Probabilistic Seasonality

Feras Saad* Todd B. Walker†

March 2026

Abstract

Seasonal adjustment is a foundational step in economic time series analysis, as it removes recurring calendar-driven fluctuations to reveal underlying trends, business cycles, and irregular components. This paper introduces a novel probabilistic approach to seasonal adjustment based on a Gaussian process (GP) framework for automated structure discovery in time series. Rather than imposing a fixed parametric form for seasonality, we treat the seasonal component as part of a flexible, data-driven GP decomposition. Seasonal adjustments emerge as full posterior distributions over latent seasonal functions, naturally accommodating non-stationary and evolving patterns while providing coherent uncertainty quantification for both adjusted values and future revisions. Monte Carlo simulations—based on a growing-sine process mimicking classic economic benchmarks such as Air Passengers Dataset—demonstrate that our approach recovers the true seasonal structure more reliably than X-13ARIMA-SEATS in weak and strong seasonality regimes, yields lower mean squared forecast errors for the observed series, and produces better-calibrated predictive intervals. In an application to monthly U.S. Advance Retail Sales (1992–2025), the method delivers stable seasonally adjusted series with posterior credible intervals that widen during recessionary periods, transparently reflecting heightened uncertainty in distinguishing seasonal from cyclical movements.

*Department of Computer Science, Carnegie Mellon University, fsaad@cmu.edu.
†Department of Economics, Indiana University, walkertb@iu.edu

1 INTRODUCTION

Many economic time series—such as retail sales, industrial production, employment, and inflation—exhibit strong and often complex seasonal patterns driven by calendar effects, weather, holidays, and institutional factors. Removing these recurring fluctuations allows analysts, policymakers, and forecasters to isolate underlying trends, business cycles, and irregular movements more clearly. Accurate seasonal adjustment is therefore essential for real-time monitoring of economic conditions, the construction of reliable leading indicators, and the evaluation of policy.

Seasonal adjustments are extraordinarily consequential as the magnitude of within-year variation in key series, such as nonfarm payroll employment, can exceed the effects of recessions (Wright, 2013). For example, the average absolute difference between seasonally adjusted and unadjusted monthly changes in payrolls dwarfs typical month-to-month variation in the adjusted data. Poorly executed adjustment risks misattributing cyclical movements to seasonality (or vice versa), leading to distorted interpretations of economic strength or weakness, larger and less predictable data revisions, and misguided forecasts.

The X-13ARIMA-SEATS program developed and maintained by the U.S. Census Bureau remains the standard for seasonal adjustment of official statistics (U.S. Census Bureau, 2024). This software combines nonparametric filtering procedures (an evolution of the classic X-11 method) with parametric ARIMA-based modeling via the SEATS procedure (Gómez and Maravall, 2001b). It has proven robust and scalable across thousands of series, yet relies on relatively rigid parametric assumptions and prespecified filter structures. When seasonal patterns are time-varying, the approach can produce suboptimal decompositions, larger revisions upon data updates, forecast errors that understate uncertainty, and persistent seasonal "echoes" in the adjusted series (Bell and Hillmer, 1984; Watson, 1987; Wright, 2013).

This paper introduces a probabilistic approach to seasonal adjustment based on a Bayesian nonparametric framework for automated structure discovery in time series [AutoGP, Saad et al. (2023)]. AutoGP works within a rich space of Gaussian process (GP) covariance kernels that are composed compositionally from simple base kernels (linear, periodic, exponential, etc.) using operators such as addition, multiplication, and change-points. A sequential Monte Carlo (SMC) algorithm efficiently explores this symbolic model space and infers posterior distributions over both kernel structures and hyperparameters.

The core contribution of this paper lies in identifying and modeling the seasonal component as an integral part of a flexible, data-driven Gaussian process (GP) decomposition. Rather than imposing a pre-specified parametric form or fixed filter for seasonality, our framework treats the entire time series as arising from a compositionally constructed GP kernel that can automatically discover latent structures, including non-stationary and evolving seasonal patterns. This novel approach allows seasonality to interact flexibly with trend and irregular components through symbolic kernel expressions (e.g., periodic kernels modulated by change-points or multiplied by slowly varying functions), all inferred probabilistically without strong prior assumptions on the form of seasonality. Seasonal adjustments then emerge naturally as posterior inferences over the latent seasonal function, yielding not only point estimates but full predictive distributions over the adjusted series and future revisions.

This probabilistic treatment offers several key advantages:

- It automatically discovers appropriate kernel structures tailored to the specific dynamics of each series, including complex interactions between seasonal and non-seasonal components that standard methods often miss or misattribute.
- It provides coherent, well-calibrated uncertainty quantification for both the seasonally adjusted values and the magnitude of future revisions, supporting more informed decision-making under ambiguity.

- It substantially reduces the need for manual intervention and model tuning while preserving interpretability—the learned kernel expressions can be inspected and linked directly to economic features, such as a periodic kernel whose amplitude or frequency shifts at a change-point corresponding to a policy reform or behavioral change.

We evaluate the method through a Monte Carlo simulation and an application to monthly U.S. Advance Retail Sales (RSAFSNA, January 1992–August 2025).

In the simulation exercise, data are generated from a growing-sine process that replicates key features of the classic Air Passengers Dataset: a linear trend, a seasonal sinusoid whose amplitude increases over time, and Gaussian noise. This design creates environments with weak, moderate, and strong seasonality (corresponding to training lengths $T = 36$, 72, and 144 months). Our approach consistently produces more accurate out-of-sample point forecasts of the observed series than X-13ARIMA-SEATS across nearly all horizons and sample sizes, approaching the performance of a correctly specified OLS benchmark at longer samples while delivering statistically significant reductions in mean squared error (often exceeding 5–10% relative improvement at $T = 144$). It also yields more efficient predictive intervals as measured by the Mean Interval Score (MIS). Performance on the latent seasonal component itself is more nuanced as our approach outperforms X-13 in short samples (where it appropriately down-weights seasonality when the signal is weak) and long samples (where its flexibility captures the growing amplitude), but X-13 holds an edge at intermediate lengths where a moderate seasonal signal rewards its more dogmatic structure. Crucially, our approach provides full posterior uncertainty over the seasonal factor while automatically discovering appropriate kernel structures.

Examining U.S. retail sales data, after pre-processing to remove trading-day effects, we show that our approach produces stable seasonally adjusted series. Time-domain diagnostics show that the adjusted series preserves the low-frequency evolution of retail activity while effectively removing within-year oscillations, with 95% posterior credible intervals that widen meaningfully during recessionary periods—reflecting state-dependent uncertainty in distinguishing seasonal from cyclical movements. Frequency-domain analysis confirms substantial attenuation of power at the annual seasonal frequency in the adjusted growth rates, while low-frequency components remain largely intact.

2 PROBABILISTIC SEASONALITY

Seasonal adjustment procedures used by government agencies—such as X-13ARIMA-SEATS—rely on a battery of frequentist specification tests to select a model prior to estimation. While each test may appear statistically valid in isolation, the overall procedure lacks a coherent inferential framework. Conversely, we show how seasonality modeled using Bayesian probabilistic model discovery avoids this incoherency by providing a unified inferential framework that jointly estimates both model structure and parameters—a natural definition of seasonality follows.

2.1 GENERAL FORMULATION

Let $\mathscr{M} = \{m_1, m_2, \ldots\}$ denote a space of possible statistical models, possibly infinite. Each model $m \in \mathscr{M}$ defines a likelihood function $f(\mathbf{y}; \theta, m)$ for observable data $\mathbf{y}$, parameterized by $\theta \in \Theta_m$. A prior distribution over models $\pi(m)$ and a prior distribution over parameters conditional on a model $\pi(\theta|m)$ induces the joint distribution

$$p(m, \theta, \mathbf{y}) = \pi(m)\pi(\theta|m) f(\mathbf{y}; \theta, m),$$

and in turn the following posterior distributions:

i. over models and parameters,

$$p(m,\theta \mid \mathbf{y}) = \frac{\pi(m)\pi(\theta \mid m) f(\mathbf{y};\theta,m)}{\sum_{m' \in \mathscr{M}} \int_{\Theta_{m'}} \pi(m')\pi(\theta' \mid m') f(\mathbf{y};\theta',m') \,\mathrm{d}\theta'} \qquad (m \in \mathscr{M},\ \theta \in \Theta_m) \tag{1}$$

ii. over the model spac,

$$p(m \mid \mathbf{y}) = \int_{\Theta_m} p(m,\theta \mid \mathbf{y}), \mathrm{d}\theta \qquad (m \in \mathscr{M}) \tag{2}$$

iii. the posterior over parameters, for specific model $m^\star \in \mathscr{M}$,

$$p(\theta \mid \mathbf{y}, m^\star) = \frac{\pi(\theta \mid m^\star) f(\mathbf{y};\theta,m^\star)}{\int_{\Theta_{m^\star}} \pi(\theta' \mid m^\star) f(\mathbf{y};\theta',m^\star) \,\mathrm{d}\theta'} \qquad (\theta \in \Theta_{m^\star}). \tag{3}$$

As the distributions of interest in Eqs. (1)–(3) are almost always intractable to compute exactly, they are instead approximated using Monte Carlo statistical methods (Robert and Casella, 2004), such as Markov chain Monte Carlo or sequential Monte Carlo. These algorithms generate a set $\mathscr{S} = \left\{\left(w^{(i)}, (m^{(i)}, \theta^{(i)})\right)\right\}_{i=1}^{N}$ of N weighted samples that form a discrete approximation of the target distribution. These samples should be properly weighted (Liu, 2004, §2.5.4) for the posterior, meaning that for any square integrable function φ over models, parameters and data,

$$\mathbb{E}\left[w^{(i)} \varphi\left(m^{(i)}, \theta^{(i)}, \mathbf{y}\right) \,\middle|\, \mathbf{y} \right] = c\, \mathbb{E}_{p(m,\theta \mid \mathbf{y})}\left[\varphi(m,\theta,\mathbf{y})\right] \qquad (1 \le i \le n),$$

where $c > 0$ is a constant common to all N samples. Equipped with properly weighted samples, consistent estimates of the expectation of φ can be computed via ratio estimation

$$\mathbb{E}_{p(m,\theta \mid \mathbf{y})}[\varphi(m,\theta;\mathbf{y})] \approx \sum_{i=1}^{N} \left(\frac{w^{(i)}}{\sum_{j=1}^{N} w^{(j)}} \right) \varphi(m^{(i)}, \theta^{(i)}, \mathbf{y}).$$

This framework enables flexible inference over both model structures and parameters and forms the basis for structural and numeric queries. A key challenge is designing a model space $\mathscr{M}$ that is both expressive and interpretable. Following Saad et al. (2023), our approach constructs $\mathscr{M}$ as a probabilistic context-free grammar (PCFG) over *Gaussian process* (GP) kernels, which describe the covariance structure of latent functions.

Gaussian Processes A Gaussian process (GP) over an index set $\mathbb{T}$ is a family $\{X(t) : t \in \mathbb{T}\}$ of random variables such that for any finite tuple $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{T}^n$ of $n \ge 1$ time indices, the random vector $\mathbf{X}(\mathbf{t}) \coloneqq (X(t_1), \ldots, X(t_n))$ follows a multivariate Gaussian distribution:

$$\mathbf{X}(\mathbf{t}) = \begin{pmatrix} X(t_1) \\ \vdots \\ X(t_n) \end{pmatrix} \sim \mathscr{N}\left(\mu(\mathbf{t}), K(\mathbf{t},\mathbf{t})\right), \tag{4}$$

where $\mu(\mathbf{t}) \coloneqq (\mu(t_1), \ldots, \mu(t_n))$, with $\mu(t) \coloneqq \mathbb{E}[X(t)]$, is the mean vector, and $K(\mathbf{t},\mathbf{t})$ is the $n \times n$ covariance matrix with entries $K_{ij} \coloneqq k(t_i,t_j) \coloneqq \mathbb{E}\left[(X(t_i) - \mu(t_i))(X(t_j) - \mu(t_j))\right]$ for all $1 \le i, j \le n$. Letting $\mathbf{x}(\mathbf{t}) \coloneqq (x(t_1), \ldots, x(t_n)) \in \mathbb{R}^n$ denote a possible realization of $\mathbf{X}(\mathbf{t})$, the log probability density is given by

$$\ln p(\mathbf{x}(\mathbf{t})) = -\tfrac{1}{2}\left[(\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t}))^\intercal K(\mathbf{t},\mathbf{t})^{-1} (\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t}))\right] - \tfrac{1}{2} \ln\det(K(\mathbf{t},\mathbf{t})) - \tfrac{n}{2}\ln(2\pi). \tag{5}$$

Gaussian distributions are closed under conditioning. In particular, conditioned on the event $\mathbf{X}(\mathbf{t}) = \mathbf{x}(\mathbf{t})$, the posterior distribution of $\mathbf{X}(\mathbf{t}')$ at new time points $\mathbf{t}' = (t'_1, \ldots, t'_m) \in \mathbb{T}^m$, is itself multivariate Gaussian:

$$\mathbf{X}(\mathbf{t}') \mid \mathbf{X}(\mathbf{t}) = \mathbf{x}(\mathbf{t}) \sim \mathcal{N}\left(\mu^{\text{post}}(\mathbf{t}'), K^{\text{post}}(\mathbf{t}', \mathbf{t}')\right), \tag{6}$$

$$\mu^{\text{post}}(\mathbf{t}') := \mu(\mathbf{t}') + K(\mathbf{t}', \mathbf{t}) K(\mathbf{t}, \mathbf{t})^{-1} (\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t})), \tag{7}$$

$$K^{\text{post}}(\mathbf{t}', \mathbf{t}') := K(\mathbf{t}', \mathbf{t}') - K(\mathbf{t}', \mathbf{t}) K(\mathbf{t}, \mathbf{t})^{-1} K(\mathbf{t}, \mathbf{t}'). \tag{8}$$

Economic and scientific data are often subject to measurement noise. We model noisy data by introducing a stochastic process $\mathbf{Y}$ that is the the sum of the latent GP and i.i.d. Gaussian noise, i.e., $\mathbf{Y}(\mathbf{t}) := \mathbf{X}(\mathbf{t}) + \boldsymbol{\epsilon}(\mathbf{t})$, where $\epsilon(t) \sim \mathcal{N}(0, \eta)$ for $t \in \mathbb{T}$. As the distribution of $\mathbf{Y}(\mathbf{t})$ is Gaussian with covariance matrix $K(\mathbf{t}, \mathbf{t}) + \eta I_n$, the log probability density is given by

$$\ln p(\mathbf{y}(\mathbf{t})) = -\tfrac{1}{2}\left[(\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t}))^{\intercal} \left[K(\mathbf{t}, \mathbf{t}) + \eta I_n\right]^{-1} (\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t}))\right] - \tfrac{1}{2} \ln \det\left(K(\mathbf{t}, \mathbf{t}) + \eta I_n\right) - \tfrac{n}{2} \ln(2\pi). \tag{9}$$

Conditioning on noisy observations $\mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t})$ yields the posterior for the latent vector $\mathbf{X}(\mathbf{t}')$:

$$\mathbf{X}(\mathbf{t}') \mid \mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t}) \sim \mathcal{N}\left(\mu^{\text{post}}(\mathbf{t}'), K^{\text{post}}(\mathbf{t}', \mathbf{t}')\right), \tag{10}$$

$$\mu^{\text{post}}(\mathbf{t}') := \mu(\mathbf{t}') + K(\mathbf{t}', \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \eta I_n]^{-1} (\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t})), \tag{11}$$

$$K^{\text{post}}(\mathbf{t}', \mathbf{t}') := K(\mathbf{t}', \mathbf{t}') - K(\mathbf{t}', \mathbf{t}) \left[K(\mathbf{t}, \mathbf{t}) + \eta I_n\right]^{-1} K(\mathbf{t}, \mathbf{t}'). \tag{12}$$

It is common in Gaussian process modeling to assume that the mean $\mu \equiv 0$, which is justified on both theoretical and practical grounds (Rasmussen and Williams, 2006, Chapter 2). First, this assumption is made without loss of generality as any systematic component of the mean can be equivalently represented by including appropriate kernels in the covariance function. Second, focusing inference entirely on the covariance simplifies the marginal likelihood and posterior expressions, avoiding potential identification issues. Finally, assigning the prior mean to zero forces inferred trends, periodicities, and other structure to be captured by the covariance rather than absorbed into an estimate of the mean.

Primitive Kernels Several covariance kernels are used in the Gaussian process literature to express priors over functions with a variety of temporal characteristics, which include the following constructs.

- **Constant** $k_{\mathsf{C}}(t, t') := \theta_{\mathsf{C}}$, where $\theta_{\mathsf{C}} > 0$ is the constant variance value. This kernel produces a full covariance matrix with every entry equal to θ_{C}, implying that all function values are perfectly correlated. Realizations from a Gaussian process with this kernel are horizontal lines.

- **Linear** $k_{\mathsf{LIN}}(t, t') = \theta_{\mathsf{LIN},2}(t - \theta_{\mathsf{LIN},1})(t' - \theta_{\mathsf{LIN},1})$ where $\theta_{\mathrm{L},1}$ is the intercept that centers the input domain, and $\theta_{\mathrm{L},2}$ controls the slope or scaling of the linear trend. This kernel generates covariance matrices whose entries grow with distance from the intercept, and draws from a Gaussian process with this kernel are sloped lines. The additive bias term is excluded here as it can be captured equivalently by the Constant kernel.

- **GammaExponential**

$$k_{\mathsf{GE}}(t, t') := \theta_{\mathsf{GE},3} \exp\left(-\left(\frac{|t - t'|}{\theta_{\mathsf{GE},1}}\right)^{\theta_{\mathsf{GE},2}}\right) \tag{13}$$

  where $\theta_{\mathsf{GE},1}$ is the lengthscale, $\theta_{\mathsf{GE},2} > 0$ is the shape parameter (often denoted γ), and $\theta_{\mathsf{GE},3}$ is the amplitude. It allows for modeling functions that vary in smoothness across different time scales. Setting $\theta_{\mathsf{GE},2} = 2$ gives

an equivalent parameterization of the SquaredExponential kernel $k_{\mathsf{SE}}(t,t') = \theta_{\mathsf{SE},2}\exp\left(-(t-t')^2/2\sigma^2_{\mathsf{SE},1}\right)$.

- **Periodic**

$$k_{\mathsf{P}}(t,t') \coloneqq \theta_{\mathsf{P},3}\exp\left(-2\sin^2\left(\frac{\pi}{\theta_{\mathsf{P},2}}\left|t-t'\right|\right)\Big/\theta^2_{\mathsf{P},1}\right) \tag{14}$$

where $\theta_{\mathsf{P},1}$ is the lengthscale controlling smoothness within each period, $\theta_{\mathsf{P},2}$ is the period of repetition, and $\theta_{\mathsf{P},3}$ is the amplitude. The kernel can be understood by warping the time points t to the unit circle using $\theta(t) \coloneqq (\cos(2\pi/\theta_{\mathsf{P},2}\,t), \sin(2\sin/\theta_{\mathsf{P},2}\,t))$ and then applying the SquaredExponential kernel to inputs $\theta(t), \theta(t')$.

Composite Kernels Composing primitive kernels enables Gaussian processes to serve as universal approximators for continuous functions on compact sets. The reproducing kernel Hilbert space (RKHS) induced by the kernel provides an infinite-dimensional basis for function approximation, allowing GPs to model complex patterns through kernel compositions without explicit basis functions (Rasmussen and Williams, 2006; Micchelli et al., 2006). The θ hyperparameters are assigned lognormal priors, as specified in the subsequent section, to ensure positivity and flexible modeling of lengthscales, amplitudes, and periods.

These primitive kernels can be combined using a set of algebraic operators, enabling the construction of rich, interpretable covariance structures. In particular, two covariance kernels k_1 and k_2 may be combined to create a new kernel through the following operations.

- **Addition** $k_+(t,t') \coloneqq k_1(t,t') + k_2(t,t')$. Represents additive structure, such as a seasonal component superimposed on a trend.
- **Multiplication** $k_\times(t,t') \coloneqq k_1(t,t') \cdot k_2(t,t')$. Allows for modulation, e.g., periodic components with time-varying amplitude.
- **ChangePoint** $k_{\mathsf{CP}}(t,t') \coloneqq (1-\sigma(t))\,k_1\left(t,t'\right)\left(1-\sigma\left(t'\right)\right) + \sigma(t)\,k_2\left(t,t'\right)\sigma\left(t'\right)$, where the sigmoid function $\sigma(u)$ acting as the transition mechanism is defined as

$$\sigma(u) \coloneqq \frac{1}{2}\left(1+\tanh\left(\frac{u-\theta_{\mathsf{CP},1}}{\theta_{\mathsf{CP},2}}\right)\right), \tag{15}$$

with $\theta_{\mathsf{CP},1}$ representing the location of the change point and $\theta_{\mathsf{CP},2} > 0$ controlling the sharpness of the transition. For small values of $\theta_{\mathsf{CP},2}$, the kernel switches abruptly from k_1 (active before the change point) to k_2 (active after the change point). Larger values of $\theta_{\mathsf{CP},2}$ yield a smoother interpolation between the two kernels. This structure allows the Gaussian process to adapt to regime changes, such as structural breaks or evolving dynamics in time series data, by modulating the influence of each kernel across time.

Automated Kernel Discovery Recent advances have significantly extended the expressiveness of Gaussian process modeling through the introduction of *algebraic kernel composition* (Duvenaud et al., 2013; Saad, 2022; Saad et al., 2023). Rather than requiring the econometrician to manually specify a kernel, these methods define a space of composite kernels via a symbolic grammar, enabling flexible model discovery through *automated* kernel search.

Together the primitive kernels and algebraic operators form a *kernel algebra*—a symbolic language of model composition—which can be defined formally as a *probabilistic context-free grammar* (PCFG), i.e., a context-free grammar whose production rules are assigned probabilities, thereby defining a prior distribution over the discrete space of kernel expressions. This grammar-based approach allows the econometrician to place a prior distribution over model structures, not just parameters, enabling Bayesian structure learning to perform inference over the grammar-defined space of Gaussian process models. By doing so, it provides both expressivity and interpretability

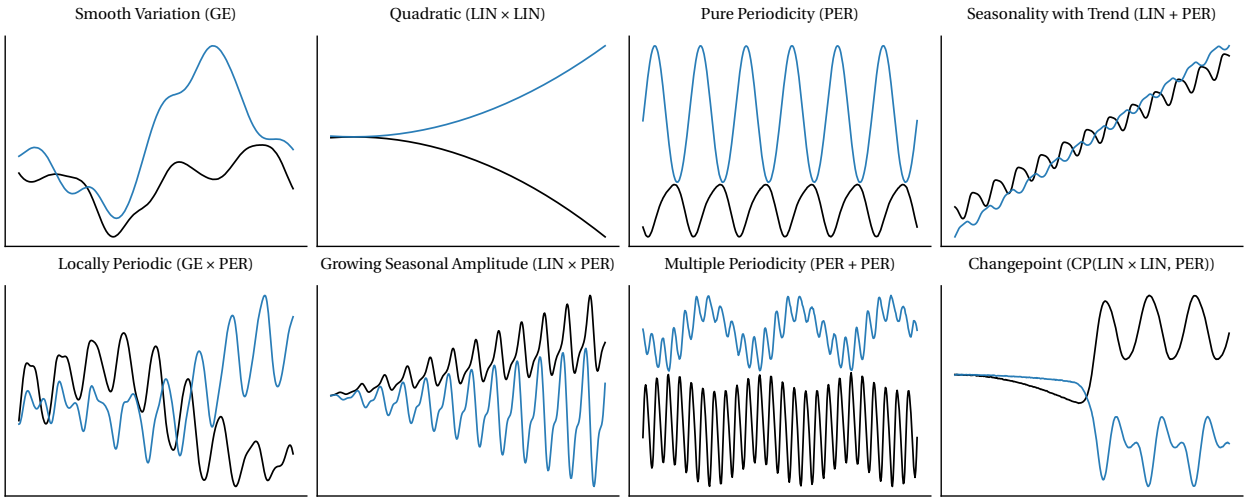


Figure 1: Gaussian process models can express a variety of time series structures, by varying the primitive covariance kernels and combining them using addition, multiplication, and changepoint operators.

as each sampled model can be inspected as a symbolic expression revealing interpretable qualitative features of the underlying time series, such as seasonal patterns, trends, change points, and their interactions. Figure 1 shows two samples of time series data that follow a Gaussian process prior using eight different covariance kernels, highlighting the flexibility of the model class $\mathscr{M}$.

For any time points $\mathbf{t} = (t_1, \ldots, t_n)$, we define a prior over model structures $m \in \mathscr{M}$ using a hierarchical Bayesian model as follows:

$$B \sim \mathsf{Constant} \mid \mathsf{Linear} \mid \mathsf{GammaExponential} \mid \mathsf{Periodic} \qquad \text{(w.p. } \alpha_1, \alpha_2, \alpha_3, \alpha_4) \tag{16}$$

$$\oplus \sim + \mid \times \mid \mathsf{CP} \qquad \text{(w.p. } \beta_1, \beta_2, \beta_3) \tag{17}$$

$$m \sim B \mid \oplus[m_1, m_2] \qquad \text{(w.p. } \gamma, 1-\gamma) \tag{18}$$

$$[\theta_1, \ldots, \theta_{d(m)}] \sim \mathrm{LogNormal}(0, 1) \tag{19}$$

$$\eta \sim \mathrm{InvGamma}(1, 1) \tag{20}$$

$$\mathbf{X}(\cdot) \sim \mathscr{N}(0, k_{m,\theta}) \tag{21}$$

$$Y(t_i) \sim \mathscr{N}(f(t_i), \eta) \qquad (i = 1, \ldots, n). \tag{22}$$

Equations (16)–(18) describe the process for generating a model structure using a PCFG over primitive and composite Gaussian process kernels, with the shown probabilities. The probabilities for the recursive rule m are chosen to ensure the prior halts almost surely (Saad et al., 2019, §4.2). In Eq. (19), $d(m)$ denotes the number of numerical parameters within the kernel structure m, e.g., $d\left(\mathsf{Periodic}\right) = 3$ and $d\left(\mathsf{Periodic} + \mathsf{Constant}\right) = 4$. In Eq. (21), $K_{m,\theta}$ denotes the covariance function defined by the structure and parameters (m, θ), where θ will be elided when clear from context, and $\mathbf{X}(\cdot)$ is the (infinite-dimensional) latent Gaussian process. Equation (22) specifies the n noisy observations $Y(t_i)$ for $i = 1, \ldots, n$.

Posterior Inference Recall that our goal is to produce properly weighted samples $\{(w_i, (m^i, \theta^i, \eta^{(i)}))\}_{i=1}^N$ for all the hidden variables given an observation $\mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t})$. The inverse gamma prior for η on Eq. (20) and Gaussian prior for $\mathbf{X}(\mathbf{t})$ are conjugate with the Gaussian likelihood for Y, enabling closed-form conditional posteriors. How-

ever, the lognormal prior for kernel parameters θ and the PCFG prior for model structures m are not conjugate, necessitating more sophisticated sampling methods, in particular Sequential Monte Carlo (SMC) (Chopin and Papaspiliopoulos, 2020) and Involutive Markov Chain Monte Carlo (Neklyudov et al., 2020) to explore the joint posterior $p(m, \theta, \eta \mid \mathbf{y}(\mathbf{t}))$. With the exception of the periodic kernel (discussed below), the LogNormal(0, 1) prior (i.e., $\ln\theta \sim \mathcal{N}(0,1)$, median of 1 and 95% prior mass roughly between 0.14 and 7.1) ensures positivity of kernel hyperparameters such as lengthscales, amplitudes, and shape parameters, while providing heavy-tailed support permitting flexibility. We use AutoGP[1] to conduct the posterior inference; see Saad et al. (2023, §3) for full technical details.

2.2 IDENTIFYING SEASONALITY

The contribution of this paper is identifying seasonality via probabilistic model discovery. The literature generally agrees that the defining characteristic of seasonality is periodicity in time series behavior, whether treated as deterministic or stochastic (Granger, 1978). Our framework aligns with this notion by defining seasonality as the existence of a Periodic kernel in the posterior over covariance structures.

Periodic Kernel A time series exhibits seasonality if the posterior distribution $p(m \mid \mathbf{y}(\mathbf{t}))$ over kernel structures assigns non-negligible posterior probability to models containing a Periodic base kernel. By grounding the notion of seasonality in the kernel algebra, we obtain an operational and model-based criterion that is directly amenable to standard Bayesian inference.

Our definition of seasonality imposes that the covariance between observations should depend only on their *phase difference* within a fixed period p, rather than the absolute time index. For example, monthly seasonality would imply strong correlation between, say, January 2020 and January 2023, while January and July should be minimally correlated. The Periodic kernel, k_{P}, is designed to capture such structure. It can be derived from the SquaredExponential kernel, which measures similarity through Euclidean distance $\Delta_{\text{eu}}(t, t') = |t - t'|$,

$$k_{\mathsf{SE}}(t, t') = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\left(\Delta_{\text{eu}}(t, t')\right)^2\right) = \sigma^2 \exp\left(-\frac{1}{2\ell^2}|t - t'|^2\right) \tag{23}$$

where ℓ is the lengthscale parameter controlling smoothness and σ^2 is the marginal variance. While this kernel is stationary and smooth, it does not encode periodicity, because correlations decay monotonically with absolute distance. To enforce periodicity, the time domain is wrapped onto the unit circle, where each time point t is mapped to an angle $\theta(t) = 2\pi t/p$. The natural distance on the circle between two time points (t, t') is given by the chord length of the corresponding points on the unit circle at angles $(2\pi t/p,\ 2\pi t'/p)$

$$\Delta_{\text{per}}(t, t') = 2\sin\left(\frac{\pi}{p}\left|t - t'\right|\right) \tag{24}$$

This mapping ensures periodic invariance, $k(t, t') = k(t + mp,\ t' + np)$, $\forall m, n \in \mathbb{Z}$. Using Δ_{per} instead of Δ_{eu} in the SquaredExponential kernel Eq. (23) yields the Periodic kernel:

$$k_{\mathsf{P}}(t, t') = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\left(\Delta_{\text{per}}(t, t')\right)^2\right) = \sigma^2 \exp\left(-\frac{2}{\ell^2}\sin^2\left(\frac{\pi}{p}\left|t - t'\right|\right)\right)$$

This kernel is parameterized by $\theta_{\mathsf{P},1} = \ell$ (lengthscale, controlling smoothness within each period), $\theta_{\mathsf{P},2} = p$ (the period of repetition), and $\theta_{\mathsf{P},3} = \sigma^2$ (the amplitude). AutoGP's Bayesian estimation procedure lets the data speak as to the values of these parameters and *how* the Periodic kernel enters the model.

[1] `https://probsys.github.io/AutoGP.jl/stable/index.html`

$(m_{\text{per}}, m_{\text{non}}) \coloneqq \text{split}[m]$

$\text{split}[\text{PER}] \coloneqq (\text{PER}, \mathbf{0})$ $\quad \text{split}[b] \coloneqq (\mathbf{0}, b),\ b \neq \text{PER}$

$\text{split}[m_1 + m_2] \coloneqq$
$\quad$ let $\text{split}[m_1] \leftarrow (a_1, c_1)$
$\quad$ let $\text{split}[m_2] \leftarrow (a_2, c_2)$
$\quad$ in $(a_1 + a_2, c_1 + c_2)$

$\text{split}[m_1 \times m_2] \coloneqq$
$\quad$ let $\text{split}[m_1] \leftarrow (a_1, c_1)$
$\quad$ let $\text{split}[m_2] \leftarrow (a_2, c_2)$
$\quad$ in $\big(a_1 \times a_2 + a_1 \times c_2 + c_1 \times a_2, c_1 \times c_2\big)$

$\text{split}[\text{CP}_{\ell,s}(m_1, m_2)] \coloneqq$
$\quad$ let $\text{split}[m_1] \leftarrow (a_1, c_1)$
$\quad$ let $\text{split}[m_2] \leftarrow (a_2, c_2)$
$\quad$ in $\big(\text{CP}_{\ell,s}[a_1, a_2], \text{CP}_{\ell,s}[c_1, c_2]\big)$

Figure 2: Recursive definition of split$[m]$, which decomposes a covariance kernel expression m into a pair $(m_{\text{per}}, m_{\text{non}})$ of kernels, where m_{per} contains the periodic structure and m_{non} contains the nonperiodic structure.

Extracting Periodicity Given a kernel symbolic expression m from the PCFG Eqs. (16)–(18), the split$[m]$ procedure in Fig. 2 decomposes the kernel as a sum two subexpressions: m_{per}, which captures the periodic structure, and m_{non}, which captures the remaining nonperiodic structure. The operator distributes over the kernel's sum-of-products expansion; this ensures that for composite structures like (LIN + PER) × GE, the resulting cross-term PER × GE is correctly attributed to the seasonal component while LIN × GE is not.

This decomposition means that the covariance function k_m for m satisfies $k_m = k_{\text{per}} + k_{\text{non}}$, which a latent decomposition as an additive Gaussian process

$$\mathbf{X} = \mathbf{S} + \mathbf{U}, \qquad \mathbf{S} \sim \text{GP}\big(0, k_m^{\text{per}}\big) \quad \mathbf{U} \sim \text{GP}\big(0, k_m^{\text{non}}\big) \tag{25}$$

and noisy observations $Y(t) = X(t) + \varepsilon(t)$, $\varepsilon(t) \sim \mathcal{N}(0, \eta)$. For time points $\mathbf{t} \coloneqq (t_1, \ldots, t_n)$, write for short $\mathbf{u} \coloneqq \mathbf{U}(\mathbf{t})$, $\mathbf{s} \coloneqq \mathbf{S}(\mathbf{t})$, $\mathbf{x} \coloneqq \mathbf{X}(\mathbf{t})$, $\mathbf{y} \coloneqq \mathbf{Y}(\mathbf{t})$, $K_m^{\text{per}} \coloneqq k_{\text{per}}(\mathbf{t}, \mathbf{t})$, and $K_m^{\text{non}} \coloneqq k_{\text{non}}(\mathbf{t}, \mathbf{t})$. The prior distribution is given by

$$\begin{bmatrix} \mathbf{s} \\ \mathbf{u} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_m^{\text{per}} & 0 & K_m^{\text{per}} \\ 0 & K_m^{\text{non}} & K_m^{\text{non}} \\ K_m^{\text{per}} & K_m^{\text{non}} & K_m^{\text{per}} + K_m^{\text{non}} + \eta I \end{bmatrix}\right), \tag{26}$$

where, as above, we have a priori independence of $\mathbf{S}$ and $\mathbf{U}$ so that $\text{Cov}(\mathbf{s}, \mathbf{u}) = 0$. Recall that for any Gaussian vector

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right), \tag{27}$$

the conditional distribution of $\mathbf{z}_1$ given $\mathbf{z}_2$ is also Gaussian, with mean and variance given by

$$\mathbb{E}[\mathbf{z}_1 \mid \mathbf{z}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{z}_2 - \boldsymbol{\mu}_2), \qquad \text{Var}(\mathbf{z}_1 \mid \mathbf{z}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{28}$$

Using this result and denoting $K \coloneqq K_m(\mathbf{t}, \mathbf{t})$ yields

$$\mathbb{E}[\mathbf{s} \mid \mathbf{y}, m, \theta, \eta] = K_m^{\text{per}}(K + \eta I)^{-1}\mathbf{y} \qquad \mathbb{E}[\mathbf{u} \mid \mathbf{y}, m, \theta, \eta] = K_m^{\text{non}}(K + \eta I)^{-1}\mathbf{y} \tag{29}$$

$$\text{Var}(\mathbf{s} \mid \mathbf{y}, m, \theta, \eta) = K_m^{\text{per}} - K_m^{\text{per}}(K + \eta I)^{-1}K_m^{\text{per}} \tag{30}$$

$$\text{Var}(\mathbf{u} \mid \mathbf{y}, m, \theta, \eta) = K_m^{\text{non}} - K_m^{\text{non}}(K + \eta I)^{-1}K_m^{\text{non}} \tag{31}$$

$$\text{Cov}(\mathbf{s}, \mathbf{u} \mid \mathbf{y}, m, \theta, \eta) = -K_m^{\text{per}}(K + \eta I)^{-1}K_m^{\text{non}} \tag{32}$$

where the last term follows from the off-diagonal block of the joint posterior covariance matrix $\mathrm{Var}\left(\mathbf{s},\mathbf{u} \mid \mathbf{y}\right)$. This term is derived by conditioning on the data $\mathbf{y}$ using the formula above $\mathrm{Var}(\mathbf{z}_1 \mid \mathbf{z}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, where $\mathbf{z}_1 = [\mathbf{s},\mathbf{u}]$ and $\mathbf{z}_2 = \mathbf{y}$.

$$\mathrm{Var}\left(\begin{bmatrix}\mathbf{s}\\ \mathbf{u}\end{bmatrix} \,\middle|\, \mathbf{y}\right) = \underbrace{\begin{bmatrix}K_m^{\mathrm{per}} & \mathbf{0}\\ \mathbf{0} & K_m^{\mathrm{non}}\end{bmatrix}}_{\boldsymbol{\Sigma}_{11}} - \underbrace{\begin{bmatrix}K_m^{\mathrm{per}}\\ K_m^{\mathrm{non}}\end{bmatrix}}_{\boldsymbol{\Sigma}_{12}} \underbrace{(K+\eta I)^{-1}}_{\boldsymbol{\Sigma}_{22}^{-1}} \underbrace{\begin{bmatrix}K_m^{\mathrm{per}} & K_m^{\mathrm{non}}\end{bmatrix}}_{\boldsymbol{\Sigma}_{21}}$$

$$= \begin{bmatrix}K_m^{\mathrm{per}} & \mathbf{0}\\ \mathbf{0} & K_m^{\mathrm{non}}\end{bmatrix} - \begin{bmatrix}K_m^{\mathrm{per}}(K+\eta I)^{-1}K_m^{\mathrm{per}} & K_m^{\mathrm{per}}(K+\eta I)^{-1}K_m^{\mathrm{non}}\\ K_m^{\mathrm{non}}(K+\eta I)^{-1}K_m^{\mathrm{per}} & K_m^{\mathrm{non}}(K+\eta I)^{-1}K_m^{\mathrm{non}}\end{bmatrix}$$

$$= \begin{bmatrix}\mathrm{Var}(\mathbf{s}\mid\mathbf{y}) & \mathrm{Cov}(\mathbf{s},\mathbf{u}\mid\mathbf{y})\\ \mathrm{Cov}(\mathbf{u},\mathbf{s}\mid\mathbf{y}) & \mathrm{Var}(\mathbf{u}\mid\mathbf{y})\end{bmatrix} = \begin{bmatrix}K_m^{\mathrm{per}} - K_m^{\mathrm{per}}(K+\eta I)^{-1}K_m^{\mathrm{per}} & -\mathbf{K_m^{per}}(\mathbf{K}+\eta\mathbf{I})^{-1}\mathbf{K_m^{non}}\\ -K_m^{\mathrm{non}}(K+\eta I)^{-1}K_m^{\mathrm{per}} & K_m^{\mathrm{non}} - K_m^{\mathrm{non}}(K+\eta I)^{-1}K_m^{\mathrm{non}}\end{bmatrix}$$

The off-diagonal block $\mathrm{Cov}(\mathbf{s},\mathbf{u}\mid\mathbf{y})$ is *nonzero*, indicating that the two components are competing to explain the same variance. Adding the means gives the full latent posterior mean $\mathbb{E}[\mathbf{x}\mid\mathbf{y}] = K(K+\eta I)^{-1}\mathbf{y}$, and the variances obey

$$\mathrm{Var}(\mathbf{f}\mid\mathbf{y}) = \mathrm{Var}(\mathbf{s}\mid\mathbf{y}) + \mathrm{Var}(\mathbf{u}\mid\mathbf{y}) + \mathrm{Cov}(\mathbf{s},\mathbf{u}\mid\mathbf{y}) + \mathrm{Cov}(\mathbf{u},\mathbf{s}\mid\mathbf{y}), \tag{33}$$

which equals the standard GP conditional variance for K. Thus, we have established the following the result.

Proposition 1 (Periodic Posterior Decomposition)**.** *Let $\mathbf{t}' := (t'_1,\ldots,t'_m)$ be a set of m test time points. The posterior of the periodic component $\mathbf{s}(\mathbf{t}') := (s(t'_1),\ldots,s(t'_m))$ identified using the* split *operator, given the observation $\mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t})$ and latent variables (m,θ,η) is Gaussian with mean and covariance*

$$\mu_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}') := K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t})\left[K_m(\mathbf{t},\mathbf{t}) + \eta I\right]^{-1}\mathbf{y}(\mathbf{t}) \tag{34}$$

$$\Sigma_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}',\mathbf{t}') := K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t}') - K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t})\left[K_m(\mathbf{t},\mathbf{t}) + \eta I\right]^{-1}K_m^{\mathrm{per}}(\mathbf{t},\mathbf{t}'). \tag{35}$$

Using the weighted posterior draws $\left\{\left(w_i,(m_i,\theta_i,\eta_i)\right)\right\}_{i=1}^{M}$, the model-averaged periodic predictive is approximated by a finite Gaussian mixture that averages over posterior uncertainty in model structure, kernel hyperparameters, and noise variance

$$p\left(\mathbf{s}(\mathbf{t}')\mid\mathbf{y}(\mathbf{t})\right) \approx \sum_{i=1}^{M}\tilde{w}_i\,\mathcal{N}\left(\mu_{\mathrm{per}}^{(i)}(\mathbf{t}'),\Sigma_{\mathrm{per}}^{(i)}(\mathbf{t}',\mathbf{t}')\right), \qquad \tilde{w}_i := w_i \Big/ \sum_{j=1}^{M}\tilde{w}_j, \tag{36}$$

where $\mu_{\mathrm{per}}^{(i)}(\mathbf{t}')$ and $\Sigma_{\mathrm{per}}^{(i)}(\mathbf{t}',\mathbf{t}')$ are the mean and covariance of the periodic component for the i^{th} posterior draw (m_i,θ_i,η_i), as defined in Equations (34) *and* (35)*. Evaluating at observed time points $\mathbf{t}' = \mathbf{t}$ yields the conditional posterior for the in-sample periodic component*

$$\mathbf{s}(\mathbf{t})\mid\mathbf{y}(\mathbf{t}),m_i,\theta_i,\eta_i \sim \mathcal{N}\left(\mu_{\mathrm{per}}^{(i)}(\mathbf{t}),\,\Sigma_{\mathrm{per}}^{(i)}(\mathbf{t},\mathbf{t})\right) \qquad (1\le i\le n). \tag{37}$$

2.3 SEASONAL ADJUSTMENTS AND FORECASTS We can now use these constructs to establish both in-sample seasonality adjustments and out-of-sample forecasts of seasonality that take full advantage of the Bayesian approach to articulate the inherent uncertainty associated with such objects. In doing so, we impose a centering convention on the seasonal component, which serves as an identifying restriction separating level and seasonality in additive decompositions. As is well known in the seasonal adjustment literature (Watson, 1987; Harvey, 1989; West

and Harrison, 1997), the decomposition $y(t) = \tau(t) + s(t) + \varepsilon(t)$ is not uniquely identified without normalization, because the likelihood is invariant to the transformation $\tau^*(t) = \tau(t) + a$ and $s^*(t) = s(t) - a$ for any constant a. Consequently, an explicit convention is required to assign the zero-frequency (or level) component to the trend rather than to seasonality.

We adopt the standard econometric convention that seasonality represents a zero-mean fluctuation around the local level. This restriction is equivalent to the familiar "sum-to-zero" constraint $\sum_{j=0}^{s-1} s_{t-j} \approx 0$ used in state-space formulations (Harvey, 1989) and to the canonical decomposition principle of Hillmer and Tiao (1982), which enforces a deterministic or stochastic zero-sum restriction on seasonal factors—corresponding to a spectral dip at frequency zero—so that long-run variation is attributed to the trend and irregular components. Watson (1987) makes this identification explicit by setting the constant term in the seasonal difference equation to zero, thereby assigning the unconditional mean of the process to the nonseasonal component. In our setting, centering the estimated seasonal component provides a finite-sample analogue of this spectral restriction. When the observation window spans an integer number of seasonal cycles, the centering is exact due to orthogonality between seasonal and low-frequency components. When the sample ends mid-cycle, the centering reflects the composition of the observed phases and may be biased in finite samples; however, this bias vanishes asymptotically.

In our implementation below, this identifying restriction is imposed operationally either by forecast-window self-centering of the seasonal forecast or, in the more elaborate procedure, by particle-wise detrending followed by the same self-centering step.

In-Sample Seasonal Decomposition Let $\mathbf{1}_n$ denote the n-vector of ones and define the centering matrix

$$C_n := I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\intercal. \tag{38}$$

For each posterior particle $i = 1, \ldots, M$, let $\mu_{\text{per}}^{(i)}(\mathbf{t})$ denote the in-sample posterior mean of the periodic component evaluated on the observed inputs $\mathbf{t}$, and let the normalized particle weights be

$$\tilde{w}_i := \frac{w_i}{\sum_{j=1}^M w_j} \qquad (1 \le i \le M).$$

We impose the identifying restriction that the seasonal component is a zero-mean fluctuation over the observed window by centering each particle-specific seasonal mean

$$\tilde{\mu}_{\text{per}}^{(i)}(\mathbf{t}) := C_n\, \mu_{\text{per}}^{(i)}(\mathbf{t}). \tag{39}$$

Thus, the in-sample seasonal component used in the implementation is the weighted particle approximation

$$\hat{\mathbf{s}}(\mathbf{t}) := \sum_{i=1}^{M} \tilde{w}_i\, \tilde{\mu}_{\text{per}}^{(i)}(\mathbf{t}). \tag{40}$$

We then define the in-sample *seasonally adjusted series* as

$$\hat{\mathbf{y}}^{\text{SA}}(\mathbf{t}) := \mathbf{y}(\mathbf{t}) - \hat{\mathbf{s}}(\mathbf{t}). \tag{41}$$

The seasonally adjusted series is represented by a weighted particle decomposition rather than by a single model-specific adjustment.

When a dispersion summary is desired, we use the cross-particle variation in the centered seasonal means

$$\Sigma_{\text{seas}}(\mathbf{t}) \coloneqq \sum_{i=1}^{M} \tilde{w}_i \big(\tilde{\mu}_{\text{per}}^{(i)}(\mathbf{t}) - \hat{\mathbf{s}}(\mathbf{t})\big)\big(\tilde{\mu}_{\text{per}}^{(i)}(\mathbf{t}) - \hat{\mathbf{s}}(\mathbf{t})\big)^{\mathsf{T}}. \tag{42}$$

This object summarizes between-particle uncertainty in the centered seasonal decomposition and therefore quantifies how strongly the posterior agrees on the shape, timing, and magnitude of seasonality in the observed sample. This object is one of the main advantages of the Bayesian approach. Rather than producing a single seasonally adjusted path with no direct measure of uncertainty, the particle approximation delivers a posterior distribution over seasonal decompositions. As a result, the analyst may ask not only for a point estimate of the seasonal component, but several other queries such as

- the posterior probability that seasonality is present,
- the posterior probability that a given month lies above or below trend,
- the posterior magnitude of seasonal amplitude, and
- the degree of disagreement across admissible model structures.

Out-of-Sample Forecast Construction We report two forecast-construction procedures that both target the econometric definition of seasonality motivated above (a zero-mean seasonal component around the local level), but differ in how they remove low-frequency leakage and how they aggregate uncertainty across the particle mixture. We examine two methodologies to demonstrate robustness of our results. Simulation results below show negligible difference in how forecasts are constructed.

Suppose there are T in-sample time points $\mathbf{t} \coloneqq (t_1, \ldots, t_T)$ recorded at regular discrete time intervals. Let $H = 12$ denote a forecast horizon of twelve time periods and let $\mathbf{t}' = (t_{T+1}, \ldots, t_{T+H})$. Out-of-sample forecasts are constructed from the particle-level predictive summaries returned by the decomposition in Proposition 1. For each posterior particle $p = 1, \ldots, M$, let

$$m_{c,T+h}^{(p)}, \qquad \ell_{c,T+h}^{(p)}, \qquad u_{c,T+h}^{(p)}, \qquad\qquad (h = 1, \ldots, H) \tag{43}$$

denote the particle-level predictive mean and $(\alpha/2, 1-\alpha/2)$ pointwise bounds for component $c \in \{0, 1, 2\}$, where $c = 0$ is the total series, $c = 1$ is the seasonal component, and $c = 2$ is the nonseasonal/trend component. Seasonal forecasts are then constructed using one of two methods.

Method A: Simple weighted aggregation with forecast-window self-centering. For each component c and horizon h, the particle outputs are aggregated by weighted averaging

$$\hat{m}_{c,T+h} = \sum_{p=1}^{M} \tilde{w}_p \, m_{c,T+h}^{(p)}, \qquad \hat{\ell}_{c,T+h} = \sum_{p=1}^{M} \tilde{w}_p \, \ell_{c,T+h}^{(p)}, \qquad \hat{u}_{c,T+h} = \sum_{p=1}^{M} \tilde{w}_p \, u_{c,T+h}^{(p)}. \tag{44}$$

Because the decomposition is identified only up to an additive constant, we impose the seasonal normalization on the forecast window itself by subtracting the mean seasonal forecast over the H horizons

$$\bar{\mu} \coloneqq \frac{1}{H} \sum_{j=1}^{H} \hat{m}_{1,T+j}. \tag{45}$$

The centered seasonal forecast summaries are then

$$\tilde{m}_{1,T+h} = \hat{m}_{1,T+h} - \bar{\mu}, \qquad \tilde{\ell}_{1,T+h} = \hat{\ell}_{1,T+h} - \bar{\mu}, \qquad \tilde{u}_{1,T+h} = \hat{u}_{1,T+h} - \bar{\mu} \qquad (h = 1, \dots, H). \tag{46}$$

These summaries yield a forecast-window analogue of the in-sample centering restriction. Forecast errors are then evaluated using the total-series forecast $\hat{m}_{0,T+h}$ and the centered seasonal forecast $\tilde{m}_{1,T+h}$.

Method B: Particle-wise detrending, mixture sampling, and forecast-window self-centering. The second procedure first removes low-frequency leakage from the particle-level seasonal forecasts before aggregating uncertainty. For each particle p, we regress the in-sample seasonal mean on a linear time trend over the training window:

$$m_{1,t}^{(p)} = a_p + b_p t + \eta_t^{(p)} \qquad (t = 1, \dots, T), \tag{47}$$

and then extrapolate the fitted linear component over the combined train–forecast span

$$L_t^{(p)} = a_p + b_p t, \qquad (t = 1, \dots, T + H). \tag{48}$$

The particle-level decomposition is then adjusted by shifting the seasonal and nonseasonal/trend components in opposite directions

$$m_{1,t}^{(p)} \leftarrow m_{1,t}^{(p)} - L_t^{(p)}, \qquad m_{2,t}^{(p)} \leftarrow m_{2,t}^{(p)} + L_t^{(p)}, \tag{49}$$

with the same shift applied to the corresponding particle-level interval bounds. This step imposes a Watson-style (1987) identifying restriction where low-frequency drift is removed from the seasonal component and reassigned to the nonseasonal/trend component.

To obtain predictive intervals that better reflect the particle mixture, we then construct pointwise mixture summaries by Monte Carlo sampling. Let J denote a particle draw with $\mathbb{P}(J = p) = \tilde{w}_p$. For each component c and horizon h, we sample

$$Y_{c,T+h}^{(r)} \mid J = p \sim \mathcal{N}\left(m_{c,T+h}^{(p)}, \sigma_{c,T+h}^{2(p)}\right) \qquad (r = 1, \dots, R), \tag{50}$$

where $\sigma_{c,T+h}^{2(p)}$ is recovered from the reported particle-level bounds under a Gaussian approximation. The mixture mean and pointwise quantiles are then computed empirically

$$\hat{m}_{c,T+h} := \frac{1}{R} \sum_{r=1}^{R} Y_{c,T+h}^{(r)}, \qquad \hat{\ell}_{c,T+h} := \text{Quantile}_{\alpha/2}\left(\left\{Y_{c,T+h}^{(r)}\right\}_{r=1}^{R}\right), \qquad \hat{u}_{c,T+h} := \text{Quantile}_{1-\alpha/2}\left(\left\{Y_{c,T+h}^{(r)}\right\}_{r=1}^{R}\right). \tag{51}$$

The same forecast-window centering restriction is imposed on the seasonal mixture summaries

$$\bar{\mu} := \frac{1}{H} \sum_{j=1}^{H} \hat{m}_{1,T+j}, \quad \tilde{m}_{1,T+h} = \hat{m}_{1,T+h} - \bar{\mu}, \quad \tilde{\ell}_{1,T+h} = \hat{\ell}_{1,T+h} - \bar{\mu}, \quad \tilde{u}_{1,T+h} = \hat{u}_{1,T+h} - \bar{\mu} \qquad (h = 1, \dots, H). \tag{52}$$

Both procedures target the same econometric definition of seasonality—a zero-mean fluctuation around the local level—but differ in how they aggregate uncertainty and in whether they explicitly remove low-frequency leakage from the seasonal component before doing so. Method A is computationally simple and stable, but its interval summaries are formed by averaging particle-level bounds. Method B is computationally more costly, but it targets the particle mixture more directly and reduces low-frequency leakage into the seasonal forecast.

3 A SIMULATION STUDY

We now demonstrate efficacy of the probabilistic approach through a simulation exercise. Because seasonal adjustment concerns the extraction of latent components, there is generally no directly observed "true" seasonal adjustment against which competing procedures can be evaluated. For this reason, simulations provide an environment in which the underlying seasonal structure is known by construction and the consequences of alternative procedures can be assessed transparently [see e.g., Ghysels and Perron (1990)].

The simulated data are designed to capture characteristics common to seasonal economic time series. To that end, consider Fig. 3, which plots the number of passengers per month on international flights on U.S. air carriers from January 1949 through December 1960. The grey shaded bars indicate NBER recessions. This time series is representative in the sense that it displays three characteristics common to economic data: [i.] a saw-tooth pattern due to seasonality as air travel peaks in summer months and troughs in the winter; [ii.] an upward trend that either slows or declines with recessions; [iii.] an increase in peak-to-trough dynamics over time.

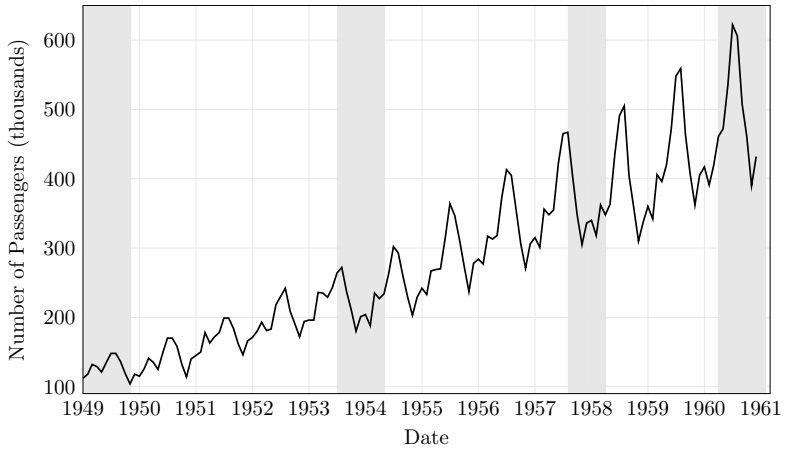


Figure 3: Passengers on International Flights: 1949-1960

The data of Fig. 3, originally sourced from the FAA Statistical Handbook of Civil Aviation and first appearing (Brown, 1963, Table C.10), has become a classic benchmark for assessing seasonality, extensively analyzed in seminal textbooks such as Box and Jenkins (1976) and used as a baseline specification for the X-13-ARIMA-SEATS algorithm (referred to as the "airline model"). It is readily available in popular repositories like the R "AirPassengers" dataset and the Time Series Data Library (TSDL) [161].

Data Generating Process. Consider a discrete-time stochastic process given by

$$X(t) := a_0 + a_1 \cdot t + (a_2 + a_4 \cdot t)\sin(\omega \cdot t) + (a_3 + a_5 \cdot t)\cos(\omega \cdot t) + \varepsilon(t) \tag{53}$$

where t is time (e.g., months) indexed by $t = 1, 2, 3, \ldots, T$, $\omega = 2\pi/12$ is the angular frequency corresponding to a 12-month (annual) seasonal cycle and $\varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is a Gaussian error term. This stochastic process captures the three quintessential characteristics of the economic time series of Figure 3 through a linear time trend $(a_1 \cdot t)$, seasonal terms $(a_2 \cdot \sin(\omega t),\ a_3 \cdot \cos(\omega t))$ whose amplitudes grow linearly in time $(a_4 \cdot t,\ a_5 \cdot t)$, which allows for increasing peak-to-trough dynamics. To ensure a constant phase, the parameters satisfy the constraint $a_5/a_4 = a_3/a_2$, reducing the effective number of parameters to five. The seasonal component can be expressed as a single

sinusoid

$$s(t) \coloneqq R(t)\sin\left(\omega \cdot t + \phi\right), \quad R(t) \coloneqq \sqrt{(a_2 + a_4 \cdot t)^2 + (a_3 + a_5 \cdot t)^2}, \quad \phi \coloneqq \arctan 2(a_3, a_2) \tag{54}$$

where $R(t)$ is the amplitude, which grows linearly with time, and ϕ is the constant phase. This model specification is discussed in Brown (1963). The model is fit to the Air Passengers data of Figure 3 using OLS. The fitted parameter values are: $\hat{a}_0 = 87.82$, $\hat{a}_1 = 2.65$, $\hat{a}_2 = -0.24$, $\hat{a}_3 = -0.57$, $\hat{a}_4 = -0.25$, and $\hat{a}_5 = -0.58$, $\hat{\sigma}_\varepsilon = 26.90$, with an adjusted R-squared of 0.95. This fitted model is assumed to be the true data generating process.

For simulation length T, we generate $T + 12$ observations $\{x(t)\}_{t=1}^{T+12}$ with the first T observations used for estimation and the final 12 observations for out-of-sample forecast evaluation. We examine sample sizes of $T = \{36, 72, 144\}$, corresponding to 3, 6, and 12 years of monthly data to assess forecasting performance across small to large training samples. For each sample size of length T, we conduct $N = 200$ Monte Carlo simulations, each consisting of $T + 12$ observations.

For a given training length T, the strength of the seasonal signal relative to the innovation noise can be summarized by

$$\xi(T) = \sqrt{\frac{\frac{1}{T}\sum_{t=1}^{T} s_t^2}{\sigma_\varepsilon^2}} \tag{55}$$

where s_t is the (true) seasonal component implied by (54). This quantity equals the root-mean-square (RMS) amplitude of the seasonal component, measured in units of the innovation standard deviation, and therefore increases with T as the seasonal amplitude grows over time. Because $R(t)$ grows over time, later portions of the sample contain larger peak-to-trough seasonal movements, and these observations enter the RMS measure $\xi(T)$ directly. The resulting increase in $\xi(T)$ implies that the seasonal component is weakly identified at short sample lengths, more clearly identifiable at intermediate sample lengths, and strongly identified in long samples. Thus, our simulation design with $T = 36$, $T = 72$, and $T = 144$ can be interpreted as weak, moderate, and strong seasonality environments, respectively.

We estimate three models with this data—the correctly specified model of (53) using OLS (we do not force the constant-phase constraint $a_5/a_4 = a_3/a_2$ on the estimated model); the probabilistic model of Section 2; and a model estimated using the X-13 ARIMA-SEATS algorithm (to be discussed below). Both parameters *and* model structure must be learned when implementing probabilistic model discovery and X-13, while standard asymptotic arguments apply to OLS estimates of the correctly-specified model; thus, the OLS specification serves as our benchmark.

We evaluate the ability of these models to forecast both the stochastic process, $x(T+h)$ and the seasonal component, $s(T+h)$. For each forecast horizon $h = 1, ..., 12$, we compute the forecast error $e(T+h) \coloneqq x(T+h) - \hat{x}(T+h)$ and the Mean Interval Score (MIS) for a given coverage level

$$\text{MIS}_{h,\alpha} = \underbrace{(U(T+h) - L(T+h))}_{\text{Interval Width}} + \underbrace{\frac{2}{\alpha}(L(T+h) - x(T+h))}_{\text{Lower Bound Penalty}} + \underbrace{\frac{2}{\alpha}(x(T+h) - U(T+h))}_{\text{Upper Bound Penalty}} \tag{56}$$

where $L(T+h)$ and $U(T+h)$ denote the lower and upper bounds of the predictive interval, corresponding to the $\alpha/2$ and $1-\alpha/2$ quantiles of the forecast distribution ($\alpha = 0.05$ below). The MIS is a key metric for evaluating probabilistic forecasts; it penalizes wide intervals via the interval width term and miscalibrated forecasts through penalties for observations outside the quantiles.

3.1 OLS MODEL The correctly specified model obeys standard asymptotics. For OLS coefficients, the intercept and seasonal terms (a_0, a_2, a_3) converge at rate $\sqrt{T}$, with asymptotic variances $\sigma_\varepsilon^2 \cdot 4/T$ for a_0 and $\sigma_\varepsilon^2 \cdot 8/T$ for a_2, a_3. Trend terms (a_1, a_4, a_5) converge at rate $T^{3/2}$, with variances $\sigma_\varepsilon^2 \cdot 12/T^3$ (a_1) and $\sigma_\varepsilon^2 \cdot 24/T^3$ (a_4, a_5), driven by the block-diagonal form of $(X^\top X)^{-1}$ whose blocks scale as T^{-1}, T^{-2}, T^{-3}. The h-step forecast error is $e(T+h) = x(T+h) - \hat{x}(T+h) = \varepsilon(T+h) - \mathbf{x}(T+h)^\intercal(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, with variance

$$\mathrm{Var}(e(T+h)) = \sigma_\varepsilon^2(1 + \mathbf{x}(T+h)^\intercal (X^\intercal X)^{-1}\mathbf{x}(T+h)) \approx \sigma_\varepsilon^2\left(1 + \frac{12}{T}\right). \tag{57}$$

The expected RMSE has two opposing finite-sample effects. First, Jensen's inequality implies $\mathbb{E}[\mathrm{RMSE}_h] = \sigma_\varepsilon c_h$ with $c_h = \sqrt{2/h}\,\Gamma((h+1)/2)/\Gamma(h/2) < 1$ (e.g. $c_5 \approx 0.9515$), so RMSE is biased *downward*. Second, variance inflation from a leverage effect multiplies this by $\sqrt{1+12/T} > 1$, an *upward* bias. We can approximate this quantity as $\mathbb{E}[\mathrm{RMSE}_h] \approx \sigma_\varepsilon\, c_h \sqrt{1+12/T}$. Depending on T and h, the net effect may place RMSE slightly below σ_ε (Jensen dominates) or above (leverage dominates). Asymptotically, with Gaussian predictive distributions, $\mathbb{E}[\mathrm{MIS}_{h,\alpha}] = \frac{4\sigma_\varepsilon}{\alpha}\phi(z_{1-\alpha/2})$ (e.g. $\approx 4.67\,\sigma_\varepsilon$ for $\alpha = 0.05$). In finite samples, predictive distributions are Student-t with $\nu = T - k$ degrees of freedom, so MIS is inflated by t-quantiles and leverage; an approximation is

$$\mathbb{E}[\mathrm{MIS}_{h,\alpha}] \approx \tfrac{4\sigma_\varepsilon}{\alpha}\, t_\nu(q)\, \tfrac{\nu+q^2}{\nu-1}\, \sqrt{1+12/T} \tag{58}$$

where $q = q_{\nu,1-\alpha/2}$ and t_ν is the t_ν density. Thus MIS is well-calibrated in large samples but wider for small T. By contrast, the seasonal component $s(T+h)$ is deterministic given parameters, so forecast error arises only from parameter uncertainty. Both RMSE and MIS for $s(T+h)$ decay at the rate $O\left(T^{-1/2}\right)$ for seasonal forecasts.

Finally, the simulation examines $N = 200$ Monte Carlo simulations for each $T = \{36, 72, 144\}$. This was the smallest number of simulations for which the OLS-distribution of RMSE and MIS remained stable.

3.2 X-13 ARIMA-SEATS The most relevant comparison model is the current standard in seasonal adjustment, X-13ARIMA-SEATS (X-13 hereafter), developed over many decades at various statistical agencies and described in U.S. Census Bureau (2024). It is a comprehensive statistical algorithm used by the vast majority of governments to de-seasonalize economic data, including the Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), Eurostat, and Statistics Canada.[2] Analysis below calls X-13ARIMA-SEATS Version 1.1 Build 61, July 10, 2024, both the Windows and Linux versions.[3] It is important to note that X-13 is an algorithm in the sense that the user can specify *automated* responses at all steps based on optimization routines and well-established statistical thresholds. This allows us to compare X-13 directly to our probabilistic model over numerous simulations.

The algorithm takes as input a discrete time series dataset $\{x(t_i), 1 \le i \le T\}$ and its frequency (e.g., monthly, quarterly), and decomposes the data into trend-cycle, seasonal, and irregular components in order to produce seasonally adjusted data. We now provide a brief overview of the algorithm and direct interested readers to U.S. Census Bureau (2024) for additional details. The algorithm can be viewed as proceeding in the following steps:

1. **Data Transformation.** A generic SARIMA model of Box and Jenkins (1976) can be written as

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D x(t) = \theta(L)\Theta(L^s)\epsilon(t) \tag{59}$$

[2] The BEA uses X-13 to seasonally adjust many of its key economic indicators, including Gross Domestic Product and all components, Personal Income and Outlays, many international transactions such as trade balances. The BLS uses X-13 to seasonally adjust nearly all employment statistics (unemployment rate, hours worked), and all price indices (Consumer Price and Producer Price Indices).

[3] JDemetra+, developed by the National Bank of Belgium (NBB) in collaboration with Eurostat and the European Central Bank (ECB), is a Java-based, open-source alternative to X-13. However, given that we focus on U.S. data and that both programs implement nearly identical methodologies, we do not include JDemetra+ in our analysis.

where $\phi(L)$ and $\theta(L)$ are non-seasonal AR and MA polynomials of orders p and q, $\Phi(L^s)$ and $\Theta(L^s)$ are seasonal AR and MA polynomials of orders P and Q, L is the lag operator, s is the seasonal period, d and D are differencing orders, and $\epsilon(t) \sim N(0, \sigma_\epsilon^2)$, abbreviated as SARIMA(p,d,q)(P,D,Q)$_s$.

A default SARIMA $(0,1,1)(0,1,1)_s$ "airline" model is initially estimated if data are monthly or quarterly. Given this baseline model, the X-13 algorithm tests nonlinear transformations of the data, motivated by Tukey (1957); Box and Cox (1964); Box and Tiao (1975), that are often applied to better conform data to the stationary and Gaussian assumptions. The `transform` specification of X-13 with argument `function=auto` estimates the preliminary SARIMA model with both original and log-transformed data. If the transformation lowers the Akaike Information Criterion (AICC) by greater than two, it is retained.

2. **Automated Model Building.** The `automdl` procedure in X-13 automates ARIMA model selection, adapting the TRAMO methodology of Gomez and Maravall (1996), Gómez and Maravall (2001a) Gómez and Maravall (2001b) and Monsell (2002, 2006). All instances of the X-13ARIMA-SEATS algorithm used in this paper call `automdl` with default maximum order of (3,2,3)(2,1,2), allowing for a constant term. The procedure comprises five stages: default model estimation (described above), differencing order identification, ARMA order identification, comparison with the default model, and final model checks. The final automatic model choice for both the Air Passengers and Revenue-Passenger Miles data is given by $(0,1,1)(0,1,1)_{12}$.

3. **Exogenous Regressors.** Exogenous regressors address deterministic calendar effects, such as trading day and holiday variations (Cleveland et al., 1983; Hillmer et al., 1983; Bell and Hillmer, 1983), capturing predictable economic shifts like increased retail sales before holidays or monthly fluctuations due to trading day counts. In X-13, trading day effects are modeled as regressors for the count of each weekday per period (e.g., number of Mondays), while holiday effects, such as Easter's variable timing, use window-based variables (e.g., days before or after). These regressors are tested via `aictest`, retained if the AICC decreases by any amount (> 0), per U.S. Census Bureau (2024, p. 153), even if individually insignificant.

   Outliers are detected by analyzing residuals $e(t) = y(t) - \hat{y}(t)$ from the SARIMA model and transformation of Steps 1 and 2, (Chang et al., 1988; Chen and Liu, 1993). Potential outliers are flagged using a t-test, $t = \hat{\omega}/\mathrm{SE}(\hat{\omega})$, against an adaptive and subjective critical value: "determined based on sample size and the median absolute deviation of the residuals" (U.S. Census Bureau, 2024, p. 142). For monthly data with 100–300 observations, this value ranges from 3.5 to 4.0. Detected outliers are added as regressors, and the model is iteratively refit until convergence.

4. **Final Model & Estimation.** The final model is a *reg*SARIMA, where *reg* acknowledges the exogenous regressors. For example, the final regSARIMA model for the log-transformed, $y(t) = \ln(x(t))$, Air Passengers data is

$$(1-L)\left(1-L^{12}\right)\left(y_t - \beta_1 \cdot \text{Weekday}_t - \beta_2 \cdot \text{Easter}_t - \beta_3 \cdot \text{AO1951.May}_t\right) = (1+\theta_1 L)\left(1+\Theta_1 L^{12}\right)\epsilon(t), \qquad (60)$$

   where Weekday_t is the trading day regressor, Easter_t is the Easter holiday indicator, AO1951.May_t is a dummy for May 1951 (1 if t = May 1951, 0 otherwise), $\beta_1 = -0.00295$, $\beta_2 = 0.01777$, $\beta_3 = 0.10016$ are the estimated coefficients, $\theta_1 = 0.11562$ and $\Theta_1 = 0.49736$ are the non-seasonal and seasonal MA parameters from the X-13 output, and $\epsilon(t) \sim N(0, 0.0009538)$ is the innovation term. Once the model is estimated, forecasting is standard. For the log-transformed International Passengers data $y(t) = \ln(x(t))$, X-13 computes forecasts recursively on the log scale as above and then optimally maps into the original scale.

5. **Seasonal Decomposition.** The observed time series is decomposed in an additive manner, $x(t) = R(t) + T(t) + S(t) + I(t)$, or log-additive (multiplicative) manner, $x(t) = R(t) \times T(t) \times S(t) \times I(t)$ or $\ln x(t) = \ln R(t) + \ln T(t) + \ln S(t) + \ln I(t)$, depending on the Data Transformation of Step 1, with a calendar/outlier component ($R(t)$), trend component ($T(t)$), seasonal component ($S(t)$), and an irregular component ($I(t)$). At this point, the user can select between a moving-average, filter-based approach (X-11) or a model-based spectral decomposition (SEATS-TRAMO). In this paper, we employ the moving-average filter.

   The X-11 algorithm iterates on moving-average filters to isolate various components of the seasonal-trend-irregular decomposition, (Shiskin et al., 1967; Dagum, 1980; Ladiray and Quenneville, 2001). The algorithm can be summarized as follows:

   5.i. Using the model developed in Steps 1–4, exogenous regressor effects identified in Step 3 are estimated and removed to obtain a prior-adjusted series: $B1(t) \coloneqq x(t) - \hat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{X}(t)$. Our notation, $B1(t)$, matches X-13 output as this "prior adjusted" data appears in Table B1.

   5.ii. An initial estimate of the trend is given by the centered 12-month moving average,

$$T(0,t) \coloneqq \sum_{k=-6}^{6} w_k B1(t+k) \tag{61}$$

   with weights $w(\pm 6) \coloneqq \frac{1}{24}$ and $w_k \coloneqq \frac{1}{12}$ for $k = -5,\ldots,5$. By averaging over roughly one year, this filter nullifies the seasonal component and isolates a trend.

   5.iii. For each month $t = 1,\ldots,12$, compute the average seasonal-irregular (SI) ratio across years using 3x3 moving average:

$$S(1,t) = \sum_{j=-1}^{1} \theta_j^{(3)} SI(1, t+12j) \tag{62}$$

   where weights $\theta_{-1}^{(3)} = \theta_1^{(3)} = \frac{1}{9}$, $\theta_0^{(3)} = \frac{3}{9} = \frac{1}{3}$, and j indexes years. The 3x3 MA spans three years (current, prior, and future), smoothing irregular fluctuations. Seasonal factors are then centered to ensure a sum of approximately zero over a year: $\widetilde{S}(1,m) = S(1,m) - \frac{1}{12}\sum_{m=1}^{12} S(1,m)$ with time indexing $\widetilde{S}(1,t)$ assigned by month (e.g., if $t = 243$ is March, then $\tilde{S}(1,3) = \tilde{S}(1,243)$). This provides a preliminary seasonal adjustment given by $SA(1,t) = B1(t) - \widetilde{S}(1,t)$.

   5.iv. An updated trend is then computed using a symmetric Henderson moving average of $SA(1,t)$

$$T(2,t) = \sum_{k=-J}^{J} h_k SA(1,t_k), \tag{63}$$

   where h_k is is the Henderson filter coefficient at lag k. The Henderson moving average corresponds to a local weighted least squares (OLS) fit of a cubic polynomial at each point t, minimizing $\sum_{k=-m}^{m} w_k(y(t+k) - \beta_0 - \beta_1 k - \beta_2 k^2 - \beta_3 k^3)^2$, where the effective weights w_k are derived to satisfy the constraint of exact reproduction of polynomials up to degree 3.

   Once the trend is updated, the process returns to step 5.ii. and iterates on 5.ii.–5.iv. until convergence.

A few points are noteworthy given the importance of outliers in our study. First, all moving-average calculations discussed above require *additional* data at beginning and end points. Backcasts and forecasts from the regSARIMA model provide these observations. This is a critical feature as end-point analysis can significantly impact data revisions. Second, extreme values can distort moving average calculations. Thus, extremes are flagged based on

deviation from the median of the irregular component: values with $|I(t) - \text{median}(I)| > 2.5\sigma_I$ are fully down-weighted (weight $w_t = 0$), and those with $1.5\sigma_I \le |I_t - \text{median}(I)| \le 2.5\sigma_I$ are partially down-weighted (weight $w_t = (2.5 - |d_t|/\sigma_I)/(2.5 - 1.5)$, where $d_t = I_t - \text{median}(I)$).

3.3 RESULTS. Before reporting results, it is worth repeating that the simulated data originate from a well-fit OLS model (adjusted R-square of 0.95) of the airline data of Figure 3. This is the same airline data that is used to fit the original "airline model" of X-13 and the same data analyzed in the ARIMA / seasonality chapter of Box and Jenkins (1976). If there is a data set that should skew towards the X-13 specification, it is this data.[4]

Figure 4 displays the average Root Mean Squared Error (RMSE) of the out-of-sample data forecasts relative to the (correctly specified) Ordinary Least Squares (OLS) benchmark across a 12-month horizon ($h = 1, \ldots, 12$), where the average is over $N = 200$ simulations. Specifically, the RMSE for model m at horizon h, aggregated across all N simulations, is given by

$$\text{RMSE}_h^{(m)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(e_{i,h}^{(m)}\right)^2} \tag{64}$$

To facilitate direct comparison and isolate the excess error introduced by model misspecification or finite-sample estimation, we evaluate performance relative to the OLS benchmark. The *Relative RMSE* plotted in the results is defined as

$$\text{Relative RMSE}_h^{(m)} = \frac{\text{RMSE}_h^{(m)}}{\text{RMSE}_h^{(\text{OLS})}} - 1 \tag{65}$$

The OLS baseline is centered exactly at zero and values above zero indicate higher (worse) error than OLS, while values below indicate lower (better) error. Circles denote the average relative RMSE for the AutoGP estimated models with training lengths $T = 36$ (green), $T = 72$ (orange), and $T = 144$ (purple). Squares denote the average relative RMSE for the X-13 model. Vertical bars connecting the X-13 and AutoGP markers serve to highlight the magnitude and direction of the head-to-head performance gap between the two models at each respective horizon.

For out-of-sample data forecasts, Figure 4 illustrates that AutoGP consistently achieves a lower RMSE than the X-13 baseline across nearly all horizons and sample sizes. Given a sufficient sample size ($T = 144$), AutoGP uniformly dominates, yielding statistically significant MSE reductions over X-13 across all twelve forecast horizons ($q < 0.05$). AutoGP approaches the OLS benchmark for $T = 144$, with an RMSE within 5% at all horizons. When data is scarce ($T = 36$), AutoGP's flexibility outperforms X-13 at all horizons beyond the first ($h \ge 2$, $q < 0.01$) and even outperforms the OLS specification at certain horizons. At the intermediate sample size ($T = 72$), the performance gap narrows. AutoGP achieves statistically significant error reductions at select intermediate horizons (e.g., $h \in \{3, 4, 5, 10, 11\}$), while performing statistically on par with X-13 elsewhere. Notably, the only instance where X-13 exhibits a slight numerical edge— the one-step-ahead forecast at $T = 72$—fails to reach statistical significance at the 5% level ($q = 0.071$).

Figure 5 reports the Mean Interval Score (MIS) for the data. For a given simulation i and forecast horizon h, let $y_{i,h}$ denote the realized true data, and let $\left[L_{i,h}^{(m)}, U_{i,h}^{(m)}\right]$ denote the central $(1-\alpha)$ prediction interval produced by model $m \in \{\text{AutoGP}, \text{X-13}, \text{OLS}\}$. The interval score for a single forecast is defined as

$$\text{IS}_{i,h}^{(m)} = \left(U_{i,h}^{(m)} - L_{i,h}^{(m)}\right) + \frac{2}{\alpha}\max\left(L_{i,h}^{(m)} - y_{i,h}, 0\right) + \frac{2}{\alpha}\max\left(y_{i,h} - U_{i,h}^{(m)}, 0\right), \tag{66}$$

where $\alpha = 0.05$ corresponds to our 95% predictive intervals. The aggregated MIS for model m at horizon h is simply

[4] Moreover, X-13 failed to converge in roughly 3% of the simulations. Assuming these observations, if modeled, would have pushed X-13 forecast errors up, the numbers reported below are biased down.

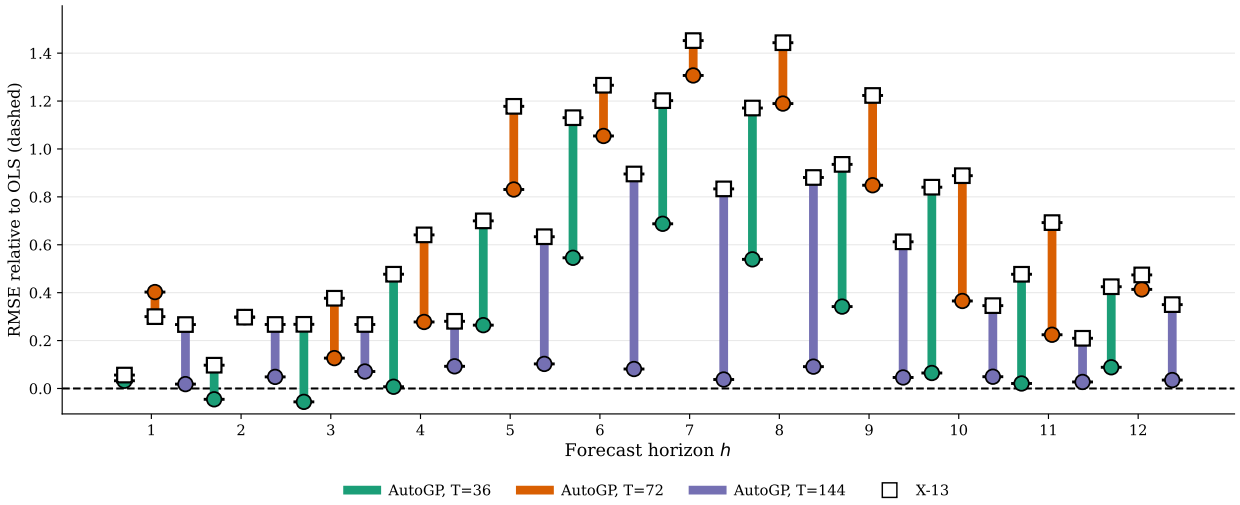


Figure 4: **Relative RMSE performance for the Data target.** The plot compares the Root Mean Squared Error (RMSE) of AutoGP and X-13-ARIMA-SEATS relative to the OLS baseline across forecast horizons $h = 1, \ldots, 12$. The vertical axis represents the relative metric, calculated as $(\text{RMSE}_{\text{model}}/\text{RMSE}_{\text{OLS}}) - 1$. The dashed line at 0 represents parity with the OLS model. Colored circles denote AutoGP performance for training sample sizes $T = 36$ (green), $T = 72$ (orange), and $T = 144$ (purple). Gray squares denote X-13 performance. Vertical bars connect the AutoGP and X-13 results for the same horizon and sample size to facilitate direct comparison.

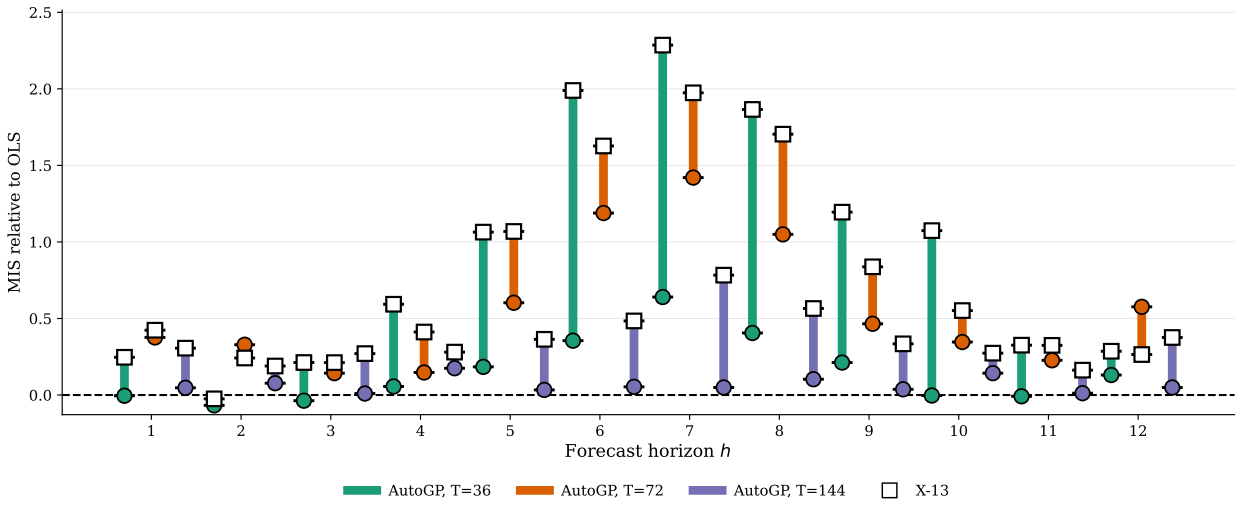


Figure 5: **Relative MIS performance for the Data target.** The plot compares the Mean Interval Score (MIS) of AutoGP and X-13 relative to the OLS baseline for the overall data. The vertical axis shows the relative MIS, defined as $(\text{MIS}_{\text{model}}/\text{MIS}_{\text{OLS}}) - 1$. Positive values indicate that the model's predictive intervals are less efficient (wider or with lower coverage) than the OLS baseline. AutoGP (circles) and X-13 (squares) are shown for horizons $h = 1, \ldots, 12$ and sample sizes $T \in \{36, 72, 144\}$. Vertical bars connect the AutoGP and X-13 markers for direct comparison within each horizon and sample size group.

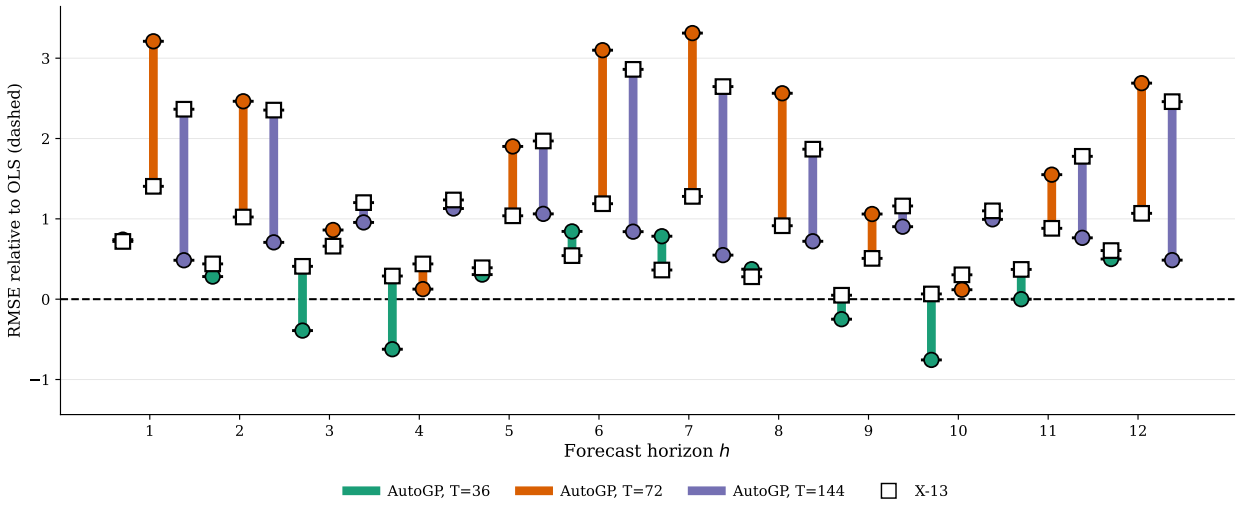


Figure 6: **Relative RMSE performance for the Seasonal target.** The plot compares the RMSE of AutoGP and X-13 relative to OLS for the extracted seasonal component. Plotting conventions follow Figure 4: the y-axis shows performance relative to OLS, where 0 is parity. AutoGP (circles) and X-13 (squares) are compared for sample sizes $T \in \{36, 72, 144\}$.

the average of these scores across all $N = 200$ Monte Carlo replications

$$\text{MIS}_h^{(m)} = \frac{1}{200} \sum_{i=1}^{200} \text{IS}_{i,h}^{(m)}. \tag{67}$$

And similar to our point forecast evaluation, we isolate the excess inefficiency of the predictive intervals by measuring performance relative to the correctly specified OLS model.

A similar pattern emerges for the quality of the interval forecasts in Figure 5. Given a sufficient training sample ($T = 144$), AutoGP produces more efficient predictive intervals, achieving a lower average MIS than X-13 across all horizons, with these reductions reaching statistical significance ($q < 0.05$) in eight of the twelve periods. Again with $T = 144$ observations, the MIS of the AutoGP model is nearly identical to the OLS baseline.

With limited data ($T = 36$), AutoGP significantly outperforming X-13 from horizons $h = 4$ through $h = 10$ ($q < 0.01$). At the intermediate sample size ($T = 72$), the comparative efficiency of the intervals is more mixed, with the two models performing statistically on par at nearly all horizons. Taken together, these results indicate that AutoGP not only provides more accurate point forecasts of the observed data but also, across most conditions, delivers more reliable and efficient predictive intervals vis-a-vis the X-13 model.

The results for the *seasonal target* are more nuanced and informative about the tradeoff between flexibility and imposed seasonality. For the short sample, $T = 36$, Figure 6 shows that AutoGP generally outperforms X-13 in seasonal RMSE (even outperforming the OLS baseline as asymptotics have yet to firmly take hold). With only three years of data, the seasonal signal is weak relative to the innovation noise, so a method that is willing to assign little or no posterior mass to a seasonal component can improve forecast performance. MSE reductions over X-13 at $h \in \{3, 4, 9, 10\}$ are statistically significant ($q < 0.05$).

The interpretation of the seasonal-target results is reinforced by an explicit posterior model-structure diagnostic. For each Monte Carlo run, we inspect the retained AutoGP particle population and classify a particle as seasonal if its kernel contains a Periodic term. For $T = 36$, the periodic component is either not modeled or given very little weight. AutoGP is correctly concluding that the evidence for seasonality is weak at that sample size. This

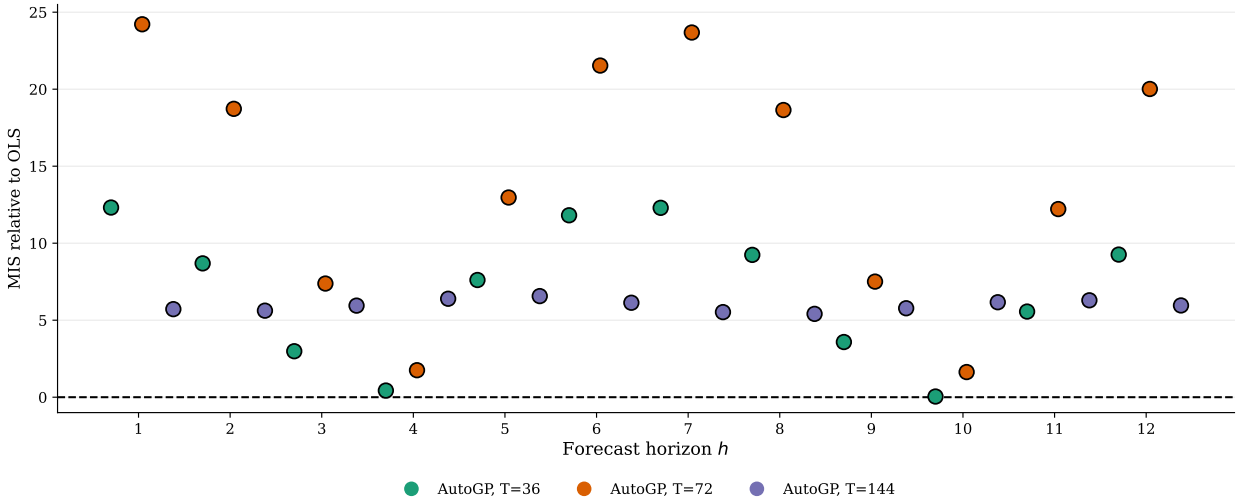


Figure 7: **Relative MIS performance for the Seasonal target.** The plot compares the MIS of AutoGP relative to OLS for the extracted seasonal component. Note that X-13 is excluded from this comparison due to the lack of valid predictive intervals for its seasonal factors.

helps explain why AutoGP performs well at the short horizon: unlike X-13, which is designed to extract seasonality from monthly data, AutoGP is free to withhold posterior mass from a seasonal component when the signal-to-noise ratio is low.

At the intermediate sample size, $T = 72$, the ranking reverses. Here X-13 outperforms AutoGP at most horizons—achieving statistically significant error reductions across ten of the twelve periods ($q < 0.05$). This suggests that six years of monthly observations provide enough information for a dogmatic prior over seasonality to be useful.

At the long sample, $T = 144$, however, the ranking reverses substantially and AutoGP clearly dominates X-13 at every horizon. Once the sample is sufficiently informative, the flexibility of probabilistic model discovery outweighs the advantage of imposing seasonality mechanically, yielding statistically significant MSE reductions for AutoGP in ten out of twelve forecast horizons ($q < 0.05$).

Figure 7 provides the interval-forecast analogue for the seasonal component. Here the comparison is necessarily limited to AutoGP because X-13 does not provide usable predictive intervals for the seasonal factor in this exercise; the corresponding intervals often collapse to zero width. This is itself informative. A central advantage of the Bayesian framework is that uncertainty about seasonality is treated as a first-order object rather than suppressed by construction. Relative to the OLS benchmark, AutoGP's seasonal MIS remains substantially larger, reflecting the difficulty of learning both the existence and the shape of seasonality from finite samples. Nevertheless, when combined with the RMSE evidence, the overall picture is clear: AutoGP dominates X-13 for forecasting the observed data across nearly all designs, and it dominates X-13 for forecasting the seasonal component in both very short and long samples. The only environment in which X-13 has a clear advantage is the intermediate sample length, where seasonality is present strongly enough to be exploitable but not yet so strongly to exploit the additional flexibility of AutoGP.

4 AN APPLICATION TO RETAIL SALES

We now apply our probabilistic seasonal-adjustment framework to U.S. Advance Retail Sales. We analyze the non-seasonally-adjusted retail-sales series (Fred handle, RSAFSNA) from January 1992 to August of 2025. This application is useful for two reasons. First, retail sales is an important statistic as it correlates strongly with consumption and hence, GDP. Second, it is among the most visibly seasonal time series. Data revisions associated with retail sales are almost certainly due to seasonal adjustments. This does not make it unique but other seasonal time series have other components that contribute to revisions (e.g., labor data and survey response rates).

4.1 PREPROCESSING AND ESTIMATION SPECIFICATIONS

Before estimating the latent seasonal component, we first remove deterministic trading-day effects. This step is necessary because data are observed only at the monthly frequency, while economically meaningful variation in retail activity often occurs at the within-month frequency. In particular, the exact number of Fridays, Saturdays, and Sundays in a given month can materially affect total monthly sales even when the underlying seasonal pattern is unchanged. A month with five Fridays and Saturdays will typically record stronger sales than an otherwise similar month with only four. If these calendar-composition effects are left in the data, they are absorbed into the monthly time series as apparent low-frequency seasonal variation. This is a classic aliasing problem as high-frequency deterministic variation generated by the arrangement of weekdays within the month is observed only through a coarser monthly aggregate, so it becomes confounded with the latent seasonal component we seek to estimate. As a result, our probabilistic model will incorrectly attribute predictable calendar variation to evolving seasonality.

For each monthly observation, we compute the exact number of Mondays through Sundays in that calendar month. From these counts, we construct six variables, $T_{\text{Mon}}, \ldots, T_{\text{Sat}}$, each defined relative to Sunday (which is omitted to avoid perfect multicollinearity). Let Y_t denote raw retail sales and let $D_{j,t}$ denote the count of weekday j in month t. We define

$$T_{j,t} = D_{j,t} - D_{\text{Sun},t}, \qquad j \in \{\text{Mon}, \text{Tue}, \text{Wed}, \text{Thu}, \text{Fri}, \text{Sat}\} \tag{68}$$

and then estimate the log-additive trading-day effect in first differences, controlling for ordinary month-of-year seasonality with monthly dummy variables. The estimated trading-day coefficients confirm that weekday composition is economically relevant in retail sales, with Friday exhibiting the largest positive effect relative to Sunday. With $y_t = \log Y_t$, the pre-processing regression is

$$\Delta y_t = \alpha + \sum_{j \in \{\text{Mon},\ldots,\text{Sat}\}} \beta_j \, \Delta T_{j,t} + \sum_{m=2}^{12} \gamma_m M_{m,t} + u_t, \tag{69}$$

where $M_{m,t}$ are month dummies. The estimated coefficients $\hat{\beta}_j$ are then used to reconstruct the trading-day component in levels,

$$\widehat{\text{TD}}_t = \sum_{j \in \{\text{Mon},\ldots,\text{Sat}\}} \hat{\beta}_j \, T_{j,t},$$

and the scrubbed series is defined by

$$Y_t^{\text{scrub}} = \exp\big(\log Y_t - \widehat{\text{TD}}_t\big). \tag{70}$$

For numerical stability, the cleaned series is modeled in log space, $z_t = \log Y_t^{\text{scrub}}$ and standardized

$$z_t^* = \frac{z_t - \mu_{\log}}{\sigma_{\log}}, \qquad \mu_{\log} = \frac{1}{T}\sum_{t=1}^{T} z_t, \qquad \sigma_{\log} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(z_t - \mu_{\log})^2} \tag{71}$$

The model is then fit to $\{z_t^*\}_{t=1}^T$ with posterior objects mapped back into log units and, when needed, into the original data scale.

The period prior of the periodic kernel is centered at 12 months by setting $p \sim \text{LogNormal}\big(\log 12,\ 0.15^2\big)$ which is more informative than the generic default and reflects prior knowledge that the dominant seasonal cycle in retail sales is annual. This does not force a seasonal component on the model.

We use 16 particles with the number of MCMC rejuvenation steps per stage set to 250 and the number of HMC updates of 25. The staged schedule is built from the sequence [12, 24, 36, 72, 144, 288404], which helps the model find temporal patterns in the data. Finally, the HMC configuration uses step sizes `eps_param=0.01` and `eps_noise=0.01` with `n_exit=4`, which are standard, Saad et al. (2023).

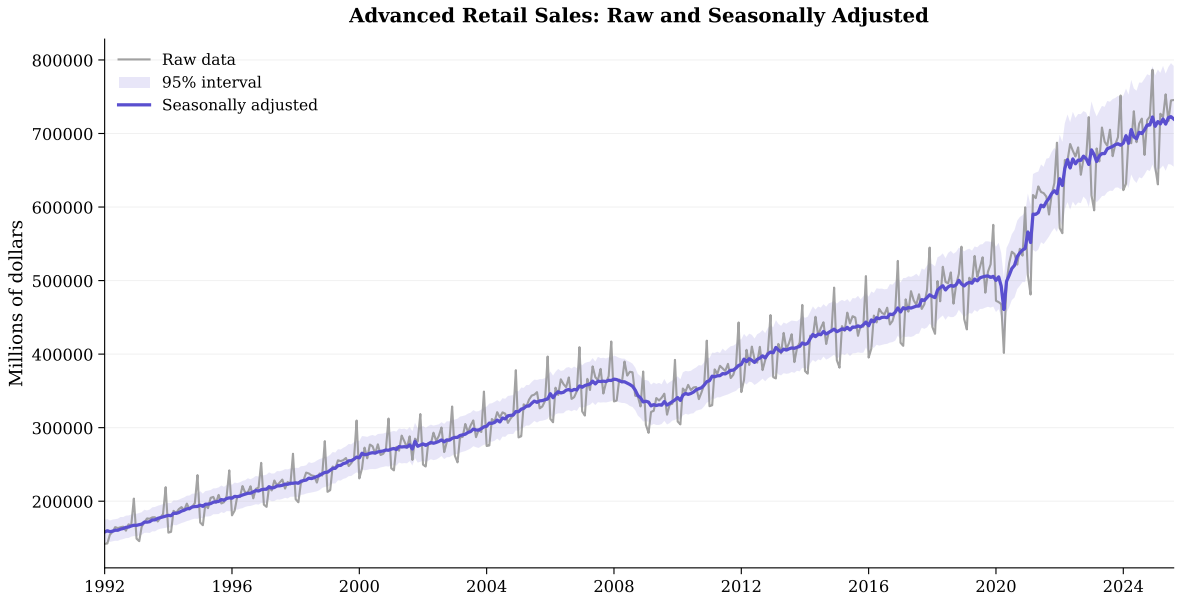


Figure 8: Seasonally adjusted and non-seasonally adjusted retail sales. The shaded region reports the 95% posterior credible interval for the adjusted series.

4.2 CONSTRUCTING THE IN-SAMPLE SEASONAL ADJUSTMENT

The in-sample seasonal-adjustment script operationalizes the theoretical decomposition of Section 2. As discussed in that section, additive decompositions are identified only up to a constant. The retail-sales implementation therefore fits a simple linear trend

$$m_{1,t}^{(p)} = a_p + b_p t + \eta_t^{(p)}, \qquad t = 1, \ldots, T, \tag{72}$$

and interprets the fitted line

$$L_t^{(p)} = \hat{a}_p + \hat{b}_p t \tag{73}$$

as low-frequency leakage that should not be attributed to the seasonal component. This leakage term is then subtracted from the seasonal component and, symmetrically, added to the nonseasonal component at the particle level.[5]

Let $\widehat{s}_t$ denote the resulting posterior mean of the log seasonal component. The seasonally adjusted retail-sales

[5] This Method A of Section 2, Method B did not yield significantly different results.

series is then constructed as

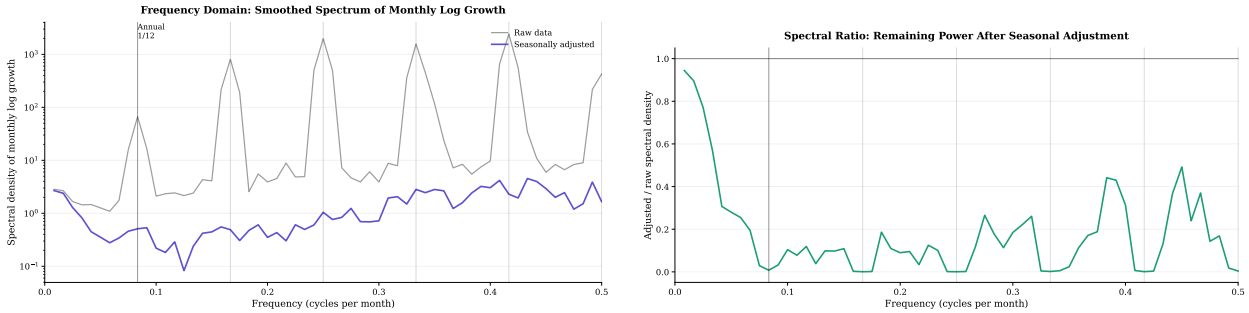
$$Y_t^{\mathrm{SA}} = \exp\big(y_t - \widehat{s}_t\big), \tag{74}$$

with pointwise lower and upper bands obtained analogously from the posterior quantiles of the seasonal component:

$$Y_{t,\mathrm{lo}}^{\mathrm{SA}} = \exp\big(y_t - s_{t,\mathrm{hi}}\big), \qquad Y_{t,\mathrm{hi}}^{\mathrm{SA}} = \exp\big(y_t - s_{t,\mathrm{lo}}\big). \tag{75}$$

4.3 TIME-DOMAIN AND FREQUENCY-DOMAIN DIAGNOSTICS

Figure 8 plots both the original and seasonally adjusted data and shows that the probabilistically seasonally adjusted series preserves the low-frequency evolution of retail sales while removing the regular within-year oscillations present in the raw data. Second, and more importantly, it displays the posterior credible interval around the adjusted series, which widens noticeably during recessionary episodes and other periods of macroeconomic stress, indicating that uncertainty about the latent seasonal adjustment itself is state dependent. This feature is central to the probabilistic approach. Rather than producing a single deterministic adjustment, our procedure yields a full posterior distribution over the seasonal component and, therefore, over the seasonally adjusted series. In tranquil periods the credible region is relatively tight, reflecting greater agreement across models. In recessions, however, the band expands, signaling that the distinction between persistent nonseasonal movements and shifts in seasonal structure becomes harder to parse. A measure of uncertainty is precisely the type of information a probabilistic seasonal-adjustment procedure is designed to deliver.



(a) Spectral density of raw and adjusted monthly log growth.

(b) Ratio of adjusted to raw spectral density.

Figure 9: Frequency-domain diagnostics for retail sales. The left panel plots the smoothed spectra of monthly log growth in the raw and seasonally adjusted series. The right panel plots the ratio of adjusted to raw spectral density.

Figure 9 provides frequency-domain evidence that seasonality is attenuated. The left panel plots the smoothed spectra of monthly log growth in the raw and adjusted series. The raw series exhibits substantial spectral power at the annual seasonal frequency $1/12$, whereas the adjusted series sharply attenuates those peaks while preserving low frequency components. The right panel makes the same point in ratio form by plotting adjusted spectral density relative to raw spectral density. Values below one in the neighborhood of the seasonal frequencies indicate that the adjustment removes power precisely where seasonal variation is concentrated. At the same time, the ratio remains much closer to one away from the seasonal band, and especially at low frequencies. Thus, employing the median of our estimated values effectively removes the dominant seasonal component from retail sales data while quantiles provide a probabilistic notion of seasonality.

5 CONCLUSION

This paper introduces a novel probabilistic framework for seasonal adjustment based on a Gaussian process approach for automated structure discovery in time series. The core contribution is the identification and modeling of the seasonal component itself as an integral part of a flexible, data-driven GP decomposition. Our approach automatically discovers compositional kernel structures that capture non-stationary and evolving seasonality, while delivering full posterior inferences over latent seasonal functions. This approach yields more stable adjusted series, smaller and better-calibrated revisions, improved forecast accuracy, and transparent uncertainty quantification—advantages that are particularly valuable when seasonal patterns shift due to economic shocks.

Monte Carlo simulations based on a growing-sine data-generating process (designed to mimic classic economic benchmarks such as Air Passengers) demonstrate that the method reliably recovers the true seasonal structure in both weak and strong seasonality environments, outperforming X-13ARIMA-SEATS in point forecasts of the observed series and providing superior predictive intervals. An application to monthly U.S. Advance Retail Sales confirms effective removal of seasonality while providing time-varying credible intervals.

Future work will apply this framework to a broader range of economic time series to better understand the properties and robustness of the resulting seasonal adjustments. Promising directions include labor market indicators (e.g., nonfarm payroll employment or unemployment rates), price indices (CPI and PPI components), industrial production, and international trade data. Such extensions would allow systematic evaluation of revision volatility across different economic sectors, the method's ability to handle abrupt regime shifts (such as those induced by pandemics or policy interventions), interactions with trading-day and holiday effects in multivariate settings, and performance under varying degrees of seasonality.

REFERENCES

Bell, William R. and Steven C. Hillmer, "Modeling Time Series with Calendar Variation," *Journal of the American Statistical Association*, 1983, *78* (383), 526–534.

_ **and** _ , "Issues Involved with Seasonal Adjustment of Economic Time Series," *Journal of Business & Economic Statistics*, 1984, *2* (4), 291–320.

Box, George E. P. and David R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society: Series B*, 1964, *26* (2), 211–252.

_ **and George C. Tiao**, "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 1975, *70* (349), 70–79.

_ **and Gwilym M. Jenkins**, *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day, 1976.

Brown, Robert G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Englewood Cliffs, NJ: Prentice Hall, 1963.

Chang, I., George C. Tiao, and Chung Chen, "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 1988, *30* (2), 193–204.

Chen, Chung and Lon-Mu Liu, "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 1993, *88* (421), 284–297.

Chopin, Nicolas and Omiros Papaspiliopoulos, *An Introduction to Sequential Monte Carlo* Springer Series in Statistics, Cham: Springer, 2020.

Cleveland, William S., Douglas M. Dunn, and Irma J. Terpenning, "SABL: A Resistant Seasonal Adjustment Procedure with Graphical Methods for Interpretation and Diagnosis," in Arnold Zellner, ed., *Seasonal Analysis of Economic Time Series*, U.S. Department of Commerce, Bureau of the Census Washington, D.C. 1983, pp. 201–245.

Dagum, Estela Bee, *The X-11-ARIMA Seasonal Adjustment Method*, Ottawa: Statistics Canada, 1980. Catalogue No. 12-564E.

Duvenaud, David, James Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani, "Structure discovery in nonparametric regression through compositional kernel search," in "Proceedings of the 30th International Conference on Machine Learning," Vol. 28 of *Proceedings of Machine Learning Research* PMLR 2013, pp. 1166–1174.

Ghysels, Eric and Pierre Perron, "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root," Research Memorandum 355, Econometric Research Program, Princeton University October 1990.

Gomez, Víctor and Agustín Maravall, *Programs TRAMO and SEATS: Instructions for the User*, Madrid: Banco de España, 1996.

Gómez, Víctor and Agustín Maravall, "Automatic Modeling Methods for Univariate Series," in D. Peña, G. C. Tiao, and R. S. Tsay, eds., *A Course in Time Series Analysis*, New York: John Wiley & Sons, 2001, pp. 171–201.

_ **and** _ , "Seasonal Adjustment and Signal Extraction in Economic Time Series," in D. Peña, G. C. Tiao, and R. S. Tsay, eds., *A Course in Time Series Analysis*, New York: John Wiley & Sons, 2001, pp. 202–246.

Granger, Clive W. J., "Seasonality: Causation, Interpretation, and Implications," in Arnold Zellner, ed., *Seasonal Analysis of Economic Time Series*, U.S. Department of Commerce, Bureau of the Census, 1978, pp. 33–46.

Harvey, Andrew C, *Forecasting, structural time series models and the Kalman filter*, Cambridge university press, 1989.

Hillmer, Steven C. and George C. Tiao, "An ARIMA-Model-Based Approach to Seasonal Adjustment," *Journal of the American Statistical Association*, 1982, *77* (377), 63–70.

_ **, William R. Bell, and George C. Tiao**, "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," in Arnold Zellner, ed., *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, 1983, pp. 74–100.

Ladiray, Dominique and Benoît Quenneville, *Seasonal Adjustment with the X-11 Method* number 158. In 'Lecture Notes in Statistics.', New York, NY: Springer-Verlag, 2001.

Liu, Jun S., *Monte Carlo Strategies in Scientific Computing* Springer Series in Statistics, New York: Springer-Verlag, 2004.

Micchelli, Charles A., Yuesheng Xu, and Haizhang Zhang, "Universal Kernels," *Journal of Machine Learning Research*, 2006, *7*, 2651–2667.

Monsell, B. C., "An Overview of the Automatic ARIMA Model Selection Procedure in X-13ARIMA-SEATS," 2002. Unpublished.

_ , "Further Details on the Automatic ARIMA Model Selection Procedure in X-13ARIMA-SEATS," 2006. Unpublished.

Neklyudov, Kirill, Max Welling, Evgenii Egorov, and Dmitry Vetrov, "Involutive MCMC: A Unifying Framework," in "Proceedings of the 37th International Conference on Machine Learning," Vol. 119 of *Proceedings of Machine Learning Research* PMLR 2020, pp. 7273–7282.

Rasmussen, Carl E. and Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2006.

Robert, Christian P. and George Casella, *Monte Carlo Statistical Methods*, 2 ed., New York: Springer, 2004.

Saad, F. A. K., "Scalable Structure Learning, Inference, and Analysis with Probabilistic Programs." PhD dissertation, Massachusetts Institute of Technology 2022.

Saad, Feras A., Brian J. Patton, Matthew D. Hoffmann, Rif A. Saurous, and V. K. Mansinghka, "Sequential Monte Carlo Learning for Time Series Structure Discovery," in "Proceedings of the 40th International Conference on Machine Learning," Vol. 202 of *Proceedings of Machine Learning Research* PMLR 2023, pp. 29473–29489.

_ **, Marco F. Cusumano-Towner, Ulrich Schaechtle, Martin C. Rinard, and Vikash K. Mansinghka**, "Bayesian Synthesis of Probabilistic Programs For Automatic Data Modeling," *Proceedings of the ACM on Programming Languages*, January 2019, *3* (POPL), 37:1–37:32.

Shiskin, Julius, Allan H. Young, and John C. Musgrave, "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper 15, U.S. Department of Commerce, Bureau of the Census, Washington, D.C. February 1967.

Tukey, John W., "On the Comparative Anatomy of Transformations," *The Annals of Mathematical Statistics*, 1957, *28* (3), 602–632.

U.S. Census Bureau, "X-13ARIMA-SEATS Reference Manual," Software Documentation, U.S. Census Bureau, Center for Statistical Research and Methodology, Washington, DC June 2024. Version 1.1, Build 61 (or current version). Available at `https://www.census.gov/data/software/x13as.html`.

Watson, Mark W., "Uncertainty in Model-Based Seasonal Adjustment Procedures and Construction of Minimax Filters," *Journal of the American Statistical Association*, 1987, *82* (398), 395–408.

West, Mike and Jeff Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed., New York: Springer, 1997.

Wright, Jonathan H., "Unseasonal Seasonals?," *Brookings Papers on Economic Activity*, 2013, pp. 65–126.