BEA Working Paper Series, WP2026–14

The Seasons They Are A-Changin': a Century of Definitions and a Way Forward

Authors Carter Bryson and Gary Cornwall, U.S. Bureau of Economic Analysis

Contacts carter.bryson@bea.gov and gary.cornwall@bea.gov

Date June 2026

Abstract What is seasonality? To date, there is not yet a consensus definition, as over a dozen unique definitions of seasonality can be found in the economics and statistics literature. This lack of agreement complicates the identification of, and adjustment for, seasonal effects in economic time series. In this paper, we propose a new definition of seasonality as well as how to detect and adjust for it, ultimately aiming to build consensus around a more unified approach. We first review the literature and identify common themes with an eye toward formulating a foundational definition—expressed in both plain language and mathematical terms—that can be used by academics, data providers, policymakers, and practitioners alike. We then propose to classify a series as seasonal if, among all its possible repeating cycles, the ones associated with the fundamental periodicity are demonstrably larger than the others. In other words, we define seasonality as a measure of relative peak dominance in the spectral density function, building on Nerlove (1964) and Granger (1978). Next, we develop a new methodology for detecting seasonality and performing seasonal adjustment that follows directly from our definition and leverages properties of the sample periodogram, the empirical counterpart to the spectral density function. We illustrate the performance of our test statistic through simulation studies and find that it is well-sized and has good power. Lastly, we introduce Stochastic Spectral Imputation (SSI), an adjustment procedure that outperforms standard methods currently used by national statistical offices when applied to prominent macroeconomic time series.

Keywords Time series, seasonality, seasonal adjustment, frequency domain

JEL codes C22, C12, C15, E37

Suggested citation:
Bryson, Carter, and Gary Cornwall. *The Seasons They Are A-Changin': a Century of Definitions and a Way Forward.* Working paper no. WP2026–14. Washington, DC: U.S. Bureau of Economic Analysis, June 2026, https://doi.org/10.66137/ZHRT5937.

1. Introduction

Seasonality is among the most recognizable features of economic time series. In many settings, it is visible to the naked eye, yet the field has seemingly never converged on a single operational definition of what it means for a series to be "seasonal." The plurality of existing definitions has legitimate origins. The phenomena economists label as seasonal are themselves heterogeneous, ranging from deterministic calendar effects and stochastic seasonal dynamics to seasonal unit roots and slowly evolving periodic patterns. Each represents a distinct object with its own internally consistent formalization.[1] Accordingly, diagnostic suites and recommended workflows used in production pipelines for seasonal adjustment—most prominently X-13ARIMA-SEATS and TRAMO-SEATS—are designed to navigate this heterogeneity.[2] The resulting patchwork of definitions, tests, and adjustment procedures reflects both the genuine heterogeneity of seasonal phenomena and the organic co-evolution of software and methods developed to capture them. The accumulation of many such tools and concepts is a rational response to a complex empirical reality.[3]

Consider, however, the structure of this accumulated toolkit from the perspective of set theory. Among all time series, there is a subset whose members exhibit seasonality. The existing battery of tests (e.g., seasonal dummies, HEGY, Canova-Hansen, spectral diagnostics, autocorrelation-based procedures) each targets a *strict subset* of this seasonal partition (e.g., deterministic seasonal means, seasonal unit roots, seasonal stability, spectral peaks, stationary seasonality).[4] These subsets partially overlap but are not equivalent; a series may be seasonal under one test and nonseasonal under another, not because either test is deficient, but because they are designed for different population objects. The field's current practice therefore answers "what *type* of seasonality?" before resolving the logically prior question: "is the series in the seasonal subset at all?"

Against this backdrop, we argue that addressing the "superset" question, particularly in an end-to-end fashion as required by national statistical offices, must satisfy three properties simultaneously: (1) an explicit population target, (2) a test statistic with known and tractable distributional properties for that target, and (3) an adjustment operator that is the logical conjugate of the definition and test. In this paper, we make progress toward this goal by developing an internally consistent definition-test-adjustment procedure for seasonality in economic time series.[5] Specifically, we define seasonality as a form of *relative peak dominance* in the spectral density function: a series is seasonal if the height of the spectral density function within a prespecified set of "seasonal" frequency intervals exceeds the height in the remaining intervals. This yields a population object that formalizes seasonality as a comparative statement

[1] For example, see Falkner (1924); Kuznets (1933); Nerlove (1964); Granger (1978); Hylleberg et al. (1990); Canova and Hansen (1995); Ghysels and Osborn (2001).

[2] These diagnostic tools and recommended workflows are further embedded in international guidance intended to promote quality, coherence, and transparency. For example, the ESS Guidelines on Seasonal Adjustment emphasize coherent practice, metadata, and transparency across domains and producers (European Commission and Eurostat, 2015; European Statistical System, Eurostat, 2024). The U.S. Census Bureau's X-13ARIMA-SEATS documentation similarly emphasizes diagnostics for residual seasonality and stability. See Dagum (1980a); Findley et al. (1998); Gómez and Maravall (2001); Maravall (2011); Lytras et al. (2007); Bell et al. (2022); McElroy and Roy (2022) for additional examples.

[3] The link between individual diagnostics and an explicit object of interest representing what is seasonal is often indirect; see Dagum (1980a); Lytras et al. (2007); Maravall (2011); Webel and Ollech (2018); Chen and Cornwall (2021) for discussion.

[4] For example, see Dagum (1980a); Findley et al. (1998); Gómez and Maravall (2001); Maravall (2011); Lytras et al. (2007); McElroy and Roy (2022).

[5] We note that while individual elements of this triad can be found in prior work, our contribution is to *simultaneously* satisfy all three criteria within a single closed-loop chain. For example, both Nerlove (1964) and Granger (1978) proposed spectral population objects; both Hylleberg et al. (1990) and Canova and Hansen (1995) developed tests with well-characterized null distributions; both Maravall (2011) and the X-13 pipeline implement adjustment operators that are optimal under their respective model assumptions.

about two *ex ante* frequency subsets, following Nerlove (1964), who first identified spectral peaks at seasonal frequencies as the defining characteristic of seasonality, and Granger (1978), who formalized the population-sample distinction for spectral seasonality.[6]

From this definition and population object, we derive a corresponding finite-sample test statistic based on the sample periodogram. We partition the frequency domain into nonoverlapping intervals and designate each as either seasonal or nonseasonal based on whether it contains prespecified frequencies. Mirroring the population characterization, our test statistic computes the distance between the maximum height of the periodogram in the seasonal intervals and the maximum height of the periodogram in the nonseasonal intervals. We then derive its limiting distribution in closed form using results from extreme value theory. It is well known that the sample periodogram is an inconsistent estimator of the spectral density function. In our approach, we exploit this inconsistency directly: under white noise, periodogram ordinates at Fourier frequencies are approximately independently exponentially distributed (Priestley, 1981). This allows us to show that our test statistic converges to a logistic distribution with known location and scale parameters. The test therefore requires no simulation-based critical values, bandwidth selection, or kernel specification.

Finally, we propose an adjustment method, *Stochastic Spectral Imputation* (SSI), that is definitionally consistent with our target object and operates on the discrete Fourier transform of the observed series. Rather than mechanically suppressing frequencies in the seasonal intervals as in notch filtering, SSI replaces their magnitudes with draws from an empirically estimated "donor" distribution constructed from the nonseasonal intervals. Phase can be optionally kept or randomized, and we enforce conjugate symmetry accordingly. The resulting adjusted series, by construction, removes exactly the excess peak power that defines seasonality in our framework while preserving the nonseasonal spectral floor.

Working in the frequency domain is not merely a technical convenience; it is motivated by four substantive considerations. First, the frequency domain characterizes periodic behavior by repetition rate rather than calendar position, making the resulting definition invariant to the choice of calendar (Gregorian, fiscal, lunar, or otherwise). Second, a frequency-domain formulation does not presuppose the *type* of seasonal process generating the data; it targets a superset property (excess spectral mass at seasonal frequencies) that is shared by deterministic, stochastic, and unit-root seasonal processes alike. Third, the spectral density function provides a natural variance decomposition by frequency, so that seasonality becomes a statement about the *disproportionate* contribution of seasonal frequencies to total variance rather than a statement about mean structure or autocovariance at specific lags. Fourth, effects that complicate time-domain definitions (e.g., trading-day variation, outliers, and moving holidays) are treated as nuisance effects in the frequency domain. Trading-day effects and outliers affect spectral estimation but not the definition itself; moving holidays, which migrate across the calendar, do not concentrate power at fixed seasonal frequencies and are therefore correctly treated as calendar effects rather than periodic structure. In practice, we address these through standard pre-filtering prior to spectral analysis.

In addition to its theoretical appeal, the approach we propose in this paper is motivated by practical considerations faced by national statistical offices and major data providers. These organizations are necessarily resource constrained: they must identify and seasonally adjust large, evolving inventories of series, maintain production schedules, support diverse user communities, and document methods in ways that are both auditable and communicable.[7] In such environments, properties such as statistical

[6] The framework applies to weakly stationary processes. Extension to the nonstationary case proceeds through standard pre-filtering (e.g., differencing); the contribution here is the closed-loop structure linking definition, test, and adjustment, not the pre-filtering step.

[7] For example, the U.S. Bureau of Economic Analysis (BEA) is primarily known for producing a measure of gross domestic product and its components that make up the National Income and Product Accounts (NIPAs). However, BEA also provides

robustness, analytic transparency, minimal parametric commitment, and scalability are not luxuries; they are design requirements (European Statistical System, Eurostat, 2024; United Nations et al., 2025). Our methodology, which includes a definition, test statistic, and adjustment procedure, fulfills these requirements, making it an ideal framework around which data providers and national statistical offices can coalesce.

We view the definition-test-adjustment chain developed in this paper as a proof of concept that epistemic closure in seasonal adjustment is achievable. We do not claim that our frequency-domain definition is the only valid formalization of seasonality; the type-specific tests and procedures developed over the past century remain valuable for characterizing *which kind* of seasonality a series exhibits. Rather, our claim is narrower and, we believe, stronger: once a definition of seasonality is adopted, the downstream test and adjustment should be its logical consequences, and the three elements should form a closed system in which each step is derivable from the preceding one. We demonstrate that this is feasible in the frequency domain. We challenge others to construct equivalent closed-loop chains in their preferred domains (time-domain, state-space, or otherwise), thereby establishing a common architecture and platform for communication even if the specific definitions differ.

The remainder of the paper proceeds as follows. Section 2 reviews over a century of academic literature on seasonality, highlighting prominent definitions and how they map (often imperfectly) into diagnostic and adjustment practice. Section 3 formalizes our population object and derives the sample statistic along with its limiting null distribution. Section 4 presents our adjustment procedure and establishes its definitional consistency and asymptotic justification. Section 5 studies finite-sample size, power, and adjustment properties under a range of seasonal and nonseasonal data-generating processes. Section 6 illustrates the performance of our testing procedure and adjustment operator on popular macroeconomic time series in comparison with published seasonally adjusted series. Section 7 concludes by discussing implications for official statistics production, documentation, and future extensions.

2. A Century of Discourse on Seasonality

Seasonality has been at the center of econometrics and the analysis of economic time series going back to at least Falkner (1924), who argued that observed movements can be decomposed into a secular trend, cyclical fluctuations, irregular deviations, and seasonal variation. In the intervening decades, the field has produced multiple internally coherent but mutually incommensurable notions of seasonality, a state of affairs captured memorably by Zellner et al. (1978):

> *"Some of us looking for answers regarding seasonal analysis may feel as if we're in a dark room looking for a black cat. That's bad, but it's not as bad as it could be. Pity the philosophers who are in a dark room looking for a black cat that isn't there." — Zellner (1978, p. 451)*

Since this volume was published nearly five decades ago, there have been additional efforts to characterize different aspects of seasonality (Hylleberg et al., 1990; Canova and Hansen, 1995; McElroy and Holan, 2009; McElroy and Roy, 2022) as well as substantial work on identifying residual seasonality in macroeconomic indicators (Wright, 2013; Rudebusch et al., 2015; Lunsford, 2017; Consolvo and Lunsford, 2019; Lunsford, 2025; Hamlette and Lunsford, 2026). In this section, our goal is to taxonomize these along

supplemental measures of the economy in satellite accounts (e.g., space economy) and, in total, produces approximately four million time series measuring economic activity. Therefore, a statistically robust, analytically transparent, scalable, and epistemologically consistent framework would not only benefit the agency and its stakeholders, but also represent an efficient use of taxpayer dollars in pursuit of their stated mission from Congress.

two broad axes, the *population object* each definition targets and the *domain* (time versus frequency) in which it is formalized. The goal is not to adjudicate which historical approach is correct; rather, it is to show that these approaches target genuinely different population objects and that this plurality propagates into tests and adjustment operators that have never been assembled into a single, closed system linking definition, test, and adjustment.

2.1. Definitions

Broadly speaking there are four categories of seasonality definitions: (1) those based on regularity and repetition, (2) seasonal stability and unit roots, (3) unobserved component or signal extraction, and (4) spectral or frequency-domain definitions. In this section, we elucidate where these categories disagree and where they provide partially overlapping or even exclusive concepts. Before turning to each in detail, three points of agreement are worth noting. First, each relies upon the concept of periodicity at known, often sub-annual frequencies. None of the definitions we examined treated the seasonal frequencies as unknown parameters to be estimated. Second, each category and definition therein views seasonality as a phenomenon separable from the long-run trend cycle and short-run irregular terms. Finally, every one of the definitions within each category either implicitly or explicitly states that seasonality, as a natural-language concept, is broader than the specific mathematical object it formalizes (see, e.g., Bell and Hillmer, 1984; European Statistical System, Eurostat, 2024, p. 111).

2.1.1. Regularity and Repetition (Calendar-Anchored Definitions)

The oldest class of definitions grounds seasonality in the observable regularity of intra-annual patterns. Falkner (1924) offered the earliest formal articulation, defining seasonal variation as the "persistent tendency for certain months of each year to be regularly higher than certain other months of the year." This phrasing is both directional (certain months are *higher*) and calendar-specific (the unit of observation is the month within a Gregorian year). Subsequent formulations relaxed the directional commitment while retaining the calendar anchor. Kallek (1978), writing from the U.S. Census Bureau, described seasonality as "regularly periodic fluctuations which recur every year with about the same timing and with the same intensity" (p. 5), a definition that imposes no claim about which months are high or low but that still centers the concept on annual recurrence at fixed calendar positions. Granger (1978) cataloged four classes of causes underlying these recurring patterns (calendar effects, timing decisions, weather, and expectations) and argued that imprecise definitions of seasonality breed improper adjustment criteria, but even his causal taxonomy takes the annual cycle as the organizing frame. Across all three formulations, the population object is a conditional mean structure (or, more loosely, a regularity in levels) within a fixed calendar partition.

The test statistics native to this category reflect this object directly. Friedman (1937) and Kruskal and Wallis (1952) test whether seasonal means or medians are equal across calendar positions under nonparametric rank assumptions, while the GLS F-test of Lytras et al. (2007) tests joint significance of seasonal dummies in a regression framework (see Table 1 below). In each case, the null hypothesis is the absence of level differences across seasons; the alternative is that at least one season differs. Because the tests partition time into discrete calendar bins and compare levels across those bins, their inferential reach is bounded by the definition that motivates them: they can detect whether seasonal levels differ, but they are silent on why or how.

Table 1. Selected tests and diagnostics for seasonality (extended taxonomy)

Procedure	Domain	Population Target	H_0	H_1	Null Dist.
Calendar-anchored					
Friedman (1937)	Time	Seasonal means	Equal ranks across seasons	Unequal ranks	χ^2
Kruskal and Wallis (1952)	Time	Seasonal medians	Equal medians across seasons	Unequal medians	χ^2
Lytras et al. (2007)	Time	Seasonal means (GLS)	No stable seasonality	Stable seasonality	F
Seasonal unit roots and stability					
Hylleberg et al. (1990)	Time	AR polynomial roots	Unit root at seasonal freq.	Stationary at seasonal freq.	DF-type (non-standard)
Canova and Hansen (1995)	Time	Seasonal intercept stability	Stable (deterministic)	Drifting (unit root)	CvM func of Brownian bridge
Busetti and Harvey (2003)	Mixed	Spectrum at seasonal freq.	No seasonal variation	Deterministic or stochastic	CvM
Spectral / frequency-domain					
Fisher (1929)	Freq.	Periodogram max / total	No hidden periodicity	Hidden periodicity	Exact (inclusion-exclusion)
Priestley (1981)	Freq.	Spectral density at seasonal freq.	Flat spectrum	Peaked spectrum	χ^2
McElroy and Holan (2009)	Freq.	Spectral slope/convexity	No peak (flat or rising)	Peak (zero slope, neg. convexity)	Normal (asymptotic)
Maravall (2012)	Time	Seasonal autocorrelation (QS)	No seasonal autocorrelation	Seasonal autocorrelation	χ^2 (asymptotic)
McElroy and Roy (2021)	Freq.	Log spectral ratio	Mild persistence	Strong persistence	fixed-b func of Brownian motion
McElroy (2021)	Freq.	AR root proximity	Roots near seasonal freq.	Roots far from seasonal freq.	Quadratic form in normals
Webel and Ollech (2025)	Mixed	Classification tree on p-values	Nonseasonal ARIMA	Seasonal ARIMA	Empirical (no analytical null)
This paper					
Bryson–Cornwall	**Freq.**	**Relative peak dominance** $\Delta(f_Z)$	**No seasonal excess**	**Seasonal excess**	**Logistic (EVT)**

Note: Adapted and extended from McElroy and Roy (2022). Nulls and alternatives are schematic. Each procedure targets a different population object (column 3), so "seasonal" conclusions are not automatically comparable across rows. Section 2.2.3 (unobserved component / signal extraction) does not appear because that tradition does not propose a standalone test for the *presence* of seasonality; it presupposes the decomposition and estimates within it. The QS statistic of Maravall (2012) is listed under the spectral heading because it targets seasonal autocorrelation, though it is computed in the time domain.

This category has proven durable in governance documents. A century after Falkner (1924), the European Statistical System (ESS) defines seasonality as "movements which recur with similar intensity in the same season each year and which, based on the past movements of the time series in question and under normal circumstances, can be expected to recur" (European Statistical System, Eurostat, 2024, §2, p. 8). The ESS definition adds a behavioral and expectational component ("under normal circumstances") that Falkner and Kallek lack, conditioning the concept on past behavior and on an undefined state of normalcy. This progression from Falkner's directional empiricism through Kallek's symmetric periodicity to the ESS behavioral formulation illustrates a pattern that recurs across the broader taxonomy: each successive version accommodates a wider class of phenomena while making the underlying primitive less precise. What counts as a "season," what "similar intensity" requires, and what constitutes "normal circumstances" are all left to the analyst. The associated tests, however, do not accommodate this ambiguity. The Friedman, Kruskal-Wallis, and GLS F statistics all test for fixed level differences across discrete calendar bins; a series whose seasonal pattern evolves smoothly over time (satisfying the ESS definition in spirit) may fail these tests in short windows and pass them in long ones, depending on the span and the rate of evolution.

The weaknesses of this definitional category follow from its calendar dependence. Because the primitive is a partition of calendar time (months, quarters, or seasons within a Gregorian year), the definitions cannot accommodate periodicities that do not align with this partition without ad hoc extension. Migrating holidays tied to non-Gregorian calendars (Ramadan in the Hijri calendar, Easter in the ecclesiastical calendar), election cycles, fiscal-year reporting schedules, and day-of-week effects in daily data all generate periodic or quasi-periodic patterns that these definitions either exclude or absorb only by stretching the meaning of "seasonal" beyond its original scope.[8]

More fundamentally, the definitions are (generally) silent on the mechanism that generates the regularity. A series of fixed seasonal dummies and a series driven by a stochastic seasonal cycle can both satisfy Kallek's definition, yet the two processes have different spectral signatures, different persistence properties, and different implications for adjustment. The calendar-anchored tradition tells the analyst *where to look* (within-year patterns at fixed calendar positions) but not *what to look for* (the mathematical object that distinguishes seasonal from nonseasonal). The tests built on this tradition inherit the same limitation: they can detect whether seasonal levels differ, but they cannot distinguish a fixed seasonal pattern from an evolving one, nor can they characterize the nature or persistence of the seasonal structure they detect.

2.1.2. Seasonal Unit Roots and Stability (ARIMA Polynomial Definitions)

The seasonal ARIMA framework provides the mechanism that the calendar-anchored definitions lack. Rather than partitioning time into discrete bins and comparing levels across them, this tradition defines seasonality through the algebraic structure of the autoregressive polynomial. The seasonal differencing operator $(1 - B^s)$, where B is the backshift operator and s the seasonal period, factors into s linear and quadratic terms whose roots lie on the unit circle at the zero and seasonal frequencies $\omega_k = 2\pi k/s$, $k = 0, 1, \ldots, \lfloor s/2 \rfloor$. A series is "seasonally integrated" if the autoregressive polynomial shares one or more of these roots, so that shocks at the corresponding frequency persist indefinitely; it is "deterministically seasonal" if the seasonal pattern is stable (coefficients on trigonometric or dummy regressors are fixed) and no such roots are present. The population object thus shifts from calendar-bin means to the location and multiplicity of roots at seasonal frequencies, and the question "is the series seasonal?" becomes a question about the polynomial structure of the data-generating process. This reframing carries a cost: it

[8] Granger (1978) was aware of this limitation; his four-class causal taxonomy includes calendar effects and timing decisions as separate categories, implicitly acknowledging that not all intra-annual regularity has the same source or structure. The ESS Guidelines handle the problem operationally by treating calendar and moving-holiday effects as pre-adjustment regressors rather than as part of the seasonal component itself (European Statistical System, Eurostat, 2024, §5.2).

presupposes that the analyst has already committed to a seasonal ARIMA representation, so that the definitions, tests, and adjustment operators that follow are all internal to that modeling framework.

Hylleberg et al. (1990)—henceforth HEGY—operationalized this framework by constructing a Lagrange-interpolation factorization of $(1 - B^s)$ around its s roots, yielding an auxiliary regression in which each seasonal frequency maps to a separate coefficient π_k on a transformed regressor $y_{k,t}$.[9] The HEGY null at each frequency is a unit root (nonstationarity); rejection implies stationarity at that frequency. Canova and Hansen (1995) reversed this logic, taking stationary (deterministic) seasonal parameters as the null and testing for instability via Lagrange multiplier statistics constructed from OLS residuals and a kernel-based long-run covariance estimator (see Table 1). In the same spirit that Canova and Hansen (1995) generalized the Kwiatkowski et al. (1992) zero-frequency stationarity framework to seasonal frequencies, the Canova-Hansen test inherits the Von Mises limiting distribution and is numerically identical whether formulated as a test for seasonal-intercept instability or as a test for unit roots at all seasonal frequencies. Subsequent work refined the HEGY side of the literature along several dimensions. Ghysels et al. (1994) introduced joint F-statistics (F_{234}, F_{1234}) and documented severe size distortion when seasonal moving-average roots approach cancellation with the unit-root factors, a pathology they characterized as "a very serious challenge to the design of tests for unit roots at seasonal frequencies" (p. 424). del Barrio Castro et al. (2012) resolved a long-standing conjecture by proving that augmented HEGY F-statistics retain pivotal limiting null distributions under general linear-process shocks driven by martingale difference innovations, placing the augmented HEGY framework on the same asymptotic footing that Chang and Park (2002) established for the augmented Dickey-Fuller test. On the efficiency frontier, Rodrigues and Taylor (2007) extended the Elliott et al. (1996) power-envelope framework to seasonal frequencies, producing GLS-detrended HEGY variants with substantially higher power than their OLS counterparts. Smith and Taylor (1999) developed likelihood-ratio variants of HEGY that fully incorporate the restrictions on deterministic parameters implied by the seasonal unit root null, and Taylor (2005) constructed bandwidth-free variance-ratio counterparts that achieve pivotal null distributions under periodic heteroskedasticity, a condition under which the regression-based HEGY statistics lose asymptotic pivotality.

Busetti and Harvey (2003) occupy a boundary position between the unit-root tradition and the spectral tradition. Their nonparametric spectral correction of the Canova-Hansen test replaces the full long-run variance matrix with estimates of the spectrum at each seasonal frequency, yielding test statistics whose Cramér-von Mises null distributions are invariant to deterministic trends (including structural breaks) and to nonseasonal unit roots. This invariance property extends the original Canova-Hansen result, which required a constant-level specification. The paper also develops a general test against *any* form of permanent seasonality (deterministic or stochastic), bridging the two camps by testing the prior question of whether a seasonal component exists at all. The methodological positioning is explicit: "Parametric tests are attractive within the context of a model-building exercise, but if the sole focus is on testing for stochastic seasonality, then there is no overwhelming case for preferring them to nonparametric tests" (Busetti and Harvey, 2003, p. 432). A companion paper by Busetti and Taylor (2003) demonstrated that both the HEGY and Canova-Hansen frameworks are compromised by unattended structural breaks: HEGY tests are unaffected at the frequency under test but the Canova-Hansen statistic converges to zero when breaks occur at frequencies not explicitly modeled, rendering the stationarity test near-powerless precisely when the data depart from the maintained hypothesis.

The HEGY/Canova-Hansen split encodes a fundamental asymmetry. HEGY takes seasonal unit roots as

[9] For quarterly data ($s = 4$), the four roots are $\{1, -1, i, -i\}$ and the auxiliary regression produces individual t-statistics t_1, t_2 at the zero and Nyquist frequencies, a joint F-statistic F_{34} at the annual harmonic pair, and joint F-statistics F_{234} and F_{1234} across seasonal and all frequencies. See Table 1 for the HEGY null–alternative structure.

its null; Canova-Hansen takes seasonal stability as its null. Each camp treats the other's maintained hypothesis as its alternative. This duality means that the two tests cannot jointly answer the same question: failure to reject in both (a plausible outcome in short samples) leaves the analyst without a decision, while rejection in both signals model misspecification rather than a characterization of the seasonal structure. More consequentially, neither framework addresses the logically prior question of whether the series is seasonal at all. Both presuppose that a seasonal structure exists and test *within* it: HEGY asks whether the seasonal pattern is driven by unit roots; Canova-Hansen asks whether the seasonal parameters are stable. A series with no seasonal component whatsoever would, under the HEGY null, appear to have seasonal unit roots (since the absence of seasonal structure is observationally equivalent to having all seasonal roots on the unit circle in a degenerate sense), and under the Canova-Hansen null, would appear to have stable seasonal parameters (since zero-valued parameters are trivially constant). The implied adjustments are similarly type-specific and not coupled to the test by construction: rejection of the HEGY null at a given frequency suggests that seasonal differencing at that frequency is unnecessary, while rejection of the Canova-Hansen null suggests that fixed seasonal factors are inadequate, but neither test prescribes a specific adjustment operator as a logical consequence of its inference.[10] The deterministic/stochastic dichotomy that organizes this literature is therefore better understood as a maintained assumption than as a discovered property, a theme that recurs in Section 2.

2.1.3. Unobserved Component / Signal Extraction Definitions

The signal-extraction tradition shifts the population object once more. Whereas the unit-root approach asks whether the autoregressive polynomial of Z_t contains roots at seasonal frequencies, the signal-extraction tradition asks a different question: what is the seasonal component, and how can it be recovered from the observed series? The object of interest is no longer a property of the data-generating process for Z_t itself but a latent component S_t in a decomposition $Z_t = T_t + S_t + I_t$, where T_t is trend, S_t is seasonal, and I_t is irregular. Two sub-traditions have developed around this framework. The model-based lineage, originating with Box et al. (1978) and formalized by Hillmer and Tiao (1982), derives the component models from a fitted ARIMA process for Z_t via a canonical decomposition that maximizes irregular variance; Bell and Hillmer (1984) articulated the identifying assumptions explicitly, and Gómez and Maravall (2001) provided the definitive tutorial for the TRAMO-SEATS implementation, proving along the way that all canonical seasonally adjusted estimators are non-invertible in the standard nonstationary case. The filter-based lineage proceeds differently: Cleveland and Tiao (1976) showed that the X-11 moving-average filters are approximately optimal under the airline model, Dagum (1980b) extended the procedure to X-11-ARIMA by appending ARIMA forecasts and backcasts before applying the filter sequence, and Burman (1980) developed the minimum signal extraction (MSX) method via partial fractions of the model spectrum, coining the term "silent spectrum" for variance lost to the cross-spectrum between components. Both sub-traditions share a structural commitment: the seasonal component exists as a well-defined object, and the task is identification and estimation, not detection.

The core technical challenge is that the decomposition $Z_t = T_t + S_t + I_t$ is not unique without identifying restrictions. Given a fitted ARIMA model for Z_t, infinitely many triples of component models can reproduce the same aggregate autocovariance structure. Box et al. (1978) and Hillmer and Tiao (1982) resolve this indeterminacy by adopting the canonical decomposition: maximize the variance of the irregular component I_t (equivalently, minimize the innovation variances of the trend and seasonal components), subject to coprimality of the component autoregressive polynomials and a fixed seasonal summation

[10] This disconnect between test and adjustment is not unique to the seasonal setting. In the nonseasonal unit-root literature, the Dickey-Fuller test does not prescribe a specific detrending method, and the KPSS test does not prescribe a specific differencing operator. The seasonal case inherits this gap and compounds it with the additional dimension of frequency-specific structure.

operator $U(B) = 1 + B + \cdots + B^{s-1}$. Granger (1978) and Zellner et al. (1978) questioned whether this minimum-extraction principle is arbitrary; Zellner noted that it "may conflict with smoothness objectives for the trend-cycle component" and asked whether the principle "should be rationalized in terms of subject matter considerations" (Zellner et al., 1978, p. 454). Bell and Hillmer (1984) acknowledged the concern directly: "Unfortunately, there is not enough information in the data to define the components, so these types of arbitrary choices must be made" (p. 111). The filter-based tradition faces the same problem in a different guise. Cleveland and Tiao (1976) demonstrated that the X-11 filters are approximately optimal under a specific ARIMA model with airline-type structure, which means the implicit identifying restriction in X-11 is model-dependent even though X-11 is presented as a nonparametric procedure. Eurostat (2018) summarized the state of affairs plainly: "The evaluation of the 'seasonal component' provided by an adjustment method is hampered by the fact that the true seasonal component remains a theoretical and imprecise concept, never liable to direct observation" (Ch. 1, p. 26).

This identification problem is where the signal-extraction tradition confronts the spectral tradition most directly. Bell and Hillmer (1984) criticized frequency-domain definitions of seasonality on the grounds that "these definitions only tell us when a series has a seasonal component, not what the seasonal component is" (p. 110, referring to Nerlove (1964) and Granger (1978)). The criticism is well taken on its own terms: a spectral peak at a seasonal frequency does not by itself determine a unique time-domain component. Yet Burman (1980), writing from within the signal-extraction tradition, conceded the complementary point: "It seems that, in the time domain, the decomposition is incapable of precise definition. However, in the frequency domain, the trend and seasonal components of a series can be more clearly defined" (p. 321). This internal tension is revealing. Even practitioners committed to recovering a time-domain component acknowledge that the frequency domain provides the sharper characterization of what "seasonal" means. The signal-extraction tradition proceeds in the time domain nonetheless, because seasonal adjustment requires a recoverable additive (or multiplicative) component, not merely a detected spectral property. The trade-off is real: clarity of definition is exchanged for recoverability of an object suitable for subtraction.

The principal weakness of this category follows from the identification problem: the population object (S_t) is defined by the model, not by the data. Different ARIMA specifications for the same observed series yield different component models, different Wiener-Kolmogorov filters, and therefore different seasonal components.[11] Gómez and Maravall (2001) proved that in the standard nonstationary case (unit roots at zero and seasonal frequencies), all canonical seasonally adjusted estimators are non-invertible: their moving-average representations contain roots on the unit circle. This result has a direct consequence for post-adjustment diagnostics. Tests based on autoregressive representations (including ROOT-type and ARMA-root diagnostics applied to the adjusted series) operate on an object whose MA structure has been contaminated by the non-invertible adjustment filter. The diagnostic and the adjustment operator therefore act on different objects. Table 1 does not contain a test statistic native to the signal-extraction category because this tradition does not propose a standalone test for the presence of seasonality; it presupposes the decomposition and estimates within it. The tests applied after adjustment (the QS statistic, visual spectrum analysis, ROOT diagnostics) belong to other categories in the taxonomy, and their inferential targets do not coincide with the latent-component object that the adjustment was designed to remove.

[11] This is not merely a finite-sample estimation issue. Two correctly specified but distinct aggregate models (for example, an airline model versus a model with an additional AR factor) produce structurally different canonical decompositions, so that the "true" seasonal component is model-contingent at the population level.

2.1.4. Spectral / Frequency-Domain Definitions

The spectral tradition defines seasonality as a property of the observed series itself, expressed through the spectral density at seasonal frequencies, rather than as a component extracted from a decomposition or a set of polynomial roots. The root is Nerlove (1964), which provided the first explicit frequency-domain definition: "In the more general case, then, we may define seasonality as that characteristic of a time series that gives rise to spectral peaks at seasonal frequencies" (§2.1, after eq. 2.6). This definition predates the model-based canonical decomposition by two decades and locates the population object in a different space entirely. The target is a functional of the spectral density $f(\lambda)$ evaluated at $\lambda_i = i/12$, $i = 1, \ldots, 6$ for monthly data. The domain is frequency; the object is the observed spectrum, not a latent component. Nerlove's paper also provided the first frequency-domain evaluation of a seasonal adjustment pipeline: cross-spectral estimation of the gain and phase of the BLS procedure revealed over-removal of nonseasonal power and severe phase distortion at low frequencies, a finding Nerlove identified as "perhaps the most important" result in the paper. That diagnosis established a template (compare spectra before and after adjustment; look for distortion in frequency bands that are not seasonal) that persists in production diagnostics today.

Granger (1978) formalized Nerlove's qualitative statement into three explicit spectral definitions. Definition 1 (Property S): the generating process X_t has property S if its population spectrum $f(\omega)$ has peaks within the seasonal frequency bands $\omega_S(\delta)$ for some small $\delta > 0$. Definition 2 (Apparent Property S): the observed series x_t apparently has property S if the estimated spectrum $\hat{f}(\omega)$ has such peaks. Definition 3 (Strong Seasonality): the integrated spectral mass over the seasonal bands accounts for nearly all power, that is, $\int_{\omega_S(\delta)} f(\omega)\, d\omega \approx A(\delta)$ with $A(\delta)$ near unity. Granger's rationale was direct: "These limitations, together with the fact that the most obvious feature of a seasonal component is its repetitiveness over a 12-month period, strongly suggest that a definition can be most naturally stated in the frequency domain, since spectral methods investigate particular frequencies and are essentially model-free" (p. 35). The formulation is notable for two reasons. First, it explicitly separates the population object $f(\omega)$ from the sample diagnostic $\hat{f}(\omega)$, a distinction that the unobserved-components tradition (Section 2.1.3) leaves implicit. Second, the bandwidth δ that defines the seasonal bands is "unfortunately arbitrary and has to be left to the individual analyst" (p. 35). Granger's bin structure (the set $\omega_S(\delta) = \bigcup_{k=1}^{5}(\omega_k - \delta, \omega_k + \delta) \cup (\omega_6 - \delta, \pi)$) is the conceptual ancestor of the partition $\mathcal{P}$ that we formalize in Section 3; what Granger left unresolved was how to make δ explicit and how to test for spectral mass within the resulting bins against a known null distribution.

The classical periodogram tests for hidden periodicity form a parallel lineage that addresses statistical inference at a single frequency rather than across a seasonal partition. Fisher (1929) derived the exact null distribution for the largest periodogram ordinate expressed as a fraction of the total: the statistic $g = x_{\max} / \sum x_i$ follows the inclusion-exclusion distribution $P = \sum_{j=0}^{k} (-1)^j \binom{n}{j} (1 - jg)^{n-1}$ under Gaussian white noise (Table 1). Hannan (1961) extended Fisher's test to a general stationary null by replacing the white-noise denominator with a kernel-smoothed spectral estimate, corrected for the inflation that arises when the estimator is contaminated by the very peak whose significance is being tested.[12] Davis and Mikosch (1999) completed the probability theory by proving that the Gumbel limit for the periodogram maximum holds under non-Gaussian, finite-variance conditions, with periodogram ordinates exhibiting "almost iid" behavior and the associated point process converging to a Poisson

[12] Hannan's correction subtracts from the kernel estimate the contribution of the periodogram ordinate at the candidate frequency, a device that anticipates the leave-one-out logic in later partition-based approaches. The resulting statistic K_n^* recovers Fisher's null distribution asymptotically under a bandwidth condition that Hannan himself described as "not very meaningful in practice" (p. 401), offered primarily as validation of the intuitive requirement that the kernel smoother not be too aggressive.

limit. These results are foundational for extreme-value-based inference on periodogram ordinates. The limitation shared across this lineage is that each test targets a single-frequency object (the largest ordinate) rather than a partition-based comparison, and each operates under a white-noise or iid null rather than under a general stationary null with a population-level seasonal object.

The production diagnostic that brought spectral peak detection into official statistics was not a formal test but a visual heuristic. The X-12-ARIMA spectrum diagnostic, developed by Soukup and Findley (1999) and described in Findley et al. (1998), estimates the spectrum of the (first-differenced) seasonally adjusted series using an AR(30) model, evaluated in decibels at the seasonal frequencies $k/12$ for $k = 1, \ldots, 6$. A peak is flagged as "visually significant" when the gap between the spectral amplitude at a seasonal frequency and the amplitudes at its two plotted neighbors exceeds a margin scaled to the global range of the log spectrum. This is a visual evaluation rather than a statistical test: it operates on frequency neighborhoods (the bins around each seasonal harmonic) but calibrates significance by eye against the overall spectral amplitude, not by reference to a null distribution. As McElroy and Holan (2009) observed, "One limitation of this approach is that it has really no statistical component: the significance is not statistical, i.e., it is not associated with a hypothesis test, and the thresholds to determine 'visual significance' are determined in an ad hoc fashion" (p. 68). The Findley et al. (1998) rejoinder (pp. 171–172) acknowledged that the false-positive rate was too high in Monte Carlo experiments on airline-model-generated series with no trading-day effects. Despite these limitations, the visual-significance diagnostic remains embedded in the X-13ARIMA-SEATS production pipeline and serves as the most widely deployed spectral seasonality check in official statistics.

The progression from heuristic to formal test was carried out in a sequence of papers by McElroy and co-authors. McElroy and Holan (2009) proposed the first formal nonparametric spectral-peak test with asymptotic theory: kernel-weighted integrals of the periodogram measure aggregate slope and convexity of the spectral density at each seasonal frequency band, and a peak is declared when slope is not significantly different from zero and convexity is significantly negative, with a sequential two-stage procedure and Hochberg family-wise error rate correction across five seasonal frequencies (Table 1). This approach does not use bins in the Granger sense; it evaluates the local shape of $f(\lambda)$ (zero slope plus negative curvature) to determine whether a peak exists. McElroy and Holan (2016) simplified the procedure to a single t-statistic T_j per seasonal frequency using cosine-windowed periodogram integration, with the five statistics jointly asymptotically independent standard normal, an implementation requiring "few lines of code." McElroy and Roy (2021) took a different route, placing the Soukup-Findley visual-significance criterion on a rigorous footing via fixed-bandwidth-fraction tapered asymptotics for a log spectral ratio at seasonal frequencies (Table 1). McElroy (2021) bridged the spectral and unit-root traditions by constructing a diagnostic based on AR polynomial root proximity to seasonal frequencies (Table 1), testing for r-persistent seasonality through the functional $g(r) = |\pi(r^{-1}e^{i\omega})|^2$ evaluated at the fitted ARMA model's AR(∞) polynomial. The comprehensive review by McElroy and Roy (2022) synthesized this body of work, evaluated the four leading residual-seasonality diagnostics (Qs, ROOT, SC, VSnew), and formally disendorsed Qs on the ground that it "fails to account for the second portion of seasonality's definition, namely, that seasonal persistence is not explained by inter-seasonal lag effects" (p. 270).

Three distinct approaches to operationalizing "peak at a seasonal frequency" thus coexist within the spectral tradition, differing in their mechanics even as they share a common conceptual target. Granger's bandwidth δ defines a peak via integrated spectral mass within a frequency band: the analyst chooses δ, and the peak is whatever falls inside. This is a bin-based approach. The Soukup-Findley visual-significance criterion defines a peak via the gap between a frequency's spectral amplitude and its neighbors' amplitudes, scaled to the global spectral range: a neighbor-comparison approach with no formal null. The McElroy-Holan slope-and-convexity test defines a peak via the local shape of $f(\lambda)$ around the candidate

frequency (zero slope plus negative convexity): a calculus-based approach that does not require bins at all. Our relative peak dominance approach returns to the bin structure of Granger but with two additions: a formal partition $\mathcal{P}$ that makes δ explicit, and a comparative construction (seasonal-bin-maximum versus nonseasonal-bin-maximum) that provides the EVT-based null distribution that Granger lacked. The strengths of the spectral category as a whole (model-free definition, calendar-invariant frequency grid) are real and have been recognized since Nerlove. The limitation is that "peak" and "neighborhood" remained under-specified for decades, and it is precisely this gap that the partition-based framework developed in Section 3 is designed to close.

2.1.5. Discussion

The four categories of definitions surveyed above target four distinct population objects: conditional mean structure within a calendar partition, polynomial roots at seasonal frequencies, a latent component identified under a model-specific restriction, and a functional of the spectral density. Each object induces a different null hypothesis when operationalized as a test (Table 1). The Friedman and Kruskal-Wallis tests ask whether seasonal levels differ; HEGY asks whether unit roots are present at seasonal frequencies; Canova-Hansen asks whether seasonal parameters are stable; the McElroy-Holan test asks whether the spectral density exhibits local curvature consistent with a peak. A series can satisfy one definition of "seasonal" while failing another, not because of estimation error but because the definitions characterize different features of the data-generating process. The calendar-anchored tradition detects level differences but is silent on persistence; the unit-root tradition detects persistence but presupposes a seasonal ARIMA structure; the signal-extraction tradition recovers a component but requires an identifying restriction that is acknowledged to be arbitrary; the spectral tradition detects peaks but, until recently, lacked a formal specification of what constitutes one.

The existing definitions in the literature are therefore best understood not as competing answers to a single question but as partial characterizations of a larger space. Each category identifies a subset of processes that it considers "seasonal," and these subsets overlap without necessarily coinciding. A series with deterministic seasonal dummies falls within the calendar-anchored, signal-extraction, and spectral categories simultaneously; a series with a stochastic seasonal cycle falls within the unit-root, signal-extraction, and spectral categories but may evade the calendar-anchored tests if the pattern evolves faster than the test's span can track. The superset, the collection of all processes that exhibit seasonal behavior in any of these senses, has no single, unified definition in the existing literature. The tests in Table 1 probe different regions of this superset, and the diagnostic batteries prescribed by the ESS Guidelines (European Statistical System, Eurostat, 2024) and the Eurostat Handbook (Eurostat, 2018) are, in effect, attempts to cover the superset by taking the union of several subsets. The cost of this approach, as documented in Section 2, is that the union is informal: no explicit population object corresponds to it, and no single null hypothesis governs it.

The framework we develop in the subsequent sections draws on each of these traditions. From the spectral category we inherit the population object: a functional of $f(\omega)$ evaluated at seasonal frequencies, following the lineage from Nerlove (1964) through Granger (1978). From Granger specifically we inherit the bin structure, the idea that seasonal and nonseasonal frequency bands can be defined by a partition of $(0, \pi]$, though we replace his unspecified δ with an explicit partition $\mathcal{P}$ parameterized by a bin count M. From the classical periodogram-extremes lineage (Fisher, Hannan, Davis-Mikosch) we inherit the use of extreme-value theory on periodogram ordinates to derive an analytical null distribution, extending it from a single-frequency object to a comparative statistic across the full partition. From the signal-extraction tradition we take seriously the demand, articulated most forcefully by Bell and Hillmer (1984), that a definition of seasonality should specify what to estimate in an adjustment procedure, not merely when a

series is seasonal; our SSI adjustment operator is the direct response to that demand, constructed as the logical conjugate of the test. From the calendar-anchored tradition we take the motivating lesson rather than the technical apparatus: by grounding the partition in repetition rates rather than in Gregorian calendar positions, the framework avoids conditioning the definition on a particular calendar system, a choice that limits the influence of institutional or cultural conventions on what counts as seasonal.[13]

3. A Way Forward: Definition and Test Statistic

In this section we pursue two objectives that mirror the themes of Section 2. First, we define seasonality in the frequency domain through a population object that makes its primitive notions explicit. Second, we derive a sample statistic and a null distribution that are tied to that same object, thereby enforcing the epistemic closure that is often absent in prevailing definition–test–adjustment workflows.

Our approach rests on two building blocks. The first is a fixed *frame of reference*: a deterministic partition of the frequency domain into disjoint bins, chosen *ex ante*. The second is an operational definition of a *peak* as a relative notion of height, defined by extrema computed within the same bins. This construction abstracts away from calendar-specific regressors (Easter, trading-day, moving holidays) and instead targets repetitive behavior as a property of spectral structure, while retaining transparency about the subjective inputs that remain—namely the designation of a seasonal frequency set. The contribution is not to eliminate all discretion, but to narrow the scope for idiosyncratic choices by anchoring both testing and (later) adjustment to the same explicitly defined subset of the spectrum.

3.1. A Population Object Defined by a Seasonal Frequency Partition

Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a real-valued, weakly stationary process with mean μ, autocovariance function

$$\gamma_Z(\tau) \;=\; \mathbb{E}[(Z_t-\mu)(Z_{t+\tau}-\mu)]\,, \tag{1}$$

and spectral density

$$f_Z(\omega) \;=\; \sum_{\tau=-\infty}^{\infty} \gamma_Z(\tau)e^{-i\omega\tau}, \qquad \omega\in[-\pi,\pi], \tag{2}$$

with the inversion formula $\gamma_Z(\tau) = (2\pi)^{-1}\int_{-\pi}^{\pi} f_Z(\omega)e^{i\omega\tau}\,d\omega$. We impose the standard regularity condition $\sum_\tau |\gamma_Z(\tau)| < \infty$, implying that f_Z is continuous on $[-\pi,\pi]$. We emphasize that the population object in this section is f_Z, not a decomposition component, seasonal dummy coefficients, or a seasonal unit root factor; the taxonomy in Section 2 clarifies that these are distinct objects, and our goal here is to commit to one explicitly.

Defining Primitives Let $P_{\max}\in\mathbb{N}$ be an *ex ante* upper bound on the longest fundamental period (in years) we are willing to regard as seasonally relevant, and let N be the number of observations per calendar year. For each fundamental p define the associated harmonic set

$$\Omega_p(N) = \left\{\frac{2\pi k}{pN} : \; k=1,2,\ldots,\; \frac{pN}{2}\right\} \subset (0,\pi), \tag{3}$$

[13] This is not a claim that calendar effects are unimportant. Trading-day, moving-holiday, and fiscal-year effects are genuine sources of periodic variation that production pipelines must handle. The claim is narrower: the *definition* of seasonality should not depend on which calendar the analyst happens to use, even if the *application* of that definition to a specific series requires specifying the seasonal period s and its harmonics. The partition $\mathcal{P}$ takes s as given; it does not take the Gregorian calendar as given.

which excludes $\omega = 0$ and (when pN is even) excludes the Nyquist frequency π. The *complete* candidate set of calendar harmonics up to $P_{\max}$ years is

$$\Omega(N; P_{\max}) = \bigcup_{p=1}^{P_{\max}} \Omega_p(N). \tag{4}$$

Given a chosen set of fundamentals $\mathcal{P} \subseteq \{1, \ldots, P_{\max}\}$ (e.g., $\mathcal{P} = \{1, 2, 4\}$ for annual, biennial, and quadrennial cycles), the induced "Gregorian" seasonal subset is

$$\Omega_G(\mathcal{P}, N) = \bigcup_{p \in \mathcal{P}} \Omega_p(N) \subseteq \Omega(N; P_{\max}). \tag{5}$$

The common special case is $\mathcal{P} = \{1\}$, but longer fundamentals ($p > 1$) can be empirically relevant (e.g., quadrennial political cycles) and to avoid any degeneracy it is assumed $\mathcal{P} \neq \emptyset$.

Partitioning the Frequency Domain Rather than introducing *ad hoc* windows $\omega_s \pm \delta$ around points in $\Omega_G(\mathcal{P}, N)$ (Granger, 1978), we fix a deterministic partition of $(0, \pi)$ into M disjoint, equal-width bins,

$$B_m = \left(\frac{\pi(m-1)}{M}, \frac{\pi m}{M} \right], \qquad m = 1, \ldots, M, \tag{6}$$

with bin maximum,

$$b_m(f_Z) := \sup_{\omega \in B_m} f_Z(\omega), \tag{7}$$

and classify bins according to whether they intersect the seasonal set:

$$\begin{aligned} \mathcal{M}_1(\mathcal{P}, N; M) &= \{m : \mathrm{int}(B_m) \cap \Omega_G(\mathcal{P}, N) \neq \emptyset\} \\ \mathcal{M}_0(\mathcal{P}, N; M) &= \{1, \ldots, M\} \setminus \mathcal{M}_1(\mathcal{P}, N; M). \end{aligned} \tag{8}$$

We treat M as a design parameter fixed *ex ante*. Since M is fixed, $T/M \to \infty$ as $T \to \infty$, ensuring that the number of Fourier frequencies per bin diverges. We also choose M so that each $\omega_s \in \Omega_G(\mathcal{P}, N)$ lies in the interior of some bin B_m; this is always possible because, for any fixed ω_s, the set of M placing ω_s on a bin boundary is countable, hence avoidable. The conceptual role of (6)–(8) is to render the frame of reference fixed *ex ante*: the bins are declared before inspecting the data, and the only remaining modeling input is $\Omega_G(\mathcal{P}, N)$, which is itself explicit and interpretable.

The Population Object We define a population separation functional

$$\Delta(f_Z; \mathcal{P}, N, M) = \underbrace{\max_{m \in \mathcal{M}_1(\mathcal{P}, N; M)} b_m(f_Z)}_{\text{largest seasonal-bin peak}} - \underbrace{\max_{m \in \mathcal{M}_0(\mathcal{P}, N; M)} b_m(f_Z)}_{\text{largest nonseasonal-bin peak}}. \tag{9}$$

We call $\{Z_t\}$ *seasonal relative to* $(\mathcal{P}, N, M)$ if $\Delta(f_Z; \mathcal{P}, N, M) > 0$. This definition is intentionally narrow: it declares seasonality as a *relative* property of spectral maxima across two explicitly defined subsets of the frequency domain. It thereby differs from mean-based definitions and from unit-root definitions in the taxonomy, and it is designed to align with the operational logic of spectral-peak diagnostics while making the underlying primitives (the partition and the peak operator) explicit.

Although motivated by "peaks" language, $\Delta(f_Z; \mathcal{P}, N, M)$ is an extreme-value contrast of upper envelopes across two frequency subsets. This includes both narrow-band seasonal peaks and diffuse seasonal energy that remains concentrated within the seasonal bins. The definition is intentionally indexed by the

design resolution M; it formalizes a particular notion of seasonality, dominance of seasonal-bin spectral power relative to the nonseasonal envelope, chosen to align exactly with the sample test statistic and adjustment procedure introduced below.

Two remarks clarify the role of M. First, M governs the resolution of the frame of reference: the bin width π/M determines how localized the notion of a "seasonal peak" is around the calendar harmonics in $\Omega_G(\mathcal{P}, N)$. Smaller bins (larger M) yield a more localized comparison, while larger bins (smaller M) aggregate across wider frequency ranges; in either case, the definition remains anchored to the same *ex ante* partition through $b_m(f_Z)$ and $\mathcal{M}_1(\mathcal{P}, N; M)$. Second, the asymptotic requirement $T/M \to \infty$ is imposed only to support the sample analog developed below: it ensures that each bin contains a diverging number of Fourier frequencies so that within-bin maxima of the periodogram admit classical extreme-value approximations. Accordingly, $\Delta(f_Z; \mathcal{P}, N, M)$ should be interpreted as a population object indexed by the fixed design choice M, rather than as a limit in M that collapses the seasonal set to points.

3.2. A Sample Statistic and Its Limiting Distribution

We now construct a statistic that targets the same object as that in Equation 9. Let $\{z_t\}_{t=1}^T$ be a finite sample from $\{Z_t\}$, and let $\bar{z} = T^{-1} \sum_{t=1}^T z_t$. Define the discrete Fourier frequencies $\omega_j = 2\pi j/T$ for $j \in \mathcal{J} := \{1, \ldots, \lfloor (T-1)/2 \rfloor\}$. Without loss of generality, we work with the demeaned sample, $\tilde{z}_t = z_t - \bar{z}$ and define the periodogram using $\tilde{z}_t$. For $j \geq 1$, this replacement does not affect the asymptotic distribution under the null. The (raw) periodogram is

$$I_T(\omega_j) = \frac{1}{T} \left| \sum_{t=1}^{T} \tilde{z}_t e^{-i\omega_j t} \right|^2, \qquad j \in \mathcal{J}. \tag{10}$$

As is well known, the sample periodogram is not a consistent estimator of $f_Z(\omega)$ for fixed ω. However, rather than attempting to correct for this inconsistency, we exploit it: under a white-noise null, the periodogram ordinates have a stable, non-degenerate limiting distribution and, crucially, their maxima are characterized by classical extreme-value approximations.

Deriving the Null Distribution Under H_0 we assume Z_t is white noise with variance σ^2, so $f_Z(\omega) \equiv \sigma^2$ on $(0, \pi)$ and, for $j \in \mathcal{J}$, the periodogram ordinates satisfy $I_T(\omega_j)/\sigma^2 \Rightarrow \mathrm{Exp}(1)$ and are asymptotically independent (Politis and McElroy, 2019), where $\mathrm{Exp}(1)$ denotes an exponential with unit rate (equivalently, unit mean). The alternative of interest is the class of processes satisfying $\Delta(f_Z; \mathcal{P}, N, M) > 0$.

Using B_m as defined earlier, let $\mathcal{J}_m := \{j \in \mathcal{J} : \omega_j \in B_m\}$ and $n_m := |\mathcal{J}_m|$. Define the within-bin periodogram maximum as

$$b_m(I_T) := \max_{j \in \mathcal{J}_m} I_T(\omega_j). \tag{11}$$

Then, we can define the sample statistic as

$$\Delta(z; \mathcal{P}, N, M) := \max_{m \in \mathcal{M}_1(\mathcal{P}, N; M)} b_m(I_T) - \max_{m \in \mathcal{M}_0(\mathcal{P}, N; M)} b_m(I_T). \tag{12}$$

Because the bins form a partition, we have that $\max_{m \in \mathcal{M}_1} b_m(I_T) = \max_{j \in \mathcal{J}_1} I_T(\omega_j)$ and similarly for $\mathcal{M}_0$, where

$$\mathcal{J}_1 := \{j : \omega_j \in B_m \text{ for some } m \in \mathcal{M}_1(\mathcal{P}, N; M)\}, \qquad \mathcal{J}_0 := \mathcal{J} \setminus \mathcal{J}_1. \tag{13}$$

Let $N_1 := |\mathcal{J}_1|$ and $N_0 := |\mathcal{J}_0|$ denote the total count of ordinates in seasonal bins and nonseasonal bins accordingly. Under H_0, $I_T(\omega_j)/\tau \Rightarrow \mathrm{Exp}(1)$ with asymptotic independence across distinct Fourier

frequencies, where $\tau = \mathbb{E}[I_T(\omega_j)] = \sigma^2$. Classical extreme-value theory (EVT) for i.i.d. exponential distributions implies that, for $g \in \{0, 1\}$,

$$\frac{\max_{j \in \mathcal{J}_g} I_T(\omega_j) - \tau \log N_g}{\tau} \Rightarrow \text{Gumbel}(0, 1), \tag{14}$$

so $\max_{j \in \mathcal{J}_g} I_T(\omega_j) \approx \text{Gumbel}(\tau \log N_g, \tau)$. Moreover, the maxima over $\mathcal{J}_1$ and $\mathcal{J}_0$ are (asymptotically) independent since they are functions of disjoint sets of ordinates. Therefore,

$$\Delta(z; \mathcal{P}, N, M) \mid H_0 \Rightarrow \text{Logistic} \quad \tau \log \frac{N_1}{N_0} \quad , \tau \quad . \tag{15}$$

In practice one may either standardize the input series to enforce $\tau = 1$, or use the plug-in estimate $\hat{\tau} = s^2$, yielding $\text{Logistic}(\hat{\tau} \log(N_1/N_0), \hat{\tau})$ for critical values.

In applied settings, we recommend the following procedure for testing. First, if the series is nonstationary (trending), apply first differences: $\delta z_t = z_t - z_{t-1}$. If the differenced series exhibits nonseasonal autocorrelation, fit a nonseasonal ARMA model by BIC and extract the residuals $\hat{e}_t$. Finally, compute $\Delta(\hat{e}; \mathcal{P}, N, M)$ and compare to the appropriate quantiles of the null distribution. Under this workflow, the Logistic null distribution applies to $\hat{e}_t$ with $\tau = \hat{\sigma}_e^2$. In our empirical experience, simple first-differencing provides adequate size control for most economic time series, obviating the need for formal ARMA pre-whitening. Section 5 provides simulation evidence supporting this claim under a range of colored-noise data-generating processes.

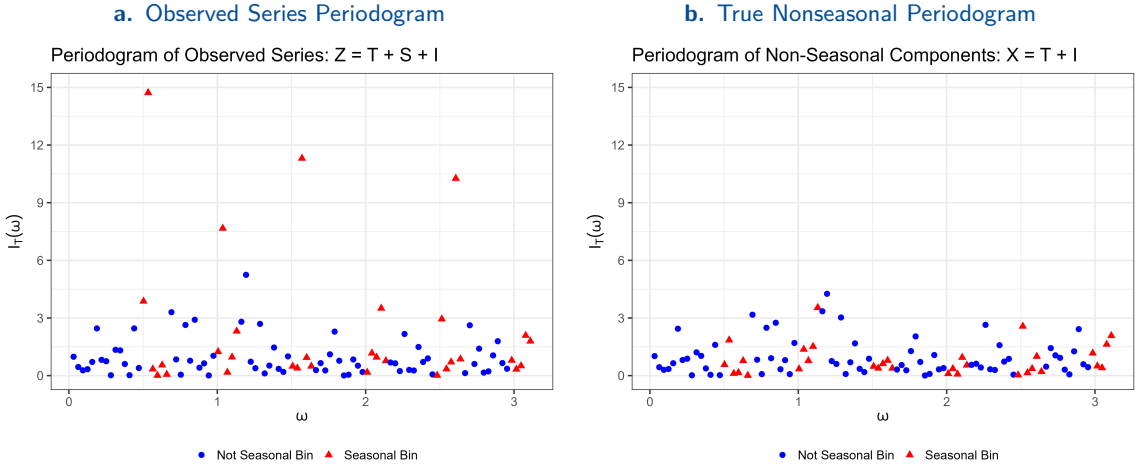
4. A Way Forward: Seasonal Adjustment

In Section 3, we formalized seasonality in the frequency domain as a property of the *relative peak dominance* over an *ex ante* partition of the frequency domain. We argued that, in the absence of seasonality, the spectral maxima in designated seasonal bins, $\mathcal{M}_1$, should be statistically indistinguishable from those in nonseasonal bins, $\mathcal{M}_0$. This definition relies on the relative height of spectral peaks, analogous to comparing the elevations of two mountain ranges relative to a common sea level. Under the null hypothesis of no seasonality, the difference in these maxima converges to a logistic distribution with known parameters, providing a tractable test statistic.

In this section, we construct a proof of concept for an adjustment operator that is the logical conjugate of this definition. Our objective is to remove the seasonal component in a manner that is (i) statistically robust, (ii) analytically transparent, (iii) agnostic to functional form, and (iv) scalable for production environments. To achieve this, we propose *Stochastic Spectral Imputation* (SSI), a non-parametric procedure that manipulates the series directly in the frequency domain. The intuition for SSI rests on the counterfactual implied by our definition. For a stationary process, if the excess power at seasonal frequencies were removed, the residual spectral density should resemble the underlying nonseasonal process.

For example, suppose we generate a series that is an additive function of trend, seasonal and irregular components: $Z = T + S + I$. We observe the series z_t with periodogram in Figure 1a, and this series comes with a nonseasonal process with periodogram outlined in Figure 1b. In the real world, we don't know what the nonseasonal process is and as such Figure 1b is hidden. However, our objective is to modify the ordinates in the seasonal bins from Figure 1a so that the result is similar to that in Figure 1b overall. To accomplish this, SSI uses ordinates in the nonseasonal bins as donors to produce a new series with a periodogram that resembles that of Figure 1b. The remainder of this section will detail the SSI algorithm along with any options one may use in its configuration.

Figure 1. Conceptual Overview of Spectral Imputation



Note: Panel (a) shows the spectral density of a process with distinct seasonal peaks. Panel (b) shows the periodogram of the true nonseasonal components, something that is typically a hidden object.

4.1. The SSI Algorithm

The Stochastic Spectral Imputation (SSI) algorithm operationalizes our definition of seasonality by treating the adjustment process as a surgical signal-extraction task in the frequency domain. Unlike filter-based methods that aggregate information across the spectrum, SSI targets only the specific frequency bands identified as seasonal during the detection phase thereby enforcing epistemic closure, a core contribution of this paper.

Let x_t be the observed series for $t = 1, \ldots, T$. We compute the discrete Fourier transform (DFT) coefficients, $d_x(\omega_j)$, such that

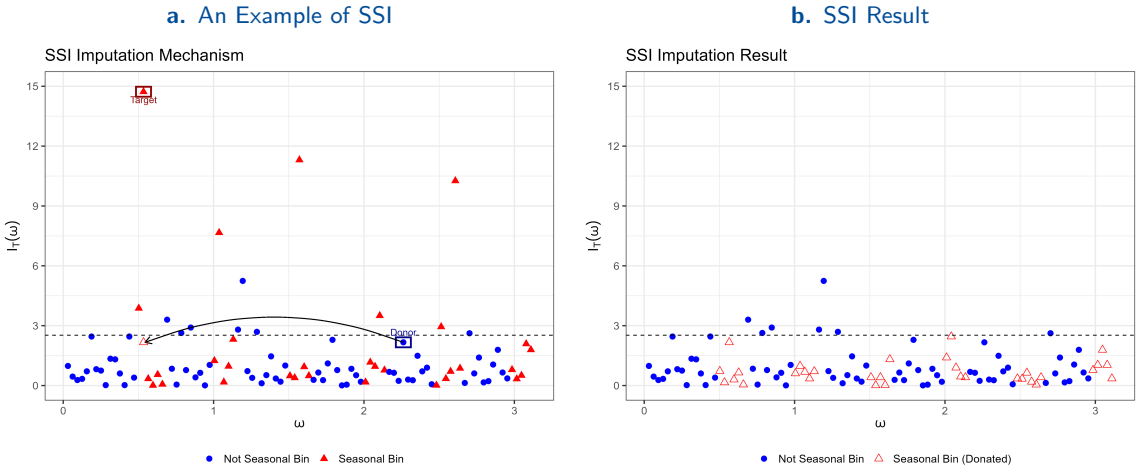
$$x_t = \frac{1}{T} \sum_{j=0}^{T_1} d_x(\omega_j) e^{i\omega_j t} \tag{16}$$

where $\omega_j = 2\pi j/T$ for $j = 0, \ldots, T-1$. Using partitions $\mathcal{M}_1$ and $\mathcal{M}_0$ established in Section 3, we define the set of seasonal Fourier indices, $\mathcal{J}_S$, as those falling within the seasonal bins, and the donor set, $\mathcal{J}_D$, as the indices corresponding to nonseasonal background frequencies.

The procedure for constructing the adjusted series x_t^* consists of four primary steps. First, we identify $\mathcal{J}_S = \{j \in \mathcal{J} : \omega_j \in B_m \text{ for some } m \in \mathcal{M}_1\}$. To ensure the donor set represents the background noise floor, we optionally filter $\mathcal{J}_D$ using a quantile threshold, q_D, to exclude any nonseasonal frequency outliers that might distort the background distribution. The second step is the imputation; for each seasonal index $j \in \mathcal{J}_S$ we replace the observed magnitude, $|d_x(\omega_j)|$ with a draw from the empirical magnitude distribution of the donor set. Specifically, we sample k uniformly from $\mathcal{J}_D$ and set $|d^*(\omega_j)| = |d_x(\omega_k)|$. This means that the power in the seasonal bins is reduced to a level that is statistically indistinguishable from the nonseasonal background.

Step three involves the phase component of the DFT coefficient. Because the seasonal component is removed, the original phase information at seasonal frequencies is often no longer meaningful. SSI allows

Figure 2. Building Intuition for SSI



Note: Panel (a) shows a basic representation of the SSI algorithm through a single target-donor pair. Panel (b) shows the end result of the SSI algorithm, a periodogram that is comparable in structure to that found in Figure 1b.

for phase ϕ_j to be kept from the original series, inherited from the donor, or drawn from $U(-\pi, \pi)$. For the Nyquist frequency and the DC component, we enforce the necessary real-value constraints. Finally, to ensure the adjusted series x_t^* remains real-valued, step four enforces conjugate symmetry such that $d^*(\omega_{T-j}) = \overline{d^*(\omega_j)}$ for all j. The final adjusted series is obtained via the inverse discrete Fourier transform (IDFT) of the modified coefficients. A pseudo-code representation of SSI is covered in Algorithm 1 (see Appendix A).

4.2. Intuition for SSI

In Figure 2 we continue our example from Figure 1 to help build intuition and illustrate the impact of the SSI adjustment process. SSI moves iteratively through ordinates found to be within seasonal bins. These are the targets, and in Figure 2a we have marked one such target with a red box. A random ordinate is chosen from the donor pool with magnitude less than q_D, given by the dashed line in Figures 2a and 2b. The magnitude of the DFT coefficient corresponding to that ordinate is used in place of the magnitude from the target ordinate as illustrated by the open circle. The goal is that this should be in a location similar to where the power of that frequency would be in the periodogram of the nonseasonal components. As mentioned earlier, the algorithm allows for the choice of the phase for that DFT coefficient to be kept from the target coefficient, its donor, or drawn randomly. Finally, conjugate symmetry is enforced for the second DFT coefficient implied by the target. This process is repeated for all ordinates inside $\mathcal{M}_1$. In Figure 2b we show the impact of a complete run through the algorithm. The open triangles represent the donated magnitudes that replaced ordinates in the seasonal bins. The final result of the SSI algorithm is the ordinates in Figure 2b, both in seasonal and nonseasonal bins, appear to be very similar to the oracle periodogram from Figure 1b. Moreover, one can visually see that there is no dominant peak in the seasonal bins compared to the nonseasonal bins. The value for Δ from Figure 1a is 9.48 with a p-value of 0.001 while the adjusted data given in Figure 2b produces $\Delta = -2.79$ with a p-value of 0.921.

The development of SSI was guided by four essential requirements for modern seasonal adjustment: (i) statistical robustness, (ii) analytical transparency, (iii) agnosticism toward functional form, and (iv)

computational scalability. By utilizing the empirical distribution of the nonseasonal bins as a donor pool, SSI functions as a frequency-domain resampling procedure analogous to the bootstrap. This approach ensures the method remains computationally efficient and avoids the parametric commitments inherent in standard model-based pipelines. While this paper establishes the asymptotic justification for SSI, the framework's flexibility allows for the construction of uncertainty intervals through repeated draws, providing a path for future extensions into fully probabilistic seasonal adjustment. In the next section our simulations will demonstrate that SSI is not merely an alternative procedure, but the logical conjugate to our frequency-domain definition.

5. Monte Carlo Simulations

This section conducts simulation studies for our proposed testing and adjustment procedures. First, we show that the empirical null distribution of our test statistic almost exactly approximates its theoretical counterpart. Then, we contrast the power of our test statistic with popular alternatives proposed by the literature and find that it achieves comparable performance across several different specifications. Last, we explore the statistical properties of SSI under a known data generating process (DGP) and illustrate its advantages relative to X-13. We first describe the setup of and assumptions used in the DGP, which follows Busetti and Harvey (2003).

5.1. Setup

The data generating process is the basic structural model used in Busetti and Harvey (2003) consisting of the four equations below.

$$z_t = \mu_t + S_t^{'}\gamma_t + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2) \tag{17}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0, \sigma_\eta^2) \tag{18}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \qquad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2) \tag{19}$$

$$\gamma_t = \boldsymbol{\alpha} + \rho\gamma_{t-1} + \kappa_t, \qquad \kappa_t \sim \text{NID}(0, \sigma_\kappa^2). \tag{20}$$

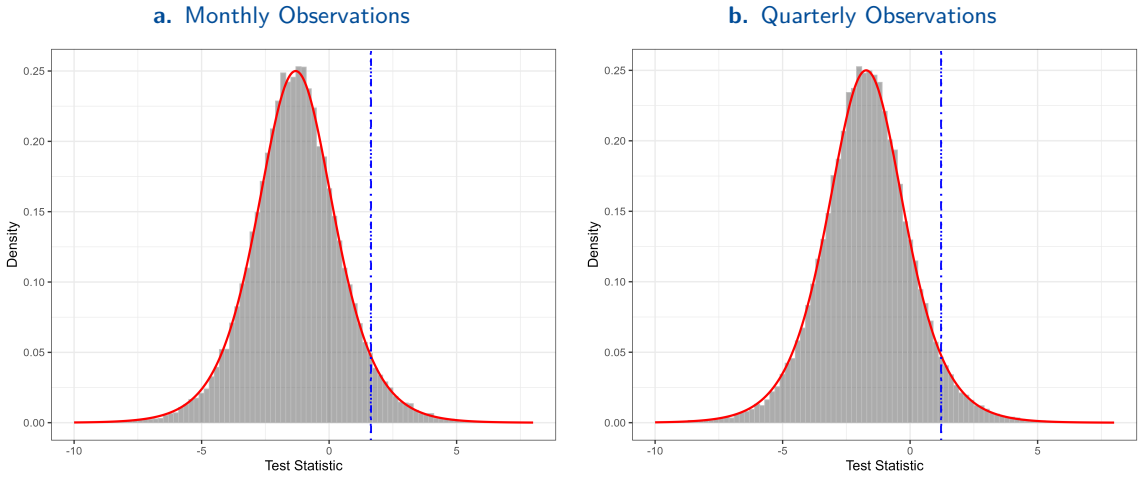
Here S_t is an $(N-1)$ vector of trigonometric seasonal variables and γ_t is an $(N-1)$ vector that controls their amplitudes. It is assumed that γ_t evolves according to an AR(1) process with some intercept $\boldsymbol{\alpha}$, an $(N-1)$ vector of constant values, and white-noise innovations κ_t. In Busetti and Harvey (2003) the value of ρ was fixed at one, meaning that the amplitudes evolved according to a random walk and, over a long sample period, the variance of the seasonal component would come to dominate that of z_t. By allowing γ_t to evolve as an AR(1) process we preserve the random walk aspect but also allow for a stationary, stochastically evolving set of amplitudes, and a fixed or memoryless evolving set of amplitudes if we so desire. We fix $T = 200$ across all simulations and report rejection rates consistent with a 5% significance level.

5.2. Evaluating the Empirical Null Distribution of Δ

Before proceeding, we first evaluate the empirical null distribution as compared to the theoretical null of a $\text{Logistic}(\tau \log(N_1/N_0), \tau)$ derived earlier.[14] To do this, we set $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\kappa^2 = 0$ and $\boldsymbol{\alpha} = 0$, thereby allowing us to generate sample data from a white-noise process. We also set $M = 2N - 1$ leading to, in

[14] For simplicity in evaluation we scale the series such that $\tau = 1$ as indicated thus removing variation due to sample variance entering the empirical null distribution.

Figure 3. A Comparison of the Empirical and Theoretical Null



Note: We fix $T = 200$ across simulations and scale the series so that $\tau = 1$. Vertical lines are consistent with nominal and empirical 5% values.

the monthly case, about two-thirds of the periodogram ordinates occurring in nonseasonal bins; this assumption assures convergence under the EVT conditions. Figure 3 plots the proposed test statistic under these conditions over the course of $50,000$ iterations. In Figure 3a, we can see that the histogram follows the theoretical density almost exactly and that the empirical 5% critical-value is nearly identical to the nominal value, producing size close to the nominal 5%. Figure 3b shows the null for quarterly observations, where we set $N = 4$ and $M = 7$. Again, we see that the empirical distribution mirrors the theoretical construct outlined in Section 3. Therefore, in finite samples the empirical null distribution converges to the nominal null distribution.[15]

It is important to note that in practical applications an observed series is generally not composed of white noise plus a seasonal component. Instead, what we tend to observe is a composite (usually additive) of a trend, seasonal, and irregular component, the latter of which is often white noise. It is therefore important to think about how to handle the presence of both a trend and non-white noise irregular component in the testing procedure. In Table 2 we explore the empirical size of our test statistic under a variety of colored-noise null generating processes with no seasonal component. In column one we report the empirical rejection rates when the test is applied directly to the raw data. From this we see that, as the persistence of the noise increases, the test generally becomes poorly sized with rejection rates falling below the nominal rate as the autoregressive parameter increases and over-rejecting when the moving average parameter increases. In columns two through four we explore three different pre-whitening approaches to handle this issue: fitting the known ARMA order via maximum likelihood and using an exhaustive search procedure to select the ARMA order via information criterion, in this case AIC and BIC.[16]

[15] However, more work is needed to understand the implications of the choice of M under finite T.

[16] There is a great deal of literature on the relative trade-offs between these two selection criteria. It is well known that AIC tends to favor larger, more complex models due to its linear in parameters penalty term. On the other hand, BIC favors more parsimonious models due to its penalty term being a multiplicative function of the number of parameters and the log of the sample size. In Mills and Prasad (1992) the authors find that BIC tends to perform better in selecting the true model in a time series context.

Table 2. Empirical Size Under Colored-Noise Null DGPs

DGP	Raw	True	AIC	BIC	Δ
Panel A: $T = 200$					
AR(1), $\phi = 0.3$	0.052	0.045	0.037	0.043	0.100
AR(1), $\phi = 0.5$	0.051	0.047	0.036	0.045	0.083
AR(1), $\phi = 0.8$	0.003	0.049	0.040	0.049	0.063
AR(1), $\phi = 0.95$	0.000	0.051	0.042	0.051	0.055
MA(1), $\theta = 0.3$	0.058	0.048	0.039	0.048	0.093
MA(1), $\theta = 0.7$	0.066	0.045	0.038	0.043	0.088
ARMA(1,1), $\phi{=}0.5$, $\theta{=}0.3$	0.056	0.042	0.036	0.043	0.071
AR(2), spectral peak	0.003	0.049	0.040	0.048	0.006
Panel B: $T = 600$					
AR(1), $\phi = 0.3$	0.049	0.050	0.042	0.050	0.116
AR(1), $\phi = 0.5$	0.040	0.050	0.042	0.049	0.091
AR(1), $\phi = 0.8$	0.001	0.051	0.047	0.051	0.061
AR(1), $\phi = 0.95$	0.000	0.050	0.043	0.048	0.055
MA(1), $\theta = 0.3$	0.066	0.048	0.042	0.047	0.092
MA(1), $\theta = 0.7$	0.067	0.047	0.043	0.047	0.097
ARMA(1,1), $\phi{=}0.5$, $\theta{=}0.3$	0.038	0.046	0.041	0.046	0.080
AR(2), spectral peak	0.000	0.043	0.038	0.043	0.001

Note: Empirical rejection rates at nominal $\alpha = 0.05$ under colored-noise null DGPs with no seasonal component. $N = 12$, $M = 23$, $P = 1$. Panel A uses $T = 200$ with 10,000 replications per cell; Panel B uses $T = 600$ with 5,000 replications per cell. Pre-whitening lanes: *Raw* applies the test directly; *True* fits the known ARMA order via ML; *AIC* and *BIC* select a nonseasonal ARMA model via exhaustive search over `auto.arima`; Δ first-differences the series. The AR(2) spectral peak places complex roots at modulus $r = 0.9$ and frequency $\omega_0 = 2\pi/4.8 \approx 1.31$ rad, near but not at the seasonal harmonic $2\pi \cdot 2/12 \approx 1.05$ rad. BIC-selected pre-whitening is well-calibrated at both sample sizes (0.043–0.051). AIC is systematically conservative and partially converges toward nominal at $T = 600$. First-differencing of stationary processes over-rejects, and the distortion worsens at larger T.

It is clear from these results that pre-whitening the data is critical to obtaining a well-sized test statistic under a colored noise series. Moreover, the choice of pre-whitening criterion matters. Overall BIC tends to provide empirical rejection rates closer to the nominal rate across a variety of ARMA specifications while AIC tends to overfit the data and produce empirical rejection rates that are too low. Therefore, we recommend relying on BIC to select the nonseasonal ARMA order for pre-whitening the data prior to applying our test statistic. Finally, in column five we report the empirical rejection rates when the test is applied to the raw data after first differencing. It is clear that this approach is strictly inferior to pre-whitening as the empirical rejection rates are generally much higher than the nominal rate.

Finally, we note that M is a choice parameter that determines the resolution of the periodogram evaluation. In Table 3 we explore the sensitivity of the empirical null distribution to different choices of M under the same white noise generating process as above. We find that the empirical null distribution is quite robust to different choices of M with the empirical critical values being close to the nominal value across a wide range of M. This is an important finding, as it suggests that the test statistic is not particularly sensitive to the choice of M under the null, a result likely being driven by the relatively fast

Table 3. Sensitivity of Empirical Size to the Bin Count M Under White Noise

T	M	Ordinates/bin	Rejection rate	SE
200	11	9.0	0.050	0.002
200	19	5.2	0.049	0.002
200	23	4.3	0.048	0.002
200	31	3.2	0.049	0.002
600	11	27.2	0.047	0.002
600	19	15.7	0.050	0.002
600	23	13.0	0.050	0.002
600	31	9.6	0.049	0.002

Note: Empirical rejection rates at nominal $\alpha = 0.05$ under a standard Gaussian white-noise null, with $N = 12$, $P = 1$. Each cell consists of $10{,}000$ replications. The column "Ordinates/bin" reports $\lfloor (T-1)/2 \rfloor / M$, the average number of Fourier frequencies per bin. Values of $M \in \{15, 27\}$ are excluded because seasonal frequencies $\pi/3$ and $2\pi/3$ fall exactly on bin boundaries at those resolutions. All rejection rates fall within the $[0.046, 0.054]$ acceptance band, indicating that the test's finite-sample size is insensitive to M across this range — even at $M = 31$, where bins contain only ≈ 3 ordinates.

convergence to the Gumbel distribution of the maximum of the periodogram ordinates under the EVT conditions. As mentioned earlier, we have found that, in general, choosing $M = 2N - 1$ leads to good performance of the test statistic in small samples.

5.3. Evaluating Empirical Power

Now, we explore the empirical power of our test statistic Δ. For comparison, we calculate rejection rates for a battery of other popular test statistics found throughout the literature: the Canova-Hansen (CH) test (Canova and Hansen, 1995), both the parametric and non-parametric (spectral) versions of the Busetti-Harvey test (BHp and BHs) (Busetti and Harvey, 2003), and the McElroy-Roy visual significance test (VS) McElroy and Roy (2021). Following Busetti and Harvey (2003) we set the lag parameter, m, for appropriate test statistics to a full year of observations. Following McElroy and Roy (2021) we set $\delta = 30$, a similar parameter to our M, and retain other suggested parameters. To make the VS test null comparable to that of Δ we jointly test the significance of all peaks in the spectrum using the Benjamini-Hochberg procedure to control the family-wise error rate (Benjamini and Hochberg, 1995).[17] For the other tests, a joint test statistic, null distribution, and critical values are all provided and we use them accordingly.

We examine empirical rejection rates for these tests under three different assumptions on the behavior of γ_t, the process that controls the amplitudes of the seasonal variables S_t.

1. γ_t evolves as a random walk ($\rho = 1$)
2. γ_t is a stationary process with high persistence ($\rho = 0.95$), where we fix $\boldsymbol{\alpha} = 0.01$
3. γ_t is a memoryless process ($\rho = 0$), where we vary $\boldsymbol{\alpha} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$

In all three cases we hold $\sigma^2_\epsilon = 1$ and $\sigma^2_\zeta = 0$ constant while varying $\sigma^2_\eta \in \{0.0, 0.5\}$ and $\sigma^2_\kappa \in \{.01, .02, .03, .04, .05\}$. For cases where $\sigma^2_\eta > 0$, we assume that the data are non-stationary and take

[17] A big thank you to Tucker McElroy who graciously provided us replication code for this test statistic in R with the recommended tuning parameters and suggested the Hochberg procedure to correct for multiple comparisons.

first differences, rather than first conducting a test for stationarity. Looking across this set of parameters shows how empirical power changes as the share of the variance of z_t that is driven by $S_t'\gamma_t$ grows. Moreover, these scenarios capture conditions commonly encountered in applied work using seasonal data, enabling us to evaluate performance in realistic settings.

Random Walk Amplitudes Table 4 shows the empirical rejection rates across tests and parameter constellations for the first scenario. Across columns, we report rejection rates under a 5% significance level for: the test statistic we introduce in this paper (Δ), the non-parametric spectral Busetti-Harvey test (BHs), the parametric Busetti-Harvey test (BHp), the Canova-Hansen test (CH), the Maravall (2012) QS test (QS), and the McElroy-Roy visual significance test (VS). In the final column, we report the proportion of total variance accounted for by the seasonal component S_t ($\%Var(S)$).

Table 4. Stochastic Seasonality ($\rho = 1$)

Rates	σ_κ	Δ	BHs	BHp	CH	QS	VS	$\%\mathrm{Var}(S)$
		Monthly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.01	0.19	0.12	0.13	0.09	0.32	0.25	0.08
	0.02	0.72	0.40	0.42	0.26	0.88	0.75	0.24
	0.03	0.95	0.75	0.77	0.53	0.99	0.96	0.41
	0.04	0.99	0.92	0.93	0.76	1.00	0.99	0.55
	0.05	1.00	0.98	0.98	0.89	1.00	1.00	0.66
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.01	0.27	0.08	0.17	0.05	0.28	0.13	0.01
	0.02	0.68	0.28	0.43	0.16	0.79	0.53	0.03
	0.03	0.91	0.58	0.71	0.34	0.97	0.84	0.06
	0.04	0.97	0.81	0.88	0.55	1.00	0.96	0.10
	0.05	0.99	0.93	0.95	0.75	1.00	0.99	0.15
		Quarterly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.01	0.07	0.09	0.08	0.08	0.11	0.15	0.02
	0.02	0.24	0.23	0.23	0.21	0.33	0.43	0.06
	0.03	0.50	0.43	0.44	0.41	0.64	0.68	0.13
	0.04	0.66	0.62	0.63	0.59	0.82	0.82	0.20
	0.05	0.79	0.74	0.76	0.72	0.92	0.88	0.28
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.01	0.23	0.09	0.18	0.09	0.12	0.09	0.00
	0.02	0.39	0.21	0.33	0.20	0.30	0.29	0.01
	0.03	0.61	0.38	0.50	0.36	0.54	0.54	0.01
	0.04	0.79	0.56	0.67	0.54	0.74	0.70	0.02
	0.05	0.88	0.70	0.79	0.68	0.86	0.81	0.04

Concentrating first on the case where $\sigma_\eta^2 = 0$ (in other words, when the trend term in the simulated process z_t is purely deterministic) and monthly simulated data, we can see that the power of all tests increases as the variance σ_κ^2 of the innovations to the scaling process γ_t grows larger. Intuitively, this makes sense as the seasonal component of the process must comprise a large enough fraction of the total variance in order for seasonality tests to detect it. However, we see that this threshold is quite low for Δ: once $\%Var(S)$ reaches about 25%, our test already achieves a power of over 70%. Moreover, Δ

strongly dominates all but the VS and QS tests across values of σ_κ; its power is in line with these two test statistics for $\sigma_\kappa \geq 0.03$.

The next panel ($\sigma_\eta^2 = 0$) shows that we obtain similar results if we allow the trend term to stochastically drift.[18] We obtain similar results whereby Δ outperforms most of the other tests and is comparable to VS and QS at higher levels of σ_κ.

For quarterly observations (bottom two panels) we obtain largely similar results. When $\sigma_\eta^2 = 0$, we find that Δ achieves comparable power to the other tests under consideration, though both QS and VS exhibit greater performance. When $\sigma_\eta^2 = 0.5$, Δ strictly dominates the other tests under consideration, including both QS and VS. We conclude that our test has good power under a process that displays nonstationary seasonality.

Stationary but Persistent Amplitudes Next, we explore the power of our test when the process that scales the amplitude of the seasonal component is stationary, yet highly persistent ($\rho = 0.95$). Table 5 displays our results for the same set of parameter values as in the previous section.

Table 5. Stochastic Seasonality ($\rho = 0.95$)

Rates	σ_κ	Δ	BHs	BHp	CH	QS	VS	$\%\mathrm{Var}(S)$
		Monthly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.01	0.49	0.06	0.06	0.05	0.90	0.70	0.19
	0.02	0.51	0.07	0.06	0.06	0.89	0.70	0.20
	0.03	0.51	0.08	0.08	0.07	0.91	0.70	0.22
	0.04	0.54	0.11	0.13	0.09	0.91	0.70	0.24
	0.05	0.56	0.16	0.17	0.13	0.92	0.71	0.27
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.01	0.63	0.05	0.11	0.04	0.82	0.42	0.02
	0.02	0.62	0.05	0.12	0.04	0.81	0.43	0.02
	0.03	0.65	0.06	0.14	0.05	0.83	0.43	0.02
	0.04	0.64	0.09	0.18	0.06	0.83	0.43	0.03
	0.05	0.66	0.10	0.22	0.07	0.85	0.46	0.03
		Quarterly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.01	0.24	0.06	0.06	0.06	0.46	0.63	0.07
	0.02	0.24	0.06	0.06	0.06	0.48	0.61	0.08
	0.03	0.26	0.08	0.08	0.08	0.51	0.62	0.08
	0.04	0.27	0.10	0.10	0.09	0.53	0.61	0.09
	0.05	0.30	0.13	0.14	0.12	0.58	0.61	0.10
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.01	0.49	0.07	0.13	0.06	0.43	0.42	0.01
	0.02	0.50	0.07	0.16	0.07	0.45	0.43	0.01
	0.03	0.50	0.08	0.17	0.08	0.46	0.42	0.01
	0.04	0.51	0.11	0.20	0.10	0.48	0.42	0.01
	0.05	0.54	0.13	0.23	0.12	0.52	0.42	0.01

For the case of monthly simulated data, we can see that Δ has reasonably good power relative to the

[18] Note that this increases (decreases) the variance of the irregular (seasonal) component by assumption.

other tests. The exception is the QS test, which strongly outperforms all others in this configuration. The reason is that the parameter values in this case induce large, positive autocorrelations in the simulated process at seasonal lags, which the QS test (a variation of the Ljung-Box test) is specifically designed to detect.

For quarterly observations with $\sigma_\eta^2 = 0$, the power of our test statistic is quite low, but it dominates the other tests under consideration besides VS and QS. With $\sigma_\eta^2 = 0.5$, Δ dominates all other test statistics under consideration in terms of power. We conclude that Δ is reasonably well-suited to detect this form of "stationary seasonality" as compared with other popular alternatives.

Memoryless Amplitudes Lastly, we consider a scenario with no persistence in the amplitude scaling process γ_t, in other words when $\rho = 0$. In this case, we vary the mean of γ_t between 0.1 and 0.5 while keeping the variance of its innovations constant. Table 6 displays the results.

Table 6. Deterministic Seasonality (Fixed $\rho = 0$)

Rates	γ_0	Δ	BHs	BHp	CH	QS	VS	$\%\mathrm{Var}(S)$
		Monthly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.10	0.12	0.05	0.05	0.05	0.25	0.19	0.06
	0.20	0.48	0.05	0.05	0.05	0.90	0.70	0.19
	0.30	0.92	0.05	0.05	0.04	1.00	0.98	0.35
	0.40	1.00	0.05	0.05	0.04	1.00	1.00	0.49
	0.50	1.00	0.05	0.05	0.05	1.00	1.00	0.60
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.10	0.24	0.05	0.10	0.03	0.25	0.10	0.01
	0.20	0.64	0.04	0.11	0.04	0.82	0.43	0.02
	0.30	0.94	0.04	0.10	0.03	1.00	0.86	0.04
	0.40	1.00	0.05	0.10	0.04	1.00	0.99	0.08
	0.50	1.00	0.04	0.11	0.04	1.00	1.00	0.12
		Quarterly Observations						
$\sigma_\eta^2/\sigma_\epsilon^2 = 0$	0.10	0.07	0.05	0.05	0.05	0.12	0.19	0.02
	0.20	0.26	0.05	0.05	0.05	0.49	0.65	0.08
	0.30	0.73	0.05	0.05	0.05	0.90	0.94	0.15
	0.40	0.97	0.06	0.05	0.05	1.00	1.00	0.25
	0.50	1.00	0.06	0.06	0.05	1.00	1.00	0.34
$\sigma_\eta^2/\sigma_\epsilon^2 = .5$	0.10	0.27	0.06	0.13	0.06	0.15	0.12	0.00
	0.20	0.50	0.06	0.13	0.06	0.45	0.44	0.01
	0.30	0.84	0.06	0.12	0.06	0.85	0.82	0.02
	0.40	0.98	0.06	0.13	0.05	0.99	0.98	0.03
	0.50	1.00	0.06	0.12	0.06	1.00	1.00	0.04

Again, our test statistic displays good power as compared with the other tests we consider. In most cases, Δ achieves performance that is either in line with or dominates the other tests. As the relative variance of S_t increases, the power of our test approaches 1 very quickly, similar to both the QS and VS tests. Collectively, we conclude from these exercises that the test statistics Δ we propose in this paper obtains good power under a wide range of scenarios, in line with the best performing alternatives in the

literature.

5.4. Test Power with Pre-Whitening

As we indicated earlier in this section, it is unlikely in practice that an observed series will be a simple combination of a white-noise irregular term and a seasonal component. To explore this possibility, we generate data with an AR(1) based irregular component and evaluate the power of our test statistic as compared to the QS statistic. Prior to applying either test, we pre-whiten the data using a nonseasonal ARMA model with order selected via BIC. Table 7 shows the results of this exercise.

Overall we note that the QS statistic is empirically undersized with a null rejection rate of approximately 1% relative to a 5% nominal. On the other hand Δ has empirical size consistent with the nominal value.[19] When a seasonal unit root is present, both statistics quickly near maximum power as σ_κ increases. When the seasonality evolves in a highly persistent but stationary manner the QS statistic exhibits greater power across the board. On the other hand, when the seasonality is deterministic ($\rho = 0$) we see Δ with higher power near the null but lower power in the larger γ_0 settings though performance can be considered roughly comparable. Overall, we find that pre-whitening the data produces a well-sized test statistic when the irregular term is not white noise, and that the power of Δ, while not strictly dominant, is largely comparable to one of the most popular alternatives.

5.5. Evaluating Adjustment with SSI

Having evaluated the empirical rejection rates of the test statistic developed in Section 3, we turn now to evaluating the efficacy of the Stochastic Spectral Imputation algorithm relative to the X-13 benchmark. As in Section 5.1, we employ a basic structural model with one critical modification: for all simulations, the variances of the trend components (σ_η^2 and σ_ζ^2) are set to zero. This produces a process z_t consisting solely of a seasonal component and an irregular component.

This constraint is central to our identification strategy. Since the true trend is deterministically zero, any trend component extracted by an adjustment procedure represents a "decomposition error" – the misidentification of low-frequency irregular variance as trend. This allows us to penalize "hallucinated" trends and treat the adjustment problem as a pure signal extraction task where the target is the precise recovery of the white-noise irregular component, ϵ_t. We assume for these simulations that the presence of seasonality has already been established.

5.5.1. Evaluation Criteria

Judging the quality of a seasonal adjustment is a multidimensional problem that we evaluate through a combination of time- and frequency-domain metrics. The first is the variance ratio (VR), defined as $\mathrm{Var}(\hat{e}_t)/\mathrm{Var}(\epsilon_t)$. This serves as a detector for "over-smoothing." A ratio significantly below unity indicates that the adjustment process has suppressed variance in the irregular component, effectively discarding signal along with noise. As noted by Maravall (1993), standard linear filters often reduce the variance of the adjusted series below that of the true nonseasonal component.

Related to variance, we evaluate the mean absolute error (MAE) of $\hat{e}_t$ relative to ϵ_t. We prioritize MAE over mean squared error (MSE) because linear filters (like X-13) mechanically alter every ordinate in the frequency domain, whereas the SSI algorithm is surgical, leaving nonseasonal ordinates largely untouched. MSE tends to be dominated by large errors, whereas MAE provides a clearer aggregate measure of these

[19] These findings echo that of Chen and Cornwall (2021) which showed the χ^2_2 is likely to be a poor approximation for the QS in finite sample settings. Moreover, we note that Chen and Cornwall (2021) also pointed out that the QS only evaluates positive autocorrelation values and is "blind" to a basic SARIMA(1) process with a negative coefficient by construction.

Table 7. Power with AR(1) Colored Noise After BIC Pre-Whitening

Signal	Test	$\phi = 0$	$\phi = 0.3$	$\phi = 0.5$	$\phi = 0.7$	$\phi = 0.9$
Panel A: Seasonal unit root ($\rho = 1$, $\sigma^2_\eta/\sigma^2_\epsilon = 0$)						
$\sigma_\kappa = 0$ (null)	Δ	0.052	0.049	0.051	0.053	0.051
	QS	0.015	0.009	0.012	0.014	0.012
$\sigma_\kappa = 0.01$	Δ	0.197	0.181	0.241	0.301	0.413
	QS	0.195	0.194	0.277	0.332	0.464
$\sigma_\kappa = 0.02$	Δ	0.730	0.645	0.747	0.815	0.867
	QS	0.857	0.848	0.908	0.941	0.967
$\sigma_\kappa = 0.03$	Δ	0.946	0.908	0.910	0.952	0.970
	QS	0.994	0.989	0.994	0.998	0.999
$\sigma_\kappa = 0.04$	Δ	0.989	0.974	0.960	0.977	0.993
	QS	1.000	1.000	0.999	1.000	1.000
$\sigma_\kappa = 0.05$	Δ	0.996	0.994	0.986	0.979	0.995
	QS	1.000	1.000	1.000	1.000	1.000
Panel B: Stochastic seasonality ($\rho = 0.95$, $\sigma^2_\eta/\sigma^2_\epsilon = 0$)						
$\sigma_\kappa = 0$ (null)	Δ	0.049	0.044	0.046	0.050	0.057
	QS	0.013	0.011	0.011	0.013	0.014
$\sigma_\kappa = 0.01$	Δ	0.494	0.408	0.556	0.652	0.783
	QS	0.833	0.839	0.940	0.977	0.996
$\sigma_\kappa = 0.02$	Δ	0.512	0.408	0.552	0.670	0.784
	QS	0.850	0.841	0.932	0.978	0.996
$\sigma_\kappa = 0.03$	Δ	0.518	0.422	0.577	0.674	0.764
	QS	0.843	0.858	0.945	0.978	0.995
$\sigma_\kappa = 0.04$	Δ	0.548	0.442	0.579	0.670	0.769
	QS	0.872	0.863	0.946	0.974	0.991
$\sigma_\kappa = 0.05$	Δ	0.568	0.463	0.579	0.676	0.754
	QS	0.892	0.885	0.948	0.977	0.989
Panel C: Deterministic seasonality ($\rho = 0$)						
$\gamma_0 = 0.00$ (null)	Δ	0.051	0.046	0.050	0.048	0.052
	QS	0.012	0.010	0.012	0.009	0.011
$\gamma_0 = 0.05$	Δ	0.065	0.057	0.065	0.070	0.077
	QS	0.028	0.027	0.032	0.031	0.045
$\gamma_0 = 0.10$	Δ	0.120	0.107	0.125	0.168	0.213
	QS	0.126	0.134	0.175	0.246	0.341
$\gamma_0 = 0.15$	Δ	0.257	0.223	0.296	0.392	0.535
	QS	0.428	0.460	0.609	0.732	0.880
$\gamma_0 = 0.20$	Δ	0.493	0.406	0.549	0.666	0.784
	QS	0.831	0.848	0.941	0.980	0.996
$\gamma_0 = 0.30$	Δ	0.911	0.787	0.875	0.947	0.975
	QS	1.000	1.000	1.000	1.000	1.000

Note: Empirical rejection rates at nominal $\alpha = 0.05$ after BIC-selected nonseasonal ARMA pre-whitening via `auto.arima(seasonal=FALSE)`. DGP: $z_t = \mu_t + S_t + n_t$, where S_t is the BSM seasonal component ($\rho = 1$ for Panel A, $\rho = 0.95$ for Panel B, $\rho = 0$ for Panel C) and $n_t \sim \text{AR}(1)$ with persistence ϕ, independent of S_t. Throughout, $\sigma^2_\eta/\sigma^2_\epsilon = 0$. $T = 200$, $N = 12$, $M = 23$; 5,000 replications per cell. The first row of each panel is the null baseline (no seasonal component; DGP reduces to AR(1) noise). Δ denotes the proposed EVT test; QS is the autocorrelation-based test of Maravall (2012). Both tests are applied to the same BIC pre-whitened residuals.

small, broadband spectral changes. Finally, in the time domain, we evaluate the correlation between $\hat{e}_t$ and ϵ_t. While VR and MAE measure scale and distance, correlation measures dynamics; a high correlation indicates that the adjustment procedure has correctly preserved the phase and timing of structural innovations.

In the frequency-domain, we also measure the spectral MAE,the mean absolute difference between periodogram ordinates of the adjusted data, $I_{\hat{e}}(\omega)$, and those of the true irregular, $I_{\epsilon}(\omega)$. Since the target irregular is white noise, this metric captures the global spectral distortion introduced by the adjustment process, with an optimal outcome of zero. Additionally, linear filters used in X-13 are known to create spectral dips or troughs at seasonal frequencies (McElroy and Roy, 2017). To quantify this, we calculate a spectral dip statistic: the average absolute difference between $I_{\hat{e}}(\omega)$ and $I_{\epsilon}(\omega)$ specifically within the designated seasonal frequency bins, $\mathcal{M}_1$. A smaller value indicates that the adjustment method successfully removed the seasonal peaks without carving out artificial valleys in the spectrum.

5.5.2. Results

Table 8 provides the results across both monthly and quarterly based series with varying levels of seasonal signal-to-noise. Across all specifications, X-13 consistently produces a Variance Ratio significantly below unity (e.g., ≈ 0.75 in monthly panels and ≈ 0.50 in quarterly panels). This confirms the well-documented tendency of linear filters to over-smooth the data by suppressing variance across the spectrum, creating a downward bias in the variance of the adjusted series. In contrast, SSI maintains a variance ratio near 1.0 across almost all specifications.

This over-smoothing behavior is supported by the spectral diagnostics, specifically the spectral bias or dip statistic. Our results show that X-13 systematically introduces deeper troughs at seasonal frequencies consistent with findings of McElroy and Roy (2017) and others. Overall, X-13 produces periodogram ordinates that are on average 50% lower than those produced by SSI relative to the ordinates from the true irregular term. By design, the filters used by X-13 are aggressive at dampening the power at seasonal frequencies in an effort to remove the seasonal signal, perhaps too aggressive given the spectral troughs it creates. Meanwhile, SSI avoids this by imputing power from a nonseasonal donor set rather than mechanically suppressing it. Because it is the logical extension of the definition of seasonality provided in Section 3, it is a local approach to removing that power. Consequently, there is less of a trough created by the process and, as shown through the spectral MAE in concert with the dip measure, SSI preserves the spectral density of the irregular component in a manner demonstrably better than the benchmark.

In the time-domain, a notable trade-off emerges between the smoothness of the series and structural fidelity. X-13 consistently achieves a lower MAE due to its variance-reducing properties. However, this smoothness comes at the cost of accuracy. SSI consistently achieves a higher correlation with the true irregular component, suggesting that while X-13 produces a "cleaner" looking series, SSI more accurately tracks the true irregular component. Thus, for users prioritizing the recovery of a true generating process (or an approximation thereof) over visual smoothness, SSI offers a robust alternative that mitigates the spectral distortions inherent in filter-based adjustments.

Table 8. Comparative Efficacy of SSI vs. X-13 Adjustment

	Monthly Observations										Quarterly Observations									
	Time Domain				Freq Domain				Diag		Time Domain				Freq Domain				Diag	
	MAE		Corr		MAE		Dip		VR		MAE		Corr		MAE		Dip		VR	
Param	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13	SSI	X13
Panel A: Seasonal Unit Root ($\rho = 1$)																				
$\sigma_\kappa = 0.01$	1.10	**1.02**	**0.93**	0.83	**0.31**	0.50	**0.22**	0.38	**0.94**	0.74	1.09	**0.95**	**0.94**	0.71	**0.32**	0.88	**0.19**	0.39	**0.93**	0.48
$\sigma_\kappa = 0.02$	1.10	**1.02**	**0.90**	0.82	**0.37**	0.52	**0.21**	0.39	**0.97**	0.74	1.09	**0.95**	**0.93**	0.71	**0.32**	0.88	**0.18**	0.40	**0.93**	0.48
$\sigma_\kappa = 0.03$	1.11	**1.02**	**0.86**	0.81	**0.43**	0.54	**0.21**	0.39	**1.02**	0.76	1.09	**0.95**	**0.93**	0.71	**0.33**	0.88	**0.20**	0.40	**0.93**	0.48
$\sigma_\kappa = 0.04$	1.13	**1.03**	**0.82**	0.79	**0.49**	0.57	**0.23**	0.40	**1.09**	0.77	1.09	**0.95**	**0.92**	0.71	**0.34**	0.88	**0.19**	0.41	**0.93**	0.48
$\sigma_\kappa = 0.05$	1.15	**1.03**	**0.79**	0.77	**0.55**	0.59	**0.23**	0.38	**1.17**	0.80	1.09	**0.95**	**0.91**	0.71	**0.35**	0.88	**0.19**	0.41	**0.94**	0.48
Panel B: Stochastic Stationary ($\rho = 0.95$)																				
$\sigma_\kappa = 0.01$	1.10	**1.02**	**0.91**	0.83	**0.34**	0.50	**0.21**	0.38	**0.95**	0.74	1.09	**0.95**	**0.94**	0.72	**0.32**	0.88	**0.19**	0.39	**0.93**	0.48
$\sigma_\kappa = 0.02$	1.11	**1.02**	**0.90**	0.83	**0.35**	0.51	**0.21**	0.38	**0.96**	0.75	1.09	**0.95**	**0.94**	0.71	**0.32**	0.88	**0.19**	0.39	**0.93**	0.48
$\sigma_\kappa = 0.03$	1.11	**1.03**	**0.90**	0.82	**0.38**	0.53	**0.21**	0.38	**0.98**	0.76	1.09	**0.95**	**0.93**	0.71	**0.33**	0.87	**0.19**	0.39	**0.93**	0.49
$\sigma_\kappa = 0.04$	1.11	**1.03**	**0.88**	0.81	**0.41**	0.54	**0.21**	0.37	**1.00**	0.78	1.09	**0.95**	**0.93**	0.71	**0.33**	0.88	**0.20**	0.40	**0.93**	0.49
$\sigma_\kappa = 0.05$	1.12	**1.04**	**0.87**	0.80	**0.45**	0.57	**0.21**	0.37	**1.03**	0.80	1.09	**0.95**	**0.93**	0.71	**0.34**	0.88	**0.19**	0.40	**0.93**	0.49
Panel C: Deterministic Seasonality ($\rho = 0$)																				
$\gamma_0 = 0.10$	1.10	**1.02**	**0.94**	0.84	**0.30**	0.49	**0.21**	0.38	**0.94**	0.74	1.09	**0.95**	**0.94**	0.72	**0.32**	0.87	**0.19**	0.39	**0.93**	0.48
$\gamma_0 = 0.20$	1.10	**1.02**	**0.91**	0.84	**0.33**	0.50	**0.21**	0.39	**0.95**	0.74	1.09	**0.95**	**0.94**	0.72	**0.32**	0.88	**0.19**	0.39	**0.93**	0.48
$\gamma_0 = 0.30$	1.11	**1.02**	**0.89**	0.84	**0.37**	0.50	**0.21**	0.39	**0.97**	0.74	1.09	**0.95**	**0.94**	0.72	**0.32**	0.87	**0.20**	0.40	**0.93**	0.48
$\gamma_0 = 0.40$	1.11	**1.02**	**0.86**	0.84	**0.41**	0.49	**0.22**	0.38	**1.00**	0.74	1.09	**0.95**	**0.94**	0.71	**0.32**	0.87	**0.20**	0.40	**0.92**	0.48
$\gamma_0 = 0.50$	1.12	**1.02**	0.84	0.84	**0.45**	0.49	**0.21**	0.38	**1.04**	0.74	1.09	**0.95**	**0.94**	0.72	**0.32**	0.88	**0.19**	0.40	**0.93**	0.48

Note: Bold values indicate better performance (Minimum MAE, Maximum Correlation, Maximum Seasonal Dip Power, or Variance Ratio closest to 1).

6. Empirical Examples

To further illustrate the performance of our adjustment procedure relative to the current practice, we compare the results of using SSI versus X-13 on popular economic time series. We download both nonseasonally adjusted and seasonally adjusted published series from the Bureau of Labor Statistics (BLS) and Census Bureau and examine the characteristics of the published nonseasonally adjusted data, the published X-13 seasonally adjusted data, and the SSI-adjusted data.

6.1. Housing Starts

First, we consider new residential construction (aka "housing starts"), which are published monthly by the Census Bureau jointly with the Department of Housing and Urban Development (HUD). Housing starts measure the number of privately owned housing units on which construction begins and serve as a key indicator of housing market and broader economic conditions. They are closely watched by home builders, economists, policymakers, and investors alike because they reflect current housing demand and help forecast future construction employment and home sales.[20]

Figure 4 shows the resulting series from our adjustment procedure as well as some common diagnostic tests for seasonality. In the top-left panel, we plot the nonseasonally adjusted and seasonally adjusted data for the full period (1990–present). Two observations stand out. First, SSI results in a series that tracks the published, X-13 adjusted series very closely. We view this as a strength of SSI: it does not produce seasonally adjusted data that differ wildly from published aggregates, yet its transparency and known statistical properties make it a desirable alternative. Second, SSI results in a series that is slightly more volatile than X-13. We also view this property as a strength of SSI: it does not *over-smooth* the underlying data, which is a concern when using filter-based methods such as X-13.[21]

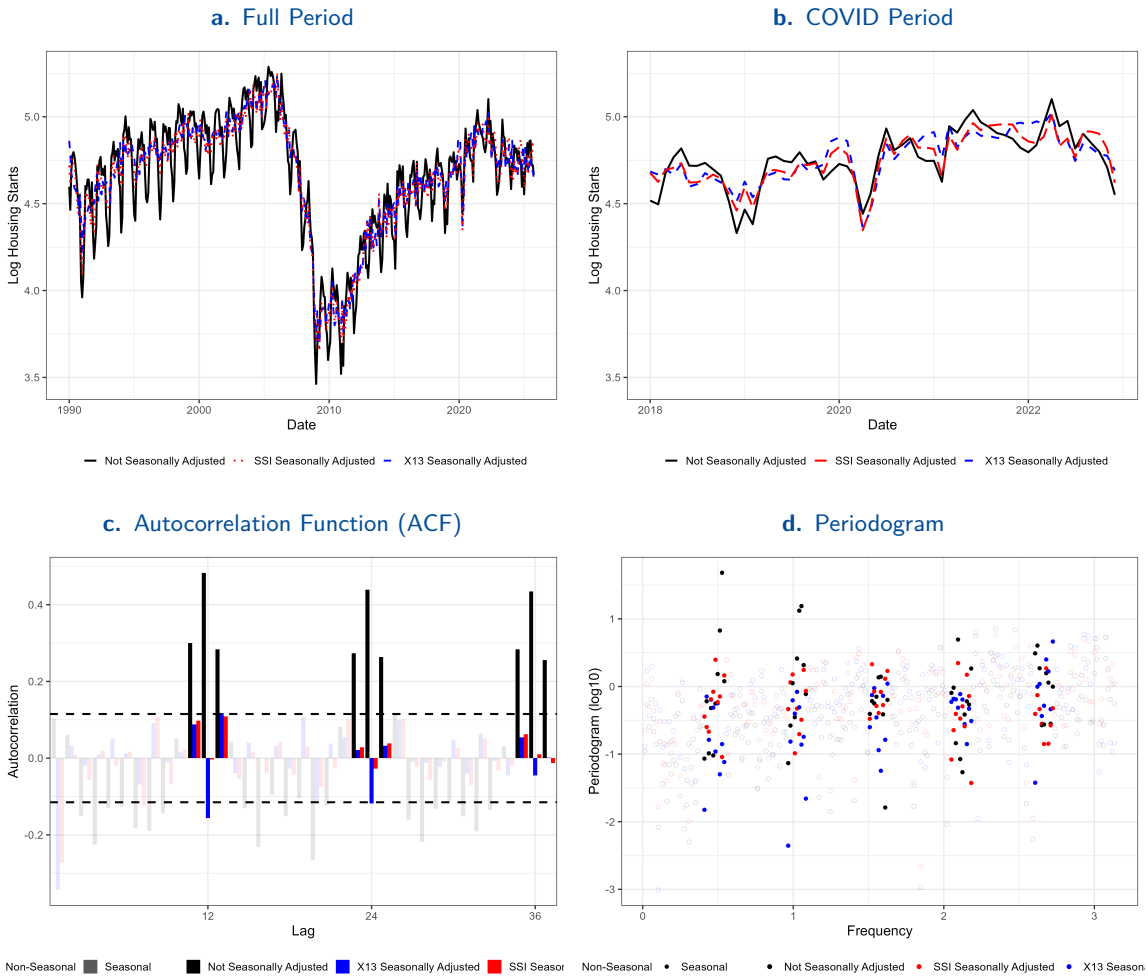
Reliably separating out seasonal patterns from the effects of underlying economic fundamentals is particularly important around business-cycle "turning points." Therefore, in the top-right panel, we zoom in on more recent data in order to examine the performance of SSI during the COVID period. We can see from the figure that housing starts declined significantly throughout 2020, but seasonal factors were clearly at play. In particular, in the first quarter of 2020, the beginning of the COVID–19 pandemic coincided with the end of the winter housing market, which is typically less active compared to the rest of the year. Therefore, some of the observed softness in housing during this period was likely due to "true" seasonal factors. We can see from the figure that both seasonally adjusted series pick up on this pattern and suggest that the underlying level of housing starts was considerably higher (about 0.10 log points) than the unadjusted data during the first few months of 2020. However, the figure shows that X-13 overstates the level of housing starts during these months relative to SSI. Moreover, since housing starts under both SSI and X-13 dropped sharply in the remainder of 2020, X-13 overstates the contraction in the housing market. This feature of seasonal adjustment—accurately removing seasonal noise to isolate changes in economic fundamentals—is especially important because it allows data users to draw reliable inferences from published data.

The bottom two panels of Figure 4 dig into the mechanics of how each seasonal adjustment procedure operates. The bottom-left panel shows the autocorrelation function (ACF), a popular tool in time series analysis, for the three series: unadjusted, X-13 adjusted, and SSI adjusted. In the figure, we highlight the "seasonal lags"—those that occur around 12, 24, and 36 months—in order to concentrate on the relevant behavior. We can see that the autocorrelation of the unadjusted series with its seasonal lags is

[20] Data are available at https://www.census.gov/econ/currentdata/ under the New Residential Construction heading. We download the Housing Units Started series.

[21] We elaborate more on this point below.

Figure 4. Housing Starts



Note: Data are from the New Residential Construction report, issued jointly by the U.S. Census Bureau and the Department of Housing and Urban Development (HUD).

quite high; it is outside of the significance bands at these intervals. In other words, this diagnostic tool deems the series to be strongly seasonal, making it a useful test case. Next, we can see that SSI results in values of the ACF that occur inside the significance bands at all seasonal lags—SSI is successful at dampening these seasonal autocorrelations. Last, we can also see that X-13 significantly reduces the magnitudes of seasonal autocorrelations in most instances, with the exception of the value at $t - 12$. In this instance, the ACF value under the X-13 adjusted series is actually significantly negative. Therefore, the X-13 procedure took a series that is strongly positively correlated with its own value 12 months ago and produced a series that is strongly *negatively* correlated with its own value 12 months ago.

To further understand the behavior of X-13 relative to SSI, we examine its periodogram in the bottom-right panel of Figure 4. We normalize the y-axis of the periodogram using a log base 10 scale and again highlight the ordinates that occur in the seasonal bins. Unsurprisingly, the periodogram of the

unadjusted series (black dots) has large spikes inside the seasonal bins.[22] For the periodogram of the SSI adjusted series, we can see that this is no longer the case: the values of the periodogram in the seasonal bins are akin to their nonseasonal counterparts. Again, X-13 displays similar behavior, but in some instances its periodogram ordinates fall below those of the other two series in the seasonal bins. Hence, X-13 negatively skews the height of the periodogram at seasonal frequencies, reminiscent of the over-smoothing behavior it displays in the time series.

6.2. Nonfarm Payrolls

Next, we repeat this exercise for nonfarm payroll employment published monthly by BLS in the Employment Situation report, which is closely followed by both policymakers and financial market participants. This series measures the number of paid jobs in the economy excluding farm workers, private household employees, unpaid family workers, and the self-employed in agriculture, covering roughly 80 percent of total employment in the U.S. economy.[23]

The top two panels of Figure 5 again show that SSI produces results very similar to X-13 in the time series. Adjusted nonfarm payrolls under SSI are less volatile than the unadjusted series, as with X-13, and SSI does not materially alter the behavior of payroll employment during the COVID period, which was marked by historically large month-to-month changes that complicated seasonal adjustment considerably. That SSI can be implemented mechanically, without judgmental adjustments, highlights a key advantage of our approach.

Next, the bottom left and right panels show the ACF and periodogram, respectively. Again, we can see that for both X-13 and SSI, the adjusted series no longer has large, positive ACF values at 12, 24, and 36 months. If anything, SSI introduces an alternating pattern in the seasonal autocorrelations that was not necessarily present in the unadjusted data, as it operates using trigonometric functions, though these values are not significant.[24] By this metric, SSI is therefore successful in removing seasonal time-dependence in a statistical sense. The same conclusion holds in the frequency domain, where the figure shows that seasonal periodogram ordinates under the SSI adjusted series are again in line with their nonseasonal counterparts. In particular, the spikes in the periodogram at seasonal frequencies in the unadjusted series (black dots) disappear under SSI (red dots). Here, we again see evidence of over-smoothing by X-13: periodogram dispersion in the seasonal bins (blue dots) is much smaller than in the nonseasonal bins. This suggests that X-13's filtering methods suppress seasonal variability more than is actually warranted.

We conclude from these empirical examples that SSI is well-suited for seasonal adjustment at scale. The selected series are among the most closely followed economic indicators produced by statistical agencies. Our findings demonstrate that SSI not only produces results qualitatively similar to current practice but also yields adjusted time series with superior statistical properties, as illustrated by the ACF and periodogram.

7. Conclusion

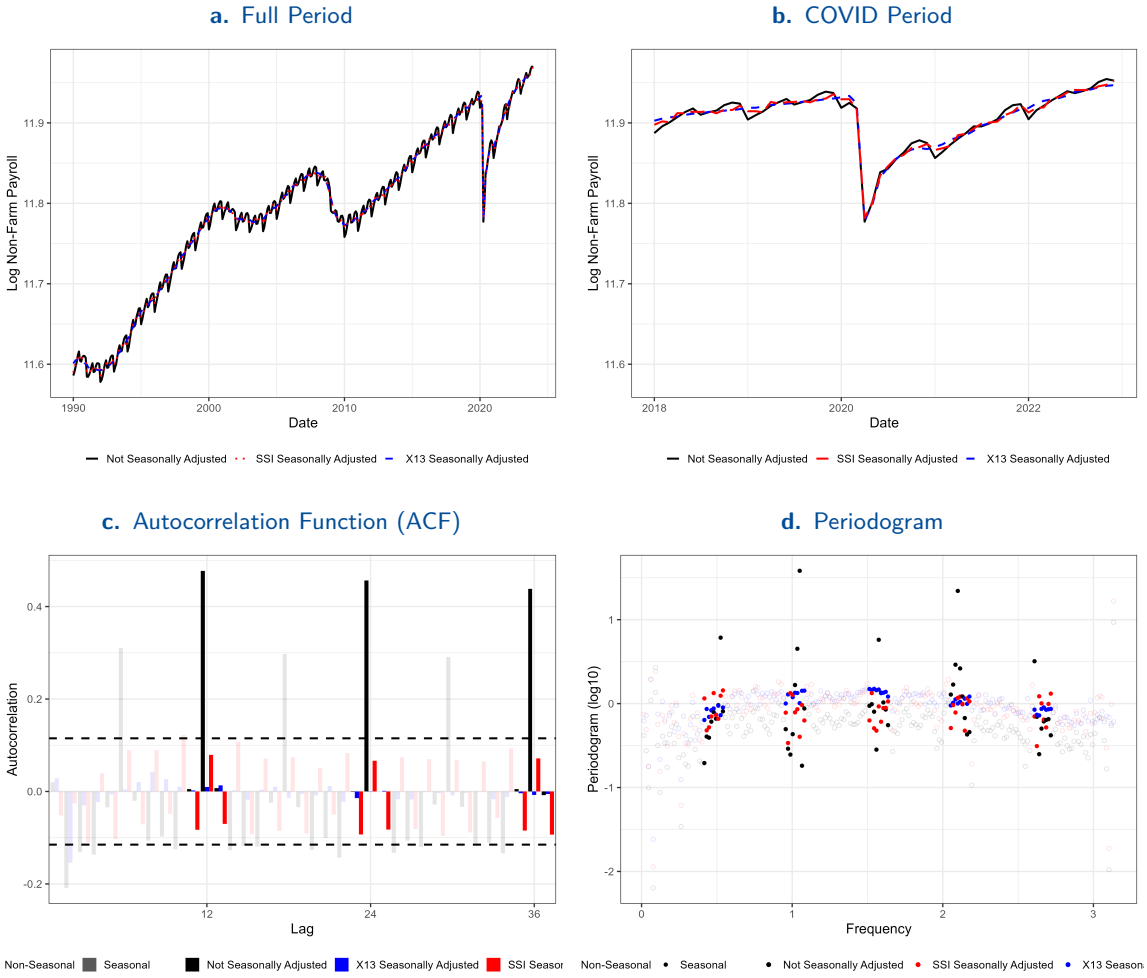
This paper argues that the central difficulty in seasonal adjustment is conceptual rather than computational. In the absence of a unified definition of seasonality, diagnostics and adjustment procedures cannot be

[22] Our test statistic that is based on this logic has a p-value of 0 to machine precision.

[23] Data are available at data.bls.gov under the Current Employment Statistics - CES heading. We download the series IDs: CES0000000001 and CEU0000000001.

[24] Observe that in lags 11, 12, and 13, the ACF values under SSI are negative, positive, and negative. A similar pattern is seen in lags 23-25 and 35-37.

Figure 5. Nonfarm Payroll Employment



Note: Data are from the Employment Situation report produced by the U.S. Bureau of Labor Statistics (BLS) using data from the Current Employment Statistics (CES) program.

coherently aligned, resulting in heterogeneous treatment across producers and making clear communication between data providers and users difficult. We address this gap by introducing a non-tautological, frequency-domain definition based on relative peak dominance that yields a well-defined population object. From this object, we derive a test statistic with a known and analytically tractable null distribution and an adjustment operator—Stochastic Spectral Imputation (SSI)—that follows directly from the same logical construct. The result is an internally consistent structure in which definition, detection, and adjustment are mutually reinforcing rather than loosely connected.

SSI operationalizes this framework by replacing ad hoc filtering with an adjustment operator that is the logical conjugate of the definition itself. Adjustment becomes a single, well-defined operation applied to a precisely stated object rather than the outcome of sequential diagnostic tuning. The resulting adjusted series closely track those produced by standard pipelines such as X-13, but the underlying structure is fundamentally different: seasonal adjustment becomes a well-posed statistical problem with a fixed

target, a known sampling distribution, and an adjustment rule dictated by first principles rather than accumulated convention. The benefit of this approach is increased transparency for academics, data providers, policymakers, and practitioners alike.

A. Algorithms

Algorithm 1 Generalized Stochastic Spectral Imputation (SSI)

Require: Observed series x_t, Seasonal params P, N, Bin param M, Options: Band, PhaseMode, q_{donor}

Compute DFT: $d_x \leftarrow \text{DFT}(x_t)$
Identify seasonal bins $\mathcal{M}_1$ and nonseasonal bins $\mathcal{M}_0$ based on $\Omega_G(P, N)$
Define initial sets: $\mathcal{J}_S^0 \leftarrow \{j : \omega_j \in \mathcal{M}_1\}$, $\mathcal{J}_D^0 \leftarrow \{j : \omega_j \in \mathcal{M}_0\}$
if Band = "between_seasonal_extremes" **then**
    Determine frequency range $[m_{\min}, m_{\max}]$ from $\mathcal{M}_1$
    Restrict $\mathcal{J}_S \leftarrow \{j \in \mathcal{J}_S^0 : \omega_j \in [B_{m_{\min}}, B_{m_{\max}}]\}$
    Restrict $\mathcal{J}_D^{\text{cand}} \leftarrow \{j \in \mathcal{J}_D^0 : \omega_j \in [B_{m_{\min}}, B_{m_{\max}}]\}$
else
    $\mathcal{J}_S \leftarrow \mathcal{J}_S^0$, $\mathcal{J}_D^{\text{cand}} \leftarrow \mathcal{J}_D^0$
end if
Filter Donors: $\tau \leftarrow \text{Quantile}(\{|d_x(\omega_k)|^2\}_{k \in \mathcal{J}_D^{\text{cand}}}, q_{\text{donor}})$
$\mathcal{J}_D \leftarrow \{k \in \mathcal{J}_D^{\text{cand}} : |d_x(\omega_k)|^2 \leq \tau\}$
Initialize adjusted coefficients: $d^* \leftarrow d_x$
for $j \in \mathcal{J}_S$ **do**
    Sample donor index $k \sim \text{Uniform}(\mathcal{J}_D)$
    Impute magnitude: $|d^*(\omega_j)| \leftarrow |d_x(\omega_k)|$
    if $\omega_j = \pi$ (Nyquist Frequency) **then**
        Determine sign s:
        if PhaseMode = "random" **then** $s \sim \text{Uniform}(\{-1, 1\})$
        else if PhaseMode = "keep" **then** $s \leftarrow \text{sgn}(\text{Re}(d_x(\omega_j)))$
        else if PhaseMode = "donor" **then** $s \leftarrow \text{sgn}(\text{Re}(d_x(\omega_k)))$
        end if
        $d^*(\omega_j) \leftarrow s \cdot |d^*(\omega_j)|$
    else
        Determine phase ϕ:
        if PhaseMode = "random" **then** $\phi \sim \text{Uniform}(-\pi, \pi)$
        else if PhaseMode = "keep" **then** $\phi \leftarrow \text{Arg}(d_x(\omega_j))$
        else if PhaseMode = "donor" **then** $\phi \leftarrow \text{Arg}(d_x(\omega_k))$
        end if
        $d^*(\omega_j) \leftarrow |d^*(\omega_j)|e^{i\phi}$
        Enforce symmetry: $d^*(\omega_{T-j}) \leftarrow \overline{d^*(\omega_j)}$
    end if
end for
return Adjusted series $x_t^* \leftarrow \text{Re}(\text{IDFT}(d^*))$

References

Bell, W. R. and Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, 2(4):98–127.

Bell, W. R., McDonald-Johnson, K. M., McElroy, T. S., Pang, O., Monsell, B. C., and Chen, B. (2022). Identifying seasonality. Interagency seasonal adjustment team, subgroup a, Bureau of Labor Statistics.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Box, G. E. P., Hillmer, S. C., and Tiao, G. C. (1978). Analysis and modeling of seasonal time series. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*, chapter V, pages 309–344. National Bureau of Economic Research, Cambridge, MA.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society. Series A (General)*, 143(3):321–337.

Busetti, F. and Harvey, A. (2003). Seasonality tests. *Journal of Business & Economic Statistics*, 21(3):420–436.

Busetti, F. and Taylor, A. M. R. (2003). Testing against stochastic trend and seasonality in the presence of unattended breaks and unit roots. *Journal of Econometrics*, 117(1):21–53.

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

Chang, Y. and Park, J. Y. (2002). On the asymptotics of adf tests for unit roots. *Econometric Reviews*, 21(4):431–447.

Chen, J. and Cornwall, G. (2021). The dark side of the moon: Searching for the other half of seasonality. Technical report, U.S. Bureau of Economic Analysis.

Cleveland, W. P. and Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the Census X-11 program. *Journal of the American Statistical Association*, 71(355):581–587.

Consolvo, V. N. and Lunsford, K. G. (2019). Residual seasonality in GDP growth remains after latest BEA improvements. Economic Commentary 2019-05, Federal Reserve Bank of Cleveland.

Dagum, E. B. (1980a). *The X-11-ARIMA seasonal adjustment method*. Statistics Canada.

Dagum, E. B. (1980b). The X-II-ARIMA seasonal adjustment method. Technical Paper Catalogue 12-564E, Statistics Canada, Ottawa.

Davis, R. A. and Mikosch, T. (1999). The maximum of the periodogram of a non-Gaussian sequence. *The Annals of Probability*, 27(1):522–536.

del Barrio Castro, T., Osborn, D. R., and Taylor, A. M. R. (2012). On augmented HEGY tests for seasonal unit roots. *Econometric Theory*, 28(5):1121–1143.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica: Journal of the Econometric Society*, pages 813–836.

European Commission and Eurostat (2015). *ESS guidelines on seasonal adjustment – 2015 edition*. Publications Office.

European Statistical System, Eurostat (2024). ESS Guidelines on Seasonal Adjustment, 2024 Edition. Technical Report KS-GQ-24-012-EN-N, European Statistical System, Eurostat, Luxembourg.

Eurostat (2018). *Handbook on Seasonal Adjustment, 2018 Edition*. Number KS-GQ-18-001-EN-N in Eurostat Manuals and Guidelines. Publications Office of the European Union, Luxembourg.

Falkner, H. D. (1924). The measurement of seasonal variation. *Journal of the American Statistical Association*, 19(146):167–179.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152.

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 125(796):54–59.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701.

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

Ghysels, E. and Osborn, D. R. (2001). *The econometric analysis of seasonal time series*. Cambridge University Press.

Gómez, V. and Maravall, A. (2001). Seasonal adjustment and signal extraction in economic time series. In Peña, D., Tiao, G. C., and Tsay, R. S., editors, *A Course in Time Series Analysis*. Wiley.

Granger, C. W. J. (1978). Seasonality: Causation, interpretation, and implications. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*, pages 33–56. National Bureau of Economic Research.

Hamlette, J. and Lunsford, K. G. (2026). Residual seasonality in some components of PCE inflation. *Federal Reserve Bank of Cleveland Economic Commentary*, (2026-04).

Hannan, E. J. (1961). Testing for a jump in the spectral function. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(2):394–404.

Hillmer, S. C. and Tiao, G. C. (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77(377):63–70.

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

Kallek, S. (1978). An overview of the objectives and framework of seasonal adjustment. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*, pages 3–32. National Bureau of Economic Research.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

Kuznets, S. (1933). *Seasonal Variations in Industry and Trade*. General series. National Bureau of Economic Research.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Lunsford, K. G. (2017). Lingering residual seasonality in GDP growth. Economic Commentary 2017-06, Federal Reserve Bank of Cleveland.

Lunsford, K. G. (2025). Residual seasonality in five measures of PCE inflation. Economic Commentary EC 2025-03, Federal Reserve Bank of Cleveland.

Lytras, D. P., Feldpausch, R. M., and Bell, W. R. (2007). Determining seasonality: A comparison of diagnostics from X-12-ARIMA. Research report, U.S. Census Bureau.

Maravall, A. (1993). Stochastic linear trends: Models and estimators. *Journal of Econometrics*, 56(1-2):5–37.

Maravall, A. (2011). Seasonality tests and automatic model identification in tramo-seats. *Bank of Spain: Madrid, Spain*.

Maravall, A. (2012). Update of seasonality tests and automatic model identification in tramo-seats. *Bank of Spain*.

McElroy, T. and Holan, S. (2009). A nonparametric test for residual seasonality. *Survey Methodology*, 35(1):67–83.

McElroy, T. and Holan, S. (2016). Using spectral peaks to detect seasonality. Research report, U.S. Census Bureau, Statistical Research Division, Washington, D.C.

McElroy, T. and Roy, A. (2017). Detection of seasonality in the frequency domain. *Statistics*, 1(2).

McElroy, T. and Roy, A. (2021). Testing for adequacy of seasonal adjustment in the frequency domain. *Journal of Statistical Planning and Inference*, 211:241–255.

McElroy, T. and Roy, A. (2022). A review of seasonal adjustment diagnostics. *International Statistical Review*, 90(2):259–284.

McElroy, T. S. (2021). A diagnostic for seasonality based upon polynomial roots of ARMA models. *Journal of Official Statistics*, 37(2):367–394.

Mills, J. A. and Prasad, K. (1992). A comparison of model selection criteria. *Econometric Reviews*, 11(2):201–234.

Nerlove, M. (1964). Spectral analysis of seasonal adjustment procedures. *Econometrica*.

Politis, D. N. and McElroy, T. S. (2019). *Time series: A first course with bootstrap starter*. Chapman and Hall/CRC.

Priestley, M. (1981). Spectral analysis and time series.

Rodrigues, P. M. M. and Taylor, A. M. R. (2007). Efficient tests of the seasonal unit root hypothesis. *Journal of Econometrics*, 141(2):548–573.

Rudebusch, G. D., Wilson, D., and Mahedy, T. (2015). The puzzle of weak first-quarter GDP growth. *FRBSF Economic Letter*, (2015-16).

Smith, R. J. and Taylor, A. M. R. (1999). Likelihood ratio tests for seasonal unit roots. *Journal of Time Series Analysis*, 20(4):453–476.

Soukup, R. J. and Findley, D. F. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. Washington, DC.

Taylor, A. M. R. (2005). Variance ratio tests of the seasonal unit root hypothesis. *Journal of Econometrics*, 124(1):33–54.

United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, and World Bank (2025). System of national accounts 2025. Technical report, United Nations.

Webel, K. and Ollech, D. (2018). An overall seasonality test based on recursive feature elimination in conditional random forests. In *Proceedings of the 5th international conference on time series and forecasting*, pages 20–31.

Webel, K. and Ollech, D. (2025). A novel tree-based combined test for seasonality. *Data Science in Science*, 4(1):2517006.

Wright, J. H. (2013). Unseasonal seasonals? *Brookings Papers on Economic Activity*, Fall 2013:65–110.

Zellner, A., Tukey, J. W., BarOn, R. R. V., and Kaufman, H. M. (1978). Retrospect and prospect. In Zellner, A., editor, *Seasonal Analysis of Economic Time Series*, pages 451–458. National Bureau of Economic Research.