BASTION: A Bayesian Framework for Trend and Seasonality Decomposition

Jason B. Cho
Department of Statistics and Data Science
Cornell University
Ithaca, United States

David S. Matteson
Department of Statistics and Data Science
Cornell University
Ithaca, United States

Abstract

We introduce BASTION (Bayesian Adaptive Seasonality and Trend DecompositION), a flexible Bayesian framework for decomposing time series into trend and multiple seasonality components. We cast the decomposition as a penalized nonparametric regression and establish formal conditions under which the trend and seasonal components are uniquely identifiable, an issue only treated informally in the existing literature. BASTION offers three key advantages over existing decomposition methods: (1) accurate estimation of trend and seasonality amidst abrupt changes, (2) enhanced robustness against outliers and time-varying volatility, and (3) robust uncertainty quantification. We evaluate BASTION against established methods, including TBATS, STR, and MSTL, using both simulated and real-world datasets. By effectively capturing complex dynamics while accounting for irregular components such as outliers and heteroskedasticity, BASTION delivers a more nuanced and interpretable decomposition. To support further research and practical applications, BASTION is available as an R package at `https://github.com/Jasoncho0914/BASTION`.

1 INTRODUCTION

Time-series decomposition is a powerful analytical method used to break down a univariate time-series into components such as trend and seasonality. Decomposition reveals deeper insights into underlying long or medium-term trends that may not be immediately apparent. Historically, the most prominent applications of the method have been on economic reporting by government agencies, where it is used to adjust for seasonality when publishing key macroeconomic indicators such as the Consumer Price Index (CPI) (of Labor Statistics, 1977) and unemployment rates (Dagum, 1979). Beyond economics, it has been widely applied in climate science (Grieser et al., 2002; Zhou et al., 2015), epidemiology (Abeku et al., 2002; Ke et al., 2016), and business and management (Gardner Jr and Diaz-Saiz, 2002; Zarnowitz and Ozyildirim, 2006; Rosselló and Sansó, 2017), where seasonal effects play a crucial role in data analysis.

More recently, time-series decomposition has received significant attention, as it has been shown to improve forecasting (Kreuzer et al., 2025). State-of-the-art time-series forecasting models such as TBATS (Alysha M. De Livera and Snyder, 2011), ETSformer (Woo et al., 2022), Prophet (Taylor and Letham, 2018), DeepFS (Jiang et al., 2022), Autoformer (Wu et al., 2022), TimeMixer (Wang et al., 2024) and MSD-mixer (Zhong et al., 2023), explicitly incorporate seasonal decomposition in their model architectures. Despite this widespread adoption, formal analysis of identifiability and principled uncertainty quantification for each component remain limited in the existing literature.

We present Bayesian Adaptive Seasonality Trend decompositION (BASTION), a flexible Bayesian framework for trend-seasonality decomposition. A key innovation is the use of a global–local shrinkage prior, which reduces noise while preserving salient signals, enabling locally adaptive yet smooth estimates for trend and seasonal components. The Bayesian formulation also enables the joint estimation of components that are typically treated outside the decomposition model. Such components include outliers or time-varying volatility. In addition, BASTION provides coherent uncertainty quantification for each compo-

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

nent, an aspect that most existing methods lack. Beyond methodology, we develop the theoretical foundations required for reliable decomposition by establishing conditions for identifiability of trend and seasonal components under penalized regression formulations.

The paper is organized as follows. Section 2 reviews related work on time-series decomposition and global-local shrinkage priors. Section 3 formulates decomposition as a penalized regression problem and develops theoretical results on identifiability of decomposition components. Section 4 presents the Bayesian formulation of the model, including the use of global–local shrinkage priors and posterior inference. Section 5 evaluates the proposed method through extensive simulation studies and real data applications, including airline passenger counts and electricity demand.

2 RELATED WORK

Much of the time-series decomposition literature treats decomposition primarily as an exploratory or preprocessing tool rather than as a statistical inference problem. As a result, existing methods face crucial limitations, including difficulty accommodating multiple seasonalities, adapting to abrupt trend changes, providing principled uncertainty quantification, handling heteroskedastic noise, and maintaining robustness to outliers. While some methods address subsets of these challenges, none resolves all of them simultaneously.

Foundational approaches include Seasonal–Trend decomposition using LOESS (STL) (Cleveland et al., 1990) and structural time-series models (Harvey and Peters, 1990). Subsequent extensions address specific limitations. Multiple STL (MSTL) (Bandara et al., 2021) accommodates multiple seasonalities but does not provide uncertainty quantification and remains sensitive to abrupt trend changes. Seasonality–Trend decomposition with Regression (STR) (Dokumentov and Hyndman, 2022) supports more complex seasonal structure and provides uncertainty quantification, but is less effective at capturing abrupt changes. Robust STL (Wen et al., 2019) improves robustness to outliers and can accommodate multiple seasonalities and abrupt changes, but does not provide uncertainty quantification. Moreover, none of the existing approaches explicitly model heteroskedastic noise, which is commonly observed in time-series data across various disciplines.

We address these gaps through a Bayesian framework with global-local shrinkage priors. These priors have become an important class of continuous shrinkage priors in high-dimensional regression (Carvalho et al., 2010; Bhadra et al., 2015) and have been extended to time-series problems including Bayesian smoothing (Kowal et al., 2019; Schafer and Matteson, 2025), regression (Cadonna et al., 2020), change-point detection (Wu et al., 2024; Banerjee, 2022), and volatility estimation (Cho and Matteson, 2026; Huber and Pfarrhofer, 2021). To our knowledge, they have not previously been used for time-series decomposition.

3 DECOMPOSITION AS A PENALIZED REGRESSION PROBLEM

We first frame the decomposition problem in the language of penalized regression. This provides a familiar framework for situating our work within the broader nonparametric regression literature and highlights connections with existing approaches by Harvey and Peters (1990) and Dokumentov and Hyndman (2022). Given a length N time-series $\boldsymbol{y} := \{y_t\}_{t=1}^N$, an additive observation equation structure is assumed, consisting of the trend component $\boldsymbol{T} := \{T_t\}_{t=1}^N$, P seasonal components $\boldsymbol{S}_i := \{S_{i,t}\}_{t=1}^N$ with corresponding seasonal periods $\{k_i\}_{i=1}^P$, and the gaussian remainder $\boldsymbol{R} := \{R_t\}_{t=1}^N$,

$$y_t = T_t + \sum_{i=1}^P S_{i,t} + R_t, \quad [R_t|\sigma_y^2] \sim N(0, \sigma_y^2). \tag{1}$$

This is a high-dimensional problem: $N(P+1)$ unknown parameters must be estimated from only N observations. Without additional structure, the maximum likelihood estimate is not identifiable. Specifically, structure on each latent component must be imposed.

Define D_T and $D_{S,i}$ be the linear operators for trend and the i-th seasonal component, and $\lambda_T, \lambda_{S,i} > 0$ be the hyperaparameter controlling the penalty. The norm of the linear operation of each component is further penalized. With this additional regularization, the maximum likelihood estimates are written as:

$$\begin{aligned}\hat{\boldsymbol{T}}, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_P = \underset{\boldsymbol{T}, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_P}{\arg\min} \Bigg\{ &\|Y - \boldsymbol{T} - \sum_{i=1}^P \boldsymbol{S}_i\|_2^2 \\ &+ \lambda_T \|D_T \boldsymbol{T}\|_2^2 + \sum_{i=1}^P \lambda_{S,i} \|D_{S,i} \boldsymbol{S}_i\|_2^2 \Bigg\}.\end{aligned} \tag{2}$$

A natural choice of the linear operator, D_T, for the trend is the second-difference operator, often denoted by Δ^2, where $\Delta^2 T_t = T_{t+1} - 2T_t + T_{t-1}$. The use of second differencing for smoothing is well established in the literature, with applications in smoothing splines (Schoenberg, 1964), Hodrick-Prescott filter (Hodrick and Prescott, 1997), ℓ_1-trend filter (Kim et al., 2009; Tibshirani, 2014), and Bayesian Trend filters (Roualdes, 2015; Kowal et al., 2019).

For the seasonality, Harvey and Peters (1990) proposes the seasonal recurrence operator: $S_t = -\sum_{m=1}^{k-1} S_{t-m}$, where k is the length of the seasonality or $(I + B + B^2 + \ldots + B^{k-1})$ as a matrix operation with B denoting the backshift operator. Another natural choice would be the k-th order backshift operator $(1 - B^k)$, where k is the length of the cycle, thereby penalizing, $S_{i,t+k} - S_{i,t}$. While not used for time-series decomposition, such operation shows up in a classical time-series modeling literature such as SARIMA type models.

3.1 Identifiability

Existing work such as Dokumentov and Hyndman (2022) has addressed the identfiability issue only informally by imposing centering constraints on seasonal terms. To our knowledge, this is the first work to formalize identifiability conditions in the context of time-series decomposition. In particular, we show that even after penalizing each component with the operators discussed above, the solution to Equation 2 need not be uniquely identified.

The first order condition of the Equation 2 gives us the following set of equations:

$$
\begin{aligned}
&A_0\hat{\boldsymbol{T}} = \boldsymbol{Y} - \sum_{i=1}^{P} \hat{\boldsymbol{S}}_i \\
&A_1\hat{\boldsymbol{S}}_1 = Y - \hat{\boldsymbol{T}} - \sum_{j\neq 1} \hat{\boldsymbol{S}}_j \\
&\ldots \\
&A_P\hat{\boldsymbol{S}}_P = Y - \hat{\boldsymbol{T}} - \sum_{j\neq P} \hat{\boldsymbol{S}}_j,
\end{aligned}
$$

where $A_0 = (I + \lambda_0 L_0)$, $L_0 = (D_{\boldsymbol{T}})' D_{\boldsymbol{T}}$, and $A_j = (I + \lambda_j L_j)$, $L_j = (D_{S,j})' D_{S,j}$ for $j = 1, \ldots P$. This means that the identifiability of the solution hinges on whether the block matrix H,

$$
H = \begin{bmatrix} A_0 & I & \ldots & I \\ I & A_1 & \ldots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \ldots & A_P \end{bmatrix},
$$

arising from this first-order conditions is full-rank, thus invertible.

Theorem 1 (Identifiability Condition)**.** *The solution in Equation 2 is uniquely identifiable if and only if*

$$
\begin{aligned}
&\dim\ker(D_T) + \sum_{j=1}^{P} \dim\ker(D_{S,j}) \\
&= \dim\Big(\ker(D_T) + \sum_{j=1}^{P} \ker(D_{S,j})\Big)
\end{aligned}
$$

Equivalently, the subspaces $\ker(D_T)$, $\ker(D_{S,1})$, *...,* *and* $\ker(D_{S,P})$ *are linearly independent*

Proof. The argument relies on expanding the first-order conditions and analyzing the induced sum map on the kernels of the D_T and D_j. A full derivation is provided in Appendix A.1. □

This theorem shows that non-identifiability can arise only from directions that are simultaneously unpenalized by all operators. Using second differencing for the trend together with either the recurrent seasonal operator of Harvey and Peters (1990) or the seasonal differencing operator for the seasonal component yields a unique solution.

Corollary 1.1 (Single seasonality)**.** *For $P = 1$, let D_T be the second–difference operator and $D_{S,1}$ be either the seasonal recurrent operator $(1 + B + \cdots + B^{k-1})$, or the seasonal differencing operator $(1 - B^k)$.Then the decomposition is uniquely identifiable under the seasonal recurrent operator, and identifiable only up to an additive constant under the seasonal differencing operator.*

Proof. Since $\dim\ker(D_T) = 2$ and $\dim\ker(D_{S,1}) = k$, we compute their intersections. For the recurrent operator, $\ker(D_T) \cap \ker(D_{S,1}) = \mathbf{0}$, so the decomposition is unique. For the differencing operator, $\ker(D_T) \cap \ker(D_{S,1}) = \mathrm{span}(\mathbf{1})$, so the decomposition is identifiable only up to a constant shift. Full details are given in Appendix A.2. □

However, identifiability is lost once multiple seasonal components are introduced. The corollary below shows that the identifiability of a model depends on the greatest common denominator of the seasonal lengths when either seasonal recurrent or differencing operator is applied to penalize each seasonal component.

Corollary 1.2 (Multiple seasonalities)**.** *For $P > 1$, with D_T given by the second differencing and each $D_{S,j}$ given by either 1) seasonal differencing or 2) seasonal recurrent operator, the decomposition is not uniquely identifiable.*

Proof. We show that the nullity of the design matrix,

$$
\begin{aligned}
\mathrm{nullity}(H_R) = &\sum_{i<j} \gcd(k_i, k_j) \\
&- \sum_{i<j<l} \gcd(k_i, k_j, k_l) + \cdots - \gcd(k_1, \ldots, k_P),
\end{aligned}
$$

for the seasonal recurrence operator and

$$
\mathrm{nullity}(H_S) = 1 + \mathrm{nullity}(H_R),
$$

for the seasonal differencing operator. $k_1, \ldots, k_P$ are the length of each season and gcd is the greatest common denominator (Appendix A.3). □

3.2 Multiple Regularization Operators

The identifiability issues arising with multiple seasonalities stem from large null spaces left unpenalized by any single operator. A natural way to address this problem is to penalize each component with multiple linear operators simultaneously, thereby reducing the intersection of their null spaces. For instance, one may penalize a seasonal component both by second differencing (to enforce smoothness) and by seasonal differencing (to enforce periodicity). Under this framework, the estimator becomes

$$\begin{aligned}\hat{\boldsymbol{T}}, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_P = \underset{\boldsymbol{T}, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_P}{\arg\min} \Big\{ &\|Y - \boldsymbol{T} - \sum_{i=1}^{P} \boldsymbol{S}_i\|_2^2 \\ &+ \sum_{m=1}^{q_T} \lambda_T^m \|D_T^m \boldsymbol{T}\|_2^2 \\ &+ \sum_{m=1}^{q_S} \lambda_{S,i}^m \sum_{i=1}^{P} \|D_{S,i}^m \boldsymbol{S}_i\|_2^2 \Big\}.\end{aligned} \tag{3}$$

q_T and q_S represent the number of penalties for the trend and seasonal component respectively. The hyperparameters, $\lambda_T^m, \lambda_{S,i}^m > 0$ are strictly positive scalar controlling the degree of penalty applied to each operator.

Theorem 2 (Identifiability with Multiple Penalties)**.** *Let $V_0 = \bigcap_{m=1}^{q_T} \ker(D_T^m)$ and $V_j = \bigcap_{m=1}^{q_S} \ker(D_{S,j}^m)$ for $j = 1, \ldots, P$. Then, the solution in Equation 3 is identifiable if and only if*

$$\sum_{j=0}^{P} \dim V_j = \dim \Big(\sum_{j=0}^{P} V_j \Big),$$

Equivalently, the subspaces $V_0, \ldots, V_P$ are linearly independent subspaces.

Proof. Similar proof is performed as in Theorem 1. A full derivation is provided in Appendix A.4. □

Corollary 2.1 (Mixed differencing 1)**.** *For $P \geq 1$, consider the trend penalized by the second-difference operator D_T and each seasonal component $\boldsymbol{S}_j$ penalized by either*

1. *a second-difference operator $D_{S,j}^1$ and a seasonal differencing operator $D_{S,j}^2$ with an additional constraint that $\mathbf{1}^T \boldsymbol{S}_j = 0$ or*
2. *a second-difference operator $D_{S,j}^1$ and a seasonal recurrence operator $D_{S,j}^2$,*

the decomposition is uniquely identifiable.

Proof. By imposing both the second and seasonal differencing penalties, each seasonal component has only a constant, span($\mathbf{1}$), in its null space. The additional centering constraint $\mathbf{1}^\top \boldsymbol{S}_j = 0$ removes this from the null space, which leaves only the trend null space. The null space reduces to 0 when the seasonal recurrence and second-difference operator are imposed. Detailed derivation is in Appendix A.5. □

This result shows that in the multiple-seasonality setting, the identifiability issues arising under seasonal differencing and recurrence penalties can be resolved by additionally imposing a second-difference operator on each seasonal component. More generally, Theorem 2 provides a simple diagnostic: for any proposed set of operators, one can compute the associated null spaces $V_0, \ldots, V_P$ and check whether they are linearly independent.

4 BAYESIAN PERSPECTIVE

Building on this foundation, we now propose **BASTION** (Bayesian Adaptive Seasonal Trend decompositION), a unified Bayesian model for trend-seasonality decomposition. We extend the penalized least squares formulation described in Section 3 by casting it in a Bayesian framework. Recall that D_T and $D_{S,i}$ for $i = 1, \ldots, P$ are the operators for the trend and seasonal components, respectively. We assign Gaussian priors:

$$\begin{aligned}[D_T \boldsymbol{T} | \sigma_T, \sigma_y] &\sim N(0, \sigma_y^2 \sigma_T^2 I), \\ [D_{S,i} \boldsymbol{S}_i | \sigma_{s,i}, \sigma_y] &\sim N(0, \sigma_y^2 \sigma_{S,i}^2 I),\end{aligned}$$

together with the observation equation described in Equation 1. Under this specification, the posterior mode of $\boldsymbol{T}, \boldsymbol{S}_i, \ldots, \boldsymbol{S}_P$ is identical to the penalized estimator described in Equation 2, with penalty weights $\lambda_T = 1/\sigma_T^2$ and $\lambda_{S,i} = 1/\sigma_{S,i}^2$.

The penalized regression with multiple operators as described in Equation 3 also admits a Bayesian interpretation. In this setting, the appropriate prior is not a product of Gaussians but a Gaussian random field, since the quadratic forms are not well defined as independent priors. With $L_T^m = (D_T^m)' D_T^m$ and $L_{S,i}^m = (D_{S,i}^m)' D_{S,i}^m$, we have:

$$P(\boldsymbol{T}) \propto \exp\Big\{ -\frac{1}{2\sigma_y^2} \boldsymbol{T}' \Big(\sum_{m=1}^{q_T} \frac{1}{\sigma_{T,m}^2} L_T^m \Big) \boldsymbol{T} \Big\}$$

$$P(\boldsymbol{S}) \propto \exp\Big\{ -\frac{1}{2\sigma_y^2} \sum_{i=1}^{P} \boldsymbol{S}_i' \Big(\sum_{m=1}^{q_T} \frac{1}{\sigma_{S,i,m}^2} L_{S,i}^m \Big) \boldsymbol{S}_i \Big\}.$$

The posterior mode under this specification also reduces to the penalized estimator described in Equation 3 with $\lambda_T^m = 1/\sigma_{T,m}^2$ $\lambda_{S,i}^m = 1/\sigma_{S,i,m}^2$.

The Bayesian formulation not only recovers the penalized estimators from the previous sections, but also broadens the framework in several directions. It provides coherent uncertainty quantification through posterior credible intervals, accommodates various prior structures that can encode smoothness, sparsity, or other domain knowledge and also incorporates additional components such as heterogeneous error variances (Kim et al., 1998) or outlier component (Wu et al., 2024).

4.1 Global-Local Shrinkage Prior

As shown in Equations 2 and 3, existing frameworks impose a uniform penalty term $\lambda_\bullet$ on each operator, which can oversmooth sharp transitions or underpenalize noise. Global–local shrinkage priors address this limitation by coupling an overall shrinkage parameter that enforces stability with time-varying local parameters that selectively relax shrinkage when large changes occur. The defining feature of BASTION is the use of such global–local shrinkage priors for both trend and seasonal components, enabling estimates that evolve smoothly while still capturing abrupt shifts. Formally, we specify:

$$
\begin{aligned}
&[D_T\boldsymbol{T}|\sigma_y, \eta_T, \tau_T] \sim N(0, \sigma_y^2\tau_T^2\eta_T^2 I),\\
&[\tau_T] \sim p(\tau_T), \quad [\eta_T] \sim p(\eta_T),\\
&[D_{S,i}\boldsymbol{S}_i|\sigma_y, \tau_{S,i}, \eta_{S,i}] \sim N(0, \sigma_y^2\tau_{S,i}^2\eta_{S,i}^2 I),\\
&[\tau_{S,i}] \sim p(\tau_{S,i}), \quad [\eta_{S,i}] \sim p(\eta_{S,i}),
\end{aligned}
$$

where each component is regularized by a global scalar parameter $\tau_\bullet$, controlling the overall shrinkage, paired with a time-varying local parameter $\eta_\bullet := [\eta_{\bullet,1}, \dots \eta_{\bullet,N}]'$ that captures local variation.

Several global-local shrinkage priors have been proposed, including the horseshoe prior Carvalho et al. (2010), horseshoe+ Bhadra et al. (2015), dynamic shrinkage processes Kowal et al. (2019), the regularized horseshoe Piironen and Vehtari (2017), and the triple Gamma prior Cadonna et al. (2020), each differing in shrinkage strength. Considering the large number of parameters to be estimated for time-series decomposition task, the horseshoe prior is chosen, as it requires the least number of parameters. Under this specification:

$$
[\tau_\bullet] \sim C^+(0, 1/N), \qquad [\eta_{\bullet,t}] \overset{iid}{\sim} C^+(0, 1).
$$

4.2 Explicit Outlier Detection

Empirical time series frequently exhibit atypical observations due to measurement error or rare events. Existing methods typically handle outliers by pre-adjusting the observed series or by applying smoothing methods that are robust to outliers, rather than directly addressing them within the decomposition framework (Findley et al., 1998; Dokumentov and Hyndman, 2022). BASTION is the first decomposition method to explicitly model outliers through the use of an extreme shrinkage-inducing prior inspired by the approach in Wu et al. (2024). To capture large deviations at isolated time points, BASTION applies the horseshoe+ prior by Bhadra et al. (2015) to ζ_t. By nesting two levels of half-Cauchy distributions, the horseshoe+ prior provides more aggressive shrinkage, allowing larger deviations from zero at a fewer number of locations when compared to the regular horseshoe prior. Specifically,

$$
\begin{aligned}
&[\zeta_t|\sigma_y, \eta_{\zeta,t}] \sim N(0, \sigma_y^2\eta_{\zeta,t}^2),\\
&[\eta_{\zeta,t}|\tau_\zeta, \xi_{\zeta,t}] \sim C^+(0, \tau_\zeta\xi_{\zeta,t}),\\
&[\tau_\zeta] \sim C^+(0, 1),\\
&[\xi_{\zeta,t}] \overset{iid}{\sim} C^+(0, 1).
\end{aligned}
$$

4.3 Handling Heteroskedasticity

While heteroskedastic noise, characterized by changing volatility over time, is often observed in data sets across various academic disciplines, no existing decomposition methods explicitly account for it. Another novel feature of BASTION is its ability to directly handle heteroskedasticity. To address this gap, BASTION introduces a stochastic volatility (SV) model to the remainder term. In addition to the overall variance parameter, σ_y, we introduce a time-varying variance component $\{\nu_t\}_{t=1}^N$, which follows a first order stochastic volatility model by Kim et al. (1998), defined as follows:

$$
\begin{aligned}
&[R_t|\sigma_y, \nu_t] \sim N(0, \sigma_y^2\nu_t^2)\\
&log(\nu_t^2) = \mu + \phi(log(\nu_{t-1}^2) - \mu) + \sigma_\nu\epsilon_t\\
&[\epsilon_t] \overset{iid}{\sim} N(0, 1).
\end{aligned}
$$

Here, μ represents the mean log-variance, ϕ is a persistence parameter that controls how strongly current volatility depends on past values, and σ_ν dictates the variability of the log-variance process. By incorporating a time-varying volatility model, BASTION provides a more flexible decomposition that adapts to heteroskedastic structures in the data. This extension allows for more accurate trend and seasonality estimation, particularly in data with substantial fluctuations in residual variance.

	Trend	Seasonality	Seasonal Length	Remainder	Outlier
DGP 1	Piecewise Linear	Fourier Pairs	12, 40	Constant	FALSE
DGP 2	Linear	Piecewise Constant	40	Constant	FALSE
DGP 3	Quadratic	Fourier Pair	50	Stochastic Volatility	FALSE
DGP 4	Piecewise Linear	Piecewise Constant & Fourier Pair	12, 40	Stochastic Volatility	TRUE

Table 1: Descriptions of data generating processes (DGP) with regards to their trend, seasonality, seasonal lengths, remainder, and outliers. The sample size of each time-series is 500, and each DGP is replicated 1000 times with random coefficients for both the trend and seasonality components. Exact specification of each simulation schemes can be found in Appendic C and example paths from each DGP can be found in Appendix D.

4.4 Implementation

Fully Bayesian inference provides calibrated uncertainty quantification by sampling from the posterior distribution. General-purpose samplers such as the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014), implemented in Stan (Carpenter et al., 2017), are popular for complex models because of their automatic tuning and robust convergence properties (Livingstone et al., 2019; Neal et al., 2011). However, for high-dimensional and structured time-series models like BASTION, these methods are often computationally infeasible due to poor scaling and slow mixing (Thomas and Tu, 2021; Sacher et al., 2021).

Instead, we adopt a Gibbs sampling strategy that scales linearly with the length of the time series, making it practical for large datasets if the goal is to sample from the exact posterior distribution. With the exception of the persistence parameter ϕ, the persistence parameter for the SV model in the remainder term, all full conditional distributions are available in closed form and admit conjugate updates. Because all the state variables, $\{T_t\}$, $\{S_t\}$, and $\{\zeta_t\}$ are conditionally Gaussian and sampled using the multivariate Gaussian sampler of Rue (2001), which is $\mathcal{O}(N)$. Parameters associated with the horseshoe prior are updated using the parameter-expanded Gibbs scheme of Makalic and Schmidt (2016), while the stochastic volatility parameters in the remainder term are sampled with the All Without a Loop (AWOL) algorithm of Kastner and Frühwirth-Schnatter (2014). Importantly, both procedures, like the Gaussian state sampler, scale linearly with the number of observations. Detailed specification of the prior distribution as well as the conditional posterior distribution for Gibbs sampling is detailed in Appendix B. Implementation of BASTION in `R` is available at `https://github.com/Jasoncho0914/BASTION`.

5 SIMULATION STUDY

5.1 Set Up

In this section, BASTION is compared against existing multiple seasonalities decomposition models: MSTL by Bandara et al. (2021), STR by Dokumentov and Hyndman (2022), and TBATS by Alysha M. De Livera and Snyder (2011) across various simulation scenarios described in Table 1. Exact simulation schemes are detailed in Appendix C. DGP 1 through 4 exhibit characteristics such as trend discontinuities, seasonal discontinuities, heteroskedastic noise, or a combination of these with additive outliers. Note that DGP 3 and DGP 4 include heteroskedastic noise, with additive outliers additionally introduced in DGP 4. The comparisons are made in terms of their ability to accurately extract trend, seasonality, and signal (trend and seasonality combined), by measuring the mean squared error (MSE). As a Bayesian methods, BASTION provides uncertainty quantification via credible region. STR, a Frequentist counterpart, is the only other decomposition method with uncertainty quantification. We compare empirical coverage and average interval width to assess the accuracy and precision of uncertainty quantification provided by BASTION's credible intervals and STR's confidence intervals.

Models are implemented in `R` R Core Team (2024). TBATS and MSTL are implemented using the `forecast` package Hyndman et al. (2024), and STR is implemented using the `stR` package Dokumentov and Hyndman (2024). For BASTION, we extend the efficient sampling approach for the Gaussian state-space model framework from Rue (2001), and the stochastic volatility model from Kastner (2016).

5.2 Results

Table 2 reports mean squared error (MSE) for BASTION, MSTL, TBATS, and STR. Across all four simulation scenarios, BASTION consistently achieves the lowest or near-lowest MSE for the trend, seasonality,

	Mean Squared Error			
	TBATS	MSTL	STR	BASTION
DGP 1				
Signal	13.878	11.760	10.829	**0.376**
Trend	3.624	13.227	13.785	**0.581**
Seasonality	11.672	2.712	2.575	**0.536**
DGP 2				
Signal	1.411	1.214	0.7020	**0.3922**
Trend	0.621	0.184	0.0878	**0.0579**
Seasonality	0.838	1.032	0.6156	**0.3412**
DGP 3				
Signal	11.307	3.041	0.706	**0.439**
Trend	10.509	0.364	**0.274**	0.293
Seasonality	0.900	2.683	0.444	**0.278**
DGP 4				
Signal	11.829	11.430	20.548	**2.877**
Trend	11.111	12.328	13.431	**5.210**
Seasonality	5.364	3.045	11.358	**2.562**

Table 2: Mean Squared Error (MSE) for Trend, Seasonality, and their combined component (Signal). Each data generating processes (DGP) is replicated 1000 times.

	Empirical Coverage		Interval Width	
	STR	BASTION	STR	BASTION
DGP 1				
Signal	0.799	**0.998**	4.048	5.306
Trend	0.616	**0.970**	1.934	1.993
Seasonality	0.815	**0.999**	2.114	3.314
DGP 2				
Signal	0.869	**0.989**	2.179	2.779
Trend	0.716	**0.976**	0.992	0.854
Seasonality	0.793	**0.956**	1.187	1.925
DGP 3				
Signal	**0.940**	0.995	3.017	4.589
Trend	0.812	**0.929**	1.621	1.959
Seasonality	0.847	**0.981**	1.396	2.630
DGP 4				
Signal	0.679	**0.981**	5.148	5.606
Trend	0.668	**0.939**	2.296	2.246
Seasonality	0.623	**0.999**	2.852	3.360

Table 3: Empirical Coverage (EC) and Interval Width for Trend, Seasonality, and their combined component (Signal). EC is calculated using 95% confidence intervals for STR and 95% credible intervals for BASTION.

and combined signal components. The improvement is most pronounced in DGP 4, which includes both heteroskedastic noise and outlier components. Because BASTION explicitly models these features unlike the competing methods, it is better able to adapt to the data generating process.

Table 3 reports empirical coverage and average interval width for BASTION and STR, the only two methods considered that provide uncertainty quantification. In addition to evaluating uncertainty for the combined signal, we assess how well each method quantifies uncertainty for the individual components. BASTION consistently maintains coverage above the nominal 95% level across all four DGPs, not only for the combined signal but also for both the trend and seasonality components. In contrast, STR attains nominal coverage only for the combined signal in DGP3 and fails to do so in DGPs1, 2, and 4. Moreover, STR substantially undercovers for the trend and seasonality components across all four DGPs, leading to comparatively narrower interval widths across all DGPs for STR.

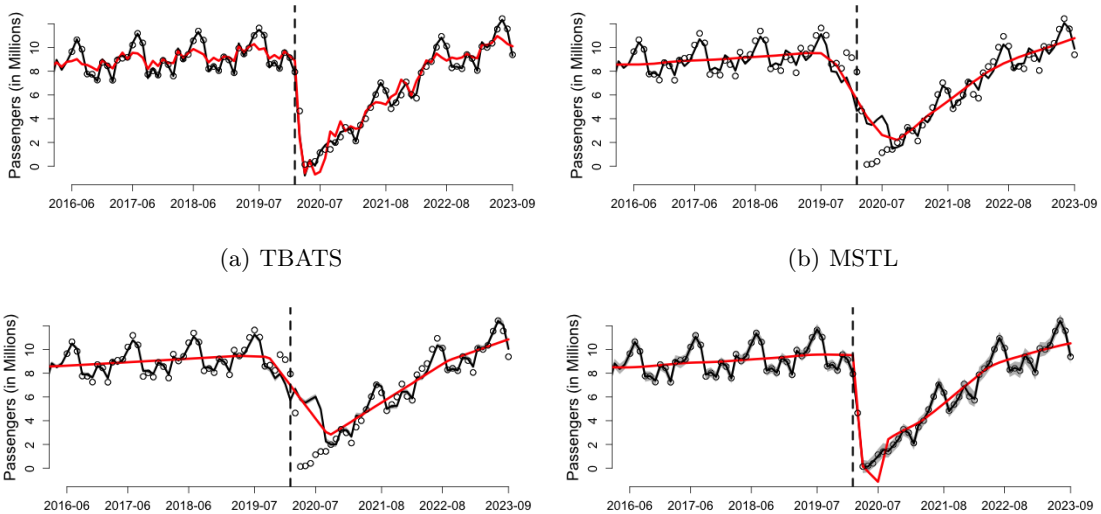
The simulation results show that BASTION reliably recovers trend and seasonal structure across a broad range of data generating processes. Its advantages are most pronounced in settings with structural breaks, heteroskedastic noise, and additive outliers, where existing methods degrade. Moreover, BASTION provides well-calibrated uncertainty quantification for both the combined signal and the individual components. These results motivate applying BASTION to real-world time series, where such features are common across various disciplines.

6 EMPIRICAL DATA ANALYSIS

In this section, we apply BASTION to two real-world datasets that exhibit complex seasonality patterns, abrupt changes, heteroskedastic noise or combination of these characteristics. The first dataset consists of monthly U.S. airline traffic from 2003 to 2023, obtained from Kaggle (Yan, 2023), which highlights sharp declines in passenger volume during significant disruptions such as the COVID-19 pandemic. For the second dataset, we analyze the average daily electricity demand (in Megawatts per hour) for the state of New York, sourced from the New York Independent System Operator (NYISO) via the U.S. Energy Information Administration (EIA, 2024). This dataset captures complexity with its long-term trend, with both weekly and yearly seasonal patterns and heteroskedastic noise.

6.1 U.S. Airline Traffic Data

Monthly international airline traffic data from 2003 to 2023 (Figure 1) exhibits a clear yearly seasonal pattern: passenger counts are lowest early in the year, rise steadily to a peak in July and August, decline sharply in the fall, and rebound moderately in December and January. As per the trend, we see a steady rise in the number of passenger with the significant drop in early 2020, corresponding to the COVID-19 pandemic and resulting travel restrictions, introducing a clear



(a) TBATS (b) MSTL

(c) STR (d) BASTION

Figure 1: Observed data (points); estimated trend + seasonality, $(\hat{\boldsymbol{T}}+\hat{\boldsymbol{S}})$, (black) with 95% credible regions; and trend $(\hat{\boldsymbol{T}})$ component alone (red), for monthly U.S. international airline traffic, 2016 - 2023. The sharp decline in early 2020 reflects COVID-19 travel restrictions.

structural change in the trend. Figure 1 compares seasonality and trend estimates across TBATS, MSTL, STR, and BASTION. BASTION provides a smooth yet adaptive trend, capturing the abrupt COVID-19 drop while preserving overall structure. MSTL and STR oversmooth the break, treating it as gradual, whereas TBATS identifies the sharp change but yields noisy estimates. This example highlights BASTION's ability to maintain smoothness while adapting to abrupt changes. The full decomposition is provided in Appendix D, Figure 6.

6.2 New York Daily Average Electricity Demand

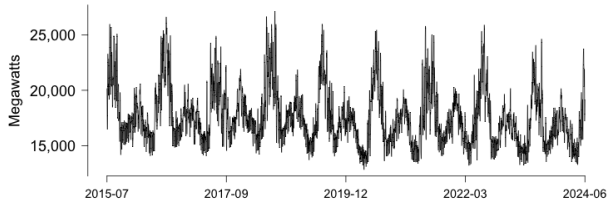


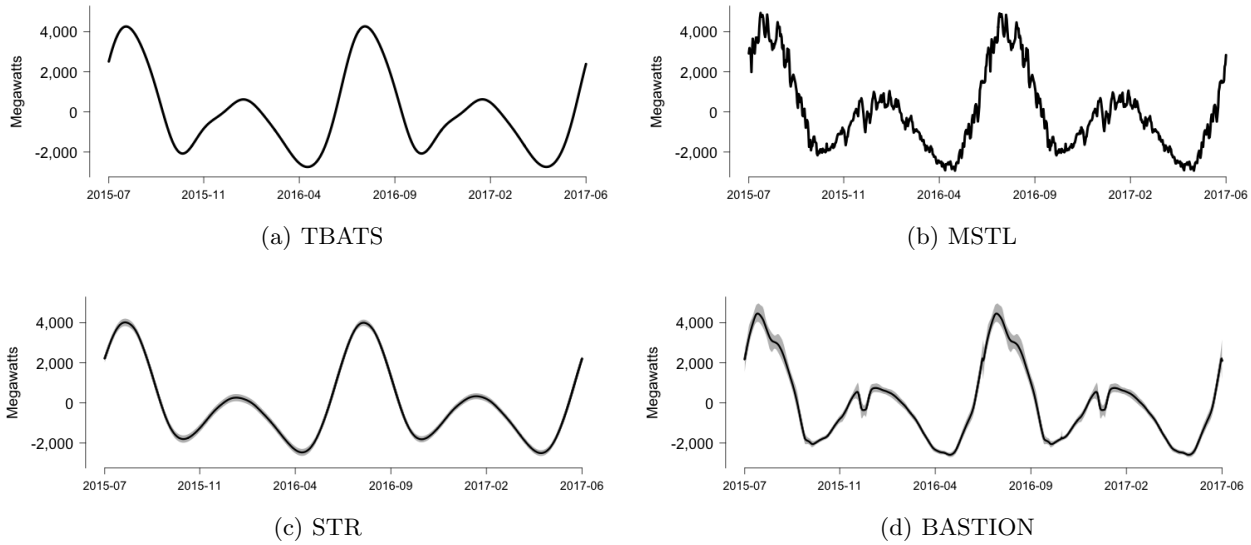
Figure 2: Daily average electricity demand in New York State from 2015-07-01 to 2024-06-30.

The daily average electricity demand data for New York, spanning from July 1, 2015, to June 30, 2024, exhibits clear yearly seasonal patterns. As shown in Figure 2, demand peaks in the summer months due to widespread air-conditioning use and increases again during winter, reflecting heating needs. Spring and fall, by contrast, experience milder temperatures and consequently lower demand.

As shown in Figure 3, all four methods accurately capture the broad seasonal patterns typical of the East Coast, with peak electricity demand during the summer months of July and August due to increased cooling needs. This is followed by a steady decline in the fall and a moderate rise during the winter, reflecting increased heating requirements.

One interesting feature visible in the BASTION and MSTL estimates is a small dip in demand lasting approximately 30 days, from mid-December to mid-January, coinciding with the holiday slowdown in commercial and industrial activity. This reduction in electricity use drives the observed drop. BASTION distinguishes itself by detecting such subtle shifts while maintaining smooth, accurate seasonal estimates. The full BASTION decomposition, including trend and weekly seasonality, is provided in Appendix D.

Another defining characteristic of BASTION, that no other decomposition model provides, is its ability to explicitly estimate time varying volatility. BASTION's estimate of time-varying volatility, depicted in Figure 4, highlights not only the heteroskedasticity of the noise term but also the presence of seasonality in the volatility itself as well. Specifically, as the estimate suggests, higher volatility is observed during the win-



(a) TBATS (b) MSTL

(c) STR (d) BASTION

Figure 3: Yearly seasonality estimates, ($\hat{\boldsymbol{S}}$), for daily average electricity demand in New York from July 2015 to June 2017. Figures (a) - (d) display estimates from TBATS, MSTL, STR, and BASTION, respectively.

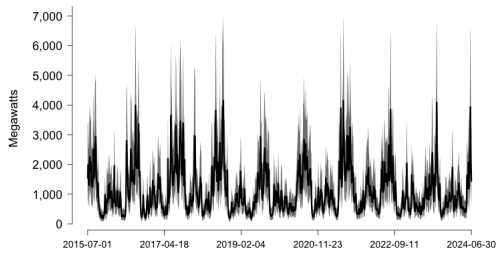


Figure 4: Volatility estimate based on BASTION for daily average electricity demand from 2015-07-01 to 2024-06-30. 95% credible regions are drawn in dark grey.

ter and summer months, while lower volatility occurs during the spring and fall. This insight is particularly relevant for electricity management, where anticipating periods of elevated volatility informs resource allocation, capacity planning, and pricing strategies.

7 CONCLUSION

We introduced BASTION, a flexible Bayesian framework for decomposing time series into trend and multiple seasonalities while explicitly modeling outliers and volatility. Unlike existing methods, BASTION is able to capture abrupt changes in both the trend and seasonality component, while also providing rigorous uncertainty quantification. Simulation results demonstrates that BASTION outperforms TBATS, STR, and MSTL in both accuracy and empirical coverage in most cases. Empirical analyses of airline passenger data and electricity demand further validated its ability to capture complex temporal dynamics and provide interpretable insights. To facilitate further research and practical use, BASTION is available as an R package at `https://github.com/Jasoncho0914/BASTION`.

Several directions for future work remain. One natural extension is to evaluate how improved decomposition translates to downstream forecasting performance, particularly in settings with structural breaks and heteroskedastic noise. Another promising direction is to incorporate additional latent structure, such as Bayesian change point detection, to explicitly identify regime shifts while accounting for evolving seasonal patterns. These extensions would further enhance the flexibility and applicability of the BASTION.

Acknowledgments

Financial support from National Science Foundation grants OAC-1940124 and DMS-2114143 is gratefully acknowledged.

References

Tarekegn A. Abeku, Sake J. De Vlas, Gerard Borsboom, Awash Teklehaimanot, Asnakew Kebede, Dereje Olana, Gerrit J. Van Oortmarssen, and J. D. F. Habbema. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of ethiopia: a simple seasonal adjustment method performs best. *Tropical Medicine & International Health*, 7(10):851–857, 2002.

Rob J. Hyndman Alysha M. De Livera and Ralph D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496): 1513–1527, 2011. doi: 10.1198/jasa.2011.tm09771.

Kasun Bandara, Rob J Hyndman, and Christoph Bergmeir. Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns, 2021.

Sayantan Banerjee. Horseshoe shrinkage methods for bayesian fusion estimation. *Computational Statistics & Data Analysis*, 174:107450, 2022. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2022.107450.

Anindya Bhadra, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard. The horseshoe+ estimator of ultra-sparse signals, 2015.

Annalisa Cadonna, Sylvia Frühwirth-Schnatter, and Peter Knaus. Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics*, 8(2):20, 2020.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76:1–32, 2017.

Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. ISSN 00063444, 14643510.

Jason B. Cho and David S. Matteson. Smoothing variances across time: Adaptive stochastic volatility. *Journal of Business & Economic Statistics*, 0(0): 1–14, 2026. doi: 10.1080/07350015.2025.2602846.

Robert B Cleveland, William S Cleveland, Jean E McRae, Irma Terpenning, et al. Stl: A seasonal-trend decomposition. *J. off. Stat*, 6(1):3–73, 1990.

Estela Bee Dagum. *On the seasonal adjustment of economic time series aggregates: A case study of the unemployment rate*. Number 31. National Commission on Employment and Unemployment Statistics, Washington, 1979.

Alexander Dokumentov and Rob Hyndman. *stR: STR Decomposition of Time Series*, 2024. R package version 0.7.

Alexander Dokumentov and Rob J. Hyndman. Str: Seasonal-trend decomposition using regression. *INFORMS Journal on Data Science*, 1(1):50–62, 2022. doi: 10.1287/ijds.2021.0004.

U.S. Energy Information Administration EIA. New York independent system operator electricity demand data, 2024. Accessed: 2024-06-16.

David F. Findley, Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen. New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152, 1998. doi: 10.1080/07350015.1998.10524743.

Everette S Gardner Jr and Joaquin Diaz-Saiz. Seasonal adjustment of inventory demand series: a case study. *International Journal of Forecasting*, 18(1): 117–123, 2002.

J. Grieser, S. Trömel, and C.-D. Schönwiese. Statistical time series decomposition into significant components and application to european temperature. *Theoretical and Applied Climatology*, 71(3): 171–183, Feb 2002. ISSN 1434-4483. doi: 10.1007/s007040200003.

Andrew C Harvey and Simon Peters. Estimation procedures for structural time series models. *Journal of forecasting*, 9(2):89–108, 1990.

Robert J. Hodrick and Edward C. Prescott. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997. ISSN 00222879, 15384616.

Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.

Florian Huber and Michael Pfarrhofer. Dynamic shrinkage in time-varying parameter stochastic volatility in mean models. *Journal of Applied Econometrics*, 36(2):262–270, 2021. doi: https://doi.org/10.1002/jae.2804.

Rob Hyndman, George Athanasopoulos, Christoph Bergmeir, Gabriel Caceres, Leanne Chhay, Mitchell O'Hara-Wild, Fotios Petropoulos, Slava Razbash, Earo Wang, and Farah Yasmeen. *forecast: Forecasting functions for time series and linear models*, 2024. R package version 8.23.0.

Song Jiang, Tahin Syed, Xuan Zhu, Joshua Levy, Boris Aronchik, and Yizhou Sun. Bridging self-attention and time series decomposition for periodic forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, page 3202–3211, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392365. doi: 10.1145/3511808.3557077.

Gregor Kastner. Dealing with Stochastic Volatility in Time Series Using the R Package stochvol. *Journal of Statistical Software*, 69(i05), 2016. doi: http://hdl.handle.net/10.

Gregor Kastner and Sylvia Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423, aug 2014. doi: 10.1016/j.csda.2013.01.002.

Guibao Ke, Yao Hu, Xin Huang, Xuan Peng, Min Lei, Chaoli Huang, Li Gu, Ping Xian, and Dehua Yang. Epidemiological analysis of hemorrhagic fever with renal syndrome in china with the seasonal-trend decomposition method and the exponential smoothing model. *Scientific reports*, 6(1):39350, 2016.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393, 1998. ISSN 00346527, 1467937X.

Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. l_1 trend filtering. *SIAM Rev.*, 51(2):339–360, 2009. ISSN 0036-1445,1095-7200. doi: 10.1137/070690274.

Daniel R. Kowal, David S. Matteson, and David Ruppert. Dynamic Shrinkage Processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(4):781–804, 05 2019. ISSN 1369-7412. doi: 10.1111/rssb.12325.

Tim Kreuzer, Jelena Zdravkovic, and Panagiotis Papapetrou. Unpacking the trend: Decomposition as a catalyst to enhance time series forecasting models. *Data Mining and Knowledge Discovery*, 39(5):54, 2025.

Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of hamiltonian monte carlo. *Bernoulli*, 25(4A):3109–3138, 2019.

Enes Makalic and Daniel F. Schmidt. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182, 2016. doi: 10.1109/LSP.2015.2503725.

Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

United States. Bureau of Labor Statistics. *The Consumer Price Index: Concepts and Content Over the Years*. Report (United States. Bureau of Labor Statistics). The Bureau, 1977.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2006.07.008.

Juho Piironen and Aki Vehtari. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018 – 5051, 2017. doi: 10.1214/17-EJS1337SI.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.

Jaume Rosselló and Andreu Sansó. Yearly, monthly and weekly seasonality of tourism demand: A decomposition analysis. *Tourism Management*, 60:379–389, 2017.

Edward A. Roualdes. Bayesian trend filtering, 2015.

Havard Rue. Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2):325–338, 2001. ISSN 13697412, 14679868.

Szymon Sacher, Laura Battaglia, and Stephen Hansen. Hamiltonian Monte Carlo for Regression with High-Dimensional Categorical Data. Papers 2107.08112, arXiv.org, July 2021.

Toryn L. J. Schafer and David S. Matteson. Locally adaptive shrinkage priors for trends and breaks in count time series. *Technometrics*, 67(1):157–167, 2025. doi: 10.1080/00401706.2024.2407316.

I. J. Schoenberg. Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences of the United States of America*, 52(4):947–950, 1964. ISSN 00278424, 10916490.

Sean J. Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018. doi: 10.1080/00031305.2017.1380080.

Samuel Thomas and Wanzhu Tu. Learning hamiltonian monte carlo in r. *The American Statistician*, 75(4):403–413, 2021. doi: 10.1080/00031305.2020.1865198. PMID: 37465458.

Ryan J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1), February 2014. ISSN 0090-5364. doi: 10.1214/13-aos1189.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting, 2024.

Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, Huan Xu, and Shenghuo Zhu. Robuststl: A robust seasonal-trend decomposition algorithm for long time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5409–5416, 2019.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, 2022.

Haoxuan Wu, Toryn L. J. Schafer, and David S. Matteson. Trend and variance adaptive bayesian changepoint analysis and local outlier scoring. *Journal of Business & Economic Statistics*, 0(0):1–12, 2024. doi: 10.1080/07350015.2024.2362269.

Yuxian Yan. U.s. airline traffic data (2003-2023), 2023. Accessed: 2024-06-16.

Victor Zarnowitz and Ataman Ozyildirim. Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 53(7):1717–1739, 2006. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2005.03.015.

Shuhan Zhong, Sizhe Song, Weipeng Zhuo, Guanyao Li, Yang Liu, and S-H Gary Chan. A multi-scale decomposition mlp-mixer for time series analysis. *arXiv preprint arXiv:2310.11959*, 2023.

Jing Zhou, Zhongyao Liang, Yong Liu, Huaicheng Guo, Dan He, and Lei Zhao. Six-decade temporal change and seasonal decomposition of climate variables in lake dianchi watershed (china): stable trend or abrupt shift? *Theoretical and Applied Climatology*, 119:181–191, 2015.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. **Yes**
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. **Yes**
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. **Yes**

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. **Yes**
   (b) Complete proofs of all theoretical results. **Yes**
   (c) Clear explanations of any assumptions. **Yes**

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). **Yes**
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). **Yes**
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). **Yes**
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). **Yes**

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. **Not Applicable**
   (b) The license information of the assets, if applicable. **Not Applicable**
   (c) New assets either in the supplemental material or as a URL, if applicable. **Not Applicable**
   (d) Information about consent from data providers/curators. **Not Applicable**
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **Not Applicable**

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. **Not Applicable**
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not Applicable**
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not Applicable**

Supplementary Materials

A PROOFS

A.1 Proof of Theorem 1

Given a length N time-series $\boldsymbol{Y} := \{Y_t\}_{t=1}^N$, the goal is to characterize the condition in which the following

$$\hat{\boldsymbol{T}}, \hat{\boldsymbol{S}}_1, \ldots, \hat{\boldsymbol{S}}_P = \underset{\boldsymbol{T}, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_P}{\arg\min} \left\{ \|Y - \boldsymbol{T} - \sum_{i=1}^P \boldsymbol{S}_i\|_2^2 + \lambda_0 \|D_T \boldsymbol{T}\|_2^2 + \sum_{i=1}^P \lambda_i \|D_{S,i} \boldsymbol{S}_i\|_2^2 \right\},$$

is uniquely identifiable.

The first order condition of the objective function gives us the following sets of equations,

$$\begin{aligned}
&A_0 \hat{\boldsymbol{T}} = \boldsymbol{Y} - \sum_{i=1}^P \hat{\boldsymbol{S}}_i \\
&A_1 \hat{\boldsymbol{S}}_1 = Y - \hat{\boldsymbol{T}} - \sum_{j \neq 1} \hat{\boldsymbol{S}}_j \\
&\ldots \\
&A_P \hat{\boldsymbol{S}}_P = Y - \hat{\boldsymbol{T}} - \sum_{j \neq P} \hat{\boldsymbol{S}}_j,
\end{aligned}$$

where $A_0 = (I + \lambda_0 L_0)$, $L_0 = (D_{\boldsymbol{T}}' D_{\boldsymbol{T}})$, and $A_j = (I + \lambda_j L_j)$, $L_j = (D_{S,j}' D_{S,j})$ for $j = 1, \ldots P$. This can be written as a linear system:

$$\begin{bmatrix} A_0 & I & \ldots & I \\ I & A_1 & \ldots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \ldots & A_P \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{T}} \\ \hat{\boldsymbol{S}}_1 \\ \vdots \\ \hat{\boldsymbol{S}}_P \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y} \\ \vdots \\ \boldsymbol{Y} \end{bmatrix}.$$

If the matrix $H = \begin{bmatrix} A_0 & I & \ldots & I \\ I & A_1 & \ldots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \ldots & A_P \end{bmatrix}$, is full rank, the unique solution exists. This is equivalent to checking the dimension of the kernel, nullity, of the matrix. If the nullity is 0, H is full ranked by the rank nullity relation.

First we show that $\boldsymbol{u} = (u_0, \ldots, u_P) \in \ker(H)$ if and only if $\lambda_j L_j u_j = 0, \forall j \in \{0, \ldots, P\}$ and $\sum_{j=0}^P u_j = 0$.

($\Rightarrow$) Given $\boldsymbol{u} := (u_0, u_1, \ldots, u_P) \in \ker(H)$, define $s := \sum_{j=0}^P u_j$. By the assumption that $\boldsymbol{u} \in \ker(H)$, $H\boldsymbol{u} = 0$. Or equivalently, for each block $\forall j \in \{0, \ldots, P\}$,

$$A_j u_j + \sum_{i \neq j}^P u_i = A_j u_j + \sum_{i \neq 0}^P u_i + (u_j - u_j) = (A_j - I) u_j + \sum_{i=0}^P u_i = (A_j - I) u_j + s = (\lambda_j L_j) u_j + s = 0 \tag{4}$$

The only solution to $s = 0$ because $\sum_{j=0}^P \lambda_j u_j' L_j u_j = \sum_{j=0}^P -u_j' s = -(\sum_{j=0}^P u_j)' s = -\|s\|_2^2 > 0$.

($\Leftarrow$) Fix $\boldsymbol{u} = (u_0, \ldots, u_P)$ such that $\lambda_j L_j u_j = 0, \forall j \in \{0, \ldots, P\}$ and $\sum_{j=0}^{P} u_j = 0$. We need to show that $H\boldsymbol{u} = 0$, which can be verified by the same derivation shown in Equation 4 but backwards. $\forall j \in \{0, \ldots, P\}$,

$$A_j u_j + \sum_{i \neq 0}^{P} u_i = (A_j - I)u_j + \sum_{i=0}^{P} u_i = \lambda_j L_j u_j + \sum_{i=0}^{P} u_i = 0.$$

Lastly, We argue that $\ker(L_0) = \ker(D_T)$ and $\ker(L_j) = \ker(D_{S,j})$ for $j \geq 1$.

($\Rightarrow$) If $u_0 \in \ker(L_T)$, $L_T u_0 = D_T' D_T u_0 = 0$. Since positive semidefinite, $u_0' D_T' D_T u_0 = (D_T u_0)' D_T u_0 \geq 0$, which is only possible if and only if $D_T u_0 = 0$. The same argument applies to $D_{S,j}$ for $j \geq 1$.

($\Leftarrow$) if $u_0 \in \ker(D_T)$, $D_T u_0 = 0$ thus, $(D_T u_0)'(D_T u_0) = u_0' L_T u_0 = 0$. Therefore, $L_T u_0 = 0$.

This proves that $\ker(H)$ consists of vectors from each null space of the operators: $D_T D_{S,1}, \ldots, D_{S,P}$, whose sum is zero. This can be characterized by $V := D_T \oplus_{j=1}^{P} \ker(D_{S,j})$ and define the sum map $\Phi : V \to \mathbb{R}^N$, where $\Phi(v_0, \ldots, v_P) = \sum_{j=0}^{P} v_j$. Therefore, $\ker(H) = \ker(\Phi)$. It follows that

$$\dim\ker(H) \;=\; \dim\ker(\Phi) \;= \dim\ker(D_T) + \sum_{j=1}^{P} \dim\ker(D_j) \;-\; \dim\Big(\ker D_T + \sum_{j=1}^{P} \ker(D_j)\Big).$$

The matrix H is uniquely identifiable if and only if the dimension of the null space is 0 by the rank-nullity relation.

A.2 Proof of Corollary 1.1

Define $2 < k < N$, where k is the length of the seasonality and N be the length of the data. $\mathbf{1} := (1, \ldots, 1)^T \in \mathbb{R}^N$. B is the back shift operator, $Bx_t = x_{t-1}$ for $t \geq 2$. For the trend $\ker(D_T) = \text{span}\{\mathbf{1}, m\}$, where $m = (1, 2, \ldots N)'$. This is because $D_T x = 0$ if and only if differences vanish i.e., $x_t = a + bt$. Two linearly independent solutions are $\mathbf{1}$ and m. For the seasonal differencing operator, $\ker(D_{S,1}) = \{x \in \mathbb{R}^N : x_t = x_{t-k}, \forall t > k\}$. The first k coordinates are free and determine all subsequent values. The shared null space is $\mathbf{1}$.

For the seasonal recurrence $D_{S,1}$, each row has its first nonzero entry in a distinct column. For example row r starts at column r, ending at $r + k$. therefore, rows are linearly independent and Rank of $D_{S,1}$ is $N - k$. By the rank nullity relation, the dimension of the null space is k. Note that $\mathbf{1} \notin \ker(D_{S,1})$ and $m \notin \ker(D_{S,1})$.

A.3 Proof of Corollary 1.2

Let $P > 1$, D_T be the second differencing operator and $D_{S,1}, \ldots D_{S,P}$ be the seasonal differencing operator,$k_1, \ldots k_P$ be the length of the seasons for each seasonal components,

$$\begin{aligned}
&\dim(\ker(D_T)) + \sum_{j=1}^{P} \dim(\ker(D_{S,j}) - \dim(\ker(D_T) + U) \\
&= 2 + \sum_{j=1}^{P} k_j - \dim(\ker(D_T) + U) \\
&= 2 + \dim(U) + (\sum_{j=1}^{P} k_j - \dim(U)) - \dim(\ker(D_T) + U) \\
&= \dim(\ker(D_T) + U) + \sum_{j=1}^{P} k_j - \dim(U)
\end{aligned}$$

As shown previously $\ker(D_T) = \text{span}\{\mathbf{1}, m\}$, where $m = (1, 2, \ldots N)'$, and $\ker(D_{S,j}) = \{x \in \mathbb{R}^N : x_t - x_{t-k_j}, \forall t > k_j\}$, where $\dim\ker(D_{S,j}) = k_j$. $\dim(\ker(D_T) + U) = 1$.

For U inclusion-exclusion gives us:

$$\dim U = \sum_{j} \dim\ker(D_{s,j}) - \sum_{i<j} \dim\ker(D_{S,i} \cap D_{S,j}) + \sum_{i<j<l} \dim\ker(D_{S,i} \cap D_{S,j} \cap D_{S,l}) - \cdots + \dim\ker(\bigcap_{j=1} D_{S,j}).$$

Note that the shared null space between two seasonal components $D_{S,i}$ and $D_{S,j}$ is $gcd(k_i, k_j)$.

$$\dim U = \sum_j k_j - \sum_{i<j} \gcd(k_i, k_j) + \sum_{i<j<l} \gcd(k_i, k_j, k_l) - \cdots + \gcd(k_1, \dots k_P).$$

Thus, the nullity of the design matrix H in Theorem 1 reduces to

$$\begin{aligned} &\dim(\ker(D_T)) + \sum_{j=1}^{P} \dim(\ker(D_{S,j}) - \dim(\ker(D_T) + U) \\ &= 1 + \sum_{i<j} \gcd(k_i, k_j) - \sum_{i<j<l} \gcd(k_i, k_j, k_l) + \cdots - \gcd(k_1, \dots k_P) \end{aligned}$$

If we assume $D_{S,1}, \dots D_{S,P}$ to be a recurrent operator, the nullity is reduced by 1 as the null space between the seasonality and the trend does not interact.

$$\begin{aligned} &\dim(\ker(D_T)) + \sum_{j=1}^{P} \dim(\ker(D_{S,j}) - \dim(\ker(D_T) + U) \\ &= \sum_{i<j} \gcd(k_i, k_j) - \sum_{i<j<l} \gcd(k_i, k_j, k_l) + \cdots - \gcd(k_1, \dots k_P) \end{aligned}$$

A.4 Proof of Theorem 2

Given a length N time-series $\boldsymbol{Y} := \{Y_t\}_{t=1}^N$, we consider the following objective function:

$$\hat{\boldsymbol{T}}, \hat{\boldsymbol{S}}_1, \dots, \hat{\boldsymbol{S}}_P = \underset{\boldsymbol{T}, \boldsymbol{S}_1, \dots, \boldsymbol{S}_P}{\arg\min} \left\{ \|Y - \boldsymbol{T} - \sum_{i=1}^{P} \boldsymbol{S}_i\|_2^2 + \sum_{m=1}^{q_T} \lambda_0^m \|D_T^m \boldsymbol{T}\|_2^2 + \sum_{m=1}^{q_S} \lambda_i^m \sum_{i=1}^{P} \|D_{S,i}^m \boldsymbol{S}_i\|_2^2 \right\},$$

Here, we have q_T and q_S now represent the number of penalties for the trend and seasonal component respectively. The hyperparameter λ_0^m for the trend and λ_i^m is strictly positive scalar controlling the degree of penalty applied to each penalty operator.

We provide a similar identifiability condition as in Theorem 1. Define $V_0 = \bigcap_{m=1}^{q_T} \ker(L_0^m)$ and $V_j = \bigcap_{m=1}^{q_S} \ker(L_j^m)$ for $j > 0$. In this proof, we show that we get a uniquely identifiable solution if:

$$\sum_{j=0}^{P} \dim \ker(V_j) \; - \; \dim \Big(\sum_{j=0}^{P} \ker(V_j) \Big) = 0$$

The same first order condition therefore the same system of equation in Theorem 1 holds, but $A_0, \dots A_P$ are defined differently: $A_0 = (I + \sum_{m=1}^{q_T} \lambda_0^m L_0^m)$, $L_0^m = (D_{\boldsymbol{T}}^m)' D_{\boldsymbol{T}}^m$, for $m = 1, \dots, q_T$ and $A_j = (I + \sum_{m=1}^{q_S} \lambda_j^m L_j^m)$, $L_j^m = (D_{S,j}^m)' D_{S,j}^m$, for $j = 1, \dots P$ and $m = 1, \dots q_S$.

$$\begin{bmatrix} A_0 & I & \dots & I \\ I & A_1 & \dots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \dots & A_P \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{T}} \\ \hat{\boldsymbol{S}}_1 \\ \vdots \\ \hat{\boldsymbol{S}}_P \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y} \\ \vdots \\ \boldsymbol{Y} \end{bmatrix}.$$

Given $H = \begin{bmatrix} A_0 & I & \dots & I \\ I & A_1 & \dots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \dots & A_P \end{bmatrix}$, its nullity is considered.

We show that $\boldsymbol{u} = (u_0, \dots, u_P) \in \ker(H)$ if and only if 1) u_0 is in the kernel of $(\sum_{m=1}^{q_T} \lambda_0^m L_0^m)$ and u_j is in the kernel of $(\sum_{m=1}^{q_s} \lambda_j^m L_j^m)$ for $j = 1, \dots, P$, and $\sum_{j=0}^{P} u_j = 0$.

($\Rightarrow$) Let $\boldsymbol{u} = (u_0, \ldots, u_P) \in \ker(H)$, and define $s := \sum_{j=0}^{P} u_j$. For the trend component:

$$A_0 u_0 + \sum_{i \neq 0}^{P} u_i = (A_0 - I)u_0 + \sum_{i=0}^{P} u_i = (A_0 - I)u_0 + s = \sum_{m=1}^{q_T} \lambda_0^m L_0^m u_0 + s = \left(\sum_{m=1}^{q_T} \lambda_0^m L_0^m \right) u_0 + s$$

and for the seasonal components, $j \in 1, \ldots P$,

$$A_j u_j + \sum_{i \neq j}^{P} u_j = \left(\sum_{m=1}^{q_s} \lambda_j^m L_j^m \right) u_j + s = 0.$$

We also show that only possible $s = 0$. We use the fact that L_j^m are positive semi-definite and $\lambda_j^m > 0$:

$$\begin{aligned}
u_0'(\lambda_0^m L_0^m) u_0 + \sum_{j=1}^{P} \sum_{m=1}^{q_s} u_j'(\lambda_j^m L_j) u_j &\geq 0 \\
u_0'(\lambda_0^m L_0^m) u_0 + \sum_{j=1}^{P} u_j' \sum_{m=1}^{q_s} (\lambda_j^m L_j) u_j &\geq 0 \\
u_0'(-s) + \sum_{j=1}^{P} u_j'(-s) = (\sum_{j=0}^{P} u_j)'(-s) &= -\|s\|_2^2 \geq 0.
\end{aligned}$$

($\Leftarrow$) Fix $\boldsymbol{u} = (u_0, \ldots, u_P)$ such that u_0 is in the kernel of $(\sum_{m=1}^{q_T} \lambda_0^m L_0^m)$ and u_j is in the kernel of $(\sum_{m=1}^{q_s} \lambda_j^m L_j^m)$ for $j = 1, \ldots, P$, and $\sum_j^P u_j = 0$. For $j = 0$,

$$A_0 u_0 + \sum_{i \neq 0}^{P} u_i = (A_0 - I)u_0 + \sum_{i=0}^{P} u_i = \left(\sum_{m=1}^{q_T} \lambda_0^m L_0^m \right) u_0 + \sum_{i=0}^{P} u_i = 0.$$

This proves that $\boldsymbol{u} = (u_0, \ldots, u_P) \in \ker(H)$ if and only if 1) u_0 is in the kernel of $(\sum_{m=1}^{q_T} \lambda_0^m L_0^m)$ and u_j is in the kernel of $(\sum_{m=1}^{q_s} \lambda_j^m L_j^m)$ for $j = 1, \ldots, P$, and $\sum_j^P u_j = 0$.

Lastly, $\ker(\sum_{m=1}^{q_T} \lambda_0^m L_0^m) = \bigcap_{m=0}^{q_T} \ker(L_0^m)$ and $\ker(\sum_{m=1}^{q_S} \lambda_j^m L_j^m) = \bigcap_{m=0}^{q_S} \ker(L_j^m)$ for $j > 1$.

($\Rightarrow$) Let's assume $u_0 \in \ker(\sum_{m=1}^{q_T} \lambda_0^m L_0^m)$. Because L_0^m is positive semi-definite, $u_0' \lambda_0^m L_0^m u_0 \geq 0$, $\forall m$. This requires that $u_0' \lambda_0^m L_0^m u_0 = 0$, $\forall m$. Thus $u_0 \in \bigcap_{m=1}^{q_T} \ker(L_0^m)$. The same argument can be made to show that $u_j \in \bigcap_{m=1}^{q_S} \ker(L_j^m) \forall j > 1$.

($\Leftarrow$) If $u_0 \in \bigcap_{m=1}^{q_T} \ker(L_0^m)$, then $u_0' \lambda_0^m L_0^m u_0 = 0$ for all m, thus $\sum_{m=1}^{q_T} u_0' \lambda_0^m L_0^m u_0 = 0$. $u_0 \in \ker(\sum_{m=1}^{q_T} \lambda_0^m L_0^m)$. The same argument applies to show $u_j \in \ker(\sum_{m=1}^{q_S} \lambda_j^m L_j^m)$ for $j > 0$.

Define $V_0 = \bigcap_{m=1}^{q_T} \ker(L_0^m)$ and $V_j = \bigcap_{m=1}^{q_S} \ker(L_j^m)$ for $j > 0$.

This proves that $\ker(H)$ consists of vectors from each null space of V_j, whose sum is zero:

$$\ker(H) = \{(u_0, \ldots, u_P) | \forall j, u_j \in V_j, \sum_{j=0}^{P} u_j = 0\}.$$

The condition above can be characterized by $V := \oplus_{j=0}^{P} V_j$ and define the sum map $\Phi : V \to \mathbb{R}^N$, where $\Phi(v_0, \ldots, v_P) = \sum_{j=0}^{P} v_j$. Therefore, $\ker(H) = \ker(\Phi)$. It follows that

$$\dim \ker(H) = \sum_{j=0}^{P} \dim \ker(V_j) \; - \; \dim \left(\sum_{j=0}^{P} \ker(V_j) \right).$$

By the nullity-rank relationship, the matrix is full-rank if and only if above equals to 0.

A.5 Proof of Corollary 2.1

Using the same notation as in Theorem 2, we first note that

$$V_0 = \ker(D_T) = \operatorname{span}\{\mathbf{1}, m\}, \quad m = (1, 2, \ldots, N)^\top,$$

$\dim V_0 = 2$. For each seasonal component S_j, the joint kernel of the second and seasonal differencing operators is

$$V_j = \ker(D^1_{S,j}) \cap \ker(D^2_{S,j}) = \operatorname{span}\{\mathbf{1}\},$$

following from Corollary 1.1 (Appendix A.2). Imposing the centering constraint $\mathbf{1}^\top S_j = 0$ eliminates this direction, so in fact $V_j = \{0\}$ for all $j \geq 1$. Hence

$$\sum_{j=0}^{P} \dim V_j = \dim V_0 = 2.$$

Moreover,

$$\sum_{j=0}^{P} V_j = V_0,$$

which also has dimension 2. Thus the condition of Theorem 2 holds with equality, and the decomposition is identifiable.

If we assume $D^2_{S_j}$ is the seasonal recurrence operator instead of the differencing operator

$$V_j = \ker(D^1_{S,j}) \cap \ker(D^2_{S,j}) = \{0\},$$

as shown in Corollary A.2. Therefore, the decomposition becomes identifiable.

B Full Conditional Distribution for Gibbs Sampling

B.1 Full Model

The exact parameterization and the prior distributions of BASTION used for both the simulation and empirical study section is described below:

Observation Equation

$$y_t = T_t + \sum_{i=1}^{P} S_{i,t} + \zeta_t + R_t, \qquad [R_t|\sigma_y, \nu_t] \overset{iid}{\sim} N(0, \sigma_y^2 \nu_t^2)$$

Trend

$$[T_t|\sigma_y, \eta_{T,t}] \sim N(0, \sigma_y^2 \eta_{T,t}^2) \qquad \forall t \in \{1, 2\},$$
$$[\Delta^2 T_t|\sigma_y, \tau_T, \eta_{T,t}] \sim N(0, \sigma_y^2 \tau_T^2 \eta_{T,t}^2) \qquad t \geq 3,$$
$$[\tau_T] \sim C^+(0, 1/N), \qquad [\eta_{T,t}] \overset{iid}{\sim} C^+(0, 1).$$

Multiple Seasonality

$$[S_{i,1}] = 0, \qquad [S_{i,2}|\sigma_y, \eta_{S_i,2}] \sim N(0, \sigma_y^2 \eta_{S_i,2}^2),$$
$$[\Delta^2 S_{i,t}|\sigma_y, \tau_{S_i}, \eta_{S_i,t}] \sim N(0, \sigma_y^2 \tau_{S_i}^2 \eta_{S_i,t}^2) \qquad \forall t \in \{3, \ldots, k_i\},$$
$$[(1-B)^{k_i} S_{i,t}|\sigma_y, \tau_{S_i}, \eta_{S_i,t}] \sim N(0, \sigma_y^2 \tau_{S_i}^2 \eta_{S_i,t}^2) \qquad \forall t \geq (k_i + 1),$$
$$[\tau_{S_i}] \sim C^+(0, 1/N), \qquad [\eta_{S_i,t}] \overset{iid}{\sim} C^+(0, 1).$$

Outliers

$$[\zeta_t|\sigma_y, \eta_{\zeta,t}] \sim N(0, \sigma_y^2 \eta_{\zeta,t}^2), \qquad [\eta_{\zeta,t}|\tau_\zeta, \xi_{\zeta,t}] \sim C^+(0, \tau_\zeta \xi_{\zeta,t})$$
$$[\tau_\zeta] \sim C^+(0, 1), \qquad [\xi_{\zeta,t}] \overset{iid}{\sim} C^+(0, 1).$$

Remainder

$$[\sigma_y^2] \sim \sigma_y^{-2} d\sigma_y^2$$
$$\log(\nu_t^2) = \mu + \phi(\log(\nu_{t-1}^2) - \mu) + \sigma_\nu \epsilon_t, \qquad [\epsilon_t] \overset{iid}{\sim} N(0, 1)$$
$$[\mu] \sim N(0, 100), \qquad [\phi] \sim \text{Beta}(5, 1.5)$$
$$[\sigma_\nu^2] \sim IG(1/2, 1/2)$$

Let's first define $\boldsymbol{Y} = [y_1, \ldots, y_N]'$ and similarly for the variables $\boldsymbol{T}, \boldsymbol{S_i}, \boldsymbol{\zeta}, \boldsymbol{\nu}$. The full conditional posterior distributions for Gibbs sampling are derived in the following subsections.

B.2 Trend

For the likelihood, we have:

$$[\boldsymbol{Y}|\boldsymbol{T}, \ldots] \sim N\left(\boldsymbol{T} + \sum_{i=1}^{P} \boldsymbol{S_i} + \boldsymbol{\zeta}, \sigma_y^2 \boldsymbol{\nu^2} I\right).$$

The prior distribution on $\boldsymbol{T}$ are imposed on its second differencing.

$$D_T = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}.$$

Thus, given $\boldsymbol{\sigma_T^2} := [\sigma_y^2\eta_{T,1}^2, \sigma_y^2\eta_{T,2}^2, \sigma_y^2\tau_T^2\eta_{T,3}^2, \ldots, \sigma_y^2\tau_T^2\eta_{T,N}^2]'$

$$\begin{aligned} [\boldsymbol{T}|\ldots] &\sim N(0, A\boldsymbol{\sigma_T^2}IA') \\ &\sim N\left(0, \left(D_T'\frac{\mathbf{1}}{\boldsymbol{\sigma_T^2}}ID_T\right)^{-1}\right). \end{aligned}$$

Let's define $Q_T := (D_T)'\frac{\mathbf{1}}{\boldsymbol{\sigma_T^2}}I(D_T)$.

$$Q_T = \frac{1}{\sigma_y^2\tau_T^2} \begin{bmatrix} \left(\frac{\tau_T^2}{\eta_{T,1}^2}+\frac{1}{\eta_{T,3}^2}\right) & -\frac{2}{\eta_{T,3}^2} & -\frac{1}{\eta_{T,3}^2} & 0 & \ldots & \ldots & 0 \\ -\frac{2}{\eta_{T,3}^2} & \left(\frac{\tau_T^2}{\eta_{T,2}^2}+\frac{4}{\eta_{T,3}^2}+\frac{1}{\eta_{T,4}^2}\right) & \frac{-2}{\eta_{T,3}^2}-\frac{2}{\eta_{T,4}^2} & -\frac{1}{\eta_{T,4}^2} & \ddots & \ddots & \vdots \\ -\frac{1}{\eta_{T,3}^2} & \frac{-2}{\eta_{T,3}^2}-\frac{2}{\eta_{T,4}^2} & \left(\frac{1}{\eta_{T,3}^2}+\frac{4}{\eta_{T,4}^2}+\frac{1}{\eta_{T,5}^2}\right) & \frac{-2}{\eta_{T,4}^2}-\frac{2}{\eta_{T,5}^2} & -\frac{1}{\eta_{T,5}^2} & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \left(\frac{1}{\eta_{T,N-2}^2}+\frac{4}{\eta_{T,N-1}^2}+\frac{1}{\eta_{T,N}^2}\right) & \left(\frac{-2}{\eta_{T,N-1}^2}-\frac{2}{\eta_{T,N}^2}\right) & -\frac{1}{\eta_{T,N}^2} \\ \vdots & \ddots & \ddots & \ddots & \left(\frac{-2}{\eta_{T,N-1}^2}-\frac{2}{\eta_{T,N}^2}\right) & \left(\frac{1}{\eta_{T,N-1}^2}+\frac{4}{\eta_{T,N}^2}\right) & -\frac{2}{\eta_{T,N}^2} \\ 0 & \ldots & \ldots & \ldots & -\frac{1}{\eta_{T,N}^2} & -\frac{2}{\eta_{T,N}^2} & \frac{1}{\eta_{T,N}^2} \end{bmatrix}$$

Therefore,

$$[\boldsymbol{T}|\boldsymbol{y}, \ldots] \sim N\left(\left(Q_T + \frac{1}{\sigma_y^2\boldsymbol{\nu^2}}I\right)^{-1}\left(\frac{\boldsymbol{Y} - \sum_{i=1}^P S_i - \boldsymbol{\zeta}}{\sigma_y^2\boldsymbol{\nu^2}}\right), \left(Q_T + \frac{1}{\sigma_y^2\boldsymbol{\nu^2}}I\right)^{-1}\right).$$

For the global parameter τ_T and the local parameter $\eta_{T,t}$ both following the half-Cauchy distribution, the parameter expansion as described in Makalic and Schmidt (2016) are used. Thus,

$$\begin{aligned} [\tau_T|\psi_{\tau_T}, \ldots] &\sim IG(1/2, 1/\psi_{\tau_T}) & [\psi_{\tau_T}] &\sim IG(1/2, 1) \\ [\eta_{T,t}|\psi_{\eta_T,t} \ldots] &\sim IG(1/2, 1/\psi_{\eta_T,t}) & [\psi_{\eta_T,t}] &\sim IG(1/2, 1) \end{aligned}$$

Since they are conjugate priors, the conditional prior distributions are as follows:

$$
\begin{aligned}
[\tau_T^2|\boldsymbol{Y},\psi_{\tau_T},\ldots] &\sim IG\left(\frac{1}{2}+\frac{N-2}{2},\frac{1}{\psi_{\tau_T}}+\frac{1}{2\sigma_y^2}\sum_{t=3}^{N}\left(\frac{\Delta^2 T_t}{\eta_{T,t}}\right)^2\right).\\
[\psi_{\tau_T}|\boldsymbol{Y},\tau_T,\ldots] &\sim IG\left(1,1+\frac{1}{\tau_T^2}\right).\\
[\eta_{T,t}^2|\boldsymbol{Y},\psi_{\eta_T,t}\ldots] &\sim IG\left(1,\frac{1}{\psi_{\eta_T,t}}+\frac{1}{2}\left(\frac{T_t}{\sigma_y}\right)^2\right), && t\in\{1,2\}.\\
&\sim IG\left(1,\frac{1}{\psi_{\eta_T,t}}+\frac{1}{2}\left(\frac{\Delta^2 T_t}{\sigma_y\tau_T}\right)^2\right), && t\geq 3.\\
[\psi_{\eta_T,t}|\boldsymbol{Y},\eta_{T,t},\ldots] &\sim IG\left(1,1+\frac{1}{\eta_{T,t}^2}\right).
\end{aligned}
$$

B.3 Seasonality

Let's fix $j\in\{1,\ldots,P\}$, and consider the seasonality term $\boldsymbol{S_j}$. Let k_j denote the length of the cycle and to simplify notation let $\boldsymbol{S}:=\boldsymbol{S}_j$. The seasonality parameter is penalized by both the seasonal differencing for the first k component and are also penalized by the seasonal differencing operator:

$$
D_S=(\begin{bmatrix} I_{k-1} & 0\\ B_3 & B_2\end{bmatrix}\begin{bmatrix} B_1 & 0\\ 0 & I_{N-k},\end{bmatrix})^{-1}
$$

where each matrix $B_{k,1}$, $B_{k,2}$ has the following structure:

$$
B_1=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0\\ 2 & 1 & 0 & \cdots & 0\\ 3 & 2 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ k-1 & k-2 & k-3 & \cdots & 1\end{bmatrix}\qquad B_2=\begin{bmatrix} I_k & 0 & 0 & \cdots & 0\\ I_k & I_k & 0 & \cdots & 0\\ I_k & I_k & I_k & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ I_k & I_k & I_k & \cdots & I_k\end{bmatrix}\qquad B_3=\begin{bmatrix} \mathbf{0}\\ I_{k-1}\\ \mathbf{0}\\ I_{k-1}\\ \mathbf{0}\\ \vdots\end{bmatrix}.
$$

B_1 is a $(k-1)\times(k-1)$ lower triangle matrix for un-differencing the second differencing operator on the first $k-1$ observations. B_2 is a $(N-k)\times(N-k)$ lower triangle matrix and B_3 is a $(N-k)\times(k-1)$ matrix for un-differencing the seasonal differencing operator on the last $T-k$ observations. Similar to the derivation of Q_T in Section B.2, $Q_S^{-1}:=\left((D_S)'\frac{\mathbf{1}}{\boldsymbol{\sigma_S^2}}ID_S\right)$, where $\boldsymbol{\sigma_S^2}:=[\sigma_y^2\eta_{S,2}^2,\sigma_y^2\tau_S^2\eta_{S,3}^2,\sigma_y^2\tau_S^2\eta_{S,4}^2,\ldots,\sigma_y^2\tau_S^2\eta_{S,N}^2]'$. The precision matrix Q_S has a block-tridiagonal structure with blocks separated by the seasonality k except for the top left $(k-1)\times(k-1)$ submatrix, which has a tridiagonal structure resulting from the second differencing.

The conditional posterior distribution for $\boldsymbol{S}$ therefore,

$$
[\boldsymbol{S}|\boldsymbol{y},\ldots]\sim N\left(\left(Q_S+\frac{1}{\sigma_y^2\boldsymbol{\nu}^2}I\right)^{-1}\left(\frac{\boldsymbol{Y}-\boldsymbol{T}-\sum_{i\neq j}\boldsymbol{S_i}-\boldsymbol{\zeta}}{\sigma_y^2\boldsymbol{\nu}^2}\right),\left(Q_S+\frac{1}{\sigma_y^2\boldsymbol{\nu}^2}I\right)^{-1}\right).
$$

The conditional posterior distributions of τ_S and $\boldsymbol{\eta_S}$ are identical to τ_T and $\boldsymbol{\eta_T}$ explored in Section B.2. With

the parameter expansion of the horseshoe prior:

$$
\begin{aligned}
[\tau_S^2|\boldsymbol{Y}, \psi_{\tau_S}, \ldots] &\sim IG\left(\frac{1}{2} + \frac{N-2}{2}, \frac{1}{\psi_{\tau_S}} + \frac{1}{2\sigma_y^2}\left(\sum_{t=3}^{k}\left(\frac{\Delta^2 S_t}{\eta_{T,t}}\right)^2 \qquad + \sum_{t=(k+1)}^{N}\left(\frac{(1-B^k)S_t}{\eta_{T,t}}\right)^2\right)\right). \\
[\psi_{\tau_S}|\boldsymbol{Y}, \tau_S, \ldots] &\sim IG\left(1, 1 + \frac{1}{\tau_S^2}\right). \\
[\eta_{S,t}^2|\boldsymbol{Y}, \psi_{\eta_S,t} \ldots] &\sim IG\left(1, \frac{1}{\psi_{\eta_S,t}} + \frac{1}{2}\left(\frac{S_t}{\sigma_y}\right)^2\right), && t = 2, \\
&\sim IG\left(1, \frac{1}{\psi_{\eta_S,t}} + \frac{1}{2}\left(\frac{\Delta^2 S_t}{\sigma_y \tau_S}\right)^2\right), && t = 3, \ldots, k, \\
&\sim IG\left(1, \frac{1}{\psi_{\eta_S,t}} + \frac{1}{2}\left(\frac{(1-B^k)S_t}{\sigma_y \tau_S}\right)^2\right), && t = (k+1), \ldots, N. \\
[\psi_{\eta_S,t}|\boldsymbol{Y}, \eta_{S,t}, \ldots] &\sim IG\left(1, 1 + \frac{1}{\eta_{S,t}^2}\right).
\end{aligned}
$$

B.4 Outliers

For the additive outlier term ζ_t, we use the horseshoe+ prior by Bhadra et al. (2015) to provide more shrinkage than the one induced by the horseshoe prior. The parameter expansion by Makalic and Schmidt (2016) can be applied to the horseshoe+ prior:

$$
\begin{aligned}
&[\zeta_t|\sigma_y, \eta_{\zeta,t}] \sim N(0, \sigma_y^2 \eta_{\zeta,t}^2), && [\eta_{\zeta,t}^2|\psi_{\eta_\zeta,t}] \sim IG(1/2, 1/\psi_{\eta_\zeta,t}), \\
&[\psi_{\eta_\zeta,t}|\tau_\zeta, \xi_{\zeta,t}] \sim IG(1/2, 1/(\tau_\zeta^2 \xi_{\zeta,t}^2)), \\
&[\tau_\zeta^2|\psi_{\tau_\zeta}] \sim IG(1/2, 1/\psi_{\tau_\zeta}), && [\psi_{\tau_\zeta}] \sim IG(1/2, 1), \\
&[\xi_{\zeta,t}^2|\psi_{\xi_\zeta,t}] \sim IG(1/2, 1/\psi_{\xi_\zeta,t}), && [\psi_{\xi_\zeta,t}] \sim IG(1/2, 1).
\end{aligned}
$$

Thus, we have the following conditional posterior distribution:

$$
\begin{aligned}
&[\boldsymbol{\zeta}|\boldsymbol{y}, \ldots] \sim N\left(\left(\frac{\boldsymbol{\eta_\zeta^2}}{\boldsymbol{\nu^2} + \boldsymbol{\eta_\zeta^2}}\right)\left(\boldsymbol{Y} - \boldsymbol{T} - \sum_{i=1}^{P}\boldsymbol{S}_i\right), \left(\frac{\sigma_y^2 \boldsymbol{\nu^2 \eta_\zeta^2}}{\boldsymbol{\nu^2} + \boldsymbol{\eta_\zeta^2}}\right)\right), \\
&[\eta_{\zeta,t}^2|\boldsymbol{y}, \ldots] \sim IG\left(1, \frac{1}{\psi_{\eta_\zeta,t}} + \frac{1}{2}\left(\frac{\zeta_t}{\sigma_y}\right)^2\right) \\
&[\psi_{\eta_\zeta,t}|\boldsymbol{y}, \ldots] \sim IG\left(1, \frac{1}{\eta_{\zeta,t}} + \frac{1}{\tau_\zeta^2 \xi_{\zeta,t}^2}\right) \\
&[\tau_\zeta^2|\boldsymbol{y}, \ldots] \sim IG\left(\frac{N+1}{2}, \sum_{t=1}^{N}\left(\frac{1}{\xi_{\zeta,t}^2 \psi_{\eta_\zeta,t}}\right) + \frac{1}{\psi_{\tau_\zeta}}\right) \\
&[\psi_{\tau_\zeta}|\boldsymbol{y}] \sim IG\left(1, 1 + \frac{1}{\tau_\zeta^2}\right) \\
&[\xi_{\zeta,t}^2|\boldsymbol{y}, \ldots] \sim IG(\left(1, \frac{1}{\tau_\zeta^2 \psi_{\eta_\zeta,t}} + \frac{1}{\psi_{\xi_\zeta,t}}\right) \\
&[\psi_{\xi_\zeta,t}|\boldsymbol{y}] \sim IG\left(1, 1 + \frac{1}{\xi_{\zeta,t}^2}\right)
\end{aligned}
$$

B.5 Remainders

Remainder term also decomposes into two components, the time-invariant σ_y, and the time-varying term ν_t. For the time-invariant term, we have

$$\sigma_y^2 \sim IG\Bigg(\frac{3N+P(N-1)}{2}, \frac{1}{2}\Bigg(\Bigg(\frac{\boldsymbol{Y}-\boldsymbol{T}-\sum_{i=1}^{P}\boldsymbol{S_i}-\zeta}{\boldsymbol{\nu}}\Bigg)^2 + \sum_{t=1}^{2}\left(\frac{T_t}{\eta_{T,t}}\right)^2 + \sum_{t=3}^{N}\left(\frac{\Delta^2 T_t}{\tau_T \eta_{T,t}}\right)^2 + \sum_{i=1}^{P}\Bigg(\left(\frac{S_{i,2}}{\eta_{S_i,2}}\right)^2 + \sum_{t=3}^{k}\left(\frac{\Delta^2 S_{i,t}}{\tau_{S_i}\eta_{S_{i,t}}}\right)^2 + \sum_{k+1}^{N}\left(\frac{(1-B)^{k_i} S_{i,t}}{\tau_{S_i}\eta_{S_{i,t}}}\right)^2\Bigg) + \sum_{t=1}^{N}\left(\frac{\zeta_t}{\eta_{\zeta,t}}\right)^2\Bigg)\Bigg)$$

For the time-varying term $\boldsymbol{\nu}$, we closely follow the sampler proposed by Kastner and Frühwirth-Schnatter (2014). Define

$$\boldsymbol{Y}^* = log\left(\frac{\boldsymbol{Y}-\boldsymbol{T}-\sum_{i=1}^{P}\boldsymbol{S_i}-\boldsymbol{\zeta}}{\sigma_y}\right)^2$$
$$\boldsymbol{h} := log(\boldsymbol{\nu}^2).$$

By transforming the likelihood, we have a linear system with a non-Gaussian error term, which may be approximated by a Gaussian mixture distribution Omori et al. (2007):

$$\begin{aligned}
&y_t^* = h_t + \log(u_t^2), && u_t \overset{iid}{\sim} N(0,1),\\
&y_t^* \approx h_t + \mu_{j_t} + \sigma_{j_t} u_t, && u_t \overset{iid}{\sim} N(0,1),\ \ j_t \overset{iid}{\sim} \text{Categorical}(\pi)\\
&h_t = \mu + \phi(h_{t-1}-\mu) + \sigma_\nu \epsilon && [\epsilon_t] \overset{iid}{\sim} N(0,1)
\end{aligned}$$

With the approximation of the likelihood, the model becomes conditionally Gaussian, allowing efficient sampling of associated parameter $\boldsymbol{h}, \mu, \phi, \sigma_\nu$, and the additional parameter $\boldsymbol{j}$.

C Simulation Schemes

C.1 Data Generating Scheme 1

Observation Equation

$$Y_t = T_t + S_t^{12} + S_t^{40} + R_t.$$

Trend

$$T_t = \begin{cases} m_1(0.04)(t-b_1)+c_1, & \text{if } 0 \le t < b_1,\\ m_2(0.04)(t-b_2)+c_2, & \text{if } b_1 \le t < b_1+b_2,\\ m_3(0.04)(t-b_3)+c_3, & \text{if } b_1+b_2 \le t < b_1+b_2+b_3,\\ m_4(0.04)(t-b_4)+c_4, & \text{if } b_1+b_2+b_3 \le t \le 500, \end{cases}$$

$$[m_1,\ldots,m_4] \overset{iid}{\sim} U(-20,20),$$
$$[c_1,\ldots,c_4] \overset{iid}{\sim} U(-10,10),$$
$$[b_1,\ldots,b_3] \overset{iid}{\sim} U(30,125).$$

Seasonal

$$
\begin{aligned}
&S_t^{12} = \gamma_1^{12} \sin\left(\frac{2\pi t}{12}\right) + \gamma_2^{12} \cos\left(\frac{2\pi t}{12}\right) \\
&[\gamma_1^{12}, \gamma_2^{12}] \overset{iid}{\sim} N(0, 4^2), \\
&S_t^{40} = \gamma_1^{40} \sin\left(\frac{2\pi t}{40}\right) + \gamma_2^{40} \cos\left(\frac{2\pi t}{40}\right) \\
&[\gamma_1^{40}, \gamma_2^{40}] \overset{iid}{\sim} N(0, 5^2).
\end{aligned}
$$

Error

$$[R_t] \overset{iid}{\sim} N(0, 2^2)$$

C.2 Data Generating Scheme 2

Observation Equation

$$Y_t = T_t + S_t^{40} + R_t.$$

Trend

$$
\begin{aligned}
T_t &= \frac{mt}{500} \\
[m] &\sim N(0, 30^2).
\end{aligned}
$$

Seasonal

$$
\begin{aligned}
&S_t = \begin{cases} \tilde{v}_1, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 1 \le j \le 10\}, \\ \tilde{v}_2, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 11 \le j \le 20\}, \\ \tilde{v}_3, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 21 \le j \le 30\}, \\ \tilde{v}_4, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 31 \le j \le 40\}, \end{cases} \\
&[v_1, \ldots, v_4] \overset{iid}{\sim} U\left(-8, 8\right), \\
&\tilde{v}_i = v_i - \frac{1}{4} \sum_{j=1}^{4} v_j.
\end{aligned}
$$

Error

$$[R_t] \overset{iid}{\sim} N\left(0, \frac{m^2}{100}\right).$$

C.3 Data Generating Scheme 3

Observation Equation

$$Y_t = T_t + S_t^{50} + R_t.$$

Trend

$$
\begin{aligned}
&T_t = b_0 + b_1 \frac{t}{500} + b_2 \left(\frac{t}{500}\right)^2 + b_3 \left(\frac{t}{500}\right)^3 \\
&[b_0] \sim U(-15, 15), \\
&[b_1, \ldots, b_4] \overset{iid}{\sim} N(0, 20^2).
\end{aligned}
$$

Seasonal

$$S_t^{50} = \gamma_1^{50} \sin\left(\frac{2\pi t}{50}\right) + \gamma_2^{50} \cos\left(\frac{2\pi t}{50}\right)$$
$$[\gamma_1^{50}, \gamma_2^{50}] \overset{iid}{\sim} N(0, 5^2).$$

Error

$$[R_t] \overset{iid}{\sim} N(0, exp(h_t))$$
$$h_t = 2.5 + 0.98(h_{t-1} - 2.5) + 0.2\epsilon_t$$
$$[\epsilon_t] \overset{iid}{\sim} N(0, 1)$$

C.4 Data Generating Scheme 4

Observation Equation

$$Y_t = T_t + S_t^{12} + S_t^{50} + O_t + R_t, \qquad t = 1, \ldots, 500.$$

Trend

$$T_t = \begin{cases} m_1(0.04)(t - b_1) + c_1, & \textbf{if } 0 \le t < b_1, \\ m_2(0.04)(t - b_2) + c_2, & \textbf{if } b_1 \le t < b_1 + b_2, \\ m_3(0.04)(t - b_3) + c_3, & \textbf{if } b_1 + b_2 \le t < b_1 + b_2 + b_3, \\ m_4(0.04)(t - b_4) + c_4, & \textbf{if } b_1 + b_2 + b_3 \le t \le 500, \end{cases}$$
$$[m_1, \ldots, m_4] \overset{iid}{\sim} U(-20, 20),$$
$$[c_1, \ldots, c_4] \overset{iid}{\sim} U(-10, 10),$$
$$[b_1, \ldots, b_3] \overset{iid}{\sim} U(30, 125).$$

Seasonal

$$S_t^{12} = \gamma_1^{12} \sin\left(\frac{2\pi t}{12}\right) + \gamma_2^{12} \cos\left(\frac{2\pi t}{12}\right)$$
$$[\gamma_1^{12}, \gamma_2^{12}] \overset{iid}{\sim} N(0, 4^2),$$
$$S_t^{50} = \begin{cases} \tilde{v}_1, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 1 \le j \le 10\}, \\ \tilde{v}_2, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 11 \le j \le 20\}, \\ \tilde{v}_3, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 21 \le j \le 30\}, \\ \tilde{v}_4, & \textbf{if } t \in \{t \mid t = 40k + j, 0 \le k \le 8, 31 \le j \le 40\}, \end{cases}$$
$$[v_1, \ldots, v_4] \overset{iid}{\sim} U\left(-20, 20\right),$$
$$\tilde{v}_i = v_i - \frac{1}{4}\sum_{j=1}^{4} v_j$$

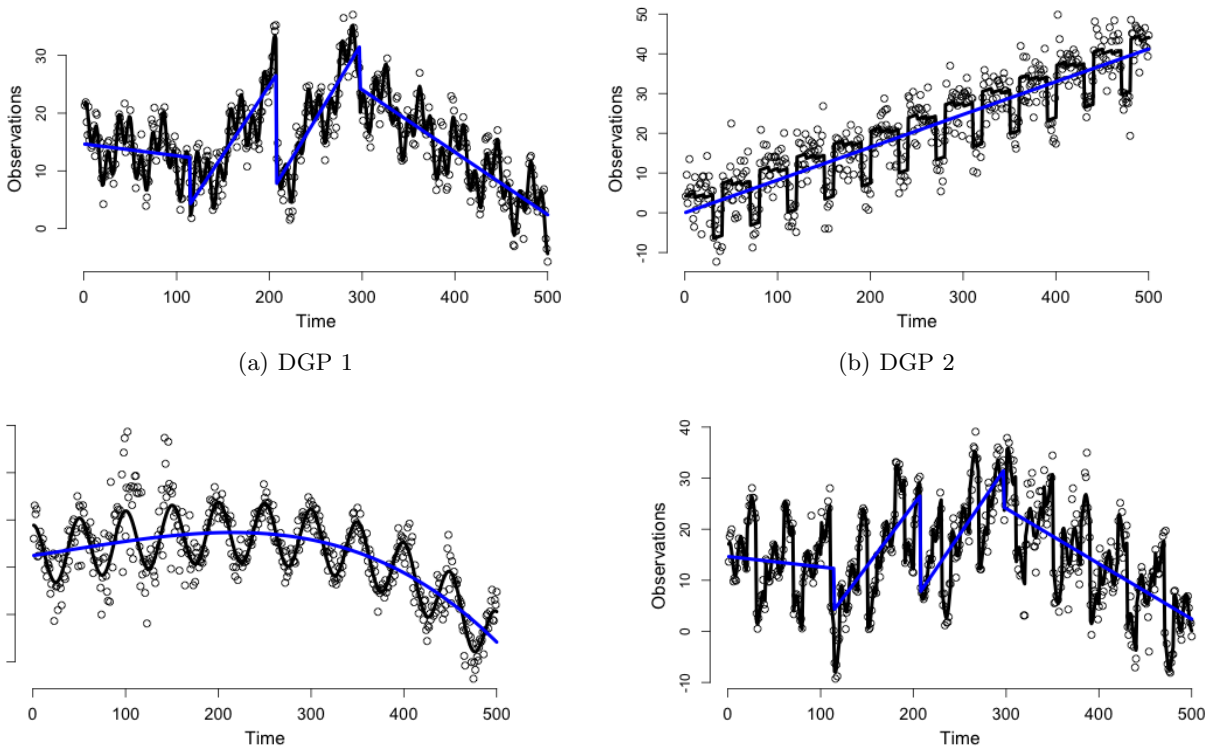
Outlier

$$O_t = \begin{cases} a_t, & \textbf{if } t \in \{j_1, j_2, \ldots, j_l\}, \\ 0, & \textbf{if } t \notin \{j_1, j_2, \ldots, j_l\}, \end{cases}$$
$$a_t \overset{iid}{\sim} N(0, 15^2),$$
$$l \sim Pois(5),$$
$$[j_1, j_2, \ldots, j_l] \sim U(1, 500) \textbf{ without replacement}.$$

Error

$$[R_t] \overset{iid}{\sim} N(0, exp(h_t))$$
$$h_t = 0.8 + 0.98(h_{t-1} - 2.5) + 0.2\epsilon_t$$
$$[\epsilon_t] \overset{iid}{\sim} N(0, 1)$$

D Additional Figures



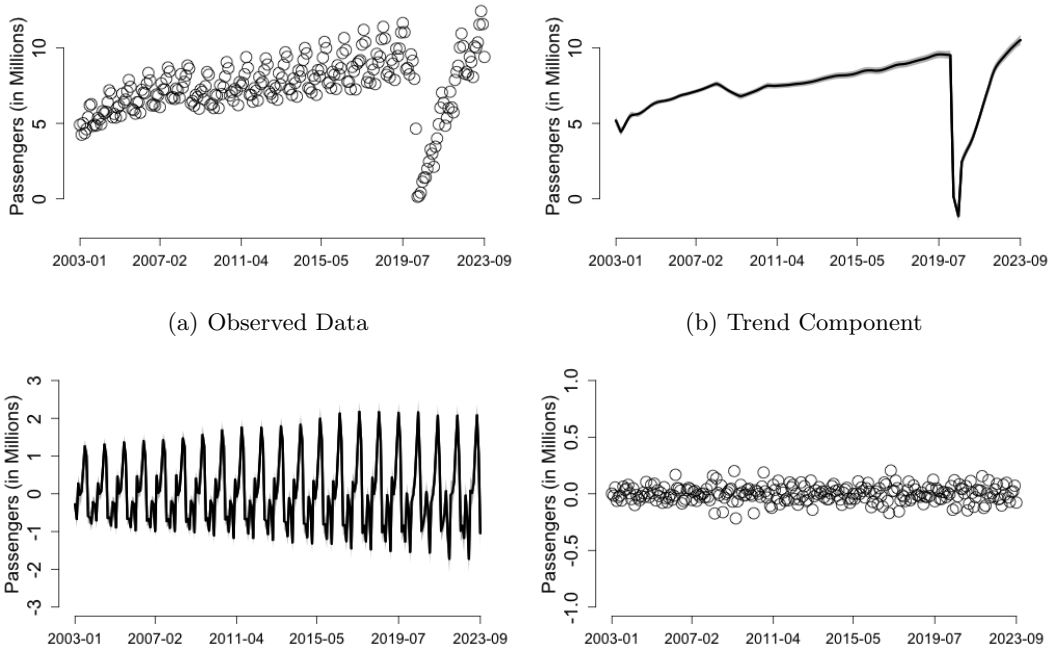
(a) DGP 1

(b) DGP 2

(c) DGP 3

(d) DGP 4

Figure 5: Synthetic time series generated by adding a trend component, T_t, seasonal components, S_t, and remainders R_t based on the descriptions in Table 1. T_t are drawn in blue and $T_t + S_t$ are drawn in black. Figures represent one replication of each DGP.



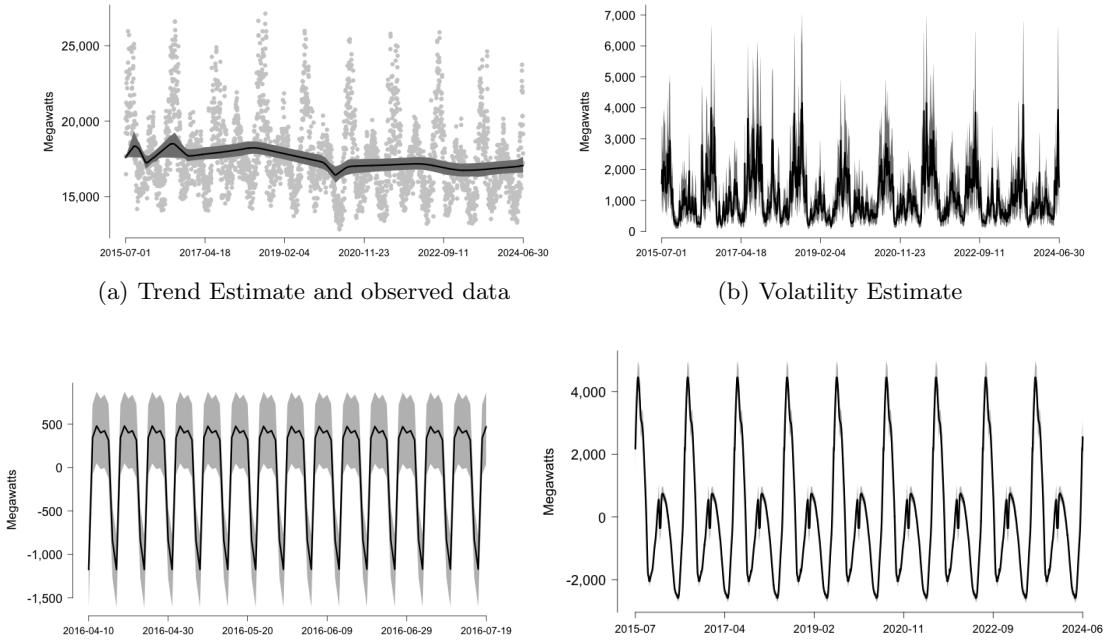
(a) Observed Data

(b) Trend Component

(c) Seasonality Component

(d) Remainder

Figure 6: Monthly international airline traffic in the U.S from 2003 to 2023 6(a), and its Trend 6(b), Seasonality 6(c), and remainder Figure 6(d) decomposition based on the proposed model, BASTION. 95% credible regions are generated and shaded in grey for the trend and seasonality components.



(a) Trend Estimate and observed data

(b) Volatility Estimate

(c) Weekly Seasonality

(d) Yearly Seasonality

Figure 7: Decomposition of the trend 7(a), volatility 4, weekly seasonality 7(c), and yearly seasonality 7(d) based on BASTION for daily average electricity demand from 2015-07-01 to 2024-06-30. 95% credible regions are drawn in dark grey.