Probabilistic Seasonality

Feras A. Saad[*] Todd B. Walker[†]

September 2026

Abstract

Seasonal adjustment is fundamental to economic analysis, but uncertain because seasonal components are inherently latent. This article introduces a probabilistic model discovery method that decomposes a time series into seasonal and nonseasonal components. The method returns a posterior distribution over the structure and parameters of a seasonal component. In simulation studies, the method can improve point forecasts, interval predictions, and recovery of seasonal components relative to X-13ARIMA-SEATS. In a study of eight U.S. macroeconomic series during the COVID-19 recession, the method surfaces significant ex-ante uncertainty about current seasonal adjustments in real time, well before many X-13 revisions reach their eventual peaks.

[*]Computer Science Department, Carnegie Mellon University, fsaad@cmu.edu.

[†]Department of Economics, Indiana University, walkertb@iu.edu

1 INTRODUCTION

Many economic time series, such as retail sales, unemployment, and inflation, exhibit strong seasonal patterns that are driven by calendar effects, weather, holidays, and institutional factors. The magnitude of this within-year variation can even exceed the impact of recessions. In nonfarm payroll employment, for example, the average absolute difference between seasonally adjusted and unadjusted monthly changes is much larger than the typical month-to-month variation in the adjusted data (Wright 2013). The removal of recurring fluctuations allows analysts, policymakers, and forecasters to distinguish underlying trends, business-cycle movements, and irregular shocks more clearly. Accurate seasonal adjustment is therefore essential for real-time monitoring of economic conditions, constructing reliable leading indicators, and evaluating policy. These issues have received renewed attention following unusually large downward revisions, released in August 2025, to U.S. nonfarm payroll employment changes for May and June by a combined 258,000 jobs (U.S. Bureau of Labor Statistics 2025), which prompted heightened public scrutiny of the production and revision of official labor-market statistics (Mutikani 2025).

Distinguishing seasonal variation from cyclical and irregular movements has nevertheless remained a fundamental problem, as formally articulated in Nerlove (1964). Several recent observations make the point. Rudebusch et al. (2015) document systematically weak first-quarter GDP growth that was subsequently attributed to residual seasonality surviving in the published seasonally adjusted aggregates (Lunsford 2017; Consolvo and Lunsford 2019). Lunsford (2025) finds a similar pattern in five measures of PCE inflation. At the onset of the COVID-19 pandemic, the collapse in economic activity coincided with the seasonal trough in the winter housing market, leading standard procedures to overstate the level of housing starts immediately before the collapse and therefore the apparent depth of the subsequent contraction (Bryson and Cornwall 2026). These examples illustrate that seasonal components are inherently latent and can exhibit significant uncertainty.

Conventional approaches to seasonal adjustment typically report only a conditional mean, with little-to-no information about revisions, turning points, and cyclical contamination. Major seasonal adjustment procedures used by government agencies—such as X-13ARIMA-SEATS—perform a battery of frequentist specification tests to identify and remove seasonality (U.S. Census Bureau 2025). While each test may appear statistically valid in isolation, the overall procedure lacks a unified inferential framework. Moreover, the procedure delivers only point estimates of seasonal factors. Once these estimates are obtained, uncertainty about the decomposition itself is typically discarded or treated as a secondary diagnostic.

We introduce an alternative approach based on a probabilistic view of seasonality. Our framework rests on Bayesian model discovery within a family of Gaussian process models (Rasmussen and Williams 2006) that can express a variety of time series patterns. The method does not impose a particular form of seasonality a-priori, but instead infers the structure and parameters of periodic patterns—such as additive, multiplicative, or locally varying periodicity—from observed data. The use of periodic structures is consistent with the usual characterization of seasonality as a component whose pattern repeats from year to year (Granger 1978). We term the resulting method, which is based on the Automatic Gaussian Process (AutoGP) method of Saad et al. (2023), *Seasonal AutoGP*.

Given a time series dataset, Seasonal AutoGP infers a posterior distribution over its decomposition into seasonal and nonseasonal components. This probabilistic representation supports real-time updates to the model as data becomes available. It also enables direct posterior queries without auxiliary specification tests or statistical re-estimation of the entire model on a per-query basis. In particular, the method (i) discovers the presence and form of periodicity rather than imposing it; (ii) returns a full posterior over seasonal paths rather than a point factor; and (iii) quantifies existence, strength, sign and timing, model disagreement, and revision risk associated with seasonal components. The framework reports seasonally adjusted values along with an ex-ante measure of how much confidence the analyst should place in the decomposition, before its eventual revision is observed.

We first assess the efficacy of the framework in a simulation study that mimics classical economic benchmarks, where the true seasonal component is known. The data-generating process is calibrated to the Box and Jenkins data of monthly international air passengers from 1949–1960. This dataset remains the canonical benchmark for seasonal adjustment and supplies the default $(0,1,1)(0,1,1)_{12}$ "airline model" specification of X-13ARIMA-SEATS itself (Box and Jenkins 1976; U.S. Census Bureau 2025), so the simulation is intentionally staged on terrain that favors X-13. Fitting a sinusoid with linearly growing amplitude to the airline data captures the upward trend, annual seasonal pattern, and widening peak-to-trough variation of the series. For the observed series, Seasonal AutoGP produces lower prediction errors than X-13 across nearly all forecast horizons and sample lengths, and its prediction intervals exhibit improved coverage and sharpness across most experimental conditions. The experimental results on predicting the seasonal component reveal an informative tradeoff between imposing seasonality (as in X-13) and discovering it from the data (as in Seasonal AutoGP). When the seasonal signal is weak, neither method dominates. At intermediate signal-to-noise ratios, the imposed seasonal structure from X-13 is more accurate. In the longest samples, where the data contain sufficient information to identify periodic structure, Seasonal AutoGP reduces prediction error at ten of the twelve forecast horizons.

We then study eight monthly U.S. macroeconomic series spanning labor markets, consumer prices, industrial production, retail spending, residential construction, manufacturing demand, and business inventories. Compared to X-13, Seasonal AutoGP produces lower twelve-month not-seasonally-adjusted point forecast errors for six of the eight series, and a lower interval forecast error in seven. The largest reductions include 33 percent for core CPI and 24 percent for retail sales. Both Seasonal AutoGP and X-13 nonetheless remove similar amounts of conventional fixed-frequency seasonal variation. In most cases, the residual seasonal-harmonic power is at most a few percent of the unadjusted power under either procedure. These results show that Seasonal AutoGP's forecast improvements do not occur at the cost of materially weaker removal of conventional fixed-frequency seasonality.

Our final application compares the extent to which X-13 and Seasonal AutoGP respond to large changes that occurred in these eight time series during the COVID-19 recession. We find that X-13 and Seasonal AutoGP produce comparable revisions, but the former produces revisions much later in the sample period. Estimates from X-13 show that the seasonal components initially assigned to the COVID-19 shock month were subsequently revised by multiples of their ordinary historical revisions, and typically only after long delays. For example, peak revisions from X-13 arrive 37 months later for payroll employment and housing starts, 33 months later for retail sales, and at least ten months later for every series except durable-goods orders. Because these revisions become observable only as later data arrive, conventional revision diagnostics identify the instability ex post.

Using Seasonal AutoGP, we exploit the probabilistic nature of seasonal adjustments to define a so-called "Revision Risk" distribution, which provides a coherent measure of the uncertainty in a seasonal adjustment at the time of release. Informally, the Revision Risk answers the question: *How large are future revisions to the current seasonal adjustment likely to be?* The Revision Risk from Seasonal AutoGP already exceeds its historical 95th percentile for four of eight series at the COVID-19 shock month itself, and for all eight just three months later, when the median index is 8.1 times the historical median. While X-13 and Seasonal AutoGP eventually agree about the seasonal pattern—with a median correlation between the seasonal components of 0.972—the timing and uncertainty of the adjustments are distinct. Seasonal AutoGP makes most of the adjustment between two and five months after the shock and reports the associated uncertainty in real time. In contrast, X-13 defers most rewriting for one to three years later, and offers no analogous ex-ante probabilistic distribution of future revisions to seasonal adjustments. This study underscores that a central contribution of the probabilistic approach is its ability to monitor the uncertainty in adjustments in real time, instead of treating uncertainty using retrospective diagnostics as in X-13.

The remainder of this paper is structured as follows: Section 2 describes probabilistic model discovery and

AutoGP; Section 3 introduces its extension to Seasonal AutoGP; Section 4 presents a simulation study on airline data; Section 5 analyzes eight U.S. economic datasets; and Section 6 concludes the paper.

2 Probabilistic Model Discovery

We begin with a description of our approach to learning probabilistic models of time series data.

2.1 Bayesian Formulation Let $\mathscr{M} = \{m_1, m_2, \dots\}$ denote a set of statistical models. Each model $m \in \mathscr{M}$ defines a likelihood function $f(\mathbf{y};\theta, m)$ for observable data $\mathbf{y}$, parameterized by $\theta \in \Theta_m$. Using a prior distribution $\pi(m)$ over models along with a prior distribution $\pi(\theta|m)$ over parameters within a model induces the joint distribution

$$p(m,\theta,\mathbf{y}) = \pi(m)\pi(\theta|m)f(\mathbf{y};\theta,m), \tag{1}$$

and in turn the following posterior distributions:

i. over models and parameters,

$$p(m,\theta \mid \mathbf{y}) = \frac{\pi(m)\pi(\theta \mid m)f(\mathbf{y};\theta,m)}{\sum_{m'\in\mathscr{M}}\int_{\Theta_{m'}}\pi(m')\pi(\theta' \mid m')f(\mathbf{y};\theta',m')\,\mathrm{d}\theta'} \qquad (m\in\mathscr{M},\ \theta\in\Theta_m) \tag{2}$$

ii. over the model space,

$$p(m \mid \mathbf{y}) = \int_{\Theta_m} p(m,\theta \mid \mathbf{y}),\mathrm{d}\theta \qquad (m\in\mathscr{M}) \tag{3}$$

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

$$p(\theta \mid \mathbf{y}, m^\star) = \frac{\pi(\theta \mid m^\star)f(\mathbf{y};\theta,m^\star)}{\int_{\Theta_{m^\star}}\pi(\theta' \mid m^\star)f(\mathbf{y};\theta',m^\star)\,\mathrm{d}\theta'} \qquad (\theta\in\Theta_{m^\star}). \tag{4}$$

As the distributions of interest in (2)–(4) are almost always intractable to compute exactly, they are instead approximated using statistical algorithms such as Markov chain Monte Carlo or sequential Monte Carlo (Robert and Casella 2004). These algorithms generate a set $\mathscr{S} = \{(w^{(i)}, (m^{(i)}, \theta^{(i)}))\}_{i=1}^{M}$ of M 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,

$$\mathbb{E}\left[w^{(i)}\varphi\left(m^{(i)},\theta^{(i)},\mathbf{y}\right) \,\middle|\, \mathbf{y}\right] = c\,\mathbb{E}_{p(m,\theta|\mathbf{y})}\left[\varphi(m,\theta,\mathbf{y})\right] \qquad (1 \le i \le M), \tag{5}$$

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

$$\mathbb{E}_{p(m,\theta|\mathbf{y})}\left[\varphi(m,\theta,\mathbf{y})\right] \approx \sum_{i=1}^{M}\left(\frac{w^{(i)}}{\sum_{j=1}^{M} w^{(j)}}\right)\varphi\left(m^{(i)},\theta^{(i)},\mathbf{y}\right). \tag{6}$$

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.

2.2 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 time points $\mathbf{t} := (t_1, \dots, t_n) \in \mathbb{T}^n$, the random vector $X(\mathbf{t}) := (X(t_1), \dots, X(t_n))$ follows a multivariate Gaussian distribution:

$$X(\mathbf{t}) = (X(t_1), \dots, X(t_n)) \sim \mathcal{N}\left(\mu(\mathbf{t}), K(\mathbf{t}, \mathbf{t})\right). \tag{7}$$

The symbol $\mu(\mathbf{t}) := (\mu(t_1), \dots, \mu(t_n))$ denotes the mean vector with $\mu(t) := \mathbb{E}[X(t)]$, and $K(\mathbf{t}, \mathbf{t})$ is the $n \times n$ covariance matrix whose entries are defined by a *covariance kernel*: $[K(\mathbf{t}, \mathbf{t})]_{ij} = k(t_i, t_j) := \mathbb{E}\left[(X(t_i) - \mu(t_i))(X(t_j) - \mu(t_j))\right]$ for all $1 \le i, j \le n$. Conditioned on $X(\mathbf{t}) = x(\mathbf{t})$, the posterior distribution of $X(\mathbf{t}')$ at new time points $\mathbf{t}' = (t'_1, \dots, t'_{n'}) \in \mathbb{T}^{n'}$, is also a multivariate Gaussian:

$$X(\mathbf{t}') \mid X(\mathbf{t}) = x(\mathbf{t}) \sim \mathcal{N}\left(\mu^{\text{post}}(\mathbf{t}'), K^{\text{post}}(\mathbf{t}', \mathbf{t}')\right), \tag{8}$$

which forms the basis of predictive inference on unobserved data given observed data $X(\mathbf{t})$. The specific forms of μ^{post} and K^{post}, as well as further background on Gaussian processes, is given in Appendix A.1.

2.3 Covariance Kernels Gaussian processes can be used to model time series models that exhibit rich temporal characteristics, by appropriate selection of the *covariance kernel* $k : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ that defines the covariance matrix K of the underlying multivariate normal. Covariance kernels provide a compact representation of a potentially infinite-dimensional feature space—rather than explicitly specifying basis functions, the covariance kernel k encodes the similarity between inputs (t, t'), inducing a rich space of functions (Rasmussen and Williams 2006). We will consider the following kernels, whose full definitions are given in Appendix A.3: **Constant** for modeling constant functions, **Linear** for modeling linear functions, **GammaExponential** for modeling smoothly varying functions, and **Periodic** for modeling periodically varying functions. Figure 1 shows representative samples from Gaussian processes with these kernels, excluding measurement noise.

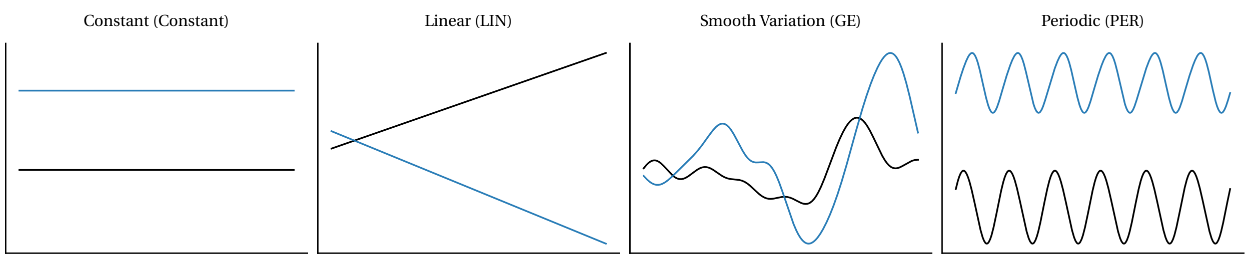

Figure 1: Samples from Gaussian process models with a single covariance kernel.

2.4 Kernel Composition More expressive Gaussian process models can be constructed by composing primitive covariance kernels to form composite kernels. Given a sufficiently expressive set of primitive kernels and composition operators, Gaussian processes can serve as universal approximators for continuous functions on compact sets (Micchelli et al. 2006). Two covariance kernels k_1 and k_2 can be combined to create a new kernel through the following operations:

- **Addition** $k_+(t, t') := k_1(t, t') + k_2(t, t')$. This operator supports additive structure, such as a seasonal component superimposed on a trend.
- **Multiplication** $k_\times(t, t') := k_1(t, t') \times k_2(t, t')$. This operator allows for modulation, such as a periodic component with time-varying amplitude.

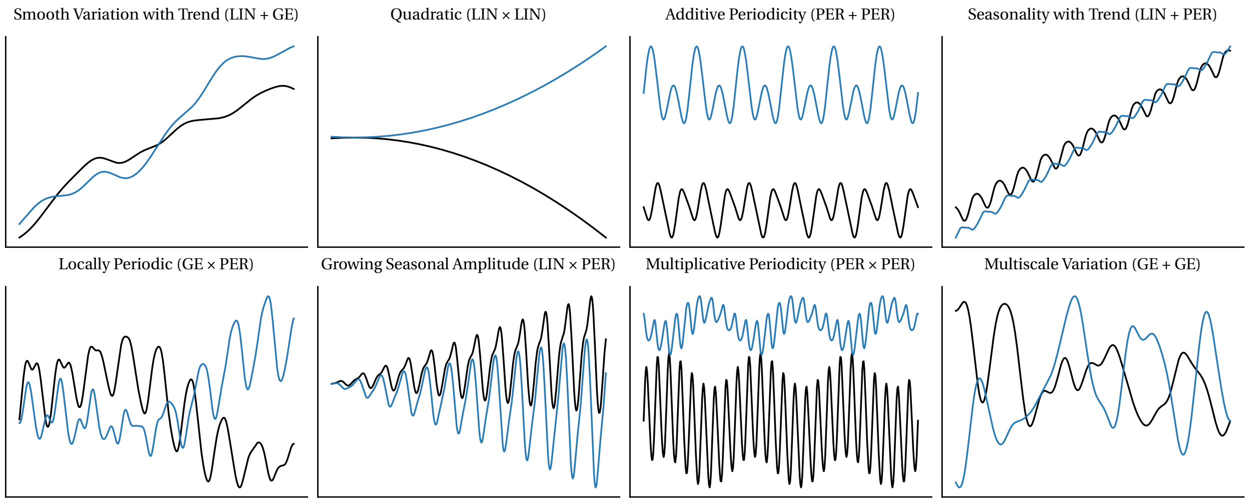

Figure 2: Samples from Gaussian process models with composite covariance kernels. The models can express a variety of time series structures, by varying the primitive covariance kernels and combining them using addition and multiplication.

Figure 2 shows examples of Gaussian process time series data that arise from using different composite covariance kernels, demonstrating a range of structural patterns that the model can express.

2.5 AUTOMATED KERNEL DISCOVERY Recent advances have made it possible to discover the structure of a time-series dataset, by automatically learning the covariance kernel of a Gaussian process model for the data (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 that defines a countable model class $\mathcal{M} = \{m_1, m_2, \dots\}$. This approach allows the econometrician to place a prior distribution over model structures, not just parameters, enabling Bayesian structure learning to perform inference over a space of Gaussian process models. By doing so, it provides both expressivity and interpretability as each sampled model can be inspected as a symbolic expression revealing qualitative features of the time series, such as seasonal variation, trends, and their interactions. Following Saad et al. (2023), we define a prior over model structures $m \in \mathcal{M}$ using a *probabilistic context-free grammar* (PCFG), whose terminal symbols are primitive covariance kernels and whose production rules leverage kernel composition operators:

$$B \sim \text{Constant} \mid \text{Linear} \mid \text{GammaExponential} \mid \text{Periodic} \qquad (\text{w.p. } \alpha_1, \alpha_2, \alpha_3, \alpha_4) \tag{9}$$

$$\oplus \sim + \mid \times \qquad (\text{w.p. } \beta, 1-\beta) \tag{10}$$

$$m \sim B \mid \oplus[m_1, m_2] \qquad (\text{w.p. } \delta, 1-\delta). \tag{11}$$

In the rule for m, the probability $\delta > 1/2$ of selecting a primitive kernel is chosen to ensure that the prior halts almost surely (Saad et al. 2019, §4.2). Equipped with the structure, we next define a prior over numerical parameters of the covariance kernel, which fully defines the prior Gaussian process:

$$\theta_i \mid m \sim p_{m,i} \qquad (1 \le i \le d(m)) \tag{12}$$

$$X(\cdot) \mid m, \theta \sim \text{GP}\left(0, k_{m,\theta}\right). \tag{13}$$

In (12), $d(m)$ denotes the number of numerical parameters within the kernel structure m, e.g., $d(\text{Periodic}) = 3$ and $d(\text{Periodic} + \text{Constant}) = 4$, and $p_{m,i}$ denotes a prior over the parameters of the ith scalar parameter within m (lognormal for all parameters, except for the $\gamma \in (0,2]$ parameter of the GammaExponential kernel which uses a logit normal). In (13), $X(\cdot)$ is the (infinite-dimensional) latent Gaussian process and $k_{m,\theta}$ denotes its covariance function, which is determined by the sampled structure m and parameters θ.

The final step is to define an observation model. Economic time series are often subject to measurement noise. We model noisy data by using a Gaussian process $Y(t) := X(t) + \epsilon(t)$ that is the sum of a latent Gaussian process and i.i.d. Gaussian noise, where $\epsilon(t) \sim \mathcal{N}(0,\eta)$ for $t \in \mathbb{T}$. Assuming an inverse gamma prior over the noise variance η, for any time points $\mathbf{t} = (t_1,\dots,t_n)$ the observable noisy data $Y(\mathbf{t}) := (Y(t_1),\dots,Y(t_n))$ is defined by

$$\eta \sim \text{InvGamma}(1,1) \tag{14}$$

$$Y(t_i) \mid X(\cdot) \sim \mathcal{N}(X(t_i),\eta) \qquad (i = 1,\dots,n). \tag{15}$$

We have thus defined all the ingredients for the joint distribution (1) over structures, parameters, and data.

2.6 POSTERIOR INFERENCE Recall that our goal is to produce samples $\{(w^{(i)},(m^{(i)},\theta^{(i)},\eta^{(i)}))\}_{i=1}^{M}$, called a particle collection, of the hidden variables given an observation $Y(\mathbf{t}) = y(\mathbf{t})$. As the model does not exhibit conjugacy that enables closed-form posterior inference, more sophisticated sampling methods are required to explore the posterior $p(m,\theta,\eta \mid y(\mathbf{t}))$. We use AutoGP[1], which is implemented in the Gen probabilistic programming system (Cusumano-Towner et al. 2019) to conduct approximate posterior inference. The implementation uses sequential Monte Carlo (Chopin and Papaspiliopoulos 2020) and involutive Markov Chain Monte Carlo (Neklyudov et al. 2020) algorithms; see Saad et al. (2023, §3) for full technical details. Our contribution extends this inference engine to enable probabilistic identification, decomposition, and adjustment of seasonal data built on top of the posterior draws, as described in Section 3 . For brevity, however, the figures and discussion below label the resulting estimates "AutoGP," because this engine generates the particles underlying every estimate we report.

3 PROBABILISTIC SEASONALITY

We now extend the AutoGP method from Section 2 to Seasonal AutoGP, which identifies seasonality from the learned Gaussian process models and supports posterior queries. Periodic structure is a defining characteristic of seasonality in time series data (Granger 1978). Our framework aligns with this notion by defining seasonality in terms of Periodic kernels in the posterior over Gaussian process models. By grounding the notion of seasonality in the kernel algebra, we obtain an operational and model-based criterion that is amenable to Bayesian inference. We first give some brief intuition about the periodic kernel, following MacKay (1998, §5.2).

3.1 PERIODIC KERNEL The periodic kernel imposes that the covariance between observations should depend only on their *phase difference* within a fixed period ρ, 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 the similarity between two time points (t,t') using the Euclidean distance $\Delta_{\text{eu}}(t,t') = |t-t'|$,

$$k_{\text{SE}}(t,t') = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\big(\Delta_{\text{eu}}(t,t')\big)^2\right) = \sigma^2 \exp\left(-\frac{1}{2\ell^2}|t-t'|^2\right), \tag{16}$$

1. https://probsys.github.io/AutoGP.jl

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 with period ρ, the time domain is wrapped onto the unit circle, where each time point t is mapped to an angle $\theta(t) = 2\pi t/\rho$. 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 $(2\pi t/\rho,\ 2\pi t'/\rho)$:

$$\Delta_{\text{per}}(t,t') = 2\left|\sin\left(\frac{\pi}{\rho}\left|t-t'\right|\right)\right|. \tag{17}$$

This mapping ensures periodic invariance, i.e., $k(t,t') = k(t+m\rho,\ t'+n\rho)$ for all integers $m,n \in \mathbb{Z}$. Using the distance Δ_{per} instead of Δ_{eu} in the SquaredExponential kernel (16) yields the Periodic kernel:

$$k_{\text{P}}(t,t') = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\left(\Delta_{\text{per}}(t,t')\right)^2\right) = \sigma^2 \exp\left(-\frac{2}{\ell^2}\sin^2\left(\frac{\pi}{\rho}\left|t-t'\right|\right)\right). \tag{18}$$

AutoGP's Bayesian inference procedure discovers likely values of the parameters (ℓ,ρ,σ) as well as how the Periodic kernel enters the overall model through appropriate composition operators in the kernel grammar.

3.2 Extracting Periodicity

Periodic structure is extracted from a Gaussian process model structure m using the split$[m]$ procedure, shown in Figure 3. It decomposes m into a sum of two subexpressions $(m_{\text{per}}, m_{\text{non}})$, where the former captures the periodic structure and the latter captures all remaining nonperiodic structure. The operator distributes over the kernel's sum-of-products expansion, which ensures that for composite structures like (LIN + PER) × GE, the resulting cross-term PER × GE is attributed to the seasonal component while LIN × GE is not.

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

$\text{split}[\text{PER}] := (\text{PER}, \mathbf{0})$	$\text{split}[m_1 + m_2] :=$	$\text{split}[m_1 \times m_2] :=$
$\text{split}[b] := (\mathbf{0}, b),\ b \neq \text{PER}$	let $\text{split}[m_1] \leftarrow (a_1, c_1)$	let $\text{split}[m_1] \leftarrow (a_1, c_1)$
	let $\text{split}[m_2] \leftarrow (a_2, c_2)$	let $\text{split}[m_2] \leftarrow (a_2, c_2)$
	in $(a_1 + a_2, c_1 + c_2)$	in $\big(a_1 \times a_2 + a_1 \times c_2 + c_1 \times a_2, c_1 \times c_2\big)$

Figure 3: 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.

The decomposition procedure in Figure 3 ensures that the covariance kernel k_m for m satisfies $k_m = k_{\text{per}} + k_{\text{non}}$, which gives an additive decomposition of the latent process X into seasonal (S) and nonseasonal (U) components:

$$X(t) = S(t) + U(t), \qquad S \sim \text{GP}\big(0, k_m^{\text{per}}\big), \quad U \sim \text{GP}\big(0, k_m^{\text{non}}\big). \tag{19}$$

For time points $\mathbf{t} := (t_1, \dots, t_n)$, we will momentarily write the random variables $\mathbf{U} := U(\mathbf{t})$, $\mathbf{S} := S(\mathbf{t})$, $\mathbf{X} := X(\mathbf{t})$, $\mathbf{Y} := Y(\mathbf{t})$, $K_m^{\text{per}} := k_{\text{per}}(\mathbf{t},\mathbf{t})$, and $K_m^{\text{non}} := k_{\text{non}}(\mathbf{t},\mathbf{t})$. The prior distribution is given by

$$\begin{bmatrix}\mathbf{S}\\ \mathbf{U}\\ \mathbf{Y}\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_m^{\text{per}} & 0 & K_m^{\text{per}} \\ 0 & K_m^{\text{non}} & K_m^{\text{non}} \\ K_m^{\text{per}} & K_m^{\text{non}} & K_m^{\text{per}} + K_m^{\text{non}} + \eta I \end{bmatrix}\right), \tag{20}$$

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

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

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

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

Using this result and denoting $K := K_m(\mathbf{t},\mathbf{t})$ yields the posterior over the seasonal and nonseasonal components:

$$\mathbb{E}\left[\mathbf{S}\,\middle|\,\mathbf{Y}, m,\theta,\eta\right] = K_m^{\mathrm{per}}\,(K+\eta I)^{-1}\mathbf{Y} \qquad \mathbb{E}\left[\mathbf{U}\,\middle|\,\mathbf{Y}, m,\theta,\eta\right] = K_m^{\mathrm{non}}\,(K+\eta I)^{-1}\mathbf{Y} \tag{23}$$

$$\mathrm{Var}\left[\mathbf{S}\,\middle|\,\mathbf{Y}, m,\theta,\eta\right] = K_m^{\mathrm{per}} - K_m^{\mathrm{per}}\,(K+\eta I)^{-1}K_m^{\mathrm{per}} \qquad \mathrm{Var}\left[\mathbf{U}\,\middle|\,\mathbf{Y}, m,\theta,\eta\right] = K_m^{\mathrm{non}} - K_m^{\mathrm{non}}\,(K+\eta I)^{-1}K_m^{\mathrm{non}} \tag{24}$$

$$\mathrm{Cov}\left(\mathbf{S},\mathbf{U}\,\middle|\,\mathbf{Y}, m,\theta,\eta\right) = -\,K_m^{\mathrm{per}}\,(K+\eta I)^{-1}K_m^{\mathrm{non}}. \tag{25}$$

The last term follows from the off-diagonal block of the joint posterior covariance matrix $\mathrm{Var}[\mathbf{S},\mathbf{U} \mid \mathbf{Y}]$. As off-diagonal block $\mathrm{Cov}(\mathbf{S},\mathbf{U} \mid \mathbf{Y})$ is generally *nonzero*, the two components compete to explain the variance. Adding the means gives the overall posterior mean $\mathbb{E}[\mathbf{X} \mid \mathbf{Y}] = K(K+\eta I)^{-1}\mathbf{Y}$, and the variances obey

$$\mathrm{Var}[\mathbf{X} \mid \mathbf{Y}] = \mathrm{Var}[\mathbf{S} \mid \mathbf{Y}] + \mathrm{Var}[\mathbf{U} \mid \mathbf{Y}] + \mathrm{Cov}(\mathbf{S},\mathbf{U} \mid \mathbf{Y}) + \mathrm{Cov}(\mathbf{U},\mathbf{S} \mid \mathbf{Y}), \tag{26}$$

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

Proposition 1 (Periodic Posterior Decomposition)**.** *For time points $\mathbf{t}' := (t'_1,\dots,t'_{n'})$, the posterior distribution of the periodic component $S(\mathbf{t}') := \left(S(t'_1),\dots,S(t'_{n'})\right)$ identified using the* split *operator, given the observation $Y(\mathbf{t}) = y(\mathbf{t})$ and latent variables (m,θ,η) satisfies*

$$S(\mathbf{t}') \mid m,\theta,\eta, Y(\mathbf{t}) = y(\mathbf{t}) \sim \mathcal{N}\left(\mu_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}'), \Sigma_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}',\mathbf{t}')\right), \tag{27}$$

where

$$\mu_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}') := K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t})\left[K_m(\mathbf{t},\mathbf{t}) + \eta I\right]^{-1}\mathbf{y}(\mathbf{t}), \tag{28}$$

$$\Sigma_{\mathrm{per}}^{(m,\theta,\eta)}(\mathbf{t}',\mathbf{t}') := K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t}') - K_m^{\mathrm{per}}(\mathbf{t}',\mathbf{t})\left[K_m(\mathbf{t},\mathbf{t}) + \eta I\right]^{-1}K_m^{\mathrm{per}}(\mathbf{t},\mathbf{t}'). \tag{29}$$

Using properly weighted posterior samples $\{(w^{(i)}, (m^{(i)},\theta^{(i)},\eta^{(i)}))\}_{i=1}^{M}$, the posterior law of the seasonal component is approximated by a finite mixture that averages over posterior uncertainty in model structure, kernel parameters, and observation noise variance:

$$\mathcal{L}\left(\mathbf{S}(\mathbf{t}') \mid Y(\mathbf{t}) = y(\mathbf{t})\right) \approx \sum_{i=1}^{M}\tilde{w}^{(i)}\mathcal{N}\left(\mu_{\mathrm{per}}^{(i)}(\mathbf{t}'), \Sigma_{\mathrm{per}}^{(i)}(\mathbf{t}',\mathbf{t}')\right), \qquad \tilde{w}^{(i)} := w^{(i)}\Big/\sum_{j=1}^{M}w^{(j)} \quad (1 \le i \le M), \tag{30}$$

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 $\left(m^{(i)},\theta^{(i)},\eta^{(i)}\right)$, as defined in (28) and (29). «

Proposition 1, whose proof is in Appendix A.4, characterizes seasonality as a posterior distribution over latent seasonal paths. This probabilistic representation is what we mean by *probabilistic seasonality.*

3.3 IDENTIFYING THE SEASONAL COMPONENT Proposition 1 gives the posterior distribution of the raw periodic component $S(\mathbf{t}')$ at an arbitrary collection of time points $\mathbf{t}'$. To interpret this component as an econometric seasonal factor, however, we must impose a normalization.

It is well known in the seasonal adjustment literature (Watson 1987; Harvey 1989), that the decomposition $Y(t) = S(t) + U(t) + \varepsilon(t)$ of a series into seasonal, nonseasonal, and residual components is not uniquely identified. For example, the likelihood is invariant to the transformation $S'(t) = S(t) + a$ and $U'(t) = U(t) - a$ for any constant a. Such identification issues appear in the Gaussian process model family from Section 2 as follows. Let k_{LIN} and k_{P} denote a linear and 12-month periodic covariance kernel, respectively, and consider a process $X \sim \mathrm{GP}(0, k_{\mathrm{LIN}} \times k_{\mathrm{P}})$ with a multiplicative periodic component S and a zero nonperiodic component U. This time series admits an alternative decomposition as a sum of a seasonal and nonseasonal Gaussian processes:

$$X \overset{\mathrm{D}}{=} X' := S' + U', \quad S' \sim \mathrm{GP}(0, k_{\mathrm{LIN}} \times (k_{\mathrm{P}} - v_A)), \quad U' \sim \mathrm{GP}(0, v_A k_{\mathrm{LIN}}), \quad S' \perp U', \tag{31}$$

where $v_A := \mathrm{Var}\left(\sum_{t=1}^{12} W(t)/12\right)$ for $W \sim \mathrm{GP}(0, k_{\mathrm{P}})$. It follows that the split procedure from Figure 3 returns a different decomposition when applied to X or X', even though they are distributionally equivalent on the monthly grid. Figure 4 demonstrates this effect. Note that the alternative representation X' is not contained in the prior (9)–(11), while the original representation X is, which underscores the role of the kernel language in identification.

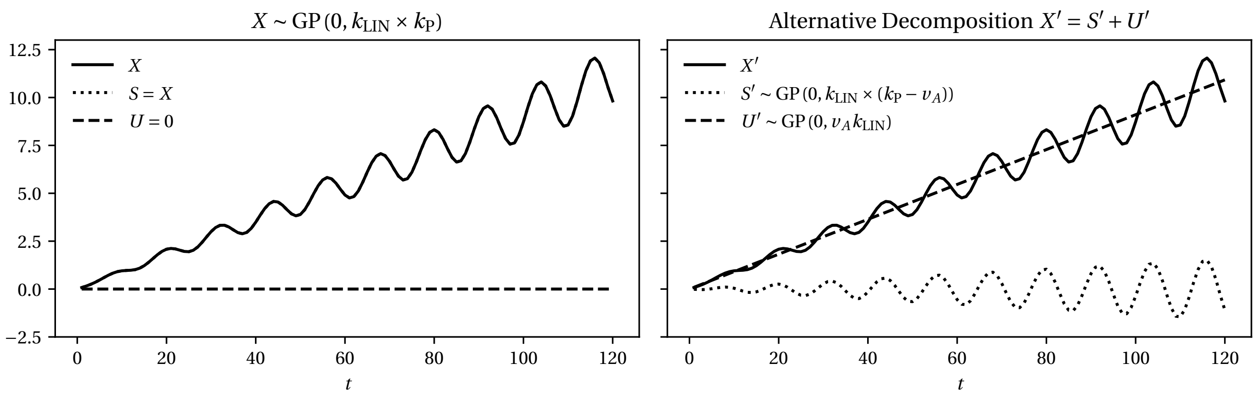

Figure 4: Seasonal components are not identified. The left panel shows a model with a multiplicative seasonal component. The right panel shows a distributionally equivalent model with a nonzero nonseasonal component.

As a unique decomposition is impossible, we follow standard econometric conventions by imposing a zero-mean normalization over a fixed reference window of time points. This step removes the additive level ambiguity by assigning the average seasonal level on that window to the nonseasonal component. The remaining allocation of persistent variation is determined by the covariance kernel representation and split rule.

Our normalization procedure operates as follows. Let $\mathbf{r} := (r_1, \dots, r_q)$ denote a finite sequence of time points on which the seasonal component is to be normalized (called the normalization window). For in-sample seasonal adjustment, $\mathbf{r} = \mathbf{t}$ is the list of time points in the observed dataset. For forecast evaluation, $\mathbf{r} = \mathbf{t}'$ is the list of time points in the forecast window. Define the $P_q := \frac{1}{q}\mathbf{1}_q\mathbf{1}_q^\top$ to be the projection operator on the constant vector and $C_q := I_q - P_q$, to be the centering operator, respectively. Thus, for any vector $v \in \mathbb{R}^q$, $C_q v$ subtracts the sample average of v over the length-q window $\mathbf{r}$. The *normalized seasonal component* on $\mathbf{r}$ is the random vector

$$\widetilde{S}_{\mathbf{r}} := C_q S(\mathbf{r}), \qquad [\widetilde{S}_{\mathbf{r}}]_j = S(r_j) - \frac{1}{q}\sum_{\ell=1}^{q} S(r_\ell) \quad (1 \le j \le q). \tag{32}$$

By construction, $\mathbf{1}_q^\top \widetilde{S}_{\mathbf{r}} = 0$, so the normalized seasonal component has zero average over the normalization window. The removed level component is reassigned to the *normalized nonseasonal component* on $\mathbf{r}$:

$$\widetilde{U}_{\mathbf{r}} := U(\mathbf{r}) + P_q S(\mathbf{r}), \qquad [\widetilde{U}_{\mathbf{r}}]_j = U(r_j) + \frac{1}{q}\sum_{\ell=1}^{q} S(r_\ell) \quad (1 \le j \le q). \tag{33}$$

These transformations preserve the latent Gaussian process X on $\mathbf{r}$, by reallocating the removed level component from the seasonal to the nonseasonal component:

$$\widetilde{S}_{\mathbf{r}} + \widetilde{U}_{\mathbf{r}} = C_q S(\mathbf{r}) + U(\mathbf{r}) + P_q S(\mathbf{r}) = S(\mathbf{r}) + U(\mathbf{r}) = X(\mathbf{r}). \tag{34}$$

Normalization only removes the constant level of the extracted seasonal component over the window, but it does not remove time variation in the periodic structure itself. For example, composite terms such as PER × GE or PER × LIN still remain part of the normalized seasonal component $\widetilde{S}_{\mathbf{r}}$, which allows periodic structure to evolve over time through changes in local smoothness or amplitude.

3.4 POSTERIOR LAW OF SEASONAL COMPONENT In our probabilistic seasonality framework, the normalized seasonal component $\widetilde{S}_{\mathbf{r}}$ is a bona fide random vector. Because the normalization step applies a linear transformation of the raw seasonal component $S(\mathbf{r})$, the posterior distribution from Proposition 1 can be pushed forward directly. Therefore, seasonal normalization, uncertainty quantification, and seasonal adjustment all live within the same inferential framework. Recall that for a fixed posterior draw $(m^{(i)}, \theta^{(i)}, \eta^{(i)})$, the raw seasonal component satisfies

$$S(\mathbf{r}) \,|\, m^{(i)}, \theta^{(i)}, \eta^{(i)}, Y(\mathbf{t}) = y(\mathbf{t}) \sim \mathcal{N}_q\left(\mu^{(i)}_{\text{per}}(\mathbf{r}), \Sigma^{(i)}_{\text{per}}(\mathbf{r},\mathbf{r})\right). \tag{35}$$

Therefore the normalized seasonal component has the Gaussian law

$$\widetilde{S}_{\mathbf{r}} \,|\, m^{(i)}, \theta^{(i)}, \eta^{(i)}, Y(\mathbf{t}) = y(\mathbf{t}) \sim \mathcal{N}_q\left(C_q\mu^{(i)}_{\text{per}}(\mathbf{r}), C_q\Sigma^{(i)}_{\text{per}}(\mathbf{r},\mathbf{r})C_q^\top\right), \tag{36}$$

which is supported on the zero-sum subspace $\{v \in \mathbb{R}^q \mid \mathbf{1}_q^\top v = 0\}$. Marginalizing over the hidden structure and parameters, the posterior distribution of $\widetilde{S}_{\mathbf{r}}$ can be approximated using the particle collection

$$\mathscr{L}\left(\widetilde{S}_{\mathbf{r}} \mid Y(\mathbf{t}) = y(\mathbf{t})\right) \approx \sum_{i=1}^{M} \widetilde{w}^{(i)} \mathcal{N}_q\left(C_q\mu^{(i)}_{\text{per}}(\mathbf{r}), C_q\Sigma^{(i)}_{\text{per}}(\mathbf{r},\mathbf{r})C_q^\top\right), \qquad \tilde{w}^{(i)} := w^{(i)} \Big/ \sum_{j=1}^{M} w^{(j)} \quad (1 \le i \le M). \tag{37}$$

The posterior mean is therefore approximated as

$$\widehat{s}_{\mathbf{r}} := \mathbb{E}\left[\widetilde{S}_{\mathbf{r}} \mid Y(\mathbf{t}) = y(\mathbf{t})\right] \approx \sum_{i=1}^{M} \widetilde{w}^{(i)} C_q \mu^{(i)}_{\text{per}}(\mathbf{r}). \tag{38}$$

When the normalization window satisfies $\mathbf{r} = \mathbf{t}$, we obtain in-sample seasonally adjusted series

$$\widehat{y}^{SA}(\mathbf{t}) := y(\mathbf{t}) - \widehat{s}_{\mathbf{t}}. \tag{39}$$

3.5 SEASONAL QUERIES The posterior law of $\widetilde{S}_{\mathbf{r}}$ given $\{Y(\mathbf{t}) = y(\mathbf{t})\}$ supports several queries that are either ill-posed or difficult to answer using standard seasonal-adjustment procedures. For compactness, we define $\widetilde{\mu}^{(i)}(\mathbf{r}) := C_q\mu^{(i)}_{\text{per}}(\mathbf{r})$ and $\widetilde{\Sigma}^{(i)}(\mathbf{r}) := C_q\Sigma^{(i)}_{\text{per}}(\mathbf{r},\mathbf{r})C_q^\top$ to be the normalized posterior mean and covariance of $\widetilde{S}_{\mathbf{r}}$, respectively, according to particle $i \in \{1, \ldots, M\}$. The following seasonality queries are solved using the inferred posterior law: no auxiliary statistical tests or per-query re-estimation procedures are required.

Existence of Seasonality. The posterior probability that seasonality is present is the posterior probability that the sampled kernel structure has a nonzero periodic component after applying the split operator from Figure 3. Let $(m_{\text{per}}, m_{\text{non}}) \coloneqq \text{split}(m)$ for a kernel structure m; the posterior probability of periodic structure is approximated as

$$\Pr\Big(m_{\text{per}} \neq 0 \mid Y(\mathbf{t}) = y(\mathbf{t})\Big) \approx \sum_{i=1}^{M} \widetilde{w}^{(i)} \mathbf{1}\Big\{m_{\text{per}}^{(i)} \neq 0\Big\}. \tag{40}$$

Strength of Seasonality. The seasonal magnitude over a normalization window $\mathbf{r}$ within a particle i is the random variable

$$A^{(i)}(\mathbf{r}) \coloneqq \frac{1}{\sqrt{q}} \left\| Z^{(i)} \right\|_2, \qquad Z^{(i)} \overset{\mathscr{L}}{=} \widetilde{S}_{\mathbf{r}} \mid m^{(i)}, \theta^{(i)}, \eta^{(i)}, Y(\mathbf{t}) = y(\mathbf{t}), \qquad (1 \leq i \leq M). \tag{41}$$

This quantity is nonnegative and in the same units as the modeled series. All summaries, such as the mean, median, or tail probabilities, of the seasonal magnitude variable are estimated from a coherent posterior distribution using the collection of M particles.

Sign and Timing. The posterior probability that the seasonal effect is positive at time r_j in the normalization window $\mathbf{r} = (r_1, \dots, r_q)$ is approximated as

$$\Pr\Big(\big[\widetilde{S}_{\mathbf{r}}\big]_j > 0 \mid Y(\mathbf{t}) = y(\mathbf{t})\Big) \approx \sum_{i=1}^{M} \mathbf{1}\left[m_{\text{per}}^{(i)} \neq 0\right] \widetilde{w}^{(i)} \Phi\left(\frac{[\widetilde{\mu}^{(i)}(\mathbf{r})]_j}{\sqrt{[\widetilde{\Sigma}^{(i)}(\mathbf{r})]_{jj}}}\right). \tag{42}$$

This probability quantifies the posterior confidence that seasonality raises the observed time series relative to the nonseasonal component at time r_j.

Model Disagreement. The posterior covariance of the normalized seasonal component decomposes into within-particle uncertainty and between-particle disagreement, by the law of total variance:

$$\text{Overall uncertainty}: \quad \text{Var}\big[\widetilde{S}_{\mathbf{r}} \mid Y(\mathbf{t}) = y(\mathbf{t})\big] \approx W_{\mathbf{r}} + B_{\mathbf{r}} \tag{43}$$

$$\text{Within particle uncertainty}: \quad W_{\mathbf{r}} \coloneqq \textstyle\sum_{i=1}^{M} \widetilde{w}^{(i)} \widetilde{\Sigma}^{(i)}(\mathbf{r}) \tag{44}$$

$$\text{Between particle uncertainty}: \quad B_{\mathbf{r}} \coloneqq \textstyle\sum_{i=1}^{M} \widetilde{w}^{(i)} \left(\widetilde{\mu}^{(i)}(\mathbf{r}) - \widehat{s}(\mathbf{r})\right)\left(\widetilde{\mu}^{(i)}(\mathbf{r}) - \widehat{s}(\mathbf{r})\right)^{\top}. \tag{45}$$

Large diagonal entries of $B_{\mathbf{r}}$ indicate time points where posterior particles disagree about the seasonal effect. The scalar $\text{tr}(B_{\mathbf{r}})/q$ summarizes the average between-particle disagreement over the window.

Revision Risk. Standard diagnostics for the revision history of seasonal adjustments operate retrospectively, by comparing an initial adjustment with new estimates obtained *after* additional observations have become available (Monsell and Findley 1984). Using Seasonal AutoGP, we instead characterize a full posterior distribution over future revisions to a current adjustment, *before* any additional data is observed.

Let $\mathbf{t} \coloneqq (t_1, \dots, t_n)$ be the current observed time points with observed data $Y(\mathbf{t}) = y(\mathbf{t})$ and $\mathbf{t}' \coloneqq (t_{n+1}, \dots, t_{n+h})$ be $h > 0$ future points. For a possible realization $y(\mathbf{t}')$ of the future data $Y(\mathbf{t}')$, define $\widehat{s}_{\mathbf{r}}(y(\mathbf{t}'))$ as the posterior mean of the normalized seasonal component $\widetilde{S}_{\mathbf{r}}$ conditioned on all $n+h$ observations $Y(\mathbf{t}, \mathbf{t}') = y(\mathbf{t}, \mathbf{t}')$.. The *revision* in the estimated seasonal adjustment at time r_j (typically $r_j = t_j$) is a deterministic function of the new data:

$$R_j(y(\mathbf{t}')) := \left[\widehat{s}_{\mathbf{r}}(y(\mathbf{t}'))\right]_j - [\widehat{s}_{\mathbf{r}}]_j \qquad (1 \le j \le q). \tag{46}$$

With probabilistic seasonality, we can quantify the full predictive distribution of the future revisions R_j conditioned on the current data $Y(\mathbf{t}) = y(\mathbf{t})$, but *before* observing the new data $Y(\mathbf{t}') = y(\mathbf{t}')$. The posterior predictive distribution of $Y(\mathbf{t}')$ is therefore

$$\mathscr{L}\left(Y\left(\mathbf{t}'\right) \mid Y(\mathbf{t}) = y(\mathbf{t})\right) \approx \sum_{i=1}^{M} \widetilde{w}^{(i)} \mathcal{N}\left(\mu^{(i)}_{\text{post}}(\mathbf{t}'), \Sigma^{(i)}_{\text{post}}(\mathbf{t}',\mathbf{t}') + \eta^{(i)} I_h\right), \tag{47}$$

where $\mu^{(i)}_{\text{post}}(\mathbf{t}')$ and $\Sigma^{(i)}_{\text{post}}(\mathbf{t}',\mathbf{t}')$ are the particle-specific posterior mean and covariance of the latent process $X(\mathbf{t}')$ conditioned on $Y(\mathbf{t}) = y(\mathbf{t})$ (Appendix A.1). This distribution induces a distribution over the future revisions:

$$\Pr\left(R_j\left(Y(\mathbf{t}')\right) \in A \mid Y(\mathbf{t}) = y(\mathbf{t})\right) = \mathbb{E}\left[\mathbf{1}\{R_j(Y(\mathbf{t}')) \in A\} \mid Y(\mathbf{t}) = y(\mathbf{t})\right] \qquad (A \subset \mathbb{R} \text{ measurable}), \tag{48}$$

which can be used to directly quantify the expected magnitude $\mathbb{E}\left[\left|R_j(Y(\mathbf{t}'))\right| \mid Y(\mathbf{t}) = y(\mathbf{t})\right]$ as well as other statistics related to future revisions.

4 A Simulation Study

We next assess the efficacy of Seasonal AutoGP through a simulation exercise. Because seasonality is a latent component with no universally-accepted definition, the "true" seasonal component of a time series is not directly observable from data. For this reason, simulations provide an environment in which the underlying seasonal structure is known by construction and methodologies can be assessed transparently (Ghysels and Perron 1993).

The simulated data are designed to capture characteristics common to seasonal economic time series. To that end, consider Figure 5, 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; and (iii) an increase in peak-to-trough dynamics over time.

The dataset in Figure 5, originally sourced from the FAA Statistical Handbook of Civil Aviation and first appearing in 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").

4.1 Methodology

4.1.1 Data Generating Process

Consider a discrete-time stochastic process given by

$$Y(t) := a_0 + a_1 \cdot t + (a_2 + a_4 \cdot t)\sin(\omega \cdot t) + (a_3 + a_5 \cdot t)\cos(\omega \cdot t) + \varepsilon(t) \tag{49}$$

where $t \in \{1,2,3,\ldots,T\}$ is a discrete time index (e.g., months); $\omega = 2\pi/12$ is the angular frequency corresponding to a 12-month (annual) seasonal cycle; and $\varepsilon(t) \sim \mathcal{N}(0,\sigma^2_\varepsilon)$ is a Gaussian error term (i.i.d.). This model captures essential features of the economic time series in Figure 5 through a linear time trend ($a_1 \cdot t$), seasonal terms ($a_2 \cdot$

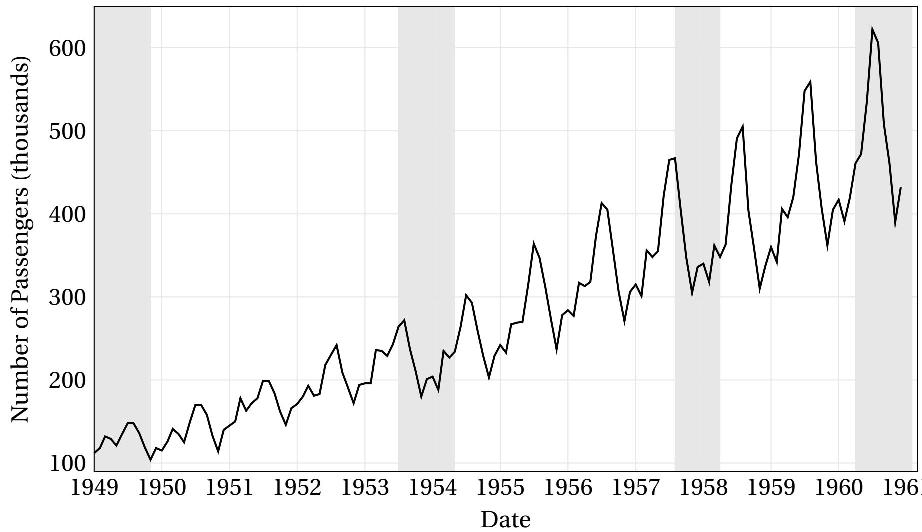

Figure 5: Passengers on International Flights 1949–1960.

$\sin(\omega t)$, $a_3 \cdot \cos(\omega t)$) 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 $a_2 a_5 = a_3 a_4$, reducing the effective number of parameters to five. The seasonal component can be expressed as a single sinusoid

$$s(t) := r(t)\sin\big(\omega \cdot t + \phi\big), \quad r(t) := \sqrt{(a_2 + a_4 \cdot t)^2 + (a_3 + a_5 \cdot t)^2}, \quad \phi := \arctan 2(a_3, a_2), \tag{50}$$

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, Chapter 4).

4.1.2 SIMULATION The model (49) is fit to the Air Passengers data of Figure 5 using ordinary least squares (OLS). The fitted parameters are $\hat{a}_0 = 87.82$, $\hat{a}_1 = 2.65$, $\hat{a}_2 = -0.24$, $\hat{a}_3 = -0.57$, $\hat{a}_4 = -0.25$, $\hat{a}_5 = -0.58$, and $\hat{\sigma}_\varepsilon = 26.90$, with an adjusted R-squared of 0.95. Note that our OLS fit does not impose the constant-phase constraint $a_2 a_5 = a_3 a_4$, but the phase change implied by the fitted parameters is negligible ($\hat{a}_2 \hat{a}_5 \approx 0.1392$ and $\hat{a}_3 \hat{a}_4 \approx 0.1425$). This fitted model is used as the true data generating process for the simulation study. For simulation length T, we generate T observations $y(\mathbf{t}) := (y(1), \dots, y(T))$ to serve as the observed data at time points $\mathbf{t} := (1, \dots, T)$, followed by 12 observations $y(\mathbf{t}') := (y(T+1), \dots, y(T+12))$ for out-of-sample forecasting at time points $\mathbf{t}' := (T+1, \dots, T+12)$. We examine sample sizes of $T \in \{36, 72, 144\}$, 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 produce $N = 200$ simulated time series, each consisting of $T + 12$ data points.

4.1.3 ESTIMATION Each simulated time-series dataset is fit using three competing approaches:

- the correctly specified model of (49) using OLS (without the constant-phase constraint $a_2 a_5 = a_3 a_4$);
- the AutoGP probabilistic seasonality approach described in Sections 2 and 3; and
- a model estimated using the X-13 ARIMA-SEATS (X-13 hereafter) algorithm.

Both model parameters *and* model structure must be learned when implementing probabilistic model discovery and X-13ARIMA-SEATS. Because OLS already uses the correctly specified model structure, standard asymptotic arguments apply to OLS estimates, which will serve as the gold-standard reference benchmark.

The most relevant comparison model is X-13, which is the current standard in seasonal adjustment developed over many decades at various statistical agencies and described in U.S. Census Bureau 2025. It is a comprehensive statistical algorithm used by the vast majority of governments to deseasonalize economic data, including the U.S. Bureau of Economic Analysis (BEA), the U.S. Bureau of Labor Statistics (BLS), Eurostat, and Statistics Canada.[2] Our analysis uses X-13ARIMA-SEATS Version 1.1 Build 61, July 10, 2024.[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[4], which enable us to compare X-13 to Seasonal AutoGP.

4.1.4 OLS GOLD STANDARD OLS serves as a gold-standard baseline because it is given the exact structural form of the simulation data-generating process (49), which is linear in the coefficients $(a_0, \dots, a_5)$. The OLS estimates exhibit two sources of uncertainty: innovation noise uncertainty and parameter uncertainty. In particular, forecasting the observed series $Y(T+h)$ requires accounting for a new innovation $\varepsilon(T+h)$, so even the correctly specified gold-standard forecast has irreducible aleatoric uncertainty. By contrast, the latent seasonal component $s(T+h)$ is deterministic, conditional on the regression coefficients, so seasonal forecast error arises only from parameter uncertainty. To make these effects explicit, write (49) as

$$Y(t) = \mathbf{z}_t^\top \beta + \varepsilon(t), \qquad \mathbf{z}_t := (1,\, t, \sin(\omega t), \cos(\omega t),\, t\sin(\omega t),\, t\cos(\omega t))^\top, \quad \beta := (a_0, a_1, a_2, a_3, a_4, a_5)^\top. \tag{51}$$

The OLS forecast error for the observed series at horizon h is

$$\widehat{Y}(T+h) - Y(T+h) = \mathbf{z}_{T+h}^\top \big(\widehat{\beta} - \beta\big) - \varepsilon(T+h). \tag{52}$$

Thus the forecast error contains both the new innovation $\varepsilon(T+h)$ and coefficient estimation error $\hat{\beta} - \beta$. The innovation does not affect the mean forecast of the observed series, but it does affect the realized forecast error and prediction intervals. In contrast, the seasonal components and OLS forecast errors are

$$s(t) = \mathbf{q}_t^\top \beta, \qquad \mathbf{q}_t := (0, 0, \sin(\omega t), \cos(\omega t),\, t\sin(\omega t),\, t\cos(\omega t))^\top, \tag{53}$$

$$\widehat{s}(T+h) - s(T+h) = \mathbf{q}_{T+h}^\top (\widehat{\beta} - \beta). \tag{54}$$

There is no irreducible innovation term in the seasonal forecast error. OLS thus provides a gold-standard reference point for assessing the cost of discovering seasonal structure (as in AutoGP and X-13) rather than being given the seasonality in advance. Performance close to OLS indicates that little accuracy is lost from structure discovery.

Our simulation study uses $N = 200$ datasets for each $T \in \{36, 72, 144\}$, which is the smallest number of simulations for which the gold-standard OLS distribution of RMSE and MIS remained stable.

4.1.5 EVALUATION For each method, we evaluate forecasts of the held-out observed data $Y(T+h)$ and a normalized seasonal component $\widetilde{s}_{\mathbf{t}'} := C_{12}\, s(\mathbf{t}')$, where the normalization centers over the forecast window $\mathbf{t}' := (T+1, \dots, T+12)$ using the centering matrix C_{12}. While this normalization does *not* alter our simulation results in any substantive manner, it is consistent with the empirical approach of the next section and serves to remove the level

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. Given that we focus on U.S. data and that both programs implement nearly identical methodologies, we do not include JDemetra+ in our analysis.

4. In the empirical and simulation exercises, our automated X-13 specification is `transformfunction=auto,regressionaictest=(td easter)`, `automdl`, `outlier`. Thus X-13 automatically chooses the transformation, tests for trading-day and Easter calendar regressors, selects the regARIMA model, detects outliers, and then applies the X-11 seasonal-adjustment decomposition.

ambiguity inherent in the trend–seasonal decomposition. For AutoGP, this normalization is applied draw-by-draw to posterior seasonal paths before posterior means and quantiles are computed.

For the data target, relative RMSE of a particular method at horizon h is given by

$$\text{Relative RMSE}_h := \frac{\text{RMSE}_h}{\text{RMSE}h^{(\text{OLS})}} - 1, \qquad \text{RMSE}_h := \left[\frac{1}{|\mathcal{I}_h|} \sum_{i \in \mathcal{I}_h} \left(\hat{y}^{(i)}(T+h) - y^{(i)}(T+h) \right)^2 \right]^{1/2} \tag{55}$$

where $\mathcal{I}_h$ is the set of replications for which the method produces a forecast at horizon h. OLS and AutoGP produce forecasts for all 200 replications at each T. X-13 fails to converge for 21 of the 600 simulated series, which gives 192, 197, and 190 replications for $T = 36, 72, 144$, respectively. The analogous seasonal RMSE is computed using the normalized target vector $\tilde{s}^{(i)}(\mathbf{t}')$ and normalized seasonal forecast $\hat{\tilde{s}}^{(i)}(\mathbf{t}')$.

We also evaluate the quality of central $(1-\alpha)$ predictive intervals with the mean interval score (MIS), which jointly evaluates the coverage and sharpness. For $\alpha = 0.05$, scalar target z, and prediction interval $[\ell, u]$, define

$$\text{MIS}_\alpha(z; \ell, u) = (u - \ell) + \frac{2}{\alpha}(\ell - z)^+ + \frac{2}{\alpha}(z - u)^+, \qquad x^+ := \max\{x, 0\}. \tag{56}$$

For the data target we have $z = Y(T+h)$, and for the seasonal target we have $z = \tilde{s}(T+h)$. X-13 produces predictive intervals for the observed data target but not usable predictive intervals for the X-11 seasonal component in this exercise; hence X-13 seasonal MIS is not reported.

Finally, we conduct paired AutoGP–X-13 tests with significance based on Benjamini–Hochberg (BH) adjusted q-values across the twelve forecast horizons within each target and sample length. For each horizon h, the unit of analysis is the matched pair of per-dataset losses $L_{i,h}^{(\text{AutoGP})}$ and $L_{i,h}^{(\text{X-13})}$ (squared error for RMSE and interval score for MIS) computed on the same simulated dataset i. We test the null hypotheses $H_{0,h}: \mathbb{E}\big[L_{i,h}^{(\text{AutoGP})} - L_{i,h}^{(\text{X-13})}\big] = 0$ with a two-sided paired t-test over the $N = 200$ replications.

4.2 RESULTS

4.2.1 DATA TARGET Figure 6 displays the relative RMSE across forecast horizons $h = 1, \ldots, 12$. 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 relative RMSE for the AutoGP estimated models with training lengths $T = 36$ (green), $T = 72$ (orange), and $T = 144$ (purple). Squares denote the relative RMSE for the X-13 model. Vertical bars connecting the X-13 and AutoGP markers highlight the magnitude and direction of the performance gap between the two models at each horizon.

Figure 6 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. At the short sample size ($T = 36$) AutoGP outperforms X-13 at all horizons ($h \geq 2$, $q < 0.01$), 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 intermediate horizons ($h \in \{3, 4, 5, 9, 10, 11\}$), while performing statistically on par with X-13, except for $h = 1$ where X-13 slightly outperforms.

Figure 7 reports the quality of the prediction intervals for the data target according to the MIS, using a similar format to the RMSE plot in Figure 6. The results shows that AutoGP produces more efficient prediction intervals. For $T = 144$, these reductions reach statistical significance in eight of the twelve horizons. For $T = 36$, AutoGP significantly improves on X-13 from $h = 4$ through $h = 10$, and at $T = 72$, the interval comparison is mixed, with one

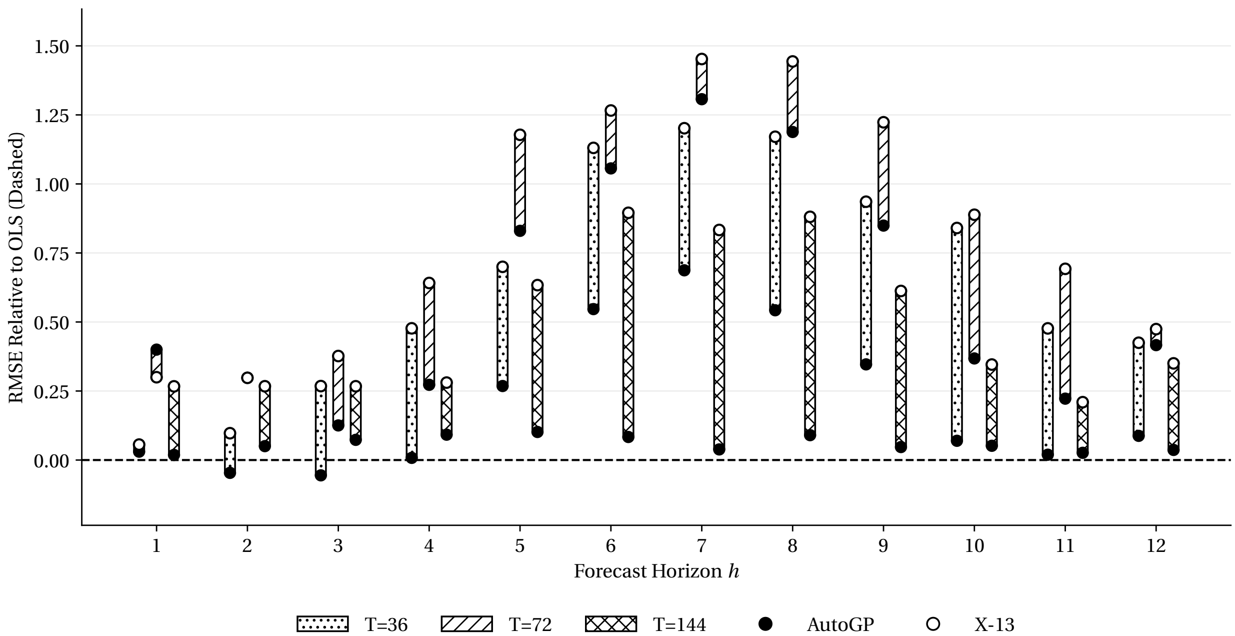

Figure 6: Relative RMSE performance for the data target.

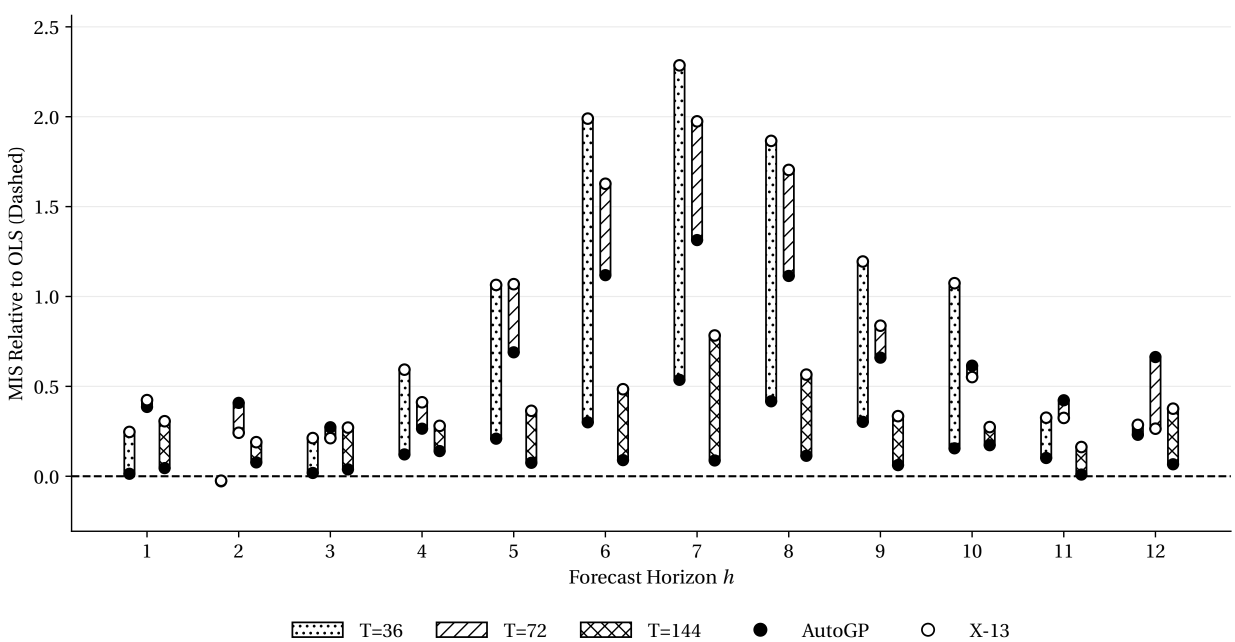

Figure 7: Relative MIS performance for the data target.

horizon favoring X-13. Taken together, these results indicate that the AutoGP model not only provides more accurate point forecasts of the observed data but also, across most conditions, delivers improved prediction intervals vis-a-vis the X-13 model.

4.2.2 Seasonal Target

We next evaluate predictions of the seasonal target. Because the seasonal amplitude $r(t)$grows over time while the innovation variance remains fixed, the seasonal signal becomes progressively stronger relative to noise as the training length increases. Thus, $T = 36$, 72, and 144 provide increasingly informative settings for recovering the seasonal component.

In Figure 8, the RMSE values for predicting the seasonal target highlight a tradeoff between flexibility and imposed seasonality. For the short sample, $T = 36$, the seasonal-target results are mixed. AutoGP significantly improves on X-13 at $h \in \{3, 4, 9, 10\}$, while X-13 significantly improves on AutoGP at six horizons. In this setting, the seasonal signal is weakest and so neither approach uniformly dominates.

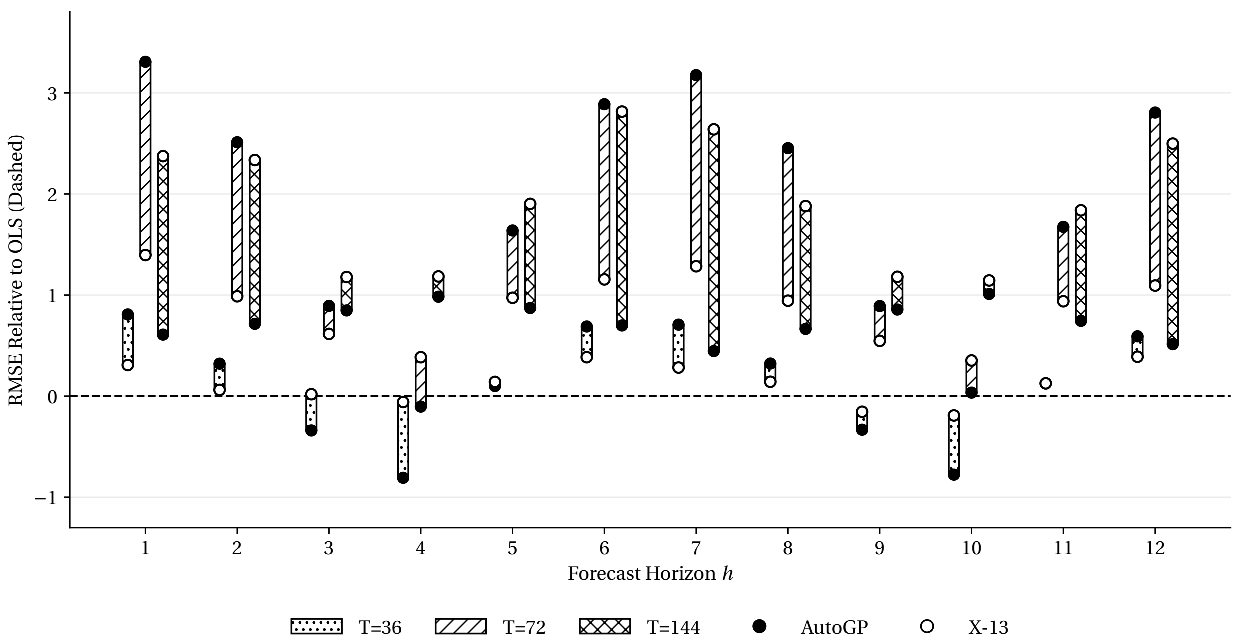

Figure 8: Relative RMSE performance for the seasonal target.

For each of the $N = 200$ datasets, we compute the posterior probability $\Pr\left(m_{\text{per}} \neq 0 \mid Y(\mathbf{t}) = y(\mathbf{t})\right)$ that seasonality is present via (40). For $T = 36$, AutoGP assigns negligible posterior mass to periodic structure, as the seasonal signal is weak relative to the noise. Conversely, as X-13 is designed to extract seasonality from data, its behavior is equivalent to a dogmatic prior heavily weighting the existence of seasonality in a probabilistic modeling sense.

At the intermediate sample size, $T = 72$, X-13 attains lower seasonal error than AutoGP at most horizons, with statistically significant reductions in ten of the twelve horizons. AutoGP significantly improves on X-13 only at $h \in \{4, 10\}$. AutoGP detects seasonality in roughly 81% of replications, but at this intermediate training length it tends to distribute the periodic variation across competing kernel structures rather than the periodic kernel. This flexibility yields a less-accurate estimate of the seasonal component, despite more accurately forecasting the data target at $T = 72$. This gap illustrates the central tradeoff of our approach. X-13 imposes a more fixed seasonal structure by construction, which enables it to identify a moderate seasonal signal relative to AutoGP.

At the long sample, $T = 144$, the ranking reverses. AutoGP significantly reduces seasonal squared error relative to X-13 in ten of the twelve horizons, with no horizon significantly favoring X-13. Once the sample is sufficiently informative, the flexibility of probabilistic model discovery outweighs the advantage of imposing seasonality ex ante, yielding statistically significant RMSE reductions.

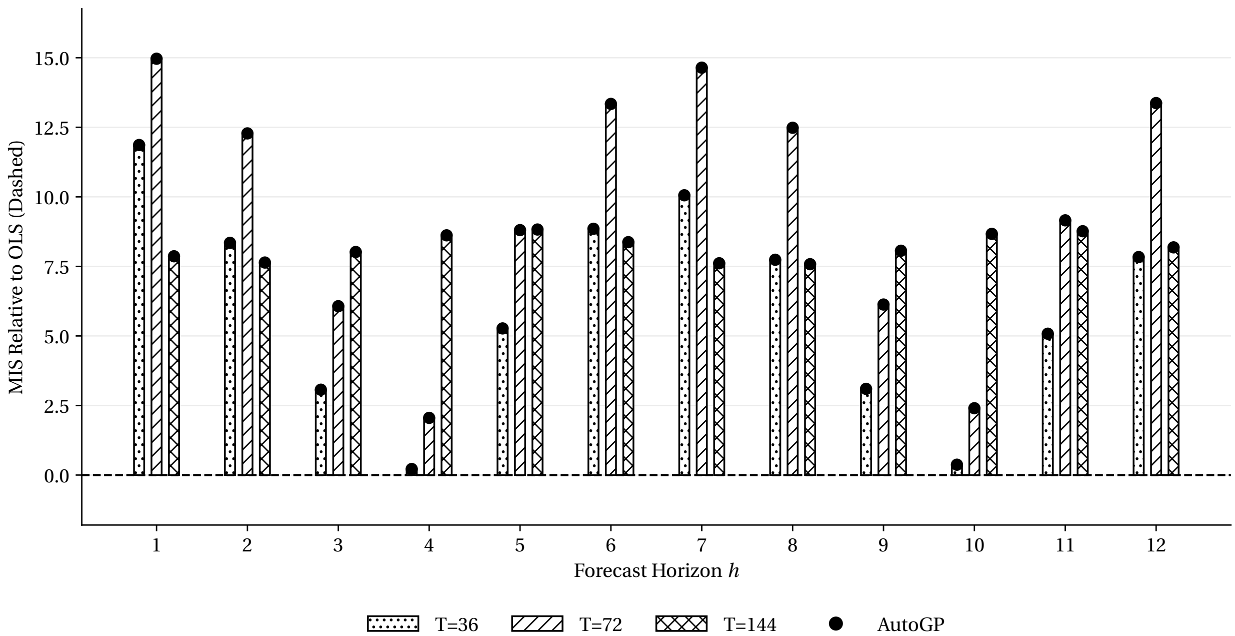

Figure 9: Relative MIS performance for the seasonal target. Note that X-13 is excluded from this comparison due to the lack of valid predictive intervals for its seasonal factors.

Figure 9 provides the interval-forecast analogue for the seasonal component using the MIS. Here the comparison is necessarily limited to AutoGP because X-13 does not provide usable predictive intervals for the seasonal factor in this exercise, and so the latter is excluded from Figure 9. Relative to the OLS gold standard, AutoGP's seasonal MIS remains substantially larger, although this result is expected given the favorable nature of OLS, which has access to the exact seasonal structure. To contrast Figure 7, where AutoGP has moderate relative MIS for the data target, with Figure 9, where AutoGP has substantially larger relative MIS for the seasonal target, it is useful to compare the OLS gold-standard forecast-error variances and corresponding intervals used in the MIS calculation. For the normalized (demeaned) seasonal target, redefine $\mathbf{q}_{T+h} \leftarrow \mathbf{q}_{T+h} - \frac{1}{12}\sum_{r=1}^{12}\mathbf{q}_{T+r}$, so that $\widetilde{s}(T+h) = \mathbf{q}_{T+h}^\top \beta$ and $\widehat{\widetilde{s}}(T+h) = \mathbf{q}_{T+h}^\top \widehat{\beta}$. We have

$$\operatorname{Var}\big[\widehat{Y}(T+h) - Y(T+h)\big] = \mathbf{z}_{T+h}^\top \operatorname{Var}\big[\widehat{\beta}-\beta\big]\,\mathbf{z}_{T+h} + \sigma_\varepsilon^2 = \sigma_\varepsilon^2\left(1 + \mathbf{z}_{T+h}^\top (Z^\top Z)^{-1}\mathbf{z}_{T+h}\right), \tag{57}$$

$$\operatorname{Var}\big[\widehat{\widetilde{s}}(T+h) - \widetilde{s}(T+h)\big] = \mathbf{q}_{T+h}^\top \operatorname{Var}\big[\widehat{\beta}-\beta\big]\,\mathbf{q}_{T+h} = \sigma_\varepsilon^2 \mathbf{q}_{T+h}^\top (Z^\top Z)^{-1}\mathbf{q}_{T+h}, \tag{58}$$

where $Z := [\mathbf{z}_1 \ldots \mathbf{z}_T]^\top$ is the OLS design matrix. The first variance is over both the in-sample innovations $\varepsilon(\mathbf{t})$, which determine $\widehat{\beta}$, and the held-out innovation $\varepsilon(T+h)$ at horizon h. The second variance is only over the in-sample innovations through $\widehat{\beta}$. The corresponding OLS 95% prediction intervals are

$$\widehat{Y}(T+h) \pm t_{T-6,0.975}\sqrt{\widehat{V}_{Y,h}}, \qquad \widehat{V}_{Y,h} := \widehat{\sigma}_\varepsilon^2\left(1 + \mathbf{z}_{T+h}^\top (Z^\top Z)^{-1}\mathbf{z}_{T+h}\right), \tag{59}$$

$$\widehat{\widetilde{s}}(T+h) \pm t_{T-6,0.975}\sqrt{\widehat{V}_{s,h}}, \qquad \widehat{V}_{s,h} := \widehat{\sigma}_\varepsilon^2 \mathbf{q}_{T+h}^\top (Z^\top Z)^{-1}\mathbf{q}_{T+h}. \tag{60}$$

Thus the data-target interval must account for the held-out innovation, whereas the seasonal-target interval reflects only parameter uncertainty, which gives a much sharper reference point. Because the relative MIS calculation normalizes by the OLS MIS separately for each target, AutoGP's relative MIS can appear much larger for the seasonal target partly because the OLS seasonal denominator is significantly smaller.

5 EMPIRICAL ANALYSIS

We now evaluate how our seasonal adjustment method performs on real-world economic data. Our analysis examines eight monthly U.S. macroeconomic series selected to span labor markets, consumer prices, industrial production, retail spending, residential construction, manufacturing demand, and business inventories, as shown in Table 1. Together, the series provide a heterogeneous panel that differs substantially in persistence, volatility, seasonal strength, and susceptibility to large economic shocks.

Table 1: Monthly macroeconomic series.

Series	NSA Series	SA Counterpart	Source	Units	Analysis Sample
Apparel CPI	`CUUR0000SAA2`	`CUSR0000SAA2`	BLS	Index (1982–84 = 100)	1990:01–2025:09
Business Inventories	`TOTBUSIMNSA`	`BUSINV`	Census	Dollars (Millions)	1992:01–2025:09
Core CPI	`CPILFENS`	`CPILFESL`	BLS	Index (1982–84 = 100)	1990:01–2025:09
Durable-Goods Orders	`UMDMNO`	`DGORDER`	Census	Dollars (Millions)	1992:02–2025:09
Housing Starts	`HOUSTNSA`	`HOUST`	Census/HUD	Units (Thousands)	1990:01–2025:09
Industrial Production	`IPB50001N`	`INDPRO`	Federal Reserve G.17	Index (2017 = 100)	1990:01–2025:09
Payroll Employment	`PAYNSA`	`PAYEMS`	BLS	Persons (Thousands)	1990:01–2025:09
Retail Sales	`RSAFSNA`	`RSAFS`	Census	Dollars (Millions)	1992:01–2025:09

Notes: The listed series are the not-seasonally-adjusted (NSA) monthly series used as model inputs. The corresponding official seasonally adjusted (SA) series are listed for comparison but are not used to estimate the forecast models.

These series include several of the most closely watched measures of U.S. economic activity. Payroll Employment and Core CPI summarize labor-market conditions and underlying consumer-price inflation. Retail Sales and Durable-Goods Orders provide information about household spending and manufacturing demand. Housing Starts and Industrial Production are important indicators of cyclical activity. The panel also spans substantial differences in seasonal behavior (see Appendix B). Seasonality is a prominent feature of Retail Sales, Housing Starts, Payroll Employment, and Durable-Goods Orders, while Core CPI and Business Inventories are smoother and more trend dominated. This variation allows us to evaluate whether the performance of Seasonal AutoGP relative to X-13 depends on the strength or shape of the seasonal component.[5]

5.1 EVALUATION SETUP

For each time series i, let $y_i^{\text{NSA}}(t)$ denote the not-seasonally-adjusted (NSA) observation in levels in month t, and define $y_i(t) := \log y_i^{\text{NSA}}(t)$. The available dates are shown in the Analysis Sample column of Table 1. For each time series i, we define the forecast origins to be the December months 2000:12, 2001:12, …, 2023:12. For every origin v, each method is estimated using data available through $t = 1, \dots, v$ and produces a twelve-month forecast $y_i(v+h)$ for $h = 1, \dots, 12$, i.e., January to December of the following year. Appendix C describes standard series-specific transformations (e.g., removing calendar-day effects and applying first differences where necessary) that ensure a direct comparison between AutoGP and X-13, and the logarithmic scale on which

5. We separately examine the relationship between seasonal adjustment and official data revisions using release-aligned ALFRED vintages. Applying X-13 to the NSA history available at each release closely reproduces official SA revisions for Industrial Production, Business Inventories, Payroll Employment, Retail Sales, and Durable-Goods Orders, but less so for Core CPI, Apparel CPI, and Housing Starts.

forecast losses are computed.[6]

5.2 Forecast Results

Table 2 reports, for each series, the AutoGP and X-13 RMSE and MIS, together with the AutoGP-to-X-13 ratio of each (ratios below one favor AutoGP). For example, Housing Starts has an AutoGP RMSE of 0.124 versus 0.157 for X-13 (a ratio of 0.789) and an AutoGP MIS of 0.778 versus 0.797 (a ratio of 0.976). A log RMSE of 0.124 for Housing Starts corresponds approximately to a relative forecast error of 12.4 percent, versus 15.7 percent for X-13. For illustration, at a level of 1.4 million starts at a seasonally adjusted annual rate, this difference corresponds to roughly 46,000 starts on the annualized scale. For Payroll Employment, an AutoGP RMSE of 0.021 on a level near 155 million corresponds to a typical error of about 3.3 million persons.

Table 2: AutoGP versus robust automated X-13 forecast distributions.

Series	AutoGP RMSE	X-13 RMSE	RMSE ratio	AutoGP MIS	X-13 MIS	MIS ratio
Apparel CPI	0.026	0.025	1.013	0.136	0.126	1.083
Business Inventories	0.029	0.031	0.929	0.182	0.201	0.907
Core CPI	0.006	0.007	0.847	0.044	0.066	0.673
Durable-Goods Orders	0.075	0.077	0.977	0.520	0.556	0.935
Housing Starts	0.124	0.157	0.789	0.778	0.797	0.976
Industrial Production	0.030	0.032	0.963	0.212	0.213	0.996
Payroll Employment	0.021	0.022	0.962	0.162	0.180	0.899
Retail Sales	0.037	0.037	1.004	0.241	0.315	0.763

Notes: The table reports RMSE and the mean interval score (MIS) for 95 percent prediction intervals, in log units, pooled over horizons $h = 1, \dots, 12$ and the 23 complete forecast origins. Ratios are AutoGP divided by X-13, so values below one favor AutoGP. Four predictions for X-13 Housing Starts are nonpositive at 2008:12 and therefore cannot be transformed into log interval, so all measurements for this incomplete forecast origin are dropped.

Table 2 shows that AutoGP has lower RMSE for six of the eight series. The exceptions are Retail Sales, where the ratio is 1.004 (essentially a tie), and Apparel CPI, where X-13 is modestly better (ratio 1.013). Similarly, AutoGP has lower MIS for seven of the eight series. The largest MIS improvements are for Core CPI (ratio 0.673, approximately 33 percent reduction) and Retail Sales (ratio 0.763), followed by Payroll Employment (0.899) and Business Inventories (0.907). Apparel CPI is the only series for which X-13 performs better on both criteria, although the differences remain moderate. Our small evaluation sample means that most individual improvements are not statistically distinguishable from zero. This effect may arise from limited statistical power rather than an absence of forecast improvement, as the loss ratios fall at or below one for nearly every series.

5.3 Residual Seasonality

We next evaluate how completely X-13 and AutoGP remove conventional fixed-frequency seasonal variation from the analyzed series. Findley et al. (2017) observe that a fundamental deficiency in seasonal adjustment is detectable seasonality after adjustment, and define various diagnostics that can be used to quantify the magnitude of residual seasonality. These diagnostics are useful for characterizing the decomposition, but they are not forecast loss functions and should not be interpreted as the primary measure of empirical performance. A procedure can, in principle, remove nearly all fixed-frequency seasonal variation without improving its prediction of future observations. Conversely, a model can leave detectable periodic variation while providing accurate and well-calibrated forecasts of the observed series.[7]

6. We use X-13ARIMA-SEATS Version 1.1 Build 62. At each origin, X-13 automatically selects the transformation and ARIMA model, applies AIC tests for trading-day and Easter regressors, and searches for additive outliers, level shifts, and temporary changes. The ARIMA residual diagnostic uses a maximum lag of 36, the forecast module produces twelve leads with 95 percent intervals, and the X-11 seasonal moving average is selected using `seasonalma=msr`.

7. Moreover, these residual-seasonality diagnostics should themselves be interpreted cautiously. For the series analyzed in log levels, the trend must also be estimated and removed before power at the seasonal frequencies can be measured. Imperfect detrending and leakage

For each series i and origin v, let $P^{\mathrm{SA}}_{i,v}$ denote the absolute spectral power of the seasonally adjusted series in frequency bands surrounding the six monthly seasonal harmonics (see Appendix D), and define $P^{\mathrm{NSA}}_{i,v}$ analogously for the not-seasonally-adjusted series. The residual seasonal power is defined as the ratio $P^{\mathrm{SA}}_{i,v}/P^{\mathrm{NSA}}_{i,v}$. We calculate this statistic across all six monthly seasonal harmonics and separately using only the annual fundamental, for both X13 and AutoGP. The all-harmonic statistic is the more comprehensive diagnostic because a monthly seasonal pattern need not be sinusoidal and can therefore place substantial power at higher seasonal harmonics. The annual-fundamental statistic computes the contribution at the 12-month fundamental only. Following Findley et al. (2017), we provide a complementary time-domain diagnostic by regressing the transformed adjusted series on the identifiable sine and cosine terms associated with the six monthly seasonal harmonics, and report the resulting R^2. Low residual power and residual harmonic R^2 indicate that the adjustment has removed seasonal-frequency variation. For Core CPI and Business Inventories, the diagnostics are computed from monthly log changes, matching the transformation used in estimation. Linearly detrended log levels are used for the remaining series.

Table 3: Residual fixed-frequency seasonal variation after seasonal adjustment.

	Residual Power Ratio (All Harmonics)		Residual Power Ratio (Annual Fundamental)		Residual Harmonic R^2 (Time Domain)	
Series	AutoGP	X-13	AutoGP	X-13	AutoGP	X-13
Apparel CPI	0.3%	0.7%	2.7%	11.3%	0.6	1.4
Business Inventories	1.4%	0.9%	6.1%	5.2%	1.7	1.5
Core CPI	3.9%	2.6%	4.3%	3.0%	3.3	3.5
Durable-Goods Orders	7.4%	4.6%	61.0%	39.0%	0.8	1.6
Housing Starts	4.5%	2.4%	1.6%	1.0%	2.2	2.2
Industrial Production	5.2%	3.8%	15.2%	9.3%	0.7	0.8
Payroll Employment	0.5%	0.3%	0.9%	0.5%	0.9	1.1
Retail Sales	1.0%	0.3%	4.0%	2.0%	1.6	1.4

Notes: Entries are percentages and are medians across the 25 December estimation origins from 2000:12 through 2024:12. For Core CPI and Business Inventories, the statistics are computed from monthly log changes, while for the remaining series, the statistics are computed from linearly detrended log levels. "All Harmonics" is the ratio of the absolute spectral power (in bands surrounding the six monthly seasonal harmonics) in the seasonally adjusted data and in the not-seasonally-adjusted data. "Annual Fundamental" is defined analogously using only the band surrounding the 12-month annual fundamental. "Residual Harmonic R^2" is the fraction of variation in the transformed adjusted series explained by the identifiable sine and cosine terms at the six monthly seasonal harmonics.

Both AutoGP and X-13 remove nearly all conventional fixed-frequency seasonal variation across the panel under both the frequency-domain and time-domain diagnostics (Table 3). The sole exceptions are Industrial Production and Durable-Goods Orders at the annual fundamental, which is not a dominant frequency for these series (Appendix D). The differences in residual power across all six harmonics are also small. Seasonal AutoGP produces lower residual power than X-13 for Apparel CPI, while X-13—which is designed to remove specific harmonics—is lower for the other seven series. The residual harmonic R^2 is consistent with these results, as the maximum difference is at most 0.8 percentage points across all of the series. These results indicate the forecast gains in Table 2 using Seasonal AutoGP coexist with largely comparable removal of conventional seasonal effects as X-13.

5.4 REVISION TIMING & RISK An advantage of Seasonal AutoGP relative to other approaches is the ability to quantify uncertainty about the seasonal adjustment in a coherent manner. Our Revision Risk query of Section 3.5 addresses an issue that is particularly problematic for X-13. The X-13 moving-average filter (dubbed X-11) estimates the seasonal pattern for a calendar month as a weighted average of nearby same-month observations. Any

from large low-frequency movements can either obscure genuine seasonal variation or generate apparent power near the seasonal harmonics. Consequently, failure to eliminate all measured seasonal-frequency variation need not imply that the predictive model has failed to learn the economically relevant seasonal structure.

extreme observation, therefore, is partially absorbed into the seasonal factors assigned to that month in adjacent years. Substantial outliers can impact seasonal adjustments for up to five years (U.S. Bureau of Labor Statistics 2026). We first document realized revisions to the X-13 seasonal adjustment over the pandemic period, demonstrating the challenges with communicating seasonality in real time using this methodology. We then show how Seasonal AutoGP offers a probabilistic assessment of such quantities.

5.4.1 X-13 SEASONAL REVISIONS Consider the fixed current-vintage history $y_i^{\mathrm{NSA}}(1),\dots,y_i^{\mathrm{NSA}}(T_i)$ of NSA levels through 2025:09 for each series i. Let c_i denote the corresponding COVID "shock month", which is May 2020 for Business Inventories and April 2020 for the remaining series. We use the rolling observations $y_i^{\mathrm{NSA}}(1),\dots,y_i^{\mathrm{NSA}}(e)$ to build an analysis run with expanding endpoints e. In particular, when moving from e to $e+1$, we hold earlier observations fixed, add one month, and re-estimate X-13, including its automatically selected transformation, regression model, outliers, and filters. X-13 identifies outliers by iteratively adding additive-outlier, level-shift, and temporary-change regressors with t-statistics exceeded a specified critical value that grows with sample size (U.S. Census Bureau 2025), and adjusts the flagged observations before the seasonal filter is applied. As these quantities change with each new observation, so does the seasonal component assigned to the shock month.

As X-13 and Seasonal AutoGP admit different seasonal decompositions, it is essential to quantify their revisions on the same scale. For the Seasonal AutoGP decomposition, we model the log NSA data $\log\big(y_i^{\mathrm{NSA}}(t)\big) = y_i(t) = s_i(t) + u_i(t) + \epsilon_i(t)$ following (19), so the extracted seasonal component $s_i(t)$ is already measured in log units. Equivalently, $s_i(t)$ from Seasonal AutoGP is the log ratio of the NSA level to the corresponding Seasonal AutoGP-adjusted level. To place the seasonal component from X-13 on the same log scale, we let $F_i^{(e)}(t)$ be the final one-centered seasonal component in a multiplicative decomposition and $C_i^{(e)}(t)$ the seasonal component in an additive decomposition (i.e., `d10`, as described in U.S. Census Bureau (2025, Section 7.19, Table 7.51)) and define the X-13 log seasonal component as

$$s_i^{(e)}(t) := \begin{cases} \log F_i^{(e)}(t) & \text{multiplicative X-13 decomposition} \\ \log\left(\dfrac{\mathrm{SA}_i^{(e)}(t) + C_i^{(e)}(t)}{\mathrm{SA}_i^{(e)}(t)}\right) & \text{additive X-13 decomposition,} \end{cases} \tag{61}$$

where $\mathrm{SA}_i^{(e)}(t)$ is the corresponding X-11 seasonally adjusted level (i.e., `d11`). Holding the shock month c_i of series i fixed, we report the absolute revision each endpoint $e = c_i + h$ in log percentage points:

$$Q_i(h) := 100\left|s_i^{(c_i+h)}(c_i) - s_i^{(c_i)}(c_i)\right|, \tag{62}$$

which for small revisions is approximately equal to the absolute percentage revision of the seasonal component. In order to contextualize these values, we also compute the historical revision magnitudes by constructing, for each pre-COVID shock month $t = 2012{:}01,\dots,2019{:}12$, the analogous $Q_{i,t}^{\mathrm{pre}}(h)$ whenever $t + h \le 2019{:}12$.

Figure 10 plots the median pre-COVID seasonal revision with the shaded gray region accounting for the 5th–95th percentile of pre-COVID revisions. The black line plots $Q_i(h)$, which measures how much X-13 has revised the seasonal component initially assigned to the shock observation after h additional releases, on the log percentage points scale. For all but Durable Goods, peak revisions occur at least 10 months after the initial shock. Peak revisions generally lie above the 95th percentile of pre-COVID revisions, with Housing Starts approximately at that percentile. For Payroll Employment, Retail Sales, and Housing Starts, the peak revision occurs more than two years after the shock. Not surprisingly, X-13 categorizes many observations as outliers during and slightly after the pandemic. For Payroll Employment, Industrial Production, and Retail Sales, X-13 initially attributes much of the

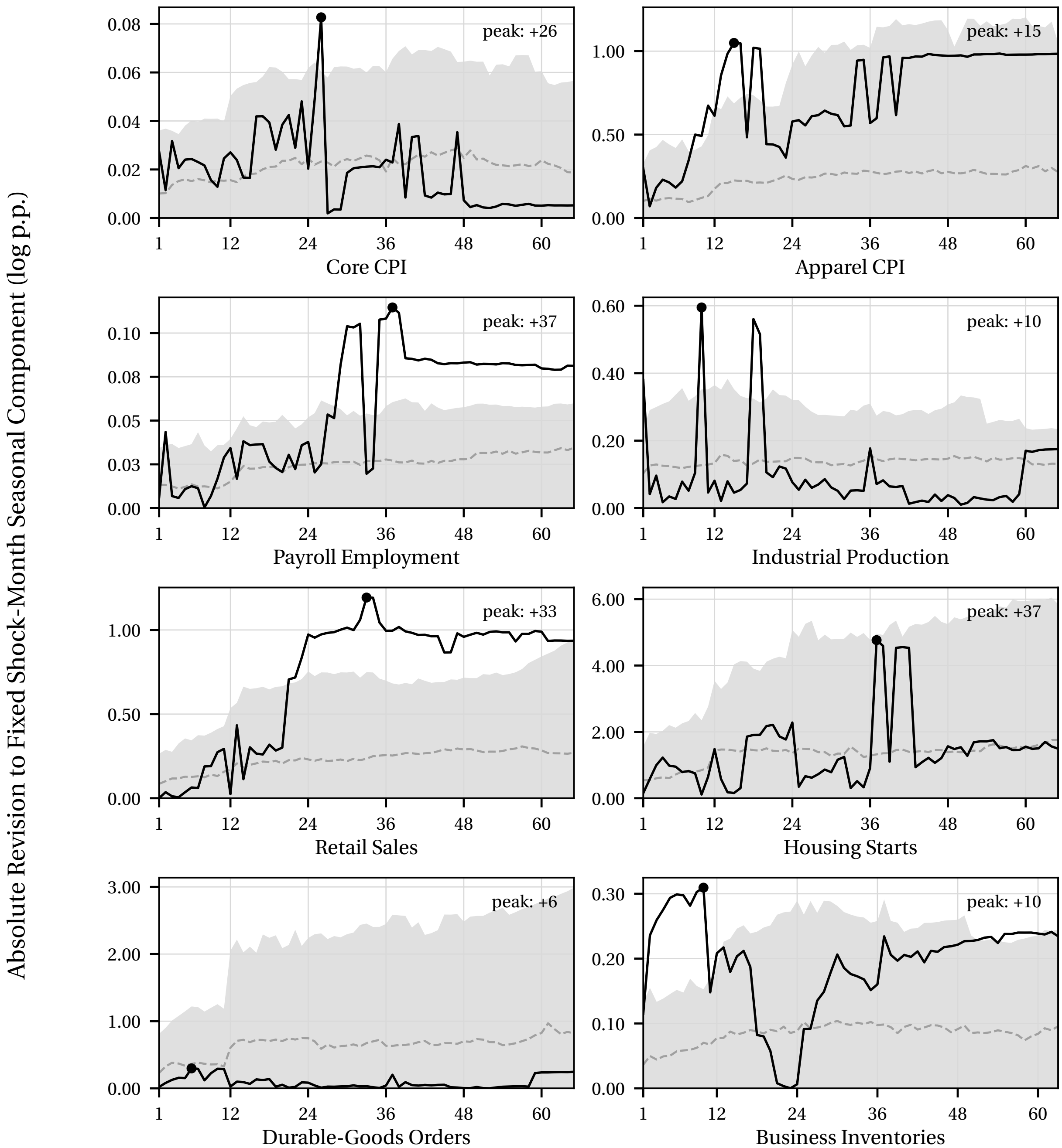

Figure 10: X-13 revision paths for the seasonal component assigned to the COVID shock month. The solid line plots the absolute revision after h additional monthly observations in log percentage points. The dashed line is the median of the corresponding pre-COVID revisions at the same horizon, and the gray region spans their 5th–95th percentiles. Peak labels report the horizon at which the COVID revision is largest.

Table 4: Seasonal AutoGP Revision Risk Index during the COVID pandemic.

Series	Shock	+3	+6	+12	Peak (origin)	Above 95th
Core CPI	**3.8**	**5.5**	**2.3**	**1.9**	**5.9** (+2)	12/13
Apparel CPI	**1.6**	**4.2**	**1.5**	1.3	**4.2** (+3)	9/13
Payroll Employment	**5.6**	**329.2**	**168.7**	**3.8**	**615.0** (+2)	13/13
Industrial Production	**1.8**	**6.3**	**8.4**	**1.6**	**20.2** (+10)	13/13
Retail Sales	1.3	**44.9**	**4.2**	**1.7**	**44.9** (+3)	12/13
Housing Starts	0.7	**55.0**	**2.1**	**2.2**	**55.0** (+3)	9/13
Durable-Goods Orders	0.9	**5.3**	**1.9**	**2.0**	**7.0** (+4)	9/13
Business Inventories	1.1	**10.0**	2.0	1.1	**13.8** (+4)	5/13

Notes: An index of one is the pre-COVID median for the same series and target age. Bold entries exceed the corresponding historical 95th percentile. Parentheses give the origin of the peak in post-shock months. The final column counts exceedances across the 13 post-COVID endpoints.

COVID shock to outliers. These fitted outlier effects are approximately 15–20 percent in absolute magnitude, while the seasonal factor assigned to the shock month differs by less than approximately 0.1 percent from that assigned to the same calendar month one year earlier. This initial allocation of observations to outliers helps explain why large X-13 revisions to the seasonal component emerge only after many additional observations arrive.

5.4.2 SEASONAL AUTOGP REVISION RISK

Rather than quantify revisions retrospectively as in X-13, the Revision Risk of Section 3.5 takes advantage of the Bayesian nature of Seasonal AutoGP to provide an estimate of how much a seasonal adjustment reported today may change in the future. At each monthly origin $e = c_i, \dots, c_i + 12$, we draw $B = 1{,}000$ possible paths $y^{(b)}_{i,e+1:e+6}$ $(b = 1, \dots, B)$ for the next six observations from the current posterior predictive distribution. For each path, we then evaluate the six-month revision defined in (46), using a fixed 24-month normalization window ending at the shock month c_i. We write $R^{(b)}_{i,e}$ for the resulting draws of the revision at the shock month associated with origin e. The random variables $R^{(1:B)}_{i,e}$ are thus i.i.d. draws from the Revision Risk distribution (48) available at origin e, before the next six observations are known.

Our first experiment addresses the following question: *In real time, how much did seasonal uncertainty increase as COVID unfolded?* For each origin e, let $W^{90}_{i,e} := 100\left(\mathrm{Quantile}_{0.95}\left(R^{1:B}_{i,e}\right) - \mathrm{Quantile}_{0.05}\left(R^{1:B}_{i,e}\right)\right)$ denote the width of its 90-percent interval. In order to normalize the magnitudes of revision risks and compare them across series, we define the post-shock *Revision Risk Index* for series i at the post-COVID origin e as

$$\mathrm{RRI}_{i,e} := \frac{W^{90}_{i,e}}{\operatorname*{median}_{u \in \{2000:12,\dots,2019:12\}} W^{90,\mathrm{pre}}_{i,u,e-c_i}} \qquad (c_i \le e \le c_i + 12). \tag{63}$$

The term $W^{90,\mathrm{pre}}_{i,u,k}$ denotes the corresponding 90-percent revision-interval width at historical December origin u for a target k months old. Thus, setting $k = e - c_i$, the denominator in (63) is the median pre-COVID revision risk for the same series and target age. The Revision Risk Index is dimensionless. A value of one gives the historical median, while values above the corresponding historical 95th percentile identify unusually high Revision Risk.

Table 4 shows that the Revision Risk rose substantially as the pandemic unfolded. At the initial shock, four of the eight series already exceeded their historical 95th percentile (bold entries). Three months later, the median Revision Risk was 8.1 and all eight series exceeded their historical 95th percentile. Revision Risk remains elevated thereafter with seven series exceeding the threshold at +6 and six still exceeding it at +12, when the median indices are 2.2 and 1.8, respectively. Overall, 82 of the 104 COVID observations lie above the corresponding pre-COVID 95th percentile, and every series does so at least once. All eight series attain their maximum after the shock origin (seven by +4, while Industrial Production peaks latest at +10). For Payroll Employment, at +2 months, its COVID interval width is 5.766 log percentage points, compared with a pre-COVID median of 0.0094, yielding an index of approximately 615. This wide ratio reflects both COVID uncertainty interval and (more importantly) the narrow pre-COVID benchmark. Table 4 demonstrates a benefit of reporting the Revision Risk statistic over this volatile post-COVID period, as it provides a coherent measure of uncertainty intrinsic to the seasonal adjustment itself.

Calibration We next assess the extent to which these higher values of Revision Risk materialized in the actual data. Our next experiment asks: *Did the realized revisions to the seasonal component assigned to the shock month appear plausible under the Revision Risk distribution available at that time?*

For each origin $e = c_i, \ldots, c_{i+6}$, we use the next six observations $y^{\mathrm{obs}}_{i,e+1:e+6}$ that actually occurred to compute the realized revision $r^{\mathrm{obs}}_{i,e}$. Using the $B = 1000$ ex-ante draws of the revision, we can test whether the realized revision lies inside the central 90-percent Revision Risk interval:

$$U_{i,e} := \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\Big\{R^{(b)}_{i,e} \le r^{\mathrm{obs}}_{i,e}\Big\} \in [0.05, 0.95] \qquad (c_i \le e \le c_i + 6). \tag{64}$$

For example, a value of $U_{i,e} = 0.50$ implies that the realized revision was near AutoGP's predictive median.

Figure 11 shows the values of $U_{i,e}$. At the shock origin, only two of the eight realized revisions fell inside the central 90-percent interval, which suggests the initial predictive distributions did not anticipate the enormous adjustments that would occur over the next six months. However, the Revision Risk diagnostic rapidly increased as the pandemic unfolded. Five realized revisions fell inside the central 90-percent interval one month later, six after two months, and seven after three months. Seven series are also covered at +4 and +5, and six at +6. Thus, while the initial observation did not generate a dispersed enough Revision Risk diagnostic, the distribution adapted as additional observations arrived. In total, 40 of the 56 rolling realized revisions lie inside their ex-ante intervals.

Comparing the seasonal revisions of X-13 and Seasonal AutoGP reveals that the two procedures arrive at similar final seasonal patterns (the median correlation between the final seasonal components is 0.972), but through different incremental revisions. For Seasonal AutoGP, the largest observed seasonal adjustment occurs between months two and five for every series; for five series (Industrial Production, Retail Sales, Housing Starts, Durable-Goods Orders, Business Inventories), X-13's largest monthly adjustment instead occurs between months 10 and 12. The timing of revisions therefore varies more than the final adjustment size. Seasonal AutoGP revisions occur as the pandemic sequence is observed, while the method reports large uncertainty about the revision in real time. In contrast, X-13 defers large revisions until much later, and does not report uncertainty with each release.

6 CONCLUSION

This paper introduces a novel probabilistic model discovery framework for seasonal adjustment of time series called Seasonal AutoGP. The method uses Gaussian process models to automatically discover seasonal and non-seasonal components through a data-driven decomposition. It is based on model hat can express flexible patterns such as nonstationary and evolving seasonality, while also delivering full posterior inferences over latent seasonal structure. This approach yields improved forecast accuracy and uncertainty estimates compared to conventional baselines, which is particularly valuable when seasonal patterns shift due to economic shocks.

In a simulation study on classic economic benchmark of airline passenger data, we find that Seasonal AutoGP produces improved point and interval forecasts compared to X-13. Its accuracy in recovering the latent seasonal component exhibits an informative tradeoff. When the periodic signal in the data is weak, using a hand-specified

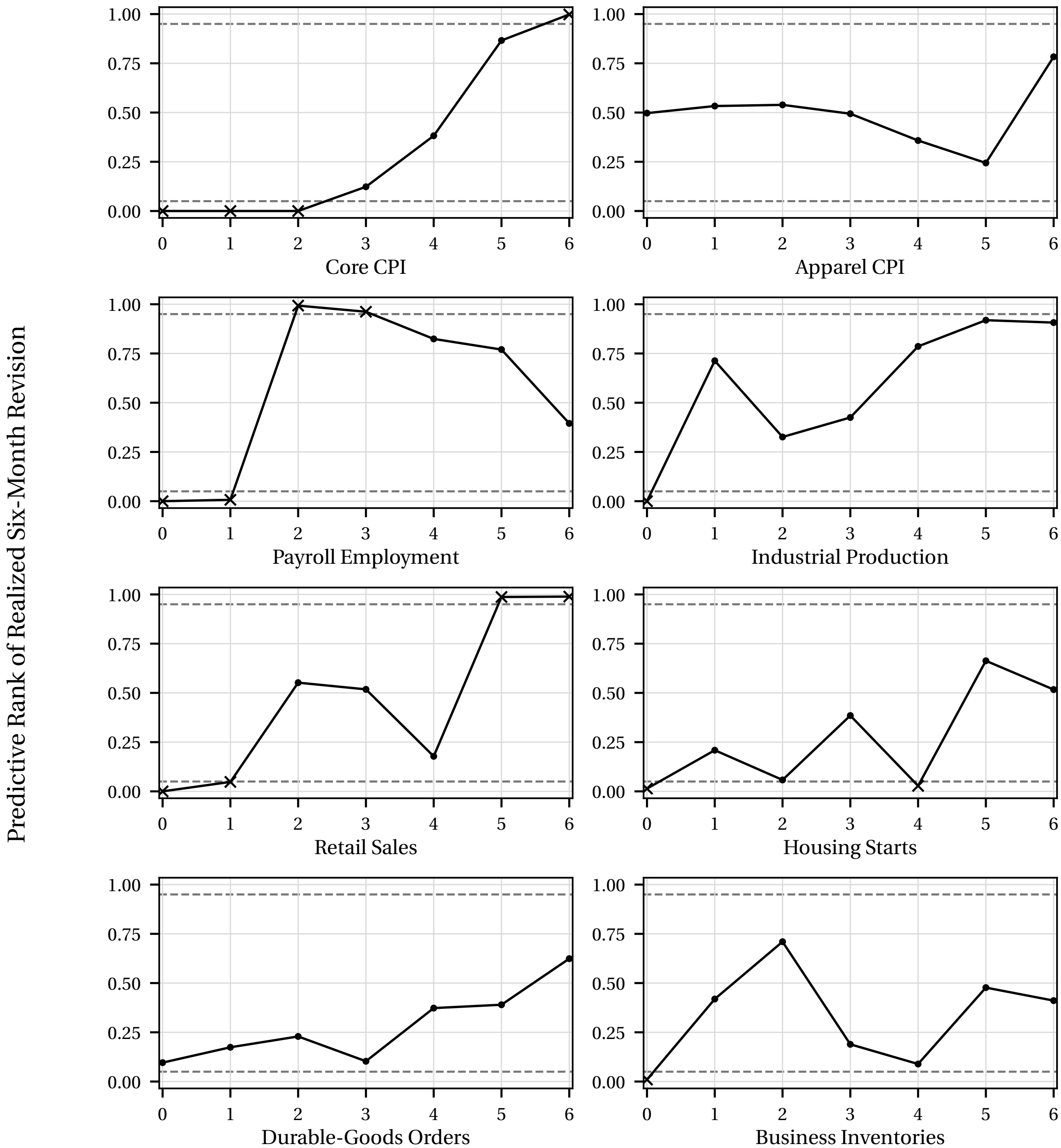

Figure 11: Rolling AutoGP Revision Risk diagnostic. At each of the seven post-COVID origins (+0 to +6 months) on the x-axes, the y-values show the predictive rank of the revision to the shock month induced by the actual next six observations within 1,000 draws from the ex-ante distribution. Crosses mark realized revisions outside the 90 percent central region (dashed lines).

imposed periodic structure as in X-13 performs better than attempting to automatically discover periodic structure using Seasonal AutoGP. However, once the samples are sufficiently informative so as to distinguish periodic structure, Seasonal AutoGP achieves significantly lower prediction errors.

Using a case study of eight U.S. macroeconomic series, we find that Seasonal AutoGP achieves lower forecast errors for six series and lower interval forecast errors for seven. Both X-13 and Seasonal AutoGP remove similar amounts of fixed-frequency seasonal variation, which shows that these forecast improvements do not occur at the expense of weaker removal of conventional seasonal effects that are common in many economic time series.

A key contribution of Seasonal AutoGP is its notion of "Revision Risk", which delivers an ex-ante distribution over the size of future revisions to a current seasonal adjustment. In a study of rolling revisions during the COVID-19 pandemic, we find that the Revision Risk already exceeds its pre-pandemic 95th percentile for four of the eight macroeconomic series, and for all eight series three months later. This experiment shows that Seasonal AutoGP exhibits significant uncertainty well before the corresponding X-13 revisions become visible, which for several series do not peak until two to three years after the shock. By treating seasonality as a probabilistic concept, the method converts the reliability of the current decomposition from a retrospective diagnostic into a quantity that can be monitored in real time, which is especially useful in high-volatility regimes.

There are limitations to the proposed framework. One limitation, inherent to all data-driven approaches to seasonal structure discovery, is non-identifiability of latent seasonal signals, even in the presence of strong periodicity. In Seasonal AutoGP, identification is aided by two modeling choices. First, the prior distribution over Gaussian process covariance kernels, and second, a normalization step that specifically assigns the average seasonal level over a fixed window to the nonseasonal component. A second limitation is that the computational cost of Gaussian process modeling scales cubically in the number of observations, which imposes limits on the length of a time series that can be efficiently handled. Instead of using a full Gaussian process, it is possible to instead leverage sparse approximations (Liu et al. 2020) with a lower computational overhead on large datasets, although these methods trade-off computational cost with modeling accuracy. A third limitation is the restriction of our framework to handling a single series at a time. By extending the method to jointly model multiple time series, it may be possible to obtain improved decompositions by discovering the cross-covariance structure between economic signals whose seasonal and nonseasonal components are highly informative of one another.

ACKNOWLEDGMENTS

F. Saad acknowledges support by the National Science Foundation under Award No. 2311983. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DECLARATION OF ARTIFICIAL INTELLIGENCE USAGE

The evaluation pipeline used for the studies in Sections 4 and 5, as well as Figures 12–15 in Appendix D, were developed with programming assistance from GPT-5.6 Codex. The underlying AutoGP.jl library is maintained without AI assistance. A draft of the manuscript was analyzed by GPT-5.6, which surfaced suggestions for improving the clarity and presentation. The authors reviewed these items, implemented a subset of those that they deemed appropriate, and take responsibility for the contents of this article.

DATA AVAILABILITY STATEMENT

A replication package for the analyses in Sections 4 and 5 is available at doi:10.5281/zenodo.22818922.

REFERENCES

Babadi, Behtash, and Emery N. Brown. 2014. "A Review of Multitaper Spectral Analysis." *IEEE Transactions on Biomedical Engineering* 61 (5): 1555–1564. https://doi.org/10.1109/TBME.2014.2311996.

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

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

Bryson, Carter, and Gary Cornwall. 2026. *The Seasons They Are A-Changin': A Century of Definitions and a Way Forward.* Working Paper WP2026-14. Washington, DC: U.S. Bureau of Economic Analysis, June. https://doi.org/10.66137/ZHRT5937.

Chopin, Nicolas, and Omiros Papaspiliopoulos. 2020. *An Introduction to Sequential Monte Carlo.* Springer Series in Statistics. Cham: Springer. https://doi.org/10.1007/978-3-030-47845-2.

Consolvo, Victoria, and Kurt G. Lunsford. 2019. *Residual Seasonality in GDP Growth Remains after Latest BEA Improvements.* Economic Commentary 2019-05. Cleveland, OH: Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-ec-201905.

Cusumano-Towner, Marco F., Feras A. Saad, Alexander K. Lew, and Vikash K. Mansinghka. 2019. "Gen: A General-Purpose Probabilistic Programming System with Programmable Inference." In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation,* 221–236. ACM. https://doi.org/10.1145/3314221.3314642.

Duvenaud, David, James Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani. 2013. "Structure Discovery in Nonparametric Regression Through Compositional Kernel Search." In *Proceedings of the 30th International Conference on Machine Learning,* 28:1166–1174. Proceedings of Machine Learning Research. PMLR.

Findley, David F., Demetra P. Lytras, and Tucker S. McElroy. 2017. *Detecting Seasonality in Seasonally Adjusted Monthly Time Series.* Research Report Series (Statistics) 2017-03. Washington, D.C.: Center for Statistical Research & Methodology, U.S. Census Bureau, February 13, 2017. https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/rrs2017-03.pdf.

Ghysels, Eric, and Pierre Perron. 1993. "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root." *Journal of Econometrics* 55 (1–2). https://doi.org/10.1016/0304-4076(93)90004-O.

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

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press. https://doi.org/10.1017/CBO9781107049994.

Liu, Haitao, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. 2020. "When Gaussian Process Meets Big Data: A Review of Scalable GPs." *IEEE Transactions on Neural Networks and Learning Systems* 31 (11): 4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109.

Liu, Jun S. 2004. *Monte Carlo Strategies in Scientific Computing.* Springer Series in Statistics. New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-76371-2.

Lunsford, Kurt G. 2017. *Lingering Residual Seasonality in GDP Growth.* Economic Commentary 2017-06. Cleveland, OH: Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-ec-201706.

———. 2025. *Residual Seasonality in Five Measures of PCE Inflation.* Economic Commentary 2025-03. Cleveland, OH: Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-ec-202503.

MacKay, David J. C. 1998. "Introduction to Gaussian Processes." In *Neural Networks and Machine Learning,* edited by Christopher M. Bishop, 168:133–166. NATO ASI Series F: Computer & Systems Sciences. Berlin: Springer.

Micchelli, Charles A., Yuesheng Xu, and Haizhang Zhang. 2006. "Universal Kernels." *Journal of Machine Learning Research* 7:2651–2667. http://jmlr.org/papers/v7/micchelli06a.html.

Monsell, Brian C., and David F. Findley. 1984. *Techniques for Determining if a Seasonal Time Series Can Be Seasonally Adjusted Reliably by a Given Seasonal Adjustment Methodology.* SRD Research Report CENSUS/SRD/RR-84/14. Washington, D.C.: U.S. Bureau of the Census, Statistical Research Division, August.

Mutikani, Lucia. 2025. "US Labor Market Cracks Widen as Job Growth Hits Stall Speed." Reuters, August 1, 2025. Accessed September 17, 2026. https://www.reuters.com/world/us/us-labor-market-cracks-widen-job-growth-hits-stall-speed-2025-08-01/.

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

Nerlove, Marc. 1964. "Spectral Analysis of Seasonal Adjustment Procedures." *Econometrica* 32 (3): 241–286. https://doi.org/10.2307/1913037.

Percival, Donald B., and Andrew T. Walden. 1993. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511622762.

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

Robert, Christian P., and George Casella. 2004. *Monte Carlo Statistical Methods.* 2nd ed. New York: Springer. https://doi.org/10.1007/978-1-4757-4145-2.

Rudebusch, Glenn D., Daniel Wilson, and Tim Mahedy. 2015. "The Puzzle of Weak First-Quarter GDP Growth." Federal Reserve Bank of San Francisco, *FRBSF Economic Letter,* nos. 2015-16 (May). https://www.frbsf.org/economic-research/publications/economic-letter/2015/may/weak-first-quarter-gdp-growth-residual-seasonality/.

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

Saad, Feras A., Marco F. Cusumano-Towner, Ulrich Schaechtle, Martin C. Rinard, and Vikash K. Mansinghka. 2019. "Bayesian Synthesis of Probabilistic Programs For Automatic Data Modeling." *Proceedings of the ACM on Programming Languages* 3, no. POPL (January): 37. https://doi.org/10.1145/3290350.

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

Thomson, David J. 1982. "Spectrum Estimation and Harmonic Analysis." *Proceedings of the IEEE* 70 (9): 1055–1096. https://doi.org/10.1109/PROC.1982.12433.

U.S. Bureau of Labor Statistics. 2025. *The Employment Situation—July 2025.* Economic News Release USDL-25-1202. U.S. Bureau of Labor Statistics, August 1, 2025. https://www.bls.gov/news.release/archives/empsit_08012025.htm.

———. 2026. *Seasonal Adjustment Methodology for National Labor Force Statistics.* Accessed September 2026. https://www.bls.gov/cps/seasonal-adjustment-methodology.htm.

U.S. Census Bureau. 2025. *X-13ARIMA-SEATS Reference Manual.* Software Documentation Version 1.1. Washington, DC: Center for Statistical Research and Methodology, U.S. Census Bureau, April. https://www2.census.gov/software/x-13arima-seats/x13as/unix-linux/documentation/docx13as.pdf.

Watson, Mark W. 1987. "Uncertainty in Model-Based Seasonal Adjustment Procedures and Construction of Minimax Filters." *Journal of the American Statistical Association* 82, no. 398 (June): 395–408. https://doi.org/10.1080/01621459.1987.10478442.

Wright, Jonathan H. 2013. "Unseasonal Seasonals?" *Brookings Papers on Economic Activity* 44 (2): 65–126. https://doi.org/10.1353/eca.2013.0017.

A GAUSSIAN PROCESS MODELS

A.1 MODEL DEFINITION We briefly overview Gaussian process models for time series data. A Gaussian process over an index set $\mathbb{T}$ is a family $\langle X(t) : t \in \mathbb{T}\rangle$ of random variables such that for any finite tuple $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{T}^n$ of $n \geq 1$ time indices, the random vector $\mathbf{X}(\mathbf{t}) := (X(t_1), \ldots, X(t_n))$ follows a multivariate Gaussian distribution:

$$\mathbf{X}(\mathbf{t}) = \begin{pmatrix} X(t_1) \\ \vdots \\ X(t_n) \end{pmatrix} \sim \mathcal{N}\left(\mu(\mathbf{t}), K(\mathbf{t}, \mathbf{t})\right), \tag{65}$$

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

$$\ln p(\mathbf{x}(\mathbf{t})) = -\tfrac{1}{2}\left[(\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t}))^\intercal K(\mathbf{t}, \mathbf{t})^{-1}(\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t}))\right] - \tfrac{1}{2}\ln\det(K(\mathbf{t}, \mathbf{t})) - \tfrac{n}{2}\ln(2\pi). \tag{66}$$

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 n' new time points $\mathbf{t}' = \left(t'_1, \ldots, t'_{n'}\right) \in \mathbb{T}^{n'}$, is itself multivariate Gaussian:

$$\mathbf{X}(\mathbf{t}') \,\big|\, \mathbf{X}(\mathbf{t}) = \mathbf{x}(\mathbf{t}) \sim \mathcal{N}\left(\mu_{\text{post}}(\mathbf{t}'), K_{\text{post}}(\mathbf{t}', \mathbf{t}')\right), \tag{67}$$

$$\mu_{\text{post}}(\mathbf{t}') := \mu(\mathbf{t}') + K(\mathbf{t}', \mathbf{t})K(\mathbf{t}, \mathbf{t})^{-1}(\mathbf{x}(\mathbf{t}) - \mu(\mathbf{t})), \tag{68}$$

$$K_{\text{post}}(\mathbf{t}', \mathbf{t}') := K(\mathbf{t}', \mathbf{t}') - K(\mathbf{t}', \mathbf{t})K(\mathbf{t}, \mathbf{t})^{-1}K(\mathbf{t}, \mathbf{t}'). \tag{69}$$

Measurement Noise. Observation noise data is modeled by using a stochastic process $\mathbf{Y}$ that is the the sum of the latent GP and i.i.d. Gaussian innovations, 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

$$\ln p(\mathbf{y}(\mathbf{t})) = -\tfrac{1}{2}\Big[(\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t}))^\intercal \left[K(\mathbf{t}, \mathbf{t}) + \eta I_n\right]^{-1}(\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t}))\Big] - \tfrac{1}{2}\ln\det\big(K(\mathbf{t}, \mathbf{t}) + \eta I_n\big) - \tfrac{n}{2}\ln(2\pi). \tag{70}$$

Conditioning on noisy observations $\mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t})$ yields the posterior distribution over the latent process:

$$\mathbf{X}(\mathbf{t}') \,\big|\, \mathbf{Y}(\mathbf{t}) = \mathbf{y}(\mathbf{t}) \sim \mathcal{N}\left(\mu_{\text{post}}(\mathbf{t}'), K_{\text{post}}(\mathbf{t}', \mathbf{t}')\right), \tag{71}$$

$$\mu_{\text{post}}(\mathbf{t}') := \mu(\mathbf{t}') + K(\mathbf{t}', \mathbf{t})[K(\mathbf{t}, \mathbf{t}) + \eta I_n]^{-1}(\mathbf{y}(\mathbf{t}) - \mu(\mathbf{t})), \tag{72}$$

$$K_{\text{post}}(\mathbf{t}', \mathbf{t}') := K(\mathbf{t}', \mathbf{t}') - K(\mathbf{t}', \mathbf{t})\left[K(\mathbf{t}, \mathbf{t}) + \eta I_n\right]^{-1}K(\mathbf{t}, \mathbf{t}'). \tag{73}$$

A.2 PRIOR MEAN It is standard in Gaussian process modeling to assume the prior mean satisfies $\mu \equiv 0$ (Rasmussen and Williams 2006, Chapter 2). This convention is largely without loss of generality for our purposes, because systematic structure can be represented using the covariance kernel rather than the mean function. For example, suppose a Gaussian process X with covariance kernel k has an affine mean function $\mu(t) = A(t - c) + B$, with unknown slope A and intercept B:

$$A \sim \mathcal{N}(0, \sigma_A^2) \tag{74}$$

$$B \sim \mathcal{N}(0, \sigma_B^2) \tag{75}$$

$$X(\cdot) \mid A, B \sim \text{GP}(A(t - c) + B, k). \tag{76}$$

Marginally, X is equivalent to a mean-zero Gaussian process with covariance kernel $k+k'$:

$$X(\cdot) \sim \mathrm{GP}\left(\mu(t)=0,\, k+k'\right) \qquad\qquad k'(t,t') := \sigma_A^2(t-c)(t'-c)+\sigma_B^2. \tag{77}$$

Here, the unknown parameters A and B are analytically marginalized out, while the hyperparameters σ_A, σ_B, and c become parameters of the covariance kernel k'. In general, affine and other structures with unknown parameters can often be represented through a composite covariance kernel, even when the prior mean is set to zero. On the other hand, deterministic components with known parameters, such as fixed offsets or calendar corrections, can either be encoded in the mean function or removed by preprocessing as in Appendix C.

As deterministic components are typically not known a priori in our model discovery setting, we set the prior mean to zero and let level, trend, periodicity, and other temporal structures be inferred using the kernel grammar and posterior inference. It should be noted that although the prior mean is zero, posterior summaries of the Gaussian process such as means, medians, and quantiles, become nonzero after conditioning on data.

A.3 COVARIANCE KERNELS A large number of Gaussian process covariance kernels are described in MacKay (1998) and Rasmussen and Williams (2006). Our model discovery framework makes use of the following subset.

Constant.

$$k_{\mathrm{C}}(t,t') := \theta_{\mathrm{C}}, \tag{78}$$

where $\theta_{\mathrm{C}} > 0$ 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.

$$k_{\mathrm{LIN}}(t,t') = \theta_{\mathrm{LIN},2}(t-\theta_{\mathrm{LIN},1})(t'-\theta_{\mathrm{LIN},1}), \tag{79}$$

where $\theta_{\mathrm{LIN},1}$ is the intercept that centers the input domain, and $\theta_{\mathrm{LIN},2}$ 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.

GammaExponential.

$$k_{\mathrm{GE}}(t,t') := \theta_{\mathrm{GE},3}\exp\left(-\left(\frac{|t-t'|}{\theta_{\mathrm{GE},1}}\right)^{\theta_{\mathrm{GE},2}}\right) \tag{80}$$

where $\theta_{\mathrm{GE},1}$ is the lengthscale, $\theta_{\mathrm{GE},2} \in (0,2]$ is the shape parameter (often denoted γ), and $\theta_{\mathrm{GE},3}$ is the amplitude. It allows for modeling functions that vary in smoothness across different time scales. Points that are close together are strongly correlated while points farther apart become less correlated. The rate at which this correlation decays depends on a parameter $\theta_{\mathrm{GE},2}$: larger values produce smoother function draws, while smaller values allow more rapidly varying functions. Setting $\theta_{\mathrm{GE},2}=2$ gives an equivalent parameterization of the SquaredExponential kernel.

Periodic.

$$k_{\mathrm{P}}(t,t') := \theta_{\mathrm{P},3}\exp\left(-2\sin^2\left(\frac{\pi}{\theta_{\mathrm{P},2}}\left|t-t'\right|\right)\Big/\theta_{\mathrm{P},1}^2\right) \tag{81}$$

where $\theta_{\mathrm{P},1}$ is the lengthscale controlling smoothness within each period, $\theta_{\mathrm{P},2}$ is the period of repetition, and $\theta_{\mathrm{P},3}$ is the amplitude. The kernel can be understood by warping the time points t to the unit circle using $\theta(t) :=$

$(\cos(2\pi/\theta_{\mathrm{P},2}t), \sin(2\pi/\theta_{\mathrm{P},2}t))$ and then applying the SquaredExponential kernel to inputs $\theta(t), \theta(t')$. Further description is given in Section 3.1.

A.4 PROOF OF PROPOSITION 1

Eqs. (24)–(25) are derived by conditioning on the data $\mathbf{Y}$ using the formula

$$\mathrm{Var}[\mathbf{Z}_1 \mid \mathbf{Z}_2] = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \tag{82}$$

where $\mathbf{Z}_1 = [\mathbf{S}, \mathbf{U}]$ and $\mathbf{Z}_2 = \mathbf{Y}$:

$$\mathrm{Var}\left[\begin{bmatrix}\mathbf{S}\\ \mathbf{U}\end{bmatrix} \middle| \mathbf{Y}\right] = \underbrace{\begin{bmatrix}K_m^{\text{per}} & \mathbf{0}\\ \mathbf{0} & K_m^{\text{non}}\end{bmatrix}}_{\mathbf{\Sigma}_{11}} - \underbrace{\begin{bmatrix}K_m^{\text{per}}\\ K_m^{\text{non}}\end{bmatrix}}_{\mathbf{\Sigma}_{12}} \underbrace{(K+\eta I)^{-1}}_{\mathbf{\Sigma}_{22}^{-1}} \underbrace{\begin{bmatrix}K_m^{\text{per}} & K_m^{\text{non}}\end{bmatrix}}_{\mathbf{\Sigma}_{21}} \tag{83}$$

$$= \begin{bmatrix}K_m^{\text{per}} & \mathbf{0}\\ \mathbf{0} & K_m^{\text{non}}\end{bmatrix} - \begin{bmatrix}K_m^{\text{per}}(K+\eta I)^{-1}K_m^{\text{per}} & K_m^{\text{per}}(K+\eta I)^{-1}K_m^{\text{non}}\\ K_m^{\text{non}}(K+\eta I)^{-1}K_m^{\text{per}} & K_m^{\text{non}}(K+\eta I)^{-1}K_m^{\text{non}}\end{bmatrix} \tag{84}$$

$$= \begin{bmatrix}K_m^{\text{per}} - K_m^{\text{per}}(K+\eta I)^{-1}K_m^{\text{per}} & -K_m^{\text{per}}(K+\eta I)^{-1}K_m^{\text{non}}\\ -K_m^{\text{non}}(K+\eta I)^{-1}K_m^{\text{per}} & K_m^{\text{non}} - K_m^{\text{non}}(K+\eta I)^{-1}K_m^{\text{non}}\end{bmatrix} \tag{85}$$

$$= \begin{bmatrix}\mathrm{Var}[\mathbf{S} \mid \mathbf{Y}] & \mathrm{Cov}(\mathbf{S}, \mathbf{U} \mid \mathbf{Y})\\ \mathrm{Cov}(\mathbf{U}, \mathbf{S} \mid \mathbf{Y}) & \mathrm{Var}[\mathbf{U} \mid \mathbf{Y}]\end{bmatrix}. \tag{86}$$

The conclusion follows.

B DATA DESCRIPTION

This appendix describes the data used for the empirical evaluation in Section 5, which are summarized in Table 1.

The empirical analysis uses eight monthly U.S. macroeconomic indicators spanning labor markets, consumer prices, industrial production, retail spending, residential construction, manufacturing demand, and business inventories. For each indicator, the not seasonally adjusted (NSA) series is the object modeled by AutoGP, while the corresponding official seasonally adjusted (SA) series provides a descriptive benchmark. Data were retrieved from Fred on July 21, 2026. The sample ends in September 2025, the latest uninterrupted month common to all 16 NSA and SA series. The sample begins in January 1990 where available. Retail Sales and Business Inventories begin in January 1992, and Durable-Goods Orders begins in February 1992.

Payroll Employment, `PAYNSA`, measures employees on total nonfarm payrolls in thousands of persons. The two price indexes provide deliberately different seasonal environments: `CPILFENS` is the broad CPI excluding food and energy, whereas `CUUR0000SAA2` is the narrower CPI for women's and girls' apparel, which has a much stronger recurring seasonal pattern. `IPB50001N` measures total industrial production, and `RSAFSNA` measures nominal advance retail and food-services sales. `HOUSTNSA` measures the monthly number of privately owned housing units started, `UMDMNO` measures manufacturers' new orders for durable goods, and `TOTBUSIMNSA` measures inventories held by manufacturers, wholesalers, and retailers. One scaling difference is important when comparing levels. The official Housing Starts series, `HOUST`, is reported at a seasonally adjusted annual rate, while `HOUSTNSA` is a monthly count in thousands of units. Consequently, the official SA level is approximately 12 times the corresponding seasonally adjusted monthly count. This constant scaling difference cancels from monthly log growth and therefore does not affect the frequency-domain or NSA-minus-SA growth comparisons below.

C DATA PREPROCESSING

Monthly economic series can be affected by variation in the number and composition of weekdays within each month. Flow variables are especially sensitive to the number of weekdays in the month. In the forecasting exercise from Section 5.2, we remove deterministic trading-day effects from the calendar-sensitive series before estimating AutoGP for four series where calendar composition has been demonstrated as relevant: Retail Sales, Industrial Production, Housing Starts, and Durable-Goods Orders. For each month t we count the occurrences of each weekday, letting $D_j(t)$ denote the number of weekday j. Taking Sunday as the omitted baseline, the trading-day regressors are $T_j(t) \coloneqq D_j(t) - D_{\text{Sun}}(t)$, for $j \in \{\text{Mon}, \text{Tue}, \text{Wed}, \text{Thu}, \text{Fri}, \text{Sat}\}$. Let $y_i(t)$ denote the log NSA values of series i at month t. We estimate the log-additive trading-day effect in first differences

$$\Delta y_i(t) \coloneqq \alpha_i + \sum_{j \in \{\text{Mon},\ldots,\text{Sat}\}} \beta_{i,j}\, \Delta T_j(t) + \sum_{l=2}^{12} \gamma_{i,l} M_l(t) + u_i(t), \tag{87}$$

where the $M_l(t)$ are month dummies. Differencing removes the trend and the month dummies absorb the seasonal mean, so the $\hat{\beta}_{i,j}$ identify the trading-day effect net of both. We then reconstruct the trading-day component in log levels and subtract it,

$$\widehat{\text{TD}}_i(t) \coloneqq \sum_{j \in \{\text{Mon},\ldots,\text{Sat}\}} \hat{\beta}_{i,j}\, T_j(t), \qquad z_i(t) \coloneqq y_i(t) - \widehat{\text{TD}}_i(t), \tag{88}$$

yielding the trading-day-scrubbed log series $z_i(t)$ on which AutoGP is estimated.

As the Core CPI and Business Inventories series are highly persistent and trend-dominated, we estimate AutoGP using monthly log differences. No preprocessing is performed for the two remaining series (Payroll Employment and Apparel CPI). The resulting data that AutoGP is estimated on is

$$z_i(t) \coloneqq \begin{cases} y_i(t) - \widehat{\text{TD}}_i(t) & i = \left\{\begin{matrix}\text{Retail Sales, Industrial Production,} \\ \text{Housing Starts, Durable-Goods Orders}\end{matrix}\right\}, \\ y_i(t) - y_i(t-1) & i \in \{\text{Core CPI, Business Inventories}\}, \\ y_i(t) & i \in \{\text{Payroll Employment, Apparel CPI}\}. \end{cases} \tag{89}$$

All forecast evaluation is performed on log-level target $y_i(t)$. For each series i, forecast origin v, and forecast horizon $h = 1, \ldots, 12$, AutoGP produces predictions $\hat{z}_i(v+h)$. The predicted log-level series $\hat{y}_i$ is reconstructed as

$$\hat{y}_i(v+h) \coloneqq \begin{cases} \hat{z}_i(v+h) + \widehat{\text{TD}}_i(v+h) & i \in \left\{\begin{matrix}\text{Retail Sales, Industrial Production,} \\ \text{Housing Starts, Durable-Goods Orders}\end{matrix}\right\}, \\ y_i(v) + \sum_{j=1}^{h} \hat{z}_i(v+j) & i \in \{\text{Core CPI, Business Inventories}\}, \\ \hat{z}_i(v+h) & i \in \{\text{Payroll Employment, Apparel CPI}\}. \end{cases} \tag{90}$$

D FREQUENCY AND TIME–DOMAIN SUMMARIES

This appendix describes the frequency analysis in Section 5.3. For series i, let $y_i^{\text{NSA}}(t)$ and $y_i^{\text{SA}}(t)$ denote the published NSA series and its SA counterpart in levels. The frequency-domain summaries use monthly log growth,

$$g_i^s(t) \coloneqq 100\left(\log y_i^s(t) - \log y_i^s(t-1)\right) \qquad\qquad (s \in \{\text{NSA}, \text{SA}\}), \tag{91}$$

which limits domination by low-frequency trend variation and sharpens concentration at seasonal frequencies.

Spectra are estimated with the DPSS multitaper method of Thomson (1982), which reduces the variance and leakage of a single-taper periodogram (Percival and Walden 1993; Babadi and Brown 2014). For a demeaned log-growth series $g^{\circ}(t) \coloneqq g(t) - \overline{g}$ of length N and DPSS taper $v^{(k)}(t)$, the plotted spectrum is the equally weighted average of the K tapered spectra,

$$\widehat{S}_{\mathrm{MT}}(f) \coloneqq \frac{1}{K}\sum_{k=0}^{K-1}\left|\sum_{t=0}^{N-1} v_t^{(k)} g^{\circ}(t)\, e^{-\mathrm{i}2\pi f t}\right|^2. \tag{92}$$

We set $NW = 2$ and $K = 2NW - 1 = 3$, which resolves the seasonal frequencies $f_j = j/12$ $(j = 1, \ldots, 6)$ while averaging three approximately uncorrelated estimates. The series is demeaned but not prewhitened, and zero-padded before the FFT for a finer plotting grid. This transformation does not increase statistical resolution.

Figures 12 and 13 show that the series exhibit heterogeneity in the strength and attenuation of seasonality. Payroll Employment, Apparel CPI, Industrial Production, Retail Sales, Durable-Goods Orders, and Business Inventories show pronounced NSA peaks at one or more seasonal frequencies, with smaller peaks in their SA counterparts. Core CPI has much less seasonal-frequency power than Apparel CPI. Housing Starts has exceptionally strong NSA seasonality, and its SA counterpart retains more seasonal-frequency power than most other SA series.

For each indicator, Figures 14 and 15 show the difference between NSA and SA monthly log growth,

$$a_i^{\mathrm{OFF}}(t) \coloneqq g_i^{\mathrm{NSA}}(t) - g_i^{\mathrm{SA}}(t), \tag{93}$$

i.e., the implied official seasonal adjustment. We avoid calling it a directly observed seasonal component because published NSA and SA series can also differ through aggregation and other agency modeling choices. The plots show that the periodic patterns vary widely across the dataset adjustments in terms of the locations, magnitudes, and regularity of the seasonal peaks.

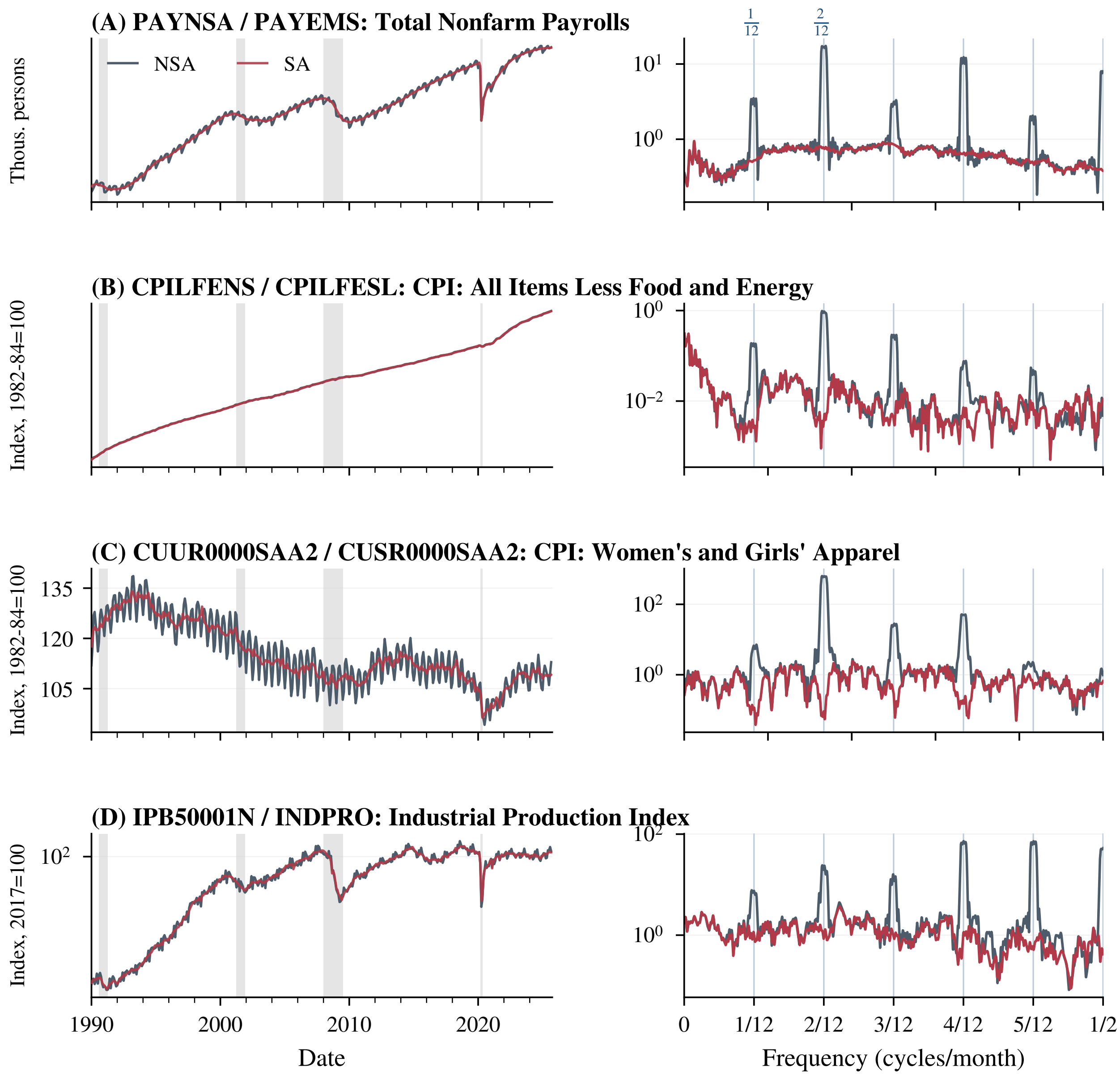

Figure 12: Levels and spectra of monthly U.S. macroeconomic series. Each row compares the NSA series (dark) with its official SA counterpart (red). The left column reports levels, using a logarithmic vertical scale where indicated; shaded regions denote NBER recessions. The right column reports the DPSS multitaper spectrum of monthly log growth, $g_t = 100\,\Delta \log y_t$, with vertical lines at $j/12$ for $j = 1, \ldots, 6$.

Notes. Spectra use $NW = 2$ and $K = 3$, with demeaning, no prewhitening, and a zero-padded FFT. The sample is monthly from 1990:01 through 2025:09. Continued in Figure 13.

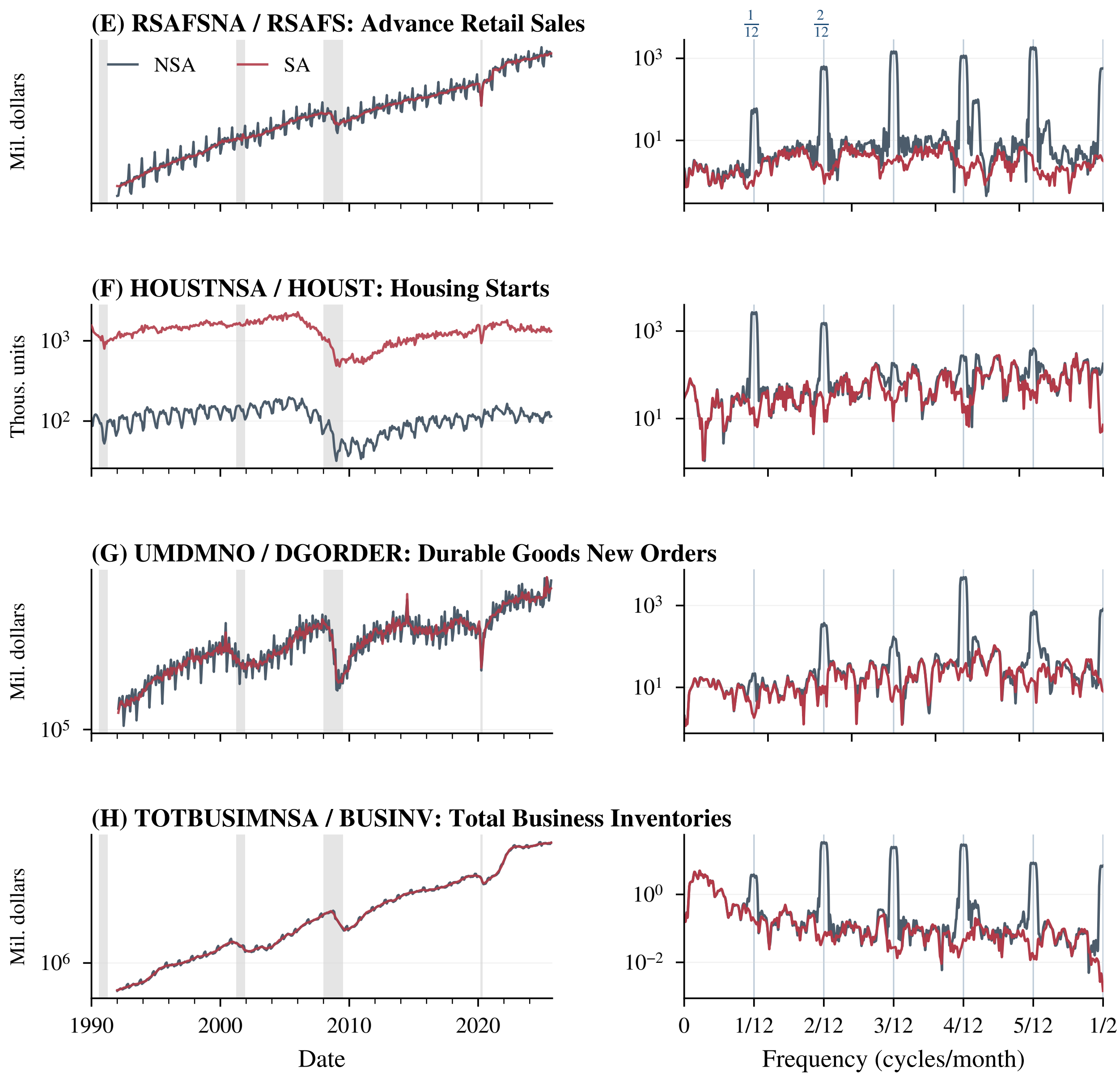

Figure 13: Levels and spectra of monthly U.S. macroeconomic series. The format follows Figure 12.

Notes. Housing Starts begins in 1990:01, Retail Sales and Business Inventories begin in 1992:01, and Durable-Goods Orders begins in 1992:02; all samples end in 2025:09. The official `HOUST` series is reported at a seasonally adjusted annual rate, so the level separation in panel (F) reflects annualization rather than residual seasonality; this constant factor cancels from the log-growth spectrum. Continued from Figure 12.

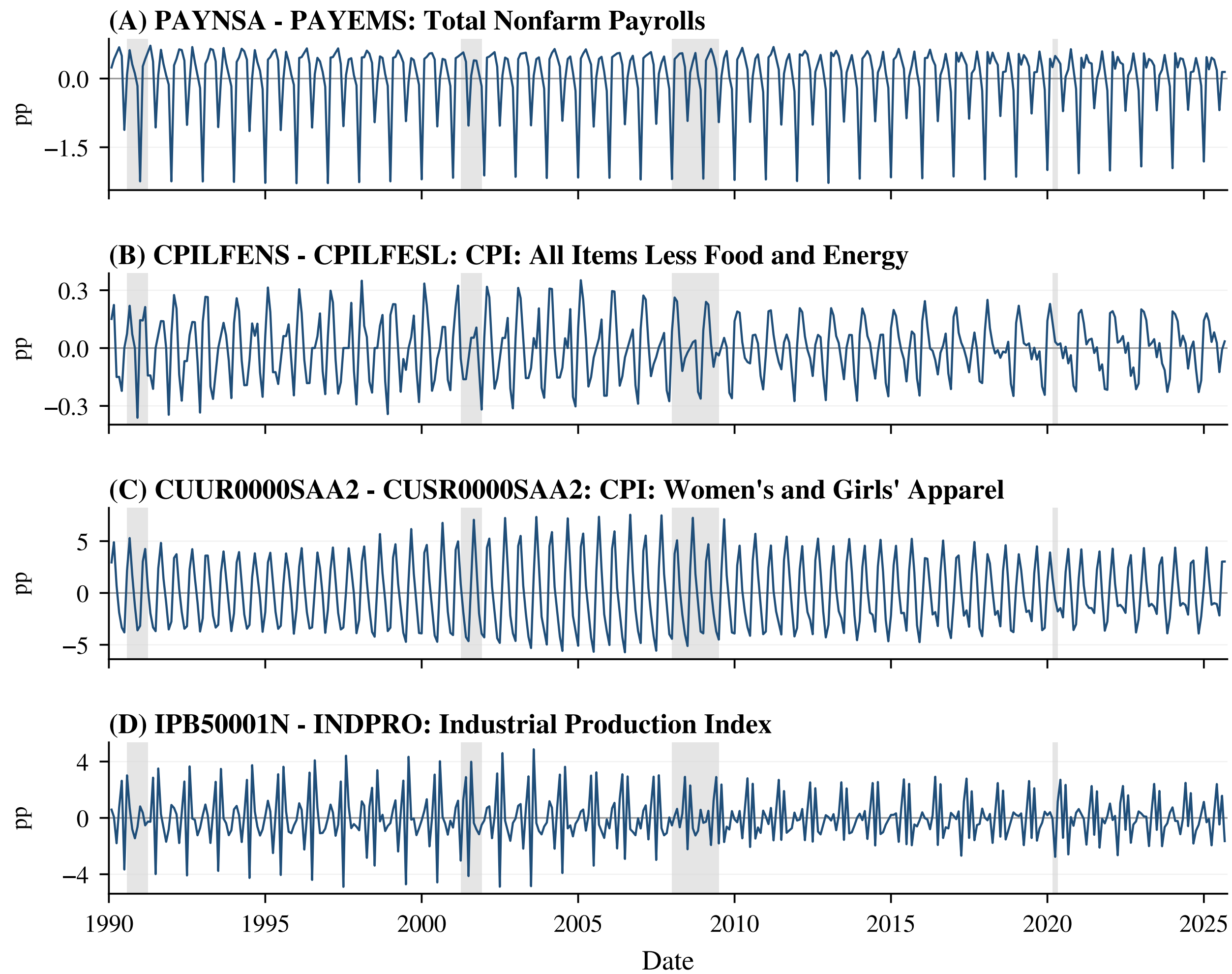

Figure 14: Implied official adjustment: NSA minus SA monthly log growth. Each panel reports $a_i^{\mathrm{OFF}}(t)$ in percentage points at a monthly rate. Shaded regions denote NBER recessions, and the horizontal line denotes no NSA–SA growth difference.

Notes. The sample is monthly from 1990:02 through 2025:09. The first month is lost when log growth is computed. Larger absolute values indicate a larger difference between the published NSA and SA monthly growth rates. Continued in Figure 15.

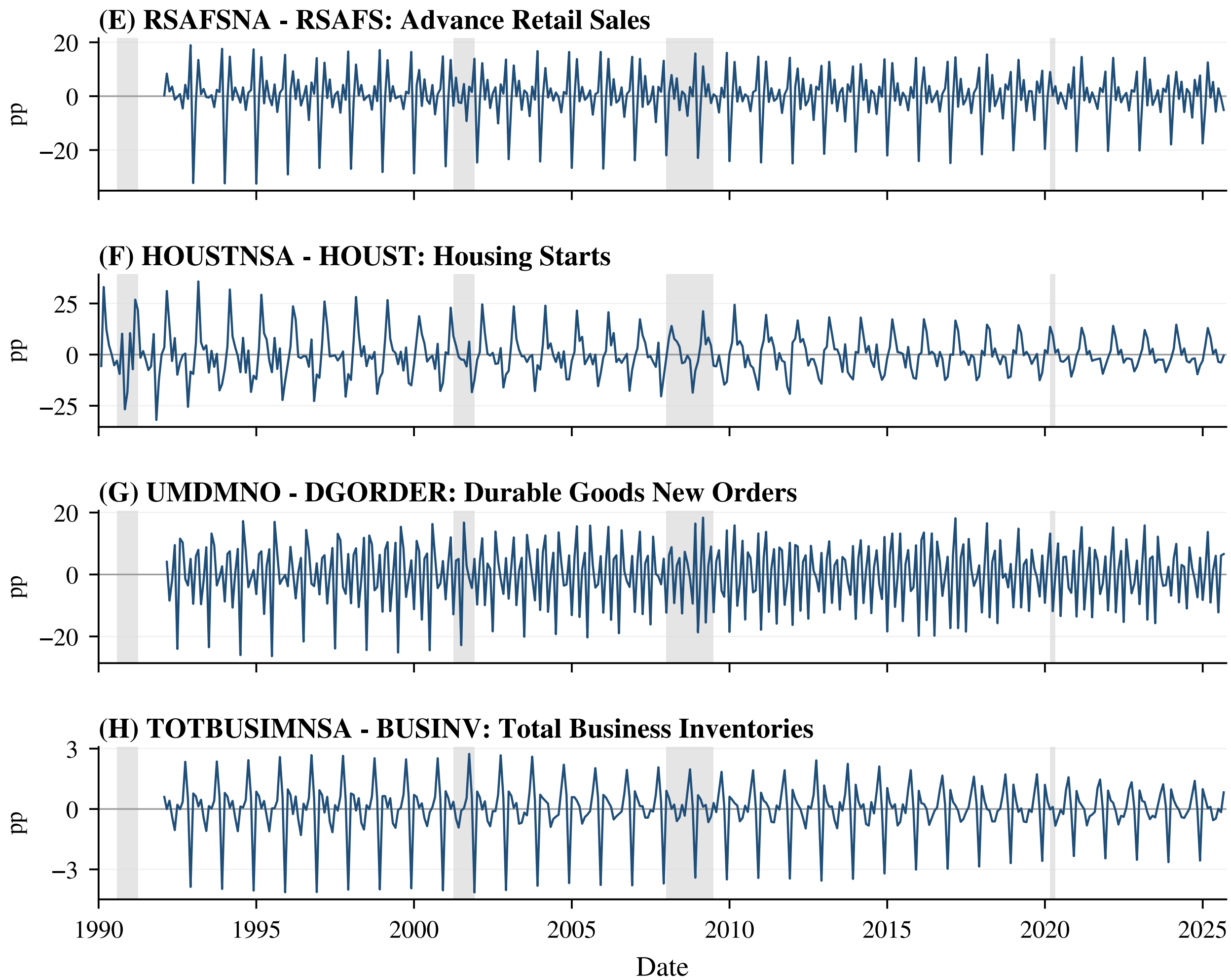

Figure 15: Implied official adjustment: NSA minus SA monthly log growth. The format follows Figure 14.

Notes. The first plotted observations are 1992:02 for Retail Sales and Business Inventories, 1990:02 for Housing Starts, and 1992:03 for Durable-Goods Orders; all samples end in 2025:09. Continued from Figure 14.