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

Ross Doppelt*

February 14, 2023

Abstract

When fitting structural vector autoregressions (VARs), macroeconomists should prefer the original, unadjusted data over the seasonally adjusted versions, even when using VARs to study non-seasonal phenomena. 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. Third, as an application, I incorporate seasonality into Baumeister and Hamilton's (2015) model of labor-market demand and supply; the models with and without seasonality produce substantially different results.

Keywords: Bayesian VARs, Structural VARs, Frequency Domain, Seasonality, Identification Through Heteroskedasticity

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

1 Introduction

Structural vector autoregressions (VARs) are one of the most commonly used tools for empirical macroeconomists, and it is standard practice to estimate VARs using seasonally adjusted time series, whenever seasonally adjusted versions are available. In part, the widespread use of seasonally adjusted variables stems from the fact that most VAR practitioners are interested in business cycles, not seasonal cycles. However, if switching from seasonally adjusted to unadjusted time series changes a model's results, then it demonstrates that conclusions about business cycles depend on assumptions about seasonality. In general, seasonal adjustment distorts identification in structural VARs, and seasonal variation can provide useful identifying information. The treatment of seasonality can therefore affect inferences about non-seasonal phenomena, such as structural parameters, impulse responses, and variance decompositions. Given the choice, macroeconomists who are estimating structural VARs should opt for the original, unadjusted variables over the seasonally adjusted versions. When doing so, it is necessary to incorporate seasonality explicitly into the model and confront the novel statistical challenges of seasonal data.

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 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 identify structural parameters.

Many macroeconomists have found that being a Bayesian ameliorates two challenges of VAR estimation: fitting densely parameterized models and dealing with unit roots (Del Negro and

Schorfheide, 2011). 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 can be challenging. 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 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 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 prior for the autoregressive coefficients favors 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. Holding fixed all assumptions about non-seasonal behavior, the seasonal model and the seasonally adjusted model produce large differences in the estimates of the structural parameters, impulse responses, and variance decompositions. One reason for the discrepancies is the identifying information embodied in seasonal heteroskedasticity. Even when the model is constrained to be homoskedastic, the seasonally adjusted and unadjusted time series lead to divergent results, suggesting that the seasonal adjustment routine is also responsible for some of the difference.

To the best of my knowledge, this paper provides the first systematic characterization of how seasonality affects identification in structural VARs, which relates to several lines of literature. Early antecedents include Wallis (1974) and Sims (1974), who discuss the role of seasonal-adjustment filters in atheoretical time-series regressions. More broadly, economists have long studied how filtering affects identification and inference in econometric models; recent examples include Hamilton (2018), who focuses on the Hodrick-Prescott filter, and Ashley and Verbrugge (2022), who focus on the bandpass filter.[1] There is some overlap between their criticisms of filters that suppress low frequencies and my criticisms of filters that suppress seasonal frequencies. In particular, all two-sided filters raise conceptual questions about how to interpret the identified shock for date t when it incorporates information about the future and can be predicted using the lags of the unfiltered data. The more novel part of my analysis is showing that seasonal adjustment creates quantitatively large distortions in the structural parameters, which is an issue independent of whether the filter is one-sided or two-sided. The framework that I develop to make this argument may also be useful for researchers who are studying the effects of other linear filters on identification schemes in structural VARs.

This paper is also the first to combine seasonal volatility with identification through heteroskedasticity in a structural VAR. Other studies using identification through heteroskedasticity often designate different volatility regimes based on the timing of market disruptions (as in Rigobon, 2003), policy announcements (as in Wright, 2012), or other turning points in economic history (as in Brunnermeier et al., 2021). In many settings, it seems plausible that some economic shocks may tend to be larger at certain times of the year, given the clear seasonality in many production technologies, household preferences, and government policies.[2] In the labor-market data that I study, incorporating volatility shifts at medium and low frequencies, along the lines of Brunnermeier et al. (2021), provides little identifying information. In contrast, incorporating seasonal volatility makes it easier to identify labor supply and labor demand, because the relative magnitudes of supply shocks and demand shocks depend on the calendar month. This kind of identification strategy provides an affirmative reason to use seasonally unadjusted variables, beyond any negative reasons for avoiding the seasonally adjusted versions.

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

[1] These authors are upfront about their proscriptions: Hamilton's work is entitled "Why You Should Never Use The Hodrick-Prescott Filter," and Ashley and Verbrugge's is entitled "Death to Regression Modeling on Bandpass-Filtered Data." My advice is more targeted. I will focus my attention on the situation where a researcher (a) wants to fit a structural VAR and (b) has the choice between using seasonally adjusted or unadjusted time series. In this paper, I will remain mostly agnostic about scenarios where either of those conditions do not apply, but I will touch on them briefly in Section 6.

[2] See Beaulieu et al. (1992) for stylized facts on seasonality in the volatility of output.

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.[3] Although those authors provide welcome discussions of seasonality in Bayesian VARs, the specific priors that they propose come with important limitations. These priors implicitly assume that deterministic seasonality exhibits negative serial correlation, whereas my framework provides a way of specifying the degree of smoothness in season-specific means. For stochastic seasonality, Raynauld and Simonato's setup makes it difficult to separate beliefs about the spectrum at seasonal and non-seasonal frequencies; my approach provides a convenient way to combine prior information about different frequency bands. Like me, Canova advocates using a prior that favors spectral peaks at seasonal frequencies, but because of a technical issue, the specific class of priors that he proposes does not typically have that property. In the appendix, I discuss in detail how my strategy addresses some of the limitations in these earlier approaches.

Finally, my results on seasonality in structural VARs provide a counterpoint to the literature on seasonality 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) offer a rationale for fitting DSGE models with seasonally adjusted variables: If the source of seasonality is misspecified in the equilibrium model, then seasonally adjusted time series can, under certain circumstances, lead to more robust inferences about the model's non-seasonal properties. Although Sims and Hansen and Sargent tend to favor seasonal adjustment when estimating DSGE models, they don't support the practice universally: "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

[3]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.

on the particulars of the equilibrium model being estimated.[4] 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.

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 contains the application. In Section 6, I summarize the main arguments and take stock of their practical implications. Appendix A contains proofs. Appendix B provides extra details on the X-11 family of seasonal adjustment algorithms. Appendix C contains computational details. Appendix D 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. To distinguish $\boldsymbol{\mu}$ from $\mathbf{s}_t$, I will adopt the normalization $\frac{1}{n_s} \sum_{t=1}^{n_s} \mathbf{s}_t = \mathbf{0}_{n\times 1}$. 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 lower-than-expected ice cream sales this summer being systematically correlated with lower-than-

[4] Sargent (1978) and Ghysels (1988) make early arguments that equilibrium models ought to be estimated with unadjusted data. 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. 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.

expected ice cream sales next summer. More generally, a process exhibits stochastic seasonality if its spectrum has peaks at seasonal frequencies.

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 $\mathbf{y}_t$; adjusting for stochastic seasonality entails filtering out seasonal oscillations in $\tilde{\mathbf{y}}_t$. Removing $\mathbf{s}_t$ 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. In contrast, stochastic seasonality is intrinsically a matter of how shocks are propagated, because unanticipated fluctuations today can persist into the same season in the years ahead. 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.

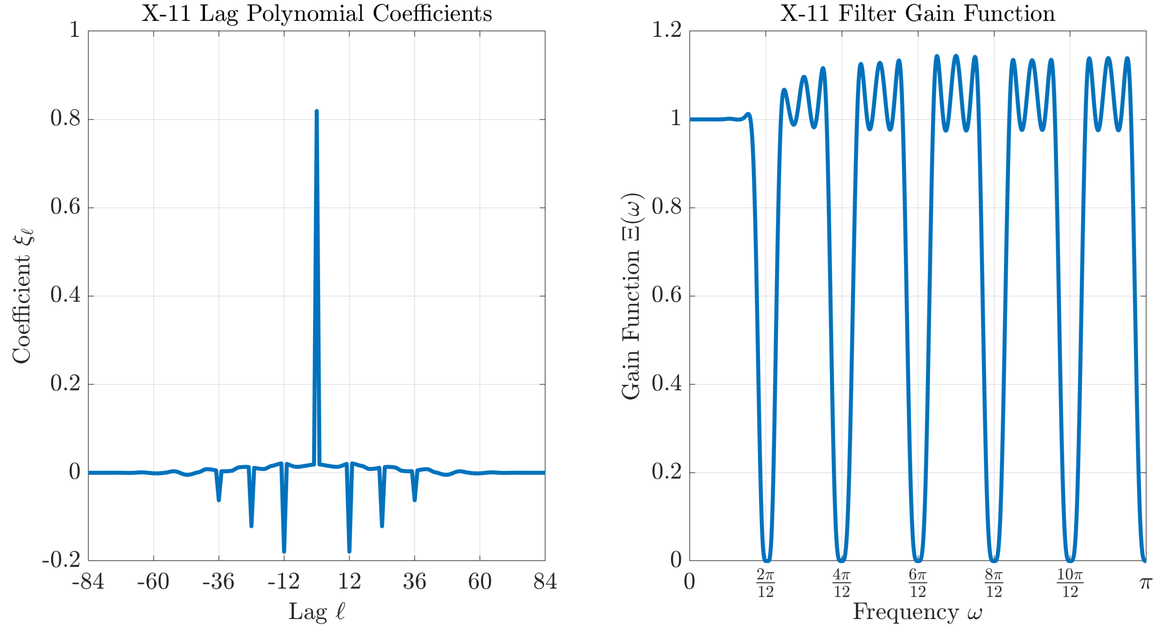
2.1 Seasonal Adjustment and the Identification Problem

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 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. 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

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 B.

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 $\mathbf{\Psi}$ and $\mathbf{\Lambda}$. 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 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 B contains some more elaboration on X-11, but I refer readers to the Census Bureau's documentation for an exhaustive explanation.[5] Assuming that deterministic terms have already been removed, the seasonally adjusted series, denoted $\mathbf{y}_t^{sa}$, is constructed by applying

[5]See Ladiray and Quenneville (2001) for a book-length treatment. Although the Census Bureau has now developed X-12 and X-13 algorithms, they consist of extensions and refinements of X-11, so it's common to refer to this entire family of algorithms as X-11 seasonal-adjustment procedures. In this section, the numerical results on the X-11 filter use the same tuning parameters as Chapter 3.4 in Ladiray and Quenneville; the appendix states these parameters explicitly and briefly discusses the alternatives.

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 B). The filter has interpretations in both the time domain and the frequency domain. In the time domain, $[1-\xi(L)]\,\mathbf{y}_t$ computes the approximate average difference between the current calendar month and all other calendar months within a window of a few years surrounding date t. Hence, $\xi(L)\,\mathbf{y}_t$ removes a local season-specific mean from $\mathbf{y}_t$. This property is meant to account for the fact that differences between the seasons may not be fixed. For example, an unusually high crop yield this summer may be positively correlated with an unusually high crop yield next summer, so the difference between summer and winter may be relatively high for a few years, and then relatively low for a few years. In the frequency domain, the filter suppresses spectral peaks at seasonal frequencies. 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:

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

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. To remove 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. According to Ramey (2016), one of the definitive characteristics of shocks is that "they should be exogenous with respect to the other current and lagged endogenous variables in the model" (pp. 74-75). 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 $\boldsymbol{\Psi}$ were known a priori: The econometrician would infer that the realized shocks were equal to $\boldsymbol{\Psi}\mathbf{e}_t^{sa}$, 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[6] relates the determinant of $\mathbf{Q}$ to the determinant of the spectral density:

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

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 the following relationship between the determinants of the matrices $\mathbf{Q}$ and $\mathbf{Q}^{sa}$, in terms of the X-11 gain function $\Xi(\omega)$ and the dimension of the time series n:

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

If only n_{sa} of the n variables are seasonally adjusted, then D should be raised to the power n_{sa}, rather than n, in the above expression. 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.[7] To put that magnitude in perspective, in a univariate setting ($n = 1$), 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. It may be tempting to 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 seasonality in the original time series is strong or mild. Note that the quantitative issue that arises in equation (5) 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

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

[7]Note that 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.

$\mathbf{y}_t^{sa}$. In equation (5), D depends on the log of the gain function; because the X-11 filter sets the gain function to zero at seasonal frequencies, the log gain goes to negative infinity. Consequently, an alternative seasonal-adjustment filter that is one-sided, but still zeros out the same frequencies as the X-11 filter, will typically lead to a large value of D as well.[8] 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.

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 $\mathbf{\Psi}^{sa\prime}\mathbf{\Lambda}^{sa}\mathbf{\Psi}^{sa} = \mathbf{Q}^{sa}$, equation (5) implies:

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

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 $\frac{\partial \mathbf{y}_t^{sa}}{\partial\left((\mathbf{\Lambda}^{sa})^{-1/2}\boldsymbol{\epsilon}_t^{sa}\right)'}$ 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

[8] For example, consider the alternative $\xi_{alt}(L) = 1 - \left(\frac{1}{q}\sum_{\ell=0}^{q-1} L^{n_s \ell}\right)\left(1 - \frac{1}{n_s}\sum_{k=0}^{n_s-1} L^k\right)$. This filter is one-sided, but retains two salient properties of the X-11 filter. First, $[1 - \xi_{alt}(L)]\,\mathbf{y}_t$ computes the average difference between the current calendar month and the annual average over the preceding q years, so $\xi_{alt}(L)\,\mathbf{y}_t$ subtracts a local season-specific average from $\mathbf{y}_t$. Second, the gain function $\Xi_{alt}(\omega) \equiv |\xi_{alt}(\exp\{i\omega\})|^2$ has zeros at the annual frequency and its harmonics $\left(\frac{2\pi}{n_s}j,\ j = 1, \ldots, \lfloor \frac{n_s}{2} \rfloor\right)$, so the filter suppresses spectral peaks at seasonal frequencies. For $\Xi_{alt}(\omega)$, the value of D in equation (5) is approximately 2.04 when $q = 3$ and approximately 1.68 when $q = 4$.

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 $\boldsymbol{\Psi}$ 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 $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^{sa}$ 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 response of variable k to shock k on impact:

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

Hence, seasonal adjustment creates large distortions in the contemporaneous 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 5. I will estimate a supply-and-demand model of the labor market using aggregate data on real wages and personhours. The model specifies:

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

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:

$$\Delta\log(\text{personhours}_t) = \eta_d \times \Delta\log(\text{real wage}_t) + \boldsymbol{\phi}^d(L)'\,\mathbf{y}_t + \epsilon_t^d \tag{9}$$

$$\Delta\log(\text{personhours}_t) = \eta_s \times \Delta\log(\text{real wage}_t) + \boldsymbol{\phi}^s(L)'\,\mathbf{y}_t + \epsilon_t^s, \tag{10}$$

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 5, 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 are allowed to 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 $t \overset{\mathrm{mod}\ n_s}{=} t'$. By extension, there will be n_s season-specific precision matrices for the reduced-form innovations, given by $\mathbf{Q}_t \equiv \mathbb{E}\left[\mathbf{e}_t \mathbf{e}_t'\right]^{-1} = \boldsymbol{\Psi}' \boldsymbol{\Lambda}_t \boldsymbol{\Psi}$. 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 k^{th} row of $\mathbf{\Psi}$ is proportional to the k^{th} eigenvector of $\mathbf{Q}_t\mathbf{Q}_{t'}^{-1}$. The novelty here is designating different seasons as different volatility regimes.

For a concrete illustration of how seasonal heteroskedasticity can inform identification, consider 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.[9] 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 relatively 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 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 simultaneously.[10] 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 5, 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.

The seasonal approach complements existing strategies for identification through heteroskedasticity. Rigobon (2003) shows that identification through heteroskedasticity is robust to some misspecification in the volatility regimes, as long as the average variance in one regime is different from the average variance in another. With a finite sample, though, applied researchers can face a practical tradeoff. Declaring too few regimes can obscure the identifying variation, by pooling together eras with different levels of volatility. Declaring too many regimes can produce less precise estimates of the variance matrices, because there will be relatively few observations per regime. Treating calendar months as volatility regimes presents a promising way of balancing these concerns: The number of observations for each calendar month grows linearly with the sample size, with approximately

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

[10] 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)

the same number of observations per regime. Seasonal and non-seasonal heteroskedasticity are not mutually exclusive, and it's an empirical question what volatility specification is most appropriate for a specific application. In Section 5, I will allow for both seasonal and non-seasonal heteroskedasticity in U.S. labor-market data, but the seasonal volatility regimes will provide more information for identifying supply and demand shocks.

3 Seasonal Priors

I will take a Bayesian approach to estimating equation (1). Section 3.1 provides a prior for the parameters that govern $\mathbf{s}_t$, Section 3.2 provides a prior for the parameters that govern $\tilde{\mathbf{y}}_t$, and Section 3.3 contains some discussion. Throughout, my goals will be twofold: (a) providing a flexible framework for articulating a range of prior beliefs about seasonality, and (b) explaining the specific prior beliefs that I will adopt for seasonal data.

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 $\frac{1}{n_s}\sum_{t=1}^{n_s}\mathbf{w}_t$ is zero, implying that $\frac{1}{n_s}\sum_{t=1}^{n_s}\mathbf{s}_t$ is zero as well. Hence, $\mathbf{s}_t$ should be interpreted as the season-specific deviation from a long-run average. Further, 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 researchers have more sharply defined beliefs about $\mathbf{s}_t$ than about $\mathbf{B}$. Eliciting a prior therefore requires a reverse-engineering exercise: What explicit prior over $\mathbf{B}$ will imply a reasonable prior predictive distribution over $\mathbf{s}_t$? For a class of priors to be useful, an econometrician needs to be able to specify beliefs about three aspects of deterministic seasonality, governed by a small number of easy-to-interpret hyperparameters. First, one should be able to declare the expected value of the time series for each season. Second, one should be able to control the prior correlation across seasons, or the expected smoothness of seasonal fluctuations. Third, one should be able to control how much deterministic seasonality is expected to contribute to the variance of the observed time series.

I will consider priors of the following form. Let $\mathbb{E}_{prior}[\cdot]$ denote expectations taken under the prior distribution. The first moments of the prior satisfy:

$$\mathbb{E}_{prior}[\mathbf{B}] = \bar{\mathbf{S}}\mathbf{C}, \tag{18}$$

where $\bar{\mathbf{S}}$ is an $n \times n_s$ matrix with columns summing to zero, and $\mathbf{C}$ is defined as the $n_s \times (n_s - 1)$ matrix whose (j,k) element is $\mathbf{C}_{j,k} \equiv \frac{\sqrt{2}}{n_s}\cos\left(\frac{2\pi}{n_s}jk - \frac{\pi}{4}\right)$. The second moments of the prior satisfy:

$$\mathbb{E}_{prior}\left[\mathrm{vec}\left(\mathbf{B} - \mathbb{E}_{prior}[\mathbf{B}]\right)\mathrm{vec}\left(\mathbf{B} - \mathbb{E}_{prior}[\mathbf{B}]\right)'\right] = \mathbf{K} \otimes \mathbf{V}_S, \tag{19}$$

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 matrices $\bar{\mathbf{S}}$, $\mathbf{K}$, and $\mathbf{V}_S$ have distinct, interpretable roles governing the level, smoothness, and magnitude of $\mathbf{s}_t$ under the prior predictive distribution.

Proposition 2. *Let $\bar{\mathbf{s}}_t$ denote the t^{th} column of $\bar{\mathbf{S}}$. Equation (18) implies:*

$$\mathbb{E}_{prior}[\mathbf{s}_t] = \bar{\mathbf{s}}_t. \tag{20}$$

Let $\mathbf{\Gamma}_u^s$ denote the prior predictive covariance between $\mathbf{s}_t$ and $\mathbf{s}_{t-u}$:

$$\mathbf{\Gamma}_u^s \equiv \mathbb{E}_{prior}\left[\frac{1}{n_s}\sum_{t=1}^{n_s}\left(\mathbf{s}_t - \bar{\mathbf{s}}_t\right)\left(\mathbf{s}_{t-u} - \bar{\mathbf{s}}_{t-u}\right)'\right]. \tag{21}$$

Equation (19) implies:

$$\mathbf{\Gamma}_u^s = \kappa_u \mathbf{V}_S, \tag{22}$$

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

Equation (20) demonstrates that the columns of $\bar{\mathbf{S}}$ determine the expected levels of deterministic seasonality. Adopting non-zero values of $\{\bar{\mathbf{s}}_t\}_{t=1}^{n_s}$ allows an econometrician to incorporate beliefs about the timing of seasonal peaks and troughs. 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. In other contexts, a researcher may want to be agnostic about the timing of seasonal cycles, which entails setting $\bar{\mathbf{S}}$ to a matrix of zeros: In that case, $\mathbb{E}_{prior}[\mathbf{s}_t]$ does not depend on t, so such a prior takes no stand on whether $\mathbf{s}_t$ is more likely to peak in summer or in winter.

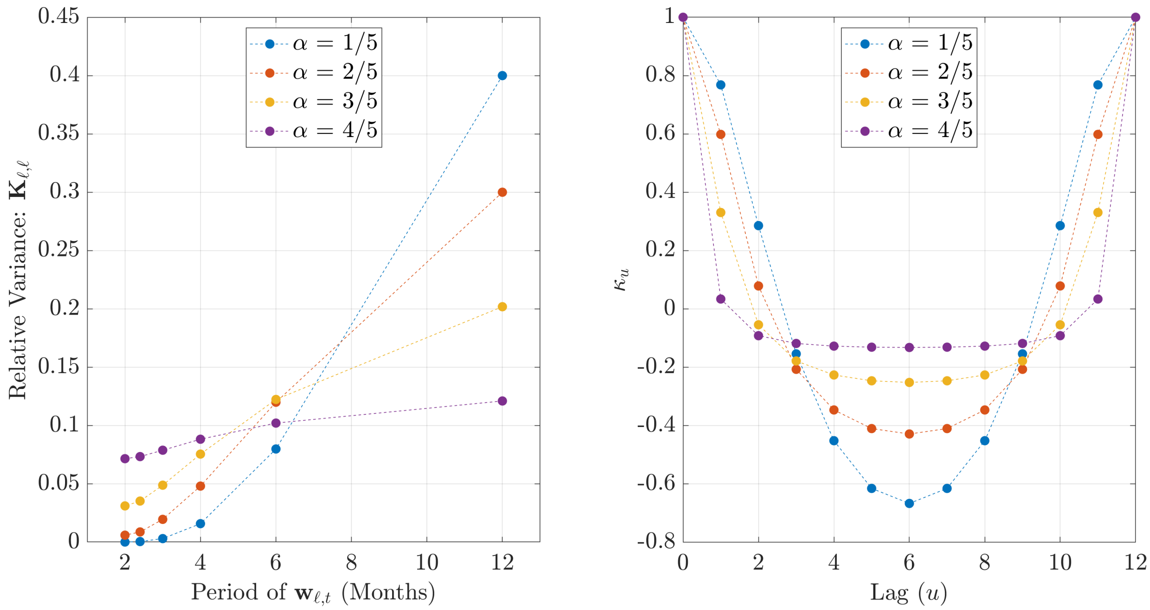
Regardless of whether the prior is agnostic about the timing of seasonal peaks and troughs, the prior will be informative about the persistence and magnitude of deterministic seasonality, as governed by $\mathbf{K}$ and $\mathbf{V}_S$. The matrix $\mathbf{K}$, by determining the κ_u coefficients in equation (22), controls the prior correlation across seasons, or the expected smoothness of deterministic seasonal cycles. Suppose for the moment that $\bar{\mathbf{S}}$ is a matrix of zeros, and assume that $\mathbf{K}$ is normalized such that $\operatorname{tr}\{\mathbf{K}\} = \kappa_0 = 1$. Then, $\{\kappa_u\}_{u\in\mathbb{Z}}$ is the prior predictive autocorrelation function for the elements of $\mathbf{s}_t$; the rate at which κ_u changes with u provides an indication of how smoothly $\mathbf{s}_t$ is expected to vary with t. My own prior is that, in many settings, adjacent seasons are more likely to relatively similar, while distant seasons are more likely to be relatively dissimilar. Two considerations inform this belief. First, meteorological factors are important for economic activity, and in many parts of the world, monthly averages of temperatures and precipitation vary smoothly over the course of the year. Second, adjustment costs and search frictions may induce households and firms to smooth out predictable seasonal changes, at least to some degree. In small samples, a smoothness prior effectively pools information across seasons: If a time series tends to be above its average in January and March, then it is likely to be above its average in February as well. In larger samples, this smoothing will play less of a role, as the data speaks about the differences between season-specific means. If $\bar{\mathbf{S}}$ is non-zero, reflecting an informed belief about the timing of seasonal cycles, then κ_u reflects the autocorrelation in $\mathbf{s}_t - \bar{\mathbf{s}}_t$, the deviation from the prior mean. Even if the path of $\bar{\mathbf{s}}_t$ is not smooth — recall the example of retail sales being expected to drop sharply after Christmas — the prior can allow $\mathbf{s}_t - \bar{\mathbf{s}}_t$ to be positively autocorrelated. In that sense, it's possible for this class of priors to pool information across seasons while simultaneously acknowledging that the seasons are expected to be different.

In practice, I will set $\mathbf{K}$ to be a diagonal matrix, with the ℓ^{th} element of the main diagonal given by:

$$\mathbf{K}_{\ell,\ell} \propto \alpha^{\ell} + \alpha^{n_s-\ell}, \tag{23}$$

normalized such that $\operatorname{tr}\{\mathbf{K}\} = \kappa_0 = 1$. 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 ℓ^{th} and $(n_s-\ell)^{th}$ columns of $\mathbf{B}$, and $\frac{n_s}{\ell}$ is the period of oscillation for the ℓ^{th} 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 equation (23). Adjacent months are expected to be positively correlated, and

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.

months that are half a year apart are expected to negatively correlated. When α is lower, $\mathbf{s}_t - \bar{\mathbf{s}}_t$ is expected to be smoother, in the sense of being more positively correlated across consecutive seasons.

The choice of $\mathbf{V}_S$ determines the expected contribution of $\mathbf{s}_t$ to the variance of $\mathbf{y}_t$. Assume that the prior treats $\mathbf{B}$ as independent of the parameters governing $\tilde{\mathbf{y}}_t$. If the prior predictive mean of $\tilde{\mathbf{y}}_t$ is zero, then 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 $\tilde{\mathbf{y}}_t$:

$$\mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t-\boldsymbol{\mu}\right)\left(\mathbf{y}_t-\boldsymbol{\mu}\right)'\right] = \mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_t'\right] + \mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t'\right]. \tag{24}$$

With the normalization $\kappa_0 = 1$, the first term on the right-hand side satisfies:

$$\mathbb{E}_{prior}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_t'\right] = \mathbf{V}_S + \frac{1}{n_s}\bar{\mathbf{S}}'\bar{\mathbf{S}}, \tag{25}$$

whenever T is divisible by n_s. (If T is not divisible by n_s, then the above expression is approximately true, and the approximation error is $O\left(T^{-1}\right)$.) 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 how a diffuse prior over $\mathbf{B}$ is not agnostic about the magnitude of seasonal cycles. 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. Instead, the choice of $\mathbf{V}_S$ should reflect expectations about the amount of variance attributable to deterministic seasonality. Let $\bar{\mathbf{\Sigma}}_y$ be a prior estimate for the variance of $\mathbf{y}_t$. I will specify $\mathbf{V}_S = \varsigma \bar{\mathbf{\Sigma}}_y$; when $\bar{\mathbf{S}}$ is zero, $\varsigma \in (0, 1)$ can be interpreted as the expected fraction of the variance of $\mathbf{y}_t$ that can be accounted for by $\mathbf{s}_t$.[11]

The most obvious alternative to using seasonal waveforms is using season-specific dummy variables, but simple priors for seasonal dummies can come with unintended, and possibly unappealing, features. In applied work, Bayesian practitioners often assume that the coefficients on seasonal dummies are mean-zero and uncorrelated under the prior. However, as explained in Appendix D, that approach implies that $\mathbf{s}_t$ is expected to exhibit negative serial correlation. 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. My prior is different: 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(p):

$$\mathbf{A}(L)\,\tilde{\mathbf{y}}_t = \boldsymbol{\epsilon}_t, \quad \mathbf{A}(L) \equiv \mathbf{\Psi} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell L^\ell, \quad \boldsymbol{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \mathbf{\Lambda}_t^{-1}\right), \tag{26}$$

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

$$\tilde{\mathbf{y}}_t = \sum_{\ell=1}^{p} \mathbf{\Psi}^{-1}\mathbf{\Phi}_\ell \tilde{\mathbf{y}}_{t-\ell} + \mathbf{e}_t, \quad \mathbf{e}_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \mathbf{Q}_t^{-1}\right), \tag{27}$$

where $\mathbf{e}_t \equiv \mathbf{\Psi}^{-1}\boldsymbol{\epsilon}_t$ and $\mathbf{Q}_t \equiv \mathbf{\Psi}'\mathbf{\Lambda}_t\mathbf{\Psi}$. Beliefs about $\mathbf{\Psi}$ depend on the specific application; here, I will simply assume that the econometrician has a marginal prior over $\mathbf{\Psi}$ and focus on the conditional prior over $\{\mathbf{\Phi}_\ell\}_{\ell=1}^{p}$ and $\{\mathbf{\Lambda}_t\}_{t=1}^{n_s}$. My prior will reflect a belief that random oscillations at seasonal

[11] If one expects deterministic seasonality to be more important for some series than others, then one can specify $\mathbf{V}_S = \mathrm{diag}(\varsigma_1, \ldots, \varsigma_n)^{\frac{1}{2}}\, \bar{\mathbf{\Sigma}}_y\, \mathrm{diag}(\varsigma_1, \ldots, \varsigma_n)^{\frac{1}{2}}$, so ς_k is the expected fraction of the variance attributable to deterministic seasonality for series k. This approach may be warranted in applications that combine strongly seasonal time series with series that are mostly aseasonal.

frequencies play an important part in accounting for the variation in $\tilde{\mathbf{y}}_t$. Again, this requires a reverse-engineering exercise: What explicit distribution over the VAR coefficients will imply a prior predictive distribution for the spectrum that favors peaks at seasonal frequencies? 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 k^{th} element of the main diagonal of $\mathbf{\Lambda}$, and let $\mathbf{\Phi} \equiv \left[\begin{array}{ccc} \mathbf{\Phi}_1 & \cdots & \mathbf{\Phi}_p \end{array} \right]$ concatenate the structural lag coefficients. One can write T observations from the process (26) as $\tilde{\mathbf{Y}}\mathbf{\Psi}' = \tilde{\mathbf{X}}\mathbf{\Phi}' + \boldsymbol{\mathcal{E}}$, where the t^{th} rows of $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{X}}$, and $\boldsymbol{\mathcal{E}}$ contain $\tilde{\mathbf{y}}_t'$, $\left(\tilde{\mathbf{y}}_{t-1}', \ldots, \tilde{\mathbf{y}}_{t-p}'\right)$, 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 $\left(\bar{\boldsymbol{\mathcal{E}}}\right)_{j,k} \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \lambda_k\right)$.[12] More explicitly, the prior implied by the dummy observations is:

$$
\begin{aligned}
\lambda_k \mid \mathbf{\Psi} &\sim \mathrm{G}\left(\alpha_\lambda, \beta_{\lambda,k}\right) && (28)\\
\mathrm{vec}\left(\mathbf{\Phi}\right) \mid \left\{\lambda_k\right\}_{k=1}^n, \mathbf{\Psi} &\sim \mathrm{N}\left(\mathrm{vec}\left(\bar{\mathbf{\Phi}}\right), \left(\bar{\mathbf{X}}'\bar{\mathbf{X}} \otimes \mathbf{\Lambda}\right)^{-1}\right) && (29)\\
\alpha_\lambda &\equiv \frac{\bar{T} - mn}{2} + 1 && (30)\\
\beta_{\lambda,k} &\equiv \frac{1}{2}\left[\left(\bar{\mathbf{Y}}\mathbf{\Psi}' - \bar{\mathbf{X}}\bar{\mathbf{\Phi}}'\right)'\left(\bar{\mathbf{Y}}\mathbf{\Psi}' - \bar{\mathbf{X}}\bar{\mathbf{\Phi}}'\right)\right]_{k,k} && (31)\\
\bar{\mathbf{\Phi}} &\equiv \mathbf{\Psi}\bar{\mathbf{Y}}'\bar{\mathbf{X}}\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}, && (32)
\end{aligned}
$$

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 processes that are close to having 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 $|z| = 1$. Any point on the complex unit circle z can be written in polar form as $z = \exp\left\{i\omega^*\right\}$, where $\omega^* \in [-\pi, \pi]$. The process $\tilde{\mathbf{y}}_t$ is said to have a *zero-frequency* unit root if $\left|\mathbf{A}\left(1\right)\right| = 0$, and $\tilde{\mathbf{y}}_t$ 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:

$$
\mathbf{f}\left(\omega\right) = \frac{1}{2\pi}\left[\mathbf{A}\left(\exp\left\{i\omega\right\}\right)' \mathbf{\Lambda}\mathbf{A}\left(\exp\left\{-i\omega\right\}\right)\right]^{-1}. \qquad (33)
$$

[12]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.

The function $\mathbf{f}(\omega)$ is a proper spectral density if $\tilde{\mathbf{y}}_t$ is stationary; otherwise, $\mathbf{f}(\omega)$ is understood to be a pseudo-spectral density. A seasonal unit root at frequency ω^* implies that $\lim_{\omega\to\omega^*}|\mathbf{f}(\omega)| = \infty$. For a univariate process, this means that the spectrum as a spike at ω^*, indicating that oscillations at frequency ω^* account for a substantial amount of the variation in the time series. If the process is stationary but with $|\mathbf{A}(\exp\{i\omega^*\})| \approx 0$, then the spectral density will have a peak in the vicinity of ω^*.

For the autoregressive coefficients, I will use stochastic linear restrictions to implement a prior belief in an approximate unit root at a specific frequency. To see what stochastic restrictions favor an approximate unit root, it is instructive to look at the restrictions that induce an exact unit root. 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 $\left\{\mathbf{\Psi}^{-1}\mathbf{\Phi}_\ell\right\}_{\ell=1}^p$ 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:

$$\mathbf{A}(\exp\{i\omega^*\}) = \mathbf{\Psi} - \sum_{\ell=1}^{p}\mathbf{\Phi}_\ell\cos(\omega^*\ell) - i\sum_{\ell=1}^{p}\mathbf{\Phi}_\ell\sin(\omega^*\ell). \tag{34}$$

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:

$$\mathbf{\Psi} = \sum_{\ell=1}^{p}\mathbf{\Phi}_\ell\cos(\omega^*\ell) \tag{35}$$

$$\mathbf{0}_{n\times n} = \sum_{\ell=1}^{p}\mathbf{\Phi}_\ell\sin(\omega^*\ell). \tag{36}$$

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:

$$\bar{\mathbf{Y}}_{\omega^*}\mathbf{\Psi}' = \bar{\mathbf{X}}_{\omega^*}\mathbf{\Phi}' + \bar{\boldsymbol{\mathcal{E}}}_{\omega^*}, \quad \left(\bar{\boldsymbol{\mathcal{E}}}_{\omega^*}\right)_{j,k} \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \lambda_k\right) \tag{37}$$

$$\bar{\mathbf{Y}}_{\omega^*} \equiv \tau_{\omega^*}\begin{bmatrix} \mathbf{I}_n \\ \mathbf{0}_{n\times n} \end{bmatrix} \tag{38}$$

$$\bar{\mathbf{X}}_{\omega^*} \equiv \tau_{\omega^*}\begin{bmatrix} \cos\left(\omega^* 1\right) & \cos\left(\omega^* 2\right) & \cdots & \cos\left(\omega^* p\right) \\ \sin\left(\omega^* 1\right) & \sin\left(\omega^* 2\right) & \cdots & \sin\left(\omega^* p\right) \end{bmatrix} \otimes \mathbf{I}_n, \tag{39}$$

where $\tau_{\omega^*} > 0$ is a scalar hyperparameter chosen by the econometrician. Equations (37)-(39) 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 $\mathbf{\Psi}$:

$$\operatorname{vec}\left(\mathbf{A}\left(\exp\left\{i\omega^*\right\}\right)\right) \mid \mathbf{\Lambda}, \mathbf{\Psi} \sim \mathrm{CN}\left(\mathbf{0}_{n^2\times 1}, \mathbf{I}_n \otimes \left(\frac{\tau_{\omega^*}^2}{2}\mathbf{\Lambda}\right)^{-1}\right). \tag{40}$$

The hyperparameter τ_{ω^*} controls the amount of prior confidence the econometrician has about the unit root at frequency ω^*.[13] In the limiting case, where $\tau_{\omega^*} \to \infty$, the prior imposes the restrictions (35) and (36) exactly. Otherwise, if equations (35) and (36) 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 (36) 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 (37)-(39) 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: one at the zero frequency and $n_s - 1$ at seasonal frequencies. Specifically, the seasonal (i.e., non-zero) roots of the operator $(1 - L^{n_s})$ take the form $\exp\left\{\pm\frac{2\pi}{n_s}ij\right\}$ for all positive integers j up to $\frac{n_s}{2}$. 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

[13] In some contexts, a researcher may have more confidence in a seasonal unit root for some variables than others. In that case, one could use the following dummy observations:

$$\bar{\mathbf{Y}}_{\omega^*} = \begin{bmatrix} \mathbf{T} \\ \mathbf{0}_{n\times n} \end{bmatrix}, \; \bar{\mathbf{X}}_{\omega^*} \equiv \begin{bmatrix} \cos\left(\omega^* 1\right) & \cos\left(\omega^* 2\right) & \cdots & \cos\left(\omega^* p\right) \\ \sin\left(\omega^* 1\right) & \sin\left(\omega^* 2\right) & \cdots & \sin\left(\omega^* p\right) \end{bmatrix} \otimes \mathbf{T},$$

with $\mathbf{T} = \operatorname{diag}\left(\tau_{\omega^*}^{(1)}, \ldots, \tau_{\omega^*}^{(n)}\right)$. This would imply:

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

of a year, etc.). Each frequency can receive its own set of dummy observations, as in equations (37)-(39), 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:

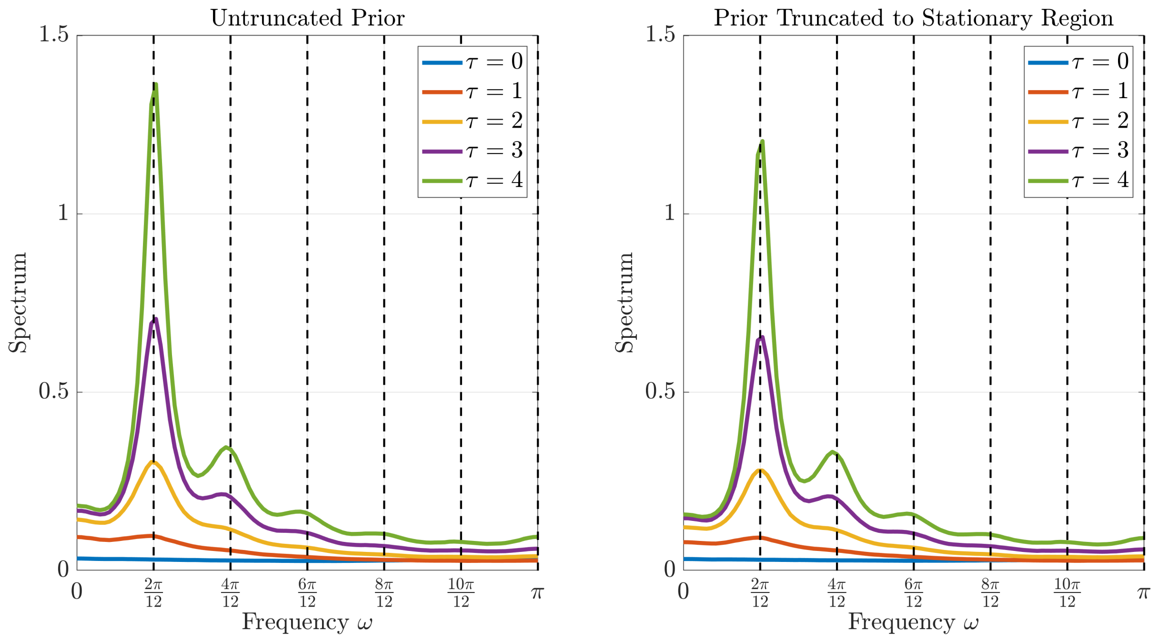
$$\bar{\mathbf{Y}} \equiv \begin{bmatrix} \bar{\mathbf{Y}}_0' & \bar{\mathbf{Y}}_{\omega_1^*}' & \cdots & \bar{\mathbf{Y}}_{\omega_{\lfloor n_s/2 \rfloor}^*}' \end{bmatrix}', \quad \bar{\mathbf{X}} \equiv \begin{bmatrix} \bar{\mathbf{X}}_0' & \bar{\mathbf{X}}_{\omega_1^*}' & \cdots & \bar{\mathbf{X}}_{\omega_{\lfloor n_s/2 \rfloor}^*}' \end{bmatrix}', \tag{41}$$

where $\bar{\mathbf{Y}}_{\omega_j^*}$ and $\bar{\mathbf{X}}_{\omega_j^*}$ take the form of equations (38) and (39), with $\omega_j^* = \frac{2\pi}{n_s} j$ and $\tau_{\omega_j^*} = \tau_S/j$ taking 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 $\text{vec}(\mathbf{\Phi})$ 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 expected 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.

Finally, I will note that the above strategy for favoring seasonal unit roots can be extended to favor seasonal cointegration, although it is not something I will apply in practice. 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 coin-

Figure 3: Prior Beliefs About the Spectrum



Notes: The figure shows the prior predictive 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.

tegrated 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 (38) and (39); like before, when $\omega^* \in \{0, \pi\}$, one would exclude the lower halves of equations (38) and (39). 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 5, my prior will favor seasonal unit roots, but not seasonal cointegration. However, the strategy described above provides a way to incorporate beliefs about seasonal cointegration for researchers whose priors or applications differ from my own.

The Seasonally Heteroskedastic Case. The following parameterization of the precision matrix allows for seasonal heteroskedasticity:

$$\log\left(\left(\boldsymbol{\Lambda}_t\right)_{k,k}\right) = \log\left(\lambda_k\right) + \boldsymbol{\rho}_k' \mathbf{w}_t, \tag{42}$$

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

$$\boldsymbol{\rho}_k \sim \mathrm{N}\left(\mathbf{0}_{(n_s-1)\times 1}, \nu_\rho \mathbf{K}\right), \tag{43}$$

independent across k, with $\mathbf{K}$ defined in equation (23). Assuming that the variance of $\boldsymbol{\rho}_k$ is proportional to $\mathbf{K}$ ensures that the seasonal component of $\log\left(\left(\boldsymbol{\Lambda}_t\right)_{k,k}\right)$ 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 $\nu_\rho > 0$, one can show that:

$$\mathbb{E}_{prior}\left[\frac{1}{n_s}\sum_{t=1}^{n_s}\left[\log\left(\left(\boldsymbol{\Lambda}_t\right)_{k,k}\right) - \log\left(\lambda_k\right)\right]^2\right] = \nu_\rho. \tag{44}$$

Hence, ν_ρ can be interpreted as the expected variance of $\log\left(\left(\boldsymbol{\Lambda}_t\right)_{k,k}\right)$ over the course of a calendar year.

The priors for $\{\lambda_k\}_{k=1}^{n}$ and $\boldsymbol{\Phi}$ 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.[14] 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:

$$\mathbf{f}\left(\omega\right) = \frac{1}{2\pi}\mathbf{A}\left(\exp\left\{-i\omega\right\}\right)^{-1}\boldsymbol{\Sigma}_\epsilon \mathbf{A}\left(\exp\left\{i\omega\right\}\right)^{-1\prime}, \quad \boldsymbol{\Sigma}_\epsilon \equiv \frac{1}{n_s}\sum_{t=1}^{n_s}\boldsymbol{\Lambda}_t^{-1}. \tag{45}$$

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

[14] More formally, let $\left\{\tilde{\mathbf{y}}_t^{(0)}\right\}_{t\in\mathbb{Z}}$ be a realization of the data-generating process given by equation (26), 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 (28)-(32), with the understanding that the $\mathbf{\Lambda}$ in equation (29) refers to $\text{diag}\,(\lambda_1, \ldots, \lambda_n)$.

3.3 Discussion

Implicit in the X-11 algorithm is a belief that seasonality is partly deterministic and partly stochastic, a belief that my prior incorporates explicitly. Given that deterministic and stochastic seasonality behave differently, it's worth recognizing how the approaches in Sections 3.1 and 3.2 complement one another. A potential limitation of treating seasonality as purely deterministic is that it would be too rigid, because the amplitude of seasonal cycles can shift. The prevalence of stochastic seasonality will determine the stability of seasonal patterns. If $\tilde{\mathbf{y}}_t$ is stationary but has spectral peaks at seasonal frequencies, then the difference in $\mathbf{y}_t$ between summer and winter will fluctuate over time; these seasonal shifts can be persistent, but not permanent, with $\mathbf{s}_t$ determining the long-run average for each season. If $\tilde{\mathbf{y}}_t$ has a seasonal unit root, then $\mathbf{y}_t$ no longer has a well defined long-run mean for each month, because the process is non-stationary. This fact raises a possible criticism of a prior that favors seasonal unit roots: An exact unit root would imply that summer and winter will eventually switch places as the process drifts.[15] For some variables, seasonality appears to be genuinely non-stationary: Wright (2013) gives the example of electricity usage, which used to peak in winter (from people using electric light) but now peaks in summer (from people using air conditioning). For other variables, it may seem more plausible that seasonal peaks and troughs tend to occur at fixed times of the year, but with the magnitudes of the peaks and troughs evolving in a stationary way. This is one reason why my prior does not enforce the seasonal unit roots dogmatically: A near-unit-root process can be a useful starting point for approximating seasonal fluctuations that shift over time, but if the data show that seasonal patterns have relatively stable long-run means, then this fact will be reflected in the likelihood function. The classic Minnesota prior, which favors zero-frequency unit roots, uses analogous reasoning: A random walk can be a serviceable local approximation to highly persistent processes, even for variables (such as the unemployment rate) that are not literally thought to be non-stationary.[16] If a researcher has a strong prior against seasonal unit roots, then one way to sidestep them is to truncate the prior to enforce stationarity, while still favoring stationary processes with near-unit roots at seasonal frequencies, as depicted in the right panel of Figure 3. It's also

[15] This property is easiest to see in the example $\tilde{y}_t = \tilde{y}_{t-12} + \epsilon_t$ for monthly data. The difference between, say, January and July will follow a random walk. More formally, Canova and Hansen (1995) demonstrate how testing for seasonal unit roots is equivalent to testing for whether season-specific intercepts in a regression model follow random walks.

[16] Doan, Litterman, and Sims (1984) write: "While we recognize that a more accurate representation of our prior beliefs would give less weight to systems with explosive roots than is implied by our symmetric distributions around this [random-walk] mean, we doubt that the gain that could be achieved by abandoning the Gaussian form for our prior would be worth the price."

possible to truncate the prior to preclude unit or explosive roots at seasonal frequencies, while still allowing for unit or explosive roots at the zero frequency. Whether truncated or not, this class of priors captures the belief that stochastic seasonal oscillations contribute to the observed variation in the data.

4 Estimation

The most general version of the model that I will estimate is:

$$\begin{aligned} \boldsymbol{\Psi}\left(\mathbf{y}_t - \mathbf{G}\mathbf{w}_t^m\right) &= \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \left(\mathbf{y}_{t-\ell} - \mathbf{G}\mathbf{w}_{t-\ell}^m\right) + \boldsymbol{\epsilon}_t & (46) \\ \boldsymbol{\epsilon}_t &\overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(\mathbf{0}_{n\times 1}, \operatorname{diag}\left(\lambda_1 \exp\left\{\boldsymbol{\rho}_1' \mathbf{w}_t^v\right\}, \ldots, \lambda_n \exp\left\{\boldsymbol{\rho}_n' \mathbf{w}_t^v\right\}\right)^{-1}\right). & (47) \end{aligned}$$

The above specification allows the deterministic variables entering the mean $\mathbf{w}_t^m$ to differ from the deterministic variables entering the variance $\mathbf{w}_t^v$. Equation (1) corresponds to the case where $\mathbf{w}_t^m \equiv (1, \mathbf{w}_t')'$ and $\mathbf{G} \equiv \begin{bmatrix} \boldsymbol{\mu} & \mathbf{B} \end{bmatrix}$. Equation (42) corresponds to the case where $\mathbf{w}_t^v$ only contains the seasonal sinusoids, but the application in Section 5 will consider other variables in $\mathbf{w}_t^v$ as well, to allow for non-seasonal heteroskedasticity.

The prior over $\mathbf{g} \equiv \operatorname{vec}(\mathbf{G})$ is normal:

$$\mathbf{g} \sim \mathrm{N}\left(\bar{\mathbf{g}}, \mathbf{V}_g\right), \quad \bar{\mathbf{g}} \equiv \begin{bmatrix} \bar{\boldsymbol{\mu}} \\ \mathbf{0}_{(n_s-1)n\times 1} \end{bmatrix}, \quad \mathbf{V}_g \equiv \begin{bmatrix} \mathbf{V}_\mu & \mathbf{0}_{n\times(n_s-1)n} \\ \mathbf{0}_{(n_s-1)n\times n} & \mathbf{K}\otimes\mathbf{V}_S \end{bmatrix}. \tag{48}$$

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 (41) 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 $\tilde{\mathbf{y}}_t \equiv \mathbf{y}_t - \mathbf{G}\mathbf{w}_t^m$ and $\tilde{\mathbf{x}}_t \equiv \left(\tilde{\mathbf{y}}_{t-1}', \ldots, \tilde{\mathbf{y}}_{t-p}'\right)'$. For $k \in \{1, \ldots, n\}$, define:*

$$\begin{aligned} \mathbf{Y}_{(k)} &\equiv \begin{bmatrix} \exp\left\{\tfrac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_1^v\right\}\tilde{\mathbf{y}}_1 & \cdots & \exp\left\{\tfrac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_T^v\right\}\tilde{\mathbf{y}}_T & \bar{\mathbf{Y}}' \end{bmatrix}' & (49) \\ \mathbf{X}_{(k)} &\equiv \begin{bmatrix} \exp\left\{\tfrac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_1^v\right\}\tilde{\mathbf{x}}_1 & \cdots & \exp\left\{\tfrac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_T^v\right\}\tilde{\mathbf{x}}_T & \bar{\mathbf{X}}' \end{bmatrix}'. & (50) \end{aligned}$$

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* k^{th} *column of* $\boldsymbol{\Phi}'$*, and let* $\boldsymbol{\psi}_k$ *denote the* k^{th} *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:*

$$\begin{aligned}
\lambda_k \mid \boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} &\sim \mathrm{G}\left(\hat{\alpha}_\lambda, \hat{\beta}_{\lambda,k}\right) && (51)\\
\boldsymbol{\phi}_k \mid \boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} &\sim \mathrm{N}\left(\hat{\boldsymbol{\phi}}_k, \left(\lambda_k \mathbf{X}_{(k)}' \mathbf{X}_{(k)}\right)^{-1}\right) && (52)\\
\hat{\boldsymbol{\phi}}_k &\equiv \left(\mathbf{X}_{(k)}' \mathbf{X}_{(k)}\right)^{-1} \mathbf{X}_{(k)}' \mathbf{Y}_{(k)} \boldsymbol{\psi}_k && (53)\\
\hat{\alpha}_\lambda &\equiv \frac{T + \bar{T} - mn}{2} + 1 && (54)\\
\hat{\beta}_{\lambda,k} &\equiv \frac{1}{2}\left(\mathbf{Y}_{(k)}\boldsymbol{\psi}_k - \mathbf{X}_{(k)}\hat{\boldsymbol{\phi}}_k\right)'\left(\mathbf{Y}_{(k)}\boldsymbol{\psi}_k - \mathbf{X}_{(k)}\hat{\boldsymbol{\phi}}_k\right), && (55)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] \quad &\propto \quad \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}[\boldsymbol{\rho}]\, \mathbb{P}[\mathbf{g}]\, \mathbb{P}[\boldsymbol{\eta}] \\
&\quad \times |\boldsymbol{\Psi}|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^n \boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^T \mathbf{w}_t^v\right)\right\} \prod_{k=1}^n \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}}, \qquad (56)
\end{aligned}$$

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:*

$$\mathbf{z}_t \equiv \left(\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-m}'\right)', \quad \tilde{\boldsymbol{\Phi}} \equiv \begin{bmatrix} \boldsymbol{\Psi} & -\boldsymbol{\Phi} \end{bmatrix}, \quad \mathbf{W}_t \equiv \begin{bmatrix} \mathbf{w}_t^m & \mathbf{w}_{t-1}^m & \cdots & \mathbf{w}_{t-m}^m \end{bmatrix}' \otimes \mathbf{I}_n. \quad (57)$$

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

$$\begin{aligned}
\mathbf{g} \mid \boldsymbol{\Phi}, \boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{\eta} &\sim \mathrm{N}\left(\hat{\mathbf{g}}, \hat{\mathbf{V}}_g\right) && (58)\\
\hat{\mathbf{V}}_g &\equiv \left(\mathbf{V}_g^{-1} + \sum_{t=1}^T \mathbf{W}_t' \tilde{\boldsymbol{\Phi}}' \boldsymbol{\Lambda}_t \tilde{\boldsymbol{\Phi}} \mathbf{W}_t\right)^{-1} && (59)\\
\hat{\mathbf{g}} &\equiv \hat{\mathbf{V}}_g \left(\mathbf{V}_g^{-1}\bar{\mathbf{g}} + \sum_{t=1}^T \mathbf{W}_t' \tilde{\boldsymbol{\Phi}}' \boldsymbol{\Lambda}_t \tilde{\boldsymbol{\Phi}} \mathbf{z}_t\right). && (60)
\end{aligned}$$

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 $\mathbf{\Phi}$, conditional on $(\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{g}, \boldsymbol{\rho})$ and $\mathbf{y}^T$.

5. Draw $\mathbf{g}$, conditional on $(\mathbf{\Phi}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\rho})$ 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 $\mathbf{\Phi}$, 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).[17] Consequently, applied researchers who have already implemented those estimation routines can incorporate seasonality with only modest modifications to their code. Appendix C contains additional computational details for implementing the posterior sampler.

5 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. Equations (8) and (46) imply a demand curve and a supply curve:

$$\begin{aligned}
\Delta \log(\text{personhours}_t) &= c_d + \eta_d \times \Delta \log(\text{real wage}_t) + \boldsymbol{\delta}_d' \mathbf{w}_t + \boldsymbol{\phi}^d(L)' \mathbf{y}_t + \epsilon_t^d && (61) \\
\Delta \log(\text{personhours}_t) &= c_s + \eta_s \times \Delta \log(\text{real wage}_t) + \boldsymbol{\delta}_s' \mathbf{w}_t + \boldsymbol{\phi}^s(L)' \mathbf{y}_t + \epsilon_t^s, && (62)
\end{aligned}$$

[17] 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 $\mathbf{\Psi}'\boldsymbol{\Lambda}\mathbf{\Psi}$ is a Wishart distribution.

where I have defined:

$$\begin{bmatrix} \boldsymbol{\phi}^d(L)' \\ \boldsymbol{\phi}^s(L)' \end{bmatrix} \equiv \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell L^\ell, \quad \begin{bmatrix} c_d \\ c_s \end{bmatrix} \equiv \left(\boldsymbol{\Psi} - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \right) \boldsymbol{\mu}, \quad \begin{bmatrix} \boldsymbol{\delta}_d' \\ \boldsymbol{\delta}_s' \end{bmatrix} \equiv \boldsymbol{\Psi}\mathbf{B} - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \mathbf{B}\mathbf{R}^\ell, \tag{63}$$

and $\mathbf{R}$ is a known rotation matrix that satisfies $\mathbf{w}_{t-1} = \mathbf{R}\mathbf{w}_t$, defined explicitly in the appendix. I will analyze two sets of estimates. First, I will use seasonally unadjusted data on wage growth and hours growth to fit equations (61) and (62); 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 (61) and (62), subject to the restriction that $\boldsymbol{\delta}_d$ and $\boldsymbol{\delta}_s$ are 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. For all non-seasonal aspects, my goal is to treat the two models as symmetrically as possible, so any discrepancies in the results will reflect the role of seasonality.

Using seasonally adjusted data, Baumeister and Hamilton (2015) assume that the shocks are homoskedastic, which renders their version of the model only partially identified. On the one hand, it's appealing that seasonal heteroskedasticity has the potential to provide full identification, as discussed in Section 2.2. On the other hand, it may not seem entirely fair to compare a heteroskedastic version of the model, which can be fully identified, with a homoskedastic version of the model, which cannot be fully identified. I will therefore allow the seasonally adjusted model to exhibit time-varying volatility, along the lines of Brunnermeier, Palia, Sastry, and Sims (2021). Those authors, based on their reading of U.S. economic history, declare seven volatility regimes during their sample from January 1973 to June 2015: oil crisis and stagflation (January 1973 to September 1979), Volcker disinflation (October 1979 to December 1982), savings and loan crisis (January 1983 to December 1989), great moderation (January 1990 to December 2007), financial crisis and great recession (January 2008 to December 2010), and zero lower bound and recovery (January 2011 to June 2015). Because my dataset runs from 1964 through 2019, I will extend the zero lower bound regime through December 2016, and I will add two additional regimes: pre-stagflation (January 1964 to December 1972) and interest-rate takeoff (January 2017 to December 2019).[18] For the seasonally adjusted model, the matrix of shock precisions $\boldsymbol{\Lambda}_t$ is constant within these regimes, but is allowed to differ across regimes. For the seasonal model, $\boldsymbol{\Lambda}_t$ is allowed to vary across low-frequency regimes and across calendar months; i.e., $\mathbf{w}_t^v$ in equation (47) contains both the seasonal sinusoids and indicator

[18] As a robustness check, I also estimated the model with only three low-frequency volatility regimes, along the lines of Carriero et al. (2021). Those authors declare the regime breaks to occur at the end of 1989 and the end of 2007. I find that the three-regime specification produces similar results to the eight-regime specification.

variables for the low-frequency regimes. As a point of reference, I will also produce estimates for homoskedastic versions of both models.

5.1 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 $\bar{\mathbf{S}}$ to zero, so the prior is agnostic about the timing of seasonal peaks and troughs. 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), report that deterministic seasonality accounts for an even higher fraction of the variance of some 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 diagonal of $\mathbf{\Lambda}_t$ coming from seasonality is expected to be .3 log points 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 $p = 13$ for monthly data. The series for hours and real wages, after being logged and differenced, are multiplied by 100 so that the elements of $\mathbf{y}_t$ can be interpreted as approximate growth rates. I will set the prior mean of $\boldsymbol{\mu}$ to $\left(\frac{1}{12}, \frac{2}{12}\right)'$, so the annual growth rates of real wages and aggregate hours are expected to be about 1% and 2%, respectively. The prior treats the elements of $\boldsymbol{\mu}$ as independent, with $\mathbb{V}_{prior}\left[\boldsymbol{\mu}_1\right] = \left(\frac{.5}{12}\right)^2$ and $\mathbb{V}_{prior}\left[\boldsymbol{\mu}_2\right] = \left(\frac{1}{12}\right)^2$, 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 (41) reflect standard priors that are commonly applied in the literature to seasonally adjusted data. Specifically:

$$\bar{\mathbf{Y}}_0 = \begin{bmatrix} \mathbf{0}_{np\times n} \\ \mathbf{1}_{\tau_\sigma \times 1} \otimes \operatorname{diag}\left(\hat{\sigma}_1, \ldots, \hat{\sigma}_n\right)^{-1} \end{bmatrix}, \quad \bar{\mathbf{X}}_0 = \begin{bmatrix} \frac{1}{\tau_0} \operatorname{diag}\left(1, \ldots, p\right) \otimes \operatorname{diag}\left(\hat{\sigma}_1, \ldots, \hat{\sigma}_n\right)^{-1} \\ \mathbf{0}_{n\times nm} \end{bmatrix}, \tag{64}$$

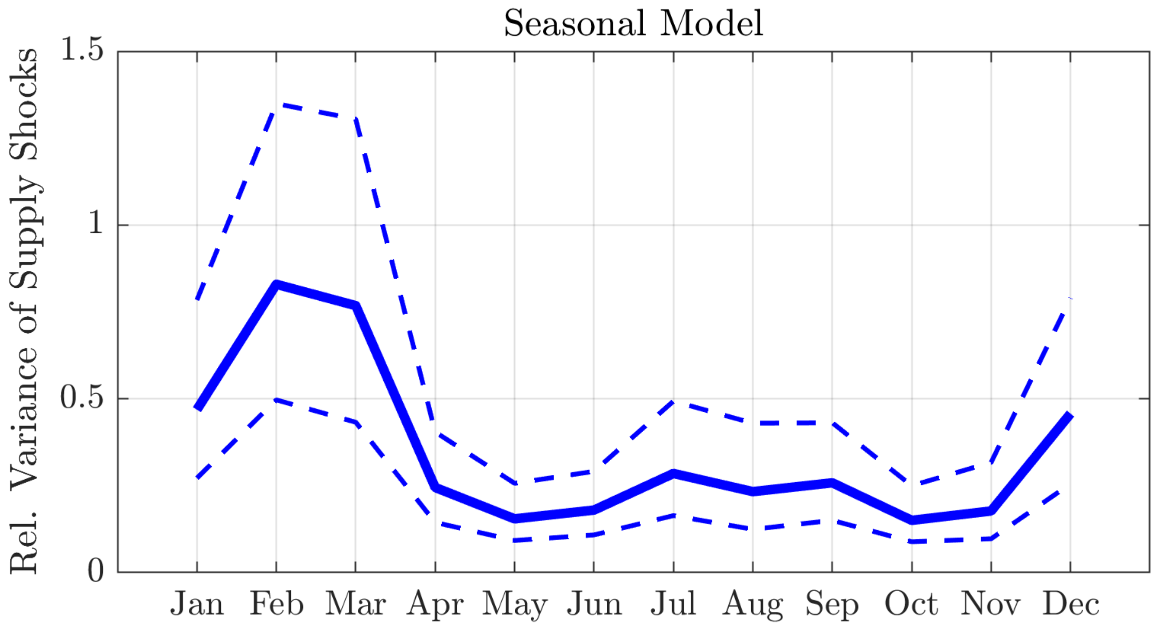
where $\hat{\sigma}_k$ is the standard deviation of the residuals from a univariate autoregression, augmented with seasonal dummy variables, fit to $\mathbf{y}_{k,t}$. (For the seasonally adjusted model, $\hat{\sigma}_k$ is the residual standard deviation from a univariate autoregression, without seasonal dummy variables, fit to the k^{th} seasonally adjusted variable.) If not for the additional dummy observations used to implement beliefs about seasonal unit roots, the first np 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 ℓ^{th} 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 $\bar{\mathbf{X}} = \bar{\mathbf{X}}_0$. For the non-seasonal heteroskedasticity, the coefficients on the low-frequency volatility regimes are uncorrelated with mean zero and variance .3.

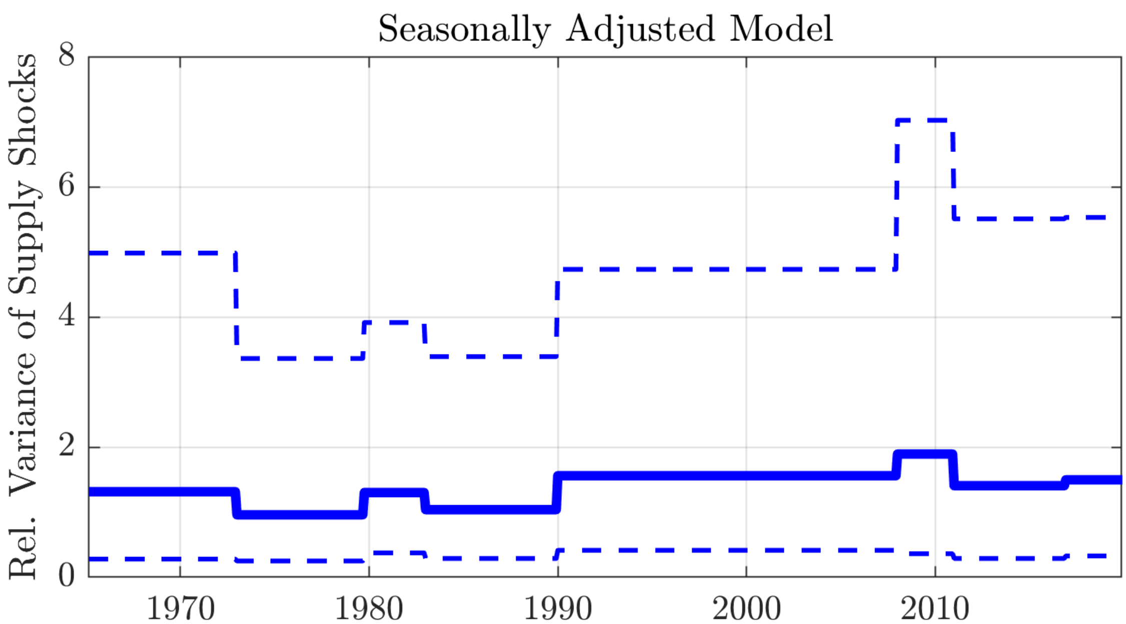
5.2 Data

The data are downloaded from FRED (fred.stlouisfed.org), and the Bureau of Labor Statistics (BLS) is the original source. 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 (CPI).[19] 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 by the BLS's attempts to account for those outliers in the seasonal adjustment routine.[20]

[19] 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 BLS only releases a seasonally adjusted version of that series, not a

Figure 4: Evidence of Identifying Information via Heteroskedasticity





Notes: Posterior estimates of $\mathbb{V}\left[\epsilon_t^s\right]/\mathbb{V}\left[\epsilon_t^d\right]$, the relative variance of supply shocks. The top panel shows the relative variance, as a function of the calendar month, in the seasonal model, holding fixed the low-frequency regime. The bottom panel shows the relative variance across low-frequency regimes in the seasonally adjusted model. The solid lines are the posterior medians; dashed lines are the 10^{th} and 90^{th} posterior quantiles.

Table 1: Estimated Structural Parameters

	η_d	η_s	$\frac{1}{T}\sum_t \mathbb{V}\left[\epsilon_t^d\right]$	$\frac{1}{T}\sum_t \mathbb{V}\left[\epsilon_t^s\right]$
Seasonal Model	-2.58 $[-3.27, -2.06]$	1.36 $[1.19, 1.55]$	2.00 $[1.49, 2.80]$	0.62 $[0.55, 0.71]$
Seasonally Adjusted Model	-1.21 $[-2.44, -0.67]$	1.59 $[0.79, 2.92]$	0.40 $[0.28, 0.94]$	0.53 $[0.30, 1.24]$

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

5.3 Results

The seasonal model and the seasonally adjusted model both allow for identification through heteroskedasticity, but in this application, seasonal volatility provides more identifying information than non-seasonal volatility. Recall that identification requires variation in the relative shock variances. Figure 4 plots $\mathbb{V}\left[\epsilon_t^s\right]/\mathbb{V}\left[\epsilon_t^d\right]$, the relative variance of supply shocks, as estimated by the two models. The top panel shows the how relative volatility changes across calendar months, holding fixed the low-frequency volatility regime, in the seasonal model. 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. The bottom panel of Figure 4 shows the relative volatility of supply shocks across low-frequency volatility regimes in the seasonally adjusted model. Although the point estimates exhibit some movement, the credible sets are very wide. In other words, the data are not very informative about how the relative volatility of the shocks changes across low-frequency regimes, so the seasonally adjusted model is only weakly identified. 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 displays the estimated elasticities and average shock variances, and there are large discrepancies in the structural parameters between the seasonal model and the seasonally adjusted model. The seasonal model's point estimate for η_d is about double 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

seasonally unadjusted version.

[20]The BLS took a number of steps to try to account for the extreme outliers in the 2020 data when performing seasonal adjustment; in particular, BLS 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 BLS to discern how to account for the extreme observations in seasonal-adjustment methods.

Table 2: Estimated Structural Parameters Under Homoskedasticity

	η_d	η_s	$\mathbb{V}\left[\epsilon_t^d\right]$	$\mathbb{V}\left[\epsilon_t^s\right]$
Seasonal Model	-2.19 $[-3.29, -1.52]$	1.46 $[1.15, 1.83]$	1.64 $[1.10, 2.85]$	0.66 $[0.55, 0.83]$
Seasonally Adjusted Model	-1.37 $[-2.28, -0.84]$	1.37 $[0.84, 2.20]$	0.49 $[0.34, 0.93]$	0.48 $[0.33, 0.86]$

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

(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.[21] 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.[22] 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 almost zero probability to $|\eta_d|$ being as low as 1.21, 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 the average variance of ϵ_t^d is five times larger than the seasonally adjusted point estimate. On the supply side, the discrepancy between the seasonal and seasonally adjusted models is not as drastic. The point estimates from the two models differ by about 15% for both η_s and $\frac{1}{T}\sum_t \mathbb{V}\left[\epsilon_t^s\right]$, with the point estimates from the seasonally adjusted model falling just outside the credible sets from the seasonal model.

The seasonal and seasonally adjusted models have two differences in their setups that can drive the differences in their results: (a) they are estimated using different time series, and (b) only the seasonal model incorporates seasonal heteroskedasticity to help identify the shocks. To shed some light on why the two models produce such divergent results, Table 2 presents estimates of the structural parameters when the models are restricted to be homoskedastic. Three observations are worth noting. First, the main qualitative patterns are unchanged: The seasonal model and the seasonally

[21] 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.

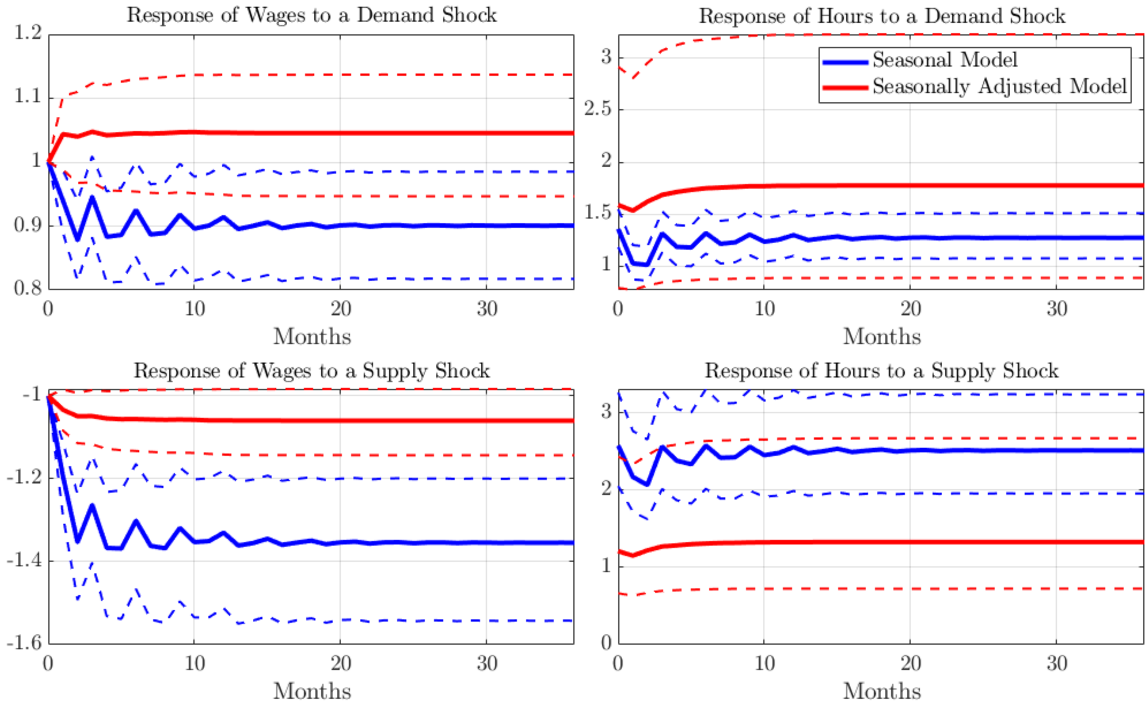
[22] Any comparison of elasticities across studies deserves the caveat that different datasets have different frequencies. Consequently, some papers are estimating how much workers and firms adjust their hours within a quarter, rather than within a month. Although the national income accounts are quarterly, Cooley and Prescott's approach is based on steady-state ratios, which are invariant to the frequency with which the data are observed. In the present application, the seasonal and seasonally adjusted models are both based on monthly time series, but produce divergent results.

adjusted model produce clear discrepancies, which are severe for the demand parameters, though more moderate for the supply parameters. Consequently, the differences between the two models in Table 1 cannot be attributed entirely to the fact that the heteroskedastic versions of the the models incorporate heteroskedasticity in different ways. Instead, the disagreements between the homoskedastic versions of the models suggest that a significant part of the discrepancy is coming from the seasonal adjustment itself. Second, in the seasonal model, allowing for heteroskedasticity modestly changes the point estimates while shrinking the credible sets for the structural parameters. This finding corroborates the earlier claim that seasonal heteroskedasticity is providing useful identifying information and facilitating more precise inferences. Third, in the seasonally adjusted model, allowing for heteroskedasticity has only a small effect on the point estimates, while inflating the credible sets. This finding likely reflects the fact that low-frequency heteroskedasticity is providing little identifying information, while adding free parameters that are imprecisely estimated.[23]

Figure 5 shows how the seasonal model and the seasonally adjusted model differ in their impulse responses. The magnitudes of the shocks are normalized to induce a one-percent change in wages on impact, and the plots show the approximate percent changes in the levels of wages and hours. In both models, by assumption, the instantaneous effect of a positive labor-demand shock is to raise hours and wages, while the instantaneous effect of a positive labor-supply shock is to raise hours and depress wages. However, the quantitative differences between the models' impulse responses are large. Aside from the contemporaneous response of wages, which is normalized to be the same across models, the seasonally adjusted point estimates fall outside the credible sets from the seasonal model, for all responses at all horizons. In both the short run and the long run, the seasonal model implies that hours are nearly twice as responsive to supply shocks, compared to the seasonally adjusted model. For the other impulse responses, the discrepancy between the models is smaller on impact but grows over time. After two years, the seasonal model's point estimate is about 14% smaller than the seasonally adjusted model's point estimate for the response of wages to a demand shock, 28% larger for the response of wages to a supply shock, 28% smaller for the response of hours to a demand shock, and 89% larger for the response of hours to a supply shock. In the medium run, the models show different dynamics: The seasonal model's impulse responses exhibit oscillations, whereas the seasonally adjusted model's impulse responses are smooth. Note that the impulse responses do not depend on the deterministic seasonal terms, so the oscillations in the impulse responses are indicative

[23] These results present an interesting contrast to Carriero et al. (2021). Using seasonally adjusted data, those authors add low-frequency volatility regimes to Baumeister and Hamilton's (2015) labor-market model. Carriero et al. find that, with their dataset, adding low-frequency heteroskedasticity can shift the point estimates while narrowing the credible sets. However, like Baumeister and Hamilton, they measure wages using the index of real hourly compensation in the nonfarm business sector, which the BLS releases exclusively in seasonally adjusted form. Because there is no seasonally unadjusted version of their data, it is not possible to assess the role of seasonality in Carriero et al.'s results.

Figure 5: Impulse Responses



Notes: The figures depict responses in the log levels of real wages and aggregate hours, multiplied by 100, so the results can be interpreted approximate percent changes. The shocks are normalized so that the contemporaneous change in wages is 1%. Solid lines are posterior median estimates; dashed lines are the 10^{th} and 90^{th} posterior quantiles.

of how the shocks are propagated, not predictable differences across the seasons.

The seasonal and seasonally adjusted models have different implications for 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 $\tilde{\mathbf{y}}_t$.[24] Table 3 displays the fraction of the variance that can be attributed to supply shocks for each 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 suggests that supply shocks are somewhat more 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

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

Table 3: Percentage of Variance Due to Supply Shocks

	Wage Growth		Hours Growth	
	Seasonal	Seasonally Adjusted	Seasonal	Seasonally Adjusted
Unconditional Variance	25 $[19, 32]$	57 $[24, 81]$	52 $[43, 60]$	43 $[19, 75]$
Low Frequencies	41 $[30, 53]$	57 $[24, 83]$	55 $[45, 64]$	42 $[18, 74]$
Business-Cycle Frequencies	40 $[29, 50]$	57 $[24, 82]$	55 $[46, 64]$	42 $[18, 75]$
Irregular Frequencies	23 $[17, 30]$	57 $[24, 81]$	51 $[43, 59]$	43 $[19, 75]$

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 10^{th} and 90^{th} 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.

that the majority of the variation in wage growth at business-cycle frequencies stems from demand shocks. The two models also exhibit differences in the variance decomposition of hours growth, though smaller than for wage growth. For all of the variance shares, the seasonal model's estimates are much more precise.

6 Taking Stock

The use of seasonally adjusted time series is so commonplace that most researchers do not articulate affirmative reasons for doing it. One rationale for feeding filtered data through a model is that the filter may be designed to eliminate certain behaviors that a particular model is simply incapable of replicating. Although that point may be relevant for some DSGE models, an advantage of VARs is their statistical flexibility for capturing rich seasonal patterns, both deterministic and stochastic. Another motivation for using seasonally adjusted time series is a desire to focus on non-seasonal phenomena, such as business cycles. That argument has a limitation: If the seasonally adjusted and unadjusted variables generate substantively similar results, then there's little reason not to use the unadjusted data — but if the seasonally adjusted and unadjusted variables generate substantively different results, then it's difficult to claim that seasonality does not matter for inferences about business cycles.

If given a choice, macroeconomists should prefer seasonally unadjusted variables over the seasonally adjusted versions when estimating structural VARs. That assertion, though broad, does not encompass two scenarios, both of which deserve further study: when researchers are not fitting VARs or when researchers do not have the option of using unadjusted data. Some form of seasonal adjustment may be convenient for visually inspecting a time series or exploring summary statistics. For econometric modeling, the relationship between seasonal adjustment and structural identification is less straightforward outside the context of VARs, especially for models that place strong (and possibly incorrect) restrictions on how the seasonality is specified. Sometimes, the most compelling and practical argument for using seasonally adjusted time series, with VARs or otherwise, is the absence of an alternative: Certain official statistics are released exclusively in seasonally adjusted form, or the unadjusted versions are only available over short time spans.[25] For example, macroeconomists studying productivity dynamics in the U.S. have little choice but to rely on seasonally adjusted variables and hope that any biases from seasonal adjustment are relatively mild for the estimates of interest; although that may not be ideal, it makes sense that researchers learn what they can from the data they have. In other cases, the availability of unadjusted data should be at least one factor in determining which time series enter a model. For example, there are well known pros and cons when deciding whether to measure inflation with the CPI or the PCE index; an advantage of the CPI is that it has long been published in seasonally unadjusted form, unlike the PCE index. One possible avenue of future research could be to assemble unadjusted versions of widely 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

Ashley, R. and R. Verbrugge (2022). Death to regression modeling on bandpass-filtered data: New results on the general undesirability of filtering dynamic regressions, with a proposed solution. Presented at The Australasian Meetings of the Econometric Society, July 7, 2022.

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

[25] The Bureau of Economic Analysis (BEA) only began publishing non-seasonally adjusted quarterly estimates of GDP and its components in 2018; the real versions of these variables are backdated to 2002, though the nominal versions are backdated to 1947. The BEA's monthly index of personal consumption expenditures (PCE) is not published in seasonally unadjusted form, nor is the associated monthly inflation index. Likewise, the productivity and costs release, constructed jointly by the BEA and BLS, exclusively contains seasonally adjusted series; such variables include labor productivity, total factor productivity, and the labor share. Fortunately, most BLS statistics derived from monthly surveys of households and firms are fully available in seasonally unadjusted from.

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. *Comminsssioner'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.

Carriero, A., M. Marcellino, and T. Tornese (2021). Heteroskedasticity as a complementary identification strategy. Technical report.

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., 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.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics 100*(5), 831–843.

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., 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.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. Published in *The Handbook of Macroeconomics*, vol. 2A, ed. J. Taylor and H. Uhlig, Elsevier, pp. 71–162.

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.

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.

Shiller, R. J. (1973). A distributed lag estimator derived from smoothness priors. *Econometrica: Journal of the Econometric Society*, 775–788.

Sims, C. A. (1974). Seasonality in regression. *Journal of the American Statistical Association 69*(347), 618–626.

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.

Wallis, K. F. (1974). Seasonal adjustment and relations between variables. *Journal of the American Statistical Association 69*(345), 18–31.

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.

Wright, J. H. (2013). Unseasonal seasonals? *Brookings Papers on Economic Activity 2013*(2), 65–126.

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 (For Online Publication)

A.1 Proof of Proposition 1

Suppose that $(\mathbf{\Psi}, \mathbf{\Lambda}) = \mathcal{I}(\mathbf{Q}) = \mathcal{I}(\mathbf{Q}^{sa})$, or in the case where the model is partially identified, $(\mathbf{\Psi}, \mathbf{\Lambda}) \in \mathcal{I}(\mathbf{Q}) \cap \mathcal{I}(\mathbf{Q}^{sa})$. 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

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

It remains to characterize $\mathbf{\Gamma}_u^s$. The vector of waveforms $\mathbf{w}_t$ has two convenient properties:

$$\frac{1}{n_s}\sum_{t=1}^{n_s}\mathbf{w}_t\mathbf{w}_t' = \mathbf{I}_{n_s-1}, \quad \mathbf{w}_{t-1} = \mathbf{R}\mathbf{w}_t, \tag{65}$$

where $\mathbf{R}$ is a known orthogonal matrix defined in Lemma 1. Define $\tilde{\mathbf{B}} \equiv \mathbf{B} - \mathbb{E}_{prior}[\mathbf{B}]$. We can write:

$$\begin{aligned}
\frac{1}{n_s}\sum_{t=1}^{n_s}(\mathbf{s}_t - \bar{\mathbf{s}}_t)(\mathbf{s}_{t-u} - \bar{\mathbf{s}}_{t-u})' &= \frac{1}{n_s}\sum_{t=1}^{n_s}\left(\mathbf{B}\mathbf{w}_t - \bar{\mathbf{S}}\mathbf{C}\mathbf{w}_t\right)\left(\mathbf{B}\mathbf{R}^u\mathbf{w}_t - \bar{\mathbf{S}}\mathbf{C}\mathbf{R}^u\mathbf{w}_t\right)' \\
&= \left(\mathbf{B} - \bar{\mathbf{S}}\mathbf{C}\right)\left(\frac{1}{n_s}\sum_{t=1}^{n_s}\mathbf{w}_t\mathbf{w}_t'\right)\mathbf{R}^{u\prime}\left(\mathbf{B} - \bar{\mathbf{S}}\mathbf{C}\right)' \\
&= \tilde{\mathbf{B}}\mathbf{R}^{u\prime}\tilde{\mathbf{B}}'.
\end{aligned} \tag{66}$$

Observe that the (j,k) element of this matrix is given by:

$$\left(\tilde{\mathbf{B}}\mathbf{R}^{u\prime}\tilde{\mathbf{B}}'\right)_{j,k} = \sum_{\ell=1}^{n_s-1}\sum_{h=1}^{n_s-1}\tilde{\mathbf{B}}_{j,h}\left(\mathbf{R}^{u\prime}\right)_{h,\ell}\left(\tilde{\mathbf{B}}'\right)_{\ell,k}. \tag{67}$$

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

$$\left(\mathbf{R}^{u\prime}\right)_{h,\ell} = \left(\mathbf{R}^u\right)_{\ell,h} = \cos\left(\frac{2\pi}{n_s}\ell u\right)\mathbb{I}[\ell = h] - \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right)\mathbb{I}[h = n_s - \ell]. \tag{68}$$

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

$$\begin{aligned}
\left(\tilde{\mathbf{B}}\mathbf{R}^{u\prime}\tilde{\mathbf{B}}'\right)_{j,k} &= \sum_{\ell=1}^{n_s-1}\sum_{h=1}^{n_s-1}\tilde{\mathbf{B}}_{j,h}\tilde{\mathbf{B}}_{k,\ell}\left(\cos\left(\frac{2\pi}{n_s}\ell u\right)\mathbb{I}[\ell = h] - \cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right)\mathbb{I}[h = n_s - \ell]\right) \\
&= \sum_{\ell=1}^{n_s-1}\tilde{\mathbf{B}}_{j,\ell}\tilde{\mathbf{B}}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u\right) - \sum_{\ell=1}^{n_s-1}\tilde{\mathbf{B}}_{j,n_s-\ell}\tilde{\mathbf{B}}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u - \frac{\pi}{2}\right).
\end{aligned} \tag{69}$$

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 (19) implies that $\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,\ell}\tilde{\mathbf{B}}_{k,\ell}\right] =$

$\mathbf{K}_{\ell,\ell}\left(\mathbf{V}_S\right)_{j,k}$. Hence:

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

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

Equation (19) implies that $\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,n_s-\ell}\tilde{\mathbf{B}}_{k,\ell}\right] = \mathbf{K}_{n_s-\ell,\ell}\left(\mathbf{V}_S\right)_{j,k}$, so:

$$\mathbb{E}_{prior}\left[\sum_{\ell=1}^{n_s-1}\tilde{\mathbf{B}}_{j,n_s-\ell}\tilde{\mathbf{B}}_{k,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\right] = \sum_{\ell=1}^{n_s-1}\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,n_s-\ell}\tilde{\mathbf{B}}_{k,\ell}\right]\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)$$
$$= \left[\sum_{\ell=1}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\right]\left(\mathbf{V}_S\right)_{j,k}. \tag{71}$$

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 $\ell > \frac{n_s}{2}$:

$$\begin{aligned}
\sum_{\ell=1}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right) &= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\sum_{\ell=\frac{n_s+1}{2}}^{n_s-1}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\\
&= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)\\
&\quad+\sum_{\ell'=1}^{\frac{n_s-1}{2}}\mathbf{K}_{\ell',n_s-\ell'}\cos\left(\frac{2\pi}{n_s}\left(n_s-\ell'\right)u-\frac{\pi}{2}\right)\\
&= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\sum_{\ell'=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell',\ell'}\cos\left(\frac{2\pi}{n_s}\ell' u+\frac{\pi}{2}\right)\\
&= \sum_{\ell=1}^{\frac{n_s-1}{2}}\mathbf{K}_{n_s-\ell,\ell}\left[\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)+\cos\left(\frac{2\pi}{n_s}\ell u+\frac{\pi}{2}\right)\right]\\
&= 0,
\end{aligned} \tag{72}$$

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 $\ell=\frac{n_s}{2}$:

$$\mathbf{K}_{n_s-\ell,\ell}\cos\left(\frac{2\pi}{n_s}\ell u-\frac{\pi}{2}\right)=\mathbf{K}_{\frac{n_s}{2},\frac{n_s}{2}}\cos\left(\pi u-\frac{\pi}{2}\right)=0, \tag{73}$$

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

$$\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T\mid\boldsymbol{\theta}\right] \;\; = \;\; & (2\pi)^{-\frac{Tn}{2}}\left|\mathbf{\Psi}\right|^T\exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n}\boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^{T}\mathbf{w}_t^v\right)\right\}\left(\prod_{k=1}^{n}\lambda_k^{\frac{T}{2}}\right)\\
&\times\exp\left\{-\sum_{k=1}^{n}\frac{\lambda_k}{2}\boldsymbol{\psi}_k'\dot{\mathbf{Y}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k\right\}\\
&\times\exp\left\{-\sum_{k=1}^{n}\frac{\lambda_k}{2}\left(-2\boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{Y}}_{(k)}\boldsymbol{\psi}_k+\boldsymbol{\phi}_k'\dot{\mathbf{X}}_{(k)}'\dot{\mathbf{X}}_{(k)}\boldsymbol{\phi}_k\right)\right\},
\end{aligned} \tag{74}$$

where the t^{th} rows of $\dot{\mathbf{Y}}_{(k)}$ and $\dot{\mathbf{X}}_{(k)}$ are $\exp\left\{\frac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_t^v\right\}\tilde{\mathbf{y}}_t'$ and $\exp\left\{\frac{1}{2}\boldsymbol{\rho}_k'\mathbf{w}_t^v\right\}\tilde{\mathbf{x}}_t'$. Lemma 3 establishes that the conditional prior for $\mathbf{\Phi}$ and $\boldsymbol{\lambda}$ can be written as:

$$\begin{aligned}
\mathbb{P}\left[\mathbf{\Phi},\boldsymbol{\lambda}\mid\boldsymbol{\eta}\right] \;\; = \;\; & \Gamma\left(\alpha_\lambda\right)^{-n}(2\pi)^{-\frac{n^2m}{2}}\left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}}\left(\prod_{k=1}^{n}\beta_{\lambda,k}^{\alpha_\lambda}\right)\\
&\times\left(\prod_{k=1}^{n}\lambda_k^{\frac{\bar{T}}{2}}\right)\exp\left\{-\sum_{k=1}^{n}\frac{\lambda_k}{2}\boldsymbol{\psi}_k'\bar{\mathbf{Y}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right\}\\
&\times\exp\left\{-\sum_{k=1}^{n}\frac{\lambda_k}{2}\left(\boldsymbol{\phi}_k'\bar{\mathbf{X}}'\bar{\mathbf{X}}\boldsymbol{\phi}_k-2\boldsymbol{\phi}_k'\bar{\mathbf{X}}'\bar{\mathbf{Y}}\boldsymbol{\psi}_k\right)\right\}.
\end{aligned} \tag{75}$$

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:

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

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:

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

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:

$$
\begin{aligned}
\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \times & \\
\mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] \quad = \quad & (2\pi)^{-\frac{Tn}{2}} \left(\frac{\Gamma\left(\hat{\alpha}_\lambda\right)}{\Gamma\left(\alpha_\lambda\right)}\right)^n \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \\
& \times \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^{T} \mathbf{w}_t^v\right)\right\} \prod_{k=1}^{n} \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{\frac{1}{2}}} \\
& \times \left[\prod_{k=1}^{n} \frac{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}}{\Gamma\left(\hat{\alpha}_\lambda\right)} \lambda_k^{\hat{\alpha}_\lambda - 1} \exp\left\{-\lambda_k \hat{\beta}_{\lambda,k}\right\}\right] \\
& \times \left[\prod_{k=1}^{n} \frac{\left|\lambda_k \mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{\frac{1}{2}}}{(2\pi)^{\frac{nm}{2}}} \exp\left\{-\frac{\lambda_k}{2}\left(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k\right)' \mathbf{X}'_{(k)}\mathbf{X}_{(k)} \left(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k\right)\right\}\right]. \quad (78)
\end{aligned}
$$

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

$$
\prod_{k=1}^{n} \mathrm{G}\left(\lambda_k \mid \hat{\alpha}_\lambda, \hat{\beta}_{\lambda,k}\right) \mathrm{N}\left(\boldsymbol{\phi}_k \mid \hat{\boldsymbol{\phi}}_k, \left(\lambda_k \mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right), \quad (79)
$$

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:

$$
\begin{aligned}
\mathbb{P}\left[\boldsymbol{\theta} \mid \mathbf{y}^T\right] \quad \propto \quad & \mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] \mathbb{P}\left[\boldsymbol{\rho}\right] \mathbb{P}\left[\mathbf{g}\right] \mathbb{P}\left[\boldsymbol{\eta}\right] \\
= \quad & (2\pi)^{-\frac{Tn}{2}} \left(\frac{\Gamma\left(\hat{\alpha}_\lambda\right)}{\Gamma\left(\alpha_\lambda\right)}\right)^n \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}\left[\boldsymbol{\rho}\right] \mathbb{P}\left[\mathbf{g}\right] \mathbb{P}\left[\boldsymbol{\eta}\right] \\
& \times \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^{T} \mathbf{w}_t^v\right)\right\} \prod_{k=1}^{n} \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{1/2}} \\
& \times \left[\prod_{k=1}^{n} \mathrm{G}\left(\lambda_k \mid \hat{\alpha}_\lambda, \hat{\beta}_{\lambda,k}\right) \mathrm{N}\left(\boldsymbol{\phi}_k \mid \hat{\boldsymbol{\phi}}_k, \left(\lambda_k \mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right)^{-1}\right)\right]. \quad (80)
\end{aligned}
$$

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

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] &= \frac{C}{\mathbb{P}\left[\mathbf{y}^T\right]} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}[\boldsymbol{\rho}]\,\mathbb{P}[\mathbf{g}]\,\mathbb{P}[\boldsymbol{\eta}] \\ &\quad \times |\boldsymbol{\Psi}|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n} \boldsymbol{\rho}_k\right)'\left(\sum_{t=1}^{T} \mathbf{w}_t^v\right)\right\} \prod_{k=1}^{n} \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}'_{(k)}\mathbf{X}_{(k)}\right|^{1/2}}, \end{aligned} \tag{81}$$

where $\mathbb{P}\left[\mathbf{y}^T\right] = \int \mathbb{P}\left[\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta} \mid \mathbf{y}^T\right] d(\boldsymbol{\rho}, \mathbf{g}, \boldsymbol{\eta})$ 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:

$$\boldsymbol{\Psi}\mathbf{y}_t - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \mathbf{y}_{t-\ell} = \boldsymbol{\Psi}\mathbf{G}\mathbf{w}_t^m - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \mathbf{G}\mathbf{w}_{t-\ell}^m + \boldsymbol{\epsilon}_t. \tag{82}$$

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

$$\boldsymbol{\Psi}\mathbf{y}_t - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \mathbf{y}_{t-\ell} = \left(\boldsymbol{\Psi}\left(\mathbf{w}_t^{m\prime} \otimes \mathbf{I}_n\right) - \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell\left(\mathbf{w}_{t-\ell}^{m\prime} \otimes \mathbf{I}_n\right)\right)\mathbf{g} + \boldsymbol{\epsilon}_t. \tag{83}$$

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

$$\mathbf{z}_t \equiv \left(\mathbf{y}'_t, \ldots, \mathbf{y}'_{t-m}\right)', \ \tilde{\boldsymbol{\Phi}} \equiv \begin{bmatrix} \boldsymbol{\Psi} & -\boldsymbol{\Phi} \end{bmatrix}, \ \mathbf{W}_t \equiv \begin{bmatrix} \mathbf{w}_t^m & \mathbf{w}_{t-1}^m & \cdots & \mathbf{w}_{t-m}^m \end{bmatrix}' \otimes \mathbf{I}_n. \tag{84}$$

We can write the likelihood as:

$$\begin{aligned}\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= \prod_{t=1}^{T} (2\pi)^{-\frac{n}{2}} \det\left(\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t\boldsymbol{\Psi}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)'\boldsymbol{\Lambda}_t\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)'\boldsymbol{\Lambda}_t\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)\right\}, \end{aligned} \tag{85}$$

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:

$$\begin{aligned}\sum_{t=1}^{T}\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right)'\boldsymbol{\Lambda}_t\left(\tilde{\boldsymbol{\Phi}}\mathbf{z}_t - \tilde{\boldsymbol{\Phi}}\mathbf{W}_t\mathbf{g}\right) &= \left(\sum_{t=1}^{T} \mathbf{z}'_t\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{z}_t\right) - 2\left(\sum_{t=1}^{T} \mathbf{z}'_t\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} \\ &\quad + \mathbf{g}\left(\sum_{t=1}^{T} \mathbf{W}'_t\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}. \end{aligned} \tag{86}$$

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

$$\mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] \propto \exp\left\{-\frac{1}{2}\left[\mathbf{g}\left(\sum_{t=1}^{T}\mathbf{W}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} - 2\left(\sum_{t=1}^{T}\mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}\right]\right\}. \tag{87}$$

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

$$\begin{aligned}\mathbb{P}[\mathbf{g}] &= (2\pi)^{-\frac{(n_s-1)n}{2}}\left|\mathbf{V}_g\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{g}-\bar{\mathbf{g}})'\,\mathbf{V}_g^{-1}(\mathbf{g}-\bar{\mathbf{g}})\right\}\\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{g}'\mathbf{V}_g^{-1}\mathbf{g} - 2\bar{\mathbf{g}}'\mathbf{V}_g^{-1}\mathbf{g}\right)\right\},\end{aligned} \tag{88}$$

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:

$$\begin{aligned}\mathbb{P}\left[\mathbf{g} \mid \mathbf{y}^T\right] &\propto \mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right]\mathbb{P}[\mathbf{g}]\\ &\propto \exp\left\{-\frac{1}{2}\left[\mathbf{g}\left(\mathbf{V}_g^{-1}+\sum_{t=1}^{T}\mathbf{W}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g} - 2\left(\bar{\mathbf{g}}'\mathbf{V}_g^{-1}+\sum_{t=1}^{T}\mathbf{z}_t'\tilde{\boldsymbol{\Phi}}'\boldsymbol{\Lambda}_t\tilde{\boldsymbol{\Phi}}\mathbf{W}_t\right)\mathbf{g}\right]\right\}\\ &= \exp\left\{-\frac{1}{2}\left(\mathbf{g}\hat{\mathbf{V}}_g^{-1}\mathbf{g} - 2\hat{\mathbf{g}}'\hat{\mathbf{V}}_g^{-1}\mathbf{g}\right)\right\}\\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{g}-\hat{\mathbf{g}})'\,\hat{\mathbf{V}}_g^{-1}(\mathbf{g}-\hat{\mathbf{g}})\right\},\end{aligned} \tag{89}$$

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 $\hat{\mathbf{V}}_g$.

A.4 Lemmata

The proofs of the propositions use the following lemmata. For brevity, I have omitted the proofs, which appeared in an earlier version of the paper and are available upon request.

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:*

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

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:*

$$(\mathbf{R}^u)_{j,k} = \cos\left(\frac{2\pi}{n_s}ju\right)\mathbb{I}[j=k] - \cos\left(\frac{2\pi}{n_s}ju - \frac{\pi}{2}\right)\mathbb{I}[k=n_s-j]. \tag{91}$$

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

$$\begin{aligned} \mathbb{P}\left[\mathbf{y}^T \mid \boldsymbol{\theta}\right] &= (2\pi)^{-\frac{Tn}{2}} \left|\boldsymbol{\Psi}\right|^T \exp\left\{\frac{1}{2}\left(\sum_{k=1}^n \boldsymbol{\rho}_k\right)' \left(\sum_{t=1}^T \mathbf{w}_t^v\right)\right\} \left(\prod_{k=1}^n \lambda_k^{\frac{T}{2}}\right) \\ &\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2} \boldsymbol{\psi}_k' \dot{\mathbf{Y}}_{(k)}' \dot{\mathbf{Y}}_{(k)} \boldsymbol{\psi}_k\right\} \\ &\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2} \left(-2\boldsymbol{\phi}_k' \dot{\mathbf{X}}_{(k)}' \dot{\mathbf{Y}}_{(k)} \boldsymbol{\psi}_k + \boldsymbol{\phi}_k' \dot{\mathbf{X}}_{(k)}' \dot{\mathbf{X}}_{(k)} \boldsymbol{\phi}_k\right)\right\}. \end{aligned} \tag{92}$$

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

$$\begin{aligned} \mathbb{P}\left[\boldsymbol{\Phi}, \boldsymbol{\lambda} \mid \boldsymbol{\eta}\right] &= \Gamma\left(\alpha_\lambda\right)^{-n} (2\pi)^{-\frac{n^2 m}{2}} \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \left(\prod_{k=1}^n \beta_{\lambda,k}^{\alpha_\lambda}\right) \\ &\quad \times \left(\prod_{k=1}^n \lambda_k^{\frac{\bar{T}}{2}}\right) \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2} \boldsymbol{\psi}_k' \bar{\mathbf{Y}}'\bar{\mathbf{Y}} \boldsymbol{\psi}_k\right\} \\ &\quad \times \exp\left\{-\sum_{k=1}^n \frac{\lambda_k}{2} \left(\boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{X}} \boldsymbol{\phi}_k - 2\boldsymbol{\phi}_k' \bar{\mathbf{X}}'\bar{\mathbf{Y}} \boldsymbol{\psi}_k\right)\right\}. \end{aligned} \tag{93}$$

B 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:

$$A(L)\,\tilde{A}\left(L^{12}\right)\left(1 - L^d\right)\left(1 - L^{12\tilde{d}}\right)\left(y_t - \boldsymbol{\beta}'\mathbf{w}_t^+\right) = M(L)\,\tilde{M}\left(L^{12}\right) e_t, \tag{94}$$

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:

$$\tilde{y}_t = y_t^s + y_t^{irr} + y_t^c. \tag{95}$$

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:

$$\xi_{2\times 12}^{MA}(L) \equiv \frac{1}{24}\left(L^{-6} + L^{-5}\right)\left(\sum_{\ell=0}^{11} L^{\ell}\right) \tag{96}$$

$$\xi_{3\times q}^{MA}(L) \equiv \frac{1}{3q}\left(L^{-12} + 1 + L^{12}\right)\left(\sum_{\ell=-\frac{q-1}{2}}^{\frac{q-1}{2}} L^{12\ell}\right). \tag{97}$$

(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:

$$\xi_q^H(L) = \frac{315}{8} \sum_{\ell=-\frac{q-1}{2}}^{\frac{q-1}{2}} \frac{\left[(m_H - 1)^2 - \ell^2\right]\left[m_H^2 - \ell^2\right]\left[(m_H + 1)^2 - \ell^2\right]\left[3m_H^2 - 16 - 11\ell^2\right]}{m_H\left(m_H^2 - 1\right)\left(4m_H^2 - 1\right)\left(4m_H^2 - 9\right)\left(4m_H^2 - 25\right)} L^{\ell}, \tag{98}$$

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:

$$\xi(L) \equiv 1 - \left[1 - \xi_{2\times 12}^{MA}(L)\right]\xi_{3\times q_2}^{MA}(L)\left[1 - \xi_{q_3}^{H}(L)\left[1 - \left[1 - \xi_{2\times 12}^{MA}(L)\right]^2 \xi_{3\times q_1}^{MA}(L)\right]\right], \tag{99}$$

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 algorithm to an additive version 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.[26]

Taken together, these ad hoc adjustments make the mapping between the original series and the seasonally adjusted series less transparent, but several of the points I raised in the body 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 k^{th} 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

[26] There is no single entity that decides how to adjust official economic statistics. The BLS, which is part of the U.S. Department 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)

the definition of D revised to be:

$$D \equiv \left(\prod_{k=1}^{n} D_k\right)^{\frac{1}{n}}, \quad D_k \equiv \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\Xi_k(\omega)\right) d\omega\right\}, \quad \Xi_k(\omega) \equiv \left|\xi_k\left(\exp\{-i\omega\}\right)\right|^2. \quad (100)$$

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).

C 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:

$$\left(\boldsymbol{\eta}^\dagger, \boldsymbol{\rho}^\dagger, \mathbf{g}^\dagger\right) \equiv \underset{(\boldsymbol{\eta},\boldsymbol{\rho},\mathbf{g})}{\operatorname{argmax}} \log\left(\mathbb{P}\left[\boldsymbol{\eta}, \boldsymbol{\rho}, \mathbf{g} \mid \mathbf{y}^T\right]\right), \quad \mathbf{H} \equiv -\left.\frac{\partial^2 \log\left(\mathbb{P}\left[\boldsymbol{\eta}, \boldsymbol{\rho}, \mathbf{g} \mid \mathbf{y}^T\right]\right)}{\partial\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)' \partial\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)}\right|_{\left(\boldsymbol{\eta}^\dagger, \boldsymbol{\rho}^\dagger, \mathbf{g}^\dagger\right)}. \quad (101)$$

Let $\mathbf{H}_\eta$ denote the square submatrix of $\mathbf{H}$ whose rows and columns correspond to the coordinates of $\boldsymbol{\eta}$ within $\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)'$. (For example, if $\boldsymbol{\eta}$ constitutes the first n_η elements of $\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)'$, 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 $\left(\boldsymbol{\eta}', \boldsymbol{\rho}', \mathbf{g}'\right)'$, 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

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 $\left\{\boldsymbol{\rho}^{(i)}, \boldsymbol{\eta}^{(i)}, \boldsymbol{\lambda}^{(i)}, \boldsymbol{\Phi}^{(i)}, \mathbf{g}^{(i)}\right\}_{i=1}^{N_0}$.

---

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:

$$p_\eta(\boldsymbol{\eta} \mid \boldsymbol{\rho}, \mathbf{g}) \equiv \left|\bar{\mathbf{X}}'\bar{\mathbf{X}}\right|^{\frac{n}{2}} \mathbb{P}[\boldsymbol{\eta}] |\boldsymbol{\Psi}|^T \prod_{k=1}^{n} \frac{\beta_{\lambda,k}^{\alpha_\lambda}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda} \left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}} \tag{102}$$

$$p_{\rho_k}(\boldsymbol{\rho}_k \mid \mathbf{g}, \boldsymbol{\eta}) \equiv \frac{\mathbb{P}[\boldsymbol{\rho}_k] \exp\left\{\frac{1}{2}\boldsymbol{\rho}_k'\left(\sum_{t=1}^T \mathbf{w}_t\right)\right\}}{\hat{\beta}_{\lambda,k}^{\hat{\alpha}_\lambda}\left|\mathbf{X}_{(k)}'\mathbf{X}_{(k)}\right|^{1/2}}, \tag{103}$$

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 50^{th} 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.

D 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 D.1 discusses existing priors for deterministic seasonality, and Appendix D.2 discusses existing priors for stochastic seasonality.

D.1 Existing Approaches to Deterministic Seasonality

D.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.[27] However, uncorrelated coefficients on seasonal dummy variables imply a prior belief that $\mathbf{s}_t$ exhibits negative serial correlation, which may be undesirable, for the reasons mentioned 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 j^{th} 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:

$$\boldsymbol{\mu} = \frac{1}{n_s}\tilde{\mathbf{B}}\mathbf{1}_{n_s\times 1}, \quad \mathbf{s}_t = \tilde{\mathbf{B}}\left(\mathbf{d}_t - \frac{1}{n_s}\mathbf{1}_{n_s\times 1}\right). \tag{104}$$

[27] 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.

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

$$\mathbb{E}_{prior}\left[\tilde{\mathbf{B}}_{j,k}\right] = \bar{\boldsymbol{\mu}}_j, \quad \mathbb{E}_{prior}\left[\left(\tilde{\mathbf{B}}_{j,k} - \bar{\boldsymbol{\mu}}_j\right)\left(\tilde{\mathbf{B}}_{h,\ell} - \bar{\boldsymbol{\mu}}_h\right)\right] = \begin{cases} \sigma^2_{\tilde{B}} & \text{if } (j,k) = (h,\ell) \\ 0 & \text{otherwise} \end{cases}. \tag{105}$$

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. (The assumption that each coefficient has the same prior variance can be relaxed; what matters is that the columns of $\tilde{\mathbf{B}}$ are uncorrelated.) As in the body of the paper, let $\boldsymbol{\Gamma}^u_s$ denote the covariance between $\mathbf{s}_t$ and $\mathbf{s}_{t-u}$ with respect to the prior predictive distribution. One can so that the above prior implies:

$$\boldsymbol{\Gamma}^s_u = \begin{cases} \sigma^2_{\tilde{B}} \frac{n_s - 1}{n_s} \mathbf{I}_n & \text{if } u \overset{\text{mod } n_s}{=} 0 \\ -\frac{\sigma^2_{\tilde{B}}}{n_s} \mathbf{I}_n & \text{otherwise} \end{cases}. \tag{106}$$

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). At a minimum, if an econometrician believes in this kind of negative serial correlation, then that belief should be stated explicitly and defended. As was the case with seasonal waveforms, adopting a diffuse prior over the seasonal coefficients (i.e., making $\sigma^2_{\tilde{B}}$ arbitrarily large) implicitly means that the magnitude of the deterministic seasonal cycle, as summarized by $\boldsymbol{\Gamma}^s_0$, is expected to be enormous.

D.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:

$$y_t = \boldsymbol{c}_t' \boldsymbol{\beta}_t + e_t, \quad \boldsymbol{\beta}_t = \tilde{\mathbf{B}} \mathbf{d}_t, \tag{107}$$

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. Building on Shiller (1973), 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:

$$\left(\tilde{B}_{t+1} - \tilde{B}_t\right) - \left(\tilde{B}_t - \tilde{B}_{t-1}\right) \sim \mathrm{N}\left(0, \zeta^2\right), \tag{108}$$

where $\tilde{B}_t$ is the t^{th} 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 (108) is equivalent to specifying $(s_{t+1} - s_t) - (s_t - s_{t-1}) \sim \mathrm{N}\left(0, \zeta^2\right)$. 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.

D.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(p) process:

$$y_t = \sum_{\ell=1}^{p} \Phi_\ell y_{t-\ell} + e_t, \quad e_t \sim \mathrm{N}\left(0, \frac{1}{\lambda}\right). \tag{109}$$

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:

$$1 = \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell) \tag{110}$$

$$0 = \sum_{\ell=1}^{p} \Phi_\ell \sin(\omega^* \ell), \tag{111}$$

which are the reduced-form, univariate analogues to equations (35) and (36). My prior treats these as stochastic constraints:

$$\sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell) \mid \lambda \sim \mathrm{N}\left(1, \frac{1}{\tau_{\omega^*}^2 \lambda}\right) \tag{112}$$

$$\sum_{\ell=1}^{p} \Phi_\ell \sin(\omega^* \ell) \mid \lambda \sim \mathrm{N}\left(0, \frac{1}{\tau_{\omega^*}^2 \lambda}\right), \tag{113}$$

with $\tau_{\omega^*} > 0$ controlling the prior confidence in a spectral peak at frequency ω^*. As in the body of the paper, equation (113) is only necessary when $\omega^* \notin \{0, \pi\}$. Appendix D.2.1 will discuss the approach to stochastic seasonality in Canova (1992, 1993), and Appendix D.2.2 will discuss the approach to stochastic seasonality in Raynauld and Simonato (1993).

D.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 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 $1 \approx \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell)$, and he does so with a stochastic linear restriction along the lines of equation (112). However, there are many processes that satisfy the condition $1 = \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell)$ exactly, but do not have a spectral peak at, or even near, frequency ω^*.[28] Importantly, Canova's prior does not favor parameters that satisfy $0 \approx \sum_{\ell=1}^{p} \Phi_\ell \sin(\omega^* \ell)$, which is necessary for a spectral peak to occur near frequency ω^*. Absent a stochastic linear restriction along the lines of equation (113), Canova's prior will not typically favor seasonal processes.

[28]Consider the example where $p = 2$ and $\omega^* = \frac{\pi}{2}$, which corresponds to the annual periodicity for quarterly data. For illustrative purposes, suppose that the restriction $1 = \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell)$ holds exactly. If $|\Phi_1| > 2$, then y_t is explosive, but if $|\Phi_1| \leq 2$, then one can show that y_t will have a unit root at some frequency — but not necessarily at ω^*. In particular, if $\Phi_1 = -\frac{\sin(2\omega^\dagger)}{\sin(\omega^\dagger)}$ for an arbitrary value $\omega^\dagger$, then y_t will have a seasonal unit root at frequency $\omega^\dagger$. In this example, adopting extreme prior confidence in Canova's restriction $1 \approx \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^* \ell)$ does not necessarily increase the prior probability of a spectral peak at (or near) the desired frequency; instead, prior beliefs about the location of the spectral peak depend on beliefs about Φ_1.

To see the issue explicitly, note that the autoregressive process can be written $A(L)\,y_t = e_t$, where $A(L) \equiv 1 - \sum_{\ell=1}^{p} \Phi_\ell L^\ell$. Define $a(\omega) \equiv A(\exp\{-i\omega\})$ as the autoregressive transfer function. The spectral density of y_t is:

$$f(\omega) = \frac{1}{|a(\omega)|^2}\frac{1}{2\pi\lambda} = \frac{1}{\Re(a(\omega))^2 + \Im(a(\omega))^2}\frac{1}{2\pi\lambda}, \tag{114}$$

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:

$$a(\omega) = \underbrace{1 - \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega\ell)}_{\Re(a(\omega))} + i\underbrace{\sum_{\ell=1}^{p} \Phi_\ell \sin(\omega\ell)}_{\Im(a(\omega))}. \tag{115}$$

Canova's restriction $1 \approx \sum_{\ell=1}^{p} \Phi_\ell \cos(\omega^*\ell)$ 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 $\Im(a(\omega^*)) \approx 0$.[29] 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 (114) 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.

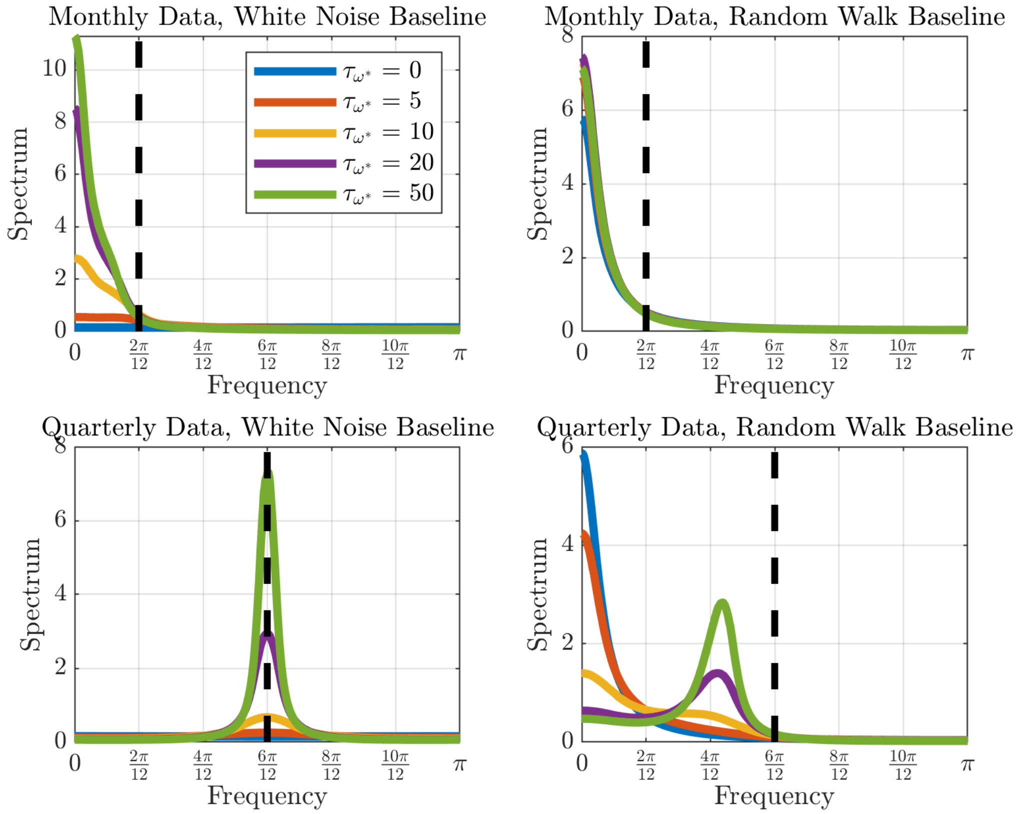
One way to illustrate this fact is to simulate from the prior predictive distribution of the spectrum, and see where the peaks in the spectrum are most likely to occur. That is, I will draw parameters draws from the prior, and then compute the spectra implied by the parameter draws using equation (114). Both Canova's prior and mine take the form:

$$(\Phi_1, \ldots, \Phi_p)' \mid \lambda \sim \mathrm{N}\left(\left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Y}}, \left(\lambda\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1}\right), \quad \bar{\mathbf{X}} = \begin{bmatrix} \bar{\mathbf{X}}_0 \\ \bar{\mathbf{X}}_{\omega^*} \end{bmatrix}, \quad \bar{\mathbf{Y}} = \begin{bmatrix} \bar{\mathbf{Y}}_0 \\ \bar{\mathbf{Y}}_{\omega^*} \end{bmatrix}, \tag{116}$$

where $\bar{\mathbf{X}}_0$ and $\bar{\mathbf{Y}}_0$ are dummy observations that implement a baseline prior about low-frequency behavior, and $\bar{\mathbf{X}}_{\omega^*}$ and $\bar{\mathbf{Y}}_{\omega^*}$ are dummy observations that are meant to capture beliefs about the spectrum at the annual periodicity, denoted $\omega^* \equiv \frac{2\pi}{n_s}$. I will consider two common choices of baseline prior. The first, which I will call the white noise prior, specifies $\Phi_\ell \mid \lambda \sim \mathrm{N}\left(0, \frac{1}{\lambda\ell^2}\right)$, which can be implemented using the dummy observations $\bar{\mathbf{Y}}_0 = \mathbf{0}_{p\times 1}$ and $\bar{\mathbf{X}}_0 = \mathrm{diag}(1, 2, \ldots, p)$. The second, which I will call the random-walk prior, specifies $\Phi_\ell \mid \lambda \sim \mathrm{N}\left(\mathbb{I}[\ell = 1], \frac{1}{\lambda\ell^2}\right)$, which can be

[29] 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 (112) 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 $\lambda = 1$.

implemented using the dummy observations $\bar{\mathbf{Y}}_0 = \left(1, \mathbf{0}_{1\times(p-1)}\right)'$ and $\bar{\mathbf{X}}_0 = \text{diag}\,(1, 2, \ldots, p)$. For $\bar{\mathbf{X}}_{\omega^*}$ and $\bar{\mathbf{Y}}_{\omega^*}$, Canova's prior uses the following dummy observations to implement beliefs about the spectrum at frequency ω^*:

$$\bar{\mathbf{Y}}_{\omega^*} = \tau_{\omega^*}, \quad \bar{\mathbf{X}}_{\omega^*} = \tau_{\omega^*} \begin{bmatrix} \cos(\omega^* 1) & \cos(\omega^* 2) & \cdots & \cos(\omega^* p) \end{bmatrix}. \tag{117}$$

My prior uses the following dummy observations instead:

$$\bar{\mathbf{Y}}_{\omega^*} = \tau_{\omega^*} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \bar{\mathbf{X}}_{\omega^*} = \tau_{\omega^*} \begin{bmatrix} \cos(\omega^* 1) & \cos(\omega^* 2) & \cdots & \cos(\omega^* p) \\ \sin(\omega^* 1) & \sin(\omega^* 2) & \cdots & \sin(\omega^* p) \end{bmatrix}. \tag{118}$$

In both cases, a higher value of τ_{ω^*} reflects greater confidence in the relevant restrictions being satisfied. Figure 6 shows what happens when the two baseline priors are augmented with Canova's dummy observations (117) and my own (118). I have plotted the median of the prior predictive distribution over the spectrum for various values of τ_{ω^*}; in each case, I have assumed that $p = n_s + 1$ 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.

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.

D.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 $\{\Phi_\ell\}_{\ell=1}^{p}$ is Gaussian and independent across ℓ with:

$$\mathbb{E}\left[\Phi_\ell\right] = \mathbb{I}\left[\ell = 1\right] + \mathbb{I}\left[\ell = n_s\right] - \mathbb{I}\left[\ell = n_s + 1\right], \quad \mathbb{V}\left[\Phi_\ell\right] = \left(\frac{\text{Tight} \times \text{SD}^{\lfloor \ell / n_s \rfloor}}{\left(\ell - (n_s - 1) \times \left\lfloor \frac{\ell}{n_s} \right\rfloor\right)^{\text{Tooth}}} \right)^2, \quad (119)$$

where Tight, SD, and Tooth are hyperparameters, and the number of lags p 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 $\mathbb{V}[\Phi_\ell]$ 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:

$$\mathbb{P}\left[(\Phi_1, \Phi_2, \ldots, \Phi_m) = \left(2\cos\left(\frac{2\pi}{n_s}\right), -1, \mathbf{0}_{1\times(p-2)}\right)\right] < \mathbb{P}\left[(\Phi_1, \Phi_2, \ldots, \Phi_p) = \mathbf{0}_{1\times p}\right]. \tag{120}$$

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. In contrast, I declare a set of beliefs about seasonal fluctuations in y_t and reverse engineer a prior over the coefficients to generate reasonable attributes for the spectral density at seasonal frequencies.