weights
Importance sampler
t
Conjugate state posteriors
¥
$Au
£
€
KrNo
Decouple ?
Simulation-based full posterior
$\{\Theta_t, \Lambda_t\}$
Observe $\mathbf{y}_t$

Decouple/Recouple: Time t decoupling

Joint posterior ⟹ decoupled conjugate margins
Variational Bayes (VB)

Decoupling at t:

Match product of conjugate forms with full posterior IS

- min Kullback-Leibler (KL) divergence (of product from full)
- IS weights ~ "uniform": good IS - sparse Γ_t - good VB

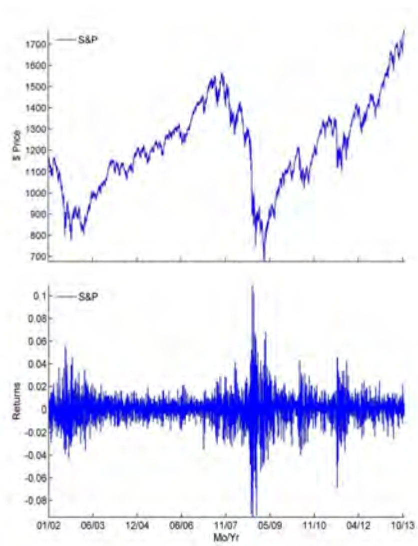
- variation in IS weights:
  ESS (effective sample size)
  Entropy (relative to uniform)
- monitor VB+IS approximation

Evolution to t+1:

- Decoupled, conjugate : Evolution improves Entropy, ESS

Example: S&P stock price series

Multiple models x portfolio utilities comparisons



***m=401**:* *400* stocks & S&P index
- daily closing prices

Q1/2002 – Q3/2013

Model (log-difference) returns

Training data : 2002

Test/prediction/portfolio decisions :
Q1/2003 – Q3/2013

S&P example – Models for comparison

M: $\mathbf{F}_{jt} = 1$

MA: $\mathbf{F}_{jt} = (1, x_{jt})'$

$x_{jt} = 0.2 \sum_{k=1:5} (y_{j,t-k} - f_{j,t-k})$

W: Traditional Wishart discount DLM

- common predictor models only
- "oversmooths" volatilities with "many" series

Various other models based on:

Lagged &/or changes of

- *TNX: 10-year T-bill rate*
- *VIX: implied 30-day volatility index*
- *S&P index*

Simultaneous parental set specification?

AIC / *IC / full Bayesian model selection / model averaging ?

Parallel stochastic search over parental sets ?

Goals? Learning/choosing "best" statistical models ✘

Multi-step forecasting ✓… but

Decisions! Portfolio outcomes (risk, return, …) ✓!

Adaptive "refresh" of parental sets
- monitor: candidates to add, remove each t -

Adaptive updating of simultaneous parental sets

Bayesian selection &/or model averaging:

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

Parallel stochastic model search

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

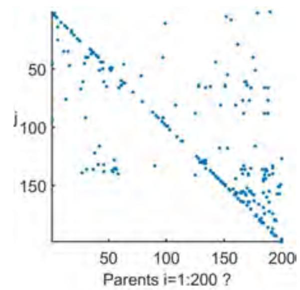
- 10 parents per series
- Adaptively modify: in/out switches

How?

Run Wishart discount DLM in parallel to SGDLM

"Warm-up" set of "top10" candidate parents:
- Evaluate over 10 days
- Merge into sp(*) or drop

"Cool-down" least significant current parents



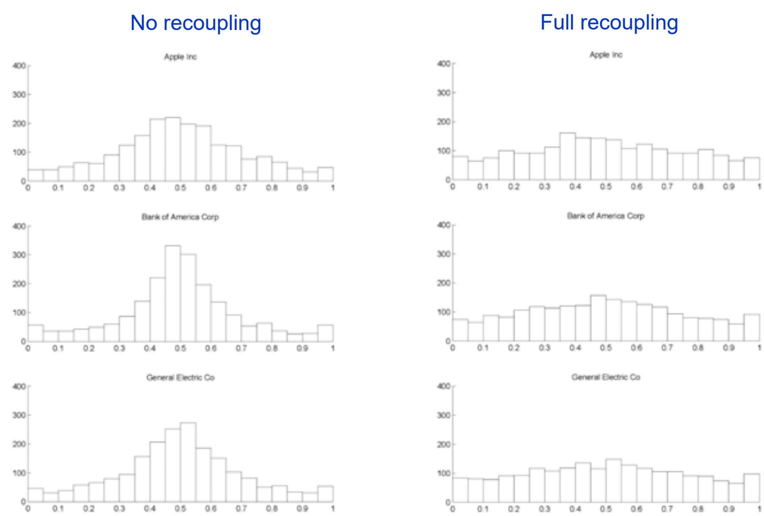
S&P example – Simultaneous parental set turnover

sp(3M)

S&P SGDLM : Can we ignore recoupling?

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

No recoupling / Full recoupling



S&P SGDLMs and portfolios

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

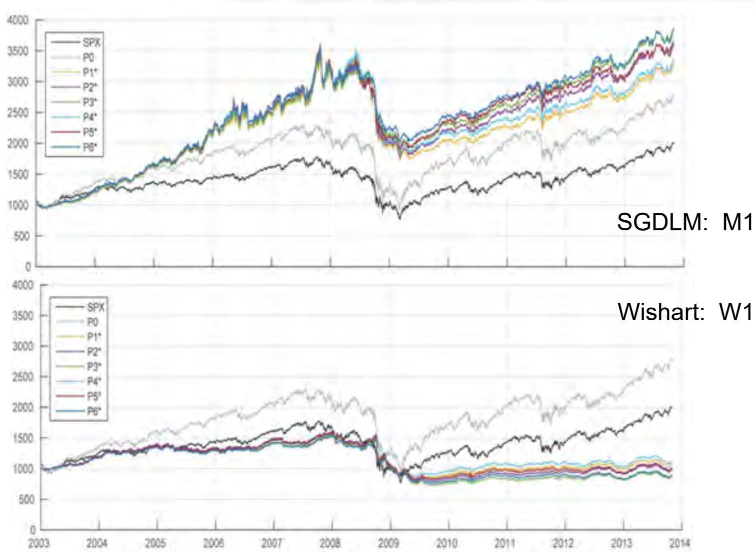
Daily trading
Hypothetical trading cost: 10bp

Strategy	Description
SPX	passive investment in the S&P 500
P0	equal weights
P1*	minimum variance
P2*	target return $\tau_t = 10\%/252$
P3*	target return $\tau_t = 15\%/252$
P4*	SPX neutral, minimum variance
P5*	SPX neutral, target return $\tau_t = 10\%/252$
P6*	SPX neutral, target return $\tau_t = 15\%/252$

S&P SGDLMs – Sharpe ratios

Model	SPX	P0	P1*	P2*	P3*	P4*	P5*	P6*
	0.31	0.41						
M1			0.71	0.79	0.86	0.71	0.76	0.84
M2			0.82	0.87	0.86	0.81	0.87	0.88
M3			0.63	0.62	0.69	0.60	0.62	0.69
M4			0.68	0.68	0.67	0.70	0.69	0.69
M5			0.73	0.74	0.72	0.68	0.68	0.67
MA1			0.72	0.71	0.69	0.71	0.74	0.73
MA2			0.77	0.74	0.70	0.76	0.72	0.68
MA3			0.80	0.76	0.68	0.78	0.72	0.68
MA4			0.83	0.78	0.75	0.83	0.78	0.75
MA5			0.88	0.86	0.69	0.91	0.88	0.77
W1			0.06	-0.01	-0.11	0.10	0.01	-0.09
W2			0.06	-0.00	-0.10	0.09	0.02	-0.08
W3			0.05	0.01	-0.08	0.09	0.03	-0.06
W4			0.05	0.04	-0.03	0.08	0.05	-0.01
W5			0.05	0.06	0.03	0.08	0.08	0.05

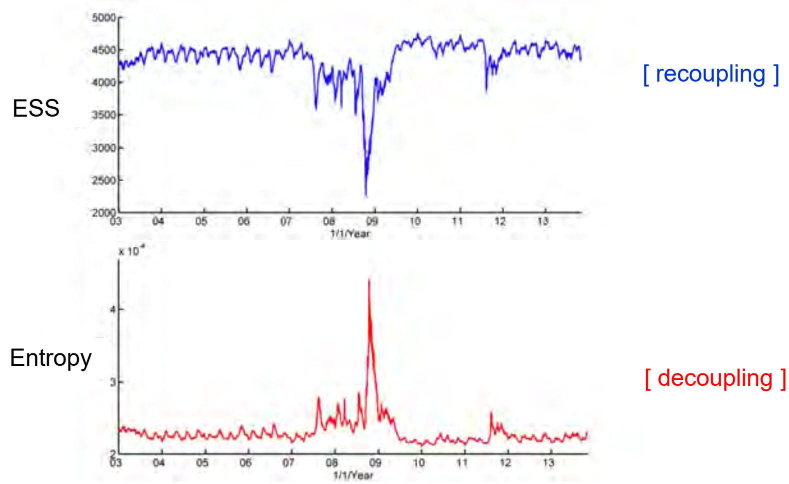
S&P400 portfolios: Examples of cumulative returns



Trading costs adjusted returns
Uniform improvements (Sharpe ratios, etc) over WDLMs

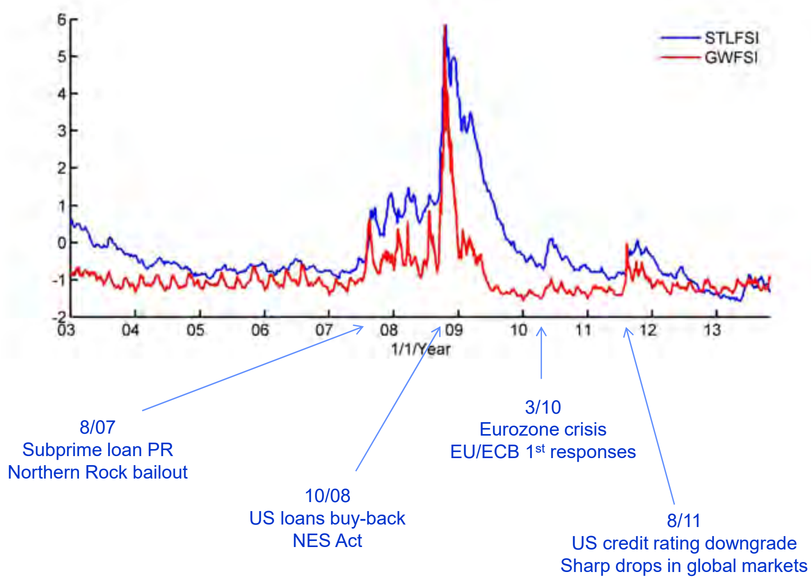
S&P example – Monitoring decoupling/recoupling

I=5,000 Monte Carlo sample size



Potential to intervene- “signal” of deterioration of ESS:
Increase sample size &/or modify model: parental sets, discounts, etc

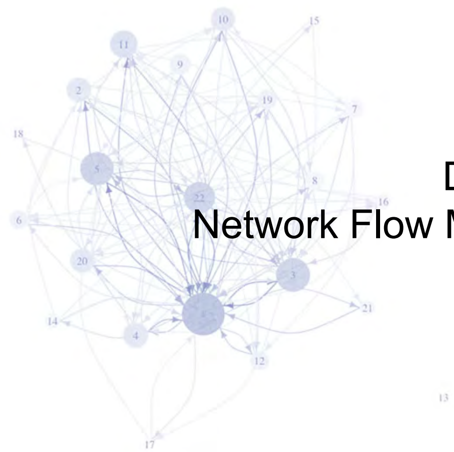
St Louis Fed “Financial Markets Stress Index”



SGDLM decoupling/recoupling and computations

- Flexible, customisable univariate models
- Flexible, sparse stochastic volatility matrix model
- No series-order dependence
- Scale-up : conceptual & computational

- On-line sequential learning: Analytics + cheap simulation + VB
  … parallelisable within each time step
- Decouple (parallel, GPU) & recouple (CPU) computation
- Decouple/Recouple IS&VB: theory, monitoring …
  by all series, subsets/sectors, each series



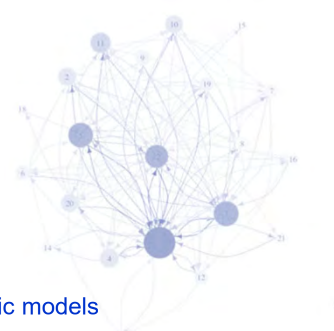
Dynamic Count Systems: Network Flow Modelling and Monitoring

Mike West
Duke University

Dynamic network data: IT, e-commerce, finance, …

Context- counts/flows between nodes:

- Sensitive tracking of flow dynamics
- Characterize "normal" patterns of time variation
- Sequential analysis: Fast? Parallelisation?
- On-line monitoring, anomaly detection
- Structured models: Node-node interactions?
- Scale up: #nodes I & time $T \uparrow$



Today:

- *Decoupled* node-pair flows: univariate dynamic models
- *Recoupling* - for inferences
- Bayesian model emulation: Map to structured dynamic models
- Sequential model monitoring and adaptation

E-advertising

Network “flows” …

- “Leisure” ⟶ “Travel”
- Hotel & car hire ads …. impact/responses?

Customized on-line ads …

- Sequential ad placement decisions
  - *auctions, timing*
- Understanding browser “experience”
- Dynamics of browser traffic flows

MAXPOINT
Valassis

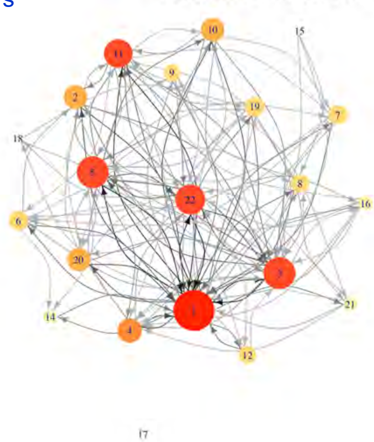
Network flow data: Fox News web domains

- US east coast browsers
- node-node counts:
  user/browser clicks
- flows in/out:
  “external” node

- 30sec intervals
- multiple days
- 1hour am, 1hour pm

- comparisons?

2015-02-23 09:25:00 - 09:25:30

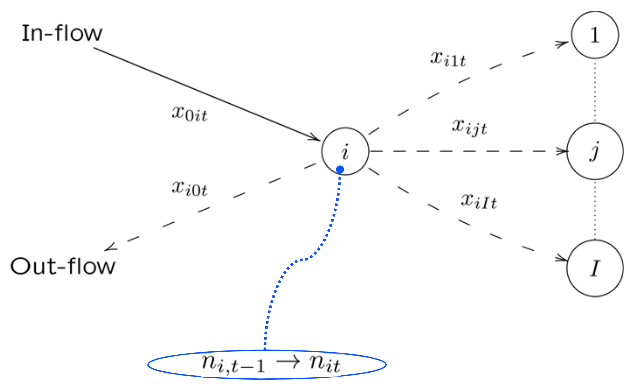


1 homepage
2 politics
3 us
4 opinion
5 entertainment
6 tech
7 science
8 health
9 travel
10 leisure
11 world
12 sports
13 shows
14 weather
15 category
16 latino
17 story
18 on-air
19 video
20 nation
21 magazine
22 others

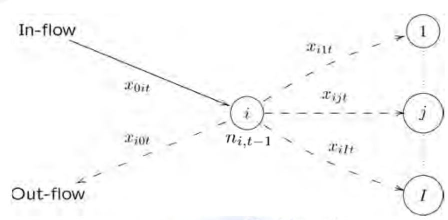
Counting?
- browser clicks
- user refreshes: stay
- user “quiet”: → external

MAXPOINT

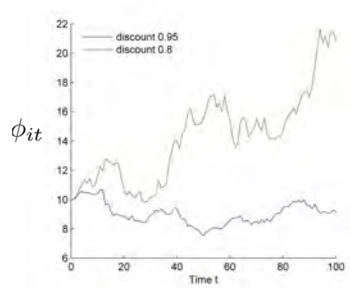
Dynamic network flows: Notation



Bayesian dynamic flow model - BDFM: (i) in-flows



$$x_{0it} \sim Poi(\phi_{it})$$



Canonical & general/flexible
Time evolution of rates?

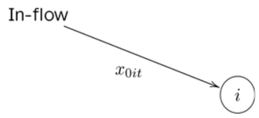
$$\phi_{it} = \phi_{i,t-1}\nu_{it}$$

- “Stochastic volatility model”
- Multiplicative random walk:

$$E[\phi_{it}|\phi_{i,t-1}] = \phi_{i,t-1}$$

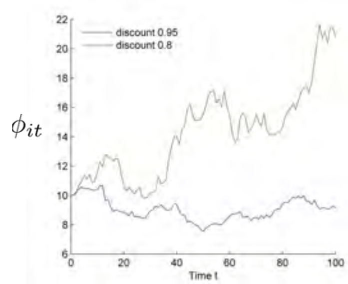
  - allows change
  - does not predict direction
- Single control parameter: *discount factor* δ_i

Bayesian dynamic flow model - BDFM: (i) in-flows



- Decoupled across nodes: Parallel models
- Fast, analytic sequential filtering
- On-line 1-step forecasting: monitoring
- Optimize discount factors in parallel
  [predictive marginal likelihood]

$$x_{0it} \sim Poi(\phi_{it})$$



Forward filtering (FF) - Gamma posteriors:

$$p(\phi_{it}|x_{0i,0:t})$$

Retrospective/backward sampling (BS):

$$p(\phi_{i,1:T}|x_{0i,0:T})$$

Monte Carlo simulation for inference

Bayesian dynamic flow model - BDFM: (i) in-flows

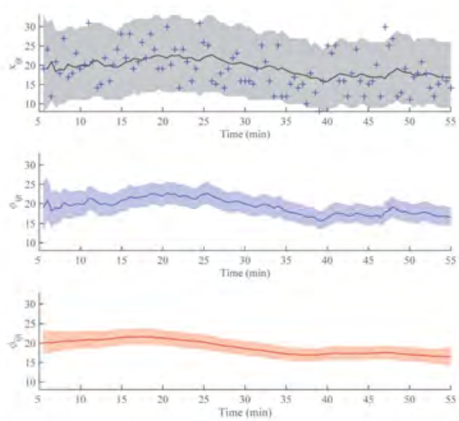
In-flows to node *i=10:* “Leisure”

+ counts $x_{0,10,t}$

1-step forecasts

on-line posteriors

retrospective posterior



Bayesian dynamic flow model – BDFM: (i) in-flows

Monitoring, intervene!

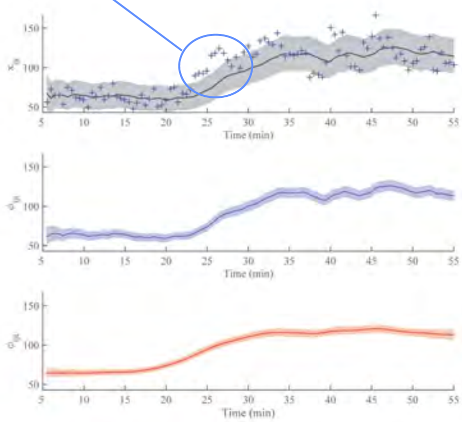
In-flows to node *i=11:* "World"

+ *counts* $x_{0,11,t}$

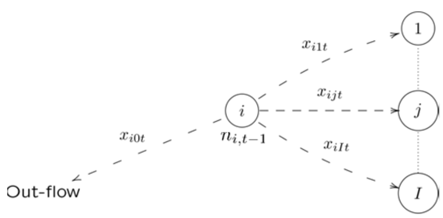
1-step forecasts

on-line posteriors

retrospective posterior



Bayesian dynamic flow model - BDFM: (ii) transitions



$$x_{i,0:I,t} \sim Mn(n_{i,t-1}, \theta_{i,0:I,t})$$

Time evolution of probabilities?

Decouple:

$$x_{ijt} \sim Poi(m_{it}\phi_{ijt})$$

$$\phi_{ijt} = \phi_{ij,t-1}\nu_{ijt}$$

- Random walking latent rates
- Node-node specific models

Recouple:

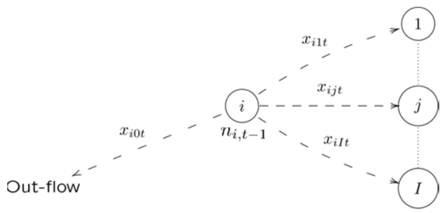
$$\theta_{ijt} = m_{it}\phi_{ijt} / \sum_{h=0:I} m_{it}\phi_{iht}$$

$$= \phi_{ijt} / \sum_{h=0:I} \phi_{iht}$$

key to decoupling:
occupancy factor

$$m_{it} = n_{i,t-1}/n_{i,t-2}$$

Bayesian dynamic flow model - BDFM: (ii) transitions



- Decoupled: Parallel over nodes
- Fast, analytic sequential filtering
- On-line monitoring
- Optimize discount factors in parallel
  [predictive marginal likelihood]

$$x_{i,0:I,t} \sim Mn(n_{i,t-1}, \theta_{i,0:I,t})$$

$$\begin{aligned} \phi_{ijt} &= \phi_{ij,t-1}\nu_{ijt} \\ \theta_{ijt} &= \phi_{ijt} / \textstyle\sum_{h=0:I} \phi_{iht} \end{aligned}$$

Filtering (FF) :

$$p(\theta_{i,0:I,t} | x_{i,0:I,1:t})$$

Retrospection:

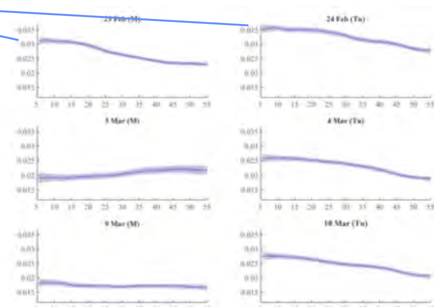
$$p(\theta_{i,0:I,1:T} | x_{i,0:I,1:T})$$

Monte Carlo simulation for inference

Bayesian dynamic flow model - BDFM: (ii) transitions

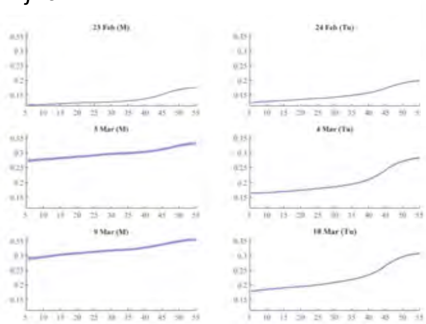
pm Feb 22: Academy Awards?

Multi-morning flows from *i=1:* "Homepage" to …



j=5: "Entertainment"

j=0: "External"



MAXPOINT

Dynamic network modelling: Perspectives

Bayesian dynamic flow model (BDFM):

- “Unstructured” adaptive models: sensitive tracking of flows
- Characterize “normal” patterns of time variation
- Full formal inference, uncertainties, comparisons
- Sequential analysis:

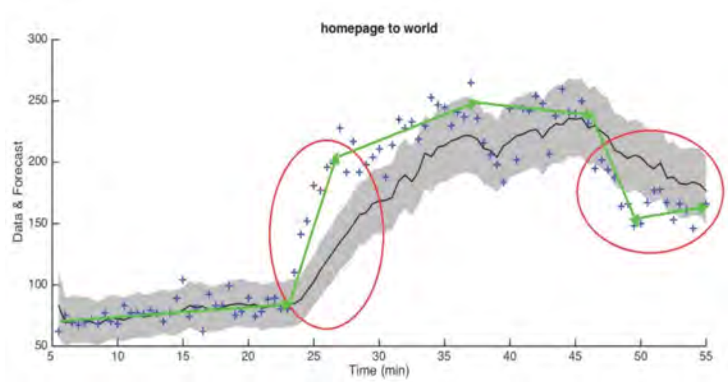
Analytic: fast, parallel, scalable

Open to:

- Sequential monitoring:
  - *formal Bayesian sequential testing*
  - *anomaly detection*
  - *impact of e-commerce interventions*
- Decoupled: scales to large networks

More general, flexible, predictive models ?

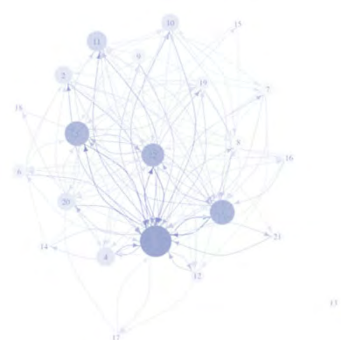
But ….



Dynamic generalized linear models (DGLMs) for flows

More structure for explanation, feed-forward modelling?

- Dynamics in node-node interactions
- Random effects: node specific, node-node specific
- Covariate information:
  - *node changes, interventions, e-ads?*



Extended class of BDFMs using decoupled sets of DGLMs

Flow Poisson DGLMs and example

Counts: $x_t \sim Poi(m_t \phi_t)$

$$\log(\phi_t) = \mathbf{F}_t' \boldsymbol{\theta}_t$$

State vector evolution:

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

Eg: local linear + regression

$$\boldsymbol{\theta}_t = \begin{pmatrix} \text{level}_t \\ \text{gradient}_t \\ \text{zcoeff}_t \end{pmatrix}$$

$$\mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \\ z_t \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Over-dispersion? Random effects extensions: Extra-Poisson variation

$$\log(\mu_t) = \mathbf{F}_t' \boldsymbol{\theta}_t + \lambda_t$$

Mean 0
Independent "shocks"

Univariate DGLM analysis

Counts: $x_t \sim Poi(m_t \phi_t)$

$$\log(\phi_t) = \mathbf{F}_t' \boldsymbol{\theta}_t$$

State vector evolution:

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

- Traditional forward filtering, forecasting & retrospective (smoothing)
- Filtering propagates & updates { mean, variance matrix } of state
  - exploits "variational Bayes" constraints to (gamma) conjugacy

    $\phi_t | \mathcal{D}_{t-1} \sim$ gamma
  - and "linear Bayes" for prior-posterior updates
- Fast, sequential and decoupled across node pairs

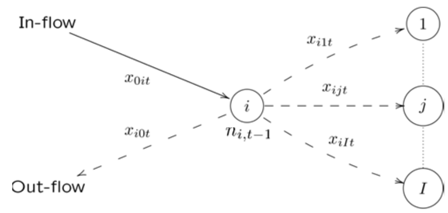
Structured dynamic network modelling: Perspectives

Deeper structure investigation/explanation?

- Dynamics in node-node interactions
- Node specific effects, node-node specific interactions

Recouple node-specific BDFMs ?

Dynamic gravity model (DGM)



$$x_{0it} \sim Poi(\phi_{it})$$
$$x_{i,0:I,t} \sim Mn(n_{i,t-1}, \theta_{i,0:I,t})$$
$$\theta_{ijt} = \phi_{ijt} / \textstyle\sum_{h=0:I} \phi_{iht}$$

covariates

$$\phi_{it} = \phi_{it}(z_{it})$$

$$\phi_{ijt} = \mu_t(z_t) \times \alpha_{it}(z_{it}) \times \beta_{jt}(z_{jt}) \times \gamma_{ijt}(z_{ijt})$$

- "Gravity": node-node attraction/affinity
- Many choices of model forms

[West 1994; transportation flows epidemic spread …]

A first dynamic gravity model

$$\phi_{ijt} = \mu_t \times \alpha_{it} \times \beta_{jt} \times \gamma_{ijt}$$

Today:

(i) First-step: Random effects model within network
- *covariates: node indices*
- *main effects=node, interaction effects=node pairs*

Full Bayesian model fitting?

- *Computationally easy, fast, analytic*
- *Scalable*
- *Sequential analysis, monitoring, …*

(ii) Link to BDFM: Bayesian model emulation
– *exploit simple, efficient BDFM to map to DGM*

Dynamic gravity model (DGM)

Log-linear model mapping

BDFM DGM

$$\phi_{ijt} = \mu_t \, \alpha_{it} \, \beta_{jt} \, \gamma_{ijt}$$

- aliasing constraints
- "centred" effects

$$\begin{aligned} \textstyle\prod_i \alpha_{it} &= 1 \\ \textstyle\prod_j \beta_{jt} &= 1 \\ \textstyle\prod_i \gamma_{ijt} &= 1 \\ \textstyle\prod_j \gamma_{ijt} &= 1 \end{aligned}$$

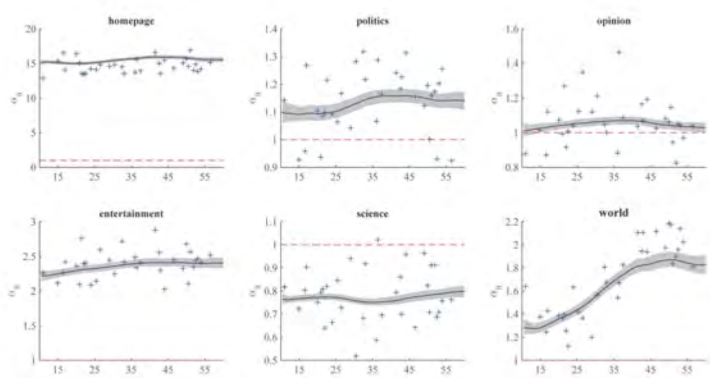
geometric means $\breve{\phi}_*$

$$\begin{aligned} \mu_t &= \breve{\phi}_{\cdot\cdot t} \\ \alpha_{it} &= \breve{\phi}_{i\cdot t}/\mu_t \\ \beta_{jt} &= \breve{\phi}_{\cdot jt}/\mu_t \\ \gamma_{ijt} &= \phi_{ijt}/(\mu_t \alpha_{it} \beta_{jt}) \end{aligned}$$

Emulation map:
BDFM Monte Carlo posterior $\Longrightarrow$ DGM effects posterior

DGM: strength of node "out-flow" transitions

February 23 am
α_{it}

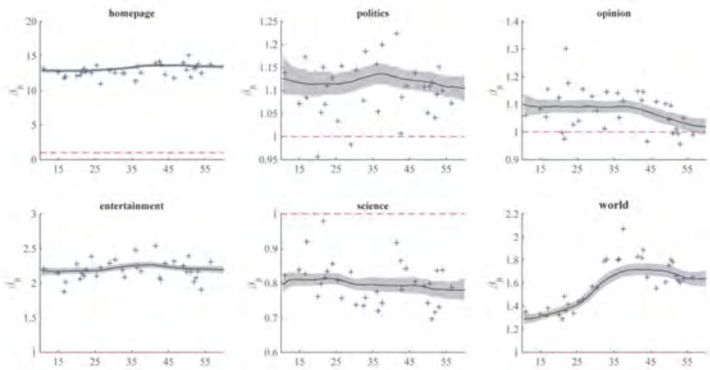


MAXPOINT

DGM: strength of node “in flow” transitions

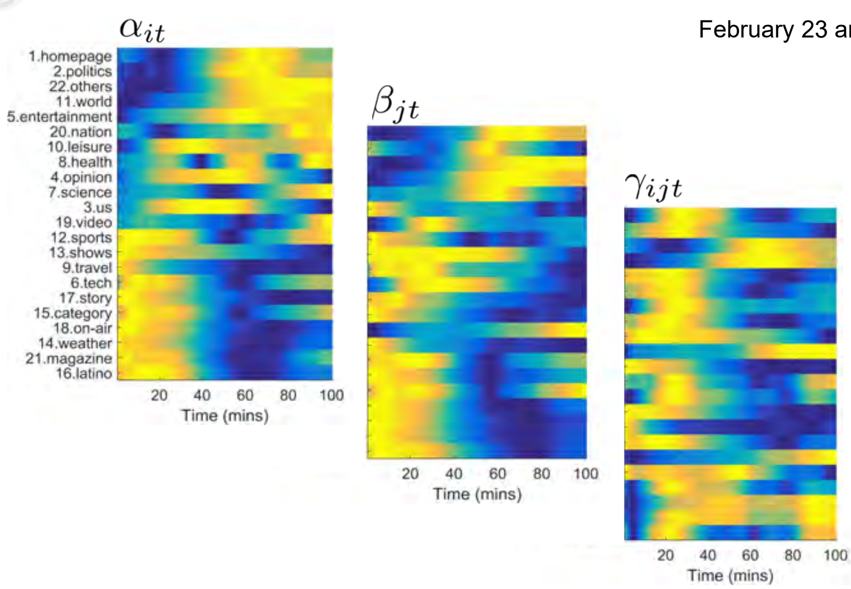
February 23 am

β_{jt}



DGM: network-wide inter-relationships

February 23 am



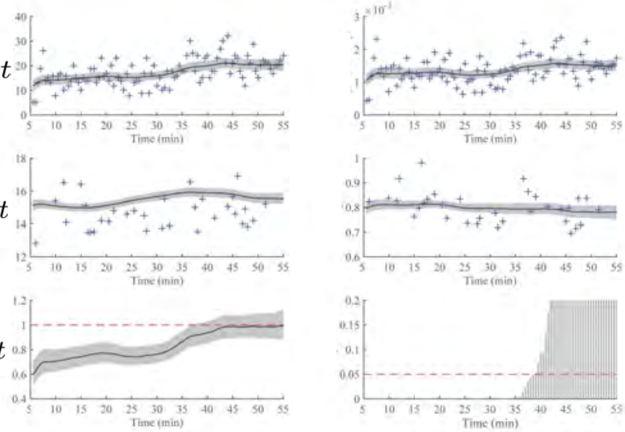
DGM: node-node affinities

February 23 am
i=1, Homepage
j=7, Science

ϕ_{ijt} θ_{ijt}

α_{it} β_{jt}

γ_{ijt} $\gamma_{ijt} \approx 1?$

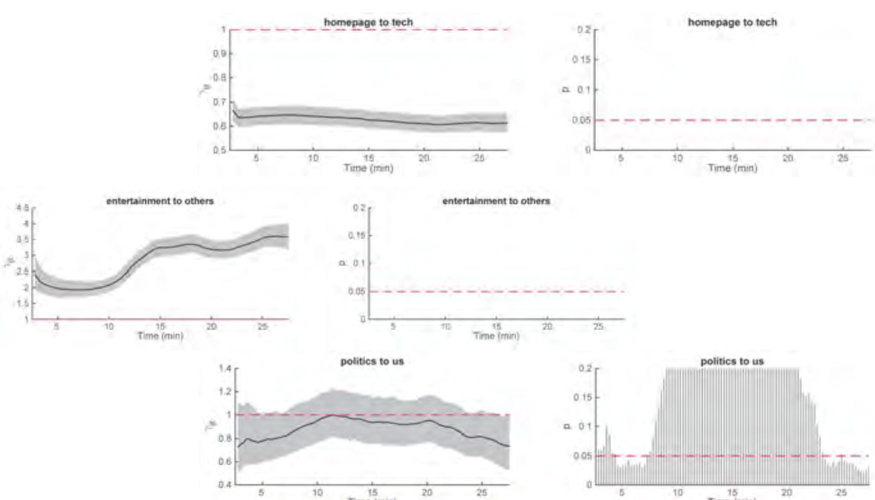


"dynamic sparsity"?

DGM: Dynamic "significance" of (interactions) affinities

$Pr(1 \in 95\% CR)$

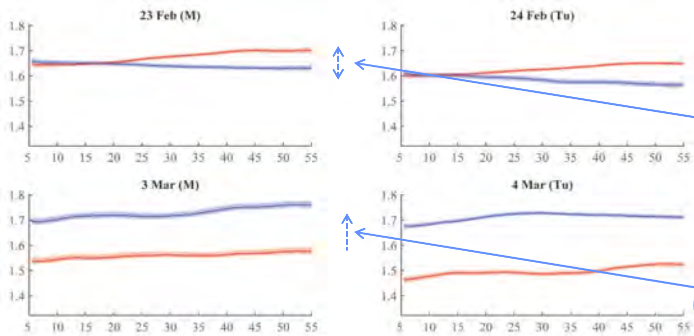
CR = posterior credible region for γ_{ijt}



DGM: day-day comparisons, differences?

μ_t

09:05-09:55
13:05-13:55



Academy Awards blogging?

Breaking news: Clinton email "scandal"?

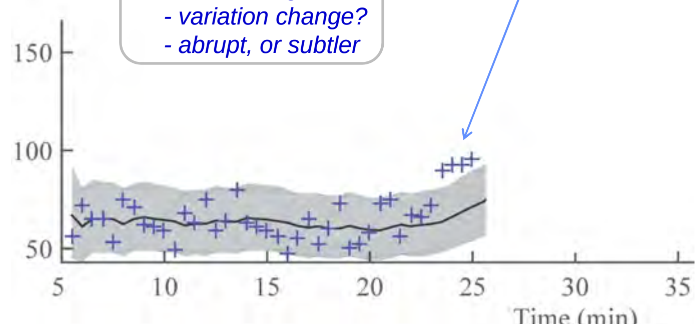
Bayesian model monitoring and adaptation

"Standard" sequential Bayesian assessment
On-line predictions

1 'discrepant' *x …*
- rare event? outlier?

2+ 'discrepancies …
- level change?
- variation change?
- abrupt, or subtler

Intervene!



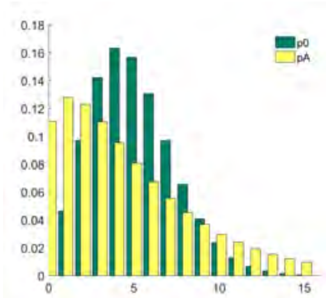
In-flows to node *i=11:* "World"

Predictive Bayes factor testing:
'standard' model
versus
suitable 'alternative'

$$H_t = \frac{p_0(x_t|x_{1:t-1})}{p_A(x_t|x_{1:t-1})}$$

$$L_t = \min_{0\le r<t} H_t H_{t-1} \cdots H_{t-r}$$

$$l_t = argmin$$

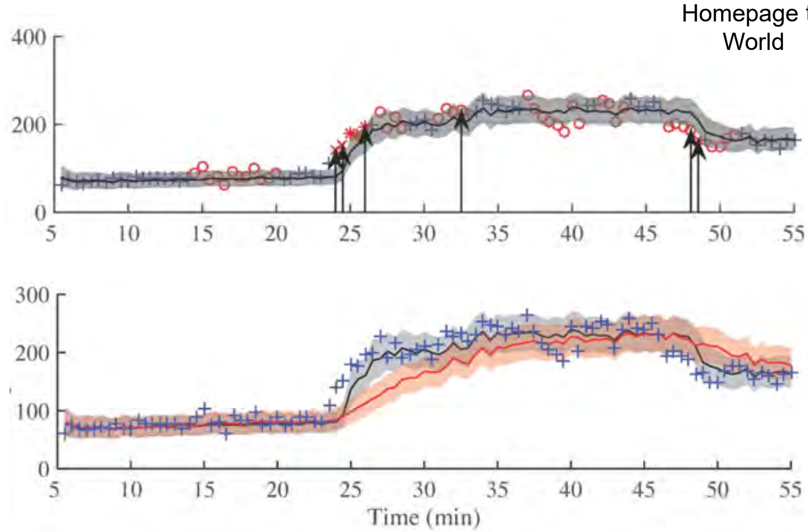


Monitor/tracking: $\{L_t, l_t\}$

- *most discrepant recent, consecutive, observations*
- thresholds
- *intervene*: ditch outlier
  decrease discount factor

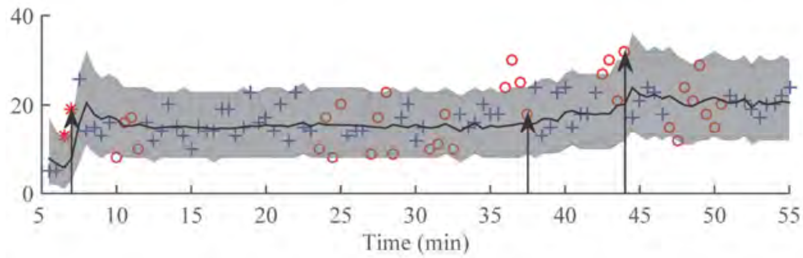
Bayesian model monitoring and adaptation

February 23 am
Homepage to
World



Bayesian model monitoring and adaptation

February 23 am
Homepage to
Science

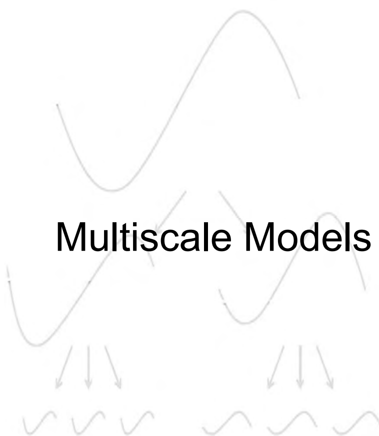


Decouple/recouple in dynamic network flow modelling

- Flexible, decoupled, parallelizable, scalable BDFMs
- Characterise network flow dynamics
- Fast sequential/forward data analysis
- Formal monitoring for anomaly/surprise - decoupled

- Currently –
Scale-up
BDFM covariates
"Local" dependencies
Zero/low counts
Other applications
…

- Recouple: structured DGMs
- Interdependencies across network
- "Impossible" computationally at scale or sequentially
- Bayesian model emulation
    - *map from flexible BDFM to DGM*



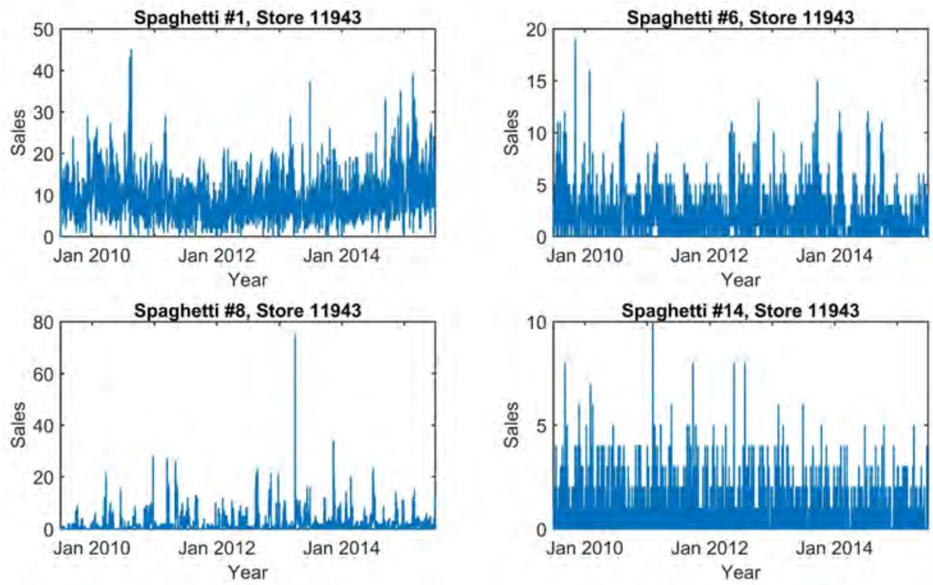
Dynamic Count Systems: Multiscale Models for Large-Scale Forecasting

Mike West
Duke University

Background & goals

- Forecast time series of non-negative counts *– item-level daily sales/store*
- Many items, many stores : shared “patterns”
  – multivariate - multi-scale concepts
- Joint forecast: *1-14 days, revised everyday, every item, every store*
- Bayesian/probabilistic: Multiple end-users, decisions
- Broad purview: ALL items (mature/new, sporadic/popular, all stores …)
  - i.e., concern for general model/analysis/forecasting frameworks
- Automation- models flexible and adaptive
- Computation feasibility: Sequential/on-line
  - later: hourly updates, “real-time” forecasting for informed decisions

Some pasta sales - One store



Some pasta - One store

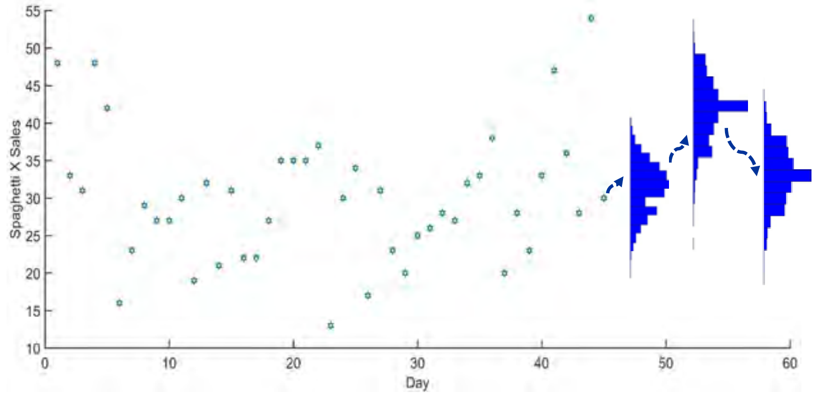
Item	Mean	Mean|Sale	SD	Pr(Sale)
1	9.18	9.31	5.28	0.99
6	2.09	2.80	2.19	0.75
8	1.74	3.86	3.64	0.45
14	0.77	1.87	1.19	0.41

Aggregates: Daily seasonal patterns, holidays, long-term trend
Local/store/brand effects; Unpredictable volatility

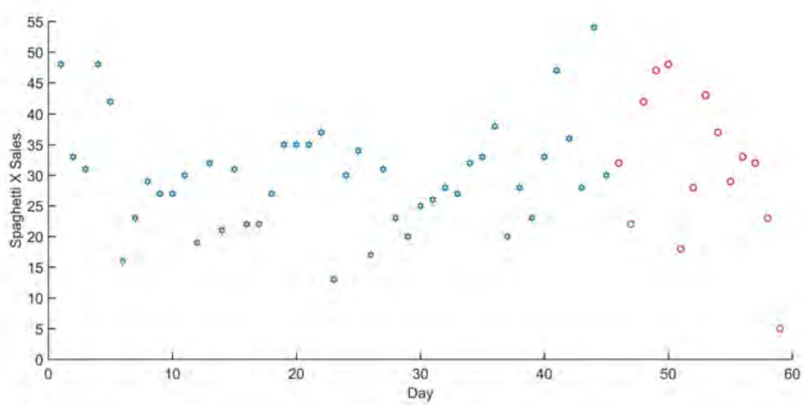
Medium selling items: Noisier/shared seasonal patterns, trends
Price, promotion, local effects
Low-medium counts; diverse non-Poisson features
(e.g. unpredictable "surges")

Low/sporadic items: Zero versus non-zero sales

Probabilistic forecasting



Probabilistic forecasting: Synthetic realities



Univariate dynamic (generalised linear) models

Higher-level aggregates (all pasta, all spaghetti)

Normal Models $\log(y_t) \sim N(\mu_t, v_t)$

$$\mu_t = \mathbf{F}_t'\boldsymbol{\theta}_t$$

Item-level covariates dummy variables

Local trend, daily seasonal, price/promotion elasticities

Residual volatility

Time-varying state-vector, volatility: assess and predict

Univariate dynamic (generalised linear) models

Non-negative integer counts: Item-level sales when sold

Shifted Poisson: $y_t = 1 + x_t,\ \ x_t \sim Po(\mu_t)$

$$\log(\mu_t) = \mathbf{F}_t'\boldsymbol{\theta}_t$$

Over dispersion?

Random effects extensions:
Extra-Poisson variation

$$\log(\mu_t) = \mathbf{F}_t'\boldsymbol{\theta}_t + \lambda_t$$

Mean 0
Independent "shocks"

Easy technically: extend state vector

Univariate dynamic (generalised linear) models

What about zero item sales?

Binary model: $\Pr(y_t > 0) = \pi_t$

$$\text{logit}(\pi_t) = \mathbf{F}_t'\boldsymbol{\theta}_t$$

DCMM: Dynamic count mixture model

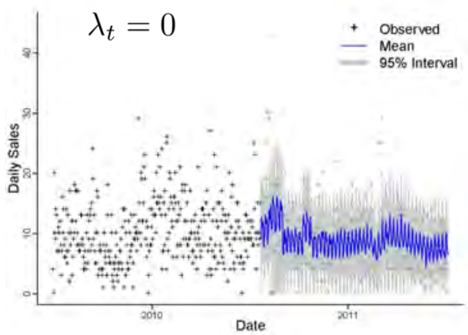
DCMMs for non-negative count time series

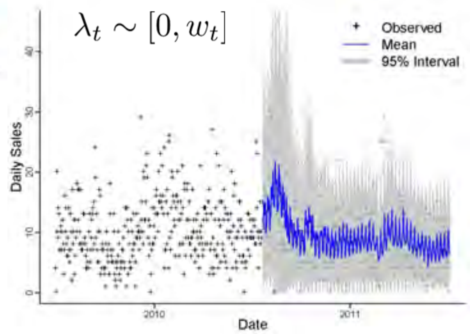
$$y_t = \begin{cases} 0, & \text{with probability } 1 - \pi_t \\ 1 + x_t, \quad x_t \sim Po(\mu_t), & \text{otherwise} \end{cases}$$

$$\begin{aligned} \text{logit}(\pi_t) &= \mathbf{F}_{0t}'\boldsymbol{\theta}_{0t} \\ \log(\mu_t) &= \mathbf{F}_{+t}'\boldsymbol{\theta}_{+t} + \lambda_{+t} \end{aligned}$$

DCMM: Role for dynamic random effects

Item 1: 1-day forecasts





Evaluating forecasts

Non-negative counts: under/over-forecasting, asymmetries

Distributional form: tail events, “value-at-risk”?

1-14 day forecasting: “Path forecasts”, cumulative forecasts

Decision analysis perspective:

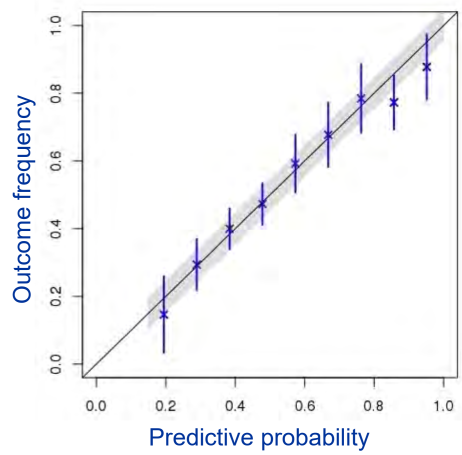
Context specific goals, decisions -
Meaningful accuracy/evaluation metrics?

Traditional point forecast metrics: rMSE, MAD, sSE, MAPE …

Probabilistic: Calibration, coverage - full forecast distributions

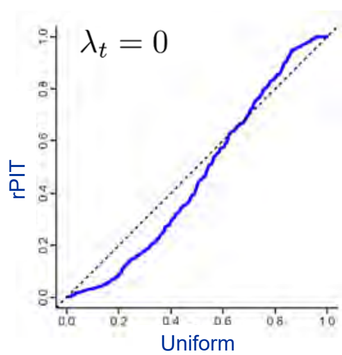
Calibration: Predicting zero/non-zero

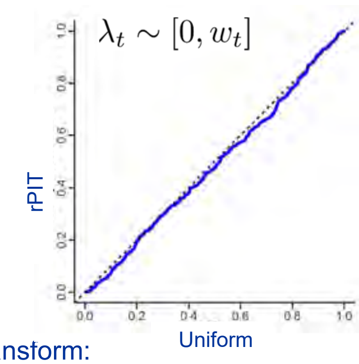
(low selling) Item 8: 1-day forecasts



Calibration: Full predictive distribution

Item 1: 1-day forecasts



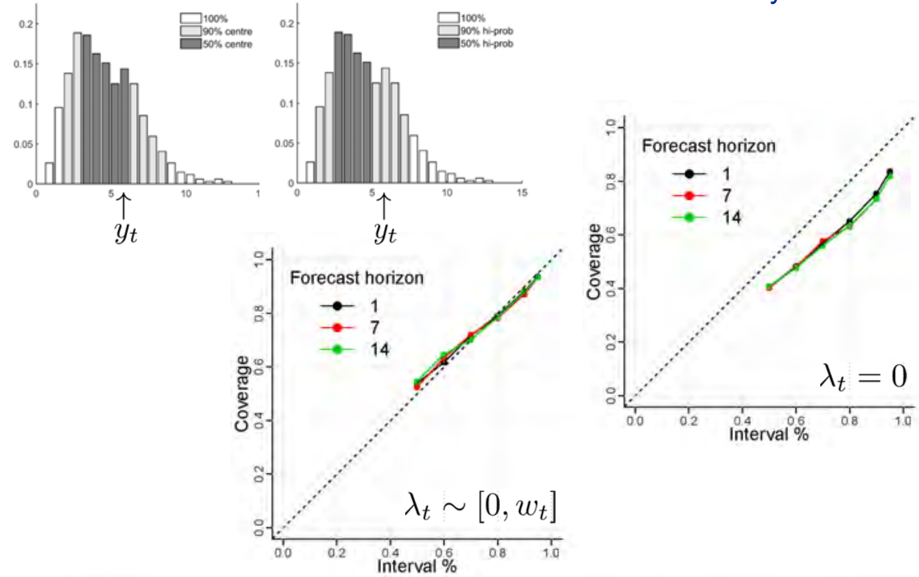


rPIT: Randomized probability integral transform:

- "probabilistic predictive residual analysis"
- ordered predictive CDF values
- good model: "looks uniform"

Coverage: Forecast intervals

Item 8: 1-14 day forecasts



Multivariate: All items

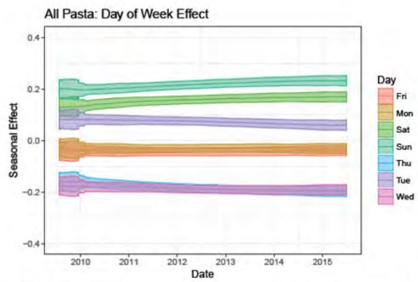
Hierarchies: spaghetti items
< { spaghetti, brand, store }
< { pasta, brand, store }
< …

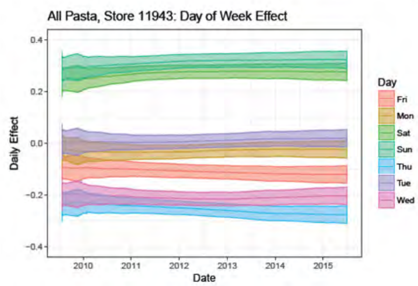
Pilot project: By store,
spaghetti items < spaghetti < pasta

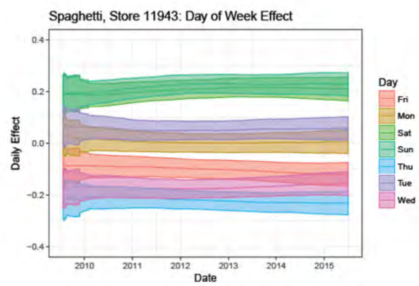
Cross-level linkages / e.g. consistency of daily seasonals?

- Related patterns: Borrowing strength ~ Hierarchical models
- Improved prediction of daily effects at aggregate level (“traffic” proxy) may improve item-level forecasts

Multi-scale thinking: Example of day/week seasonality



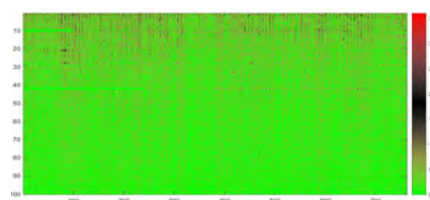




Decouple/recouple : Multi-scale forecasting

Many time series sharing "similar patterns"
- Dynamic hierarchical models ? Factor models ?
- Not scalable

Consumer sales/supermarkets
- *m* ~ 10,000s-100,000s



DECOUPLE – univariate DCMMs

RECOUPLE – direct simulation
- "Common" seasonal factors from external model
- Series-specific mapping of common factors
- New approach to dynamic factor modelling

Multi-scale thinking: Example of day/week seasonality

Concept: predict “common” effects at aggregate level
project to item-level with item-specific elasticities

Top level: $\mathcal{M}_0$

Fine level: $\mathcal{M}_j$

Multi-scale thinking: Example of day/week seasonality

$$\log(\text{spaghetti}_t) \leftarrow \{ \text{trend}_t, \ \text{seas}_t, \ \text{store features}_t, \ \text{etc}_t \}$$

$$\log(\text{item}_{i,t}) \leftarrow \{ \text{item/trend}_{i,t}, \ \beta_{i,t}\text{seas}_t, \ \text{item/price}_{i,t}, \ \text{etc}_{i,t} \}$$

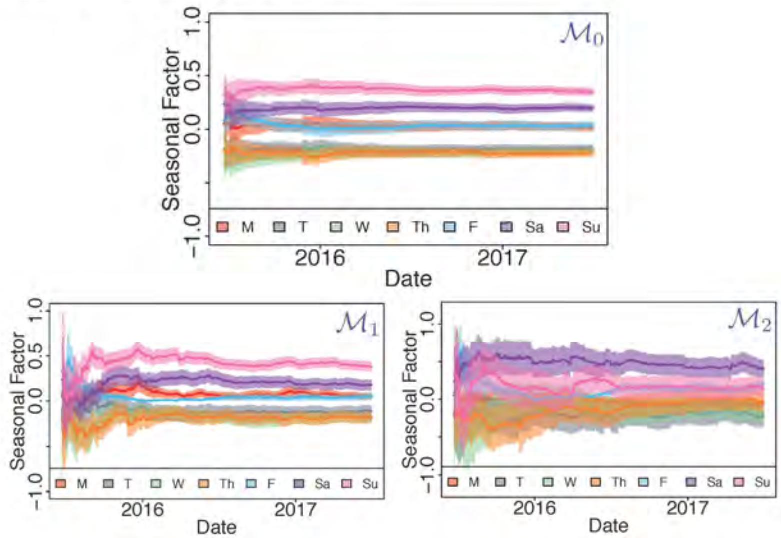
common factor

item-specific effect

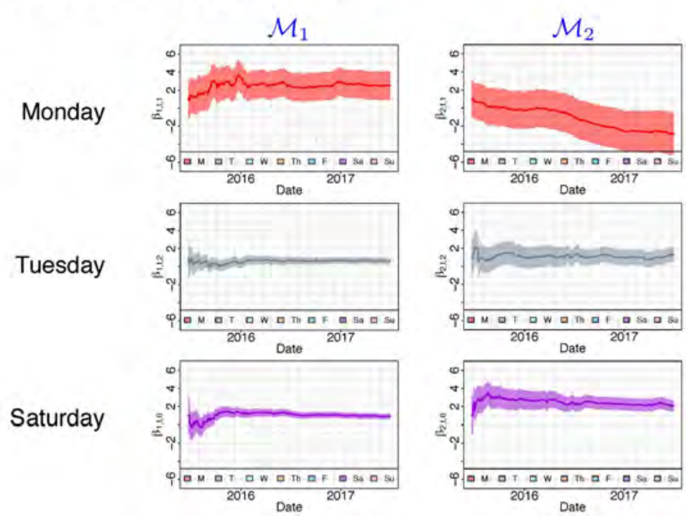
- Projection: simulate- fully probabilistic, accounts for all uncertainties

- Sequential analysis: Parallel, scalable -
  - *adds 1 external/higher level/aggregate DGLM*

Multi-scale and fine-level factors: Day-of-week



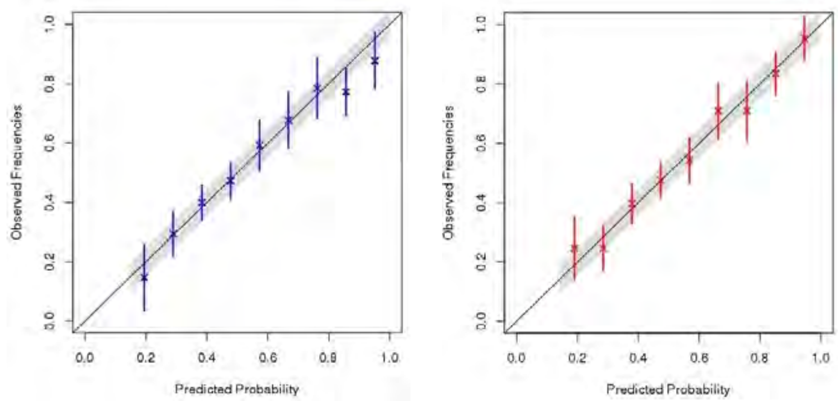
Fine-level factor loadings: Day-of-week



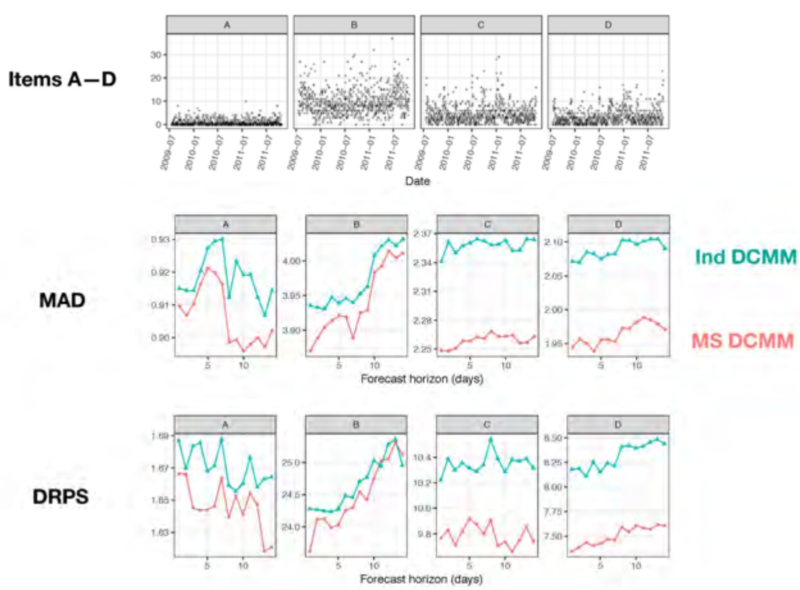
Calibration: Predicting zero/non-zero

(low selling) Item 8: 1-day forecasts

uni-scale multi-scale

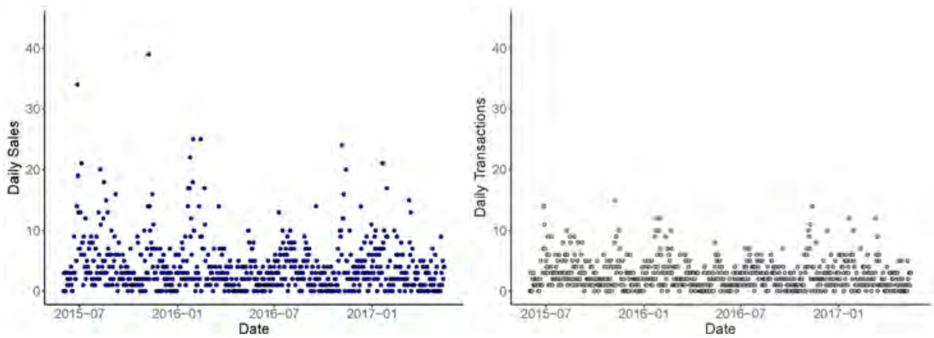


Multi-scale forecasting metrics



More data: Transactions and sales

Item 6

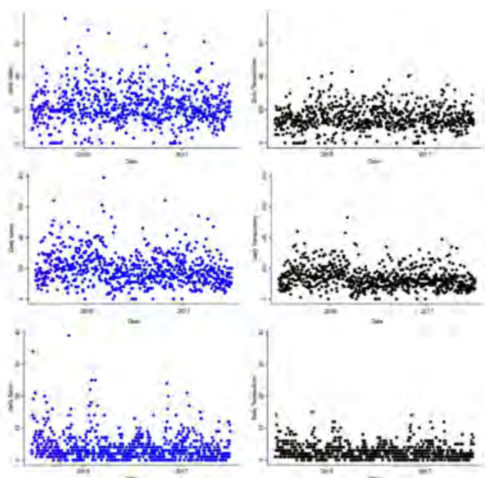


Resolving heterogeneity? Coupled transactions and sales

Item A: high seller

Item B: high/middle seller

Item C: low/mid seller



	Daily transactions			Sales-per-transaction		
Item	Mean	Median	Variance	Mean	Median	% < 5
A	22.84	21	100.52	1.46	1	98.9
B	19.75	18	101.15	1.44	1	99.0
C	4.66	3	18.70	1.53	1	98.4

New dynamic models for transactions/sales time series

Potential to (partially) resolve heterogeneity

Transactions :

- DCMM: dynamic count mixture model
- reduced heterogeneity/random effect impact

Sales per transaction :

- new DBCM: Dynamic binary cascade model
- customized to item, admitting "rare" events

DBCM: Dynamic binary cascade model

$$y_t = \sum_{r=1:d} r(n_{r-1,t} - n_{r,t}) + e_t$$

e.g. $d = 4$

transactions with $> r$ units

"excess" - rare events -

transactions $b_t = n_{0,t}$

Cascade of ("stable") binary DGLMs:

$$n_{r,t} | n_{r-1,t} \sim Bin(n_{r-1,t}, \pi_{r,t})$$

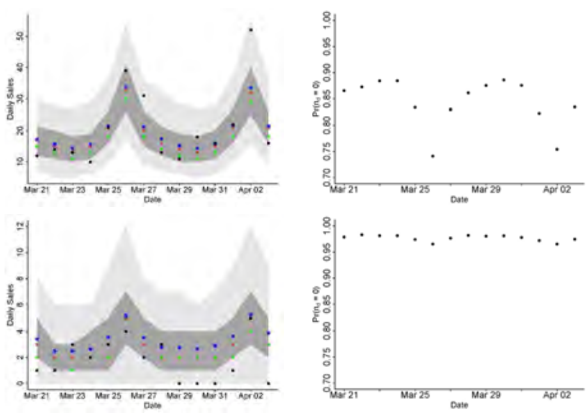
Excess? Rare events …

i. unspecified: conditional forecast $p(y_t | e_t = 0)$ and $Pr(e_t = 0)$
ii. nonparametric Bayes: empirical past excesses

DBCM : Examples

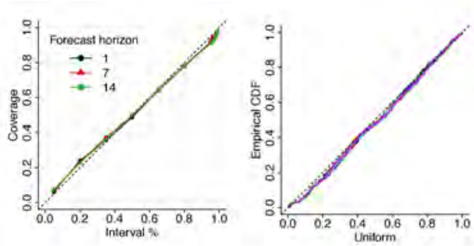
2 items, 1 day
1-14 day forecasts

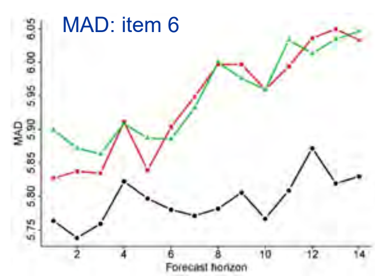
empirical excess strategy | $Pr(e_t = 0)$ | conditional on excess=0



1-median, median, mean
● outcome

DBCM compared to DCCM : Example and summaries



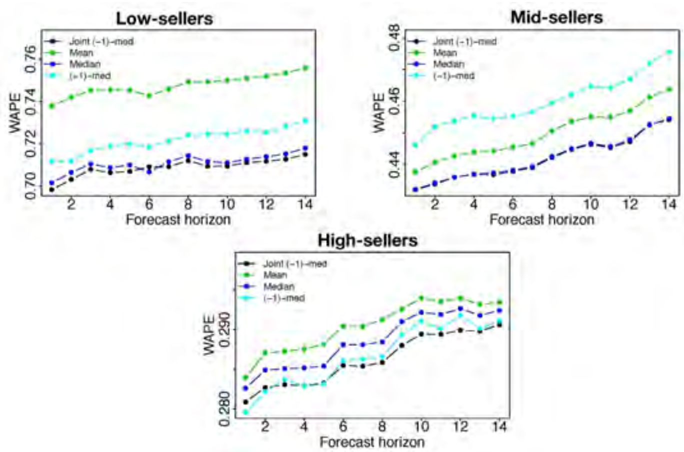


uni-scale DCMM
uni-scale DBCM
multi-scale DBCM

Coverage and calibration:

- Generally as good or better with lower levels of random effects
- DCMM “better” for transactions than sales
- Sales heterogeneity resolved- to a degree- by DBCM
- Improvements greater at higher levels of sales
- Multi-scale model increasedly important for DBCM: traffic ~ transactions

Larger scale comparison: Multiple items, several stores



Industry standard
~ mean ~

Other metrics
- similar or better -

Differences?
$$$$$$

Forecasting heterogeneous collections count time series

Flexible, adaptable dynamic models

- *Founded in traditional Bayesian forecasting model contexts*
  - *Forecasting: control and intervention*
- *Novel binary/non-negative integer structures*
- *Multi-scale innovations: hierarchies, decouple/recouple, scalability*
- *Cascade & random effects concepts help resolve heterogeneity*

Consumer sales forecasting

- *Multi-scale model benefits: many items, some times, by most metrics*
- *Focus on decision goals in choosing forecast summaries*
- *Integrate in-house forecasting, management & decision systems*
- *Other covariates (refine "promotions", local store information, …)*
- *Improve excess forecast model in DBCM*
- *Multiple hierarchies: item < { spaghetti, brand, store … }*

Other potential areas

- *"Epidemics" (e.g., health, crime, social mobility)*
  - *spatial and other covariates*
- *Economics & finance (e.g. regional bankruptcies, stock events)*