Dynamic Count Mixture Models (DCMM)

Summary and discussion of Berry, L. R., & West, M. (2019). Bayesian forecasting of many count-valued time series. *Journal of Business & Economic Statistics*, 1-15.

Team: Ed Tam, Yunran Chen, Alessandro Zito

Motivation

- Product sales/demand forecasting
- Meaningful for inventory management/production planning/…

Features:

- Multiple time granularities
- Many individual products
- Across multiple outlets

Ideal Models:

- Efficient
- Effective
- Flexible
- Scalable

Decouple/Recouple Idea

- **Decouple:** Product-level model (univariate DCMMs)
  - parallel computation
  - Flexible to handle features
  - Effective and efficient computationally

- **Recouple:** multivariate forecasting framework
  - Integration across potentially many products
  - Scalability to large-scale problems

pasta + ?

- Challenges
- Model

DGLM -> DCMM -> DCMM + RE

- Motivations
- Idea: Random effect model
- Analysis framework

Features of Product-level Data and Challenges

- Nonnegative counts
- Seasonality, holiday, price/promotion, …
- High-frequency: High variability + extreme values (across time and products) -> overdispersion
- Fine-scale resolution: many 0's (intermittent demand) and low counts -> Main issue

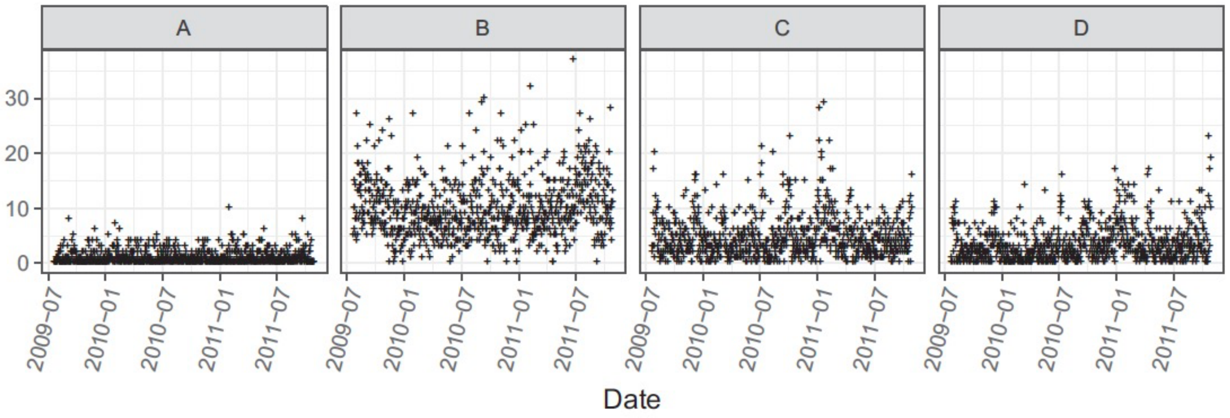


Figure 7. Daily unit sales (in counts per day) of four spaghetti items A–D in one store from July 22, 2009 to October 29, 2011.

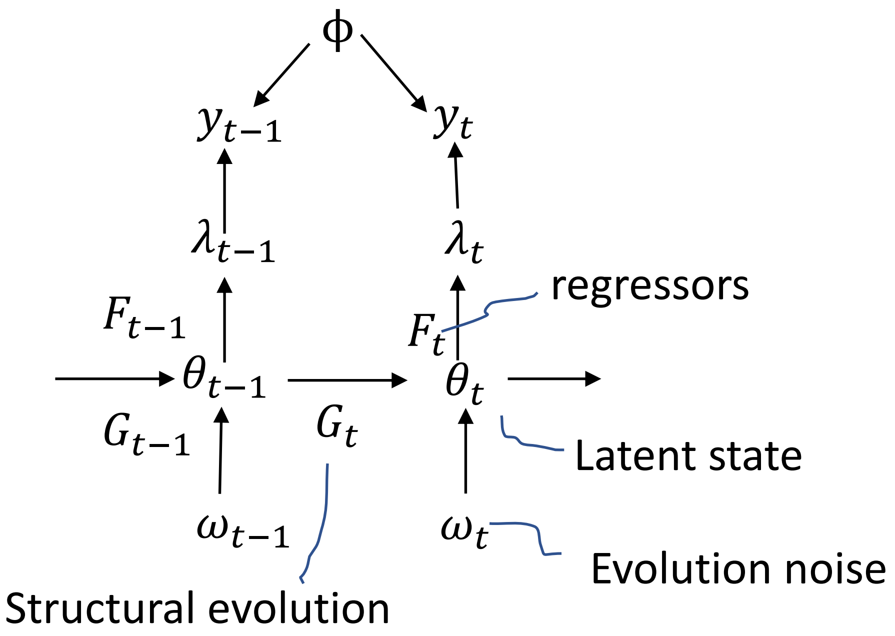
Table 1. Some summaries of daily pasta sales data by item.

Item	Mean	Median	Variance	% 0 sales
A	1.0	0	1.8	51.8
B	9.9	9	29.4	1.3
C	4.7	4	15.6	5.9
D	3.4	2	10.6	14.4

Dynamic Count Mixture Models (DCMMs)

Data: $y_t \quad t = 1, \ldots, T.$
Available Info at t: $\mathcal{D}_t = \{y_t, \mathcal{D}_{t-1}, \mathcal{I}_{t-1}\}$
Additional Info at t-1: $\mathcal{I}_{t-1}$

Dynamic Generalized Linear Models (DGLMs)



$$p(y_t \mid \eta_t, \phi) = b(y_t, \phi) \exp[\phi\{y_t\eta_t - a(\eta_t)\}],$$

$$\lambda_t = \mathbf{F}_t'\boldsymbol{\theta}_t \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim (\mathbf{0}, \mathbf{W}_t)$$

- Maintain pros of DGLMs
- Flexibility:
  - Allowing different F0 and F+
  - Varying freq of 0 sales across items and over time

Overdispersion ? -> DCMMs + random effect (discount factor)

Dynamic Count Mixture Models (DCMMs)

Bernoulli DGLMs

$$z_t \sim \mathrm{Ber}(\pi_t) \qquad z_t = \mathbb{1}(y_t > 0)$$

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

Conditionally Poisson DGLMs

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

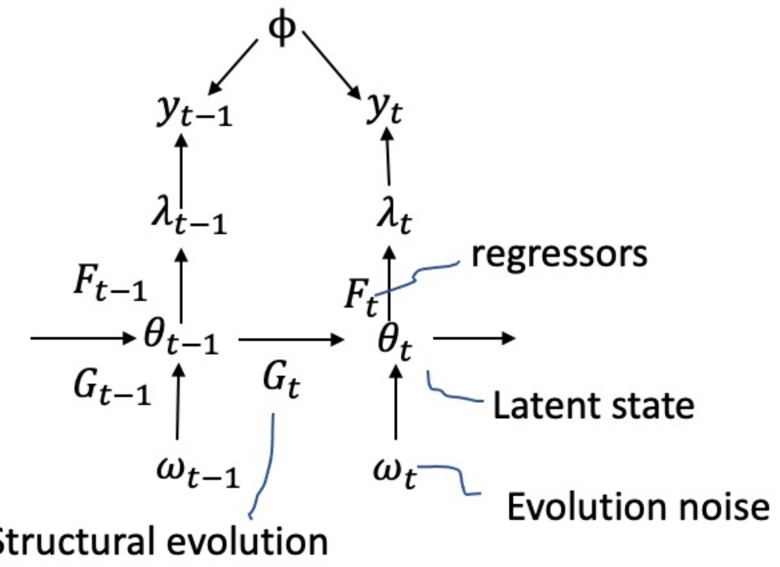
$$\log(\mu_t) = \mathbf{F}_t^{+\prime}\boldsymbol{\theta}_t$$

DCMM Random Effects Extension

- Overdispersion -> underestimate uncertainty

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

$$\log(\mu_t) = \mathbf{F}_t^{+\prime}\boldsymbol{\theta}_t \; + \zeta_t$$



- Reparameterization:

$$\log(\mu_t) = \mathbf{F}_{t,0}'\boldsymbol{\theta}_{t,0} + \zeta_t$$

$$\boldsymbol{\theta}_t = (\zeta_t, \boldsymbol{\theta}_{t,0}')'$$

$$\mathbf{F}_t = (1, \mathbf{F}_{t,0}')'$$

$$\log(\mu_t) = \mathbf{F}_t'\boldsymbol{\theta}_t = \mathbf{F}_{t,0}'\boldsymbol{\theta}_{t,0} + \zeta_t$$

- Discount factor:

$$q_{t,0} \equiv \mathrm{V}[\mathbf{F}_{t,0}'\boldsymbol{\theta}_{t,0} \mid \mathcal{D}_{t-1}, \mathcal{I}_{t-1}] = \mathbf{F}_{t,0}'\mathbf{R}_{t,0}\mathbf{F}_{t,0}.$$

$$v_t \equiv \mathrm{V}[\zeta_t \mid \mathcal{D}_{t-1}, \mathcal{I}_{t-1}] = q_{t,0}(1-\rho)/\rho.$$

$$\rho \in (0, 1]$$

- Close to 0: increased dispersion
- Set to 1: Poisson DGLM

Motivation for Multivariate Forecasting

pasta + ?

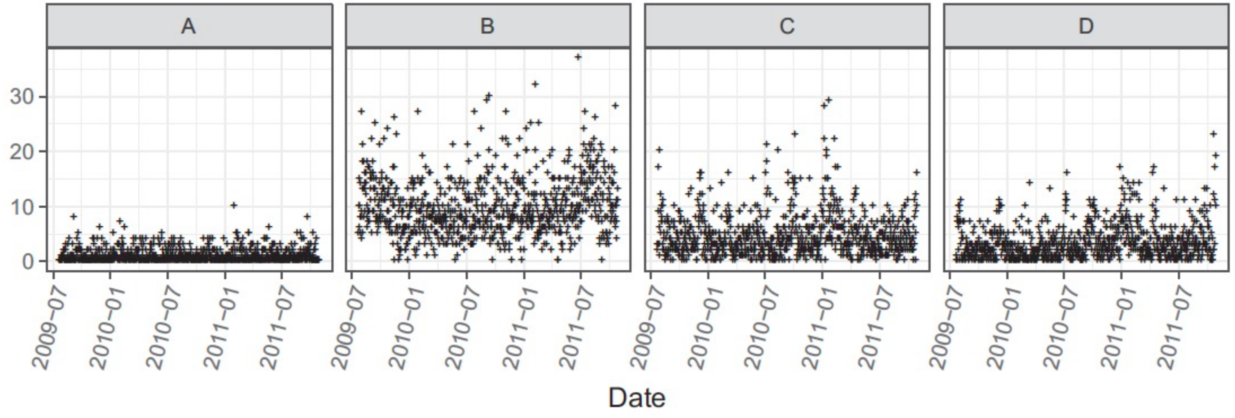


Figure 7. Daily unit sales (in counts per day) of four spaghetti items A–D in one store from July 22, 2009 to October 29, 2011.

- Information sharing:
  - product family, brand and store location
  - similar trends or seasonal patterns
  - aggregated level: more accurate and less noisy
- Important for sporadic data:
  - nonnegligible 0s and low count
  - more evident seasonality for high level sales items
  - gain more accurate predict for sporadic data

Multivariate Forecasting Framework (Top-down Recoupling)

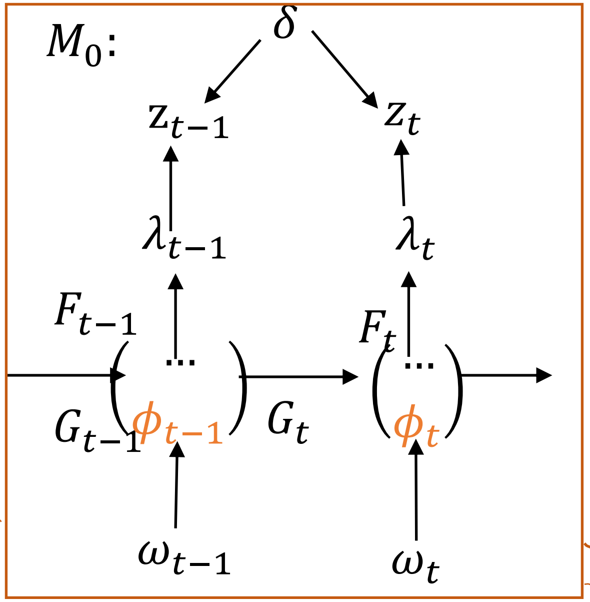
Hierarchical random effects model:

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

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

Common factor (latent factor)

Series-specific effects

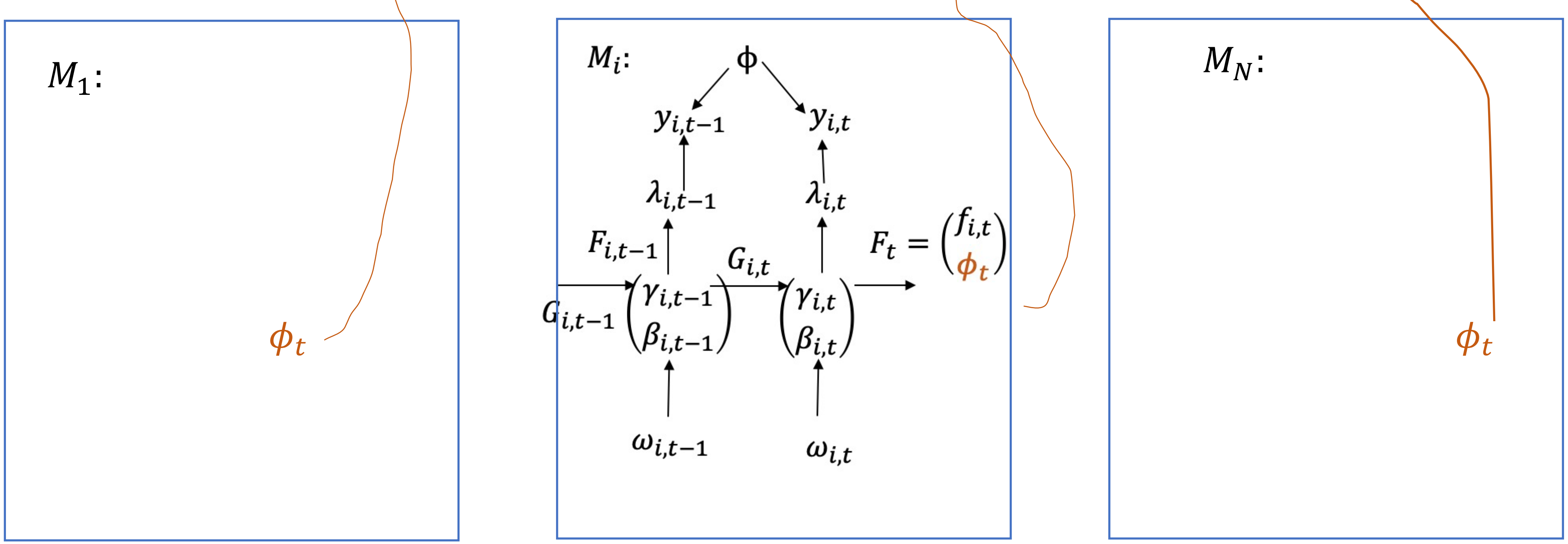


From M_0: $\boldsymbol{\phi}^s_{t:t+k} \sim (\boldsymbol{\phi}_{t:t+k}|\mathcal{D}_{0,t-1},\mathcal{I}_{0,t-1})$

Send $\boldsymbol{\phi}^s_{t:t+k}$. to M_i's: forecast $y_{i,t:t+k}|\boldsymbol{\phi}^s_{t:t+k},\mathcal{D}_{i,t-1},\mathcal{I}_{i,t-1}$

Posterior prob of ϕ_t

Posterior prob of $\boldsymbol{\theta}_{i,t}|\boldsymbol{\phi}^s_t, y_{i,t},\mathcal{D}_{i,t-1},\mathcal{I}_{i,t-1}$ marginalized over $\boldsymbol{\phi}^s_t$



Performance of DCMMs

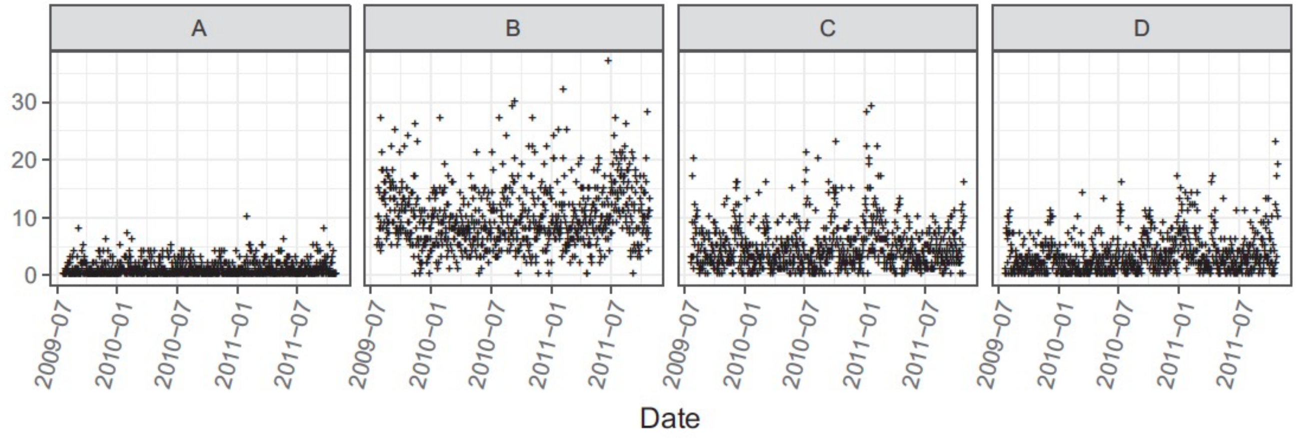


Figure 7. Daily unit sales (in counts per day) of four spaghetti items A–D in one store from July 22, 2009 to October 29, 2011.

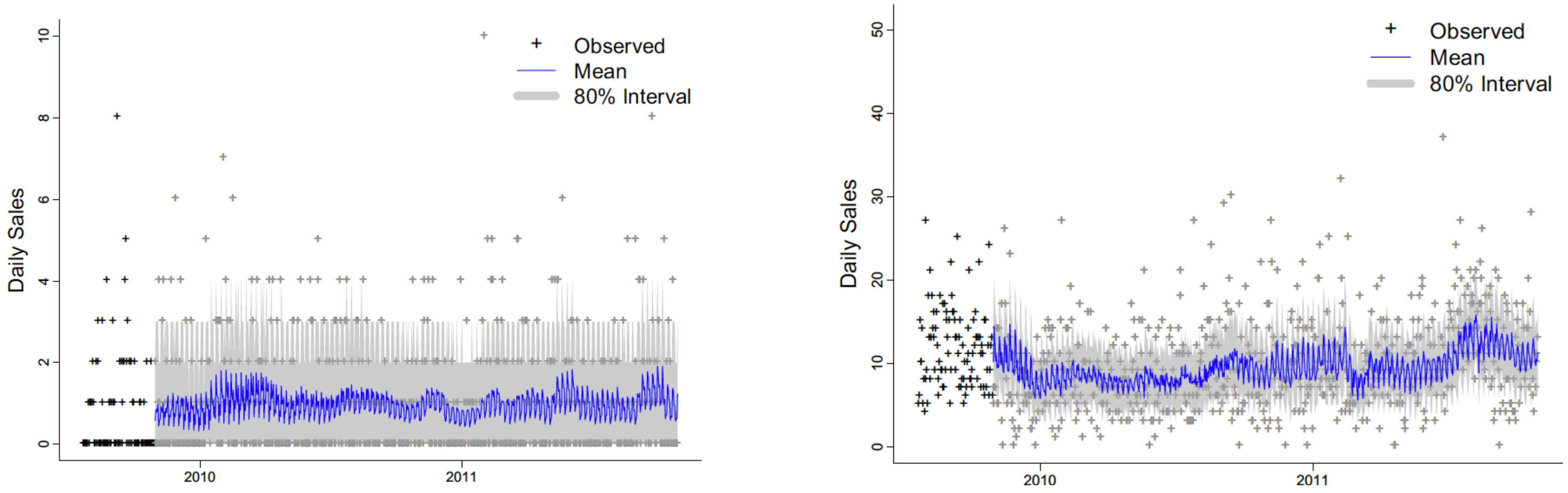


Figure 1. Data and aspects of 1-step ahead forecast distributions for items A (upper) and B (lower). Shading: 80% predictive credible intervals; full line: predictive mean.

Performance of DCMMs

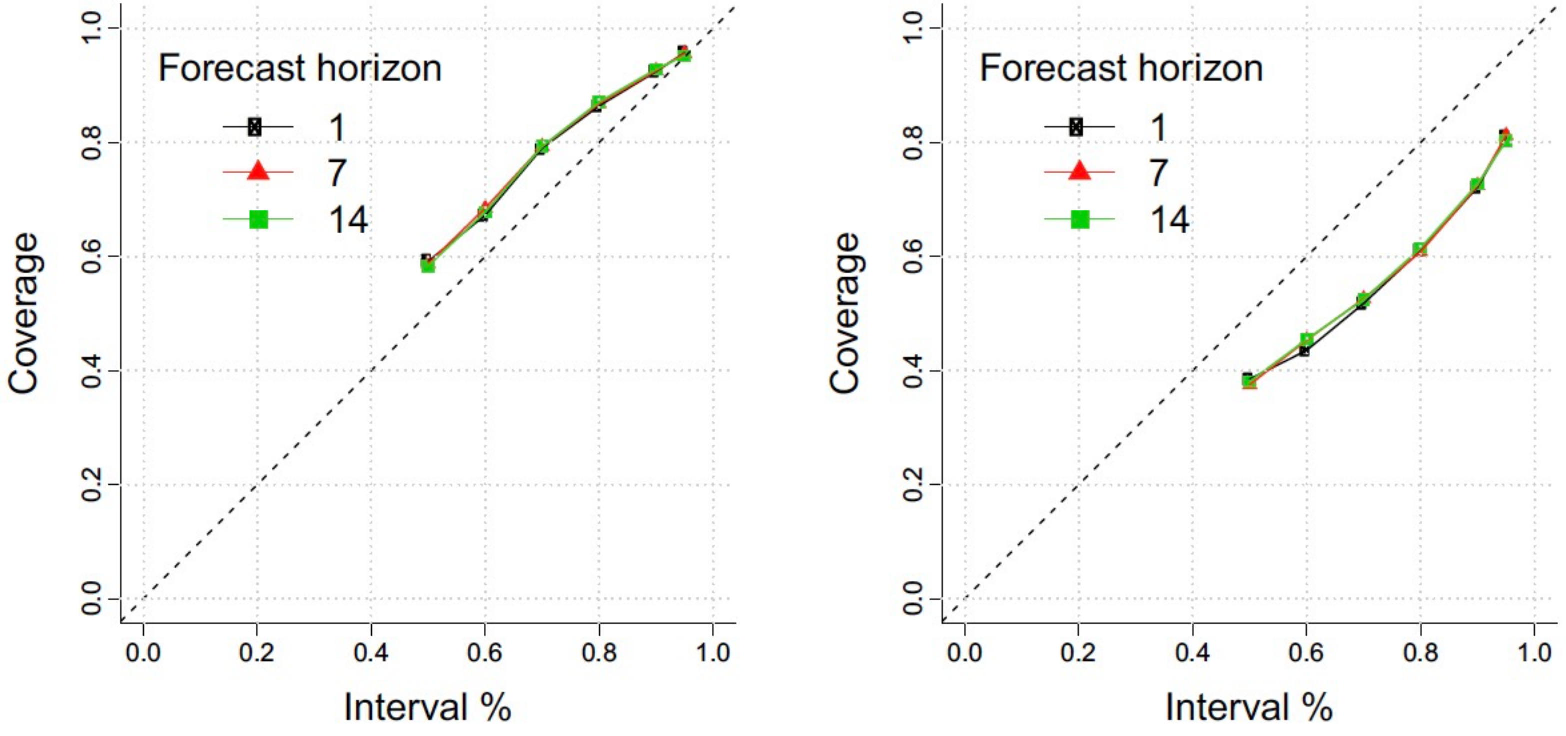


Figure 2. Coverage plots for items A (left) and B (right) from 1-, 7-, and 14-day ahead forecasts.

Empirical coverage over
the full year of forecasting for HPD CI

- Ideal: close to the line
- A: slight over-coverage
- B: evident under-coverage

(infrequent higher sales)

Performance of DCMMs

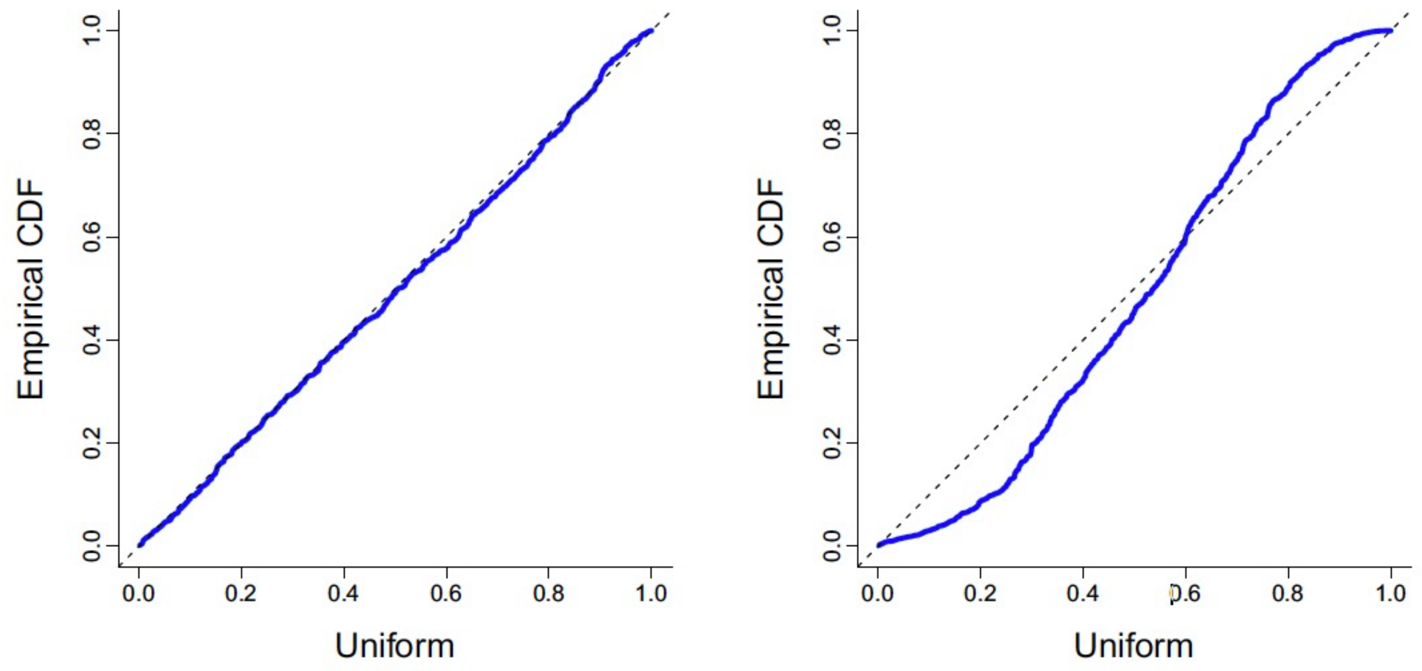


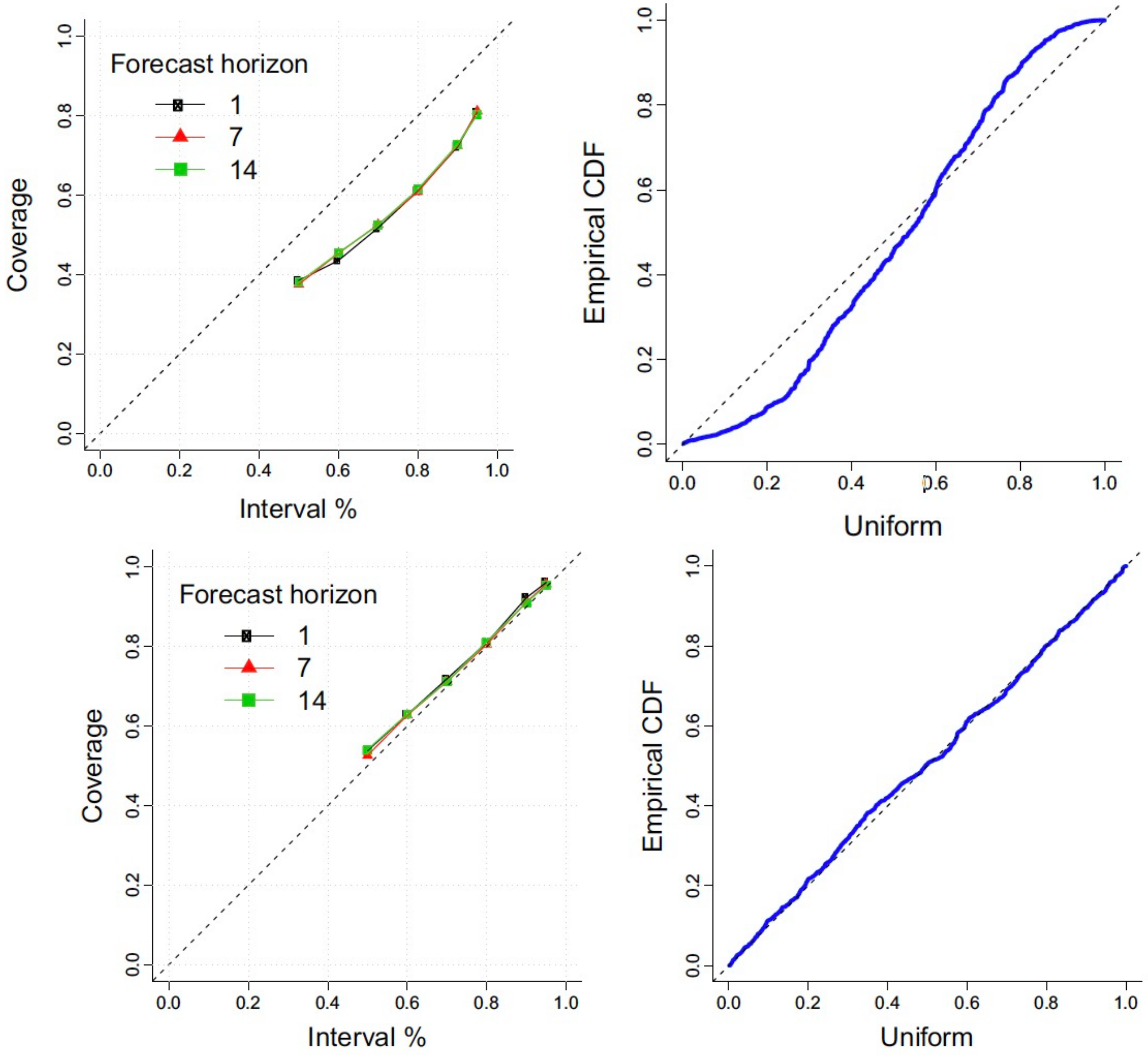
Figure 3. Randomized PIT plots from 1-day ahead forecasting of items A (left) and B (right). Full line: ordered randomized PIT values; dashed: 45° line.

Probabilistic integral transform (PIT):
General residual plot based on predictive cdf

- Ideal: close to the line
- A: strong uniformity
- B: significantly nonuniformity

(higher level of variations exist)

DCMMs+RE improved DCMMs



DCMMs

DCMMs+RE

Figure 6. Empirical coverage plot (left) and randomized PIT plot (right) for 1-day ahead forecasting of item B using the DCMM with random effects extension. Compare with the results under the basic DCMM in the right-hand frames of Figures 2 and 3.

Multivariate DCMMs improved univariate DCMMs

Lower values: Better performance

Multivariate DCMMs outperforms especially for moderate selling items

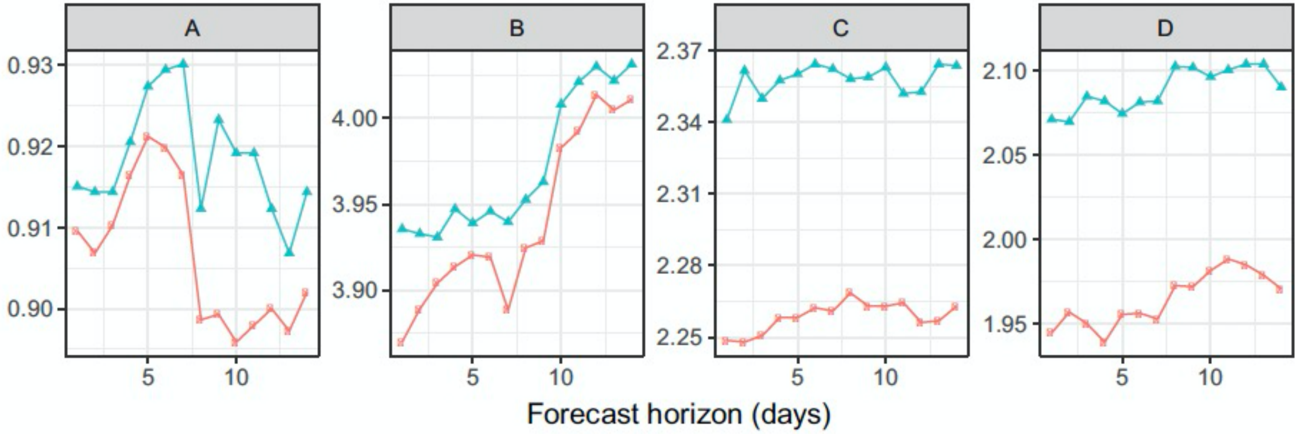


Figure 8. Mean absolute deviation (MAD) versus forecast horizon (days) for items A–D from the multiscale (orange circles) and benchmark (blue triangles) models.

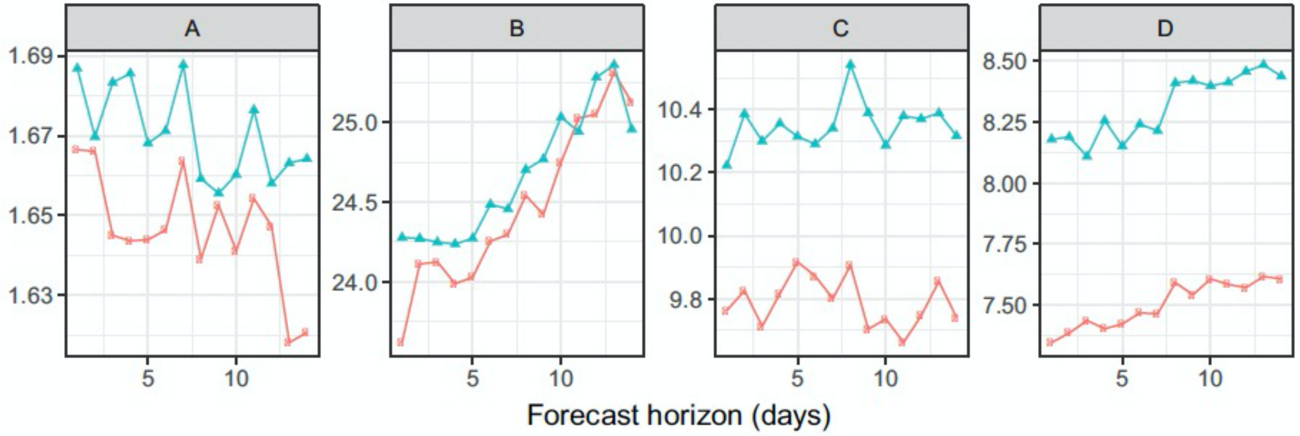


Figure 9. Mean rank probability score (MRPS) versus forecast horizon (days) for items A–D from the multiscale (orange circles) and benchmark (blue triangles) models.

Q & A