Bayesian Dynamic Modeling and Forecasting of Count Time Series

Lindsay Berry
Berry Consultants

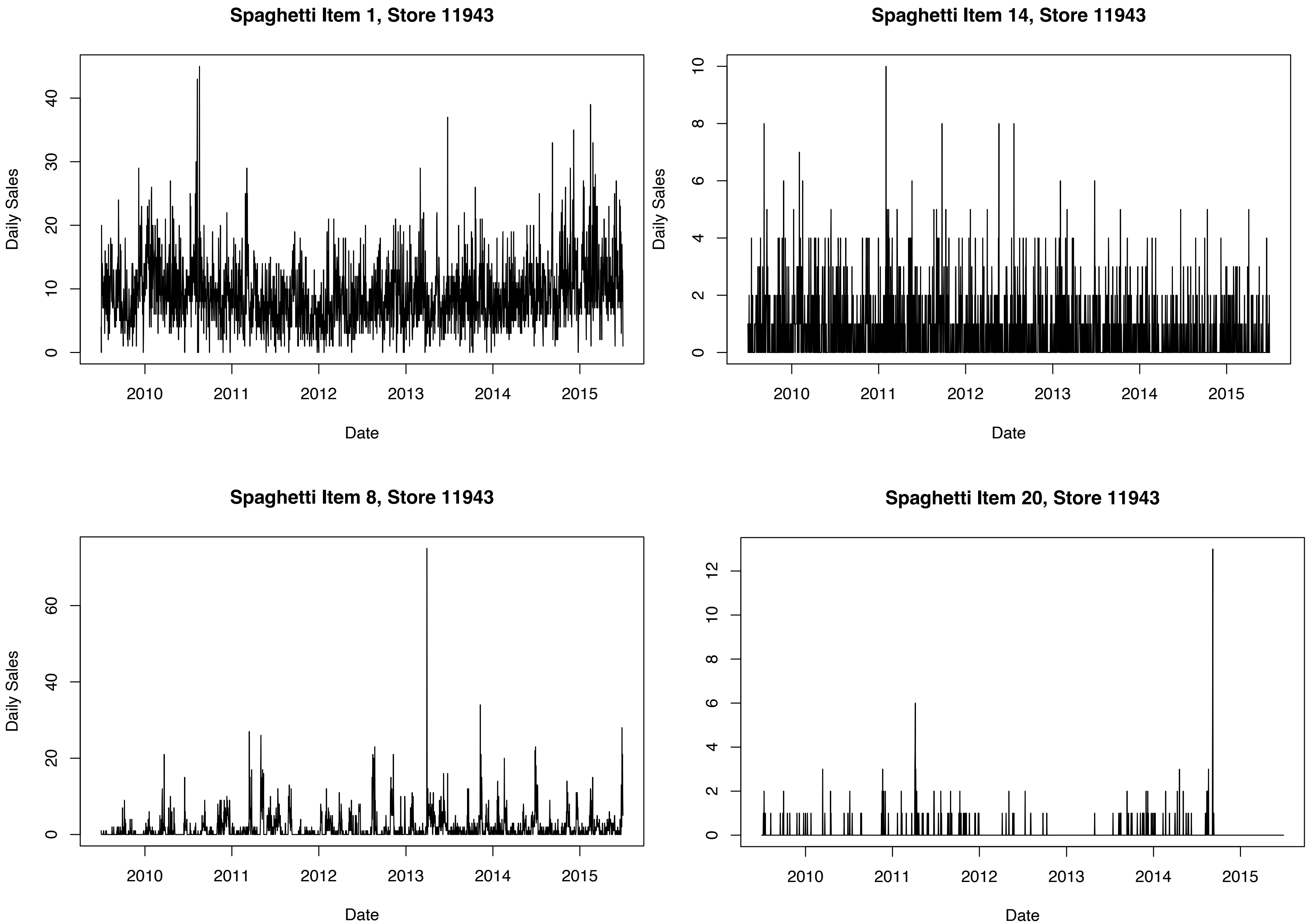
JSM 2020
August 3, 2020

Duke Department of
STATISTICAL SCIENCE

Berry Consultants
Statistical Innovation

Background

- Forecast time series of non-negative **counts**
- Broad purview: general framework for count data
- **Automation**: models flexible and adaptive
- Many series with **hierarchical** structure
- Joint forecast: multi-step ahead (1:14 days)
- Computational feasibility: **on-line estimation**
- Forecasts feed into many decisions



- Varying demand
- Seasonality
- Variability/ extreme values
    - Price/promo/ holidays
- Many zeroes
- All effects potentially time-varying

Item	Mean	Mean $\mid$ Sale	Var	$\% = 0$
1	9.18	9.31	27.88	0.01
8	1.74	3.86	13.25	0.55
14	0.77	1.87	1.42	0.59
20	0.09	1.47	0.22	0.94

Contributions

- Flexible **dynamic models** of counts
  - Built from traditional Bayesian forecasting contexts
  - Coupled binary/non-negative integer models
  - **Multi-scale** extension: scalable, decouple/recouple
  - Cascade and random effect: heterogeneity, rare events
- **Decision analysis**: loss functions, optimal forecasts
- **Consumer sales forecasting**
  - Multi-scale model beneficial for many items, forecast horizons, and metrics
  - DBCM: large-scale reduction in key error metrics
  - Probabilistic forecasting: uncertainty, decision making

Dynamic Count Mixture Model (DCMM)

y_t : non-negative count

$z_t = I(y_t > 0)$

$$z_t \sim Ber(\pi_t)$$

$$y_t \mid z_t = \begin{cases} 0, & \text{if } z_t = 0 \\ 1 + x_t, \quad x_t \sim Po(\mu_t), & \text{if } z_t = 1 \end{cases}$$

DGLMs:

$$\text{logit}(\pi_t) = \mathbf{F}_t^0 \boldsymbol{\xi}_t$$

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

Random effect extension

$$p(y_{t+k} \mid D_t, \pi_{t+k}) = (1 - \pi_{t+k})\delta_0(y_{t+k}) + \pi_{t+k}\left\{Nb\left(\alpha_t^+(k), \frac{\beta_t^+(k)}{1 + \beta_t^+(k)}\right) + 1\right\}$$

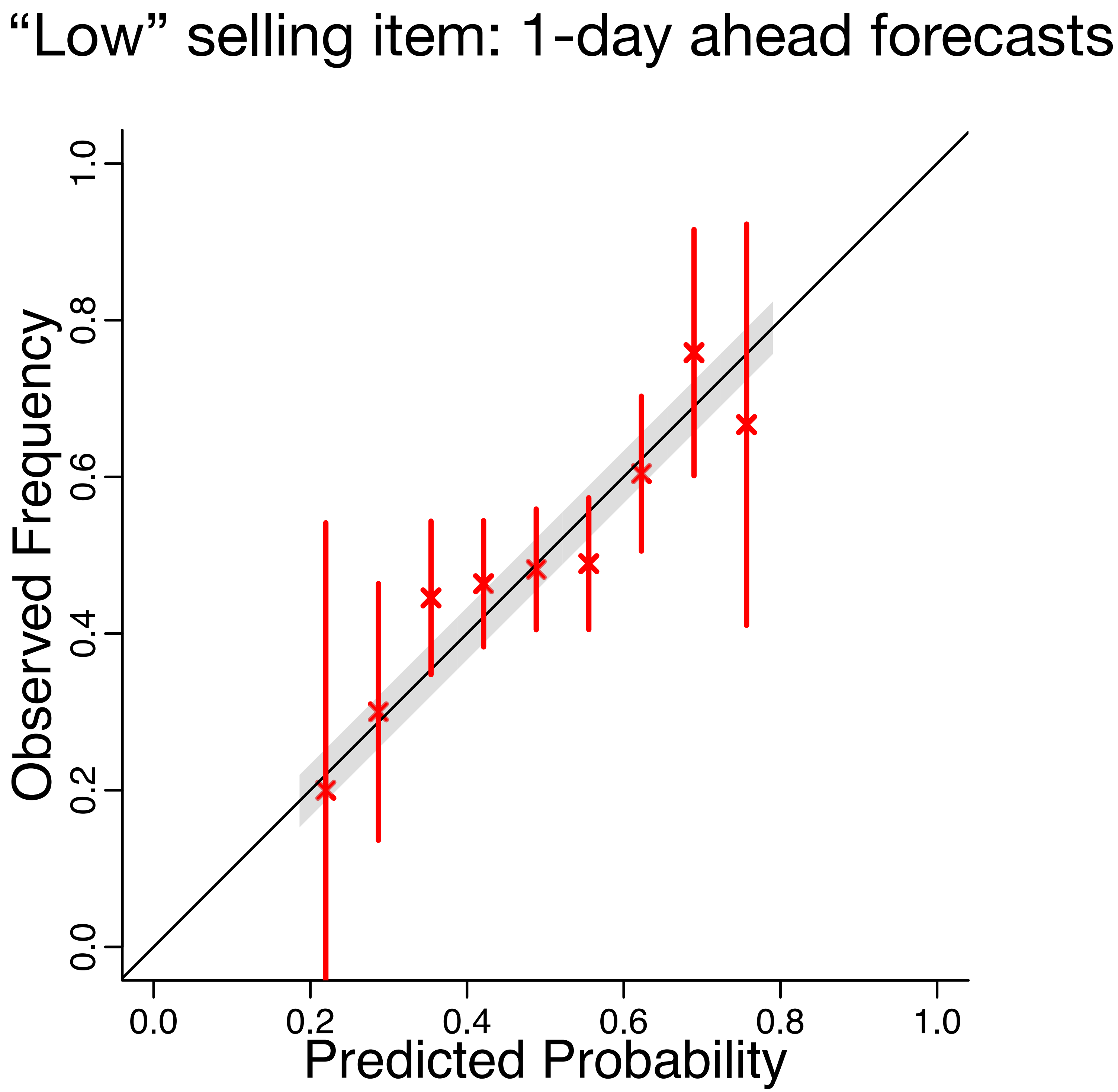
$$(\pi_{t+k} \mid D_t) \sim B(\alpha_t^0(k), \beta_t^0(k))$$

[Berry and West (2018)]

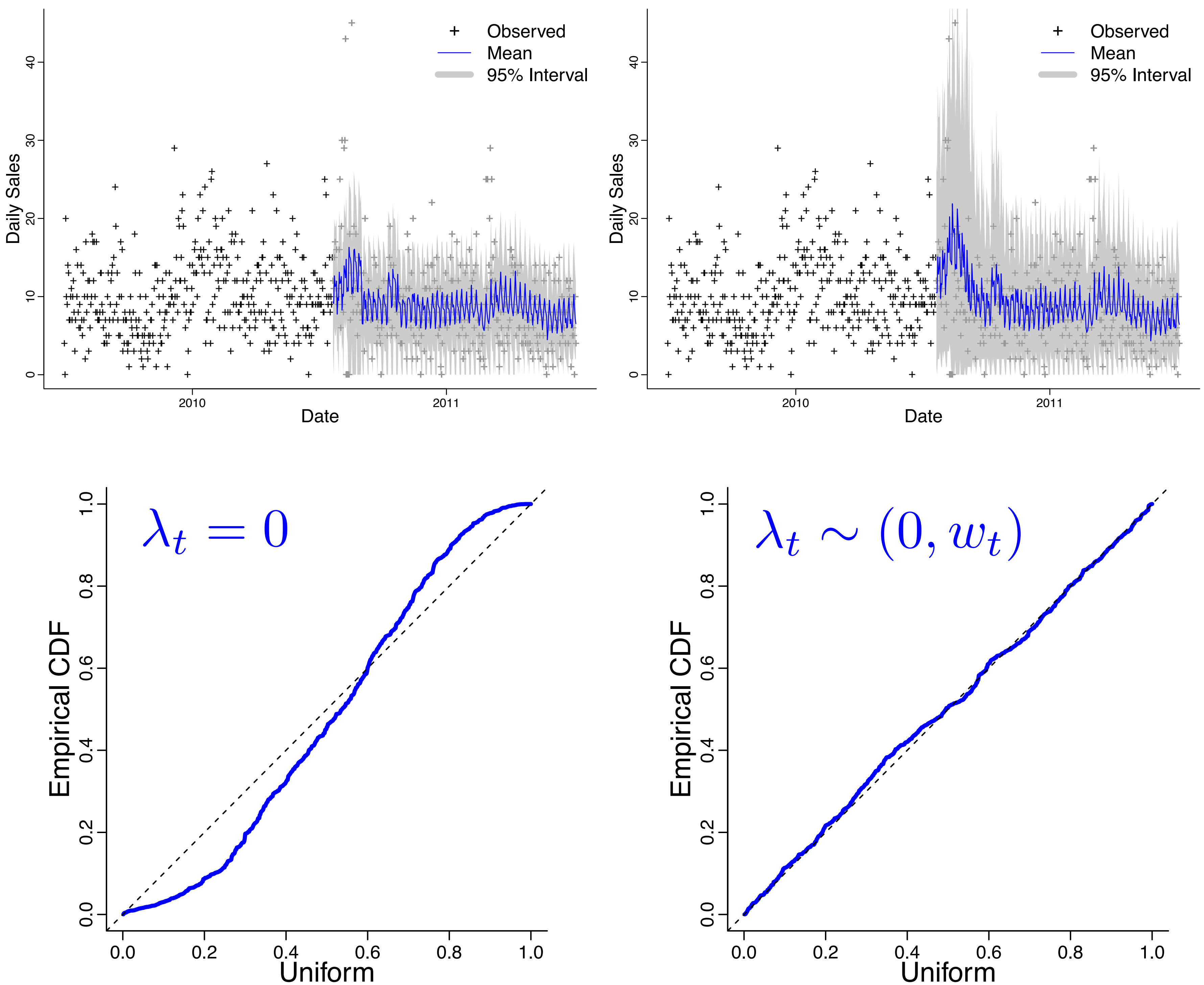
Count forecast evaluation

- Traditional point forecast metrics: RMSE, MAD, MAPE
  - Possibly undefined/infinite/misleading for low counts
- Full predictive distribution rather than point forecasts
  - Probabilistic evaluation: Calibration, coverage
  - 1:14 day forecasting: cumulative forecasts
- Decision analysis:
  - Rationalize accuracy metrics in specific context
  - Asymmetries, tail events
  - Stock-outs vs overstocking

Binary calibration: Predicting zero/non-zero



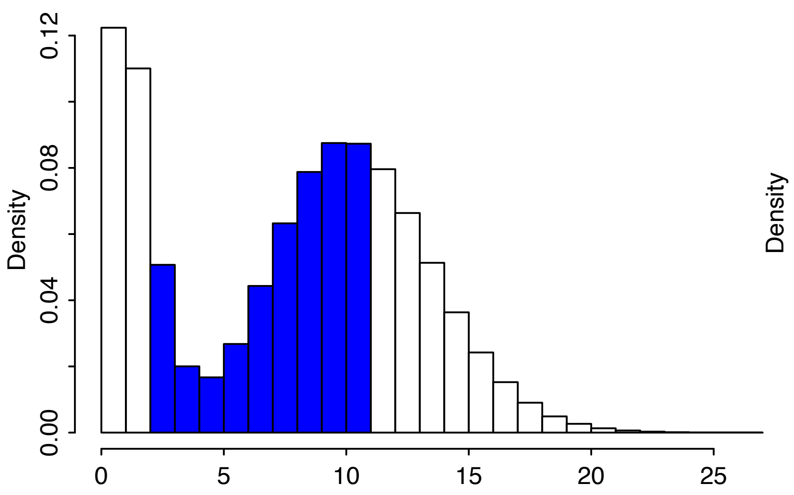
Probabilistic calibration and random effects

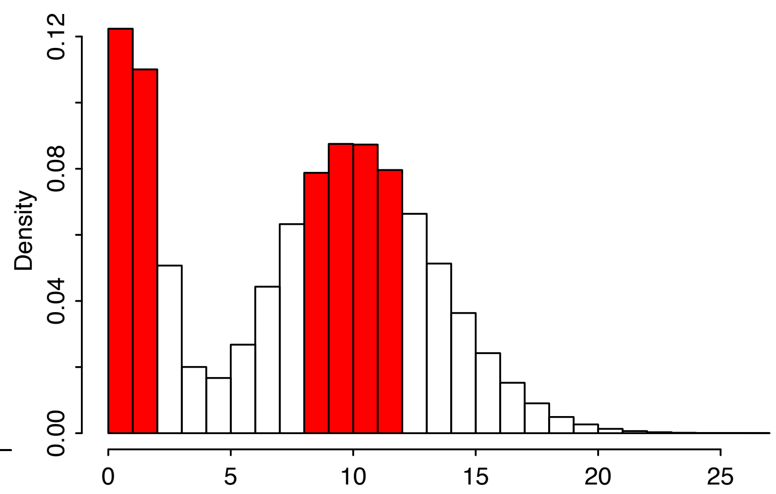


rPIT: Randomized probability integral transform:

- ordered predictive CDF values
- good model: "looks uniform"
- reveals areas of forecast deficiency

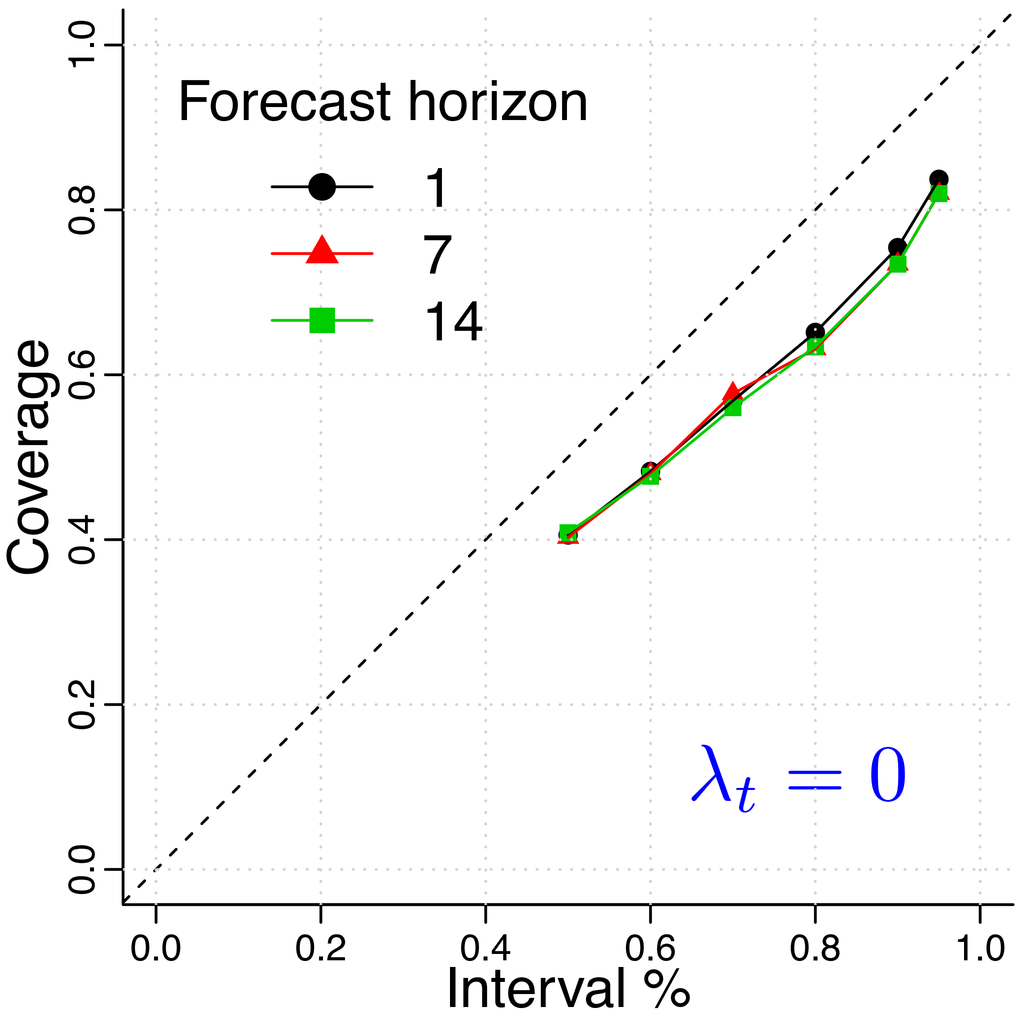
Coverage and random effects

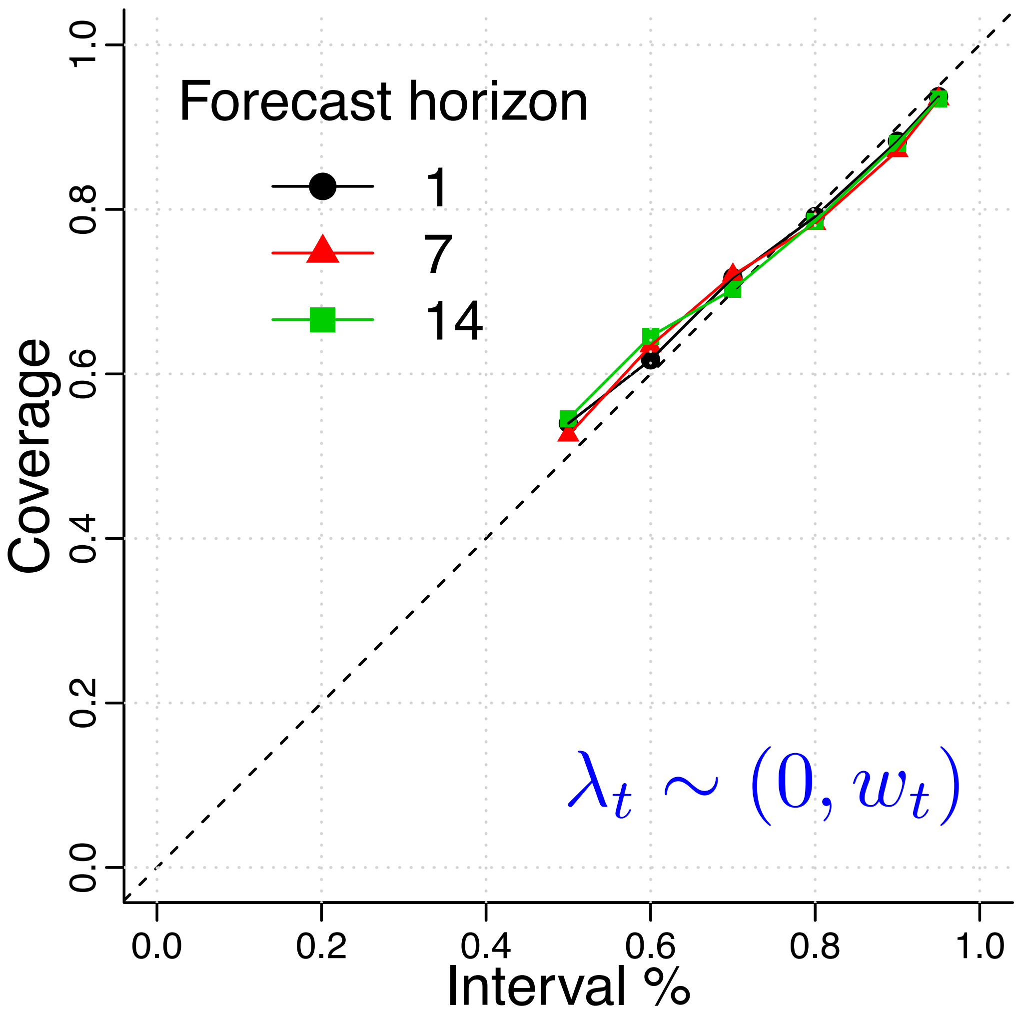




Low-count forecast uncertainty:

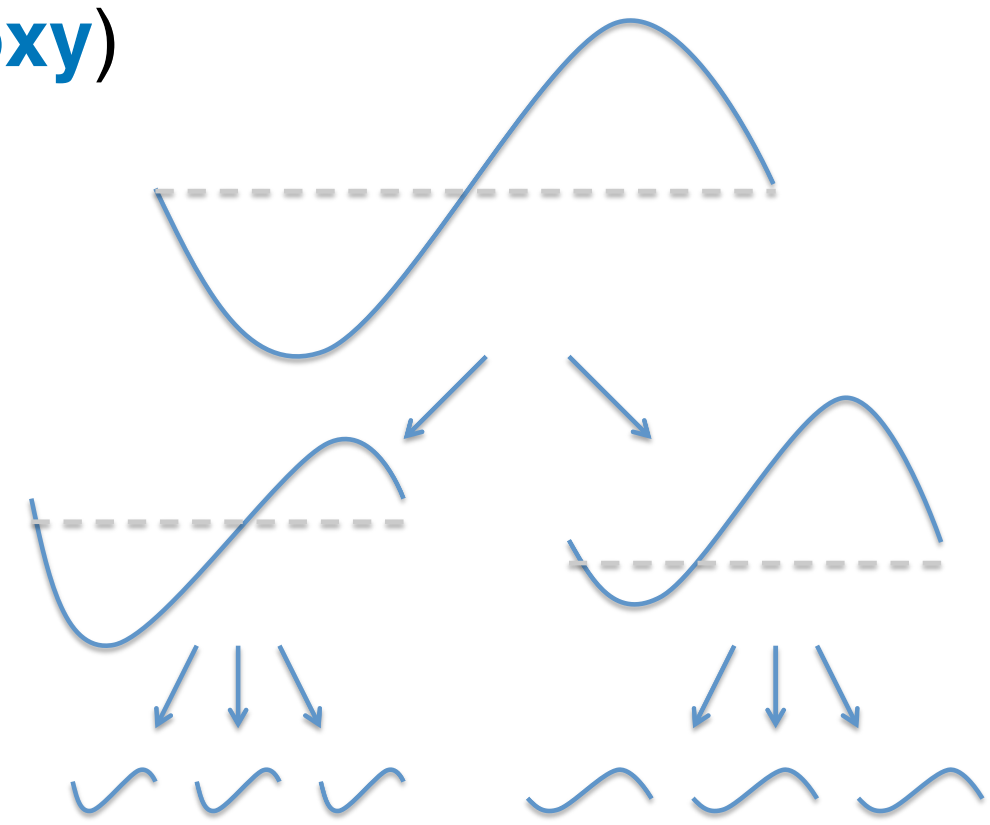
- **Central intervals**
- **HPD regions**





Multi-scale forecasting

- Improve item-level forecasting through improved prediction of aggregate daily effects (**“traffic” proxy**)
- Borrow strength on shared patterns
- Maintain **efficiency/scalability**
  - Avoid MCMC/SMC
- **Decouple**: univariate DCMMs
- **Recouple**: direct simulation
  - “Common” latent factor from external model
  - Series-specific adjustment of common factors



DCMM multi-scale framework

$y_{i,t}$: set of N series
$\mathcal{M}_i$: individual DCMMs
$\boldsymbol{\phi}_t$: latent factor (ex: seasonal, brand effect)

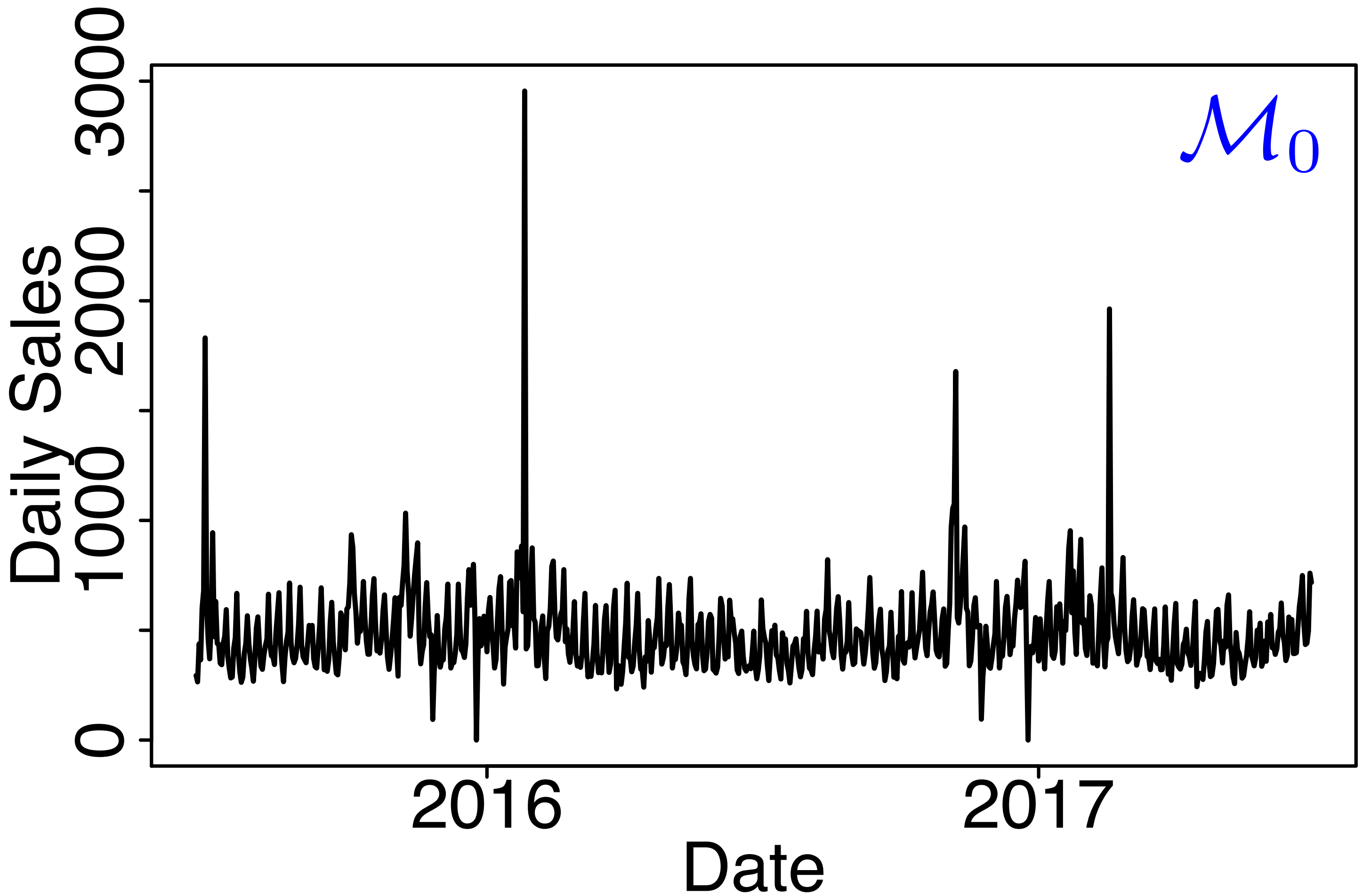
$$\mathcal{M}_i : \quad \boldsymbol{\theta}_{i,t} = \begin{pmatrix} \boldsymbol{\gamma}_{i,t} \\ \boldsymbol{\beta}_{i,t} \end{pmatrix}, \quad \mathbf{F}_{i,t} = \begin{pmatrix} \mathbf{f}_{i,t} \\ \boldsymbol{\phi}_t \end{pmatrix}, \quad i = 1, \ldots, N$$

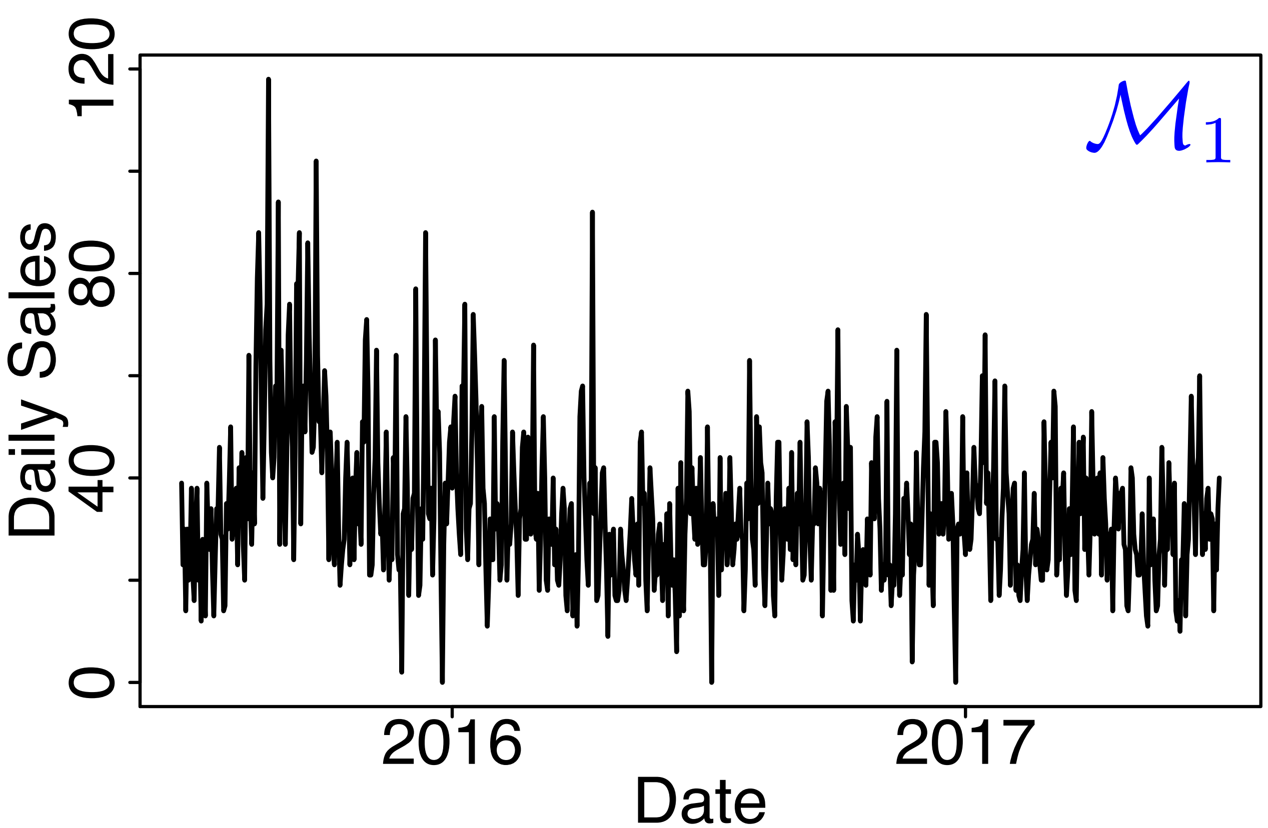
$$\lambda_{i,t} = \boldsymbol{\gamma}'_{i,t}\mathbf{f}_{i,t} + \boldsymbol{\beta}'_{i,t}\boldsymbol{\phi}_t$$

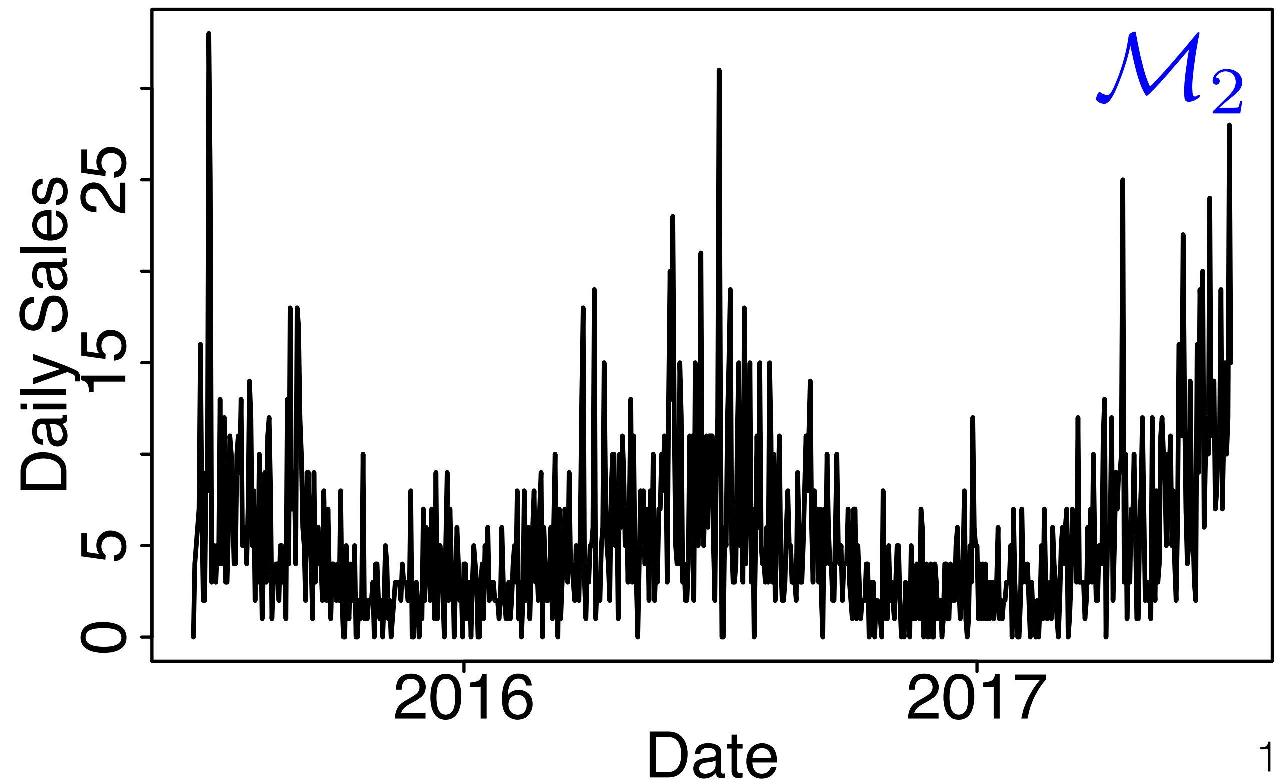
$$\mathcal{M}_0 : \text{ external DLM depending on } \boldsymbol{\phi}_t$$

- Simulation enables fully probabilistic use of external information
- Example $\mathcal{M}_0$: log-normal DLM of aggregate sales
- Parallel analysis of item-level DCMMs

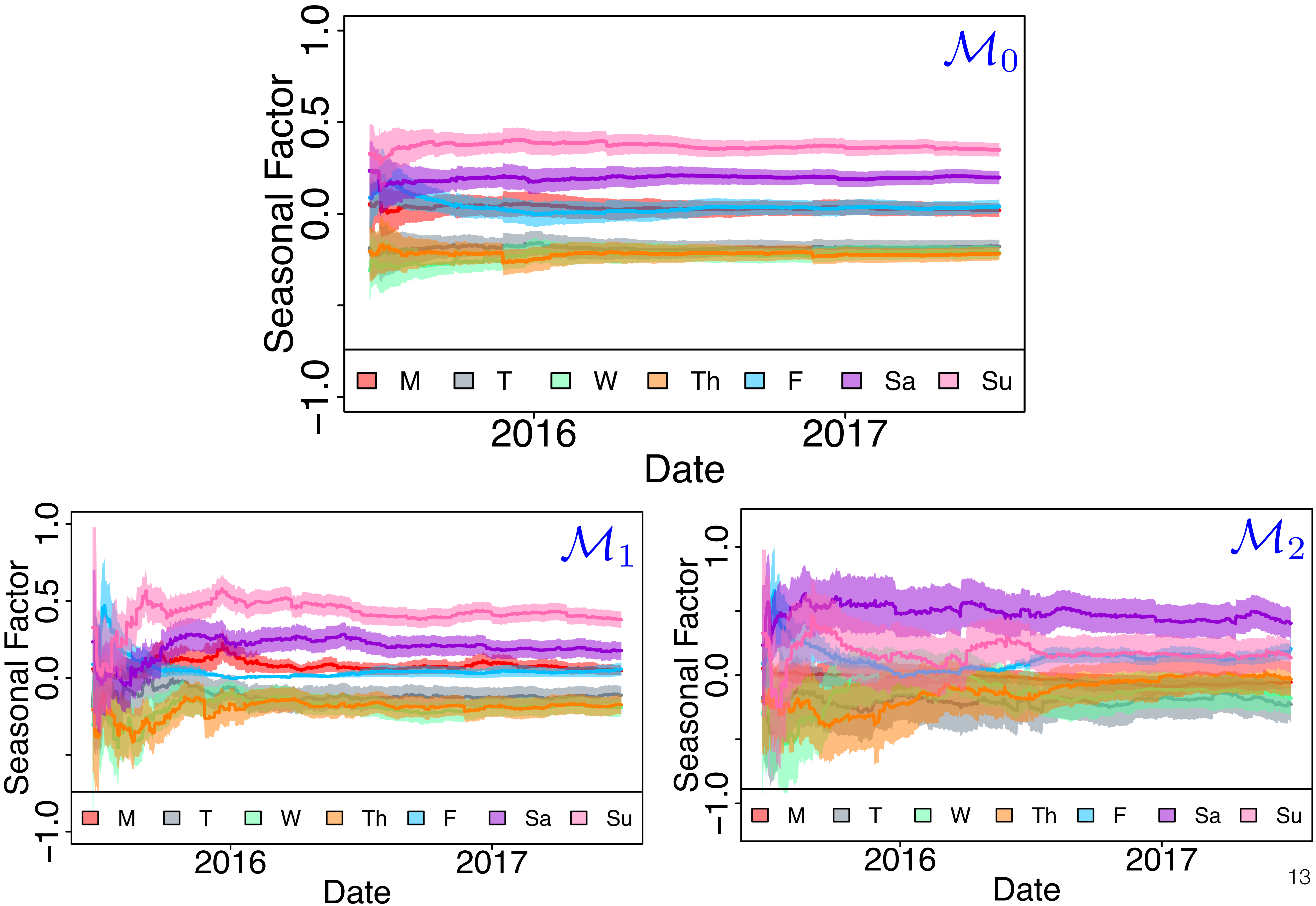
Multi-scale framework: aggregate and items



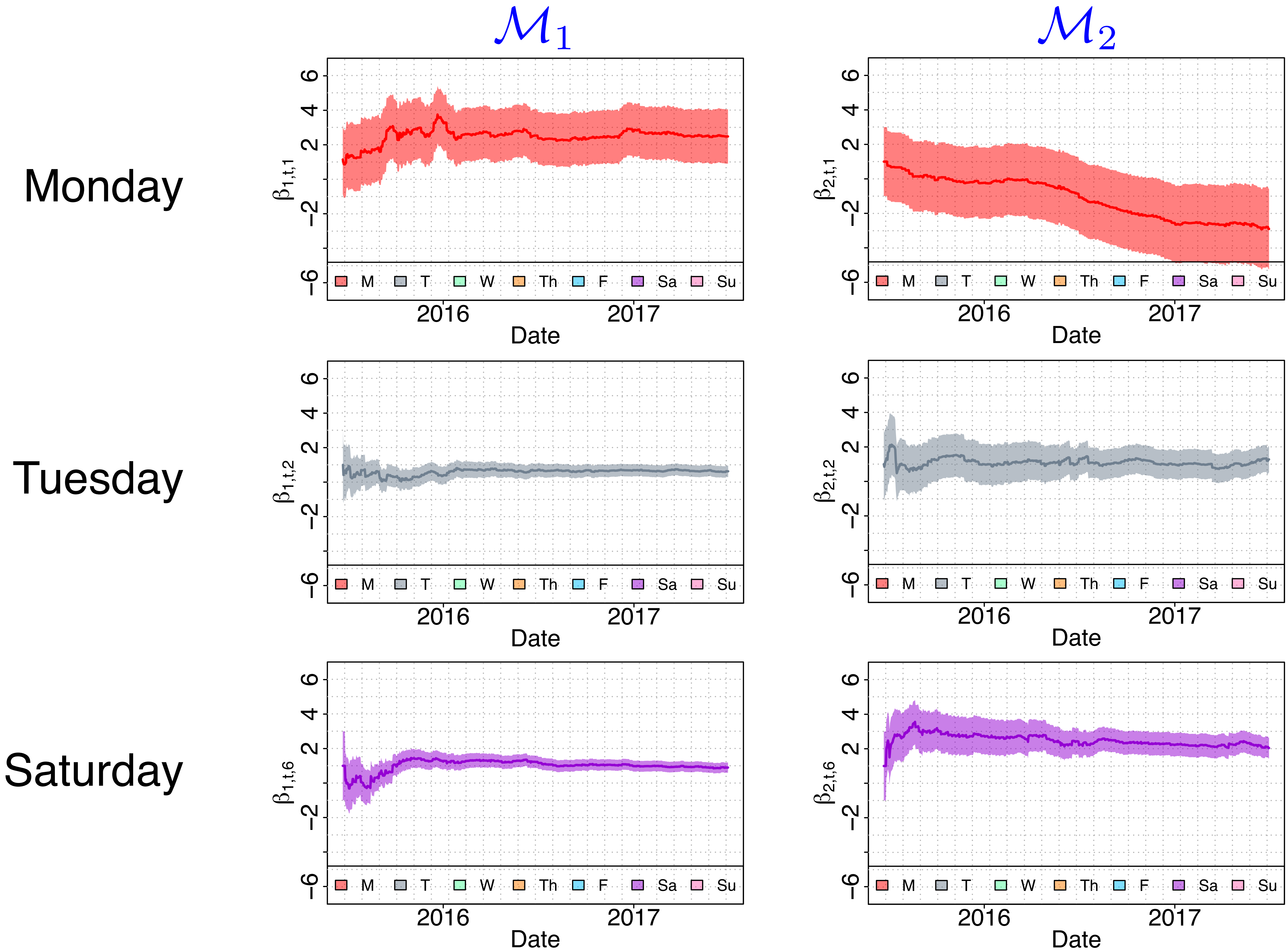




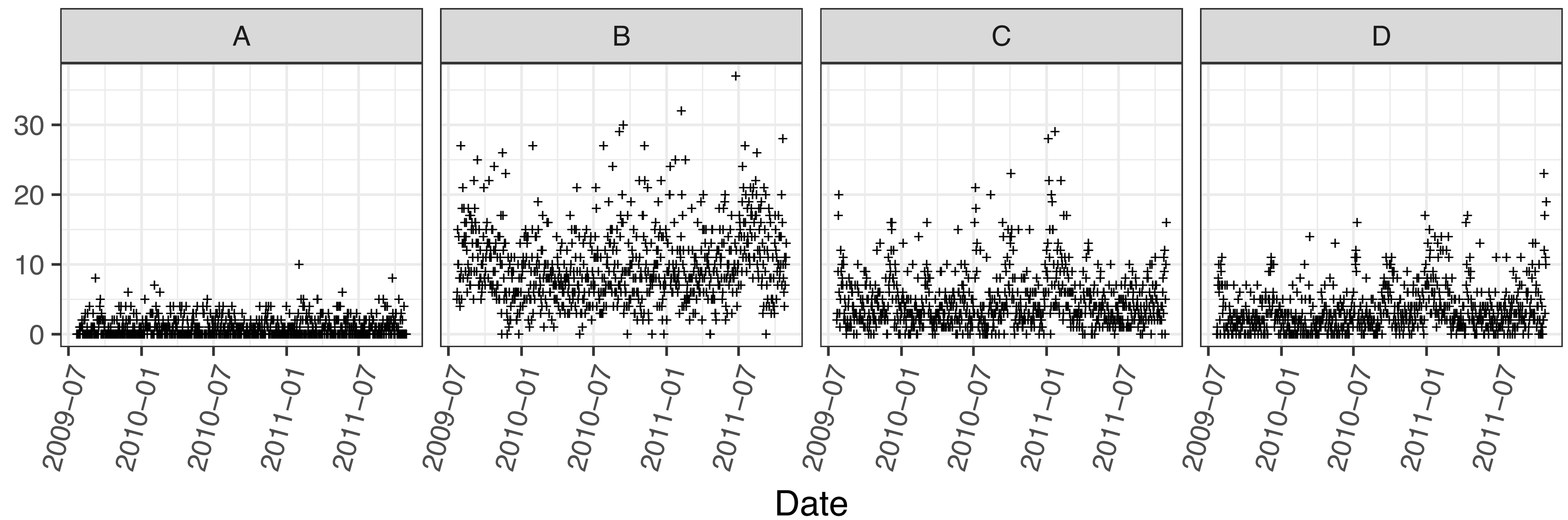
Multi-scale seasonal factors



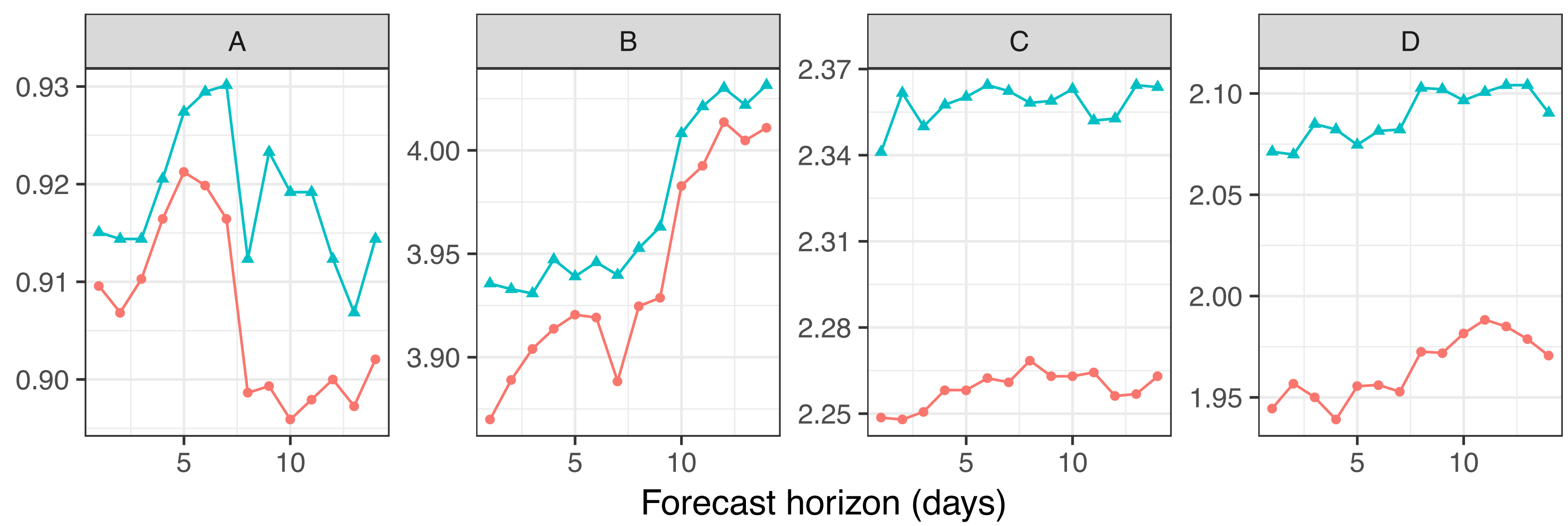
Multi-scale seasonal coefficients



Items A—D



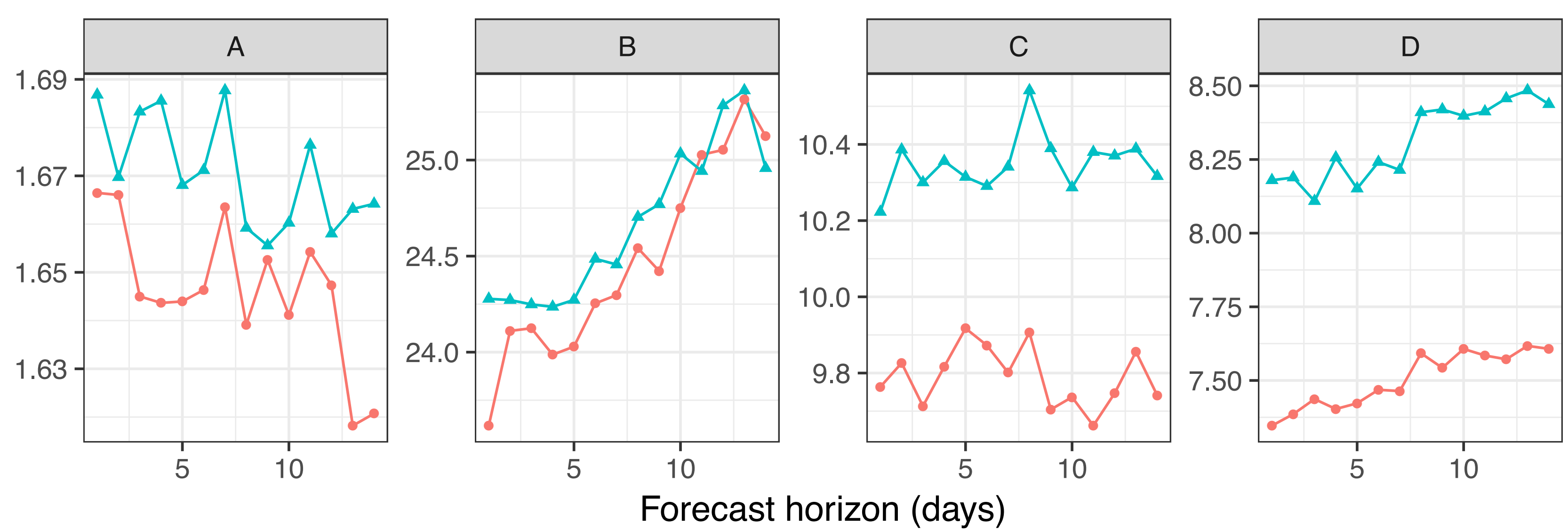
MAD



Ind DCMM

MS DCMM

DRPS

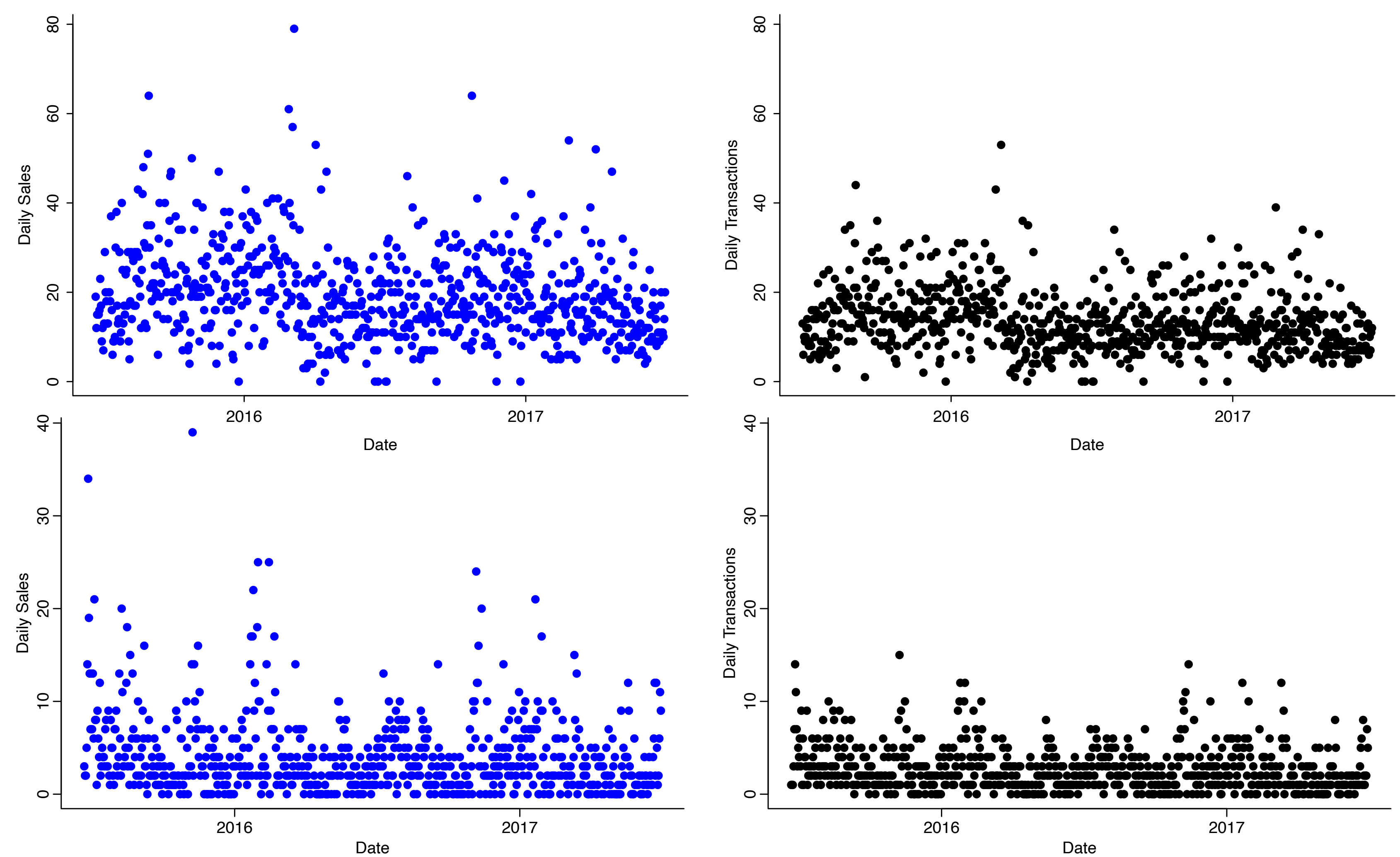


Dynamic models for transactions/sales time series

Sales | Transactions

Item A: high/mid seller

Item B: low/mid seller



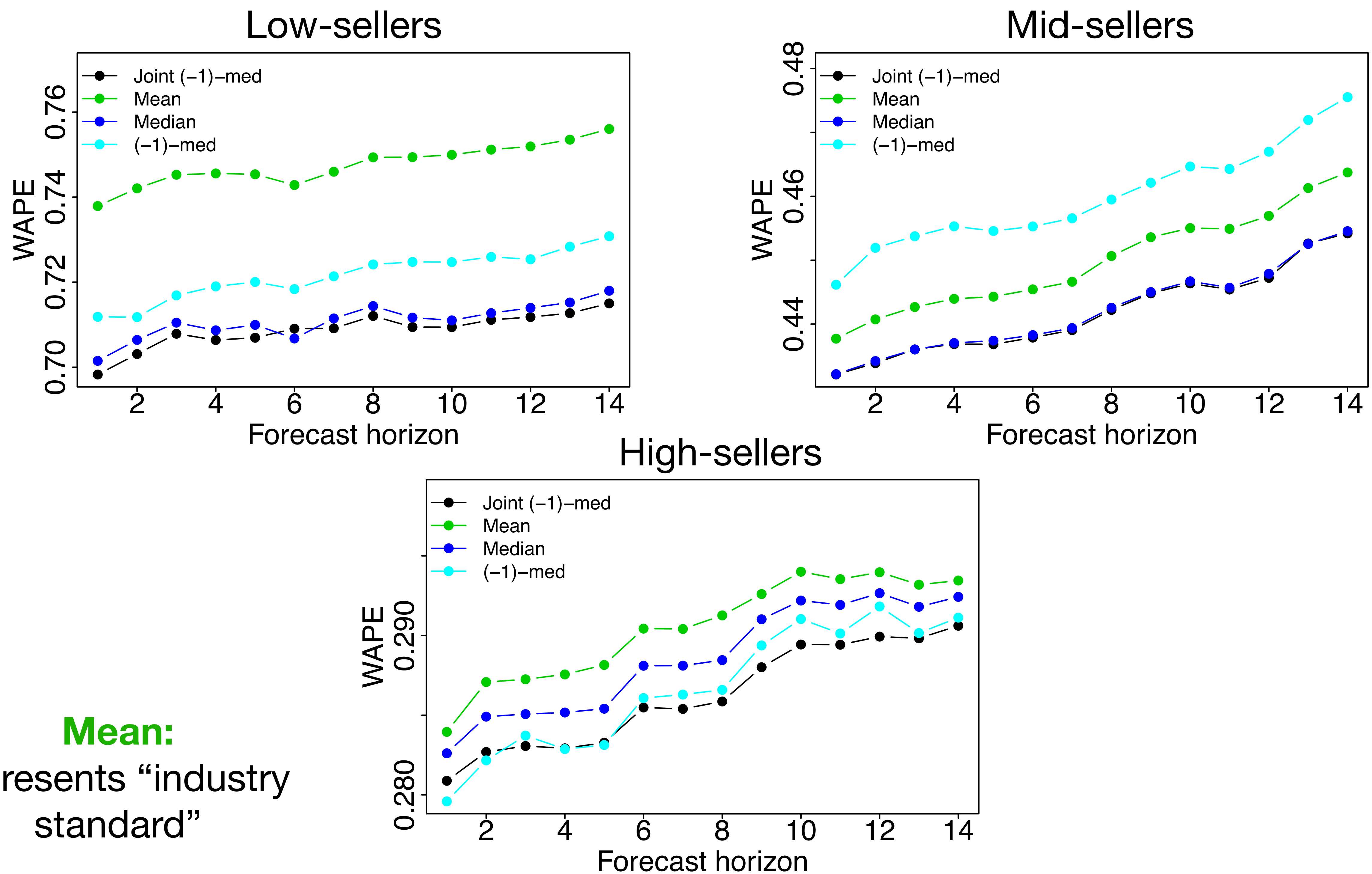
Item	Daily transactions			Sales-per-transaction		
	Mean	Median	Variance	Mean	Median	% < 5
A	19.75	18	101.15	1.44	1	99.0
B	4.66	3	18.70	1.53	1	98.4

Dynamic models for transactions/sales time series

- **Daily transactions:**
  - Multi-scale DCMM
  - Reduced random effect impact
- **Sales-per-transaction:**
  - Dynamic binary cascade model (DBCM)
    - Number of units sold per transaction
    - Sequence of conditional probabilities (‘cascade’)
      - “Given a customer buys >1 unit, what’s the probability they buy >2 units”
    - Customized to item
    - Admits “rare” events — e.g. 100 units/transaction

DBCM multi-item comparison

- 3 supermarkets, 32 UPCs = 96 store-UPC series
- Models trained for 365 days, forecast 1:14 days ahead for next 332 days
- Weighted Absolute Percent Error (WAPE) by demand group



Mean:
represents "industry standard"

Additional Resources:

- Berry, L. and West, M. (2019), "Bayesian forecasting of many count-valued time series," *Journal of Business & Economic Statistics*.
- Berry, L., Helman, P., and West, M. (2020), "Probabilistic forecasting of heterogeneous consumer transaction-sales time series," *International Journal of Forecasting*.
- PyBATS: Python package for Bayesian Analysis of Time Series by Isaac Lavine, Andrew Cron
    - Code: https://github.com/lavinei/pybats
    - Documentation: https://lavinei.github.io/pybats/

References

Berry, L. and West, M. (2019), “Bayesian forecasting of many count-valued time series,” *Journal of Business & Economic Statistics*. 1-34.

Berry, L., Helman, P., and West, M. (2020), “Probabilistic forecasting of heterogeneous consumer transaction-sales time series,” *International Journal of Forecasting*. 36(2), 552-569.

Hyndman, R. J. and Koehler, A. B. (2006), “Another look at measures of forecast accuracy,” *International Journal of Forecasting*, 22, 679-688.

Kolassa, S. (2016), “Evaluating predictive count data distributions in retail sales forecasting,” *International Journal of Forecasting*, 32, 788-803.

Snyder, R. D., Ord, J. K., and Beaumont, A. (2012), “Forecasting the intermittent demand for slow-moving inventories: a modeling approach,” *International Journal of Forecasting*, 28, 485-496.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York, Inc, 2nd ed.

West, M. and Harrison, P. J., and Migon, H. S. (1985), “Dynamic generalized linear models and Bayesian forecasting,” *Journal of the American Statistical Association*, 80, 73-83.