Chapter Two

Macro Data Calculation and Handling

Tingguo Zheng (zhengtg@gmail.com)
Department of Statistics and Data Science, School of Economics, Xiamen University

(For Undergraduate Program, 2025-2026, Spring Semester)

- **Outline:**
  - Stock vs. Flow variables
  - Growth calculation
  - Data processing
  - Outlier detection and handling
  - Seasonal adjustment

1. Stock vs. Flow Variables

Definition 1 (Stock Variables). Measured at a specific point in time, they represent the "state" of an economic system at that instant.

- Stock variables are independent of the length of the time interval and reflect cumulative values.

- Analogy (类比): A snapshot of a reservoir's water level at a given moment.

- Examples: National wealth (total assets on a specific date), capital stock (value of machinery, equipment, and infrastructure), money supply (M2 at the end of a month), government debt balance, foreign exchange reserves, and unemployment count.

Definition 2 (Flow Variables). Measured over a specific time period, they represent the “flow” or “activity” of economic processes during that interval.

- Flow variables depend on the length of the time period and reflect changes over time.

- Analogy: The volume of water flowing into a reservoir over an hour.

- Examples: Gross Domestic Product (GDP) in a year, consumption expenditure in a quarter, investment volume, government purchases, net exports, tax revenue, and new loan amounts.

Key Relationships and Implications

- Interaction: Flows change stocks, and stocks influence future flows. For instance, annual investment (a flow) increases the capital stock (a stock), while a larger capital stock (a stock) boosts future production capacity and output (a flow).

- Critical Distinction: Confusing stocks and flows leads to misinterpretations of economic dynamics. For example, confusing "government debt (stock)" with "government deficit (flow)" can distort assessments of fiscal sustainability.

- Aggregation Differences: When aggregating data across time (e.g., converting monthly to quarterly data), stock variables use end-of-period values, while flow variables use summation.

2. Growth Rate Calculation

Growth rates are core indicators for measuring short-term fluctuations and long-term trends of economic variables.

Three common types of growth rates are widely used:

- month-on-month (MoM) growth rate
- year-on-year (YoY) growth rate
- cumulative year-on-year growth rate

Let Y_t denote the observed value of a variable at time t, where $Y_t = Y_{ks+i}$ represents the value in the i-th month (or quarter) of the k-th year. Let s be the number of periods in a year: $s = 12$ for monthly data and $s = 4$ for quarterly data.

© 2026 Tingguo Zheng

2.1 Month-on-Month (MoM) Growth Rate

Definition 3 (**Month-on-Month (MoM) Growth Rate).** It measures the rate of change between the current period and the immediately preceding period, reflecting short-term fluctuations.

Its formula is given by

$$R_t = \frac{Y_t}{Y_{t-1}} - 1$$

Example 1. If the retail sales in February 2024 are 120 million yuan and 100 million yuan in January 2024, the MoM growth rate in February is:

$$R_t = \left(\frac{120}{100} - 1 \right) \times 100\% = 20\%$$

Properties:

- High sensitivity to short-term changes: Ideal for tracking real-time trends (e.g., monthly economic indicators, quarterly financial reports).
- Susceptibility to seasonal interference（易受季节性干扰）: For example, summer air conditioning sales are naturally higher than winter, leading to volatile MoM growth rates.
- Dependence on the base period: A small base value in the previous period can cause sharp fluctuations in the growth rate even with minor changes.

2.2 Year-on-Year (YoY) Growth Rate

Definition 4 (**Year-on-Year (YoY) Growth Rate).** It measures the rate of change between the current period and the same period of the previous year, eliminating seasonal effects.

Its formula is given by

$$R_t(s) = \frac{Y_t}{Y_{t-s}} - 1 = \frac{Y_{ks+i}}{Y_{(k-1)s+i}} - 1$$

Example 2. If the retail sales in February 2024 are 120 million yuan and 100 million yuan in February 2023, the YoY growth rate in February 2024 is:

$$R_t(12) = \left(\frac{120}{100} - 1\right) \times 100\% = 20\%$$

Relationship with MoM Growth Rate: The YoY growth rate is equivalent to the cumulative effect of s consecutive MoM growth rates:

$$R_t(s) = (1 + R_t)(1 + R_{t-1}) \times (1 + R_{t-s+1}) - 1$$

Properties:

- Elimination of seasonal effects: Compares data from the same season, avoiding distortions caused by seasonal factors.
- Reflection of long-term trends: Reduces the impact of short-term fluctuations, suitable for evaluating long-term growth potential.
- Lag in capturing cyclical changes: May not immediately reflect economic turning points (e.g., YoY growth may remain negative in the early stages of economic recovery).

© 2026 Tingguo Zheng

2.3 Cumulative Year-on-Year Growth Rate

Definition 5 (**Cumulative Year-on-Year Growth Rate).** It measures the rate of change between the cumulative data from the start of the year to the current period and the cumulative data of the same period in the previous year, smoothing short-term volatility.

It formula is given by

$$\overrightarrow{R}_t(s) = \frac{\sum_{j=1}^{\ell} Y_{ks+j}}{\sum_{j=1}^{\ell} Y_{(k-1)s+j}} - 1$$

where ℓ is the number of periods from the start of the year to the current period (e.g., $\ell = 3$ for the first quarter).

Example 3. If the cumulative retail sales from January to March 2024 are 350 million yuan and 300 million yuan in the same period of 2023, the cumulative YoY growth rate for Q1 2024 is:

$$\overrightarrow{R}_t(s) = \left(\frac{350}{300} - 1\right) \times 100\% = 16.67\%$$

Properties:

- Smoothing of short-term fluctuations: Reduces the impact of outlier values in individual months/quarters, better reflecting overall trends.
- Tracking of annual progress: Suitable for monitoring the completion of annual targets (e.g., cumulative fiscal revenue vs. annual budget).
- Stronger lag: Less responsive to recent policy or market changes compared to YoY growth rates.

2.4 Logarithmic Growth Rates

Logarithmic transformation simplifies growth rate calculations and has favorable statistical properties (e.g., additivity), making it widely used in time series analysis and financial econometrics.

(1) Logarithmic Month-on-Month Growth Rate: It is defined as

$$r_t = \ln(1 + R_t) = \ln(Y_t/Y_{t-1}) = \ln Y_t - \ln Y_{t-1} = y_t - y_{t-1}$$

where $y_t = \ln Y_t$.

Interpretation: Known as the "continuous compound growth rate" in finance. For example, if Y_t is an asset price, r_t represents the continuous compound return from $t - 1$ to t:

$$Y_t = Y_{t-1} e^{r_t}$$

(2) Logarithmic Year-on-Year Growth Rate: Sum of s consecutive logarithmic MoM growth rates (additivity property):

$$r_t(s) = \ln(1 + R_t(s)) = \ln(Y_t/Y_{t-s}) = y_t - y_{t-s} = \sum_{i=0}^{s-1} r_{t-i}$$

Advantage: Converts multiplicative growth into additive growth, simplifying the analysis of long-term cumulative effects.

(3) Logarithmic Cumulative Year-on-Year Growth Rate

Using the geometric mean to approximate the arithmetic mean of cumulative values,

$$\ell^{-1}\sum_{j=1}^{\ell} Y_{ks+j} \approx \exp\left\{\ell^{-1}\sum_{j=1}^{\ell}\ln Y_{ks+j}\right\}$$

the logarithmic cumulative YoY growth rate can be approximated as the average of logarithmic YoY growth rates over the same period:

$$\begin{aligned}
\overrightarrow{r}_t(s) = \ln(1+\overrightarrow{R}_t(s)) &= \ln\left(\frac{\sum_{j=1}^{\ell} Y_{ks+j}}{\sum_{j=1}^{\ell} Y_{(k-1)s+j}}\right) = \overrightarrow{r}_{ks+\ell} \\
&\approx \ln\left(\frac{\exp\{\ell^{-1}\sum_{j=1}^{\ell}\ln Y_{ks+j}\}}{\exp\{\ell^{-1}\sum_{j=1}^{\ell}\ln Y_{(k-1)s+j}\}}\right) \\
&= \ell^{-1}\left(\sum_{j=1}^{\ell}(\ln Y_{ks+j} - \ln Y_{(k-1)s+j})\right)
\end{aligned}$$

Hence, we have the following approximation

$$\overrightarrow{r}_t(s) \approx \frac{1}{\ell} \sum_{j=1}^{\ell} r_{ks+j}$$

where r_{ks+j} is the logarithmic YoY growth rate in the j-th period of the k-th year.

For small growth rates, the simple cumulative YoY growth rate can be approximated as:

$$\begin{aligned} \overrightarrow{R}_t(s) = \overrightarrow{R}_{ks+\ell} &\approx \exp\left\{ \ell^{-1} \sum_{j=1}^{\ell} \ln(1 + R_{ks+j}) \right\} - 1 \\ &\approx \ell^{-1} \sum_{j=1}^{\ell} R_{ks+j} \end{aligned}$$

© 2026 Tingguo Zheng

3. Data Processing

3.1 Data Transformation

Data transformation improves the statistical properties of the data (e.g., normality, variance stability) to enhance the reliability of model estimation and inference.

(1) Logarithmic transformation

- Form: $y_t = \log(x_t)$ or $y_t = \ln(x_t)$;
- Purpose:
  - Stabilizes variance: Reduces heteroscedasticity in data with exponential growth (e.g., GDP, asset prices).
  - Converts multiplicative relationships to additive ones: Simplifies the analysis of growth rate interactions.
  - Improves normality: Makes skewed data more symmetric (e.g., financial market volatility).
- Application: take a log of seasonal unadjusted GDP; take a log of realized volatility.

 © 2026 Tingguo Zheng

(2) Box-Cox transformation

- Form: Proposed by Box & Cox (1964), applicable to positive data:

$$y_t = h(x_t, \lambda) = \begin{cases} (x_t^{\lambda} - 1)/\lambda, & \text{if } \lambda \neq 0; \\ \log x_t, & \text{if } \lambda = 0. \end{cases} \tag{1}$$

  for $x_t > 0$, $t = 1, \ldots, n$.

- Purpose: Finds the optimal λ to make the data as close to normally distributed as possible. When $\lambda = 0$, it reduces to the logarithmic transformation.

- Advantage: Flexible adaptation to different data distributions, improving the robustness of statistical analysis.

(3) Variance stabilizing transformation

- Form: Proposed by Bartlett (1947), eliminates the correlation between the mean and variance of the data. For a random variable X with $E(X) = \mu$ and $Var(X) = v(\mu)$, the transformation function is:

$$f(\mu) = \int [v(\mu)]^{-1/2} d\mu$$

Table 1 The variance stabilizing transformation of some exponential family distributions

	Normal	Lognormal	Gamma	Beta	Poisson	Binomial
Parameter	$(\mu, \sigma^2)'$	$(\mu, \sigma^2)'$	$(\alpha, \alpha/\mu)'$	$(\tau\mu, \tau(1-\mu))'$	μ	μ/n
$h(y)$	y	$\ln y$	$\ln y$	$\arcsin\sqrt{y}$	$\sqrt{y}$	$\arcsin\sqrt{y/n}$
$\mathrm{Var}[h(y)]$	σ^2	σ^2	$\psi_1(\alpha)$	$\approx 1/4(1+\tau)$	$\approx 1/4$	$\approx 1/4n$

Note: $\psi()$ and $\psi_1()$ denote the digamma and trigamma functions, respectively.

3.2 Detrending (Trend/Cycle decomposition)

Assume that a variable y_t can be decomposed as:

$$y_t = T_t + c_t,$$

where T_t and c_t denote trend component and cyclical component.

This is often used for decomposizing the GDP and unemployment rate.

Five univariate methods are commonly used to extract cyclical components in time series.

- Linear trend filter or quadratic trend filter.
- BN (Beveridge-Nelson, 1981) decomposition.
- HP (Hodrick-Prescott, 1981) filter.
- Frequency filtering techniques, for example BK (Baxter-King, 1999) filter, CF (Christiano-Fitzgerald, 2003) filter.
- Unobservable components models, for example Harvey (1985), Clark (1987), Harvey and Jaeger (1993).

See also in 郑挺国和王霞(《经济研究》2010年第10期)

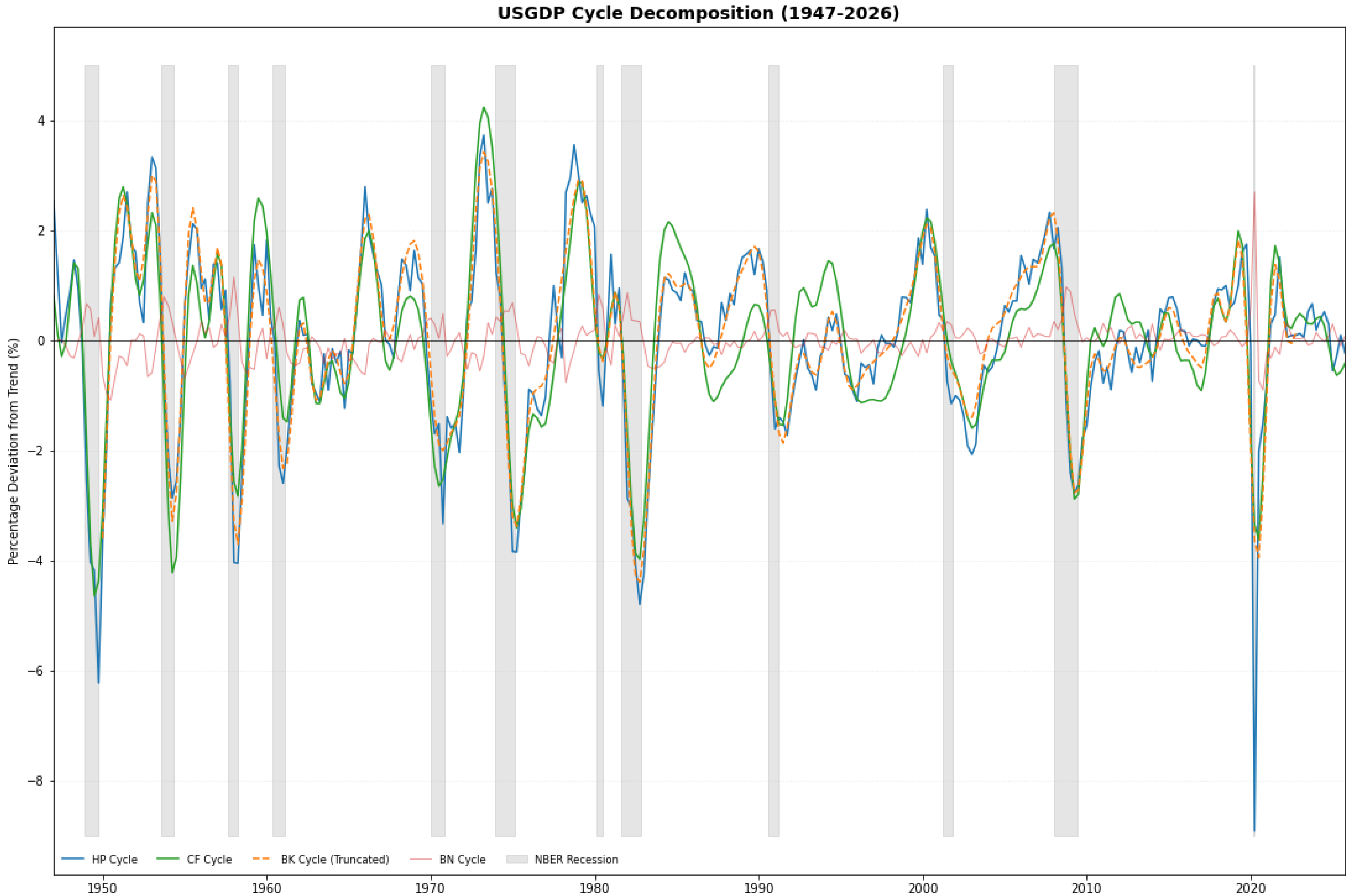

Figure 1 Trend/Cycle decomposition for the U.S. GDP (in log $\times 100$) using quadratic filter, HP filter, BK filter, and CF filter.

- Edward Christian Prescott (1940/12/26 – 2022/12/6) was an American economist. He received the Nobel Memorial Prize in Economics in 2004, sharing the award with Finn E. Kydland, “for their contributions to dynamic macroeconomics: the time consistency of economic policy and the driving forces behind business cycles”. This research was primarily conducted while both Kydland and Prescott were affiliated with the Graduate School of Industrial Administration (now Tepper School of Business) at Carnegie Mellon University. According to the IDEAS/RePEc rankings, he was the 19th most widely cited economist in the world in 2013.

The Hodrick – Prescott Filter

One popular example is the ad hoc trend extraction filter proposed by Hodrick and Prescott (1981→1997).

HP define the estimator of the trend as the minimiser of the penalised least square criterion:

$$PLS = \sum_{t=1}^{T}(y_t - \mu_t)^2 + \lambda \sum_{t=3}^{T}(\Delta^2 \mu_t)^2,$$

where the first summand measures fidelity (精确度) and the second roughness (粗糙度); λ is the smoothness parameter governing the trade-off between them ($\lambda = 1600$ for quarterly series; $= 14400$ for monthly series).

Unobservable components models

A classic trend/cycle decomposition model is given as follows:

$$y_t = \tau_t + c_t \tag{2}$$

$$\tau_t = \tau_{t-1} + g_{t-1} + u_t \tag{3}$$

$$g_t = gc + \lambda g_{t-1} + w_t, \tag{4}$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + v_t \tag{5}$$

where y_t is log GDP; τ_t is its *trend component* follows a *random walk* with a *stochastic drift* or growth rate which is an autoregressive process; c_t is the *cycle component*.

When $\lambda = \sigma_w = 0$, this becomes the Watson (1986) model.

When $gc = 0$ and $\lambda = 1$, this becomes the Clark (1987) model.

3.3 Lag operator and Differencing

- Lag operator (L) or backshift operator (B). For example, given some time series $\{y_1, y_2, \ldots\}$, then

$$Ly_t = y_{t-1} \quad \text{for all } t > 1.$$

- Lag operator can be raised to arbitrary integer powers so that

$$L^{-1}y_t = y_{t+1}, \quad \text{and} \quad L^k y_t = y_{t-k}.$$

- First difference operator $\Delta = (1 - L)$ is a special case of lag polynomial. Then $\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$.

 © 2026 Tingguo Zheng

For difference operator, we have the following results:

- The first difference operator: $\Delta y_t = (1-L)y_t$;
- The second difference operator: $\Delta^2 y_t = (1-L)^2 y_t = (1-2L+L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$;
- The above approach generalises to the i-th difference operator $\Delta^i y_t = (1-L)^i y_t$;
- Log-First-Difference: $\Delta \ln y_t = \ln y_t - \ln y_{t-1}$.

 © 2026 Tingguo Zheng

Now if consider the change between y_t and y_{t-k}, we can use notation of difference operator by $\Delta_k y_t = (1 - L^k)y_t$. Similarly, we have the following results:

- The seasonal difference operator: $\Delta_{12} y_t = (1 - L^{12})y_t$ if monthly data; $\Delta_4 y_t = (1 - L^4)y_t$ if quarterly data;

- Year over year change: $\Delta_{12} \ln y_t = (1 - L^{12}) \ln y_t = \ln(y_t / y_{t-12})$ if y_t is an aggregated monthly time series.

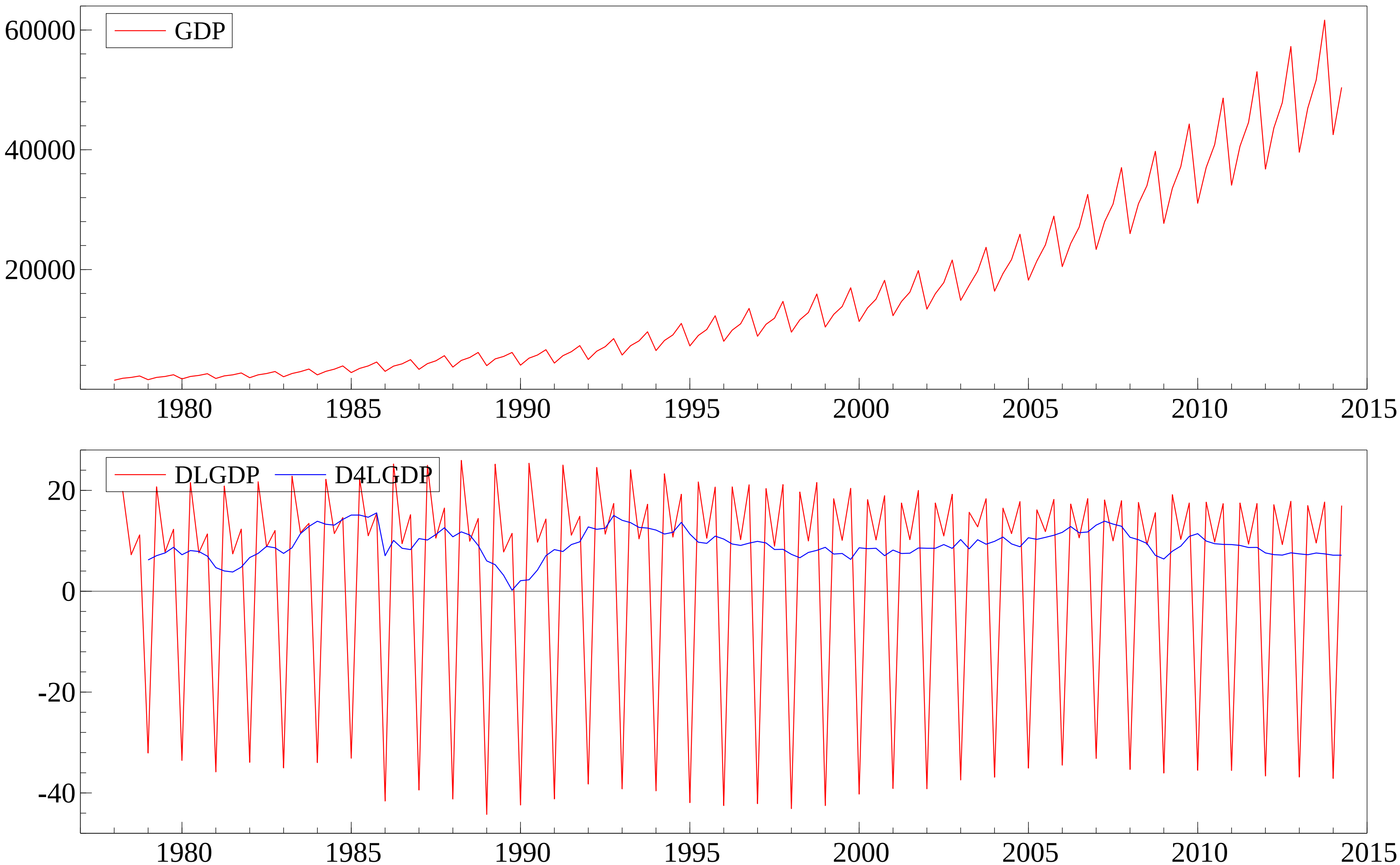

Figure 2 Comparison of $\Delta \log(GDP_t)$ and $\Delta_4 \log(GDP_t)$ for China's quarterly GDP.

4. Outlier Detection and Handling

Outliers—observations that deviate drastically from the general pattern of a dataset—are common in macroeconomic data.

They may arise from measurement errors (e.g., statistical survey biases), extreme events (e.g., financial crises, natural disasters, pandemics), policy shocks (e.g., sudden tax reforms, trade wars), or rare economic phenomena (e.g., unprecedented commodity price spikes).

Unaddressed outliers can distort parameter estimates, bias statistical inference, and mislead economic interpretations (e.g., overestimating volatility, falsifying causal relationships).

4.1 Outlier Detection Methods

(1) Standard Deviation Method (Z-Score Method)

- Core Idea: For normally distributed data, observations outside a certain number of standard deviations from the mean are considered outliers.
- Calculate the mean: $\bar{y} = \frac{1}{T} \sum_(t = 1)^T y_t$
- Calculate the sample standard deviation:

$$s = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \bar{y})^2} \quad \text{(divide by } T-1 \text{ for unbiased estimation)}$$

- Compute the Z-score for each observation: $Z_t = \dfrac{y_t - \bar{y}}{s}$
- Define outliers: Observations where $|Z_t| > k$, where k is a threshold.

Common Thresholds:

- $k = 2$: Covers 95% of normal data (identifies mild outliers).
- $k = 3$: Covers 99.7% of normal data (standard for strict outlier detection in macroeconomics).
- $k = 3.29$: Corresponds to the 99.9th percentile (used for rare events in financial or high-stakes policy analysis).

Advantages:

- Simple to calculate and interpret—intuitive.
- Effective for normally distributed data (e.g., seasonally adjusted GDP growth, interest rates).

Limitations:

- Sensitive to the mean and standard deviation, which are themselves distorted by outliers.
- Inappropriate for non-normal data (common in macroeconomics, e.g., unemployment rates, fiscal deficits).
- Fails for skewed distributions (e.g., asset returns with fat tails).

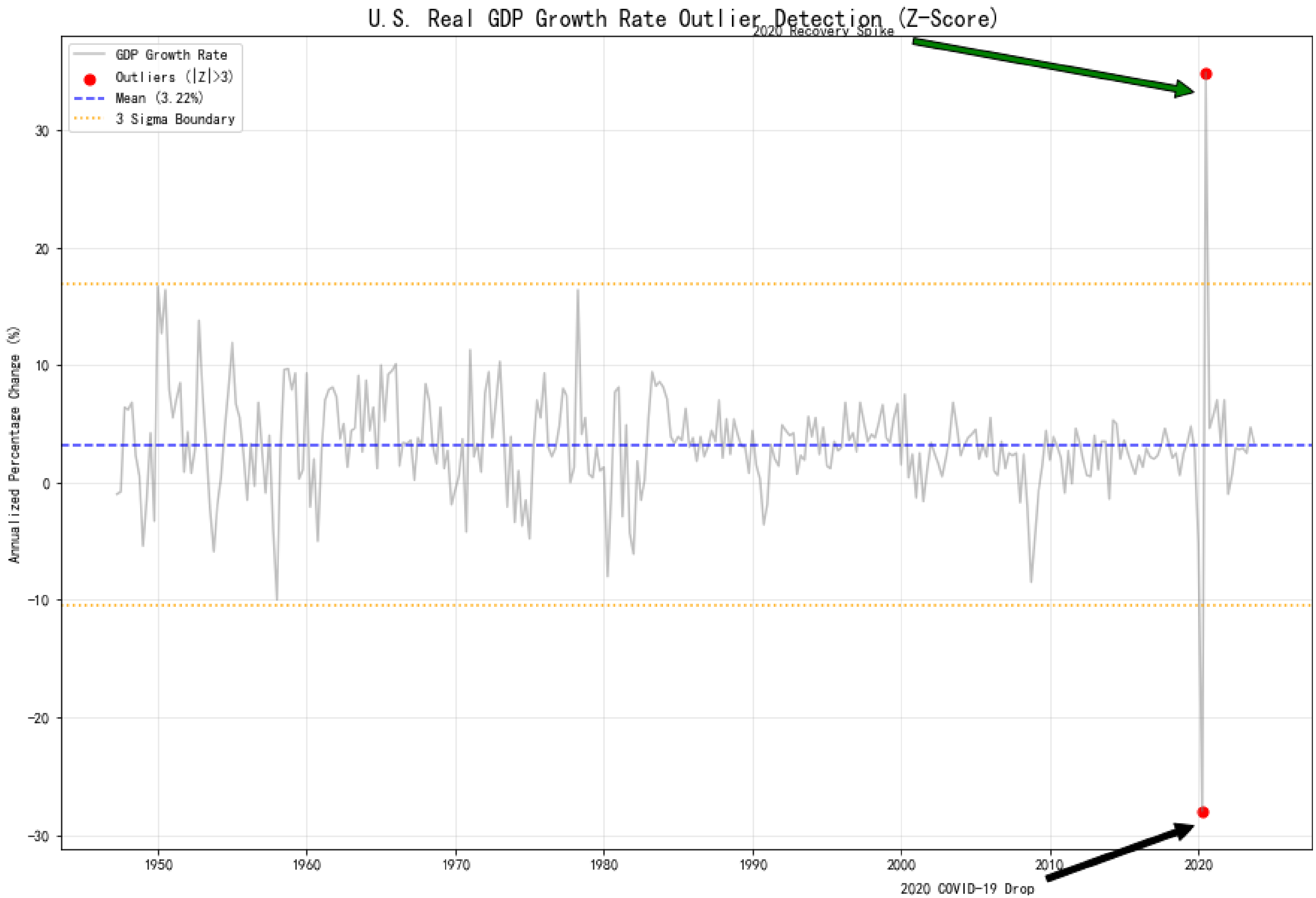

Figure 3 Outlier detection of US real GDP growth with Z-Score method.

(2) Modified Z-Score Method (Median Absolute Deviation, MAD)

- Core Idea: Addresses the standard deviation method’ s sensitivity to outliers by using the median (robust to extreme values) and median absolute deviation (MAD) instead of the mean and standard deviation.

- Calculate the median: Med $=$ middle value of sorted y_t

- Calculate MAD: MAD $=$ Median$(|y_t -$ Med$|)$

- Compute the modified Z-score: $Z_t^* = 0.6745 \times \dfrac{y_t - \text{Med}}{\text{MAD}}$
(the constant 0.6745 aligns MAD with the standard deviation for normal data).

- Define outliers: $|Z_t^*| > k$ (typically $k = 3.5$ for macroeconomic data).

© 2026 Tingguo Zheng

Advantages:

- Robust to outliers (median and MAD are not affected by extreme values).
- Still applicable to approximately normal data but more reliable than the standard Z-score.

Limitations:

- Still assumes approximate normality; less effective for highly skewed data.
- Less intuitive for users familiar with the standard Z-score.

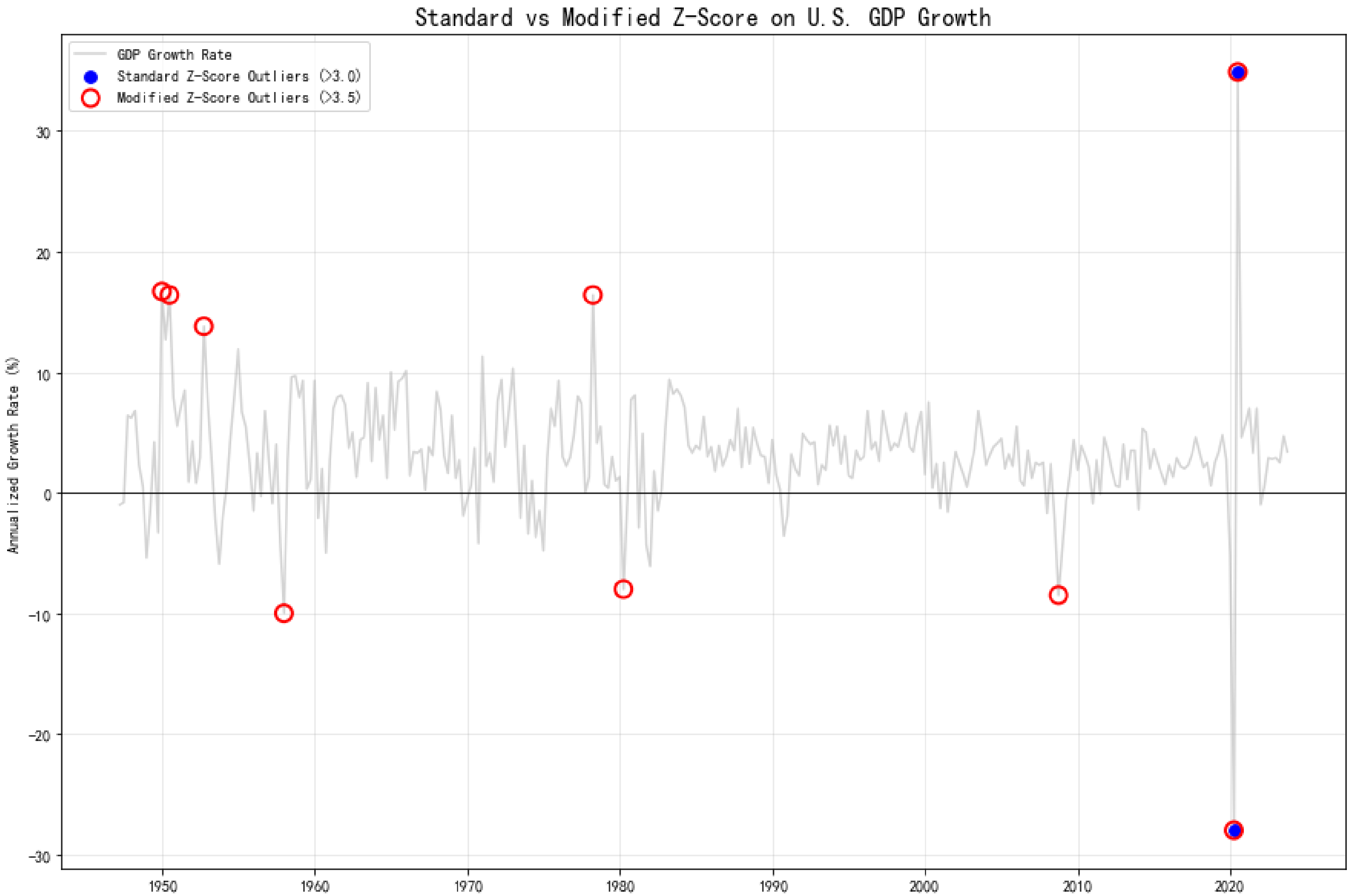

Figure 4 Outlier detection of US real GDP growth with Modified Z-Score method.

(3) Interquartile Range (IQR) Method

- Core Idea: Uses quartiles to define the ”normal range” of the data, as quartiles are robust to outliers.

- Sort the data in ascending order: $y_((1)) \leq y_((2)) \leq \cdots \leq y_((T))$

- Calculate quartiles:
  - First quartile (Q_1): 25th percentile (median of the lower half of the data). For T even: $Q_1 = y_{(T/4)}$; for T odd: $Q_1 = y_{(((T+1))/4)}$.
  - Third quartile (Q_3): 75th percentile (median of the upper half of the data).

- Compute IQR: $IQR = Q_3 - Q_1$ (measures the spread of the middle 50% of the data).

- Define the outlier range (fences):
  - Lower fence: $Q_1 - k \times IQR$
  - Upper fence: $Q_3 + k \times IQR$
- Identify outliers: Observations outside the fences.

Common Thresholds (k)**:**

- $k = 1.5$: Identifies "mild outliers" (standard for most macroeconomic applications).
- $k = 3$: Identifies "extreme outliers" (used for highly volatile data, e.g., financial market returns).

Advantages:

- Completely distribution-agnostic—works for skewed, fat-tailed, or non-normal data.
- Robust to extreme values (quartiles are not influenced by outliers).
- Intuitive interpretation (focuses on the middle 50% of the data).

Limitations:

- Less sensitive to small deviations from the norm (may miss mild outliers in large datasets).
- Not ideal for high-dimensional data (works best for univariate time series).

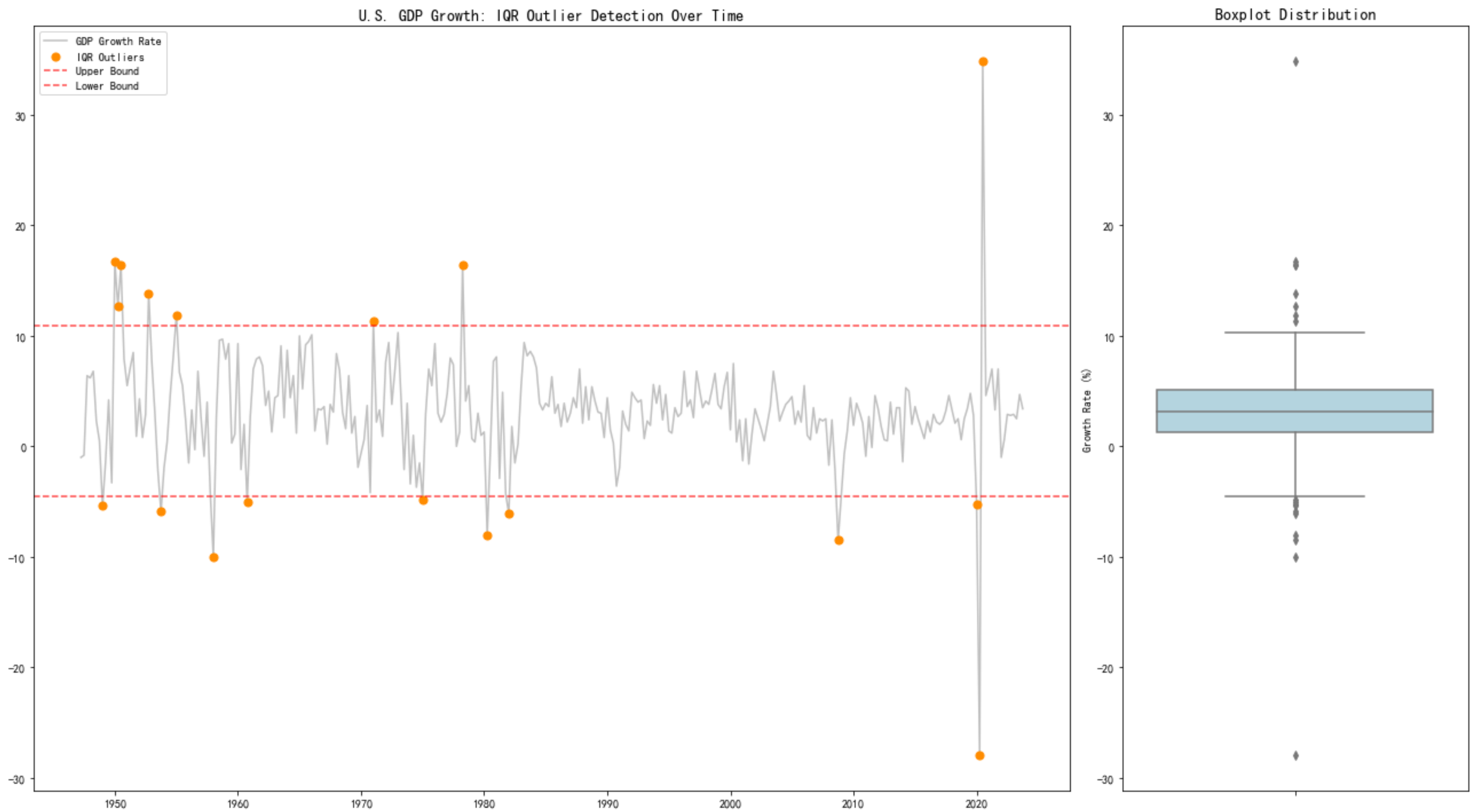

Figure 5 Outlier detection of US real GDP growth with Interquartile Range (IQR) method.

4.2 Outlier Correction Methods

(1) Winsorization (Capping/Flooring)

- Core Idea: Replace extreme outliers with the nearest non-outlier value (caps the data at a specified threshold).

- Identify the outlier thresholds (e.g., $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$ for IQR-based winsorization).

- Replace values below the lower threshold with the lower threshold.

- Replace values above the upper threshold with the upper threshold.

(2) Interpolation

- Core Idea: Estimate the outlier's value using adjacent observations (assumes the data follows a smooth trend between non-outlier points).

- Linear Interpolation: Estimates the outlier as a linear combination of the previous and next observations:

$$\hat{y}_t = y_{t-1} + \frac{y_{t+1} - y_{t-1}}{(t+1-(t-1)} \times (t-(t-1)) = \frac{y_{t-1} + y_{t+1}}{2}$$

Example: If April GDP growth (outlier = 5.8%) is between March (2.1%) and May (2.5%), linear interpolation gives (2.1 + 2.5)/2 = 2.3%.

- Moving Average Interpolation: Uses the average of a window of adjacent observations (e.g., 3-month moving average):

$$\hat{y}_t = \frac{y_{t-2} + y_{t-1} + y_{t+1} + y_{t+2}}{4}$$

  Example: For April's outlier, use Feb, Mar, May, Jun growth rates.

- Cubic Spline Interpolation: Fits a smooth cubic curve through non-outlier points and estimates the outlier's value from the curve—more flexible for non-linear trends.

5. Seasonal Adjustment

Seasonal adjustment or deseasonalization is a statistical method for removing the seasonal component of a time series.

It is usually done when wanting to analyse the trend, and cyclical deviations from trend, of a time series independently of the seasonal components.

Many economic phenomena have seasonal cycles, such as agricultural production, (crop yields fluctuate with the seasons) and consumer consumption (increased personal spending leading up to Christmas).

It is necessary to adjust for this component in order to understand underlying trends in the economy, so official statistics are often adjusted to remove seasonal components.

5.1 Time series components

The investigation of many economic time series becomes problematic due to seasonal fluctuations.

Time series are made up of four components:

- S_t: The seasonal component;
- T_t: The trend component;
- C_t: The cyclical component;
- I_t: The error, or irregular component.

The difference between seasonal and cyclic patterns:

- Seasonal patterns have a fixed and known length, while cyclic patterns have variable and unknown length.
- Cyclic pattern exists when data exhibit rises and falls that are not of fixed period (duration usually of at least 2 years).
- The average length of a cycle is usually longer than that of seasonality.
- The magnitude of cyclic variation is usually more variable than that of seasonal variation.

The relation between decomposition of time series components

- Additive decomposition: $Y_t = S_t + T_t + C_t + E_t$, where Y_t is the data at time t.

- Multiplicative decomposition: $Y_t = S_t \cdot T_t \cdot C_t \cdot E_t$.

- Logs turn multiplicative relationship into an additive relationship: $Y_t = S_t \cdot T_t \cdot C_t \cdot E_t \rightarrow \log Y_t = \log S_t + \log T_t + \log C_t + \log E_t$:

- An additive model is appropriate if the magnitude of seasonal fluctuations does not vary with level.

- If seasonal fluctuations are proportional to the level of the series, then a multiplicative model is appropriate. Multiplicative decomposition is more prevalent with economic series.

5.2 Adjustment methods

Different statistical research groups have developed different methods of seasonal adjustment, for example

- *X-11-ARIMA, X-12-ARIMA* and *X-13-ARIMA-SEATS* in many softwares such as SAS, Eviews, OxMetrics, R and Python developed by the United States Census Bureau (美国人口调查局);
- *TRAMO/SEATS* in Eviews developed by the Bank of Spain;
- *STAMP* in OxMetrics developed by a group led by S. J. Koopman.
- *STL* in R and Python developed by Cleveland, et al. (1990), “STL: A Seasonal-Trend Decomposition Procedure Based on Loess”, Journal of Official Statistics, 6(1), 3-73.

(1) **X-12-ARIMA** (Gold Standard for Official Statistics)

- Data Preparation: Ensure the series is in a regular frequency (monthly/quarterly) with no major structural breaks (if breaks exist, split the series).
- Specify regARIMA Model: Choose ARIMA orders (p, d, q) to fit the trend-cyclical component (e.g., ARIMA(1,1,0) for GDP).
- Calendar Adjustment: Include regressors for trading days (e.g., number of Sundays) and moving holidays.
- Outlier Correction: Identify additive/multiplicative outliers and replace them with ARIMA-predicted values.
- Extend Series: Use ARIMA to forecast 6–12 periods ahead and backcast 1–2 periods to handle end-point bias.
- Apply X-11 Filter: Decompose the extended series into T_t, C_t, S_t, I_t.
- Generate Seasonally Adjusted Data: Remove S_t from the original series.

(2) **TRAMO/SEATS** (European Standard)

- TRAMO: Preprocesses the data (regression with ARIMA noise) to handle outliers, missing data, and calendar effects.
  - Fits a regression model with ARIMA errors to the series, including regressors for calendar effects and outliers.
  - Identifies four types of outliers: additive (AO), innovational (IO), level shift (LS), and temporary change (TC).

- SEATS: Extracts seasonal components using ARIMA-based signal extraction (decomposes the series into trend, seasonal, and irregular components via spectral analysis).
  - Uses the ARIMA model's spectral density to separate the series into “signal” (trend + cyclical) and “noise” (seasonal + irregular).
  - Estimates the seasonal component by isolating frequencies corresponding to 12-month cycles (for monthly data).

(3) STL (Seasonal and Trend Decomposition using Loess)

- STL is a non-parametric method that uses local weighted regression (Loess) to decompose the series. It is flexible, handles non-linear trends and time-varying seasonality, and is ideal for high-frequency data (daily/weekly) and data with outliers.

- Implementation Steps
  - Set Parameters: Choose s (seasonal window) and t (trend window) based on data frequency (e.g., monthly data: $s = 13$, $t = 15$).
  - Initial Trend Estimation: Apply Loess smoothing to the original series to get an initial trend estimate.
  - Extract Seasonal-Irregular Component: Subtract the initial trend from the original series.

- Estimate Seasonal Component: Average the seasonal-irregular component for each period (e.g., average all January values) to get the seasonal factor.
- Refine Trend: Subtract the seasonal component from the original series and apply Loess smoothing again to get the final trend.
- Iterate: Repeat steps 3 – 5 until the decomposition converges (typically 2 – 3 iterations).

(4) BSM model in STAMP*

The basic structural model (BSM) is formulated in terms of trend, seasonal, and irregular components. It can be represented by

$$y_t = \tau_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2), \tag{6}$$

where τ_t is a stochastic trend, γ_t is a stochastic seasonal component, and ε_t is an irregular component; see Harvey (1989).

The trend is specified in the following way:

$$\tau_t = \tau_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim iidN(0, \sigma_\eta^2) \tag{7}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim iidN(0, \sigma_\zeta^2), \tag{8}$$

where β_t is the slope, η_t and ζ_t are assumed to be mutually independent.

The seasonal component γ_t is often specified in two ways such as dummy-variable seasonality and trigonometric seasonality, see Harvey et al. (1998) and Commandeur and Koopman(2007).

A time-varying dummy seasonal specification is given by

$$\sum_{i=0}^{s-1} \gamma_{t-i} = e_t, \tag{9}$$

where s is the seasonal length, e_t is an i.i.d. normal variable with mean zero and variance σ_e^2, i.e., $e_t \sim iidN(0, \sigma_e^2)$.

- In the limiting case $\sigma_e^2 = 0$, the seasonal effects are fixed over time.

- Commonly, $s = 4$ for quarterly data and $s = 12$ for monthly data.

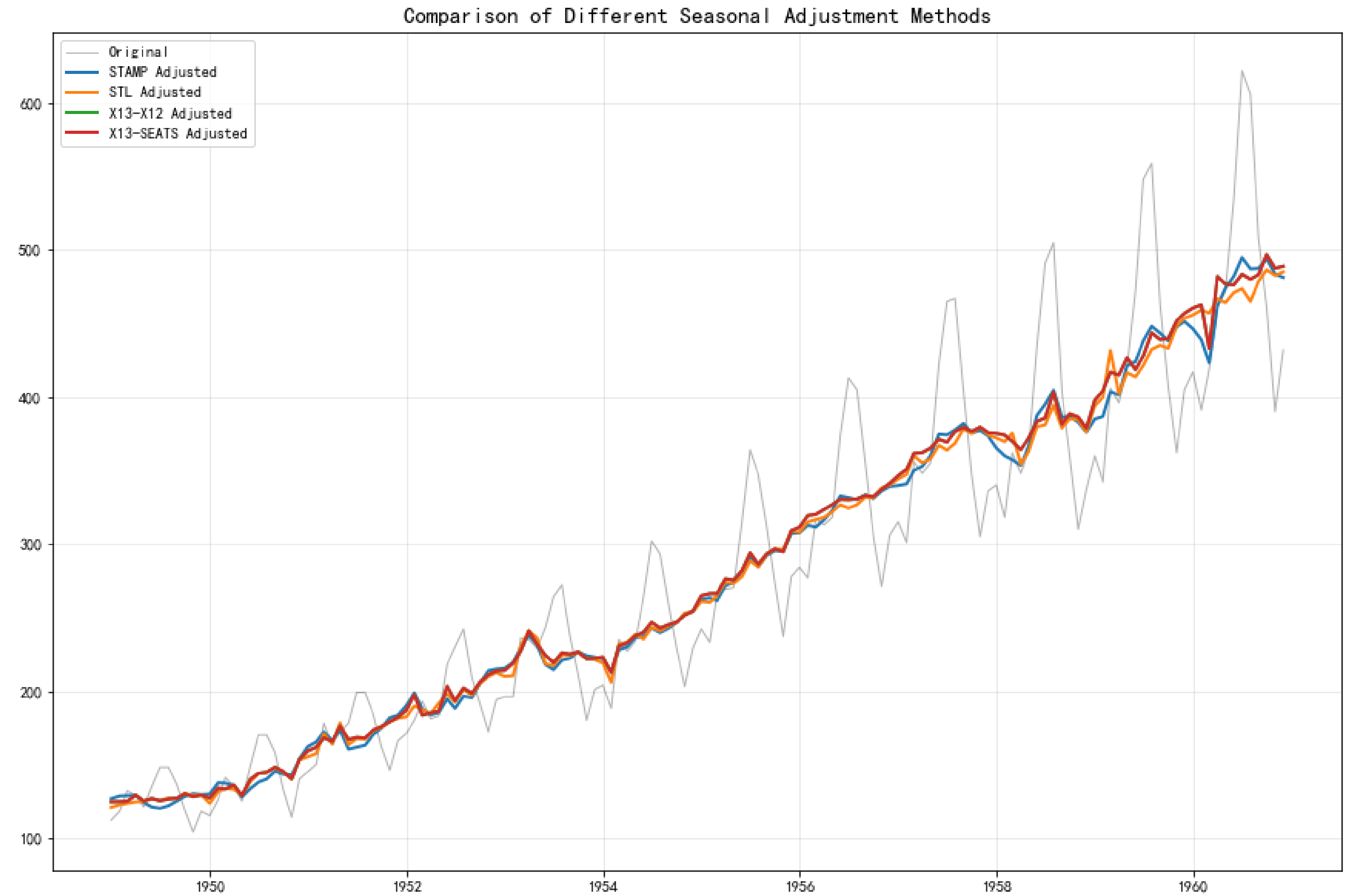

Figure 6 Seasonal adjustment of international airline-passengers with four methods ranged from January 1949 to December 1960, see more details in Box, Jenkins and Reinsel (1994).

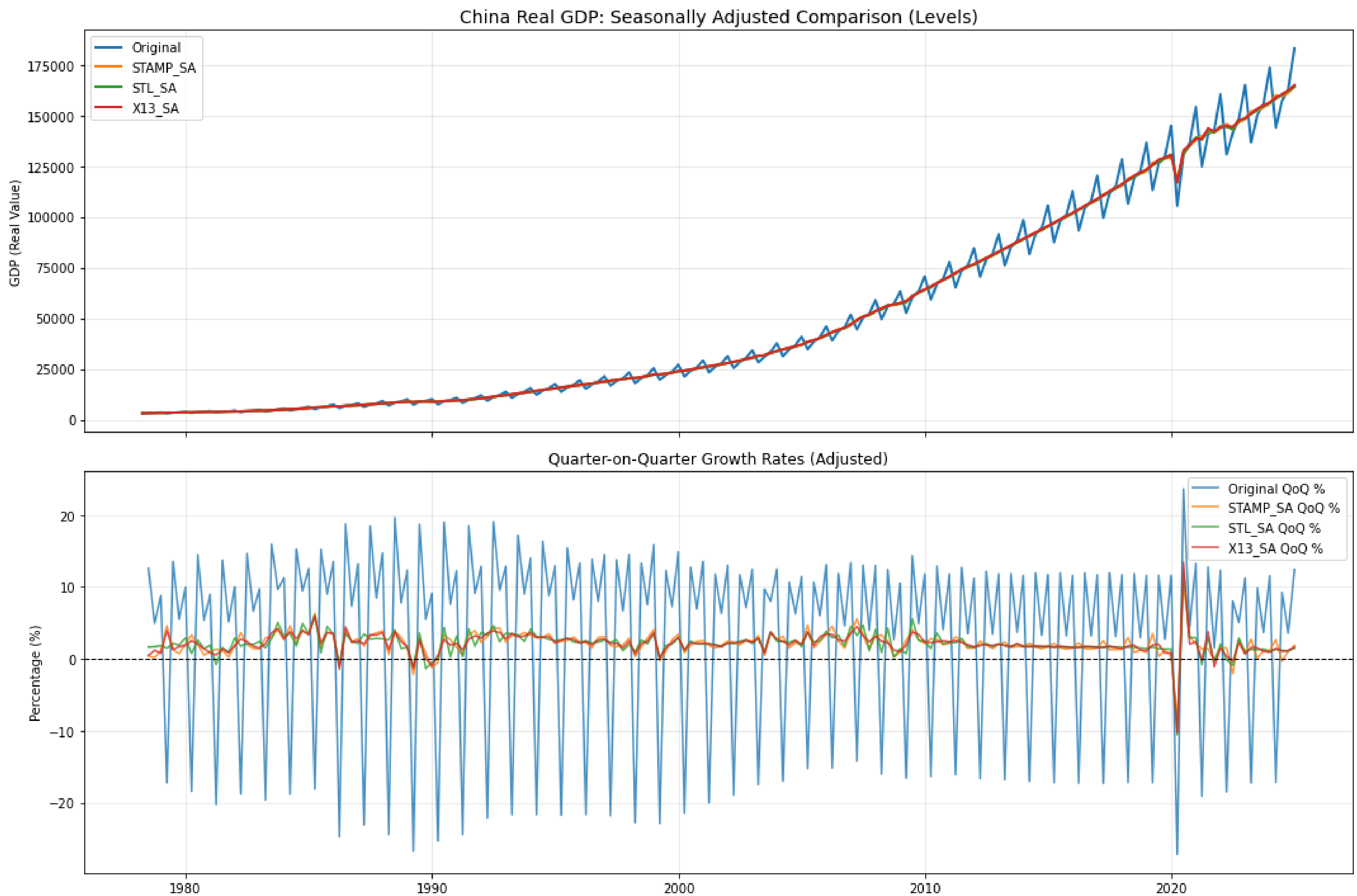

Figure 7 Seasonal adjustment of China's GDP with four methods sampled from 1978Q1 to 2024Q4.