Two-Part Forecasting for Time-Shifted Metrics
Supplementary Material

Harrison Katz, Erica Savage, Kai Thomas Brusch

March 2025

1 A Bayesian Dirichlet Auto-Regressive Moving Average Model (B-DARMA)

1.1 Data Model

A J-component multivariate compositional time series is observed:

$$\mathbf{y}_t = (y_{t1}, \ldots, y_{tJ})',$$

indexed by $t = 1, \ldots, T$, where $0 < y_{tj} < 1$ and $\sum_{j=1}^{J} y_{tj} = 1$. A Dirichlet observation model is assumed:

$$\mathbf{y}_t \,\Big|\, \boldsymbol{\mu}_t,\, \phi_t \;\sim\; \text{Dirichlet}\big(\phi_t\, \boldsymbol{\mu}_t\big), \tag{1}$$

with $\boldsymbol{\mu}_t = (\mu_{t1}, \ldots, \mu_{tJ})'$, $\mu_{tj} > 0$, and $\sum_{j=1}^{J} \mu_{tj} = 1$. The scale parameter is $\phi_t > 0$. This leads to

$$f(\mathbf{y}_t \mid \boldsymbol{\mu}_t, \phi_t) \;\propto\; \prod_{j=1}^{J} y_{tj}^{\phi_t\, \mu_{tj} - 1}.$$

1.2 Additive Log-Ratio (alr) Link

Since $\boldsymbol{\mu}_t$ and $\mathbf{y}_t$ lie in the $(J-1)$-dimensional simplex, the additive log-ratio (alr) transform is applied:

$$\boldsymbol{\eta}_t = (\boldsymbol{\mu}_t) \;=\; \Big(\log \frac{\mu_{t1}}{\mu_{tj^*}},\; \ldots,\; \log \frac{\mu_{t,J-1}}{\mu_{tj^*}}\Big), \tag{2}$$

where j^* is a chosen reference component. Its inverse recovers μ_{tj}.

1.3 B-DAR(p, q) Process in alr-Space

A vector ARMA process is defined for $\boldsymbol{\eta}_t = (\boldsymbol{\mu}_t)$. In the most general form,

$$\boldsymbol{\eta}_t \;=\; \sum_{p=1}^{P} \mathbf{A}_p \left((\mathbf{y}_{t-p}) - \mathbf{X}_{t-p}\, \boldsymbol{\beta}\right) \;+\; \sum_{q=1}^{Q} \mathbf{B}_q \left((\mathbf{y}_{t-q}) - \boldsymbol{\eta}_{t-q}\right) \;+\; \mathbf{X}_t\, \boldsymbol{\beta}, \tag{3}$$

for $t = m+1, \ldots, T$, $m = \max(P, Q)$. The matrices $\mathbf{A}_p, \mathbf{B}_q$ and vector $\boldsymbol{\beta}$ are unknown parameters; $\mathbf{X}_t$ is a known matrix of covariates. In the paper, a B-DAR(1) model (i.e. $P = 1, Q = 0$) is used.

1.4 Scale Parameter ϕ_t

The scale parameter follows a log link:

$$\phi_t \;=\; \exp(\mathbf{z}_t\, \boldsymbol{\gamma}), \tag{4}$$

where $\mathbf{z}_t$ are additional covariates and $\boldsymbol{\gamma}$ is an unknown parameter.

1.5 Posterior Inference and Forecasting

Posterior inference proceeds by combining a prior for $\boldsymbol{\theta} = (\mathbf{A}_p, \mathbf{B}_q, \boldsymbol{\beta}, \boldsymbol{\gamma})$ with the likelihood from (1) and (3). MCMC sampling in `Stan` is performed. For forecasting, posterior draws of $\boldsymbol{\theta}$ generate future $\boldsymbol{\eta}_t$ and ϕ_t, and thus future compositions.

2 Airbnb Data Analysis: Model Implementation

2.1 B-DAR(1) for Lead-Time Distributions

In the Airbnb example:

- $J = 13$ lead-time buckets (0–12 months).
- $P = 1, Q = 0$ for B-DAR(1).
- $\mathbf{X}_t$ includes an intercept and 11 Fourier terms (monthly/annual seasonality), as well as a linear trend.
- $\mathbf{z}_t$ incudes the same seasonal terms, trend, and intercept as $\mathbf{X}_t$ for modeling ϕ_t via (4).

2.2 Priors and MCMC

Weakly informative normal priors are used for $\mathbf{A}_1$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$. HMC sampling in `Stan` provides posterior draws and posterior predictive samples.

2.3 Forecasting Steps

1. Fit a univariate model for total bookings (Part 1 in the paper).
2. Compute observed proportions $\mathbf{y}_t$ for each lead-time bucket.
3. Fit B-DAR(1) to these $\mathbf{y}_t$ (Part 2).
4. Generate out-of-sample composition forecasts.
5. Multiply total bookings by forecasted proportions, shifting by lead time to the trip-date axis.

3 Monthly Prophet Benchmark

A monthly-level Prophet approach was also tested to compare with the daily bottom-up Prophet benchmark. Monthly bookings were computed by summing daily counts for each of the 13 lead-time buckets (0–12 months). Each lead-time series was then modeled with Prophet, after which forecasts were summed to obtain total monthly bookings.

3.1 Implementation Details

- **Aggregation.** For each month m,

  $$\text{Bookings}_{m,\ell} = \sum_{t \in m} \text{DailyBookings}_{t,\ell},$$

  where $\ell \in \{0, 1, \ldots, 12\}$.

- **Prophet Setup.** Annual seasonality and holiday/event effects were included, matching the daily Prophet model where possible, but using monthly frequency.

- **Forecast Window.** January–December 2019 served as the test period, with 2014–2018 used for training.

3.2 Results and Comparison

3.2.1 Daily vs. Monthly Prophet vs. B-DARMA

Table 1 presents a consolidated view of the three forecasting methods in the test window of 2019: (1) Two-Part B-DARMA, (2) Daily Bottom-up Prophet, and (3) Monthly Prophet. The table includes mean absolute error (MAE), mean absolute percentage error (MAPE), and normalized L1 distance for lead-time distributions.

Table 1: Comparison of Daily Bottom-up Prophet, Two-Part (B-DARMA), and Monthly Prophet Forecasts (Test Window: 2019)

City	Method	MAE	MAPE	Normalized L1
A	B-DARMA (Two-Part)	5083	4.80%	0.0229
A	Daily Prophet (Bottom-up)	5336	5.07%	0.0389
A	Monthly Prophet (Botom-up)	7522	7.15%	0.0395
B	B-DARMA (Two-Part)	1406	3.07%	0.0300
B	Daily Prophet (Bottom-up)	1455	3.15%	0.0499
B	Monthly Prophet (Bottom-up)	2047	4.41%	0.0498

3.2.2 Discussion of Monthly Results

For City A, the Monthly Prophet approach yields an MAE of 7522 and MAPE of 7.15%, both of which exceed those of Daily Prophet (5336, 5.07%), indicating higher total bookings error. The monthly approach also has a normalized L1 of 0.0395, which is slightly *higher* than the 0.0389 from the daily model, suggesting slightly worse lead-time distribution accuracy.

For City B, the Monthly Prophet approach results in an MAE of 2047 and a MAPE of 4.41%, again higher than the daily model's MAE (1455) and MAPE (3.15%). However, the normalized L1 for City B is 0.0498, which is almost identical to the daily model's 0.0499. Hence, monthly aggregation for City B offers no appreciable gain in compositional accuracy and comes at the cost of reduced accuracy in total demand.

Across both markets, the Two-Part B-DARMA method remains the top performer, showing notably lower MAE, MAPE, and normalized L1 compared to either the daily or monthly Prophet benchmarks.

Figures 1–2 shows the monthly Prophet and B-DARMA forecasts by lead-time bucket and the normalized L1 distance.

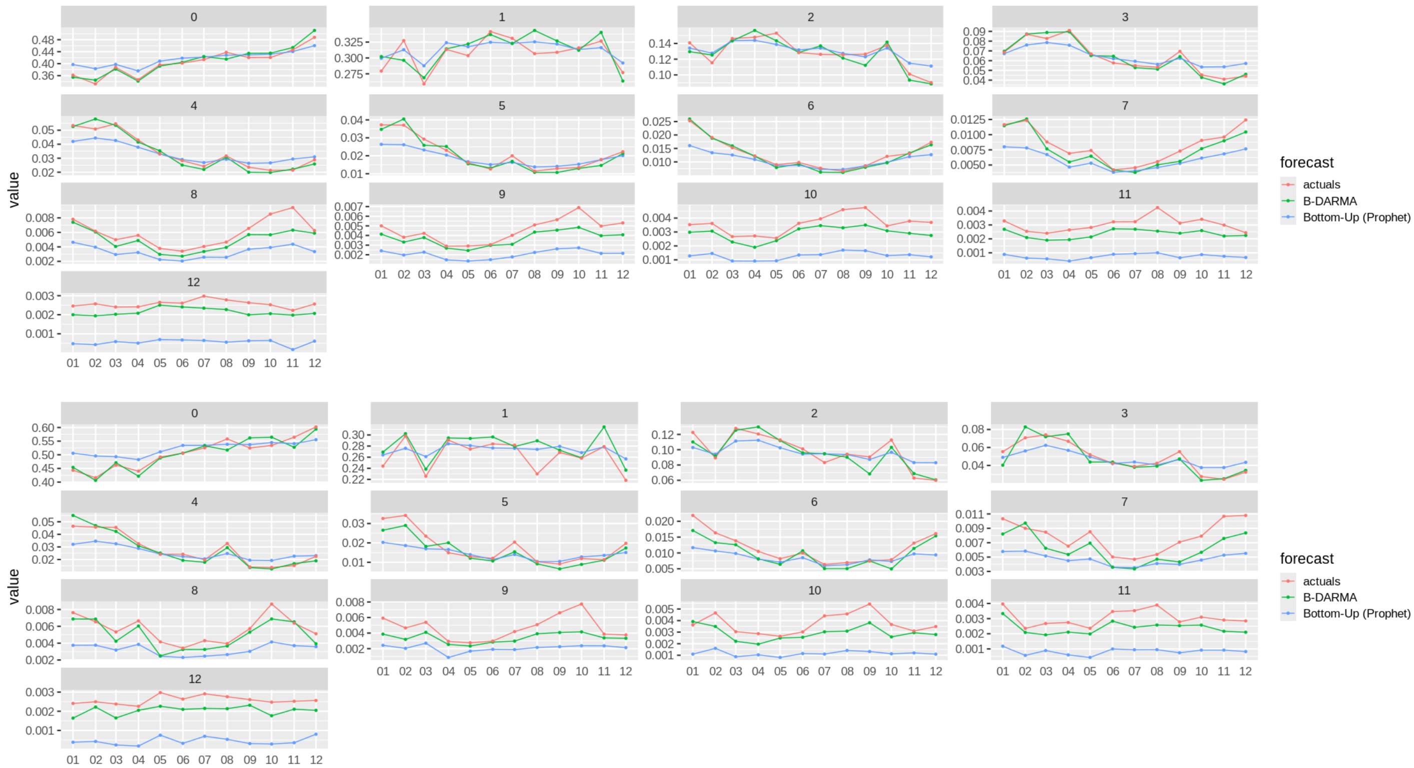

Figure 1: Monthly Prophet forecasts (lines) and actuals (points) by lead-time bucket (0–12).

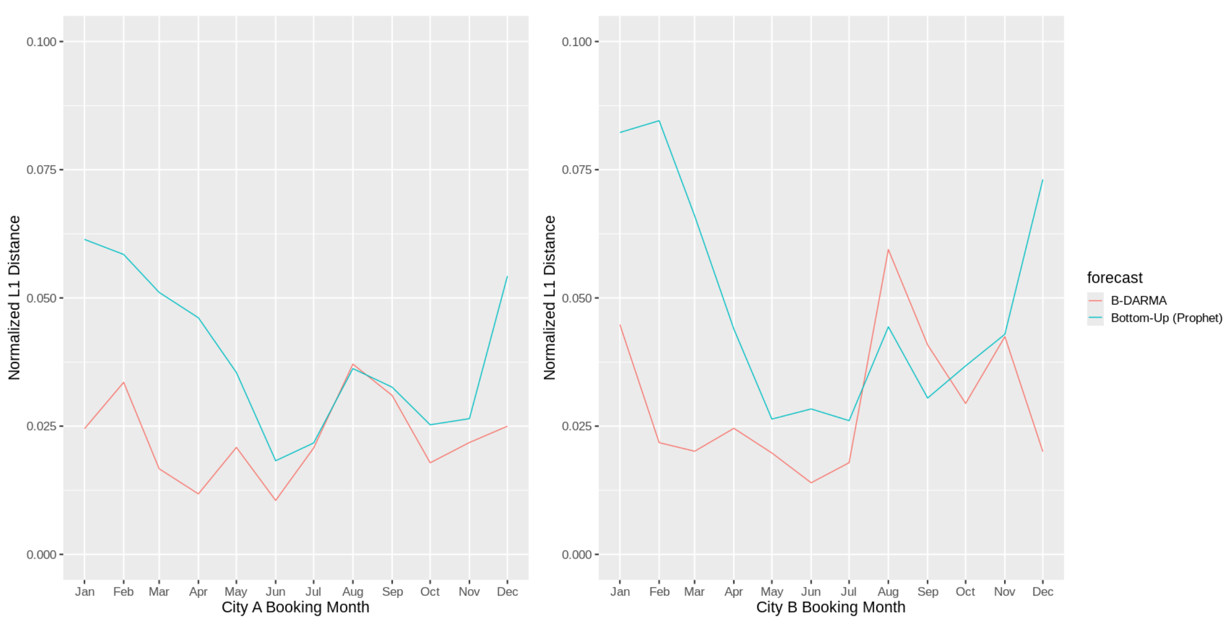

Figure 2: Normalized L1 distance for monthly Prophet forecasts. Lower values indicate closer agreement with actual lead-time proportions.

References (Selected)

- Katz, H., Brusch, K. T., & Weiss, R. E. (2024). A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times. *International Journal of Forecasting*, 40(4), 1556–1567.
- Zheng, T., & Chen, R. (2017). Dirichlet ARMA models for compositional time series. *Journal of Multivariate Analysis*, 158, 31–46.
- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45.