

How the ONS assesses statistical outputs for residual seasonality

Outlines the ONS's approach for the assessment of residual seasonality, including descriptions of all the tests applied.

Contact:
Gross Domestic Product team
gdp@ons.gov.uk
+44 1633 455284

Release date:
12 May 2026

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [How the ONS approaches seasonal adjustment](#)
3. [Residual seasonality and why it can occur](#)
4. [Moving seasonality](#)
5. [How we detect and monitor residual seasonality](#)
6. [Case Study – General Government, Education \(volume estimates\)](#)
7. [Aggregate residual seasonality test results: quarterly and monthly GDP](#)
8. [Testing for residual seasonality](#)
9. [Related links](#)

1 . Overview

Seasonal adjustment is a fundamental statistical process that underpins many of the Office for National Statistics' (ONS's) most widely used economic series.

Prominent aggregates, such as gross domestic product (GDP) and its components, along with a wide range of labour market indicators and other economic indicators, are adjusted to remove regular seasonal effects. By estimating and removing systematic calendar-driven fluctuations, seasonally adjusted estimates allow users to identify the underlying trends and non-seasonal movements in important economic indicators.

In a dynamic economy like the UK's, the seasonal pattern of economic activity can evolve as the economy, and the forces shaping it, change. The process for identifying regular seasonal variations requires a number of years of data to identify a pattern. When the seasonal pattern of activity changes, it can take between three and five years before one has enough observations to establish fully what the new seasonal pattern is. In periods where this occurs, close monitoring of the seasonally adjusted outputs, in line with ONS practice, is essential.

In recent years, we have seen changes in the seasonal patterns of the raw components that we use to estimate GDP. For example, prior to adjustment, the data that firms report to us in our monthly business survey now show much stronger growth at the start of the year. Our dynamic seasonal adjustment process allows us to capture seasonality as it changes.

In response, the ONS has expanded the range of tests it routinely carries out to assess whether there is any "residual seasonality" in the GDP estimates that are constructed from the seasonally adjusted components. Residual seasonality is where identifiable seasonality remains in the aggregate GDP estimate after seasonally adjusting the components of GDP.

Drawing on this expanded range of tests, we have enhanced our internal quality assurance processes, and we review how we seasonally adjust our headline aggregates on a systematic basis before every publication. Our internal "curiosity sessions," combined with methodological developments and new automated diagnostic tools, play an important role in maintaining the integrity and quality of seasonally adjusted outputs.

Recently, the Office for Statistics Regulation (OSR) completed a [compliance review on the treatment of seasonality in quarterly GDP statistics](#). The OSR found that: "ONS has followed internationally recognised best practice in its approach to seasonal adjustment, and current statistical tests show no evidence of significant residual seasonality in quarterly or monthly GDP. However, because emerging seasonal patterns can take several years to detect using standard methods, there remains a risk that early signs of change may not yet be visible in statistical tests.

It is therefore important for producers, users and commentators to keep an open mind about the possibility of newly evolving trends. To strengthen confidence and manage this risk, ONS should continue increasing transparency around its methods and uncertainties, rebuild and stabilise the specialist team responsible for seasonal adjustment, and seek external assurance, particularly on the detection of emerging seasonal signals. Doing so will ensure that its approach remains robust, timely and benefits from challenge."

OSR also made a number of recommendations, to which the ONS responded positively in a [letter to OSR on the treatment of seasonality in quarterly GDP](#). The ONS has already made progress on these recommendations. In line with both the recommendations and ongoing engagement on this topic with important users, including the Bank of England, the ONS will soon publish non-seasonally adjusted GDP data at the component level to complement the non-seasonally adjusted series for headline GDP that is already published. This will allow users to see how seasonal patterns are changing at the component level. This article is itself also a response to the OSR recommendations.

This article summarises the concept of residual seasonality, describes how the ONS addresses it in practice, outlines the statistical tests currently in use, and discusses the challenges posed by evolving and moving seasonal patterns.

Other countries face similar challenges on residual seasonality. The Bureau of Economic Analysis (USA) provide detail on their approach in their [SCB, Seasonal Adjustment in the National Income and Product Accounts, August 2018 update](#). We have engaged with them and follow very similar principles and use similar tests.

Our approach to testing for residual seasonality, applied to indirectly seasonally adjusted aggregate series, is summarised as the following (where tests are described in Section 8):

1. Apply the model-based F-test to the full span of the series and the last five years, flagging any p-value less than 0.01.
2. Apply the QS test to the full span and the last five years, flagging any p-value less than 0.01.
3. If both tests flag an issue with residual seasonality for either the full or the five-year span, explore measures that could be taken to address the issue.

In this methodology article, [Section 8: Testing for residual seasonality](#) is authored by Professor Paul A Smith from the University of Southampton. It provides independent commentary on the state of the literature and the methodological tools available. We are grateful to Professor Paul A Smith for his reflections, which highlight the relative strengths and weaknesses of the various tests and have helped shape and inform our approach.

2 . How the ONS approaches seasonal adjustment

Scope of seasonally adjusted outputs

The Office for National Statistics (ONS) maintains seasonal adjustment models for around 50 different output areas (covering economic and social time series), and a total of around 10,000 individual time series.

The majority of these seasonal adjustment models are reviewed annually as new data become available. Those reviewed less frequently are usually not seasonal or they contribute a small proportion to headline aggregates. For the main components of national accounts, an annual review takes place during the window in which data are revised and compiled for the annual [Blue Book](#) and [Pink Book](#) publications. These seasonal adjustment model specifications are then used for the following year, unless quality assurance identifies an issue during the regular production process.

In addition to annual reviews, monthly and quarterly GDP have additional reviews as required, where a subset of series is identified for consideration. Monthly and quarterly residual seasonality checks are run by experts to identify any causes for concern, and these checks are used to pinpoint components that require attention outside of the regular annual reviews.

The review process involves careful consideration of a range of model diagnostics and possible tools to achieve the highest quality seasonal adjustment, while also aiming to minimise revisions. Events occurring in specific series that could cause significant distortion to the estimated seasonal factors are brought to the attention of time series methodologists so that models can be revised as necessary in real time before publication. Prior adjustments, for example additive outliers, level shifts, seasonal breaks, trading day and other effects, are reviewed and updated where relevant. It should be noted that the X-11 process, which is used to estimate the seasonal factors in the X-13-ARIMA-SEATS package, is also reasonably robust to the presence of outliers as the algorithm adjusts for outliers present in the input data.

Indirect and direct seasonal adjustment

In the UK, seasonal adjustment of GDP follows a well-established indirect approach where individual component series are separately seasonally adjusted to account for different seasonal patterns of the components. These component-level seasonal adjustments are then aggregated to produce higher-level seasonally adjusted estimates. This indirect approach ensures that contributions from components remain consistent with the behaviour of the aggregate estimates. In domains such as economic statistics, particularly the National Accounts, users place importance on the consistency of a dataset in an aggregation framework. Consistent frameworks can aid interpretation, particularly where changes and impacts at an aggregated level can be traced back to more detailed data.

The alternative, direct seasonal adjustment, fits the most appropriate seasonal adjustment model directly to the non-seasonally adjusted aggregate series. In most circumstances, this does not lead to a seasonally adjusted aggregate series that is consistent with the aggregated seasonally adjusted component time series. This situation may have a range of possible causes, for example different effects and outliers being found significant at different levels of aggregation or in different components, application of constraining or benchmarking techniques, or even different decompositions or moving averages being found optimal for different series.

In practice, there is no clear agreement on whether indirect or direct seasonal adjustment is preferred since neither theoretical nor empirical evidence uniformly favours one approach over the other. More detail is available in Section 5.4, page 23 in the [European Statistical Service guidelines on seasonal adjustment – 2024 edition](#).

One of the implications of indirect seasonal adjustment is that additional quality checks, for example, the assessment for residual seasonality, need to be completed on the aggregate seasonally adjusted estimates before publication. The ONS performs these additional checks as part of our routine quality assurance for each monthly and quarterly GDP publication.

3 . Residual seasonality and why it can occur

Residual seasonality is the presence of statistically significant seasonal patterns in a series after it has been either indirectly or directly seasonally adjusted.

For a directly seasonally adjusted time series, a well-specified seasonal adjustment procedure should remove these systematic calendar-related patterns entirely. Any persistence in a directly seasonally adjusted time series may indicate either that the underlying model for seasonal adjustment is misspecified, or that seasonal patterns have changed too rapidly for the seasonal adjustment method to adequately cope.

For a time series that is indirectly seasonally adjusted, residual seasonality in the aggregate estimates can indicate that there are issues that need to be identified and resolved at a more detailed level. These issues could occur in the directly seasonally adjusted component time series, or separately as part of the aggregation process such as chain-linking or benchmarking. Sometimes, these issues can be obvious, such as a misspecification of the seasonal adjustment parameters. In other cases, they can be more subtle, where seasonal effects are not statistically significant on a case-by-case basis, but may become statistically significant when aggregated. Therefore, close monitoring is required to both assess and resolve any issues in the aggregate estimates.

The ONS actively manages the risk of residual seasonality, by scheduling annual reviews of all seasonally adjusted outputs to ensure models are still appropriate given the latest data added to time series. Rapid intervention is also used as part of regular monthly and quarterly production cycles before publication.

4 . Moving seasonality

A distinction must be drawn between residual seasonality, which is stable or recurring patterns that are still present in the seasonally adjusted estimates, and the concept of moving seasonality (also called dynamic or evolving seasonality).

Moving seasonality arises when seasonal patterns change gradually over time, either in timing, duration, or intensity. These changes may result from economic restructuring, policy reforms, measurement improvements, or shifts in reporting behaviour. Unlike residual seasonality, where existing models fail to remove stable seasonal features, evolving seasonality requires dynamic modelling approaches capable of capturing changing seasonal structures.

From a practical perspective, seasonality is hardly ever constant or fixed over time and can evolve gradually or abruptly, depending on the real-world events causing it. Figure 1 shows how the non-seasonally adjusted data for aggregate quarterly GDP (in volume terms) is both strongly seasonal but has also evolved in recent years. The published seasonally adjusted estimates are also presented. These aggregate seasonally adjusted GDP estimates are calculated on an indirect basis using detailed seasonally adjusted data. Separately, Figure 2 shows the quarterly growth rates for both the non-seasonally adjusted and seasonally adjusted data including a restricted span to focus on the most recent years after the coronavirus (COVID-19) pandemic period.

These two charts together show that prior to the pandemic, GDP increased strongly in the final quarter of the year before falling sharply in the first quarter of the following year. For example, in the two years before the pandemic, the average of non-seasonally adjusted GDP growth was negative 3.4% in the first quarter (Figure 2). This pattern has changed substantively with a more muted recovery in GDP in the final quarter of the year and smaller falls in the first quarter. The average of the non-seasonally adjusted growth in the first quarter of the last two years was negative 0.9%, 2.6 percentage points higher than before the pandemic.

Comparing our non-seasonally adjusted data to our seasonally adjusted data, on average, seasonally-adjusted growth was 0.3% in the first quarter two years before the pandemic, and 0.7% for the past two years. This means that seasonally adjusted growth for the past two years was 0.4 percentage points higher than it was in the first quarter two years before the pandemic. This gap is around 85% lower in percentage-point terms than for the non-seasonally adjusted data. This example shows how our dynamic seasonal adjustment has adapted to capture additional detail in quarterly growth figures, as seen recently at the start of 2026.

However, as we explain in this article, it can take three to five years of data to establish what the regular seasonal pattern is and what is a more idiosyncratic variation in growth over the course of a year. Given that it takes time for seasonal factors to fully reflect new patterns, as more data become available, we will continue to keep our seasonal adjustments under close review. With each further quarter of data, we have more information to discern what the regular pattern has been and, therefore, more evidence to update our seasonal adjustments.

Figure 1: Quarterly gross domestic product at market prices: Chained volume measure – non-seasonally adjusted and seasonally adjusted time series

Notes:

1. The data span from Quarter 1 (Jan to Mar) 2018 to Quarter 4 (Oct to Dec) 2025, a restricted time span of the whole series.
2. The data show quarterly gross domestic product at market prices: Chained volume measure.
3. The data show both the non-seasonally adjusted data and the published indirect seasonally adjusted data.

Figure 2: Quarterly growth rates for gross domestic product: Chained volume measure - non-seasonally adjusted and seasonally adjusted time series

Notes:

1. The data spans from Quarter 2 (Apr to June) 1955 to Q4 (Oct to Dec) 2025, with data highlighted from Quarter 1 (Jan to Mar) 2018 onwards.
2. The data show quarterly growth rates for the gross domestic product at market prices: Chained volume measure.
3. The data show both the non-seasonally adjusted data and the published indirect seasonally adjusted data.

A different way to view the seasonal patterns within the quarterly GDP estimate is shown by Figure 3, which gives an illustration of how seasonality evolves over time. For this example, we have used a direct seasonal adjustment of the aggregate non-seasonally adjusted GDP data. This will not match the published indirect outputs but is useful to illustrate the evolution of seasonality in different quarters. This shows that seasonality is never constant, and in recent years, activity in Quarter 1 (Jan to Mar) has increased, while other quarters have decreased (for example, Quarter 2 (Apr to June) and Quarter 4 (Oct to Dec)). The standard seasonal adjustment packages allow us to adequately capture this evolution. Conceptually, seasonality is neutral over a year, so if activity in one period increases, we would expect to see an offset in other periods.

Figure 3: Illustration of the evolution of the quarterly GDP seasonal factors obtained by applying a direct seasonal adjustment to aggregate quarterly GDP

Notes:

1. Data cover the period from Quarter 1 (Jan to Mar) 1955 to Quarter 4 (Oct to Dec) 2025.
2. Calculated by using JDemetra+ v3.7.1, using a direct seasonal adjustment of the aggregate non-seasonally adjusted quarterly GDP estimates. This provides an illustration of how seasonality evolves over time.
3. The figure shows the seasonal x irregular ratios, where the seasonal x irregular factor (D8) and the seasonal factor (D10) are calculated for a direct seasonal adjustment of UK GDP quarterly volume.
4. In figure: the blue line is the estimated seasonal factor, and the green dots represent the irregular combined with the seasonal component.
5. This is an illustrative seasonally adjusted estimate for quarterly GDP, which takes the non-seasonally adjusted aggregate dataset and seasonally adjusts these data directly. Importantly, this will not give exactly the same results as the official indirect seasonal adjustment, which uses more detailed calculations to account for different seasonal patterns in the component time series.
6. The download data are grouped by quarter and ordered by year.

The aim of seasonal adjustment is to estimate and remove systematic and calendar-related patterns, including seasonality that evolves over time. However, when seasonality changes abruptly, this can require expert intervention to apply a "seasonal break" prior correction to ensure the seasonal factor estimation is not adversely affected. There are statistical tests for seasonal breaks, and when a significant result is obtained this type of prior correction can be applied as part of a regular seasonal adjustment review. This approach requires expert assessment to ensure prior corrections are applied correctly to treat the change in seasonal pattern and typically requires at least three years of data and a valid real-world reason. Evidence of new seasonal patterns accumulates slowly, and one or two data points do not give sufficient evidence on their own to identify a change.

As new data become available each month or quarter, this can affect the seasonally adjusted estimates, including historically, as new patterns can emerge. For example, where there is an emergence of level shifts, changes in seasonality, or large extreme impacts, the estimation of the seasonal factor will be affected. It is generally difficult to distinguish changes in the seasonal pattern from the random variation in a series, so it is only as evidence accumulates over multiple periods (multiple observations in the same month or quarter of the year for seasonality) that any change in the pattern becomes discernible.

Understanding and knowledge of real-world events is crucial to inform interpretation of seasonally adjusted estimates. This also provides confidence in applying corrections when there are real-world reasons. For example, changes in policy may affect timing of economic activity as businesses and users adapt, and one-off events such as weather, or additional Bank holidays will also influence behaviour. While gradual changes can be captured as part of evolving seasonality, patterns that emerge need to be closely monitored and intervention applied if needed, particularly at the current end of the time series. The illustrative example in Figure 3 shows economic activity increasing in Quarter 1, but with a similar offsetting reduction in Quarter 2. This could suggest that business behaviour is changing in recent periods, as activity moves between those quarters, although the historical seasonal factor path shows similar movements. This is precisely where tests for residual seasonality are useful. Our tests in [Section 7](#) (Table 2) show there is no residual seasonality in monthly or quarterly GDP, over both the full span and more recent five-year span, even though seasonality continues to evolve.

The nature of seasonal adjustment means that back data can also be revised once new information is available, for example, the latest data (for example, for May) will help us also understand previous years' data for that same month or quarter. For ONS GDP estimates, there is a transparent [National Accounts revision policy](#), which is updated as required. The revisions policy allows us to update the back series where new data points add to the evidence that seasonality is changing. In recent periods, the ONS has opened up additional time periods to allow the previously published seasonally adjusted estimates to be revised.

5 . How we detect and monitor residual seasonality

The Office for National Statistics (ONS) has strengthened its approach to detecting and addressing residual seasonality over recent years, as reported in our [Assessing residual seasonality in published outputs methodology](#) and in discussion with the [Office for Statistics Regulation \(OSR\)](#). Each output already has an established quality assurance process, and we have strengthened this with the inclusion of additional formal statistical tests for residual seasonality on directly and indirectly adjusted series. Regular reviews of headline aggregates are now embedded within internal quality assurance processes, where analysts examine external contextual evidence alongside statistical diagnostics to identify any emerging seasonal anomalies as part of structured "curiosity sessions."

Seasonal adjustment updates occur during the annual National Accounts cycle. Changes applied through the Blue Book and Pink Book revisions, such as methodological improvements, new data sources, or rebalancing of historical estimates, necessitate the reestimation of seasonal adjustment parameters. The ONS now publishes an [annual assessment of residual seasonality](#) in GDP components to accompany these revisions, ensuring transparency for users and enabling consistent monitoring from year to year.

There are many tests available, as shown in [Section 8: Testing for residual seasonality](#) to detect the presence of residual seasonality, and it is not straightforward to choose a critical value for statistical significance at which intervention should take place. Intervening too frequently may compromise the statistical properties of the time series and the principle of parsimony (or avoiding over-adjustment). There is a risk that signs of (what will prove to be) change are not visible in statistical tests (through lack of power in the test leading to type II error). There is also a risk that apparent signs of change are just part of the natural variation in the series (so that significant test results are actually Type I errors). Deciding on suitable criteria to balance these two alternatives is difficult.

Our approach is very similar to that laid out by the United States Bureau of Economic Analysis (BEA) in their [2018 methodology article](#) describing their criteria for assessing residual seasonality.

We are currently using two tests that are readily available in existing seasonal adjustment software. Where Section 8 describes the different tests, these are:

- Model-based F-test
- QS test

Intervention is likely to be necessary if the p-value for both tests is less than an established critical p-value value. As recommended in the [recent review by the Office for Statistics Regulation](#), we will also use a second, greater p-value, below which the results will be reported. These results may be interpreted as an early sign that seasonal patterns are changing and may require intervention at a later stage. These settings are:

- critical p-value for intervention: 0.01
- critical p-value to report: 0.05

For monthly time series, we test for residual seasonality in the last five years and in the full span. For quarterly series, [US Census Bureau guidelines](#) recommend a longer minimum span of 15 years because of the smaller number of data points. However, to provide reassurance to users that more recent seasonal changes will be picked up, we test five-year spans as well as the full span. Testing additional intermediate spans is avoided, as this increases the probability of a false positive result and then intervening unnecessarily.

These tests have been built into a reproducible automatic process to be run on the top-level aggregates before each statistical GDP-related release:

- Average GDP
- Output approach to measuring GDP
- Expenditure approach to measuring GDP
- Income approach to measuring GDP
- Monthly GDP
- Index of Services
- Index of Production (and manufacturing)
- Construction

The TRAMO-SEATS algorithm implemented in RJDemetra+ is used for the trend identification necessary to obtain test results. We use the results for the linearised series, which means that deterministic effects, such as outliers and trading days, are removed before applying the tests. We therefore now take a transparent and consistent approach, which is similar to that taken by the US BEA.

6 . Case Study – General Government, Education (volume estimates)

Within the ONS, and in particular within the National Accounts, seasonal adjustment processing is applied to many thousands of time series.

To show the type of intervention that may be applied in practice, we describe in detail an example series where a change to the model specification has been made after routine residual seasonality checks. This intervention reduced the residual seasonality detected by the statistical tests used.

The general government education series had historically been treated as non-seasonal; therefore, no seasonal adjustment was applied. From 2020, the data supply relating to pupil attendance began to change and by 2021, a monthly data feed was available, so quarterly estimates could be used in place of an annual series. While running routine residual seasonality checks, this component was identified as contributing to residual seasonality in an aggregate component series. A possible seasonal break was tested and found significant and therefore confirmed as the best treatment to use. There are three full years of data following the seasonal break and a clear stable seasonal pattern, which is sufficient to estimate the seasonal factors adequately for this period.

Table 1 shows the results of the two statistical tests for residual seasonality described above. The p-value before intervention was below the level set where intervention is required. The test results show that no issues remain after the intervention.

Table 1: Residual seasonality tests for case study – general government, education (CVM)

Tests for seasonality on linearized (SA series) - TRAMO-SEATS (five-year span)				
	Test	Test statistic	p-value	Findings
Before intervention	F-test	38.7421	<0.0001	Residual seasonality
	QS	23.2935	<0.0001	Residual seasonality
After intervention	F-test	0.4694	0.7083	No issues
	QS	0	1	No issues

Source: ONS calculations

Notes

1. These test results are calculated using the linearised time series from TRAMO-SEATS (using the RSA Full setting in the software JDemetra+). A p-value greater than 0.01 indicates there is no significant effect.
2. The F-test used is the model-based F-test, described in Section 8 and in JDemetra+ documentation.
3. The QS test is as described in Section 8 and in JDemetra+ documentation.
4. "Before intervention" shows the test results using the previous approach of not applying seasonal adjustment.
5. "After intervention" shows the test results after applying a seasonal break to seasonally adjust from Quarter 3 (July to Sept) 2022 onwards.

The following plots show a limited span of the series to illustrate these recent features and the seasonal break applied. Figure 4 shows the original series and the new seasonally adjusted series. The new seasonally adjusted series successfully addresses the regular peaks observed in Quarter 3 (July to Sept) of the original series.

Figure 4: General government, Education – Original and seasonally adjusted time series

Notes:

1. This time series spans from Quarter 1 (Jan to Mar) 1995 to Quarter 4 (Oct to Dec) 2025. The reduced range in the plot (Quarter 1 2015 to Quarter 4 2025) has been chosen to show more clearly the introduction of a seasonal break in Quarter 3 (July to Sept) 2022.
2. The data show the GDP contributions of general government education.
3. These data show both: the non-seasonally adjusted (Non-SA) data (equivalent to the previous seasonal adjustment specification); and the seasonally adjusted data (using the recently implemented specification).
4. "Original Non-SA" is the original non-seasonally adjusted time series, which is what was used as the seasonally adjusted time series before the seasonal break was introduced.
5. "SA" is the seasonally adjusted time series using the new specification.
6. The y-axis has been restricted to make the recent seasonality visible. This causes the values of several outliers in the coronavirus (COVID-19) pandemic not to be shown, but these values are available in the data download.

Figure 5 shows the new seasonal break applied at Quarter 3 of 2022. A seasonal break is appropriate when there is an abrupt and sustained change in the seasonal pattern, ideally when the reason for this change is known.

The points represent the seasonal and irregular components combined, after the estimated trend has been subtracted (in this series an additive relationship of the components was used, so the seasonal factor is centred around zero). The green line shows the estimated seasonal factors, which track the change in the seasonal-irregular factors closely, and adequately capture the abrupt change in seasonality.

Figure 5 Seasonal-irregular factors and estimated seasonal factors after the seasonal break has been applied.**Notes:**

1. This time series spans from Quarter 1 (Jan to Mar) 1995 to Quarter 4 (Oct to Dec) 2025. The reduced range in the plot (Quarter 1 2015 to Quarter 4 2025) has been chosen to show more clearly the introduction of a seasonal break in Quarter 3 (July to Sept) 2022.
2. These are the combined seasonal and irregular (SI) factors, and the estimated seasonal factor for a direct seasonal adjustment of general government education GDP quarterly volume.
3. The seasonal adjustment has been carried out treating the series as additive. Therefore, the SI factors are calculated as: seasonal + irregular.
4. "SI Factor d8 (corrected)" is the combined seasonal and irregular components of the time series, after fitting the specified time series model using X13-ARIMA-SEATS.
5. The d8 table output by X13-ARIMA-SEATS does not show the step change in the SI factors, as the seasonal break is implemented by applying a prior adjustment to each quarter up to the seasonal break. This prior adjustment has been reversed to show the SI factors as they would be in the original series.
6. In the model specification, several data points during the coronavirus (COVID-19) pandemic are identified as additive outliers, which are treated separately, therefore the Seasonal and Irregular component shows less variation than might be expected when looking at the original series.
7. "Seasonal Factor (d10)" is the estimated seasonal factor calculated by the X-11 algorithm, after prior adjustments (for example additive outliers and the seasonal break) have been accounted for.

7 . Aggregate residual seasonality test results: quarterly and monthly GDP

[Section 8: Testing for residual seasonality](#) describes a wide range of different tests. However, we have decided to place the greatest emphasis on the model-based F-test and QS test, using the criteria described in [Section 6: Case Study – general government, education \(volume estimates\)](#). This is aligned with the approach of the United States Bureau of Economic Analysis.

In [previous Office for National Statistics publications](#), results for the Kruskal-Wallis, Friedman, and periodogram tests were included for completeness. However, the findings from [University of Southampton](#) highlight several weaknesses of those specific tests and the disadvantages of using too many tests. While there are currently no causes for concern in the results from these other tests that are not shown in Table 2, we have adapted our approach to focus on a set of standard tests and spans. These are shown in Table 2 with an example for aggregate quarterly and monthly gross domestic product (GDP). The tests show no residual seasonality is present.

Table 2: Residual seasonality test results for quarterly and monthly GDP using different spans

	Test	Full span		Five-year span	
		Test statistic	p-value	Test statistic	p-value
Quarterly GDP to Q4 2025 (run on 17 March 2026)	F-test	0.2018	0.8951	1.5735	0.2401
	QS	4.0869	0.1296	0.59	0.7445
Monthly GDP to Feb 2026 (run on 10 April 2026)	F-test	0.6328	0.8007	0.613	0.8081
	QS	0	1	0	1.0000

Notes

1. These tests are calculated using the linearised time series from TRAMO-SEATS (using the RSA Full setting in the software JDemetra+). A p-value greater than 0.01 indicates there is no significant effect.
2. The F-test used is the model-based F-test, described in Section 8 and in JDemetra+ documentation.
3. The QS test is as described in Section 8 and in JDemetra+ documentation.

8 . Testing for residual seasonality

The commentary in this section has been contributed by Paul A Smith, Professor of Official Statistics at the University of Southampton. His reflections have informed our approach.

Tests for seasonality

There are many tests for seasonality of different types, which are used as part of the regular seasonal adjustment processes. The Office for National Statistics (ONS) [has previously described the tests](#) that it uses for this purpose; there is more information in Section 2.1 of McElroy and Roy's 'A review of seasonal adjustment diagnostics'. For tests used within the JDemetra+ seasonal adjustment package (endorsed by Eurostat) see the [JDemetra+ Github page on Seasonality tests](#).

In this section, we describe some of the available tests for seasonality, and what types of seasonality they may be able to detect, including how many of the tests work. McElroy and Roy give a longer list that is focused on testing for seasonality, rather than specifically for residual seasonality. We also discuss several additional factors requiring careful consideration, including what spans should be tested and multiple testing.

Tests for residual seasonality

Tests for seasonality are not necessarily appropriate for testing for residual seasonality. This is because the characteristics of a series that has already been seasonally adjusted are different from those of a completely unadjusted series. In particular, it should not have any seasonal unit roots but may have some small seasonality if it has been indirectly adjusted.

In general, these tests must be applied to stationary series with no trend, so series need to be seasonally differenced appropriately or have the trend component removed. Sometimes this is accomplished by applying the tests to the seasonal \times irregular (SI) ratios, but this is essentially equivalent.

Tests for residual seasonality based on analysis of variance

Seasonal dummy regression

Fit a regression model with dummy variables for $S-1$ seasonal effects (by comparison with the S th effect, which acts as a reference category; S is the number of periods in a year). A naïve application assumes that the errors are independent and identically distributed, which is unlikely to be true in a seasonally adjusted (SA) time series. Ordinary least squares (OLS) and generalised least squares (GLS) approaches to fitting have been proposed. GLS allows more complex variance structures, which can make the test more robust to departures from the assumption of independent and identically distributed errors.

Friedman test (or Stable Seasonality test)

A Friedman test is a nonparametric analysis of variance (ANOVA) (Friedman 1937), where the observations within blocks are given ranks. The blocks are years, and the S periods are the treatments; in each year the smallest value is given rank 1, the next smallest is rank 2, up to rank S , and this repeats in the next year. This tests whether there is a difference between the periods, based on the allocated ranks. Using the ranks avoids the assumption that the errors are normally distributed but does require that the treatments are independent and identically distributed. This may be more likely to be true in assessing residual seasonality. However, in general, the independence assumption is not met in time series, so the application of this method is questionable (McElroy and Roy 2022).

Kruskal-Wallis test

The Kruskal-Wallis test is another nonparametric ANOVA (Kruskal and Wallis 1952). It is based on ranking all observations in the detrended series (instead of ranking within years). So, the smallest observation gets rank 1, the next smallest gets rank 2, up to rank T , across the length of the whole time series. It tests the differences in the ranks between the S seasonal periods, which is a general nonparametric test that is applied to time series data. This test makes the same assumptions as the Friedman test, and so its use is also questionable with time series data (McElroy and Roy 2022).

Model-based F

A model-based F (MBF) tests for the joint significance of the set of seasonal dummies in a seasonal regARIMA model, using a Wald test (with a correction to account for estimating the residual variance). X-13-ARIMA-SEATS has a standard implementation of the MBF based on an ARIMA model that is suitable for the series and contains S-1 dummies for the S periods per year. The ARIMA Model part allows for the errors to be serially correlated, so this method is appropriate for time series data.

With direct seasonal adjustment, the seasonal filter removes any stable seasonal pattern that is present across the whole span of a series. This means that "any testing for residual seasonality in directly adjusted series should include only a subspan of the data", according to Bell and others' [Identifying seasonality article \(PDF, 2.3MB\)](#).

D8 F-test

The D8 F-test is one of the standard tests in X-13-ARIMA-SEATS (named after the table in which it appears). It is also a test for the joint significance of seasonal dummies (as in the MBF). However, this time it uses the detrended series and does not account for the serial correlation (which is induced by detrending, even if it is not present in the original series, according to Lytras and others' [Determining seasonality: a comparison of diagnostics from X-12-ARIMA article \(PDF, 264KB\)](#)). It is usual to make the test conservative to account for the serial correlation by taking a critical value of 7. However, Lytras and others show that this gives quite variable significance and power properties, depending on the underlying model. Therefore, the MBF is preferred.

M7

M7 is another test from X-13-ARIMA-SEATS, which combines the D8 F-test and a similar test for moving seasonality that is based on a two-way ANOVA for differences in the detrended series between years, accounting for differences in periods. These are scaled in such a way that values less than 1 indicate the presence of seasonality. If the value is greater than 1, then there is either no seasonality or there is seasonality that cannot be detected by the test.

Tests for residual seasonality based on the spectrum

The spectrum is a decomposition of the autocovariance function into a sum of sinusoids (waves) of different frequencies. If there is a distinct peak in the spectrum at a seasonal frequency, it suggests seasonality in the series. This leads to the following tests. The p-values should be adjusted to account for the multiple testing at several seasonal frequencies using the family-wise error rate (FWER).

Spectral peak

Spectral peak (SP) tests for a seasonal peak as a jump in the spectrum of the seasonally adjusted (SA) series against the background variance of the estimated spectrum. For a more detailed description, see Lytras and others' [Determining seasonality: a comparison of diagnostics from X-12-ARIMA article](#). It is possible to construct equivalent tests in the frequency domain (McElroy and Roy 2022).

Visual significance

Visual significant (VS) looks for seasonal peaks in the spectrum of the SA series. A peak is identified by comparing the value of the spectrum at the seasonal frequencies with the values of neighbouring areas of the spectrum. A nonparametric approach to obtain a p-value is derived in McElroy and Roy (2022). If multiple frequencies are to be tested, then the p-values should be adjusted to control the FWER.

Spectral convexity

The spectral convexity (SC) test also looks for seasonal peaks in the spectrum. However, in this case, the presence of a peak is detected using the slope and curvature of the spectrum. This is done by calculating the first and second derivatives of a kernel density estimate (with an appropriate bandwidth) of the spectrum, and then testing whether the first derivative is significantly different from zero; if it is not, then there is further testing of whether the second derivative is significantly smaller than zero.

Tests for residual seasonality based on the autocovariances

QS

The QS test is based on the idea that a seasonal series has positive autocorrelation at the seasonal frequencies, which is necessarily true. However, having seasonality is not the only reason that such autocorrelations may arise. The test is based on an assessment of whether the autocorrelations at lag S and $2S$ are both significantly greater than zero. The test can therefore assess that there is no evidence that seasonality is present. This is not the same as a test that "seasonality is not present" – a significant test result indicates only that there is significant autocorrelation at the seasonal frequency, which may or may not be because of seasonality.

A better logic may be obtained by a one-sided test of the autocorrelation parameter at lag S ; the null hypothesis is that some relatively large value that is consistent with seasonality is present (McElroy and Roy (2022) suggest this value is 0.7). Then if the test result is that the observed autocorrelation is significantly smaller, we can conclude that there is evidence that there is no seasonality. However, this is not the QS test.

ROOT

In the ROOT test "seasonality is assessed through oscillations in the autocovariance sequence", according to Chen and others (2022). This can be converted to a Wald test of the solutions of a complex valued equation (using the Wold representation of the series) at each seasonal frequency. These solutions are for specific values of the seasonal persistence, which must be chosen. It is therefore necessary to test at many different values of ρ . Chen and others (2022) suggest using $\hat{\rho} \in [0.98, 1]$. Then if all results are not significant, we may conclude that there is no seasonality of such persistence present.

In finite samples, the test may be biased but the bias tends to zero as the sample size increases (though not in a simple way). As a result, McElroy and Roy (2022) recommend that at least 20 years of data are used when implementing this test.

Other tests for residual seasonality

Lunsford (2017) approach

Lunsford (2017) suggests an approach for assessing residual seasonality in United States gross domestic product (GDP). It has two phases:

1. detrending the original series using a relatively straightforward filter
2. forming the means of the resulting series by period (S)

The test is whether the means of periods 1 to S are different from zero, using a variance estimate derived from a low frequency model of the detrended series. This variance can account for autocorrelation, so the approach is similar in concept to the MBF. However, the properties of this test have not been investigated in the context of residual seasonality. An adjustment of the size of the test to control the FWER for multiple testing of the S periods would be appropriate but has not apparently been implemented.

HEGY test

The HEGY test (Hylleberg and others 1990) is a test for unit roots, allowing for separate testing of the seasonal and non-seasonal unit roots. This test is appropriate for series that are nonstationary. It is therefore not suitable for testing for residual seasonality, which is based on series that have already been made stationary by differencing and/or seasonal differencing.

Summary of tests

Many of these tests for seasonality have shortcomings, either in general because they require assumptions of independent and identically distributed errors that are unlikely to be appropriate for time series data, or specifically for residual seasonality because of the properties of the already seasonally adjusted series. For example, the HEGY test is for a type of seasonality that must be removed during seasonal adjustment.

However, several tests are appropriate to this task, including MBF (and the Lunsford approach which seems analogous), VS and SC, QS and ROOT, though even in some of these cases some doubts have been expressed by McElroy and Roy (2022) and Chen and others (2022). Only the tests based on spectral peaks or the autocovariance functions are suitable for testing moving seasonality. Though all the tests are for the same property (seasonality), they actually test quite different statistics, so different results from the same input series are a plausible outcome.

Considerations in testing for residual seasonality

Tests for residual seasonality are also affected by some properties of the procedures and the tests. The topics in this subsection give an overview of the relevant features and considerations.

Edge effects at the beginning and end of the time series

The X-11 seasonal adjustment programs use Henderson moving averages. Extending the series by forecasting with an appropriate ARIMA model before applying the Henderson filter improves the seasonal adjustment (which was the basis of the extension to X-11-ARIMA). However, the forecasting may induce local non-stationarity in the ends of the adjusted time series (Chen and others 2022, page 405), which means that tests for residual seasonality in these periods are not appropriate. So, tests involving the first or last three to five years of a series should be interpreted more cautiously.

Aggregation effects

Some series are collected, processed and seasonally adjusted monthly, but included in higher level statistical series (such as GDP) as quarterly aggregates. It is sometimes possible for the monthly series aggregated to quarters to show residual seasonality, in a similar way to the aggregation of component (quarterly) series in indirect seasonal adjustment. Any of the tests can be applied to the aggregated series; the QS test of the aggregated series is included in X-13-ARIMA-SEATS and so is readily available.

Seasonal unit roots

A series with a seasonal unit root has non-stationary seasonality and requires differencing at the seasonal frequency to become stationary. This means that it is a process with a long memory, so the seasonality gradually evolves as new observations are added. This is a type of moving (or evolving) seasonality (for more information see Section 13.4 in Chatfield (2003)). Seasonally adjusted series will have been differenced appropriately in the seasonal part, so there should be no seasonal unit roots, and most tests assume they are absent.

Periods for assessing residual seasonality

Tests for residual seasonality over short spans are likely to be misleading, as there is insufficient data to draw firm conclusions. The United States [Census Bureau guidelines](#) recommend that evaluation of residual seasonality needs 60 observations (5 years for a monthly series or 15 years for a quarterly one). While this may seem conservative, it shows the challenges in ensuring a reliable assessment in a practical context.

Chen and others (2022) review the analyses of United States National Income and Product Accounts data, aimed at detecting whether there is residual seasonality; they found that different studies draw different conclusions. They assess that "overall, the findings in each case appear to be sensitive to the methods and sample spans used" (Chen and others 2022, page 400).

Chen and others (2022) then assess the case for residual seasonality in GDP and its components in the United States (which are quarterly series). They are very careful to explain exactly what test results relate to; for example "there is significant residual seasonality in series X for the period Jan 2000 to Dec 2015", and not just "there is significant residual seasonality in series X". The Office for National Statistics (ONS) should similarly be careful with the description of any test results.

Size and power

Lytras and others undertake an extensive evaluation of different tests for seasonality, using the D8 F-test, M7, SP, and MBF. Of these, only the MBF maintains the nominal type I error across a range of seasonal models. Most of the tests have good power to detect seasonality across a range of models. The power is naturally lowest when the seasonality is smaller than the noise in the series. The SP has lower power than the other tests considered.

Some more investigation of power properties of tests specifically in the situation of residual seasonality would be valuable.

Significance and multiple testing

The ONS seasonally adjusts many time series. So, under the null hypothesis, we expect some significant results that are actually false positives (that is, the test result is significant even though there is no residual seasonality). If there were 10,000 series – all conforming with the null hypothesis and using p equals 0.01 as the critical value – from a statistical perspective, we would expect 100 false positives. When making multiple tests like this, it is usual to adjust the p-values to account for this and control the FWER.

In a similar way, some of the tests for residual seasonality involve multiple assessments, either at different frequencies or for different periods. So it is also appropriate to adjust the p-values. McElroy and Roy (2022) give an example of where this is done.

Finally, the assessment of residual seasonality as described in [Section 5: How we detect and monitor residual seasonality](#) also involves repeated testing of the same series as each new data point is added. Each test is strongly correlated with the same test at the preceding period. It is not known how to adjust the p-values for multiple testing in this situation, but a simple strategy is to use a conservative p-value – this is reflected in the ONS's approach.

References

Chatfield C (2003), 'The analysis of time series: an introduction', 6th edition, CRC Press, Boca Raton.

Chen B, McElroy T and Pang O (2022), 'Assessing residual seasonality in the US national income and product accounts aggregates', *Journal of Official Statistics*, Volume 38, pages 399 to 428.

Friedman, M (1937), 'The use of ranks to avoid the assumption of normality implicit in the analysis of variance', *Journal of the American Statistical Association*, Volume 32, Issue 200, pages 675 to 701.

Hylleberg S, Engle R, Granger C, and Yoo B (1990), 'Seasonal integration and cointegration', *Journal of Econometrics*, Volume 44, pages 215 to 238.

Kruskal W and Wallis W (1952), 'Use of ranks in one-criterion variance analysis'. *Journal of the American Statistical Association*, Volume 47, Issue 260, pages 583 to 621.

Lunsford K (2017), 'Lingering residual seasonality in GDP growth', *Economic Commentary*, Volume 2017, Issue 06.

McElroy T and Roy A (2022), 'A review of seasonal adjustment diagnostics', *International Statistical Review*, Volume 90, pages 259 to 284.

Quartier-la-Tente A, Michalek A, Palate J, and Baeyens R (2024), 'RJDemetra: Interface to 'JDemetra+' Seasonal Adjustment Software', R package version 0.2.8, <https://CRAN.R-project.org/package=RJDemetra>.

9 . Related links

[Assessing residual seasonality in published outputs](#)

Article | Released 30 September 2025

How the ONS uses detailed statistical methods and analytical tools to assess the impact of residual seasonality on published time series.

[GDP monthly estimate, UK: February 2026](#)

Bulletin | Released 16 April 2026

Gross domestic product (GDP) measures the value of goods and services produced in the UK. It estimates the size of the economy and its growth.