Title

Advancing Bayesian Forecasting: A Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity Model, a Bayesian Dirichlet Auto-Regressive Moving Average Model, and Other Innovations

Author

Katz, Harrison

Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Advancing Bayesian Forecasting:
A Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity Model, a Bayesian Dirichlet Auto-Regressive Moving Average Model, and Other Innovations

A dissertation submitted in partial satisfaction

of the requirements for the degree of Doctor of Philosophy in Statistics

by

Harrison Katz

2025

ABSTRACT OF THE DISSERTATION

Advancing Bayesian Forecasting:
A Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity Model,
a Bayesian Dirichlet Auto-Regressive Moving-Average Model,
and Other Innovations

by

Harrison Katz

Doctor of Philosophy in Statistics
University of California, Los Angeles
2025

Professor Robert E. Weiss, Co-Chair
Professor Ying Nian Wu, Co-Chair

This dissertation introduces new Bayesian time series models for compositional and high-dimensional data with dynamic structure and potential heteroskedasticity. It comprises four papers that offer methodological advances, simulation results, and applications in hospitality and finance.

Paper 1 introduces the Bayesian Dirichlet Auto-Regressive Moving Average (B-DARMA) model, motivated by the need to forecast the proportion of future fees recognized in future monthly intervals using daily Airbnb data. By embedding Auto-Regressive Moving Average (ARMA) components on an additive log-ratio scale within a Dirichlet likelihood, the model enforces compositional constraints and yields reasonable forecasts. Simulation studies highlight B-DARMA's predictive performance, and empirical analysis shows more accurate lead-time predictions compared with standard VARMA-based methods—vector auto-regressive moving average models that jointly capture relationships among multiple time

series—thereby guiding resource allocation and strategic planning.

Paper 2 extends B-DARMA to a Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity (B-DARCH) model by incorporating a Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH)-like process for the Dirichlet precision parameter. Empirical analysis of Airbnb's currency-fee data demonstrates that B-DARCH achieves higher forecast accuracy than both B-DARMA and standard VARMA-based methods. Simulation studies further confirm that modeling time-varying volatility significantly improves predictive coverage relative to simpler B-DARMA or transformed VARMA models.

Paper 3 conducts a sensitivity analysis of B-DARMA under several priors—normal, Laplace, horseshoe, spike-and-slab, and hierarchical. Six simulation studies highlight shrinkage as crucial for pruning unneeded parameters while showing that prior choice alone cannot fix model misspecification. An application to S&P 500 sector allocations illustrates how prior-based shrinkage manages complexity in high-dimensional or limited-sample scenarios.

Paper 4 examines high-dimensional vector auto-regressive processes, comparing horseshoe, lasso, and hierarchical priors with ridge and nonparametric shrinkage methods in three distinct simulations. In Canadian macroeconomic data, horseshoe priors outperform other approaches by shrinking smaller coefficients while retaining major signals, enhancing forecast accuracy.

Collectively, these four papers form a cohesive suite of Bayesian methods for compositional and high-dimensional time series, addressing interpretability, over-parameterization, and volatility. They offer theoretically grounded, empirically tested frameworks for accurate inference, improved risk management, and deeper strategic insights in hospitality, macroeconomics, and finance.

The dissertation of Harrison Katz is approved.

Frederic Schoenberg

Oscar Madrid-Padilla

Robert E. Weiss, Committee Co-Chair

Ying Nian Wu, Committee Co-Chair

University of California, Los Angeles

2025

This dissertation is lovingly dedicated to those who have been an unwavering source of support throughout this journey. To my friends and family, whose belief in me never faltered. To Allo for the meaningful connection and lessons that have shaped me during pivotal moments in my life, for helping ground me, and being my thought partner. To Blaine and my mother Deborah, whom I wish I could share this with, for the years of pushing me and guiding me. To Tomo for your peculiarities, unconditional love, and the joy you brought into my life during the hardest moments of this journey. Thank you all for your love, patience, and belief in me. I could not have reached this milestone without you.

Contents

List of Figures

List of Tables

I gratefully acknowledge those whose expertise and partnership made this research possible.

- Professor Rob Weiss, my co-advisor, for incisive guidance, patient mentorship, and friendship.
- Professor Yingnian Wu, my co-advisor, for reminding me to approach scholarship—and life—with optimism and curiosity.
- Thomas and my Ph.D. cohort, for camaraderie, challenge, and countless discussions that sharpened this work.
- Sean, whose mastery of Stan and tireless willingness to troubleshoot code elevated every chapter.

Your collective wisdom and support were indispensable to reaching this milestone.

VITA / BIOGRAPHICAL SKETCH

Harrison Katz

Education:

- Bachelor of Arts (B.A.) in *Political Science*, University of California, Los Angeles (UCLA)
- Master of Science (M.S.) in *Mathematics & Statistics*, Georgetown University

Professional Experience:

- **Research Analyst (Risk Analysis), Federal Reserve (2016-2018)**
  Responsible for working alongside economists to conduct rigorous data analytics, mathematical modeling, data visualizations, and research in the contexts of risk analysis in financial markets with the primary focus being on the credit-default-swap market (CDS).
- **Staff Data Scientist, Airbnb (2019 - present)**
  Designed, built, and maintained company wide forecast models for metrics from Nights Booked to Revenue. Developed a new class of Bayesian Compositional Time Series models to forecast lead times during COVID19.

Publications and Presentations:

- Katz, H., K. T. Brusch, and R. E. Weiss (2024). A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times. International Journal of Forecasting 40 (4), 1556–1567
- Katz, H., Savage, E., & Brusch, K. T. (in press). Two-part forecasting for time-shifted metrics. Foresight: The International Journal of Applied Forecasting.

Chapter 1

Introduction

This dissertation develops and applies new Bayesian time series models for compositional and high-dimensional data that exhibit dynamic structure and potential heteroskedasticity. It comprises four self-contained papers—each with its own introduction, methodology, empirical application, and conclusion—offering methodological innovations, simulation evidence, and real-world applications in hospitality and finance.

Paper 1 introduces the *Bayesian Dirichlet Auto-Regressive Moving Average* model (**B-DARMA**). The main outcome is the distribution of fees to be recognized in 12 consecutive future monthly intervals; daily fractional shares, which sum to one, form a compositional time series derived from five years of Airbnb data in which fee accrual and recognition occur at different times. By embedding Auto-Regressive Moving Average (ARMA) components on an additive log-ratio scale within a Dirichlet likelihood, the model enforces compositional constraints and produces reliable forecasts. Simulation studies highlight B-DARMA's consistent predictive performance, while empirical analysis shows more accurate lead-time predictions than standard *vector autoregressive moving average* (**VARMA**) methods. In addition to forecasting, the B-DARMA framework facilitates inference on large compositional vectors and provides efficiency gains through the choice of priors, making it well-suited for settings with complex or limited data. These improvements guide both resource allocation

and strategic planning in hospitality operations.

Paper 2 extends B-DARMA to a *Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity* model (**B-DARCH**). It incorporates a generalized autoregressive conditional heteroskedasticity (GARCH)-like process for the Dirichlet precision parameter to capture time-varying volatility, particularly useful when economic shocks occur such as during the COVID-19 pandemic. Empirical studies with Airbnb currency fee proportions highlight substantial gains in forecast accuracy, as the model flexibly accommodates volatility clustering and structural breaks. Simulation findings confirm that B-DARCH captures evolving uncertainty more effectively than B-DARMA or a transformed normal VARMA model. By maintaining valid predictions on the simplex through the B-DARCH's Dirichlet likelihood, B-DARCH further enhances volatility modeling while preserving interpretability. This expanded capacity to model volatility is especially beneficial for high-frequency financial applications where robust risk management is essential.

Paper 3 conducts a sensitivity analysis of the B-DARMA model using five priors—normal, Laplace, horseshoe, spike-and-slab, and hierarchical—across three simulation scenarios: (i) a *correctly specified* model, (ii) an *overfitted* model including extraneous parameters, and (iii) an *underfitted* model missing key ARMA terms. Under correct specification, all priors perform similarly in terms of bias and predictive accuracy, although the horseshoe and hierarchical priors exhibit slightly better shrinkage properties. In the overfitted scenario, strong shrinkage priors (particularly the horseshoe) effectively reduce spurious complexity by zeroing out unnecessary parameters. However, in the underfitted scenario, none of the priors can compensate for structural omissions in the ARMA dynamics, underscoring that prior choice alone cannot remedy fundamental misspecification. An application to *Standard & Poor's 500* (**S&P 500**) sector allocations further illustrates how prior-based shrinkage manages model complexity and preserves forecasting quality in high-dimensional or limited-sample contexts. These findings emphasize the pivotal interplay between thoughtful prior selection and careful structural modeling in compositional time-series analysis.

Paper 4 investigates high-dimensional *vector autoregressive* processes (VAR) by comparing three Bayesian shrinkage priors—horseshoe, lasso, and hierarchical—to frequentist regularization methods such as ridge and nonparametric shrinkage. Simulation scenarios include low- and high-dimensional over-parameterization, as well as sparse data, capturing diverse challenges faced in real-world forecasting. In an application to Canadian macroeconomic data, local-global priors like the horseshoe adaptively shrink small coefficients while preserving significant signals, leading to more accurate forecasts. Results across scenarios indicate that local-global approaches generally outperform alternatives by offering both efficient parameter estimation and robust uncertainty quantification. These advantages are increasingly relevant in environments where dimension and lag order may be large or uncertain.

Viewed together, these four standalone papers introduce a cohesive suite of Bayesian methods for compositional and high-dimensional time series. By tackling interpretability, over-parameterization, and volatility through theoretical insights and applied examples, the dissertation provides practical frameworks that support accurate inference, risk management, and strategic planning in hospitality, macroeconomics, and finance.

Chapter 2

A Bayesian Dirichlet Auto-Regressive Moving Average Model for Forecasting Lead Times

Provenance. This chapter is a version of: Katz et al., 2024.

Author contributions. Harrison Katz conceived the study, developed the model, implemented the `Stan` code, analysed the data, and drafted the manuscript. Kai Thomas Brusch supervised the research. Robert E. Weiss offered methodological insight and contributed to manuscript revisions.

2.1 Introduction

Travel, entertainment, and hospitality businesses earn fees each day, however these fees cannot be recognized until later. The *lead time* is the amount of time until fees earned on a given day can be recognized. Future dates that fees may be recognized are allocated into regular intervals forming a partition of the future, often weekly, monthly or annual time intervals. Each day, another vector of fee allocations is observed. The distribution of fees into future intervals can be analyzed separately from the total amount of fees; fractional

allocations into future intervals sum to one and form a compositional time series. The basic compositional observation is a continuous vector of probabilities that a dollar earned today can be recognized in each future interval. We wish to understand the process generating the compositional time series and forecast the series on into the future. This information is used by businesses for the allocation of resources, for business planning, and for staffing.

We analyze five years of daily fees billed at Airbnb that will be recognized in the future. Future fees are allocated to one of 12 consecutive monthly intervals; fees beyond this range are small and ignored in this analysis. More granular inference on lead time of revenue recognition would not improve business planning or resource allocation though it substantially complicates modeling, model fitting, and communication of results. To forecast a current day's allocations we have all prior days' data and we have deterministic characteristics of days such as day of the week, season, and sequential day of the year.

A lead time compositional time series $\mathbf{y}_t = (y_{t1}, \ldots, y_{tJ})'$ is a multivariate J-vector time series with observed data y_{tj} where $t = 1, \ldots, T$, indexes consecutive days or other time units, $j = 1, \ldots, J$ indexes the J future revenue recognition intervals, $0 < y_{tj} < 1$ and $\sum_{j=1}^{J} y_{tj} = 1$ for all t. A natural model for compositional data is the Dirichlet distribution which is in the exponential family of distributions and thus Dirichlet time series are special cases of generalized linear time series. Benjamin et al. (2003) proposed a univariate Generalized ARMA data model in a frequentist framework.

There are numerous books on Bayesian Analysis of Time Series Data (Barber et al., 2011; Berliner, 1996; Pole et al., 2018; Koop and Korobilis, 2010; West, 1996; Prado and West, 2010) as well as papers on Bayesian vector auto-regressive (AR) (VAR) moving average (MA) (VMA) (ARMA)(VARMA) time series models (Spencer, 1993; Uhlig, 1997; Bańbura et al., 2010; Karlsson, 2013); and on Bayesian generalized linear time series models (Brandt and Sandler, 2012; Roberts and Penny, 2002; Nariswari and Pudjihastuti, 2019; Chen and Lee, 2016; McCabe and Martin, 2005; Berry and West, 2020; Nariswari and Pudjihastuti, 2019; Fukumoto et al., 2019; Silveira de Andrade et al., 2015; West, 2013).

Dirichlet time series data are less commonly modeled in the literature. Grunwald et al. (1993) proposed a Bayesian compositional state space model with data modeled as Dirichlet given a mean vector, with the current mean given the prior mean also modeled as Dirichlet. Grunwald does not use Markov chain Monte Carlo (MCMC) for model fitting. Similarly, da Silva et al. (2011) proposed a state space Bayesian model for a time series of proportions, extended in da Silva and Rodrigues (2015) to Dirichlets with a static scale parameter. Zheng and Chen (2017) propose a frequentist Dirichlet ARMA time series. Much of the prior work on modeling compositional data and compositional time series transforms $\mathbf{y}_t$ from the original J-dimensional simplex to a Euclidean space where the data is now modeled as normal (Aitchison, 1982; Cargnoni et al., 1997; Ravishanker et al., 2001; Silva and Smith, 2001; Mills, 2010; Barcelo-Vidal et al., 2011; Koehler et al., 2010; Kynčlová et al., 2015; Snyder et al., 2017; AL-Dhurafi et al., 2018).

It would seem preferable to not transform the raw data before modeling and instead model the data $\mathbf{y}_t$ directly as Dirichlet distributed. Thus we propose a new class of Bayesian Dirichlet Auto-Regressive Moving Average models (B-DARMA) for compositional time series. We model the data as Dirichlet given the mean and scale, then transform the J-dimensional mean parameter vector to a $J-1$-dimensional vector. The distributional parameters are then modeled with vector auto-regressive moving average structure. We also model the Dirichlet scale parameter as a log linear function of time-varying predictors.

We give a general framework, and present submodels motivated by our Airbnb lead time data. The B-DARMA model can be applied to data sets with large compositional vectors or few observations, offers efficiency gains through choice of priors and/or submodels, and provides sensible forecasts on the Airbnb data.

We consider normal and horseshoe priors for the VAR and VMA coefficients, and for regression coefficients. Normal priors have long been a default for coefficients while the horseshoe is a newer choice which allows for varying amounts of shrinkage (Carvalho et al., 2009, 2010a; Huber and Feldkircher, 2019a; Kastner and Huber, 2020; Bańbura et al., 2010)

depending on the magnitude of the coefficients.

The next section presents the B-DARMA model. Section 3.3 presents simulation studies comparing the B-DARMA to both a frequentist DARMA data model and a transformed data normal VARMA model. Section 3.4 presents analysis of the Airbnb data. The paper closes with a short discussion.

2.2 A Bayesian Dirichlet Auto-Regressive Moving Average Model

2.2.1 Data model

We observe a J-component multivariate compositional time series $\mathbf{y}_t = (y_{t1}, \ldots, y_{tJ})'$, observed at consecutive integer valued times $t = 1$ up to the most recent time $t = T$, where $0 < y_{tj} < 1$, $\mathbf{1}'\mathbf{y}_t = 1$ where $\mathbf{1}$ is a $J-$vector of ones. We model $\mathbf{y}_t$ as Dirichlet with mean vector $\boldsymbol{\mu}_t = (\mu_{t1}, \ldots, \mu_{tJ})'$, with $0 < \mu_{tj} < 1$, $\mathbf{1}'\boldsymbol{\mu}_t = 1$, and scale parameter $\phi_t > 0$

$$\mathbf{y}_t | \boldsymbol{\mu}_t, \phi_t \sim \text{Dirichlet}(\phi_t \boldsymbol{\mu}_t), \tag{2.1}$$

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

We model $\boldsymbol{\mu}_t$ as a function of prior observations $\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}$, prior means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{t-1}$ and known covariates $\boldsymbol{x}_t$ in a generalized linear model framework. As $\boldsymbol{\mu}_t$ and $\mathbf{y}$ are constrained, we model $\boldsymbol{\mu}_t$ after reducing dimension using the *additive log ratio* (alr) link

$$\boldsymbol{\eta}_t \equiv \text{alr}(\boldsymbol{\mu}_t) = \left(\log\left(\frac{\mu_{t1}}{\mu_{tj^*}}\right), \ldots, \log\left(\frac{\mu_{tJ}}{\mu_{tj^*}}\right) \right) \tag{2.2}$$

where j^* is a chosen reference component $1 \leq j^* \leq J$, and the element of $\boldsymbol{\eta}_t$ that would correspond to the j^*th element $\log(\mu_{j^*}/\mu_{j^*}) = 0$ is omitted. The *linear predictor* $\boldsymbol{\eta}_t$ is a $J-1$-vector taking values in $\mathbb{R}^{J-1}$. Given $\boldsymbol{\eta}_t$, $\boldsymbol{\mu}_t$ is defined by the inverse of equation

(3.3) where $\mu_{tj} = \exp(\eta_{tj})/(1 + \sum_{j=1}^{J-1} \exp(\eta_{tj}))$ for $j = 1, \ldots, J$, $j \neq j^*$ and for $j = j*$, $\mu_{tj*} = 1/(1 + \sum_{j=1}^{J-1} \exp(\eta_{tj}))$.

We model $\boldsymbol{\eta}_t$ as a Vector Auto-Regressive Moving Average (VARMA) process

$$\boldsymbol{\eta}_t = \sum_{p=1}^{P} \boldsymbol{A}_p(\text{alr}(\mathbf{y}_{t-p}) - \mathbf{X}_{t-p}\boldsymbol{\beta}) + \sum_{q=1}^{Q} \boldsymbol{B}_q(\text{alr}(\mathbf{y}_{t-q}) - \boldsymbol{\eta}_{t-q}) + \mathbf{X}_t\boldsymbol{\beta} \tag{2.3}$$

for $t = m+1, \ldots, T$ where $m = \max(P, Q)$, $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ are $(J-1)\times(J-1)$ coefficient matrices of the respective Vector Auto-Regressive (VAR) and Vector Moving Average (VMA) terms, $\boldsymbol{X}_t$ is a known $(J-1) \times r_\beta$ matrix of deterministic covariates including an intercept, and including seasonal variables and trend as needed, and $\boldsymbol{\beta}$ is an $r_\beta \times 1$ vector of regression coefficients. The form of an intercept in $\mathbf{X}_t$ is the $(J-1) \times (J-1)$ identity matrix I_{J-1} as $J-1$ columns in $\mathbf{X}_t$. Given an $r_0 \times 1$ vector of covariates $\boldsymbol{x}_t$ for day t, the simplest form of $\mathbf{X}_t$ is $\mathbf{X}_t = I_{J-1} \otimes \boldsymbol{x}_t$ and $r_\beta = (J-1) * r_0$. The additive log ratio in (3.4) is a multivariate logit link and leads to elements of matrices $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ and vector $\boldsymbol{\beta}$ that are log odds ratios. Given $\boldsymbol{\eta}_t$, $\boldsymbol{\mu}_t$ gives the expected allocations of fees to components.

Scale parameter ϕ_t is modeled with log link as a function of an r_γ-vector of covariates $\boldsymbol{z}_t$,

$$\phi_t = \exp(\boldsymbol{z}_t\boldsymbol{\gamma}), \tag{2.4}$$

where $\boldsymbol{\gamma}$ is an r_γ-vector of coefficients. In the situation of no covariates for ϕ_t, $\log \phi_t = \gamma$ for all t.

Define the consecutive observations $\mathbf{y}_{a:b} = (\mathbf{y}_a, \ldots, \mathbf{y}_b)'$ for positive integers $a < b$. To be well defined, linear predictor (3.4) requires having m previous observations $\mathbf{y}_{(t-m):(t-1)}$, and corresponding linear predictors $\boldsymbol{\eta}_{t-m}, \ldots, \boldsymbol{\eta}_{t-1}$. In computing posteriors, we condition on the first m observations $\mathbf{y}_{1:m}$ which then do not contribute to the likelihood. For the corresponding first m linear predictors, on the right hand side of (3.4), we set $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_m$ equal to $\text{alr}(\mathbf{y}_1), \ldots, \text{alr}(\mathbf{y}_m)$ which effectively omits the VMA terms $\boldsymbol{B}_l(\text{alr}(\mathbf{y}_{t-l}))$ from (3.4) when $t-l \leq m$. In contrast, in (3.4) the VAR terms and $\mathbf{X}_t\boldsymbol{\beta}$ are well defined for $t = 1, \ldots, m$.

Define the C-vector $\boldsymbol{\theta}$ of all unknown parameters $\boldsymbol{\theta} = (\boldsymbol{A}_{prs}, \boldsymbol{B}_{qrs}, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$, where $r, s = 1, \ldots, J-1$ index all elements of matrices $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$, $p = 1, \ldots, P$, $q = 1, \ldots, Q$ and $C = (P+Q) * (J-1)^2 + r_\beta + r_\gamma$. Prior beliefs $p(\boldsymbol{\theta})$ about $\boldsymbol{\theta}$ are updated by Bayes' theorem to give the posterior

$$p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m})}{p(\mathbf{y}_{(m+1):T}|\mathbf{y}_{1:m})},$$

where $p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m}) = \prod_{t=m+1}^{T} p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{y}_{(t-m):(t-1)})$, $p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{y}_{(t-m):(t-1)})$ is the density of the Dirichlet in (4.1), and the normalizing constant

$$p(\mathbf{y}_{(m+1):T}|\mathbf{y}_{1:m}) = \int p(\boldsymbol{\theta})p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m})d\boldsymbol{\theta}.$$

We wish to forecast the next S observations $\mathbf{y}_{(T+1):(T+S)}$. These have joint predictive distribution

$$p(\mathbf{y}_{(T+1):(T+S)}|\mathbf{y}_{1:T}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}_{(T+1):(T+S)}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_{1:T}).$$

The joint predictive distribution $p(\mathbf{y}_{(T+1):(T+S)}|\mathbf{y}_{1:T})$ can be summarized for example by the mean or median against time $t \in (T+1) : (T+S)$ to communicate results to business managers.

Our data model can be viewed as a Bayesian multivariate Dirichlet extension of the Generalized ARMA model by Benjamin et al. (2003). Similarly, Zheng and Chen (2017) propose a DARMA data model whose link function in (3.3) does not have an analytical inverse, so needs to be approximated numerically. They view our data model as an approximation to theirs, noting that the resulting noise sequence from having the analytical inverse isn't a martingale difference sequence (MDS). This lack of an MDS complicates the investigation of the probabilistic properties of the series and the asymptotic behavior of their estimators.

From a Bayesian perspective, the need for an MDS for making inference is circumvented.

Bayesian inference is based on the posterior distribution of the parameters, which combines the likelihood function (dependent on the data) and the prior distribution (representing our prior beliefs about the parameters). This allows us to make inferences regardless of whether the data form an MDS.

Furthermore, the concept of a "noise sequence" or residuals is somewhat different in the context of generalized linear models (GLMs) compared to location-scale models like linear regression. In GLMs, residuals are not explicitly defined, and the noise sequence discussed in the context of Zheng and Chen's model is an approximation. In our Bayesian approach, we do not need to define or consider the noise sequence at all, which simplifies the model and the analysis.

2.2.2 Choice of link function

The link function in (3.3) and (3.4) can be replaced with other common simplex transformations such as the Centered Log-Ratio (CLR)

$$\text{clr}(\mathbf{y}) = \left[\ln\left(\frac{y_1}{g(\mathbf{y})}\right), \ln\left(\frac{y_2}{g(\mathbf{y})}\right), \ldots, \ln\left(\frac{y_J}{g(\mathbf{y})}\right)\right]$$

where $g(\mathbf{y}) = (y_1 \times y_2 \times \ldots \times y_J)^{1/J}$. Alternatively, the Isometric Log-Ratio (ILR) transformation can be used, with its j-th component defined as:

$$y_j = \sqrt{\frac{r_j}{r_j + 1}} \ln\left(\frac{g_j(\mathbf{y})}{\left(\prod_{i \in H_j} y_i\right)^{1/r_j}}\right).$$

Here, $g_j(\mathbf{y})$ is the geometric mean of a subset S_j of $\mathbf{y}$, H_j is the complement of S_j, and r_j is the number of elements in H_j. The choice of subsets S_j and H_j can vary depending on the specific problem and interpretation requirements. Egozcue et al. (2003) show that the coefficient matrices of the ALR, CLR, and ILR are linear transformations of each other. This means that the DARMA data models with these three link functions are equivalent,

provided the same transformation is applied to the priors.

2.2.3 Model selection

For model selection, we use the approximate calculation of the leave-future-out (LFO) estimate of the expected log pointwise predictive density (ELPD) for each model which measures predictive performance (Bernardo and Smith, 2009; Vehtari and Ojanen, 2012),

$$\text{ELPD}_{\text{LFO}} = \sum_{t=L}^{T-M} \log p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t})$$

where

$$p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{1:t}) d\boldsymbol{\theta}, \tag{2.5}$$

where M is the number of step ahead predictions and L is a chosen minimum number of observations from the time series needed to make predictions. We use Monte-Carlo methods to approximate (2.5) with S^* random draws from the posterior and estimate $p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t})$ as

$$p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t}) \approx \frac{1}{S^*} \sum_{s=1}^{S^*} p(\mathbf{y}_{(t+1):(t+M)}|\mathbf{y}_{1:t}, \boldsymbol{\theta}_{1:t}^{(s)})$$

for $s = 1, \ldots, S^*$ draws from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}_{1:t})$. Calculating the ELPD LFO requires refitting the model for each $t \in (L, \ldots, T-M)$, to get around this, we use approximate M-step ahead predictions using Pareto smoothed importance sampling (Bürkner et al., 2020).

2.2.4 Priors

A useful vague prior for the individual coefficients in $\boldsymbol{\theta}$ is a proper independent normal $N(V_0, V)$ with varying V_0's and V's depending on prior beliefs about the coefficients. We

take $V_0 = .4, .1$ in our data analysis when we believe a priori that the coefficients will be positive and $V_0 = 0$ when we are unsure. For elements a_{prs} and b_{qrs} of $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ matrices, we take $V = .5^2$ as we expect those elements to be between $[-1, 1]$. For elements of $\boldsymbol{\beta}$, we might for example let V vary with the standard deviation of the covariate.

It may be thought that many elements of $\boldsymbol{\theta}$ might be at or near zero. In this case, rather than a normal prior, we might consider a shrinkage prior like that proposed by Carvalho et al. (2010a), who propose a horseshoe prior of the form

$$\begin{aligned} \theta_c|\tau, \lambda_{c^*} &\sim N(0, \tau^2\lambda_{c^*}^2) \\ \lambda_{c^*} &\sim C^+(0, 1). \end{aligned}$$

for some $c^* = 1, \ldots, C^*$, $c = 1, \ldots, C$, and $C^* \leq C$. Each θ_c belongs to one of $c^* = 1, \ldots, C^*$ groups and each group gets its own shrinkage parameter λ_{c^*} applied to it. We apply this with component specific shrinkage parameters in our data analysis.

2.3 Simulation study

In two simulation studies, we study the B-DARMA model's capacity to accurately retrieve true parameter values, a critical aspect for reliable forecasting. By comparing B-DARMA with established models on a number of metrics, we provide a robust assessment of both its estimation and forecasting potential. Each simulation study has $L = 400$ data sets indexed by $l = 1, \ldots, L$ with $T = 540$ observations $\mathbf{y}_t^{(l)}$, $t = 1, \ldots, T$ with $J = 3$ components, $j^* = 3$ as the reference component for all models, and $\mathbf{Y}^l = (\mathbf{y}_1^{(l)}, \ldots, \mathbf{y}_T^{(l)})$. We split the 540 observations into a training set of the first 500 observations and a test set of the last 40 observations, the holdout data, which the model is not trained on and which is used to show the models' forecasting performance.

In both studies, we compare B-DARMA to a non-Bayesian DARMA data model and to a non-Bayesian transformed-data normal VARMA (tVARMA) model that transforms $\mathbf{y}_t^{(l)}$ to

$\text{alr}(\mathbf{y}_t^{(l)}) \in \mathbb{R}^{J-1}$ and models $\text{alr}(\mathbf{y}_t) \sim N_{J-1}(\boldsymbol{\eta}_t, \boldsymbol{\Sigma})$, where Σ is an unknown $J-1 \times J-1$ positive semi-definite matrix, and $\boldsymbol{\eta}_t$ is defined in (3.4). We use the VARMA function from the MTS package (Tsay et al., 2022) in R (R Core Team, 2022) to fit the tVARMA models with maximum likelihood. For the non-Bayesian DARMA data model, the BFGS algorithm as implemented in optim in R is used for optimization. To compute the parameters' standard errors, the negative inverse Hessian at the mode is calculated.

The data generating model (DGM) is a DARMA model in simulation 1 and a tVARMA model in simulation 2. We set $p = 1$, $q = 1$ for a DARMA$(1, 1)$ (simulation 1) or tVARMA$(1, 1)$ (simulation 2) data generating model.

To keep the parameterization of the B-DARMA(P, Q) model consistent with the parameterization of the MTS package, we remove $-\mathbf{X}_{t-p}\boldsymbol{\beta}$, $p = 1, \ldots, P$ from the VAR term in (3.4) and set $\boldsymbol{\eta}_t = \sum_{p=1}^{P} \boldsymbol{A}_p(\text{alr}(\mathbf{y}_{t-p})) + \sum_{q=1}^{Q} \boldsymbol{B}_q(\text{alr}(\mathbf{y}_{t-q}) - \boldsymbol{\eta}_{t-q}) + \mathbf{X}_t\boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^* = \boldsymbol{\beta} - \sum_{p=1}^{P} \boldsymbol{A}_p\boldsymbol{\beta}$ and we drop the asterisk on $\boldsymbol{\beta}$ for the remainder of this section.

All B-DARMA models are fit with STAN (Stan Development Team, 2022) in R. We run 4 chains with 2000 iterations each with a warm up of 1000 iterations for a posterior sample of 4000. Initial values are selected randomly from the interval $[-1, 1]$.

In all simulations, the covariate matrix $\mathbf{X}_t = \boldsymbol{I}_2$, the 2×2 identity matrix. Unknown coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have dimension 2×2. The shared parameters of interest for the DARMA, B-DARMA, and tVARMA are $\boldsymbol{\beta} = (\beta_1, \beta_2)'$, $\boldsymbol{B} = (b_{rs})$, and $\boldsymbol{A}_p = (a_{prs}), r, s = 1, 2$, $p = 1$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. We use the out of sample mean Forecast Residual Sum of Squares, mFRSS_j, and Forecast Mean Absolute Error, FMAE_j for each component j as a measure of forecasting performance

$$\text{mFRSS}_j = \left(\frac{1}{400}\right)\left(\frac{1}{40}\right)\sum_{l=1}^{400}\sum_{t=501}^{540}(y_{tj}^l - \bar{\mu}_{tj}^l)^2$$

$$\text{FMAE}_j = \left(\frac{1}{400}\right)\left(\frac{1}{40}\right)\sum_{l=1}^{400}\sum_{t=501}^{540}|y_{tj}^l - \bar{\mu}_{tj}^l|$$

where $\bar{\mu}_{tj}^l$ is the posterior mean of μ_{tj} or the maximum likelihood estimate in the lth data set.

The FMAE gives an indication of the average error size, while the FRSS provides information about the variability of the forecast errors. The tVARMA model has additional unknown covariance matrix $\mathbf{\Sigma} = (\Sigma_{rs}), r, s = 1, 2$ a function of 3 unknown parameters, two standard deviations $\sigma_1 = \Sigma_{11}^{1/2}$, $\sigma_2 = \Sigma_{22}^{1/2}$, and correlation $\rho = \Sigma_{12}/(\sigma_1 * \sigma_2)$, while the B-DARMA and DARMA have a single unknown scale parameter ϕ. For the tVARMA data generating simulations we set $\sigma_1 = \sigma_2 = .05$, $\rho = .30$, and for DARMA generating models, we set $\phi = 1000$.

Priors in simulations 1 and 2 for the B-DARMA$(1, 1)$ are independent $N(0, .5^2)$ for all coefficients in β, A, B, and a priori $\gamma = \log \phi \sim$ Gamma$(25/7, 5/7)$ with mean 5 and variance 7.

For each parameter, generically θ, with true value θ_{true} in simulation l, for Bayesian models we take the posterior mean $\bar{\theta}^l$ as the point estimate. For the 95% Credible Interval (CI) $(\theta^l_{\text{low}}, \theta^l_{\text{upp}})$ we take the endpoints to be 2.5% and 97.5% quantiles of the posterior. For the tVARMA and DARMA models, we use the maximum likelihood estimates (MLE) as the parameter estimate, and 95% confidence intervals (CI) calculated as maximum likelihood estimate plus or minus 1.96 standard errors.

For each parameter θ in turn, for the B-DARMA, tVARMA, and DARMA models, we assess bias, root mean squared error (RMSE), length of the 95% CI (CIL), and the fraction of simulations % IN where θ_{true} falls within the 95% CI

$$\begin{aligned}
\text{bias} &= \frac{1}{400}\sum_{l=1}^{400}(\bar{\theta}^l - \theta_{\text{true}}) \\
\text{RMSE} &= \left(\frac{1}{400}\sum_{l=1}^{400}(\bar{\theta}^l - \theta_{\text{true}})^2\right)^{1/2} \\
\text{CIL} &= \frac{1}{400}\sum_{l=1}^{400}\left(\theta^l_{\text{upp}} - \theta^l_{\text{low}}\right) \\
\%\ \text{IN} &= \frac{1}{400}\sum_{l=1}^{400}\mathbf{1}\{\theta^l_{\text{low}} \le \theta_{\text{true}} \le \theta^l_{\text{upp}}\}
\end{aligned}$$

for the B-DARMA models and replace $\bar{\theta}^l$ with the MLE for the tVARMA and DARMA models.

For all simulations, elements of $\boldsymbol{\beta}, \boldsymbol{A}, \boldsymbol{B}$ are set to be

$$\beta_1 = -0.07, \beta_2 = 0.10,$$
$$a_{11} = 0.95, a_{12} = -0.18, a_{21} = 0.3, a_{22} = 0.95,$$
$$b_{11} = 0.65, b_{12} = 0.15, b_{21} = 0.2, b_{22} = 0.65.$$

2.3.1 Comparing B-DARMA, DARMA, and tVARMA Models in simulations 1-2

Supplementary tables 2.1 and 2.2 provide summarized parameter recovery results, while table 2.5 details the mFRSS_j and FMAE_j for each component. When the data generating model is a DARMA(1,1), B-DARMA consistently outperforms tVARMA, with the B-DARMA(1,1) yielding an RMSE averaging at 40% smaller, especially for the $\boldsymbol{B}$ matrix coefficients. The B-DARMA's performance aligns with that of the frequentist DARMA, each excelling in half of the ten considered coefficients. Both DARMA models showcase superior coverage to the tVARMA.

For a tVARMA(1,1) data generating model, the tVARMA generally exhibits the smallest RMSE for all coefficients, barring b_{11}. The difference in RMSE is marginal, averaging at 2%. The B-DARMA and frequentist DARMA compare similarly for $\boldsymbol{\beta}$ and $\boldsymbol{A}$, but the latter performs worse for $\boldsymbol{B}$. B-DARMA has better coverage than both tVARMA and frequentist DARMA for most coefficients, with exceptions for b_{21} and b_{22}.

The out-of-sample prediction results are summarized in table 2.5. When the DGM is a DARMA(1,1), the B-DARMA outperforms both the frequentist DARMA model and the tVARMA in mFRSS and FMAE across all components, most notably for y_3. When the DGM is a tVARMA, the B-DARMA performs comparably to the tVARMA and outperforms the

DARMA for y_1 and y_2, while significantly outperforming both the tVARMA and DARMA for y_3.

2.4 Airbnb Lead Time Data Analysis

The Airbnb lead time data, $\mathbf{y}_t$, is a compositional time series for a specific single large market where each component is the proportion of fees booked on day t that are to be recognized in 11 consecutive 30 day windows and 1 last consecutive 35 day window to cover 365 days of lead times.

Figures 2.1 and 2.2 plot the Airbnb lead time data from 01/01/2015 to 01/31/2019 and 03/26/2015 to 05/14/2015 respectively. There is a distinct repeated yearly shape in figure 2.1 and a clear weekly seasonality in figure 2.2. Figure 2.1 shows a gradual increase/decrease in component sizes, most notably a decrease for the first window of $[0, 29)$ days. The levels are driven by the attractiveness of certain travel periods with guests booking earlier and paying more for peak periods like summer and the December holiday season. The weekly variation shows a pronounced contrast between weekdays and weekends. Thus we want to include weekly and yearly seasonal variables and a trend variable in the predictors $\mathbf{X}_t$ in (3.4). We model each component with its own linear trend and use Fourier terms for our seasonal variables, pairs of $\left(\sin \frac{2k\pi t}{w_{\text{season}}}, \cos \frac{2k\pi t}{w_{\text{season}}}\right)$ for $k = 1, \ldots, K_{\text{season}} \leq \frac{w_{\text{season}}}{2}$ where we take $w_{\text{season}} = w_{\text{week}} = 7$ for weekly seasonality and $w_{\text{season}} = w_{\text{year}} = 365.25$ for yearly seasonality. The orthogonality of the Fourier terms helps with convergence which is why we prefer it to other seasonal representations.

We train the model on data from 01/01/2015 to 1/31/2019, choose a forecast window of 365, and use 02/01/2019 to 01/31/2020 as the test set. All B-DARMA models are fit with STAN using the R interface where we run 4 chains with 3000 iterations each with a warm up of 1500 iterations for a total of 6000 posterior samples. Initial values are selected randomly from the interval $[-1, 1]$.

2.4.1 Model

We use the approximate LFO ELPD of candidate models to first decide on the number of yearly Fourier terms, K_{year} given initial choices of P and Q, then given the number of Fourier terms we decide on the orders P and Q. We fix $K_{\text{season}} = K_{\text{week}} = 3$, the pairs of Fourier terms for modeling weekly seasonality, for all models. We use an intercept and the same seasonal variables and linear trend for ϕ_t. To define LFO ELPD, we take the step ahead predictions to be $M = 1$, the minimum number of observations from the time series to make predictions to be $L = 365$, and the threshold for the Pareto estimates to be .7 (Bürkner et al., 2020; Vehtari et al., 2015) which resulted in needing to refit models at most 12 times.

For all of the model selection process, we set independent $N(0, .5^2)$ priors for each $(a_{prs}), r, s = 1, \ldots, 11, p = 1, 2$ and $(b_{qrs}), r, s = 1, \ldots, 11, q = 1$, independent $N(0, 1)$ priors on the Fourier coefficients, $N(0, .1)$ priors on the linear trend coefficients, and a $N(0, 4)$ prior on the intercepts. We use these same priors for the seasonal, intercept and trend coefficients in γ in (3.5).

We fixed $P = 1$, $Q = 0$ and let $K_{\text{season}} = K_{\text{year}}$ take on increasing values starting with 3 then sequentially $K_{\text{year}} = 6, 8, 9, 10$ with increasing values of LFO ELPD until 10 had worse LFO ELPD, so we stopped and took $K_{\text{year}} = 9$ (table 2.6). We fixed $K_{\text{year}} = 9$ and similarly took $(P, Q) = (1, 0)$ first then $(P, Q) = (1, 1)$ and $(P, Q) = (2, 0)$ with $(1, 0)$ performing best (table 2.6).

We compare four different B-DAR(1) models plus a DAR(1) model and a tVAR(1) model. Model 1 (Horseshoe Full) has a horseshoe ($\tau = 1$) prior on the coefficients in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with a separate shrinkage parameter λ_{c^*} for the elements of $\boldsymbol{\theta}$ that correspond to each η_j or to ϕ.

As we expect the AR a_{rs} elements to diminish in magnitude as the time difference $|r - s|$ increases, the other three models have varying prior mean and sd for a_{rs} as functions of $|r-s|$. Model 2 (Normal Full) has an $a_{rr} \sim N(.4, .5^2)$ prior on its diagonal elements $r = 1, \ldots, 11$, a $a_{r(r+1)}, a_{r(r-1)} \sim N(.1, .5^2)$ prior on its two nearest neighbors, $r = 1, \ldots, 10$ and $r = 2, \ldots, 11$ respectively and an $N(0, .5^2)$ prior on the remaining elements of the $\boldsymbol{A}$ matrix.

In contrast to models 1 and 2, where all parameters a_{rs} in $\boldsymbol{A}$ are allowed to vary, models 3 and 4 fix some of the parameters a_{rs} to 0. Model 3 (Normal Nearest Neighbor) only allows $a_{r(r+1)}, r = 1, \ldots, 10, a_{r(r-1)}, r = 2, \ldots, 11$ and $a_{rr}, r = 1, 2, \ldots, 11$ to vary and model 4 (Normal Diagonal) only allows $a_{rr}, r = 1, 2, \ldots, 11$ to vary. The remaining elements in model 3 $a_{r(r+k)}$ for $k \geq 2$ or $k \leq -2$ in the $\boldsymbol{A}$ matrix are set equal to 0. For model 4 all off diagonal elements, a_{rs} $r \neq s$, are set to 0. The non-zero coefficients have the same priors as the Normal Full.

All 3 Normal models have the same priors on coefficients $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$: $N(0, 1)$ and $N(0, .1^2)$ for the coefficients of the Fourier and trend terms respectively. The scale model predictors include the same seasonal and trend terms and the corresponding priors on the coefficients are the same as for the $\boldsymbol{\eta}$'s. The prior on the intercepts in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is $N(0, 2^2)$.

Model 5 (DAR(1)) is the frequentist counterpart to our B-DAR(1) Normal Full model with the same seasonal and trend terms for η_j and ϕ. Model 6 (tVAR(1)) is a frequentist transformed VAR model which models $\text{alr}(\mathbf{y}_t) \sim N_{J-1}(\boldsymbol{\eta}_t, \boldsymbol{\Sigma})$. It has the same seasonal and trend variables in $\boldsymbol{\eta}_t$. Both models 5 and 6 are fit in R, model 5 with the BFGS algorithm as implemented in optim and model 6 with the VARX function in the MTS package.

2.4.2 Results

In the training set, the Horseshoe Full model has the largest LFO ELPD (supplementary table 2.8). The LFO ELPD for the Normal Full model and the Normal Nearest Neighbor model are close to the Horseshoe Full model with the Normal Diagonal model having a much worse LFO ELPD than the other models.

For the test set, the Forecast Residual Sum of Squares (FRSS_j) and the Forecast Mean Absolute Error (FMAE_j) for each component is in table 2.7 as well as the total and average Forecast Residual Sum of Squares and average Forecast Mean Absolute Error for each model. The Normal Full model performs best on the test set for most components and has the lowest total Forecast Residual Sum of Squares and smallest average Forecast Mean Absolute Error,

with the Horseshoe Full model about 1.5% and .1% worse respectively. The differences between the two full models (Model 1 and 2) and Models 3 and 4 are largest for the larger components as the FRSS's for the smaller components are closer. For example, for the largest component y_1, the Normal Diagonal model performs over 10% worse than the Normal Full model. The Normal Nearest Neighbor model performs well considering it has 91 fewer parameters than the Normal Full model, performing 5% worse in total than the Normal Full model.

When compared to the frequentist counterparts, the Normal Full and Horseshoe Full models exhibit superior performance, particularly in the larger components y_1 through y_3, having a smaller average FMAE and a smaller RSS_j. However, it's worth noting that the differences in FRSS_j for the smaller components y_4 through y_{12} among the models are subtler, although the MAE values continue to show a more pronounced difference across all components. The DAR(1) does perform better than the subsetted B-DAR(1) models with a total FRSS roughly 5% smaller than the Normal Nearest-Neighbor model and 10% less than the Normal Diagonal model. The tVAR(1) performs significantly worse than all the DAR models, with a total FRSS about 50% larger than the worst performing DAR data model, suggesting the time invariant nature of $\mathbf{\Sigma}$ is inappropriate for the data.

Figures 2.3 and 2.4 plot out of sample forecasts ($\bar{\mu}_{tj}$) and residuals of ($\text{alr}(\mathbf{y}_t) - \hat{\boldsymbol{\eta}}_t$) for the Normal Full model. The model captures most of the yearly seasonality and the residual terms exhibit no consistent positive or negative bias for any of the components as they are all centered near 0. Much of the remaining residual structure may be explained by market specific holidays which are not incorporated in the current model.

The estimated yearly and weekly seasonality for each of the $\boldsymbol{\eta}_j$'s (ALR scale) and ϕ (log scale) in the Normal Full model are detailed in supplementary figures 2.5 and 2.6. Yearly variation is pronounced in larger components, with $\boldsymbol{\eta}_1$ peaking in late December and troughing before the new year, while ϕ exhibits a simpler pattern with two high and one low period. Weekly seasonality displays a consistent weekday versus weekend behavior across all $\boldsymbol{\eta}_j$ com-

ponents, showing higher values on weekends. This pattern extends to booking lead times, with weekday bookings typically falling within 0 to 30 days and weekend bookings having longer lead times; this variability is also mirrored in the scale parameter ϕ. Supplementary table 2.9 summarizes the posterior statistics for the intercepts and linear growth rates of $\boldsymbol{\eta}_j$s and $\log(\phi)$, showing a monotonic decrease in the intercepts and larger growth rates for smaller components. The growth rate of ϕ indicates less compositional variability over time. Lastly, the posterior densities for the elements a_{rs} of the $\boldsymbol{A}$ matrix, shown in supplementary figure 2.7, reveal that larger components have pronounced coefficients on their own lag, a_{rr}, with diminishing magnitude for distant trip dates, while smaller components exhibit more uncertainty but generally maintain a strong positive coefficient on their own lag.

2.5 Discussion

There are distinct differences in the computational costs in fitting our DARMA model using frequentist or Bayesian methodologies, as well as the tVARMA model using the MTS package in R. Firstly, fitting the tVARMA model using the MTS package was fastest, attributable to its implementation of conditional maximum likelihood estimation with a multivariate Gaussian likelihood. This approach benefits from the mathematical simplicity of estimation under the normal model and its unconstrained optimization problem, resulting in significantly less computational overhead. This is in addition to the built-in performance optimizations found in specialized packages like MTS.

In contrast, the Bayesian DARMA model, implemented using Hamiltonian Monte Carlo (HMC), had higher computational costs. This is to be expected, given that HMC requires running many iterations to generate a representative sequence of samples from the posterior distribution. However, the B-DARMA fit faster than the frequentist DARMA. The frequentist DARMA model, fitted using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization method, was the most computationally demanding in our study. The complex-

ity arises from the calculation of the Hessian matrix estimates, the computation of gamma functions in the Dirichlet likelihood, and the constraint of positive parameters. Notably, around 25% of model fitting attempts in the simulation studies initially failed and required regenerating data and refitting the model, underscoring the computational challenges of this approach.

This paper presented a new class of Bayesian compositional time series models assuming a Dirichlet conditional distribution of the observations, with time varying Dirichlet parameters which are modeled with a VARMA structure. The B-DARMA outperforms the frequentist tVARMA and DARMA when the underlying data generating model is a DARMA and does comparably well to the tVARMA and outperforms the DARMA when tVARMA is the data generating model in the simulation studies. By choice of prior and model subsets, we can reduce the number of coefficients needing to be estimated and better handle data sets with fewer observations. This class of models effectively models the fee lead time behavior at Airbnb, outperforming the DAR and tVAR with the same covariates, and provides an interpretable, flexible framework.

Further development is possible. Common compositional time series have components with no ordering while the 12 lead time components in the Airbnb data set are time ordered. This suggests $\boldsymbol{\theta}$ can be modeled hierarchically. The construction of the Nearest Neighbor and Diagonal models for $\boldsymbol{A}_1$ are examples of using time ordering in model specification. While the distributional parameters $\boldsymbol{\mu}_t$ and ϕ_t are time-varying, the elements of $\boldsymbol{\theta}$ are static and thus the seasonality is constant over time, which may not be appropriate for other data sets. We model ϕ_t with exogenous covariates, but our approach is flexible enough to allow for it to have a more complex AR structure itself. While not an issue for the Airbnb data set, the model's inability to handle 0's would need to be addressed for data sets with exact zeroes.

Figures and tables

par	model	true	RMSE	bias	coverage	length
β_1	B-DARMA	-.07	0.0083	-0.0019	0.9450	0.0312
	DARMA		0.0088	-0.0015	0.9225	0.0312
	tVARMA		0.0112	-0.0038	0.9000	0.0354
β_2	B-DARMA	.10	0.0081	-0.0012	0.9425	0.0313
	DARMA		0.0083	-0.0007	0.9425	0.0312
	tVARMA		0.0096	-0.0035	0.9425	0.0361
a_{11}	B-DARMA	.95	0.0113	-0.0021	0.9250	0.0385
	DARMA		0.0108	-0.0025	0.9600	0.0395
	tVARMA		0.0145	-0.0036	0.8950	0.0428
a_{12}	B-DARMA	-.18	0.0074	-0.0014	0.9400	0.0287
	DARMA		0.0079	-0.0009	0.9400	0.0294
	tVARMA		0.0133	-0.0013	0.8500	0.0316
a_{21}	B-DARMA	.30	0.0101	-0.0020	0.9500	0.0373
	DARMA		0.0101	-0.0009	0.9500	0.0382
	tVARMA		0.0127	-0.0041	0.9225	0.0436
a_{22}	B-DARMA	.95	0.0086	-0.0021	0.9300	0.0311
	DARMA		0.0084	-0.0018	0.9300	0.0316
	tVARMA		0.0108	-0.0028	0.9050	0.0323
b_{11}	B-DARMA	.65	0.0309	-0.0064	0.9450	0.1244
	DARMA		0.0308	-0.0011	0.9500	0.1286
	tVARMA		0.1012	-0.0759	0.5025	0.1593
b_{21}	B-DARMA	.20	0.0314	0.0003	0.9475	0.1168
	DARMA		0.0297	0.0018	0.9425	0.1162
	tVARMA		0.0907	-0.0485	0.5650	0.1768
b_{12}	B-DARMA	.15	0.0286	0.0010	0.9525	0.1145
	DARMA		0.0289	0.0031	0.9525	0.1139
	tVARMA		0.0900	-0.0514	0.5425	0.1645
b_{22}	B-DARMA	.65	0.0321	-0.0040	0.9500	0.1283
	DARMA		0.0341	0.0013	0.9350	0.1272
	tVARMA		0.1135	-0.0868	0.4550	0.1575

Table 2.1: Simulation study 1 results (RMSE, bias, coverage and length of the 95% credible or confidence interval) for B-DARMA(1,1) tVARMA(1,1) and DARMA(1,1) when the data generating model is a DARMA(1,1) for the two regression coefficients in $\boldsymbol{\beta}$, the four a_{rs} elements of the A matrix, and four b_{rs} elements of the B matrix.

par	model	true	RMSE	bias	coverage	length
β_1	B-DARMA	-.07	0.0063	-0.0012	0.9650	0.0254
	DARMA		0.0065	-0.0008	0.9275	0.0336
	tVARMA		0.0061	-0.0013	0.9450	0.0238
β_2	B-DARMA	.10	0.0069	-0.0010	0.9625	0.0254
	DARMA		0.0065	-0.0007	0.9400	0.0334
	tVARMA		0.0062	-0.0010	0.9550	0.0241
a_{11}	B-DARMA	.95	0.0125	-0.0035	0.9525	0.0461
	DARMA		0.0125	-0.0038	0.9300	0.0604
	tVARMA		0.0115	-0.0022	0.9525	0.0431
a_{12}	B-DARMA	-.18	0.0086	-0.0004	0.9650	0.0344
	DARMA		0.0091	-0.0007	0.9225	0.0595
	tVARMA		0.0084	-0.0007	0.9600	0.0321
a_{21}	B-DARMA	.30	0.0115	-0.0010	0.9600	0.0454
	DARMA		0.0112	-0.0009	0.9375	0.0453
	tVARMA		0.0111	-0.0007	0.9500	0.0437
a_{22}	B-DARMA	.95	0.0090	-0.0025	0.9475	0.0347
	DARMA		0.0096	-0.0040	0.9275	0.0456
	tVARMA		0.0086	-0.0016	0.9450	0.0325
b_{11}	B-DARMA	.65	0.0342	-0.0050	0.9475	0.1320
	DARMA		0.0365	-0.0015	0.9300	0.1728
	tVARMA		0.0348	0.0002	0.9350	0.1287
b_{21}	B-DARMA	.20	0.0342	-0.0027	0.9275	0.1318
	DARMA		0.0357	0.0032	0.9150	0.1718
	tVARMA		0.0337	-0.0007	0.9475	0.1315
b_{12}	B-DARMA	.15	0.0325	-0.0011	0.9600	0.1320
	DARMA		0.0346	-0.0005	0.9125	0.1707
	tVARMA		0.0319	0.0012	0.9475	0.1278
b_{22}	B-DARMA	.65	0.0354	-0.0077	0.9375	0.1317
	DARMA		0.0366	0.0003	0.8850	0.1717
	tVARMA		0.0344	-0.0008	0.9475	0.1318

Table 2.2: Simulation study 2 results (RMSE, bias, coverage and length of the 95% credible or confidence interval) for B-DARMA(1,1), tVARMA(1,1) and DARMA(1,1) when the data generating model is a tVARMA(1,1) for the two regression coefficients in $\boldsymbol{\beta}$, the four a_{rs} elements of the A matrix, and four b_{rs} elements of the B matrix.

Parameter	Model	Net Bias	RMSE Ratio	Net Coverage	Length Ratio
β_1	DARMA	-0.0004	1.06	-0.0225	1.00
	tVARMA	-0.0019	1.35	-0.0450	1.13
β_2	DARMA	0.0005	1.02	0.0000	0.99
	tVARMA	-0.0023	1.18	0.0000	1.15
a_{11}	DARMA	-0.0004	0.96	0.0350	1.03
	tVARMA	-0.0015	1.28	-0.0300	1.11
a_{12}	DARMA	0.0005	1.07	0.0000	1.02
	tVARMA	0.0001	1.80	-0.0900	1.10
a_{21}	DARMA	0.0011	1.00	0.0000	1.02
	tVARMA	-0.0021	1.26	-0.0275	1.17
a_{22}	DARMA	0.0003	0.98	0.0000	1.02
	tVARMA	-0.0007	1.26	-0.0250	1.04
b_{11}	DARMA	0.0053	1.00	0.0050	1.03
	tVARMA	-0.0695	3.27	-0.4425	1.28
b_{21}	DARMA	0.0006	0.95	-0.0050	0.99
	tVARMA	-0.0482	2.89	-0.3825	1.51
b_{12}	DARMA	0.0001	1.01	0.0000	0.99
	tVARMA	-0.0504	3.15	-0.4100	1.44
b_{22}	DARMA	0.0073	1.06	-0.0150	0.99
	tVARMA	-0.0828	3.54	-0.4950	1.23

Table 2.3: Simulation study 1 results (Net Bias, RMSE Ratio, Net Coverage, and Credible/Confidence Interval Length Ratio) comparing B-DARMA(1,1) with tVARMA(1,1) and DARMA(1,1) when the data generating model is a DARMA(1,1) for the two regression coefficients in $\boldsymbol{\beta}$, the four a_{rs} elements of the A matrix, and four b_{rs} elements of the B matrix. RMSE and Length ratios greater than 1 and negative net coverage differences favor the B-DARMA model.

Parameter	Model	Net Bias	RMSE Ratio	Net Coverage	Length Ratio
β_1	DARMA	-0.0004	1.03	-0.0375	1.32
	tVARMA	-0.0001	0.97	-0.0200	0.94
β_2	DARMA	0.0003	0.94	-0.0225	1.31
	tVARMA	0.0000	0.90	-0.0075	0.95
a_{11}	DARMA	-0.0003	1.00	-0.0225	1.31
	tVARMA	0.0013	0.92	0.0000	0.93
a_{12}	DARMA	-0.0003	1.06	-0.0425	1.73
	tVARMA	-0.0003	0.98	-0.0050	0.93
a_{21}	DARMA	0.0001	0.97	-0.0225	1.00
	tVARMA	0.0003	0.97	-0.0100	0.96
a_{22}	DARMA	-0.0015	1.07	-0.0200	1.31
	tVARMA	0.0009	0.96	-0.0025	0.94
b_{11}	DARMA	0.0035	1.07	-0.0175	1.31
	tVARMA	0.0052	1.02	-0.0125	0.98
b_{21}	DARMA	0.0059	1.04	-0.0125	1.30
	tVARMA	0.0020	0.98	0.0200	1.00
b_{12}	DARMA	0.0006	1.06	-0.0475	1.29
	tVARMA	0.0023	0.98	-0.0125	0.97
b_{22}	DARMA	0.0080	1.03	-0.0525	1.30
	tVARMA	0.0069	0.97	0.0100	1.00

Table 2.4: Simulation study 2 results (Net Bias, RMSE Ratio, Net Coverage, and Length Ratio) comparing B-DARMA(1,1) with tVARMA(1,1) and DARMA(1,1) when the data generating model is a tVARMA(1,1) for the two regression coefficients in $\boldsymbol{\beta}$, the four a_{rs} elements of the A matrix, and four b_{rs} elements of the B matrix. RMSE and Length ratios greater than 1 and negative net coverage differences favor the B-DARMA model.

	DGM:	DARMA(1,1)			tVARMA(1,1)		
	Model:	B-DARMA	DARMA	tVARMA	B-DARMA	DARMA	tVARMA
mFRSS	y_1	0.0111	0.0119	0.0113	0.0032	0.0063	0.0032
	y_2	0.0180	0.0187	0.0184	0.0056	0.0060	0.0056
	y_3	0.0103	0.0117	0.0216	0.0030	0.0041	0.0062
FMAE	y_1	0.0770	0.0777	0.0800	0.0432	0.0633	0.0432
	y_2	0.1004	0.1009	0.1020	0.0574	0.0652	0.0574
	y_3	0.0766	0.0775	0.1102	0.0422	0.0529	0.0617

Table 2.5: Simulation study mean Forecast Residual Sum of Squares (FmRSS) and Forecast Mean Absolute Error (FMAE) on the test set ($T = 40$). DGM is data generating model. Simulation study 1 is on the top left and study 2 on the top right.

(P,Q)	pairs of Fourier terms	ELPD diff	LFO ELPD
(1,0)	9	0.0	70135.3
(1,0)	10	-23.3	70112.0
(1,0)	8	-83.2	70052.0
(1,0)	6	-594.7	69540.6
(1,0)	3	-1190.1	68945.2
(1,1)	9	-101.7	70033.6
(2,0)	9	-123.7	70011.5

Table 2.6: Airbnb data analysis - Leave-future-out expected log pointwise predictive density (LFO ELPD) and the LFO ELPD differences between the best performing B-DARMA model and candidate B-DARMA models. The first five candidate models fix (P, Q) and vary pairs of Fourier terms for yearly seasonality and the last two candidate models fix the Fourier terms at 9 pairs and vary (P, Q).

		Normal Full	Horseshoe Full	Normal Nearest-Neighbor	Normal Diagonal	DAR(1)	tVAR(1)
FRSS	y_1	0.1625	0.1654	0.1708	0.1842	0.1663	0.3300
	y_2	0.0540	0.0557	0.0608	0.0639	0.0559	0.0769
	y_3	0.0369	0.0371	0.0389	0.0409	0.0370	0.0507
	y_4	0.0324	0.0325	0.0346	0.0359	0.0324	0.0438
	y_5	0.0137	0.0136	0.0139	0.0147	0.0137	0.0210
	y_6	0.0092	0.0091	0.0095	0.0098	0.0092	0.0113
	y_7	0.0050	0.0051	0.0055	0.0058	0.0050	0.0064
	y_8	0.0030	0.0031	0.0033	0.0034	0.0030	0.0039
	y_9	0.0020	0.0020	0.0020	0.0020	0.0020	0.0024
	y_{10}	0.0018	0.0018	0.0019	0.0019	0.0018	0.0020
	y_{11}	0.0022	0.0020	0.0023	0.0023	0.0022	0.0022
	y_{12}	0.0017	0.0017	0.0018	0.0018	0.0017	0.0032
	Total	0.3244	0.3291	0.3453	0.3666	0.3302	0.5540
	Average	0.0270	0.0274	0.0288	0.0306	0.0275	0.0462
FMAE	y_1	0.0170	0.0171	0.0174	0.0181	0.0172	0.0247
	y_2	0.0090	0.0094	0.0099	0.0111	0.0096	0.0113
	y_3	0.0076	0.0079	0.0081	0.0083	0.0079	0.0090
	y_4	0.0067	0.0068	0.0070	0.0079	0.0070	0.0085
	y_5	0.0048	0.0049	0.0049	0.0051	0.0049	0.0059
	y_6	0.0039	0.0039	0.0042	0.0042	0.0039	0.0044
	y_7	0.0027	0.0027	0.0027	0.0030	0.0027	0.0032
	y_8	0.0021	0.0021	0.0022	0.0023	0.0022	0.0025
	y_9	0.0018	0.0018	0.0018	0.0019	0.0018	0.0019
	y_{10}	0.0016	0.0016	0.0016	0.0016	0.0016	0.0017
	y_{11}	0.0017	0.0017	0.0017	0.0017	0.0017	0.0018
	y_{12}	0.0016	0.0016	0.0016	0.0017	0.0016	0.0022
	Total	0.0605	0.0615	0.0631	0.0669	0.0621	0.0771
	Average	0.0050	0.0051	0.0053	0.0056	0.0052	0.0064

Table 2.7: Airbnb data analysis - Forecast Residual Sum of Squares (RSS) and Forecast Mean Absolute Error (MAE) for the 12 components, their total, and their average for the test set from Feb 1, 2019 to Jan 31, 2020.

model	ELPD diff	LFO ELPD
Horseshoe Full	0.0	70149.2
Normal Full	-8.9	70140.3
Normal Nearest-Neighbor	-37.3	70112.0
Normal Diagonal	-186.2	69963.1

Table 2.8: Airbnb data analysis - Leave-future-out expected log pointwise predictive density (LFO ELPD) and the LFO ELPD differences between the best performing B-DARMA model and candidate B-DARMA models.

	par	Mean	SD	Q2.5	Q97.5
η_1	intercept	-0.228	0.055	-0.337	-0.121
η_2		-0.526	0.071	-0.666	-0.389
η_3		-0.719	0.091	-0.897	-0.545
η_4		-1.033	0.112	-1.254	-0.811
η_5		-1.483	0.136	-1.750	-1.215
η_6		-2.409	0.169	-2.740	-2.070
η_7		-2.791	0.209	-3.200	-2.384
η_8		-2.819	0.249	-3.299	-2.330
η_9		-2.999	0.301	-3.592	-2.417
η_{10}		-3.064	0.349	-3.743	-2.380
η_{11}		-4.879	0.354	-5.572	-4.176
ϕ		6.746	0.023	6.701	6.791
η_1	linear growth	0.238	0.080	0.085	0.395
η_2		0.607	0.100	0.411	0.804
η_3		0.778	0.125	0.533	1.024
η_4		1.104	0.154	0.800	1.402
η_5		1.674	0.189	1.306	2.050
η_6		2.235	0.239	1.766	2.702
η_7		2.669	0.304	2.077	3.268
η_8		2.627	0.366	1.905	3.341
η_9		2.404	0.449	1.529	3.285
η_{10}		3.547	0.520	2.531	4.564
η_{11}		7.297	0.512	6.295	8.314
ϕ		6.824	0.268	6.297	7.347

Table 2.9: Airbnb data analysis - summary coefficients for the normal full model. Parameter (par), posterior mean (Mean), standard deviation (SD), and 95 % CI. The linear growth rates are multiplied by 10^4.

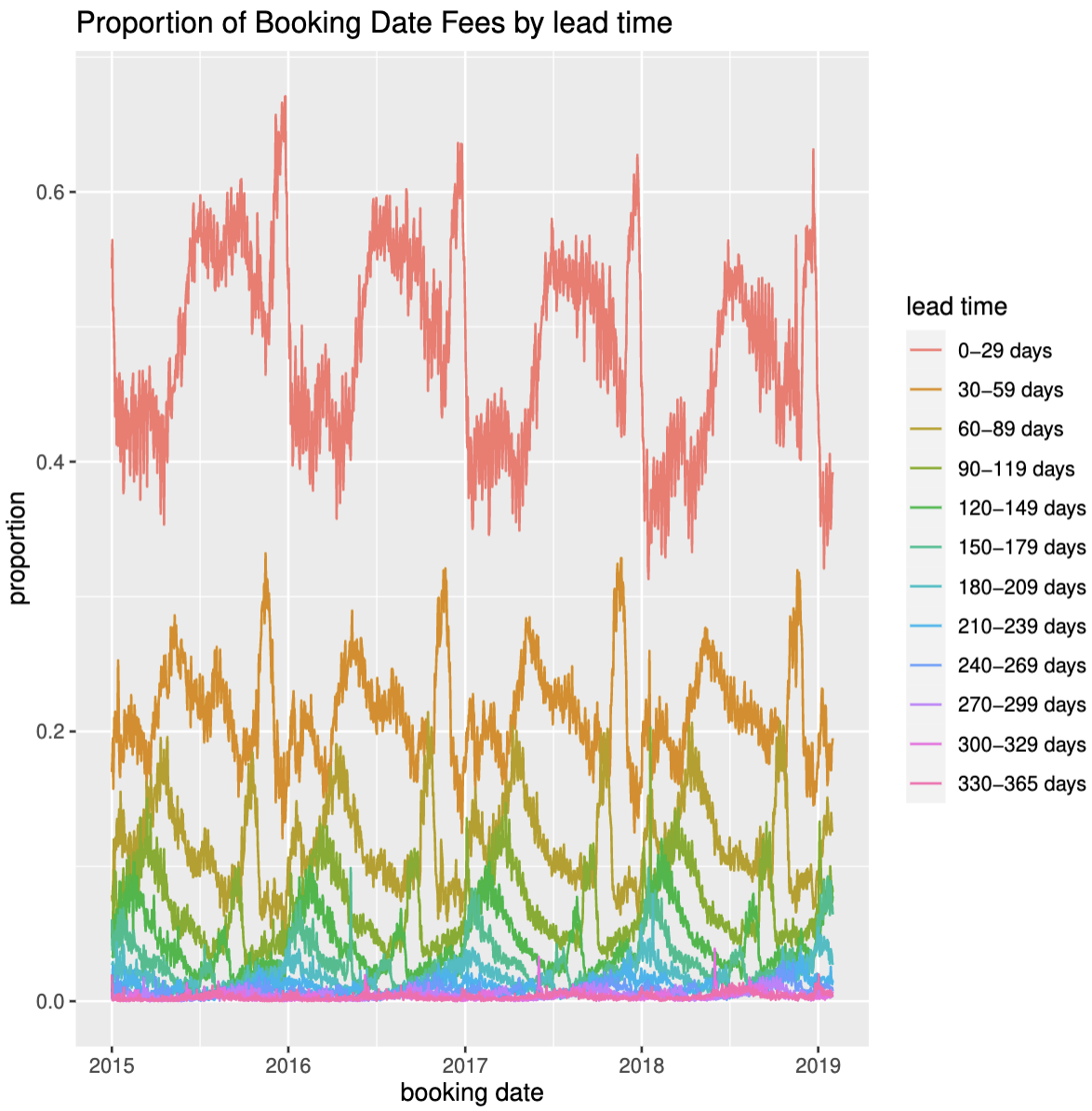

Figure 2.1: Airbnb data analysis - proportion of fees by lead time for a single large market from Jan 1, 2015 to Jan 31, 2019.

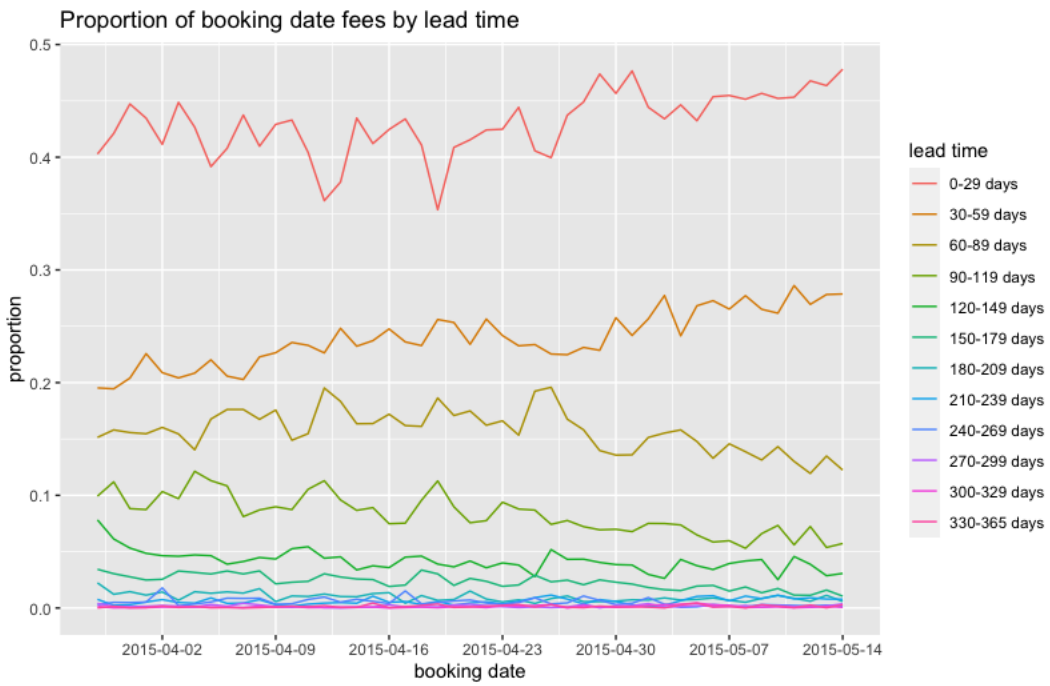

Figure 2.2: Airbnb data analysis- Proportion of fees for a single large market: weekly seasonal behavior.

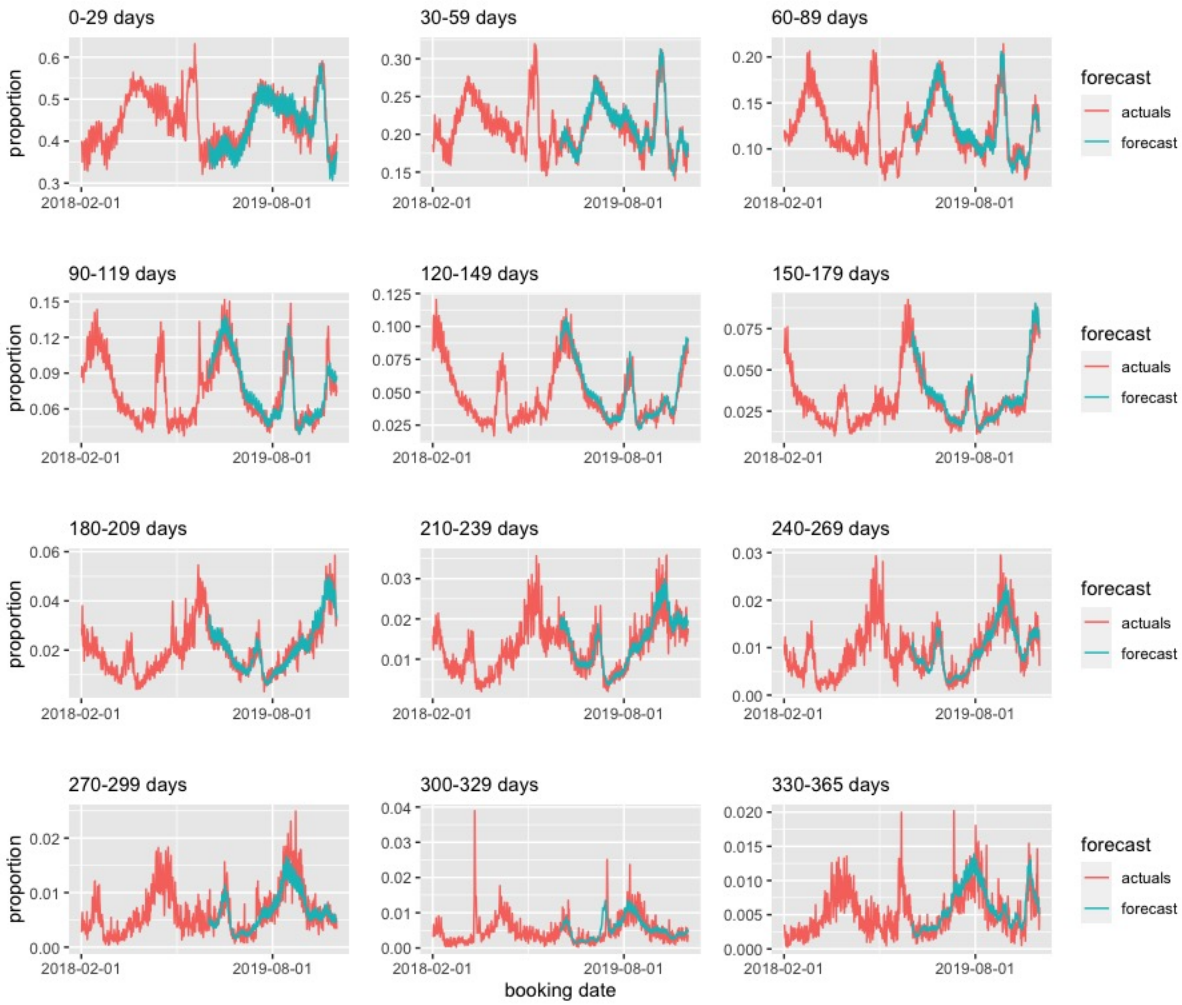

Figure 2.3: Airbnb data analysis - Normal Full Model: one year of predictions (blue) from Feb 1, 2019 to Jan 31, 2020 and two years of actuals (red) for each of the 12 components.

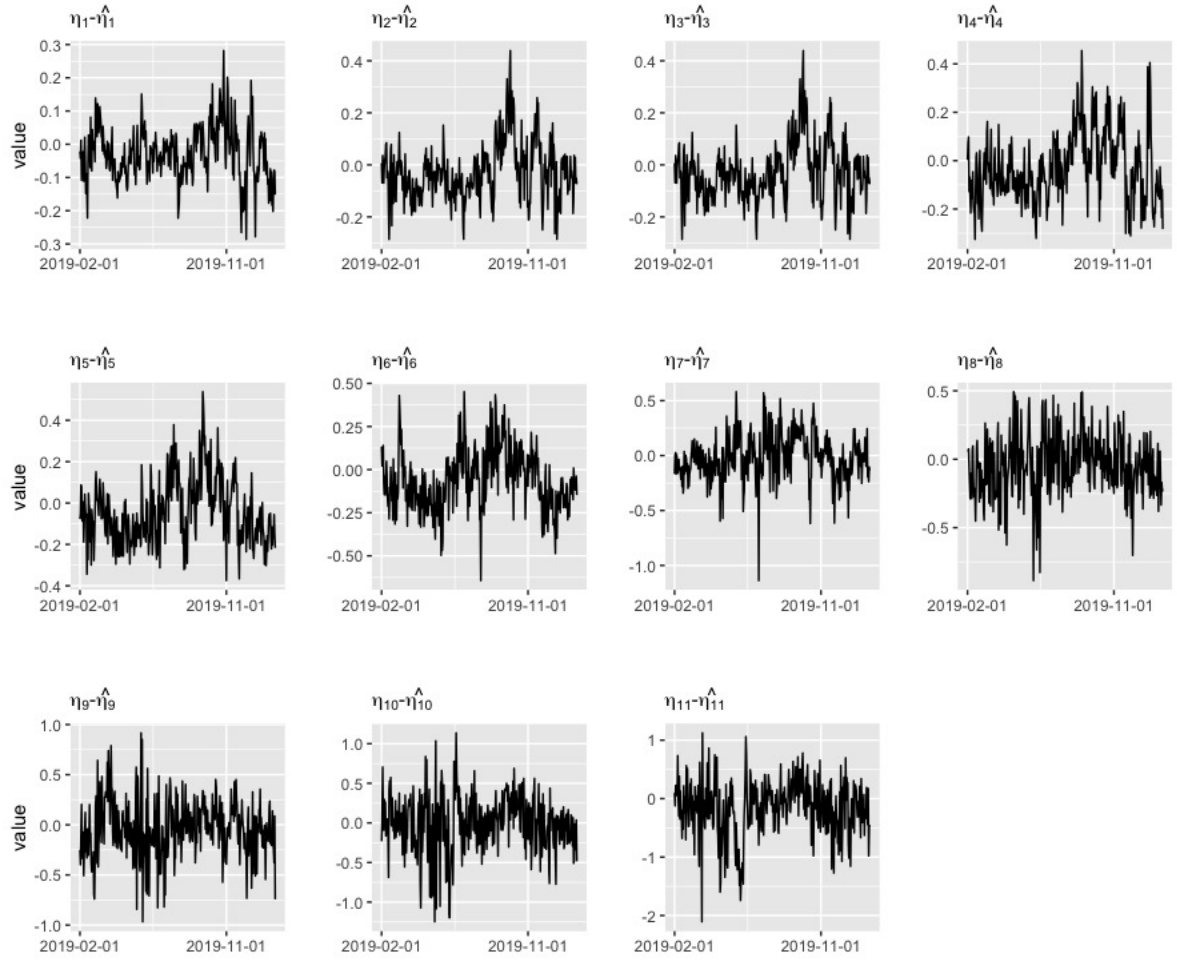

Figure 2.4: Airbnb data analysis - Normal full model residuals on the additive log ratio scale, $\boldsymbol{\eta}_t - \hat{\boldsymbol{\eta}}_t$, for the test set, Feb 1, 2019 to Jan 31, 2020.

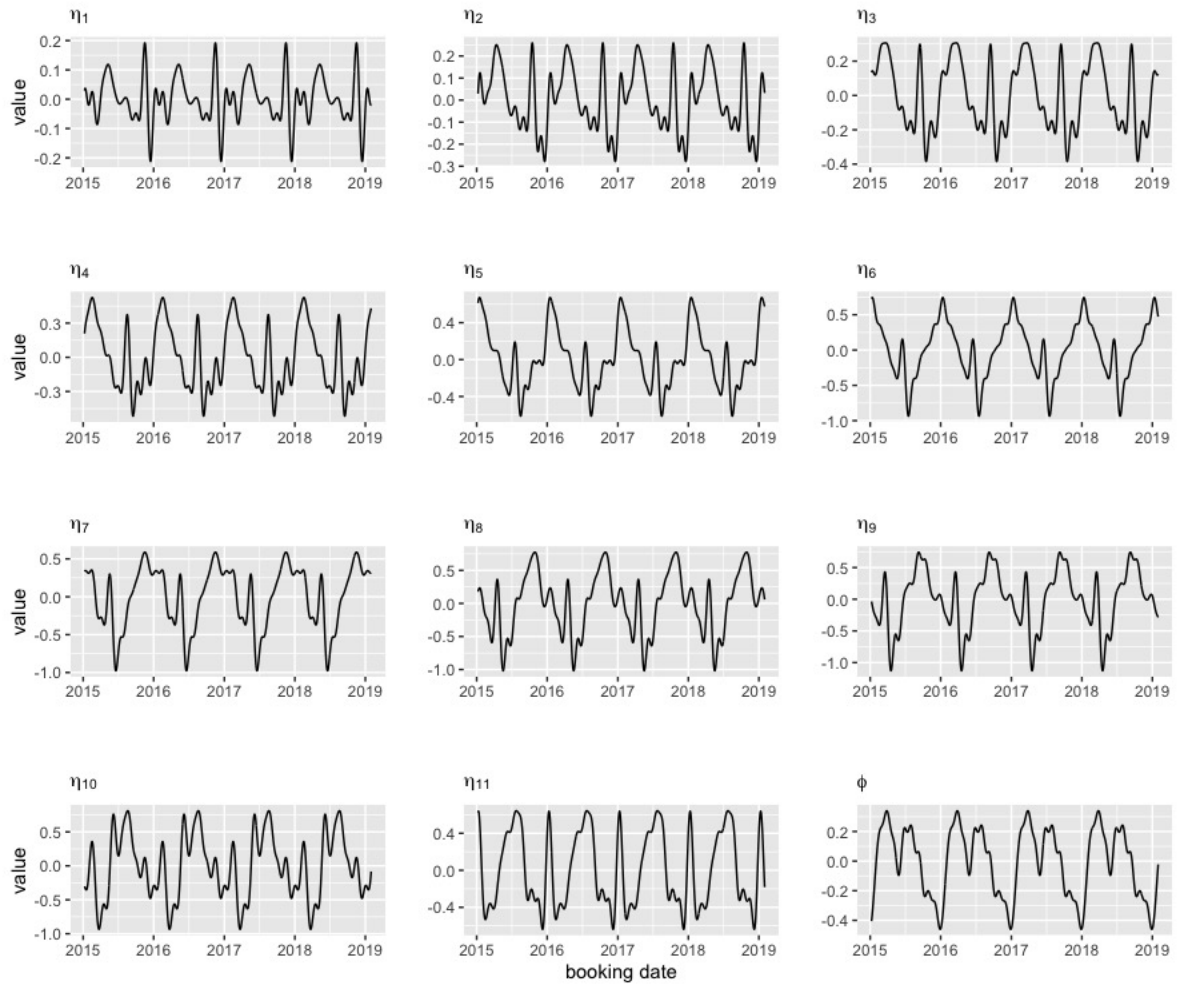

Figure 2.5: Airbnb data analysis - plot of the posterior mean yearly seasonal variation of the 11 components on the ALR scale and of the yearly seasonal variation for ϕ on the log scale for the normal full model Jan 1, 2015 to Jan 31, 2019.

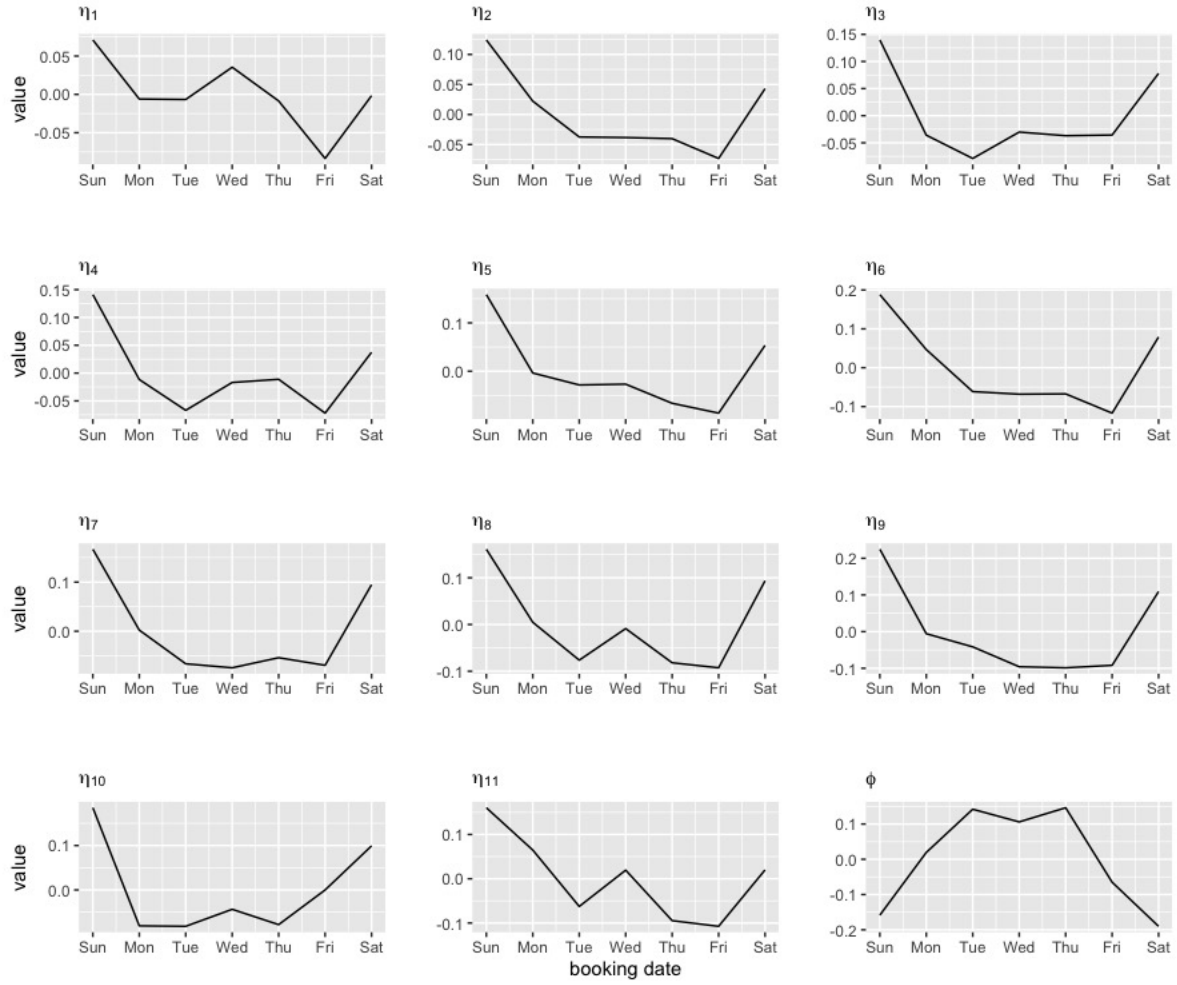

Figure 2.6: Airbnb data analysis - plot of the posterior mean weekly seasonality (η ALR scale, ϕ log scale) for the normal full model.

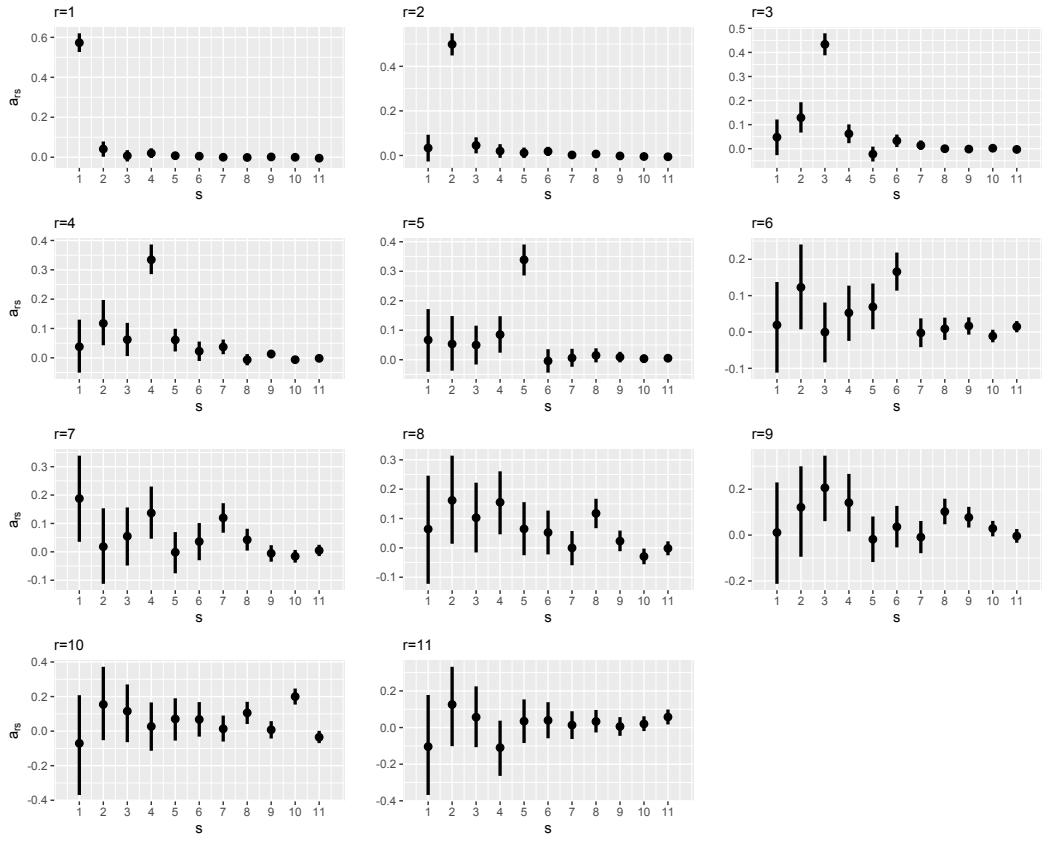

Figure 2.7: Airbnb data analysis - posterior density of elements a_{rs} in VAR A_1 matrix. The center point of each density plot is the median of the posterior distribution and line segments are 95% credible intervals.

Chapter 3

A Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity Model for Compositional Time Series

3.1 Introduction

Airbnb operates a two-sided marketplace with millions of listings worldwide, and it accommodates transactions in over 100 currencies. While individual hosts set their own nightly rates, the platform must still *forecast* these prices to predict total daily revenue. Complicating matters, travelers can pay in a currency of their choosing (often their local currency), resulting in a *dynamic compositional time series* of daily fee shares across USD, EUR, CAD, and other currencies. This currency composition is crucial for accurately converting booked stays into a consolidated U.S. Dollar (USD) amount—a requirement for company-wide financial reporting, treasury operations, and broader risk management.

To forecast revenue in USD, one must know (i) the nightly prices that hosts set in their local currency, and (ii) the fraction of total bookings ultimately paid in each currency. Even if EUR comprises only a small percentage of daily transactions, that fraction can swing

markedly from day to day, amplified by large booking volumes and exchange-rate shifts. Such changes directly affect net USD receipts and thus *any* revenue forecast that fails to anticipate compositional shifts may lead to significant errors. Moreover, once daily compositions are known, treasury teams can plan FX conversions more efficiently, hedge against unexpected fluctuations, or analyze how macroeconomic conditions influence guest payment preferences.

Each day's currency proportions must lie in a *simplex* (they are nonnegative and sum to one), a constraint that complicates standard time-series methods. One classical approach is to use log-ratio transformations, as pioneered by Aitchison (1982) and extended in works such as Cargnoni et al. (1997), Ravishanker et al. (2001), and Mills (2010), among others. Transformed additive log-ratio VARMA ($tVARMA$) models have been fruitfully applied to compositional data (Brunsdon and Smith, 1998; Mills, 2009, 2010).

In contrast, the Dirichlet distribution models compositional data directly on the simplex. Benjamin et al. (2003) proposed frequentist Generalized ARMA models that could be extended to the simplex. Contributions to Dirichlet time series include Bayesian compositional state space models by Grunwald et al. (1993), da Silva et al. (2011), and da Silva and Rodrigues (2015); the Bayesian Dirichlet Auto-Regressive Moving Average (B-DARMA) model by Katz et al. (2024); and frequentist Dirichlet ARMA models by Zheng and Chen (2017).

Seminal works in Bayesian time series analysis include West (1996), Prado and West (2010), Barber et al. (2011), Berliner (1996), Pole et al. (2018), and Koop and Korobilis (2010). Bayesian methods have been applied to Vector Auto-Regression (VAR) and Vector Auto-Regression Moving Average (VARMA) models (Spencer, 1993; Uhlig, 1997; Bańbura et al., 2010; Karlsson, 2013). Bayesian generalized linear models (GLMs) for time series data, suitable for counts and categorical responses, have been explored by Brandt and Sandler (2012), Roberts and Penny (2002), Nariswari and Pudjihastuti (2019), Chen and Lee (2016), McCabe and Martin (2005), Berry and West (2020), Fukumoto et al. (2019), Silveira de Andrade et al. (2015), and West (2013).

The generalized auto-regressive conditional heteroskedasticity (GARCH) model (Boller-

slev, 1986) is fundamental in modeling financial time series volatility. Extensions include non-normal likelihoods to capture heavy tails and skewness (Nelson, 1991; Engle, 2001). Bauwens et al. (2006) surveyed advancements in multivariate GARCH models.

Motivated by the need to capture both shifting means and evolving volatility in compositional time series—especially under sudden disruptions such as COVID-19—we develop a new class of models called **Bayesian Dirichlet Auto-Regressive Moving Average with Dirichlet Auto-Regressive Conditional Heteroskedasticity (B-DARMA-DARCH)**. Unlike existing Dirichlet ARMA methods that assume a fixed or purely deterministic precision parameter, our framework introduces a *Dirichlet Auto-Regressive Conditional Heteroskedasticity (DARCH)* component. This addition endows the standard Bayesian Dirichlet ARMA model with a GARCH-like recursion on the precision ϕ_t, enabling it to:

1. Model volatility clustering by letting ϕ_t depend stochastically on recent innovations and prior volatility, thereby reflecting abrupt regime changes and prolonged uncertainty.

2. Accommodate severe shocks (e.g., the onset of COVID-19) that induce persistent or recurrent surges in variance—phenomena not well captured by constant-precision or deterministic-variance models.

3. Preserve compositional constraints via the Dirichlet likelihood, ensuring forecasted currency shares remain valid proportions over time.

4. Leverage a full Bayesian framework to deliver predictive distributions (and credible intervals) that adapt automatically to time-varying volatility, resulting in more robust risk assessments and interval coverage.

These innovations allow B-DARMA-DARCH to simultaneously handle both systematic and stochastic changes in compositional data, providing a flexible, unified approach to *dynamic heteroskedasticity on the simplex*—a capability hitherto missing in Bayesian compositional time-series models.

We apply the B-DARMA-DARCH model to Airbnb's currency fee proportions across different regions, demonstrating its effectiveness in capturing complex temporal dynamics and volatility patterns in real-world compositional data. We conduct simulation studies comparing the B-DARMA-DARCH model with the standard B-DARMA and Bayesian transformed VARMA (B-tVARMA) models. The B-DARMA-DARCH model consistently outperforms the other models in forecast accuracy and residual diagnostics, highlighting its robustness and applicability in scenarios with heteroskedasticity, structural breaks, and regime changes. By directly modeling compositional data within the simplex and incorporating time-varying volatility, our proposed B-DARMA-DARCH model offers a powerful tool for analyzing dynamic compositional time series.

The next section presents the B-DARMA-DARCH model. Section 3.3 presents simulation studies comparing the B-DARMA-DARCH to both the B-DARMA data model and a Bayesian transformed data normal VARMA model. Section 3.4 presents analysis of the Airbnb data. The paper closes with a short discussion.

3.2 A Bayesian Dirichlet Auto-Regressive Moving Average - Dirichlet Auto-Regressive Conditional Heteroskedasticity Model

3.2.1 DARMA Data model

Consider a J-component compositional time series $\{\mathbf{y}_t\}_{t=1}^{T}$ of length T indexed by $t = 1, \ldots, T$, where each observation $\mathbf{y}_t = (y_{t1}, \ldots, y_{tJ})^\top$ lies in the simplex

$$\mathcal{S}^{J-1} = \left\{ \mathbf{y} \in (0,1)^J : \sum_{j=1}^{J} y_j = 1 \right\}.$$

We model $\mathbf{y}_t$ using the Dirichlet distribution with mean vector $\boldsymbol{\mu}_t$ and precision parameter $\phi_t > 0$

$$\mathbf{y}_t \mid \boldsymbol{\mu}_t, \phi_t \sim \text{Dirichlet}\left(\phi_t \boldsymbol{\mu}_t\right), \tag{3.1}$$

where $\boldsymbol{\mu}_t = (\mu_{t1}, \ldots, \mu_{tJ})^\top \in \mathcal{S}^{J-1}$ with probability density function (pdf)

$$p(\mathbf{y}_t \mid \boldsymbol{\mu}_t, \phi_t) = \frac{\Gamma\left(\phi_t\right)}{\prod_{j=1}^{J} \Gamma\left(\phi_t \mu_{tj}\right)} \prod_{j=1}^{J} y_{tj}^{\phi_t \mu_{tj} - 1}, \tag{3.2}$$

for $\mathbf{y}_t \in \mathcal{S}^{J-1}$.

Modeling the Mean using the Additive Log-Ratio Transformation

We model $\boldsymbol{\mu}_t$ as a function of prior observations $\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}$, prior means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{t-1}$ and known covariates $\boldsymbol{x}_t$ in a generalized linear model framework. As $\boldsymbol{\mu}_t$ and $\mathbf{y}$ are constrained, we model $\boldsymbol{\mu}_t$ after reducing dimension using the *additive log ratio* (alr) link

$$\boldsymbol{\eta}_t \equiv \text{alr}(\boldsymbol{\mu}_t) = \left(\log\left(\frac{\mu_{t1}}{\mu_{tj^*}}\right), \ldots, \log\left(\frac{\mu_{tJ}}{\mu_{tj^*}}\right) \right) \tag{3.3}$$

where $\boldsymbol{\eta}_t$ is the linear predictor, a $J-1$-vector taking values in $\mathbb{R}^{J-1}$, j^* is a chosen reference component $1 \leq j^* \leq J$, and the element of $\boldsymbol{\eta}_t$ that would correspond to the j^*th element $\log(\mu_{j^*}/\mu_{j^*}) = 0$ is omitted. Given $\boldsymbol{\eta}_t$, $\boldsymbol{\mu}_t$ is defined by the inverse of equation (3.3) where $\mu_{tj} = \exp(\eta_{tj})/(1 + \sum_{j=1}^{J-1} \exp(\eta_{tj}))$ for $j = 1, \ldots, J$, $j \neq j^*$ and for $j = j*$, $\mu_{tj*} = 1/(1 + \sum_{j=1}^{J-1} \exp(\eta_{tj}))$.

The alr-transformed mean vector η_t is modeled as a Vector Auto-Regressive Moving

Average (VARMA) process

$$\eta_t = \sum_{p=1}^{P} \boldsymbol{A}_p(\text{alr}(\mathbf{y}_{t-p}) - \mathbf{X}_{t-p}\boldsymbol{\beta}) + \sum_{q=1}^{Q} \boldsymbol{B}_q(\text{alr}(\mathbf{y}_{t-q}) - \eta_{t-q}) + \mathbf{X}_t\boldsymbol{\beta}, \tag{3.4}$$

where P is the VAR lag, Q is the VMA lag, and $t = m+1, \ldots, T$ where $m = \max(P, Q)$, $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ are $(J-1) \times (J-1)$, $\mathbf{X}_t \in \mathbb{R}^{J-1 \times r_\beta}$ is a matrix of deterministic covariates, and $\boldsymbol{\beta} \in \mathbb{R}^{r_\beta}$ is a vector of regression coefficients.

Precision Parameter ϕ_t

In the B-DARMA model, precision parameter ϕ_t is modeled with log link as a function of an r_γ-vector of covariates $\boldsymbol{z}_t$,

$$\phi_t = \exp(\boldsymbol{z}_t\boldsymbol{\gamma}), \tag{3.5}$$

where $\boldsymbol{\gamma}$ is an r_γ-vector of coefficients. If there are no covariates, $\log \phi_t = \gamma$ for all t.

3.2.2 DARCH Process for Precision Parameter ϕ_t

We introduce a DARCH process for the precision parameter ϕ_t, modeled with a log link

$$\begin{aligned}\log(\phi_t) = &\sum_{l=1}^{L} \alpha_l(\log(\phi_{t-l}) - \mathbf{z}'_{t-l}\boldsymbol{\gamma}) + \sum_{k=1}^{K} \tau_k(\text{alr}(\mathbf{y}_{t-k}) - \boldsymbol{\eta}_{t-k})'(\text{alr}(\mathbf{y}_{t-k}) - \boldsymbol{\eta}_{t-k}) \\ &+ \mathbf{z}'_t\boldsymbol{\gamma},\end{aligned} \tag{3.6}$$

where $\alpha_l \in \mathbb{R}$ are auto-regressive coefficients for the log-precision, $\tau_k \in \mathbb{R}$ are coefficients associated with past innovations, and $\mathbf{z}_t \in \mathbb{R}^{r_\gamma}$ are known covariates with coefficients $\boldsymbol{\gamma} \in \mathbb{R}^{r_\gamma}$.

For ease of notation, we set $P = L$ and $Q = K$, but this assumption is easily relaxed as shown in section 3.3. The DARCH component models the volatility in the precision parameter ϕ_t, addressing non-constant variability over time. By capturing volatility clustering

and external influences, the DARCH process enhances the model's ability to understand compositional data dynamics, improving forecasting and risk management.

Alternative log-ratio transformations, such as the centered log-ratio (CLR) or the isometric log-ratio (ILR), can also be used. A detailed discussion of these transformations is provided in the Supplementary Material (Section 3.5).

3.2.3 Joint Predictive Distribution

Define the consecutive observations $\mathbf{y}_{a:b} = (\mathbf{y}_a, \ldots, \mathbf{y}_b)'$ for positive integers $a < b$. To be well defined, linear predictor (3.4) requires having m previous observations $\mathbf{y}_{(t-m):(t-1)}$, and corresponding linear predictors $\boldsymbol{\eta}_{t-m}, \ldots, \boldsymbol{\eta}_{t-1}$. In computing posteriors, we condition on the first m observations $\mathbf{y}_{1:m}$ which then do not contribute to the likelihood. For the corresponding first m linear predictors, on the right hand side of (3.4), we set $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_m$ equal to $\text{alr}(\mathbf{y}_1), \ldots, \text{alr}(\mathbf{y}_m)$ which effectively omits the VMA terms $\boldsymbol{B}_l(\text{alr}(\mathbf{y}_{t-l}))$ from (3.4) and the MA terms τ_l $(\text{alr}(\mathbf{y}_{t-l}) - \boldsymbol{\eta}_{t-l})'(\text{alr}(\mathbf{y}_{t-l}) - \boldsymbol{\eta}_{t-k})$ from (3.6) when $t - l \leq m$. In contrast, in (3.4) the VAR terms and $\mathbf{X}_t\boldsymbol{\beta}$ and in (3.6) the AR terms and $\mathbf{z}_t'\boldsymbol{\gamma}$ are well defined for $t = 1, \ldots, m$.

Define the C-vector $\boldsymbol{\theta}$ of all unknown parameters $\boldsymbol{\theta} = (\boldsymbol{A}_{prs}, \boldsymbol{B}_{qrs}, \boldsymbol{\beta}', \boldsymbol{\gamma}', \tau_j, \alpha_i)'$, where $r, s = 1, \ldots, J-1$ index all elements of matrices $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$, $p = 1, \ldots, P$, $q = 1, \ldots, Q$, $i = 1, \ldots, L$, $j = 1, \ldots, K$ and $C = (P+Q) * (J-1)^2 + L + K + r_\beta + r_\gamma$. Our prior beliefs about $\boldsymbol{\theta}$ ($p(\boldsymbol{\theta})$) are updated by Bayes' theorem to give the posterior

$$p(\boldsymbol{\theta}|\mathbf{y}_{1:T}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m})}{p(\mathbf{y}_{(m+1):T}|\mathbf{y}_{1:m})},$$

where

$$p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m}) = \prod_{t=m+1}^{T} p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{y}_{(t-m):(t-1)}),$$

$p(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{y}_{(t-m):(t-1)})$ is the density of the Dirichlet in (3.2), and the normalizing constant $p(\mathbf{y}_{(m+1):T}|\mathbf{y}_{1:m}) = \int p(\boldsymbol{\theta})p(\mathbf{y}_{(m+1):T}|\boldsymbol{\theta}, \mathbf{y}_{1:m})d\boldsymbol{\theta}$. Our objective is to forecast the upcoming

S observations, denoted as $\mathbf{y}_{(T+1):(T+S)}$. The joint predictive distribution for these future observations is

$$p(\mathbf{y}_{(T+1):(T+S)} \mid \mathbf{y}_{1:T}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}_{(T+1):(T+S)} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T})\, d\boldsymbol{\theta}.$$

3.2.4 Differences in B-DARMA, B-tVARMA, and B-DARCH

The differences among the Bayesian-ALR-transformed VARMA (B-tVARMA), B-DARMA, and B-DARCH models lie in their handling of compositional constraints and volatility. The B-tVARMA model transforms compositional data to an unconstrained space, which may introduce distortions if model assumptions like normality are violated. In contrast, the B-DARMA and B-DARCH models directly model data within the simplex using the Dirichlet distribution. For volatility, the B-tVARMA model assumes constant variance unless extended to include heteroskedasticity. The B-DARMA model allows a time-varying precision parameter ϕ_t modeled by exogenous factors but lacks a stochastic structure, potentially missing stochastic volatility patterns. The B-DARCH model introduces a dynamic ϕ_t with an ARMA structure, capturing both systematic and stochastic changes in volatility, such as clustering and sudden shifts. While the B-tVARMA model is simpler, it may inadequately represent compositional data and volatility dynamics. The B-DARMA model offers a balance but may miss important volatility features. The B-DARCH model, though more complex, comprehensively accounts for mean dynamics and volatility patterns, enhancing model fit and forecasting accuracy when heteroskedasticity is present.

3.3 Simulation Study

3.3.1 Overview

We conducted six simulation studies to evaluate the performance of the B-DARMA(1,0)-DARCH(1,1) model compared to the B-DARMA(1,0) and B-tVARMA(1,0) models. Each study generates data from either a DARMA, DARMA-DARCH, or tVARMA and introduces random shocks to simulate misreported observations (1-3) or redraws values for all parameters (4-6) to simulate a regime shift. Each study simulates 50 datasets, and evaluates out-of-sample forecast accuracy and in-sample residual autocorrelation.

3.3.2 Simulation Studies Setup

In all six studies, datasets with five compositional components were simulated over $T = 100$ time points. The first 60 observations were used for training, and the remaining 40 for testing. For each simulation, a new set of parameters was drawn, giving different intercept vectors, auto-regressive matrices, and precision parameters across simulations.

All six simulation studies have the same underlying mean vector formulation for the latent ALR-transformed process

$$\boldsymbol{\eta}_t = \boldsymbol{\beta} + \boldsymbol{A}(\text{alr}(\mathbf{y}_{t-1}) - \boldsymbol{\beta}).$$

The auto-regressive coefficient matrix $\boldsymbol{A}$ was a 4×4 matrix with elements drawn from a uniform distribution, $A_{ij} \sim \mathcal{U}(-0.75, 0.75)$, independently for each simulation. For each of the 50 simulations, the elements of the intercept vector $\boldsymbol{\beta}$ were generated by drawing five elements $\beta_j^*, j = 1, \ldots, 5$ from $\mathcal{N}(0.2, 0.03^2)$, then normalized to sum to 1 ($\beta_j^* = \frac{\beta_j^*}{\sum_{j^*=1}^{5} \beta_j^*}$), transformed using the ALR transformation to give the 4×1 intercept vector $\boldsymbol{\beta} = \text{alr}(\boldsymbol{\beta}^*)$.

For each simulation, the initial values were

$$y_{1j} = .2 + \epsilon_j,\ \epsilon_j \sim N(0, .01^2),\ j = 1, \ldots, 5$$

followed by $y_1 = \frac{y_1}{\sum y_1}$.

Simulation Studies 1 & 4: DARMA DGP

In Simulations 1 and 4, the data-generating process (DGP) was a DARMA model, $\mathbf{y}_t|\mu_t, \phi_t \sim$ Dirichlet$(\phi_t\mu_t)$ with a constant precision parameter $\log(\phi_t) = \phi_0$. The precision parameter ϕ_0 was drawn independently for each simulation from a uniform distribution between 6 and 7.5, $\phi_0 \sim \mathcal{U}(6, 7.5)$.

Simulation Studies 2 & 5: DARCH DGP

The data in the second and fifth simulations were generated using a DARMA-DARCH process, where ϕ_t was generated

$$\log(\phi_t) = \phi_0 + \alpha(\log(\phi_{t-1}) - \phi_0) + \tau \sum_{i=1}^{4} (\text{alr}(y_{t-1,i}) - \eta_{t-1,i})^2.$$

The true intercept ϕ_0 was drawn independently for each simulation from a uniform distribution between 6 and 7.5, $\phi_0 \sim \mathcal{U}(6, 7.5)$. We fixed parameters $\alpha = 0.8$ and $\tau = -0.95$ to simulate data with high volatility persistence and strong responsiveness to past innovations, creating heteroskedasticity patterns that test the B-DARCH model's ability to capture dynamic volatility.

Simulation Studies 3 & 6: tVARMA DGP

The data in the third and sixth simulations were generated using a tVARMA process with

$$\boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{\beta} + \boldsymbol{A}(\text{alr}(\mathbf{y}_{t-1}) - \boldsymbol{\beta}), \sigma^2\boldsymbol{\Sigma})$$

$$\mathbf{y}_t = \text{alr}^{-1}(\boldsymbol{\eta}_t).$$

A single precision parameter σ was drawn independently for each simulation from $\sigma \sim \mathcal{U}(0.05, 0.5)$. The covariance matrix $\boldsymbol{\Sigma}$ was generated to be positive-definite with complex interdependencies $\boldsymbol{\Sigma} = \mathbf{M}^\top \mathbf{M}$, where $\mathbf{M}$ was a 4×4 matrix with elements drawn independently from $\mathcal{U}(-0.3, 0.3)$ for each simulation.

Simulated Shocks

In simulations 1-3, random shocks were introduced into the training data to simulate misreported observations. Time points were selected by sampling intervals from a Poisson(6) distribution and cumulatively summing them. At these times l, $\mathbf{y}_t$ was replaced with a random composition generated by independently drawing each component from a Uniform(0,1) distribution and normalizing the vector to sum to 1,

$$y_{lj} \sim \mathcal{U}(0,1),\ \mathbf{y}_{\text{misreported}[l]} = \frac{y_{lj}}{\sum_{j=1}^{5} y_{lj}}, j = 1, \ldots, 5.$$

Regime Shifts

In Simulation 4-6, we introduced regime shifts at a random time point t_{shift} uniformly selected from [10, 50] by redrawing the intercept vector $\boldsymbol{\beta}$, the auto-regressive matrix A and the precision parameters (either ϕ_0, σ, Σ) for the DARMA/DARCH models or tVARMA model. For $t = t_{\text{shift}}$ to $t = t_{\text{shift}} + 9$, data were generated using these new parameters, simulating a regime shift in the data-generating process. After this period, parameters reverted to their original values, allowing us to evaluate the models' robustness to temporary structural changes.

3.3.3 Model Comparisons

We fitted three models to each simulated dataset for comparison: the B-DARMA(1,0)-DARCH(1,1), the Bayesian transformed VARMA (B-tVARMA(1,0)), and the B-DARMA(1,0) models. All models shared the following priors: the intercepts $\boldsymbol{\beta}$ were given Normal$(0, 0.3^2)$ priors, and the elements of the auto-regressive matrices $\boldsymbol{A}_p$ had Normal$(0, 1)$ priors.

For the B-DARMA(1,0)-DARCH(1,1) model, the initial precision parameter ϕ_0 was assigned a Normal$(7, 1.5^2)$ prior, and the AR and MA terms for ϕ_t had Normal$(0.35, 0.5^2)$ and Normal$(-0.75, 0.5^2)$ priors, respectively.

In the B-tVARMA(1,0) model $\boldsymbol{\Sigma} = \sigma^2\Omega$, where σ is a scale parameter and Ω is a correlation matrix that reflects the dependencies between variables. The correlation matrix Ω is defined through its Cholesky factor L_Ω, which ensures that Ω is positive-definite and properly structured. We assigned a prior to L_Ω using the LKJ_Corr_Cholesky(3) distribution, which is a prior over Cholesky factors of correlation matrices with a shape parameter of 3. The scale parameter σ was given a half-Normal prior, $\sigma \sim$ Half-Normal$(0, 0.5^2)$, to ensure positivity.

All models were fit using Stan with four chains of 1,000 iterations each (500 warm-up), yielding 2,000 posterior samples.

3.3.4 Metrics

We assess model performance using **Forecast Root Mean Squared Error (FRMSE)** and **Forecast Mean Absolute Error (FMAE)** for each of the five components. These metrics quantify the out-of-sample forecast accuracy

$$\text{FRMSE}_j = \left(\frac{1}{40} \sum_{t=61}^{100} (y_{tj} - \bar{\mu}_{tj})^2 \right)^{\frac{1}{2}}$$

$$\text{FMAE}_j = \frac{1}{40} \sum_{t=61}^{100} |y_{tj} - \bar{\mu}_{tj}|$$

where $\bar{\mu}_{tj}$ represents the posterior mean of the predicted value at time t for component j. We then summarize by averaging them across all five components $\text{FRMSE}_{mean} = \frac{1}{5}\sum_{j=1}^{5} \text{FRMSE}_j$ and $\text{FMAE}_{mean} = \frac{1}{5}\sum_{j=1}^{5} \text{FMAE}_j$.

Additionally, we assess the **Partial Autocorrelation Function (PACF)** (see Supplementary Section 3.5) of the sums of squared standardized residuals to evaluate how well each model captures residual dynamics during the in-sample period. For each model, residuals are standardized according to the model's underlying distribution (see Supplementary Section 3.5).

3.3.5 Results

Simulation Studies 1-3: Shocks

The left side of table 3.1 shows the forecast performance FRMSE and FMAE metrics for the B-DARMA, B-DARCH, and B-tVARMA models across 50 simulated datasets in three different scenarios.

Across simulation studies 1–3, the B-DARCH model consistently had the lowest FRMSE and FMAE values compared to the second-best model, the B-tVARMA. In simulation study 1, FRMSE and FMAE were 1.6% and 1.4% lower, respectively. In simulation study 2, they were 12.6% and 7.7% lower. In simulation study 3, FRMSE and FMAE were 5.9% and 7.8% lower than B-tVARMA.

Figure 3.S1 shows the mean PACF values across all simulations for B-DARCH, B-DARMA, and B-tVARMA models. The mean PACF values for the B-DARCH model consistently remain closer to zero across the majority of lags, with fewer pronounced PACF values. Notably, the B-DARMA and B-tVARMA models show large negative PACF values, particularly for early lags (lags 1-5). These large values indicate that residuals in these models have substantial autocorrelation, suggesting less efficient capture of the underlying dependencies across time within the data. In contrast, B-DARCH shows minimal autocorrelation for early lags, reflecting an overall improvement in residual behavior and a better fit to the

true data-generating process.

Simulation Studies 4-6: Regime Shifts

The right side of Table 3.1 shows the forecast performance metrics FRMSE and FMAE for the B-DARMA, B-DARCH, and B-tVARMA models across 50 simulations with regime shifts. The B-DARCH model consistently had the lowest FRMSE and FMAE values, outperforming the second-best model (B-tVARMA) with 14.6% lower FRMSE and 8.4% lower FMAE in simulation study 4, 4.7% and 4.6% lower in simulation study 5, and 10.5% and 4.8% lower in simulation study 6. These results highlight the B-DARCH model's comparative advantage in adapting to sudden changes in the underlying dynamics.

Figure 3.S2 shows the mean PACF values for simulations 4-6 for B-DARCH, B-DARMA, and B-tVARMA models. The B-DARCH model has positive peaks for early lags, with autocorrelation generally settling closer to zero across subsequent lags. Notably, the magnitude of the PACF values for B-DARCH is smaller than those of the other models, except for Simulation 4, where the B-DARMA model shows a smaller magnitude at lag 1.

In contrast, both B-DARMA and B-tVARMA have strong positive PACF values for the first few lags, which then gradually decrease to negative values at later lags, indicating greater persistence of autocorrelation effects. This pattern suggests that B-DARMA and B-tVARMA models struggle more with handling shifts in the underlying process, resulting in larger autocorrelations in early lags compared to the B-DARCH model.

These results indicate that the B-DARCH model's dynamic volatility component enables it to adapt more effectively to structural breaks, capturing both systematic and stochastic changes in volatility. In contrast, the B-DARMA and B-tVARMA models, lacking a stochastic volatility structure, were less effective in handling the regime shifts, leading to higher forecast errors.

Model	(DARMA Shock)		(DARMA Regime Shift)	
	FRMSE	**FMAE**	**FRMSE**	**FMAE**
B-DARMA	38.6	2.94	25.3	2.58
B-DARCH	38.0	2.87	18.1	2.19
B-tVARMA	38.6	2.91	21.2	2.37
Model	**(DARCH Shock)**		**(DARCH Regime Shift)**	
	FRMSE	**FMAE**	**FRMSE**	**FMAE**
B-DARMA	11.2	1.84	20.4	2.28
B-DARCH	9.79	1.69	18.3	2.09
B-tVARMA	11.3	1.83	19.2	2.19
Model	**(tVARMA Shock)**		**(tVARMA Regime Shift)**	
	FRMSE	**FMAE**	**FRMSE**	**FMAE**
B-DARMA	25.6	2.43	32.5	2.59
B-DARCH	24.1	2.24	22.9	2.18
B-tVARMA	24.3	2.31	25.7	2.29

Table 3.1: Summary of Model Performance Metrics times 100 on test set ($T = 40$) across 50 simulations for Simulation Studies 1–6. Average FRMSE is the average Forecasted Root Mean Squared Error across all five components; Average Forecasted Mean Absolute Error is the average FMAE across all five components.

3.4 Airbnb Data Analysis

The dataset comprises daily records from January 1, 2017, to December 31, 2020, detailing the proportion of fees in U.S. Dollars (USD) originating from various billing currencies. The data is categorized into four anonymized diverse geographic regions, referred to as Regions 1-4. Within each region, the fees are broken down by the top five currencies specific to that region, with the remaining currencies grouped into an "other" category. The top five currencies for each region are fixed for the entire period. The training period spans from January 1, 2017, to June 30, 2020. The subsequent three months (July 1, 2020 to September 30, 2020) are used for validation and model specification, while the final 3 months (October 1, 2020, to December 31, 2020) constitute the test set.

3.4.1 Exploratory Data Analysis

Regional Currency Dynamics

The currency proportions for each region are shown in Figures 3.1 and Supplementary Figures 3.S3 and 3.S4. Using the U.S. Dollar (USD) as the reference category, 30-day rolling ALR means are shown in Supplementary Figures 3.S5 and 3.S6 while 30-day rolling ALR variances are shown in Supplementary Figure 3.S7 and Figure 3.2. The correlations of the ALR values with their lag 1 values is shown in Supplementary Figure 3.S8.

Across all four regions, currencies exhibit seasonal cycles and trends relative to USD. Regions 1 and 3 display pronounced yearly seasonality for multiple currencies, with Region 4 exhibiting the strongest seasonal patterns. Pre-COVID-19 (before 2020), currencies in Regions 1 and 2 generally showed steady levels or consistent declines relative to USD, while in Region 3, the Australian Dollar (AUD) and other currencies displayed distinct upward trends. During the COVID-19 period, starting around January or February of 2020, significant shifts are seen: in Region 1, all currencies except the Canadian Dollar (CAD) saw large declines in their means at the expense of the USD; in Region 2, the Euro (EUR) and Swiss Franc (CHF) experienced the largest increases as the USD decline; in Region 3, AUD continued its sharp rise, while the New Zealand Dollar (NZD) initially declined but rapidly recovered; in Region 4, the Brazilian Real (BRL) experienced sharp increases in its rolling mean before being overtaken by USD, and the Mexican Peso (MXN) peaked before a drastic decline.

Lag 1 correlations show strong autocorrelations within currencies, with high self-correlations observed in EUR and GBP in Region 1 and in BRL, Chilean Peso (CLP), and MXN in Region 4. Positive lag 1 cross correlations were found among currencies in Region 2, except for CAD, while in Region 3, NZD and AUD were strongly correlated, and EUR was negatively correlated with other currencies. In Region 4, negative correlations between BRL and EUR, as well as between CLP and EUR, suggest opposing movements between these currency pairs.

Figure 3.2 and Supplementary Figure 3.S7 show the 30-day rolling variance of each currency's ALR-transformed proportions. Rather than a single abrupt change, we observe *recurrent* episodes of elevated variability, often persisting for several weeks before subsiding. For instance, Region 1 (top left) shows multiple surges for EUR and GBP, culminating in early 2020 and tapering mid-year, while Region 2 (top right) begins with a high-variance phase (notably in CAD and CHF), settles to moderate levels around 0.1–0.2, and then rises again at the onset of 2020. Region 3 (bottom left) experiences shorter bursts around 2018 and late 2019, each followed by another peak in early 2020, suggesting that once volatility increases, it may linger rather than immediately revert. Finally, Region 4 (bottom right) transitions from moderate day-to-day swings to abrupt climbs in BRL and CLP near 2019, with additional spikes in BRL and MXN emerging after 2020. In every region, these heightened-variance intervals emerge at distinct times and exhibit varying durations, indicating that volatility in currency shares is shaped by more than one-off shocks, and can arise repeatedly under evolving market conditions. Such patterns underscore the importance of accommodating dynamic, recurrent shifts in variance for effective modeling and forecasting.

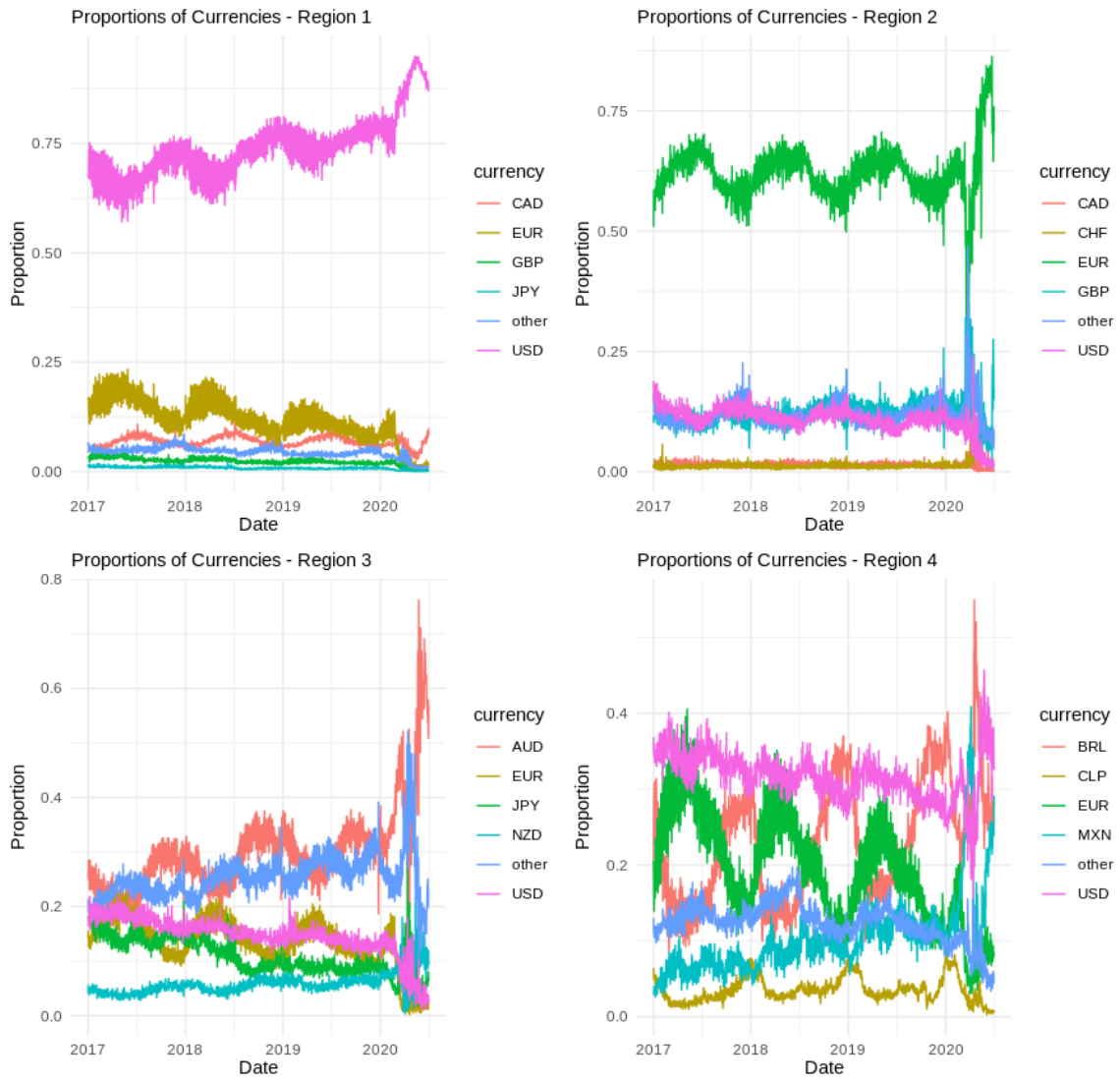

Figure 3.1: Airbnb data analysis - proportion of fees by billing currency for four regions from Jan 1, 2017 to June 30, 2020. AUD is the Australian dollar, BRL is the Brazillian Real, CAD is the Canadian Dollar,CHF is the Swiss Franc, CLP is the Chilean Peso,EUR is the European Euro, GBP is the Great British Pound, MXN is the Mexican Peso, NZD is the New Zealand Dollar, and USD is the US Dollar.

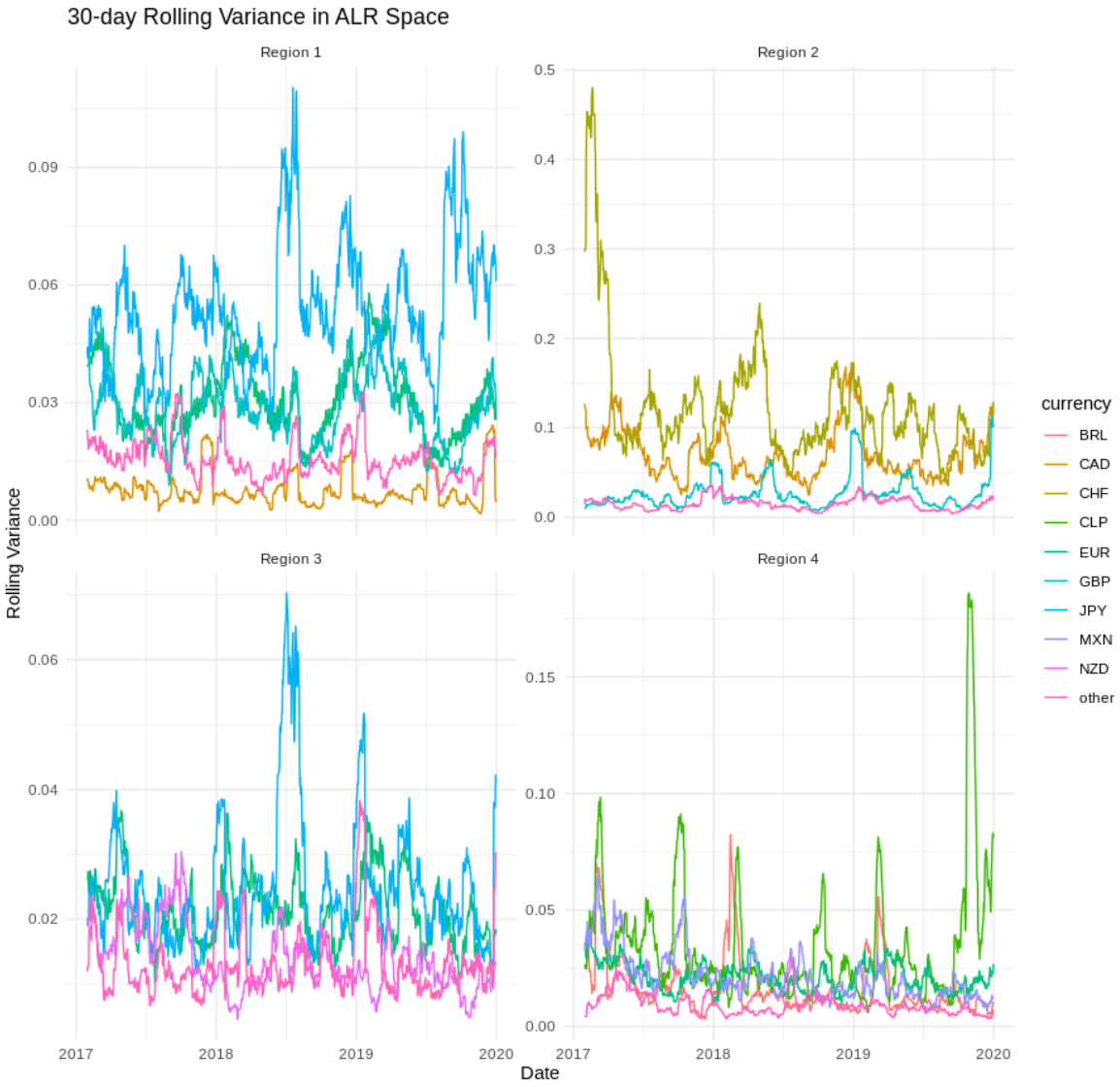

Figure 3.2: Airbnb data analysis - proportion of fees by currency- 30-day rolling ALR variance for four regions from Jan 1, 2017 to Dec 31, 2019. USD is the reference component.

3.4.2 Objective

The primary goal of this analysis is to compare the performance of three model classes: Bayesian Dirichlet Auto-Regressive Moving Average, Bayesian transformed Vector Auto-Regressive Moving Average, and Bayesian Dirichlet Auto-Regressive Moving Average with Dirichlet Auto-Regressive Conditional Heteroskedasticity.

3.4.3 Comparison Models

The B-DARCH and B-DARMA models share the same model for $\boldsymbol{\eta}$ but differ in their treatment of the precision parameter. In the B-DARCH model, the precision parameter has its own AR and MA components, allowing for dynamic heteroskedasticity that evolves over time. In contrast, the B-DARMA model does not include these components in the precision parameter, relying solely on auto-regressive and moving average terms in the mean structure to account for time dependence.

The third model, B-tVARMA, is a Bayesian normal VARMA model applied to transformed data. It transforms the observed multivariate time series $\mathbf{y}_t$ using the additive log-ratio (alr) transformation, resulting in $\text{alr}(\mathbf{y}_t) \in \mathbb{R}^{J-1}$. The transformed data is then modeled as $\text{alr}(\mathbf{y}_t) \sim N(\boldsymbol{\eta}_t, \boldsymbol{\Sigma})$, where $\boldsymbol{\eta}_t$ has time-varying mean defined in (3.4), and $\boldsymbol{\Sigma}$ is an unknown $J-1 \times J-1$ positive definite covariance matrix. The covariance matrix $\boldsymbol{\Sigma}$ is modeled through its Cholesky factor L_Σ, scaled by σ, $\boldsymbol{\Sigma} = \sigma^2\Omega$.

All models are fit with STAN (Stan Development Team, 2022) in R. We run 4 chains with 2000 iterations each with a warm up of 1000 iterations for a posterior sample of 4000. Initial values for all parameters are selected randomly from the interval $[-1, 1]$.

3.4.4 Seasonal and Trend Terms

Figure 3.1 and Supplementary figure 3.S3 show varying levels of yearly and weekly seasonality in the currency data in different regions. There is a recurring yearly pattern driven by popular

travel periods and a clear weekly seasonality reflecting variations between days of the week. Consequently, we include weekly and yearly seasonal variables and a trend variable in the predictors $\mathbf{X}_t$ and the same weekly and yearly seasonal variables in $\boldsymbol{z}_t$. Both ϕ_t and the elements of $\boldsymbol{\eta}_t$ are modeled with a linear trend and Fourier terms for seasonal variables, using pairs of $\left(\sin \frac{2k\pi t}{w_{\text{season}}}, \cos \frac{2k\pi t}{w_{\text{season}}}\right)$ for $k = 1, \ldots, K_{\text{season}} \leq \frac{w_{\text{season}}}{2}$, with $w_{\text{season}} = 7$ for weekly seasonality and $w_{\text{season}} = 365.25$ for yearly seasonality. The orthogonality of the Fourier terms aids in model convergence. The B-DARCH, B-DARMA, and B-tVARMA all share the same $\mathbf{X}_t$ while the B-DARCH and B-DARMA share the same $\boldsymbol{z}_t$.

3.4.5 Model Specification and Validation

The models are fit using data from January 1, 2017, to June 30, 2020, as the training set. The validation set, spanning July 1, 2020, to September 30, 2020, is used to determine optimal lags and exogenous covariates. The test set from October 1, 2020, to December 31, 2020, is used for final model comparison.

For the initial specification, we fix $P = 1$, $Q = 0$, and $K_{\text{week}} = 3$ for each region, allowing the intra-year Fourier terms to vary. The optimal number of Fourier terms (K_{year}) is selected based on the model's performance on the validation set. With the optimal (K_{year}) established, we vary P and Q to identify the best-performing orders of the auto-regressive (AR) and moving average (MA) components on the validation set. Using these specifications, the models are retrained up until September 30, 2020, and tested on the independent test set for the B-DARMA, B-DARMA-DARCH, and B-tVARMA models.

3.4.6 Priors for B-DARCH, B-DARMA, and B-tVARMA

For all models (B-DARCH, B-DARMA, and B-tVARMA), we assume each component's auto-regressive (AR) and moving average (MA) elements, a_{rs} and b_{rs}, are strongest with their own lags. Thus, we assign weakly informative normal priors $N(.4, .5^2)$ to a_{rrp}, b_{rrq} for diagonal elements ($r = 1, \ldots, 4, q = 1, \ldots, 3$), and $N(0, .5^2)$ for non-diagonal elements in the

$\boldsymbol{A}$ and $\boldsymbol{B}$ matrices.

The priors for the Fourier, trend, and intercept regression coefficients are $N(0, 1)$, $N(0, .1^2)$, and $N(0, 2^2)$ respectively. For the B-DARMA and B-DARCH, the precision model predictors $\boldsymbol{z}_t$ have the same seasonal terms as the mean models, with corresponding priors for the coefficients in $\boldsymbol{\gamma}$ matching those of the $\boldsymbol{\eta}$ terms. The B-DARCH has a $N(0, 1)$ prior on its AR α and MA τ terms.

For the B-tVARMA, we impose a half normal prior $N^+(0, .5^2)$ on the scale parameter σ and the correlation matrix Ω has a LKJ(3) prior on its Cholesky factor L_Ω.

3.4.7 Results

Model Specification

Out-of-sample validation results (Supplementary Table 3.S1) show that Region 1 performed best with $K_{\text{year}} = 8$ and $(P, Q) = (1, 0)$, yielding a average FMAE of 0.0036 and total FRSS of .0173. Region 2 achieved optimal performance with $K_{\text{year}} = 10$ and $(P, Q) = (1, 1)$, resulting in a average FMAE of .0127 and tRSS of .2018. In Region 3, the optimal model had $K_{\text{year}} = 10$ and $(P, Q) = (1, 1)$, with a average FMAE of .0283 and total FRSS of 1.1395. Region 4 was best modeled with $K_{\text{year}} = 8$ and $(P, Q) = (3, 2)$, with a FMAE of .0116 and FRSS of .1553.

Performance on Test Set

The Forecast Mean Absolute Error (FMAE) and Forecast Residual Sum of Squares (FRSS) for each region and major currency are shown in Table 3.2, while Figures 3.3–3.6 display the corresponding out-of-sample forecasts $\bar{\mu}_{tj}$. When errors are aggregated over all currencies within each region, the B-DARCH model consistently delivers the lowest total FMAE and FRSS. For example, in Region 1 the total FMAE (FRSS) for B-DARCH is 2.84 (3.15), which is markedly lower than 7.65 (19.11) for B-DARMA and 5.20 (10.62) for B-tVARMA. In Region 2, the advantage of B-DARCH is even more pronounced: its total FMAE is

13.55 (with an average of 2.26) and total FRSS is 73.93 (average 12.32), compared to 14.71 (2.45) and 81.87 (13.65) for B-DARMA, and 17.13 (2.86) and 104.43 (17.41) for B-tVARMA. Regions 3 and 4 exhibit similar aggregate advantages for B-DARCH.

At the individual-currency level, however, there are still several cases in which either B-DARMA or B-tVARMA yield lower FMAE—particularly in Regions 3 and 4. In Region 2, by contrast, the B-DARCH yields the smallest FMAE across *all* major currencies (CAD, CHF, EUR, GBP, USD, and "other"). In Regions 3 and 4, there remain a few currencies that favor one of the alternative models. For instance, in Region 3, B-DARMA produces slightly lower FMAE than B-DARCH for JPY (2.17 vs. 2.20) and NZD (2.91 vs. 3.57), while in Region 4, B-DARMA outperforms B-DARCH for EUR (0.58 vs. 1.29) but not for most of the other currencies. Notably, B-tVARMA is especially competitive for certain currencies in Region 1 (e.g., EUR), likely because its assumption of constant variance aligns well with periods of relatively stable volatility. Nonetheless, when the full basket of currencies is considered, the time-varying volatility framework of B-DARCH leads to the most accurate overall forecasts.

Supplementary Table 3.S2 shows the empirical coverage rates of the 95% credible intervals for the forecasts across regions and currencies. The B-DARCH model exhibits relatively stable coverage that is generally closer to the nominal 95% level, with mean coverage rates of 0.92, 0.91, 0.87, and 0.91 in Regions 1, 2, 3, and 4, respectively. By contrast, the B-DARMA model shows greater variability, with mean coverages ranging from 0.82 in Region 1 down to 0.55 in Region 3, and 0.82 in Region 2 and 0.84 in Region 4 B-tVARMA tends to undercover, with mean rates of 0.63, 0.86, 0.69, and 0.85 across Regions 1–4. These results suggest that although B-tVARMA can occasionally yield competitive point forecasts, its assumption of time-invariant volatility often leads to overconfident interval estimates. In contrast, the dynamic volatility specification in B-DARCH not only improves point forecast accuracy but also delivers more reliable interval estimates.

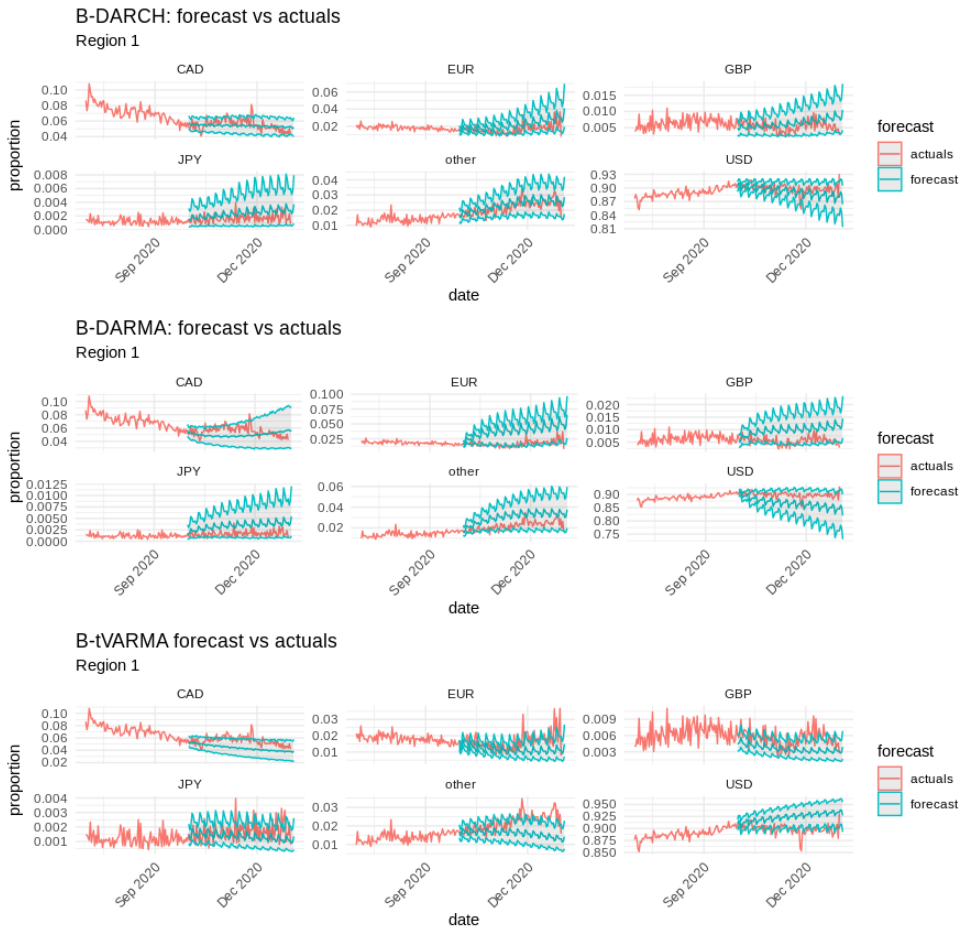

Figure 3.3: Airbnb data analysis - Region 1: 92-day forecasts (blue) with 95% confidence intervals, represented by the shaded regions around the forecast lines, for each of the six currencies, from October 1, 2020, to December 31, 2020, compared to the actual values (red) observed over the preceding six months. The models compared include B-DARCH, B-DARMA, and B-tVARMA.

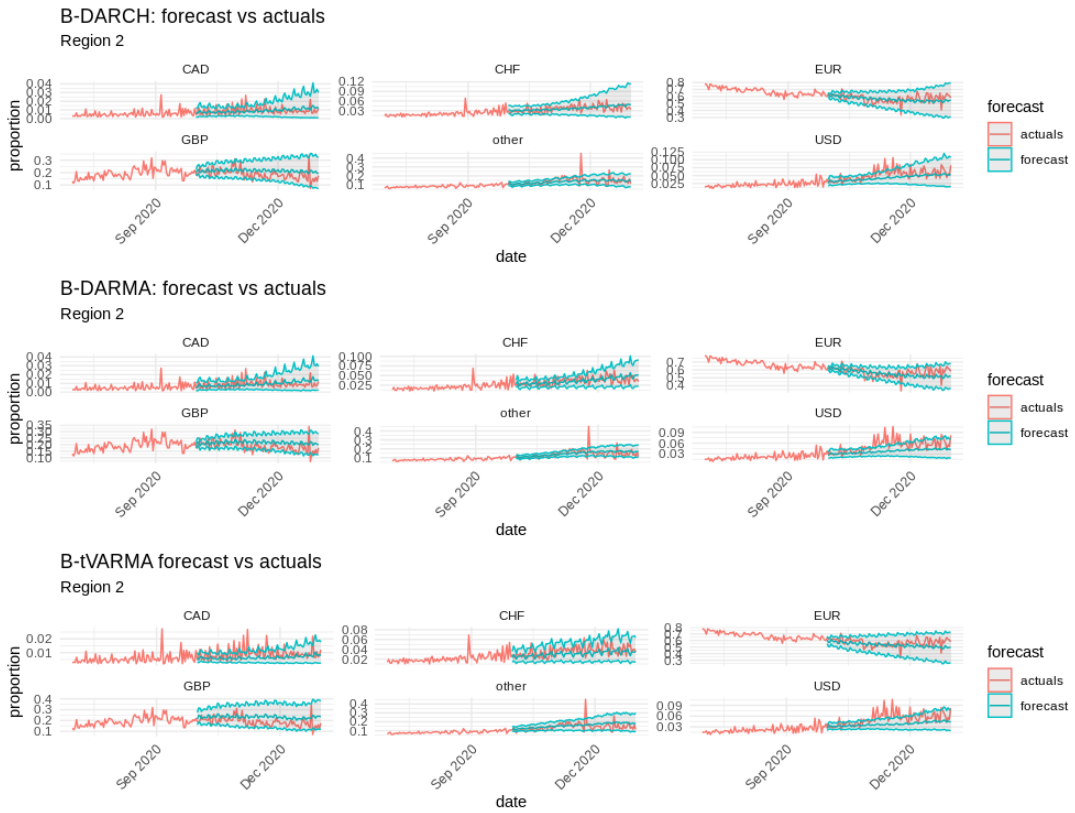

Figure 3.4: Airbnb data analysis - Region 2: 92-day forecasts (blue) with 95% confidence intervals, represented by the shaded regions around the forecast lines, for each of the six currencies, from October 1, 2020, to December 31, 2020, compared to the actual values (red) observed over the preceding six months. The models compared include B-DARCH, B-DARMA, and B-tVARMA.

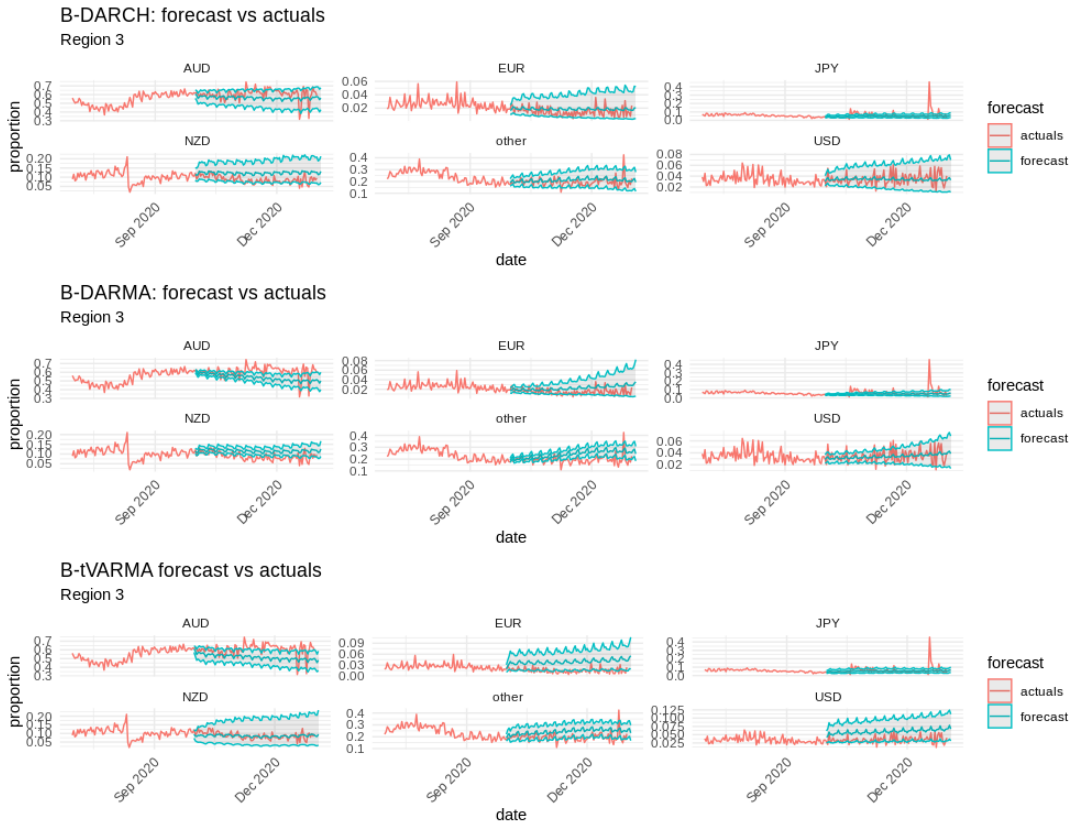

Figure 3.5: Airbnb data analysis - Region 3: 92-day forecasts (blue) with 95% confidence intervals, represented by the shaded regions around the forecast lines, for each of the six currencies, from October 1, 2020, to December 31, 2020, compared to the actual values (red) observed over the preceding six months. The models compared include B-DARCH, B-DARMA, and B-tVARMA.

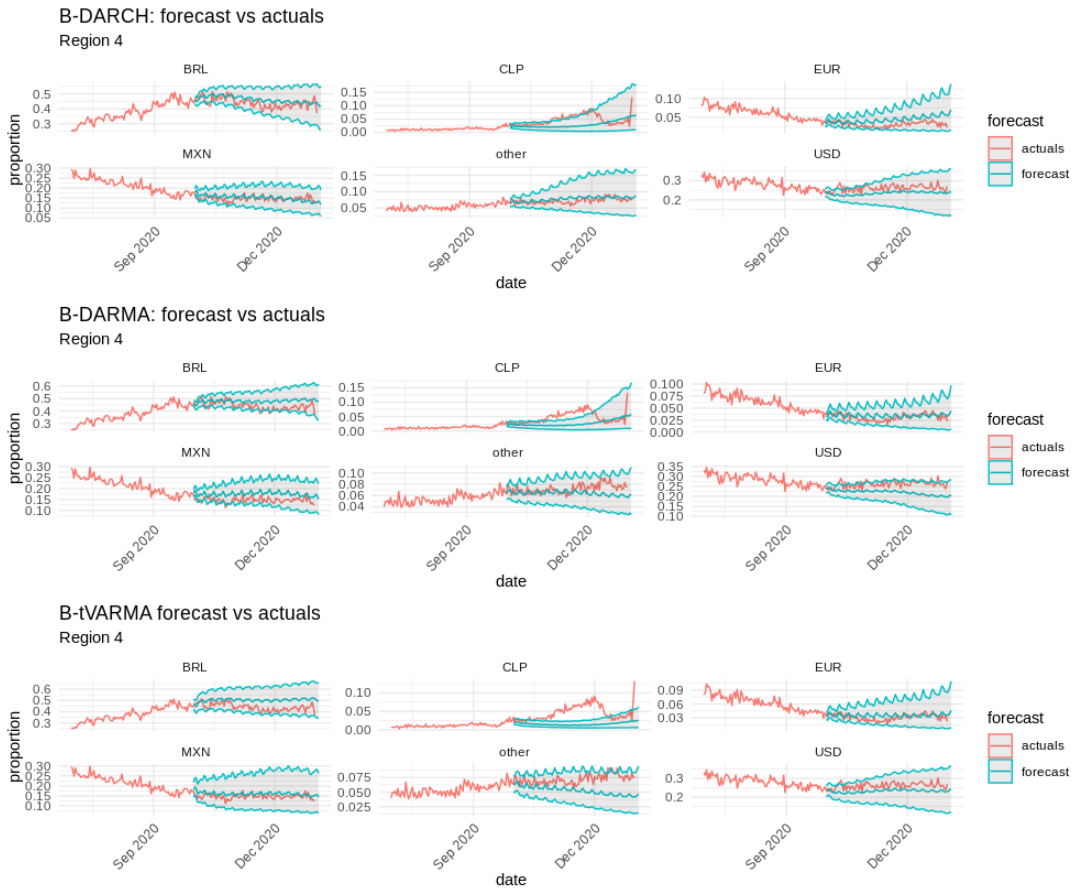

Figure 3.6: Airbnb data analysis - Region 4: 92-day forecasts (blue) with 95% confidence intervals, represented by the shaded regions around the forecast lines, for each of the six currencies, from October 1, 2020, to December 31, 2020, compared to the actual values (red) observed over the preceding six months. The models compared include B-DARCH, B-DARMA, and B-tVARMA.

Region	Currency	FMAE			FRSS		
		B-DARCH	**B-DARMA**	**B-tVARMA**	**B-DARCH**	**B-DARMA**	**B-tVARMA**
1	CAD	0.55	0.87	1.21	0.49	1.07	1.75
1	EUR	0.61	2.19	0.55	0.53	5.23	0.55
1	GBP	0.18	0.43	0.21	0.05	0.22	0.06
1	JPY	0.09	0.18	0.06	0.01	0.04	0.01
1	USD	1.08	3.13	2.52	1.91	11.70	7.65
1	other	0.33	0.85	0.66	0.17	0.85	0.60
1	**total**	2.84	7.65	5.20	3.15	19.11	10.62
1	**average**	0.47	1.27	0.87	0.53	3.19	1.77
2	CAD	0.35	0.38	0.37	0.22	0.24	0.27
2	CHF	0.83	0.89	0.96	0.97	1.14	1.36
2	EUR	4.51	4.85	5.51	29.70	34.50	42.20
2	GBP	4.06	4.13	5.32	23.00	24.00	35.30
2	USD	1.28	1.68	1.68	3.04	4.59	4.60
2	other	2.52	2.78	3.29	17.00	17.40	20.70
2	**total**	13.55	14.71	17.13	73.93	81.87	104.43
2	**average**	2.26	2.45	2.86	12.32	13.65	17.41
3	AUD	6.29	8.24	7.78	57.50	88.50	80.80
3	EUR	0.63	0.93	0.87	0.52	1.09	0.99
3	JPY	2.20	2.17	2.34	23.70	24.10	22.70
3	NZD	3.57	2.91	3.57	14.60	9.96	14.70
3	USD	0.82	0.79	1.02	1.00	0.98	1.52
3	other	2.69	5.09	3.06	13.30	34.80	15.30
3	**total**	16.20	20.12	18.64	110.62	159.43	136.01
3	**average**	2.70	3.35	3.11	18.44	26.57	22.67
4	BRL	2.84	4.36	6.38	10.80	23.60	46.20
4	CLP	2.32	2.35	3.09	7.76	8.45	13.00
4	EUR	1.29	0.58	0.64	2.08	0.51	0.60
4	MXN	1.27	2.85	1.54	2.36	8.71	2.98
4	USD	2.70	4.14	2.94	8.68	19.60	9.93
4	other	0.79	0.97	1.94	0.87	1.46	4.93
4	**total**	11.21	15.25	16.53	32.55	62.33	77.64
4	**average**	1.87	2.54	2.76	5.43	10.39	12.94

Table 3.2: Airbnb data analysis: Forecast Mean Absolute Error (FMAE) and Forecast Residual Sum of Squares (FRSS) for Regions 1 to 4 and top currencies. FMAE and FRSS values are multiplied by 10^2.

Residual Analysis

The partial autocorrelation function (PACF) of the sum of standardized squared residuals across all components on the test set was computed for each region and model (Supplementary Figures 3.S9–3.S12). All models exhibit a large peak at lag 1, reflecting immediate persistence in the residuals. However, the magnitude of this spike is noticeably smaller for the B-DARCH model. We consider a PACF value to be significant if it exceeds a threshold of 0.2; any autocorrelation values above this level indicate meaningful residual dependence.

Beyond lag 1, the B-DARCH model shows only one PACF value exceeding the 0.2 significance threshold, suggesting effective capture of short-term dependencies and minimal residual autocorrelation at higher lags. In contrast, both B-DARMA and B-tVARMA have PACF values surpassing the 0.2 threshold at multiple lags beyond lag 1, particularly between lags 2 and 10, indicating that residual dependencies persist across a broader range of lags in these models.

These suggest that the B-DARCH model achieves better residual independence and stability, with remaining residual structure potentially explained by unmodeled market-specific holidays, events, and policy implementations.

Parameter Estimates.

A complete discussion of the posterior densities for the DARCH parameters (auto-regressive α and moving average τ coefficients) in each region appears in the Supplementary Material (Section 3.5). Briefly, Regions 2 and 3 exhibit higher α values indicating more persistent volatility, while Region 1 shows moderate persistence ($0.25 < \alpha < 0.75$), and Region 4 displays a more complex multi-lag structure. These region-specific patterns underscore the flexibility of our B-DARMA-DARCH approach to accommodate various volatility regimes.

3.5 Discussion

We introduced the Bayesian Dirichlet Auto-Regressive Moving Average with Dirichlet Auto-Regressive Conditional Heteroskedasticity (B-DARMA-DARCH) model, a novel approach for modeling and forecasting dynamic compositional time series with time-varying volatility. By incorporating a DARCH component into the B-DARMA framework, the model effectively captures both mean dynamics and heteroskedasticity inherent in compositional data.

Simulation studies demonstrated the superior performance of the B-DARMA-DARCH model across six scenarios, including data generated from DARMA, DARCH, and tVARMA processes with random shocks simulating misreported observations and regime changes. The B-DARMA-DARCH model consistently achieved the lowest forecast error metrics and exhibited minimal residual autocorrelation, indicating its robustness in capturing complex volatility dynamics, especially in the presence of heteroskedasticity.

Applying the B-DARMA-DARCH model to Airbnb's currency fee proportions across different regions, we showed that it effectively captures temporal dynamics and volatility patterns in real-world data. The model generally outperformed the standard B-DARMA and B-tVARMA models in terms of forecast accuracy and residual diagnostics. Notably, the inclusion of the DARCH component allowed the model to account for significant disruptions, such as those caused by the COVID-19 pandemic, highlighting its robustness in the face of structural breaks and extreme events.

While the B-DARMA-DARCH model introduces additional complexity due to the dynamic modeling of the precision parameter, this complexity is justified by the improved model fit, empirical coverage rates, and forecasting accuracy in the presence of heteroskedasticity. The model offers a comprehensive framework that accounts for both mean dynamics and volatility patterns, which is particularly important in financial applications where volatility significantly influences decision-making and risk assessment.

However, the computational demands are higher, necessitating careful tuning of sampling parameters to ensure convergence and efficiency. Future research could focus on optimiz-

ing computational strategies or exploring approximate inference methods to mitigate these demands.

Several limitations warrant consideration. The current model models each region independently, even though some currencies are shared across multiple regions. This assumption of independence may not hold in practice, as economic events affecting a shared currency could simultaneously impact multiple regions. Incorporating hierarchical structures or accounting for inter-regional dependencies could enhance model performance by capturing these cross-regional correlations, albeit at the cost of increased computational complexity. Additionally, extending the model to handle zero-inflated compositional data or integrating external covariates that influence both the mean and volatility components could broaden its applicability

Code and Data Availability

The primary dataset used in our study, A Bayesian Dirichlet Auto-Regressive Conditional Heteroskedasticity Model for Forecasting Currency Shares, is not publicly available due to confidentiality constraints. However, the Stan code for the Bayesian Dirichlet Auto-Regressive Moving Average Dirichlet Auto-Regressive Conditional Heteroskedasticity Model (B-DARMA-DARCH) model is available for public access. It can be found at our GitHub repository: GitHub repository.

Supplementary Materials

Alternative Log-Ratio Transformations

Although we have used the additive log-ratio (ALR) link in equations (3.3), (3.4), and (3.6), other transformations to and from the simplex may be employed. A common choice is the *centered log-ratio* (clr), defined as

$$\mathrm{clr}(\mathbf{y}) \;=\; \Big(\ln\!\big(\tfrac{y_1}{g(\mathbf{y})}\big),\; \ln\!\big(\tfrac{y_2}{g(\mathbf{y})}\big),\; \ldots,\; \ln\!\big(\tfrac{y_J}{g(\mathbf{y})}\big) \Big),$$

where $g(\mathbf{y}) = (\, y_1\, y_2\, \cdots\, y_J)^{1/J}$ is the geometric mean of $\mathbf{y}$. The clr mapping places all J components in a transformed $\mathbb{R}^J$ space but introduces the constraint that their sum is zero.

Another option is to use the Isometric Log-Ratio (ilr) transformation. For each index j, the ilr-transformed coordinate can be expressed as

$$z_j = \sqrt{\frac{r_j}{r_j + 1}}\, \ln\!\left(\frac{g_j(\mathbf{y})}{\big(\prod_{i \in H_j} y_i\big)^{1/r_j}} \right),$$

where $g_j(\mathbf{y})$ is defined as the geometric mean of a chosen subset $S_j \subseteq \mathbf{y}$, and H_j is the complement of S_j in $\mathbf{y}$. The value r_j is the number of elements in H_j. The particular way in which S_j and H_j are selected depends on the specific application and how one wishes to interpret the ilr coordinates.

These mappings are mutually related by linear transformations (Egozcue et al., 2003) and each provides a bijection from the $(J-1)$-dimensional simplex onto a $(J-1)$-dimensional subset of $\mathbb{R}^J$ or $\mathbb{R}^{J-1}$. Therefore, if the same priors and model assumptions are consistently adapted to each log-ratio space, the resulting compositional models remain equivalent—differing only in how one embeds the data in a Euclidean domain.

Definition of the Partial Autocorrelation Function (PACF)

The PACF at lag k measures the partial correlation between the SSR at time t and the SSR at time $t-k$, controlling for the effects of intermediate lags 1 to $k-1$. It is defined as the last coefficient ϕ_{kk} in the autoregressive model of order k

$$\text{SSR}_t = \phi_{k1}\text{SSR}_{t-1} + \phi_{k2}\text{SSR}_{t-2} + \cdots + \phi_{k,k-1}\text{SSR}_{t-(k-1)} + \phi_{kk}\text{SSR}_{t-k} + \varepsilon_t \tag{3.7}$$

where ε_t is the white noise error term at time t. The coefficient ϕ_{kk} represents the PACF at lag k and can be calculated recursively using the Yule-Walker equations

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j}$$

Here, r_k is the autocorrelation at lag k, and $\phi_{k-1,j}$ are the PACF coefficients from the previous lag $k-1$.

By examining the PACF of the SSRs, we can identify the presence of residual autocorrelation at various lags, which indicates whether the model has adequately captured the temporal dependencies in the data.

Standardized Residuals

For the B-DARMA and B-DARCH models, which are based on the Dirichlet distribution as defined in Section 3.2, we compute standardized residuals using the standard deviations derived from the Dirichlet parameters.

At each time t, the residuals are calculated as $\text{residual}_{tj} = y_{tj} - \mu_{tj}, \quad \text{for } j = 1, \ldots, J$. The variance of each component y_{tj} under the Dirichlet distribution, $\text{Dirichlet}(\phi_t \boldsymbol{\mu}_t)$, is $\text{Var}[y_{tj}] = \frac{\mu_{tj}(1-\mu_{tj})}{\phi_t+1}$, with standard deviation $\sigma_{tj} = \sqrt{\frac{\mu_{tj}(1-\mu_{tj})}{\phi_t+1}}$.

The standardized residuals are then computed as $\text{standardized\_residual}_{tj} = \frac{y_{tj}-\mu_{tj}}{\sigma_{tj}}$ and

the sum of squared standardized residuals at time t is $\text{SSR}_t = \sum_{j=1}^{J} \left(\text{standardized\_residual}_{tj}\right)^2$.

For the B-tVARMA model, residuals are computed in the additive log-ratio (alr) transformed space, $\text{residual}_{tj} = \text{alr}(\mathbf{y}_{tj}) - \text{alr}(\hat{\mathbf{y}}_{tj})$, where $j = 1, \ldots, 4$. These residuals are standardized using the inverse of the Cholesky factor L_Σ of the estimated covariance matrix Σ,$\text{standardized\_residual}_t = L\Sigma^{-1} \times \text{residual}_t$. The sum of squared standardized residuals at time t is then calculated as $\text{SSR}_t = \sum_{j=1}^{J-1}(\text{standardized\_residual}_{tj})^2$.

By examining the PACF of $\{\text{SSR}\}_{t=1}^{T}$ we assess the presence of autocorrelation in the residuals. A well-specified model should produce residuals with minimal autocorrelation, indicating that temporal dependencies have been adequately captured.

Parameter Estimates

Supplementary Figure 3.S13 shows the posterior densities of the auto-regressive (α) and moving average (τ) coefficients for each region. In Region 1, the posterior of α is concentrated between 0.25 and 0.75, peaking around 0.45, indicating moderate volatility persistence; shocks have a noticeable but not prolonged effect on volatility. In Region 2, the α coefficient has posterior mean around 0.8, reflecting strong dependence of current volatility on past volatility, and the τ coefficient centers around -2.75, indicating that large deviations from expected currency proportions significantly increase / decrease future volatility—characteristic of volatility clustering. Region 3 shows α tightly concentrated between 0.85 and 0.90 and τ centered around -1.2 suggesting very strong volatility persistence. In contrast, Region 4 exhibits complex dynamics: α_1 is centered around 0.15, indicating small persistence from the first lag; α_2 is skewed positive and centered around 0.35, suggesting a stronger influence from the second lag; α_3 peaks slightly above 0.5, indicating moderate persistence; and the τ parameter is concentrated near 0.02, implying minimal influence of past innovations.

Overall, the higher α values in Regions 2 and 3 indicate that volatility in currency fee proportions is more persistent in these regions. The negative τ coefficients suggest that large deviations—such as those caused by abrupt travel restrictions, policy changes, and fluctuat-

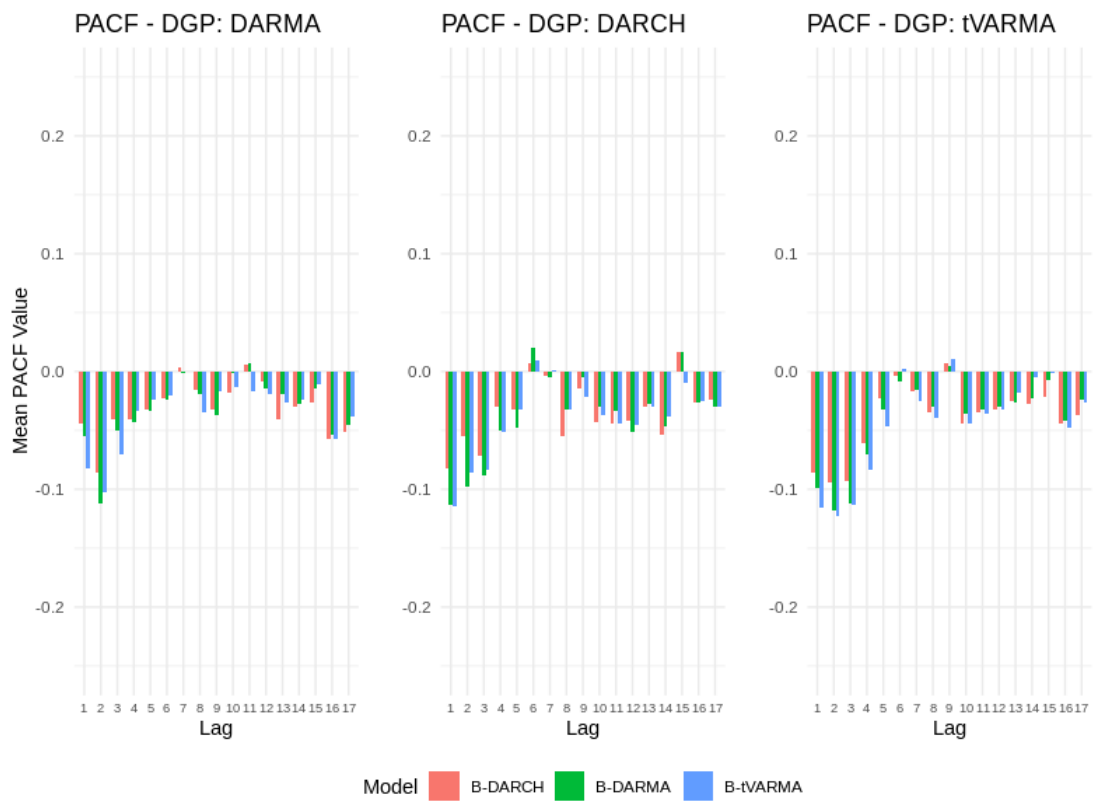

Figure 3.S1: Average PACF of the sum of squared standardized residual values across all 50 simulations for studies 1-3. B-DARMA-DARCH consistently exhibits lower PACF values beyond lag 1, indicating reduced residual autocorrelation and suggesting that its dynamic precision component better captures volatility shocks in these simulated compositional time series.

ing foreign exchange (FX) rates lead to increased future volatility. This implies that Regions 2 and 3 are more susceptible to external shocks affecting travel and currency markets, resulting in prolonged volatility in currency fee proportions. In contrast, Region 1, with moderate volatility persistence, tends to return to stability more rapidly after shocks, possibly due a more stable currency environment or a larger share of domestic travel. Region 4's dynamics, characterized by less emphasis on immediate past volatility and innovations, may be attributed to regional factors such as diversified travel sources, differing policy responses, or economic structures.

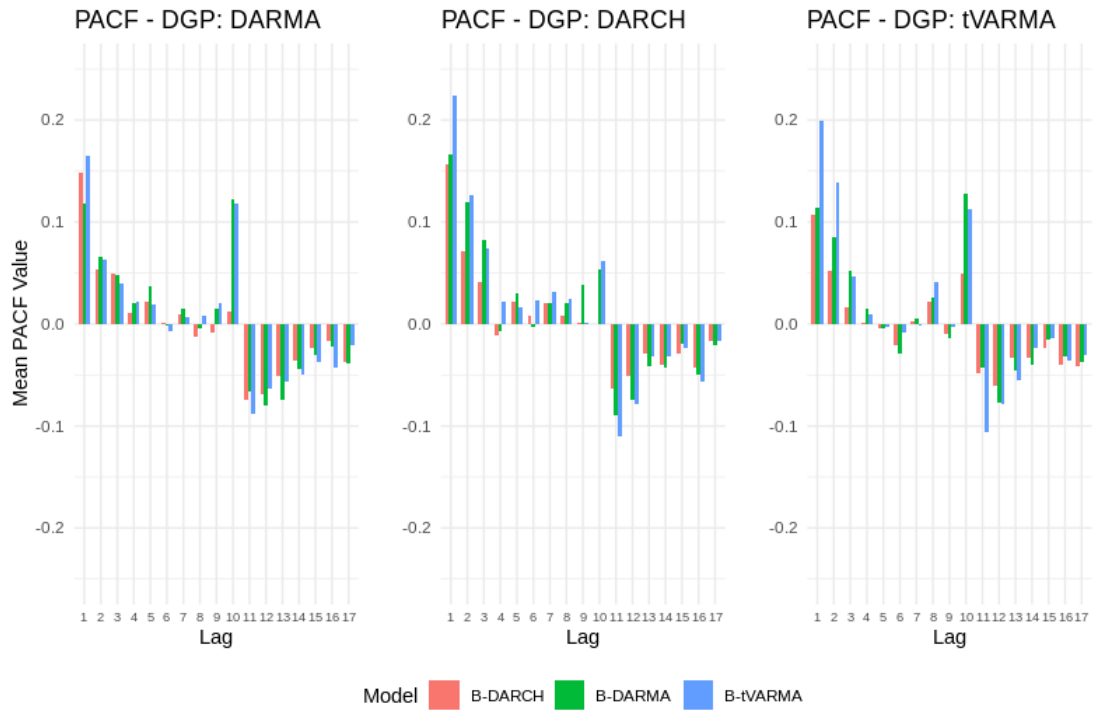

Figure 3.S2: Average PACF of the sum of squared standardized residual values across all 50 simulations for studies 4-6. The B-DARMA-DARCH model (red) again shows smaller autocorrelation at higher lags, whereas B-DARMA (green) and B-tVARMA (red) leave more persistent structures in the residuals. This result underlines the advantage of modeling time-varying volatility when faced with abrupt changes in compositional data.

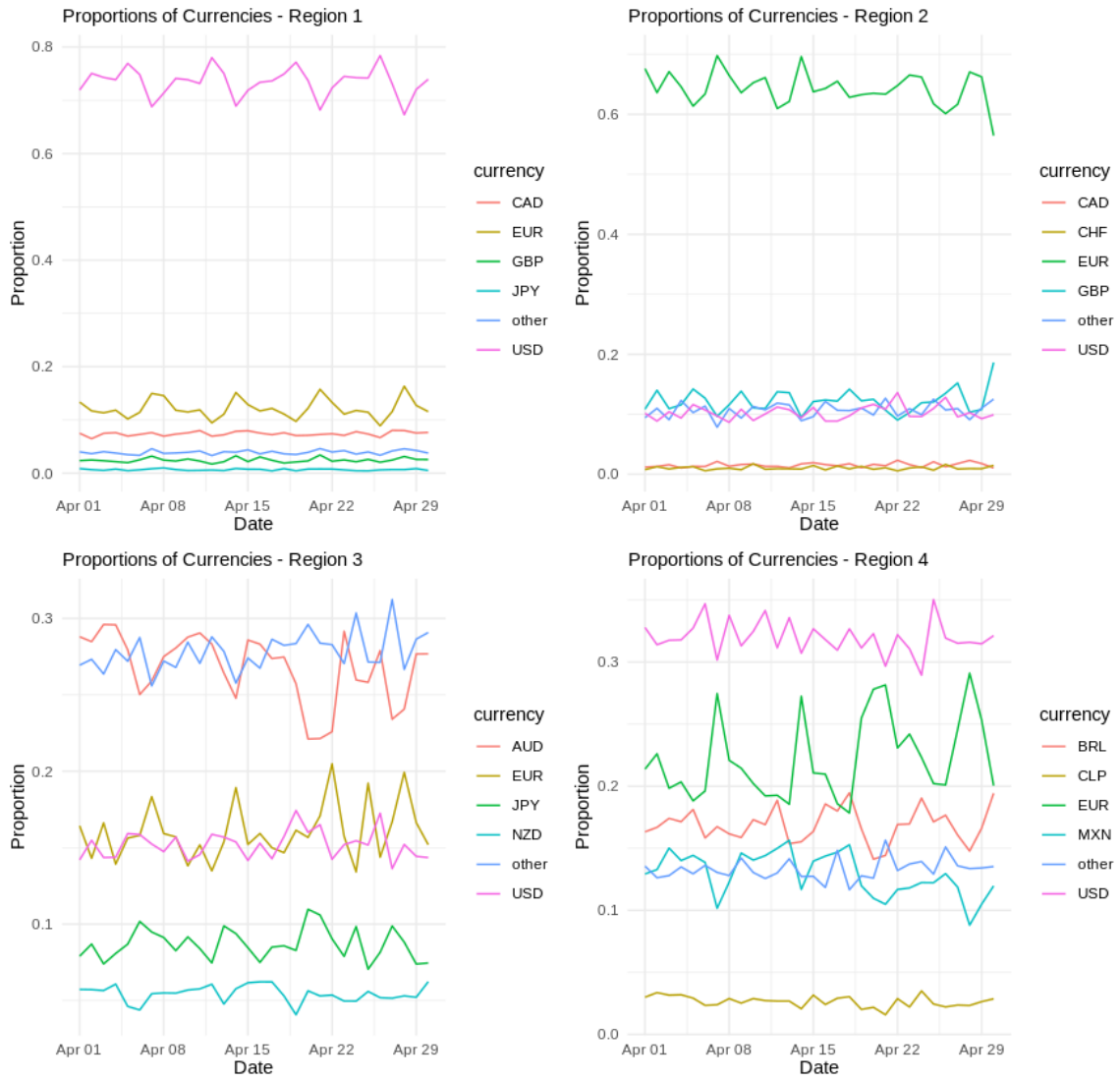

Figure 3.S3: Airbnb data analysis- Proportion of fees by currency for four regions: weekly seasonal behavior. AUD is the Australian dollar, BRL is the Brazillian Real, CAD is the Canadian Dollar,CHF is the Swiss Franc, CLP is the Chilean Peso,EUR is the European Euro, GBP is the Great British Pound, MXN is the Mexican Peso, NZD is the New Zealand Dollar, and USD is the US Dollar. Strong weekly cycles appear in many currencies, motivating the inclusion of weekly seasonal terms in our model's mean structure.

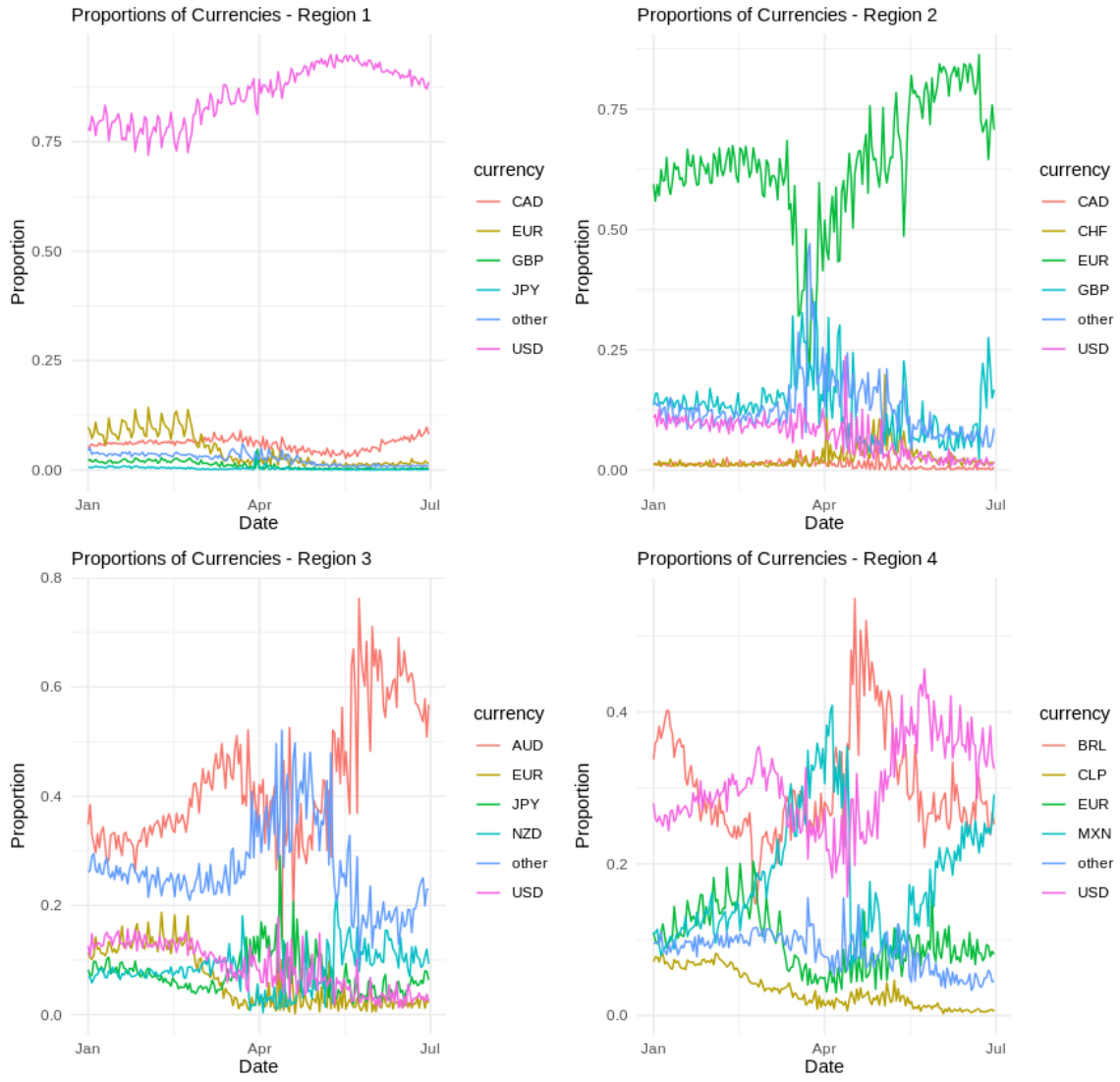

Figure 3.S4: Airbnb data analysis - proportion of fees by billing currency for four regions from Jan 1, 2020 to June 30, 2020. AUD is the Australian dollar, BRL is the Brazillian Real, CAD is the Canadian Dollar,CHF is the Swiss Franc, CLP is the Chilean Peso,EUR is the European Euro, GBP is the Great British Pound, MXN is the Mexican Peso, NZD is the New Zealand Dollar, and USD is the US Dollar. Distinct spikes and dips across multiple currencies (e.g., sharp increase in BRL or EUR) reflect significant disruptions.

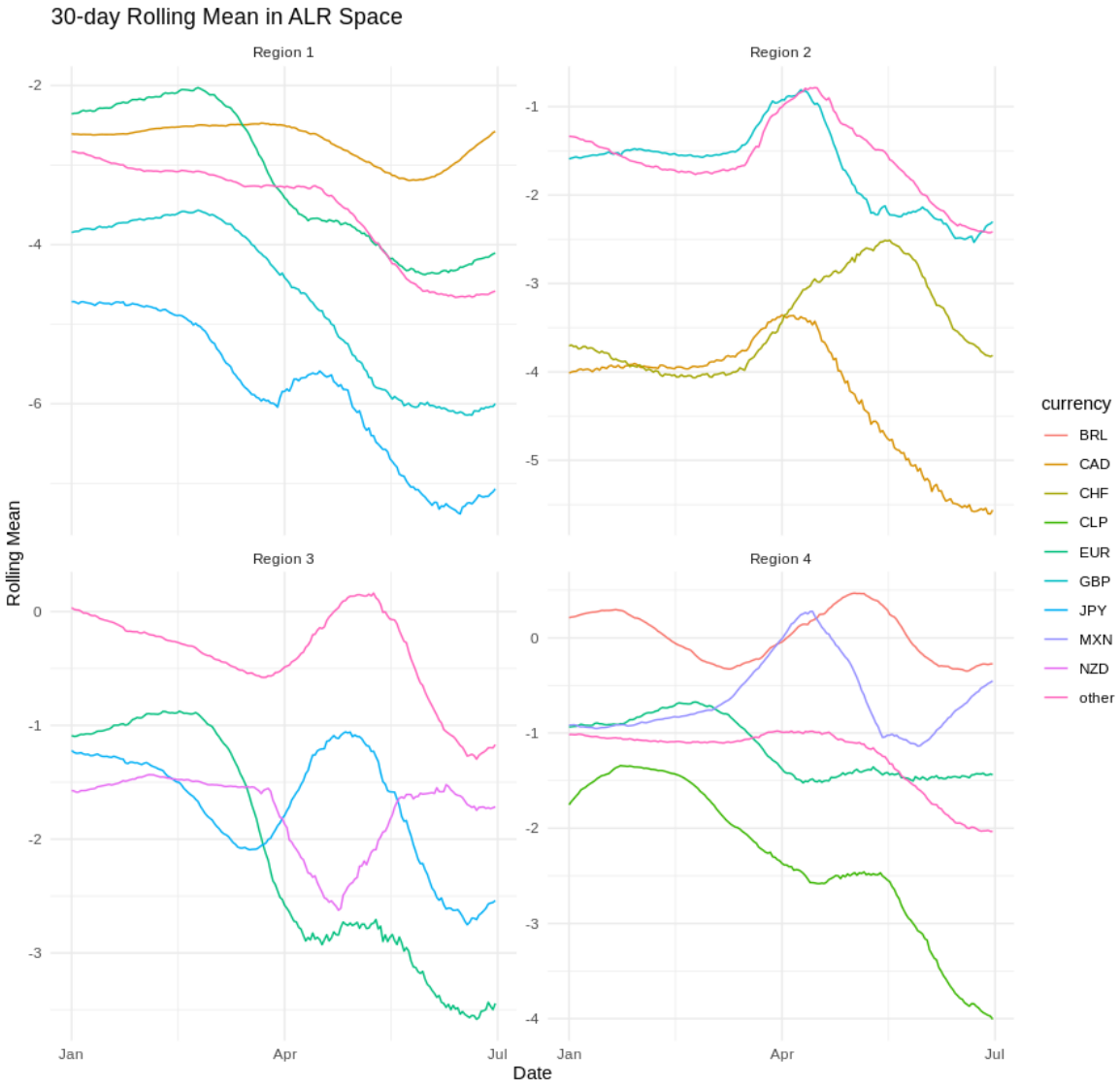

Figure 3.S5: Airbnb data analysis - proportion of fees by currency- 30-day rolling ALR means for four regions from Jan 1, 2020 to June 30, 2020. USD is the reference component.

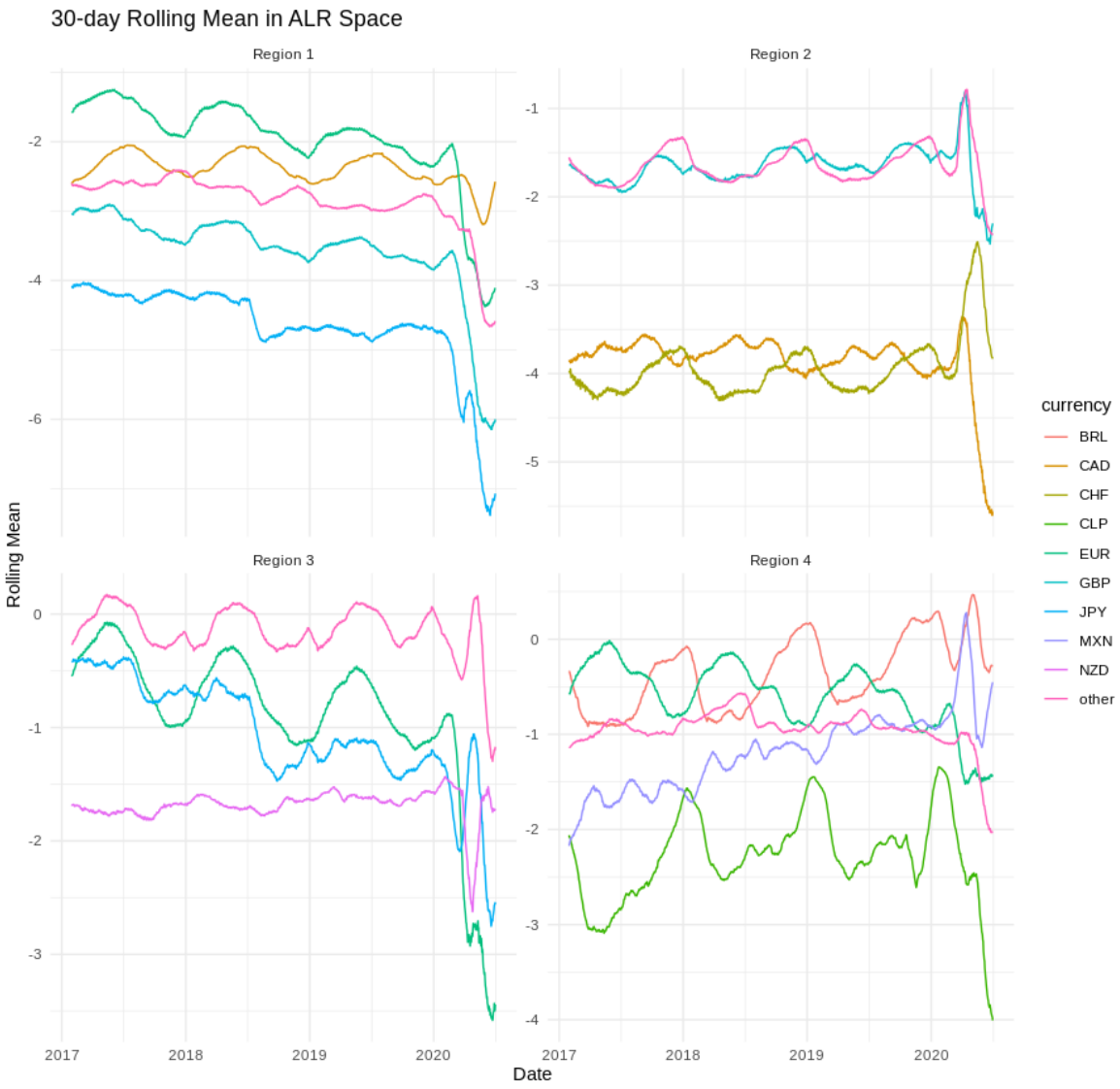

Figure 3.S6: Airbnb data analysis - proportion of fees by currency- 30-day rolling ALR means for four regions from Jan 1, 2017 to June 30, 2020. USD is the reference component. Apart from seasonal patterns, we see notable multi-year trends and cyclical behaviors (e.g., increasing share of EUR in Region 2, upward drift in AUD in Region 3), underlining the presence of both trend and seasonality in compositional dynamics.

Figure 3.S7: Airbnb data analysis - proportion of fees by currency- 30-day rolling ALR variance for four regions from Jan 1, 2017 to June 30, 2020. USD is the reference component.The data display periods of high variance that sometimes persist for weeks, demonstrating "volatility clustering." These episodes justify the GARCH-like specification in our B-DARMA-DARCH model, which allows the precision parameter ϕ_t to track spikes in uncertainty over time.

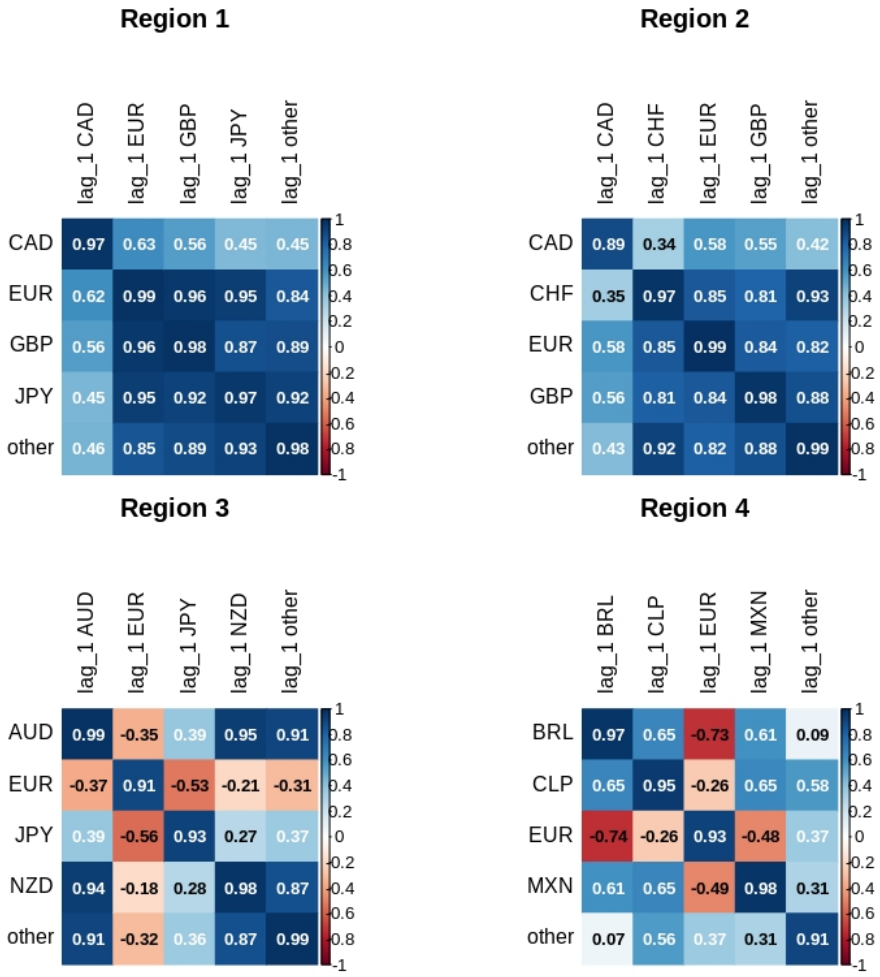

Figure 3.S8: Airbnb data analysis- correlation between ALR and lagged ALR values for each currency in the four regions. USD is the reference component. High positive autocorrelations in many currencies (e.g., EUR, GBP) highlight the strong temporal persistence. Negative cross-lag correlations for some pairs (e.g., EUR vs. BRL in Region 4) indicate that certain currencies move in opposite directions. Both effects reinforce the need for a multi-component time-series framework with auto- and cross-dependence.

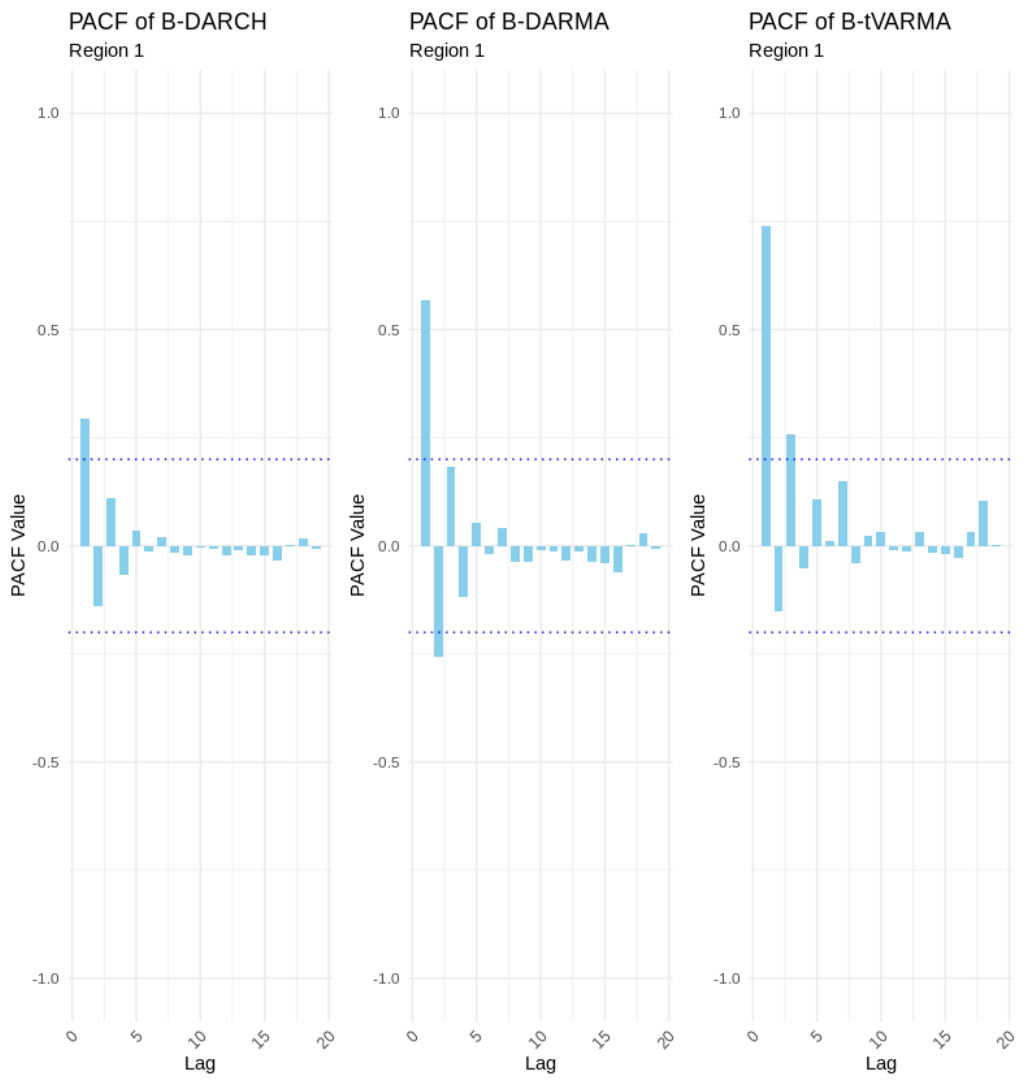

Figure 3.S9: Airbnb data analysis: Partial Autocorrelation Function (PACF) values for the sum of squared standardized residuals (SSR) on the test set from October 1, 2020, to December 31, 2020, for Region 1.

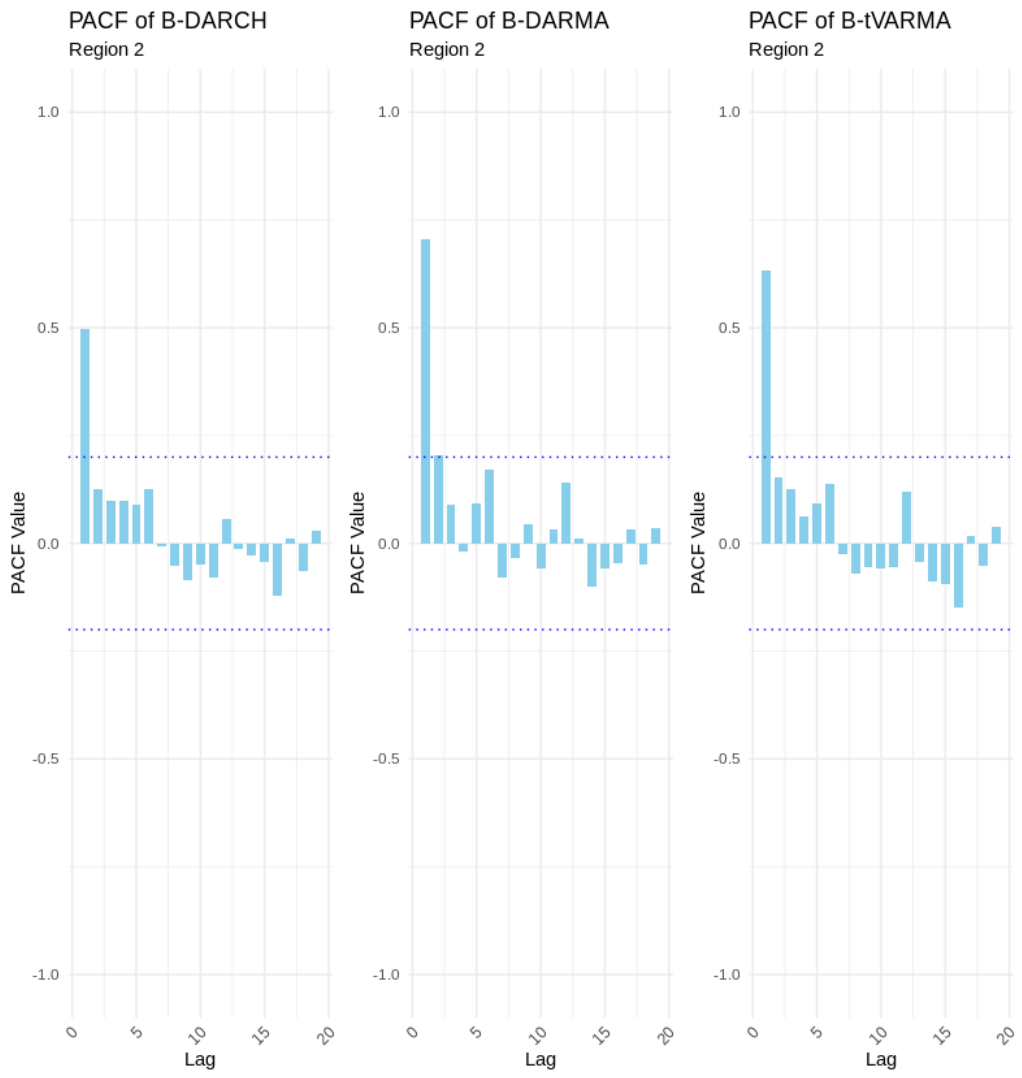

Figure 3.S10: Airbnb data analysis: Partial Autocorrelation Function (PACF) values for the sum of squared standardized residuals (SSR) on the test set from October 1, 2020, to December 31, 2020, for Region 2.

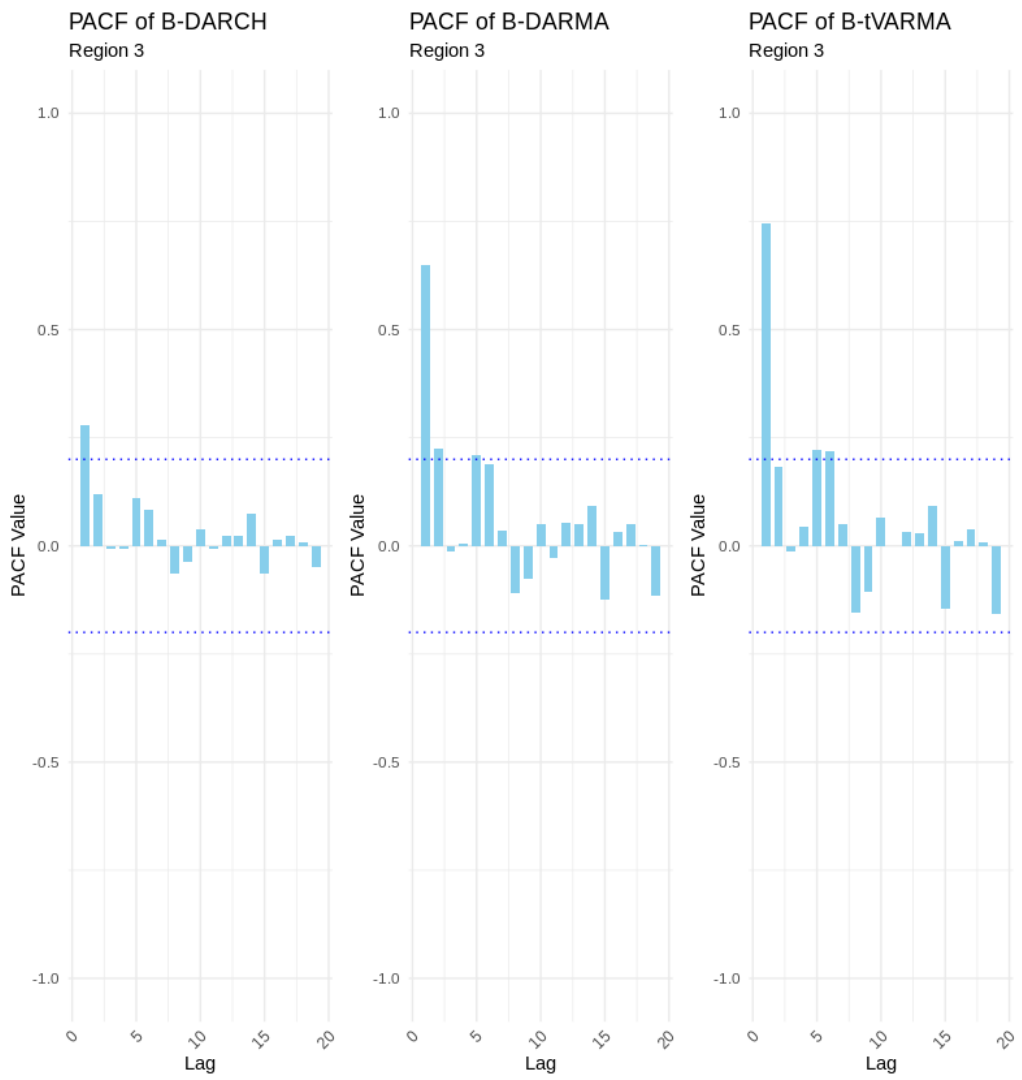

Figure 3.S11: Airbnb data analysis: Partial Autocorrelation Function (PACF) values for the sum of squared standardized residuals (SSR) on the test set from October 1, 2020, to December 31, 2020, for Region 3.

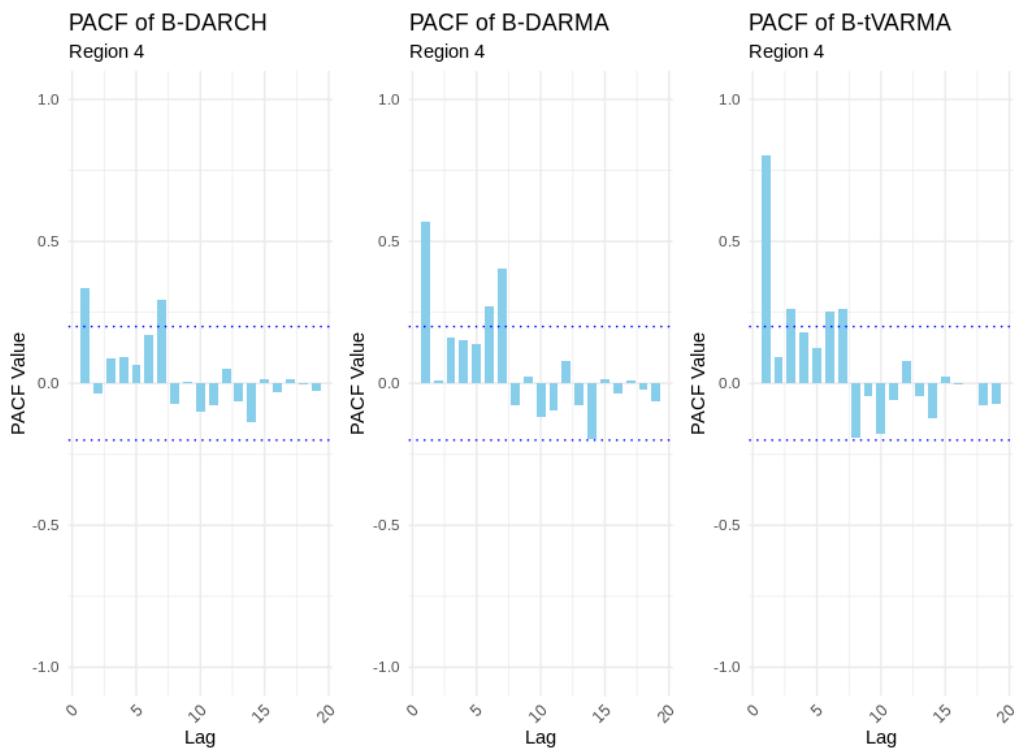

Figure 3.S12: Airbnb data analysis: Partial Autocorrelation Function (PACF) values for the sum of squared standardized residuals (SSR) on the test set from October 1, 2020, to December 31, 2020, for Region 4.

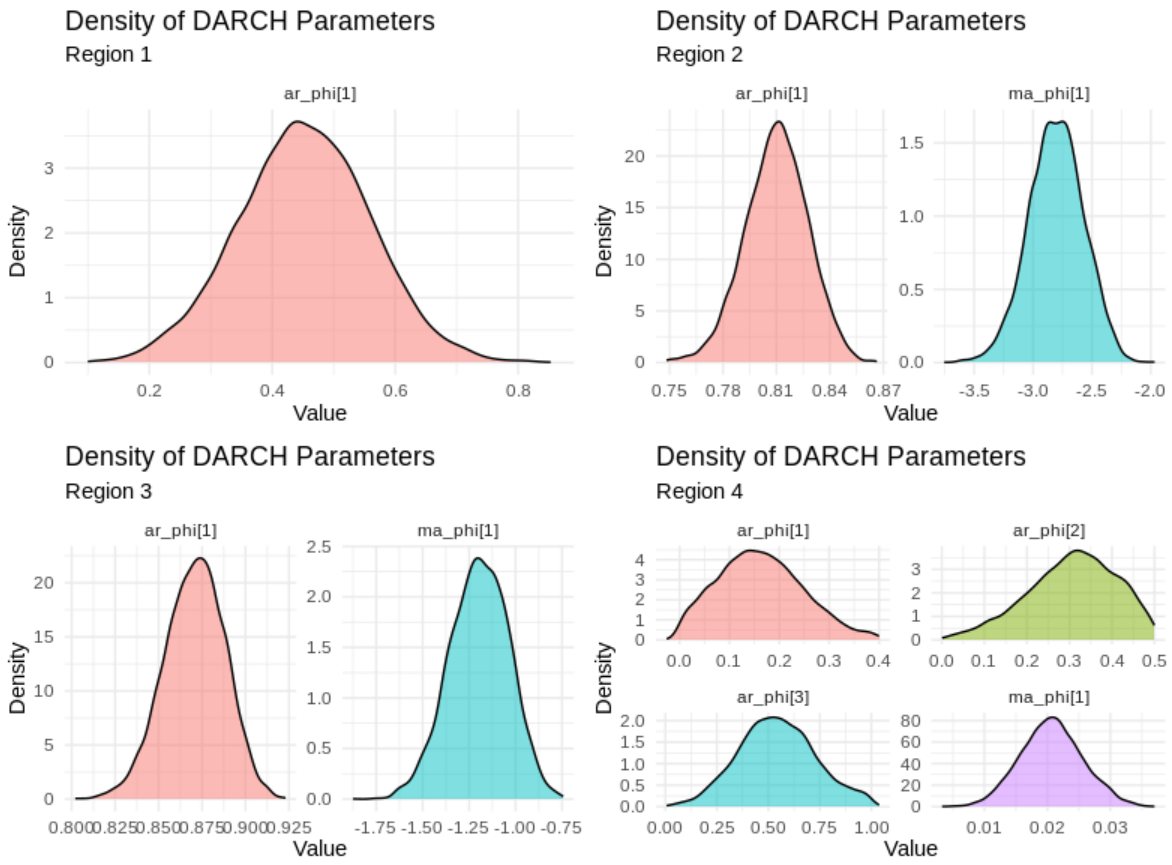

Figure 3.S13: Airbnb data analysis: Densities of DARCH model coefficients for all 4 regions. Region 1 was modeled with a DARCH(1,0), Regions 2-3 modeled with a DARCH(1,1), and Region 4 modeled with a DARCH(3,1). Each plot shows the estimated density of a specific α_l (ar_phi) or τ_k (ma_phi) parameter. Regions 2 and 3 exhibit high α indicating persistent volatility once spikes occur, whereas Region 4's coefficients suggest moderate or delayed response

Supplementary Tables

P	Q	Fourier	FMAE	FRSS
1	0	6	0.38	1.75
1	0	8	0.36	1.73
1	0	10	0.45	2.65
1	0	12	6.40	442.25
1	0	14	18.66	3662.84
1	0	16	12.42	1638.80
1	0	18	18.87	3809.00
0	1	8	5.81	404.83
1	1	8	1.05	14.60
1	2	8	0.98	15.02
1	3	8	1.03	16.12
2	1	8	0.40	3.96
2	2	8	6.57	716.43
2	3	8	2.09	60.38
3	1	8	8.17	984.45
3	2	8	8.91	1209.30
3	3	8	1.17	16.30

Region 1

P	Q	Fourier	FMAE	FRSS
1	0	6	2.15	49.95
1	0	8	2.10	48.46
1	0	10	1.76	38.08
1	0	12	6.46	343.58
1	0	14	13.31	1621.76
1	0	16	9.23	832.27
1	0	18	13.97	1892.17
0	1	10	4.74	193.91
1	1	10	1.27	20.18
1	2	10	1.28	23.08
1	3	10	2.81	48.12
2	1	10	3.08	263.17
2	2	10	2.43	91.08
2	3	10	2.38	73.18
3	1	10	5.84	344.45
3	2	10	2.67	106.77
3	3	10	5.88	380.70

Region 2

P	Q	Fourier	FMAE	FRSS
1	0	6	3.55	149.23
1	0	8	3.49	144.21
1	0	10	3.40	136.33
1	0	12	5.78	360.13
1	0	14	11.49	1281.57
1	0	16	9.37	862.47
1	0	18	12.80	1541.94
0	1	10	10.79	982.42
1	1	10	2.83	113.95
1	2	10	3.22	153.84
1	3	10	3.19	150.57
2	1	10	2.89	226.67
2	2	10	3.12	148.21
2	3	10	2.98	131.34
3	1	10	3.94	222.45
3	2	10	3.55	190.52
3	3	10	3.42	166.84

Region 3

P	Q	Fourier	FMAE	FRSS
1	0	6	4.72	216.17
1	0	8	4.64	212.57
1	0	10	4.92	236.31
1	0	12	8.00	490.97
1	0	14	12.05	1058.80
1	0	16	9.57	750.15
1	0	18	12.15	1148.46
0	1	8	9.60	693.31
1	1	8	2.66	76.77
1	2	8	1.63	31.03
1	3	8	1.86	38.07
2	1	8	1.43	47.14
2	2	8	1.50	26.37
2	3	8	1.49	24.93
3	1	8	1.33	20.92
3	2	8	1.16	15.53
3	3	8	1.90	40.07

Region 4

Table 3.S1: Airbnb data analysis: B-DARCH model performance across 4 regions – average Forecast Mean Absolute Error (FMAE) and total Forecast Residual Sum of Squares (FRSS) for the top 6 currencies for the validation set from July 1, 2020 to Sep 30, 2020. FMAE and FRSS values are multiplied by 10^2.

Table 3.S2: Coverage rates (%) of 95% credible intervals for each region and currency under B-DARCH, B-DARMA, and B-tVARMA. The mean coverage is computed across currencies in each region. The B-DARMA-DARCH coverage tends to be closest to the nominal rate across regions (e.g., 92% in Region 1, 95% in Region 4), whereas B-tVARMA often undercovers due to fixed variance assumptions, and B-DARMA coverage can be inconsistent when volatility spikes.

Region	**Currency**	**B-DARCH**	**B-DARMA**	**B-tVARMA**
1	CAD	0.85	0.88	0.72
	EUR	0.83	0.60	0.65
	GBP	0.96	0.69	0.56
	JPY	1.00	0.94	0.82
	USD	0.90	0.94	0.44
	other	0.96	0.84	0.57
	Mean	0.92	0.82	0.63
2	CAD	0.913	0.913	0.783
	CHF	0.967	0.891	0.924
	EUR	0.957	0.913	0.967
	GBP	0.870	0.772	0.924
	USD	0.859	0.641	0.663
	other	0.880	0.761	0.870
	Mean	0.91	0.82	0.86
3	AUD	0.81	0.29	0.43
	EUR	0.93	0.77	0.79
	JPY	0.71	0.73	0.78
	NZD	0.83	0.33	0.43
	USD	0.95	0.81	0.81
	other	0.97	0.39	0.89
	Mean	0.87	0.55	0.69
4	BRL	0.97	0.87	0.97
	CLP	0.69	0.54	0.30
	EUR	0.97	1.00	1.00
	MXN	0.98	0.90	0.99
	USD	0.89	0.74	0.96
	other	1.00	0.99	0.90
	Mean	0.91	0.84	0.85

Chapter 4

Sensitivity Analysis of Priors in the Bayesian Dirichlet Auto-Regressive Moving Average Model

4.1 Introduction

Compositional time series, in which observations are vectors of proportions constrained to sum to one, arise in a wide range of applications. For instance, market researchers track evolving shares of competing products (Boonen et al., 2019), ecologists monitor species composition over time, and sociologists or political scientists follow shifting demographic or budgetary profiles (Lipsmeyer et al., 2019). In each case, the data lie within a simplex, making the Dirichlet model a common starting point (Aitchison, 1986; Greenacre, 2021). However, when temporal dependence is also present—for example, today's composition influences tomorrow's—the Dirichlet framework must be extended to account for dynamics in a way that respects the simplex constraints.

Several approaches have emerged to handle dynamic compositional data, notably by coupling the Dirichlet with Auto-Regressive Moving Average (ARMA)-like structures (Zheng

and Chen, 2017), using a logistic-normal representation (Casarin et al., 2021), or adopting new families of innovation distributions (Makgai et al., 2021). Others have adapted the model to cope with zeros (Dong et al., 2025) or extreme heavy-tailed behaviors. In particular, the *Bayesian Dirichlet Auto-Regressive Moving Average* (B-DARMA) model (**?**) addresses key compositional modeling challenges by introducing Vector Auto-Regression (VAR) and Vector Moving Average (VMA) terms for multivariate compositional data. Under B-DARMA, each day's, or time point's, composition is Dirichlet-distributed with parameters that evolve with a VARMA process in the additive log-ratio space, capturing both compositional constraints and serial correlation.

Although B-DARMA provides a flexible foundation, practitioners must still specify priors for potentially high-dimensional parameter spaces, which can be prone to overfitting or omitted-lag bias. The growing literature on Bayesian shrinkage priors offers numerous solutions: from global–local shrinkage frameworks (Griffin and Brown, 2021) to hierarchical approaches (Bitto and Frühwirth-Schnatter, 2019a), along with classic spike-and-slab (Mitchell and Beauchamp, 1988; Follett and Yu, 2019) and Laplace priors (Park and Casella, 2008). These priors can encourage sparsity and suppress extraneous lags in over-parameterized models, a crucial feature for many real-world compositional applications in which the number of possible lags or covariates exceeds sample size. The horseshoe prior (Carvalho et al., 2010b; Polson and Sokolov, 2019) has shown particular promise in forecasting studies where only a minority of parameters matter and many are effectively zero. At the same time, hierarchical shrinkage (Bitto and Frühwirth-Schnatter, 2019a) facilitates partial pooling across related coefficient blocks, an appealing property when working with multi-sector or multi-species compositions.

In this paper, we systematically investigate how five different prior families—informative normal, horseshoe, Laplace, spike-and-slab, and hierarchical shrinkage—affect parameter recovery and predictive accuracy in the B-DARMA model. We compare their performance across three main scenarios using simulated data: (i) *correct specification*, where the model

order matches the true process, (ii) *overfitting*, where extraneous VAR/VMA terms inflate model dimensionality, and (iii) *underfitting*, where key VAR/VMA terms are missing altogether. Our findings confirm that shrinkage priors—especially horseshoe and hierarchical variants—can successfully rein in overfitting, providing more robust parameter estimates and improved forecasts. Conversely, no amount of shrinkage compensates for omitted terms, highlighting the need for careful model specification (Makgai et al., 2021).

To demonstrate practical impact, we also apply B-DARMA to daily S&P 500 sector trading data, a large-scale compositional time series characterized by multiple seasonalities and long-lag behavior. Consistent with the simulation insights, we find that more aggressive shrinkage priors significantly reduce spurious complexity and improve forecast accuracy, especially for volatile sectors. These outcomes reinforce that while B-DARMA provides a natural scaffolding for compositional dependence, judicious prior selection and meaningful lag choices are pivotal.

In the remainder of the paper, we first review the B-DARMA model (Section 2) and outline how our five prior families (informative, horseshoe, Laplace, spike-and-slab, and hierarchical) encode distinct shrinkage behaviors. We then present the design and results of three simulation studies (Sections 3-4) before turning to the empirical S&P 500 application (Section 5). We conclude with recommendations for practitioners modeling compositional time series with complex temporal dynamics and large parameter counts.

4.2 Background

4.2.1 Compositional Data and the Dirichlet Distribution

Compositional data consist of vectors of proportions, each strictly between zero and one and summing to unity (Aitchison, 1986). Formally, let

$$\mathbf{y}_t \;=\; \big(y_{t1}, y_{t2}, \ldots, y_{tJ}\big)', \quad t = 1, \ldots, T,$$

where each $y_{tj} > 0$ and $\sum_{j=1}^{J} y_{tj} = 1$. The vector $\mathbf{y}_t$ resides in the $(J-1)$-dimensional simplex.

A natural choice for modeling such compositional vectors is the *Dirichlet* distribution. In its basic form, a Dirichlet random vector $\mathbf{x} = (x_1, \ldots, x_K)$ is parameterized by a concentration vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ with $\alpha_k > 0$. The probability density function is

$$p(\mathbf{x} \mid \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} x_k^{\alpha_k - 1},$$

where

$$B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)},$$

and $\Gamma(\cdot)$ is the Gamma function. This parameterization captures both the support of compositional data (the simplex) and the potential correlation structure among components.

4.2.2 B-DARMA Model

To capture temporal dependence in compositional data, the Bayesian Dirichlet Auto-Regressive Moving Average (B-DARMA) model (**?**) augments the Dirichlet framework with VAR and VMA dynamics. Specifically, for each time $t = 1, \ldots, T$, let $\mathbf{y}_t$ be the observed composition. We assume

$$\mathbf{y}_t \;\Big|\; \boldsymbol{\mu}_t,\ \phi_t \sim \text{Dirichlet}\big(\phi_t\, \boldsymbol{\mu}_t\big), \tag{4.1}$$

with density $f(\mathbf{y}_t|\boldsymbol{\mu}_t, \phi_t) \propto \prod_{j=1}^{J} y_{tj}^{\phi_t \mu_{tj} - 1}$, where $\boldsymbol{\mu}_t = (\mu_{t1}, \ldots, \mu_{tJ})'$ is the mean composition, and $\phi_t > 0$ is a precision parameter. Both $\boldsymbol{\mu}_t$ and ϕ_t may vary with time.

ALR link. Because each $\boldsymbol{\mu}_t$ lies in the simplex, we map it to an unconstrained $(J-1)$-dimensional vector via the *additive log-ratio* transform

$$\operatorname{alr}(\boldsymbol{\mu}_t) \;=\; \Big(\ln \tfrac{\mu_{t1}}{\mu_{tJ}}, \ldots, \ln \tfrac{\mu_{t,J-1}}{\mu_{tJ}}\Big).$$

We denote

$$\boldsymbol{\eta}_t \;=\; \operatorname{alr}(\boldsymbol{\mu}_t) \;\in\; \mathbb{R}^{J-1}.$$

VARMA structure. To incorporate serial dependence, we assume $\boldsymbol{\eta}_t$ follows a vector VARMA process in the transformed space

$$\boldsymbol{\eta}_t \;=\; \sum_{p=1}^{P} \mathbf{A}_p \Big[\operatorname{alr}(\mathbf{y}_{t-p}) - \mathbf{X}_{t-p}\,\boldsymbol{\beta}\Big] \;+\; \sum_{q=1}^{Q} \mathbf{B}_q \Big[\operatorname{alr}(\mathbf{y}_{t-q}) - \boldsymbol{\eta}_{t-q}\Big] \;+\; \mathbf{X}_t\,\boldsymbol{\beta}, \tag{4.2}$$

for $t = m+1, \ldots, T$, where $m = \max(P, Q)$. In this notation

- $\mathbf{A}_p$ and $\mathbf{B}_q$ are each $(J-1) \times (J-1)$ coefficient matrices.
- $\mathbf{X}_t$ is a known $(J-1) \times r_\beta$ covariate matrix including any intercepts, trends, or seasonality.
- $\boldsymbol{\beta} \in \mathbb{R}^{r_\beta}$ is a vector of regression coefficients shared across the $(J-1)$ components.

Precision parameter. The Dirichlet precision ϕ_t can also evolve over time. For an r_γ-vector of covariates $\mathbf{z}_t$, we set

$$\phi_t \;=\; \exp\Big(\mathbf{z}_t\,\boldsymbol{\gamma}\Big), \tag{4.3}$$

with $\boldsymbol{\gamma} \in \mathbb{R}^{r_\gamma}$. In the absence of covariates, we simply have $\log \phi_t = \gamma$ for all t, so ϕ_t becomes a constant.

Parameter vector. We gather all unknown parameters into a vector of length C,

$$\boldsymbol{\theta} \;=\; \big(\, \mathbf{A}_{prs},\; \mathbf{B}_{qrs},\; \boldsymbol{\beta}',\; \boldsymbol{\gamma}' \big)',$$

where $p = 1, \ldots, P$, $q = 1, \ldots, Q$, and $r, s = 1, \ldots, J-1$, and θ_j is the j-th element of $\boldsymbol{\theta}$. The total number of free parameters is thus

$$C \;=\; (P+Q)\,(J-1)^2 \;+\; r_\beta \;+\; r_\gamma.$$

Bayesian inference begins by positing a prior distribution, $p(\boldsymbol{\theta})$, over the model parameters $\boldsymbol{\theta}$. Bayes' theorem updates this prior to form the posterior

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \;=\; \frac{p(\boldsymbol{\theta})\, p(\mathbf{y}_{(m+1):T} \mid \boldsymbol{\theta}, \mathbf{y}_{1:m})}{p(\mathbf{y}_{(m+1):T} \mid \mathbf{y}_{1:m})},$$

where

$$p(\mathbf{y}_{(m+1):T} \mid \boldsymbol{\theta}, \mathbf{y}_{1:m}) \;=\; \prod_{t=m+1}^{T} p\big(\mathbf{y}_t \;\big|\; \boldsymbol{\theta},\, \mathbf{y}_{(t-m)!:,(t-1)}\big),$$

and $p(\mathbf{y}_{(m+1):T} \mid \mathbf{y}_{1:m})$ is the normalizing constant obtained by integrating over $\boldsymbol{\theta}$. Each $p(\mathbf{y}_t \mid \boldsymbol{\theta}, \mathbf{y}_{(t-m):(t-1)})$ follows the Dirichlet likelihood (4.1).

Next, to generate predictions for the subsequent S time points, $\mathbf{y}_{(T+1):(T+S)}$, we construct the joint predictive distribution

$$p\big(\mathbf{y}_{(T+1):(T+S)} \mid \mathbf{y}_{1:T}\big) \;=\; \int p\big(\mathbf{y}_{(T+1):(T+S)} \mid \boldsymbol{\theta}\big)\, p\big(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}\big)\, d\boldsymbol{\theta}.$$

In practice, analysts often summarize this distribution at future time points $t \in (T+1) : (T+S)$ by reporting measures such as the posterior mean or median.

4.2.3 Bayesian Shrinkage Priors for B-DARMA Coefficients

In a fully Bayesian approach, all unknown parameters in the B-DARMA model — including the VARMA coefficients in $\mathbf{A}_p$ and $\mathbf{B}_q$, the regression vector $\boldsymbol{\beta}$, and the precision-related parameters $\boldsymbol{\gamma}$ — require prior distributions. Different shrinkage priors can produce significantly different outcomes, particularly in high-dimensional or sparse settings where many coefficients may be small. This section discusses five popular priors (normal, horseshoe, Laplace, spike-and-slab, and hierarchical), highlighting how each encodes shrinkage or sparsity assumptions.

An *informative normal prior* serves as a straightforward baseline. We model each coefficient θ_j by $\theta_j \sim \mathcal{N}(a, b^2)$. The mean a often defaults to zero unless prior knowledge indicates a different center. The variance b^2 determines shrinkage strength; smaller b^2 yields tighter concentration around a. In B-DARMA applications, it may be sensible to set $b = 1$ for the VARMA parameters if they are believed to be small on average, whereas for covariates $\boldsymbol{\beta}$, a smaller prior variance such as $b^2 = 0.01$ can reflect stronger beliefs that regression effects are modest.

The *horseshoe prior* (Carvalho et al., 2010b) is well-suited for sparse problems. We model each coefficient ν as $\nu \mid \tau, \lambda_\nu \sim \mathcal{N}(0, \tau^2\lambda_\nu^2)$ with a global scale $\tau \sim \text{Cauchy}_+(0, 1)$ and local scales $\lambda_\nu \sim \text{Cauchy}_+(0, 1)$. Large local scales allow some coefficients to remain sizable, whereas most are heavily shrunk. This can be beneficial if the B-DARMA model includes many possible lags or covariates, only a small subset of which are expected to matter for compositional forecasting.

A *Laplace (double-exponential) prior* (Park and Casella, 2008) employs the density $p(\nu \mid b) = \frac{1}{2b}\exp(-|\nu|/b)$. This enforces an ℓ_1-type penalty that can drive many coefficients close to zero while still allowing moderate signals to persist. The scale b can be chosen a priori or assigned its own hyperprior, such as a half-Cauchy, so that the data adaptively determine the shrinkage level. In a B-DARMA context, choosing a smaller b for high-dimensional VARMA terms can prevent spurious estimates from inflating the parameter space.

A *spike-and-slab prior* (Mitchell and Beauchamp, 1988) can introduce explicit sparsity by placing a point mass at zero. Let $\tau_j \sim \text{Beta}(1,1)$ be the mixing parameter for the jth coefficient. Then each coefficient θ_j follows the mixture

$$\theta_j \;\sim\; \tau_j \,\mathcal{N}(0,1) \;+\; (1-\tau_j)\,\delta_0,$$

where δ_0 is the Dirac measure at zero. Coefficients drawn from the spike component remain exactly zero, effectively excluding them from the model, while coefficients from the slab remain freely estimated. This setup allows the B-DARMA specification to adapt by discarding irrelevant lags or covariates.

A *hierarchical shrinkage prior* (Polson and Scott, 2012) can encourage partial pooling across groups of coefficients. We model each coefficient ν by $\nu \mid \sigma \sim \mathcal{N}(0, \sigma^2)$ and then place a half-Cauchy prior on σ. In B-DARMA, one could assign separate group scales to the AR, MA, and regression blocks, thereby allowing the model to learn an appropriate overall variability for each group of parameters.

These priors differ in how strongly they push coefficients toward zero and whether they favor a few large coefficients or moderate shrinkage for all. The normal prior provides a baseline continuous shrinkage, the horseshoe excels when only a minority of parameters are truly large, the Laplace induces an ℓ_1-type penalty that can zero out many coefficients, spike-and-slab explicitly discards some parameters, and the hierarchical alternative enables group-level learning of shrinkage scales. The next sections illustrate how these choices affect both parameter recovery and compositional forecasting in various simulation and real-data scenarios.

4.2.4 Posterior computation

All B-DARMA models are fit with STAN (Stan Development Team, 2022) in R using Hamiltonian Monte Carlo. We run 4 chains, each with 500 warm-up and 750 sampling iterations,

yielding 3,000 posterior draws. The sampler uses `adapt_delta` = 0.85, `max_treedepth` = 11, and random initial values drawn uniformly from $[-.25, .25]$.

4.2.5 Stationarity and structural considerations

Open theoretical gap. B-DARMA inherits the difficulty noted by Zheng and Chen (2017): after the additive-log-ratio (ALR) link the innovation is *not* a martingale-difference sequence (MDS). Consequently, classical stationarity results for VARMA (p,q) processes do not transfer directly, and a full set of strict- or weak-stationarity conditions remains an open problem (Katz et al., 2024, IJF, §2.2).

Why Bayesian inference remains valid. Bayesian estimation proceeds via the joint posterior $p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$, which does not require the likelihood-error sequence to be an MDS (Gelman et al., 2020). We therefore regard B–DARMA as a flexible likelihood-based filter for compositional time series rather than assume the fitted parameters represent a strictly stationary data-generating process.

4.3 Simulation Studies

We conduct three simulation studies to investigate how different priors affect parameter inference and forecasting in a B-DARMA model. All studies use the same sparse DARMA(2,1) data-generating process (DGP) but vary the fitted model to be correctly specified, deliberately overfitted, or underfitted. We first describe the DGP and the priors considered, then outline the study designs and performance metrics.

4.3.1 Data-Generating Process

We simulate a six-dimensional compositional time series $\mathbf{y}_t$, each component constrained to sum to one. The true process is DARMA(2,1) with fixed precision $\phi = 500$, $\boldsymbol{\beta} =$

$(0.1, -0.05, 0.03, -0.02, 0.04)'$, and $\mathbf{X}_t = I_5$, a 5×5 identity matrix. We set our VAR and VMA matrices to,

$$\mathbf{A}_1 = \begin{bmatrix} 0.80 & 0.05 & -0.04 & -0.05 & -0.05 \\ -0.01 & 0.70 & -0.03 & 0.02 & -0.01 \\ 0.02 & 0.00 & 0.90 & 0.02 & 0.04 \\ -0.03 & -0.07 & -0.02 & 0.85 & -0.01 \\ 0.04 & -0.02 & 0.01 & -0.01 & 0.75 \end{bmatrix}$$

$$\mathbf{A}_2 = \begin{bmatrix} -0.30 & 0.03 & 0.02 & 0.05 & -0.04 \\ 0.02 & -0.20 & -0.01 & -0.02 & 0.01 \\ -0.01 & 0.05 & -0.25 & -0.01 & 0.01 \\ -0.01 & 0.04 & 0.01 & -0.15 & 0.00 \\ 0.06 & 0.00 & -0.11 & -0.02 & -0.20 \end{bmatrix},$$

$$\mathbf{B}_1 = \begin{bmatrix} 0.50 & -0.02 & 0.03 & 0.00 & 0.03 \\ 0.05 & 0.40 & 0.03 & -0.01 & 0.02 \\ 0.02 & 0.01 & 0.45 & -0.02 & 0.13 \\ -0.01 & 0.10 & 0.05 & 0.35 & 0.01 \\ -0.01 & 0.04 & -0.11 & 0.10 & 0.40 \end{bmatrix}.$$

We initialize $\mathbf{y}_1$ and $\mathbf{y}_2$ from Dirichlet($\mathbf{1}$), simulate $T = 100$ observations, and replicate this procedure 50 times. We focus on posterior inference for $\mathbf{A}_p$, $\mathbf{B}_q$, and $\boldsymbol{\beta}$, as well as forecasting accuracy.

4.3.2 Prior Distributions and Hyperparameters

We specify five candidate priors for the coefficient vectors: Normal (mean 0, variance 1), Horseshoe (global and local Cauchy$_+(0, 1)$ scales), Laplace (scale $b = 1$), spike-and-slab (with local mixing parameters $\tau_j \sim \text{Beta}(1, 1)$), and hierarchical Normal (with a half-Cauchy prior

on the group-level scale). Although these priors remain the same across simulations, they are applied to different sets of coefficients: $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1$, and $\boldsymbol{\beta}$ in Simulation 1; $\mathbf{A}_1$–$\mathbf{A}_4, \mathbf{B}_1$–$\mathbf{B}_2$, and $\boldsymbol{\beta}$ in Simulation 2; and $\mathbf{A}_1$ and $\boldsymbol{\beta}$ in Simulation 3. For the Dirichlet precision parameter, we use $\gamma_\phi \sim \mathcal{N}(7, 1.5)$. The Normal prior on $\boldsymbol{\beta}$ is $\mathcal{N}(0, 0.1)$, and the Laplace and hierarchical Normal versions adopt smaller scales to impose stronger shrinkage on intercepts.

4.3.3 Study Designs

We fit three B-DARMA specifications to the same DARMA(2,1) data

- **Study 1 (Correct Specification).** B-DARMA(2,1) matches the true DGP.
- **Study 2 (Overfitting).** B-DARMA(4,2) includes extraneous higher-order VAR and VMA terms.
- **Study 3 (Underfitting).** B-DARMA(1,0) omits the second VAR lag and the MA(1) term.

Each configuration is paired with each of the five priors, yielding 15 total fitted models. We repeat the simulation for 50 synthetic datasets of length T.

4.3.4 Evaluation Metrics

We assess both parameter recovery and forecasting performance. Let θ_{true} be a true parameter and $\hat{\theta}^{(s)}$, the posterior mean from simulation s. We compute

$$\text{Bias}_j = \frac{1}{50}\sum_{s=1}^{50}\big(\hat{\theta}_j^{(s)} - \theta_{j,\text{true}}\big), \quad \text{RMSE}_j = \sqrt{\frac{1}{50}\sum_{s=1}^{50}\big(\hat{\theta}_j^{(s)} - \theta_{j,\text{true}}\big)^2},$$

along with 95% credible-interval coverage and interval length. If a parameter is omitted (as in underfitting), we exclude it from these summaries.

For forecasting, we use the first 80 points for training and the remaining 20 points for testing. Let $\hat{\mathbf{y}}_t^{(s)}$ be the posterior mean forecast at time t in simulation s. We define

$$\text{RMSE}_{\text{forecast}} = \sqrt{\frac{1}{50 \times 20 \times 6} \sum_{s=1}^{S} \sum_{t=1}^{20} \sum_{k=1}^{6} \left(y_{t,k}^{(s)} - \hat{y}_{t,k}^{(s)}\right)^2},$$

where $\mathbf{y}_t^{(s)}$ denotes the true composition in the test set for the s-th simulation. Each of the three study designs is then evaluated under each of the five priors, illuminating how prior choice interacts with model misspecification.

4.4 Results

The summarized parameter estimation results of the three simulation studies are shown in Tables 4.1-4.3 and the summarized forecast results are shown in Tables 4.4-4.6.

4.4.1 Study 1: Correct Model Specification (DARMA(2,1))

Parameter Estimation

Table 4.1 includes results for the correctly specified DARMA(2,1). All priors yield estimates that align with the true parameter values, and both bias and RMSE remain modest. Certain shrinkage priors, including the Horseshoe and Laplace, produce slightly lower RMSE and narrower intervals than the normal prior. Hierarchical priors also reduce parameter uncertainty to some extent, although the differences relative to other shrinkage methods are not pronounced. Under spike-and-slab, small but non-zero signals can be set to zero, occasionally lowering coverage (for example, a coverage rate of 0.8240 for the intercept β).

Forecasting Performance

In Table 4.4 (Sim 1 columns), the average forecast RMSE spans 0.031–0.032, with minimal variation among priors. The Horseshoe, Laplace, and Hierarchical prior provide slight gains

in predictive accuracy compared to the Informative Normal or Spike-and-Slab approaches.

4.4.2 Study 2: Overfitting Scenario (B-DARMA(4,2))

Parameter Estimation

When B-DARMA(4,2) is applied to data generated by a DARMA(2,1) process, additional VAR and VMA terms that should be zero are introduced. In Table 4.2, relatively large biases and inflated RMSE values appear under the informative normal prior for these extraneous coefficients. Horseshoe priors shrink many of those coefficients near zero, reflected in low RMSE (for instance, 0.0176 or 0.0161 for A_3 and A_4) and coverage rates close to the nominal level. Spike-and-slab also suppresses unwanted terms, although borderline signals may be excessively reduced.

Forecasting Performance

Forecast RMSE under overfitting conditions is listed in Table 4.4 (Sim 2 columns). Higher errors (0.0341) are observed for the Informative Normal prior, while the Horseshoe produces the lowest RMSE (0.0315). Laplace and Hierarchical approaches also outperform the weakly regularized alternative. Ratios in Table 4.5 indicate that Horseshoe's forecast RMSE in the overfitted model differs little from the correctly specified case.

4.4.3 Study 3: Underfitting Scenario (B-DARMA(1,0))

Parameter Estimation

In Table 4.3, broad increases in RMSE and reduced coverage are noted when the second VAR lag and the VMA(1) term are omitted, regardless of the prior used. Horseshoe, Laplace, and Spike-and-Slab cannot compensate for missing structural components, which leads to biased AR(1) estimates and coverage gaps. The hierarchical prior exhibits comparable issues.

Forecasting Performance

Forecast RMSE remains near 0.032–0.033, indicating a 3–4% rise over the correct specification. This shortfall is consistent across priors (Table 4.5) confirming that underspecification cannot be resolved by shrinkage.

4.4.4 Overview of Results

These simulations illustrate how model specification and prior choice jointly affect B-DARMA outcomes. Under a correctly specified DARMA(2,1), all priors capture the process well, although Horseshoe and Hierarchical priors yield slightly lower RMSE and narrower intervals. In overfitted scenarios, shrinkage priors—especially Horseshoe—readily suppress spurious parameters and maintain robust forecasts. Underfitting cannot be mitigated by any prior, as exclusion of critical VAR or VMA terms inflates bias and diminishes coverage. Further exploration of alternative coefficient matrices, described in the Supplementary Material, reinforces these findings: shrinkage priors protect against overfitting but cannot repair underspecified models. Horseshoe, Laplace, and Hierarchical priors yield stable forecasts across various parameter settings, whereas the Informative Normal approach is more susceptible to inflated estimates in large or sparse models.

4.5 Application to S&P 500 Sector Trading Values

4.5.1 Motivation and Data Description

A fixed daily trading value in the S&P 500 is allocated across different sectors, giving rise to compositional data that captures how investors distribute their capital each day. Tracking these proportions over time can reveal macroeconomic trends, sector rotation, and shifts in investor sentiment. In our analysis, we examined the daily proportions of eleven S&P 500 sectors from January 2021 through December 2023. These sectors include:

- **Technology**: software, hardware, and related services
- **Healthcare**: pharmaceuticals, biotechnology, and healthcare services
- **Financials**: banking, insurance, and investment services
- **Consumer Discretionary**: non-essential goods and services
- **Industrials**: manufacturing, aerospace, defense, and machinery
- **Consumer Staples**: essential goods, such as food and household items
- **Energy**: oil, gas, and renewable resources
- **Utilities**: public services such as electricity and water
- **Real Estate**: REITs and property management
- **Materials**: chemicals, metals, and construction materials
- **Communication Services**: media, telecommunication, and internet services

Let V_{kt} denote the dollar trading value executed in sector k on trading day t for $k = 1, \ldots, K$ ($K = 11$ sectors). The market-wide total is

$$g_t \;=\; \sum_{k=1}^{K} V_{kt},$$

and the composition we analyse is the vector of sector shares

$$\mathbf{y}_t \;=\; (y_{1t}, \ldots, y_{Kt})^{\mathsf{T}}, \qquad y_{kt} = V_{kt}/g_t, \qquad \sum_{k=1}^{K} y_{kt} = 1.$$

The stationarity (or non-stationarity) of the compositional time series $\mathbf{y}_t$ is independent of that of the gross total g_t: g_t may drift without affecting the stationarity of the shares, and, conversely, the shares could be non-stationary even if g_t is itself stationary.

We used January 1, 2021, to June 30, 2023, as our training data as shown in figure 4.1. Forecast evaluation was conducted on the following 126 trading days, spanning July 1 through December 31, 2023.

Short-term variability, cyclical tendencies, and gradual long-term shifts emerge across the series. Technology and Financials consistently hold the largest fractions of the fixed daily trading value, whereas Utilities and Real Estate occupy much smaller shares. Day-to-day fluctuations tend to be more pronounced in Consumer Discretionary and Communication Services, indicating greater sensitivity to rapidly changing market conditions. Seasonal patterns and moderate differences across weekdays are also evident, as illustrated by Figures 4.2–4.4, further underscoring the multifaceted nature of sector-level trading dynamics.

Model specification A B–DARMA(P=10, Q=0) model is fitted *by design* with a lag order that exceeds the horizon over which sector reallocations are generally thought to propagate. Ten trading days ($\approx$ two calendar weeks) therefore represent a deliberate over-fit, allowing us to evaluate how strongly the global–local shrinkage priors suppress redundant dynamics. Each sector's additive-log-ratio is regressed on its own composition at lags $1, \ldots, 10$, yielding 1,000 VAR coefficients in the matrices $A_1, \ldots, A_{10}$.

Seasonality is modelled with Fourier bases that are *fixed* regardless of the chosen lag: two sine–cosine pairs capture the 5-day trading cycle (Monday–Friday) and five pairs capture the annual cycle of roughly 252 trading days. These 14 terms, together with a sector-specific intercept, form the design matrix for the linear predictor; the same seasonal structure enters the model for the Dirichlet precision ϕ_t.

In total, the specification estimates 1,165 parameters: 1,000 VAR coefficients, 140 seasonal coefficients, 10 intercepts, and 15 precision-related terms. The intentionally generous lag order thus functions as a stress test.

4.5.2 Priors and Hyperparameters

For all priors, the intercept in γ_ϕ is given a $\mathcal{N}(7, 1.5)$ prior, and all seasonal Fourier terms in γ_ϕ receive $\mathcal{N}(0, 0.1)$.

Under the *informative normal* prior, every element of $\boldsymbol{\beta}$ is modeled with a $\mathcal{N}(0, 0.1)$ prior, and each element of $\mathbf{A}_p$ has a $\mathcal{N}(0, 1)$ prior.

The *Laplace* prior (Park and Casella, 2008) is implemented via an exponential mixture: each coefficient θ_i satisfies $\theta_i \sim \mathcal{N}(0, \nu)$ with $\nu \sim \text{Exponential}(1/b)$, leading to an overall Laplace$(0, b)$ prior. We set $b_\beta = 1.0$ and $b_A = 1.0$ for $\boldsymbol{\beta}$ and $\mathbf{A}_p$ respectively.

For the *spike-and-slab* prior (Mitchell and Beauchamp, 1988), each coefficient θ_i is drawn from a mixture $\theta_i = 0$ with probability $(1 - \tau_i)$ or $\theta_i \sim \mathcal{N}(0, 1)$ with probability τ_i, where $\tau_i \sim \text{Beta}(1, 1)$ is the mixing weight.

The *horseshoe* prior (Carvalho et al., 2010b) introduces one global scale parameter τ and local scales λ_i, all drawn from Cauchy$^+(0, 1)$. Each coefficient θ_i then follows $\mathcal{N}(0,\, \tau\, \lambda_i)$.

Finally, under the *hierarchical* prior (Polson and Scott, 2012), the coefficients are partitioned into three groups: (i) the elements of $\boldsymbol{\beta}$, (ii) the diagonal entries of each lag matrix $\mathbf{A}_p$, and (iii) the off-diagonal entries of each $\mathbf{A}_p$. Then the coefficients in $\boldsymbol{\beta}$ are modeled as $\beta_i \mid \sigma_\beta \sim \mathcal{N}(0, \sigma_\beta)$, while for each lag p in $\mathbf{A}_p$, the diagonal entries follow $\mathcal{N}(0.5, \sigma_A)$ and the off-diagonal entries follow $\mathcal{N}(0, \sigma_{A,\text{off}})$.

4.5.3 Evaluation Metrics and Forecasting

A B-DARMA(10,0) model was fit with each of the five priors using the training data. A 126 day forecast was then generated for the test period using the mean of the joint predictive distribution. Mean absolute error (MAE) was used as a measure of average discrepancy between predicted and observed shares, while root mean squared error (RMSE) weighs larger deviations more explicitly.

4.5.4 Results

In Figure 4.5, forecasts (in turquoise) are shown together with actual daily proportions (in red) for each sector over the 126-day test interval. Hierarchical and Horseshoe priors led to predictions that aligned more closely with the observed values compared to the informative prior, particularly in volatile sectors such as Energy and Technology. Under the Spike-and-Slab prior, smaller signals were at times set to zero, which occasionally delayed responses to abrupt shifts in sector composition.

Sector-level RMSE values under each prior appear in Figure 4.6, and average RMSE and MAE are listed in Table 4.7. Hierarchical and spike-and-slab priors yielded the smallest RMSE in multiple sectors, and Horseshoe priors were found to be effective for larger fluctuations in Consumer Cyclical and Financial Services. The informative prior generally had higher errors, reflecting the value of regularization in a high-dimensional parameter space.

Summary of Findings

The findings in this real-data application align with our earlier simulation results. In the large-scale B-DARMA model, shrinkage priors reduced extraneous complexity and lowered forecast errors. Horseshoe, Laplace, and spike-and-slab priors limited the number of active coefficients and delivered stable predictions. Hierarchical priors performed similarly, in part because they *partially pool* information across related parameters—shrinking them toward a common distribution while still allowing differences where the data support them. Sectors with higher volatility, such as Technology, Communication Services, and Consumer Cyclical, appeared to benefit most from strong shrinkage, whereas more stable sectors like Utilities and Basic Materials performed comparably under all priors. Overall, these outcomes highlight the value of effective regularization in models with multiple lags and complex seasonality, especially in markets that experience rapid fluctuations.

4.6 Discussion

4.6.1 Comparisons Across Simulations and Real Data

Our three simulation studies demonstrate how prior selection interacts with model specification in B-DARMA. When the model order matches the true data-generating process (DARMA(2,1)), all priors yield acceptable results, though Horseshoe and Hierarchical priors produce slightly lower RMSE and better interval coverage. Overfitting by fitting B-DARMA(4,2) confirms that strong shrinkage—particularly Horseshoe—helps suppress spurious higher-order coefficients and avoids inflating forecast errors. Conversely, underfitting remains impervious to prior choice, as omitting critical AR(2) and MA(1) terms leads to uniformly higher biases and coverage shortfalls across all priors. All four shrinkage priors neutralise redundant lags, and any of them is preferable to an under-specified alternative.

These lessons translate directly to real data, where we intentionally specified a large-lag B-DARMA model for S&P 500 sector trading. Just as in the overfitting simulation, Horseshoe, Laplace, and Spike-and-Slab priors effectively shrank extraneous parameters and preserved forecast accuracy. Hierarchical priors provided comparably strong performance through partial pooling, whereas the nformative prior yielded higher forecast errors in volatile sectors such as Energy and Technology. These empirical patterns mirror the simulation findings, reinforcing that the combination of a potentially over-parameterized model and minimal shrinkage can lead to unstable estimates and suboptimal forecasts.

Computation time. Using identical `Stan` settings (4 chains, 1 250 iterations, 500 warm-up), the informative prior completed fastest in about 10 min per chain and serves as our baseline. Laplace added roughly 20 % to wall-clock time (12 min), reflecting the heavier double-exponential tails that require more leapfrog steps. Horseshoe and Hierarchical priors were slower by 40 % (14–15 min) because their global–local scale hierarchies force smaller step sizes during HMC integration. Spike-and-Slab was the clear outlier, nearly doubling

run-time (20 min) owing to the latent inclusion indicators that create funnel-shaped geometry. Practitioners with tight compute budgets might prefer the Laplace prior, while more aggressive sparsity (Horseshoe, Hierarchical, Spike-and-Slab) simply warrants proportionally longer runs.

4.6.2 Implications and Guidelines

Overall, our results underscore three main points relevant to both simulated and real-world compositional time-series modeling:

- **Shrinkage priors mitigate overfitting.** Horseshoe priors are especially adept at handling sparse dynamics and large parameter spaces, as shown by the minimal performance degradation in both simulated and real overfitting scenarios.
- **Hierarchical priors offer robust partial pooling.** They attain performance comparable to Horseshoe while retaining smooth shrinkage across correlated parameters, making them a flexible choice for multi-sector or multi-component series.
- **Model mis-specification overshadows prior advantages.** Underfitting in simulations, or failing to include essential lags in practice, leads to systematic bias and reduced coverage that shrinkage alone cannot remedy. Model identification and appropriate lag selection remain critical for accurate inference.

The S&P 500 data analysis underscores these points: shrinkage priors consistently outperformed the informative prior in managing a high-dimensional model, yet sector-specific volatility still caused reduced accuracy. Sectors with greater day-to-day variability, such as Technology or Consumer Cyclical, benefited more from aggressive regularization than stable ones (e.g., Utilities, Basic Materials).

Incorporating dynamic selection of lag structures or exploring alternative priors may further improve compositional forecasting in settings with complex seasonalities or extremely

large parameter spaces. Meanwhile, practitioners should pair robust prior modeling with careful model diagnostics (e.g., residual checks, information criteria) to ensure that underfitting and overfitting are detected early.

Overall, these findings reinforce the need for carefully chosen priors—especially in high-dimensional compositional time-series contexts. Horseshoe, Laplace, and Hierarchical priors successfully protect against overfitting while preserving essential signals, as illustrated by both simulation and real-data (S&P 500) analyses. Nonetheless, fundamental model adequacy remains paramount, as even the best prior cannot rescue a structurally under-specified model. We recommend that analysts combine thorough sensitivity checks of priors with well-informed decisions about lag order, covariate inclusion, and potential seasonal effects in B-DARMA applications.

Code Availability

All `R` scripts and `Stan` model files used in this study are publicly available at https://github.com/harrisonekatz/bdarma_sensitivity_analysis. All results and figures in this manuscript can be reproduced by running the scripts found in that repository.

Tables & Figures

Table 4.1: **Parameter Estimation Summary for Simulation Study 1 (Correct Specification).** We show mean bias, RMSE, average credible interval length, and coverage for each prior, summarizing over all parameters in each of β, A_1, A_2, and B_1. Lower RMSE and shorter intervals typically indicate more effective shrinkage, while coverage near the nominal 0.95 is desirable.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
β	Informative	-0.014	0.040	0.210	0.976
	Horseshoe	-0.016	0.042	0.181	0.948
	Laplace	-0.012	0.050	0.251	0.972
	Spike-Slab	-0.012	0.074	0.283	0.824
	Hierarchical	-0.014	0.042	0.182	0.948
A_1	Informative	-0.037	0.179	0.644	0.889
	Horseshoe	-0.007	0.084	0.306	0.969
	Laplace	-0.007	0.111	0.523	0.978
	Spike-Slab	-0.010	0.208	0.589	0.804
	Hierarchical	-0.032	0.139	0.511	0.937
A_2	Informative	0.041	0.164	0.562	0.876
	Horseshoe	0.008	0.069	0.257	0.955
	Laplace	0.007	0.094	0.442	0.982
	Spike-Slab	0.009	0.180	0.497	0.791
	Hierarchical	0.020	0.110	0.439	0.957
B_1	Informative	0.021	0.170	0.684	0.941
	Horseshoe	-0.017	0.100	0.395	0.961
	Laplace	-0.005	0.130	0.599	0.966
	Spike-Slab	0.008	0.241	0.642	0.754
	Hierarchical	0.002	0.122	0.580	0.975

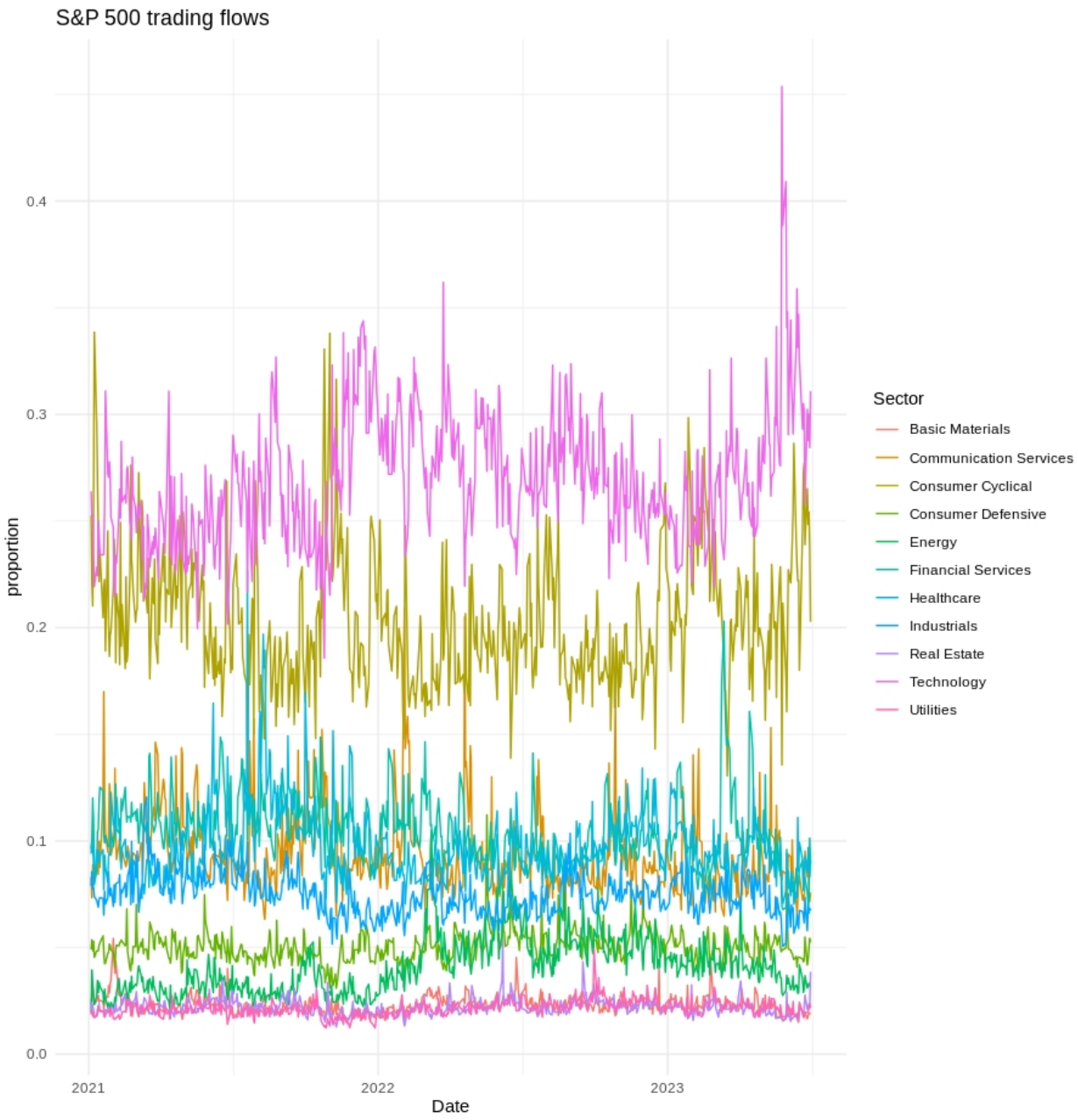

Figure 4.1: Time series of S&P 500 trading flows by sector from January 2021 to June 2023. Technology (magenta) and Financial Services (goldenrod) are consistently among the largest proportions.

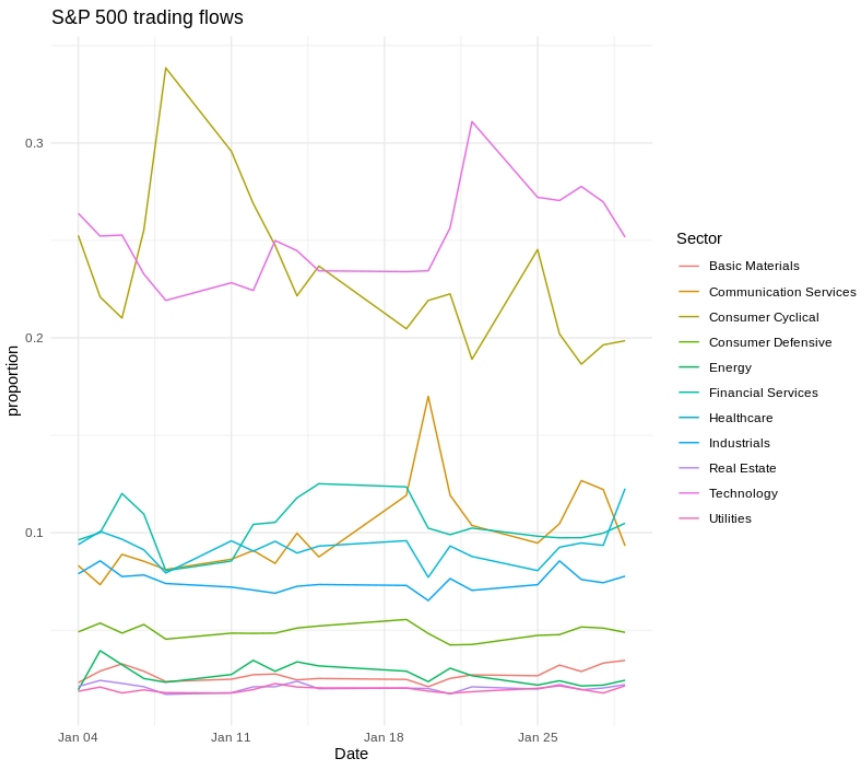

Figure 4.2: Daily S&P 500 trading flows by sector for January 2021. The shortened time window highlights intramonth volatility, especially for the Consumer Cyclical (orange) and Communication Services (yellow) sectors.

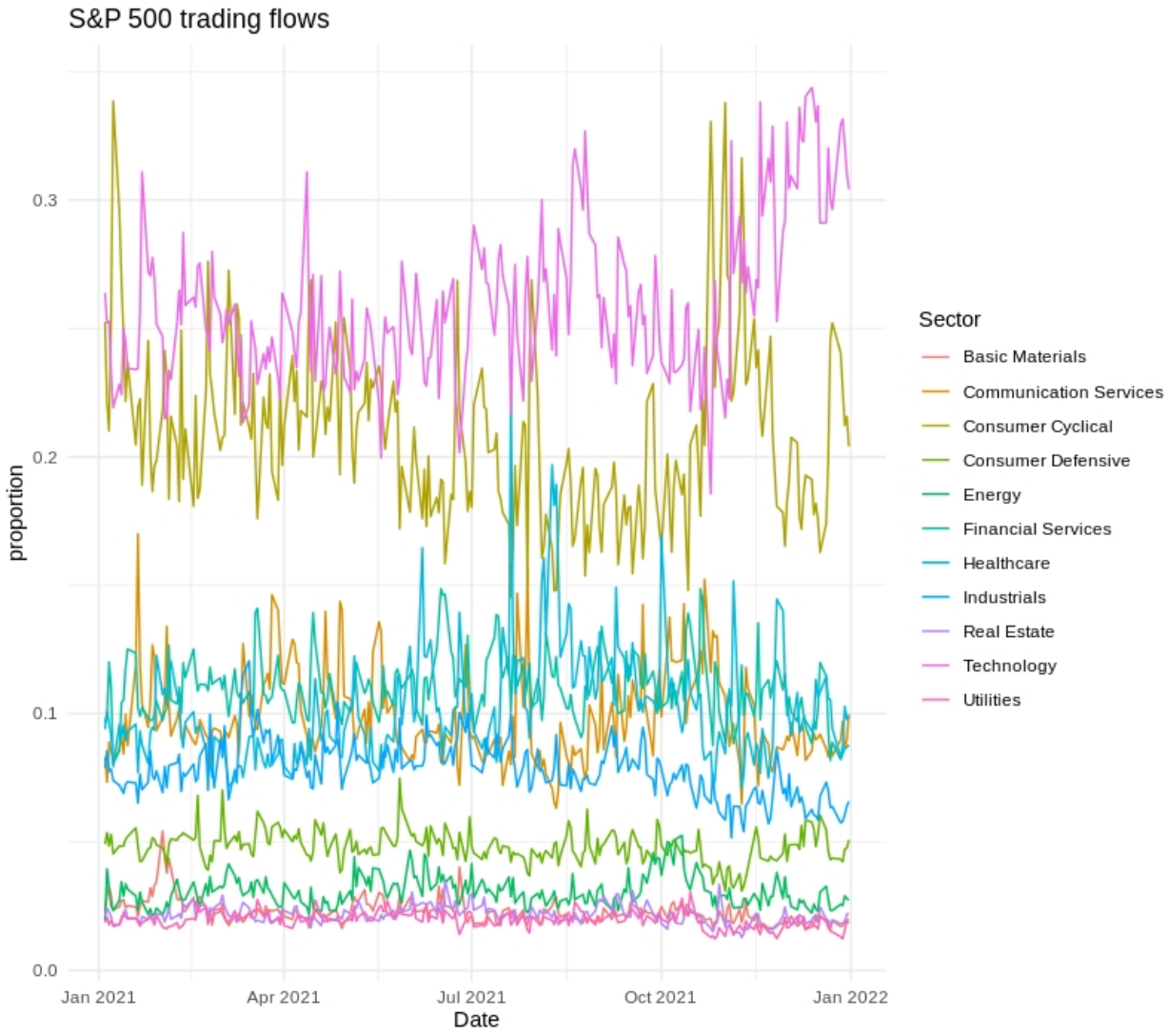

Figure 4.3: Time series of S&P 500 trading flows from Jan 2021 to Jan 2022. Despite daily and yearly fluctuations, Technology remains among the top in proportion of total trading value.

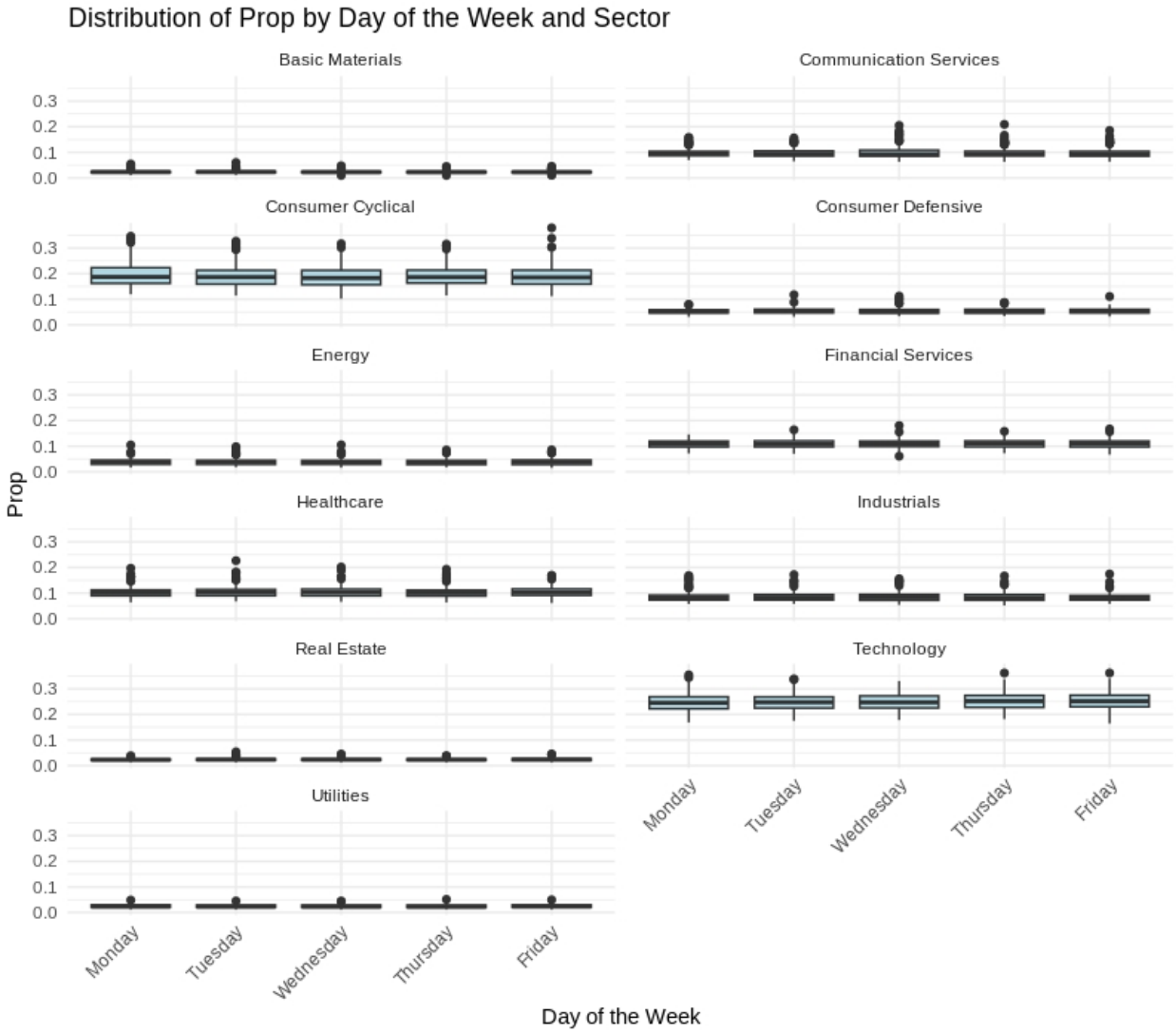

Figure 4.4: Boxplots of S&P 500 sector trading proportions by day of the week. Outliers primarily occur for Consumer Cyclical and Technology, though smaller sectors also display occasional spikes.

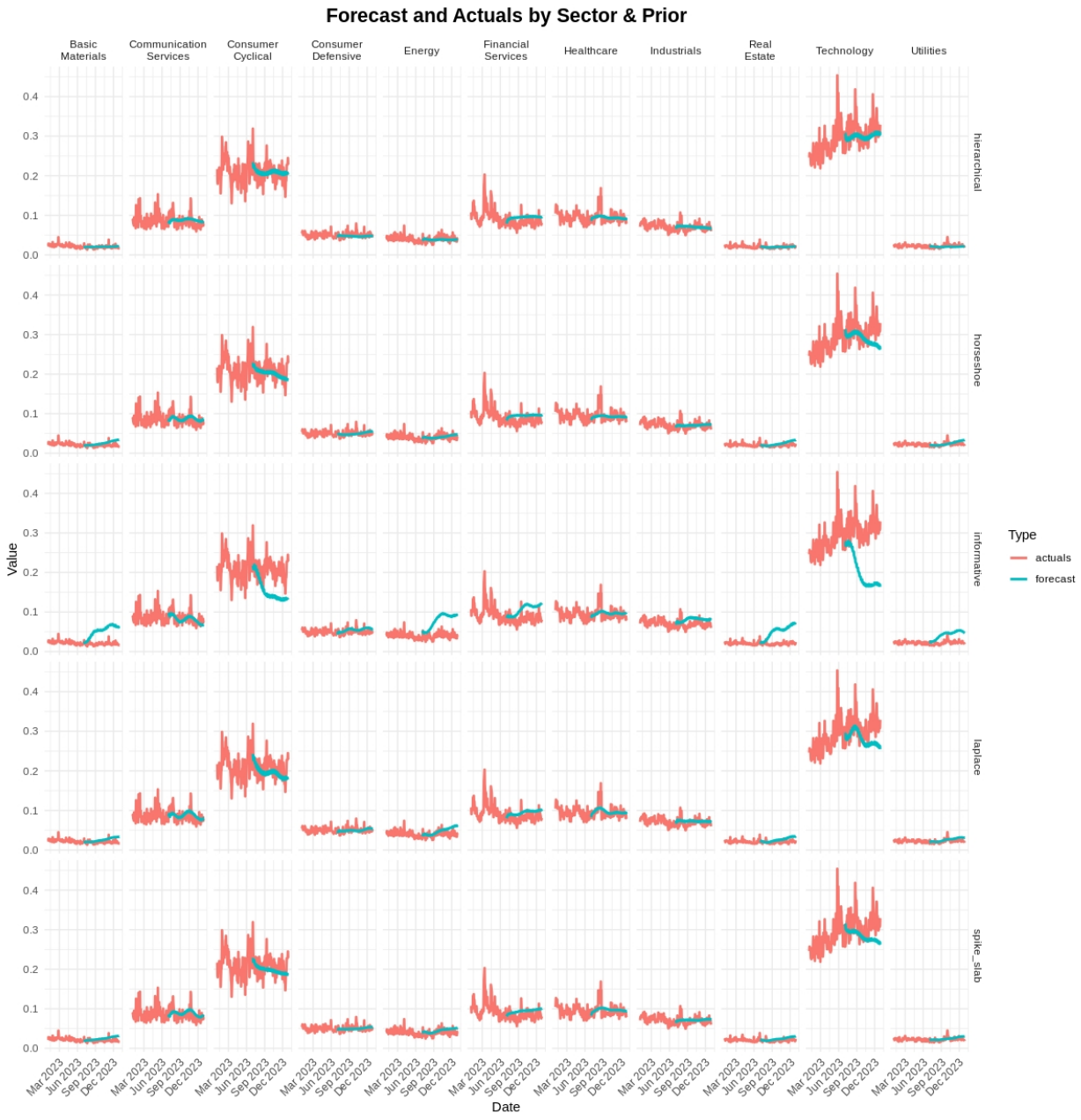

Figure 4.5: Forecast and Actuals by Sector and Prior for S&P 500 Data. Red lines indicate actual daily sector proportions; turquoise lines are the forecasts from the fitted B-DARMA model. Horseshoe and hierarchical priors tend to track actual series more tightly than the informative prior in many sectors.

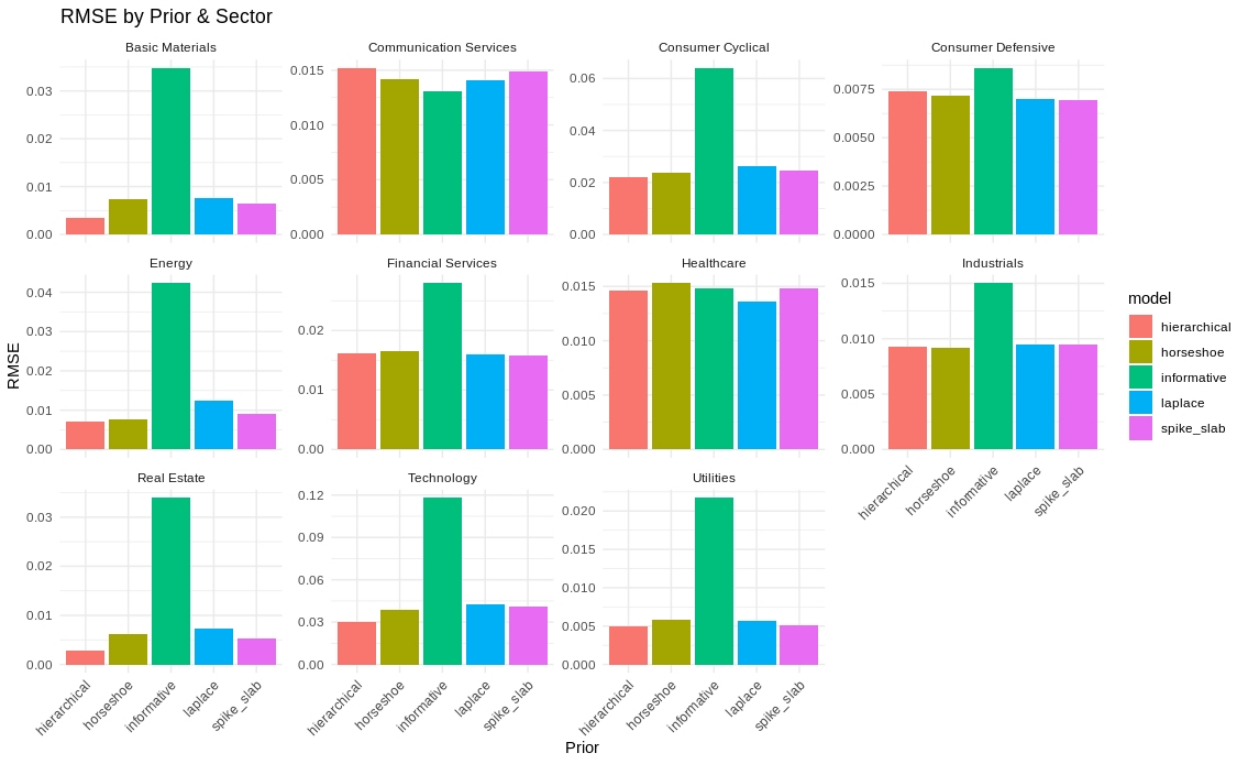

Figure 4.6: RMSE by Prior and Sector for S&P 500 forecasting. Each facet shows one sector's RMSE across the five priors. Shrinkage priors (horseshoe, laplace, spike_slab) and hierarchical strategies generally outperform the more loosely regularized informative prior.

Table 4.2: Parameter Estimation Summary for Simulation Study 2 (Overfitting). This table reflects the setting where the fitted model (B-DARMA(4,2)) exceeds the true DARMA(2,1) order. We report mean bias, RMSE, credible interval length, and coverage for each prior and parameter.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
	Informative	-0.012	0.043	0.205	0.956
	Horseshoe	-0.010	0.042	0.176	0.944
β	Laplace	-0.008	0.053	0.219	0.932
	Spike-Slab	-0.004	0.099	0.303	0.740
	Hierarchical	-0.011	0.042	0.192	0.960
	Informative	-0.061	0.224	0.751	0.809
	Horseshoe	-0.053	0.157	0.310	0.915
A_1	Laplace	-0.046	0.178	0.657	0.954
	Spike-Slab	-0.022	0.217	0.684	0.805
	Hierarchical	0.072	0.157	0.158	0.650
	Informative	0.033	0.176	0.756	0.925
	Horseshoe	0.032	0.091	0.225	0.898
A_2	Laplace	0.017	0.128	0.611	0.958
	Spike-Slab	0.002	0.203	0.691	0.810
	Hierarchical	-0.078	0.180	0.197	0.683
	Informative	0.017	0.143	0.670	0.966
	Horseshoe	-0.004	0.022	0.168	1.000
A_3	Laplace	0.002	0.100	0.514	0.986
	Spike-Slab	0.004	0.183	0.600	0.818
	Hierarchical	0.035	0.093	0.172	0.962
	Informative	0.002	0.113	0.473	0.949
	Horseshoe	-0.002	0.017	0.137	0.999
A_4	Laplace	-0.005	0.081	0.363	0.974
	Spike-Slab	-0.006	0.135	0.409	0.805
	Hierarchical	-0.008	0.035	0.118	0.990
	Informative	0.051	0.231	0.845	0.854
	Horseshoe	0.026	0.141	0.419	0.946
B_1	Laplace	0.040	0.202	0.769	0.949
	Spike-Slab	0.012	0.233	0.749	0.812
	Hierarchical	-0.096	0.187	0.237	0.756
	Informative	0.055	0.231	0.862	0.861
	Horseshoe	0.014	0.072	0.349	0.996
B_2	Laplace	0.028	0.171	0.755	0.969
	Spike-Slab	0.019	0.216	0.753	0.851
	Hierarchical	-0.005	0.023	0.236	1.000

Table 4.3: **Parameter Estimation Summary for Simulation Study 3 (Underfitting).** This table reports key metrics (mean bias, RMSE, average CI length, coverage) when crucial AR(2) and MA(1) terms are omitted. All priors suffer from higher errors and coverage shortfalls, indicating that structural misspecification is the dominant source of inaccuracy.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
β	Informative	-0.017	0.042	0.229	0.980
	Horseshoe	-0.018	0.046	0.224	0.944
	Laplace	-0.017	0.051	0.354	0.988
	Spike-Slab	-0.013	0.064	0.837	1.000
	Hierarchical	-0.017	0.044	0.222	0.964
A_1	Informative	-0.032	0.214	0.361	0.657
	Horseshoe	-0.031	0.195	0.315	0.708
	Laplace	-0.029	0.210	0.346	0.693
	Spike-Slab	-0.026	0.233	0.390	0.656
	Hierarchical	-0.036	0.215	0.358	0.653

Table 4.4: Forecast Performance Summary Across Simulations. M-RMSE is the mean (across simulations) of the root mean squared error on the test set, and SD-RMSE is its standard deviation.

Prior	Sim 1: True DGP		Sim 2: Overfitting		Sim 3: Underfitting	
	M-RMSE	SD RMSE	M-RMSE	SD RMSE	M-RMSE	SD RMSE
Informative	0.0313	0.0039	0.0324	0.0039	0.0322	0.0041
Horseshoe	0.0310	0.0036	0.0305	0.0035	0.0323	0.0041
Laplace	0.0313	0.0039	0.0314	0.0040	0.0326	0.0043
Spike-Slab	0.0323	0.0044	0.0327	0.0042	0.0331	0.0048
Hierarchical	0.0312	0.0038	0.0305	0.0033	0.0322	0.0041

Table 4.5: Forecasting Performance Ratios for Mean RMSE and SD RMSE. "S2" = Overfitting scenario, "S3" = Underfitting, "S1" = Correct DGP. Columns show how each simulation compares in terms of Mean RMSE (left) and SD RMSE (right). Ratios > 1 indicate worse performance relative to the denominator; < 1 indicates better.

	Mean RMSE Ratio				SD RMSE Ratio			
Prior	**S2/S1**	**S3/S1**	**S3/S2**	**S2/S3**	**S2/S1**	**S3/S1**	**S3/S2**	**S2/S3**
Informative	1.091	1.030	0.945	1.058	1.385	1.051	0.759	1.318
Horseshoe	1.016	1.042	1.026	0.975	1.278	1.139	0.891	1.122
Laplace	1.060	1.042	0.983	1.017	1.385	1.103	0.796	1.256
Spike-Slab	1.042	1.025	0.984	1.016	1.227	1.091	0.889	1.125
Hierarchical	1.022	1.032	1.010	0.990	1.211	1.079	0.891	1.122

Table 4.6: Forecasting Performance Ratios Within Simulations (Best Model as Denominator)

Prior	**Sim 1: True DGP**		**Sim 2: Overfitting**		**Sim 3: Underfitting**	
	Mean RMSE	SD RMSE	Mean RMSE	SD RMSE	Mean RMSE	SD RMSE
Informative	1.010	1.083	1.083	1.174	1.000	1.000
Horseshoe	1.000	1.000	1.000	1.000	1.003	1.000
Laplace	1.010	1.083	1.053	1.174	1.012	1.049
Spike-Slab	1.042	1.222	1.071	1.174	1.029	1.171
Hierarchical	1.007	1.056	1.011	1.000	1.000	1.000

Table 4.7: **Mean RMSE and MAE by Model for the S&P 500 Analysis.** We aggregate forecast errors across all 11 sectors under each prior. Lower RMSE and MAE indicate better overall predictive performance, revealing that hierarchical and horseshoe strategies often outperform the more permissive informative prior.

Model	**RMSE**	**MAE**
Informative	0.0358	0.03190
Horseshoe	0.0138	0.01060
Laplace	0.0148	0.01180
Spike-Slab	0.0140	0.01110
Hierarchical	0.0121	0.00910

Supplementary Material: Additional Simulation Studies

Here we present three additional simulation studies conducted under the same general framework as in the main text. The only difference is the use of alternative coefficient matrices for the data-generating process (DGP). As before, each study is based on 50 simulated datasets, and we examine the same set of priors: Informative Normal, Horseshoe, Laplace, Spike-and-Slab, and Hierarchical. We focus on the same three scenarios: (1) correct model specification, (2) intentional overfitting, and (3) intentional underfitting.

Data-Generating Process

In these supplementary simulations, we use a DARMA(2,1) model with the following coefficient matrices:

$$\mathbf{A}_1 = \begin{bmatrix} 0.80 & -0.08 & -0.08 & -0.09 & -0.08 \\ -0.06 & 0.70 & -0.08 & 0.06 & 0.06 \\ -0.06 & 0.07 & 0.90 & 0.05 & -0.09 \\ 0.07 & 0.09 & 0.07 & 0.85 & 0.08 \\ 0.05 & -0.09 & -0.07 & -0.09 & 0.75 \end{bmatrix}$$

$$\mathbf{A}_2 = \begin{bmatrix} -0.30 & -0.07 & -0.06 & -0.08 & 0.09 \\ 0.06 & -0.20 & -0.07 & -0.07 & -0.09 \\ 0.05 & -0.07 & -0.25 & 0.07 & -0.05 \\ -0.06 & 0.08 & -0.09 & -0.15 & 0.07 \\ -0.05 & 0.06 & 0.10 & 0.07 & -0.20 \end{bmatrix},$$

$$\mathbf{B}_1 = \begin{bmatrix} 0.50 & -0.06 & -0.06 & 0.05 & 0.09 \\ 0.07 & 0.40 & 0.07 & 0.08 & -0.06 \\ -0.10 & -0.05 & 0.45 & 0.08 & 0.10 \\ 0.09 & 0.08 & -0.09 & 0.35 & 0.10 \\ 0.09 & -0.08 & 0.06 & -0.09 & 0.40 \end{bmatrix}.$$

As in the main study, we set $\phi = 500$ and use a 5-component vector for the covariate effect $\boldsymbol{\beta}^\top = (0.1, -0.05, 0.03, -0.02, 0.04)$. The initial two observations of the compositional series $\mathbf{y}_t$ are generated from a Dirichlet($\mathbf{1}$) distribution to start the recursion.

Study S1: Correct Model Specification (DARMA(2,1))

Parameter Estimation

When fitting the correctly specified DARMA(2,1) model under these new coefficient matrices, the patterns are similar to those in the main text. Horseshoe and Laplace priors successfully induce shrinkage on small coefficients while preserving large signals, resulting in relatively low RMSE and near-nominal coverage for both VAR and VMA parameters. Hierarchical priors again strike a balance, offering moderate shrinkage without overly broad credible intervals. Spike-and-Slab priors tend to more aggressively zero out parameters, sometimes overshrinking moderate-size coefficients. Informative Normal priors perform adequately but do not penalize unnecessary complexity as strongly.

Forecasting Performance

Forecasting accuracy remains high for the correctly specified model. Forecast RMSE values are similar across priors, with Horseshoe and Hierarchical priors again offering slight improvements in predictive accuracy. The differences are subtle, as an appropriate model structure provides a strong baseline. Shrinkage priors mainly help by stabilizing parameter estimates and reducing uncertainty.

Study S2: Intentional Overfitting (B-DARMA(4,2))

Parameter Estimation

With overfitting (fitting B-DARMA(4,2) instead of DARMA(2,1)), the key challenge is the presence of additional VAR and VMA terms that are truly zero. As in the main simulations, Horseshoe priors excel at shrinking these extraneous coefficients to near zero, yielding low bias and RMSE and appropriate coverage for the truly null parameters. Spike-and-Slab also mitigates overfitting but may overly penalize some coefficients on the borderline between zero and small nonzero values. Informative Normal priors, lacking strong shrinkage, more frequently attribute nonzero mass to spurious terms, resulting in inflated RMSE and wider intervals. Laplace and Hierarchical priors offer intermediate levels of shrinkage, improving upon non-informative priors but not matching the Horseshoe's consistently strong performance.

Forecasting Performance

Forecasting under intentional overfitting highlights the value of shrinkage priors. Horseshoe priors minimize the performance degradation caused by unnecessary lags, showing lower RMSE relative to Informative Normal priors. Laplace and Hierarchical priors also help restrain overfitting, improving forecasts compared to non-informative priors. Spike-and-Slab's aggressive thresholding can stabilize forecasts but may slightly reduce predictive performance if it eliminates coefficients too aggressively. Overall, priors that induce sparsity lead to more stable, interpretable, and accurate forecasts under overfitting conditions.

Study S3: Intentional Underfitting (B-DARMA(1,0))

Parameter Estimation

Underfitting by fitting a B-DARMA(1,0) model to data generated from DARMA(2,1) again demonstrates that no prior can compensate for missing VAR(2) and VMA(1) terms. All

priors show higher bias and RMSE for the parameters that cannot be properly identified due to structural omissions in the model. Coverage rates are reduced, and credible intervals often fail to capture the true values adequately. This pattern is consistent with the main text results and confirms that when the model is fundamentally misspecified, prior choice plays a limited role in improving inference.

Forecasting Performance

In the underfitting scenario, forecasting accuracy deteriorates uniformly across priors. The lack of necessary model terms impairs the capture of true temporal dynamics, leading to systematically higher RMSE and suboptimal predictive log-likelihoods. None of the considered priors meaningfully improves forecasts when the underlying process complexity is not well represented by the chosen model order.

Summary of Supplementary Simulations

These three supplementary studies corroborate and reinforce the findings from the main simulations. Under correct model specification, all priors perform reasonably well, with Horseshoe, Laplace, and Hierarchical priors offering slight advantages in estimation and prediction. Under overfitting, Horseshoe and other shrinkage-oriented priors effectively suppress extraneous terms, preventing inflated uncertainty and degrading forecasts. Underfitting remains impervious to prior-based corrections, with all priors exhibiting similarly poor performance when key dynamics are omitted.

In combination, the supplementary results confirm that the main conclusions are robust to changes in the underlying coefficient matrices. Priors that induce shrinkage or sparsity consistently help navigate overfitting challenges, while none can overcome the fundamental limitations imposed by an underspecified model. As a result, these supplementary studies bolster our recommendation to match model complexity to the data-generating process and to employ shrinkage priors when structural uncertainty or risk of overfitting is present.

Table 4.8: **Parameter Estimation Summary for Simulation Study 4 (Correct Specification).** We show mean bias, RMSE, average credible interval length, and coverage for each prior, summarizing over all parameters in each of β, A_1, A_2, and B_1. Lower RMSE and shorter intervals typically indicate more effective shrinkage, while coverage near the nominal 0.95 is desirable.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
β	Informative	0.000	0.042	0.203	0.972
	Horseshoe	-0.002	0.042	0.184	0.956
	Laplace	0.003	0.052	0.249	0.964
	Spike-Slab	0.011	0.087	0.302	0.812
	Hierarchical	-0.000	0.044	0.185	0.952
A_1	Informative	-0.038	0.170	0.616	0.897
	Horseshoe	-0.008	0.092	0.344	0.949
	Laplace	-0.011	0.105	0.520	0.987
	Spike-Slab	-0.010	0.190	0.589	0.818
	Hierarchical	-0.030	0.123	0.507	0.958
A_2	Informative	0.043	0.156	0.533	0.885
	Horseshoe	0.009	0.079	0.286	0.938
	Laplace	0.006	0.088	0.436	0.989
	Spike-Slab	0.005	0.163	0.495	0.800
	Hierarchical	0.016	0.102	0.435	0.971
B_1	Informative	0.017	0.162	0.677	0.952
	Horseshoe	-0.018	0.106	0.426	0.960
	Laplace	-0.008	0.121	0.613	0.984
	Spike-Slab	-0.003	0.227	0.665	0.795
	Hierarchical	-0.006	0.108	0.584	0.990

Table 4.9: Parameter Estimation Summary for Simulation Study 5 (Overfitting). This table reflects the setting where the fitted model (B-DARMA(4,2)) exceeds the true DARMA(2,1) order. We report mean bias, RMSE, credible interval length, and coverage for each prior and parameter.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
	Informative	-0.012	0.043	0.207	0.956
	Horseshoe	-0.009	0.043	0.176	0.940
β	Laplace	-0.009	0.054	0.223	0.932
	Spike-Slab	-0.004	0.100	0.310	0.748
	Hierarchical	-0.010	0.042	0.194	0.960
	Informative	-0.060	0.224	0.754	0.813
	Horseshoe	-0.055	0.160	0.308	0.914
A_1	Laplace	-0.046	0.178	0.661	0.960
	Spike-Slab	-0.020	0.215	0.680	0.804
	Hierarchical	0.073	0.158	0.157	0.645
	Informative	0.033	0.177	0.761	0.925
	Horseshoe	0.032	0.092	0.222	0.895
A_2	Laplace	0.019	0.128	0.613	0.962
	Spike-Slab	0.002	0.203	0.686	0.810
	Hierarchical	-0.079	0.182	0.197	0.682
	Informative	0.017	0.143	0.674	0.968
	Horseshoe	-0.004	0.022	0.166	1.000
A_3	Laplace	0.001	0.100	0.517	0.987
	Spike-Slab	0.005	0.184	0.598	0.814
	Hierarchical	0.035	0.093	0.172	0.962
	Informative	0.003	0.113	0.475	0.950
	Horseshoe	-0.002	0.017	0.137	0.999
A_4	Laplace	-0.004	0.081	0.366	0.975
	Spike-Slab	-0.006	0.134	0.409	0.808
	Hierarchical	-0.008	0.035	0.118	0.990
	Informative	0.051	0.232	0.849	0.856
	Horseshoe	0.029	0.145	0.419	0.946
B_1	Laplace	0.040	0.203	0.770	0.951
	Spike-Slab	0.010	0.231	0.748	0.815
	Hierarchical	-0.096	0.187	0.237	0.756
	Informative	0.054	0.230	0.867	0.864
	Horseshoe	0.014	0.073	0.349	0.996
B_2	Laplace	0.026	0.173	0.758	0.971
	Spike-Slab	0.016	0.217	0.753	0.846
	Hierarchical	-0.005	0.023	0.236	1.000

Table 4.10: **Parameter Estimation Summary for Simulation Study 6 (Underfitting).** This table reports key metrics (mean bias, RMSE, average CI length, coverage) when crucial VAR(2) and VMA(1) terms are omitted. All priors suffer from higher errors and coverage shortfalls, indicating that structural misspecification is the dominant source of inaccuracy.

Coefficient	Prior	Mean Bias	Mean RMSE	Mean CI Length	Coverage
β	Informative	-0.006	0.044	0.228	0.972
	Horseshoe	-0.009	0.047	0.221	0.940
	Laplace	-0.006	0.054	0.333	0.984
	Spike-Slab	-0.001	0.068	0.719	0.996
	Hierarchical	-0.009	0.047	0.216	0.944
A_1	Informative	-0.035	0.210	0.342	0.609
	Horseshoe	-0.037	0.198	0.307	0.619
	Laplace	-0.034	0.206	0.332	0.628
	Spike-Slab	-0.029	0.223	0.372	0.616
	Hierarchical	-0.040	0.210	0.339	0.602

Table 4.11: Forecast Performance Summary Across Supplementary Simulations. M-RMSE is the mean (across simulations) of the root mean squared error on the test set, and SD-RMSE is its standard deviation.

Prior	Sim 4: True DGP		Sim 5: Overfitting		Sim 6: Underfitting	
	M-RMSE	SD RMSE	M-RMSE	SD RMSE	M-RMSE	SD RMSE
Informative	0.0378	0.0060	0.0414	0.0079	0.0394	0.0085
Horseshoe	0.0379	0.0063	0.0381	0.0057	0.0394	0.0085
Laplace	0.0381	0.0063	0.0420	0.0073	0.0399	0.0086
Spike-Slab	0.0396	0.0064	0.0412	0.0072	0.0408	0.0085
Hierarchical	0.0376	0.0063	0.0385	0.0062	0.0394	0.0083

Table 4.12: Forecasting Performance Ratios for Mean RMSE (M-RMSE) and its Standard Deviation (SD-RMSE). "S5" = Overfitting scenario, "S6" = Underfitting, "S4" = Correct DGP. Columns show how each simulation compares in terms of M-RMSE and SD-RMSE. Ratios > 1 indicate worse performance (higher RMSE) relative to the denominator; < 1 indicates better (lower RMSE).

Prior	M-RMSE Ratio				SD-RMSE Ratio			
	5/4	6/4	6/5	5/6	5/4	6/4	6/5	5/6
Informative	1.091	1.030	0.945	1.058	1.385	1.051	0.758	1.319
Horseshoe	1.016	1.042	1.026	0.974	1.278	1.139	0.892	1.121
Laplace	1.060	1.042	0.983	1.017	1.385	1.103	0.796	1.256
Spike-Slab	1.042	1.025	0.984	1.016	1.227	1.091	0.889	1.125
Hierarchical	1.022	1.032	1.010	0.990	1.211	1.079	0.891	1.122

Table 4.13: Forecasting Performance Ratios Within Simulations (Best Model as Denominator, rounded to two digits). M-RMSE is the mean root mean squared error and SD-RMSE is the standard deviation of the root mean squared errors on the test set across simulations.

Prior	**Sim 4: True DGP**		**Sim 5: Overfitting**		**Sim 6: Underfitting**	
	M-RMSE	**SD RMSE**	**M-RMSE**	**SD RMSE**	**M-RMSE**	**SD RMSE**
Informative	1.01	1.08	1.08	1.17	1.00	1.00
Horseshoe	1.00	1.00	1.00	1.00	1.00	1.00
Laplace	1.01	1.08	1.05	1.17	1.01	1.05
Spike-Slab	1.04	1.22	1.07	1.17	1.03	1.17
Hierarchical	1.01	1.06	1.01	1.00	1.00	1.00

Chapter 5

Bayesian Shrinkage in High-Dimensional VAR Models: A Comparative Study

5.1 Introduction

Vector autoregressive (VAR) models remain a cornerstone of multivariate time series analysis (Sims, 1980; Chan, 2020; Koop and Korobilis, 2013). A d-dimensional VAR(p) posits that an observed time series $\{\mathbf{y}_t\}_{t=1}^T$ satisfies a regression model where $\mathbf{y}_t$ given p previous observations $\mathbf{y}_{t-p}, \ldots, \mathbf{y}_{t-1}$ is modeled as

$$\mathbf{y}_t \;=\; \boldsymbol{A}_1\mathbf{y}_{t-1} \;+\; \boldsymbol{A}_2\mathbf{y}_{t-2} \;+\; \cdots \;+\; \boldsymbol{A}_p\mathbf{y}_{t-p} \;+\; \boldsymbol{\varepsilon}_t. \tag{5.1}$$

When each $\boldsymbol{A}_i$ is a $d \times d$ coefficient matrix, and we assume that

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon),$$

is normally distributed white noise, that is, $\boldsymbol{\varepsilon}_t$ has mean $\mathbf{0}$ and covariance Σ_ε and the $\boldsymbol{\varepsilon}_t$'s are uncorrelated over time. VAR models see extensive use in macroeconomic data analysis, where the assumption of correlated vector time series is natural; see, for instance, Stock and Watson (2002); Crump et al. (2021); Carriero et al. (2022); Zhou and Chan (2023) for discussions on the ongoing empirical suitability and methodological refinements of VAR structures in economics.

Standard VAR theory relies on the stationarity of the time series. If the characteristic polynomial of $(I_d - \boldsymbol{A}_1 z - \cdots - \boldsymbol{A}_p z^p)$ does not have all its roots outside the unit circle, the process may be nonstationary or cointegrated. In such cases, one typically either differences or applies error-correction forms to restore stationarity, or else fits a specialized variant (such as a vector error-correction model).

A single VAR(p) of dimension d has d^2p parameters, so with increasing d or p, the parameter space quickly becomes high dimensional (Bańbura et al., 2010; Koop, 2013; Korobilis and Pettenuzzo, 2019). Consequently, regularization (often referred to as "shrinkage" in the VAR literature) is crucial not only to prevent overfitting but also to control the high variability in estimated parameters, which can undermine model stability and predictive accuracy (Stock and Watson, 2002; Huber and Feldkircher, 2017; Chan, 2021). In this context, *shrinkage* describes the penalization of numerous parameters toward zero or toward a simpler structure, so as to enforce parsimony. For VAR models, shrinkage is particularly needed when the number of time series variables (d) or the chosen lag order (p) is large relative to the available sample size, a common scenario in macroeconomic forecasting (Bańbura et al., 2010; Bai et al., 2022).

A variety of shrinkage techniques have been developed in both Bayesian and frequentist frameworks. Bayesian local-global priors, such as the horseshoe (Carvalho et al., 2010b) and Bayesian lasso (Park and Casella, 2008), aggressively shrink small or irrelevant coefficients toward zero while allowing large signals to remain relatively unshrunk. The spike-and-slab prior (George and McCulloch, 1997) offers an alternative Bayesian framework, combining

sparsity with flexibility in signal detection. Hierarchical shrinkage priors, like the horseshoe and its variants (Makalic and Schmidt, 2016; Prüser, 2021), adaptively adjust shrinkage intensity and have shown strong performance in time series applications (Prüser, 2021). Other recent Bayesian shrinkage methods in time series have been explored (Huber and Koop, 2023; Kowal et al., 2019; Katz et al., 2024). Frequentist approaches such as ridge regression (Doan et al., 1984) penalize large coefficients to prevent overfitting, while nonparametric shrinkage approaches can directly estimate an optimal shrinkage factor—often via James–Stein type estimators—and adjust the sample covariance-based estimates accordingly (Giannone et al., 2015; Del Negro and Giannoni, 2015).

Recent literature has refined global-local priors and explored flexible hierarchical frameworks tailored to VARs. Huber and Feldkircher (2019b) propose an adaptive shrinkage scheme with a Normal-Gamma prior that learns the degree of shrinkage across coefficients, showing improved forecasts in large VAR systems. Bitto and Frühwirth-Schnatter (2019b) apply a Horseshoe-type prior in time-varying parameter models, highlighting the ability of heavy-tailed priors to shrink aggressively while preserving large signals. In the realm of block-specific shrinkage, Aprigliano (2020) demonstrates that treating different coefficient groups with distinct shrinkage intensities can further boost forecast accuracy. Recent computational strategies, such as variational approximations, also facilitate faster Bayesian inference for high-dimensional VARs, as discussed in Gefang et al. (2023).

Lasso-based methods (Tibshirani, 1996) and elastic net (Zou and Hastie, 2005) also appear in the high-dimensional VAR toolkit, providing sparse solutions when many coefficients are truly zero or near zero. Recent innovations, such as dynamic shrinkage models (Griffin and Brown, 2017) and time-varying parameter (TVP) VARs (Nakajima and West, 2011), add further flexibility for capturing structural breaks or changing relationships over time.

Beyond ridge and nonparametric shrinkage, recent research has proposed more flexible or structured penalties to handle the large parameter spaces typical of VAR models. Group lasso and hierarchical lag penalties can exploit the grouped structure of lagged predictors

by shrinking or eliminating entire blocks of coefficients (Nicholson et al., 2017; Basu et al., 2019; Nicholson, 2020). Non-convex penalties such as the smoothly clipped absolute deviation (SCAD) and the minimax concave penalty (MCP) balance aggressive shrinkage of small coefficients with reduced bias for large signals (Song and Bickel, 2019; Chen and Chen, 2021). Adaptive lasso variants further improve selection consistency by penalizing large coefficients less, based on preliminary estimates (Kock and Callot, 2015). Lastly, time series–specific modifications—like rolling cross-validation or penalization that accounts for serial dependence (Nicholson et al., 2017; Medeiros et al., 2021)—help tailor these methods to dynamic settings, often improving stability and forecast performance in high-dimensional VARs.

In this paper, we compare five shrinkage approaches for high-dimensional VAR estimation: three Bayesian priors (horseshoe, lasso, and normal) and two frequentist estimators (ridge and nonparametric shrinkage). We evaluate each method's handling of overparameterization in both low- and high-dimensional settings, using root mean squared error, interval coverage, interval length, and RMSE to assess parameter recovery and forecast performance.

We find that local-global priors—particularly the horseshoe—strike a strong balance between parsimony and flexibility, delivering accurate estimation and consistent coverage even in heavily overfitted scenarios. While ridge regression frequently provides competitive point forecasts, it underestimates uncertainty when the parameter space grows large. Meanwhile, nonparametric shrinkage, though computationally efficient, suffers from undercoverage in complex models.

Application to Canadian macroeconomic time series illustrates how each method behaves with various lag choices and a relatively small sample. These empirical findings corroborate the simulation evidence: local-global priors, especially the horseshoe, remain robust even when the chosen lag order exceeds what is strictly necessary.

The remainder of the paper is structured as follows. Section 2 defines the VAR(p) framework and outlines both Bayesian and frequentist shrinkage estimators. Section 3 describes

our simulation designs and metrics, Section 4 presents the simulation results, and Section 5 provides the real-data application to Canadian macroeconomic variables. Finally, we offer concluding remarks.

5.2 Bayesian and Frequentist Approaches to VAR(p)

We study a d-dimensional VAR(p) of the form

$$\mathbf{y}_t \;=\; \mathbf{A}_1\,\mathbf{y}_{t-1} \;+\; \cdots \;+\; \mathbf{A}_p\,\mathbf{y}_{t-p} \;+\; \boldsymbol{\varepsilon}_t, \tag{5.2}$$

where each $\mathbf{A}_i$ is a $d \times d$ coefficient matrix, and $\boldsymbol{\varepsilon}_t$ is a white-noise process following

$$\boldsymbol{\varepsilon}_t \;\sim\; \mathcal{N}\big(\mathbf{0},\, \Sigma_\varepsilon\big), \quad \mathrm{Cov}\big(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s\big) \;=\; \mathbf{0} \text{ for } t \neq s.$$

We treat each $\mathbf{y}_t$ as a $d \times 1$ column vector.

Vectorizing the Coefficients. Let

$$\mathbf{B} \;=\; \big[\mathbf{A}_1\;\mathbf{A}_2\;\cdots\;\mathbf{A}_p\big] \;\in\; \mathbb{R}^{d\times(dp)},$$

i.e., the horizontal concatenation of the p coefficient matrices with elements β_j, $j = 1, \ldots, d\times dp$. Then define the *lagged-regressor* vector

$$\mathbf{X}_t \;=\; \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix} \;\in\; \mathbb{R}^{dp\times 1},$$

so that

$$\mathbf{B}\,\mathbf{X}_t \;\in\; \mathbb{R}^{d\times 1}.$$

The VAR(p) model in (5.1) can thus be written as

$$\mathbf{y}_t \;=\; \mathbf{B}\,\mathbf{X}_t \;+\; \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}\big(\mathbf{0},\, \Sigma_\varepsilon\big).$$

5.2.1 Bayesian Shrinkage Priors

In a fully Bayesian treatment, we specify priors for both the coefficient matrix and the error covariance. We gather the coefficients into a matrix $\mathbf{B} \in \mathbb{R}^{d\times(dp)}$, so that

$$\mathbf{y}_t \;\sim\; \mathcal{N}\big(\mathbf{X}_t\,\mathbf{B}',\; \Sigma_\varepsilon\big),$$

where $\mathbf{X}_t$ is the row vector of the lagged responses at time t. To ensure Σ_ε is positive-definite, we use a Cholesky-factor parameterization:

$$\Sigma_\varepsilon \;=\; \mathbf{L}\,\mathbf{L}^\top, \quad \mathbf{L} = \mathrm{diag}(\sigma)\,\mathbf{L}_\Omega,$$

where $\mathbf{L}_\Omega$ is the Cholesky factor of a correlation matrix with an LKJ prior (Lewandowski et al., 2009), and each component of σ follows a half-Cauchy prior. This flexible structure permits correlation among the d error components.

We then place shrinkage priors on each coefficient in $\mathbf{B}$. Below, we detail three such priors—normal (ridge), horseshoe, and Bayesian lasso—all of which can be combined with the same LKJ-based prior for Σ_ε:

Normal Prior (Bayesian Ridge). A normal (Gaussian) prior imposes a global ℓ_2 penalty. For each coefficient β_j we set priors,

$$\beta_j \;\sim\; \mathcal{N}\big(0,\, 1\big), \quad j = 1, \ldots, d^2p,$$

so that most coefficients are moderately shrunk towards zero. This parallels the frequentist ridge penalty, and one can include an additional scale factor if stronger or weaker global shrinkage is desired depending on the data set.

Horseshoe Prior. The horseshoe prior (Carvalho et al., 2010b; Makalic and Schmidt, 2016) introduces more adaptive shrinkage via a local-global hierarchy. Each β_j is modeled as $\beta_j = B_{\text{raw},j}\,\lambda_j\,\tau$, where $B_{\text{raw},j} \sim \mathcal{N}(0,1)$, $\lambda_j \sim \text{C}^+(0,1)$ (local scale), and $\tau \sim \text{C}^+(0,1)$ (global scale). Small coefficients are heavily shrunk by small local scales, while the heavy-tailed Cauchy priors allow some large signals to remain.

Bayesian Lasso Prior. Finally, the Bayesian lasso (Park and Casella, 2008) imposes a Laplace (double-exponential) prior,

$$\beta_j|\eta \;\sim\; \text{Laplace}(0,\eta),$$

which corresponds to an ℓ_1 penalty in a frequentist setting. As with the horseshoe, this encourages coefficients to be near zero, possibly leading to sparsity in $\boldsymbol{\beta}$.

In all three cases, the covariance matrix Σ_ε is handled by the same LKJ-based prior, thus capturing potential correlations in the innovation process. The posterior distribution factors as

$$\pi\big(\mathbf{B}, \boldsymbol{\Sigma}_\varepsilon \mid \mathbf{y_t}\big) \;\propto\; \ell\big(\mathbf{y_t} \mid \mathbf{B}, \boldsymbol{\Sigma}_\varepsilon\big)\,\pi(\mathbf{B})\,\pi(\boldsymbol{\Sigma}_\varepsilon),$$

where ℓ is the Gaussian likelihood induced by the VAR model, and $\pi(\mathbf{B})$, $\pi(\Sigma_\varepsilon)$ encode the chosen shrinkage and covariance priors, respectively. By jointly estimating $\mathbf{B}$ and Σ_ε, this framework avoids the assumption of uncorrelated errors and allows us to examine how different shrinkage priors influence coefficient estimation in a fully multivariate setting.

5.2.2 Frequentist Methods

Ridge Regression. Classical ridge regression for a VAR(p) solves

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{d^2p}} \sum_{t=p+1}^{T} \left\| \mathbf{y}_t - \sum_{i=1}^{p} \boldsymbol{A}_i \, \mathbf{y}_{t-i} \right\|^2 + \lambda \, \|\mathbf{B}\|_2^2 , \tag{5.3}$$

where $\mathbf{B}$ is just the vectorized collection of $\{\boldsymbol{A}_i\}$. We set the regularization parameter $\lambda = .1$ in our analysis.

Nonparametric Shrinkage (NS). We use a James–Stein-like shrinkage approach for VAR coefficients (Giannone et al., 2015; Del Negro and Giannoni, 2015), implemented in `R` via `VARshrink` with `method="ns"`. Instead of explicitly solving (5.3), the NS method estimates the necessary sample covariances of $\mathbf{y}_{t-i}, \mathbf{y}_t$ and then applies a closed-form shrinkage rule to these covariance estimates, thereby producing a shrunk solution for $\mathbf{B}$. In this paper, we rely on the default choice for the shrinkage parameter, which `VARshrink` selects via a moment-based (empirical Bayes) formula akin to Stein's unbiased risk estimate.

5.3 Simulation Studies

5.3.1 Data-Generating Processes

We design three VAR data scenarios to examine both low- and high-dimensional settings and the effects of overfitting the lag order. Each dataset is simulated with a 50-observation burn-in (discarded), followed by $T = 200$ observations. Of these 200, the first 180 are the training set used for estimation, and the remaining 20 are held out for forecasting. Each scenario is replicated $N = 50$ times with independent parameter draws of β_j for each simulation.

Scenario 1 (low dimension, overfit lag). We generate a *3-dimensional* VAR(1) process here but intentionally *fit* a VAR(4). To form the 3×3 transition matrix A, each of its 9 entries

is set to zero with probability 0.7 (so on average across simulations, 3 are nonzero) and the nonzeros are drawn uniformly from $(-0.4, 0.4)$. We then rescale A so its largest eigenvalue in modulus is strictly below 1 by dividing by 1.1 times the maximum absolute eigenvalue, and set the noise covariance to $\Sigma = 0.05\, I_3$. As a result, there are 4×3^2 parameters to estimate in $\mathbf{B}$, even though only 3 of them are non-zero on average.

Scenario 2 (high dimension, correct lag). We generate a *20-dimensional* VAR(1) process and also fit a VAR(1). The 20×20 matrix A is constructed where each of its 400 entries is zero with probability 0.7 (so roughly 120 are nonzero), with nonzeros drawn randomly from $(-0.4, 0.4)$. Afterward, we ensure stationarity by dividing by 1.1 times the maximum absolute eigenvalue. We set $\Sigma = 0.1\, I_{20}$.

Scenario 3 (high dimension, overfit lag). We use the same data-generating scheme as in Scenario 2 (thus a sparse VAR(1) in 20 dimensions), but this time *fit* a VAR(4). As a result, there are $4 \times 20^2 = 1600$ parameters to estimate, even though only 120 of them are truly nonzero on average across simulations.

5.3.2 Estimation Methods

Design Matrix Setup. To estimate a VAR(p) in a linear regression framework, we arrange lagged responses into a design matrix $\mathbf{X} \in \mathbb{R}^{(T_{\text{train}}-p)\times(d\,p)}$. For $t = p+1, \ldots, T_{\text{train}}$, the t-th row of $\mathbf{X}$ (denoted $\mathbf{X}_t^\top$) is formed by horizontally concatenating the transposes of the p lagged column vectors

$$\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_{t-p} \quad (\text{each } d \times 1).$$

Hence, each row of $\mathbf{X}$ is a $1 \times (d\,p)$ vector. Likewise, the t-th row of the response matrix $\mathbf{Y} \in \mathbb{R}^{(T_{\text{train}}-p)\times d}$ is simply the transpose $\mathbf{y}_t^\top \in \mathbb{R}^{1\times d}$. Once $\mathbf{X}$ and $\mathbf{Y}$ are formed, any penalized or Bayesian regression method can be applied directly, and the estimated coefficient matrix is then reshaped to match $\mathbf{B} \in \mathbb{R}^{d\times(d\,p)}$ as defined in Section 5.2.

Frequentist fits and block-bootstrapped standard errors. We estimate the VAR coefficients in Ridge (`glmnet` with $\alpha = 0$) and NS (`VARshrink` with `method="ns"`) by penalized least squares, and then obtain empirical standard errors via a *block bootstrap* to better respect local time dependence. Specifically we

1. **Partition into blocks:** We group the training data $\{X, Y\}$ into non-overlapping consecutive blocks of length 4. We choose a block size of 4 because our VAR models use up to 4 lags, so each block captures the short-range autocorrelation structure of interest.

2. **Sample blocks with replacement:** To form a bootstrap dataset of the same size as the original, we randomly select blocks *with replacement* until we have at least T_{train} observations in total.

3. **Refit the model:** We refit the Ridge or NS model on this resampled dataset and record the estimated coefficients.

4. **Repeat and aggregate:** Steps (2)–(3) are repeated for 30 bootstrap replications (`n_boot = 30`). The empirical standard error for each coefficient is then taken to be the sample standard deviation of its estimates across these replications.

This procedure retains within-block autocorrelations (up to 4 lags) while randomly mixing which blocks are selected, preserving important time-series structure better than naive row-wise (i.i.d.) resampling. As a result, the resulting intervals yield more realistic coverage for dependent data.

Bayesian fits. We fit the three Bayesian models by calling `Stan` with 4 parallel Markov chains, each run for 2000 total iterations (the first 500 of which are warm-up). We fix `seed=123` for reproducibility and use $\{$`adapt_delta` $= 0.9$, `max_treedepth` $= 12\}$. In Stan's Hamiltonian Monte Carlo (HMC) framework, `adapt_delta` is the target acceptance probability, and increasing it to 0.9 aims for smaller step sizes and more conservative sampling.

The `max_treedepth` parameter caps the depth of the binary tree in each iteration's leapfrog integrator, preventing extremely long trajectories.

For each chain, we obtain posterior draws of the coefficient vector $\boldsymbol{\beta}$. We summarize each coefficient by its posterior mean and 95% central credible interval (2.5% and 97.5% quantiles).

5.3.3 Performance Metrics

We evaluate each method along two dimensions: *parameter estimation* and *forecast* performance.

Parameter estimation. Let $\boldsymbol{\beta}_{\text{true}}$ denote the true parameters in $\mathbf{B}$. Each frequentist method (Ridge or Nonparametric Shrinkage) estimates $\boldsymbol{\beta}$ by minimizing a penalized least squares criterion, whereas each Bayesian method (Normal/Ridge, Lasso, Horseshoe) uses the posterior mean from MCMC samples as $\widehat{\boldsymbol{\beta}}$. We then compute the root mean squared error (RMSE),

$$\text{RMSE} \;=\; \sqrt{\frac{1}{d^2p}\sum_{j=1}^{d^2p}\big(\widehat{\beta}_j - \beta_{j,\text{true}}\big)^2},$$

to measure how closely $\widehat{\boldsymbol{\beta}}$ matches $\boldsymbol{\beta}_{\text{true}}$.

Next, we construct 95% intervals for each coefficient by applying a block bootstrap to estimate standard errors and forming approximate normal intervals of the form $\widehat{\beta}_j \pm z_{0.975}\,\text{SE}_j$ for the frequentist approaches, whereas for the Bayesian methods we use the 2.5% and 97.5% posterior quantiles from the MCMC samples. We record the *empirical coverage* (the percentage of intervals that contain the true value) and the *average interval length* to assess how well each approach quantifies uncertainty.

Forecasting performance. To assess predictive accuracy, we reserve the final 20 observations as a test set. Each method then produces sequential one-step-ahead forecasts by estimating $\mathbf{y}_{t+1}$ at time t based on all data up to $\mathbf{y}_t$, avoiding the accumulation of multi-step

errors. We compute the average root mean squared forecast error (RMSE)

$$\text{Forecast RMSE} \;=\; \sqrt{\frac{1}{20\,d}\sum_{t=T_{\text{train}}+1}^{T_{\text{train}}+20} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2},$$

where $\hat{\mathbf{y}}_t$ is the forecast at time t. After 50 replications per scenario, we summarize the average forecasting RMSE and coverage to compare each method's predictive capabilities.

5.3.4 Simulation Results

The results from the three simulation studies are shown in tables 5.1–5.3 and figures 5.1–5.4. The percentage of replications in which each method has the lowest forecast RMSE or parameter RMSE is shown in table 5.4.

Scenario 1 (Low-Dimension, Overfit Lag)

Forecasting The top block of Table 5.1 has each method's mean forecast RMSE. Horseshoe has the smallest value (0.211), followed by *ns* and Ridge (0.213), Lasso (0.214), and Normal (0.215). Table 5.4 has the proportion of replications in which each method has the best forecast: Horseshoe leads with 60%, *ns* has 20%, Normal 10%, Ridge 6%, and Lasso 4%.

Parameter Estimation Horseshoe has the lowest overall parameter RMSE (0.0434) and exceeds average nominal coverage (97.2%), with intervals about 8% shorter than those of the next-best method. Lasso (0.0803) and Normal (0.0838) have higher RMSEs but maintain coverage near 94–95%. Both *ns* (0.0693) and Ridge (0.0730) occupy a middle tier; *ns* has coverage of 85.7% but yields narrower intervals (mean length 0.204). Horseshoe has the best parameter RMSE in all replications (100%), giving strong shrinkage without sacrificing coverage.

Scenario 2 (High-Dimension, Correct Lag)

Forecasting All methods have similar forecasting accuracy (middle block of Table 5.1). Horseshoe has the smallest mean forecast RMSE (0.325), followed by Lasso (0.326) and Normal, *ns*, and Ridge (0.327). Horseshoe is the top forecaster in 48% of replications, Lasso and *ns* each in 20%, Ridge in 10%, and Normal in 2% (Table 5.4).

Parameter Estimation

Horseshoe again has the lowest parameter RMSE (0.0536). Lasso, Normal, *ns*, and Ridge cluster between 0.0568 and 0.0598. Coverage remains high (94–95%) for Horseshoe, Lasso, Normal, and *ns*, but dips to 84% for Ridge. Table 5.2 has results for zero coefficients, where Horseshoe has an RMSE of 0.0357 and 99.0% coverage. The *ns* method handles zero parameters well but sometimes underperforms on nonzeros. Horseshoe has a nonzero RMSE of 0.0596, higher than Lasso and Normal (0.0571–0.0576), while maintaining overall coverage of 92.9%. Horseshoe has the best parameter RMSE in 90% of replications, followed by Ridge in 10% (Table 5.4).

Scenario 3 (High-Dimension, Overfit Lag)

Forecasting In the bottom block of Table 5.1, Horseshoe has the smallest mean forecast RMSE (0.342), followed by *ns* and Ridge (0.365–0.366), Lasso (0.404), and Normal (0.418). Horseshoe is the top forecaster in 100% of replications (Table 5.4), indicating a strong ability to handle overfitting.

Parameter Estimation Horseshoe has the lowest parameter RMSE (0.0394), with better-than-nominal average coverage (97.5%) and intervals about 6% shorter than those of the next-best method. Lasso (0.104) and Normal (0.117) have higher RMSEs but maintain nominal coverage (95–96%) through wider intervals (0.432–0.464). Both *ns* and Ridge have moderate RMSEs (0.0619–0.0635) but show lower coverage (88.2% and 84.5%) and narrower

intervals (0.18–0.20). Horseshoe remains the top performer in parameter RMSE for 100% of the replications.

Overall, Horseshoe consistently has excellent forecast accuracy and parameter recovery, including the lowest RMSE, high coverage, and moderate interval lengths. Ridge occasionally has strong forecasts but frequently undercovers in higher dimensions. Lasso and Normal have intermediate performance for both forecasting and parameter estimation, ensuring reasonable coverage by using somewhat larger intervals. The *ns* approach is computationally efficient and sometimes has precise point estimates, but coverage can be volatile due to overly narrow intervals. These findings reinforce the advantages of local-global shrinkage (Horseshoe) in moderate- and high-dimensional VAR contexts, especially when the lag order is inflated.

5.4 Data Analysis

We illustrate our methods on the `Canada` dataset from the `R` package **vars** (Pfaff, 2008), which provides quarterly macroeconomic observations on four Canadian variables spanning $T = 84$ quarters (1980Q1–2000Q4): employment (e, in log-index form), productivity ($prod$, in log-index form measuring labor productivity), real wages (rw, in log-index form), and the unemployment rate (U, in percent). Economic considerations suggest these variables are jointly dependent, making a Vector Auto Regression (VAR)-based approach suitable.

Differencing and Stationarity. To reduce nonstationarity, we difference each series once

$$\Delta \mathbf{y}_t \;=\; \mathbf{y}_t \;-\; \mathbf{y}_{t-1} \quad (t = 2, \ldots, T).$$

We then estimate the VAR on these differenced observations. To obtain forecasts on the original scale, we *invert* the differencing by recursively summing the predicted differences

$$\widehat{\mathbf{y}}_{T+1} \;=\; \mathbf{y}_T \;+\; \widehat{\Delta \mathbf{y}}_{T+1}, \quad \widehat{\mathbf{y}}_{T+2} \;=\; \widehat{\mathbf{y}}_{T+1} \;+\; \widehat{\Delta \mathbf{y}}_{T+2}, \quad \text{etc.}$$

This ensures the final forecasts reflect the original scale of the data, while the differencing step helps achieve stationarity and limit spurious trends when fitting the VAR.

Lag Orders from $p = 1$ to $p = 12$. We fit VAR(p) models for $p \in \{1, 2, \ldots, 12\}$ to assess how each shrinkage method adapts to varying lag choices. Specifically, we train each VAR on the first 80 quarters of the differenced data, then produce rolling 1-step-ahead forecasts for quarters 81–84 (the holdout set). After each forecast, we *roll forward* by incorporating the newly observed actual data. We measure forecast accuracy in the original scale by computing both root mean squared error (RMSE), $\text{RMSE} = \sqrt{\frac{1}{4}\sum_{t=81}^{84}(y_t - \hat{y}_t)^2}$ and mean absolute percentage error (MAPE), $\text{MAPE} = \frac{100}{4}\sum_{t=81}^{84}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$ over the 4 test observations.

5.4.1 Results

Figure 5.5 (left) shows the original Canadian macroeconomic variables, while the right panel displays their first-differenced versions used for estimation.

Aggregate Performance ($p = 1, \ldots, 12$)

Each method's forecast accuracy (RMSE, MAPE) for varying lag orders (p) is shown in Figures 5.6 and 5.7. Table 5.5 reports the mean and standard deviation of these metrics across all 12 fitted VAR(p) models and all four series.

Overall, *Horseshoe* achieves the most consistent and accurate forecasts, attaining the lowest mean RMSE (0.51). *Ridge* (RMSE = 0.56) and *NS* (RMSE = 0.60) provide somewhat intermediate performance, while *Lasso* (RMSE = 0.60) and *Normal* (RMSE = 0.63) tend to yield slightly larger prediction errors. In terms of variability, Horseshoe's standard deviation of RMSE (0.22) is comparable to that of Lasso and Ridge, whereas Normal exhibits the largest SD (0.26).

For MAPE, Horseshoe again stands out with an average of 0.71%, followed by NS (0.97%), Ridge (1.06%), Lasso (1.24%), and Normal (1.37%). This pattern suggests that Horseshoe

effectively suppresses many small coefficients without overshrinking the larger signals, yielding robust relative-error forecasts even as p grows. By contrast, Normal's wider prior and Lasso's strong ℓ_1 shrinkage can lead to higher MAPE in certain lag settings (see Figure 5.7), and NS remains susceptible to moderate forecast deterioration for larger p.

Shrinkage patterns for a few representative lag orders ($p = 3, 6, 9, 12$) appear in Figure 5.8. Horseshoe and Lasso consistently display heavy shrinkage toward zero for small coefficients, Normal has moderate Gaussian-like shrinkage, NS exhibits a broader coefficient spread, and Ridge pulls estimates closer to zero but never to an exact zero. Across the different p values, Horseshoe's local-global prior structure appears to adapt more flexibly, resulting in better overall forecasting metrics.

Case Study: $p = 11$

To illustrate performance at a higher lag, we examine the VAR(11) specification, which delivers the lowest average forecasting errors among the tested orders. Table 5.6 shows each method's RMSE and MAPE on the final four quarters of the holdout set. Notably, *Horseshoe* again achieves the best performance on both measures (RMSE = 0.51, MAPE = 0.60%). The next closest method is *Ridge* (RMSE = 0.61, MAPE = 1.66%) and *NS* (RMSE = 0.65, MAPE = 1.26%), while Lasso and Normal both exhibit slightly higher errors (RMSE = 0.70–0.78, MAPE = 1.66–1.81%). This example highlights Horseshoe's ability to preserve large coefficients and aggressively shrink small ones, maintaining strong predictive accuracy even at high lag orders.

The 1-step-ahead forecasts of the observed data across the four holdout quarters are shown in figure 5.9. Horseshoe, ns, and Ridge all track the actual values fairly closely. Lasso and Normal lag behind somewhat. Overall, these VAR(11) results echo our broader simulation findings: Horseshoe's adaptive local-global prior can maintain strong performance at high lag orders, and Ridge remains reasonably robust as well, whereas ns, Lasso, and Normal can become less accurate or more variable depending on the specific error metric.

5.5 Discussion

Our simulation results lead to several important takeaways about shrinkage estimation in VAR models under varying dimensionality and lag orders. First, the Horseshoe prior stands out for consistently achieving the lowest parameter RMSE and near-nominal coverage, particularly in the most challenging high-dimensional or overfit scenarios. This local-global prior structure successfully suppresses small coefficients while preserving truly large effects, thereby producing stable estimates and competitive forecasts across the board. By contrast, Lasso and Normal priors often deliver mid-range forecast accuracy and parameter estimation, but they maintain coverage near or above the 95% target, albeit with wider intervals in some cases.

Ridge regression remains effective for forecasting in low- to moderate-dimensional scenarios (where the ratio of parameters to observations is not excessively large), frequently ranking second or third in terms of forecast RMSE. However, it underestimates parameter uncertainty in high-dimensional settings e.g., when the dimension-to-sample-size ratio is particularly large, leading to undercoverage. Similarly, Nonparametric Shrinkage (ns) provides very short intervals and can achieve strong point forecasts, but it exhibits markedly low coverage in the same high-dimensional regimes, suggesting that its narrower intervals are overly optimistic about uncertainty in heavily over-parameterized models.

In the Canadian macro-economic data application, similar patterns emerge: Horseshoe and Ridge each exhibit strong one-step-ahead forecast accuracy, while Lasso, Normal, and ns occasionally lag behind, particularly when the model order is large. Overall, these findings reinforce the benefits of using local-global shrinkage to adapt to large model spaces, especially for practitioners seeking reliable inference and coverage. Frequentist options like Ridge can still perform competitively in lower-dimensional or less overfit settings but risk severe undercoverage when the parameter space grows.

Taken together, these results underscore that when parameter interpretation and interval validity are paramount, Horseshoe or other local-global Bayesian priors are well-suited to

handle high-dimensional or inflated-lag VAR models. If short-term predictive performance alone is the principal goal, Ridge can remain attractive, provided one is willing to accept somewhat lower coverage in complex settings. The Lasso and Normal priors offer middle-ground alternatives, balancing coverage and moderate forecasting performance without fully matching Horseshoe's combination of shrinkage strength and coverage reliability.

Code Availability

All `R` scripts and `Stan` model files used in this study are publicly available at https://github.com/harrisonekatz/BayesVAR-SimStudy. In particular, the main simulation script `var_three_sim_script.R` (which orchestrates data generation, frequentist and Bayesian estimation, and result collation) may be found in the repository's `R/` directory. The repository also includes each of the `Stan` model files (`var_normal.stan`, `var_lasso.stan`, `var_horseshoe.stan`), along with examples illustrating their usage. All results and figures in this manuscript can be reproduced by running the scripts found in that repository.

Figures and tables

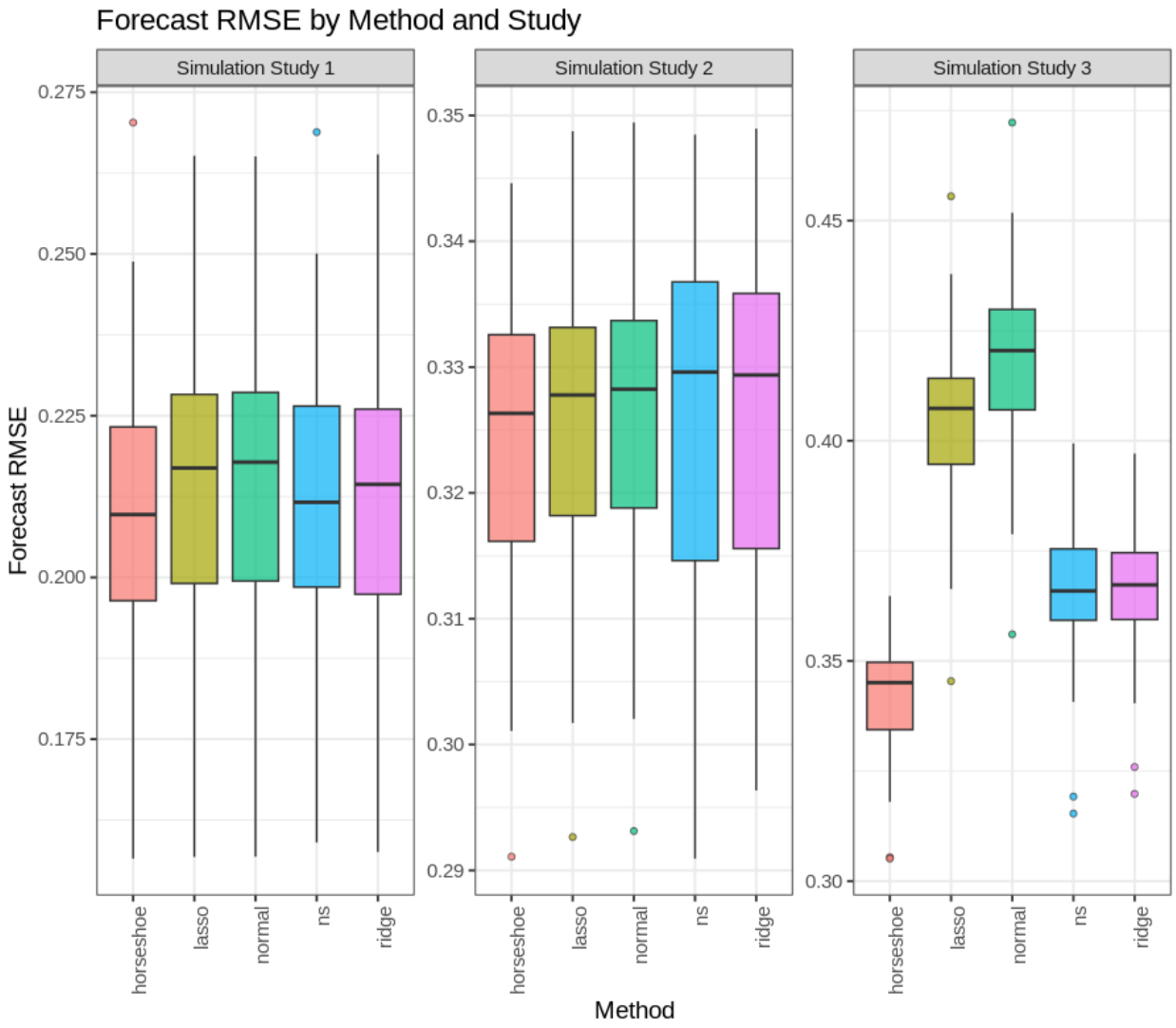

Figure 5.1: **Forecast RMSE by Method and Study (All Coefficients).** Boxplots reflect the distribution of one-step-ahead RMSE across the 50 replications. Horseshoe achieves or ties for the lowest forecast error, especially in the high-dimension overfit scenario (Study 3).

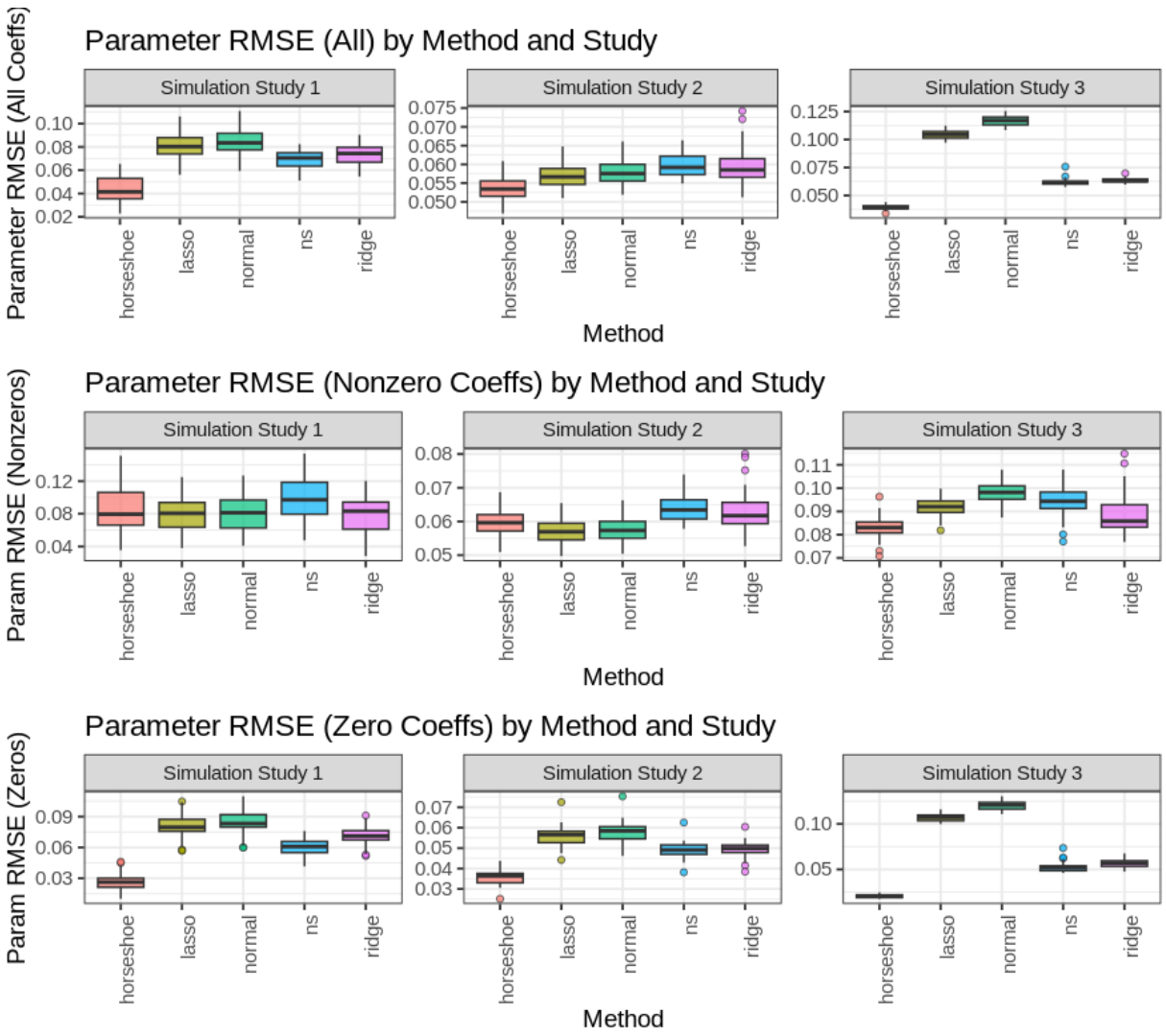

Figure 5.2: **Parameter RMSE by Method and Study.** Horseshoe is consistently lowest in overall parameter RMSE, while NonparamShrink (*ns*) occasionally performs well but can exhibit greater variance or undercoverage.

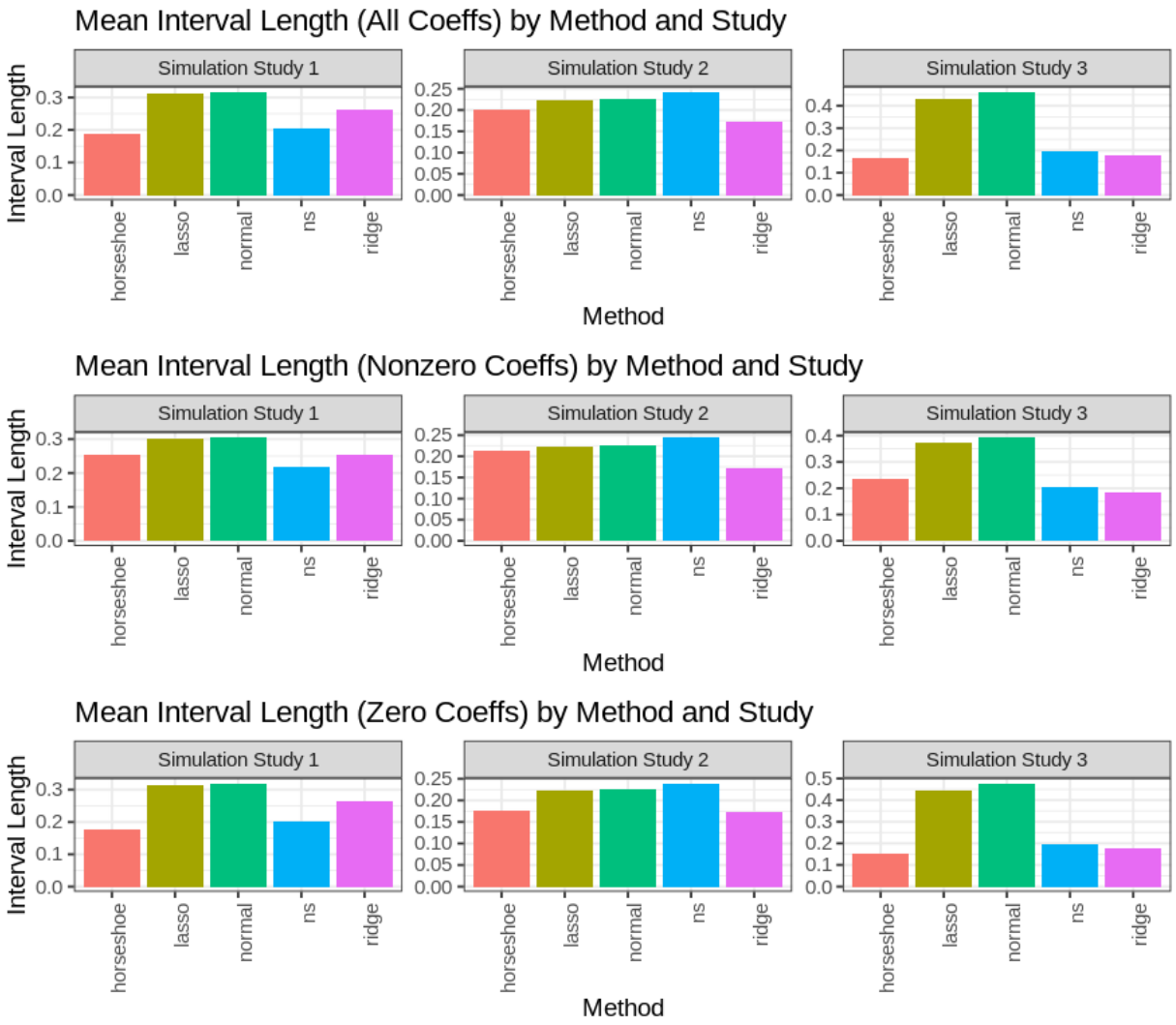

Figure 5.3: **Mean Interval Length by Method and Study.** Shorter intervals may indicate overconfidence if coverage is below the nominal 95%; for instance, *ns* has narrower intervals but lower coverage in some scenarios.

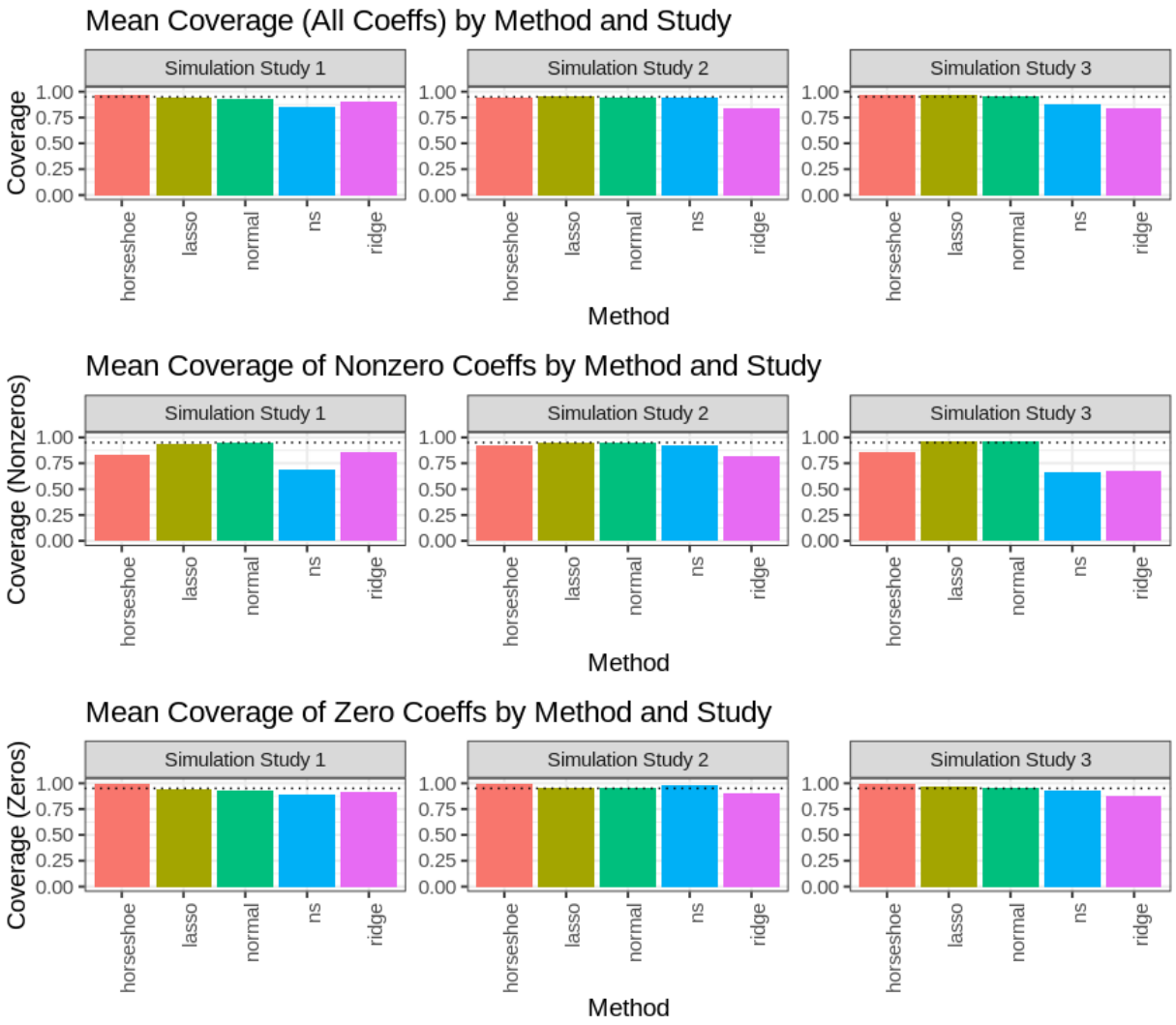

Figure 5.4: **Coverage by Method and Study.** A dotted line at 0.95 indicates the nominal coverage target. Horseshoe, Lasso, and Normal usually achieve near 95%, while *ns* and Ridge can dip below this level for high-dimensional or overfit scenarios.

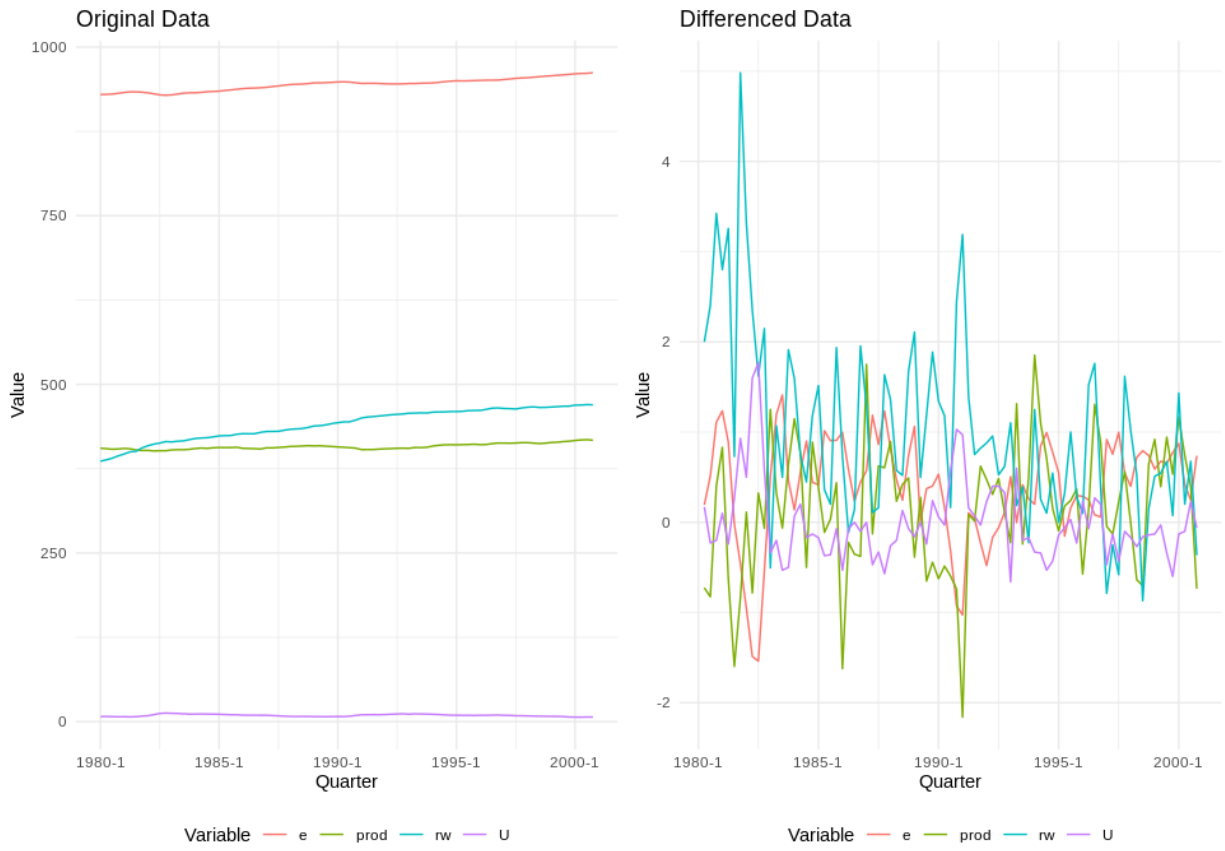

Figure 5.5: The four Canadian macroeconomic variables in their original form (left) and once-differenced (right). Differencing helps remove trends and stabilize the series prior to VAR model estimation.

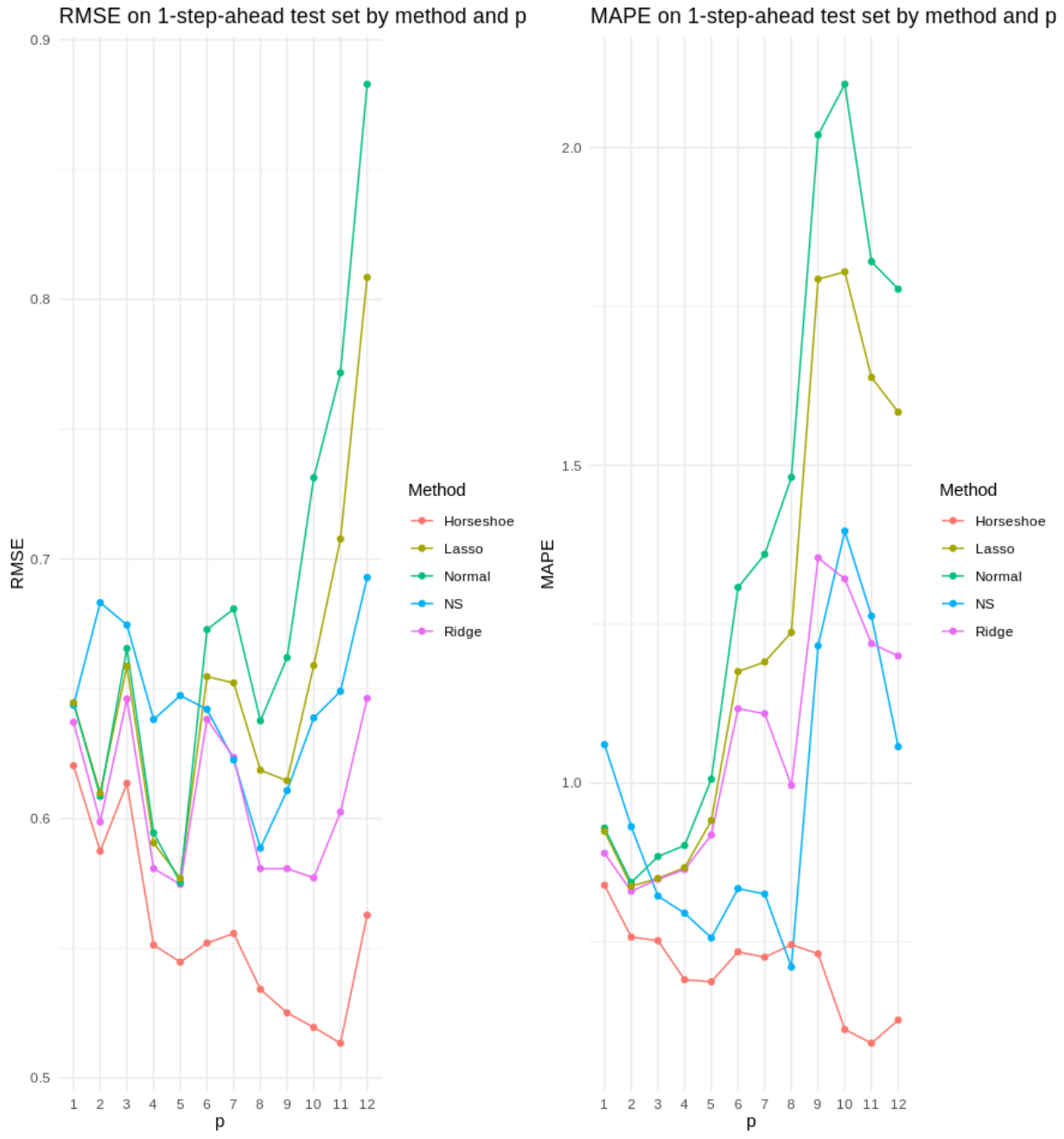

Figure 5.6: **Forecast performance on Canadian data with VAR(p), $p = 1, \ldots, 12$.** *Left:* Out-of-sample RMSE. *Right:* Out-of-sample MAPE (%). Methods evaluated: Horseshoe, Lasso, ns, Normal, and Ridge. Horseshoe and Ridge consistently achieve lower RMSEs, while Lasso and Normal show moderate performance. ns experiences higher errors, especially at larger p.

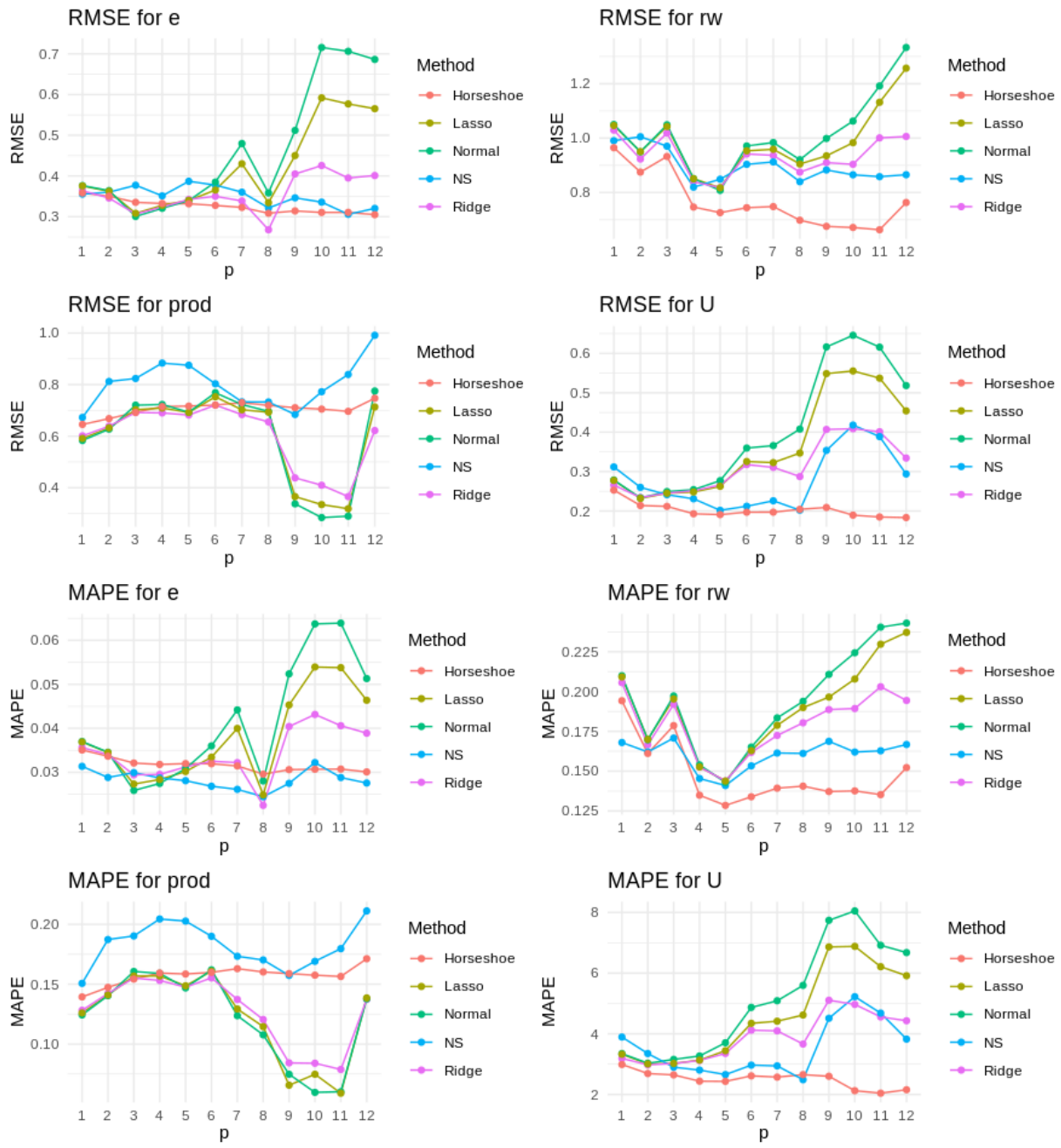

Figure 5.7: Out-of-sample forecasting accuracy for each Canadian macroeconomic variable—employment (e), real wages (rw), productivity ($prod$), and unemployment (U)—across increasing VAR orders ($p = 1, \ldots, 12$). The top row shows the Root Mean Squared Error (RMSE), and the bottom row shows the Mean Absolute Percentage Error (MAPE). Each colored line corresponds to one of five shrinkage methods (Horseshoe, Lasso, Normal, ns, and Ridge). Overall, Horseshoe achieves the lowest MAPE and exhibits relatively stable performance as p increases. Lasso and Ridge perform well but show greater variability at higher orders.

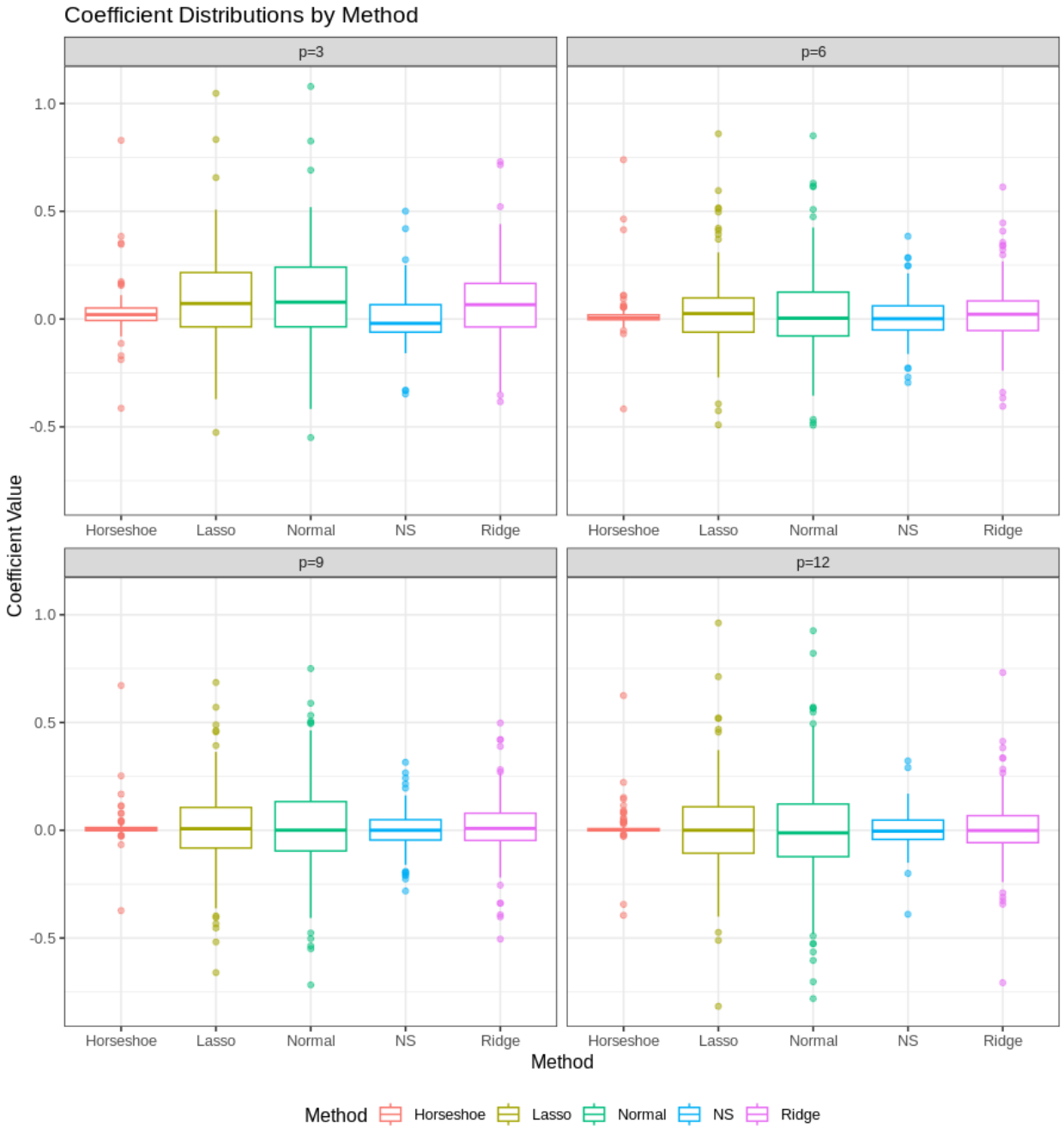

Figure 5.8: Distribution of estimated VAR coefficients by method. Each point represents one of the $(4 \times 4 \times p)$ parameters, highlighting the degree of shrinkage for each prior.

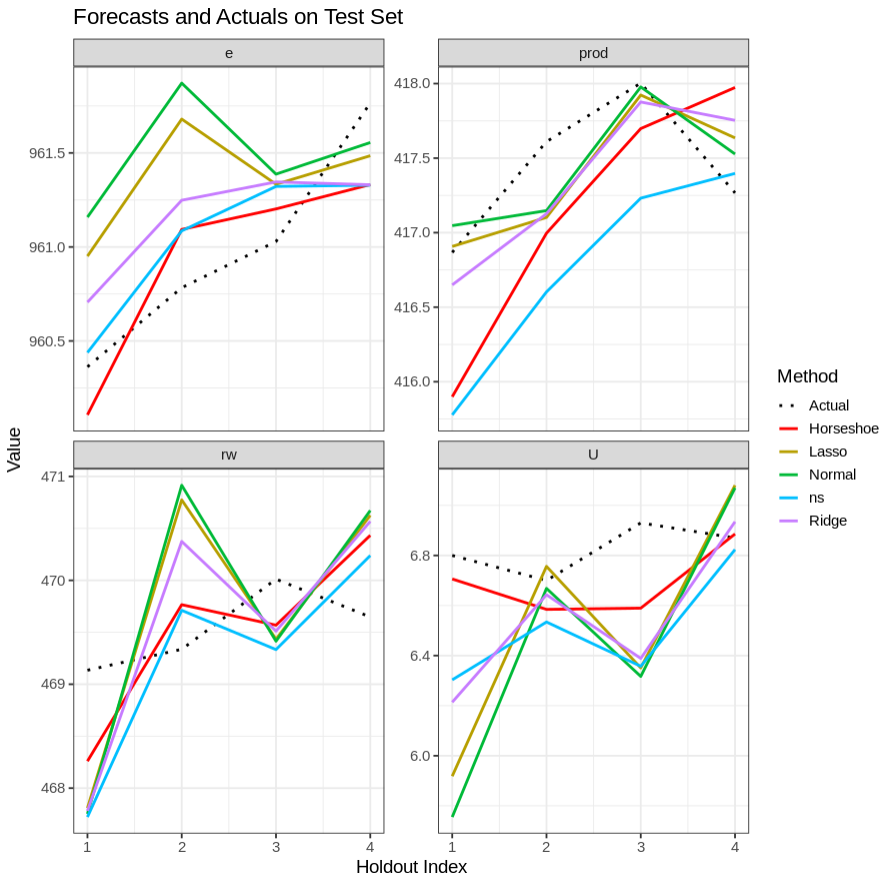

Figure 5.9: Out-of-sample 1 step ahead forecasts from VAR(11) and actuals for each Canadian macroeconomic variable—employment (e), real wages (rw), productivity ($prod$), and unemployment (U).

Table 5.1: **Overall Performance (All Coefficients).** Mean and standard deviation (SD) of forecast RMSE (FRMSE), mean and SD of parameter RMSE (PRMSE), mean coverage (Cov), and mean interval length (Int. Length) across the three studies.

Scenario	**Method**	**FRMSE Mean**	**FRMSE SD**	**PRMSE Mean**	**PRMSE SD**	**Cov**	**Int. Length**
				Study 1			
	Horseshoe	0.211	0.0208	0.0434	0.0105	0.972	0.189
	Lasso	0.214	0.0217	0.0803	0.0111	0.943	0.311
Study 1	Normal	0.215	0.0218	0.0838	0.0115	0.936	0.318
	ns	0.213	0.0212	0.0693	0.00781	0.857	0.204
	Ridge	0.213	0.0215	0.0730	0.00897	0.904	0.261
				Study 2			
	Horseshoe	0.325	0.0122	0.0536	0.00330	0.947	0.202
	Lasso	0.326	0.0124	0.0568	0.00319	0.951	0.223
Study 2	Normal	0.327	0.0125	0.0577	0.00326	0.949	0.225
	ns	0.327	0.0132	0.0598	0.00319	0.944	0.243
	Ridge	0.327	0.0129	0.0591	0.00487	0.840	0.173
				Study 3			
	Horseshoe	0.342	0.0132	0.0394	0.00229	0.975	0.167
	Lasso	0.404	0.0194	0.104	0.00368	0.963	0.432
Study 3	Normal	0.418	0.0208	0.117	0.00446	0.955	0.464
	ns	0.365	0.0164	0.0619	0.00291	0.882	0.197
	Ridge	0.366	0.0153	0.0635	0.00206	0.845	0.181

Table 5.2: **Performance on Zero Coefficients Only.** Mean and SD of parameter RMSE (PRMSE) for zero coefficients, mean coverage (Cov), and mean interval length (Len).

Scenario	**Method**	**PRMSE Mean**	**PRMSE SD**	**Cov**	**Len**
		Study 1			
Study 1	Horseshoe	0.0266	0.0082	0.999	0.176
	Lasso	0.0798	0.0119	0.945	0.313
	Normal	0.0839	0.0124	0.935	0.320
	ns	0.0606	0.0086	0.890	0.201
	Ridge	0.0711	0.0096	0.912	0.262
		Study 2			
Study 2	Horseshoe	0.0357	0.0034	0.990	0.177
	Lasso	0.0558	0.0047	0.956	0.221
	Normal	0.0578	0.0049	0.950	0.225
	ns	0.0489	0.0039	0.978	0.240
	Ridge	0.0495	0.0037	0.904	0.173
		Study 3			
Study 3	Horseshoe	0.020	0.002	1.000	0.152
	Lasso	0.107	0.004	0.965	0.445
	Normal	0.120	0.005	0.955	0.478
	ns	0.052	0.005	0.928	0.196
	Ridge	0.056	0.004	0.882	0.180

Table 5.3: **Performance on Nonzero Coefficients Only.** Mean and SD of parameter RMSE (PRMSE) for nonzero coefficients, mean coverage (Cov), and mean interval length (Len).

Scenario	**Method**	**PRMSE Mean**	**PRMSE SD**	**Cov**	**Len**
		Study 1			
Study 1	Horseshoe	0.085	0.027	0.84	0.26
	Lasso	0.079	0.022	0.94	0.30
	Normal	0.080	0.023	0.94	0.31
	ns	0.098	0.028	0.70	0.22
	Ridge	0.078	0.023	0.86	0.25
		Study 2			
Study 2	Horseshoe	0.0596	0.00391	0.929	0.213
	Lasso	0.0571	0.00327	0.949	0.224
	Normal	0.0576	0.00329	0.949	0.225
	ns	0.0639	0.00396	0.929	0.245
	Ridge	0.0626	0.00587	0.814	0.173
		Study 3			
Study 3	Horseshoe	0.083	0.0049	0.86	0.24
	Lasso	0.092	0.0040	0.96	0.37
	Normal	0.098	0.0044	0.96	0.39
	ns	0.094	0.0063	0.67	0.20
	Ridge	0.088	0.0082	0.67	0.18

Table 5.4: **Times Each Method Is "Best" in Forecast or Parameter RMSE.** For each scenario and replication (50 total), we identify which method attains the lowest forecast RMSE or lowest parameter RMSE. Columns show the percentage of replications in which each method is best.

Best in Forecast RMSE		
Scenario	**Method**	**% of Replications**
Study 1	Horseshoe	60%
	ns	20%
	Normal	10%
	Ridge	6%
	Lasso	4%
Study 2	Horseshoe	48%
	Lasso	20%
	ns	20%
	Ridge	10%
	Normal	2%
Study 3	Horseshoe	100%

Best in Parameter RMSE (All Coefficients)		
Scenario	**Method**	**% of Replications**
Study 1	Horseshoe	100%
Study 2	Horseshoe	90%
	Ridge	10%
Study 3	Horseshoe	100%

Method	**Mean RMSE**	**SD RMSE**	**Mean MAPE (%)**	**SD MAPE**
Horseshoe	0.51	0.22	0.71	0.94
Lasso	0.60	0.25	1.24	1.87
NS	0.60	0.25	0.97	1.39
Normal	0.63	0.26	1.37	2.12
Ridge	0.56	0.24	1.06	1.51

Table 5.5: **Forecast error summaries for VAR(p), $p = 1, \ldots, 12$.** Shown are the mean and standard deviation of RMSE and MAPE (%) across the 12 lag choices. Horseshoe achieves the smallest mean RMSE and MAPE, while Normal exhibits the largest mean RMSE and MAPE. Ridge, NS, and Lasso provide intermediate performance.

Method	**RMSE**	**MAPE (%)**
Horseshoe	0.51	0.60
Lasso	0.70	1.66
Normal	0.78	1.81
ns	0.65	1.26
Ridge	0.61	1.66

Table 5.6: Forecasting accuracy on the Canada data for the VAR(11) model, evaluated on the final four observations. Lower RMSE and MAPE values indicate better performance.

Bibliography

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological) 44*(2), 139–160.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

AL-Dhurafi, N. A., N. Masseran, and Z. H. Zamzuri (2018). Compositional time series analysis for air pollution index data. *Stochastic Environmental Research and Risk Assessment 32*(10), 2903–2911.

Aprigliano, V. (2020). A large bayesian var with a block-specific shrinkage: A forecasting application for italian industrial production. *Journal of Forecasting 39*(8), 1291–1304.

Bai, Y., A. Carriero, T. E. Clark, and M. Marcellino (2022). Macroeconomic forecasting in a multi-country context. *Journal of Applied Econometrics 37*(6), 1230–1255.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Barber, D., A. T. Cemgil, and S. Chiappa (2011). *Bayesian Time Series Models*. Cambridge University Press.

Barcelo-Vidal, C., L. Aguilar, and J. A. Martín-Fernández (2011). Compositional VARIMA time series. *Compositional Data Analysis: Theory and Applications*, 87–101.

Basu, S., A. Shojaie, and G. Michailidis (2019). Low dimensional representations for high dimensional vector autoregressions with structured penalties. *Statistica Sinica 29*(4), 2099–2121.

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics 21*(1), 79–109.

Bańbura, M., D. Giannone, and M. Lenza (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association 98*(461), 214–223.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods*, pp. 15–22. Springer.

Bernardo, J. M. and A. F. Smith (2009). *Bayesian theory*, Volume 405. John Wiley & Sons.

Berry, L. R. and M. West (2020). Bayesian forecasting of many count-valued time series. *Journal of Business & Economic Statistics 38*(4), 872–887.

Bitto, A. and S. Frühwirth-Schnatter (2019a). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics 210*(1), 75–97.

Bitto, A. and S. Frühwirth-Schnatter (2019b). Sparse bayesian time-varying parameter models using the horseshoe prior. *Journal of Econometrics 210*(1), 75–97.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Boonen, T. J., M. Guillén, and M. Santolino (2019). Forecasting compositional risk allocations. *Insurance: Mathematics and Economics 84*, 79–86.

Brandt, P. T. and T. Sandler (2012). A Bayesian Poisson vector autoregression model. *Political Analysis 20*(3), 292–315.

Brunsdon, T. M. and T. Smith (1998). The time series analysis of compositional data. *Journal of Official Statistics 14*(3), 237.

Bürkner, P.-C., J. Gabry, and A. Vehtari (2020). Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation 90*(14), 2499–2523.

Cargnoni, C., P. Müller, and M. West (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association 92*, 640–647.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022). Corrigendum to: "practical bayesian var forecasting". *International Journal of Forecasting 38*(4), 1580–1582.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2009). Handling sparsity via the horseshoe. In *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, pp. 73–80. Proceedings of Machine Learning Research 5.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010a). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010b). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Casarin, R., S. Grassi, F. Ravazzolo, and H. K. van Dijk (2021). A bayesian dynamic compositional model for large density combinations in finance. Technical report, Tinbergen Institute Discussion Paper 21-016/III. 51 pages. Forthcoming journal publication.

Chan, J. C. (2020). *Large Bayesian vector autoregressions*. Springer.

Chan, J. C. (2021). Minnesota-type adaptive hierarchical priors for large bayesian vars. *International Journal of Forecasting 37*(3), 1212–1226.

Chen, C. W. and S. Lee (2016). Generalized Poisson autoregressive models for time series of counts. *Computational Statistics & Data Analysis 99*, 51–67.

Chen, M. and X. Chen (2021). Penalized methods for high-dimensional time series regression under heavy-tailed distributions. *Journal of Time Series Analysis 42*(1), 54–72.

Crump, R. K., S. Eusepi, D. Giannone, E. Qian, and A. M. Sbordone (2021, August). A large bayesian var of the united states economy. Technical Report 976, Federal Reserve Bank of New York. Available at SSRN: 10.2139/ssrn.3908154.

da Silva, C. Q., H. S. Migon, and L. T. Correia (2011). Dynamic Bayesian beta models. *Computational Statistics & Data Analysis 55*(6), 2074–2089.

da Silva, C. Q. and G. S. Rodrigues (2015). Bayesian dynamic Dirichlet models. *Communications in Statistics-Simulation and Computation 44*(3), 787–818.

Del Negro, M. and M. P. Giannoni (2015). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics 177*(2), 199–223.

Doan, T., R. B. Litterman, and C. A. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*(1), 1–100.

Dong, Z., H. L. Shang, F. Hui, and A. Bruhn (2025). A compositional approach to modelling cause-specific mortality with zero counts. *Annals of Actuarial Science*. Forthcoming.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelo-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical geology 35*(3), 279–300.

Engle, R. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research w8554*.

Follett, L. and C. Yu (2019). Achieving parsimony in bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics 11*, 130–144.

Fukumoto, K., A. Beger, and W. H. Moore (2019). Bayesian modeling for overdispersed event-count time series. *Behaviormetrika 46*(2), 435–452.

Gefang, D., G. Koop, and A. Poon (2023). Forecasting using variational bayesian inference in large vector autoregressions with hierarchical shrinkage. *International Journal of Forecasting 39*(1), 346–363.

Gelman, A. et al. (2020). *Bayesian Data Analysis* (4th ed.). CRC Press.

George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*(2), 339–373.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 436–451.

Greenacre, M. (2021). Compositional data analysis in practice. Book.

Griffin, J. E. and P. J. Brown (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis 12*(1), 135–159.

Griffin, J. E. and P. J. Brown (2021). Bayesian global-local shrinkage methods for regularisation in the high-dimensional linear model. *Chemometrics and Intelligent Laboratory Systems 210*, 104255.

Grunwald, G. K., A. E. Raftery, and P. Guttorp (1993). Time series of continuous proportions. *Journal of the Royal Statistical Society: Series B (Methodological) 55*(1), 103–116.

Huber, F. and M. Feldkircher (2017). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics 35*(4), 599–614.

Huber, F. and M. Feldkircher (2019a). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics 37*(1), 27–39.

Huber, F. and M. Feldkircher (2019b). Adaptive shrinkage in bayesian vector autoregressive models. *Journal of Business & Economic Statistics 37*(1), 27–39.

Huber, F. and G. Koop (2023). Subspace shrinkage in conjugate bayesian vector autoregressions. *Journal of Applied Econometrics 38*(4), 556–576.

Karlsson, S. (2013). Forecasting with Bayesian vector autoregression. *Handbook of Economic Forecasting 2*, 791–897.

Kastner, G. and F. Huber (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting 39*(7), 1142–1165.

Katz, H., K. T. Brusch, and R. E. Weiss (2024). A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times. *International Journal of Forecasting 40*(4), 1556–1567.

Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics 186*(2), 325–344.

Koehler, A. B., R. D. Snyder, J. K. Ord, A. Beaumont, et al. (2010). Forecasting compositional time series with exponential smoothing methods. Technical report, Monash University, Department of Econometrics and Business Statistics.

Koop, G. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics 28*(2), 177–203.

Koop, G. and D. Korobilis (2010). *Bayesian Multivariate Time Series Methods for Empirical Macroeconomics*. Now Publishers Inc.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198. Dynamic Econometric Modeling and Forecasting.

Korobilis, D. and D. Pettenuzzo (2019). Adaptive hierarchical priors for high-dimensional vector autoregressions. *Journal of Econometrics 212*(1), 241–271.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology 81*(4), 781–804.

Kynčlová, P., P. Filzmoser, and K. Hron (2015). Modeling compositional time series with vector autoregressive models. *Journal of Forecasting 34*(4), 303–314.

Lewandowski, D., D. Kurowicka, and H. Joe (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Lipsmeyer, C. S., A. Q. Philips, A. Rutherford, and G. D. Whitten (2019). Comparing dynamic pies: A strategy for modeling compositional variables in time and space. *Political Science Research and Methods 7*(3), 523–540.

Makalic, E. and D. F. Schmidt (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*(1), 179–184.

Makgai, S., A. Bekker, and M. Arashi (2021). Compositional data modeling through dirichlet innovations. *Mathematics 9*(19), 2477.

McCabe, B. P. and G. M. Martin (2005). Bayesian predictions of low count time series. *International Journal of Forecasting 21*(2), 315–330.

Medeiros, M. C., G. Vasconcelos, Â. Veiga, and E. Zilberman (2021). Forecasting macroeconomic variables in data-rich environments. *International Journal of Forecasting 37*(2), 1073–1097.

Mills, T. C. (2009). Forecasting obesity trends in England. *Journal of the Royal Statistical Society Series A: Statistics in Society 172*(1), 107–117.

Mills, T. C. (2010). Forecasting compositional time series. *Quality & Quantity 44*(4), 673–690.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

Nakajima, J. and M. West (2011). Dynamic factor models with macroeconomic applications. *Bayesian Analysis 6*(3), 465–496.

Nariswari, R. and H. Pudjihastuti (2019). Bayesian forecasting for time series of count data. *Procedia Computer Science 157*, 427–435.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Nicholson, W. B. (2020). *bigVAR: Tools for Modelling Sparse High-Dimensional Multivariate Time Series*. Comprehensive R Archive Network (CRAN). R package version 1.0.2.

Nicholson, W. B., D. S. Matteson, and J. Bien (2017). Varx-l: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting 33*(3), 627–651.

Park, T. and G. Casella (2008). The Bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Pfaff, B. (2008). Var, svar and svec models: Implementation within r package vars. *Journal of Statistical Software 27*(4), 1–32.

Pole, A., M. West, and J. Harrison (2018). *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC.

Polson, N. G. and J. G. Scott (2012). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian Analysis 7*(3), 501–538.

Polson, N. G. and V. Sokolov (2019). Bayesian regularization: From tikhonov to horseshoe. *Wiley Interdisciplinary Reviews: Computational Statistics 11*(4), e1463.

Prado, R. and M. West (2010). *Time series: Modeling, Computation, and Inference*. Chapman and Hall/CRC.

Prüser, J. (2021). The horseshoe prior for time-varying parameter vars and monetary policy. *Journal of Economic Dynamics and Control 129*, 104188.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ravishanker, N., D. K. Dey, and M. Iyengar (2001). Compositional time series analysis of mortality proportions. *Communications in Statistics-Theory and Methods 30*(11), 2281–2291.

Roberts, S. J. and W. D. Penny (2002). Variational Bayes for generalized autoregressive models. *IEEE Transactions on Signal Processing 50*(9), 2245–2257.

Silva, D. and T. Smith (2001). Modelling compositional time series from repeated surveys. *Survey Methodology 27*(2), 205–215.

Silveira de Andrade, B., M. G. Andrade, and R. S. Ehlers (2015). Bayesian GARMA models for count data. *Communications in Statistics: Case Studies, Data Analysis and Applications 1*(4), 192–205.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.

Snyder, R. D., J. K. Ord, A. B. Koehler, K. R. McLaren, and A. N. Beaumont (2017). Forecasting compositional time series: A state space approach. *International Journal of Forecasting 33*(2), 502–512.

Song, P. and P. J. Bickel (2019). Large-vector autoregressions with nonconvex penalization for big time series. *Journal of Forecasting 38*(7), 642–658.

Spencer, D. E. (1993). Developing a Bayesian vector autoregression forecasting model. *International Journal of Forecasting 9*(3), 407–421.

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.5.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

Tsay, R. S., D. Wood, and J. Lachmann (2022). *MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models*. R package version 1.2.1.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica: Journal of the Econometric Society*, 59–73.

Vehtari, A. and J. Ojanen (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys 6*, 142–228.

Vehtari, A., D. Simpson, A. Gelman, Y. Yao, and J. Gabry (2015). Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*.

West, M. (1996). Bayesian time series: Models and computations for the analysis of time series in the physical sciences. In *Maximum Entropy and Bayesian Methods*, pp. 23–34. Springer.

West, M. (2013). Bayesian dynamic modelling. *Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith*, 145–166.

Zheng, T. and R. Chen (2017). Dirichlet ARMA models for compositional time series. *Journal of Multivariate Analysis 158*, 31–46.

Zhou, X. and J. C. Chan (2023). Factor-augmented high-dimensional vars with structured priors: New perspectives for macroeconomic forecasting. *Journal of Econometrics 226*, 146–168.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(2), 301–320.