Time Series Analysis of Consumer Price Index Products and Weights November 2024

MoonJung Cho*

Key Words: Discrete Fourier Transform, Periodicity, Trend, Wavelets.

1. Introduction

The Bureau of Labor Statistics (BLS) Consumer Price Index for all urban consumers (CPI-U) is a measure of the average change over time in the prices paid by urban consumers for a representative basket of consumer goods and services. It measures inflation as experienced by consumers in their day-to-day living expenses. The CPI-U is used to adjust income eligibility levels for government assistance, federal tax brackets, federally mandated cost-of-living increases, private sector wage and salary increases, and consumer and commercial rent escalations. Consequently, the BLS CPI products directly affects hundreds of millions of Americans (Bureau of Labor Statistics, 2023, Handbook of Methods, Consumer Price Index).

The estimation of the CPI is divided into two stages: lower-level and upper-level processing. Lower-level processing calculates basic item-area indexes. The Consumer Expenditure (CE) household survey serves as the source for sampled outlets, and sampled quotes are weighted based on the CE survey. The same lower-level basic indexes are used for all CPI products. Upper-level aggregate index formulas apply different weights to these lower-level indexes to create a final index product. CPI products differ by their corresponding weights and index formulas. The Chained CPI (C-CPI-U) uses monthly weights and the Tornqvist index formula, which is a geometric average where the weights are a two-month moving average for the corresponding index month. The C-CPI-U is a superlative index where both prices and weights are from the corresponding current and previous periods. However, monthly weights are available for index estimation approximately one year after the publication of the CPI. Due to the lag in the availability of monthly weights, the C-CPI-U is calculated retrospectively and is then revised quarterly, with up to four revisions (Bureau of Labor Statistics, 2023, Handbook of Methods, Consumer Price Index).

Unofficial CPI research series use different reference period weights and weight revisions occur at different frequencies. For example, the Quarterly CPI (CPI-q), Annual CPI (CPI-a), and Biennial CPI (CPI-b), use quarterly, annual, and biennial weights, respectively. The different weights correspond to varying lengths of time between the period when household data is collected and when it is used in index calculation. This time period is referred to as the reference period weight lag. The biennial weight lags the index by an average of three years, the annual weight by two

*Office of Survey Methods Research, U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

years, and the quarterly weight by one year (Klick and Park, 2022). In addition, CPI-q, CPI-a, and CPI-b apply the CPI Lowe formula, a modified Laspeyres formula, to calculate their cost weights.

CPI cost weights are the product of indexes and weights. The month-to-month price change is the ratio of cost weights from month t to $t-1$. The following is the CPI cost weight Lowe formula where quarterly, annual, and biennial weights are used:

$$CPI_{t,0}^{LO} = \sum_{k \in j} \left(IX_{tk} \times \hat{\mathbf{P}}_{\alpha\mathbf{k}} \hat{\mathbf{Q}}_{\beta\mathbf{k}} \right)$$

where

$$\begin{aligned}
k &: \text{ basic level element,} \\
j &: \text{ aggregate level,} \\
t &: \text{ period,} \\
IX_{tk} &: \text{ basic level index,} \\
\hat{\mathbf{P}}_{\alpha\mathbf{k}} \hat{\mathbf{Q}}_{\beta\mathbf{k}} &: \text{ fixed quantity weights,} \\
\alpha &: \text{ index average for the reference period,} \\
\beta &: \text{ associated weight (e.g., biennial, annual, or quarterly weight).}
\end{aligned}$$

For a detailed description of the CPI Lowe formula and the Chained CPI formula, see Klick and Park (2022) and references therein.

Our focus is to analyze the seasonal components of Chained CPI and CPI research series. Our analysis is limited to the urban population as CPI-U or C-CPI-U unless otherwise indicated. The initial investigation is conducted through frequency analysis using the Discrete Fourier Transform. We also examine CPI series with regard to trends, rate changes, jump discontinuities, and outliers.

2. Seasonality of Indexes and Weights of Major Groups

In the Consumer Price Index, urban areas in the United States are divided into 32 geographic regions, called index areas. The set of all goods and services purchased by consumers is divided into 211 categories, known as item strata: 209 commodity and service (CS) item strata, and 2 housing item strata. The number of basic items used for the calculation of aggregate indexes is larger, at 243, because entry-level items are used for the calculation of basic cells for health insurance retained earnings (item code SEME) rather than the higher item stratum level. This results in 7,776 (32 $\times$ 243) item-area combinations. The CPI calculates subaggregate indexes by averaging across subsets of item-area combinations. (Bureau of Labor Statistics, 2023, Handbook of Methods, Consumer Price Index). It is notable the CPI is additive so that the sum of component cost weights equals the final cost weights.

The data series consists of 108 monthly indexes from January 2014 to December 2022, aggregated into eight major groups: Apparel (A), Education&Communication (E), Food (F), Other

Goods (G), Housing (H), Medical (M), Recreation (R), and Transportation (T). We conducted analysis on the Chained CPI, CPI-q, CPI-a, and CPI-b, with their associated weights: monthly, quarterly, annual, and biennial, respectively. In this section, we present some results from the Chained CPI and its monthly weights for the major groups.

Table 1 presents the correlations of the Chained CPI between the major groups and shows that the Food, Other Goods, Housing, and Medical groups are highly positively correlated with each other. Education&Communication, however, is negatively correlated with all other major groups, except Apparel. Figure 1 shows the following for the Chained CPI of the major groups: the Food, Other Goods, Housing, and Medical groups move together in a similar manner; Recreation shows a moderate increase; Education&Communication remains flat with a mild decrease; Apparel exhibits seasonality; and Transportation displays high variability.

Figure 2 plots the associated monthly weights of the Chained CPI major groups. These weights sum to 1 and are used when computing the final weight for All Items.

2.1 Discrete Fourier Transform

The Fourier transform is a mathematical formula that transforms a signal sampled in time to the same signal sampled in temporal frequency. In signal processing, the Fourier transform can reveal the frequency components of the signal. If the series is a sum of several different frequency components, the Fourier transform will show all the frequencies. The Discrete Fourier Transform (DFT) takes a vector of length n and transforms it into a vector of length n. More precisely, it is a linear transformation from R^n to C^n that preserves the inner product. In applications, the input X is a vector of n values representing a signal sampled at times $t = 0, \dots, n-1$. The transform of X, denoted as $T(X)$, is a vector of n complex values. The j-th component of $T(X)$ is given by:

$$\sum_k \omega^{kj} X_k$$

where $\omega = \exp(2\pi i/n)$ is the n-th root of unity, and i is the imaginary unit defined by its property $i^2 = -1$. The j-th component of $T(X)$ represents the frequency j component of X. It is a complex number, whose absolute value relates to the strength of the component, and the square of the absolute value is referred to as the power. Power as a function of frequency is a common metric used in signal processing. Despite the presence of noise, the signal's frequencies can still be identified due to the spikes in power. Instead of using the power values in their original scale, we divide the power by the sum of the power to improve comparability. We then plot the scaled power against the period, where the period is the reciprocal of the frequency.

In Matlab, `T = fft(X)` computes the DFT of X using a fast Fourier transform (FFT) algorithm. T is the same size as X. If X is a vector, `fft(X)` returns the Fourier transform of the vector. If X is a matrix, `fft(X)` treats the columns of X as vectors and returns the Fourier transform of each column (Matlab, 2024).

For each major group, we apply the Discrete Fourier Transform (DFT) to the Chained CPI and plot the scaled power in the upper part of the graph. Additionally, we apply the DFT to the monthly weight and plot the scaled power in the lower part of the graph. As we are analyzing the seasonal components of Chained CPI and CPI research series, we limit our examination to periods up to 12 months in the graph.

2.2 Data Analysis

For the Chained CPI of major groups, we did not observe much seasonality, except in Apparel. It is consistent with our observation in Figure 1 where the Chained CPI of Apparel exhibits seasonality.

The scaled-power plot of the Chained CPI for Apparel, shown in the upper part of Figure 3, exhibits a significant peak at a 6-month period. Although the plot does not provide information about the specific months causing the peak, one may identify these months by pairing calendar months with a 6-month interval (e.g., January and July, February and August, and so on), and comparing the six paired groups. Figure 4 displays boxplots where the pair of April and October has the largest median, followed by the pair of March and September. In each boxplot, the central mark indicates the median, while the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' marker symbol (Matlab, 2024). We also note that the pair of April and October has a longer box than the pair of March and September, indicating greater variability. The pair of March and September has the largest mean, as shown in Table 2.

Note that weights are relative to the sum of all item weights. The scaled-power plot of the monthly weight of Apparel, shown in the lower part of Figure 3, reveals a notable peak at a 12-month period. Figure 5 displays boxplots of monthly weights, where December has the largest median. Table 3 shows that December has the largest mean as well as the largest median.

From Table 1, we observe that Education&Communication is negatively correlated with all other major groups except Apparel. Another interesting aspect of Education&Communication is that between November 2018 and June 2020, CPI-q, CPI-a, and CPI-b appear to move in the opposite direction to the Chained CPI, as shown in Figure 6. Table 4 also shows that CPI-q, CPI-a, and CPI-b have very low or negative correlations with the Chained CPI, while CPI-q, CPI-a, and CPI-b have highly positive correlations between each other.

The scaled-power plot of the Chained CPI for Education&Communication, shown in the upper part of Figure 7, reveals minimal seasonality in the Chained CPI. However, the scaled-power plot of the monthly weight for Education&Communication in the lower part of Figure 7 displays a notable peak at a 6-month period. As with the Apparel Chained CPI, we identify these months by pairing calendar months with a 6-month interval, and comparing the six paired groups. Figure 8 shows that the pair of February and August has the largest median. The pair of February and August also has the largest mean, as shown in Table 5.

The Chained CPI for Transportation exhibits the most variability among the major groups, as measured by the coefficient of variation. The scaled-power plot of the Chained CPI for Transportation, shown in the upper part of Figure 9, reveals minimal seasonality in the Chained CPI. However, the scaled-power plot of the monthly weight for Transportation, in the lower part of Figure 9, shows a significant peak at a 12-month period. We note that July has the largest median, followed by August (Figure 10), while August has the largest mean, as shown in Table 6. Figure 11 of monthly weights for July and August shows that the monthly weight for August has mostly been larger than for July, except in 2018, 2019, and 2022.

For CPI-q, minimal seasonality is observed in most major groups, except for Apparel. For quarterly weights, mild to considerable seasonality is observed in Apparel, Education&Communication, Food, and Transportation.

For CPI-a, minimal seasonality is observed in most major groups, except for Apparel. For annual weights, minimal seasonality is observed in all major groups.

3. CPI Products at the Final Aggregation Level

We now analyze CPI values at the final aggregation level. Figure 12 shows that, except for the first few months, the Chained CPI remains the lowest, followed by CPI-q, CPI-a, and CPI-b, in that order. In other words, the CPI values are consistently ordered from largest to smallest: CPI-b $\geq$ CPI-a $\geq$ CPI-q $\geq$ Chained CPI. Correlation analysis reveals that they are highly positively correlated each other, with each correlation value being 1 to two decimal places. Singular Value Decomposition (SVD) analysis also indicates high correlation among them: the first singular value is exceptionally large, while the others are close to zero. In general, the values of Chained CPI, CPI-q, CPI-a, and CPI-b increase steadily up to approximately two-thirds of the series, and then rise sharply thereafter.

3.1 Linear Fitting of Chained CPI

To locate the slope change point, we fit the Chained CPI linearly against time (month), compute the residuals, and locate the lowest point as shown in Figures 13, 14, and 15. Since the CPIs are highly correlated, one may use CPI-q, CPI-a, or CPI-b instead of Chained CPI to locate the slope change point. Figure 15 shows a magenta line indicating the slope change point, May 2020. Figure 16 displays both the magenta line, indicating the slope change point, and a cyan line, indicating the official COVID-19 shutdown point, March 2020.

Table 7 presents the coefficient values of Chained CPI and CPI research series before and after the slope change month. We observe that the slope coefficients are significantly larger after the slope change month than before. Overall, the slope coefficients of CPI-b and CPI-a are larger than those of CPI-q and the Chained CPI.

Figure 17 displays the differences between the CPI research series and the Chained CPI. Note that the Chained CPI is considered the gold standard. The differences seem to increase after the slope change month. The histogram in Figure 18 also shows that the differences between the CPI research series and the Chained CPI increase after the slope change month. We compute the Relative Absolute Difference (RAD) of the CPI research series from the Chained CPI, and compare the values before and after the slope change month. For example, consider the difference between

CPI-b and Chained CPI. Let d represent the difference between CPI-b and Chained CPI:

$$\begin{aligned} \text{norm}(d) &= \sum |d|, \\ \text{RAD}(d) &= \frac{\text{norm}(d)}{\text{norm(Chained CPI)}}. \end{aligned}$$

Table 8 shows that the RAD value of CPI-b is the largest, while CPI-q has the smallest RAD value. RAD values are twice as large after the slope change month compared to before.

3.2 Wavelet Application

A wavelet means a small wave. That is, it is a wave-like function of time that vanishes outside a finite closed interval. Though some wavelets are not exactly zero, they become essentially zero as time approaches positive or negative infinity. Examples include the Mexican hat, Gabor, and Morlet wavelets. This property, localized in time, contrasts with Fourier basis functions, which extend infinitely in time. Wavelets are localized in frequency. They oscillate and have average values of zero:

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0.$$

The Fourier basis, consisting of sines and cosines, is perfectly localized in frequency but does not decay as time approaches positive or negative infinity. A time series with frequency changes over time, abrupt jump discontinuities, and nonstationary variances is better analyzed using suitable wavelets rather than Fourier basis functions. Wavelets are constructed from a function $\psi(t)$ called the “mother wavelet” by dilation and translation in time and frequency:

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right),$$

where the parameters for dilation, a, and translation, b. In wavelet applications, the scale parameter controls the stretching or compressing of the function. A smaller scale factor compresses the wavelet, while a larger scale stretches it. As the scale increases, the wavelet becomes wider and includes more of the time series, but finer details become less distinct. The “optimal” choice of a wavelet basis depends on the application.

Figure 19 displays a few examples of wavelets, which decay to zero as time approaches positive or negative infinity, with average values of zero and norms of 1. The first example, (1,1), in Figure 19 is the Laplacian wavelet. We apply the Laplacian wavelet to the residuals from the linear fit of the Chained CPI. Its shape allows it to capture breakpoints well where the means on the left and right differ from the central mean.

We consider the Laplacian wavelet applied to the residuals from the linear fitting where the scale base is $s = 2^{\frac{6}{n}}$, and $n = 32$. s^i represents the window size, where $i = 1, \ldots, 6$. Naturally, the maximum window size is determined by the number of data points, which is 108 months in our case. As i goes from 1 to 6, the window size increases at each iteration: $1.30, 1.68, 2.18, \ldots, 64.00$.

Figure 21 shows the Laplacian wavelet applied to the residuals, with a color bar. The lighter the color, the greater the intensity. We observe the color gets lighter starting at the 25th level, and becomes intensely lighter in the later stages. Note the concentrated brightness around the slope change month.

The Haar wavelet, located below the Laplacian wavelet at (2,1) in Figure 19, is applied to the differences of the Chained CPI. The Haar wavelet is known for capturing jump points well, where the right mean differs from the left mean. A series with a break point as in our data example becomes a series with a jump point when the difference of the series is taken. We take the difference of the Chained CPI, $X_t - X_{t-1}$, as shown in Figure 20. For the Haar wavelet application on differences, Figure 22 shows that the color becomes lighter starting at the 20th level, intensifying in the later stages.

In both the Laplacian and Haar applications, we do not find any other abnormalities except in the area of the slope change point.

4. Summary

Fourier transform can be useful to detect the seasonality of time series data such as CPI products and weights. Methods can be applied at any levels of aggregation. Wavelet analysis is particularly well-suited for analyzing time series with characteristics like changing frequencies over time, sudden jumps (discontinuities), and nonstationary variances because it provides a localized time-frequency representation. It allows us to examine how different frequencies behave at specific points in time, making it ideal for capturing transient events and variations in signal behavior across different scales.

Acknowledgment

I would like to express sincere gratitude to Joshua Klick and Jessica Park who generously shared their time and insights throughout this study.

REFERENCES

Bureau of Labor Statistics (2023), *Handbook of Methods*, “Consumer Price Index”, Available at: https://www.bls.gov/opub/hom/cpi/.

Klick, J., and Park, J. (2022), “Analysis of Quarterly Revisions of Fixed Quantity Consumer Price Index Products.” *Proceedings of the American Statistical Association*, 2022. Available at: https://www.bls.gov/osmr/research-papers/.

Martinez, W.L., and Cho, M.J. (2015), *Statistics in Matlab: A Primer*, CRC Press, New York.

The MathWorks Inc. (2024), *Statistics and Machine Learning Toolbox Documentation*, Natick, Massachusetts: The MathWorks Inc. Available at: https://www.mathworks.com/help/stats/index.html.

Table 1: **Chained CPI Correlation Coefficients between Major Groups**

	A	E	**F**	**G**	**H**	**M**	R	T
A	1.00	0.30	-0.19	-0.28	-0.29	-0.38	-0.09	0.20
E	0.30	1.00	-0.57	-0.70	-0.72	-0.75	-0.51	-0.38
F	-0.19	-0.57	**1.00**	**0.96**	**0.95**	**0.90**	0.97	0.81
G	-0.28	-0.70	**0.96**	**1.00**	**1.00**	**0.98**	0.91	0.72
H	-0.29	-0.72	**0.95**	**1.00**	**1.00**	**0.98**	0.90	0.71
M	-0.38	-0.75	**0.90**	**0.98**	**0.98**	**1.00**	0.83	0.58
R	-0.09	-0.51	0.97	0.91	0.90	0.83	1.00	0.88
T	0.20	-0.38	0.81	0.72	0.71	0.58	0.88	1.00

Table 2: **(Apparel) Paired Chained CPI Average**

	(1,7)	(2,8)	**(3,9)**	**(4,10)**	(5,11)	(6,12)
Median	98.38	100.73	103.05	**103.23**	101.58	99.03
Mean	98.05	99.96	**102.10**	102.04	100.51	98.42

Table 3: **(Apparel) Monthly Weight Average**

	Jan	Feb	Mar	Apr	May	Jun
Median	0.034	0.025	0.026	0.029	0.028	0.027
Mean	0.033	0.026	0.028	0.028	0.029	0.027
	Jul	Aug	Sep	Oct	Nov	**Dec**
Median	0.026	0.030	0.029	0.028	0.033	**0.038**
Mean	0.027	0.029	0.029	0.029	0.033	**0.037**

Table 4: **(Education&Communication) Correlation Coefficients between Indexes**

	Chained CPI	*CPI-q*	*CPI-a*	*CPI-b*
Chained CPI	**1.000**	0.248	0.018	-0.183
CPI-q	**0.248**	1.000	0.924	0.861
CPI-a	**0.018**	0.924	1.000	0.977
CPI-b	**-0.183**	0.861	0.977	1.000

Table 5: **(Education&Communication) Paired Monthly Weight Average**

	(1,7)	**(2,8)**	(3,9)	(4,10)	(5,11)	(6,12)
Median	0.070	**0.078**	0.070	0.061	0.060	0.062
Mean	0.072	**0.078**	0.073	0.063	0.060	0.064

Table 6: **(Transportation) Monthly Weight Average**

	Jan	Feb	Mar	Apr	May	Jun
Median	0.156	0.154	0.161	0.165	0.168	0.168
Mean	0.154	0.153	0.159	0.165	0.167	0.169
	Jul	**Aug**	Sep	Oct	Nov	Dec
Median	**0.176**	0.176	0.169	0.164	0.159	0.155
Mean	0.174	**0.176**	0.170	0.164	0.161	0.157

Table 7: **CPI Coefficients (before and after slope change month)**

	Overall		*Before*		*After*	
	B0	**B1**	B0	**B1**	B0	**B1**
Chained CPI	96.77	**0.19**	99.36	**0.11**	105.39	**0.61**
CPI-q	96.78	**0.21**	99.24	**0.13**	106.25	**0.64**
CPI-a	96.57	**0.22**	99.17	**0.14**	107.06	**0.66**
CPI-b	96.57	**0.22**	99.13	**0.15**	107.26	**0.66**

Table 8: **Relative Absolute Difference from Chained CPI (before and after slope change month)**

	Overall	*Before*	*After*
CPI-q	0.008	0.006	**0.011**
CPI-a	0.012	0.009	**0.021**
CPI-b	0.014	0.010	**0.022**

Figure 1: **Chained CPI by Major Group**

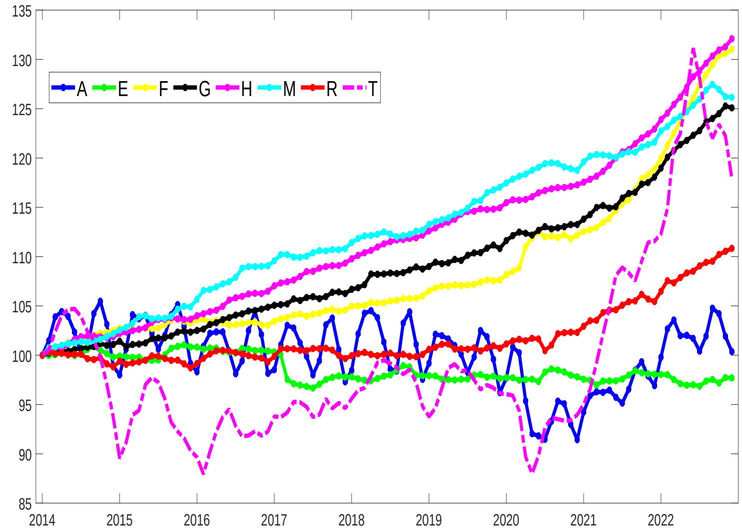

Figure 2: **Monthly Weight by Major Group**

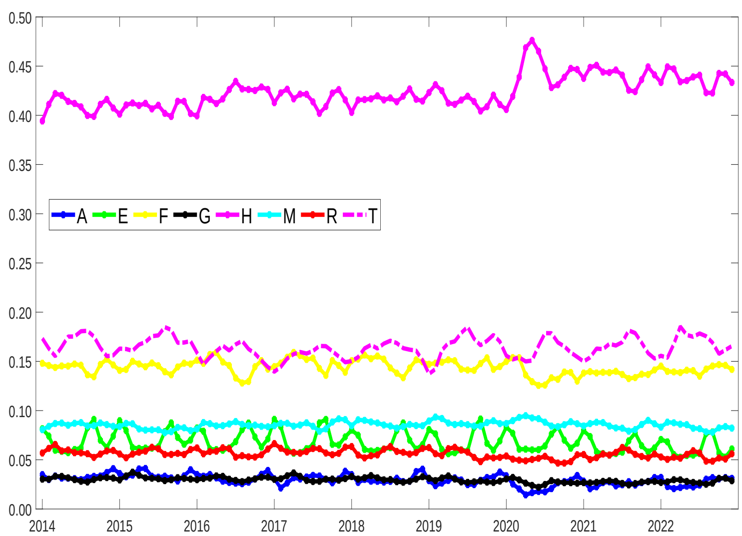

Figure 3: (Apparel) Scaled Power Plot

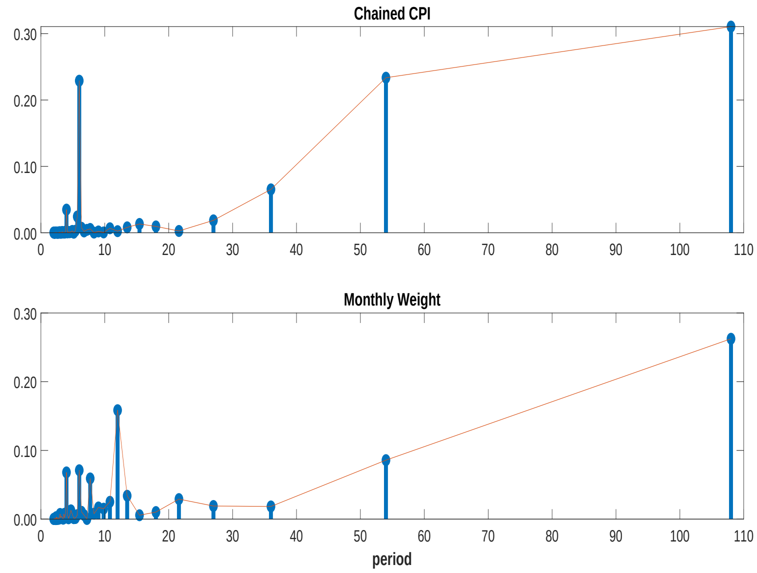

Figure 4: (Apparel) Boxplot of Paired Chained CPI

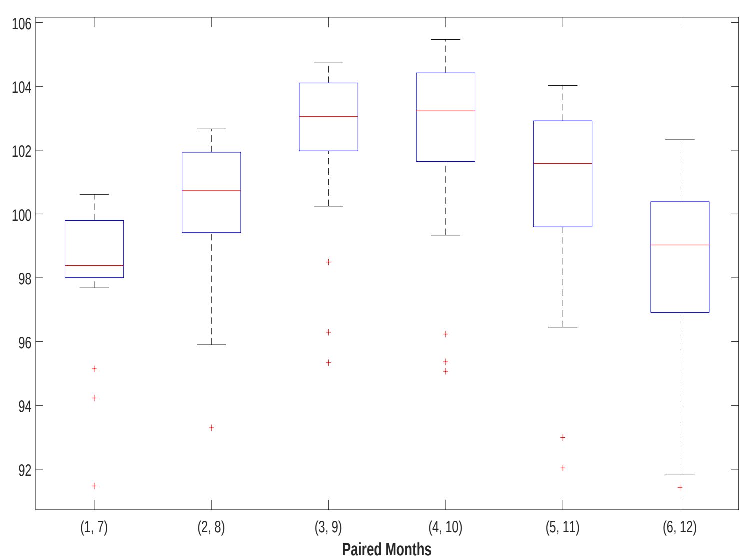

Figure 5: (Apparel) Boxplot of Monthly Weight

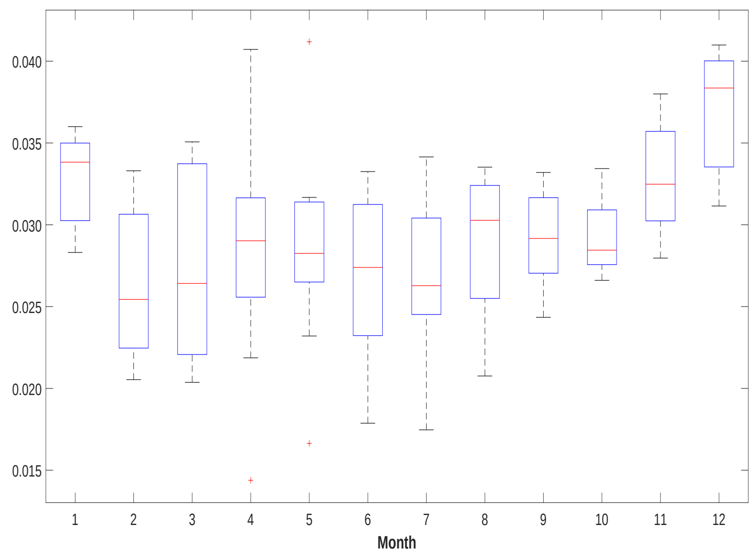

Figure 6: (Education&Communication) Chained CPI and CPI Research Series

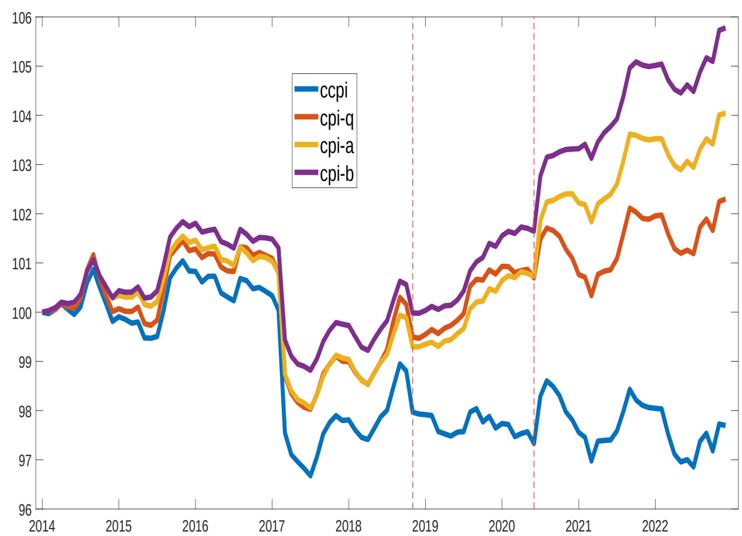

Figure 7: **(Education&Communication) Scaled Power Plot**

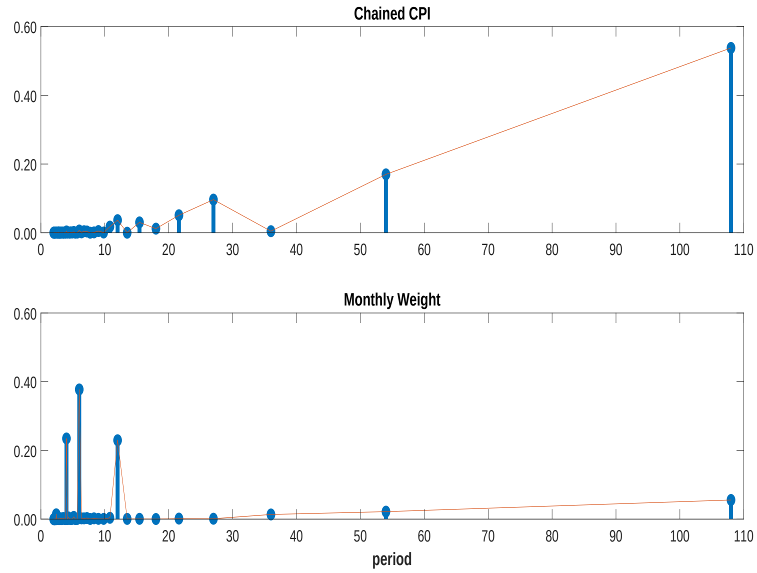

Figure 8: **(Education&Communication) Boxplot of Paired Monthly Weight**

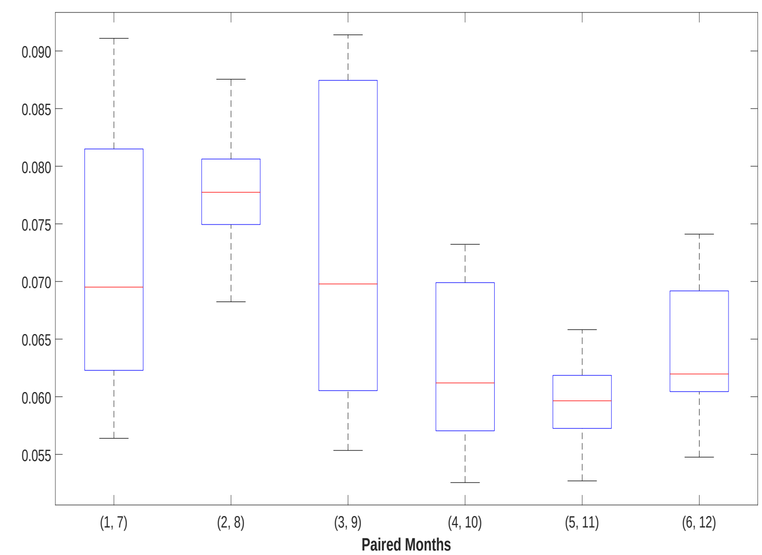

Figure 9: (Transportation) Scaled Power Plot

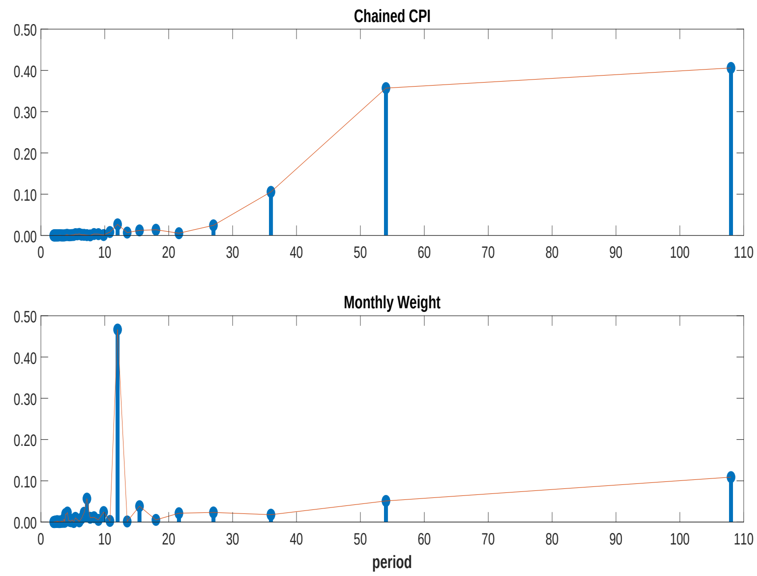

Figure 10: (Transportation) Boxplot of Monthly Weight

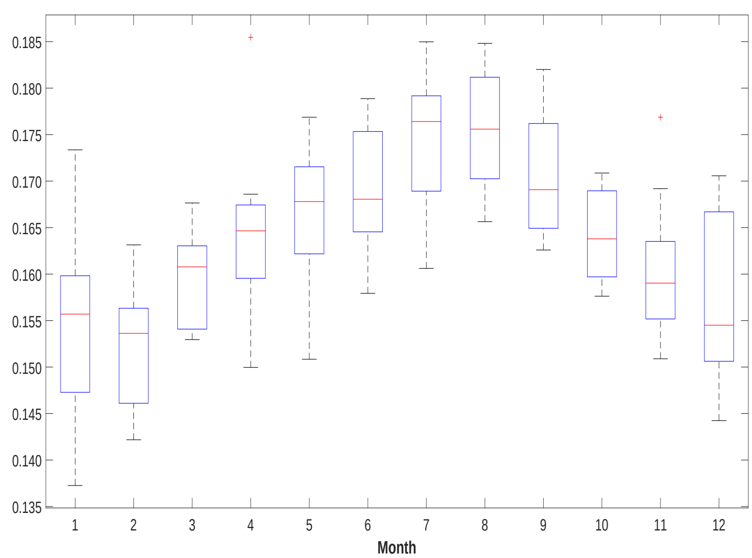

Figure 11: **(Transportation) Monthly Weight of July and August**

Figure 12: **Chained CPI and CPI Research Series at the Final Aggregation Level**

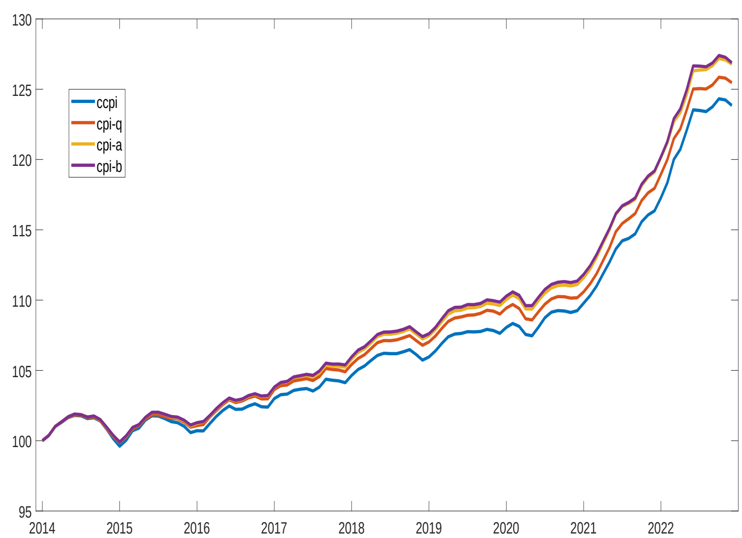

Figure 13: Chained CPI Linear Fit

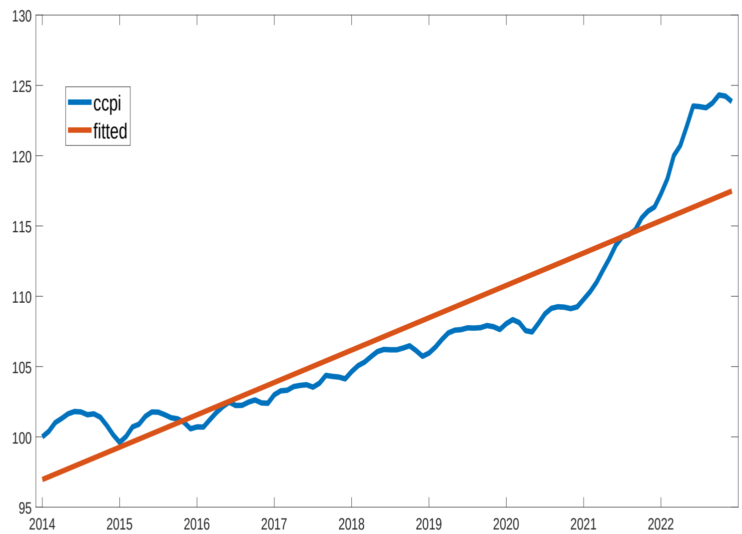

Figure 14: Residual

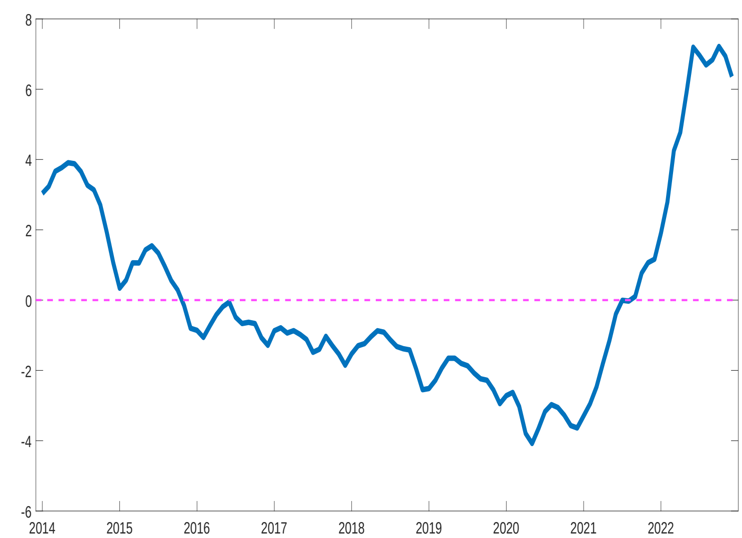

Figure 15: Locating Slope Change Point

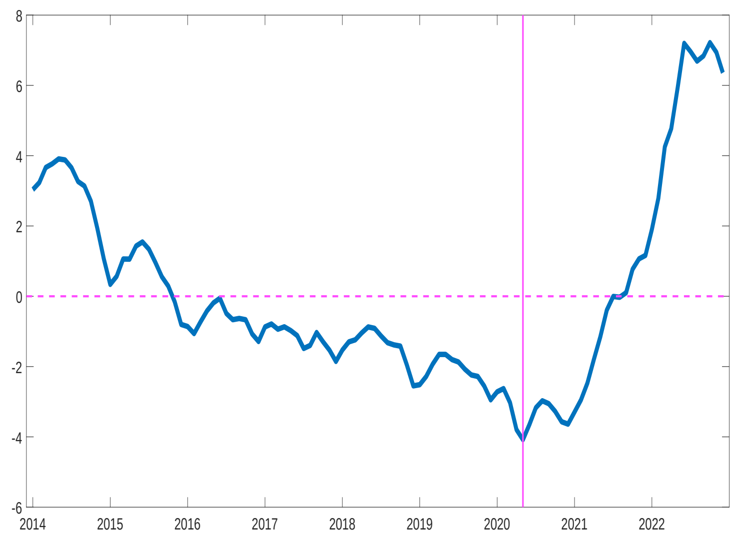

Figure 16: Chained CPI and CPI Research Series (with COVID and Slope Change Lines)

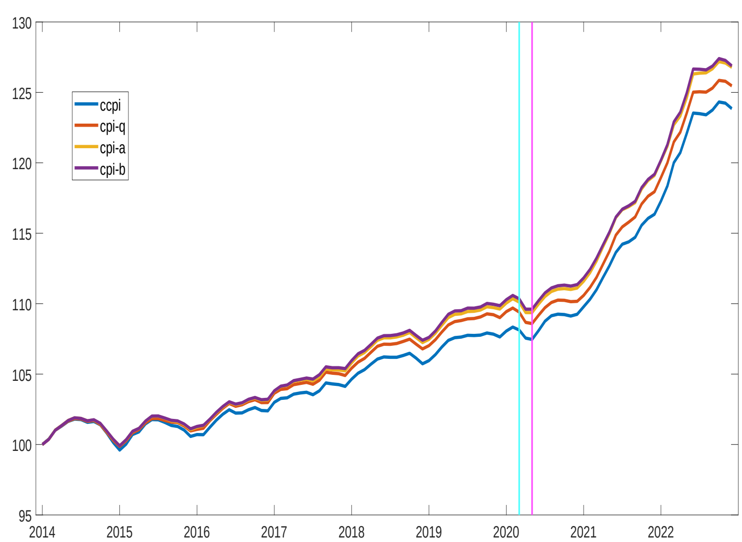

Figure 17: **Differences between Chained CPI and CPI research series (with a slope change point)**

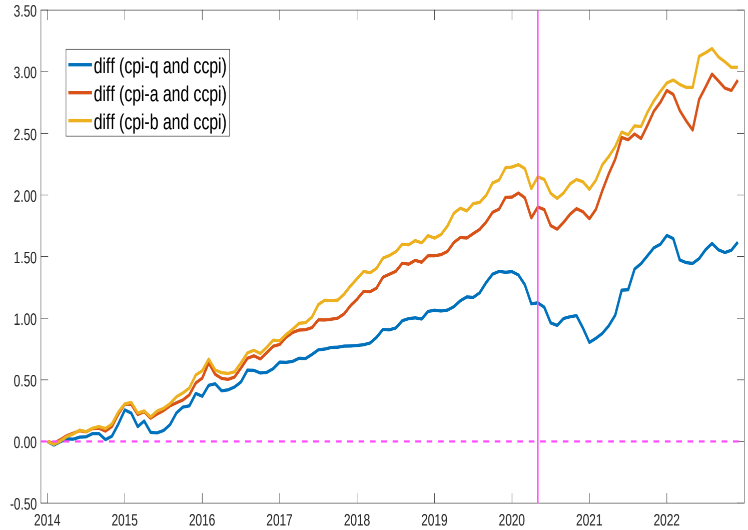

Figure 18: **Histogram of Differences between Chained CPI and CPI research series**

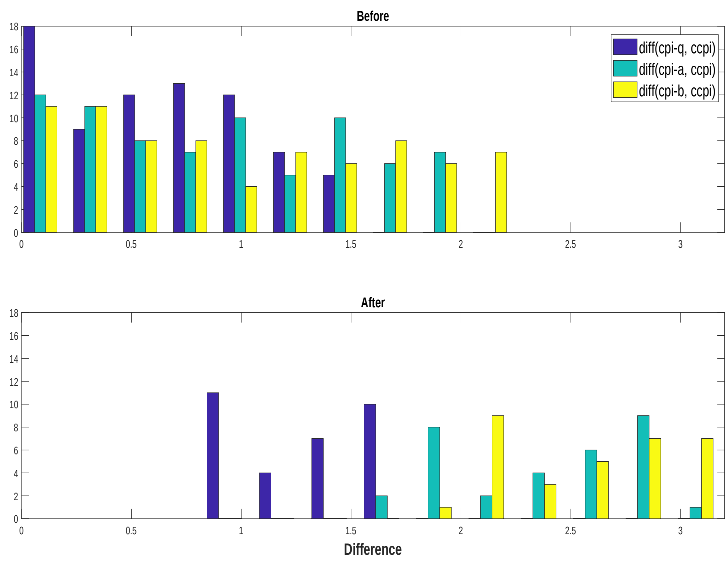

Figure 19: **Plots of Wavelets**

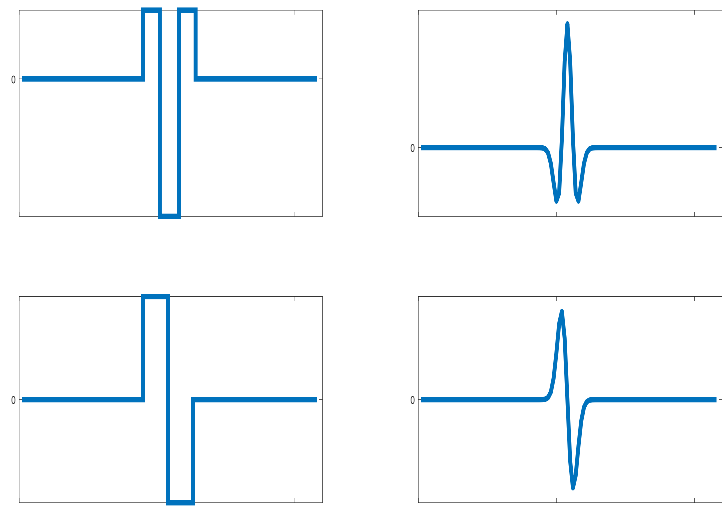

Figure 20: **Difference of Chained CPI:** $X_t - X_{t-1}$

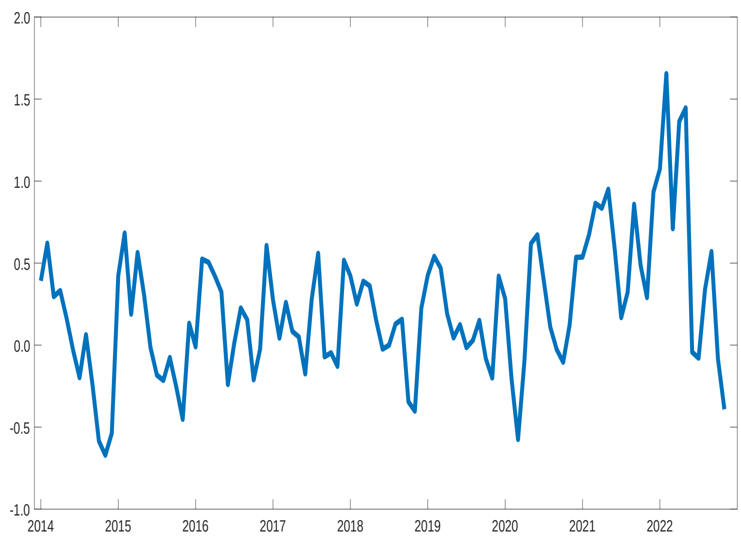

Figure 21: **Laplacian Wavelet on Residual**

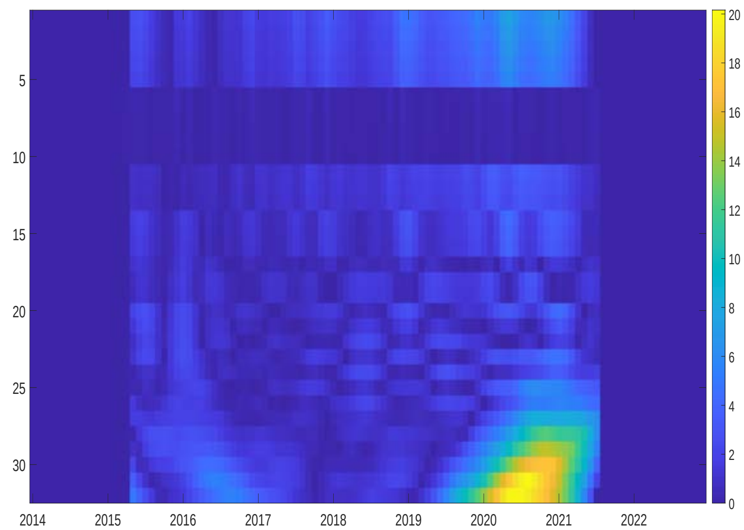

Figure 22: **Haar Wavelet on Difference**