PHY408 Final report

By: Sadeed Ahmed

INTRODUCTION

The CPI (Consumer Price Index) measures the average change in prices paid by urban consumers for a basket of consumer goods and services. It includes detailed price data for various categories such as food, energy, shelter etc.

The CPI is used for a variety of reasons including, but not limited to,:

- As an economic indicator
- As a deflator of other economic series
- As a means of adjusting dollar values: The index affects the income of more than 90 million people because of statutory action. It is also used to adjust the federal income tax structure. [1]

The dataset that I am analyzing is the 12-month-percentage change in the CPI of the U.S. for selected categories over the past 20 years. The goal of this analysis is to consider the relationship between different categories in the CPI and how they influence each other using analytical tools from topics such as Convolution and impulse response, Discrete Fourier transform, Filter design and stochastic processes etc. My analysis mainly consider 3 questions:

1) What is the relationship between the CPI of natural gas and the CPI of energy and how change in one influences the other?
2) What has been the response of the shelter CPI (Housing and rental prices) to the recessions in the past (Namely the Great Recession of 2007 and the COVID-19 recession)?
3) What are the seasonal patterns (i.e. recurring patterns or cycles) in the pricing of the food CPI?

This analysis uses two datasets. The first, as mentioned before, uses the table data of the CPI from this link. The second dataset is the civilian unemployment rate 2020 (fetched using the table data from this link). The reason for using the unemployment data is to try to get an idealized model of the Great recession (Since any recession is marked by a significant increase in unemployment rate).

In order to retrieve the table data on these pages, I had to use a web scraper since there was no direct way of downloading the table from the page.

---

1 https://www.investopedia.com/terms/c/consumerpriceindex.asp#:~:text=discounted%20by%20economists,-,The%20Bottom%20Line,the%20Bureau%20of%20Labor%20Statistics.

ANALYSIS AND RESULTS

1) Energy and Gas relationship: (analysis.py)

The first part of my analysis aims to investigate if there is any correlation between the CPI of natural gas and energy given the monthly percentage change in these categories over the last 20 years. The suggestion for a relation between these two becomes more evident when looking at the graphs of the energy and natural gas CPI separately as shown in figure 1.1 below:

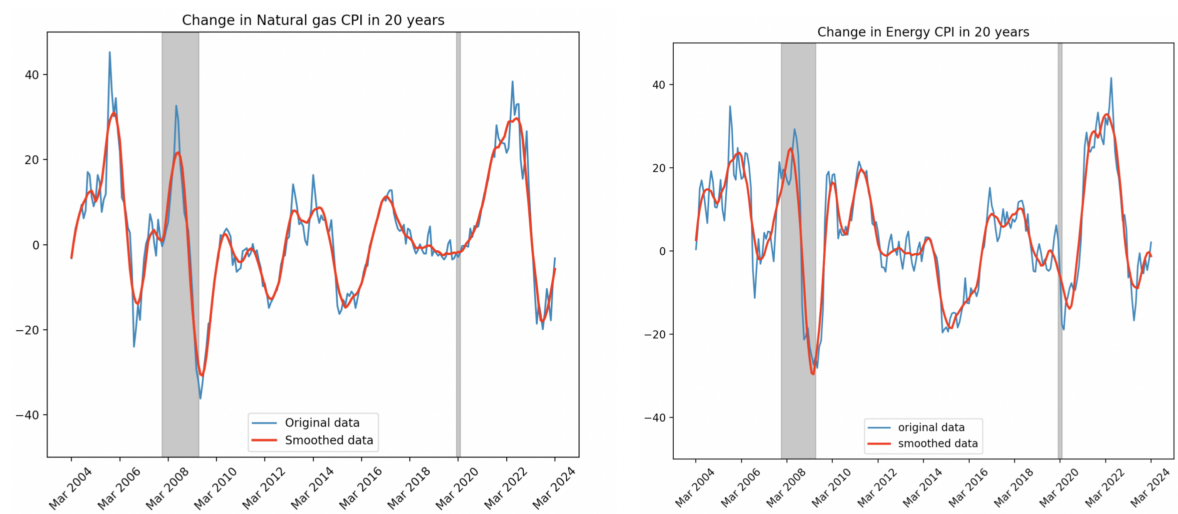


Fig 1.1: Comparison of change in CPI between Natural Gas and Energy

The gray blocks in each graph indicate the two recessions in the last two years i.e, the Great recession and the COVID-19 recession. The smoothed data in red for each data set is obtained by using a Savitzky-Golay (savgol) filter from the scipy library. The SAVGOL filter smooths out the data using convolution, by fitting successive subsets of adjacent data points with a low-degree polynomial by the method of linear least squares.

Looking at the smooth natural gas and energy CPI, the relation between the two seems really similar and change in one seems to be correlated to change in the other at nearly the same time. However, my analysis aims to use the correlation between the natural gas and the energy CPI to indicate that change in the former results in a lagged change in the latter i.e., when the gas price gets expensive or cheaper, there is a lag that has to occur for the price change to occur in the energy.

To examine this relationship, my goal was to calculate the cross-correlation between the natural gas and energy CPI. In order to use cross-correlation, we first need to ensure that both the natural gas and energy CPI time series are stationary i.e.

- They both have a fairly constant mean and variance that is not affected by time
- They don't have any seasonality (periodic behavior over time that is predictable)

Looking at the raw time series of energy and natural gas, we can visually ascertain that the mean and the variance for the most part doesn't fluctuate much over time. Moreover, there is also no seasonality in either of the time series thus it's safe to assume that both our time series are stationary and thus we can compute the time-lagged cross-correlation between them.

In order to compute the cross-correlation, we used the FFT method from question 1 of lab 4. Code for this computation is in the function titled energy_and_gas_relation() in the Analysis.py file. The resultant cross-correlation values of the two time series over the entire range of possible lags (both positive and negative) is shown below in Fig 1.2:

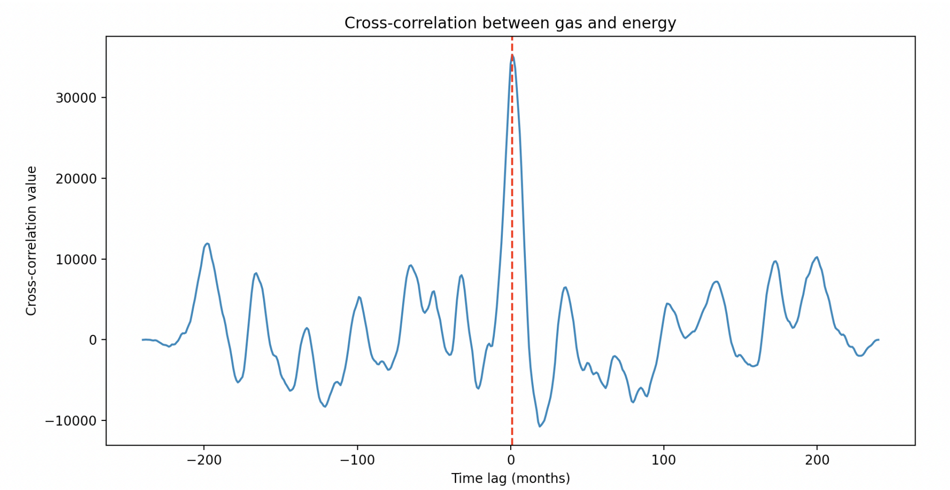


Fig 1.2: Cross-correlation between gas and energy

The peak in the cross-correlation between gas prices and energy prices at lag = 1.9 indicates that the change in gas prices has a strong correlation with the change in energy prices shortly thereafter. More notably, the lag at 1.9 indicates that nearly 2 months after the increase in gas prices, the price of total energy goes up. This significant correlation can suggest a potential causal relationship. This causal relationship is confirmed by the BLS in this article (https://www.bls.gov/opub/ted/2007/jul/wk5/art01.htm) according to which "..a double-digit decrease in utility (piped) natural-gas prices was the main factor behind the deceleration in energy prices…). The topic of this article is the decrease in energy prices between 2005 and 2006. The article further confirms this causal relationship by addressing the fact that "..During 2006, as natural-gas production capacity was restored and as supplies recovered following the previous years' hurricanes, natural-gas prices decreased 14.2 percent, after 30.2 percent in 2005."

2) Impulse response modeling of Shelter CPI: (recession_impact.py)

The next part of my analysis aims to explore the impulse response of the shelter CPI where the event of a recession is considered as an impulse.

In order to model the recession as a mathematical function, we use unemployment data from the last 20 years. This is because every recession is marked by a significant increase in unemployment rates. To make this modeling of the recession more accurate, we could have considered various other economic indexes such as GDP decline, industrial production etc and then normalize those indexes to get a more comprehensive impulse input. However, doing so would involve working with multiple other data sets so for the purposes of simplicity, our impulse (a recession) is simply modeled by the unemployment rates at the time of the great recession. The time frame that we consider for the recession includes the time a few months before the recession and a few months after the recession in order to capture a fuller image of the recession's economic impact. The effect of recession on the shelter CPI is made much more apparent by taking a look at the unemployment rates and the shelter CPI side by side after the onset of the great recession shown in fig 2.1 below:

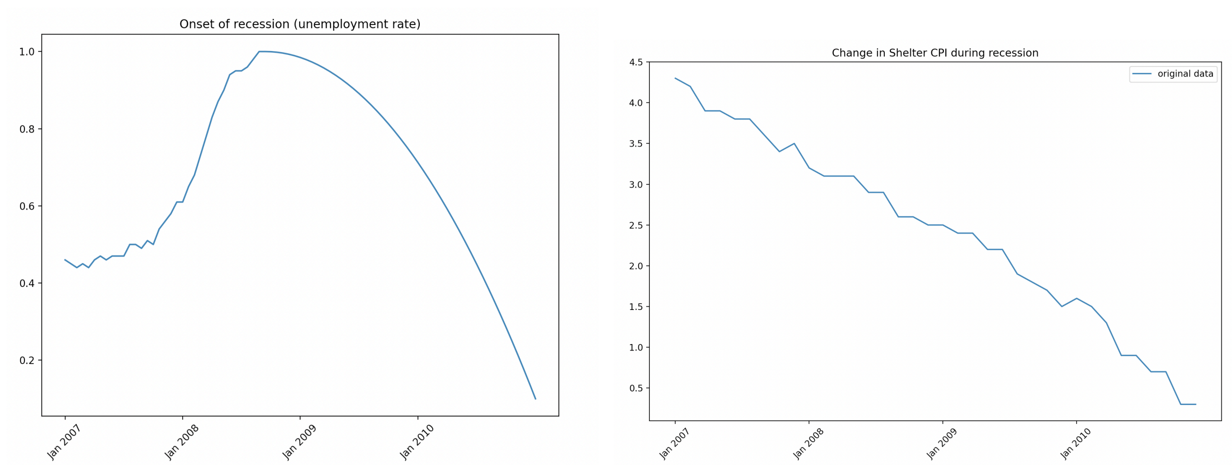
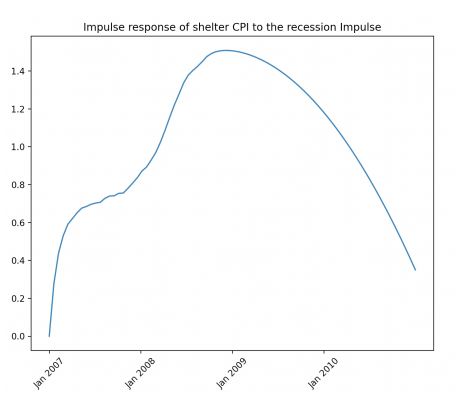


Fig 2.1: Onset of recession and the original decline in Shelter CPI

The hypothetical response of the shelter CPI to the event of recession as an impulse is modeled as an exponential decay function. Moreover, the image of recovery (decline of the unemployment rates after the end of recession) is estimated as a declining curve. The actual drop in unemployment rates are much more gradual. However, since we are not considering a normalized version of all the economic indexes constituting a recession, we can make this estimation.

After setting up the impulse and the impulse response of the shelter CPI, we convolve the response of the shelter CPI with the recession impulse to get the following result:



Even though the original shelter CPI was already in decline before the start of the great recession, the continued decline of the shelter CPI after the onset of the great recession suggests the direct impact of the recession on the housing and shelter prices. This is further reified by historical data of the shelter CPI and this effect is also more pronounced following the COVID-19 recession.

1) Seasonality patterns in the food CPI: (analysis.py)

The last part of my report attempts to analyze seasonal fluctuations that occur in the food prices using DFT and filtering. To mitigate the influence of the 2 recessions in the last 20 years, we limit our time series to include data between January 2010 and Jan 2020 shown in figure 3.1:

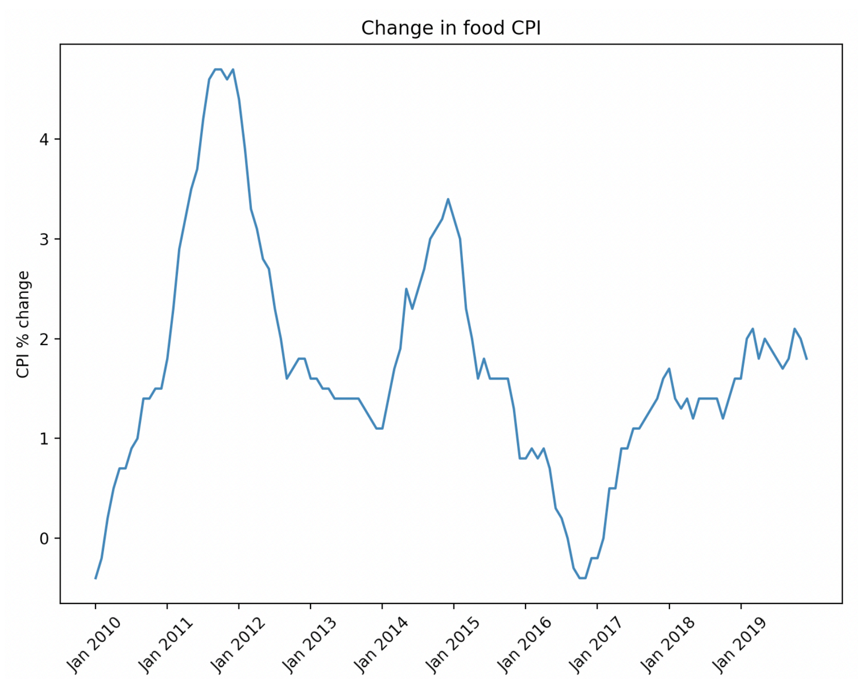


Fig 3.1: Change in food CPI over 10 years

Since our food data is non-stationary (because of the assumption that there is seasonality in the food prices), it can contain long term trends that can interfere with our frequency analysis by creating low-frequency components that dominate our Fourier Transform. For this reason, we use detrend from the scipy library to remove the least squares line fit from the data.

After removing the long-term trend from the food cpi, we perform discrete fourier transform on the food CPI data to get information about the underlying frequencies in the signal and thus the seasonality. After setting up the sampling frequency and performing the DFT, we take only the positive frequencies from our result. This is because our original signal is real-valued and thus the FFT result would be conjugate symmetric. The result of performing FFT on our signal is shown below in figure 3.2:

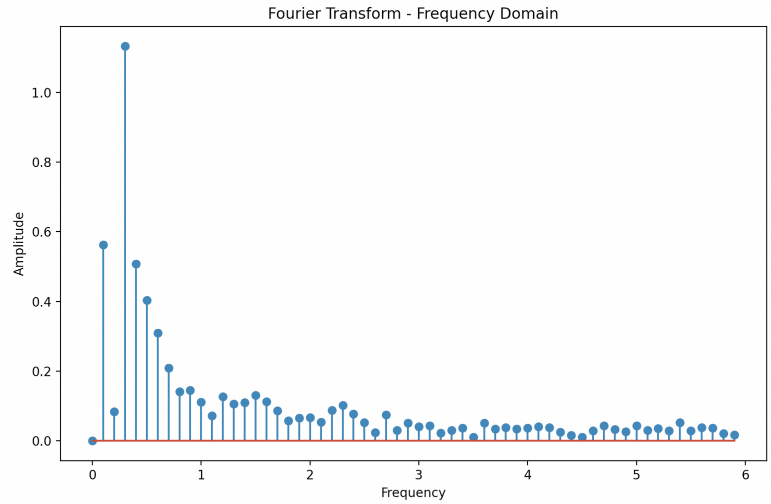


Fig 3.2: Result of FFT on food CPI

The peaks on the frequencies between 0 and 1 indicate cyclical patterns that cannot be predicted. For that reason, we only consider the peaks at the frequencies between 1 and 2 which would indicate annual or bi-annual patterns in the food prices.

To further study these annual and bi-annual patterns, we create a frequency mask that zeros out all frequencies that are not between the range 0 and 1 in the function seasonal_frequencies(). This allows us to isolate the frequency components associated with seasonality. We then perform an IFFT on these frequency components to map them back into the time domain. The plot of the real values of this signal mapped back into the time domain is shown below in figure 3.3:

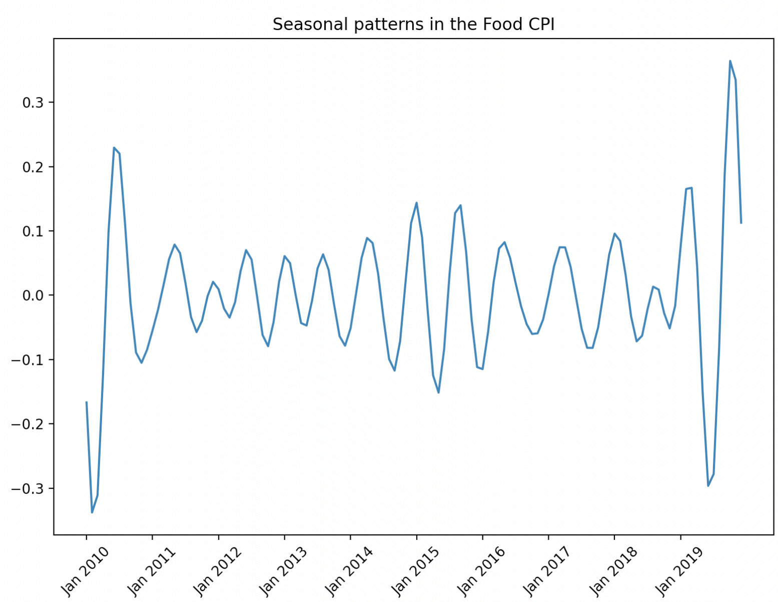


Fig 3.3: Seasonal patterns in the food CPI

The peaks and troughs at semi-regular intervals indicate a recurring, predictable pattern in the food CPI. In particular, in the span of every 12 months, there is at least one point in the year (usually in summer to late summer months) when the food prices are at their highest for that year. Moreover, towards the end of the year, there is a recurring pattern of the food prices being at their all time low for that year.

DISCUSSION

The combination of the three analyses done above aims to draw a relation between some of the categories of the CPI and how change in one category could have an impact on other categories. It also aims to discuss the possibility of seasonality and predictability in a particular CPI category. With more advanced forecasting tools and with more in depth analysis of the inter-dependencies of these categories on each other from the past, these analyses can be broadened to forecast future changes in a category of interest and thus be used as an investment tool and a strong predictor for future inflation trends. Given the reliable and monthly reporting of this data, we don’t have to worry about missing any external links and we can work with the data as a given.

The caveat for the analyses laid out in this report is that for the purposes of conciseness, many simplifying assumptions had to be made (for example the tapering of unemployment rate after the Great Recession). In reality, a more robust analysis would consider various factors or economic indexes that compound to a recession and thus that would result in a neater analysis of the impulse response of various categories in the CPI.