Should Macroeconomists Use Seasonally Adjusted Time Series? Structural Identification and Bayesian Estimation in Seasonal Vector Autoregressions

Ross Doppelt*

December 22, 2021

Abstract

When fitting structural vector autoregressions (VARs), macroeconomists should always prefer the original, unadjusted data over the seasonally adjusted versions. Seasonal adjustment will distort inferences about non-seasonal phenomena, such as structural parameters, impulse responses, and variance decompositions. This paper makes three contributions. First, I characterize how seasonal adjustment interferes with identification schemes in structural VARs, and how seasonal variation provides useful identifying information. Second, I provide a framework for Bayesian inference in seasonal VARs; the prior favors positive autocorrelation in season-specific means and spectral peaks at seasonal frequencies. Third, as an application, I incorporate seasonality into Baumeister and Hamilton's (2015) model of labor-market demand and supply. The model without seasonality is only partially identified, but the model with seasonality is fully identified and produces dramatically different empirical results.

Keywords: Bayesian VARs, Structural VARs, Frequency Domain, Seasonality

*Contact: ross.doppelt@psu.edu. I thank Fabio Canova and Jonathan Wright for helpful comments, and Spencer Wang for research assistance. Any errors are my own.

1 Introduction

Many economists hold two preconceptions about seasonality in time series. The first is that seasonal fluctuations are extraneous, because macroeconomists focus more on business-cycle fluctuations and the shocks that cause them. The second is that accounting for seasonality is trivial, because government statistical agencies release seasonally adjusted versions of most economic indicators. These preconceptions are in fact misconceptions. Seasonal adjustment distorts identification in structural vector autoregressions (VARs), and attempts to abstract from seasonality can discard useful identifying information. The treatment of seasonality can therefore affect inferences about non-seasonal phenomena, such as structural parameters, impulse responses, and variance decompositions. These facts make it necessary to develop time-series tools that can overcome the novel statistical challenges of seasonal data, even for researchers who are not studying seasonality itself.

Seasonality and seasonal adjustment matter for identification in structural VARs. Adjusting for seasonality involves more than just removing season-specific means. In the U.S., government statistical agencies also run the data through a two-sided linear filter, so the date-t value of a seasonally adjusted time series is a rolling moving average of past, present, and future observations. One problem with this procedure is conceptual: The identified shocks extracted from seasonally adjusted time series will not be orthogonal to the lagged values values of the unadjusted data, so these "shocks" can be predicted in advance. A separate problem is quantitative. All identification schemes for structural VARs entail factorizing the variance matrix of reduced-form VAR residuals. Using frequency-domain tools, I show that the seasonal-adjustment filter creates large distortions in the variance of the reduced-form residuals, which carry over into the structural parameters. Given infinite samples, an econometrician using seasonally adjusted time series will never agree about the structural parameters with an econometrician who applies the same identification scheme to the original, unadjusted data. If the model is only partially identified, then the two econometricians' identified sets will be disjoint. Rather than trying to strip seasonality from the data, econometricians should exploit the seasonality to identify economic shocks. The literature on identification through heteroskedasticity provides useful results for uncovering structural parameters by comparing different volatility regimes. Analogously, allowing the VAR to exhibit seasonal heteroskedasticity can help econometricians fully identify structural parameters in models that would otherwise be only partially identified.

With a finite sample, frequentist methods for seasonality have significant limitations. When using seasonally adjusted variables, many macroeconomists have found that being a Bayesian ameliorates two challenges of VAR estimation: fitting densely parameterized models and dealing with unit roots. Seasonality exacerbates both challenges. Suppose that $\mathbf{y}_t$ is a monthly time series. The simplest

form of seasonality is when $\mathbf{y}_t$ is a stationary process, plus a month-specific mean. Adding month-specific means to a VAR makes the parameter vector longer and, in a sense, makes the sample even shorter: A macroeconomist with 50 years of monthly data has 600 total observations, but it's necessary to estimate the January-specific mean with only 50 January observations, which may be highly correlated. Taking the year-over-year difference $\mathbf{y}_t - \mathbf{y}_{t-12}$ eliminates the seasonal means, but introduces a new complication: If $\mathbf{y}_t - \mathbf{y}_{t-12}$ is stationary, then $\mathbf{y}_t$ can have up to 12 unit roots at different locations on the unit circle. Overcoming these obstacles as a frequentist is daunting. In small samples, classical tests can struggle to distinguish between season-specific means and seasonal unit roots, and if there are any unit roots, then the asymptotic theory completely changes, relative to the stationary case.

Bayesian methods provide an appealing way of negotiating these statistical complications. A Bayesian can apply the same inferential procedure to stationary data and non-stationary data, which obviates the need for seasonal unit-root testing. In many ways, this approach is much simpler than the frequentist alternative. However, the cost of being a Bayesian is eliciting a sensible prior and implementing a feasible computational strategy. My prior has two main features. First, season-specific means are expected to exhibit smoothness, or positive autocorrelation, across months. That way, information about January and March is useful for updating beliefs about February. Second, the autoregressive coefficients are expected to exhibit seasonal unit roots, but the unit roots are not imposed dogmatically. A natural way to think about seasonality is in the frequency domain, and my strategy is equivalent to expecting peaks in the spectrum at seasonal frequencies. To conduct inference, I show that the posterior has a semi-conjugate form, which is amenable to a tractable posterior sampling algorithm. Although my approach to seasonality is novel, one of my goals is to design tools that are highly compatible with existing methods for Bayesian VARs. My prior, which favors spectral peaks at seasonal frequencies, is a strict generalization of a conventional Minnesota prior, which favors a spectral peak at the zero frequency. Likewise, my posterior sampler generalizes the estimation routines in Baumeister and Hamilton (2015) and Villani (2009).

To demonstrate the empirical relevance of seasonality for structural VARs, I incorporate seasonality into Baumeister and Hamilton's (2015) model of labor-market demand and supply. Baumeister and Hamilton's version of the model is only partially identified, but by exploiting seasonal heteroskedasticity, I am able to obtain full identification. Empirically, the choice between seasonally adjusted and unadjusted time series makes a large difference for estimates of the structural parameters, impulse responses, and variance decompositions.

To the best of my knowledge, this paper provides the first systematic characterization of how seasonality affects identification in structural VARs. My results provide a counterpoint to the lit-

erature on seasonality and identification in dynamic stochastic general equilibrium (DSGE) models. Rational-expectations models presume that households and firms are looking at the same data as econometricians. That assumption is plainly violated if households and firms inhabit an economy with seasonality, while econometricians use deseasonalized data. Sims (1993b) and Hansen and Sargent (1993) lay out the best-case scenario for fitting DSGE models with seasonally adjusted variables: If the source of seasonality is misspecified in the equilibrium model, then seasonally adjusted time series might, under certain circumstances, lead to more robust inferences about the model's non-seasonal properties. Ultimately, Sims and Hansen and Sargent agree that the pros of seasonal adjustment often outweigh the cons for DSGE models, but their recommendations for using seasonally adjusted variables are not universal: "Use of unadjusted data and a correctly specified model of seasonal variation is always the best option" (Sims, 1993b, p. 19). Misspecification is much less of a concern in VARs, relative to DSGE models. VARs are designed to be statistically flexible, and some researchers prefer structural VARs for the express purpose of sidestepping the much stronger assumptions required for DSGE modeling. Overall, the literature on seasonality in DSGE models has produced uneven results: The bias from seasonal adjustment can be either large or small, depending on the particulars of the equilibrium model being estimated.[1] In contrast, all identification schemes in structural VARs share certain properties, so I can derive general results about the effects of seasonality, rather than taking a case-by-case approach.

A large toolkit already exists for Bayesian estimation of VARs, but few of these tools are intended for seasonal data, despite the prevalence of seasonality in macroeconomic variables. The literature has made recent advances in judiciously incorporating prior information into both structural VARs (e.g., Baumeister and Hamilton, 2015, 2018, 2019) and reduced-form VARs (e.g., Giannone, Lenza, and Primiceri, 2015, 2019, 2021). However, it is standard practice to use the seasonally adjusted versions of time series whenever they are available, so priors for VARs rarely incorporate beliefs about seasonality. The main exceptions are Canova (1992, 1993) and Raynauld and Simonato (1993), who use reduced-form VARs as forecasting tools.[2] Although those authors provide welcome discussions of seasonality in Bayesian VARs, the specific priors that they propose come with important limitations and, in some cases, do not necessarily favor seasonal processes over non-seasonal processes. In the

[1] Sargent (1978) and Ghysels (1988) make early arguments that equilibrium models ought to be estimated with unadjusted data. However, Sims (1993b), Hansen and Sargent (1993), and Christiano and Todd (2002) provide examples of equilibrium models where the bias from using seasonally adjusted time series is small. More recently, Saijo (2013) examines the roll of seasonality in a modern New Keynesian DSGE model, and he finds that using seasonally adjusted time series can lead to sizable distortions in estimates of the structural parameters.

[2] There are, of course, many time-series models besides VARs that have been developed for seasonal data. Here, I am focusing exclusively on VARs, because they are so widely used to identify interpretable economic shocks. See Franses et al. (1997) for Bayesian counterparts to frequentist tests for seasonal unit roots in univariate time series. See Gersovitz and MacKinnon (1978) for an early Bayesian treatment of deterministic seasonality in a single-equation regression. See Ghysels et al. (2006) for a broader review of time-series methods for seasonal data, with an emphasis on forecasting.

appendix, I discuss in detail how my strategy addresses some of the limitations in these earlier approaches.

I will proceed as follows. Section 2 explains how seasonality affects identification. Section 3 develops the seasonal prior, and Section 4 explains posterior inference. Section 5 elaborates on the advantages of the Bayesian approach. Section 6 contains the application. Appendices A and B contain proofs. Appendix C provides extra details on X-11 seasonal adjustment. Appendix D contains computational details. Appendix E describes a modified version of the prior that may be appealing in other applications. Appendix F contains a detailed discussion of existing priors in the literature.

Notation. If $\mathbf{v}$ is an n-dimensional vector, then $\operatorname{diag}(\mathbf{v})$ denotes the $n \times n$ matrix with $\mathbf{v}$ on the main diagonal. The floor operator is denoted $\lfloor \cdot \rfloor$; i.e., $\lfloor x \rfloor$ is the largest integer that is weakly less than x. If a, b, and c are integers, then $a \overset{\text{mod } c}{=} b$ means that $a - b$ is an integer multiple of c. If $z \in \mathbb{C}$, then $|z|$ is understood to be the modulus of z; if $\mathbf{M}$ is a square matrix, then $|\mathbf{M}|$ is understood to be the determinant of $\mathbf{M}$. The circularly symmetric complex normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ is denoted $\mathrm{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The lag operator is denoted L. The indicator function is denoted $\mathbb{I}[\cdot]$. All other notation is standard.

2 Seasonality and Identification in Structural Models

Let $\mathbf{y}_t$ be an $n \times 1$ vector of observed time series, and let n_s be the number of seasons, or the number of calendar periods in a year. For instance, $n_s = 4$ for quarterly data, and $n_s = 12$ for monthly data. I will consider processes of the following form:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{s}_t + \tilde{\mathbf{y}}_t, \tag{1}$$

where $\boldsymbol{\mu}$ is an $n \times 1$ vector of parameters, $\mathbf{s}_t$ is a sequence that repeats deterministically every n_s time periods, and $\tilde{\mathbf{y}}_t$ is a purely non-deterministic stochastic process. There are two types of seasonality, deterministic and stochastic. Deterministic seasonality, captured by $\mathbf{s}_t$, allows for patterns like ice cream sales being higher in summer than in winter. Stochastic seasonality, incorporated in $\tilde{\mathbf{y}}_t$, allows for patterns like unusually low ice cream sales this summer being systematically correlated with next summer's ice cream sales.

The structural VAR literature studies how unforecasted innovations in $\mathbf{y}_t$ are indicative of interpretable economic shocks, and it is standard practice to identify those shocks using seasonally adjusted time series. Adjusting for deterministic seasonality entails estimating $\mathbf{s}_t$ and subtracting it

from $\mathbf{y}_t$. Doing so can be a statistical problem, but it does not affect the identification of shocks, because deterministic seasonality is inherently forecastable and shocks are inherently unforecastable. What undermines identification is "adjusting" for stochastic seasonality, which entails filtering out seasonal oscillations in $\tilde{\mathbf{y}}_t$. Section 2.1 defines the identification problem formally and elaborates on the pitfalls of trying to identify shocks using seasonally adjusted variables. Section 2.2 explains how seasonality in the conditional variance provides useful identifying information.

2.1 Seasonal Adjustment and the Identification Problem

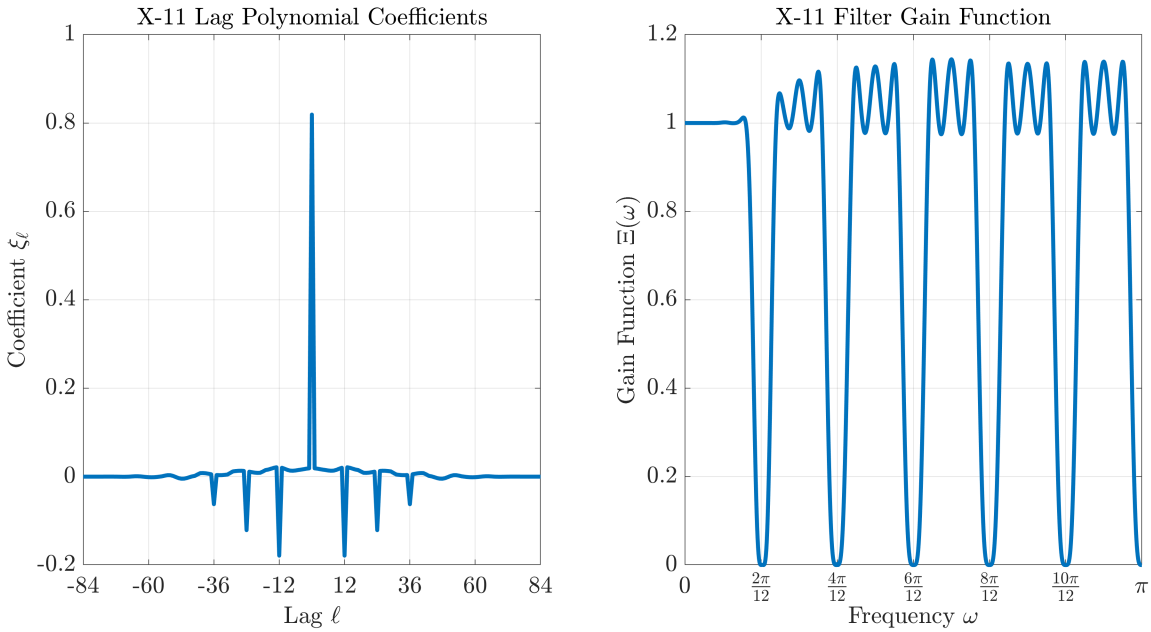
For simplicity, suppose that $\mathbf{y}_t$ is stationary and $\boldsymbol{\mu} = \mathbf{s}_t = \mathbf{0}_{n\times 1}$, so in the notation of equation (1), $\mathbf{y}_t = \tilde{\mathbf{y}}_t$. These assumptions will make make it easier to see the effects of adjusting for stochastic seasonality, but I will relax them later when going to the data. We can always represent such a time series as $\mathbf{y}_t = \hat{\mathbf{y}}_t + \mathbf{e}_t$, where $\hat{\mathbf{y}}_t$ denotes the projection of $\mathbf{y}_t$ on its history $\{\mathbf{y}_{t-\ell}\}_{\ell=1}^{\infty}$, and $\mathbf{e}_t \equiv \mathbf{y}_t - \hat{\mathbf{y}}_t$ is white noise. By construction, $\mathbf{e}_t$ is orthogonal to all lags of $\mathbf{y}_t$. Let $\mathbf{Q} \equiv \mathbb{E}\left[\mathbf{e}_t\mathbf{e}_t'\right]^{-1}$ denote the precision of the projection residuals. Whereas the central question for reduced-form modeling is how to estimate $\hat{\mathbf{y}}_t$ and $\mathbf{Q}$, the central question for structural identification is how to factorize the matrix $\mathbf{Q}$. Assume that the reduced-form residual $\mathbf{e}_t$ is an invertible linear function of a vector of structural economic shocks $\boldsymbol{\epsilon}_t$. That is, $\mathbf{e}_t = \boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon}_t$ for some invertible matrix $\boldsymbol{\Psi}$. The shocks are assumed to be white noise with precision matrix $\boldsymbol{\Lambda} \equiv \mathbb{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right]^{-1}$, and the shocks are assumed to be uncorrelated with one another, meaning $\boldsymbol{\Lambda}$ is diagonal. These conditions imply $\mathbb{V}\left[\mathbf{e}_t\right] = \mathbb{V}\left[\boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon}_t\right]$, or $\boldsymbol{\Psi}'\boldsymbol{\Lambda}\boldsymbol{\Psi} = \mathbf{Q}$. Formally, the identification problem can be stated as follows.

Definition 1. Let $\mathcal{Q}$ be the set of $n\times n$ positive-definite matrices, let $\mathcal{P}$ be the set of $n\times n$ invertible matrices, and let $\mathcal{L}$ be the set of $n\times n$ positive-definite diagonal matrices. An *identification scheme* is a mapping $\mathcal{I} : \mathcal{Q} \to \mathcal{P}\times\mathcal{L}$ with the property that if $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}) = \mathcal{I}(\mathbf{Q})$, then $\boldsymbol{\Psi}'\boldsymbol{\Lambda}\boldsymbol{\Psi} = \mathbf{Q}$. If the mapping $\mathcal{I}(\cdot)$ is a single-valued function, then the model is *fully identified*; if the mapping $\mathcal{I}(\cdot)$ is a set-valued correspondence, then the model is *partially identified*.

The above definition could be amended to allow $\mathcal{I}(\cdot)$ to depend on the reduced-form projection coefficients; I have excluded those arguments to avoid cluttering the notation. The structural VAR literature has produced a variety of identification schemes for factorizing $\mathbf{Q}$ into $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$. Regardless of the specific way this factorization is performed, standard approaches to seasonal adjustment will produce biased estimates of $\mathbf{Q}$, which will create distortions in an econometrician's beliefs about $\mathcal{I}(\mathbf{Q})$.

Government statistical agencies in the U.S. perform seasonal adjustment using variants of the Census Bureau's X-11 algorithm, which has two main components: removing deterministic terms

Figure 1: The X-11 Filter



Notes: The left panel shows the lag weights for the X-11 filter, and the right panel shows the gain function. The filter is constructed using the settings described in Appendix C.

and applying a two-sided filter. In practice, the Census Bureau's procedure is quite complicated, because it contains several user-defined settings and allows for numerous ad hoc adjustments. I will focus on the linear filter at the heart of the additive X-11 algorithm, under default settings for monthly data. Appendix C contains some more elaboration on X-11, but I refer readers to the Census Bureau's documentation for an exhaustive explanation.[3] Assuming that deterministic terms have already been removed, the seasonally adjusted series, denoted $\mathbf{y}_t^{sa}$, is constructed by applying a two-sided linear filter to the raw data $\mathbf{y}_t$:

$$\mathbf{y}_t^{sa} = \xi(L)\,\mathbf{y}_t, \quad \xi(L) \equiv \sum_{\ell=-m_\xi}^{m_\xi} \xi_\ell L^\ell, \tag{2}$$

where m_ξ is a positive integer, and each ξ_ℓ is a scalar. The numerical values for the weights $\{\xi_\ell\}_{\ell=-m_\xi}^{m_\xi}$ are plotted in the left panel of Figure 1 (and defined explicitly in Appendix C). Let $\mathbf{f}(\omega)$ denote the spectral density of $\mathbf{y}_t$ at frequency ω. The spectral density of $\mathbf{y}_t^{sa}$, denoted $\mathbf{f}^{sa}(\omega)$, is given by:

$$\mathbf{f}^{sa}(\omega) = \Xi(\omega)\,\mathbf{f}(\omega), \quad \Xi(\omega) \equiv \left|\xi\left(\exp\{-i\omega\}\right)\right|^2. \tag{3}$$

[3]See Ladiray and Quenneville (2001) for a book-length treatment. Although the Census Bureau has now developed X-12 and X-13 algorithms, they are essentially refinements of X-11, so it's common to refer to this entire family of algorithms as X-11 seasonal-adjustment procedures.

The function $\Xi(\omega)$ is the gain of the X-11 seasonal-adjustment filter, and it is plotted in the right panel of Figure 1. Because the filter is designed to suppress stochastic seasonal oscillations, the gain $\Xi(\omega)$ is equal to zero at the annual frequency $\left(\frac{2\pi}{12}\right)$ and all harmonics $\left(\frac{2\pi}{12}j,\ j=1,2,\ldots\right)$; by extension, the spectrum of the seasonally adjusted process $\mathbf{f}^{sa}(\omega)$ is equal to zero at seasonal frequencies.

Consider the problem facing an econometrician who wants to fit a structural VAR using an arbitrarily long sample of the seasonally adjusted series $\mathbf{y}_t^{sa}$. Let $\hat{\mathbf{y}}_t^{sa}$ be the projection of $\mathbf{y}_t^{sa}$ on its history $\left\{\mathbf{y}_{t-\ell}^{sa}\right\}_{\ell=1}^{\infty}$; let $\mathbf{e}_t^{sa} \equiv \mathbf{y}_t^{sa} - \hat{\mathbf{y}}_t^{sa}$ denote the projection residual; and let $\mathbf{Q}^{sa} \equiv \mathbb{E}\left[\mathbf{e}_t^{sa}\mathbf{e}_t^{sa\prime}\right]^{-1}$ denote the precision of the projection residuals. There are at least two issues with trying to apply a structural identification scheme to seasonally adjusted variables.

First, the X-11 filter creates a conceptual problem for the realized shocks that an econometrician extracts from seasonally adjusted time series. Suppose for a moment that the model is fully identified, and let $(\mathbf{\Psi}^{sa}, \mathbf{\Lambda}^{sa}) = \mathcal{I}(\mathbf{Q}^{sa})$. Define $\boldsymbol{\epsilon}_t^{sa} \equiv \mathbf{\Psi}^{sa}\mathbf{e}_t^{sa}$, which is the value for the date-t shock that the econometrician would infer by looking at the seasonally adjusted series. Because the moving-average filter in equation (2) is two-sided, $\mathbf{y}_t^{sa}$ combines past, present, and future values of $\mathbf{y}_t$. Likewise, because $\boldsymbol{\epsilon}_t^{sa}$ is in the span of $\left\{\mathbf{y}_{t-\ell}^{sa}\right\}_{\ell=0}^{\infty}$, $\boldsymbol{\epsilon}_t^{sa}$ is an amalgamation of past, present, and future values of $\boldsymbol{\epsilon}_t$. By construction, $\boldsymbol{\epsilon}_t^{sa}$ is orthogonal to lagged values of $\mathbf{y}_t^{sa}$, but $\boldsymbol{\epsilon}_t^{sa}$ is *not* orthogonal to lagged values of $\mathbf{y}_t$. Consequently, the "shocks" extracted from $\mathbf{y}_t^{sa}$ can be predicted using the history of the actual, unadjusted data. This fact makes it difficult to reconcile any rational expectations model with the results from a structural VAR when the data have been seasonally adjusted. Notice that the above criticisms would still be valid if $\mathbf{\Psi}$ were known a priori: The econometrician would infer that the realized shocks were equal to $\mathbf{\Psi}\mathbf{e}_t^{sa}$, which is also an amalgamation of past, present, and future values of $\boldsymbol{\epsilon}_t$ that can be predicted using lags of $\mathbf{y}_t$.

Second, the X-11 filter creates a quantitatively large discrepancy between $\mathbf{Q}$ and $\mathbf{Q}^{sa}$, which will create a discrepancy between $\mathcal{I}(\mathbf{Q})$ and $\mathcal{I}(\mathbf{Q}^{sa})$. For any stationary process, Kolmogorov's prediction-error formula[4] states that the precision of the projection residuals satisfies:

$$|\mathbf{Q}| = \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(|2\pi\mathbf{f}(\omega)|\right)d\omega\right\}. \tag{4}$$

The same expression applies to the seasonally adjusted process, with $\mathbf{Q}^{sa}$ and $\mathbf{f}^{sa}(\cdot)$ replacing $\mathbf{Q}$ and $\mathbf{f}(\cdot)$. Combining this fact with equation (3) implies:

$$|\mathbf{Q}^{sa}| = D^n|\mathbf{Q}|, \quad D \equiv \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\Xi(\omega)\right)d\omega\right\}. \tag{5}$$

[4]See, e.g., Theorem 3‴ in Chapter 3 of Hannan (1970, p. 162).

The term D represents the distortion that the filter creates in the precision of the projection residuals. Numerically, the distortion is large: $D \approx 2.83$ for the X-11 filter summarized in Figure 1.[5] To put that magnitude in perspective, in a univariate setting, the variance of the residuals is 2.83 times larger when the projection is performed with the unadjusted data, compared to the seasonally adjusted version. Naïvely, one might think that seasonal adjustment has little effect when the underlying data do not exhibit much seasonality, but that isn't the case: D only depends on the filter, not the data, so seasonal adjustment will create large distortions, regardless of whether the original time series is strongly seasonal or completely aseasonal. Ultimately, by biasing $\mathbf{Q}$, seasonal adjustment affects the way the identification scheme factorizes $\mathbf{Q}$ into structural parameters.

Proposition 1. *When the model is fully identified, $\mathcal{I}(\mathbf{Q}) \neq \mathcal{I}(\mathbf{Q}^{sa})$, and when the model is partially identified, $\mathcal{I}(\mathbf{Q}) \cap \mathcal{I}(\mathbf{Q}^{sa}) = \emptyset$.*

In other words, an econometrician who uses seasonally adjusted variables will never agree about the structural parameters with an econometrician who uses the original, unadjusted data. When the model is partially identified, an econometrician looking at $\mathbf{y}_t$ can believe that infinitely many configurations of structural parameters are compatible with $\mathbf{Q}$, and an econometrician looking at $\mathbf{y}_t^{sa}$ can believe that infinitely many configurations of structural parameters are compatible with $\mathbf{Q}^{sa}$ — yet there will be no overlap between the two identified sets. The issue highlighted by Proposition 1 is distinct from the conceptual issue mentioned earlier, that the two-sided filter $\xi(L)$ combines past, present, and future values of $\mathbf{y}_t$ to synthesize $\mathbf{y}_t^{sa}$. The only thing that matters for the proof of Proposition 1 is that $D \neq 1$. Consequently, even if one replaces the two-sided lag polynomial depicted in Figure 1 with a one-sided lag polynomial, seasonal adjustment can still distort the identification scheme.

In some applications, the main objects of interest are the impulse responses, rather than the structural parameters themselves. Suppose the model is fully identified, and let $(\mathbf{\Psi}, \mathbf{\Lambda}) = \mathcal{I}(\mathbf{Q})$. The matrix of normalized contemporaneous impulse responses is given by $\frac{\partial \mathbf{y}_t}{\partial \left(\mathbf{\Lambda}^{-1/2}\boldsymbol{\epsilon}_t\right)'} = \mathbf{\Psi}^{-1}\mathbf{\Lambda}^{-1/2}$; that is, the (j,k) element of $\mathbf{\Psi}^{-1}\mathbf{\Lambda}^{-1/2}$ is the response of $\mathbf{y}_{j,t}$ to a one-standard-deviation perturbation in $\boldsymbol{\epsilon}_{k,t}$. An econometrician using seasonally adjusted variables would infer that the structural parameters are $(\mathbf{\Psi}^{sa}, \mathbf{\Lambda}^{sa}) = \mathcal{I}(\mathbf{Q}^{sa})$, the identified shocks are $\boldsymbol{\epsilon}_t^{sa} = \mathbf{\Psi}^{sa}\mathbf{e}_t^{sa}$, and the normalized impulse responses are $\frac{\partial \mathbf{y}_t^{sa}}{\partial \left((\mathbf{\Lambda}^{sa})^{-1/2}\boldsymbol{\epsilon}_t^{sa}\right)'} = (\mathbf{\Psi}^{sa})^{-1}(\mathbf{\Lambda}^{sa})^{-1/2}$. Because $\mathbf{\Psi}'\mathbf{\Lambda}\mathbf{\Psi} = \mathbf{Q}$ and $\mathbf{\Psi}^{sa\prime}\mathbf{\Lambda}^{sa}\mathbf{\Psi}^{sa} = \mathbf{Q}^{sa}$,

[5] As an aside, equation (5) applies to any linear filter, with the specification of $\xi(L)$ determining the value of D. For the Hodrick-Prescott filter, $D \approx 2.45$ when the smoothing parameter is $1,600$. For a high-pass, low-pass, or band-pass filter, $D = \infty$, because $\Xi(\omega) = 0$ for a positive-measure interval of frequencies. In contrast, for the X-11 filter, $\Xi(\omega)$ is only equal to zero for a measure-zero set of points in the interval $[-\pi, \pi]$, so one can show that D is finite.

equation (5) implies:

$$\left|\frac{\partial \mathbf{y}_t^{sa}}{\partial\left(\left(\mathbf{\Lambda}^{sa}\right)^{-1/2}\boldsymbol{\epsilon}_t^{sa}\right)'}\right| = \left|D^{-\frac{1}{2}}\cdot\frac{\partial \mathbf{y}_t}{\partial\left(\mathbf{\Lambda}^{-1/2}\boldsymbol{\epsilon}_t\right)'}\right|. \tag{6}$$

Evidently, seasonal adjustment creates as much bias in the determinant of the impulse response matrix as multiplying the true impulse responses by $D^{-1/2} \approx .59$. Of course, as a statement about the determinant, equation (6) does not necessarily mean that the (j,k) element of $\frac{\partial \mathbf{y}_t^{sa}}{\partial\left(\left(\mathbf{\Lambda}^{sa}\right)^{-1/2}\boldsymbol{\epsilon}_t^{sa}\right)'}$ is equal to the (j,k) element of $D^{-1/2}\cdot\frac{\partial \mathbf{y}_t}{\partial\left(\mathbf{\Lambda}^{-1/2}\boldsymbol{\epsilon}_t\right)'}$. Nevertheless, for equation (6) to hold, it must be the case that at least some elements of the impulse response are subject to large biases.

Equation (5) is convenient because it summarizes the magnitude of the filter's distortions with a single scalar D. A limitation is that, as far as I can tell, there is not a simple characterization of the element-by-element difference between $\mathbf{Q}$ and $\mathbf{Q}^{sa}$. To get a better sense of how distortions in $\mathbf{Q}$ can manifest in impulse responses and identified sets, consider the following pen-and-paper examples.

Example 1. Many papers assume $\mathbf{\Psi}$ to be lower-triangular and normalize each shock to have unit variance. That is, $\mathcal{I}(\mathbf{Q}) = \left(\text{chol}\left(\mathbf{Q}^{-1}\right)^{-1}, \mathbf{I}_n\right)$, where $\text{chol}(\cdot)$ denotes the lower Cholesky factor. Because $\mathbf{\Psi}$ and $\mathbf{\Psi}^{sa}$ are lower-triangular, their determinants are the products of their main diagonal elements. Equation (6), after taking logs and rearranging terms, is therefore indicative of the average error that seasonal adjustment creates when computing the contemporaneous response of variable k to shock k:

$$\frac{1}{n}\sum_{k=1}^{n}\left[\log\left(\frac{\partial \mathbf{y}_{k,t}}{\partial \boldsymbol{\epsilon}_{k,t}}\right) - \log\left(\frac{\partial \mathbf{y}_{k,t}^{sa}}{\partial \boldsymbol{\epsilon}_{k,t}^{sa}}\right)\right] = \frac{1}{2}\log(D) \approx .52. \tag{7}$$

Hence, seasonal adjustment creates large distortions in the impulse response function: $\frac{\partial \mathbf{y}_{k,t}}{\partial \boldsymbol{\epsilon}_{k,t}}$ is, on average, .52 log points larger than $\frac{\partial \mathbf{y}_{k,t}^{sa}}{\partial \boldsymbol{\epsilon}_{k,t}^{sa}}$, corresponding to a discrepancy of 68%.

Example 2. This example is based on Baumeister and Hamilton (2015) and will form the basis of the empirical application in Section 6. I will estimate a supply-and-demand model of the labor market using aggregate data on real wages and personhours. The model specifies:

$$\mathbf{y}_t = \begin{bmatrix} \Delta\log(\text{real wage}_t) \\ \Delta\log(\text{personhours}_t) \end{bmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_t^d \\ \epsilon_t^s \end{bmatrix}, \quad \mathbf{\Psi} = \begin{bmatrix} -\eta_d & 1 \\ -\eta_s & 1 \end{bmatrix}, \tag{8}$$

where ϵ_t^d is a demand shock, ϵ_t^s is a supply shock, η_d is the (short-run) elasticity of labor demand, and η_s is the (short-run) elasticity of labor supply. Equation (8), combined with the reduced-form

projection $\mathbf{y}_t = \hat{\mathbf{y}}_t + \mathbf{e}_t$, implies a demand curve and a supply curve:

$$\begin{aligned} \Delta \log \left(\text{personhours}_t\right) &= \eta_d \times \Delta \log \left(\text{real wage}_t\right) + \boldsymbol{\phi}^d \left(L\right)' \mathbf{y}_t + \epsilon_t^d & (9) \\ \Delta \log \left(\text{personhours}_t\right) &= \eta_s \times \Delta \log \left(\text{real wage}_t\right) + \boldsymbol{\phi}^s \left(L\right)' \mathbf{y}_t + \epsilon_t^s, & (10) \end{aligned}$$

where $\boldsymbol{\phi}^d(L)$ and $\boldsymbol{\phi}^s(L)$ are backward-looking vector-valued lag polynomials. Economic theory implies the sign restrictions $\eta_d < 0$ and $\eta_s > 0$, so the demand curve slopes down and the supply curve slopes up. The identified set is:

$$\mathcal{I}(\mathbf{Q}) = \left\{ (\boldsymbol{\Psi}, \boldsymbol{\Lambda}) \middle| \boldsymbol{\Psi} = \begin{bmatrix} -\eta_d & 1 \\ -\eta_s & 1 \end{bmatrix}, \ \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_d & 0 \\ 0 & \lambda_s \end{bmatrix}, \ \boldsymbol{\Psi}'\boldsymbol{\Lambda}\boldsymbol{\Psi} = \mathbf{Q}, \ (\eta_d, \eta_s, \lambda_d, \lambda_s) \in \mathbb{R}_- \times \mathbb{R}_+^3 \right\}. \tag{11}$$

The model is partially identified, and there are many configurations of structural parameters consistent with the precision matrix of the reduced-form residuals. Given $\mathbf{Q}$, suppose that $(\eta_d, \eta_s, \lambda_d, \lambda_s)$ satisfies the restrictions on the structural parameters in equation (11), and given $\mathbf{Q}^{sa}$, suppose that $(\eta_d^{sa}, \eta_s^{sa}, \lambda_d^{sa}, \lambda_s^{sa})$ satisfies the analogous set of restrictions for $\mathcal{I}(\mathbf{Q}^{sa})$. Note that:

$$|\mathbf{Q}| = \lambda_d \lambda_s \left(|\eta_d| + |\eta_s|\right)^2, \quad |\mathbf{Q}^{sa}| = \lambda_d^{sa} \lambda_s^{sa} \left(|\eta_d^{sa}| + |\eta_s^{sa}|\right)^2. \tag{12}$$

It's possible that $(\lambda_d, \lambda_s) = (\lambda_d^{sa}, \lambda_s^{sa})$, but in that case, equation (5) would imply:

$$\frac{|\eta_d^{sa}| + |\eta_s^{sa}|}{|\eta_d| + |\eta_s|} = D \approx 2.83, \tag{13}$$

so the magnitudes of the elasticities consistent with the seasonally adjusted time series would be, on average, about 2.83 times larger than the elasticities consistent with the original, unadjusted data. Alternatively, it's possible that $(\eta_d, \eta_s) = (\eta_d^{sa}, \eta_s^{sa})$, but in that case, equation (5) would imply:

$$\frac{1}{2}\left(\left[\log\left(\frac{1}{\lambda_d^{sa}}\right) - \log\left(\frac{1}{\lambda_d}\right)\right] + \left[\log\left(\frac{1}{\lambda_s^{sa}}\right) - \log\left(\frac{1}{\lambda_s}\right)\right]\right) = -\log(D) \approx -1.04, \tag{14}$$

so the shock variances consistent with the seasonally adjusted time series would be, on average, about 1.04 log points lower than the shock variances consistent with the original, unadjusted data. It's clear that seasonal adjustment will create large distortions in an econometrician's beliefs about some structural parameters of interest, although it's not obvious ex ante whether the distortions will manifest more in the elasticities, the shock variances, or some combination. In Section 6, I will investigate empirically the differences between estimating this model with seasonally adjusted and unadjusted time series.

2.2 Identifying Information in Seasonal Heteroskedasticity

When identifying structural VARs, econometricians should exploit seasonality, rather than filter it out. Time series can exhibit seasonal patterns in their conditional variances, as well as their conditional means, and seasonal heteroskedasticity can provide a useful source of identifying information. Assume that the structural shocks have a separate precision matrix for each calendar season, i.e., $\mathbb{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right]^{-1} = \boldsymbol{\Lambda}_t$, where $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_{t'}$ if $t \overset{\text{mod } n_s}{=} t'$. By extension, there will be n_s season-specific precision matrices for the reduced-form innovations, given by $\mathbf{Q}_t \equiv \mathbb{E}\left[\mathbf{e}_t\mathbf{e}_t'\right]^{-1} = \boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}$. Definition 1 is amended as follows.

Definition 2. An *identification scheme with seasonal heteroskedasticity* is a mapping $\mathcal{I} : \mathcal{Q}^{n_s} \to \mathcal{P} \times \mathcal{L}^{n_s}$ with the property that if $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_{n_s}) = \mathcal{I}(\mathbf{Q}_1, \ldots, \mathbf{Q}_{n_s})$, then $\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi} = \mathbf{Q}_t$, $\forall t \in \{1, \ldots, n_s\}$.

This formulation makes it possible to exploit results from the literature on identification through heteroskedasticity. Prominent examples of this approach include Rigobon (2003), Rigobon and Sack (2003, 2004), Wright (2012), and Brunnermeier et al. (2021). It's well known that identification through heteroskedasticity allows the structural parameters to be identified (up to a scaling factor for each row of $\boldsymbol{\Psi}$) if the relative variances of the structural shocks change between two points in time: $\mathbf{Q}_t\mathbf{Q}_{t'}^{-1} = \boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_{t'}^{-1}\boldsymbol{\Psi}'^{-1}$, so if the diagonal elements of $\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_{t'}^{-1}$ are distinct, then the kth row of $\boldsymbol{\Psi}$ is proportional to the kth eigenvector of $\mathbf{Q}_t\mathbf{Q}_{t'}^{-1}$. Often, identification through heteroskedasticity relies on designating different volatility regimes based on the timing of policy interventions or large market disruptions. Identification through seasonal heteroskedasticity simply relies on the notion that different types of shocks may tend to be more volatile at different times of the year. Beaulieu et al. (1992) document stylized facts about seasonality in the volatility of output, and those authors discuss a theoretical model that can rationalize seasonal heteroskedasticity. However, to the best of my knowledge, no one has combined seasonal volatility with identification through heteroskedasticity in a structural VAR.

Let's return to the model of labor supply and labor demand from Example 2. Baumeister and Hamilton (2015) assume that the shocks are homoskedastic, so their version of the model is only partially identified. The model becomes fully identified if there is seasonal heteroskedasticity and the ratio of variances between supply shocks and demand shocks changes across seasons.[6] If the relative variance of demand shocks is higher in summer than in winter, then movements in wages and hours that occur in summertime are more likely to stem from demand shocks. In that case, summertime variation in the data would help trace out the shape of the supply curve, and wintertime variation

[6] As noted above, in the general case, each row of $\boldsymbol{\Psi}$ is identified up to a scaling factor; for this model, the scaling indeterminacy is resolved by the right column of $\boldsymbol{\Psi}$ in equation (8) containing all ones.

in the data would help trace out the shape of the demand curve. For this reason, Rigobon (2003) likens identification through heteroskedasticity to using a "probabilistic instrument" (p. 777). Note that this approach is different from a more conventional instrumental-variables strategy, along the lines discussed in Miron and Beaulieu (1996): Seasonal dummy variables can be valid instruments if there are deterministic seasonal shifts in either supply or demand, but not both.[7] However, in the context of labor markets, there is reason to believe that there are seasonal components to both demand (e.g., the need for more retail workers before Christmas) and supply (e.g., the preference of workers to take vacations during the summer). In Section 6, I will augment equations (9) and (10) to allow both supply and demand to have deterministic seasonal components, while still being able to identify the structural parameters through seasonal heteroskedasticity.

3 Seasonal Priors

I will take a Bayesian approach to estimating equation (1). Section 3.1 provides a prior for the parameters that govern $\mathbf{s}_t$, and Section 3.2 and provides a prior for the parameters that govern $\tilde{\mathbf{y}}_t$. Section 3.3 touches on how the strategy can be extended to seasonal cointegration.

3.1 Deterministic Seasonality

I will model $\mathbf{s}_t$ as a linear combination of $n_s - 1$ waveforms:

$$\mathbf{s}_t = \mathbf{B}\mathbf{w}_t \tag{15}$$

$$\mathbf{w}_t \equiv (\mathbf{w}_{1,t}, \ldots, \mathbf{w}_{n_s-1,t})' \tag{16}$$

$$\mathbf{w}_{j,t} \equiv \sqrt{2}\cos\left(\frac{2\pi}{n_s}jt - \frac{\pi}{4}\right), \quad j = 1, \ldots, n_s - 1. \tag{17}$$

One can show that it's always possible to write $(1, \mathbf{w}_t')'$ as a linear combination of n_s seasonal dummy variables, and vice versa. In that regard, working with $\mathbf{w}_t$ comes without loss of generality when describing the population behavior of $\mathbf{s}_t$. The advantages of this formulation come in eliciting a sensible prior and performing finite-sample inference. Many economic time series appear to have only one or two seasonal peaks and troughs per year. In such cases, using a full slate of season-specific dummy variables would be gratuitous for summarizing deterministic seasonal fluctuations.

[7] To motivate the use of seasonal dummy variables as instruments, Miron and Beaulieu (1996) present an equilibrium model with two exogenous disturbances, one for a representative household's preferences and one for a representative firm's technology. The authors argue that it's reasonable to assume that preferences have a deterministic seasonal component, but technology does not. Miron and Beaulieu add the caveat: "In certain circumstances, a researcher may not like the assumption that the orthogonality condition holds in all months. A simple remedy is to pare the list of instruments to those months where the orthogonality condition is likely to hold." (p. 56)

Instead, the regular ups and downs that occur each year can often be well approximated using one or two sinusoids with annual or semi-annual periodicities. Any sinusoid with a period of $1/j$ years can be expressed as a linear combination of $\mathbf{w}_{j,t}$ and $\mathbf{w}_{n_s-j,t}$. The higher-frequency components of $\mathbf{w}_t$ make it possible for $\mathbf{s}_t$ to replicate more minute differences between seasons, but the associated coefficients can introduce additional estimation uncertainty, while running the risk of over-fitting. This suggests shrinking the coefficients on the higher-frequency elements of $\mathbf{w}_t$ toward zero. By suppressing rapid oscillations, the prior will favor smoothness in $\mathbf{s}_t$: The average value of a time series in February is likely to be between the average value in January and the average value in March. The question becomes how much shrinkage is reasonable to apply to each frequency.

To elicit a specific prior, it's helpful to connect beliefs about the coefficients to beliefs about the dynamics of $\mathbf{s}_t$. Define $\mathbf{\Gamma}_u^s$ as the uth autocovariance of $\mathbf{s}_t$ in a sample of length T:

$$\mathbf{\Gamma}_u^s \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_{t-u}'. \tag{18}$$

Placing a prior over $\mathbf{B}$ induces a prior for $\mathbf{\Gamma}_u^s$. A reasonable prior should have two attributes. First, the expected correlation between adjacent seasons should be higher than the expected correlation between distant seasons. The rate at which $\mathbf{\Gamma}_u^s$ changes with u provides an indication of how smoothly $\mathbf{s}_t$ varies with t. Second, $\mathbf{s}_t$ should account for a plausible fraction of the variance of the observed series $\mathbf{y}_t$. That is, the prior expectation of $\mathbf{\Gamma}_0^s$ should not be vastly disproportionate to the prior predictive variance of $\mathbf{y}_t$.

The vector of waveforms $\mathbf{w}_t$ has three properties that are convenient for analyzing and interpreting the implied prior over $\mathbf{\Gamma}_u^s$. If the number of observations T is divisible by the number of seasons n_s, then one can show that:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_t = \mathbf{0}_{(n_s-1)\times 1} \tag{19}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_t\mathbf{w}_t' = \mathbf{I}_{n_s-1} \tag{20}$$

$$\mathbf{w}_{t-1} = \mathbf{R}\mathbf{w}_t, \tag{21}$$

where $\mathbf{R}$ is a known orthogonal matrix defined in the appendix.[8] Equation (19) implies that the sample mean of $\mathbf{s}_t$ is zero, so $\mathbf{s}_t$ should be interpreted as the season-specific deviation from a long-run average. Equations (20) and (21) make it easy to write the autocovariance of $\mathbf{s}_t$ in terms of the

[8] Assuming that T is divisible by n_s makes some of the algebra cleaner but comes with minimal loss of generality: If T were not divisible by n_s, then equation (21) would still hold, and equations (19) and (20) would hold approximately, with the order of approximation error being $O\left(T^{-1}\right)$.

coefficient matrix $\mathbf{B}$:

$$\mathbf{\Gamma}_u^s = \mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}'. \tag{22}$$

The first and second moments of the prior distribution of $\mathbf{B}$ determine the prior expectation of $\mathbf{\Gamma}_u^s$. Let $\mathbf{b} \equiv \text{vec}(\mathbf{B})$, and let $\mathbb{E}_{prior}[\cdot]$ denote expectations taken under the prior distribution. I will assume that the prior satisfies:

$$\mathbb{E}_{prior}[\mathbf{b}] = \mathbf{0}_{(n_s-1)n\times 1}, \quad \mathbb{E}_{prior}[\mathbf{b}\mathbf{b}'] = \mathbf{K} \otimes \mathbf{V}_S, \tag{23}$$

where $\mathbf{K}$ and $\mathbf{V}_S$ are symmetric, positive-definite matrices of dimension $(n_s - 1) \times (n_s - 1)$ and $n \times n$, respectively. The Kronecker structure of the variance comes with two advantages. First, this specification limits the number of hyperparameters, relative to using an unrestricted variance matrix. Second, the matrices $\mathbf{K}$ and $\mathbf{V}_S$ will have distinct, interpretable rolls.

Proposition 2. *Assume that the prior satisfies equation (23). Then, the prior expectation of the uth autocovariance of $\mathbf{s}_t$ is:*

$$\mathbb{E}_{prior}[\mathbf{\Gamma}_u^s] = \kappa_u \mathbf{V}_S, \tag{24}$$

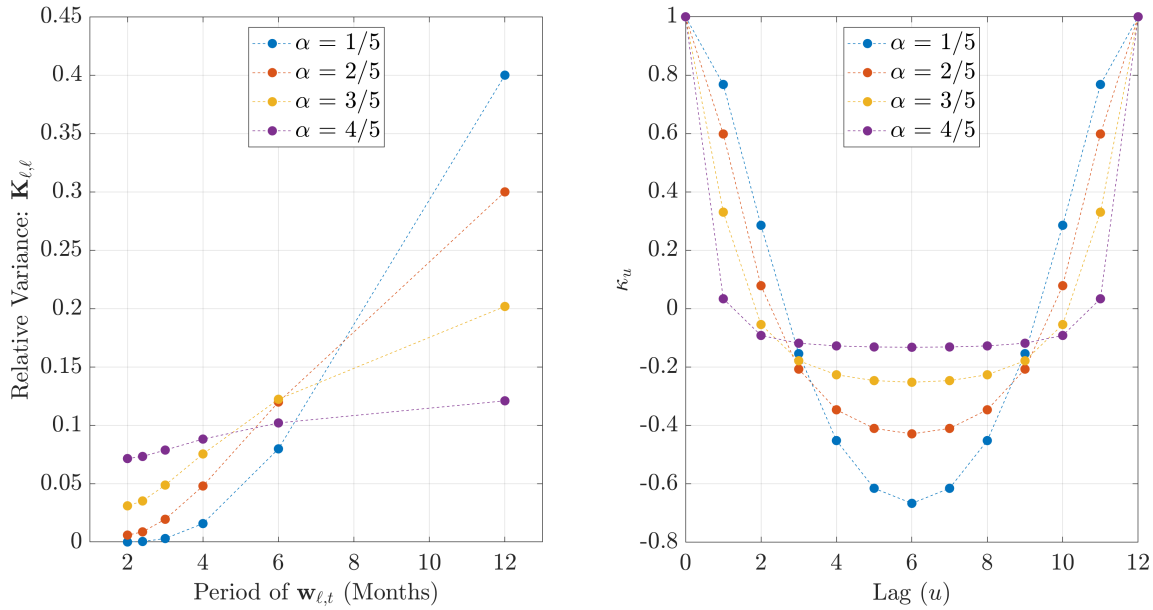
where κ_u is a scalar defined as $\kappa_u \equiv \sum_{\ell=1}^{n_s-1} \mathbf{K}_{\ell,\ell} \cos\left(\frac{2\pi}{n_s}\ell u\right)$.

The matrix $\mathbf{K}$ governs beliefs about how $\mathbf{s}_t$ is correlated across seasons, whereas the matrix $\mathbf{V}_S$ governs beliefs about how the elements of $\mathbf{s}_t$ are correlated across variables. In practice, I will set $\mathbf{K}$ to be a diagonal matrix, with the ℓth element of the main diagonal given by:

$$\mathbf{K}_{\ell,\ell} \propto \alpha^{\ell} + \alpha^{n_s-\ell}, \tag{25}$$

normalized such that $\text{tr}\{\mathbf{K}\} = \kappa_0 = 1$. The hyperparameter $\alpha \in (0, 1)$ governs the expected smoothness of deterministic seasonal fluctuations, by determining how aggressively to squeeze high-frequency oscillations in $\mathbf{w}_t$ toward zero. Figure 2 illustrates the role of α when $n_s = 12$. The left panel plots $\mathbf{K}_{\ell,\ell}$ as a function of $\frac{n_s}{\ell}$. Note that $\mathbf{K}_{\ell,\ell}$ controls the prior variance of the ℓth and $(n_s - \ell)^{th}$ columns of $\mathbf{B}$, and $\frac{n_s}{\ell}$ is the period of oscillation for the ℓth and $(n_s - \ell)^{th}$ elements of $\mathbf{w}_t$. Adopting a lower value of α applies more shrinkage to the components of $\mathbf{w}_t$ with short periods, while loosening the prior associated with long periods. Proposition 2 demonstrates how this prior across frequencies translates into smoothness across seasons. The right panel of Figure 2 plots the κ_u coefficients implied by the choice of $\mathbf{K}$ in equation (25). Adjacent seasons are expected to be relatively similar, and seasons that are half a year apart are expected to be dissimilar. When α is lower, $\mathbf{s}_t$ is expected to be smoother, in the sense of being more positively correlated across

Figure 2: Shrinkage for Deterministic Seasonality



Notes: For $n_s = 12$ and various values of α, the left panel plots the diagonal elements of the matrix $\mathbf{K}$, and the right panel plots the κ_u coefficients. $\mathbf{K}_{\ell,\ell}$ is plotted as a function of $\frac{n_s}{\ell}$, because $\mathbf{K}_{\ell,\ell}$ controls the variance of the coefficients on $\mathbf{w}_{\ell,t}$ and $\mathbf{w}_{n_s-\ell,t}$, which are both sinusoids with period $\frac{n_s}{\ell}$. Because $\mathbf{K}_{\ell,\ell} = \mathbf{K}_{n_s-\ell,n_s-\ell}$, only the first six elements of the diagonal of $\mathbf{K}$ are depicted.

consecutive seasons.

Given $\mathbf{K}$, the choice of $\mathbf{V}_S$ determines the expected contribution of $\mathbf{s}_t$ to the variance of $\mathbf{y}_t$. Assuming that the prior treats $\mathbf{B}$ as independent of the parameters governing $\tilde{\mathbf{y}}_t$, the expected sample variance of $\mathbf{y}_t$ can be decomposed into contributions from the deterministic component $\mathbf{s}_t$ and contributions from the stochastic component $\tilde{\mathbf{y}}_t$:

$$\mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'\right] = \kappa_0 \mathbf{V}_S + \mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\right]. \tag{26}$$

It may be tempting to adopt a diffuse prior over $\mathbf{B}$ for the sake of being "agnostic" about deterministic seasonality, but the above decomposition shows why doing so would be a bad idea. If $\mathbf{V}_S$ were arbitrarily set to a matrix of very large numbers, then it would imply a prior expectation that $\mathbf{s}_t$ has an enormous variance. By extension, an extremely diffuse prior for $\mathbf{B}$ would imply a prior belief that $\mathbf{y}_t$ has an enormous variance, stemming largely from deterministic seasonal fluctuations. The choice of $\mathbf{V}_S$ should reflect expectations about the fraction of the variance attributable to deterministic seasonality. Let $\bar{\boldsymbol{\Sigma}}_y$ be a prior estimate for the variance of $\mathbf{y}_t$. I will specify $\mathbf{V}_S = \varsigma\bar{\boldsymbol{\Sigma}}_y$, where $\varsigma \in (0,1)$. Because $\kappa_0 = 1$, the hyperparameter ς can be interpreted as the expected fraction of the

variance of $\mathbf{y}_t$ that can be accounted for by $\mathbf{s}_t$.

Although the prior in equation (23) is informative about the magnitude and persistence of deterministic seasonality, it is agnostic about the timing of seasonal peaks and troughs. Because $\mathbf{B}$ has zero mean, $\mathbb{E}_{prior}[\mathbf{s}_t]$ does not depend on t, so the prior takes no stand on whether $\mathbf{s}_t$ is more likely to peak in summer or in winter. Whether this property is a virtue or a limitation depends on the context of the application and the beliefs of the researcher. For example, if the goal is to model retail sales, then it may be desirable to incorporate a prior belief that $\mathbf{s}_t$ spikes in December and crashes in January, based on the timing of Christmas. Appendix E contains a modified prior that allows the econometrician to specify a separate prior mean for each $\{\mathbf{s}_t\}_{t=1}^{n_s}$, and the prior exhibits positive autocorrelation in $\mathbf{s}_t - \mathbb{E}_{prior}[\mathbf{s}_t]$.

The most obvious alternative to using seasonal waveforms is using season-specific dummy variables, but those have some unappealing features. In applied work, Bayesian practitioners often assume that the coefficients on seasonal dummies are uncorrelated under the prior. However, as explained in Appendix F, that approach implies that $\mathbf{s}_t$ is expected to exhibit *negative* serial correlation. Such beliefs are usually difficult to defend. Suppose that a time series (say, inflation) tends to be low in January and March. Conditional on January and March being relatively low-inflation months, an econometrician using uncorrelated dummy variables would expect February to be a relatively *high*-inflation month. In contrast, the implications of equation (23) are more reasonable: If January and March are low-inflation months, then inflation is probably low in the winter and high in the summer, so February is expected to be similar to January and March.

3.2 Stochastic Seasonality

I will model the stochastic component of $\mathbf{y}_t$ as a VAR(m):

$$\mathbf{A}(L)\,\tilde{\mathbf{y}}_t = \boldsymbol{\epsilon}_t, \quad \mathbf{A}(L) \equiv \boldsymbol{\Psi} - \sum_{\ell=1}^{m} \boldsymbol{\Phi}_\ell L^\ell, \quad \boldsymbol{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \boldsymbol{\Lambda}_t^{-1}\right), \tag{27}$$

where each $\boldsymbol{\Lambda}_t$ is diagonal, and $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_{t'}$ whenever $t \overset{\text{mod } n_s}{=} t'$. The reduced-form representation is:

$$\tilde{\mathbf{y}}_t = \sum_{\ell=1}^{m} \boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}_\ell \tilde{\mathbf{y}}_{t-\ell} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \mathbf{Q}_t^{-1}\right), \tag{28}$$

where $\mathbf{e}_t \equiv \boldsymbol{\Psi}^{-1}\boldsymbol{\epsilon}_t$ and $\mathbf{Q}_t \equiv \boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}$. Beliefs about $\boldsymbol{\Psi}$ depend on the specific application; here, I will simply assume that the econometrician has a marginal prior over $\boldsymbol{\Psi}$ and focus on the conditional prior over $\{\boldsymbol{\Phi}_\ell\}_{\ell=1}^{m}$ and $\{\boldsymbol{\Lambda}_t\}_{t=1}^{n_s}$. My prior will reflect a belief that random oscillations at seasonal frequencies play an important part in accounting for the variation in $\tilde{\mathbf{y}}_t$. First, I will explain the

prior under the assumption of homoskedasticity ($\mathbf{\Lambda}_t = \mathbf{\Lambda}$, $\forall t$). Doing so will make it easier to articulate prior beliefs about the frequency-domain properties of $\tilde{\mathbf{y}}_t$. Then, I will discuss the prior with seasonal heteroskedasticity.

The Homoskedastic Case. Let λ_k denote the kth element of the main diagonal of $\mathbf{\Lambda}$, and let $\mathbf{\Phi} \equiv \begin{bmatrix} \mathbf{\Phi}_1 & \cdots & \mathbf{\Phi}_m \end{bmatrix}$ concatenate the structural lag coefficients. One can write T observations from the process (27) as $\tilde{\mathbf{Y}}\mathbf{\Psi}' = \tilde{\mathbf{X}}\mathbf{\Phi}' + \boldsymbol{\mathcal{E}}$, where the tth rows of $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{X}}$, and $\boldsymbol{\mathcal{E}}$ contain $\tilde{\mathbf{y}}_t'$, $\left(\tilde{\mathbf{y}}_{t-1}', \ldots, \tilde{\mathbf{y}}_{t-m}'\right)$, and $\boldsymbol{\epsilon}_t'$. My prior will take a normal-gamma form. It's well known that such a prior can be implemented by augmenting the observed data with $\bar{T}$ dummy observations of the form $\bar{\mathbf{Y}}\mathbf{\Psi}' = \bar{\mathbf{X}}\mathbf{\Phi}' + \bar{\boldsymbol{\mathcal{E}}}$, with $\left(\bar{\boldsymbol{\mathcal{E}}}\right)_{j,k} \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \lambda_k\right)$.[9] More explicitly, the prior implied by the dummy observations is:

$$\lambda_k \mid \mathbf{\Psi} \sim \mathrm{G}\left(\alpha_\lambda, \beta_{\lambda,k}\right) \tag{29}$$

$$\mathrm{vec}\left(\mathbf{\Phi}\right) \mid \left\{\lambda_k\right\}_{k=1}^n, \mathbf{\Psi} \sim \mathrm{N}\left(\mathrm{vec}\left(\bar{\mathbf{\Phi}}\right), \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \mathbf{\Lambda}\right)^{-1}\right) \tag{30}$$

$$\alpha_\lambda \equiv \frac{\bar{T} - mn}{2} + 1 \tag{31}$$

$$\beta_{\lambda,k} \equiv \frac{1}{2}\left[\left(\bar{\mathbf{Y}}\mathbf{\Psi}' - \bar{\mathbf{X}}\bar{\mathbf{\Phi}}'\right)'\left(\bar{\mathbf{Y}}\mathbf{\Psi}' - \bar{\mathbf{X}}\bar{\mathbf{\Phi}}'\right)\right]_{k,k} \tag{32}$$

$$\bar{\mathbf{\Phi}} \equiv \mathbf{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}, \tag{33}$$

with $\left\{\lambda_k\right\}_{k=1}^n$ independent across k. The specification of $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$ determines the substance of the prior.

My prior will favor seasonal unit roots. With $\mathbf{A}\left(L\right)\tilde{\mathbf{y}}_t = \boldsymbol{\epsilon}_t$, the process $\tilde{\mathbf{y}}_t$ has a unit root if $\left|\mathbf{A}\left(z\right)\right| = 0$ for some $z \in \mathbb{C}$ such that $\left|z\right| = 1$. Any point on the complex unit circle z can be written in polar form as $z = \exp\left\{i\omega^*\right\}$, where $\omega^* \in \left[-\pi, \pi\right]$. Much of the literature uses "unit root" synonymously with "*zero-frequency* unit root," meaning $\left|\mathbf{A}\left(1\right)\right| = 0$. However, the distinction between zero-frequency unit roots and other unit roots is important in the context of seasonality. The process $\tilde{\mathbf{y}}_t$ is said to have a *seasonal* unit root at frequency ω^* if $\left|\mathbf{A}\left(\exp\left\{i\omega^*\right\}\right)\right| = 0$, where $\omega^* \neq 0$. The connection between seasonality and the location of the root on the unit circle is most apparent in the frequency domain. When the shocks are homoskedastic, the spectrum of $\tilde{\mathbf{y}}_t$ is:

$$\mathbf{f}\left(\omega\right) = \frac{1}{2\pi}\left[\mathbf{A}\left(\exp\left\{i\omega\right\}\right)'\mathbf{\Lambda}\mathbf{A}\left(\exp\left\{-i\omega\right\}\right)\right]^{-1}. \tag{34}$$

The function $\mathbf{f}\left(\omega\right)$ is a proper spectral density if $\tilde{\mathbf{y}}_t$ is stationary; otherwise, $\mathbf{f}\left(\omega\right)$ is understood to be a pseudo-spectral density. For a univariate process, a seasonal unit root at frequency ω^*

[9] See, e.g., Del Negro and Schorfheide (2011). The *dummy observations* used to implement a normal-gamma prior should not be confused with the *dummy variables* (or indicator variables) that other authors use to model deterministic seasonality.

implies that the spectrum has a peak at ω^*, with $\lim_{\omega \to \omega^*} f(\omega) = \infty$. Consequently, oscillations at frequency ω^* account for a substantial amount of the variation in the time series. More generally, for a multivariate process, a seasonal unit root at frequency ω^* implies that $|\mathbf{f}(\omega)|$ tends to infinity as ω approaches ω^*.

A prior belief in a seasonal unit root at a specific frequency ω^* can be implemented using linear restrictions on the autoregressive coefficients. A sufficient condition for $\mathbf{y}_t$ to have a seasonal unit root at frequency ω^* is $\mathbf{A}(\exp\{i\omega^*\}) = \mathbf{0}_{n\times n}$. Following Litterman (1986), many macroeconometricians adopt priors centered on the reduced-form autoregressive coefficients $\left\{\mathbf{\Psi}^{-1}\mathbf{\Phi}_\ell\right\}_{\ell=1}^m$ being diagonal. When the reduced-form autoregressive coefficients are diagonal, the condition $\mathbf{A}(\exp\{i\omega^*\}) = \mathbf{0}_{n\times n}$ is both necessary and sufficient for each individual series $\mathbf{y}_{j,t}$ to have a seasonal unit root at frequency ω^*. Observe that:

$$\mathbf{A}(\exp\{i\omega^*\}) = \mathbf{\Psi} - \sum_{\ell=1}^m \mathbf{\Phi}_\ell \cos(\omega^*\ell) - i\sum_{\ell=1}^m \mathbf{\Phi}_\ell \sin(\omega^*\ell). \tag{35}$$

For the left-hand side of the above expression to be zero, both the real part and the imaginary part of the right-hand side must be zero, which implies two sets of linear restrictions:

$$\mathbf{\Psi} = \sum_{\ell=1}^m \mathbf{\Phi}_\ell \cos(\omega^*\ell) \tag{36}$$

$$\mathbf{0}_{n\times n} = \sum_{\ell=1}^m \mathbf{\Phi}_\ell \sin(\omega^*\ell). \tag{37}$$

Imposing the above restrictions dogmatically would force $\tilde{\mathbf{y}}_t$ to have a unit root at frequency ω^*. Instead, I will use dummy observations to implement stochastic constraints, along the lines of Theil and Goldberger (1961). Doing so will simply favor areas of the parameter space where $\tilde{\mathbf{y}}_t$ is close to having a unit root at frequency ω^*. Without loss of generality, we can restrict our attention to unit roots at frequencies weakly between 0 and π, because the roots must come in conjugate pairs. First, consider the case where $\omega^* \in (0, \pi)$. A belief in a seasonal unit root at frequency ω^* can be expressed via dummy observations of the form:

$$\bar{\mathbf{Y}}_{\omega^*}\mathbf{\Psi}' = \bar{\mathbf{X}}_{\omega^*}\mathbf{\Phi}' + \bar{\boldsymbol{\mathcal{E}}}_{\omega^*}, \quad \left(\bar{\boldsymbol{\mathcal{E}}}_{\omega^*}\right)_{j,k} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \lambda_k) \tag{38}$$

$$\bar{\mathbf{Y}}_{\omega^*} \equiv \tau_{\omega^*}\begin{bmatrix} \mathbf{I}_n \\ \mathbf{0}_{n\times n} \end{bmatrix} \tag{39}$$

$$\bar{\mathbf{X}}_{\omega^*} \equiv \tau_{\omega^*}\begin{bmatrix} \cos(\omega^*1) & \cos(\omega^*2) & \cdots & \cos(\omega^*m) \\ \sin(\omega^*1) & \sin(\omega^*2) & \cdots & \sin(\omega^*m) \end{bmatrix} \otimes \mathbf{I}_n, \tag{40}$$

where $\tau_{\omega^*} > 0$ is a scalar hyperparameter chosen by the econometrician. Equations (38)-(40) imply that each element of $\mathbf{A}\left(\exp\left\{i\omega^*\right\}\right)$ has mean zero and follows a complex normal distribution, conditional on $\mathbf{\Lambda}$ and $\mathbf{\Psi}$:

$$\operatorname{vec}\left(\mathbf{A}\left(\exp\left\{i\omega^*\right\}\right)\right) \mid \mathbf{\Lambda}, \mathbf{\Psi} \sim \mathrm{CN}\left(\mathbf{0}_{n^2\times 1}, \mathbf{I}_n \otimes \left(\frac{\tau_{\omega^*}^2}{2}\mathbf{\Lambda}\right)^{-1}\right). \tag{41}$$

The hyperparameter τ_{ω^*} controls the amount of prior confidence the econometrician has about the unit root at frequency ω^*. In the limiting case, where $\tau_{\omega^*} \to \infty$, the prior imposes the restrictions (36) and (37) exactly. Otherwise, if equations (36) and (37) hold only approximately, then the spectrum will have a peak that's close to ω^*, though not necessarily a seasonal unit root. Now, consider the case where $\omega^* \in \{0, \pi\}$. Because $\sin(0) = \sin(\pi) = 0$, equation (37) will be satisfied exactly for any coefficient values, so one would omit the latter n lines of $\bar{\mathbf{Y}}_{\omega^*}$ and $\bar{\mathbf{X}}_{\omega^*}$. If $\omega^* = 0$, the first n lines of equations (38)-(40) would be equivalent to the dummy observations used to implement the sum-of-coefficients prior, originally due to Doan et al. (1984), for zero-frequency unit roots.

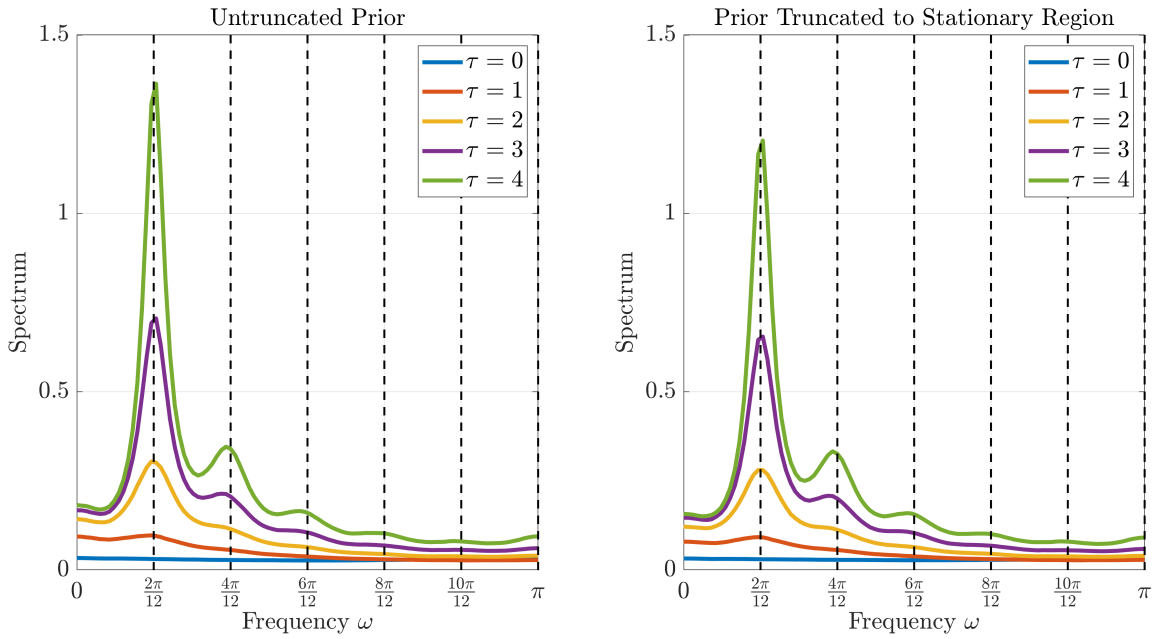
The preceding discussion considers a unit root at a single frequency, but the same line of reasoning can incorporate prior beliefs about multiple frequencies. If the seasonal difference $(1 - L^{n_s})\,\tilde{\mathbf{y}}_t$ is stationary, then $\tilde{\mathbf{y}}_t$ can have up to n_s unit roots. Specifically, the roots of the seasonal differencing operator $(1 - L^{n_s})$ take the form $\exp\left\{\pm\frac{2\pi}{n_s}ij\right\}$ for all integers j between zero and $\frac{n_s}{2}$. Let $\omega_j^* \equiv \frac{2\pi}{n_s}j$, for $j \in \left\{1, 2, \ldots, \left\lfloor\frac{n_s}{2}\right\rfloor\right\}$. These frequencies correspond to an annual periodicity, plus all harmonics (periods of half a year, a third of a year, etc.). Each frequency can receive its own set of dummy observations, as in equations (38)-(40), with its own prior precision $\tau_{\omega_j^*}$. In practice, I will set $\tau_{\omega_j^*} = \tau_S/j$. The hyperparameter $\tau_S > 0$ governs the econometrician's prior confidence that the process has a unit root at some seasonal frequency. The fact that $\tau_{\omega_j^*}$ is decreasing in j reflects a belief that stochastic seasonality is most likely to manifest at annual periodicities, somewhat likely to manifest at semi-annual periodicities, and least likely to manifest at very high frequencies.

The dummy observations provide a convenient way to combine prior beliefs about seasonality with prior beliefs about non-seasonal aspects of the time series. In the literature on Bayesian VARs, most priors are intended for seasonally adjusted variables and often reflect beliefs about low-frequency behavior. Let $\bar{\mathbf{Y}}_0$ and $\bar{\mathbf{X}}_0$ denote the dummy observations that implement this kind of baseline prior. One can concatenate the dummy observations for the baseline prior to the dummy observations for seasonal unit roots:

$$\bar{\mathbf{Y}} \equiv \begin{bmatrix} \bar{\mathbf{Y}}_0' & \bar{\mathbf{Y}}_{\omega_1^*}' & \cdots & \bar{\mathbf{Y}}_{\omega_{\lfloor n_s/2\rfloor}^*}' \end{bmatrix}', \quad \bar{\mathbf{X}} \equiv \begin{bmatrix} \bar{\mathbf{X}}_0' & \bar{\mathbf{X}}_{\omega_1^*}' & \cdots & \bar{\mathbf{X}}_{\omega_{\lfloor n_s/2\rfloor}^*}' \end{bmatrix}', \tag{42}$$

where $\bar{\mathbf{Y}}_{\omega_j^*}$ and $\bar{\mathbf{X}}_{\omega_j^*}$ take the form of equations (39) and (40), with $\omega_j^* = \frac{2\pi}{n_s}j$ and $\tau_{\omega_j^*} = \tau_S/j$ taking

Figure 3: Prior Beliefs About the Spectrum



Notes: The figure shows the prior median of the spectrum for various values of τ_S, assuming univariate monthly data with 13 lags. The prior conditions on $\Psi = 1$. The right panel truncates the prior to the stationary region of the parameter space; the left panel is the untruncated prior. Vertical dashed lines indicate seasonal frequencies $\frac{2\pi}{12}j$, $j = 1, \ldots, 6$. Each solid line is generated by taking $10,000$ draws from the conditional prior distribution for $\mathbf{\Phi}$ and λ, computing the spectrum associated with each parameter draw, and computing the median value of the spectrum across draws.

the place of ω^* and τ_{ω^*}. Because the conditional mean of $\mathbf{\Phi}$ is $\mathbf{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}$, the seasonal dummy observations shift the prior mean away from the baseline prior in the direction of the parameter space that exhibits seasonal unit roots. Because the conditional variance of $\text{vec}\left(\mathbf{\Phi}\right)$ is $\left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \mathbf{\Lambda}\right)^{-1}$, the dummy observations also induce non-trivial correlations between the coefficients: Conditional on some elements of $\mathbf{\Phi}$ deviating from the prior mean, one would expect the other elements to deviate in a way that maintains the spectral peaks near the seasonal frequencies.

Figure 3 demonstrates the prior's implications for the frequency-domain properties of $\tilde{\mathbf{y}}_t$. In the left panel, for different values of τ_S, I have plotted the prior median of the spectrum for univariate monthly data under the prior described above. The dashed vertical lines correspond to the seasonal frequencies $\left\{\frac{2\pi}{12}j\right\}_{j=1}^{6}$. Higher values of τ_S make the spectral peaks at seasonal frequencies more pronounced. In some contexts, one may want to truncate the prior to ensure that $\tilde{\mathbf{y}}_t$ is stationary, so in the right panel, Figure 3 shows the prior median of the spectrum under the truncated prior. Assuming stationarity precludes exact seasonal unit roots, but the prior still favors parameter values that imply spectral peaks near seasonal frequencies.

The Seasonally Heteroskedastic Case. When allowing for seasonal heteroskedasticity, I will parameterize the precision matrix as:

$$\log\left(\left(\mathbf{\Lambda}_t\right)_{k,k}\right) = \log\left(\lambda_k\right) + \boldsymbol{\rho}_k'\mathbf{w}_t, \tag{43}$$

with $\mathbf{w}_t$ defined in equations (16) and (17). Because the long-run average of $\mathbf{w}_t$ is zero, $\log(\lambda_k)$ can be interpreted as the average log precision of the kth shock, and $\boldsymbol{\rho}_k'\mathbf{w}_t$ captures seasonal heteroskedasticity in the kth shock. For $\boldsymbol{\rho}_k$, I will assume the prior:

$$\boldsymbol{\rho}_k \sim \mathrm{N}\left(\mathbf{0}_{(n_s-1)\times 1}, \nu_\rho \mathbf{K}\right), \tag{44}$$

independent across k, with $\mathbf{K}$ defined in equation (25). Assuming that the variance of $\boldsymbol{\rho}_k$ is proportional to $\mathbf{K}$ ensures that the seasonal component of $\log\left(\left(\mathbf{\Lambda}_t\right)_{k,k}\right)$ exhibits the same kind of smoothness as described in Section 3.1: The shock variances are expected to be similar in adjacent seasons, but not in seasons that are half a year apart. To interpret the hyperparameter $\nu_\rho > 0$, note that:

$$\mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\left[\log\left(\left(\mathbf{\Lambda}_t\right)_{k,k}\right) - \log\left(\lambda_k\right)\right]^2\right] = \nu_\rho \operatorname{tr}\left\{\mathbf{K}\right\}, \tag{45}$$

which follows from equation (20) and the fact that $\mathbb{E}_{prior}\left[\boldsymbol{\rho}_k'\boldsymbol{\rho}_k\right] = \operatorname{tr}\left\{\mathbb{E}_{prior}\left[\boldsymbol{\rho}_k\boldsymbol{\rho}_k'\right]\right\}$. My specification for $\mathbf{K}$ normalizes $\operatorname{tr}\{\mathbf{K}\} = 1$, so ν_ρ can be interpreted as the expected variance of $\log\left(\left(\mathbf{\Lambda}_t\right)_{k,k}\right)$.

The priors for $\{\lambda_k\}_{k=1}^{n}$ and $\mathbf{\Phi}$ described earlier continue to apply to the seasonally heteroskedastic case, but as a technical matter, the definition of the spectrum needs to be revised, because the shock variances depend on the calendar date. To analyze unconditional moments in the presence of seasonality, Hansen and Sargent (1993) propose treating the time series as a realization of the data-generating process, but with the calendar month of date $t = 0$ chosen at random.[10] Under this randomization scheme, the autocovariance function of $\tilde{\mathbf{y}}_t$ does not depend on the calendar date, and the spectrum is given by:

$$\mathbf{f}(\omega) = \frac{1}{2\pi}\mathbf{A}\left(\exp\{-i\omega\}\right)^{-1}\mathbf{\Sigma}_\epsilon\mathbf{A}\left(\exp\{i\omega\}\right)^{-1\prime}, \quad \mathbf{\Sigma}_\epsilon \equiv \frac{1}{n_s}\sum_{t=1}^{n_s}\mathbf{\Lambda}_t^{-1}. \tag{46}$$

Like before, the peaks in the spectrum are determined by the roots of the polynomial $|\mathbf{A}(L)|$, so the same dummy observations can be used to express prior beliefs about stochastic seasonality. Hence,

[10] More formally, let $\left\{\tilde{\mathbf{y}}_t^{(0)}\right\}_{t\in\mathbb{Z}}$ be a realization of the data-generating process given by equations (47) and (48), and for $k \in \{1,\ldots,n_s-1\}$, let $\left\{\tilde{\mathbf{y}}_t^{(k)}\right\}_{t\in\mathbb{Z}}$ be defined such that $\tilde{\mathbf{y}}_t^{(k)} \equiv \tilde{\mathbf{y}}_{t+k}^{(0)}$. Now, assume that $\{\tilde{\mathbf{y}}_t\}_{t\in\mathbb{Z}}$ is equal to $\left\{\tilde{\mathbf{y}}_t^{(k)}\right\}_{t\in\mathbb{Z}}$ with probability $\frac{1}{n_s}$, for each $k \in \{0,1,\ldots,n_s-1\}$.

the priors for $\{\lambda_k\}_{k=1}^n$ and $\mathbf{\Phi}$ are still given by equations (29)-(33), with the understanding that the $\mathbf{\Lambda}$ in equation (30) refers to $\text{diag}\left(\lambda_1, \ldots, \lambda_n\right)$.

3.3 A Remark on Seasonal Cointegration

It's possible to extend the strategy from Section 3.2 to express a prior belief in seasonal cointegration, in addition to seasonal unit roots. Seasonal cointegration, initially developed by Hylleberg, Engle, Granger, and Yoo (1990), extends the traditional approach to cointegration, pioneered by Engle and Granger (1987). The process $\tilde{\mathbf{y}}_t$ is seasonally cointegrated at frequency ω^* if $\tilde{\mathbf{y}}_t$ has a seasonal unit root at frequency ω^* but $\mathbf{\Upsilon}'\tilde{\mathbf{y}}_t$ does not, where $\mathbf{\Upsilon}$ is a full-column-rank $n \times r$ matrix with $r < n$. One implication is that $\mathbf{A}\left(\exp\left\{i\omega^*\right\}\right)\mathbf{\Upsilon}_\perp = \mathbf{0}_{n\times(n-r)}$, where $\mathbf{\Upsilon}_\perp$ is an orthogonal complement of $\mathbf{\Upsilon}$. Given a prior belief in a particular set of seasonal cointegrating relationships $\mathbf{\Upsilon}$, an econometrician can implement stochastic linear restrictions by replacing $\mathbf{I}_n$ with $\mathbf{\Upsilon}_\perp'$ in equations (39) and (40); like before, when $\omega^* \in \{0, \pi\}$, one would exclude the lower halves of equations (39) and (40). Such a prior, rather than imposing seasonal cointegration exactly, simply shrinks the coefficients toward the region of the parameter space where the seasonal cointegrating relationships are close to being satisfied. When $\omega^* = 0$ and $r = 1$, this approach nests as a special case Sims's (1993a) co-persistence dummy observations, which compensate for the traditional Minnesota prior's bias against zero-frequency cointegration. For the application in Section 6, my prior will favor seasonal unit roots, but not seasonal cointegration. However, the strategy described above provides a way to incorporate seasonal cointegration for researchers whose priors or applications differ from my own.

4 Estimation

Consolidating equations (1), (27), and (43) allows us to write the model as:

$$\mathbf{\Psi}\left(\mathbf{y}_t - \mathbf{G}\mathbf{w}_t^+\right) = \sum_{\ell=1}^{m} \mathbf{\Phi}_\ell\left(\mathbf{y}_{t-\ell} - \mathbf{G}\mathbf{w}_{t-\ell}^+\right) + \boldsymbol{\epsilon}_t \tag{47}$$

$$\boldsymbol{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{N}\left(\mathbf{0}_{n\times 1}, \text{diag}\left(\lambda_1 \exp\left\{\boldsymbol{\rho}_1'\mathbf{w}_t\right\}, \ldots, \lambda_n \exp\left\{\boldsymbol{\rho}_n'\mathbf{w}_t\right\}\right)^{-1}\right), \tag{48}$$

where I have defined $\mathbf{w}_t^+ \equiv \left(1, \mathbf{w}_t'\right)'$ and $\mathbf{G} \equiv \begin{bmatrix} \boldsymbol{\mu} & \mathbf{B} \end{bmatrix}$. The prior over $\mathbf{g} \equiv \text{vec}\left(\mathbf{G}\right)$ is normal:

$$\mathbf{g} \sim \text{N}\left(\bar{\mathbf{g}}, \mathbf{V}_g\right), \quad \bar{\mathbf{g}} \equiv \begin{bmatrix} \bar{\boldsymbol{\mu}} \\ \mathbf{0}_{(n_s-1)n\times 1} \end{bmatrix}, \quad \mathbf{V}_g \equiv \begin{bmatrix} \mathbf{V}_\mu & \mathbf{0}_{n\times(n_s-1)n} \\ \mathbf{0}_{(n_s-1)n\times n} & \mathbf{K}\otimes\mathbf{V}_S \end{bmatrix}. \tag{49}$$

I will assume that the matrix $\boldsymbol{\Psi}$ is parameterized by a vector $\boldsymbol{\eta}$, which has prior density $\mathbb{P}[\boldsymbol{\eta}]$. The dummy observations $\bar{\mathbf{Y}}_0$ and $\bar{\mathbf{X}}_0$ that enter into $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$ in equation (42) are allowed to depend on $\boldsymbol{\eta}$ as well. All other components of the prior are described in Sections 3.1 and 3.2.

Let $\boldsymbol{\theta}$ collect all of the model's parameters. Given the prior and data $\mathbf{y}^T \equiv \{\mathbf{y}_t\}_{t=1}^T$, the goal is to sample random draws from the posterior $\mathbb{P}\left[\boldsymbol{\theta} \mid \mathbf{y}^T\right]$ that can be used to form Monte Carlo inferences. To construct a sampling algorithm, it is necessary to characterize some analytical features of the posterior.

Proposition 3. *Let $\tilde{\mathbf{x}}_t \equiv \left(\tilde{\mathbf{y}}'_{t-1}, \ldots, \tilde{\mathbf{y}}'_{t-m}\right)'$. For $k \in \{1, \ldots, n\}$, define:*

$$\mathbf{Y}_{(k)} \equiv \left[\begin{array}{cccc} \exp\left\{\frac{1}{2}\boldsymbol{\rho}'_k \mathbf{w}_1\right\}\tilde{\mathbf{y}}_1 & \cdots & \exp\left\{\frac{1}{2}\boldsymbol{\rho}'_k \mathbf{w}_T\right\}\tilde{\mathbf{y}}_T & \bar{\mathbf{Y}}' \end{array}\right]' \tag{50}$$

$$\mathbf{X}_{(k)} \equiv \left[\begin{array}{cccc} \exp\left\{\frac{1}{2}\boldsymbol{\rho}'_k \mathbf{w}_1\right\}\tilde{\mathbf{x}}_1 & \cdots & \exp\left\{\frac{1}{2}\boldsymbol{\rho}'_k \mathbf{w}_T\right\}\tilde{\mathbf{x}}_T & \bar{\mathbf{X}}' \end{array}\right]'. \tag{51}$$

Let $\boldsymbol{\rho} \equiv \{\boldsymbol{\rho}_k\}_{k=1}^n$, and let $\boldsymbol{\lambda} \equiv \{\lambda_k\}_{k=1}^n$. Let $\boldsymbol{\phi}_k$ denote the kth column of $\boldsymbol{\Phi}'$, and let $\boldsymbol{\psi}_k$ denote the kth column of $\boldsymbol{\Psi}'$. Let $\bar{T}$ denote the number of rows in $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$. The conditional posterior for $\boldsymbol{\Phi}$ and $\boldsymbol{\lambda}$ is:

$$\lambda_k \mid \boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \sim \mathrm{G}\left(\hat{\alpha}_\lambda, \hat{\beta}_{\lambda,k}\right) \tag{52}$$

$$\boldsymbol{\phi}_k \mid \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \sim \mathrm{N}\left(\hat{\boldsymbol{\phi}}_k, \left(\lambda_k \mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right) \tag{53}$$

$$\hat{\boldsymbol{\phi}}_k \equiv \left(\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\mathbf{X}'_{(k)}\mathbf{Y}_{(k)}\boldsymbol{\psi}_k \tag{54}$$

$$\hat{\alpha}_\lambda \equiv \frac{T + \bar{T} - mn}{2} + 1 \tag{55}$$

$$\hat{\beta}_{\lambda,k} \equiv \frac{1}{2}\left(\mathbf{Y}_{(k)}\boldsymbol{\psi}_k - \mathbf{X}_{(k)}\hat{\boldsymbol{\phi}}_k\right)'\left(\mathbf{Y}_{(k)}\boldsymbol{\psi}_k - \mathbf{X}_{(k)}\hat{\boldsymbol{\phi}}_k\right), \tag{56}$$

independent across $k \in \{1, \ldots, n\}$. The posterior kernel for $(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$ is:

$$\begin{aligned} \mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] \propto & \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}[\boldsymbol{\rho}]\,\mathbb{P}[\mathbf{g}]\,\mathbb{P}[\boldsymbol{\eta}] \\ & \times |\boldsymbol{\Psi}|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^n \boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^T \mathbf{w}_t\right)\right\}\prod_{k=1}^n \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{1/2}}, \end{aligned} \tag{57}$$

where $\boldsymbol{\Psi}$, $\bar{\mathbf{X}}$, $\bar{\mathbf{Y}}$, and $\left\{\beta_{\lambda,k}, \hat{\beta}_{\lambda,k}, \mathbf{X}_{(k)}, \mathbf{Y}_{(k)}\right\}_{k=1}^n$ are understood to be functions of $(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$. Define:

$$\mathbf{z}_t \equiv \left(\mathbf{y}'_t, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-m}\right)', \quad \tilde{\boldsymbol{\Phi}} \equiv \left[\begin{array}{cc} \boldsymbol{\Psi} & -\boldsymbol{\Phi} \end{array}\right], \quad \mathbf{W}_t \equiv \left[\begin{array}{cccc} \mathbf{w}_t^+ & \mathbf{w}_{t-1}^+ & \cdots & \mathbf{w}_{t-m}^+ \end{array}\right]' \otimes \mathbf{I}_n. \tag{58}$$

The conditional posterior for $\mathbf{g}$ *is:*

$$\mathbf{g} \mid \boldsymbol{\Phi}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\eta} \sim \mathrm{N}\left(\hat{\mathbf{g}}, \hat{\mathbf{V}}_g\right) \tag{59}$$

$$\hat{\mathbf{V}}_g \equiv \left(\mathbf{V}_g^{-1} + \sum_{t=1}^{T} \mathbf{W}_t' \tilde{\boldsymbol{\Phi}}' \boldsymbol{\Lambda}_t \tilde{\boldsymbol{\Phi}} \mathbf{W}_t\right)^{-1} \tag{60}$$

$$\hat{\mathbf{g}} \equiv \hat{\mathbf{V}}_g \left(\mathbf{V}_g^{-1} \bar{\mathbf{g}} + \sum_{t=1}^{T} \mathbf{W}_t' \tilde{\boldsymbol{\Phi}}' \boldsymbol{\Lambda}_t \tilde{\boldsymbol{\Phi}} \mathbf{z}_t\right). \tag{61}$$

The results from Proposition 3 suggest using a Markov chain Monte Carlo (MCMC) algorithm. I will sample the parameters in the following blocks:

1. Draw $\boldsymbol{\eta}$, conditional on $(\boldsymbol{\rho}, \mathbf{g})$ and $\mathbf{y}^T$.
2. Draw $\boldsymbol{\rho}$, conditional on $(\mathbf{g}, \boldsymbol{\eta})$ and $\mathbf{y}^T$.
3. Draw $\boldsymbol{\lambda}$, conditional on $(\boldsymbol{\eta}, \mathbf{g}, \boldsymbol{\rho})$ and $\mathbf{y}^T$.
4. Draw $\boldsymbol{\Phi}$, conditional on $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{g}, \boldsymbol{\rho})$ and $\mathbf{y}^T$.
5. Draw $\mathbf{g}$, conditional on $(\boldsymbol{\Phi}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\rho})$ and $\mathbf{y}^T$.

Iterating on the above steps produces a Markov chain with invariant distribution $\mathbb{P}\left[\boldsymbol{\theta} \mid \mathbf{y}^T\right]$. The algorithm's structure has several convenient attributes. This is a partially collapsed sampler: Blocks 1 and 2 marginalize over $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$, which typically account for most of the model's parameters. Van Dyk and Park (2008) show that a collapsed sampler will have better convergence properties than an ordinary sampler, and they recommend integrating out as many parameters as possible. The distributions in Blocks 1 and 2 are not standard parametric families, so it's necessary to sample $\boldsymbol{\eta}$ and $\boldsymbol{\rho}$ using Metropolis steps. Fortunately, the computational cost of evaluating the posterior kernel is minimal. Furthermore, because $\{\boldsymbol{\rho}_k\}_{k=1}^n$ are assumed to be independent across k under the prior, $\{\boldsymbol{\rho}_k\}_{k=1}^n$ are independent across k under the posterior, conditional on $(\mathbf{g}, \boldsymbol{\eta})$. As Robert and Casella (2004) point out, blocked samplers typically perform best when the blocks are "as independent as possible," while parameters within each block are correlated. I will therefore propose each $\boldsymbol{\rho}_k$ in a separate Metropolis step; doing so will allow me to break up Block 2 into n relatively low-dimensional proposals, rather than a single high-dimensional proposal, without adversely affecting the sampler's convergence. Conveniently, Blocks 3, 4, and 5 only require draws from standard parametric distributions. This algorithm generalizes the samplers proposed by Baumeister and Hamilton (2015) and Villani (2009).[11] Consequently, applied researchers who have already implemented those estimation

[11] Baumeister and Hamilton's (2015) sampling routine applies to the special case where $\mathbf{g}$ and $\boldsymbol{\rho}$ are constrained to be zero. Villani's (2009) sampling routine applies to the special case where $\boldsymbol{\rho}$ is constrained to be zero and the prior over $\boldsymbol{\Psi}'\boldsymbol{\Lambda}\boldsymbol{\Psi}$ is a Wishart distribution.

routines can incorporate seasonality with only modest modifications to their code. Appendix D contains additional computational details for implementing the posterior sampler.

5 Advantages of a Bayesian Approach

For both practical and conceptual reasons, many empirical macroeconomists prefer Bayesian methods for fitting VARs, and virtually all of the advantages of being a Bayesian become more pronounced when working with seasonal time series. Rather than just testing for a single unit root, a frequentist econometrician working with the unadjusted data needs to know how many unit roots and at what locations on the unit circle: It's necessary to test for n_s unit roots, corresponding to the roots of the seasonal differencing operator $(1 - L^{n_s})$. Ghysels et al. (1994) present Monte Carlo evidence that frequentist tests for seasonal unit roots can suffer from low power and finite-sample size distortions, and Hylleberg (1995) shows that the results are often sensitive to ancillary specification choices, such as the number of lags included. When seasonal unit-root tests produce errors, they don't just affect inferences about stochastic seasonality; Abeysinghe (1991, 1994) and Franses et al. (1995) demonstrate how failing to account for seasonal unit roots can lead to spurious inferences about deterministic seasonality.

Unfortunately, different frequentist tests can produce conflicting results: The outcome can depend on whether the econometrician tests for the n_s unit roots individually or jointly, and the order in which the tests are performed. Furthermore, two frequentists may arrive at different conclusions, depending on whether the null hypothesis is the presence of seasonal unit roots or the absence of seasonal unit roots (Canova and Hansen, 1995; Hylleberg, 1995). In that respect, Bayesians and frequentists both make subjective decisions based on the ex ante plausibility of seasonal unit roots, but the Bayesian approach provides a transparent way to articulate and quantify those beliefs. Moreover, Bayesians can apply the same inferential theory to time series with or without unit roots. That's one reason why Sims (1988) argues that econometricians shouldn't fixate on whether an autoregressive root falls exactly on the unit circle, or simply near the unit circle. Although Sims's argument focuses exclusively on zero-frequency unit roots, many of his observations are equally applicable to seasonal unit roots.

Another rationale for applying shrinkage is to reduce estimation noise, and informative priors can be useful with short samples. Season-specific means and shock variances add $2(n_s - 1)n$ extra parameters to an already densely parameterized model. Practically speaking, the definition of a large sample depends more on the number of calendar years than the total number of observations, for the reason mentioned in the introduction: Although a sample with 50 years of monthly data

contains 600 observations, there are only 50 (potentially correlated) observations for each month. Consequently, frequentist estimates of season-specific means can easily be noisy. The problem is likely to be even worse for the season-specific shock variances, because the shocks themselves are estimated, not observed. The smoothness prior introduced in Section 3 effectively pools information across months by positing that adjacent months are likely to be similar.

6 Application: Demand and Supply in Labor Markets

Building on Example 2 from Section 2, I will incorporate seasonality into Baumeister and Hamilton's (2015) model of labor-market demand and supply. The full model is defined by equations (8), (47), and (48). Those three equations, combined with equation (21), imply a demand curve and a supply curve:

$$\Delta \log \left(\text{personhours}_t\right) = c_d + \eta_d \times \Delta \log \left(\text{real wage}_t\right) + \boldsymbol{\delta}_d' \mathbf{w}_t + \boldsymbol{\phi}^d \left(L\right)' \mathbf{y}_t + \epsilon_t^d \tag{62}$$

$$\Delta \log \left(\text{personhours}_t\right) = c_s + \eta_s \times \Delta \log \left(\text{real wage}_t\right) + \boldsymbol{\delta}_s' \mathbf{w}_t + \boldsymbol{\phi}^s \left(L\right)' \mathbf{y}_t + \epsilon_t^s, \tag{63}$$

where $\epsilon_t^d \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \lambda_d^{-1} \exp\left\{-\boldsymbol{\rho}_d' \mathbf{w}_t\right\}\right)$ is a demand shock, $\epsilon_t^s \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \lambda_s^{-1} \exp\left\{-\boldsymbol{\rho}_s' \mathbf{w}_t\right\}\right)$ is a supply shock, and where I have defined:

$$\begin{bmatrix} \boldsymbol{\phi}^d \left(L\right)' \\ \boldsymbol{\phi}^s \left(L\right)' \end{bmatrix} \equiv \sum_{\ell=1}^{m} \boldsymbol{\Phi}_\ell L^\ell, \quad \begin{bmatrix} c_d \\ c_s \end{bmatrix} \equiv \left(\boldsymbol{\Psi} - \sum_{\ell=1}^{m} \boldsymbol{\Phi}_\ell\right) \boldsymbol{\mu}, \quad \begin{bmatrix} \boldsymbol{\delta}_d' \\ \boldsymbol{\delta}_s' \end{bmatrix} \equiv \boldsymbol{\Psi}\mathbf{B} - \sum_{\ell=1}^{m} \boldsymbol{\Phi}_\ell \mathbf{B}\mathbf{R}^\ell. \tag{64}$$

Because they use seasonally adjusted data, Baumeister and Hamilton effectively assume that $\boldsymbol{\delta}_d$, $\boldsymbol{\delta}_s$, $\boldsymbol{\rho}_d$, and $\boldsymbol{\rho}_s$ are zero. The latter two restrictions are relevant for identification. Setting $\boldsymbol{\rho}_d$ and $\boldsymbol{\rho}_s$ to zero renders Baumeister and Hamilton's version of the model only partially identified, but based on the argument presented in Section 2.2, my version of the model will be fully identified if $\boldsymbol{\rho}_d \neq \boldsymbol{\rho}_s$.

I will analyze two sets of estimates. First, I will use seasonally unadjusted data to fit equations (62) and (63); I will refer to this as the *seasonal model*. Second, I will use the seasonally adjusted versions of wage growth and hours growth to fit equations (62) and (63), while restricting $\boldsymbol{\delta}_d$, $\boldsymbol{\delta}_s$, $\boldsymbol{\rho}_d$, and $\boldsymbol{\rho}_s$ to be zero; I will refer to this as the *seasonally adjusted model*. That label is a bit of a misnomer: In the seasonally adjusted model, it's the data that have been seasonally adjusted, whereas the model itself is aseasonal. However, one of the points that I want to highlight is that the results change, depending on whether one uses seasonally adjusted or unadjusted time series, while holding fixed all assumptions about the non-seasonal aspects of the model.

6.1 Prior

Beliefs About Demand and Supply Elasticities. I will adopt the same priors as Baumeister and Hamilton (2015) for η_d and η_s. The prior for η_d is a Student's t distribution, truncated such that $\eta_d < 0$, with location parameter $-.6$, scale parameter .6, and 3 degrees of freedom. The prior for η_s is a Student's t distribution, truncated such that $\eta_s > 0$, with location parameter .6, scale parameter .6, and 3 degrees of freedom. Baumeister and Hamilton's rationale, based on their review of the literature, is to adopt a compromise between micro and macro estimates of these parameters: Microeconomists tend to find that both demand and supply are fairly inelastic in labor markets, whereas macroeconomists often favor larger elasticities. The modal values of η_d and η_s are in the middle of the range of earlier estimates, and the priors are sufficiently diffuse to put non-trivial weight on the values favored by both micro studies and macro studies.

Beliefs About Seasonality. I will set $\alpha = \frac{1}{2}$. With monthly data, this implies that κ_u is positive for $u = \pm 1$ and approximately zero for $u = \pm 2$. In other words, $\mathbf{s}_t$ in any calendar month is expected to be positively correlated with the preceding calendar month and the following calendar month. I will set $\mathbf{V}_S = \varsigma \bar{\mathbf{\Sigma}}_y$ with $\varsigma = .3$ and $\bar{\mathbf{\Sigma}}_y$ equal to the sample variance of $\mathbf{y}_t$. This specification means that the variance of $\mathbf{s}_t$ is expected to be about 30% of the total variance of $\mathbf{y}_t$. Other authors, such as Beaulieu et al. (1992), argue that deterministic seasonality accounts for an even higher fraction of the variance of many macroeconomic time series; however, their results may overstate the role of deterministic seasonality if there are seasonal unit roots, as Franses et al. (1995) explain. I will set $\nu_\rho = .3$, so the variance of the log diagonal of $\mathbf{\Lambda}_t$ is expected to be .3 over the course of the year. I will set $\tau_S = 1$, so there is effectively one dummy observation for each constraint used to implement the seasonal unit roots at the annual periodicity.

Beliefs About Non-Seasonal Behavior. I will set the lag order of the VAR to $m = 13$ for monthly data. The logged series for hours and real wages will be multiplied by 100 so that the elements of $\mathbf{y}_t$ can be interpreted as approximate growth rates. I will set $\bar{\boldsymbol{\mu}} = \left(\frac{1}{12}, \frac{2}{12}\right)'$, so the annual growth rates of real wages and aggregate hours are expected to be about 1% and 2%, respectively. I will set $\mathbf{V}_\mu$ to be diagonal, with $(\mathbf{V}_\mu)_{1,1} = \left(\frac{.5}{12}\right)^2$ and $(\mathbf{V}_\mu)_{2,2} = \left(\frac{1}{12}\right)^2$, so the prior means are two prior standard deviations from zero. The dummy observations $\bar{\mathbf{X}}_0$ and $\bar{\mathbf{Y}}_0$ that appear in equation (42) reflect standard priors that are commonly applied in the literature to seasonally adjusted data.

Specifically:

$$\bar{\mathbf{Y}}_0 = \begin{bmatrix} \mathbf{0}_{nm\times n} \\ \mathbf{1}_{\tau_\sigma\times 1} \otimes \operatorname{diag}\left(\hat{\sigma}_1, \ldots, \hat{\sigma}_n\right)^{-1} \end{bmatrix}, \quad \bar{\mathbf{X}}_0 = \begin{bmatrix} \frac{1}{\tau_0} \operatorname{diag}\left(1, \ldots, m\right) \otimes \operatorname{diag}\left(\hat{\sigma}_1, \ldots, \hat{\sigma}_n\right)^{-1} \\ \mathbf{0}_{n\times nm} \end{bmatrix}, \tag{65}$$

where $\hat{\sigma}_k$ is the standard deviation of the residuals from a univariate autoregression, augmented with seasonal dummy variables, fit to $\mathbf{y}_{k,t}$. (For the seasonally adjusted model, $\hat{\sigma}_k$ is the residual standard deviation from a univariate autoregression, without seasonal dummy variables, fit to the kth seasonally adjusted variable.) If not for the additional dummy observations used to implement beliefs about seasonal unit roots, the first nm rows of $\bar{\mathbf{Y}}_0$ and $\bar{\mathbf{X}}_0$ would implement a prior belief that the reduced-form autoregressive coefficients have mean zero, with the variance of the ℓth lag proportional to $1/\ell^2$; the latter $\tau_\sigma n$ rows of $\bar{\mathbf{Y}}_0$ and $\bar{\mathbf{X}}_0$ would implement a prior belief that the reduced-form innovation variances are well approximated by the innovation variances of n univariate models. The parameter $\tau_0 > 0$ controls the overall tightness of the baseline prior on the autoregressive coefficients; following Litterman (1986), I set $\tau_0 = .2$. The hyperparameter $\tau_\sigma \in \mathbb{Z}_+$ controls the confidence in the baseline prior's beliefs about the innovation variances; following Baumeister and Hamilton (2015), I set $\tau_\sigma = 2$. For the seasonally adjusted model, I exclude the dummy observations that favor seasonal unit roots, in which case $\bar{\mathbf{Y}} = \bar{\mathbf{Y}}_0$ and $\bar{\mathbf{X}} = \bar{\mathbf{X}}_0$.

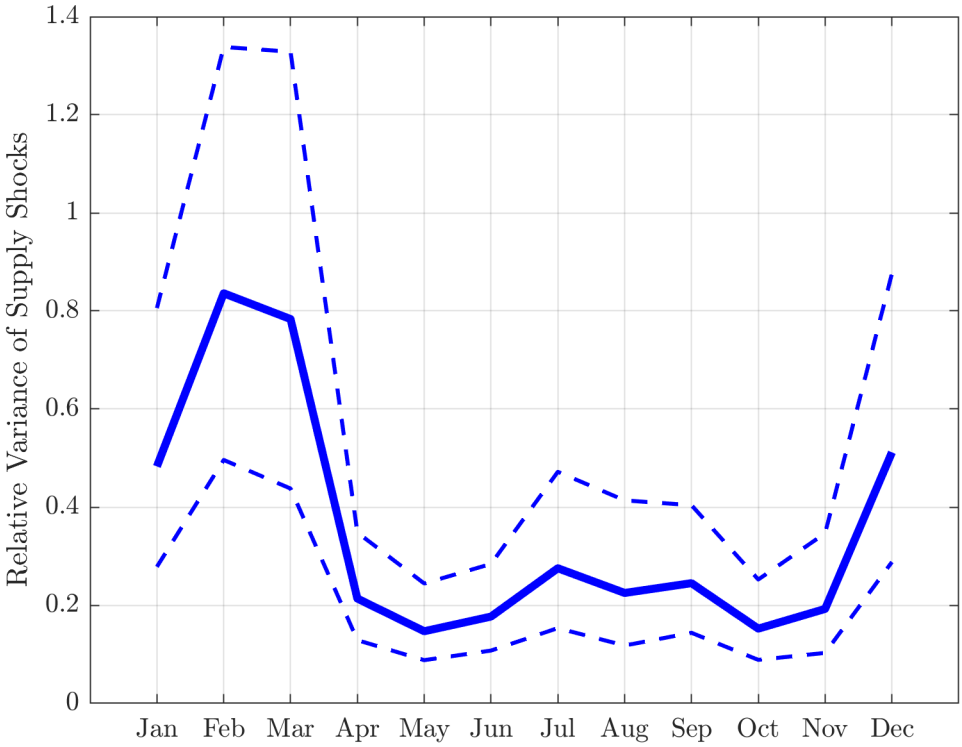
6.2 Data

The series for personhours is the index of aggregate hours of production and non-supervisory employees in the United States, and the series for the real wage is the average hourly earnings of those workers, deflated by the Consumer Price Index.[12] After taking log differences, both series are multiplied by 100. The sample period is January 1964 to December 2019. I have excluded the pandemic-era data to show how seasonal adjustment can affect the results of a structural VAR under "normal" circumstances, so the results will not be driven by the 2020 outliers (nor the Bureau of Labor Statistics's attempts to account for those outliers in the seasonal adjustment routine).[13]

[12] The data are downloaded from FRED (fred.stlouisfed.org). The FRED codes for the non-seasonally adjusted data on hours, nominal wages, and prices are CEU0500000034, CEU0500000008, and CPIAUCNS. The codes for the seasonally adjusted counterparts are AWHI, AHETPI, and CPIAUCSL. Baumeister and Hamilton (2015) measure real wages using the index of real hourly compensation in the nonfarm business sector (COMPRNFB), but the Bureau of Labor Statistics only releases a seasonally adjusted version of that series, not a seasonally unadjusted version.

[13] The Bureau of Labor Statistics took a number of steps to try to account for the extreme outliers in the 2020 data when performing seasonal adjustment; in particular, Bureau statisticians did not want their seasonal adjustment routines to attribute the abrupt recession in March and April of 2020 to seasonal factors (Bureau of Labor Statistics, 2020). However, as Lucca and Wright (2021) point out, "there are no easy answers to seasonal adjustment in this [pandemic] environment," and it may take more time for the Bureau to discern how to account for the extreme observations in seasonal-adjustment methods.

Figure 4: Seasonal Heteroskedasticity



Notes: Estimates of $\mathbb{V}\left[\epsilon_t^s\right]/\mathbb{V}\left[\epsilon_t^d\right]$, the relative variance of supply shocks, as a function of the calendar month. The solid line is the posterior median; dashed lines are the 10th and 90th posterior quantiles.

6.3 Results

Whereas the seasonally adjusted model is only partially identified, the seasonal model is fully identified as long as $\boldsymbol{\rho}_s \neq \boldsymbol{\rho}_d$, meaning that the relative variances of the shocks change over the course of the year. The prior is conservative in the sense that the most likely values of $\boldsymbol{\rho}_s$ and $\boldsymbol{\rho}_d$ are zero, corresponding to homoskedastic shocks. The data, however, contain evidence of seasonal heteroskedasticity. Figure 4 plots $\mathbb{V}\left[\epsilon_t^s\right]/\mathbb{V}\left[\epsilon_t^d\right]$, the relative variance of supply shocks, as a function of the calendar month. Supply shocks are relatively more volatile in the winter, so wintertime variation in the data helps trace out the slope of the demand curve. Conversely, summertime variation in the data is more helpful for tracing out the slope of the supply curve. Attempting to remove seasonality (or simply ignoring it) therefore discards useful identifying information. With the seasonally adjusted model, Baumeister and Hamilton (2015) show that an econometrician will never learn the model's structural parameters, even with an infinite amount of data. This fact likely contributes to the empirical finding that, in many of the results below, the seasonal model produces more precise estimates than the seasonally adjusted model.

Table 1: Estimated Structural Parameters

	η_d	η_s	λ_d^{-1}	λ_s^{-1}
Seasonal Model	-2.57 $[-3.32, -2.03]$	1.34 $[1.16, 1.54]$	1.92 $[1.46, 2.79]$	0.57 $[0.50, 0.65]$
Seasonally Adjusted Model	-1.36 $[-2.22, -0.81]$	1.38 $[0.85, 2.32]$	0.49 $[0.33, 0.90]$	0.49 $[0.33, 0.93]$

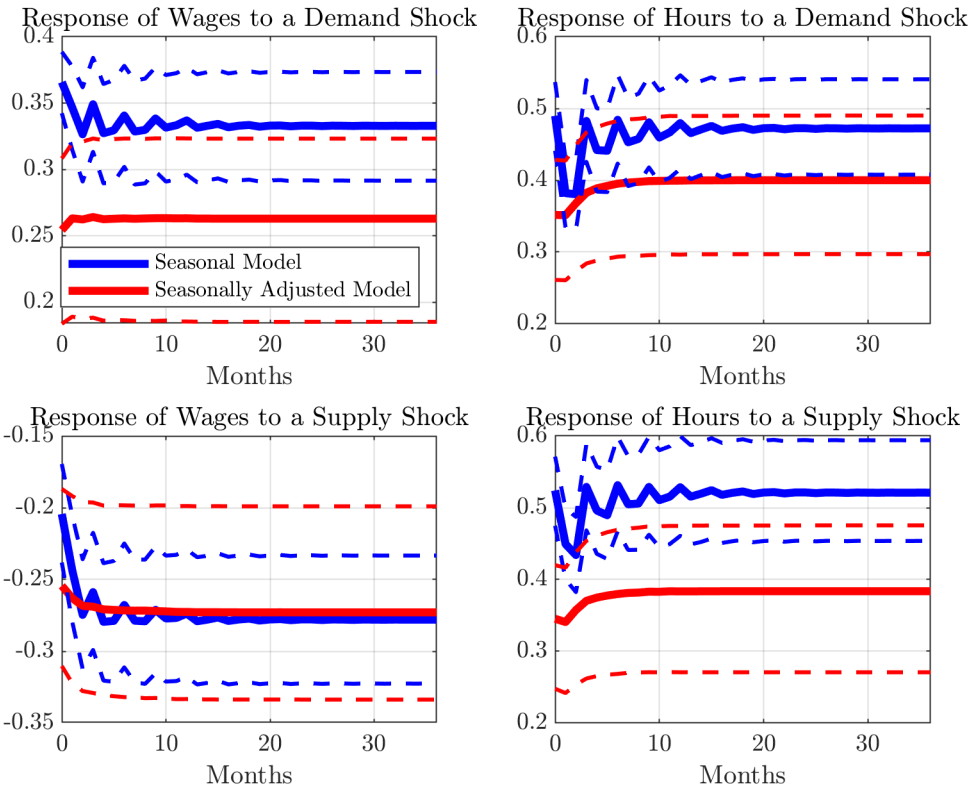
Notes: Point estimates are posterior medians. Numbers in brackets are the 10th and 90th posterior quantiles.

To show how seasonal adjustment can bias the estimates of the structural parameters, Table 1 displays the estimates of $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$. The seasonal model's point estimate for η_d is about 89% greater in magnitude than the point estimate produced by the seasonally adjusted model. This difference is large, both economically and statistically. Earlier studies have generated debate about plausible elasticities for labor demand. Cooley and Prescott (1995) provide a canonical example of how to calibrate a DSGE model to aggregate data, and their approach implies that a representative firm has a labor-demand elasticity of -2.5.[14] In contrast, Hamermesh (1996), in a survey of the applied micro literature, finds estimates for the demand elasticity between $-.15$ and $-.75$. The seasonal model produces an estimate for η_d that is similar to Cooley and Prescott's number, despite taking a completely different measurement strategy. The seasonally adjusted model produces an estimate of $|\eta_d|$ that is somewhat higher than the range suggested by applied micro studies, but far below the values common in DSGE models. Furthermore, the seasonal model assigns only .03% probability to $|\eta_d|$ being as low as 1.36, the posterior median of the seasonally adjusted model. Besides suggesting that demand is much more sensitive to wages, the seasonal model also suggests that demand is much more volatile: The seasonal model's point estimate for λ_d^{-1} is nearly four times larger than the seasonally adjusted point estimate. On the supply side, the seasonal and seasonally adjusted models produce more comparable point estimates for η_s and λ_s^{-1}, but the seasonal model produces much narrower credible sets.

Figure 5 shows how seasonal adjustment changes both the size and the shape of the impulse responses. In both models, by assumption, the contemporaneous effect of a positive labor-demand shock is to raise hours and wages, while the contemporaneous effect of a positive labor-supply shock is to raise hours and depress wages. Note, however, the magnitudes of impulse responses on impact: At horizon zero, the seasonal model's point estimate is about 44% larger than the seasonally adjusted model's point estimate for the response of wages to a demand shock, 20% smaller for the response

[14] Cooley and Prescott consider a representative firm with the production function $\text{output}_t = \text{TFP}_t \times (\text{hours}_t)^{1-\vartheta} (\text{capital}_t)^{\vartheta}$, which implies a labor-demand elasticity of $-1/\vartheta$. Based on income shares from the national accounts, the authors calibrate $\vartheta = .40$. Countless papers follow Cooley and Prescott's calibration strategy and choose ϑ between .30 and .40, and many estimated DSGE models adopt the prior that ϑ is in this range.

Figure 5: Impulse Responses



Notes: Estimates of how the log levels of real wages and aggregate hours, multiplied by 100, respond to one-standard-deviation shocks to demand and supply. Solid lines are posterior median estimates; dashed lines are the 10th and 90th posterior quantiles.

of wages to a supply shock, 39% larger for the response of hours to a demand shock, and 52% larger for the response of hours to a supply shock. For each short-run response, the seasonally adjusted model's point estimate falls outside the seasonal model's credible set. In the medium run, the models show different dynamics for how the shocks are propagated: The seasonal model's impulse responses exhibit oscillations, whereas the seasonally adjusted model's impulse responses are smooth. Finally, for three of the four impulse responses, there is a large discrepancy between the long-run effects implied by the seasonal model and those implied by the seasonally adjusted model.

Seasonal adjustment alters the relative importance of supply shocks and demand shocks when accounting for the variance of the data, especially for wage growth. The posterior probability of $\mathbf{A}(L)$ having explosive roots is essentially zero, so it's possible to compute the unconditional variance of $\tilde{\mathbf{y}}_t$.[15] Table 2 displays the fraction of the variance that can be attributed to supply shocks for each

[15]For both the seasonal model and the seasonally adjusted model, each one of the posterior draws generated by the MCMC routine had the property that all roots of the polynomial $|\mathbf{A}(L)|$ were outside the complex unit circle. Consequently, had I truncated the conditional prior of $\mathbf{\Phi}$ to the stationary region of the parameter space, the results would look identical. I will therefore proceed under the assumption that $\mathbf{A}(L)$ is non-explosive with probability one,

Table 2: Percentage of Variance Due to Supply Shocks

	Wage Growth		Hours Growth	
	Seasonal	Seasonally Adjusted	Seasonal	Seasonally Adjusted
Unconditional Variance	25 [18, 33]	50 [27, 74]	52 [44, 61]	49 [25, 72]
Low Frequencies	41 [29, 54]	52 [28, 76]	55 [45, 65]	48 [24, 71]
Business-Cycle Frequencies	39 [29, 50]	51 [28, 75]	55 [46, 65]	48 [25, 71]
Irregular Frequencies	23 [17, 31]	50 [27, 73]	52 [43, 61]	49 [26, 72]

Notes: Point estimates are posterior median estimates, rounded to the nearest integer, for the percentage of the variance of $\tilde{\mathbf{y}}_t$ attributable to supply shocks within each frequency band. Numbers in brackets are the 10th and 90th posterior quantiles. Business-cycle frequencies refer to periodicities between 1.5 and 8 years. Low frequencies (irregular frequencies) are all periodicities longer (shorter) than business cycles.

variable, after removing deterministic seasonality. The seasonal model implies that demand shocks are three times more important than supply shocks in accounting for the unconditional variance of the stochastic component of wage growth; in contrast, the seasonally adjusted model implies that the two shocks are equally important. Qualitatively, this pattern appears across all frequency bands for wage growth. Hence, by looking at seasonally adjusted variables, a researcher who is interested in business cycles would fail to detect that the majority of the variation in wage growth at business-cycle frequencies stems from demand shocks. For the variance decomposition of hours growth, the discrepancy in point estimates between the seasonal model and the seasonally adjusted model is less pronounced, but the variance shares from the seasonal model are much more precisely estimated.

7 Conclusion

When fitting structural VARs, macroeconomists should always prefer seasonally unadjusted variables over the seasonally adjusted versions. Even if a researcher is using a structural VAR to analyze a non-seasonal phenomenon, that's not a valid reason for using seasonally adjusted time series. Unfortunately, in some contexts, researchers may need to use seasonally adjusted variables, because government statistical agencies give no choice: In the U.S., some official statistics are released exclusively in seasonally adjusted form, or the unadjusted versions are only available over very short time spans. One avenue of future research could be to assemble unadjusted versions of widely

so I can treat the variance of $\tilde{\mathbf{y}}_t$ as finite.

used economic indicators. Another could be to reassess empirical results in the structural VAR literature using unadjusted data. In either case, seasonality demands more attention from empirical macroeconomists.

References

Abeysinghe, T. (1991). Inappropriate use of seasonal dummies in regression. *Economics Letters 36*(2), 175–179.

Abeysinghe, T. (1994). Deterministic seasonal models and spurious regressions. *Journal of Econometrics 61*(2), 259–272.

Baumeister, C. and J. D. Hamilton (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica 83*(5), 1963–1999.

Baumeister, C. and J. D. Hamilton (2018). Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations. *Journal of Monetary Economics 100*, 48 – 65.

Baumeister, C. and J. D. Hamilton (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review 109*(5), 1873–1910.

Beaulieu, J. J., J. K. MacKie-Mason, and J. A. Miron (1992). Why do countries and industries with large seasonal cycles also have large business cycles? *The Quarterly Journal of Economics 107*(2), 621–656.

Bell, W. and B. Monsell (1992). X-11 symmetric linear filters and their transfer functions. Technical report, Bureau of the Census. Research Report n. RR 92: 15. https://www.census.gov/srd/papers/pdf/rr92-15.pdf.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021). Feedbacks: Financial markets and economic activity. *American Economic Review 111*(6), 1845–79.

Bureau of Economic Analysis (2015). A Snapshot of the Seasonal Adjustment Process for GDP. *The BEA Wire*. June 10, 2015. https://www.bea.gov/news/blog/2015-06-10/snapshot-seasonal-adjustment-process-gdp.

Bureau of Labor Statistics (2020). The Challenges of Seasonal Adjustment during the COVID-19 Pandemic. *Comminssioner's Corner*, December 21, 2020.

https://blogs.bls.gov/blog/2020/12/21/the-challenges-of-seasonal-adjustment-during-the-covid-19-pandemic/.

Canova, F. (1992). An alternative approach to modeling and forecasting seasonal time series. *Journal of Business & Economic Statistics 10*(1), 97–108.

Canova, F. (1993). Forecasting time series with common seasonal patterns. *Journal of Econometrics 55*(1-2), 173–200.

Canova, F. and B. E. Hansen (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics 13*(3), 237–252.

Christiano, L. J. and R. M. Todd (2002). The conventional treatment of seasonality in business cycle analysis: Does it create distortions? *Journal of Monetary Economics 49*(2), 335–364.

Cooley, T. F. and E. C. Prescott (1995). Economic Growth and Business Cycles. Published in *Frontiers of Business Cycle Research*, ed. Thomas F. Cooley and Edward C. Prescott, Princeton University Press, pp. 1–38.

Del Negro, M. and F. Schorfheide (2011). Bayesian Macroeconometrics. Published in *The Oxford Handbook of Bayesian Econometrics*, ed. J. Geweke, G. Koop, and H. Van Dijk, Oxford University Press, pp. 293–389.

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

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 251–276.

Franses, P. H., H. Hoek, and R. Paap (1997). Bayesian analysis of seasonal unit roots and seasonal mean shifts. *Journal of Econometrics 78*(1), 359–380.

Franses, P. H., S. Hylleberg, and H. S. Lee (1995). Spurious deterministic seasonality. *Economics Letters 48*(3-4), 249–256.

Gersovitz, M. and J. G. MacKinnon (1978). Seasonality in regression: An application of smoothness priors. *Journal of the American Statistical Association 73*(362), 264–273.

Ghysels, E. (1988). A study toward a dynamic theory of seasonality for economic time series. *Journal of the American Statistical Association 83*(401), 168–172.

Ghysels, E., H. S. Lee, and J. Noh (1994). Testing for unit roots in seasonal time series: Some theoretical extensions and a Monte Carlo investigation. *Journal of Econometrics 62*(2), 415–442.

Ghysels, E., D. R. Osborn, and P. M. Rodrigues (2006). Forecasting seasonal time series. *Handbook of Economic Forecasting 1*, 659–711.

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

Giannone, D., M. Lenza, and G. E. Primiceri (2019). Priors for the long run. *Journal of the American Statistical Association 114*(526), 565–580.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Hamermesh, D. S. (1996). *Labor Demand*. Princeton University Press.

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons.

Hansen, L. P. and T. J. Sargent (1993). Seasonality and approximation errors in rational expectations models. *Journal of Econometrics 55*(1-2), 21–55.

Hylleberg, S. (1995). Tests for seasonal unit roots General to specific or specific to general? *Journal of Econometrics 69*(1), 5–25.

Hylleberg, S., R. F. Engle, C. W. Granger, and B. S. Yoo (1990). Seasonal integration and cointegration. *Journal of Econometrics 44*(1-2), 215–238.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method*, Volume 158 of *Lecture Notes in Statistics*. Springer Science & Business Media.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Lucca, D. and J. Wright (2021). Reasonable Seasonals? Seasonal Echoes in Economic Data after COVID-19. Federal Reserve Bank of New York, *Liberty Street Economics*, March 25, 2021. https://libertystreeteconomics.newyorkfed.org/2021/03/reasonable-seasonals-seasonal-echoes-in-economic-data-after-covid-19.html.

Miron, J. and J. J. Beaulieu (1996). What have macroeconomists learned about business cycles from the study of seasonal cycles? *Review of Economics and Statistics 78*(1), 54–66.

Raynauld, J. and J.-G. Simonato (1993). Seasonal BVAR models: A search along some time domain priors. *Journal of Econometrics 55*(1-2), 203–229.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

Rigobon, R. and B. Sack (2003). Measuring the reaction of monetary policy to the stock market. *The Quarterly Journal of Economics 118*(2), 639–669.

Rigobon, R. and B. Sack (2004). The impact of monetary policy on asset prices. *Journal of Monetary Economics 51*(8), 1553–1575.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer.

Saijo, H. (2013). Estimating DSGE models using seasonally adjusted and unadjusted data. *Journal of Econometrics 173*(1), 22–35.

Sargent, T. J. (1978). Comments on "Seasonal Adjustment and Multiple Time Series Analysis," by K. F. Wallis. In A. Zellner (Ed.), *Seasonal Analysis of Economic Time Series*, pp. 361–364. U.S. Department of Commerce, Bureau of the Census.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control 12*(2-3), 463–474.

Sims, C. A. (1993a). A Nine-Variable Probabilistic Macroeconomic Forecasting Model. Published in *Business Cylces, Indicators, and Forecasting*, ed. James H. Stock and Mark W. Watson, University of Chicago Press, pp. 179-212.

Sims, C. A. (1993b). Rational expectations modeling with seasonally adjusted data. *Journal of Econometrics 55*(1-2), 9–19.

Theil, H. and A. S. Goldberger (1961). On pure and mixed statistical estimation in economics. *International Economic Review 2*(1), 65–78.

Van Dyk, D. A. and T. Park (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association 103*(482), 790–796.

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics 24*(4), 630–650.

Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound? *The Economic Journal 122*(564), F447–F466.

Young, A. H. (1968). Linear approximations to the Census and BLS seasonal adjustment methods. *Journal of the American Statistical Association 63*(322), 445–471.

A Proofs of Propositions

A.1 Proof of Proposition 1

This is a proof by contradiction. Suppose $(\mathbf{\Psi}, \mathbf{\Lambda}) \in \mathcal{I}(\mathbf{Q})$ and $(\mathbf{\Psi}, \mathbf{\Lambda}) \in \mathcal{I}(\mathbf{Q}^{sa})$ (or, in the case where the model is fully identified, $(\mathbf{\Psi}, \mathbf{\Lambda}) = \mathcal{I}(\mathbf{Q})$ and $(\mathbf{\Psi}, \mathbf{\Lambda}) = \mathcal{I}(\mathbf{Q}^{sa})$). That would imply $\mathbf{Q} = \mathbf{\Psi}'\mathbf{\Lambda}\mathbf{\Psi} = \mathbf{Q}^{sa}$, implying $|\mathbf{Q}| = |\mathbf{Q}^{sa}|$, which contradicts $|\mathbf{Q}^{sa}| = D^n |\mathbf{Q}|$, because $D \neq 1$.

A.2 Proof of Proposition 2

In light of equation (22), the prior expectation of the uth sample autocovariance of $\mathbf{s}_t$ is equal to the prior expectation of $\mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}'$. Observe that the (j,k) element of this matrix is given by:

$$(\mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}')_{j,k} = \sum_{\ell=1}^{n_s-1}\sum_{h=1}^{n_s-1} \mathbf{B}_{j,h}\left(\mathbf{R}^{u\prime}\right)_{h,\ell}\left(\mathbf{B}'\right)_{\ell,k}. \tag{66}$$

Lemma 1 establishes that the (h,ℓ) element of $\mathbf{R}^{u\prime}$ is:

$$\left(\mathbf{R}^{u\prime}\right)_{h,\ell} = \left(\mathbf{R}^{u}\right)_{\ell,h} = \cos\left(\frac{2\pi}{n_s}\ell u\right)\mathbb{I}\left[\ell = h\right] - \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right)\mathbb{I}\left[h = n_s - \ell\right]. \tag{67}$$

The expression for $\mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}'$ therefore reduces to:

$$\begin{aligned}(\mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}')_{j,k} &= \sum_{\ell=1}^{n_s-1}\sum_{h=1}^{n_s-1}\mathbf{B}_{j,h}\mathbf{B}_{k,\ell}\left(\cos\left(\frac{2\pi}{n_s}\ell u\right)\mathbb{I}\left[\ell = h\right] - \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right)\mathbb{I}\left[h = n_s - \ell\right]\right)\\ &= \sum_{\ell=1}^{n_s-1}\mathbf{B}_{j,\ell}\mathbf{B}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u\right) - \sum_{\ell=1}^{n_s-1}\mathbf{B}_{j,n_s-\ell}\mathbf{B}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right).\end{aligned} \tag{68}$$

In turn, I will show that prior expectation of the first sum in the above expression is equal to the (j,k) element of $\kappa_u\mathbf{V}_S$, and then I will show that the prior expectation of the second sum is zero.

The Kronecker structure of the variance in equation (23) implies that $\mathbb{E}_{prior}\left[\mathbf{B}_{j,\ell}\mathbf{B}_{k,\ell}\right] = \mathbf{K}_{\ell,\ell}\left(\mathbf{V}_S\right)_{j,k}$.

Hence:

$$\begin{aligned}\mathbb{E}_{prior}\left[\sum_{\ell=1}^{n_s-1}\mathbf{B}_{j,\ell}\mathbf{B}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u\right)\right] &= \sum_{\ell=1}^{n_s-1}\mathbb{E}_{prior}\left[\mathbf{B}_{j,\ell}\mathbf{B}_{k,\ell}\right]\cos\left(\frac{2\pi}{n_s}\ell u\right)\\ &= \left[\sum_{\ell=1}^{n_s-1}\mathbf{K}_{\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u\right)\right]\left(\mathbf{V}_S\right)_{j,k}, \qquad (69)\end{aligned}$$

and the term in square brackets in the final expression coincides with the definition of κ_u in the statement of the proposition.

Equation (23) implies that $\mathbb{E}_{prior}\left[\mathbf{B}_{j,n_s-\ell}\mathbf{B}_{k,\ell}\right] = \mathbf{K}_{n_s-\ell,\ell}\left(\mathbf{V}_S\right)_{j,k}$, so:

$$\begin{aligned}\mathbb{E}_{prior}\left[\sum_{\ell=1}^{n_s-1}\mathbf{B}_{j,n_s-\ell}\mathbf{B}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\right] &= \sum_{\ell=1}^{n_s-1}\mathbb{E}_{prior}\left[\mathbf{B}_{j,n_s-\ell}\mathbf{B}_{k,\ell}\right]\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\\ &= \left[\sum_{\ell=1}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\right]\left(\mathbf{V}_S\right)_{j,k}. \qquad (70)\end{aligned}$$

I will show that the term in square brackets on the last line of the above expression is zero, because the matrix $\mathbf{K}$ must be symmetric. For now, assume that n_s is odd; I will return momentarily to the case where n_s is even. With n_s being odd, we can break the sum up into terms corresponding to $\ell < \frac{n_s}{2}$ and $\ell > \frac{n_s}{2}$:

$$\begin{aligned}\sum_{\ell=1}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right) &= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\sum_{\ell=\frac{n_s+1}{2}}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\\ &= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\\ &\quad+\sum_{\ell'=1}^{\frac{n_s-1}{2}}\mathbf{K}_{\ell',n_s-\ell'}\cos\left(\frac{2\pi}{n_s}\left(n_s-\ell'\right)u-\frac{\pi}{2}\right)\\ &= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\sum_{\ell'=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell',\ell'}\cos\left(\frac{2\pi}{n_s}\ell' u+\frac{\pi}{2}\right)\\ &= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\left[\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\cos\left(\frac{2\pi}{n_s}\ell u+\frac{\pi}{2}\right)\right]\\ &= 0, \qquad (71)\end{aligned}$$

where the second equality replaces ℓ with $\ell' \equiv n_s - \ell$; the third equality uses the symmetry of $\mathbf{K}$ and the fact that $\cos\left(\frac{2\pi}{n_s}\left(n_s-\ell'\right)u-\frac{\pi}{2}\right) = \cos\left(\frac{2\pi}{n_s}\ell' u+\frac{\pi}{2}\right)$ for any integers ℓ' and u; the fourth equality consolidates terms across the sums; and the final equality uses the fact that $\cos\left(\omega-\frac{\pi}{2}\right)+$

$\cos\left(\omega + \frac{\pi}{2}\right) = 0$ for any ω. Now, assume that n_s is even. The only part of the preceding argument that needs modification is that the sum $\sum_{\ell=1}^{n_s-1} \mathbf{K}_{n_s-\ell,\ell} \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right)$ contains a term with $\ell = \frac{n_s}{2}$, in addition to the terms with $\ell < \frac{n_s}{2}$ and $\ell > \frac{n_s}{2}$ that appeared earlier. However, when $\ell = \frac{n_s}{2}$:

$$\mathbf{K}_{n_s-\ell,\ell} \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right) = \mathbf{K}_{\frac{n_s}{2},\frac{n_s}{2}} \cos\left(\pi u - \frac{\pi}{2}\right) = 0, \tag{72}$$

where the latter equality comes from the fact that $\cos\left(\pi u - \frac{\pi}{2}\right) = 0$ for any $u \in \mathbb{Z}$. Hence, with n_s being even, all other steps in equation (71) remain valid, except with $\frac{n_s}{2} - 1$ replacing $\frac{n_s-1}{2}$ in the limits of summation.

A.3 Proof of Proposition 3

First, I will derive the joint posterior kernel of $(\boldsymbol{\Phi}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$ in order to characterize the conditional posterior of $(\boldsymbol{\lambda}, \boldsymbol{\Phi})$ and the marginal posterior kernel of $(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$. Lemma 2 establishes that the likelihood can be written as:

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= (2\pi)^{-\frac{Tn}{2}} \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^{T} \mathbf{w}_t\right)\right\} \left(\prod_{k=1}^{n} \lambda_k^{\frac{T}{2}}\right) \\
&\quad \times \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2} \boldsymbol{\psi}_k' \dot{\mathbf{Y}}_{(k)}' \dot{\mathbf{Y}}_{(k)} \boldsymbol{\psi}_k\right\} \\
&\quad \times \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2} \left(-2\boldsymbol{\phi}_k' \dot{\mathbf{X}}_{(k)}' \dot{\mathbf{Y}}_{(k)} \boldsymbol{\psi}_k + \boldsymbol{\phi}_k' \dot{\mathbf{X}}_{(k)}' \dot{\mathbf{X}}_{(k)} \boldsymbol{\phi}_k\right)\right\},
\end{aligned} \tag{73}$$

where the tth rows of $\dot{\mathbf{Y}}_{(k)}$ and $\dot{\mathbf{X}}_{(k)}$ are $\exp\left\{\frac{1}{2}\boldsymbol{\rho}_k' \mathbf{w}_t\right\} \tilde{\mathbf{y}}_t'$ and $\exp\left\{\frac{1}{2}\boldsymbol{\rho}_k' \mathbf{w}_t\right\} \tilde{\mathbf{x}}_t'$. Lemma 3 establishes that the conditional prior for $\boldsymbol{\Phi}$ and $\boldsymbol{\lambda}$ can be written as:

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] &= \Gamma\left(\alpha_\lambda\right)^{-n} (2\pi)^{-\frac{n^2 m}{2}} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left(\prod_{k=1}^{n} \beta_{\lambda,k}^{\alpha_\lambda}\right) \\
&\quad \times \left(\prod_{k=1}^{n} \lambda_k^{\frac{\bar{T}}{2}}\right) \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2} \boldsymbol{\psi}_k' \bar{\mathbf{Y}}'\bar{\mathbf{Y}} \boldsymbol{\psi}_k\right\} \\
&\quad \times \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2} \left(\boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{X}} \boldsymbol{\phi}_k - 2\boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{Y}} \boldsymbol{\psi}_k\right)\right\}.
\end{aligned} \tag{74}$$

The product of $\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right]$ and $\mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right]$ can therefore be written as:

$$\begin{aligned}
&\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \times \\
&\mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] \quad = \quad \Gamma\left(\alpha_\lambda\right)^{-n} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left(\prod_{k=1}^{n} \beta_{\lambda,k}^{\alpha_\lambda}\right) \\
&\qquad \times (2\pi)^{-\frac{Tn}{2}} \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^{T} \mathbf{w}_t\right)\right\} \\
&\qquad \times \left(\prod_{k=1}^{n} \lambda_k^{\frac{T+\bar{T}}{2}}\right) \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2} \boldsymbol{\psi}_k' \mathbf{Y}_{(k)}' \mathbf{Y}_{(k)} \boldsymbol{\psi}_k\right\} \\
&\qquad \times (2\pi)^{-\frac{n^2 m}{2}} \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2}\left(-2\boldsymbol{\phi}_k' \mathbf{X}_{(k)}' \mathbf{Y}_{(k)} \boldsymbol{\psi}_k + \boldsymbol{\phi}_k' \mathbf{X}_{(k)}' \mathbf{X}_{(k)} \boldsymbol{\phi}_k\right)\right\}, \quad (75)
\end{aligned}$$

where $\mathbf{Y}_{(k)}$ is the vertical concatenation of $\dot{\mathbf{Y}}_{(k)}$ and $\bar{\mathbf{Y}}$, and $\mathbf{X}_{(k)}$ is the vertical concatenation of $\dot{\mathbf{X}}_{(k)}$ and $\bar{\mathbf{X}}$. Completing the square for each quadratic function of $\boldsymbol{\phi}_k$ that appears in the last line, the above is algebraically equivalent to:

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] \quad = \quad & \Gamma\left(\alpha_\lambda\right)^{-n} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left(\prod_{k=1}^{n} \beta_{\lambda,k}^{\alpha_\lambda}\right) \\
& \times (2\pi)^{-\frac{Tn}{2}} \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^{T} \mathbf{w}_t\right)\right\} \\
& \times \left(\prod_{k=1}^{n} \lambda_k^{\frac{T+\bar{T}}{2}}\right) \exp\left\{-\sum_{k=1}^{n} \lambda_k \hat{\beta}_{\lambda,k}\right\} \\
& \times (2\pi)^{-\frac{n^2 m}{2}} \exp\left\{-\sum_{k=1}^{n} \frac{\lambda_k}{2}\left(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k\right)' \mathbf{X}_{(k)}' \mathbf{X}_{(k)} \left(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k\right)\right\}, \quad (76)
\end{aligned}$$

where I have invoked the definitions of $\hat{\boldsymbol{\phi}}_k$ and $\hat{\beta}_{\lambda,k}$. Multiplying and dividing $\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right]\mathbb{P}\left[\boldsymbol{\Phi},\boldsymbol{\lambda} \mid \boldsymbol{\eta}\right]$ by $\prod_{k=1}^{n}\left|\lambda_k \mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{1/2}\frac{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}}{\Gamma(\hat{\alpha}_\lambda)}$, we get:

$$
\begin{aligned}
&\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \times \\
&\quad \mathbb{P}\left[\boldsymbol{\Phi},\boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] \;=\; (2\pi)^{-\frac{Tn}{2}}\left(\frac{\Gamma\left(\hat{\alpha}_\lambda\right)}{\Gamma\left(\alpha_\lambda\right)}\right)^n \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \\
&\qquad \times\left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n}\boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^{T}\mathbf{w}_t\right)\right\}\prod_{k=1}^{n}\frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{\frac{1}{2}}} \\
&\qquad \times\left[\prod_{k=1}^{n}\frac{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}}{\Gamma\left(\hat{\alpha}_\lambda\right)}\lambda_k^{\hat{\alpha}_\lambda-1}\exp\left\{-\lambda_k\hat{\beta}_{\lambda,k}\right\}\right] \\
&\qquad \times\left[\prod_{k=1}^{n}\frac{\left|\lambda_k\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{\frac{1}{2}}}{(2\pi)^{\frac{nm}{2}}}\exp\left\{-\frac{\lambda_k}{2}\left(\boldsymbol{\phi}_k-\hat{\boldsymbol{\phi}}_k\right)'\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\left(\boldsymbol{\phi}_k-\hat{\boldsymbol{\phi}}_k\right)\right\}\right]. \quad (77)
\end{aligned}
$$

Notice that we can write the last two lines of the above expression as:

$$
\prod_{k=1}^{n}\mathrm{G}\left(\lambda_k \mid \hat{\alpha}_\lambda,\hat{\beta}_{\lambda,k}\right)\mathrm{N}\left(\boldsymbol{\phi}_k \mid \hat{\boldsymbol{\phi}}_k,\left(\lambda_k\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right), \quad (78)
$$

where $\mathrm{G}\left(\lambda_k \mid \hat{\alpha}_\lambda,\hat{\beta}_{\lambda,k}\right)$ denotes the density of a gamma distribution with parameters $\left(\hat{\alpha}_\lambda,\hat{\beta}_{\lambda,k}\right)$ evaluated at λ_k, and $\mathrm{N}\left(\boldsymbol{\phi}_k \mid \hat{\boldsymbol{\phi}}_k,\left(\lambda_k\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right)$ denotes the density of a normal distribution with parameters $\left(\hat{\boldsymbol{\phi}}_k,\left(\lambda_k\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right)$ evaluated at $\boldsymbol{\phi}_k$. We can now write the full posterior kernel as:

$$
\begin{aligned}
\mathbb{P}\left[\boldsymbol{\theta} \mid \mathbf{y}^T\right] \;&\propto\; \mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right]\mathbb{P}\left[\boldsymbol{\Phi},\boldsymbol{\lambda} \mid \boldsymbol{\eta}\right]\mathbb{P}\left[\boldsymbol{\rho}\right]\mathbb{P}\left[\mathbf{g}\right]\mathbb{P}\left[\boldsymbol{\eta}\right] \\
&= (2\pi)^{-\frac{Tn}{2}}\left(\frac{\Gamma\left(\hat{\alpha}_\lambda\right)}{\Gamma\left(\alpha_\lambda\right)}\right)^n\left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}}\mathbb{P}\left[\boldsymbol{\rho}\right]\mathbb{P}\left[\mathbf{g}\right]\mathbb{P}\left[\boldsymbol{\eta}\right] \\
&\quad \times\left|\boldsymbol{\Psi}\right|^T\exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n}\boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^{T}\mathbf{w}_t\right)\right\}\prod_{k=1}^{n}\frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{1/2}} \\
&\quad \times\left[\prod_{k=1}^{n}\mathrm{G}\left(\lambda_k \mid \hat{\alpha}_\lambda,\hat{\beta}_{\lambda,k}\right)\mathrm{N}\left(\boldsymbol{\phi}_k \mid \hat{\boldsymbol{\phi}}_k,\left(\lambda_k\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right)\right]. \quad (79)
\end{aligned}
$$

Notice that $\{\lambda_k,\boldsymbol{\phi}_k\}_{k=1}^{n}$ only appear in the final line of the above expression. This fact has two implications. First, the conditional distribution of each $(\lambda_k,\boldsymbol{\phi}_k)$, given $(\boldsymbol{\rho},\mathbf{g},\boldsymbol{\eta})$, is a normal-gamma distribution, independent across k. Second, we can integrate out $\{\lambda_k,\boldsymbol{\phi}_k\}_{k=1}^{n}$ to obtain the posterior

for $(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$:

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] &= \frac{C}{\mathbb{P}\left[\mathbf{y}^T\right]} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}\left[\boldsymbol{\rho}\right] \mathbb{P}\left[\mathbf{g}\right] \mathbb{P}\left[\boldsymbol{\eta}\right] \\ &\quad \times \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^n \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^T \mathbf{w}_t\right)\right\} \prod_{k=1}^n \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}}, \end{aligned} \tag{80}$$

where $\mathbb{P}\left[\mathbf{y}^T\right] = \int \mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] d\left(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta}\right)$ and $C \equiv (2\pi)^{-\frac{Tn}{2}} \left(\frac{\Gamma(\hat{\alpha}_\lambda)}{\Gamma(\alpha_\lambda)}\right)^n$ are constants that do not depend on the model's parameters.

Now, I will characterize the conditional posterior for $\mathbf{g}$, given $(\boldsymbol{\rho}, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{\Phi})$. Note that we can write the data-generating process as:

$$\boldsymbol{\Psi}\mathbf{y}_t - \sum_{\ell=1}^m \boldsymbol{\Phi}_\ell \mathbf{y}_{t-\ell} = \boldsymbol{\Psi}\mathbf{G}\mathbf{w}_t^+ - \sum_{\ell=1}^m \boldsymbol{\Phi}_\ell \mathbf{G}\mathbf{w}_{t-\ell}^+ + \boldsymbol{\epsilon}_t. \tag{81}$$

Notice also $\mathbf{G}\mathbf{w}_t^+ = \text{vec}\left(\mathbf{I}_n \mathbf{G}\mathbf{w}_t^+\right) = \left(\mathbf{w}_t^{+\prime} \otimes \mathbf{I}_n\right)\mathbf{g}$. We can therefore represent the data-generating process as:

$$\boldsymbol{\Psi}\mathbf{y}_t - \sum_{\ell=1}^m \boldsymbol{\Phi}_\ell \mathbf{y}_{t-\ell} = \left(\boldsymbol{\Psi}\left(\mathbf{w}_t^{+\prime} \otimes \mathbf{I}_n\right) - \sum_{\ell=1}^m \boldsymbol{\Phi}_\ell \left(\mathbf{w}_{t-\ell}^{+\prime} \otimes \mathbf{I}_n\right)\right)\mathbf{g} + \boldsymbol{\epsilon}_t. \tag{82}$$

More succinctly, $\tilde{\boldsymbol{\Phi}}\mathbf{z}_t = \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g} + \boldsymbol{\epsilon}_t$, where I have defined:

$$\mathbf{z}_t \equiv \left(\mathbf{y}_t', \ldots, \mathbf{y}_{t-m}'\right)', \ \tilde{\boldsymbol{\Phi}} \equiv \begin{bmatrix} \boldsymbol{\Psi} & -\boldsymbol{\Phi} \end{bmatrix}, \ \mathbf{W}_t \equiv \begin{bmatrix} \mathbf{w}_t^+ & \mathbf{w}_{t-1}^+ & \cdots & \mathbf{w}_{t-m}^+ \end{bmatrix}' \otimes \mathbf{I}_n. \tag{83}$$

We can write the likelihood as:

$$\begin{aligned}\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= \prod_{t=1}^T (2\pi)^{-\frac{n}{2}} \det\left(\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)' \boldsymbol{\Lambda}_t \left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^T \left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)' \boldsymbol{\Lambda}_t \left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)\right\}, \end{aligned} \tag{84}$$

where the factor of proportionality abstracts from terms that do not depend on $\mathbf{g}$. We can write the quadratic term in the above expression as:

$$\begin{aligned}\sum_{t=1}^T \left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)' \boldsymbol{\Lambda}_t \left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right) &= \left(\sum_{t=1}^T \mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{z}_t\right) - 2\left(\sum_{t=1}^T \mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} \\ &\quad + \mathbf{g}\left(\sum_{t=1}^T \mathbf{W}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}. \end{aligned} \tag{85}$$

The first sum on the right-hand side does not depend on $\mathbf{g}$, so the likelihood is proportional to:

$$\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \propto \exp\left\{-\frac{1}{2}\left[\mathbf{g}\left(\sum_{t=1}^{T}\mathbf{W}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} - 2\left(\sum_{t=1}^{T}\mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}\right]\right\}. \tag{86}$$

The prior density for $\mathbf{g}$ is:

$$\begin{aligned}\mathbb{P}\left[\mathbf{g}\right] &= (2\pi)^{-\frac{(n_s-1)n}{2}}\left|\mathbf{V}_g\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\mathbf{g}-\bar{\mathbf{g}}\right)'\mathbf{V}_g^{-1}\left(\mathbf{g}-\bar{\mathbf{g}}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{g}'\mathbf{V}_g^{-1}\mathbf{g} - 2\bar{\mathbf{g}}'\mathbf{V}_g^{-1}\mathbf{g}\right)\right\},\end{aligned} \tag{87}$$

where the second line expands the quadratic, and the factor of proportionality abstracts from all terms that do not depend on $\mathbf{g}$. The conditional posterior kernel for $\mathbf{g}$ is therefore:

$$\begin{aligned}\mathbb{P}\left[\mathbf{g} \mid \mathbf{y}^T\right] &\propto \mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right]\mathbb{P}\left[\mathbf{g}\right] \\ &\propto \exp\left\{-\frac{1}{2}\left[\mathbf{g}\left(\mathbf{V}_g^{-1} + \sum_{t=1}^{T}\mathbf{W}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} - 2\left(\bar{\mathbf{g}}'\mathbf{V}_g^{-1} + \sum_{t=1}^{T}\mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left(\mathbf{g}\hat{\mathbf{V}}_g^{-1}\mathbf{g} - 2\hat{\mathbf{g}}'\hat{\mathbf{V}}_g^{-1}\mathbf{g}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{g}-\hat{\mathbf{g}}\right)'\hat{\mathbf{V}}_g^{-1}\left(\mathbf{g}-\hat{\mathbf{g}}\right)\right\},\end{aligned} \tag{88}$$

where the third line invokes the definitions of $\hat{\mathbf{g}}$ and $\hat{\mathbf{V}}_g$, and the final line completes the square of the quadratic. The above is proportional to the density of a normal distribution with mean $\hat{\mathbf{g}}$ and variance $\hat{\mathbf{V}}_g$.

B Lemmata (For Online Publication)

First, I will state all of the lemmata used in the paper and in the proofs of the propositions. Then, in the following subsections, I will furnish proofs.

Lemma 1. *The lag of the vector of seasonal waveforms is given by* $\mathbf{w}_{t-1} = \mathbf{R}\mathbf{w}_t$*, where* $\mathbf{R}$ *is the* $(n_s-1)\times(n_s-1)$ *matrix whose* (j,k) *element is defined as:*

$$\mathbf{R}_{j,k} \equiv \cos\left(\frac{2\pi}{n_s}j\right)\mathbb{I}\left[j=k\right] - \cos\left(\frac{2\pi}{n_s}j - \frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right]. \tag{89}$$

The matrix $\mathbf{R}$ *is orthogonal; i.e.,* $\mathbf{R}'\mathbf{R} = \mathbf{I}_{n_s-1}$*. For any* $u \in \mathbb{Z}$*, the* u^{th} *power of* $\mathbf{R}$ *is given by:*

$$\left(\mathbf{R}^u\right)_{j,k} = \cos\left(\frac{2\pi}{n_s}ju\right)\mathbb{I}\left[j=k\right] - \cos\left(\frac{2\pi}{n_s}ju - \frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right]. \tag{90}$$

Lemma 2. *For $k \in \{1, \ldots, n\}$, define $\dot{\mathbf{y}}_t^{(k)} \equiv \exp\left\{\frac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_t\right\}\tilde{\mathbf{y}}_t$, $\dot{\mathbf{x}}_t^{(k)} \equiv \exp\left\{\frac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_t\right\}\tilde{\mathbf{x}}_t$, $\dot{\mathbf{Y}}_{(k)} \equiv \begin{bmatrix} \dot{\mathbf{y}}_1^{(k)} & \cdots & \dot{\mathbf{y}}_T^{(k)} \end{bmatrix}'$, and $\dot{\mathbf{X}}_{(k)} \equiv \begin{bmatrix} \dot{\mathbf{x}}_1^{(k)} & \cdots & \dot{\mathbf{x}}_T^{(k)} \end{bmatrix}'$. The likelihood can be written as:*

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= (2\pi)^{-\frac{Tn}{2}} |\boldsymbol{\Psi}|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^n \boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^T \mathbf{w}_t\right)\right\}\left(\prod_{k=1}^n \lambda_k^{\frac{T}{2}}\right) \\
&\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2}\boldsymbol{\psi}_k'\dot{\mathbf{Y}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k\right\} \\
&\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2}\left(-2\boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k + \boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{X}}_{(k)}\boldsymbol{\phi}_k\right)\right\}. \qquad (91)
\end{aligned}$$

Lemma 3. *The joint conditional prior density of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ can be written:*

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] &= \Gamma\left(\alpha_\lambda\right)^{-n}(2\pi)^{-\frac{n^2 m}{2}}\left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}}\left(\prod_{k=1}^n \beta_{\lambda,k}^{\alpha_\lambda}\right) \\
&\quad \times \left(\prod_{k=1}^n \lambda_k^{\frac{\bar{T}}{2}}\right)\exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2}\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right\} \\
&\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2}\left(\boldsymbol{\phi}_k'\bar{\mathbf{X}}'\bar{\mathbf{X}}\boldsymbol{\phi}_k - 2\boldsymbol{\phi}_k'\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right)\right\}. \qquad (92)
\end{aligned}$$

B.1 Proof of Lemma 1

I will begin by showing $\mathbf{R}^u\mathbf{w}_t = \mathbf{w}_{t-u}$, where $\mathbf{R}^u$ is given by equation (90). The claim that $\mathbf{R}\mathbf{w}_t = \mathbf{w}_{t-1}$, with $\mathbf{R}$ given by equation (89), will follow immediately by setting $u = 1$. The claim that $\mathbf{R}$ is orthogonal will follow by setting $u = -1$.

Observe that the jth element of $\mathbf{R}^u\mathbf{w}_t$ is given by:

$$\begin{aligned}
\left(\mathbf{R}^u\mathbf{w}_t\right)_j &= \sum_{k=1}^{n_s-1}\left(\mathbf{R}^u\right)_{j,k}\mathbf{w}_{k,t} \\
&= \sum_{k=1}^{n_s-1}\cos\left(\frac{2\pi}{n_s}ju\right)\mathbb{I}\left[j=k\right]\sqrt{2}\cos\left(\frac{2\pi}{n_s}kt - \frac{\pi}{4}\right) \\
&\quad - \sum_{k=1}^{n_s-1}\cos\left(\frac{2\pi}{n_s}ju - \frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right]\sqrt{2}\cos\left(\frac{2\pi}{n_s}kt - \frac{\pi}{4}\right) \\
&= \sqrt{2}\cos\left(\frac{2\pi}{n_s}ju\right)\cos\left(\frac{2\pi}{n_s}jt - \frac{\pi}{4}\right) - \sqrt{2}\cos\left(\frac{2\pi}{n_s}ju - \frac{\pi}{2}\right)\cos\left(\frac{2\pi}{n_s}\left(n_s - j\right)t - \frac{\pi}{4}\right) \\
&= \sqrt{2}\left[\cos\left(\frac{2\pi}{n_s}ju\right)\cos\left(\frac{2\pi}{n_s}jt - \frac{\pi}{4}\right) - \cos\left(\frac{2\pi}{n_s}ju - \frac{\pi}{2}\right)\cos\left(\frac{2\pi}{n_s}jt + \frac{\pi}{4}\right)\right]. \qquad (93)
\end{aligned}$$

Because $\cos(\omega) = \frac{\exp\{i\omega\}+\exp\{-i\omega\}}{2}$ for any ω, we can write $\cos\left(\frac{2\pi}{n_s}ju\right)\cos\left(\frac{2\pi}{n_s}jt-\frac{\pi}{4}\right)$ as:

$$\begin{aligned}
\cos\left(\frac{2\pi}{n_s}ju\right)\cos\left(\frac{2\pi}{n_s}jt-\frac{\pi}{4}\right) &= \frac{\exp\left\{i\frac{2\pi}{n_s}ju\right\}+\exp\left\{-i\frac{2\pi}{n_s}ju\right\}}{2} \\
&\quad\times\frac{\exp\left\{i\left(\frac{2\pi}{n_s}jt-\frac{\pi}{4}\right)\right\}+\exp\left\{-i\left(\frac{2\pi}{n_s}jt-\frac{\pi}{4}\right)\right\}}{2} \\
&= \frac{\exp\left\{i\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right)\right\}+\exp\left\{-i\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right)\right\}}{4} \\
&\quad+\frac{\exp\left\{i\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right)\right\}+\exp\left\{-i\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right)\right\}}{4} \\
&= \frac{1}{2}\cos\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right)+\frac{1}{2}\cos\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right). \qquad (94)
\end{aligned}$$

Similarly:

$$\begin{aligned}
\cos\left(\frac{2\pi}{n_s}ju-\frac{\pi}{2}\right)\cos\left(\frac{2\pi}{n_s}jt+\frac{\pi}{4}\right) &= \frac{\exp\left\{i\left(\frac{2\pi}{n_s}ju-\frac{\pi}{2}\right)\right\}+\exp\left\{-i\left(\frac{2\pi}{n_s}ju-\frac{\pi}{2}\right)\right\}}{2} \\
&\quad\times\frac{\exp\left\{i\left(\frac{2\pi}{n_s}jt+\frac{\pi}{4}\right)\right\}+\exp\left\{-i\left(\frac{2\pi}{n_s}jt+\frac{\pi}{4}\right)\right\}}{2} \\
&= \frac{\exp\left\{-i\left(\frac{2\pi}{n_s}j(t-u)+\frac{3\pi}{4}\right)\right\}+\exp\left\{i\left(\frac{2\pi}{n_s}j(t-u)+\frac{3\pi}{4}\right)\right\}}{4} \\
&\quad+\frac{\exp\left\{-i\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right)\right\}+\exp\left\{i\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right)\right\}}{4} \\
&= -\frac{\exp\left\{-i\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right)\right\}+\exp\left\{i\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right)\right\}}{4} \\
&\quad+\frac{1}{2}\cos\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right) \\
&= -\frac{1}{2}\cos\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right) \\
&\quad+\frac{1}{2}\cos\left(\frac{2\pi}{n_s}j(t+u)-\frac{\pi}{4}\right), \qquad (95)
\end{aligned}$$

where the penultimate equality uses the fact that $\exp\{\pm i\pi\} = -1$. Combining the above results:

$$\begin{aligned}
\left(\mathbf{R}^u\mathbf{w}_t\right)_j &= \sqrt{2}\left[\cos\left(\frac{2\pi}{n_s}ju\right)\cos\left(\frac{2\pi}{n_s}jt-\frac{\pi}{4}\right)-\cos\left(\frac{2\pi}{n_s}ju-\frac{\pi}{2}\right)\cos\left(\frac{2\pi}{n_s}jt+\frac{\pi}{4}\right)\right] \\
&= \sqrt{2}\cos\left(\frac{2\pi}{n_s}j(t-u)-\frac{\pi}{4}\right) = \mathbf{w}_{j,t}. \qquad (96)
\end{aligned}$$

To verify that $\mathbf{R}$ is orthogonal, we can evaluate the result for $\mathbf{R}^u$ at $u=-1$:

$$\begin{aligned}
\left(\mathbf{R}^{-1}\right)_{j,k} &= \cos\left(\frac{2\pi}{n_s}j\right)\mathbb{I}\left[j=k\right]-\cos\left(-\frac{2\pi}{n_s}j-\frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right] \\
&= \cos\left(\frac{2\pi}{n_s}k\right)\mathbb{I}\left[k=j\right]-\cos\left(-\frac{2\pi}{n_s}\left(n_s-k\right)-\frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right] \\
&= \cos\left(\frac{2\pi}{n_s}k\right)\mathbb{I}\left[k=j\right]-\cos\left(\frac{2\pi}{n_s}k-\frac{\pi}{2}\right)\mathbb{I}\left[k=n_s-j\right]=\mathbf{R}_{k,j}, \qquad (97)
\end{aligned}$$

so $\mathbf{R}^{-1}=\mathbf{R}'$.

B.2 Proof of Lemma 2

The likelihood is:

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= \prod_{t=1}^{T}(2\pi)^{-\frac{n}{2}}\left|\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\right|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)'\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)\right\} \\
&= (2\pi)^{-\frac{Tn}{2}}\left(\prod_{t=1}^{T}\left|\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\right|^{\frac{1}{2}}\right) \\
&\quad\times\exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)'\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)\right\}. \qquad (98)
\end{aligned}$$

We can write the determinant term in the likelihood as:

$$\begin{aligned}
\prod_{t=1}^{T}\left|\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\right|^{\frac{1}{2}} &= \left|\boldsymbol{\Psi}\right|^T\left(\prod_{t=1}^{T}\left|\boldsymbol{\Lambda}_t\right|\right)^{\frac{1}{2}} \\
&= \left|\boldsymbol{\Psi}\right|^T\left(\prod_{t=1}^{T}\prod_{k=1}^{n}\lambda_k\exp\left\{\boldsymbol{\rho}_k'\mathbf{w}_t\right\}\right)^{\frac{1}{2}} \\
&= \left|\boldsymbol{\Psi}\right|^T\exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n}\boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^{T}\mathbf{w}_t\right)\right\}\left(\prod_{k=1}^{n}\lambda_k^{\frac{T}{2}}\right). \qquad (99)
\end{aligned}$$

We can write the quadratic term in the likelihood as:

$$\begin{aligned}
\sum_{t=1}^{T}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)'\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\left(\tilde{\mathbf{y}}_t-\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right) &= \sum_{t=1}^{T}\left(\boldsymbol{\Psi}\tilde{\mathbf{y}}_t-\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right)'\boldsymbol{\Lambda}_t\left(\boldsymbol{\Psi}\tilde{\mathbf{y}}_t-\boldsymbol{\Phi}\tilde{\mathbf{x}}_t\right) \\
&= \sum_{t=1}^{T}\sum_{k=1}^{n}\lambda_k\exp\left\{\boldsymbol{\rho}_k'\mathbf{w}_t\right\}\left(\boldsymbol{\psi}_k'\tilde{\mathbf{y}}_t-\boldsymbol{\phi}_k'\tilde{\mathbf{x}}_t\right)^2 \\
&= \sum_{k=1}^{n}\lambda_k\sum_{t=1}^{T}\left(\boldsymbol{\psi}_k'\dot{\mathbf{y}}_t^{(k)}-\boldsymbol{\phi}_k'\dot{\mathbf{x}}_t^{(k)}\right)^2, \qquad (100)
\end{aligned}$$

where the second equality uses the fact that $\boldsymbol{\Lambda}_t$ is diagonal, and the final line invokes the definitions $\dot{\mathbf{y}}_t^{(k)}$ and $\dot{\mathbf{x}}_t^{(k)}$. We can write the inner sum as:

$$\begin{aligned}
\sum_{t=1}^{T}\left(\boldsymbol{\psi}_k'\dot{\mathbf{y}}_t^{(k)}-\boldsymbol{\phi}_k'\dot{\mathbf{x}}_t^{(k)}\right)^2 &= \sum_{t=1}^{T}\left(\boldsymbol{\psi}_k'\dot{\mathbf{y}}_t^{(k)}-\boldsymbol{\phi}_k'\dot{\mathbf{x}}_t^{(k)}\right)\left(\dot{\mathbf{y}}_t^{(k)\prime}\boldsymbol{\psi}_k-\dot{\mathbf{x}}_t^{(k)\prime}\boldsymbol{\phi}_k\right)\\
&= \boldsymbol{\psi}_k'\left(\sum_{t=1}^{T}\dot{\mathbf{y}}_t^{(k)}\dot{\mathbf{y}}_t^{(k)\prime}\right)\boldsymbol{\psi}_k-2\boldsymbol{\phi}_k'\left(\sum_{t=1}^{T}\dot{\mathbf{x}}_t^{(k)}\dot{\mathbf{y}}_t^{(k)\prime}\right)\boldsymbol{\psi}_k\\
&\quad+\boldsymbol{\phi}_k'\left(\sum_{t=1}^{T}\dot{\mathbf{x}}_t^{(k)}\dot{\mathbf{x}}_t^{(k)\prime}\right)\boldsymbol{\phi}_k\\
&= \boldsymbol{\psi}_k'\dot{\mathbf{Y}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k-2\boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k\\
&\quad+\boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{X}}_{(k)}\boldsymbol{\phi}_k. \qquad (101)
\end{aligned}$$

where the final line invokes the definitions of $\dot{\mathbf{Y}}_{(k)}$ and $\dot{\mathbf{X}}_{(k)}$. Combining the above equations shows that the likelihood takes the form of equation (91).

B.3 Proof of Lemma 3

Let $\bar{\boldsymbol{\phi}}_k \equiv \left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k$. Notice that, in the prior for λ_k, the hyperparameter β_λ can be written as:

$$\begin{aligned}
\beta_{\lambda,k} &= \frac{1}{2}\left[\left(\bar{\mathbf{Y}}\boldsymbol{\Psi}'-\bar{\mathbf{X}}\bar{\boldsymbol{\Phi}}'\right)'\left(\bar{\mathbf{Y}}\boldsymbol{\Psi}'-\bar{\mathbf{X}}\bar{\boldsymbol{\Phi}}'\right)\right]_{k,k}\\
&= \frac{1}{2}\left(\boldsymbol{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\Psi}'-\boldsymbol{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{X}}\bar{\boldsymbol{\Phi}}'-\bar{\boldsymbol{\Phi}}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\Psi}'+\bar{\boldsymbol{\Phi}}\bar{\mathbf{X}}'\bar{\mathbf{X}}\bar{\boldsymbol{\Phi}}'\right)_{k,k}\\
&= \frac{1}{2}\left(\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k-2\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{X}}\bar{\boldsymbol{\phi}}_k+\bar{\boldsymbol{\phi}}_k'\bar{\mathbf{X}}'\bar{\mathbf{X}}\bar{\boldsymbol{\phi}}_k\right)\\
&= \frac{1}{2}\left(\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k-\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right). \qquad (102)
\end{aligned}$$

The conditional prior density for $\boldsymbol{\lambda}$ can therefore be written as:

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{\lambda}\mid\boldsymbol{\eta}\right] &= \prod_{k=1}^{n}\frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\Gamma\left(\alpha_\lambda\right)}\lambda_k^{\alpha_\lambda-1}\exp\left\{-\beta_{\lambda,k}\lambda_k\right\}\\
&= \Gamma\left(\alpha_\lambda\right)^{-n}\left(\prod_{k=1}^{n}\beta_{\lambda,k}^{\alpha_\lambda}\right)\\
&\quad\times\left(\prod_{k=1}^{n}\lambda_k^{\frac{\bar{T}-mn}{2}}\right)\exp\left\{-\sum_{k=1}^{n}\frac{\lambda_k}{2}\left(\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k-\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right)\right\} \qquad (103)
\end{aligned}$$

Let $\dot{\boldsymbol{\Lambda}} \equiv \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. The conditional prior density for $\boldsymbol{\Phi}$ can be written as:

$$\mathbb{P}[\boldsymbol{\Phi} \mid \boldsymbol{\lambda}, \boldsymbol{\eta}] = (2\pi)^{-\frac{n^2 m}{2}} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \operatorname{vec}\left(\boldsymbol{\Phi} - \bar{\boldsymbol{\Phi}}\right)' \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right) \operatorname{vec}\left(\boldsymbol{\Phi} - \bar{\boldsymbol{\Phi}}\right)\right\}. \tag{104}$$

We can write the determinant term in $\mathbb{P}[\boldsymbol{\Phi} \mid \boldsymbol{\lambda}, \boldsymbol{\eta}]$ as:

$$\left|\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right|^{\frac{1}{2}} = \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left|\dot{\boldsymbol{\Lambda}}\right|^{\frac{mn}{2}} = \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left(\prod_{k=1}^{n} \lambda_k^{\frac{mn}{2}}\right). \tag{105}$$

We can write the quadratic term in $\mathbb{P}[\boldsymbol{\Phi} \mid \boldsymbol{\lambda}, \boldsymbol{\eta}]$ as:

$$\begin{aligned}
\operatorname{vec}\left(\boldsymbol{\Phi} - \bar{\boldsymbol{\Phi}}\right)' \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right) \operatorname{vec}\left(\boldsymbol{\Phi} - \bar{\boldsymbol{\Phi}}\right) &= \operatorname{vec}(\boldsymbol{\Phi})' \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right) \operatorname{vec}(\boldsymbol{\Phi}) \\
&\quad -2 \operatorname{vec}\left(\bar{\boldsymbol{\Phi}}\right)' \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right) \operatorname{vec}(\boldsymbol{\Phi}) \\
&\quad + \operatorname{vec}\left(\bar{\boldsymbol{\Phi}}\right)' \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \dot{\boldsymbol{\Lambda}}\right) \operatorname{vec}\left(\bar{\boldsymbol{\Phi}}\right) \\
&= \operatorname{tr}\left\{\boldsymbol{\Phi}\bar{\mathbf{X}}'\bar{\mathbf{X}}\boldsymbol{\Phi}'\dot{\boldsymbol{\Lambda}}\right\} - 2\operatorname{tr}\left\{\boldsymbol{\Phi}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\Psi}'\dot{\boldsymbol{\Lambda}}\right\} \\
&\quad + \operatorname{tr}\left\{\boldsymbol{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\Psi}'\dot{\boldsymbol{\Lambda}}\right\} \\
&= \sum_{k=1}^{n} \lambda_k \boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{X}}\boldsymbol{\phi}_k - 2\sum_{k=1}^{n} \lambda_k \boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k \\
&\quad + \sum_{k=1}^{n} \lambda_k \boldsymbol{\psi}_k' \bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k,
\end{aligned} \tag{106}$$

where the second equality invokes the definition of $\bar{\boldsymbol{\Phi}}$. Combining the above expressions establishes that the joint conditional prior density of $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ takes the form of equation (92).

C X-11 Seasonal Adjustment (For Online Publication)

This appendix provides a brief sketch of the X-11 approach to seasonal adjustment for a scalar time series y_t that is observed monthly. Ladiray and Quenneville (2001) provide further details. As in the body of the paper, I am using "X-11" as shorthand for the family of algorithms that includes the original X-11 algorithm, as well as the X-12 and X-13 refinements. The main components of the procedure are fitting a parametric model to remove deterministic seasonality and applying a filter to remove stochastic seasonality.

The first step is using maximum likelihood to fit a seasonal ARIMA model with deterministic terms:

$$A(L)\,\tilde{A}\left(L^{12}\right)\left(1 - L^{d}\right)\left(1 - L^{12\tilde{d}}\right)\left(y_t - \boldsymbol{\beta}'\mathbf{w}_t^{++}\right) = M(L)\,\tilde{M}\left(L^{12}\right) e_t, \tag{107}$$

where e_t is white noise; $A(\cdot)$, $\tilde{A}(\cdot)$, $M(\cdot)$, and $\tilde{M}(\cdot)$ are polynomials of order p, $\tilde{p}$, q, and $\tilde{q}$; and $\mathbf{w}_t^{++}$ is a vector of deterministic variables. One purpose for fitting the parametric model is to estimate the deterministic component $\boldsymbol{\beta}'\mathbf{w}_t^{++}$. The other purpose is to forecast and backcast y_t at the beginning and end of the sample, which will make it feasible to apply a two-sided filter at each date in the sample. (The analysis in the body of the paper focused on the population properties of the filter, applied to infinite series, so the endpoints were not a concern.)

The additive version of the filter assumes that $\tilde{y}_t \equiv y_t - \boldsymbol{\beta}'\mathbf{w}_t^{++}$ is the sum of a seasonal component y_t^s, an irregular component y_t^{irr}, and a combination trend/cycle component y_t^c:

$$\tilde{y}_t = y_t^s + y_t^{irr} + y_t^c. \tag{108}$$

The seasonal component is assumed to be given by $y_t^s = (1 - \xi(L))\,\tilde{y}_t$, where $\xi(L)$ is a two-sided lag polynomial, and the seasonally adjusted series is $y_t^{sa} \equiv \tilde{y}_t - y_t^s = \xi(L)\,\tilde{y}_t$. The following construction of the polynomial $\xi(L)$ is based on the treatment presented in Chapter 3.4 of Ladiray and Quenneville (2001). Let $\xi_{2\times 12}^{MA}(L)$ denote the lag polynomial associated with a 2×12 moving average, and let $\xi_{3\times q}^{MA}(L)$ denote the lag polynomial associated with a $3 \times q$ moving average, where q is a positive, odd integer:

$$\xi_{2\times 12}^{MA}(L) \equiv \frac{1}{24}\left(L^{-6} + L^{-5}\right)\left(\sum_{\ell=0}^{11} L^{\ell}\right) \tag{109}$$

$$\xi_{3\times q}^{MA}(L) \equiv \frac{1}{3q}\left(L^{-12} + 1 + L^{12}\right)\left(\sum_{\ell=-\frac{q-1}{2}}^{\frac{q-1}{2}} L^{12\ell}\right). \tag{110}$$

(The 2×12 moving average is called as such because it is the average of two overlapping averages of 12 consecutive months; similarly, the $3\times q$ moving average takes the average of the same calendar month across q consecutive years, and then takes the average of those averages across three consecutive years.) Let $\xi_q^H(L)$ denote the lag polynomial associated with a q-term Henderson trend:

$$\xi_q^H(L) = \frac{315}{8}\sum_{\ell=-\frac{q-1}{2}}^{\frac{q-1}{2}} \frac{\left[(m_H-1)^2 - \ell^2\right]\left[m_H^2 - \ell^2\right]\left[(m_H+1)^2 - \ell^2\right]\left[3m_H^2 - 16 - 11\ell^2\right]}{m_H\left(m_H^2-1\right)\left(4m_H^2-1\right)\left(4m_H^2-9\right)\left(4m_H^2-25\right)} L^{\ell}, \tag{111}$$

where $m_H \equiv \frac{q+3}{2}$. (The Henderson filter is designed to minimize the variance of the third difference of a series; increasing q leads to a smoother trend.) The lag polynomial that performs X-11 seasonal

adjustment is assumed to take the form:

$$\xi(L) \equiv 1 - \left[1 - \xi_{2\times 12}^{MA}(L)\right] \xi_{3\times q_2}^{MA}(L) \left[1 - \xi_{q_3}^{H}(L) \left[1 - \left[1 - \xi_{2\times 12}^{MA}(L)\right]^2 \xi_{3\times q_1}^{MA}(L)\right]\right], \tag{112}$$

where q_1 and q_2 are parameters that control the number of years that enter into the seasonal moving averages, and q_3 controls the number of terms used to estimate the low-frequency Henderson trend. The results presented in the body of the paper use $(q_1, q_2, q_3) = (3, 5, 13)$, which are the default settings described in Chapter 3.4 of Ladiray and Quenneville (2001).

The above summarizes the basic version of the X-11 algorithm; the version used by government statistical agencies can be more complicated along several dimensions. First, the parameters (q_1, q_2, q_3) need not be fixed at the above values; they can be either user-specified or selected based on the ratio of variances between the components y_t^s, y_t^{irr}, and y_t^c. Second, the above summarizes the additive version of the X-11 algorithm; there are also log-additive, multiplicative, and pseudo-additive versions. For example, the multiplicative version specifies $\tilde{y}_t = y_t^s \times y_t^{irr} \times y_t^c$, and the seasonal component is estimated by taking the ratio of the data to a two-sided moving average. Third, the vector of deterministic variables $\mathbf{w}_t^{++}$ can contain more than just monthly indicator variables; it can also include variables reflecting the timing of holidays, the number of trading days, and adjustments for outlier observations.

The Bureau of Labor Statistics (BLS) acknowledges the role of discretion when choosing the filter's settings: "But seasonal adjustment also involves some art in addition to science. The art comes in when we use our judgment about outliers in the data or when we decide whether an additive or multiplicative model more closely reflects seasonal variation in economic measures" (Bureau of Labor Statistics, 2020). Different settings may be applied to different series, and for any individual series, the settings may change over time. For example, the BLS has switched from a multiplicative version of the X-11 algorithm to an additive version of the X-11 algorithm for several of its headline employment numbers (Bureau of Labor Statistics, 2020). For many official statistics, government agencies do not publish the precise settings of the seasonal adjustment algorithm being applied.[16]

Taken together, these ad hoc adjustments make the mapping between the original series and the seasonally adjusted series less transparent, but several of the criticisms I raised in the body

[16]There is no single entity that decides how to adjust official economic statistics. The BLS, which is part of the U.S. Depart of Labor, is responsible for data on employment and wages, along with the Consumer Price Index and the Producer Price Index; the Bureau of Economic Analysis (BEA), which is part of the U.S. Department of Commerce, is responsible for the national accounts, which include GDP and its components. The X-11 family of algorithms is designed by the U.S. Census Bureau, whose documentation catalogues the algorithms' adjustable parameters and user-defined settings. However, separate teams of statisticians at the BLS and the BEA are responsible for choosing which settings to apply. The BLS tends to provide more details than the BEA about the discretionary settings that it uses for seasonal adjustment. Partly, this reflects the fact that the BEA relies on disaggregated data from other agencies, and the some of data that the BEA receives is first seasonally adjusted by the source agencies. (Bureau of Economic Analysis, 2015)

of the paper still apply. For all versions of the X-11 algorithm, the seasonally adjusted series is a function of both leads and lags of the original data. Consequently, it will be possible to predict the shocks extracted from seasonally adjusted time series using the history of the original, unadjusted data. Quantitatively, the arguments about the filter's distortionary effects can be extended to the case where different series are subjected to different specifications of the additive X-11 algorithm. Suppose that $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$, and the kth element of the seasonally adjusted vector $\mathbf{y}_t^{sa}$ is $y_{k,t}^{sa} = \xi_k(L)\, y_{k,t}$, where $\xi_k(L)$ is the lag polynomial for the additive X-11 algorithm using parameters $\left(q_1^{(k)}, q_2^{(k)}, q_3^{(k)}\right)$. All of the expressions containing D in Section 2.1 continue to hold, except with the definition of D revised to be:

$$D \equiv \left(\prod_{k=1}^{n} D_k\right)^{\frac{1}{n}}, \quad D_k \equiv \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\Xi_k(\omega)\right) d\omega\right\}, \quad \Xi_k(\omega) \equiv \left|\xi_k\left(\exp\{-i\omega\}\right)\right|^2. \tag{113}$$

The D_k terms can range from 1.42 to 7.37, depending on the values of the parameters $\left(q_1^{(k)}, q_2^{(k)}, q_3^{(k)}\right)$. (The algorithm allows q_1 and q_2 to be selected from the set $\{1, 3, 5, 9, 15\}$ and q_3 to be selected from the set $\{7, 9, 13, 17, 23, 33\}$.) If the seasonal adjustment is performed with the log-additive X-11 algorithm, then the analysis is entirely unchanged, provided that the variables of interest enter $\mathbf{y}_t$ in logs. The results in Section 2.1 are built on linear filtering theory, but if the data are adjusted using the multiplicative version of the X-11 algorithm, then the filter is non-linear. However, Young (1968) argues that the additive and multiplicative routines often produce similar results. For this reason, Census Bureau statisticians have viewed the frequency-domain properties of the additive X-11 algorithm as relevant for approximating the behavior of the multiplicative X-11 algorithm; see Bell and Monsell (1992).

D Computational Details (For Online Publication)

The posterior sampling algorithm outlined in the body of the paper entails several specifications. In particular, it's necessary to declare the initial point for the Markov chain, and it's necessary to construct proposal distributions for the Metropolis steps. I will initialize the sampler with the maximum a posteriori estimate of the parameters, and I will use a Gaussian approximation to the posterior to propose the Metropolis draws.

Let $\left(\boldsymbol{\eta}^\dagger, \boldsymbol{\rho}^\dagger, \mathbf{g}^\dagger\right)$ denote the values of the parameters that maximize the posterior (having marginalized out $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$), and let $\mathbf{H}$ denote the negative Hessian of log posterior, evaluated at the maximum:

$$\left(\boldsymbol{\eta}^\dagger, \boldsymbol{\rho}^\dagger, \mathbf{g}^\dagger\right) \equiv \underset{(\boldsymbol{\eta},\boldsymbol{\rho},\mathbf{g})}{\operatorname{argmax}} \log\left(\mathbb{P}\left[\boldsymbol{\eta}, \boldsymbol{\rho}, \mathbf{g} \mid \mathbf{y}^T\right]\right), \quad \mathbf{H} \equiv -\left.\frac{\partial^2 \log\left(\mathbb{P}\left[\boldsymbol{\eta}, \boldsymbol{\rho}, \mathbf{g} \mid \mathbf{y}^T\right]\right)}{\partial\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)' \partial\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)}\right|_{\left(\boldsymbol{\eta}^\dagger, \boldsymbol{\rho}^\dagger, \mathbf{g}^\dagger\right)}. \tag{114}$$

Let $\mathbf{H}_\eta$ denote the square submatrix of $\mathbf{H}$ whose rows and columns correspond to the coordinates of $\boldsymbol{\eta}$ within $(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}')'$. (For example, if $\boldsymbol{\eta}$ constitutes the first n_η elements of $(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}')'$, then $\mathbf{H}_\eta$ would be the first n_η rows and first n_η columns of $\mathbf{H}$.) Let $\{\mathbf{H}_{\rho_k}\}_{k=1}^n$ be defined analogously. Define $\mathbf{V}_\eta \equiv \mathbf{H}_\eta^{-1}$ and $\mathbf{V}_{\rho_k} \equiv \mathbf{H}_{\rho_k}^{-1}$ for $k \in \{1, \ldots, n\}$. Whereas $\mathbf{H}^{-1}$ approximates the posterior variance of $(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}')'$, the matrix $\mathbf{V}_\eta$ approximates the conditional posterior variance of $\boldsymbol{\eta}$, given $(\boldsymbol{\rho}, \mathbf{g})$; likewise, the matrix $\mathbf{V}_{\rho_k}$ approximates the conditional posterior variance of $\boldsymbol{\rho}_k$, given $(\boldsymbol{\eta}, \mathbf{g})$. The Metropolis steps for $\boldsymbol{\eta}$ and each $\boldsymbol{\rho}_k$ will use Gaussian random-walk proposals with variances given by $c_\eta \mathbf{V}_\eta$ and $c_\rho \mathbf{V}_{\rho_k}$, where c_η and c_ρ are scalar tuning parameters chosen to target a good acceptance rate. For brevity, define:

$$p_\eta(\boldsymbol{\eta} \mid \boldsymbol{\rho}, \mathbf{g}) \equiv \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}[\boldsymbol{\eta}] \left|\boldsymbol{\Psi}\right|^T \prod_{k=1}^n \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}} \tag{115}$$

$$p_{\rho_k}(\boldsymbol{\rho}_k \mid \mathbf{g}, \boldsymbol{\eta}) \equiv \frac{\mathbb{P}[\boldsymbol{\rho}_k] \exp\left\{\frac{1}{2}\boldsymbol{\rho}_k' \left(\sum_{t=1}^T \mathbf{w}_t\right)\right\}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}}, \tag{116}$$

where $\bar{\mathbf{X}}$, $\mathbf{X}_{(k)}$, $\beta_{\lambda,k}$, and $\hat{\beta}_{\lambda,k}$ are understood to be functions of $\boldsymbol{\eta}$, $\mathbf{g}$, and $\boldsymbol{\rho}$. (These terms are also functions of the data, but I have suppressed that argument.) Note that $p_\eta(\boldsymbol{\eta} \mid \boldsymbol{\rho}, \mathbf{g}) \propto \mathbb{P}\left[\boldsymbol{\eta} \mid \boldsymbol{\rho}, \mathbf{g}, \mathbf{y}^T\right]$ and $p_{\rho_k}(\boldsymbol{\rho}_k \mid \mathbf{g}, \boldsymbol{\eta}) \propto \mathbb{P}\left[\boldsymbol{\rho}_k \mid \boldsymbol{\eta}, \mathbf{g}, \mathbf{y}^T\right]$, so these functions are proportional to the target distributions in the Metropolis steps.

Algorithm 1 summarizes the Monte Carlo routine. Step 2b exploits the fact that $\{\boldsymbol{\rho}_k, \lambda_k, \boldsymbol{\phi}_k\}_{k=1}^n$ are independent across k, conditional on $(\boldsymbol{\eta}, \mathbf{g})$, and this step can be executed in parallel across k. Although parallelizing step 2b across multiple cores is unlikely to reduce the algorithm's running time for a bivariate model, it may speed computation when applying the algorithm to a larger model. I set $N_0 = 500,000$ and $N = 500,000$, meaning that I run the sampler for one million iterations and discard the first half as a burn in. To reduce serial correlation across the subsequent half million draws, I retain only every 50th draw, leaving me with $10,000$ posterior draws. I set $c_\eta = 1.50$ and $c_\rho = 0.50$ to target average acceptance rates between 25% and 50% for the Metropolis steps.

E A Modified Prior for Deterministic Seasonality (For Online Publication)

The prior for deterministic seasonality in the body of the paper was designed to incorporate prior beliefs about the magnitude and smoothness of fluctuations in $\mathbf{s}_t$. However, because the prior mean of $\mathbf{B}$ was assumed to be zero, that prior was agnostic about the specific time of year when each

Algorithm 1 Posterior Sampler

1. Initialize $\left(\boldsymbol{\eta}^{(0)}, \boldsymbol{\rho}^{(0)}, \mathbf{g}^{(0)}\right)$ to $\left(\boldsymbol{\eta}^{\dagger}, \boldsymbol{\rho}^{\dagger}, \mathbf{g}^{\dagger}\right)$.
2. For $i = 1, \ldots, N_0 + N$:
   (a) Propose $\boldsymbol{\eta}^{prop} \sim \mathrm{N}\left(\boldsymbol{\eta}^{(i-1)}, c_\eta \mathbf{V}_\eta\right)$. With probability $\min\left\{\frac{p_\eta\left(\boldsymbol{\eta}^{(i)}|\boldsymbol{\rho}^{(i-1)},\mathbf{g}^{(i-1)}\right)}{p_\eta\left(\boldsymbol{\eta}^{(i-1)}|\boldsymbol{\rho}^{(i-1)},\mathbf{g}^{(i-1)}\right)}, 1\right\}$, set $\boldsymbol{\eta}^{(i)} = \boldsymbol{\eta}^{prop}$. With residual probability, set $\boldsymbol{\eta}^{(i)} = \boldsymbol{\eta}^{(i-1)}$.
   (b) For $k = 1, \ldots, n$:
      i. Propose $\boldsymbol{\rho}_k^{prop} \sim \mathrm{N}\left(\boldsymbol{\rho}_k^{(i-1)}, c_\rho \mathbf{V}_{\rho_k}\right)$. With probability $\min\left\{\frac{p_{\rho_k}\left(\boldsymbol{\rho}_k^{(i)}|\mathbf{g}^{(i-1)},\boldsymbol{\eta}^{(i)}\right)}{p_{\rho_k}\left(\boldsymbol{\rho}_k^{(i-1)}|\mathbf{g}^{(i-1)},\boldsymbol{\eta}^{(i)}\right)}, 1\right\}$, set $\boldsymbol{\rho}_k^{(i)} = \boldsymbol{\rho}_k^{prop}$. With residual probability, set $\boldsymbol{\rho}_k^{(i)} = \boldsymbol{\rho}_k^{(i-1)}$.
      ii. Compute $\mathbf{Y}_{(k)}$, $\mathbf{X}_{(k)}$, $\hat{\beta}_{\lambda,k}$, and $\hat{\boldsymbol{\phi}}_k$ as functions of $\left(\boldsymbol{\rho}_k^{(i)}, \boldsymbol{\eta}^{(i)}, \mathbf{g}^{(i-1)}\right)$.
      iii. Draw $\lambda_k^{(i)} \sim \mathrm{G}\left(\hat{\alpha}_\lambda, \hat{\beta}_{\lambda,k}\right)$.
      iv. Draw $\boldsymbol{\phi}_k^{(i)} \sim \mathrm{N}\left(\hat{\boldsymbol{\phi}}_k, \left(\lambda_k \mathbf{X}_{(k)}' \mathbf{X}_{(k)}\right)^{-1}\right)$
   (c) Compute $\hat{\mathbf{g}}$ and $\hat{\mathbf{V}}_g$ as functions of $\left(\boldsymbol{\rho}^{(i)}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\lambda}^{(i)}, \boldsymbol{\Phi}^{(i)}\right)$, and draw $\mathbf{g}^{(i)} \sim \mathrm{N}\left(\hat{\mathbf{g}}, \hat{\mathbf{V}}_g\right)$.
3. Discard the burn-in $\left\{\boldsymbol{\rho}^{(i)}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\lambda}^{(i)}, \boldsymbol{\Phi}^{(i)}, \mathbf{g}^{(i)}\right\}_{i=1}^{N_0}$.

element of $\mathbf{s}_t$ experiences its peaks and troughs:

$$\mathbb{E}_{prior}\left[\mathbf{s}_t\right] = \mathbb{E}_{prior}\left[\mathbf{B}\right]\mathbf{w}_t = \mathbf{0}_{n\times 1}, \ \forall t \in \{1, \ldots, n_s\}. \tag{117}$$

By adopting a non-zero prior mean for $\mathbf{B}$, one can incorporate prior beliefs about the expectation of $\mathbf{s}_t$ for each season, not just the variance and autocovariance of $\mathbf{s}_t$.

Claim. Let $\mathbf{C}$ be defined as the $n_s \times (n_s - 1)$ matrix whose (j,k) element is $\mathbf{C}_{j,k} \equiv \frac{\sqrt{2}}{n_s}\cos\left(\frac{2\pi}{n_s}jk - \frac{\pi}{4}\right)$. Assume that the prior satisfies:

$$\mathbb{E}_{prior}\left[\mathbf{B}\right] = \bar{\mathbf{S}}\mathbf{C}, \quad \mathbb{E}_{prior}\left[\mathrm{vec}\left(\mathbf{B} - \mathbb{E}_{prior}\left[\mathbf{B}\right]\right)\mathrm{vec}\left(\mathbf{B} - \mathbb{E}_{prior}\left[\mathbf{B}\right]\right)'\right] = \mathbf{K} \otimes \mathbf{V}_S, \tag{118}$$

where $\bar{\mathbf{S}}$ is an $n \times n_s$ matrix specified by the econometrician, and $\mathbf{K}$ and $\mathbf{V}_S$ are defined as before. Let $\bar{\mathbf{s}}_t$ denote the tth column of $\bar{\mathbf{S}}$, and assume that $\frac{1}{n_s}\sum_{t=1}^{n_s}\bar{\mathbf{s}}_t = \mathbf{0}_{n\times 1}$. Then:

$$\mathbb{E}_{prior}\left[\mathbf{s}_t\right] = \bar{\mathbf{s}}_t \tag{119}$$

$$\mathbb{E}_{prior}\left[\boldsymbol{\Gamma}_u^s\right] = \kappa_u \mathbf{V}_S + \frac{1}{n_s}\sum_{t=1}^{n_s}\bar{\mathbf{s}}_t\bar{\mathbf{s}}_{t-u}', \tag{120}$$

where κ_u is defined as before, and $t-u$ in the sum is understood to be $\mod n_s$.

Remark. If $\bar{\mathbf{s}}_t$ exhibits abrupt shifts between seasons, then $\mathbb{E}_{prior}\left[\mathbf{\Gamma}_u^s\right]$ does not exhibit as much smoothness as it otherwise would. That behavior might be appropriate for a series like retail sails, which spikes in December and crashes in January. Nevertheless, $\mathbf{s}_t - \bar{\mathbf{s}}_t$ will exhibit smoothness, because equation (120) implies:

$$\mathbb{E}_{prior}\left[\frac{1}{n_s}\sum_{t=1}^{n_s}\left(\mathbf{s}_t-\bar{\mathbf{s}}_t\right)\left(\mathbf{s}_{t-u}-\bar{\mathbf{s}}_{t-u}\right)'\right]=\kappa_u\mathbf{V}_S. \tag{121}$$

Such a prior can allow retail sales to have a much higher unconditional prior mean in December than in November and January. However, conditional on retail sales in November and January exceeding their unconditional prior means, retail sales in December would also be expected to exceed its unconditional prior mean. Consequently, the prior still pools information across adjacent months, while simultaneously recognizing that the months are expected to be different.

Proof. Observe that the prior mean of $\mathbf{s}_t$ is $\mathbb{E}_{prior}\left[\mathbf{B}\mathbf{w}_t\right]=\bar{\mathbf{S}}\mathbf{C}\mathbf{w}_t$. Note that the jth row of $\mathbf{C}$ is equal to $\frac{1}{n_s}\mathbf{w}_j'$. One can show that $\mathbf{w}_j'\mathbf{w}_t = n_s\mathbb{I}\left[j \overset{\mod n_s}{=} t\right]-1$. This implies that j^{th} element of the $n\times 1$ vector $\mathbf{C}\mathbf{w}_t$ is equal to $\mathbb{I}\left[j \overset{\mod n_s}{=} t\right]-\frac{1}{n_s}$. Hence, $\bar{\mathbf{S}}\mathbf{C}\mathbf{w}_t$ is equal to $\bar{\mathbf{s}}_t-\frac{1}{n_s}\sum_{k=1}^{n_s}\bar{\mathbf{s}}_k$, which is equal to $\bar{\mathbf{s}}_t$, because $\frac{1}{n_s}\sum_{k=1}^{n_s}\bar{\mathbf{s}}_k$ is assumed to be zero.

Now, observe that:

$$\mathbb{E}_{prior}\left[\mathbf{\Gamma}_u^s\right]=\mathbb{E}_{prior}\left[\mathbf{B}\mathbf{R}^{u\prime}\mathbf{B}'\right]=\mathbb{E}_{prior}\left[\left(\mathbf{B}-\mathbb{E}_{prior}\left[\mathbf{B}\right]\right)\mathbf{R}^{u\prime}\left(\mathbf{B}-\mathbb{E}_{prior}\left[\mathbf{B}\right]\right)'\right]+\bar{\mathbf{S}}\mathbf{C}\mathbf{R}^{u\prime}\mathbf{C}'\bar{\mathbf{S}}'. \tag{122}$$

Proposition 2 establishes that the first term on the right-hand side is $\kappa_u\mathbf{V}_S$. For the second term, notice that the jth row of $\mathbf{C}\mathbf{R}^{u\prime}$ is equal to $\frac{1}{n_s}\mathbf{w}_j'\mathbf{R}^{u\prime}$, which we know to be equal to $\frac{1}{n_s}\mathbf{w}_{j-u}'$ from Lemma 1. (The $j-u$ is understood to be $\mod n_s$.) Likewise, the kth column of $\mathbf{C}'$ is $\frac{1}{n_s}\mathbf{w}_k$. Hence, the (j,k) element of $\mathbf{C}\mathbf{R}^{u\prime}\mathbf{C}'$ is $\frac{1}{n_s^2}\mathbf{w}_{j-u}'\mathbf{w}_k=\frac{1}{n_s}\mathbb{I}\left[j-k \overset{\mod n_s}{=} u\right]-\frac{1}{n_s^2}$. This implies:

$$\begin{aligned}
\left(\bar{\mathbf{S}}\mathbf{C}\mathbf{R}^{u\prime}\mathbf{C}'\bar{\mathbf{S}}'\right)_{j,k} &= \sum_{t=1}^{n}\sum_{h=1}^{n}\left(\bar{\mathbf{S}}\right)_{j,t}\left(\mathbf{C}\mathbf{R}^{u\prime}\mathbf{C}'\right)_{t,h}\left(\bar{\mathbf{S}}'\right)_{h,k}\\
&= \sum_{t=1}^{n}\sum_{h=1}^{n}\left(\bar{\mathbf{S}}\right)_{j,t}\left(\frac{1}{n_s}\mathbb{I}\left[t-h \overset{\mod n_s}{=} u\right]-\frac{1}{n_s^2}\right)\left(\bar{\mathbf{S}}\right)_{k,h}\\
&= \frac{1}{n_s}\sum_{t=1}^{n}\left(\bar{\mathbf{S}}\right)_{j,t}\left(\bar{\mathbf{S}}\right)_{k,t-u}-\left[\frac{1}{n_s}\sum_{t=1}^{n}\left(\bar{\mathbf{S}}\right)_{j,t}\right]\left[\frac{1}{n_s}\sum_{h=1}^{n}\left(\bar{\mathbf{S}}\right)_{k,h}\right].
\end{aligned} \tag{123}$$

The first term on the last line is equal to the (j,k) element of $\frac{1}{n_s}\sum_{t=1}^{n_s}\bar{\mathbf{s}}_t\bar{\mathbf{s}}_{t-u}'$, and the second term is the product of the jth row of $\frac{1}{n_s}\sum_{t=1}^{n_s}\bar{\mathbf{s}}_t$ and the kth row of $\frac{1}{n_s}\sum_{t=1}^{n_s}\bar{\mathbf{s}}_t$, both of which are assumed

to be zero. □

F Existing Priors (For Online Publication)

Few other papers attempt to develop seasonal priors for Bayesian VARs. The main exceptions are Canova (1992, 1993) and Raynauld and Simonato (1993), who work with reduced-form VARs, rather than structural VARs. Although these papers provide interesting starting points, they also come with some important limitations. In this appendix, I will discuss how my approach improves upon these existing priors for seasonality in autoregressive models. I will also touch on an earlier paper by Gersovitz and MacKinnon (1978), who propose a prior for deterministic seasonality in a (non-autoregressive) single-equation regression. Appendix F.1 discusses existing priors for deterministic seasonality, and Appendix F.2 discusses existing priors for stochastic seasonality.

F.1 Existing Approaches to Deterministic Seasonality

F.1.1 Seasonal Dummy Variables with Uncorrelated Coefficients

The most obvious way to capture deterministic seasonality is to augment the VAR with seasonal dummy variables. Canova (1992, 1993) and Raynauld and Simonato (1993) use seasonal dummy variables with coefficients that are uncorrelated under the prior.[17] However, uncorrelated coefficients on seasonal dummy variables imply a prior belief that $\mathbf{s}_t$ exhibits *negative* serial correlation, which is usually undesirable, for the reasons presented in the body of the paper. To see the issue explicitly, let $\mathbf{d}_t$ be an $n_s \times 1$ vector of season-specific dummy variables; i.e., the jth element of $\mathbf{d}_t$ is $\mathbf{d}_{j,t} \equiv \mathbb{I}\left[t \overset{\text{mod } n_s}{=} j\right]$. Specifying $\mathbf{y}_t = \tilde{\mathbf{B}}\mathbf{d}_t + \tilde{\mathbf{y}}_t$, where $\tilde{\mathbf{B}}$ is an $n \times n_s$ coefficient matrix, is equivalent to specifying $\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{s}_t + \tilde{\mathbf{y}}_t$ with:

$$\boldsymbol{\mu} = \frac{1}{n_s}\tilde{\mathbf{B}}\mathbf{1}_{n_s \times 1}, \quad \mathbf{s}_t = \tilde{\mathbf{B}}\left(\mathbf{d}_t - \frac{1}{n_s}\mathbf{1}_{n_s \times 1}\right). \tag{124}$$

Suppose that the prior for $\tilde{\mathbf{B}}$ satisfies:

$$\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,k}\right] = \bar{\boldsymbol{\mu}}_j, \quad \mathbb{E}_{prior}\left[\left(\tilde{\mathbf{B}}_{j,k} - \bar{\boldsymbol{\mu}}_j\right)\left(\tilde{\mathbf{B}}_{h,\ell} - \bar{\boldsymbol{\mu}}_h\right)\right] = \begin{cases} \sigma^2_{\tilde{B}} & \text{if } (j,k) = (h,\ell) \\ 0 & \text{otherwise} \end{cases}. \tag{125}$$

[17] I have chosen to model deterministic seasonality with season-specific means. Canova (1992, 1993) and Raynauld and Simonato (1993) use season-specific intercepts, but my critiques are broadly applicable to both setups. Furthermore, priors for season-specific means are easier to elicit and interpret than priors for season-specific intercepts, mirroring Villani's (2009) argument that steady-state VARs are preferable for modeling unconditional means.

Under this specification, the prior mean of $\boldsymbol{\mu}$ is $\bar{\boldsymbol{\mu}}$, and the $\tilde{\mathbf{B}}$ coefficients that govern deterministic seasonality are uncorrelated under the prior. Again, for simplicity, I will assume that T is divisible by n_s when computing $\boldsymbol{\Gamma}_u^s$.

Claim. If the prior for $\tilde{\mathbf{B}}$ satisfies equation (125), then the prior expectation for the ACF of $\mathbf{s}_t$ is:

$$\mathbb{E}_{prior}\left[\boldsymbol{\Gamma}_u^s\right] = \begin{cases} \sigma_{\tilde{B}}^2 \frac{n_s-1}{n_s}\mathbf{I}_n & \text{if } u \overset{\text{mod } n_s}{=} 0 \\ -\frac{\sigma_{\tilde{B}}^2}{n_s}\mathbf{I}_n & \text{otherwise} \end{cases}. \tag{126}$$

Remark. The prior with uncorrelated coefficients for dummy variables implies that each season is negatively correlated with all other seasons, regardless of whether the seasons are adjacent (e.g., January and February) or half a year apart (e.g., January and July). This fact would not change if different elements of $\tilde{\mathbf{B}}$ had different prior variances; all that matters is that the columns of $\tilde{\mathbf{B}}$ are uncorrelated.

Proof. Observe that:

$$\boldsymbol{\Gamma}_u^s = \frac{1}{T}\sum_{t=0}^{T-1}\mathbf{s}_t\mathbf{s}_{t-u} = \tilde{\mathbf{B}}\left[\frac{1}{T}\sum_{t=0}^{T-1}\left(\mathbf{d}_t - \frac{1}{n_s}\mathbf{1}_{n_s\times 1}\right)\left(\mathbf{d}_{t-u} - \frac{1}{n_s}\mathbf{1}_{n_s\times 1}\right)'\right]\tilde{\mathbf{B}}'. \tag{127}$$

Define $\boldsymbol{\Gamma}_u^d$ as the matrix that appears in square brackets above, which is the uth autocovariance of the seasonal indicator variables. The above expression implies that the (j,k) element of $\boldsymbol{\Gamma}_u^s$ can be written as:

$$\left(\boldsymbol{\Gamma}_u^s\right)_{j,k} = \sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\left(\boldsymbol{\Gamma}_u^d\right)_{h,\ell}. \tag{128}$$

To compute the (h,ℓ) element of $\boldsymbol{\Gamma}_u^d$, first note that:

$$\begin{aligned}\left(\mathbf{d}_{h,t} - \frac{1}{n_s}\right)\left(\mathbf{d}_{\ell,t-u} - \frac{1}{n_s}\right) &= \left(\mathbb{I}\left[t \overset{\text{mod } n_s}{=} h\right] - \frac{1}{n_s}\right)\left(\mathbb{I}\left[t-u \overset{\text{mod } n_s}{=} \ell\right] - \frac{1}{n_s}\right) \\ &= \mathbb{I}\left[t \overset{\text{mod } n_s}{=} h\right]\mathbb{I}\left[h \overset{\text{mod } n_s}{=} \ell+u\right] - \frac{1}{n_s}\mathbb{I}\left[t \overset{\text{mod } n_s}{=} h\right] \\ &\quad -\frac{1}{n_s}\mathbb{I}\left[t \overset{\text{mod } n_s}{=} \ell+u\right] + \frac{1}{n_s^2}.\end{aligned} \tag{129}$$

Summing across t from 0 to $T-1$ and dividing by T, we get:

$$\left(\boldsymbol{\Gamma}_u^d\right)_{h,\ell} = \frac{1}{T}\sum_{t=0}^{T-1}\left(\mathbf{d}_{h,t} - \frac{1}{n_s}\right)\left(\mathbf{d}_{\ell,t-u} - \frac{1}{n_s}\right) = \frac{1}{n_s}\left(\mathbb{I}\left[h \overset{\text{mod } n_s}{=} \ell+u\right] - \frac{1}{n_s}\right). \tag{130}$$

Hence:

$$
\begin{aligned}
\left(\mathbf{\Gamma}_u^s\right)_{j,k} &= \sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\left(\mathbf{\Gamma}_u^d\right)_{h,\ell} \\
&= \sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\frac{1}{n_s}\left(\mathbb{I}\left[h \overset{\text{mod } n_s}{=} \ell+u\right]-\frac{1}{n_s}\right). \\
&= \frac{1}{n_s}\sum_{\ell=1}^{n_s}\tilde{\mathbf{B}}_{j,\ell+u}\tilde{\mathbf{B}}_{k,\ell}-\frac{1}{n_s^2}\sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}, \qquad (131)
\end{aligned}
$$

where $\ell+u$ in the sum is understood to be mod n_s. It remains to compute the expectation of the above expression under the prior. The assumptions about the prior means, variances, and covariances of the elements of $\tilde{\mathbf{B}}$ imply that:

$$
\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\right]=\mathbb{I}\left[j=k\right]\mathbb{I}\left[h=\ell\right]\sigma_{\tilde{B}}^2+\bar{\boldsymbol{\mu}}_j\bar{\boldsymbol{\mu}}_k. \qquad (132)
$$

It follows that:

$$
\begin{aligned}
\frac{1}{n_s}\sum_{\ell=1}^{n_s}\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,\ell+u}\tilde{\mathbf{B}}_{k,\ell}\right] &= \frac{1}{n_s}\sum_{\ell=1}^{n_s}\left(\mathbb{I}\left[j=k\right]\mathbb{I}\left[\ell+u \overset{\text{mod } n_s}{=} \ell\right]\sigma_{\tilde{B}}^2+\bar{\boldsymbol{\mu}}_j\bar{\boldsymbol{\mu}}_k\right) \\
&= \mathbb{I}\left[j=k\right]\mathbb{I}\left[u \overset{\text{mod } n_s}{=} 0\right]\sigma_{\tilde{B}}^2+\bar{\boldsymbol{\mu}}_j\bar{\boldsymbol{\mu}}_k \qquad (133) \\
\frac{1}{n_s^2}\sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\right] &= \frac{1}{n_s^2}\sum_{h=1}^{n_s}\sum_{\ell=1}^{n_s}\left(\mathbb{I}\left[j=k\right]\mathbb{I}\left[h=\ell\right]\sigma_{\tilde{B}}^2+\bar{\boldsymbol{\mu}}_j\bar{\boldsymbol{\mu}}_k\right) \\
&= \frac{1}{n_s}\mathbb{I}\left[j=k\right]\sigma_{\tilde{B}}^2+\bar{\boldsymbol{\mu}}_j\bar{\boldsymbol{\mu}}_k. \qquad (134)
\end{aligned}
$$

Thus, $\mathbb{E}_{prior}\left[\left(\mathbf{\Gamma}_u^s\right)_{j,k}\right]=\mathbb{I}\left[j=k\right]\left(\mathbb{I}\left[u \overset{\text{mod } n_s}{=} 0\right]-\frac{1}{n_s}\right)\sigma_{\tilde{B}}^2$. □

F.1.2 Seasonal Dummy Variables with Correlated Coefficients

When using seasonal dummy variables, generating smoothness in season-specific means requires the coefficients to be correlated. Gersovitz and MacKinnon (1978) consider regressions with season-specific coefficients:

$$
y_t=\boldsymbol{c}_t'\boldsymbol{\beta}_t+e_t, \quad \boldsymbol{\beta}_t=\tilde{\mathbf{B}}\mathbf{d}_t, \qquad (135)
$$

where y_t is a scalar time series, e_t is white noise, $\boldsymbol{c}_t$ is a vector of observed covariates, and $\mathbf{d}_t$ is an $n_s\times 1$ vector of seasonal dummy variables. Those authors propose a prior over $\tilde{\mathbf{B}}$ with the goal of generating smoothness in the regression coefficients $\{\boldsymbol{\beta}_t\}_{t=1}^{n_s}$. Gersovitz and MacKinnon leave the distribution of $\boldsymbol{c}_t$ unmodeled, and they do not consider the case where $\boldsymbol{c}_t$ contains lagged values of y_t. Consequently, their analysis excludes autoregressive processes and is silent on stochastic seasonality.

However, when $\boldsymbol{c}_t = 1$, their framework implies a prior over season-specific means. Hereafter, assume $\boldsymbol{c}_t = 1$, so $\tilde{\mathbf{B}}$ is a row vector of season-specific means. Gersovitz and MacKinnon adopt the following stochastic restrictions over the parameters:

$$\left(\tilde{B}_{t+1} - \tilde{B}_t\right) - \left(\tilde{B}_t - \tilde{B}_{t-1}\right) \sim \mathrm{N}\left(0, \zeta^2\right), \tag{136}$$

where $\tilde{B}_t$ is the tth element of $\tilde{\mathbf{B}}$, and $t \pm 1$ is understood to be $\mod\ n_s$. (To avoid collinearlity in the restrictions, one can only impose the above condition for $t = 1, \ldots, n_s - 1$. The prior is improper, because it entails only $n_s - 1$ stochastic restrictions for n_s coefficients.) The deterministic seasonal deviation of y_t from its long-run average is $s_t = \tilde{\mathbf{B}}\left(\mathbf{d}_t - \frac{1}{n_s}\mathbf{1}_{n_s \times 1}\right)$, so equation (136) is equivalent to $(s_{t+1} - s_t) - (s_t - s_{t-1}) \sim \mathrm{N}\left(0, \zeta^2\right)$. The hyperparameter ζ controls beliefs about smoothness in the season-specific means: If ζ is small, then the second difference of s_t is expected to be small, meaning that seasonal fluctuations are unlikely to exhibit abrupt increases or decreases.

In many settings, Gersovitz and MacKinnon's prior seems more reasonable than treating the elements of $\tilde{\mathbf{B}}$ as uncorrelated, but their approach also has some limitations. In particular, there's no way to separate the smoothness of deterministic seasonality from the magnitude of deterministic seasonality, because both are controlled by the single hyperparameter ζ. Consider the limiting case, where ζ approaches zero. This implies that the restriction $\tilde{B}_{t+1} - \tilde{B}_t = \tilde{B}_t - \tilde{B}_{t-1}$ holds exactly for $t = 1, \ldots n_s - 1$. However, those restrictions imply that s_t does not change with t, meaning that there is no seasonality at all. Hence, an implementation of Gersovitz and MacKinnon's prior that favors smoother seasonal patterns will also favor smaller seasonal patterns. In contrast, my prior can accommodate distinct beliefs about the expected persistence of deterministic seasonality (captured by $\mathbf{K}$) and the expected magnitude of deterministic seasonality (captured by $\mathbf{V}_S$). More generally, because Gersovitz and MacKinnon's prior over $\tilde{\mathbf{B}}$ is improper, the prior predictive distribution over s_t is not defined.

F.2 Existing Approaches to Stochastic Seasonality

In this subsection, to make the comparison between the other authors' priors and my prior as transparent as possible, I will specialize to a homoskedastic, univariate AR(m) process:

$$y_t = \sum_{\ell=1}^{m} \Phi_\ell y_{t-\ell} + e_t, \quad e_t \sim \mathrm{N}\left(0, \frac{1}{\lambda}\right). \tag{137}$$

I have abstracted from deterministic terms in order to focus on prior beliefs about stochastic seasonality. By setting $\Psi = 1$, I have abstracted from the structural identification problem to focus on

reduced-form behavior. I have also abstracted from the possibility of time-varying parameters. The linear restrictions that imply a seasonal unit root in y_t at frequency ω^* are:

$$1 = \sum_{\ell=1}^{m} \Phi_\ell \cos(\omega^* \ell) \tag{138}$$

$$0 = \sum_{\ell=1}^{m} \Phi_\ell \sin(\omega^* \ell), \tag{139}$$

which are the reduced-form, univariate analogues to equations (36) and (37). My prior treats these as stochastic constraints:

$$\sum_{\ell=1}^{m} \Phi_\ell \cos(\omega^* \ell) \mid \lambda \sim \mathrm{N}\left(1, \frac{1}{\tau_{\omega^*}^2 \lambda}\right) \tag{140}$$

$$\sum_{\ell=1}^{m} \Phi_\ell \sin(\omega^* \ell) \mid \lambda \sim \mathrm{N}\left(0, \frac{1}{\tau_{\omega^*}^2 \lambda}\right), \tag{141}$$

with $\tau_{\omega^*} > 0$ controlling the prior confidence in a spectral peak at frequency ω^*. As in the body of the paper, equation (141) is only necessary when $\omega^* \notin \{0, \pi\}$. Appendix F.2.1 will discuss the approach to stochastic seasonality in Canova (1992, 1993), and Appendix F.2.2 will discuss the approach to stochastic seasonality in Raynauld and Simonato (1993).

F.2.1 Canova (1992, 1993)

Canova (1992, 1993) argues that a reasonable prior should favor spectral peaks at seasonal frequencies. On a conceptual level, I agree: Seasonal unit roots (or near unit roots) imply spectral peaks at seasonal frequencies. On a technical level, though, the prior restriction that Canova proposes, even if imposed dogmatically, does not imply a spectral peak at the desired frequency. Given a prior belief that y_t has a spectral peak at frequency ω^*, Canova (1992) recommends squeezing the parameters toward the region of the parameter space where $1 \approx \sum_{\ell=1}^{m} \Phi_\ell \cos(\omega^* \ell)$, and he does so with a stochastic linear restriction along the lines of equation (140). However, there are many processes that satisfy the condition $1 = \sum_{\ell=1}^{m} \Phi_\ell \cos(\omega^* \ell)$ exactly, but do not have a spectral peak at, or even near, frequency ω^*. Importantly, Canova does not have an analogue to my equation (139), which is necessary for a seasonal unit root at the desired frequency. Although an exact unit root isn't necessary for a spectral peak at a seasonal frequency, a spectral peak requires equation (139) to hold at least approximately. Absent a stochastic linear restriction along the lines of equation (141), Canova's prior restriction will not typically favor seasonal processes.

To see the issue explicitly, note that the autoregressive process can be written $A(L)\, y_t = e_t$,

where $A(L) \equiv 1 - \sum_{\ell=1}^{m} \Phi_\ell L^\ell$. Define $a(\omega) \equiv A(\exp\{-i\omega\})$ as the autoregressive transfer function. The spectral density of y_t is:

$$f(\omega) = \frac{1}{|a(\omega)|^2} \frac{1}{2\pi\lambda} = \frac{1}{\Re(a(\omega))^2 + \Im(a(\omega))^2} \frac{1}{2\pi\lambda}, \tag{142}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts. Canova's stated goal is to favor parameters such that $1/|a(\omega^*)|^2$ is large for some desired frequency ω^*. Notice that:

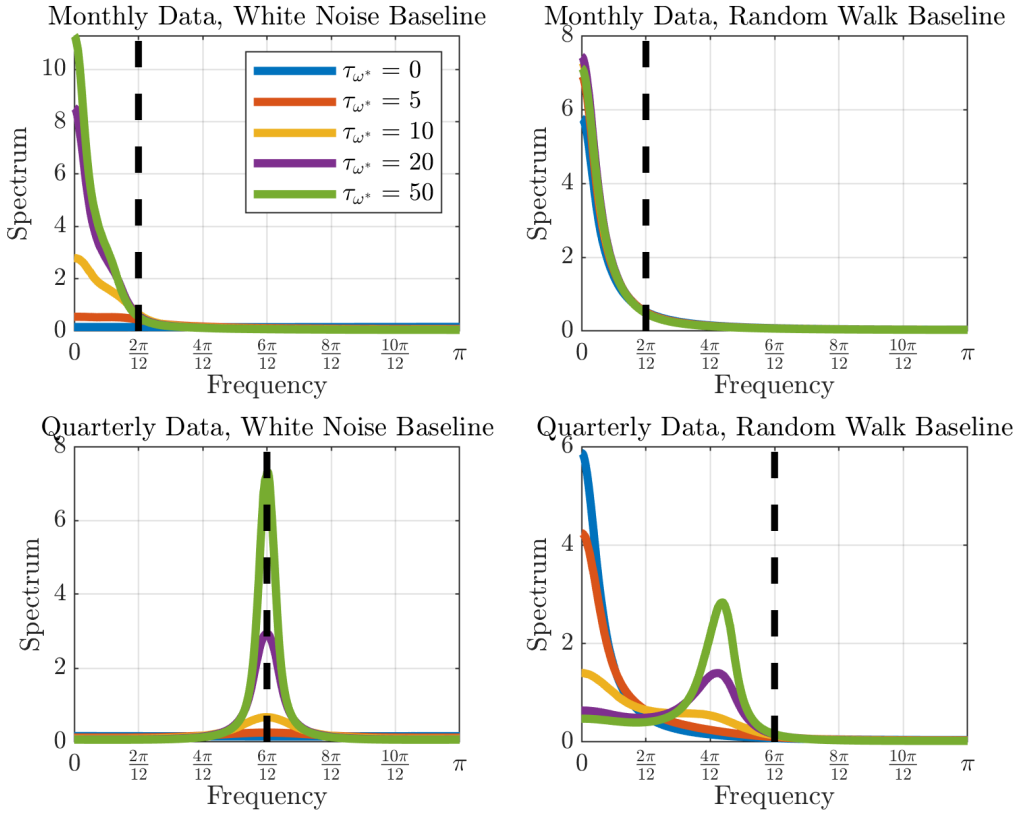
$$a(\omega) = \underbrace{1 - \sum_{\ell=1}^{m} \Phi_\ell \cos(\omega\ell)}_{\Re(a(\omega))} + i \underbrace{\sum_{\ell=1}^{m} \Phi_\ell \sin(\omega\ell)}_{\Im(a(\omega))}. \tag{143}$$

Canova's restriction $1 \approx \sum_{\ell=1}^{m} \Phi_\ell \cos(\omega^*\ell)$ is equivalent to adopting a prior belief that $\Re(a(\omega^*)) \approx 0$, but his prior, unlike mine, does not incorporate a prior belief that $\Im(a(\omega^*)) \approx 0$.[18] In other words, Canova's restriction is necessary for a spectral peak at frequency ω^*, but outside of a few special cases, it's not sufficient: As equation (142) demonstrates, $f(\omega^*)$ will not be large unless both $\Re(a(\omega^*))$ and $\Im(a(\omega^*))$ are small in absolute value.

Because Canova's approach does not incorporate any explicit beliefs about $\Im(a(\omega^*))$, the seasonal properties of Canova's prior are sensitive to ostensibly non-seasonal assumptions, and there will not necessarily be a spectral peak at the desired frequency. Figure 6 shows what happens when various baseline priors are augmented with Canova's stochastic restriction (140), but not the additional stochastic restriction that my prior employs (141). The white noise baseline prior refers to $\Phi_\ell \mid \lambda \sim \mathrm{N}\left(0, \frac{1}{\lambda\ell^2}\right)$, and the random walk baseline prior refers to $\Phi_\ell \mid \lambda \sim \mathrm{N}\left(\mathbb{I}[\ell = 1], \frac{1}{\lambda\ell^2}\right)$. In each case, I have assumed that $m = n_s + 1$ and $\omega^* = \frac{2\pi}{n_s}$, which is the annual periodicity. With monthly data, Canova's prior restriction does not generate a spectral peak at frequency ω^*; instead, the spectral peak appears at frequency zero. To be fair, in the special case where Canova's restriction is used to augment a white-noise prior for quarterly data, the prior appears to have the intended behavior of favoring a spectral peak at frequency ω^*. However, changing the baseline prior from white noise to a random walk — which is only supposed to be relevant for low frequencies — changes the spectrum at seasonal frequencies. For a random walk baseline prior with quarterly data, Canova's restriction appears to favor a spectral peak corresponding to a periodicity of about 5.5 quarters. In contrast, my prior always implies that the spectrum is expected to become more sharply peaked at frequency ω^* as τ_{ω^*} increases.

[18] His prior and mine will agree in the special case where $\omega^* = \pi$, because the condition $\sum_{\ell=1}^{m} \Phi_\ell \sin(\pi\ell) = 0$ will be satisfied for any coefficient values. For other values of ω^*, however, the prior restriction $\Im(a(\omega^*)) \approx 0$ is important for generating the spectral peak at the desired frequency.

Figure 6: Spectra Implied by Canova's Prior



Notes: The figure shows the prior median of the spectrum when Canova's stochastic linear restriction (140) is used to augment various baseline priors. The vertical dashed lines indicate the annual periodicity ($\omega^* = \frac{2\pi}{12}$ for monthly data, and $\omega^* = \frac{2\pi}{4}$ for quarterly data). Each solid line is generated by taking $10,000$ draws from the conditional prior distribution for $\mathbf{\Phi}$, computing the spectrum associated with each parameter draw, and computing the median value of the spectrum across draws. Each prior conditions on $\lambda = 1$.

Canova (1993) extends the prior in Canova (1992) from univariate to multivariate models; because Canova (1993) nests Canova (1992) as a special case, the same critiques apply.

F.2.2 Raynauld and Simonato (1993)

My strategy for eliciting a prior over the VAR parameters is to articulate beliefs about the behavior of y_t, and then construct stochastic linear restrictions on the coefficients to implement that behavior. An alternative approach would be declaring a prior directly in terms of the coefficients, as advocated by Raynauld and Simonato (1993). However, seemingly reasonable prior assumptions about the coefficients can obscure the prior's implications for the seasonal properties of y_t.

Raynauld and Simonato's prior over $\{\Phi_\ell\}_{\ell=1}^m$ is Gaussian and independent across ℓ with:

$$\mathbb{E}\left[\Phi_\ell\right] = \mathbb{I}\left[\ell = 1\right] + \mathbb{I}\left[\ell = n_s\right] - \mathbb{I}\left[\ell = n_s + 1\right], \quad \mathbb{V}\left[\Phi_\ell\right] = \left(\frac{\text{Tight} \times \text{SD}^{\lfloor \ell / n_s \rfloor}}{\left(\ell - (n_s - 1) \times \left\lfloor \frac{\ell}{n_s} \right\rfloor\right)^{\text{Tooth}}}\right)^2, \quad (144)$$

where Tight, SD, and Tooth are hyperparameters, and the number of lags m is assumed to be at least n_s+1. The prior mean is centered on a so-called "airline process" $(1-L)(1-L^{n_s})\,y_t = e_t$, which has seasonal unit roots at the annual frequency and the harmonic frequencies $\left(\frac{2\pi}{n_s}j,\ j = 1, \ldots, n_s\right)$, in addition to frequency zero. The hyperparameter Tight > 0 controls the overall tightness of the prior. The hyperparameter SD $\in [0, 1)$ controls "seasonal decay," in the sense that the first year's worth of lags have higher variances than the second year's worth of lags, and so on. The hyperparameter Tooth ≥ 1 controls how much to increase the relative prior variance on the seasonal lags (Φ_{n_s}, Φ_{2n_s}, etc.). When plotting $\mathbb{V}\left[\Phi_\ell\right]$ as a function of ℓ, the graph appears to have "teeth" spiking up at the seasonal lags: The function $\left(\ell - (n_s - 1) \times \left\lfloor \frac{\ell}{n_s} \right\rfloor\right)^{-\text{Tooth}}$ is n_s-periodic and strictly decreasing in ℓ for $\ell \in \{1, \ldots, n_s - 1\}$. With monthly data, this means that the coefficients on the first 11 lags are increasingly likely to be close to zero, but the coefficient on y_{t-12} is expected to be relatively large in absolute value.

Although some of these features may initially appear reasonable, this prior comes with three important limitations. First, the prior doesn't necessarily favor seasonal processes over non-seasonal processes. Assume that $n_s \geq 4$, which is the case in almost all applications. For y_t to have a spectral peak at a seasonal frequency, it is neither necessary nor sufficient for Φ_{n_s}, the coefficient on the seasonal lag y_{t-n_s}, to be large in absolute value. For example, the process $\left(1 - 2\cos\left(\frac{2\pi}{n_s}\right)L + L^2\right)y_t = e_t$ has a seasonal unit root at the annual frequency, even though $\Phi_{n_s} = 0$. Conversely, increasing the prior variance of Φ_{n_s} increases the probability of this coefficient being negative, which could create a spectral *trough* at the annual frequency, rather than a spectral peak. Furthermore, one can show that, under Raynauld and Simonato's prior:

$$\mathbb{P}\left[(\Phi_1, \Phi_2, \ldots, \Phi_m) = \left(2\cos\left(\frac{2\pi}{n_s}\right), -1, \mathbf{0}_{1\times(m-2)}\right)\right] < \mathbb{P}\left[(\Phi_1, \Phi_2, \ldots, \Phi_m) = \mathbf{0}_{1\times m}\right]. \quad (145)$$

Hence, the prior treats the seasonal unit-root process $\left(1 - 2\cos\left(\frac{2\pi}{n_s}\right)L + L^2\right)y_t = e_t$ as less plausible than pure white noise, which doesn't exhibit any seasonality at all.

Second, it is difficult to separate prior beliefs about seasonal behavior from beliefs about non-seasonal behavior. In Raynauld and Simonato's setup, the natural way to express more confidence in stochastic seasonality is to make the overall prior tighter. (The alternative would be to increase the

relative prior variance of Φ_{n_s}, but for the reasons discussed above, that is not entirely satisfying.) Applying more shrinkage, by adjusting the hyperparameter Tight, makes the estimated spectral density resemble an airline process at *all* frequencies, not just the seasonal frequencies. In contrast, taking $\tau_{\omega_*} \to \infty$ makes my prior collapse on the submanifold of the parameter space that features a seasonal unit root at frequency ω^*, but my prior restriction still allows the model to fit the data at non-seasonal frequencies.

Third, treating the coefficients as independent doesn't acknowledge the fact that the seasonal properties of the process depend on all of the coefficients jointly. Conditional on some coefficients deviating from the prior mean, it's reasonable to expect the other coefficients to deviate in a way that maintains the spectral peaks near the seasonal frequencies.

Components of Raynauld and Simonato's approach may be appropriate if an econometrician genuinely does prefer an airline process over other seasonal processes, or if an econometrician genuinely is uncertain about the coefficient on y_{t-n_s}. (Within my framework, one could incorporate such beliefs into the dummy observations $\bar{\mathbf{X}}_0$ and $\bar{\mathbf{Y}}_0$.) Conceptually, Raynauld and Simonato choose to declare their beliefs about the coefficients themselves, whereas I declare a set of beliefs about seasonal fluctuations in y_t. With my approach, the prior over the coefficients is simply instrumental to generating reasonable attributes for the spectral density at seasonal frequencies.