Identifying Seasonality[1] January 2022

Interagency Seasonal Adjustment Team, Subgroup A

William R. Bell, Kathleen M. McDonald-Johnson, Tucker S. McElroy, and Osbert Pang
U.S. Census Bureau

Brian C. Monsell
U.S. Bureau of Labor Statistics

Baoline Chen
U.S. Bureau of Economic Analysis

January 21, 2022

---

[1] Disclaimer: Any views expressed here are those of the authors and not necessarily those of the U.S. Census Bureau, the U.S. Bureau of Labor Statistics, or the U.S. Bureau of Economic Analysis. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance applied. (Approval ID: CBDRB-FY20-ESMD001-005)

Contents

1 Introduction

Seasonal adjustment, the process of removing seasonal patterns from time series data, generally involves three steps. The analyst should:

1. Determine if the time series is seasonal. If not, stop.
2. If the time series is seasonal, perform seasonal adjustment.
3. Determine if the seasonal adjustment performed is adequate. If so, stop. If not, return to Step 2, modify the seasonal adjustment procedure (for example, choose different program options), and continue this process until obtaining an adequate seasonal adjustment.

If, after several modifications, no seasonal adjustment is acceptable, then the analyst must decide either (a) not to seasonally adjust the series or (b) to accept one of the seasonal adjustments (allowing something previously regarded as inadequate).

Various statistical techniques – graphs, hypothesis tests, and other diagnostics – aid in deciding how to proceed through Steps 1 and 3. An important assessment at Step 3 involves determining whether seasonality can be found in the seasonally adjusted series, a phenomenon known as “residual seasonality.” Since the goal of seasonal adjustment is to remove seasonal patterns from the time series, it is natural to examine an adjusted series for residual seasonality.

This paper documents seasonality diagnostics available in the X-13ARIMA-SEATS seasonal adjustment software and illustrates their application to several real time series. The illustrations include both testing for seasonality in an unadjusted time series (known as pretesting) and checking for residual seasonality in a seasonally adjusted time series (known as posttesting). An appendix provides technical details on the seasonality diagnostics available in X-13ARIMA-SEATS. Note that X-13ARIMA-SEATS contains other diagnostics not related to checking for seasonality, such as diagnostics related to seasonal adjustment revisions; we do not discuss these other diagnostics here. See the X-13ARIMA-SEATS reference manual (U.S. Census Bureau 2020) for more information on software, terminology, and specific options.

The diagnostics available in X-13ARIMA-SEATS for detecting seasonality in an original, unadjusted series (Step 1) are essentially the same diagnostics available for detecting residual seasonality in a seasonally adjusted series (Step 3), as well as in the model residuals, or the estimated irregular. The only difference might be a modification to compute the diagnostics over a reduced span of recent series values when checking for residual seasonality. The use of the same diagnostics in these different contexts is common practice, but that does not mean it is an optimal approach. In fact, determining whether an unadjusted series is seasonal is really a different problem from determining whether a seasonally adjusted series contains residual seasonality. This is because, even apart from seasonality, the statistical properties of unadjusted series differ from those of seasonally adjusted series. One implication of this is that one would ideally like to have different diagnostics tailored to the specific problems of detecting seasonality in original unadjusted series versus detecting residual seasonality in seasonally adjusted series and other estimated series components, or at least have different interpretations for a given diagnostic depending on the context in which it is used. Unfortunately, knowledge to provide such refined tailoring of seasonality diagnostics is currently lacking, so analysts tend to use the same diagnostics in both contexts. Developing improved diagnostics, as well as improving the understanding of the differences in the two cases, is an active area of research. The remainder of the document is organized as follows. Section 2 provides a brief discussion of the direct and indirect seasonal adjustment methods for aggregate series; this topic has important implications for the testing of seasonality and residual seasonality in the aggregate series. Section 3 describes some common software for seasonal adjustment, and Section 4 lists diagnostics currently available in X-13ARIMA-SEATS for determining whether a series exhibits seasonality. Section 5 provides examples of time series and their associated diagnostics. The Appendix provides a general overview of seasonality in time series as well as technical details about the diagnostics.

Note that this document provides examples of adjustments and diagnostics from publicly available time series solely to illustrate what is available from the software; consequently, these adjustments might not match published estimates.

2 Direct and Indirect Seasonal Adjustment of Aggregate Series

Once published, analysts often view individual time series together with other related series, and most series that are adjusted at statistical agencies are components of some greater aggregate. Although it is easiest to seasonally adjust each single series in isolation, the ramifications of each adjustment's adequacy extend far beyond its own analysis.

Consider a group of related time series to be seasonally adjusted, where it is of interest to also produce a seasonally adjusted aggregate series. In this context, seasonally adjusting each single series, as well as the aggregate series, is called direct seasonal adjustment (of the aggregate series). Aggregating the seasonally adjusted component series to form the aggregate seasonal adjustment is called indirect seasonal adjustment. Often the aggregation is a sum of component series, although other combinations are possible, such as aggregating from monthly to quarterly frequency. Other examples include balance-of-trade series (exports minus imports) and ratio series, such as inventories compared to sales.

It can happen that many components exhibit no detectable seasonality according to pretesting, and yet the aggregation of them is seasonal. See McElroy (2018) for details. Hence, pretesting of the components is not adequate for determining the presence of seasonality in indirect adjustments of aggregate series. When the indirect adjustment of an aggregate series has residual seasonality, an alternate adjustment might be necessary. One obvious potential alternative is direct adjustment of the aggregate series. Another option is to review the results for the component time series and try to improve the model (particularly for SEATS seasonal adjustments), or, in the case of X-11 seasonal adjustments, check whether using different seasonal filters would help. Unfortunately, if the diagnostics for the component time series indicate no evidence of model failure or of residual seasonality, it is unclear what changes might improve the indirect adjustment of the aggregate. Probably the best series to start with are the largest in value or those with marginal signs of residual seasonality.

A related phenomenon can occur with aggregation of a monthly series to yield a quarterly series. While seasonal adjustment of a monthly series provides an indirect adjustment of the corresponding quarterly aggregate series, it is possible that the indirect adjustment manifests residual seasonality in the quarterly aggregate series, even when the original monthly seasonal adjustment appears adequate. This phenomenon is discussed in Moulton and Cowan (2016) and McElroy et al. (2019).

3 Seasonal Adjustment Software

3.1 X-11 Method as Implemented in X-13ARIMA-SEATS

X-11 is a seasonal adjustment procedure that the U.S. Census Bureau developed in the 1950s and 1960s. Statistics Canada enhanced the method with the addition of ARIMA modeling for forecast extension to reduce seasonal adjustment revisions, effectively improving on the original asymmetric X-11 moving averages used to estimate the seasonal component for the most recent time points. The X-11 procedure separates a time series into a trend-cycle component, a seasonal component, and an irregular component by iteratively filtering the original (or transformed) time series, using moving averages. Users (or automatic provisions in the software) choose the moving average filters from a set of fixed (precoded) options. X-13ARIMA-SEATS has an X11 specification that implements the X-11 seasonal adjustment method. Details on the X-11 procedure are available in Ladiray and Quenneville (2001) and in the X-13ARIMA-SEATS reference manual (U.S. Census Bureau 2020).

3.2 SEATS Method as Implemented in X-13ARIMA-SEATS

SEATS (**S**ignal **E**xtraction in **A**RIMA **T**ime **S**eries) is the seasonal adjustment part of the program TRAMO-SEATS that Agustín Maravall and Victor Gómez developed while at the Bank of Spain. TRAMO (**T**ime series **R**egression with **A**RIMA noise, **M**issing observations, and **O**utliers) fits regARIMA models (short for regression models with ARIMA errors, as described in Findley et al. (1998)), and SEATS uses the fitted model to estimate the components for seasonal adjustment. Specifically, SEATS uses the canonical ARIMA model-based approach of Hillmer and Tiao (1982) and Burman (1980) to decompose a time series into a trend-cycle component, a seasonal component, and an irregular component, each of which follows its own underlying ARIMA model.

X-13ARIMA-SEATS has an automatic regARIMA modeling procedure that uses modified versions of the TRAMO algorithms, and a SEATS specification that implements the SEATS seasonal adjustment method.

4 Diagnostics for Identifying Seasonality

Before examining seasonality diagnostics, users should begin their analysis by graphing the original time series. Graphing the series across consecutive time points, as well as year over year, will help in determining whether the series has a seasonal pattern (see Section 5 for examples). In addition, graphs can illuminate additional patterns of series behavior as well as unusual points or subspans within a series. For any analysis, it helps to know as much as possible about the series. For some series, large outliers or other effects might obscure the series' patterns. If that is the case, one should graph the outlier-adjusted or prior-adjusted series as well.

It is also worth examining the autocorrelation function (ACF) of the original and first differenced series to generally assess the dependence over time in the data, including seasonal dependence. For a seasonal monthly series, one would expect large, positive ACF values at lag 12, and multiples of 12 (24, 36, etc.). For a seasonal quarterly series, one would expect large, positive ACF values at the seasonal lags 4, 8, 12, etc. If the seasonal lag autocorrelations are small and statistically insignificant, particularly the lag-12 autocorrelation for monthly data or lag-4 autocorrelation for quarterly data, this reflects lack of evidence of seasonality in the series. The analysis can be pursued further with the diagnostics discussed below.

Appendix A provides some cautions needed when examining ACFs for evidence of seasonality. First, series with strong nonseasonal dependence can have substantial autocorrelation at lag 12 (monthly) or lag 4 (quarterly) that is not reflective of seasonality. In such cases, autocorrelations around the seasonal lag will be as large as or larger than the seasonal lag autocorrelation. Also, mild to moderate values of seasonal lag autocorrelations will not produce strong seasonal patterns in data, which is what we ordinarily think of as "seasonality." Whether it may be worth seasonally adjusting a series with just mild or moderate seasonal dependence is something of an open question that we do not address here.

Below, we provide a description for three seasonality diagnostics that are available in X-13ARIMA-SEATS; see Lytras (2015) for a more detailed discussion with a focus upon quarterly series. The discussion presumes a knowledge of background material that is provided in Appendices A and B. We remark that other diagnostics are available in X-13ARIMA-SEATS software that are not discussed here. This includes two diagnostics pertaining to seasonality – the stable F test from Table D8 of X11 and the M7 quality control statistic (Ladiray and Quenneville 2001). Both the D8 F test and the M7 diagnostic are flawed in various ways (see Findley and Monsell (1986) for a discussion of the M1-M11 diagnostics, and Lytras et al. (2007) for a critique of the D8 F test).

4.1 Regression Model-Based F Test

Brief Description

The regression model-based F test checks for evidence of a stable seasonal pattern in the original series. In our consideration of diagnostics, it replaces the stable F test from Table D8 of X-11, which erroneously assumes independence over time of the estimated seasonal-irregular component (also called the “SI ratio,” “SI difference,” or detrended series).

Although the regression model-based F test has been shown (Lytras et al. 2007) to be reliable and accurate when testing for whether the original (unadjusted) series is seasonal, research into applying this diagnostic to test for residual seasonality in a seasonally adjusted series (treating the seasonally adjusted data as the observed time series) has not been as promising (Findley et al. 2017). Investigations into different methods of applying the test continue, but we currently do not advise using it as a residual seasonality test. Analysts who nevertheless want to use the diagnostic for posttest purposes should keep in mind that direct seasonal adjustment annihilates any stable seasonal pattern that is present over the full adjustment span, so any testing for residual seasonality in directly adjusted series should include only a subspan of the data.

To apply the F test requires building a model with fixed seasonal effect regressors. Because regARIMA models built for use in seasonal adjustment (whether for ARIMA model-based adjustment from SEATS or for use in forecast extension, trading day adjustment, etc. with X-11 adjustment) typically are not of this form, the F test usually requires at least one additional run of the program. The model can be built by examining ACFs for the original and differenced series taking out fixed seasonal effects (and any other known regression effects, such as perhaps trading day) to determine a nonseasonal ARIMA model for the regression residuals, or possibly also allowing for a restricted (e.g., first order) seasonal AR or MA part. Alternatively, automatic model selection can be used with the same model restrictions.

If one already has a seasonal ARIMA model for a series and wants to apply the F test, one can often use a related regARIMA model with fixed seasonal effects, at least as a starting point. Often seasonal ARIMA models have a seasonal (0 1 1) part. For such models, Bell (1987) noted that when the seasonal MA parameter is 1.0, the seasonal difference and seasonal MA(1) operator can be cancelled, yielding an equivalent model if one also includes fixed seasonal effects and a trend constant.[2] Thus, if one has a regARIMA model with a seasonal (0 1 1) MA term, a reasonable alternative is a model with the same nonseasonal ARIMA part, plus fixed seasonal regressors and a trend constant (and retaining any regressors in the original model, such as for trading day effects). The model might be augmented with a seasonal AR(1) or MA(1) term, especially if the seasonal MA parameter estimate from the original model is not so close to 1.0, making the cancellation of the seasonal difference and seasonal MA operator not exact. The usual model diagnostics (residual plots, residual ACF, etc.) should be examined for any evidence of substantial inadequacies, modifying the model if this seems warranted. By the same token, one could build a regARIMA model with fixed seasonal effects as a starting point for building a model with a seasonal ARIMA part, keeping the regressors and nonseasonal part of the original model and adding a seasonal (0 1 1) part.

Unlike the stable F test from Table D8 of X-11, the regression model-based F test does not depend on the seasonal adjustment settings, but results vary depending on the other regARIMA model parameters and the span chosen for testing.

The null hypothesis for the regression F test is that the series does not have fixed seasonal effects, which is also called stable seasonality. To be more inclusive of series that might be seasonal (pretest), or to be less tolerant of seasonality in the final seasonally adjusted series (posttest), one may choose a relatively high significance level (*p*-value) such as 0.1. If one desires to be more careful to adjust only series with very strong evidence of seasonality – or to identify only clear cases of the presence of residual seasonality – one may choose a low significance level of 0.01. The significance level might depend on: (1) the purpose of the test (particularly whether it is pretest or posttest), (2) the priority of the adjustment, and (3) the degree to which the statistical agency wants to avoid publishing any adjustment that fails quality or stability diagnostics.

Table 1 shows an example of the F test as it appears in the X-13ARIMA-SEATS output. The *p*-value of the F test is 0.00, so the null hypothesis that all months have the same mean effect can be rejected. This result yields strong evidence that the series has stable seasonality.

This result is from a test of the full span of the original series (not yet seasonally adjusted): monthly U.S. Retail Sales of Office Supply, Stationery, and Gift Stores,

---

[2] The trend constant is needed because the seasonal difference, $(1 - B^s) = (1 - B)(1 + B + \ldots + B^{s-1})$, includes the nonseasonal difference $1 - B$, which will annihilate a constant term. If the original model already has a trend constant, then with cancellation of the seasonal difference the trend constant should be augmented to a linear trend (i.e., a ramp effect over the entire series). Trend constants in models with both a seasonal and nonseasonal difference (implying $(1 - B)^2$) are generally to be avoided and so should be rare. But if the original seasonal ARIMA model had only a seasonal difference, the model might well have a trend constant, in which case this consideration becomes relevant.

from 2002 through 2017 (Source: Monthly Retail Trade and Food Services, U.S. Census Bureau).[3] **Figure 1** and **Figure 2** show the original series, which has an apparent seasonal pattern; in particular, the series has consistent increases in activity from July to August and from November to December, as well as a drop in activity in February, April, and June. In year-over-year graphs such as **Figure 2**, month 13 (or quarter 5 in subsequent graphs) is the first month (or quarter) of the next year, to fully portray each period's change. Although the visual information from graphs is important, it does not provide a formal or rigorous comparison of series values nor is it a statistical test for differences between the values.

Summary

- The null hypothesis of the test is that the series does not have stable seasonality.
- This test typically requires at least one additional run of X-13ARIMA-SEATS to develop the regARIMA model with fixed seasonal effects.

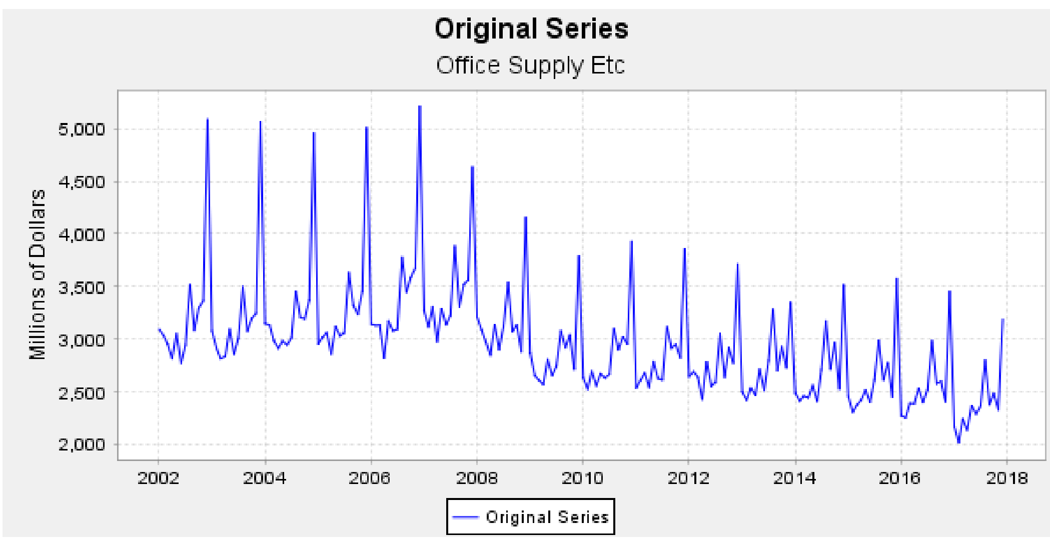

Figure 1: The original series (not seasonally adjusted) of monthly U.S. retail sales of office supply, stationery, and gift stores, in millions of dollars, 2002 – 2017. **Source:** Monthly Retail Trade and Food **Services**, U.S. Census Bureau (census.gov/retail/)

[3] The values are estimates from surveys and are subject to sampling and nonsampling error. They are not adjusted for price changes. For more information about the data collection and estimation, access the description online (census.gov/retail/how_surveys_are_collected.html).

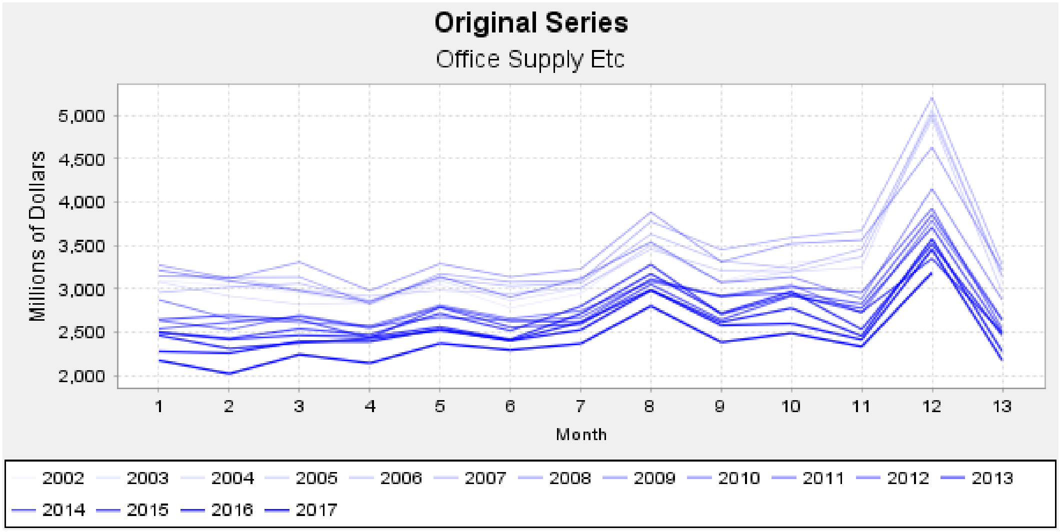

Figure 2: The original series (not seasonally adjusted) of monthly U.S. retail sales of office supply, stationery, and gift stores, year over year. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Table 1: F test of the seasonal regressors, from the X-13ARIMA-SEATS main output file of retail sales of office supply, stationery, and gift stores, with the associated degrees of freedom (DF) and *p*-value. The *p*-value indicates that at the 0.01 (or smaller) level we would reject the null hypothesis that all months have the same effect. This result provides strong evidence that the series is seasonal. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Regressor	DF	F Statistic	*P*-Value
Seasonal	11, 170	369.04	0.00

4.2 QS

Brief Description

Agustín Maravall originally developed the QS diagnostic for the TRAMO-SEATS program (Gómez and Maravall 1997). The QS diagnostic is a function of the first two seasonal lag autocorrelations – those for lags 12 and 24 for monthly data, and for lags 4 and 8 for quarterly data – and it attempts to test the null hypothesis that these first two seasonal autocorrelations are zero. Appendix B gives the formula for QS. The rationale behind QS is that a series with seasonality, or residual seasonality, should exhibit substantial positive autocorrelation at these lags. Seasonal autocorrelation may extend beyond lags 24 or 8, but such higher lags are not used by QS.

If the autocorrelation at the first seasonal lag is zero or negative, then QS is set to 0 and the *p*-value is set to 1. It is worth noting that if a series is adequately adjusted, then we would expect negative seasonal autocorrelation in the (differenced) seasonally adjusted series and the irregular. Thus, values of 0 for

QS are not unusual. If QS is not zero, the *p*-value associated with the QS is obtained on an assumption that, if the null hypothesis that the first two seasonal lag autocorrelations are zero is true, the asymptotic distribution of QS in large samples is approximately that of a chi square with two degrees of freedom. Appendix B notes some issues with this assumption.

The null hypothesis for the test is that the series is not seasonal. Maravall (2012) recommends testing at the 1% level, i.e., seeking a *p*-value less than or equal to 0.01. A significant test result, leading to rejecting the null hypothesis, is taken as evidence that the series has seasonality – or in the case of the posttesting, evidence of residual seasonality.

The QS diagnostic has limitations for pretesting. (Findley et al. 2017.) The main problem is that it may detect minor amounts of seasonal autocorrelation in a series, especially with a long series, and minor, though nonzero, amounts of seasonal autocorrelation do not reflect what we ordinarily think of as seasonality. Such indications are more readily interpreted in posttesting, for evidence of residual seasonality. If, in pretesting, QS is significant while the model-based F test is not, one should examine the entire autocorrelation function to check whether a significant QS statistic is triggered by minor seasonal autocorrelation, and whether this may reflect nonseasonal dependence. At the other extreme, strong seasonal dependence, including strong fixed seasonal effects, should trigger a significant QS statistic with high probability.

The QS statistic is available for the

- (Differenced) original series (pretest)
- (Differenced) original series adjusted for extreme values (Table B1 from the X-11 method of seasonal adjustment) (pretest)
- RegARIMA model residuals (neither a pretest nor a posttest, as we have defined them)
- (Differenced) seasonally adjusted series (posttest)
- (Differenced) seasonally adjusted series adjusted for extreme values and outliers in the model (posttest)
- Irregular component (posttest)
- Irregular component adjusted for extreme values (posttest)

In addition, the diagnostic is available for the indirect seasonally adjusted series and the indirect seasonally adjusted series adjusted for extreme values.

QS is also available for a shortened span of each of these seven series. For testing for seasonal autocorrelation in the original series, one would generally use the full adjustment span. Testing over the shortened span may be reasonable for the seasonally adjusted series and irregular (testing for residual seasonality), especially if there are concerns about possible changes in the manifestation of seasonality.

X-13ARIMA-SEATS uses the same subspan for the shortened series tests as selected for the spectral diagnostic by default, i.e., the most recent 96 observations (or eight years of a monthly series).

A difference between results of the full and shortened series can indicate potential problems. For instance, an adjustment that does not capture a change in seasonal pattern could be less than adequate, especially for the ends of the series, and interest typically is highest for the most recent part of the series.

For any seasonal adjustment, the results of the various QS diagnostics can differ. The ideal posttest result is that all the QS statistics for residual seasonality are nonsignificant, but at a minimum, the QS diagnostic for the full seasonally adjusted series should be nonsignificant.

For monthly series, X-13ARIMA-SEATS can provide additional QS diagnostic information for the *quarterly* versions of the series. This additional diagnostic is valuable for review of monthly time series that are aggregated (or “collapsed”) and included in quarterly statistics; for example, monthly economic indicators are inputs to quarterly gross domestic product estimates. Reviewers of monthly series who see residual seasonality in the quarterly QS diagnostic will want to modify the adjustment settings of the monthly series to improve the results for the end users of the seasonally adjusted estimates. The default series type is flow, where the quarterly version is the sum of each calendar quarter’s three monthly values (e.g., the first quarter’s value is the sum of January, February, and March values). One should set the series type to “stock” if the series is a stock series (such as an inventory series), and the program in this case will test the diagnostic for the quarterly series comprised of the third month’s value of each calendar quarter (e.g., the first quarter’s value is the March value).

Table 2 and **Table 3** show the QS diagnostics (for the monthly time series and for the quarterly version of the series, respectively) as they appear in the X-13ARIMA-SEATS output, for Retail Sales of Office Supply, Stationery, and Gift Stores, a flow series, with a span of 2002 through 2017 (data not adjusted for price changes). Because it is a flow series, the software based the quarterly diagnostics on the calendar-quarter sums. Note that in the main output file, these tables appear in separate sections.

For this series, all the pretest *p*-values for the QS statistics for the original series and original series adjusted for extremes, both for the full span and shortened span, are less than 0.01, so the null hypothesis of no seasonal autocorrelation can be rejected. In other words, the original series shows evidence of seasonal dependence. The tests of the quarterly aggregate of this series also indicate seasonal dependence.

For the regARIMA model residuals, seasonally adjusted series, and irregular component, the QS statistics are small. *P*-values that are 1 indicate either cases where the first seasonal autocorrelation was negative, so the QS statistic was set to zero, or cases where the seasonal autocorrelation was so small that the QS value rounded to zero for the digits shown. The posttest *p*-values for this seasonal adjustment are thus all close to 1 (or have been set equal to 1). Although this diagnostic by itself is not sufficient to judge the overall quality of the seasonal adjustment, this lack of evident residual seasonality is compatible with the seasonal adjustment being adequate.

Summary

- The null hypothesis of the test is that the series does not have seasonal autocorrelation (at the first two seasonal lags).
- Maravall, the developer of the diagnostic, used a significance level of 1%. A *p*-value less than or equal to 0.01 thus leads to rejection of the null hypothesis, i.e., it provides evidence that the original series has seasonal autocorrelation (pretest) or that the adjusted series has residual seasonality (posttest).

Table 2: QS diagnostics from the X-13ARIMA-SEATS main output file, pretest and posttest (as well as regARIMA model residuals), full series and shortened series (starting in January of 2010), for the **monthly series** of retail sales of office supply, stationery, and gift stores. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)](census.gov/retail/).

Series	Span	QS	*P*-Value
Original Series	Full Series	307.31	0.0000
Original Series (extreme value adjusted)	Full Series	322.95	0.0000
Residuals	Full Series	0.00	1.0000
Seasonally Adjusted Series	Full Series	0.00	0.9995
Seasonally Adjusted Series (extreme value adjusted)	Full Series	0.00	1.0000
Irregular Series	Full Series	0.00	1.0000
Irregular Series (extreme value adjusted)	Full Series	0.00	1.0000
Original Series	Subspan	132.56	0.0004
Original Series (extreme value adjusted)	Subspan	149.70	0.0000
Residuals	Subspan	0.00	1.0000
Seasonally Adjusted Series	Subspan	0.17	0.9170
Seasonally Adjusted Series (extreme value adjusted)	Subspan	0.00	1.0000
Irregular Series	Subspan	0.00	1.0000
Irregular Series (extreme value adjusted)	Subspan	0.00	1.0000

Table 3: QS diagnostics from the X-13ARIMA-SEATS main output file, pretest and posttest (as well as regARIMA model residuals), full series and shortened series (starting in January of 2010), for the **quarterly (sums of the monthly) series**, for retail sales of office supply, stationery, and gift stores. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Series	Span	QS	*P*-Value
Original Series	Full Series	99.78	0.0000
Original Series (extreme value adjusted)	Full Series	103.47	0.0000
Seasonally Adjusted Series	Full Series	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Full Series	0.00	1.0000
Original Series	Subspan	42.33	0.0000
Original Series (extreme value adjusted)	Subspan	45.44	0.0000
Seasonally Adjusted Series	Subspan	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Subspan	0.00	1.0000

4.3 Peaks at Seasonal Frequencies in Spectral Plots

Brief Description

Currently available in X-13ARIMA-SEATS for monthly series only (not for quarterly series), spectral plots aid in identifying evidence of seasonality in the

- Original series (pretest), note that the default input series is the prior-adjusted original series to remove effects from outliers, trading day, and moving holidays that might interfere with identification of seasonal effects
- RegARIMA model residuals (not a pretest or posttest as we have defined them)
- Modified seasonally adjusted series[4] (posttest)
- Modified irregular (posttest)
- (When applicable) indirectly seasonally adjusted (posttest)
- (When applicable) modified irregular component of the indirect seasonal adjustment (posttest)

[4] The tests are applied to the differenced modified seasonally adjusted series, though for brevity the output file label does not include the word *modified*.

For computing the spectrum reliably, at least 96 months of data (i.e., eight years) should be used. In fact, X-13ARIMA-SEATS produces the spectral plots using a default span of the last eight years of the series, as is the case with most of the residual seasonality diagnostics from the program, although software users can change the span.

Spectral plots highlight the computed spectral estimates at the seasonal frequencies 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12. These frequencies correspond to seasonal effects recurring several times per year, according to the numerator. For instance, a feature that repeats once over 12 months (one year) corresponds to frequency 1/12, whereas a feature that repeats four times over 12 months (quarterly) corresponds to frequency 4/12. Each seasonal frequency corresponds to the number of cycles per year, whether that is once (1/12), twice (2/12), or as high as six (6/12); more details are given in Appendix A. Large, or *visually significant,* peaks at the seasonal frequencies indicate evidence of seasonality. The current thresholds for visual significance of a peak are a height that is

- Six *stars* above the taller of the two nearest-neighbor frequencies on the plot
- Above the median height of all the plotted frequencies

The unit of measure, the *star*, is 1/52 of the range of the spectral values, based on the ASCII text representation of the graph. Some graphing programs, such as Win X-13 and X-13-Graph (Lytras 2017, 2013), show indicators (S for seasonal and T for trading day) in the plot to make it clear that a peak meets the thresholds for visual significance. Some output shows the frequencies as S1 for 1/12, S2 for 2/12, etc. This notion of significance does not consider the statistical variability in spectral estimates; see McElroy and Roy (2021) for discussion.

The spectral plots also indicate the trading day frequencies, 0.348 and 0.432 cycles per month for monthly series[5], for which the spectrum estimates are plotted with a column of Ts. As noted in Section 6.1 of the X-13-ARIMA-SEATS documentation (U.S. Census Bureau 2020), "In the case of trading day peaks, a peak (especially one at the lower of the two trading day frequencies) shows the need for trading day estimation if this was not done, and otherwise shows that the trading day regression model used is inadequate for the time interval used for spectrum estimation." Since this report focuses on procedures for detecting seasonality, we shall not focus on detection of trading day effects here. For more information on using the spectral diagnostic to detect trading day effects, see the X-13-ARIMA-SEATS documentation, Section 2.1 of Findley et al. (1998), and Soukup and Findley (1999).

---

[5] Spectrum estimates are plotted at the 61 frequencies k/120 for k = 1, …, 60 except that, for the two k/120 values closest to the two trading day frequencies, the spectrum estimates shown are those for the two trading day frequencies.

X-13ARIMA-SEATS provides a warning of a visually significant peak at any of the first five seasonal frequencies, but in practice, if no other seasonal frequency has a visually significant peak, we typically discount peaks at 5/12 because it is not a natural division of the months where we would expect to see activity (there is no known economic explanation for such a phenomenon). If 5/12 is the only seasonal frequency with a visually significant peak, we would not take it as strong evidence of a seasonal series or seasonality, but it **can** support the evidence of visually significant peaks at other seasonal frequencies. The software does not provide a warning if the 6/12 frequency has a peak, even though this is a seasonal frequency. Theoretical results indicate that the spectral estimate has a higher variance at frequency 6/12 than at other seasonal frequencies. Also, empirical findings are that visually significant peaks at 6/12 occur too often in the spectra of seasonally adjusted and irregular series that have few or no other visually significant seasonal peaks.

Users might not consider a single visually significant peak to be sufficient evidence of seasonality. Instead, they might require at least two visually significant peaks before accepting that the diagnostic indicates seasonality, particularly if the plot does not have a visually significant peak at 1/12. Some users consider 1/12 to be the most important frequency, as it indicates a recurring pattern every 12 months, i.e., it is the fundamental characteristic of seasonality.

Some users keep track of peaks marginally below the visual significance thresholds, to be aware of the potential for future warnings as additional series estimates become available.

The spectrum of the seasonally adjusted series is expected to show troughs at the seasonal frequencies, but pending further research into this issue, the recommendation for posttest interpretation remains the same as for the pretest spectral graphs.

Figure 3 and Figure 4 below are the spectral plot of the original series (pretest), indicating six visually significant seasonal peaks (at 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12); this is strong evidence that the series is seasonal.[6]

Figure 3 shows the spectral plot as it appears in the X-13ARIMA-SEATS main output file, in text characters. The character *S* indicates the seasonal frequencies,

---

[6] With ARIMA model-based seasonal adjustment from SEATS, the optimal (minimum mean squared error) estimator of the nonseasonal component (the seasonally adjusted series using the true model) will generally have troughs in its spectrum at the seasonal frequencies. Since these troughs would, obviously, not indicate a problem with the (optimal) adjustment, in practice troughs at seasonal frequencies in the spectrum of an adjusted series should not generally be taken as indicating any problem with the adjustment and should even be expected. (Note **Figure 5**.) This tendency towards spectral troughs should, ideally, be taken into account when looking for spectral peaks at seasonal frequencies for seasonally adjusted series (or trend or irregular estimates), but specific methods for doing this have yet to be implemented.

and the character *T* indicates the trading day frequencies. (In the output file, the seasonal frequency 6/12 also appears, but the software does not check for peaks at that frequency.) In the text representation, users can count the number of stars of each peak; for example, the height of the peak at 1/12 is 13 stars, well above the threshold of six. The peaks at 1/12, 2/12, 3/12, 4/12, and 5/12 are the tallest of the plot except for the 6/12 frequency, so they are certainly above the median level.

By contrast, **Figure 4** does not show the specific number of stars of each peak, but it shows the letter *S* above each visually significant peak at a seasonal frequency. It also provides an indicator bar to the right of the plot, with the median level (the bottom of the bar) and the height of six stars (the vertical length of the bar), so users have the visual significance measures (median level and six-star height) to use in interpreting the plot.

Figure 5 shows the spectral plot of the seasonally adjusted series (posttest), which has no visually significant peaks, so by this measure the adjustment shows no evidence of residual seasonality.

Table 4 shows abbreviated spectral peak information from the diagnostics tables of Win X-13, providing some of the same information that is in the graphs. For instance, it indicates that 1/12 (s1), 2/12 (s2), 3/12 (s3), 4/12 (s4), and 5/12 (s5) are visually significant peaks in the spectrum of the original series. It also indicates that a visually nonsignificant peak occurs in the spectrum of the model residuals, being 1.1 stars in height. It does not provide the frequencies of the nonsignificant peaks. Users can check the plots to identify the frequencies that correspond to those peaks, if desired. For this purpose, note that bars are plotted for 61 spectral ordinates from 0 to 6/12, including the endpoints, so each interval (except for those that involve the trading day frequencies) on the horizontal axis is 0.5/60 = 0.00833.

Summary

- Peaks that are (1) six or more stars tall and (2) above the median level are visually significant; visually significant peaks at seasonal frequencies are evidence of seasonality (or evidence of trading day effects if they occur at those frequencies).
- In practice, we often use 1/12, 2/12, 3/12, and 4/12 only, and discount peaks at frequency 5/12. The results for 6/12 are not considered reliable, and the software does not flag significant peaks at the frequency 6/12, although it is a seasonal frequency.

G 0 10*LOG(SPECTRUM) of the differenced, transformed Prior Adjusted Series (Table B1)

Spectrum estimated from 2010.Jan to 2017.Dec.

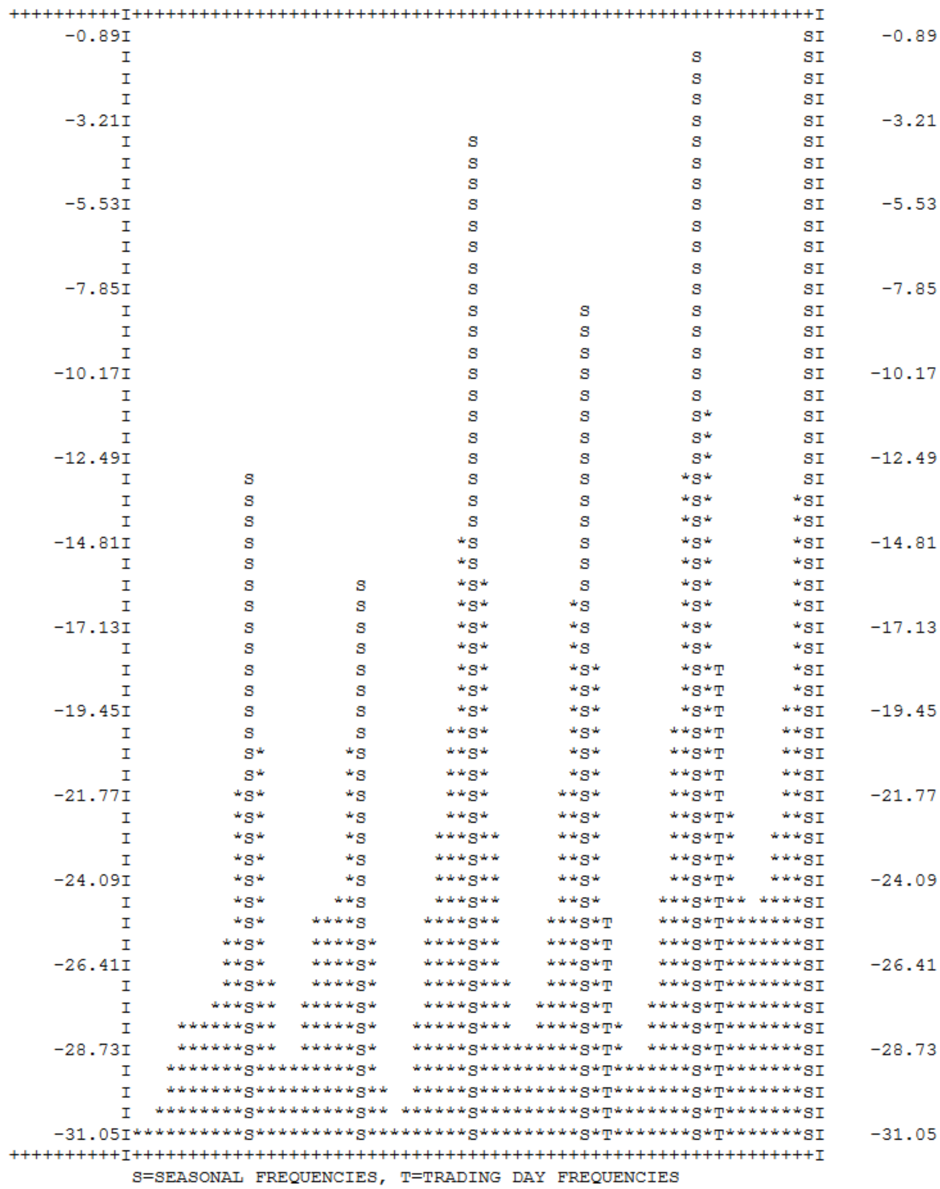

Figure 3: From the main output file of an X-13ARIMA-SEATS run, the text version of the plot; pretest, spectral plot of the prior-adjusted original series (Table B1) providing evidence that the series is seasonal. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

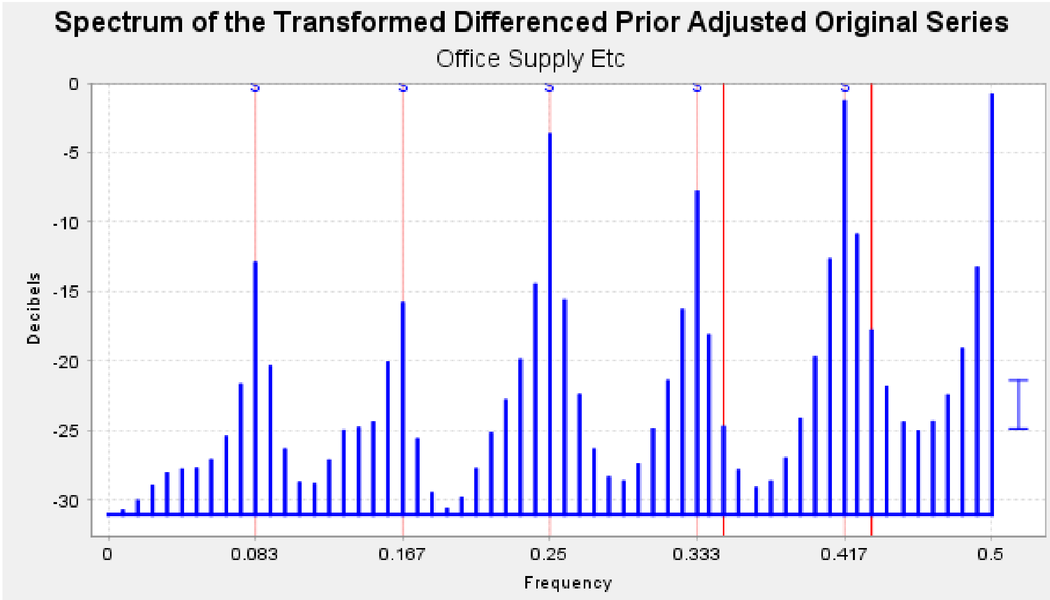

Figure 4: Pretest, spectral plot of the original series providing evidence that the series is seasonal, note the bold **S** label at frequencies 1/12, 2/12, 3/12, 4/12, and 5/12. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

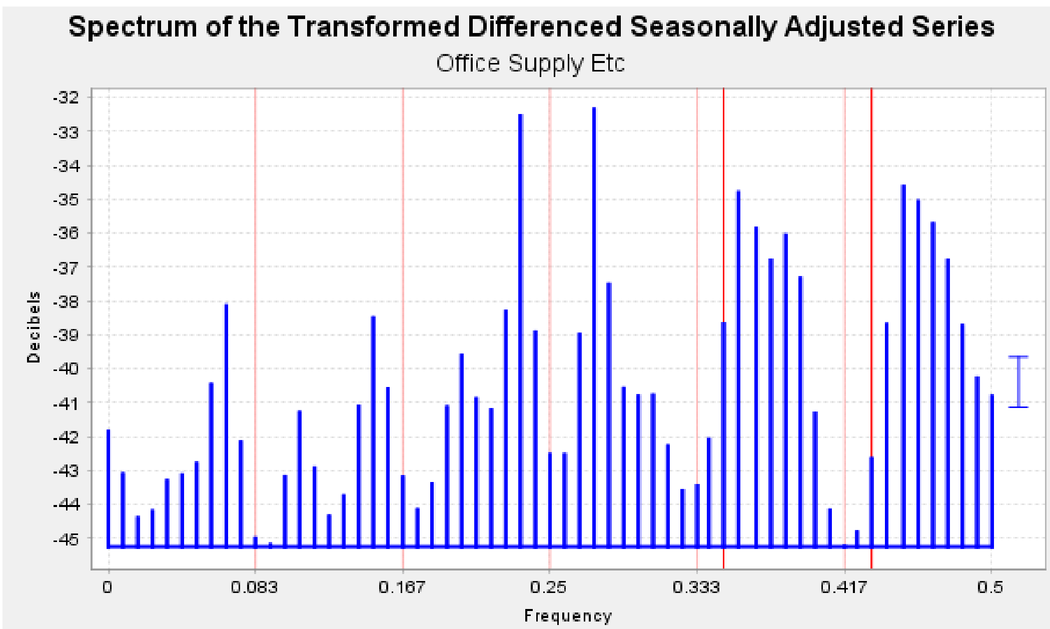

Figure 5: Posttest, spectral plot of the seasonally adjusted series providing no evidence of residual seasonality. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Table 4: Abbreviated spectral plot information from Win X-13, the same information as in the graphs. The abbreviated column headings indicate the (1) series name, (2) visually significant peaks in the spectrum of the original series (including trading day peaks, if present), (3) visually significant peaks in the spectrum of the model residuals, (4) visually significant peaks in the spectrum of the seasonally adjusted series, (5) visually significant peaks in the spectrum of the modified irregular, (6) not visually significant peaks (less than six stars in height or below the median level) in any of the spectra at seasonal frequencies, and (7) not visually significant peaks in any of the spectra at trading day frequencies. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Series Name	Sig Ori Peaks	Resid Peaks	Sig SAdj Peaks	Sig Irr Peaks	Nonsig Seasonal Peaks	Nonsig TD Peaks
Office Supply Etc	s1 s2 s3 s4 s5				rsd [1.1]	

General Remarks on Pretesting and Posttesting

It is important to keep in mind that the model-based F test is testing something very different from what QS and the spectral diagnostic are testing. The F test is testing for the presence of a fixed seasonal pattern. Even if seasonality is evolving over time, there will typically be a fixed pattern underlying the evolving seasonal component. This is the case for ARIMA model-based seasonal adjustment when the ARIMA model includes seasonal differencing, as is typical. It is also implicitly the case for X-11 seasonal adjustment, since all the X-11 seasonal adjustment filters will annihilate, and X-11 seasonal filters will reproduce, fixed seasonal effects (Bell 2012). Detecting a fixed seasonal pattern is thus a strong indication of a seasonal time series for which seasonal adjustment should be considered.

In contrast, QS and the spectral diagnostic are testing for any seasonal dependence in the form of positive seasonal autocorrelations or peaks in the spectrum at seasonal frequencies. If fixed seasonal effects are present, they will typically produce large estimated seasonal autocorrelations and strong peaks at seasonal frequencies in the estimated spectrum that will flag significance for QS and the spectral diagnostic. However, we note above and in the Appendices that QS and the spectral diagnostic may detect moderate or even mild seasonal autocorrelation that would not necessarily produce discernible seasonal patterns in the data, and thus may not suggest seasonal adjustment. Increasing the length of the time series being analyzed will generally make this more likely to occur. QS may also detect nonzero seasonal autocorrelation in original series that arises from purely nonseasonal dependence. While QS and the spectral diagnostic applied to original and first differenced series can be examined as part of pretesting, their limitations should be kept in mind, and they should not be used uncritically in deciding whether to seasonally adjust a time series. The model-based F test is more appropriate for this purpose.

The situation with posttesting is different. Since X-11 and ARIMA model-based seasonal adjustment filters (from models with seasonal differencing) annihilate fixed seasonal effects, the F test is useless when applied to the full span of a seasonally adjusted series or irregular estimate. Findley et al. (2017) examines applying the F test to a reduced span, such as the last eight years, of an adjusted series. This is akin to looking for a change in seasonal pattern in the reduced span compared to earlier years, though this task can be accomplished more directly, and probably with better statistical calibration, by fitting a regARIMA model to the original series that includes seasonal regime change regressors at the start of the reduced span.

On the other hand, mild to moderate positive seasonal autocorrelation can be taken as indicating residual seasonality, the task for which QS and the spectral diagnostic, along with examining individually the lag 12 (monthly series) or lag 4 (quarterly series) seasonal autocorrelation, are better suited. Limitations of the statistical inference procedures of QS and the visual significance spectral test discussed above should be kept in mind. This includes the complication noted that one generally expects troughs at seasonal frequencies in the spectrum of a seasonally adjusted series or irregular estimate, and that specific methods to account for this effect when examining a spectrum for residual seasonality have yet to be implemented.

As a final remark, a significant posttest result might indicate a need to modify the regARIMA model (if using the SEATS method) or modify the seasonal adjustment options (if using the X-11 method) to find settings that more thoroughly remove the seasonal effects.

5 Examples

Note that the seasonal adjustments in these examples might differ from official estimates because of revisions to the published original series or because some of the adjustment settings or series spans are different. These examples are for purposes of illustrating the diagnostics only and are not official estimates of these time series values. As mentioned previously, the informal visual comparison of series values in graphs is not a rigorous statistical test for differences between the values.

5.1 Example 1: Direct Seasonal Adjustment of a Monthly Time Series, X-11 Adjustment Method

We first consider a monthly time series that is directly seasonally adjusted: the number of new privately owned single-family housing units authorized by building permits in permit-issuing places in the Northeast region of the U.S.[7] Graphs of the original series, over consecutive time intervals as in **Figure 6** or in the year-over-year graph as in **Figure 7**, show that the number of single-family building permits are higher in the summer months and lower in the winter months. There is some inconsistency year-to-year in which specific month is the top of the peak and which is the bottom of the trough, but the summer-winter difference is clear. The seasonality diagnostics confirm that the series is seasonal. Because of the large change in level in the late 2000s, we limited the span of **Figure 7** to 2009 and later. Graphing the series on the log scale or after removing outliers and/or other regression effects also might be beneficial when looking for seasonal patterns.

[7] The estimates are subject to sampling and nonsampling error. More information about data collection and estimation methods for the Building Permits Survey is available online (census.gov/construction/bps/).

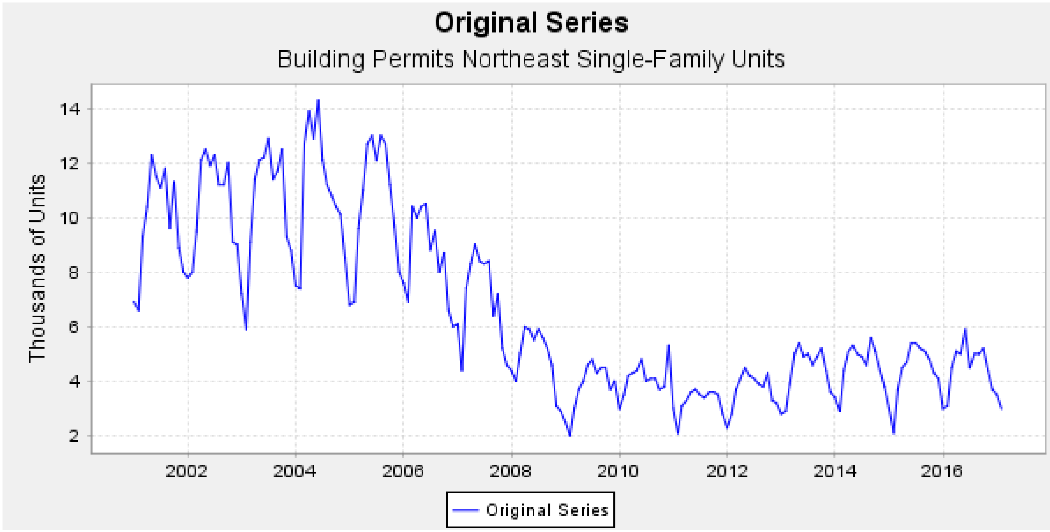

Figure 6: Building permits for single-family units (in thousands), original series, Northeast, January 2001 – February 2017. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

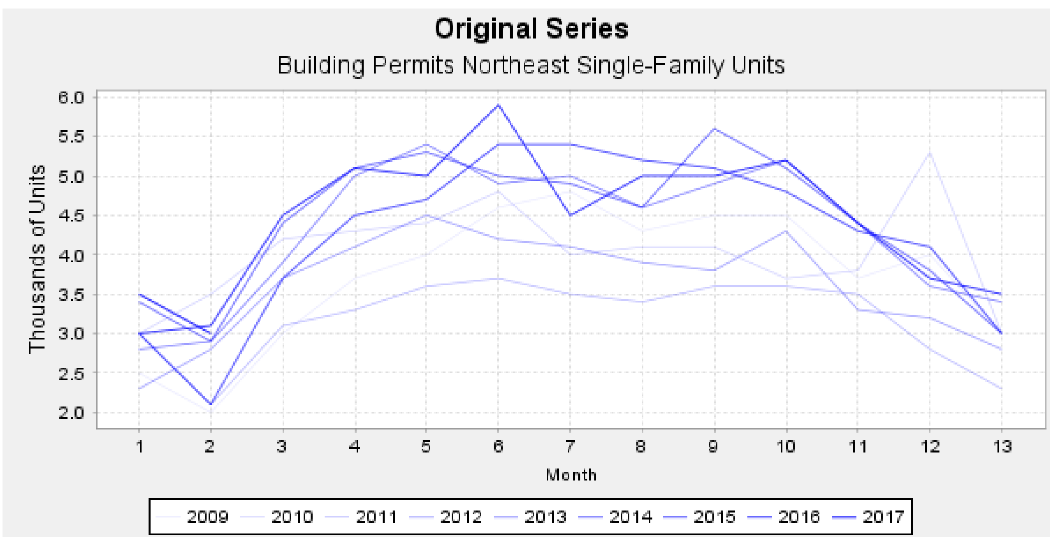

Figure 7: Building permits for single-family units (in thousands), Northeast, year over year. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

Pretest: autocorrelation function plots

An informal examination of seasonality involves inspecting the sample autocorrelation plots for large values at seasonal lags. **Figure 8**, the plot with no differencing, shows peaks in the autocorrelation function at seasonal lags 12, 24, and 36, a behavior that is consistent with the presence of seasonality. However, the high autocorrelation at the first lag – and the persistent autocorrelation over the succeeding lags – indicates it may be wise to difference the data. **Figure 9** displays autocorrelations resulting from a first difference (and no seasonal difference); the large values at seasonal lags 12, 24, and 36, in contrast to neighboring values, are further informal evidence of seasonality. If we go further and apply seasonal differencing, the seasonality has been removed to such a degree that negative correlation is obtained in the sample autocorrelations at lag 12, as shown in **Figure 10**.

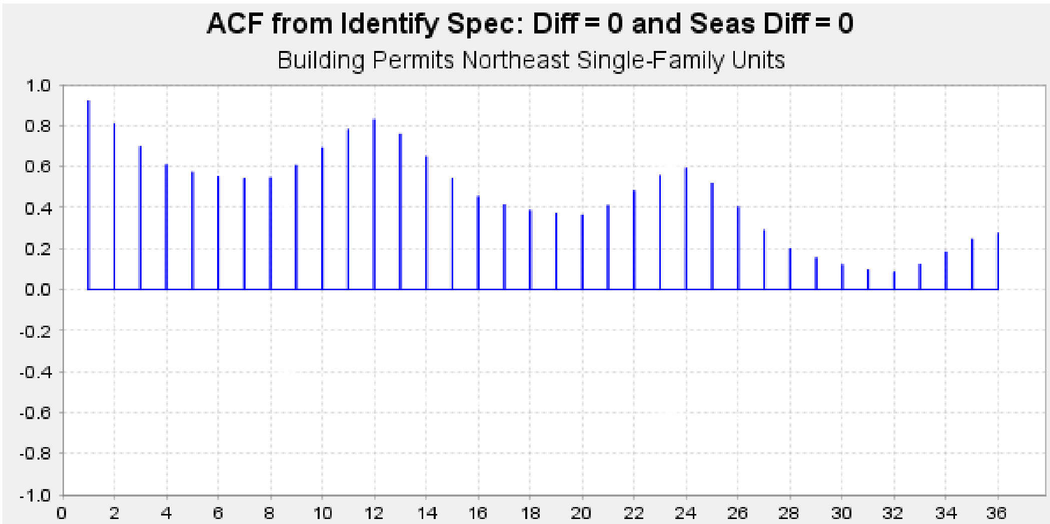

Figure 8: Autocorrelation function of Northeast single-family building permits time series with no differencing. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

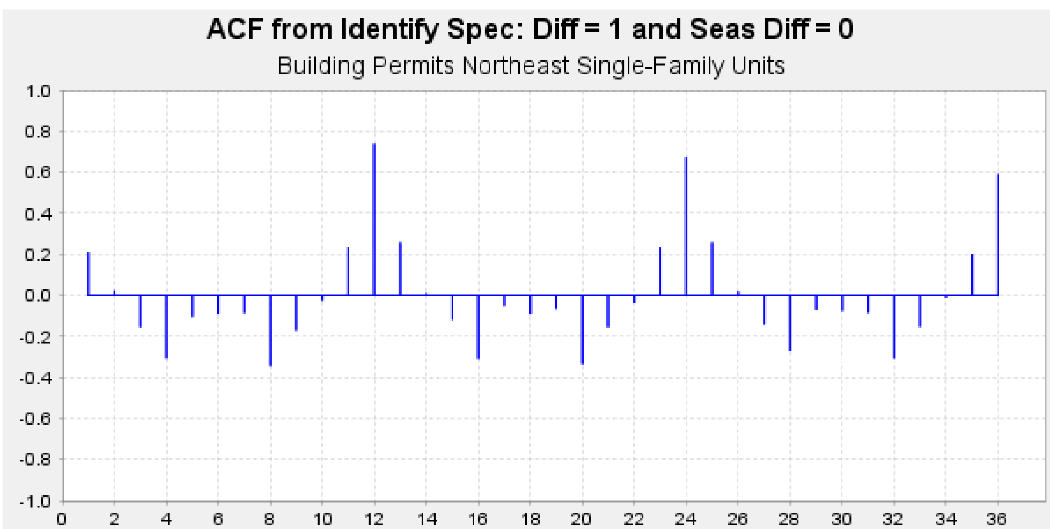

Figure 9: Autocorrelation function of Northeast single-family building permits time series with one first difference and no seasonal difference. **Source**: Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

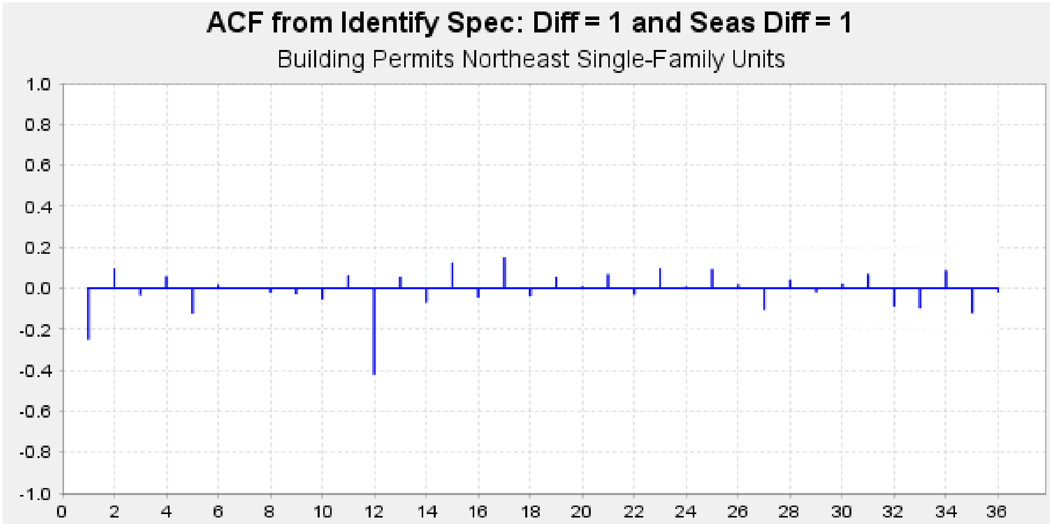

Figure 10: Autocorrelation function of Northeast single-family building permits time series with one first difference and one seasonal difference. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Pretest: regression model-based F test

For the pretest, we fit a regARIMA model together with the seasonal regressors to the full span of the series that is subject to seasonal adjustment.

A regARIMA model is available from an earlier analysis: it is an ARIMA (1 1 0)(0 1 1) model with regressors for the one-coefficient trading day effect and three outlier effects. To perform the model-based F test to determine if the series is seasonal, we removed the seasonal component of the ARIMA model and added regressors for the trend constant and for the seasonal effects, retaining the other regressors that were part of the original model.

Example spec file for the seasonal regression pretest:

```
series{
 file="BPNESingleFamily.dat" format="datevalue"
 span=(2001.1,)
}
spectrum{qcheck=yes savelog=all}
transform{function=log}
regression{
 variables=(const seasonal td1coef AO2010.Dec LS2011.Feb TC2015.Feb)
}
outlier{types=(AO LS TC) lsrun=3}
arima{model=(1 1 0)}
#comment to remove the previous model
#arima{model=(1 1 0)(0 1 1)}
estimate{print=(roots regcmatrix acm) savelog=all}
check{print=all savelog=all}
```

Table 5 shows the F statistic and *p*-value for the group of seasonal regressors fit to the full series from January 2001 to February 2017. The seasonal regressor group was highly significant at the 0.01 level (the 0.01 critical value is only 2.35). We rejected the null hypothesis, which was that the series is not seasonal.

Table 5: Degrees of freedom, F statistic, and *p*-value for the group of 11 seasonal regressors for the Northeast single-family building permits time series. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Regressor	DF	F Statistic	*P*-Value
Seasonal	11, 177	77.55	0.00

We also computed the F test and *p*-value for the group of seasonal regressors using automatic modeling instead of the previous ARIMA model, specifying the seasonal regressors as part of the model. The ARIMA model chosen was again the (1 1 0) model, though without the trend constant, as it was not significant. The resulting F statistic was nearly identical at 77.79 with (11,178) degrees of freedom. So automatic model selection led to the same conclusion.

Summary: The regression model-based F test provides evidence that the original series is seasonal.

Pretest: spectral plot peaks at seasonal frequencies

Because this is a monthly series, spectral plots are available. **Figure 11** and **Figure 12**, respectively, show the (same estimate of the) spectrum of the original series (adjusted for outliers and trading day effects), from the main output file in text form and from Win X-13 and X-13-Graph in higher resolution. The spectrum is from the last eight years of the time series, the default length. The spectral plot shows visually significant peaks at seasonal frequencies 1/12, 2/12, 4/12, and 5/12. A peak occurs at seasonal frequency 3/12, and it is above the median level of all the frequencies, but it is not taller than the larger of the values at the two neighboring frequencies by six stars, so it does not meet this criterion for visual significance.

For this example series, the peaks at 1/12, 2/12, 4/12, and 5/12 are obviously above the median level of all frequencies of the plot and so much taller than their nearest neighboring frequencies that their visual significance is apparent without a count of the stars. Note, however, that according to **Figure 11** from the main output file, the peak at 1/12 is 21 stars, the 2/12 peak is 27 stars, the 3/12 peak is 5 stars, the 4/12 peak is 14 stars, and the 5/12 peak is 7 stars. **Figure 12**, from X-13-Graph, shows a bold **S** at 1/12, 2/12, 4/12, and 5/12, indicating that they are visually significant. In addition, the bottom of the scale bar on the right side of the graph indicates the median level, and the vertical length of the bar indicates a measure equivalent to six stars, so the graph clearly shows that the peak at 3/12 is above the median but is less than six stars in height. If 5/12 were the only seasonal frequency with a visually significant peak, it would not be enough evidence of seasonality, but it supports the evidence of the visually significant peaks at other seasonal frequencies.

G 0 10*LOG(SPECTRUM) of the differenced, transformed Prior Adjusted Series (Table B1)

Spectrum estimated from 2009.Mar to 2017.Feb.

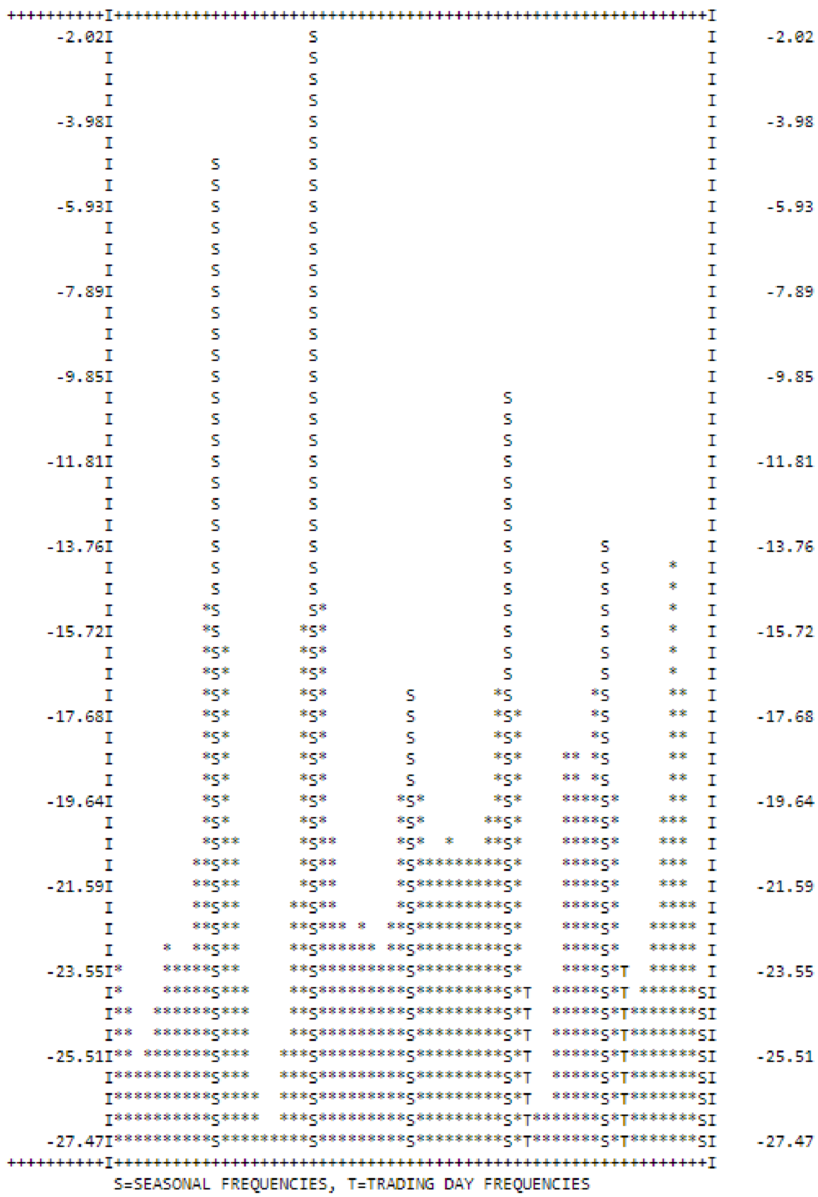

Figure 11: Spectrum of the differenced, transformed prior-adjusted Northeast single-family building permits time series (Table B1), with visually significant peaks at seasonal frequencies 1/12, 2/12, 4/12, and 5/12, from the X-13ARIMA-SEATS main output file (screen capture). **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)](census.gov/construction/bps/)

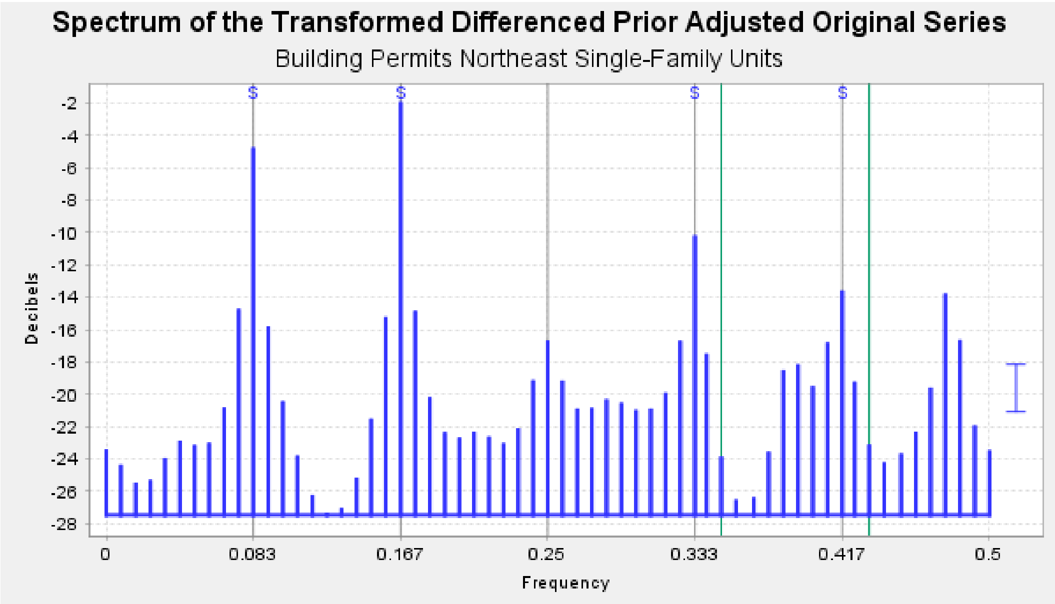

Figure 12: Spectral graph from X-13-Graph for the Northeast single-family building permits time series, showing visually significant peaks at seasonal frequencies 1/12, 2/12, 4/12, and 5/12. The peak at 3/12 does not meet the criteria of visual significance. A bold **S** label appears above each visually significant seasonal peak. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Win X-13 lists S1, S2, S4, and S5 under **Sig Ori Peaks** (Significant Peaks for the Prior-Adjusted Original Series), providing the same information apparent in the graph but condensing it for easy capture. **Table 6** shows this snippet from the Win X-13 tables. If the spectrum had visually significant peaks at the trading day frequencies, Win X-13 tables would label those as T1 and/or T2.

Table 6: From Win X-13, the list of visually significant peaks at seasonal frequencies of the prior-adjusted original series, Northeast single-family building permits. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series Name	Sig Ori Peaks
Building Permits Northeast Single-Family Units	s1 s2 s4 s5

Summary: The spectral plot provides evidence that the original series is seasonal.

The F test, spectrum, and autocorrelation function plots provide evidence that the original series is seasonal, and hence it can be considered for seasonal adjustment.

Posttest: QS diagnostics

We checked the results of the posttest QS statistics for seasonality. The software generates the QS for every direct seasonal adjustment. The QS statistics and corresponding *p*-values are available from a table in the main output file (**Table 7** below shows the posttest QS statistics) and the *p*-values appear in the diagnostics tables of Win X-13. Recall that if the estimate of the seasonal autocorrelation is negative or zero, the software sets the QS statistic to zero and the *p*-value to one.

We suggest checking the shortened series for residual seasonality in the seasonally adjusted series, the seasonally adjusted series adjusted for extreme values, the irregular series, and the irregular series adjusted for extreme values. The results for this example adjustment, shown in **Table 7**, indicate that we would not reject the null hypothesis that the adjusted series has no residual seasonality.

We suggest checking the results of the quarterly version of the series, especially if users might convert the monthly series to quarterly for further analysis. We used `qcheck=yes` in the `spectrum` spec to generate the QS diagnostics for the quarterly aggregates of the monthly series (the same time series aggregated to quarterly values). As **Table 8** shows, with a testing level of 0.05, all the quarterly QS statistics are nonsignificant; we would not reject the null hypothesis that the adjusted quarterly series has no residual seasonality. Notice, though, the difference between the *p*-values in **Table 8** with and without the extreme-value adjustment is remarkable and merits further investigation.

Table 7: QS statistics for the test for seasonality in the seasonally adjusted series and irregular component, for the full series and subspan (starting in March of 2009), of the Northeast single-family building permits time series; some tests are for the series adjusted for extreme values. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series	Span	QS	*P*-Value
Seasonally Adjusted Series	Full Series	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Full Series	0.00	1.0000
Irregular Series	Full Series	0.00	1.0000
Irregular Series (extreme value adjusted)	Full Series	0.00	1.0000
Seasonally Adjusted Series	Subspan	0.09	0.9540
Seasonally Adjusted Series (extreme value adjusted)	Subspan	0.00	1.0000
Irregular Series	Subspan	0.00	1.0000
Irregular Series (extreme value adjusted)	Subspan	0.00	1.0000

Table 8: QS statistics for the test for seasonality in the quarterly aggregate of the monthly seasonally adjusted series of Northeast single-family building permits, for the full series and shortened series (starting in March of 2009); some tests are for the series adjusted for extreme values. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series	Span	QS	*P*-Value
Seasonally Adjusted Series	Full Series	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Full Series	4.64	0.0983
Seasonally Adjusted Series	Subspan	0.04	0.9788
Seasonally Adjusted Series (extreme value adjusted)	Subspan	2.16	0.3393

Summary: The QS results provide no evidence of residual seasonality at the .05 statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies

For this direct adjustment, spectral diagnostics are available from the main output file itself and the Win X-13 diagnostics tables and graphs. We looked for visually significant peaks in the spectrum of the seasonally adjusted series and irregular series. (As a model fit diagnostic, the spectrum of the regARIMA model residuals also is available.)

[**Table 9**](#) shows a small selection of diagnostics from Win X-13's spectral peak information. These table cells provide the posttest results **Sig SAdj Peaks** for the modified seasonally adjusted series, and **Sig Irr Peaks** for the modified irregular component). Win X-13 also provides information on peaks that are not visually significant. Under the heading **Nonsig Seasonal Peaks**, Win X-13 indicates "sa irr [4]," meaning that at least one peak occurred in each of the spectra of the modified seasonally adjusted series (sa) and of the modified irregular component (irr) that did not meet the criteria of visual significance, with the tallest of the peaks having four stars. No visually significant seasonal peaks occurred in the spectra of the seasonally adjusted series or the irregular component, so those table cells are blank.

Table 9: Spectral peak information from Win X-13; no visually significant peaks. Peaks that are not visually significant occurred in the spectra of the seasonally adjusted series and the irregular component; Northeast single-family building permits.**Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series Name	Sig SAdj Peaks	Sig Irr Peaks	Nonsig Seasonal Peaks
Building Permits Northeast Single-Family Units			sa irr [4]

We checked the spectral graphs to find out what the nonsignificant peaks were. For this series, **Figure 13**, the spectrum of the seasonally adjusted series, indicates that the peak at frequency 2/12 is only three stars. Another nonsignificant seasonal peak occurred in the spectrum of the irregular. See **Figure 14**, the spectrum of the modified irregular component, indicating a nonsignificant peak of four stars at seasonal frequency 2/12. The same information is available from the higher-resolution graphs from Win X-13 and X-13-Graph.

G 1 10*LOG(SPECTRUM) of the differenced, transformed Seasonally Adjusted Data (Table E2).

Spectrum estimated from 2009.Mar to 2017.Feb.

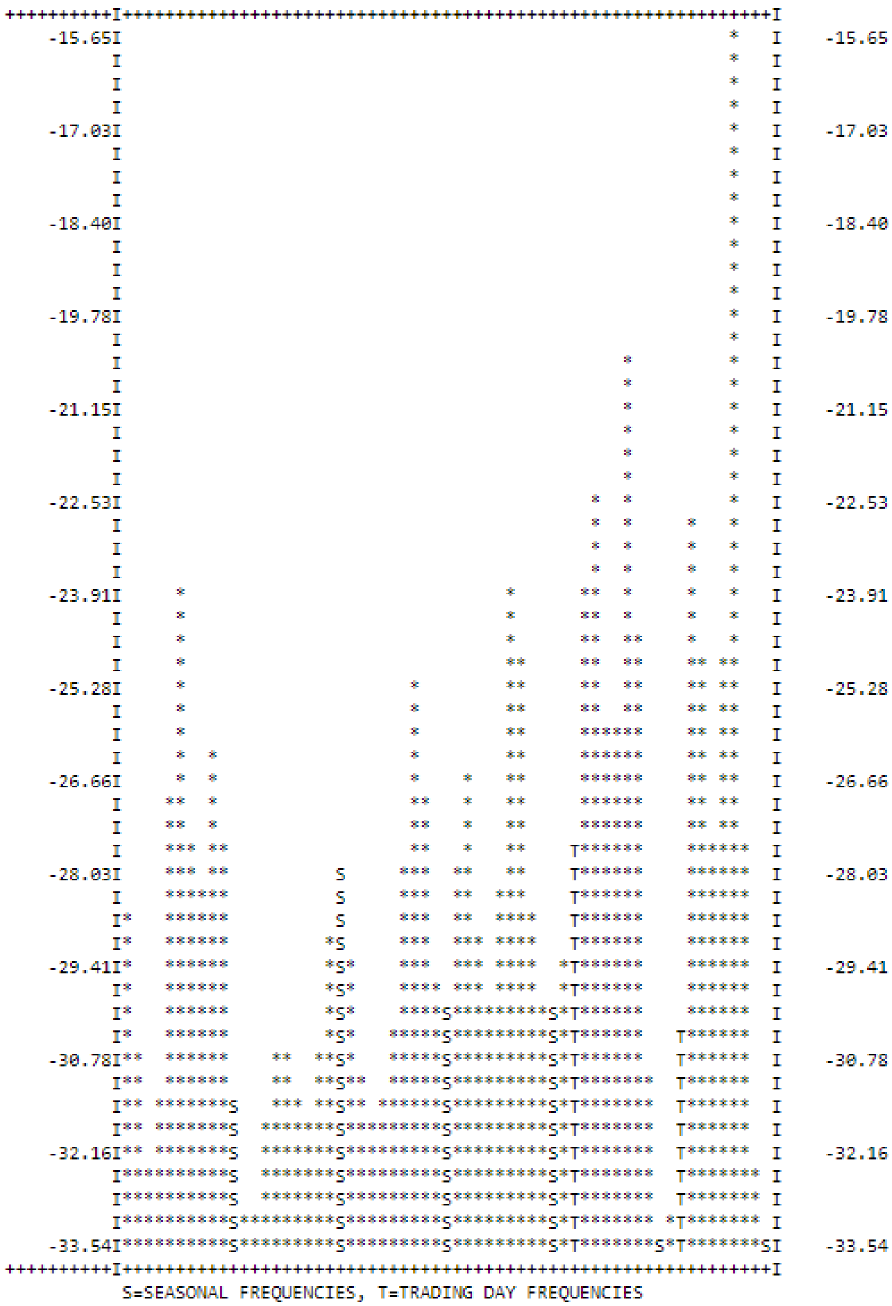

Figure 13: Spectrum of the differenced, transformed seasonally adjusted series, from the main output file for Northeast single-family building permits; note that the plot shows no visually significant peaks and has one nonsignificant peak at frequency 2/12.
Source: [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)](census.gov/construction/bps/)

G 2 10*LOG(SPECTRUM) of the Modified Irregular (Table E3).

Spectrum estimated from 2009.Mar to 2017.Feb.

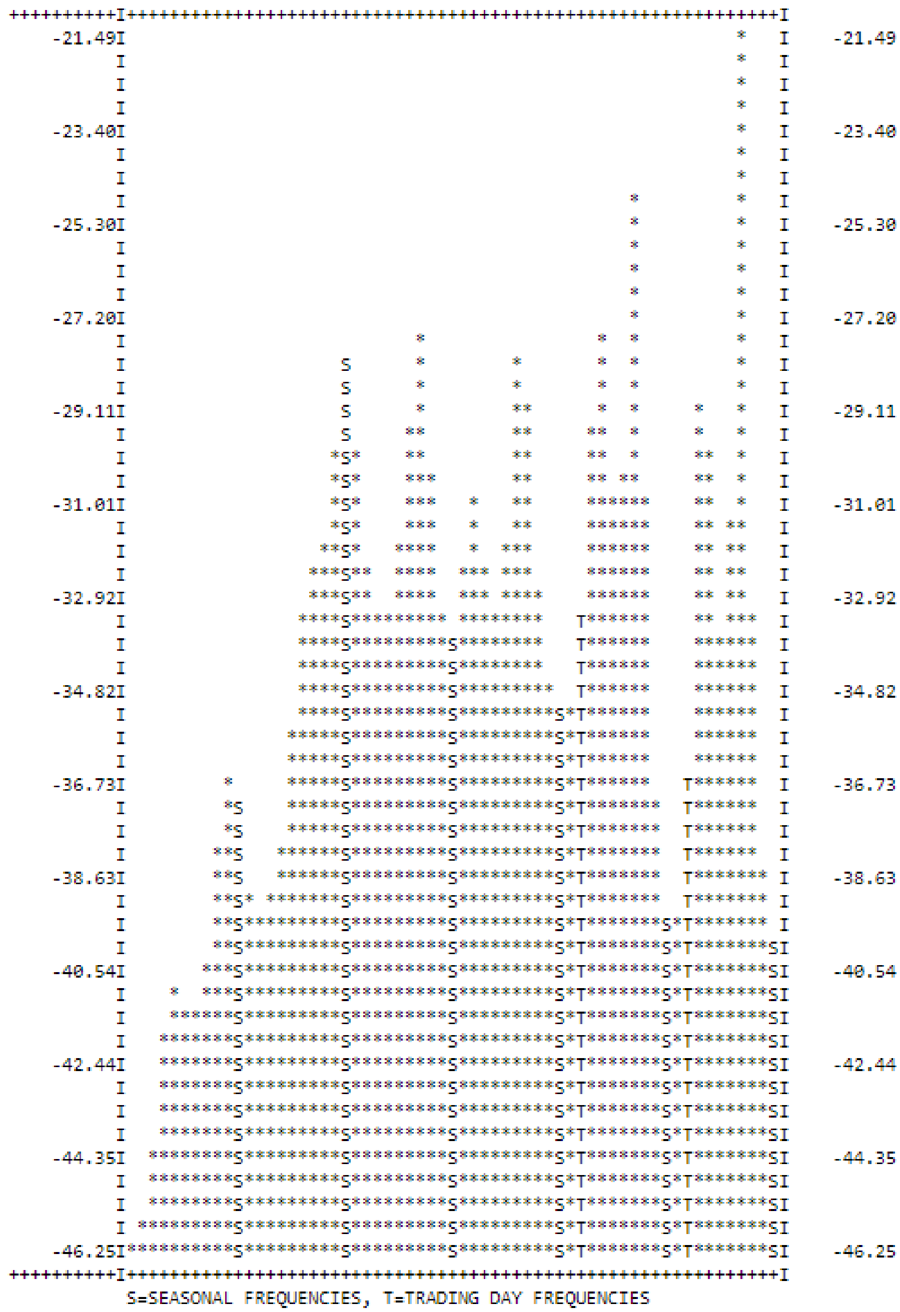

Figure 14: Spectrum of the modified irregular component series, from the main output file, Northeast single-family building permits; note that the plot shows no visually significant peaks and has one nonsignificant peak at frequency 2/12. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)](census.gov/construction/bps/)

Summary: The spectral plot results provide no evidence of residual seasonality using established significance levels.

The QS statistics and the spectral diagnostics provide no evidence of residual seasonality at established significance levels. If the adjustment meets other established quality measures, it is adequate.

5.2 Example 2: Direct Seasonal Adjustment of a Monthly Time Series, SEATS Adjustment Method

Like the Northeast building permits time series, the time series of building permits issued for single-family housing units in the West region of the U.S.[8] appears to be seasonal. Graphs of the original series, depicted in **Figure 15** and **Figure 16**, indicate a pattern similar to that of the Northeast, that is, the activity is higher in the summer months and lower in the winter months, with some inconsistencies as to the highest and lowest months each year.

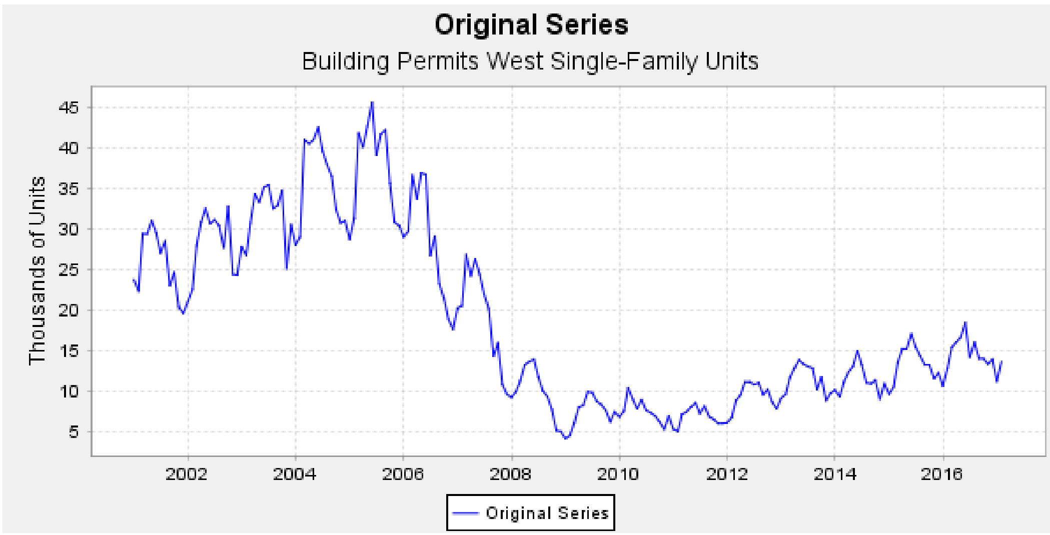

Figure 15: Building permits for single-family units (in thousands), original series, West, January 2001 – February 2017. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

[8] The estimates are subject to sampling and nonsampling error. More information about data collection and estimation methods for the Building Permits Survey is available online (census.gov/construction/bps/).

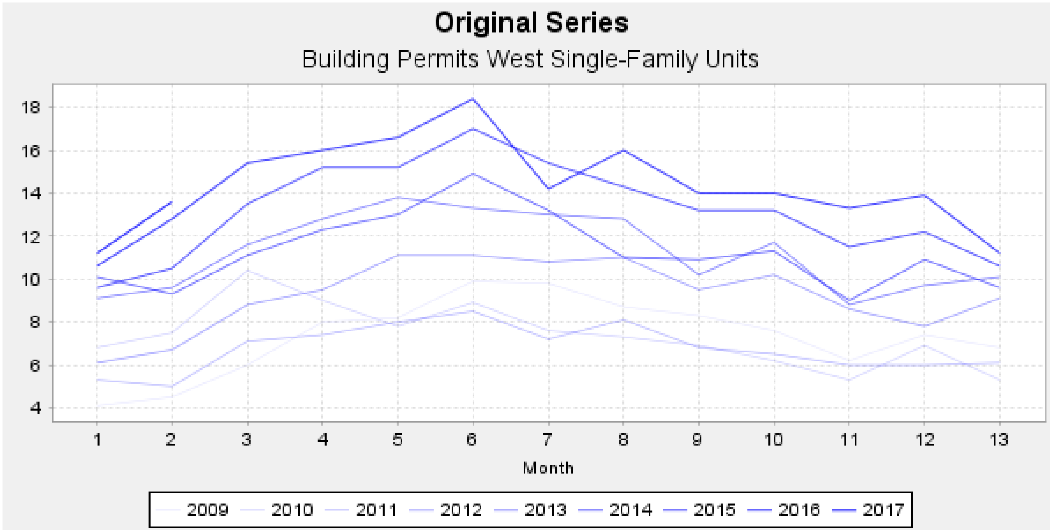

Figure 16: Building permits for single-family units (in thousands), West, year over year.
Source: [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)](census.gov/construction/bps/)

Pretest: autocorrelation function plots

As in Example 1, the sample autocorrelation plots might provide some evidence in the determination of seasonality. **Figure 17**, the plot with no differencing, shows very slight rises in the autocorrelation function at seasonal lags 12 and 24, providing some evidence that the series might be seasonal, although these are not as strong as the peaks seen from the Northeast time series. As with the Northeast building permits time series, the persistent autocorrelation over all the lags indicates differencing is necessary. **Figure 18** shows the autocorrelations resulting from a first difference (and no seasonal difference); the large values at seasonal lags 12, 24, and 36, in contrast to neighboring values, are further informal evidence of seasonality. If we apply seasonal differencing, as we saw with the Northeast building permits series, the seasonality has been removed to such a degree that negative correlation is obtained in the sample autocorrelations at lag 12, as shown in **Figure 19**.

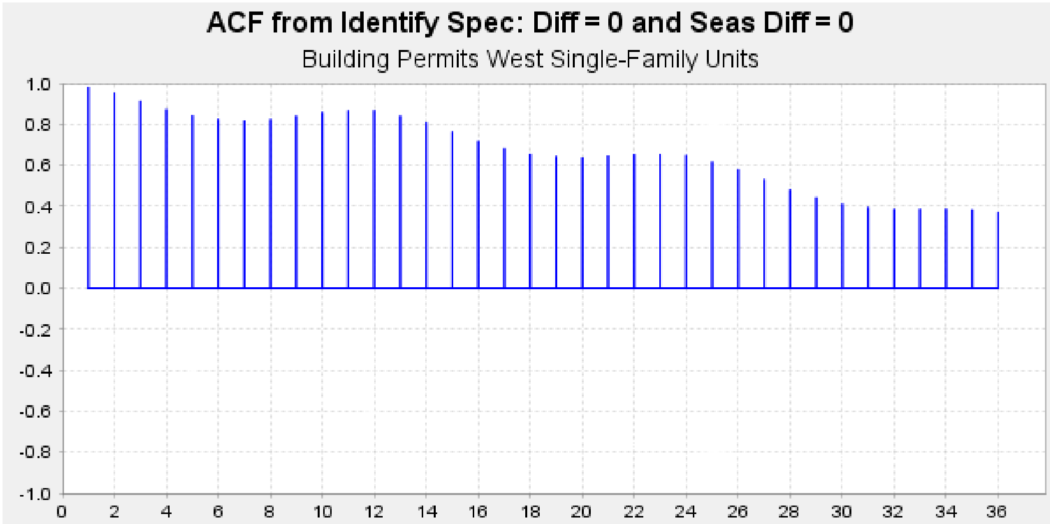

Figure 17: Autocorrelation function of West single-family building permits time series with no differencing. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

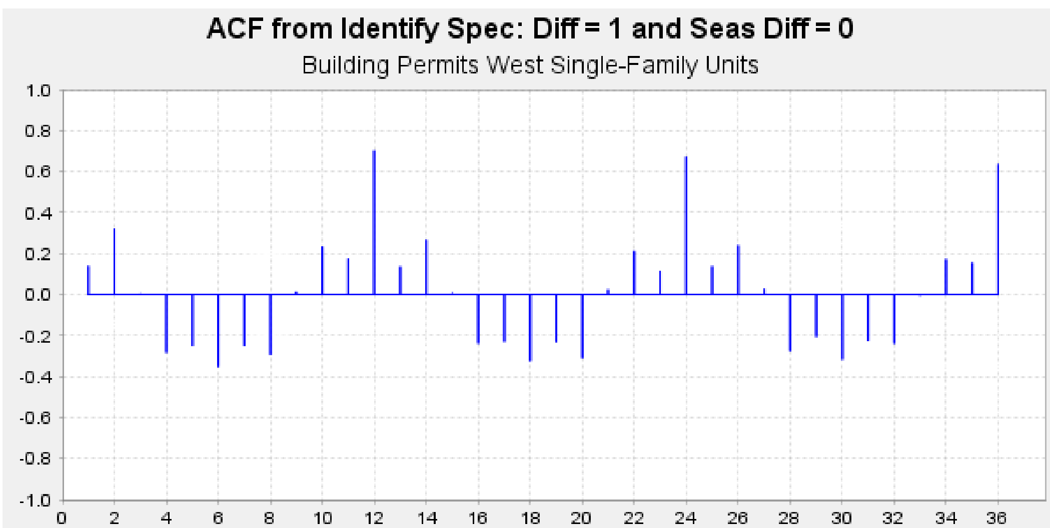

Figure 18: Autocorrelation function of West single-family building permits time series with one first difference and no seasonal difference. **Source**: [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

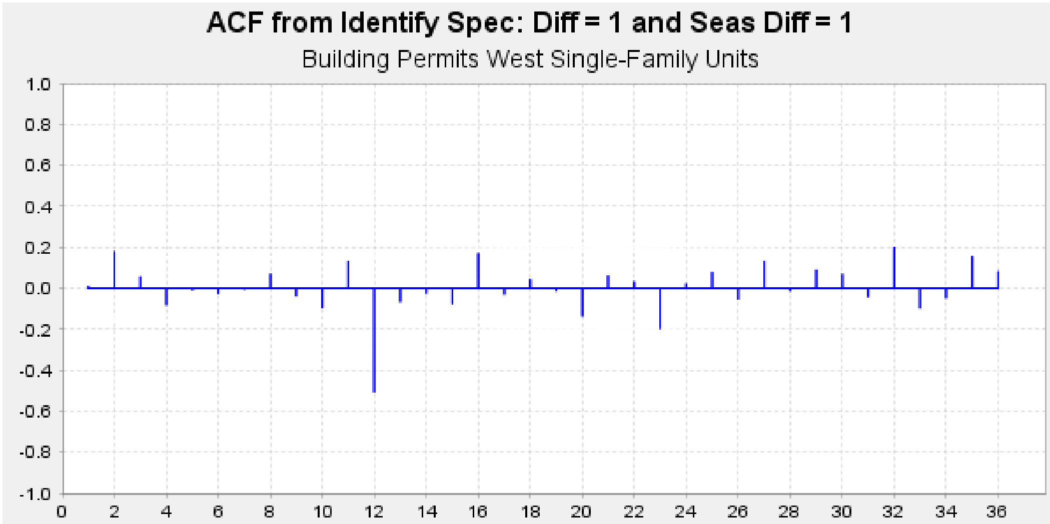

Figure 19: Autocorrelation function of West single-family building permits time series with one first difference and one seasonal difference. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Pretest: regression model-based F test

For the pretest, we fit a regARIMA model together with the seasonal regressors to the full span of the series that is subject to seasonal adjustment.

For this test, we chose to use automatic modeling settings. To perform the model-based F test to determine if the series is seasonal, we entered regressors for the trend constant and for the seasonal effects, and we used automatic testing for trading day and Easter regression effects.

Example spec file for the seasonal regression pretest:

```
series{
 file="BPWSingleFamily.dat" format="datevalue"
 span=(2001.1,)
}
spectrum{qcheck=yes savelog=all}
transform{function=log}
regression{
 variables=(const seasonal)
 aictest=(td easter) savelog=aictest
}
outlier{types=(AO LS TC) lsrun=3}
automdl{maxorder=(3 1) maxdiff=(1 0)}
estimate{print=(roots regcmatrix acm) savelog=all}
check{print=all savelog=all}
```

Table 10 shows the F statistic and *p*-value for the group of seasonal regressors fit to the full series from January 2001 to February 2017. The seasonal regressor group was highly significant at the 0.01 level. We rejected the null hypothesis, which was that the series is not seasonal.

Table 10: Degrees of freedom, F statistic, and *p*-value for the group of 11 seasonal regressors for the West single-family building permits time series. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

Regressor	DF	F Statistic	*P*-Value
Seasonal	11, 179	49.74	0.00

Summary: The regression model-based F test provides evidence that the original series is seasonal.

Pretest: spectral plot peaks at seasonal frequencies

Because this is a monthly series, spectral plots are available. **Figure 20** shows the spectrum of the original series (adjusted for outliers and trading day effects) from X-13-Graph. The spectrum is from the last eight years of the time series, the default length. The spectral plot shows visually significant peaks at seasonal frequencies 1/12, 4/12, and 5/12. A small peak occurs at seasonal frequency 2/12, and it is above the median level of all the frequencies, but it is not taller than the larger of the values at the two neighboring frequencies by six stars, so it does not meet both criteria for visual significance. A bold **S** appears above the visually significant peaks.

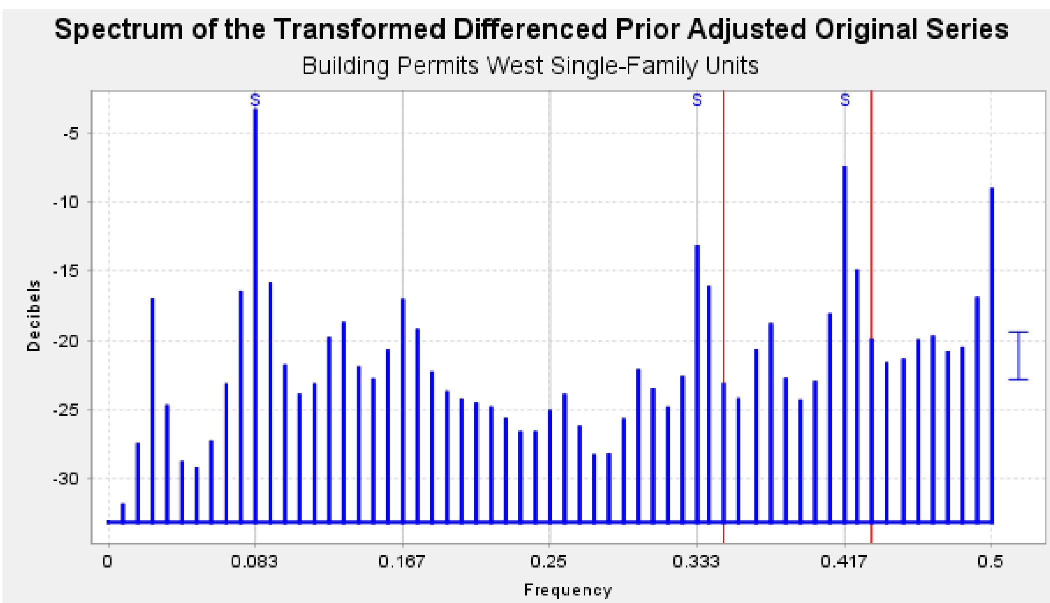

Figure 20: Spectral graph from X-13-Graph for the West single-family building permits time series, showing visually significant peaks at seasonal frequencies 1/12, 4/12, and 5/12. The peak at 2/12 does not meet the criteria of visual significance. A bold **S** label appears above each visually significant seasonal peak. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

Summary: The spectral plot provides evidence that the original series is seasonal.

Both of the more formal pretests and the autocorrelation function plots provide evidence that the original series is seasonal, and hence it can be considered for seasonal adjustment.

Posttest: QS diagnostics

We chose to seasonally adjust the series using the SEATS ARIMA model-based method, but although the adjustment method is different from Example 1, we used the same posttest diagnostics. We checked the results of the posttest QS statistics for seasonality. The software generates the QS for every direct seasonal adjustment. The QS statistics and corresponding *p*-values are available from a table in the main output file (**Table 11** below shows the posttest QS statistics) and the *p*-values appear in the diagnostics tables of Win X-13.

As in Example 1, we checked the shortened series for residual seasonality in the seasonally adjusted series, the seasonally adjusted series adjusted for extreme values, the irregular series, and the irregular series adjusted for extreme values. The results for this example adjustment, shown in **Table 11**, indicate that we would not reject the null hypothesis that the adjusted series has no residual seasonality.

As before, we used `qcheck=yes` in the `spectrum` spec to generate the QS diagnostics for the quarterly aggregates of the monthly series (the same time series aggregated to quarterly values). As **Table 12** shows, with a testing level of 0.05, all of the quarterly QS statistics are nonsignificant; we would fail to reject the null hypothesis that the adjusted quarterly series has no residual seasonality. Notice, though, just as in Example 1, the difference between the *p*-values in **Table 12** with and without the extreme-value adjustment merits further investigation.

Table 11: QS statistics for the test for seasonality in the seasonally adjusted series and irregular component, for the full series and subspan (starting in March of 2009), of the West single-family building permits time series; some tests are for the series adjusted for extreme values. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series	Span	QS	*P*-Value
Seasonally Adjusted Series	Full Series	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Full Series	0.00	1.0000
Irregular Series	Full Series	0.00	1.0000
Irregular Series (extreme value adjusted)	Full Series	0.00	1.0000
Seasonally Adjusted Series	Subspan	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Subspan	0.00	1.0000
Irregular Series	Subspan	0.04	0.9802
Irregular Series (extreme value adjusted)	Subspan	0.00	1.0000

Table 12: QS statistics for the test for seasonality in the quarterly aggregate of the monthly seasonally adjusted series of West single-family building permits, for the full series and shortened series (starting in the second quarter of 2009); some tests are for the series adjusted for extreme values. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Table	Span	QS	*P*-Value
Seasonally Adjusted Series	Full Series	0.83	0.6601
Seasonally Adjusted Series (extreme value adjusted)	Full Series	3.45	0.1786
Seasonally Adjusted Series	Subspan	0.00	1.0000
Seasonally Adjusted Series (extreme value adjusted)	Subspan	4.48	0.1066

Summary: The QS results provide no evidence of residual seasonality at the .05 statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies

We looked for visually significant peaks in the spectrum of the seasonally adjusted series and irregular series. (As a model fit diagnostic, the spectrum of the regARIMA model residuals also is available, and because we used the SEATS seasonal adjustment method, the test of the model residuals is pertinent.)

[Table 13](#) shows a small selection of diagnostics from Win X-13's spectral peak information. These table cells provide the posttest results **Sig SAdj Peaks** for the modified seasonally adjusted series, and **Sig Irr Peaks** for the modified irregular component. Win X-13 also provides information on peaks that are not visually significant. Under the heading **Nonsig Seasonal Peaks**, Win X-13 indicates "rsd [1.8]," meaning that at least one peak occurred in the spectrum of the model residuals that did not meet the criteria of visual significance, with the tallest of the peaks being 1.8 stars. No visually significant seasonal peaks occurred in the spectra of the seasonally adjusted series or the irregular component, so those table cells are blank.

Table 13: Spectral peak information from Win X-13; no visually significant peaks. Peaks that are not visually significant occurred in the spectra of the seasonally adjusted series and the irregular component; West single-family building permits. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Series Name	Sig SAdj Peaks	Sig Irr Peaks	Nonsig Seasonal Peaks
Building Permits West Single-Family Units			rsd [1.8]

We looked at the spectral graphs. **Figure 21** shows the spectrum of the regARIMA model residuals. No seasonal frequencies have visually significant peaks; at 3/12, the spectrum has a nonsignificant peak. **Figure 22** shows the spectrum of the modified seasonally adjusted series, and **Figure 23** shows the spectrum of the modified irregular component series; they show no peaks at seasonal frequencies.

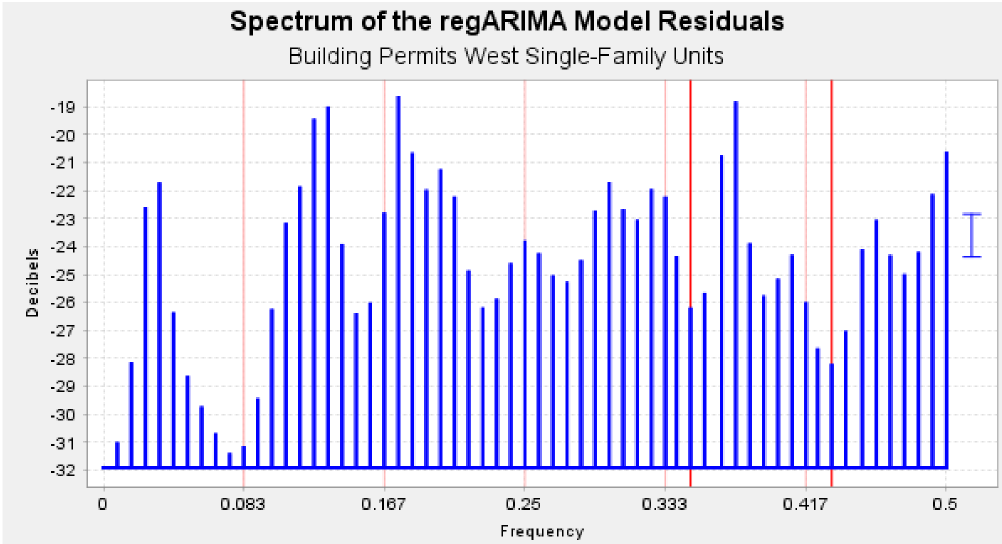

Figure 21: Spectrum of the regARIMA model residuals, from X-13-Graph, for West single-family building permits; note that the plot shows no visually significant peaks and has one nonsignificant peak at frequency 3/12. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

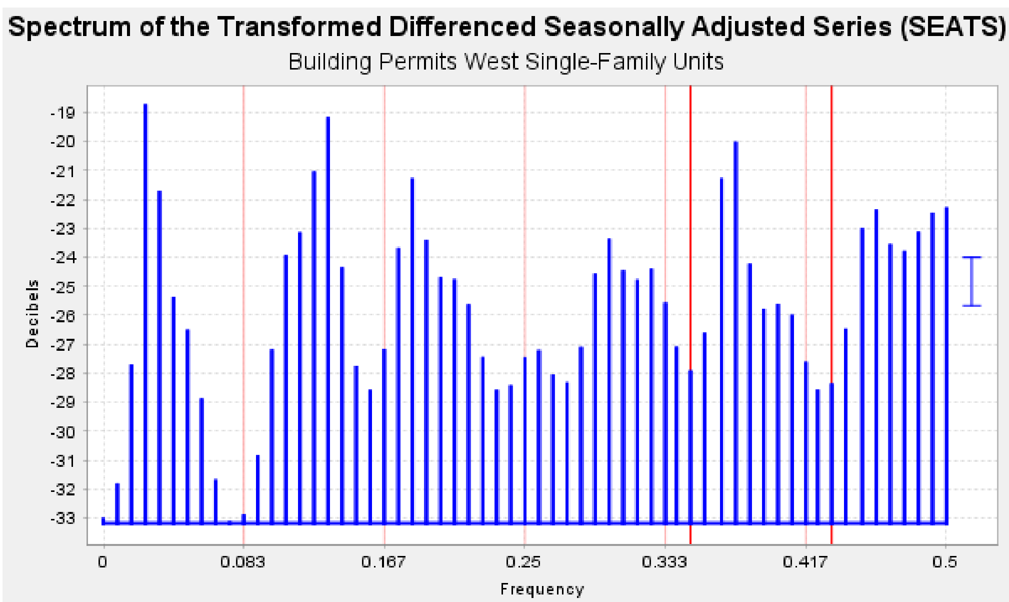

Figure 22: Spectrum of the modified seasonally adjusted series, from X-13-Graph, West single-family building permits; note that the plot shows no visually significant peaks. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

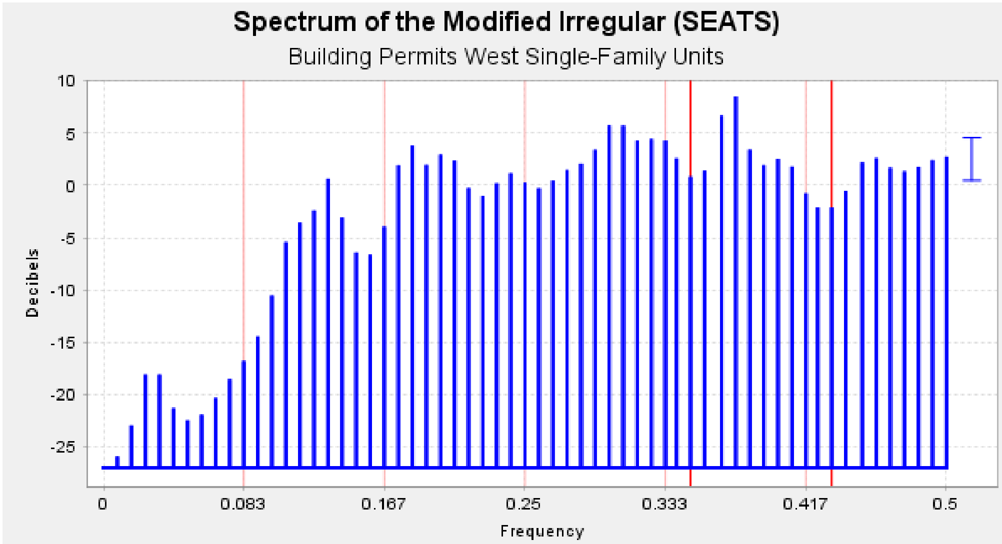

Figure 23: Spectrum of the modified irregular component series, from X-13-Graph, West single-family building permits; note that the plot shows no visually significant peaks.
Source: Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Summary: The spectral plot results provide no evidence of residual seasonality using established significance levels.

The QS statistics and the spectral diagnostics provide no evidence of residual seasonality at established significance levels. If the adjustment meets other established quality measures, it is adequate.

5.3 Example 3: Indirect Seasonal Adjustment

The U.S. Single-Family Building Permits are the sum of the single-family permits for the four U.S. regions: Northeast, Midwest, South, and West. For this example, we assume that the seasonal adjustments of the four regions' series are acceptable, and we aggregated the values to the U.S. single-family total.[9] We used the X-11 method of seasonal adjustment for the Northeast, Midwest, and South time series; we used the SEATS method of seasonal adjustment for the West. Although pretest diagnostics are available for the U.S. total, the primary interest for the aggregate series is in posttesting for the indirect seasonal adjustment.

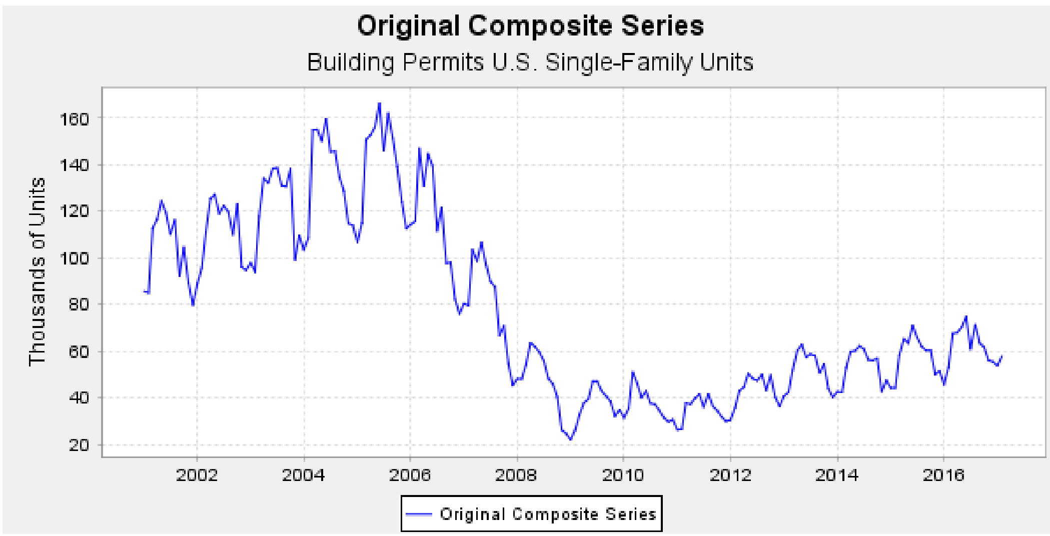

Figure 24: Monthly single-family building permits, U.S. total, original composite series, sum of the original series of the four U.S. regions, January 2001 – February 2017.
Source: Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

[9] These series estimates, from the U.S. Census Bureau's Building Permits Survey are subject to sampling and nonsampling error. More information about the data collection and estimation for the Building Permits Survey is available online (census.gov/construction/bps/).

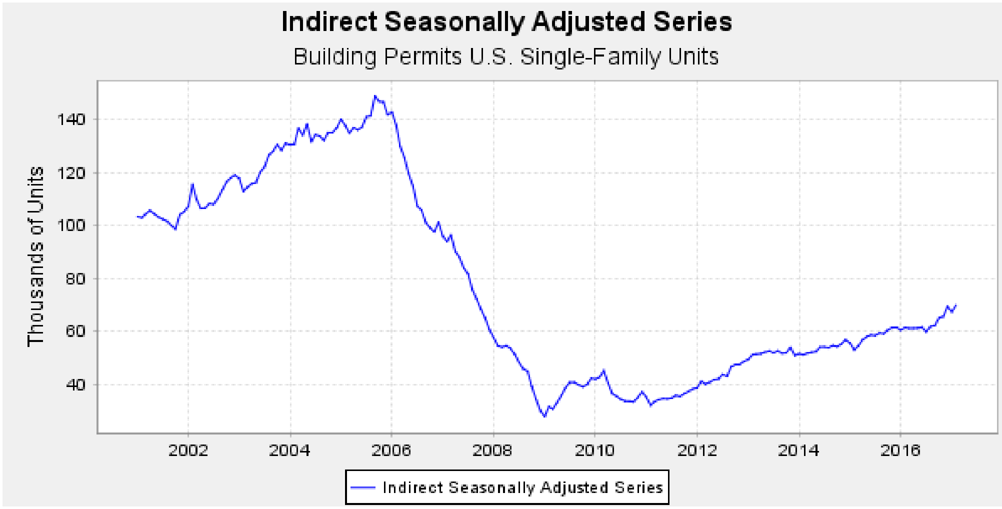

Figure 25: Monthly indirect seasonally adjusted single-family building permits, U.S. total, composite series, sum of the seasonally adjusted series of the four U.S. regions. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

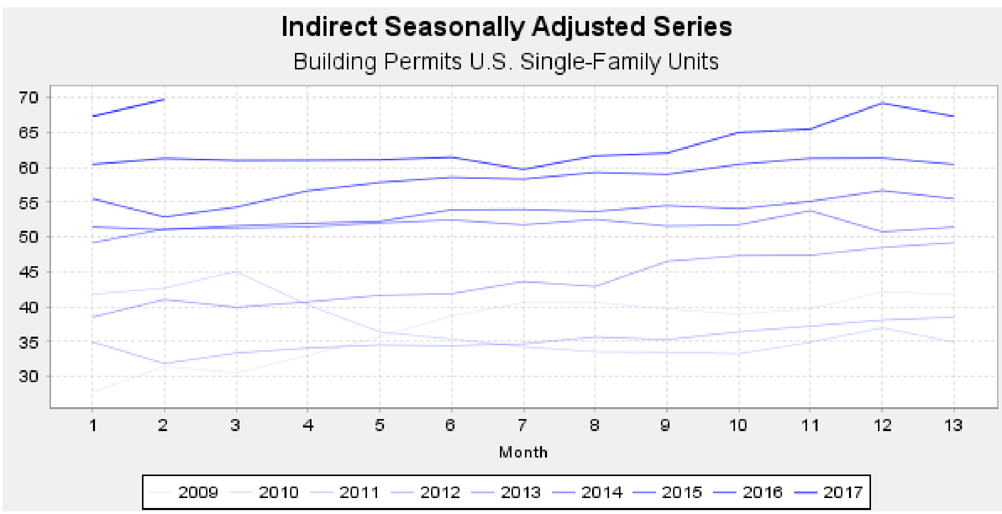

Figure 26: Monthly indirect seasonally adjusted single-family building permits, year over year, U.S. total, composite series, sum of the seasonally adjusted series of the four U.S. regions. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)]

Figure 24 shows the original composite series, the aggregate of the four regions' original series. It appears seasonal. **Figure 25** and **Figure 26** show the indirect (composite) seasonal adjustment, across consecutive months of the full span of adjustment and year over year, respectively. No seasonality is visually apparent from these plots of the seasonally adjusted series.

Posttest: QS diagnostics

We checked the results of the QS statistics for residual seasonality. Users must take care if checking these in the main output file because the first tables of QS diagnostics are for the direct adjustment. The tables for the indirect adjustment are lower in the output file. The *p*-values associated with the QS statistics are also available from the Win X-13 diagnostics tables and from the diagnostics file (file with extension "udg"). The quarterly QS diagnostics from the **`qcheck`** software feature are not available for the indirect seasonal adjustment of monthly time series. In addition, because the indirect adjustment includes a SEATS adjustment, the software does not compute an indirect irregular component, so the only QS diagnostics for the indirect adjustment are for the seasonally adjusted series.

Table 14 shows the QS residual seasonality tests, and for each statistic, the QS statistic is 0 with a corresponding *p*-value of 1, so we fail to reject the null hypothesis that the indirectly adjusted series is not seasonal, i.e., we fail to find evidence of residual seasonality in the indirectly adjusted series.

Table 14: QS Statistics for the test for seasonality (indirect adjustment), for the U.S. single-family building permits full time series and subspan (starting in March 2009). Some tests are for the series adjusted for extreme values; We do not reject the null hypothesis that the series is not seasonal. **Source:** Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)

Table	Span	QS	*P*-Value
Indirect Seasonally Adjusted Series	Full Series	0.00	1.0000
Indirect Seasonally Adjusted Series (extreme value adjusted)	Full Series	0.00	1.0000
Indirect Seasonally Adjusted Series	Subspan	0.00	1.0000
Indirect Seasonally Adjusted Series (extreme value adjusted)	Subspan	0.00	1.0000

Summary: The QS results provide no evidence of residual seasonality in the composite seasonally adjusted series at the .05 statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies

As with the QS statistics, when looking for the spectral diagnostics of the indirectly adjusted series in the main output file, users must make sure they are not looking at the diagnostics for the direct adjustment. A higher resolution graph of the spectrum of the indirect seasonal adjustment is available from Win X-13 and from X-13-Graph. The spectral diagnostics also are available from the Win X-13 diagnostics tables and from the diagnostics file.

Figure 27 shows the spectrum of the differenced indirectly seasonally adjusted series. No indirect irregular component series is available, so there is no corresponding spectrum. Interpret the graph from the indirect adjustments in the same way as for direct adjustments. The graph provides no evidence of residual seasonality in the indirect seasonally adjusted series, although it indicates a residual trading day effect, so the adjustments of the component time series warrant further review.

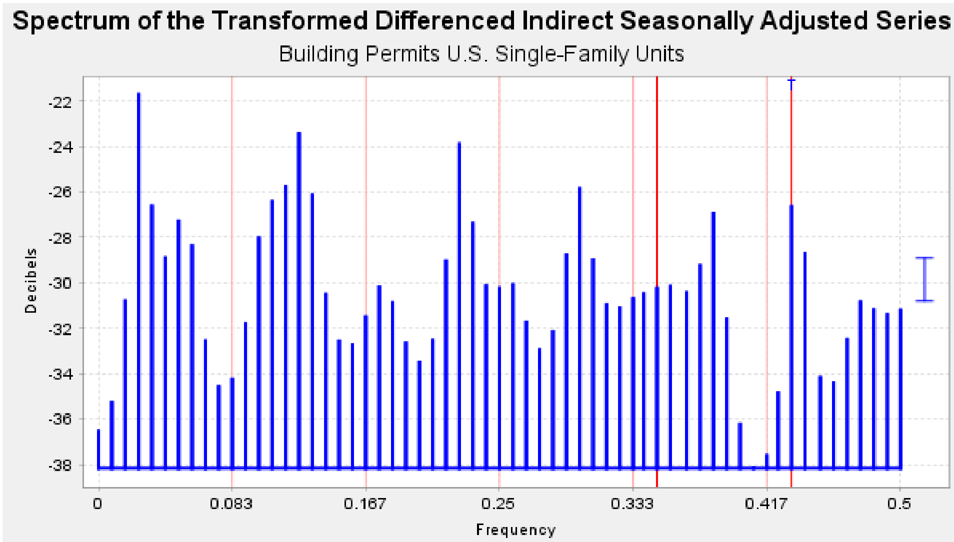

Figure 27: Spectrum of the differenced, transformed, indirect seasonally adjusted U.S. single-family building permits series from X-13-Graph; no peaks occur at seasonal frequencies. **Source:** [Building Permits Survey, U.S. Census Bureau (census.gov/construction/bps/)](census.gov/construction/bps/)

Summary: The spectral plot results provide no evidence of residual seasonality in the composite seasonally adjusted series using established significance levels.

In this example, the posttest diagnostics agreed that the seasonally adjusted series no longer has identifiable seasonality, although the spectrum of the indirect seasonally adjusted series indicates possible residual trading day effects. Disagreement among the diagnostics or agreement in evidence of residual seasonality might lead to reconsidering adjustment settings. For an indirect adjustment, that might mean revisiting each component to determine if other adjustments for one or more of those could lead to a better total adjustment.

Summary: The QS statistics and the spectral diagnostics provide no evidence of residual seasonality in the composite seasonally adjusted series at established significance levels, but the spectral diagnostic indicates the possibility of residual trading day effects. This adjustment warrants further review, although the indirect adjustment appears to have removed seasonal effects.

5.4 Example 4: Direct Seasonal Adjustment, Quarterly

The previous examples looked at monthly series; the same process applies for quarterly series. In this example, the time series is Net Sales, Receipts, and Operating Revenue for Clothing and General Merchandise Stores in the United States, obtained from the Quarterly Financial Report, U.S. Census Bureau.[10] Graphs of the original series in millions of dollars, over the entire time span as in **Figure 28**, or in the year-over-year graph in **Figure 29**, revealed a very consistent seasonal pattern. The graph indicates that activity rises sharply in the fourth quarter and exhibits a secondary, less pronounced increase in the second quarter. The graphs suggested that this series is strongly seasonal, but we checked the seasonality diagnostics to confirm the visual impression. Note that since the series is quarterly, X-13ARIMA-SEATS will not produce a spectrum, so we have no visual significance diagnostic.

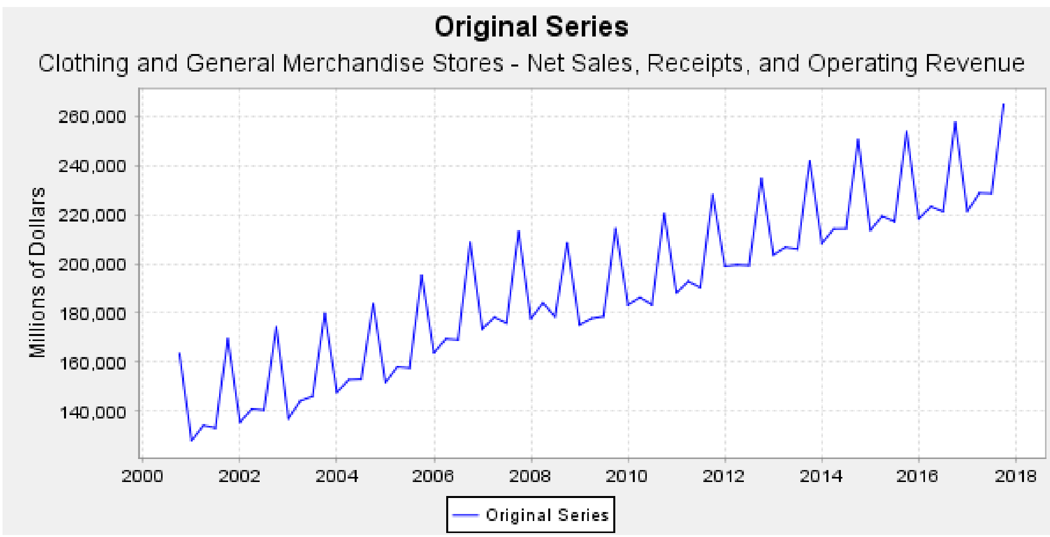

Figure 28: Net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total (in millions of dollars), 2000 quarter 4 – 2017 quarter 4. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

[10] These estimates are subject to sampling and nonsampling error. They are not adjusted for price changes. More information about the data collection process and estimation for the Quarterly Financial Report is available online (census.gov/econ/qfr/collection.html).

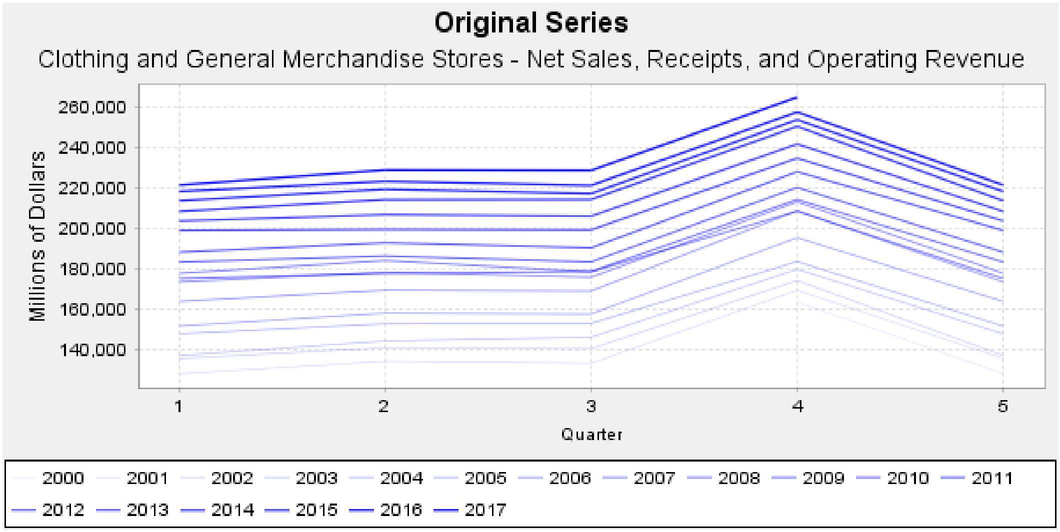

Figure 29: Net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total (in millions of dollars), original series year over year. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

Pretest: autocorrelation function plots

We obtained an informal examination of seasonality by examining the sample autocorrelation plots. Because the series has a strong upward trend, seasonal effects are somewhat masked in the plot of **Figure 30**, although we did find that at lag 4 the autocorrelation is stronger than at neighboring lags. After seasonal differencing of the series, all evidence of seasonality has been annihilated, as seen in **Figure 31**, because at lag 4 the values are close to zero. If instead we only apply a trend difference, then some seasonality remains, as indicated in **Figure 32**, with large autocorrelation at lags 4 and 8. Finally, applying both a regular and seasonal difference induces a substantial form of antiseasonality, (a negative seasonal relationship, exhibited by the negative correlation at lags 4 and 8, see **Figure 33**). All these results are consistent with a strong, dynamic, seasonal effect's presence in the original data.

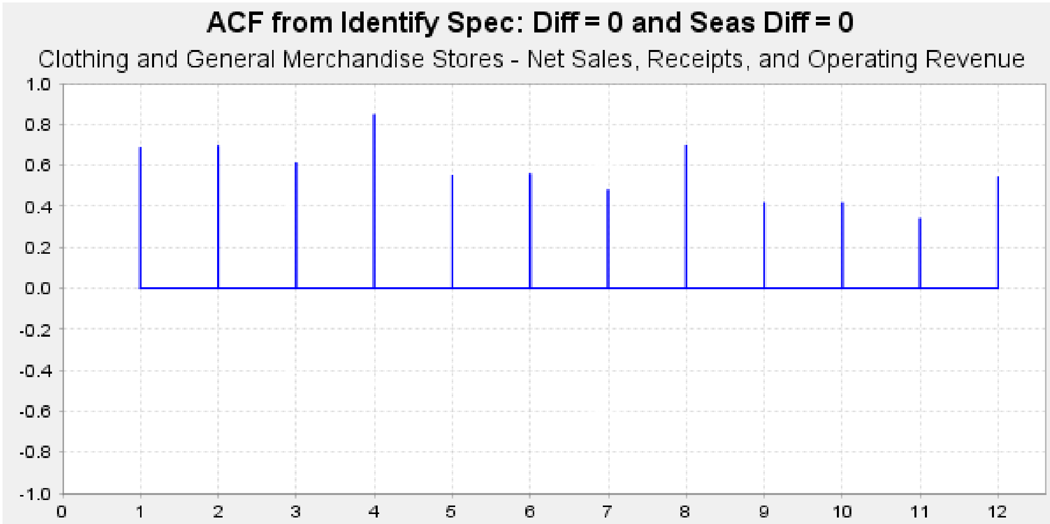

Figure 30: Autocorrelation function of net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total with no differencing. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

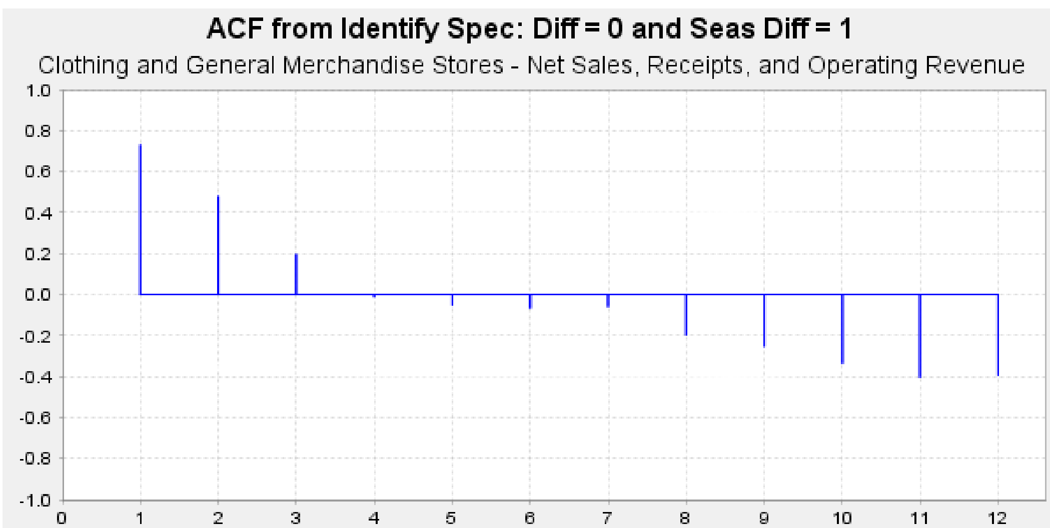

Figure 31: Autocorrelation function of net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total with no first difference and one seasonal difference. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

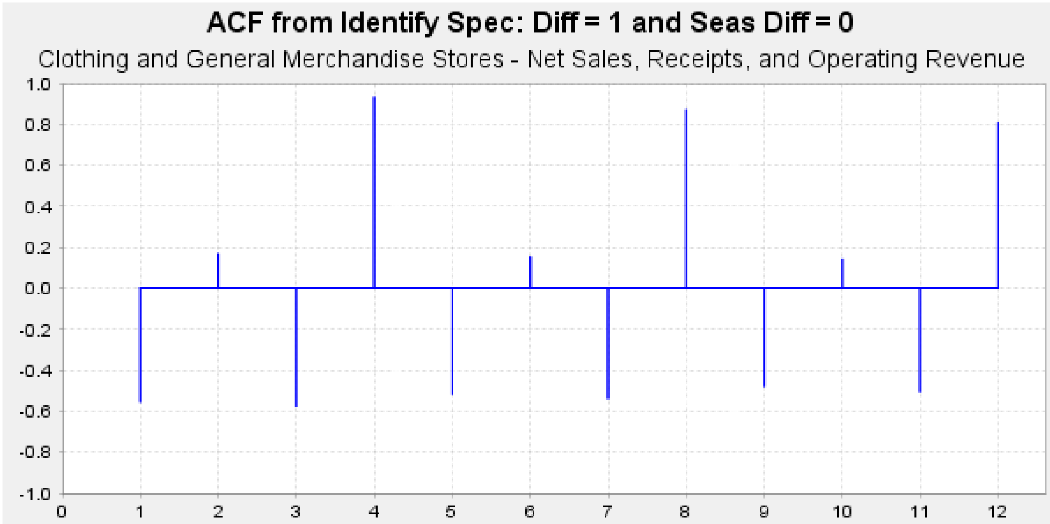

Figure 32: Autocorrelation function of net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total with one first difference and no seasonal difference. **Source:** [Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)](census.gov/econ/qfr/)

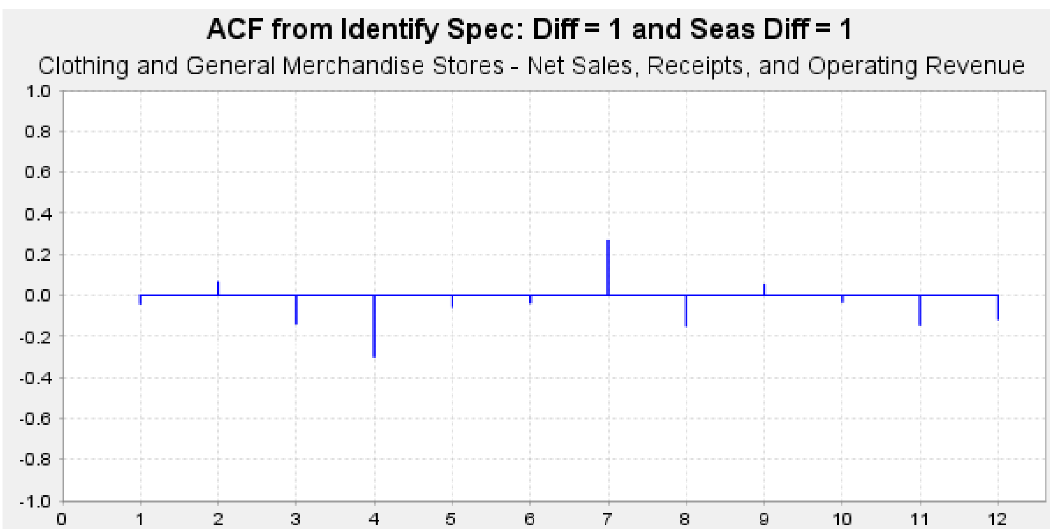

Figure 33: Autocorrelation function of net sales, receipts, and operating revenue for clothing/general merchandise stores, U.S. total with one first difference and one seasonal difference. **Source:** [Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)](census.gov/econ/qfr/)

Pretest: regression model-based F test

For the pretest, we fit the regression to the full span of the series that is subject to seasonal adjustment, fitting a regARIMA model with the seasonal regressors.

An initial pass through X-13ARIMA-SEATS suggested that the series did not require a transformation. To perform the seasonal testing, we ran automatic modeling with regressors for fixed seasonal effects. The automatic ARIMA model choice was (0 1 1), and the trend constant regressor was significant, with t=7.36.

Example spec file for the seasonal pretest:

```
series{
 file="qfr_clothing_net_sales.txt" format=”datevalue”
 period=4
 span=(2000.4,)
}
spectrum{start=2010.1}
transform{function=none}
regression{
 variables=(const seasonal)
}
outlier{types=(AO LS)}
automdl{maxorder=(2 1) savelog=amd}
estimate{print=(roots regcmatrix acm) savelog=all}
check{print=all savelog=all}
```

Table 15 shows the F test and *p*-value for the group of seasonal regressors fit to the full series from 2000.4 to 2017.4. The seasonal regressor group was highly significant at the 0.01 level (the one percent critical value is just 4.1). We rejected the null hypothesis, which was that the series is not seasonal.

Table 15: Degrees of freedom, F statistic, and *p*-value for the group of three seasonal regressors when fit with automatic ARIMA modeling (and the trend constant, which was significant); net sales, receipts, and operating revenue for clothing/general merchandise stores. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

Regressor	DF	F Statistic	*P*-Value
Seasonal	3, 64	2,666.83	0.00

Summary: The regression model-based F test provides evidence that the original series is seasonal.

The F test and the autocorrelation function plots provide evidence that the original series is seasonal, and hence it can be considered for seasonal adjustment.

Posttest: QS diagnostics

Check the results of the QS statistics for seasonality from the main output file (Table 16 below shows the posttest QS statistics) or from the diagnostics tables of Win X-13.

Recall that the test for the regARIMA model residuals is helpful for determining model adequacy. This test is important for model-based adjustments, such as SEATS, but is less relevant for X-11 adjustments.

For quarterly time series, under default settings, the software provides QS diagnostics for a shortened span only if the series is at least 24 years long (96 quarters). Because the example time series is only 17 years long, by default, the software provides QS tests only for the full series. Users can set a shortened span for testing with the **`start`** argument of the **`spectrum`** spec. Although the spectrum is not produced for quarterly series, this setting is available for the purpose of checking QS statistics. We suggest checking for residual seasonality in (1) the seasonally adjusted series, (2) seasonally adjusted series adjusted for extreme values, (3) irregular series, and (4) irregular series adjusted for extreme values. **Table 16** indicates that these QS statistics are close to zero, with correspondingly high *p*-values. Hence, we do not reject the null hypothesis that the adjusted series has no residual seasonality. The tests were similarly nonsignificant for a shortened span from 2010 through 2017.

Table 16: QS statistics for the test for seasonality in the seasonally adjusted series and modified irregular component, for the full series; net sales, receipts, and operating revenue for clothing/general merchandise stores. Some tests are for the series adjusted for extreme values. **Source:** Quarterly Financial Report, U.S. Census Bureau (census.gov/econ/qfr/)

Table	QS	*P*-Value
Seasonally Adjusted Series	0.19	0.9096
Seasonally Adjusted Series (extreme value adjusted)	0.02	0.9889
Irregular Series	0.52	0.7717
Irregular Series (extreme value adjusted)	0.24	0.8877

Summary: The QS results provide no evidence of residual seasonality at the .05 statistical test level. If the adjustment meets other established quality measures, it is adequate.

5.5 Example 5: Inadequate Direct Seasonal Adjustment With Stable Seasonal Filters

This example series is monthly U.S. Retail Sales of General Merchandise Stores in millions of dollars.[11] Graphs of the original series, over the entire time span as in **Figure 34**, or in the year-over-year graph as in **Figure 35**, reveal a consistent seasonal pattern. While the overall level of the series increases over time, activity spikes appear obvious in Decembers, as are low values in Januaries and Februaries. The series seems strongly seasonal, but we will check the seasonality diagnostics for confirmation.

For the initial run, as an experiment, we performed an X-11 seasonal adjustment with stable seasonal filters. If the series has an evolving seasonal pattern, this seasonal adjustment might exhibit residual seasonality. We looked for that in the posttest.

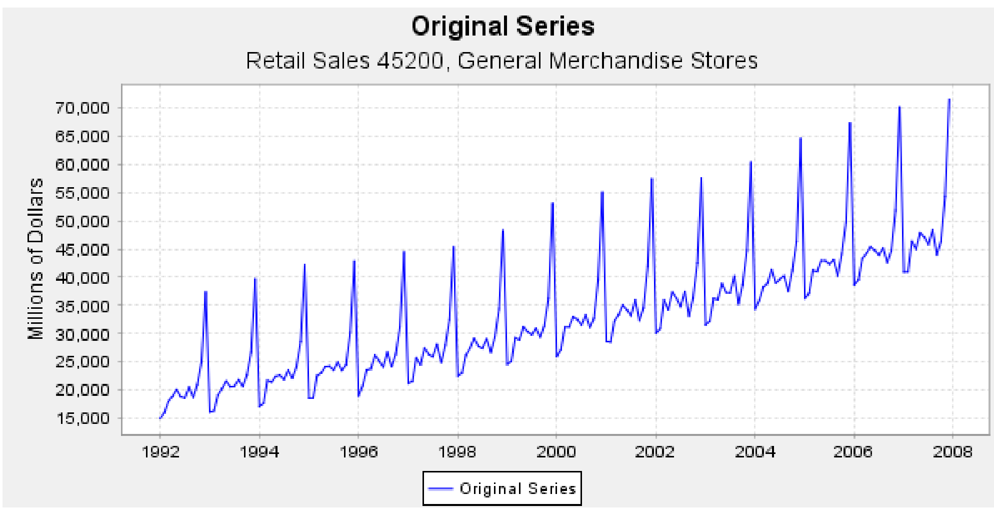

Figure 34: Monthly retail sales of general merchandise stores, original series (in millions of dollars), 1992–2007. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

[11] The estimates are subject to sampling and nonsampling error. They are not adjusted for price changes. More information about the data collection and estimation methods for the Monthly Retail Trade and Food Services Survey is available online (census.gov/retail/how_surveys_are_collected.html).

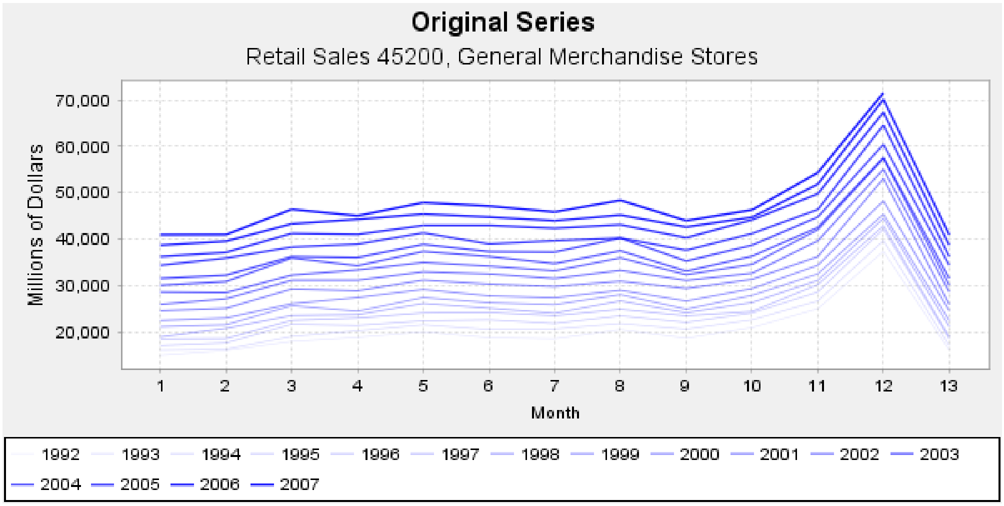

Figure 35: Monthly retail sales of general merchandise stores, original series (in millions of dollars), year over year, 1992–2007. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Pretest: autocorrelation function plots

For completeness, we looked at the autocorrelation function plots, and our initial impressions regarding the strong seasonality are confirmed. The sample autocorrelations of the time series (with no differencing at all) in **Figure 36** shows a slow decay, along with some peaks at seasonal lags, which is consistent with strong trend growth and dynamic seasonality. A single regular difference reduces the slow decay in autocorrelations, unmasking the seasonality – see the large peaks in **Figure 38**. If instead a seasonal difference is used, as **Figure 37** shows, both trend and seasonality have been annihilated. Finally, this story is confirmed by **Figure 39**, where the application of both a regular and a seasonal difference has generated negative correlation at lags 1 and 12.

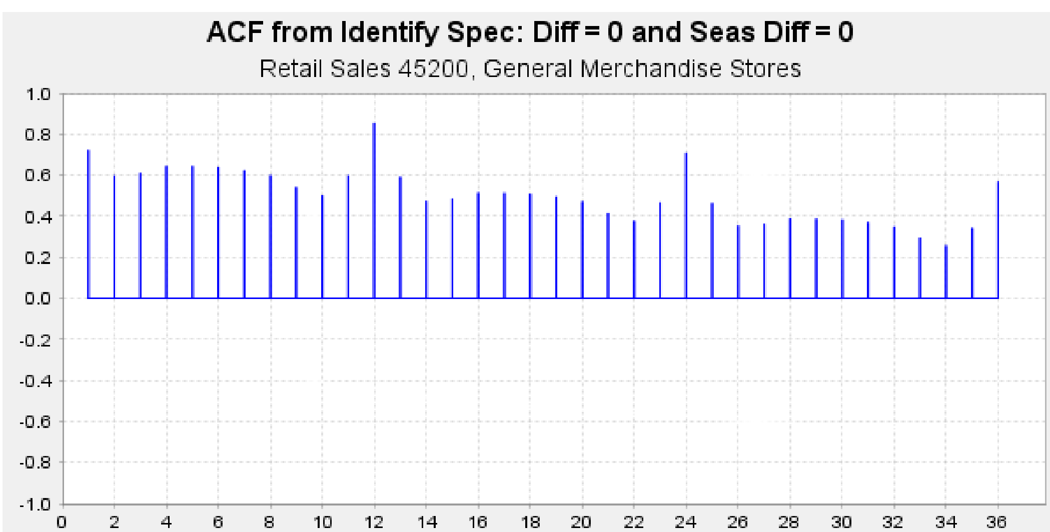

Figure 36: Autocorrelation function of retail sales of general merchandise stores, U.S. total, with no differencing. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

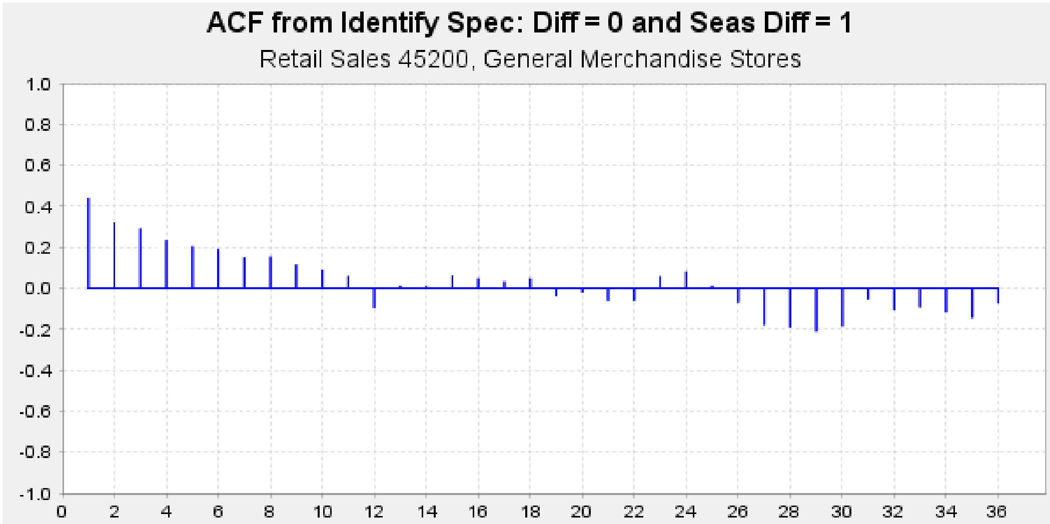

Figure 37: Autocorrelation function of retail sales of general merchandise stores, U.S. total, with no first difference and one seasonal difference. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

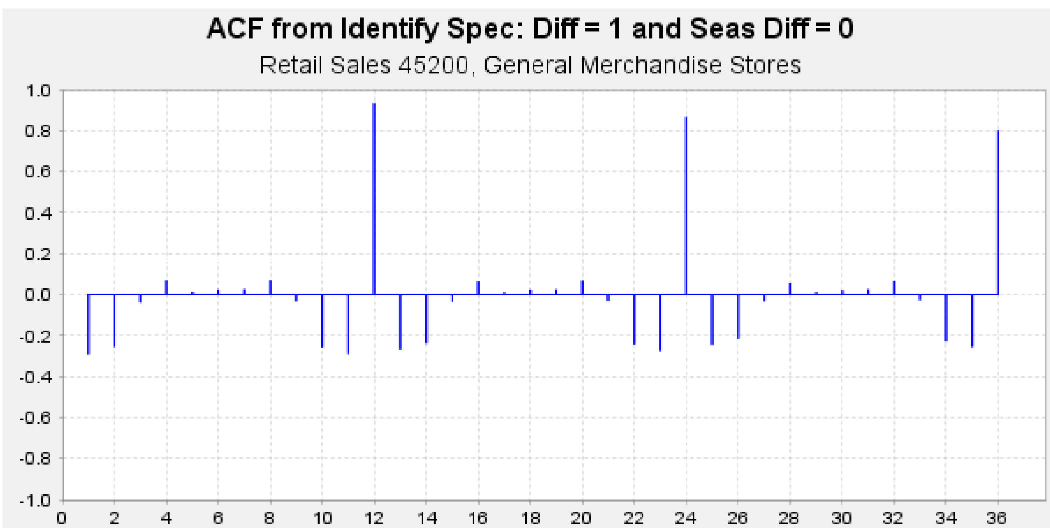

Figure 38: Autocorrelation function of retail sales of general merchandise stores, U.S. total, with one first difference and no seasonal difference. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

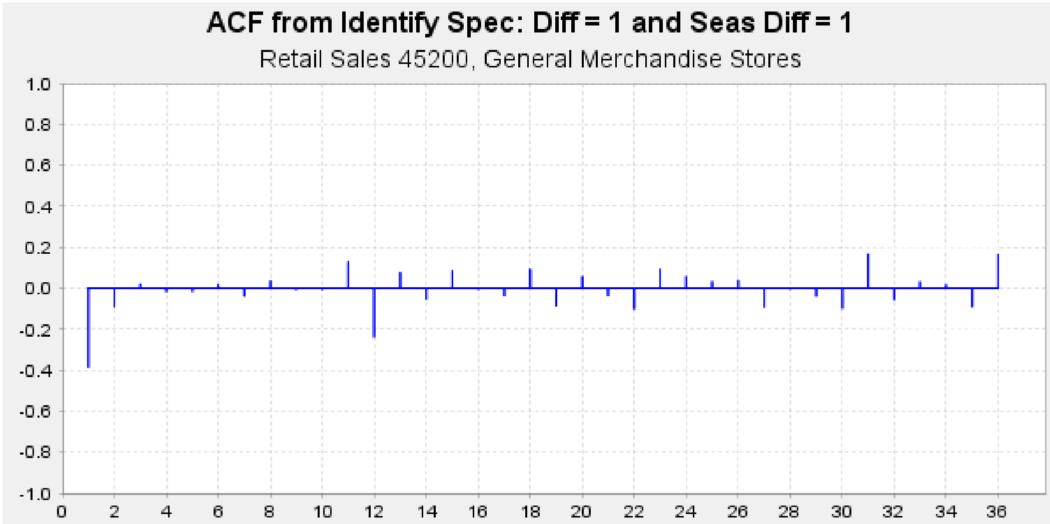

Figure 39: Autocorrelation function of retail sales of general merchandise stores, U.S. total, with one first difference and one seasonal difference. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

Pretest: regression model-based F test

For the pretest, we fit the regression to the full span of the series that is subject to seasonal adjustment and fit a regARIMA model with the seasonal regressors.

An initial pass through X-13ARIMA-SEATS suggested that the series requires no transformation, but closer review revealed that the values of Akaike's Information Criteria Corrected or Sample Size (AICC) for no transformation and for the log transformation were similar, and without a transformation, recent December values were identified as outliers or extreme values. This result indicated that a log transformation was better for the purpose of seasonal adjustment. Even though the default setting for the automatic test has a slight preference toward the log transformation, when the result is close, it is always good to look more closely when deciding between the log transformation and no transformation.

To determine if the series is seasonal, we ran the automatic modeling procedure with regressors for fixed seasonal effects. The automatic model choice was ARIMA (2 1 0)(1 0 0), with six-coefficient trading day regression and Easter regression effect with an eight-day window.

Example spec file for the seasonal pretest:

```
series {
 title="Retail Sales 45200, General Merchandise Stores - Pretest"
 file='sales452.dat' format='datevalue'
}
spectrum {qcheck=yes savelog=all}
transform {function=log}
identify{diff=(0 1) sdiff=(0)}
regression {
 variables=(const seasonal)
```

```
 aictest=( td easter ) savelog=aictest
}
outlier{types=(AO LS TC) lsrun=3}
automdl{maxorder=(3 1) savelog=amd}
# comment out the model to run automdl instead
#arima{model=(0 1 1)(1 1 0)}
estimate {print=(roots regcmatrix acm) savelog=all}
check {print=all savelog=all}
```

Table 17 shows the F test and *p*-value for the group of seasonal regressors fit to the full series from January 1992 to December 2007. The group of seasonal regressors was highly significant at the 0.01 level (the one percent critical value being just 2.35). We rejected the null hypothesis, which was that the series is not seasonal.

Table 17: Degrees of freedom, F statistic, and *p*-value for the group of 11 seasonal regressors fit to monthly sales of general merchandise store sales. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Regressor	DF	F Statistic	*P*-Value
Seasonal	11, 173	38.40	0.00

Summary: The regression model-based F test provides evidence that the original series is seasonal.

Pretest: spectral plot peaks at seasonal frequencies

Because this is a monthly series, spectral plots were available. We generated the spectral plots from a run with automatic model selection. The resulting regARIMA model was ARIMA (0 1 1)(1 1 0) with a six-day trading day regression and an Easter regressor with a 15-day window. The Easter effect differed from the model-based F test, but for purposes of identifying seasonality, this is not a problem. For this example series, the peaks at 2/12, 3/12, 4/12, and 5/12 were obviously above the median level of all frequencies of the plot and so much taller than their nearest neighboring frequencies that their visual significance was apparent without a count of the star units. Note, however, that **Figure 40**, from X-13-Graph, shows a bold **S** at 2/12, 4/12, 4/12, and 5/12, indicating that the peaks at those frequencies are visually significant.

Win X-13 lists S2, S3, S4, and S5 under **Sig Ori Peaks** (Significant Peaks for the Prior-Adjusted Original Series), providing the same information apparent in the graph but condensing it for easy capture. **Table 18** shows this snippet from the Win X-13 tables.

Table 18: From Win X-13, the list of visually significant peaks at seasonal frequencies of the prior-adjusted original series; for monthly retail sales of general merchandise stores. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Series Name	Sig Ori Peaks
Sales452: Retail Sales, General Merchandise Stores	s2 s3 s4 s5

Summary: The spectral plot provides evidence that the original series is seasonal.

Both of the more formal pretests and the autocorrelation function plots provide evidence that the original series is seasonal, and hence it can be considered for seasonal adjustment.

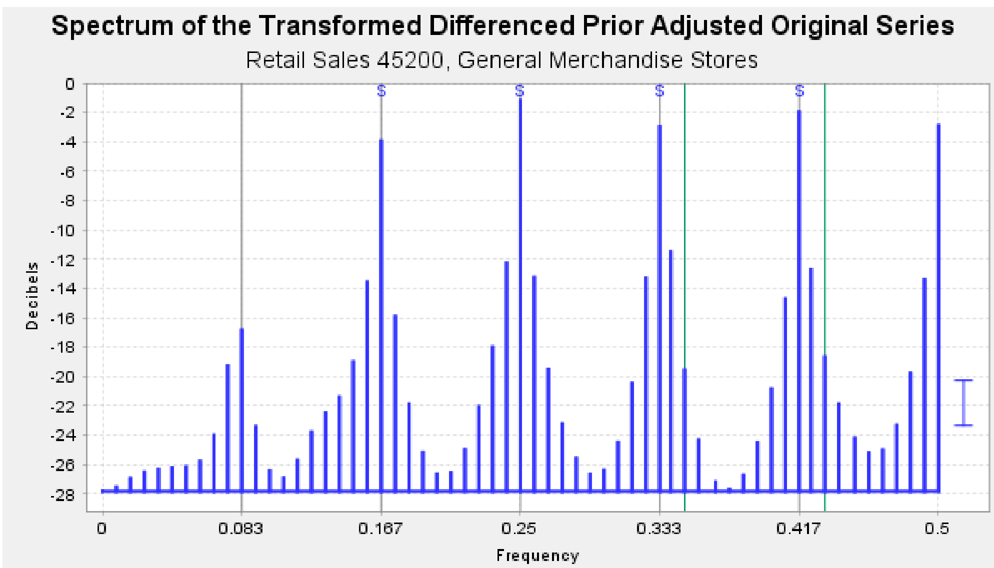

Figure 40: Monthly retail sales of general merchandise stores; spectrum of the differenced, transformed prior-adjusted series (Table B1) with visually significant peaks at seasonal frequencies 2/12, 3/12, 4/12, and 5/12. The peak at 1/12 does not meet the criteria of visual significance; it is above the median level but not tall enough compared to the taller of its nearest neighboring frequencies. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Note: For this example, we seasonally adjusted the time series using the X-11 method with stable seasonal filters. In other words, the seasonal factors differ from one month to another but are the same across years for a given month. That is, all Januaries have the same factor; all Februaries have the same factor, etc. Most time series do not have a **perfectly** repeating seasonal pattern, so this approach to the adjustment is rarely adequate.

Posttest: QS diagnostics

We checked the results of the QS statistics for residual seasonality. That meant checking the full series and shortened series for residual seasonality in the seasonally adjusted series, seasonally adjusted series adjusted for extreme values, irregular series, and irregular series adjusted for extreme values.

Table 19 indicates that regardless of the test length, the QS statistics for the seasonally adjusted series and modified irregular are high, with *p*-values equal to 0.0000. Each of the test statistics called for rejecting the null hypothesis that the adjusted series has no seasonality, so the adjustment requires more analysis.

Table 19: QS statistics for the test for seasonality in the model residuals and in the seasonally adjusted series, for the full series and subspan (from January of 2000), for monthly retail sales of general merchandise stores. Some tests are for the series adjusted for extreme values. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Table	Span	QS	*P*-Value
Residuals	Full Series	3.71	0.1568
Seasonally Adjusted Series	Full Series	138.00	0.0000
Seasonally Adjusted Series (extreme value adjusted)	Full Series	54.85	0.0000
Irregular Series	Full Series	187.72	0.0000
Irregular Series (extreme value adjusted)	Full Series	94.74	0.0000
Residuals	Subspan	0.58	0.7478
Seasonally Adjusted Series	Subspan	56.24	0.0000
Seasonally Adjusted Series (extreme value adjusted)	Subspan	29.12	0.0000
Irregular Series	Subspan	81.53	0.0000
Irregular Series (extreme value adjusted)	Subspan	38.61	0.0000

Summary: The QS results provide evidence of residual seasonality at the .05 statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies

For this adjustment, check the spectral diagnostics from one of the various sources.

Table 20 shows spectral peak information from Win X-13 that indicates the spectra of the seasonally adjusted series and the modified irregular component each have a peak (or peaks) at seasonal frequencies that are visually significant. Additionally, the spectra of the modified irregular component and of the model residuals each have a nonsignificant peak at seasonal frequencies. The abbreviated diagnostic information does not indicate which frequencies have the nonsignificant peaks and does not indicate which spectrum or peak has the height of 1.5, but users can check the spectral graphs for more information about the peak heights and at which frequencies the peaks occur.

Table 20: Spectral peak information for monthly retail sales of general merchandise stores from Win X-13; peaks occurred in the spectra of the seasonally adjusted series and the irregular component, but they are not visually significant. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

Series Name	Sig SAdj Peaks	Sig Irr Peaks	Nonsig Seasonal Peaks
General Merchandise Store Sales	**s1 s2**	**s2**	irr rsd [1.5]

For this example series, **Figure 41**, the spectrum of the seasonally adjusted series, and **Figure 42**, the spectrum of the modified irregular component, both show a visually significant peak at the 2/12 frequency, indicating residual seasonality, as the Win X-13 abbreviated information had indicated. The same visual information is available from the text graphs in the main output file.

Summary: The spectral plot results provide evidence of residual seasonality using established significance levels.

The QS statistics and spectral diagnostics provide evidence of residual seasonality. Multiple tests that indicate the same result strengthens the diagnosis.

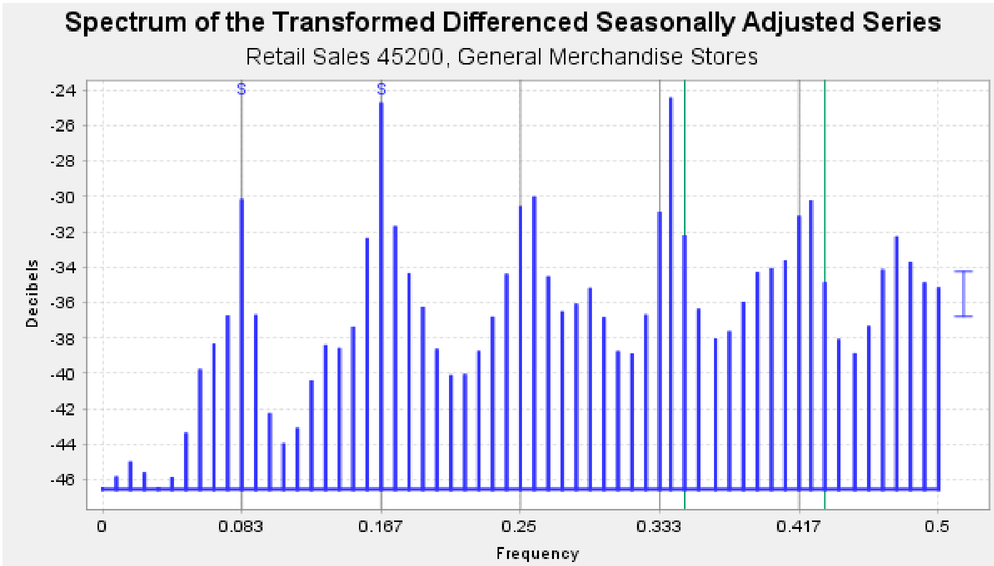

Figure 41: Spectrum of the differenced, transformed seasonally adjusted series of monthly retail sales of general merchandise stores, from X-13-Graph; visually significant peaks occur at seasonal frequencies 1/12 and 2/12 and have bold **S** indicators at the peaks, indicating residual seasonality. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

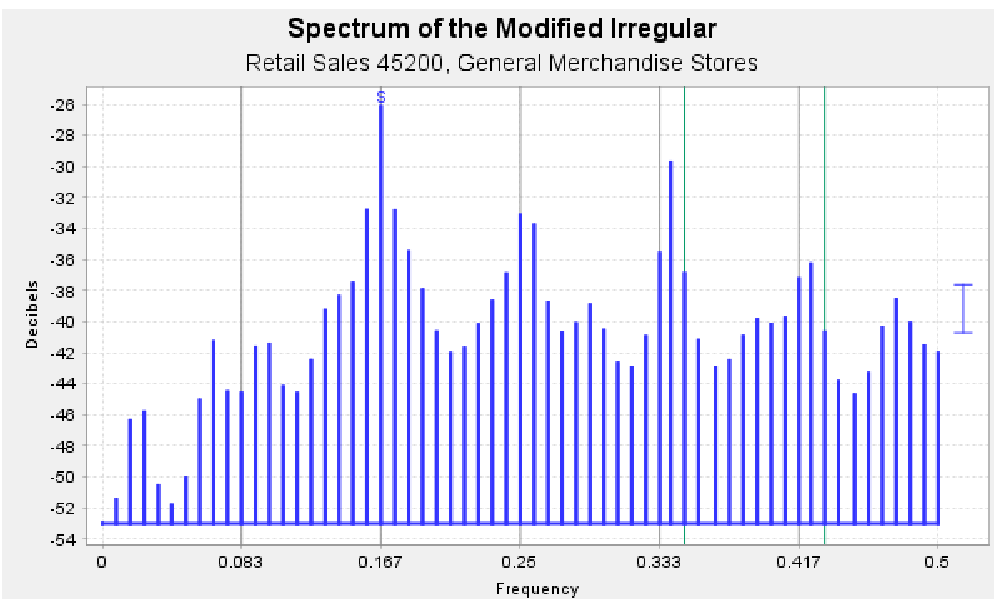

Figure 42: Spectrum of the modified irregular component series of monthly retail sales of general merchandise stores, from X-13-Graph; the peak at seasonal frequency 2/12 is visually significant and has a bold **S** at the peak, indicating residual seasonality. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

We revisited the seasonal adjustment using more appropriate filter choices with the X-11 method of seasonal adjustment and examined results from the SEATS method.

Posttest: QS diagnostics (X-11 Method of Seasonal Adjustment)

For the X-11 method of seasonal adjustment, we used a relatively short 3×3 filter for all months.

All eight QS diagnostics (for seasonally adjusted series and irregular component series, with and without the extreme value adjustment, and for the shortened test span) were 0, and so had *p*-values of 1.

Summary (X-11 method of seasonal adjustment): The QS results provide no evidence of residual seasonality at the .05 (or .01) statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies (X-11 Method of Seasonal Adjustment)

With our updated seasonal adjustment, using 3×3 seasonal filters, the spectral plots, shown in **Figure 43** and **Figure 44**, show no peaks at seasonal frequencies.

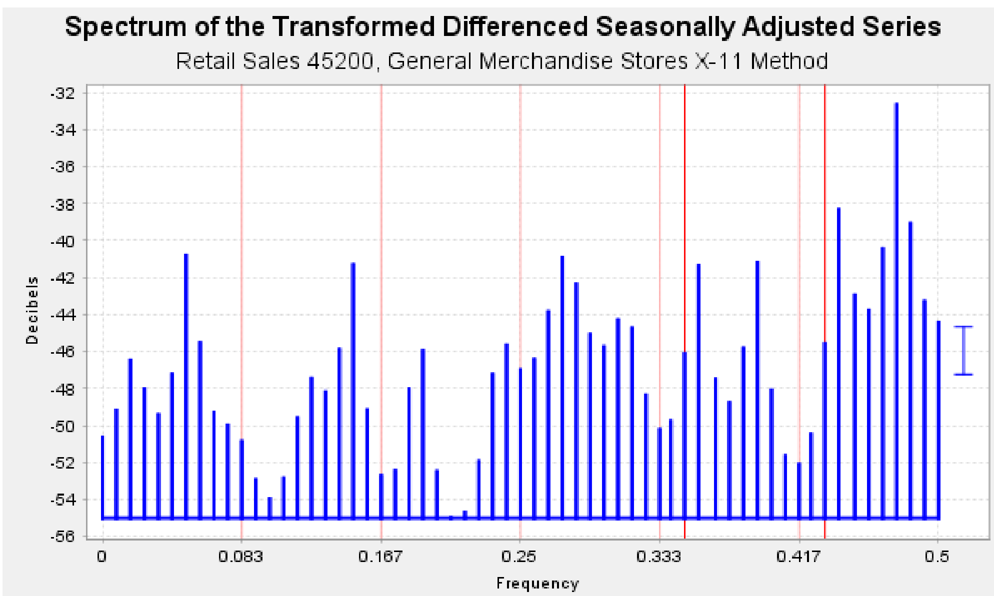

Figure 43: Spectrum of the differenced, transformed seasonally adjusted series of monthly retail sales of general merchandise stores from the revisited adjustment with the X-11 method, from X-13-Graph; no visually significant peaks occur at seasonal frequencies. **Source:** Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)

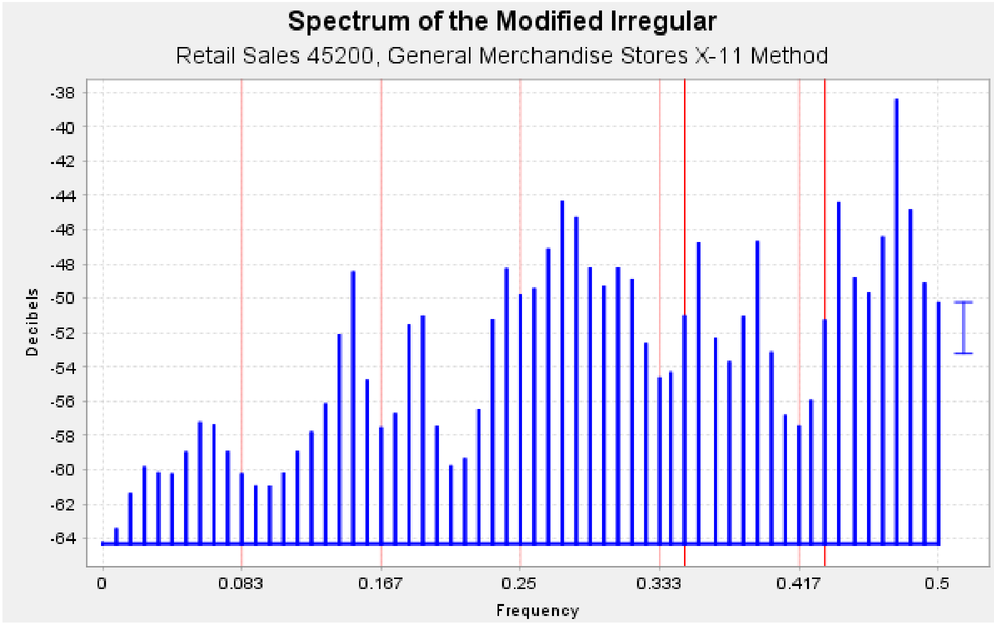

Figure 44: Spectrum of the modified irregular component series of monthly retail sales of general merchandise stores from the revisited adjustment with the X-11 method, from X-13-Graph; no visually significant peaks occur at seasonal frequencies.
Source: [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)](census.gov/retail/)

Summary (X-11 method of seasonal adjustment): The spectral plot results provide no evidence of residual seasonality using established significance levels.

The QS statistics and the spectral diagnostics provide no evidence of residual seasonality at established significance levels. If the adjustment meets other established quality measures, it is adequate.

Posttest: QS diagnostics (SEATS Method of Seasonal Adjustment)

With the SEATS method of seasonal adjustment, just as with the X-11 method, all eight QS diagnostics (for seasonally adjusted series and irregular component series, with and without the extreme value adjustment, and for the shortened test span) were 0, and thus had *p*-values of 1.

Summary (SEATS method of seasonal adjustment): The QS results provide no evidence of residual seasonality at the .05 (or .01) statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies (SEATS Method of Seasonal Adjustment)

The revisited seasonal adjustment using the SEATS method also results in spectral plots, shown in **Figure 45** and **Figure 46**, with no peaks at seasonal frequencies.

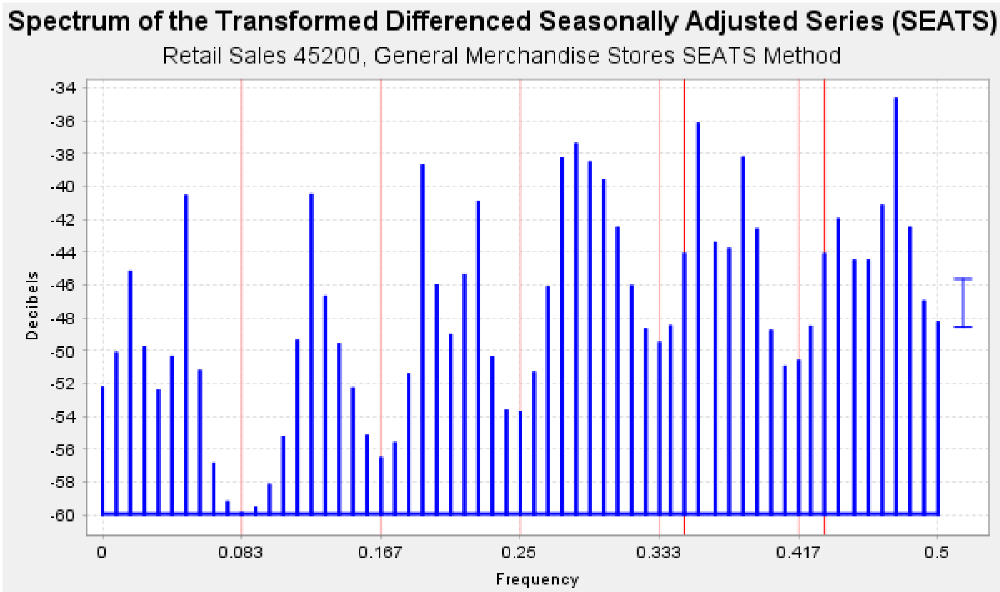

Figure 45: Spectrum of the differenced, transformed seasonally adjusted series of monthly retail sales of general merchandise stores from the revisited adjustment with the SEATS method, from X-13-Graph; no visually significant peaks occur at seasonal frequencies. **Source:** [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

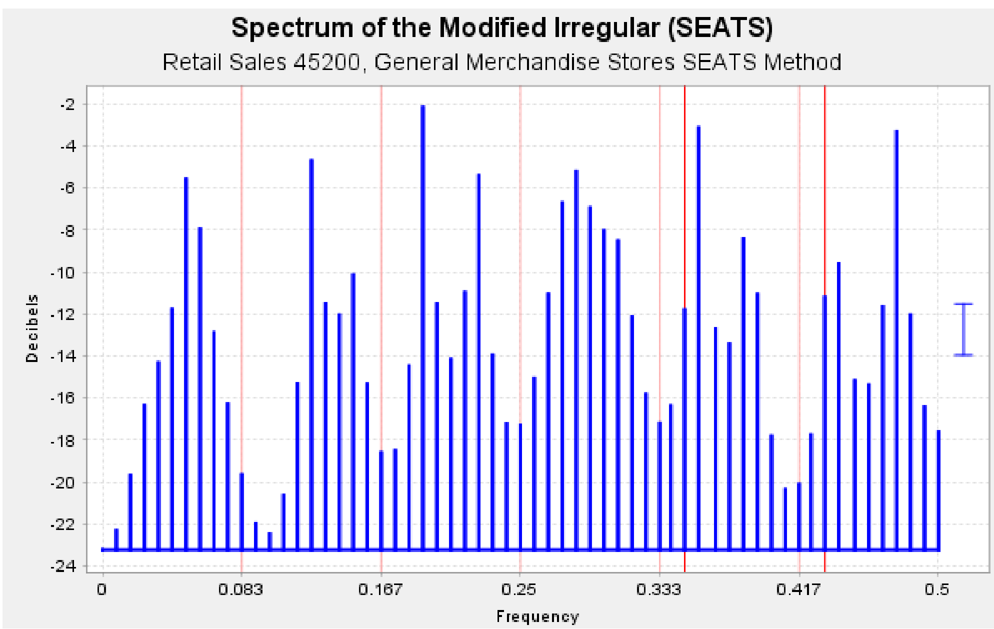

Figure 46: Spectrum of the modified irregular component series of monthly retail sales of general merchandise stores from the revisited adjustment with the SEATS method, from X-13-Graph; no visually significant peaks occur at seasonal frequencies.
Source: [Monthly Retail Trade and Food Services, U.S. Census Bureau (census.gov/retail/)]

Summary (SEATS method of seasonal adjustment): The spectral plot results provide no evidence of residual seasonality using established significance levels.

The QS statistics and the spectral diagnostics provide no evidence of residual seasonality at established significance levels. If the adjustment meets other established quality measures, it is adequate.

5.6 Example 6: Inadequate Indirect Seasonal Adjustment

This example shows a seasonally adjusted series whose adjustment may be inadequate according to some diagnostics; that is, we see evidence of residual seasonality in the indirectly adjusted time series. The series is the seasonally adjusted National Defense component of quarterly gross domestic product (GDP), in chained 2009 billions of dollars, published before the Bureau of Economic Analysis (BEA) performed their comprehensive update, which revised previous seasonally adjusted publications.[12]

This time series warranted review because BEA had concerns that it was exhibiting residual seasonality. The series is a composite of time series that BEA collects. Most, if not all, of the components are already seasonally adjusted when received. We reviewed a subspan of the time series, from 1980 – 2015.

Note: Because we only have the seasonally adjusted time series, we treated it as the **original series** to check for residual seasonality. Even though our review of the series occurs in a posttesting phase, we employed pretesting methods because we only have access to the seasonally adjusted data.

The GDP National Defense seasonally adjusted time series did not show a strong seasonal pattern when viewed over consecutive quarters, as in **Figure 47**. When we graphed all years together in the year-over-year graph (log scale), as in **Figure 48**, we still did not see a consistent seasonal pattern, but it is hard to discern series behavior with so many years in the graph. The subspan from 2008 – 2015, shown in **Figure 49**, did show a possible seasonal pattern, as decreases appear to occur often from the third to fourth quarter, and from the fourth to first quarter.

We checked the available diagnostics, including the autocorrelation function and QS statistic(s). With automatic modeling, X-13ARIMA-SEATS selected ARIMA (1 1 2)(0 1 1). The seasonal MA coefficient was 0.99980.

[12] The time series is available from first quarter 1999 through first quarter 2018 from the BEA data archive (apps.bea.gov/histdata/). The time series we reviewed is from the National Accounts (NIPA) links, third vintage of the first quarter 2018, published June-29-2018.

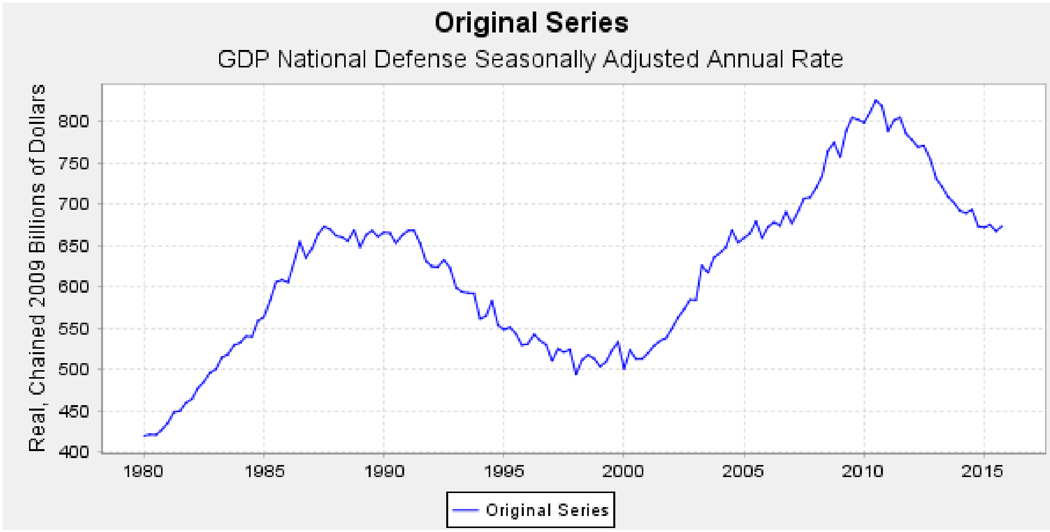

Figure 47: GDP national defense, seasonally adjusted annual rate (in real, chained 2009 billions of dollars), labeled as "original series" because the available estimates are seasonally adjusted, first quarter 1980 – fourth quarter 2015. **Source:** [Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)]

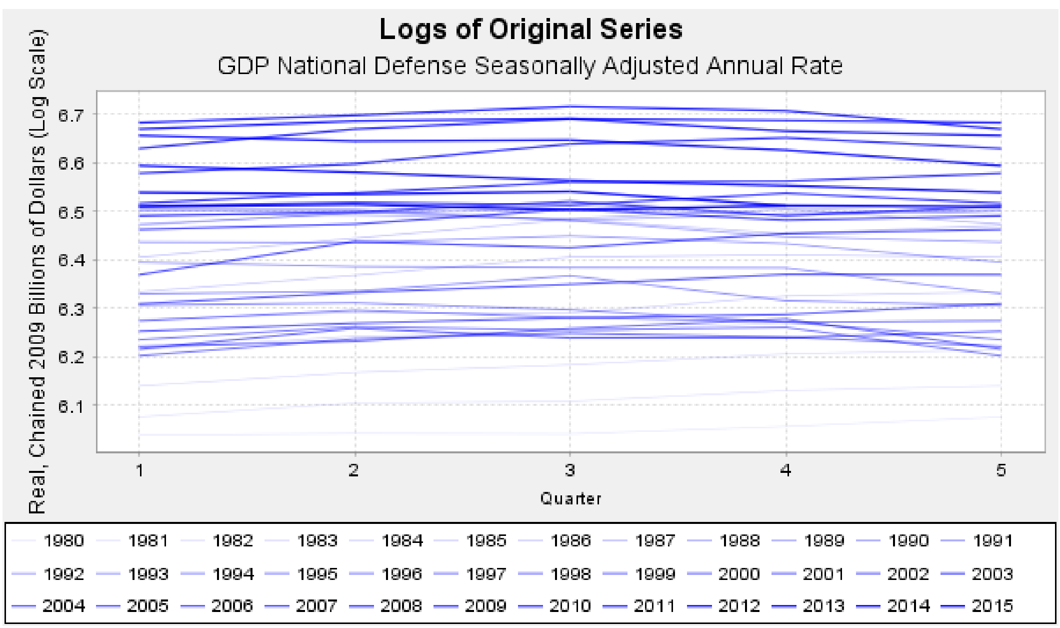

Figure 48: GDP national defense, seasonally adjusted annual rate (in real, chained 2009 billions of dollars), labeled as "original series" because the available estimates are seasonally adjusted, log scale, first quarter 1980 – fourth quarter 2015, year over year. **Source:** [Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)]

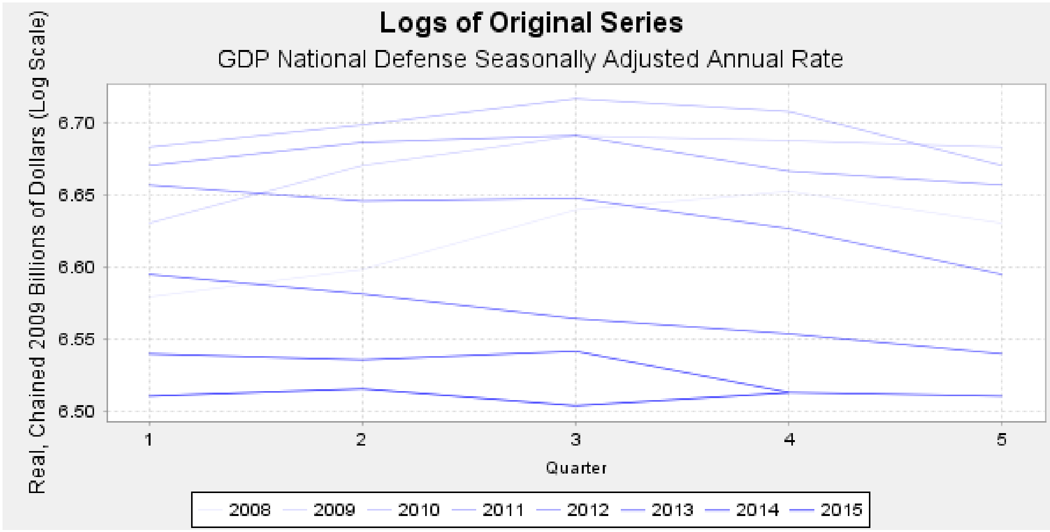

Figure 49: GDP national defense, seasonally adjusted annual rate (in real, chained 2009 billions of dollars), labeled as "original series" because the available estimates are seasonally adjusted, log scale, first quarter 2008 – fourth quarter 2015, year over year. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

Posttest: autocorrelation function plots

Once again, we examined the sample autocorrelation plots for informal evidence of seasonality. Without any differencing, the autocorrelations decay slowly, making it difficult to perceive whether any seasonality is present (**Figure 50**). After seasonal differencing, there is still a substantial degree of autocorrelation remaining, and it is still difficult to see whether the seasonality has been fully repressed (**Figure 51**). **Figure 52** shows the autocorrelation function of the GDP national defense time series with one first difference and no seasonal difference. Although the autocorrelations are not large, the spikes at seasonal lags 4, 8, and 12 are the largest of all lags, and they are all positive. This seems to confirm that seasonality is at least weakly present. **Figure 53** shows the results of both a regular and a seasonal difference, and the negative correlation at lags 1 and 4 noted in previous monthly series examples at lags 1 and 12. These results are not definitive but do provide an indication that we should investigate further.

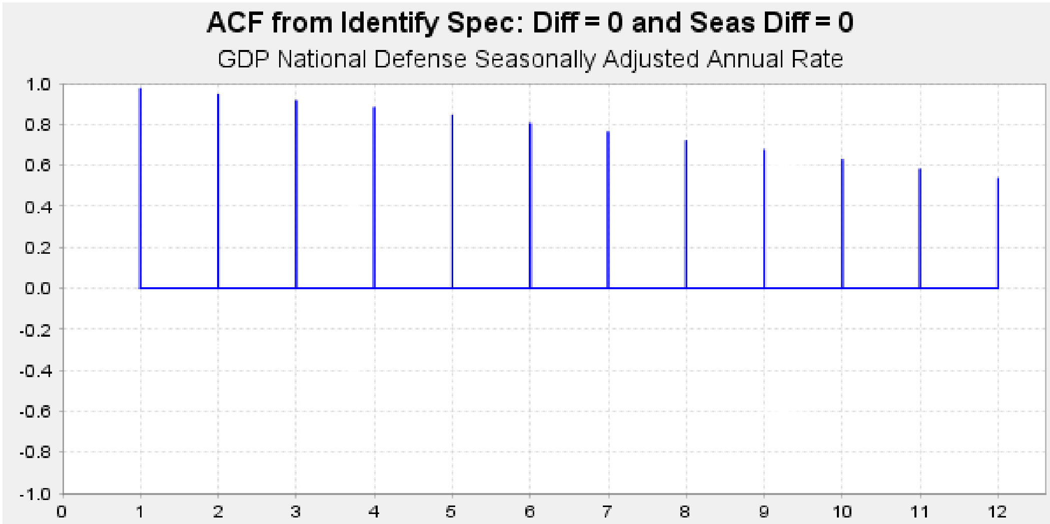

Figure 50: Autocorrelation function of GDP national defense, seasonally adjusted annual rate (series expressed in real, chained 2009 billions of dollars), with no differencing. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

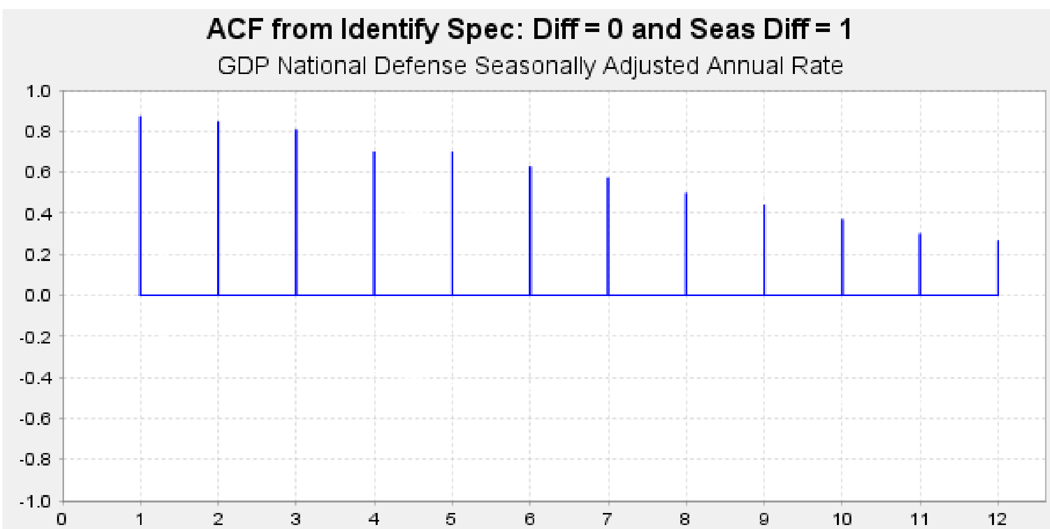

Figure 51: Autocorrelation function of GDP national defense, seasonally adjusted annual rate (series expressed in real, chained 2009 billions of dollars), with no first difference and one seasonal difference. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

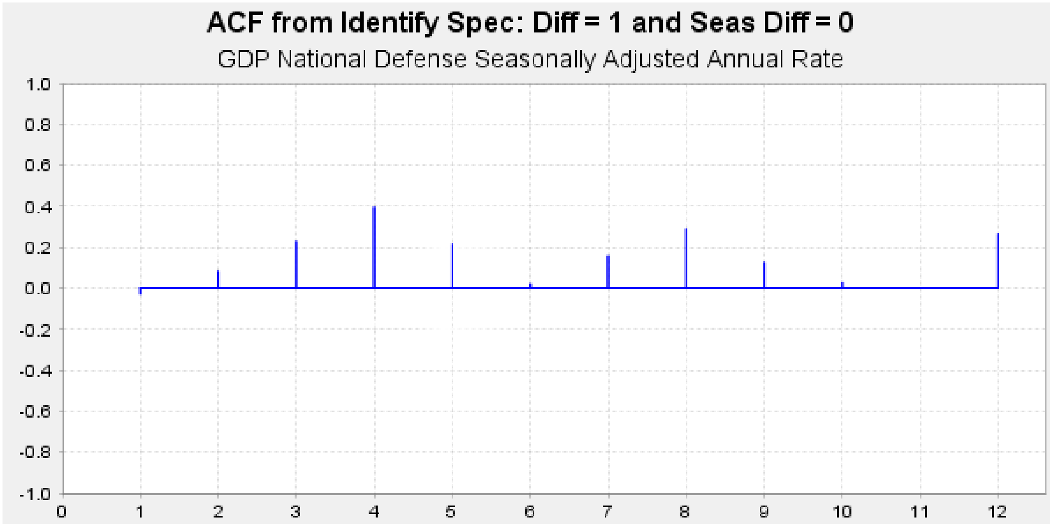

Figure 52: Autocorrelation function of GDP national defense, seasonally adjusted annual rate (series expressed in real, chained 2009 billions of dollars), with one first difference and no seasonal difference. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

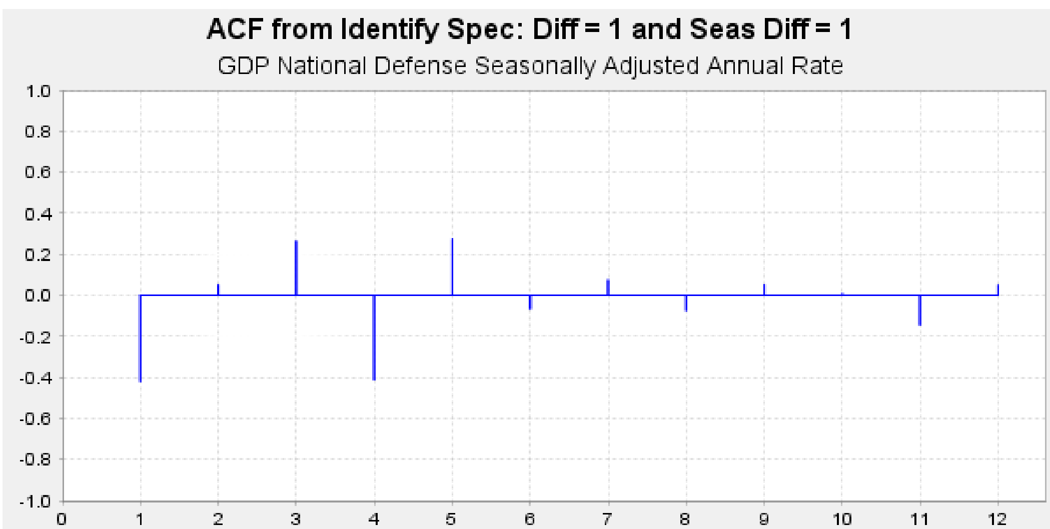

Figure 53: Autocorrelation function of GDP national defense, seasonally adjusted annual rate (series expressed in real, chained 2009 billions of dollars), with one first difference and one seasonal difference. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

Posttest: QS diagnostics

Again, because our "original series" is the seasonally adjusted series, we turned to the QS statistics for the original series for the GDP national defense time series. Given the ARIMA model from automatic selection, the QS statistics, shown in **Table 21**, provide some evidence of residual seasonality in the seasonally adjusted series at the 0.05 level, although not at the 0.01 level.

Table 21: QS statistics for the test for seasonality in the seasonally adjusted annual rate time series (labeled as "Original Series"), for the full series and subspan (starting in the first quarter of 1992), GDP national defense (series expressed in real, chained 2009 billions of dollars). Some tests are for the series adjusted for extreme values. **Source:** Data Archive: National Accounts (NIPA), Bureau of Economic Analysis (apps.bea.gov/histdata/)

Table	Span	QS	*P*-Value
Original Series	Full Series	6.00	0.0498
Original Series (extreme value adjusted)	Full Series	6.00	0.0498
Original Series	Subspan	7.33	0.0256
Original Series (extreme value adjusted)	Subspan	7.33	0.0256

Summary: The QS results provide evidence of residual seasonality at the .05 statistical significance level. Autocorrelation function plots indicate possible seasonality also. When concerns of residual seasonality persist for an indirect seasonal adjustment, it is best to look if the components' adjustments might be improved, if they are available.

5.7 Example 7: Weak Seasonality

The previous examples exhibited test results that we could classify as adequate or inadequate with confidence. This example, however, shows a series with weak seasonality, such that some diagnostics may indicate that it is not a candidate for seasonal adjustment. The series is Leather and Allied Products Work in Process Inventories, as provided by the Manufacturers' Shipments, Inventories, and Orders (M3) survey, U.S. Census Bureau.[13] Graphs of the original series in millions of dollars, on the log scale, over the entire time span as in **Figure 54**, or in the year-over-year graph as in **Figure 55** do not show a clear seasonal pattern. The large change in level of the series between 2000 and 2012 made the year-over-year graph hard to interpret, so in **Figure 55** we show a subspan of the full series, starting at 2012. Seasonality might be present, as the series seems to have a consistent decrease from May to June and usually a decrease in October followed by an increase in November. We checked the seasonality diagnostics for evidence.

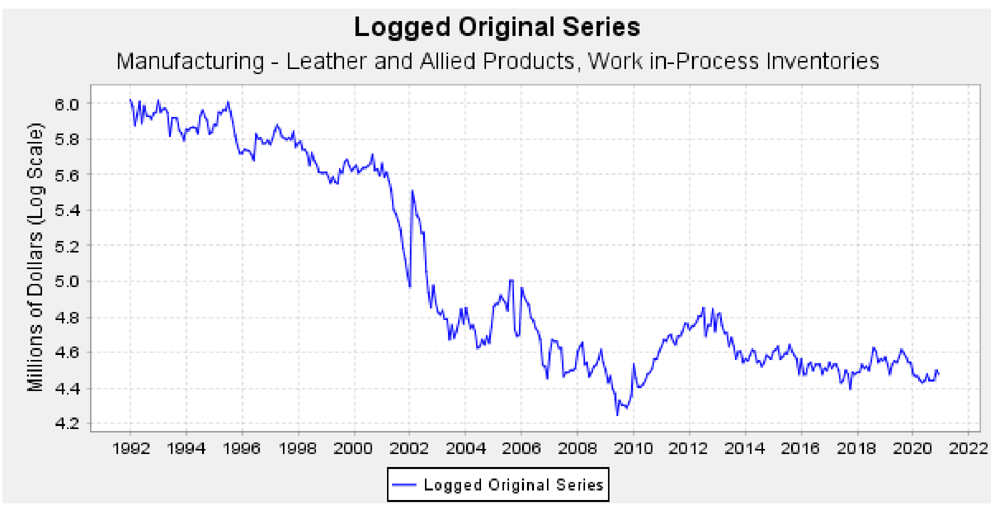

Figure 54: Manufacturing of leather and allied products, work in-process inventories (in millions of dollars), log scale, original series 1992–2020. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

[13] These estimates are subject to measurement error. Statistical significance is not measurable for this survey. The Manufacturers' Shipments, Inventories, and Orders estimates are not based on a probability sample, so the sampling error of these estimates cannot be measured, nor can the confidence intervals be computed. The estimates are not adjusted for price changes. More information about the data collection and estimation for the Manufacturers' Shipments, Inventories, and Orders Survey is available online (census.gov/manufacturing/m3/ how_the_data_are_collected/).

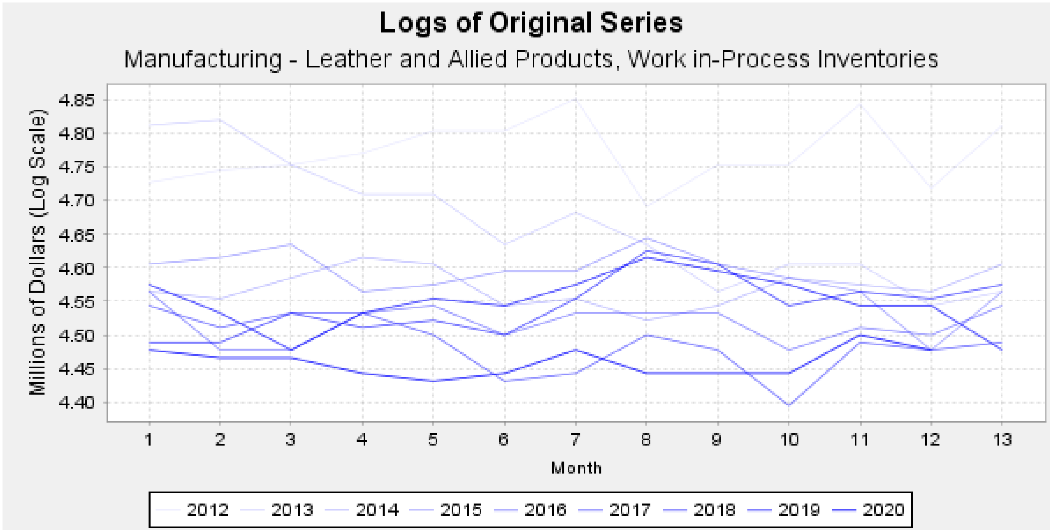

Figure 55: Manufacturing of leather and allied products, work in-process inventories (in millions of dollars), original series 1992 – 2020, year over year. **Source:** [Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)]

Pretest: autocorrelation function plots

Plots of the autocorrelation function provide an informal test of whether the series is seasonal. The autocorrelation function plots of **Figure 56**, **Figure 57**, **Figure 58**, and **Figure 59** show the results of combinations of nonseasonal differencing and seasonal differencing of orders 0 and 1. With a first difference and no seasonal difference, the autocorrelation function of **Figure 58** shows a spike at lag 12 that is only slightly larger than the other autocorrelation estimates, indicating weak seasonality. The values at seasonal lags 24 and 36 are negligible. Note that no seasonality is apparent from the plot based on the original series (**Figure 56**), and this is because the strong trend behavior completely overwhelms any signals about seasonality that the autocorrelation might convey. Similarly, **Figure 57** (with no regular difference and one seasonal difference) confirms the hypothesis of weak seasonality. Finally, **Figure 59** shows negative lag 12 autocorrelation, which is consistent with the impact of applying both a regular and a seasonal difference.

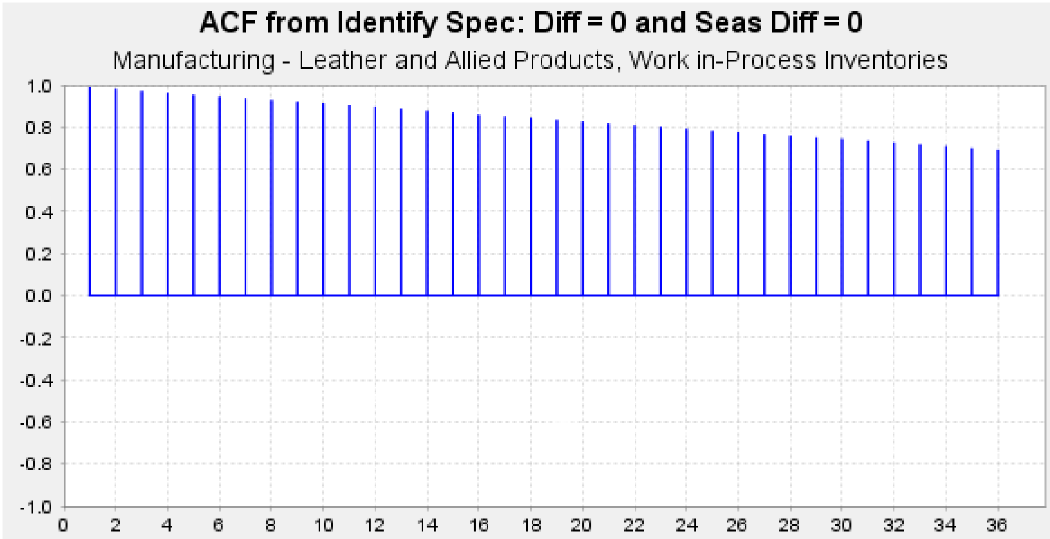

Figure 56: Autocorrelation function of manufacturing of leather and allied products work in-process inventories time series, with no differencing. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

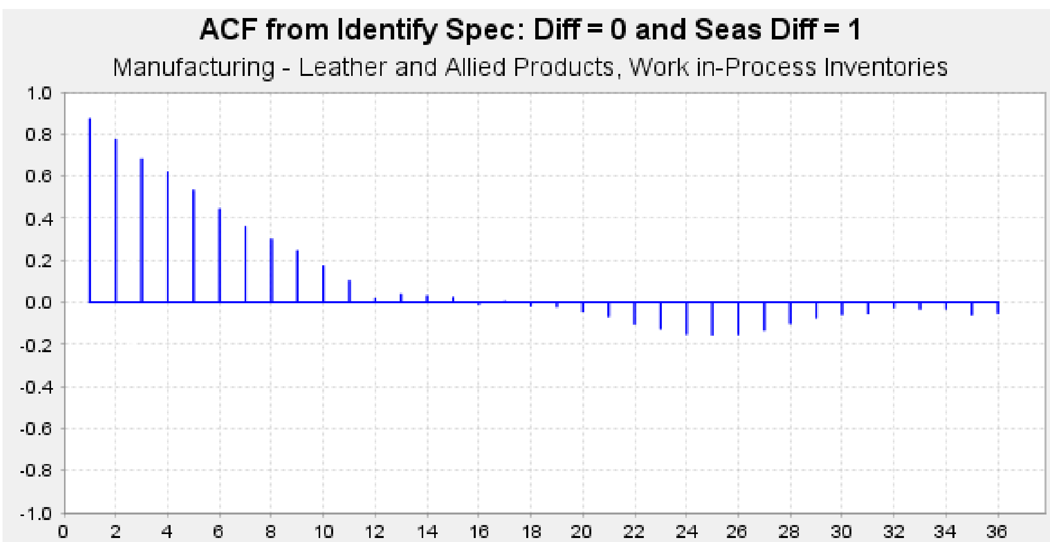

Figure 57: Autocorrelation function of manufacturing of leather and allied products work in-process inventories time series, with no first difference and one seasonal difference; **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

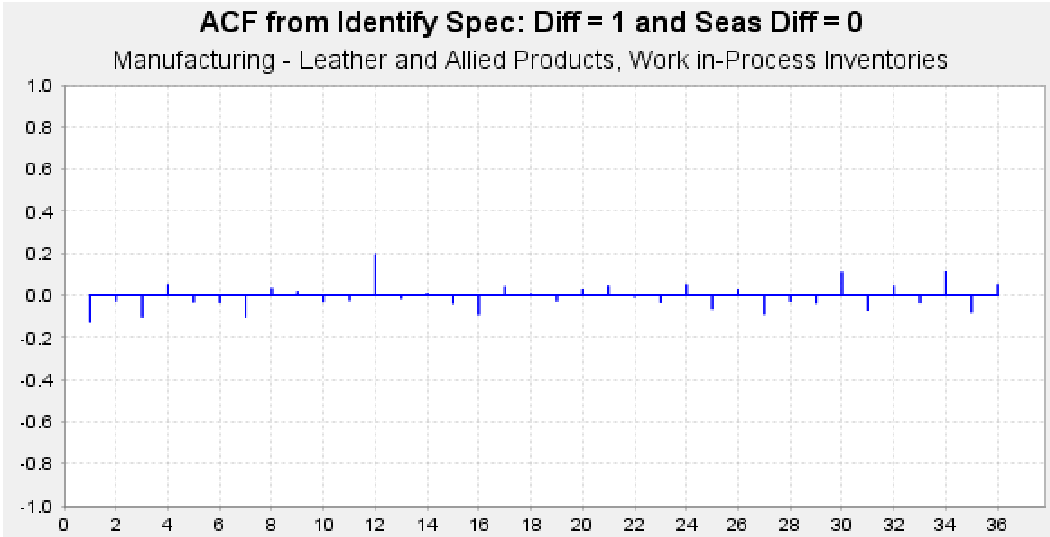

Figure 58: Autocorrelation function of manufacturing of leather and allied products work in-process inventories time series, with one first difference and no seasonal difference. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

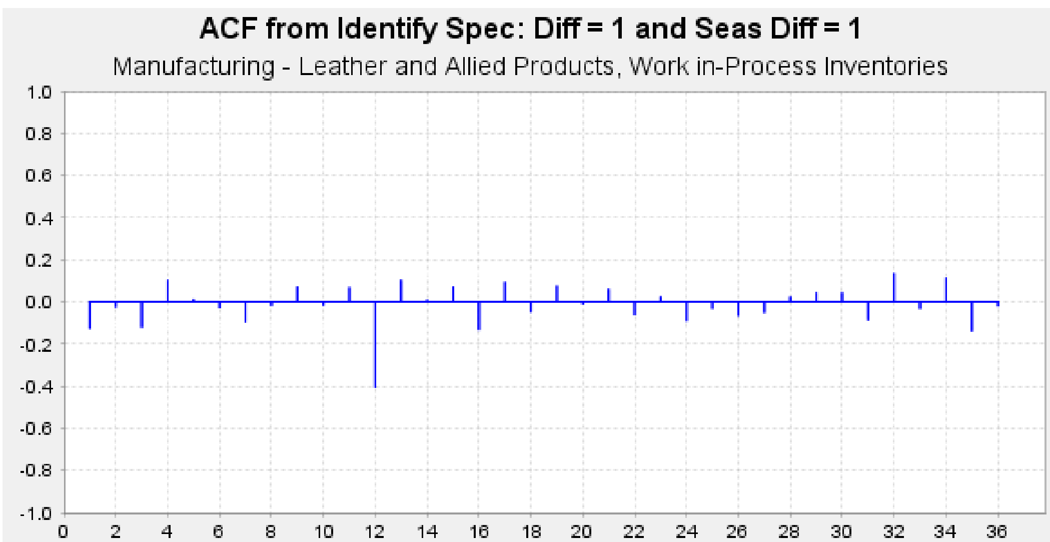

Figure 59: Autocorrelation function of manufacturing of leather and allied products work in-process inventories time series, with one first difference and one seasonal difference. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

Pretest: regression model-based F test

For the regression model-based F test, we ran automatic model selection with the fixed seasonal regressors. The resulting model was ARIMA (1 1 0), no trend constant (it was removed during model selection because it was not significant), fixed seasonal regressors, and five automatically identified level shifts, in February 2002, August 2002, October 2005, January 2006, and July 2007.

As **Table 22** shows, the test was significant with a *p*-value of 0.00, although the F statistic, at 6.65 is smaller than what we have seen with previous examples.

Table 22: Degrees of freedom, F statistic, and *p*-value for the group of 11 seasonal regressors fit to monthly manufacturing of leather and allied products work in-process inventories time series. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

Regressor	DF	F-statistic	*P*-Value
Seasonal	11, 331	6.65	0.00

Summary: The regression model-based F test provides evidence that the original series is seasonal.

Pretest: spectral plot peaks at seasonal frequencies

Continuing with the pretesting, we checked for peaks at seasonal frequencies in the spectral plot of the transformed, differenced, prior-adjusted original series. The plot, **Figure 60**, has one visually significant peak at seasonal frequency 4/12. With most seasonal time series, we see peaks at more than one seasonal frequency, and usually at frequencies 1/12 and 2/12, but even this lone visually significant peak does provide evidence that the series is seasonal.

Summary: The spectral plot provides (weak) evidence that the original series is seasonal.

Both of the more formal pretests and the autocorrelation function plots provide some evidence that the original series is seasonal, and hence it can be considered for seasonal adjustment.

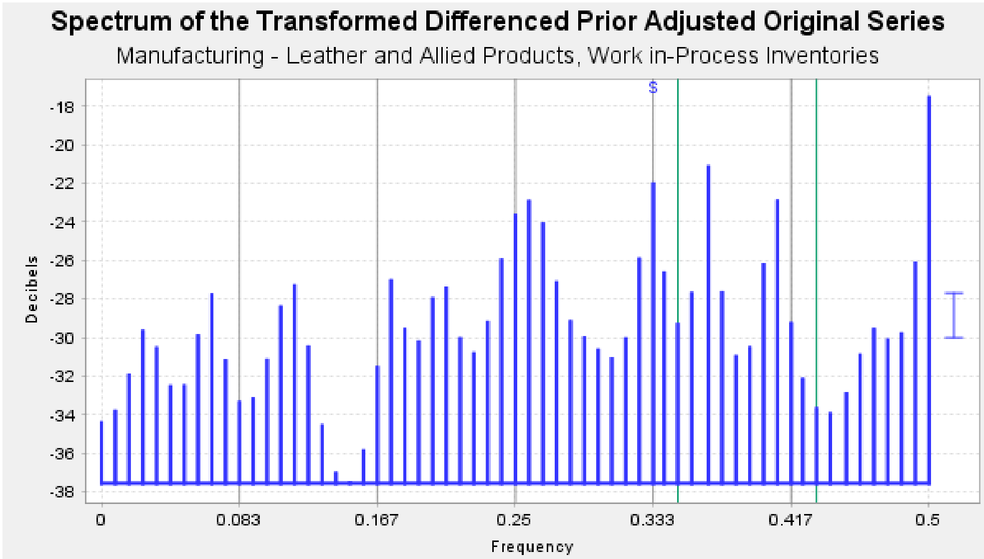

Figure 60: Monthly manufacturing of leather and allied products work in-process inventories time series; spectrum of the differenced, transformed prior-adjusted series (Table B1) with one visually significant peak at seasonal frequency 4/12. **Source:** [Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)](census.gov/manufacturing/m3/)

Note: Many concerns factor into a decision to seasonally adjust a series that exhibits only weak seasonality, particularly if not adjusting results in evidence of residual seasonality in an aggregate that includes the series as a component. Alternatively, the series might have a long history of seasonal adjustment and ceasing adjustment might not be an immediate option.

Without much background information about this time series, we seasonally adjusted using automatic and default settings and considered the diagnostics for residual seasonality.

Posttest: QS diagnostics

For both the full series and the subspan starting at January 2013, the QS statistics for the seasonally adjusted series, seasonally adjusted series (extreme value adjusted), irregular series, and irregular series (extreme value adjusted) were each 0.00, with associated *p*-values of 1.0000. We can test the derived quarterly seasonally adjusted series as well, using **`qcheck=yes`** in the **`spectrum`** spec. Because the series is a measure of inventory (stock series), we needed to set **`type=stock`** in the **`series`** spec to produce the appropriate quarterly values. This setting assumes end-of-month and end-of-quarter stock measure dates. (If inventory dates are not at the end of the month, users must derive the quarterly seasonally adjusted series outside of the software and then test the derived series in a separate run.) The test results are in the output table labeled "QS Statistics for (quarterly) seasonality." As with the monthly series, for the full series, for the shortened series, and for each of those adjusted for extreme values, the *p*-values for the QS statistics were each greater than 0.5.

Summary: The QS results provide no evidence of residual seasonality at the .05 statistical significance level.

Posttest: spectral plot peaks at seasonal frequencies

Figure 61 and **Figure 62** both exhibit dips at seasonal frequencies, not peaks. The plots indicate no evidence of residual seasonality in the seasonally adjusted series or in the irregular component.

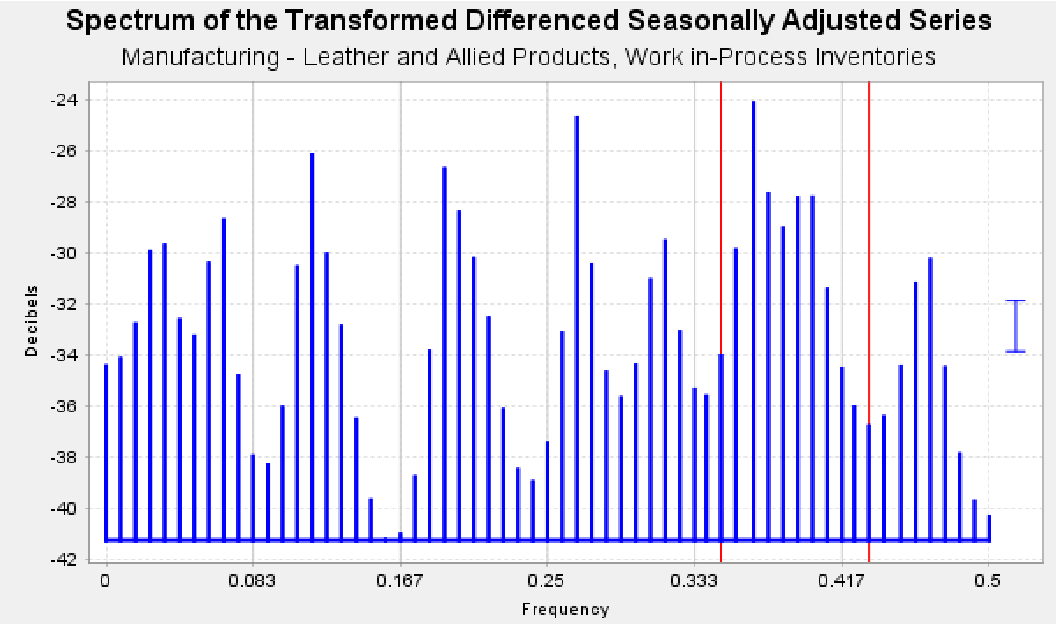

Figure 61: Monthly manufacturing of leather and allied products work in-process inventories time series; spectrum of the transformed, differenced seasonally adjusted series with no peaks, and instead, visual dips, at seasonal frequencies. **Source:** Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)

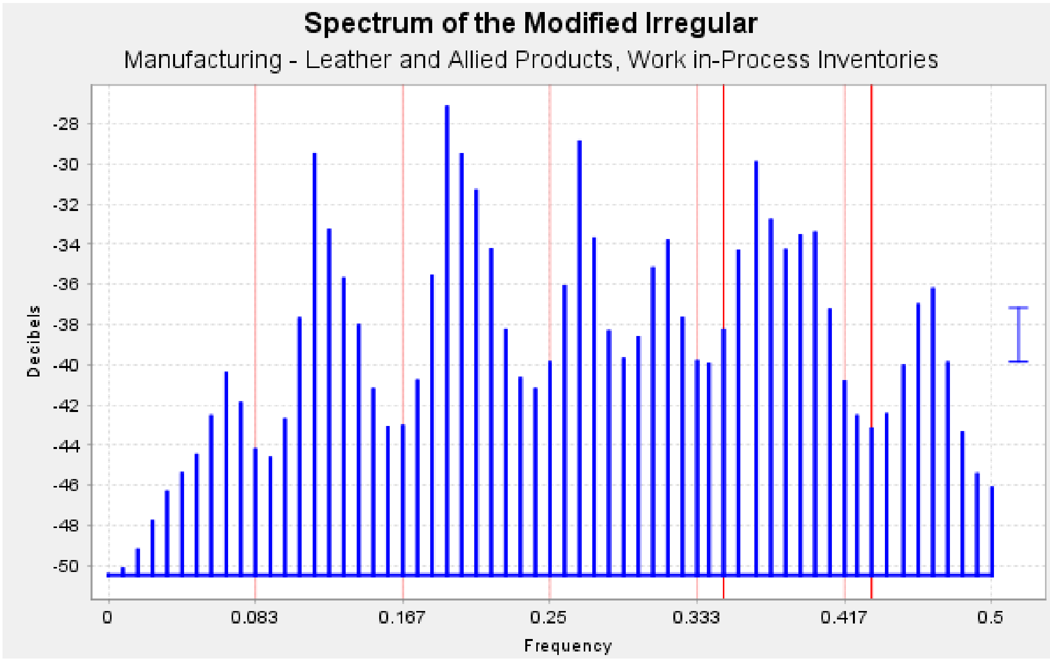

Figure 62: Monthly manufacturing of leather and allied products work in-process inventories time series; spectrum of the modified irregular component with no peaks, and instead, visual dips, at seasonal frequencies. **Source:** [Manufacturers' Shipments, Inventories, and Orders Survey, U.S. Census Bureau (census.gov/manufacturing/m3/)](census.gov/manufacturing/m3/)

Summary: The spectral plot results provide no evidence of residual seasonality using established significance levels.

The QS statistics and the spectral diagnostics provide no evidence of residual seasonality at established significance levels. If the adjustment meets other established quality measures, it is adequate.

Appendix

A: General Background on Seasonality in Time Series

Seasonality diagnostics stem from the notion that seasonality in a time series corresponds to positive association at seasonal lags, after accounting for association at nonseasonal lags. That is, for a time series $\{X_t\}$ with seasonal period s, high positive association of X_t with X_{t-s}, for any t, indicates seasonality, given that associations of X_t with X_{t-j} for j not equal to s have been accounted for. A common measure of association is correlation; more generally, if one variable is useful for predicting another, we say there is an association.

These vague notions, without further development, are not useful for generating statistical diagnostics. Seasonality is an economic and scientific phenomenon, so giving it a strict mathematical definition (as we do for variance or probability densities) is inappropriate, as that would confine the potential scope of the concept too narrowly. However, mathematical definitions can explain facets of seasonality, as they might manifest in different statistical paradigms. Differing formulations of the seasonality concept yield different diagnostics, each of which is not sufficiently broad to cover all manifestations of the phenomenon, but these diagnostics, taken together, can be helpful. See Bell and Hillmer (1984, 1985) for a historical overview, and a model-based approach to definitions and methodology.

General background material on time series is available in many texts. McElroy and Politis (2020) provides an introductory treatment of stochastic processes that is accessible to readers with an undergraduate degree in statistics. Here we distinguish between stationary and nonstationary time series; the former have shift-invariant marginal distributions, indicating that the stochastic behavior in one temporal locale is identical, probabilistically, with any other locale. As a result, a stationary time series with finite variance has an autocovariance function that depends only on lag, namely $\gamma_h = Cov[X_{t+h}, X_t]$, for any t. The autocorrelation function (ACF) is then defined via $\rho_h = \gamma_h / \gamma_0$. Nonstationary processes, in contrast, have autocovariance functions that depend upon the local time t.

A stationary time series can still exhibit seasonality, according to the broad concept discussed above, if the association is sufficiently strong. However, strong seasonality is typically associated with a nonstationary time series, as either a deterministic or stochastic phenomenon. Consider the following two examples, which illustrate a continuum of behavior from weak seasonality to strong seasonality.

Example A.1: A Seasonal Process

This example draws heavily from material in Findley, et al. (2015). Consider the first order seasonal autoregression (SAR), defined as

$$X_t = \phi_s X_{t-s} + \epsilon_t,$$

where $\{\epsilon_t\}$ is i.i.d. with mean zero and variance σ^2. The parameter $\phi_s \in (-1,1)$ to ensure stationarity of $\{X_t\}$, with larger positive values indicating a higher degree of seasonality. This relationship is immediate from our discussion of association because the optimal predictor s steps ahead, given time series $X_1, \dots, X_T$, is $\phi_s X_T$. (The h-step ahead forecast filter, based on an infinite past, is $\phi_s^{n+1} B^{s-k}$, where $h = ns + k$ for $k = 1, 2, \dots, s$ and B is the backshift operator. McElroy and McCracken (2017) provide details of the verification.) The autocorrelation function (assuming $h \geq 0$) in the case of the SAR equals ϕ_s^nif $h = ns$, and is zero otherwise. (The MA(∞) representation is $X_t = \Psi(B)\epsilon_t$, with $\Psi(B) = (1 - \phi_s B^s)^{-1} = \sum_{j\geq 0} \phi_s^j B^{js}$ and the autocovariance function is $\gamma_h = \sum_{k\geq 0} \psi_k \psi_{k+h} \sigma^2$, where $\psi_k = \phi_s^j$ if $k = js$ and is zero otherwise. Hence, $\psi_k \psi_{k+h}$ for each k can be nonzero only if both ψ_k and ψ_{k+h} are nonzero, which is impossible unless h = ns for some n.) Therefore, if ϕ_s is large, correlation is high at seasonal lags, and no correlation occurs at nonseasonal lags.

Figure A.1 depicts various values of the autocorrelation function for the SAR model, showing that only the values at the seasonal lags (*s*, 2*s*, 3*s*, …) are non-zero. **Figure A.2** depicts four simulated series for the case of monthly data (*s* = 12), the four series corresponding to each value of ϕ_s used in **Figure A.1**, to show the visual correspondence. These simulations were done by setting the first 13 values of the series to a scaled sine wave, with points indicated by the dots in the plots. The plots were done this way to show the extent to which the sine wave pattern of the first year persists through the subsequent years plotted, and how this depends on the value of ϕ_s.[14]

[14] The scaling was multiplication by 1.96, the .975 quantile of the standard normal distribution, and the innovations ϵ_t are i.i.d. $N(0, 1 - \phi_s^2)$. This choice of innovation variance gives X_t an unconditional variance of 1.0, while the scaling factor 1.96 puts the maximum and minimum of the sine wave at ±1.96 so that the range of the sine wave covers the middle 95% of the distribution of X_t, making the range of variation of the two similar. One set of innovations was used to generate all four series, just with rescaling to get the desired innovation variance.

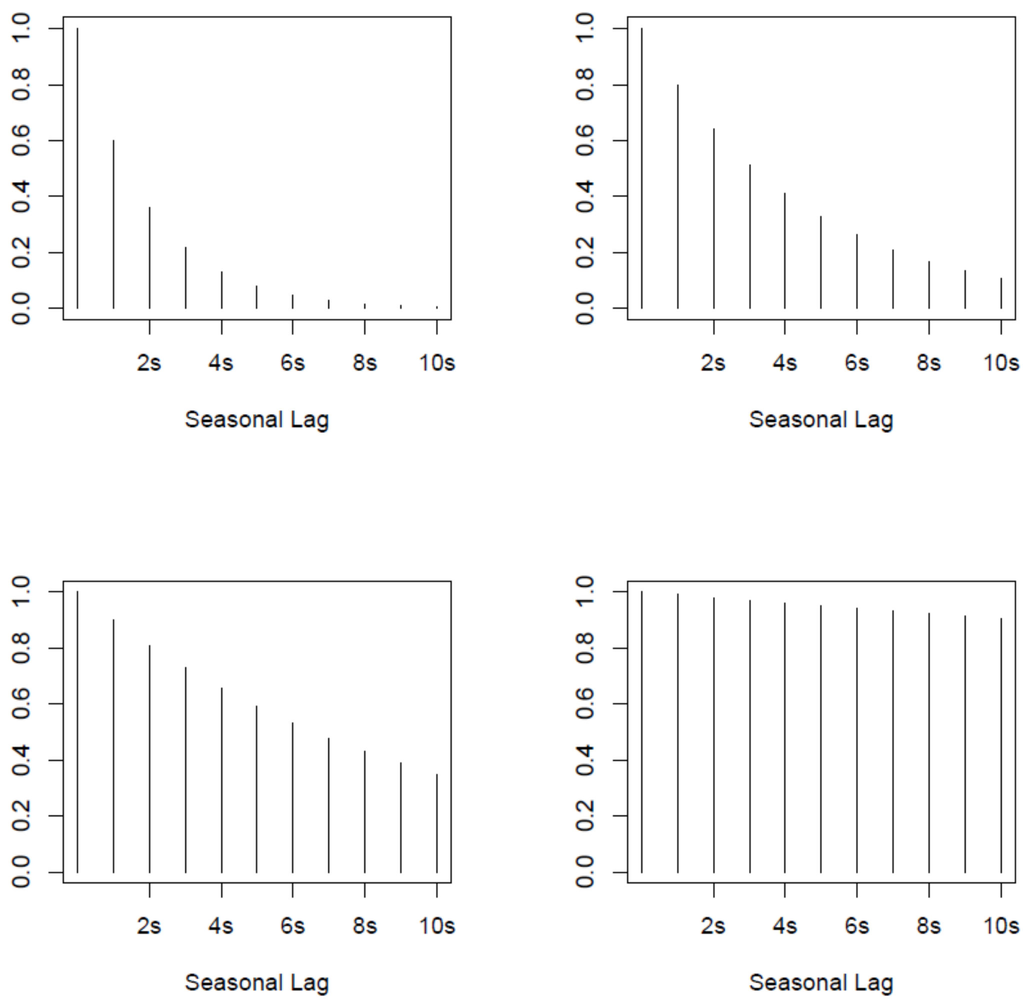

Figure A.1: Autocorrelation plots for the SAR with parameter $\phi_s = 0.6$ (upper left panel), $\phi_s = 0.8$ (upper right panel), $\phi_s = 0.9$ (lower left panel), $\phi_s = 0.99$ (lower right panel). Lag is in units of years. **Source:** Direct calculation.

The plot for $\phi_s = 0.6$ shows that the sine wave of the first year visibly persists for only about one additional year. The plots for $\phi_s = 0.8$ and $\phi_s = 0.9$ show persistence of a shape something like the sine wave for about 5 years, with obvious degradations from the noise introduced. The plot for $\phi_s = 0.99$ shows strong persistence over time, with just slight alterations due to the innovation noise, which has standard deviation of just $(1 - 0.99^2)^{0.5} = 0.141$. While redoing the simulations would produce results that vary some in appearance, they would make the general point that persistent appearance of a seasonal pattern requires a large value of ϕ_s. The autocorrelation plots of Figure A.1 reflect this behavior, with very little persistence of seasonal autocorrelations when $\phi_s = 0.6$, some persistence of seasonal autocorrelations with $\phi_s = 0.8$ or 0.9, and strong seasonal persistence of autocorrelation apparent in the case of $\phi_s = 0.99$.

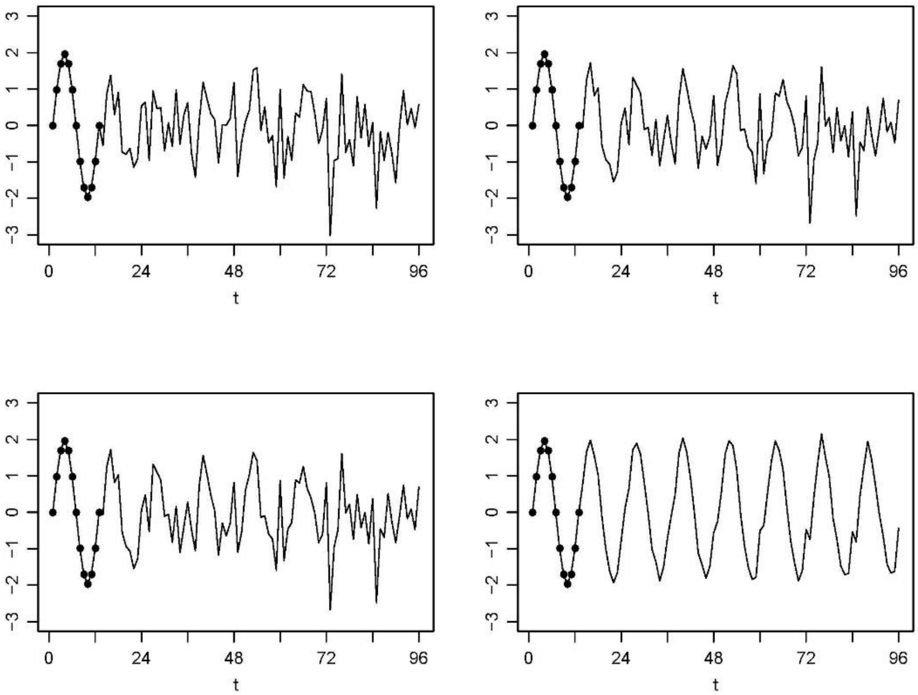

Figure A.2: Simulation sample paths for monthly series simulated from the SAR model with parameter $\phi_s = 0.6$ (upper left panel), $\phi_s = 0.8$ (upper right panel), $\phi_s = 0.9$ (lower left panel), $\phi_s = 0.99$ (lower right panel). **Source:** Simulated series.

To see why association at seasonal lags is not sufficient to describe seasonality, suppose that correlation was high at nonseasonal lags. Then it could happen that the aggregation of high season-to-season correlation generates a high seasonal correlation. The next example develops this observation explicitly.

Example A.2: A Nonseasonal Process

Consider the regular autoregression (AR), defined as

$$X_t = \phi_1 X_{t-1} + \epsilon_t$$

where $\{\epsilon_t\}$ is i.i.d. with mean zero and variance σ^2. The parameter $\phi_1 \in (-1, 1)$ to ensure stationarity of $\{X_t\}$, with larger positive values indicating a stronger persistency. The autocorrelation function is $\rho_h = \phi_1^h$. Hence, $\rho_s = \phi_1^s$, which could be large if ϕ_1 is close to unity. Although there would then be a high seasonal-lag correlation, the correlation at nonseasonal lags is also high, and this nullifies the seasonal effect. Figure A.3 depicts the autocorrelation functions in the quarterly case $s = 4$, for four values of ϕ_1. Figure A.4 then depicts simulated series corresponding to each value of ϕ_1, to show the visual correspondence. We chose the values of ϕ_1 so that the autocorrelation at seasonal lags exactly matches the values from the SAR, i.e., ϕ_1 increases from 0.880 to 0.946, 0.974, and 0.997. Clearly, the simulations indicate no seasonal-lag associations, and the behavior of the autocorrelation plot is quite different from that of the SAR.

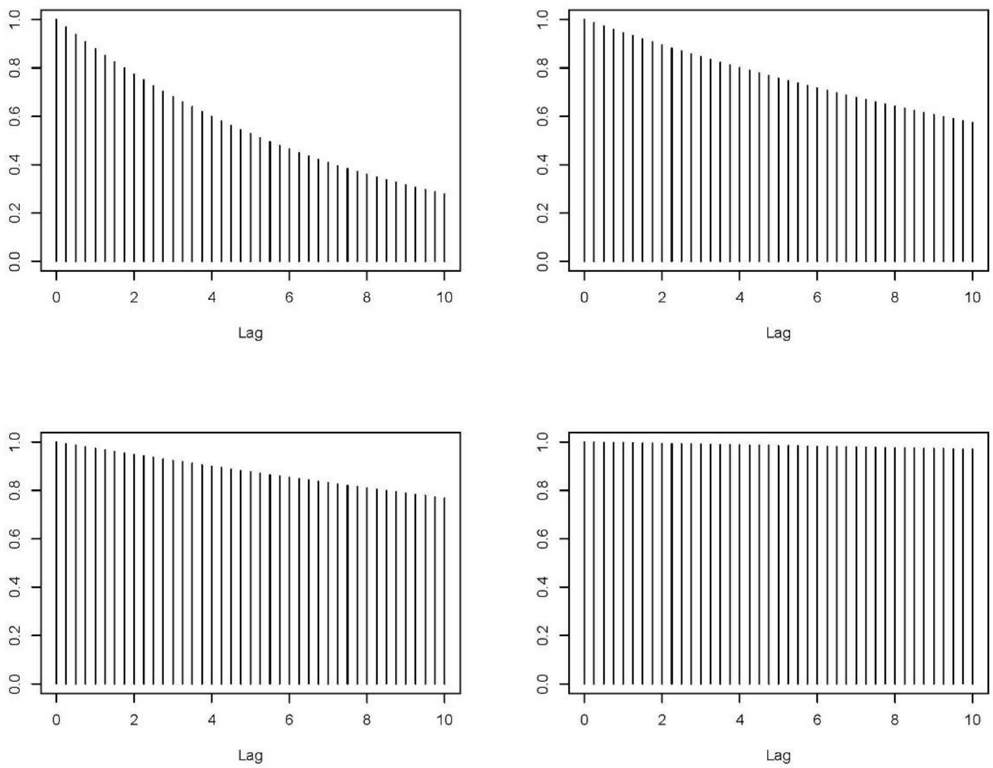

Figure A.3: Autocorrelation plots for the AR with parameter $\phi_1 = 0.9457$ (upper left panel), $\phi_1 = 0.9740$ (upper right panel), $\phi_1 = 0.9873$ (lower left panel), $\phi_1 = 0.9975$ (lower right panel). **Source:** Simulated series

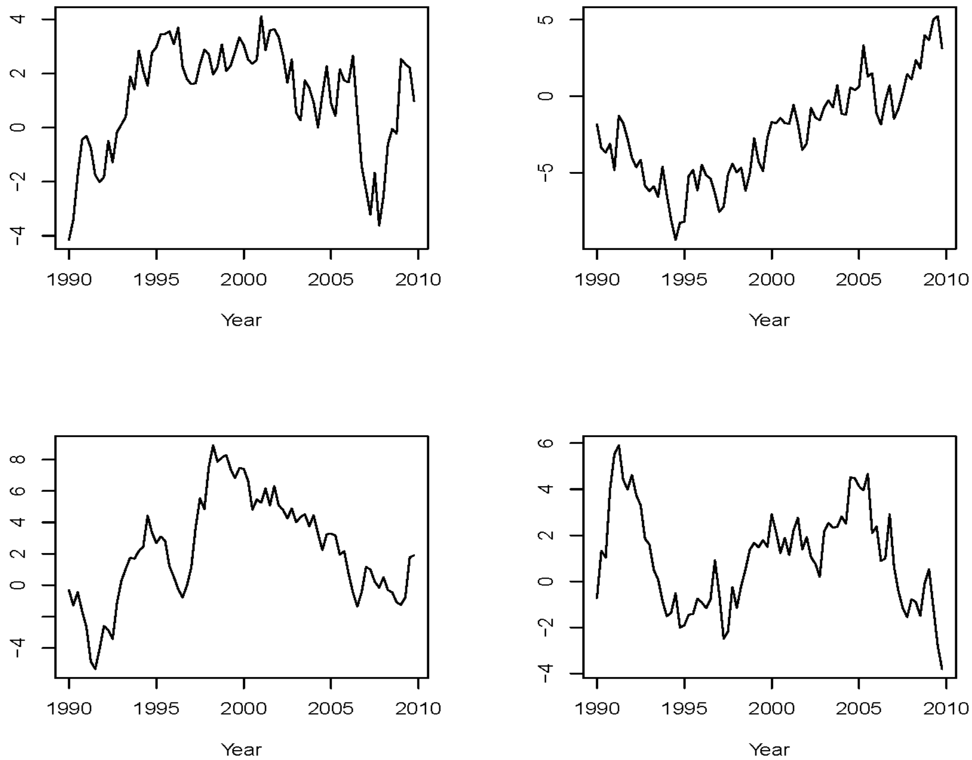

Figure A.4: Simulation sample paths for the AR with parameter $\phi_1 = 0.9457$ (upper left panel), $\phi_1 = 0.9740$ (upper right panel), $\phi_1 = 0.9873$ (lower left panel), $\phi_1 = 0.9975$ (lower right panel). **Source:** Simulated series

Stationary versus Nonstationary Seasonality

Some scholars argue that seasonality should be formulated only in terms of nonstationary time series, because stationarity implies that the associations from year to year are not everlastingly persistent. Others have argued that a correlation does not need to be perfect (that is, equal to positive or negative one) for a meaningful association to exist. Many of the diagnostics in common use are from a stationary seasonality paradigm, so we must allow for the possibility of stationary seasonality as we develop the technical foundations.

If the autocovariance function is absolutely summable, then the spectral density can be defined as

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_h \, e^{-i\,h\,\lambda}$$

for any $\lambda \in (-\pi, \pi)$. Because the autocovariance function is symmetric in *h*, the spectral density is real, and we can write it as

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_h \cos(h\lambda).$$

If a stationary time series exhibits seasonality, γ_h is high for *h* of the form *ns*. Cosine is largest when its argument equals a multiple of 2π, so λ of the form $2\pi/s$, or integer multiples of such, correspond to cos($ns\lambda$) of large value, and the spectral density will have a local maximum at those frequencies. However, for this argument to be valid, it is necessary that γ_h take low values at lags that are neighboring the seasonal lags, as otherwise the local maxima will differ.

For instance, in Example A.1 the spectral density is

$$f(\lambda) = \sigma^2(1 + \phi_s^2 + 2\phi_s \cos(s\lambda))^{-1}.$$

The maximizer is the frequency $2\pi/s$, or any integer multiple thereof. The frequencies $2\pi j/s$ for *j* = 1, 2, …, *s* are known as seasonal frequencies, as they refer to phenomena of period s/j, which occur *j* times within the year. In contrast, the spectral density for Example A.2 is

$$f(\lambda) = \sigma^2(1 + \phi_1^2 + 2\phi_1 \cos(s\lambda))^{-1},$$

which has a maximizer at frequency zero. This is not a seasonal process, and the spectral density has no peaks at seasonal frequencies.

We can formulate nonstationary processes that generalize the stationary case by having unit correlation at certain lags. In other words, there is a full association between X_t and X_{t-s}. To do this, we can generalize the autoregressive equation

$$X_t = \phi X_{t-s} + \epsilon_t$$

of Examples A.1 and A.2 by allowing $\phi = 1$. If we maintain $\sigma^2 > 0$ then the process will be stochastic (and nonstationary), but if we shrink σ^2 to zero, then the process will be deterministic. In the stochastic case, applying seasonal differencing yields

$$(1 - B^s)X_t = X_t - X_{t-s} = \epsilon_t$$

and $\{\epsilon_t\}$ is a stationary process. In the deterministic case the same equation holds, but now $\epsilon_t = 0$. In other words, seasonal differencing annihilates the process. A deterministic process can still be stationary, though the autocovariances need no longer be absolutely summable, and the spectral density does not exist – instead, we have recourse to the spectral distribution function, which is the antiderivative of the spectral density in the case that the latter exists. Stationary time series data consistently allow estimation of the spectral distribution function, and jump discontinuities correspond to periodic (seasonal) effects. The spectral distribution function could be the basis of a diagnostic, but no published research seems to exist on this approach.

The more conventional approach to seasonality diagnostics starts with a stochastic formulation of nonstationarity, broadly referred to as unit-root processes. Authors have proposed various models to capture stochastic nonstationarity, including the ARIMA and SARIMA classes (Box and Jenkins 1976). The simplest such model involves setting $\phi_s = 1$ in Example A.1.

B: Additional Technical Background on Seasonality Diagnostics

This part of the appendix focuses upon four commonly used seasonality diagnostics: (i) the autocorrelation function (ACF); (ii) the QS diagnostic; (iii) the regression model-based F test (MBF); (iv) the spectral Visual Significance (VS) criterion. Each is presented in the above context of stationary and nonstationary processes, with a discussion of the underlying assumptions necessary for the diagnostic's validity. An additional reference for the four diagnostics is Findley et al. (2017).

ACF Diagnostic

Based on the ideas of association, we say that seasonality is present if ρ_s surpasses a given threshold τ. This definition presumes a stationary time series, because otherwise ρ_s is not well defined. Clearly, setting $\tau = 0$ sets a very low bar – recall from **Figure A.1** and **Figure A.2** that any value of $\tau < 0.8$ would correspond to a quite weak seasonal autocorrelation. A second problem is that a high value of ρ_s does not necessarily require the presence of seasonality – recall the discussion of Example A.2. On the other hand, a low value of ρ_s does preclude the possibility of seasonality's presence.

In summary, we can use seasonal autocorrelations to test whether seasonality is *not* present but cannot reliably use them to test whether seasonality *is* present; the endeavor requires the examination of autocorrelation at other lags. If the series has a stochastic or deterministic trend, then it must be differenced to stationarity first by applying 1 − B to the series. Let $X_1, \ldots, X_T$ denote the resulting sample. Consider a construction of the test statistic nonparametrically via the sample autocovariance:

$$\hat{\gamma}_s = T^{-1} \sum_{t=1}^{T-s} (X_t - \bar{X})(X_{t+s} - \bar{X}),$$

where $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t$. Then the sample autocorrelation is $\hat{\rho}_s = \hat{\gamma}_s / \hat{\gamma}_0$. McElroy and Politis (2020) give the asymptotic theory of this estimator; it is asymptotically unbiased for ρ_s (as T tends to infinity with s fixed) and is asymptotically normal under fairly broad conditions. The asymptotic variance depends on other autocovariances and can be determined by plug-in estimators. Another approach is to fit a parametric model, such as an ARMA model, and estimate the autocovariances through the postulated model; however, this has the disadvantage of predicating the diagnostic on a fitted model. If the model is incorrect, it will corrupt the quantification of the uncertainty.

QS Diagnostic

Maravall (2012) introduced the QS statistic, presuming a stationary time series. The starting point is the adoption of $\tau = 0$ in the preceding ACF diagnostic, followed by proceeding with the fallacy that large values of ρ_s and ρ_{2s}indicate seasonality; in fact, large values of ρ_s and ρ_{2s} in no way preclude a nonseasonal process (recall Example A.2 of Appendix A). To focus on upper one-sided alternatives, one can square the positive part of the sample autocorrelations – if $\hat{\rho}_s$ or $\hat{\rho}_{2s}$ are negative, they are discarded. If $\hat{\rho}_s$ is less than or equal to zero, then QS is set equal to zero, but otherwise it is given by the formula

$$QS = T(T+2)\left(\frac{\max(0,\hat{\rho}_s)^2}{T-s} + \frac{\max(0,\hat{\rho}_{2s})^2}{T-2s}\right)$$

where T is the sample size (after differencing, as needed, to remove trend nonstationarity in the series). The normalization of the two terms is intended to balance the two contributions and generate a null distribution closely resembling the χ^2 distribution (with 2 degrees of freedom) under a variety of null processes. No attempt to rigorously establish an asymptotic theory for QS is published (see discussion in Findley et al. (2017)). Some consider the QS methodological foundation to have severe problems for the reasons noted above, viz. large values of the seasonal autocorrelations can occur with nonseasonal processes.

MBF Diagnostic

Lytras et al. (2007) proposed the model-based F statistic to test for deterministic (fixed) seasonality in unadjusted time series, and compared it to other approaches, including the spectral diagnostic and X-11 F statistic. The model-based F test fits a model with fixed seasonal effects and tests if the corresponding regression coefficients are all zero, thus taking as the null hypothesis that deterministic seasonality is **not** present. The model is of the form

$$X_t = Y_t + \beta' z_t$$

where $\{Y_t\}$ is a mean-zero time series, and $\{z_t\}$ is a vector of regressors that includes the regressors for fixed seasonal effects, with β the corresponding parameter vector. (The seasonal effect regressors are s − 1 seasonal contrast dummies, though seasonal trigonometric regressors could alternatively be used to yield the same result.) The method proceeds by postulating an ARIMA model for $\{Y_t\}$, identified according to standard modeling craft (ACF plots, information criteria, etc.). This model can have stationary seasonal AR and/or MA operators. Seasonal differencing is not permitted since this, in combination with the fixed seasonal effects, would lead to a singularity. Because of the presence of the regression mean effect, such a model is called a regARIMA model (Bell 2004).

X-13ARIMA-SEATS estimates regARIMA models by maximum likelihood, and inferences about the regression coefficients are made using Generalized Least Squares (GLS) results conditional on the estimated parameters of the ARIMA model. (The maximum likelihood parameter estimates satisfy the GLS results.) See Pierce (1971) and Findley et al. (1998). With the fitted regARIMA model available, the MBF diagnostic tests the null hypothesis

$$H_0: \beta_s = 0$$

where β_s is the subvector of β corresponding to the fixed seasonal effects (all of β if there are no other regressors). A Wald test statistic of H_0 is computed; such a statistic is asymptotically χ^2 with s − 1 degrees of freedom, under fairly broad conditions. McElroy and Holan (2014) provides details. This statistic is then rescaled by multiplying it by $(T-d-k)/\big((s-1)(T-d)\big)$, which yields an F statistic with $(s-1, T-d-k)$ degrees of freedom, where *d* is the order of (nonseasonal) differencing and *k* is the total number of regression parameters in the model (11 for monthly seasonal effects plus 6 for TD if present, 1 for an Easter effect if present, etc.) Significant rejections indicate that deterministic seasonality is indeed present. For further details, simulation results, and interpretation of the MBF, and discussion of the contrast between stable and dynamic seasonality, see Lytras et al. (2007) and Findley et al. (2017).

VS Diagnostic

The discussion in Appendix A indicates that peaks in the spectral density at seasonal frequencies indicate the presence of seasonality. Teasing out this general notion into a specific criterion is extremely subtle. As discussed at length in McElroy and Roy (2021), the pattern of seasonal autocorrelation is tied not only to whether the spectrum has a peak, but also to the convexity of the spectral density on the left and right sides of the peak, the degree to which the peak extends above the main spectral mass, and the width of the peak. Based on empirical investigations, the original VS criterion (Soukup and Findley 1999) adopted a comparison of *f* (in log scale) at a seasonal frequency λ to two neighboring values, to the left and right, separated by a distance of $\pi/60$. This comparison can be expressed as

$$\max\left\{\log f\,(\lambda) - \log f\,\left(\lambda - {}^{\pi}/_{60}\right), \log f\,(\lambda) - \log f\,\left(\lambda + {}^{\pi}/_{60}\right)\right\}.$$

This width of $\pi/60$ was determined as appropriate for a monthly time series. If these discrepancies are deemed sufficiently large relative to the overall dynamic range of the spectral density (in log scale), the peak is considered *visually significant*. Because the spectral density is not observed, it must be estimated – this presumes the series is stationary, because the concept of spectral density is not well defined for nonstationary processes. (Although there is an analogous concept of a pseudo-spectral density, we will not discuss it here.) Findley and Soukup (2000) use an autoregressive spectral estimator of order 30, which has the nice property of super-consistency for unit roots if an autoregression has been fitted (via OLS) to a unit-root process. Despite the popularity of this diagnostic, it warrants several cautions: (i) the original VS formulation provided no distribution theory; (ii) using the AR(30) can generate spurious peaks; (iii) the criterion width $\pi/60$ is insufficient to obtain correctly sized statistics given common time series sample sizes; (iv) the VS threshold requires careful calibration to notions of seasonal association.

McElroy and Roy (2021) establish a distribution theory for the VS statistic that is based on a nonparametric spectral estimator; no theory is currently available for the case of an autoregressive spectral estimator. The authors tabulated the excessive incidents of type I error resulting from the failure to account for statistical uncertainty in spectral density estimates. Use of the new distribution theory enables correct size at sample sizes $T = 600$ for the small width of $\pi/60$, but if the width is widened to $\pi/15$ adequate size can be obtained for $T = 120$. Furthermore, empirical work suggests that by fitting an AR(*p*) spectral estimator with *p* increasing from 12 to 30, one can find spectral peaks that appear or disappear at seasonal frequencies, purely as an artifact of over-fitting. We advocate selecting *p* according to the Akaike Information Criterion (AIC) to mitigate such artifacts of misspecification. If moving from monthly to quarterly data, the width used in the criterion should be multiplied by three, as the numerical work of McElroy and Roy (2021) shows. As to the VS threshold, the default VS threshold corresponds to roughly a value of $\phi_s = 0.9$ in Example A.1, and therefore seems to be compatible with a high degree of seasonal association.

Acknowledgments

We are indebted and deeply grateful to Thomas Evans, Erica Filipek, Bonnie Kegan, James Livsey, Demetra Lytras, Amy Newman-Smith, Scott Scheleur, Katherine Jenny Thompson, Richard Tiller, and Colt Viehdorfer for their careful reviews of drafts of this document.

References

Bell, W. R. 1987. A note on overdifferencing and the equivalence of seasonal time series models with monthly means and models with $(0,1,1)_{12}$ seasonal parts when Θ = 1. *Journal of Business and Economic Statistics*, **5**, 383–387.

Bell, W.R. 2004. On RegComponent time series models and their applications. In *State Space and Unobserved Component Models: Theory and Applications (Proceedings of a Conference in Honour of James Durbin)*. A. Harvey, S.J. Koopman, and N. Shephard (Eds.), 248–283. Cambridge: Cambridge University Press.

Bell, W. R. 2012. Unit root properties of seasonal adjustment and related filters (jos.nu/Articles/abstract.asp?article=283441). *Journal of Official Statistics*, **28**, 441-461.

Bell, W.R. and Hillmer, S.C. 1984. Issues involved with the seasonal adjustment of economic time series. (Invited paper with discussion.) *Journal of Business and Economic Statistics* **2**, 299–320.

Bell, W.R. and Hillmer, S.C. 1985. A reply (to comments by Sims in Bell and Hillmer (1984)). *Journal of Business and Economic Statistics* **3**, 95–97.

Box, G.E.P. and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control (2nd Ed.).* San Francisco, CA: Holden-Day.

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

Findley, D.F., Lytras, D.P., and Maravall, A. 2015. Illuminating model-based seasonal adjustment with the first order seasonal autoregressive and airline models. Census Bureau Research Report: RRS2015/02 (census.gov/content/dam/Census/library/working-papers/2015/adrm/rrs2015-02.pdf).

Findley, D.F., Lytras, D.P., and McElroy, T.S. 2017. Detecting seasonality in seasonally adjusted monthly time series. Census Bureau Research Report: RRS2017/03 (census.gov/content/dam/Census/library/working-papers/2017/adrm/rrs2017-03.pdf).

Findley, D.F. and Monsell, B.C. 1986. New techniques for determining if a time series can be seasonally adjusted reliably. In *Regional Economic Modeling*. M.R. Perryman and J.R. Schmidt (Eds.), Chapter 8, 195-228. Amsterdam: Kluwer-Nijhoff.

Findley, D.F., Monsell, B.C., Bell, W. R., Otto, M.C. and Chen, B.-C. 1998. New capabilities of the X-12-ARIMA seasonal adjustment program (with discussion) (census.gov/content/dam/Census/library/working-papers/1998/adrm/jbes98.pdf). *Journal of Business and Economic Statistics* **16**. 127–177.

Findley, D.F. and Soukup, R.J. 2000. Modeling and model selection for moving holidays (census.gov/content/dam/Census/library/working-papers/2000/adrm/asa00-eas.pdf). In Proceedings of the American Statistical Association, Business and Economic Statistics Section. Alexandria, VA. 102–107.

Gómez, V. and Maravall, A. 1997. Programs TRAMO and SEATS: Instructions for the User (researchgate.net/publication/247274792_Programs_TRAMO_and_SEATS_instructions_for_the_user). Mimeo, Banco de España.

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

Ladiray, D. and Quenneville, B. 2001. *Seasonal Adjustment with the X-11 Method*. (springer.com/us/book/9780387951713). Springer.

Lytras, D. 2013. X-13-Graph Java documentation, version 2.1 (census.gov/data/software/x13as/x13graph/x13gb-java.html).

Lytras, D.P. 2015. On seasonality: Comparing X-13ARIMA-SEATS diagnostics for quarterly series (census.gov/content/dam/Census/library/working-papers/2015/adrm/lytras2015.pdf). In JSM Proceedings, Business and Economic Statistics Section. Alexandria, VA: American Statistical Association.

Lytras, D. 2017. Win X-13 version 2.5: A Windows interface for X-13ARIMA-SEATS (census.gov/data/software/x13as/winx13/documentation.html).

Lytras, D.P., Feldpausch, R.M., and Bell, W.R. 2007. Determining seasonality: a comparison of diagnostics from X-12-ARIMA (census.gov/content/dam/Census/library/working-papers/2007/adrm/ices2007dpl.pdf). In Proceedings of the Third International Conference on Establishment Surveys. Alexandria, VA: American Statistical Association.

Maravall, A. 2012. Update of seasonality tests and automatic model identification in TRAMO-SEATS. Banco de España. (November 2012 draft.)

McElroy, T.S. 2018. Seasonal adjustment subject to accounting constraints. *Statistica Neerlandica* **72**, 574-589.

McElroy, T.S. and Holan, S.H. 2014. Asymptotic theory of cepstral random fields. *The Annals of Statistics* **42**(1), 64–86.

McElroy, T.S. and McCracken, M.W. 2017. Multistep ahead forecasting of vector time series. *Econometric Reviews* **36**(5), 495–513.

McElroy, T.S., Pang, O., and Monsell, B. 2019. Seasonal adjustment subject to frequency aggregation constraints. In Proceedings of the American Statistical Association, Business and Economic Statistics Section. Alexandria, VA.

McElroy, T.S. and Politis, D.N. 2020. *Time Series: A First Course with Bootstrap Starter*. New York: Chapman Hall.

McElroy, T.S. and Roy, A. 2021. Testing for adequacy of seasonal adjustment in the frequency domain. *Journal of Statistical Planning and Inference* **221**, 241—255. Also see Census Bureau Research Report: RRS2017/01 (census.gov/content/dam/Census/library/working-papers/2017/adrm/rrs2017-01.pdf).

Moulton, B.R. and Cowan, B.D. 2016. Residual seasonality in GFP and GDI: Findings and next steps (apps.bea.gov/scb/pdf/2016/07%20July/0716_residual_seasonality_in_gdp_and_gdi.pdf). *Survey of Current Business*, 1-6, July 2016.

Pierce, D. A. 1971. Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika* **58**, 299–312.

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 (census.gov/content/dam/Census/library/working-papers/1999/adrm/rr9903s.pdf). In Proceedings of the American Statistical Association. Alexandria, VA: American Statistical Association.

U.S. Census Bureau. 2020. X-13ARIMA-SEATS reference manual, version 1.1 (census.gov/software/x-13arima-seats/x13as/windows/documentation/docx13ashtml.pdf).