How Has News Sentiment Affected Housing Prices in China?

By: Richard Reynard (14170539)

Subject : SC3901 - Special Topic 1
Supervisor : Dr. Tan Sook Rei
Date of Submission : 13 October 2023
Final Word Count : 2750

Table of Contents

1. Introduction

Sentiment analysis is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes (Liu, 2015). Sentiment Analysis is also known as subjectivity analysis, opinion mining, and appraisal extraction, with some connections to affective computing (Pang and Lee, 2008). In sentiment analysis, the result is classified to different polarity scores, ranging from positive, neutral, and negative. Sentiment analysis is often used in the study of social media, consumer products, financial market, and politics (Feldman, 2013). Despite having been used for decades to study subjective texts, sentiment analysis is only applied into news in recent years. However, news articles are mostly objective as journalists often refrain from using subjective vocabulary to maintain their independence and therefore it is harder for the sentiment to be identified as they are not expressed lexically (Balahur et al., 2013).

The past decade has presented the pubic with various news and events, one of which is the COVID-19. In 2023 itself, most countries are slowly recovering from the global economic downturn caused by the pandemic. In China, the government had implemented "zero-COVID" policy in response to the rapidly rising cases in various cities. This has resulted in a fall in domestic demand due to poor domestic and foreign confidence in China's economy, especially in the property market (Sutherland, 2023). The China's government is aiming for 5% GDP growth while the IMF predicts that the economy could rise by 4.5% to 5% if Beijing were to implement economic transformations and stimulus packages to boost domestic expenditure.

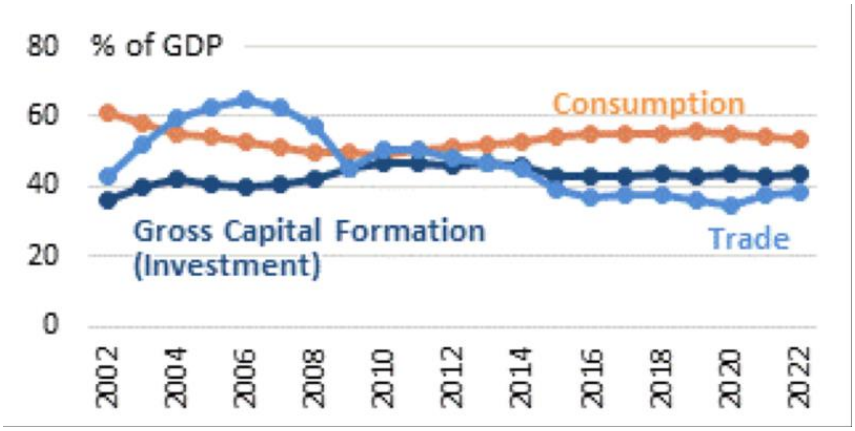


Figure 1. *Composition of China's Economy (Sutherland, 2023)*

Measures had been taken by the government to recover the economy. Value Added Tax (VAT) rebates, subsidies for electric vehicles and electronics, investment in transportation, clean energy, infrastructure, manufacturing, and agriculture. However, the measures taken were insufficient as the underlying issue of the economic challenges started before COVID-19 itself. There has been declining economic growth, increasing cost of production, trade war with U.S., uncertainty and increasing corporate and government debt, including non-financial debt which accumulated to 297% of GDP in 2022.

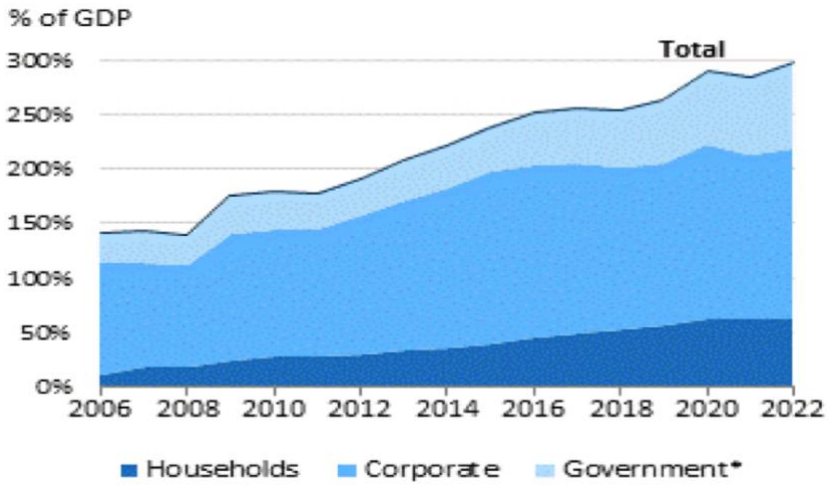


Figure 2. *Nonfinancial Debt as Share of China's GDP (Sutherland, 2023)*

The economy is further exacerbated by the default by major property developers such as China Evergrande Group and Country Garden as the revenue generated from this industry is the main source of income for the local government (Sutherland, 2023). This event has not only made it

into local newspaper but also international news channel namely CNN, BBC, The Guardian, and Channel NewsAsia.

In the modern world, aided with the adoption of digital distribution, the public are served with various news hourly. According to a study, news coverage and future economic performances are correlated in one way or another. The study reveals that published news can sway the public's opinion towards the economy and political behavior (Soroka et al., 2014). Another study by Mutz and Soss (1997) suggests that although there is limited influence in the public's opinion by the media coverage, the news organizations are able to steer the public's confidence towards the future events.

Therefore, this paper aims to study the relationship between the media coverage and housing price index in China and observe if there have been any changes in the price of property and real estate given the recent waves of instability surrounding the housing market in China. The initial hypothesis is that there has been a decline in the property prices due to the debt crisis facing some of China's biggest property developers.

2. Methods

2.1. Data

2.1.1. Sentiment Analysis

In this study, the dataset is obtained from several sources throughout the steps. News articles are obtained from Channel NewsAsia (CNA), a Singapore multinational news channel owned by public broadcaster Mediacorp. This is because the topic of the study centers around China's housing market and according to a study by Natarajan and Xiaoming (2003), CNA puts emphasis on events happening in Asia and is classified as "high credibility" and "least biased" (Channel News Asia (CNA) – Bias and Credibility, 2023), and therefore is suitable to serve as

source of dataset. 105 most recent articles about the housing market in China will be web-scrapped to determine the sentiment of each of the articles.

2.1.2. Housing Price Index (HPI)

The HPI is obtained from CEIC Data (https://www.ceicdata.com/en/indicator/china/real-residential-property-price-index) starting from March 1998 to September 2023 with 2018 as the base year.

2.2. Exploratory Analysis

The study utilizes Python 3, an interpreted and general-purpose programming language, as it is versatile and can be used for different tasks such as data manipulation, machine learning, and the possibility of presenting the result in application format. Furthermore, it is also more readable and understandable over other statistical programming language such as R (Ozgur et al., 2017).

2.2.1. Data Cleaning / Pre-processing

Raw dataset will then be pre-processed through data cleaning and manipulation by omitting the stop words such as “a”, “an”, “the”, “in” and the others to transform the data into a format that computer understand and ensure that more focus can be given to words which has higher relevance to the meaning and sentiment of the text. Advertisement will also be omitted should it exist during the web scraping process. Clean dataset will then be sorted based on date in ascending order and exported as Excel and comma-separated values (CSV) file for further study.

2.2.2. Sentiment Analysis

Each of the dataset will be scored using SentimentIntensityAnalyzer which is imported from NLTK library. The compound score is classified to 3 classes, to determine the polarity – positive (>0.05), negative (<-0.05), and neutral (between -0.05 and 0.05). The result will then be plotted to show the trend.

2.2.3. Housing Price Index

Data is plotted to show the return, trend, and initial hypothesis. Natural logarithm will be utilized as it accounts for the compounding effect of returns over time, hence increasing the accuracy of the percentage change over the time series analysis. Augmented Dickey Fuller Test will then be conducted to check for non-stationarity with 5% significance level. The null hypothesis, H_0, is that the time series is stationary and the alternative hypothesis, H_1, is that it is non-stationary. Should the result show non-stationarity, series transformation will then be conducted. Some examples of series transformations include power transformations (Box-Cox, Yeo-Johnson) and order differencing (seasonal, n^{th} order differencing). Power transformations are typically used to make data more Gaussian-like (normalized) and used in the case of scatter plot. Differencing, on the other hand, is more commonly used technique for autoregressive modelling. It can help with stabilizing the mean by removing the changes in the level of a time series, reducing the trend and seasonality. As this study is adopting time series, order differencing will be used for the case of non-stationarity.

3. Result

3.1. Sentiment Analysis

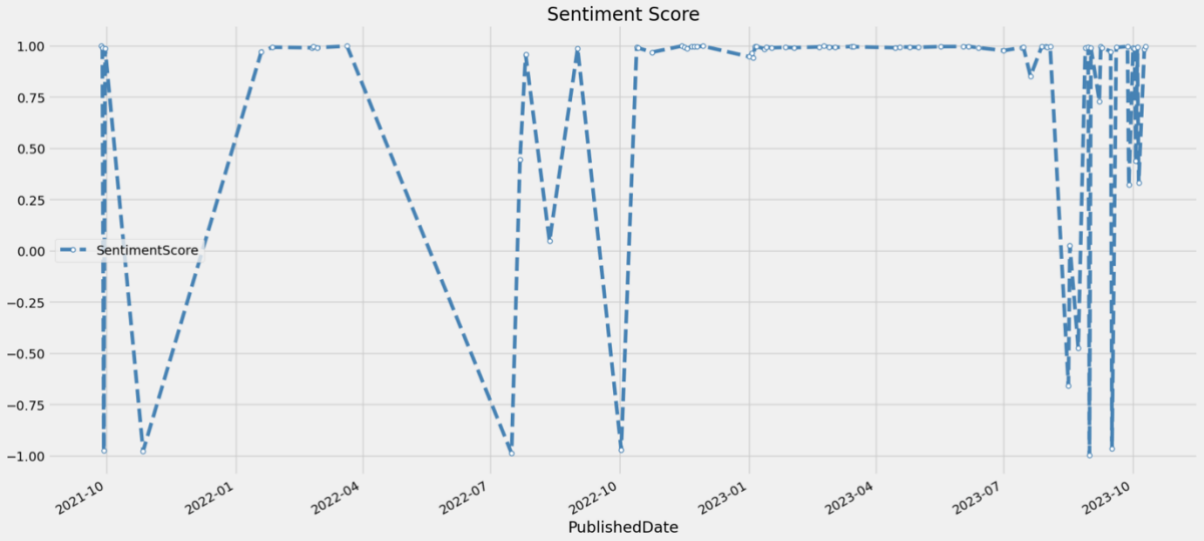


Figure 3. *Sentiment Score of Articles*

The result shows some variations in the sentiment index for news between September 2021 to October 2023. While there are fluctuations in most of the month of 2022, there are some consistencies from October 2022 until July 2023. However, the fiscal year of 2023 shows that there are significant fluctuations that occurs monthly, and in some cases daily. Sentiment is also more varied, in which there are some days with strong positives, weak positives, and strong negatives.

3.2. Housing Price Index

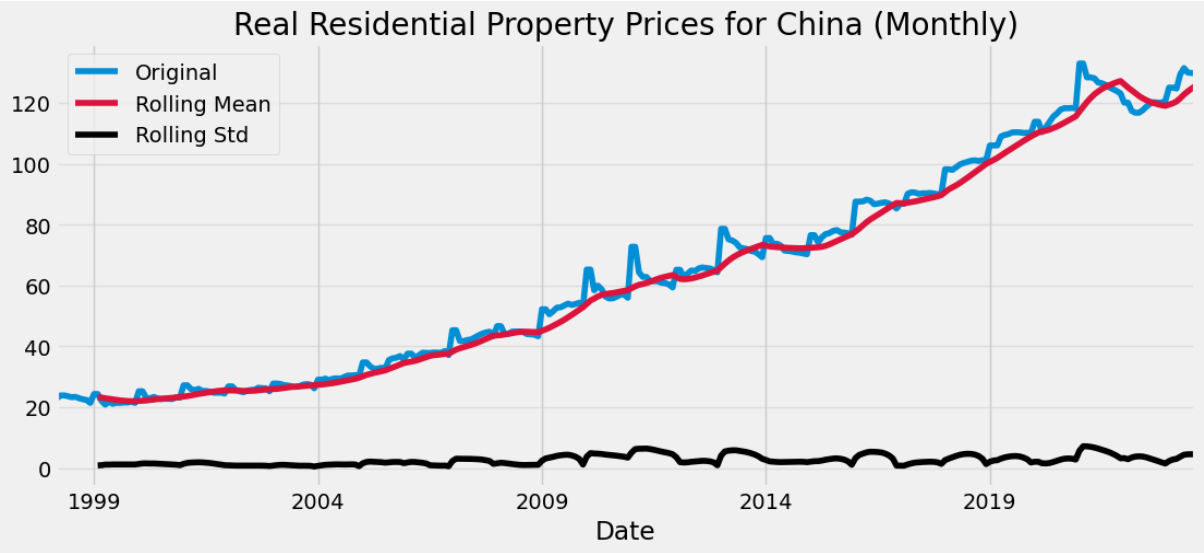


Figure 4. *Housing Price Index and Trend (China Real Residential Property Price Index, 2023)*

The graph above shows the housing price index from March 1998 to August 2023, with the year 2018 as the base year. Generally, there has been an increase in the average price of properties in China, as denoted by the rolling mean (red line) on the figure. However, there are some instances where the prices fall, with the year 2020 and 2022. The rolling standard deviation (black line) can also be used to identify the trends and patterns of the data, in which the "bumps" indicate the periods of volatility.

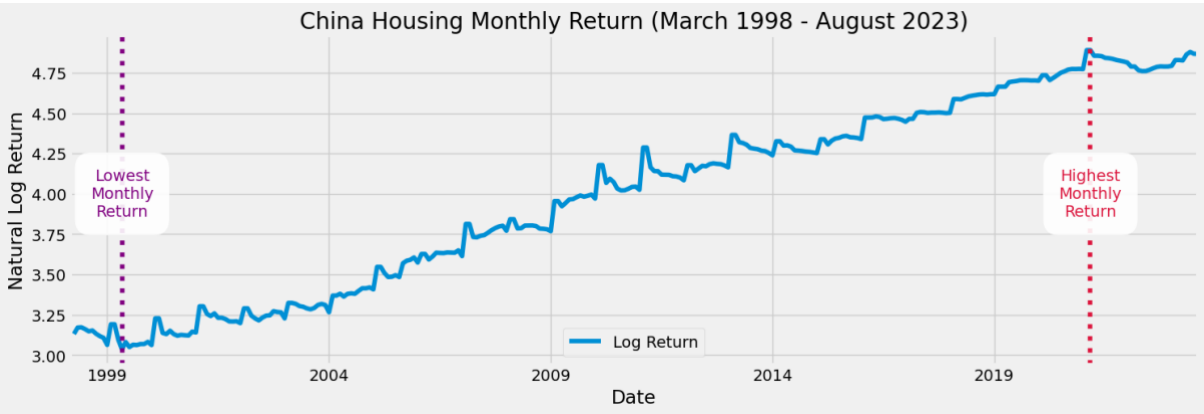


Figure 5. *China Housing Return*

The figure above shows the monthly return in the price of houses. While there are circumstances in which the returns are decreasing, the general trend indicates that housing

returns are increasing, with the monthly return averages between 3.06% (May 1999) to 4.89% (March 2021).

```
======= Augmented Dickey-Fuller Test Results =======

1. ADF Test Statistic: 1.004735
2. P-value: 0.994316
3. Used Lags: 13
4. Used Observations: 292
5. Critical Values:
        1%: -3.452945
        5%: -2.871490
        10%: -2.572071
```

Figure 6. *Augmented Dickey-Fuller Test Results*

However, as seen from the Augmented Dickey-Fuller Test above, the p-value is 0.994316, which is significantly larger than the significance level of 0.05, and therefore the series fails to reject the null hypothesis (H_0), the data has a unit root and is non-stationary. Therefore, the series will be transformed by taking the first-order difference.

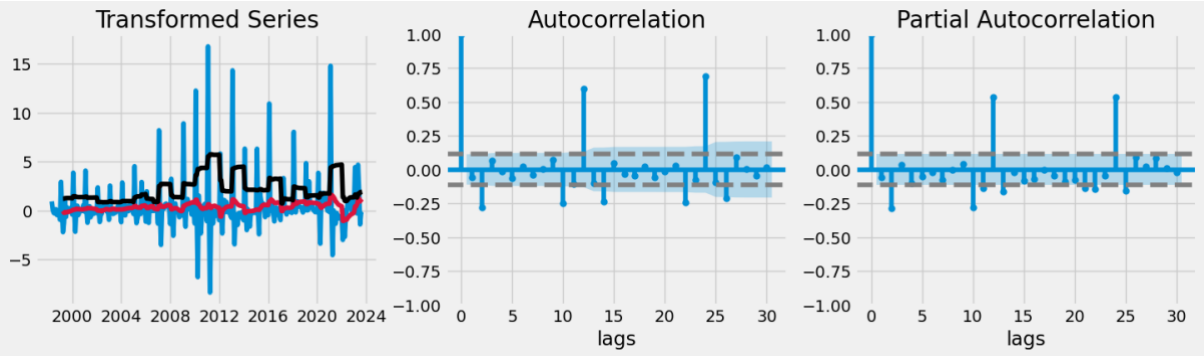


Figure 7. *First-Order Differencing of Housing Price Index*

The result above shows less trend as compared to the original series in Figure 4. The p-value is found to be 0.00077, which is significantly lower than 0.05. Hence, series transformation will stop at this stage as the data is likely to be stationary. Further transformations will not show a strong trend and the series should not be over-differenced. Implementing another differencing order will increase the standard deviation, which can result in additional complexity.

3.3. Correlation between Housing Price Index and News Sentiment

Figure 8. *Housing Price Index and Sentiment Score, October 2021 to October 2023*

The figure above is the plotted result that shows both the HPI and sentiment score for the period of October 2021 to October 2023. An empirical observation on the figure proves that sentiment index does affect the housing prices, suggesting that there is a correlation between the two variables. The correlation is calculated to be 0.13072, suggesting that there is a weak positive relationship between them.

```
corr = si_mean_monthly["SentimentScore"].corr(si_mean_monthly["PropertyIndex"])
print("Correlation between Sentiment Score and Property Index is: ",␣
↪round(corr, 5))
```

```
Correlation between Sentiment Score and Property Index is:  0.13072
```

Figure 9. *Calculated Correlation Value*

4. Discussions

4.1. Sentiment Analysis

Based on the figure produced, it is evident that the news sentiment has remained largely positive in the past years, with few occasions in which it reaches extreme negative. This can

be seen in July 2022 and October 2022, which may be caused by the real estate developers postponing their debt-restructuring measures in anticipation that the Communist Party Congress would be able to stabilize the property sector, creating uncertainty and lack of confidence ("Analysis-Chinese property developers on tenterhooks ahead of Communist Party Congress", 2022). The first half of 2023 also shows highly positive sentiment, which may be induced by the Chinese government introducing expansionary demand-side policy such as tax cuts, stimulus packages and credit easing to the public when buying properties. Property developers have also gain higher business confidence, denoted by their willingness to set goals to revamp their debt and having the ability to repay their debt. However, the Q3 of 2023 sees a more significant fluctuations in the sentiment score which ranges from strong positive to strong negative. Some of the positive indices are evident to happen when the government introduces policy measures to make housing more affordable, such as easing mortgage rules ("Chinese cities ease mortgage rules in bid to revive property sector", 2023), while negative indices are induced by poor economic performance such as the inability to purchase houses due to unemployment and solely relying on transfer payment for daily necessities (Chia & Du, 2023).

4.2. Housing Price Index

From Figure 4, the housing price index increases year on year. This may be caused by the demand-side regulations from the government that are not balanced by the supply factor (Sheng, 2021). Sheng explains that the Chinese government viewed houses as necessity and not investment tool. In addition, despite not having land shortages, there had been increase in the demand for housing due to the fear of future regulatory changes that may restrict them from buying houses. Property tax was also proven to have little to no effect on the buyers due to the challenges during its implementation in Shanghai, putting upward pressure on the housing prices. However, the index decreased in some years, with the most significant one being in the year of 2022. This slump first began due to the government's contractionary policy in which it

makes the financing for property developers much more difficult (Huang, 2023). This was done with the aim to manage the risk caused by high level of debt by the real estate developers. However, the issue spiraled as developers had been running a business model that relies on high project turnover which was heavily influenced by borrowings from their financial channels, including banks. To counter this, the government introduced "three red lines" policy, a financial regulatory that sets limit on the ratio of debt to cash, equity, and assets ("China may ease 'three red lines' property rules – Bloomberg News", 2023). China's zero-COVID policy effect had also worsened the situation as the nationwide lockdowns contracts the demand for properties and default their payments. These combined effects led to surplus in the inventory of unsold houses which forced the developers to halt the construction projects. Buyers who had started to make their payments found themselves to be disadvantaged and refused to make further transactions, leading to subsequent fall in the housing prices.

4.3. Correlation between Housing Price Index and News Sentiment

Figure 8 shows the relationship between the housing price index and sentiment score, from which we can infer that sentiment index does affect the housing prices. When sentiment score is becoming more positive (green area), house prices increases and when sentiment score is becoming more negative (red area), house prices decrease. Evidently, from January 2022 to July 2023, the sentiment score has always been positive, in which housing price index also sees and increase from 117 to 132. While from July 2023 to October 2023, sentiment score is less positive and therefore housing prices fall from 132 to 130. Though the correlation is 0.13, which indicates weak relationship between sentiment and house index, overall, there is a positive correlation between them.

5. Limitations and recommendation for future study

Although the study tried to execute the empirical study precisely, there are still some limitations during the process, but may serve as future reference to improve the reliability and accuracy. Firstly, the study only leverages on 105 datasets from one news publisher. Therefore, further study may scrap more articles which dates to a few years ago, and from several sources to bridge the gap such that there is no sudden spike in the sentiment score which may alter the perception of the reader. Secondly, during the data cleaning process, there are some "impurities" that may still persist, such as strange chain of strings, which is usually due to different decoding of punctuation mark such as apostrophe (‘). Hence, such impurities may be further processed as they may affect the sentiment score. Thirdly, while the study tried to use several keywords "China+housing+prices", there may be other related keywords that are not identified. Hence, future study may adopt a wider choice of keywords to account for a more precise articles to be scrapped. Lastly, there may be more than one factors which may influence the Housing Price Index in real life. Hence, to increase the accuracy, multiple linear regression may be implemented in which additional independent variables (such as disposable income, income inequality, education, and population size) that may affect the housing prices. Hence, each variable can be assigned a weight to indicate the relative significance in determining the housing prices. If the data can be updated, future researchers will be able to improve on the current study to create a more reliable and comprehensive result.

6. References

Analysis-Chinese property developers on tenterhooks ahead of Communist Party Congress. (2022, October 14). *CNA*. http://bit.ly/3rLCVrV

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... & Belyaeva, J. (2013). Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.

China Real Residential Property Price Index (2023, October 3). *CEIC Data*. https://www.ceicdata.com/en/indicator/china/real-residential-property-price-index

Channel News Asia (CNA) – Bias and Credibility (2023, May). In *Media Bias Fact Check*. https://mediabiasfactcheck.com/channel-news-asia-cna/

Chia, L. & Du, W. (2023, September 9). She has a master's but no job and lives on discount coupons. In China, there are many like her. *CNA*. https://bit.ly/3SeJ2zV

China may ease 'three red lines' property rules – Bloomberg News. (2023, January 6). *Reuters*. https://bit.ly/3PVpzkD

Chinese cities ease mortgage rules in bid to revive property sector. (2023, August 30). *CNA*. https://bit.ly/3RYeTVk

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82-89.

Huang, T. (2023). Why China's housing policies have failed. *Peterson Institute for International Economics Working Paper*, (23-5).

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. The Cambridge University Press.

Mutz, D. C., & Soss, J. (1997). Reading public opinion: The influence of news coverage on perceptions of public sentiment. *Public Opinion Quarterly*, 431-451.

Natarajan, K., & Xiaoming, H. (2003). An Asian voice? A comparative study of Channel News Asia and CNN. *Journal of Communication*, *53*(2), 300-314.

Ozgur, C., Colliau, T., Rogers, G., & Hughes, Z. (2017). MatLab vs. Python vs. R. *Journal of data Science*, *15*(3), 355-371.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, 2(1-2):1–135.

Sheng, S. C. (2021). Why Chinese house prices keep going up and up. *China Europe International Business School*. https://bit.ly/45pcV37

Soroka, S. N., Stecula, D. A., & Wlezien, C. (2015). It's (change in) the (future) economy, stupid: economic indicators, the media, and public opinion. *American Journal of Political Science*, *59*(2), 457-474.

Sutherland, M. D. (2023). *China's economy: current trends and issues.* Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11667.

Appendix

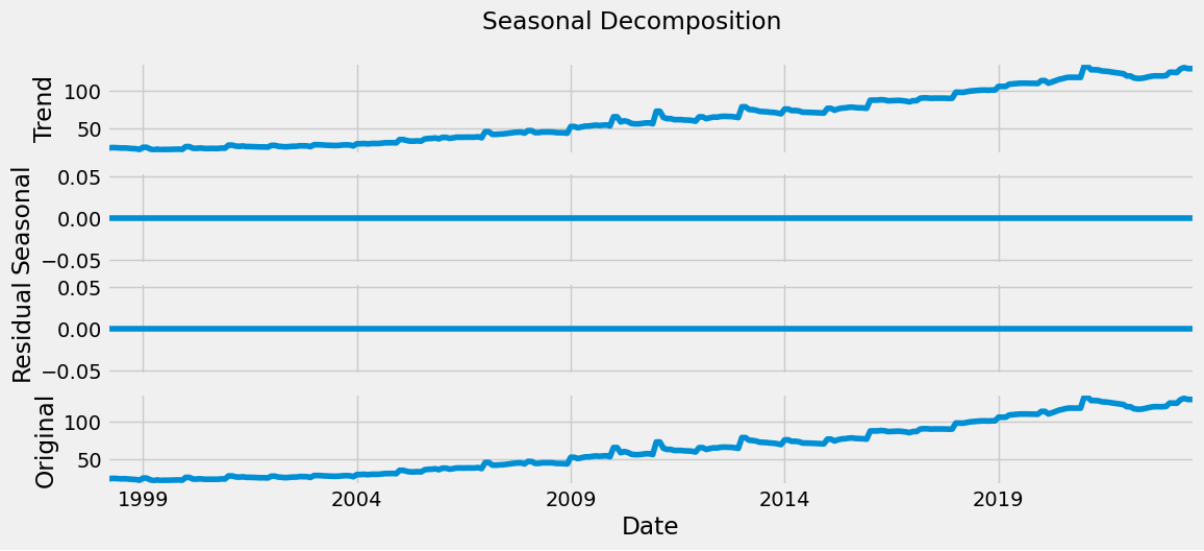


Figure 10. *Residual Plot of Housing Price Index*

SpecialTopic_Combined

October 13, 2023

1 Sentiment Analysis

[1]:
```
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
!pip install htmldate
from htmldate import find_date
import urllib.request
import re
from urllib.request import urlopen
from urllib.error import HTTPError
import csv
```

```
Requirement already satisfied: htmldate in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (1.5.1)
Requirement already satisfied: dateparser>=1.1.2 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from htmldate)
(1.1.8)
Requirement already satisfied: python-dateutil>=2.8.2 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from htmldate)
(2.8.2)
Requirement already satisfied: lxml==4.9.2 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from htmldate)
(4.9.2)
Requirement already satisfied: charset-normalizer>=3.2.0 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from htmldate)
(3.2.0)
Requirement already satisfied: urllib3<3,>=1.26 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from htmldate)
(1.26.12)
Requirement already satisfied: pytz in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from
dateparser>=1.1.2->htmldate) (2022.6)
Requirement already satisfied: regex!=2019.02.19,!=2021.8.27 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from
dateparser>=1.1.2->htmldate) (2022.10.31)
Requirement already satisfied: tzlocal in
```

```
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from
dateparser>=1.1.2->htmldate) (5.0.1)
Requirement already satisfied: six>=1.5 in
/Users/richardreynard/miniconda3/lib/python3.9/site-packages (from python-
dateutil>=2.8.2->htmldate) (1.16.0)
```

[2]:
```
article_data = pd.read_csv('file.csv')
article_data.head()
```

[2]:

	Unnamed: 0	link
0	Article 1	https://www.channelnewsasia.com/business/count…
1	Article 2	https://www.channelnewsasia.com/commentary/chi…
2	Article 3	https://www.channelnewsasia.com/asia/china-eco…
3	Article 4	https://www.channelnewsasia.com/cna-insider/ch…
4	Article 5	https://www.channelnewsasia.com/commentary/chi…

[3]:
```
news_links = list(article_data["link"])
```

[4]:
```
count = 0
articles_array = []
for link in news_links:
    try:
        # Get article link
        url = news_links[count]
        webUrl = urllib.request.urlopen(url)
        data = webUrl.read()

        # Get title of the article
        page = urlopen(url)
        html = page.read().decode("utf-8")
        pattern = "<title.*?>.*?</title.*?>"
        match_results = re.search(pattern, html, re.IGNORECASE)
        title = match_results.group()
        title = re.sub("<.*?>", "", title) # Remove HTML tags

        # Get the article's content
        page = urlopen(url)
        html = page.read().decode("utf-8")
        soup = BeautifulSoup(html, "html.parser")
        content = soup.get_text().replace('\n', '')
        text = ' '.join(content.split())

        # Get article's published date
        date = find_date(data)

        # Combine all details of an article
        article = {}
```

```
        article['Title'] = title
        article['Date'] = date
        article['Link'] = url
        article['Text'] = text
        articles_array.append(article)
    except HTTPError:
        pass

    count = count + 1
```

[5]:
```
# Save article as CSV-file
try:
    f = csv.writer(open('cna_dataset.csv', 'w', encoding='utf-8'))
    f.writerow(['Title','PublishedDate', 'Link', 'Text'])
    for article_details in articles_array:
        title = article_details['Title']
        date = article_details['Date']
        link = article_details['Link']
        text = article_details['Text']
        f.writerow([title, date, link, text])
except Exception as e: print(e)
```

[6]:
```
import numpy as np
import pandas as pd
import nltk
from nltk.tokenize import sent_tokenize
from nltk.corpus import stopwords
from nltk.cluster.util import cosine_distance
import networkx as nx
import matplotlib.pyplot as plt
from PIL import Image
from wordcloud import WordCloud
import spacy
import datetime
```

[7]:
```
df = pd.read_csv('/Users/richardreynard/Downloads/SC3901/ChinaHousingMarket/
↪cna_dataset.csv')
df.head()
```

[7]:

	Title	PublishedDate	Link	\
0	Country Garden, Sunac debt deals bring respite…	2023-09-19		
1	Commentary: Why China's real estate crisis sho…	2023-09-16		
2	China's economic data shows signs of life amid…	2023-09-15		
3	She has a master's but no job and lives on dis…	2023-09-09		
4	Commentary: Is China finally getting serious a…	2023-09-08		

```
0  https://www.channelnewsasia.com/business/count…
1  https://www.channelnewsasia.com/commentary/chi…
2  https://www.channelnewsasia.com/asia/china-eco…
3  https://www.channelnewsasia.com/cna-insider/ch…
4  https://www.channelnewsasia.com/commentary/chi…

                                                Text
0  Country Garden, Sunac debt deals bring respite…
1  Commentary: Why China's real estate crisis sho…
2  China's economic data shows signs of life amid…
3  She has a master's but no job and lives on dis…
4  Commentary: Is China finally getting serious a…
```

[8]:
```
print(f"Total number of rows: {len(df)}")
```

```
Total number of rows: 105
```

[9]:
```
advertisement_rows = sum(len(str(row['Text']).split()) < 200 for _, row in df.
 ↪iterrows())
print(f"Number of rows that might be an advertisement: {advertisement_rows}")
```

```
Number of rows that might be an advertisement: 0
```

[10]:
```
import csv

# specify the column to check and update
column_to_check = 'Text'

# open the input CSV file
with open('/Users/richardreynard/Downloads/SC3901/ChinaHousingMarket/
 ↪cna_dataset.csv', 'r') as input_file:
    # create a CSV reader object
    reader = csv.DictReader(input_file)

    # open the output CSV file
    with open('new_cna_dataset.csv', 'w', newline='') as output_file:
        # create a CSV writer object
        writer = csv.DictWriter(output_file, fieldnames=reader.fieldnames)
        writer.writeheader()

        # iterate over the rows in the input CSV
        for row in reader:
            # check the length of the column
            if len(row[column_to_check].split()) < 200:
                continue

            # read the specified column
```

```
            text = row[column_to_check]

            # remove extra spaces between words and newlines
            text = ' '.join(text.split())
            text = text.replace('\n', ' ')

            # update the row with the cleaned text
            row[column_to_check] = text

            # remove duplicated rows
            df = df[df.shift() != df].dropna()

            # write the updated row to the output CSV
            writer.writerow(row)
```

[11]:
```
df = pd.read_csv('/Users/richardreynard/Downloads/SC3901/ChinaHousingMarket/
 ↪new_cna_dataset.csv')
df.head()
```

[11]:

	Title	PublishedDate	Link	Text
0	Country Garden, Sunac debt deals bring respite…	2023-09-19	https://www.channelnewsasia.com/business/count…	Country Garden, Sunac debt deals bring respite…
1	Commentary: Why China's real estate crisis sho…	2023-09-16	https://www.channelnewsasia.com/commentary/chi…	Commentary: Why China's real estate crisis sho…
2	China's economic data shows signs of life amid…	2023-09-15	https://www.channelnewsasia.com/asia/china-eco…	China's economic data shows signs of life amid…
3	She has a master's but no job and lives on dis…	2023-09-09	https://www.channelnewsasia.com/cna-insider/ch…	She has a master's but no job and lives on dis…
4	Commentary: Is China finally getting serious a…	2023-09-08	https://www.channelnewsasia.com/commentary/chi…	Commentary: Is China finally getting serious a…

[12]:
```
df['Text'] = df['Text'].apply(lambda txt: txt.lower())
stop_words=stopwords.words('english')
df['Text'] = df['Text'].apply(lambda txt: ' '.join([word for word in txt.
 ↪split() if word not in stop_words]))
df['Text'] = df['Text'].apply(lambda txt: sent_tokenize(txt))
df['Text'] = df['Text'].apply(lambda txt: ' '.join(txt))
```

[13]:
```
df.head()
```

[13]:

	Title	PublishedDate	Link	Text
0	Country Garden, Sunac debt deals bring respite…	2023-09-19	https://www.channelnewsasia.com/business/count…	country garden, sunac debt deals bring respite…
1	Commentary: Why China's real estate crisis sho…	2023-09-16	https://www.channelnewsasia.com/commentary/chi…	commentary: china's real estate crisis make gl…
2	China's economic data shows signs of life amid…	2023-09-15	https://www.channelnewsasia.com/asia/china-eco…	china's economic data shows signs life amid pr…
3	She has a master's but no job and lives on dis…	2023-09-09	https://www.channelnewsasia.com/cna-insider/ch…	master's job lives discount coupons. china, ma…
4	Commentary: Is China finally getting serious a…	2023-09-08	https://www.channelnewsasia.com/commentary/chi…	commentary: china finally getting serious huko…

[14]:
```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
sid = SentimentIntensityAnalyzer()
```

[15]:
```
df['score'] = df['Text'].apply(lambda txt: sid.polarity_scores(txt))
```

[16]:
```
df['negative'] = df['score'].apply(lambda txt: txt['neg'])
df['neutral'] = df['score'].apply(lambda txt: txt['neu'])
df['positive'] = df['score'].apply(lambda txt: txt['pos'])
df['compound'] = df['score'].apply(lambda txt: txt['compound'])
```

[17]:
```
def polarity_score(compound):
    if compound > 0.05:
        return "positive"
    elif compound < -0.05:
        return "negative"
    elif compound >= -0.05 and compound < 0.05:
        return "neutral"
```

[18]:
```
df['sentiment'] = df['compound'].apply(lambda val: polarity_score(val))
df
```

[18]:

	Title	PublishedDate
0	Country Garden, Sunac debt deals bring respite…	2023-09-19
1	Commentary: Why China's real estate crisis sho…	2023-09-16

```
2    China’s economic data shows signs of life amid…    2023-09-15
3    She has a master’s but no job and lives on dis…    2023-09-09
4    Commentary: Is China finally getting serious a…    2023-09-08
..                                                 …           …
100  Commentary: China needs bolder moves for growt…    2021-10-27
101  Commentary: Evergrande woes show China’s overr…    2021-09-30
102  Commentary: China’s leaders risk a dangerous m…    2021-09-29
103  China asking state-backed firms to pick up Eve…    2021-09-28
104  Evergrande pain spreads to wealthy investors a…    2021-09-27

                                                  Link  \
0    https://www.channelnewsasia.com/business/count…
1    https://www.channelnewsasia.com/commentary/chi…
2    https://www.channelnewsasia.com/asia/china-eco…
3    https://www.channelnewsasia.com/cna-insider/ch…
4    https://www.channelnewsasia.com/commentary/chi…
..                                                 …
100  https://www.channelnewsasia.com/commentary/chi…
101  https://www.channelnewsasia.com/commentary/eve…
102  https://www.channelnewsasia.com/commentary/eve…
103  https://www.channelnewsasia.com/business/china…
104  https://www.channelnewsasia.com/business/everg…

                                                  Text  \
0    country garden, sunac debt deals bring respite…
1    commentary: china’s real estate crisis make gl…
2    china’s economic data shows signs life amid pr…
3    master’s job lives discount coupons. china, ma…
4    commentary: china finally getting serious huko…
..                                                 …
100  commentary: china needs bolder moves growth en…
101  commentary: evergrande woes show china’s overr…
102  commentary: china’s leaders risk dangerous mis…
103  china asking state-backed firms pick evergrand…
104  evergrande pain spreads wealthy investors inte…

                                                 score  negative  neutral  \
0    {'neg': 0.068, 'neu': 0.809, 'pos': 0.123, 'co…     0.068    0.809
1    {'neg': 0.109, 'neu': 0.805, 'pos': 0.087, 'co…     0.109    0.805
2    {'neg': 0.088, 'neu': 0.806, 'pos': 0.106, 'co…     0.088    0.806
3    {'neg': 0.071, 'neu': 0.834, 'pos': 0.095, 'co…     0.071    0.834
4    {'neg': 0.059, 'neu': 0.829, 'pos': 0.112, 'co…     0.059    0.829
..                                                 …       …       …
100  {'neg': 0.14, 'neu': 0.723, 'pos': 0.138, 'com…     0.140    0.723
101  {'neg': 0.082, 'neu': 0.798, 'pos': 0.121, 'co…     0.082    0.798
102  {'neg': 0.125, 'neu': 0.765, 'pos': 0.11, 'com…     0.125    0.765
103  {'neg': 0.046, 'neu': 0.852, 'pos': 0.103, 'co…     0.046    0.852
```

```
104  {'neg': 0.087, 'neu': 0.765, 'pos': 0.148, 'co…     0.087    0.765

     positive  compound sentiment
0       0.123    0.9917  positive
1       0.087   -0.9653  negative
2       0.106    0.9718  positive
3       0.095    0.9910  positive
4       0.112    0.9956  positive
..        …        …        …
100     0.138   -0.9780  negative
101     0.121    0.9867  positive
102     0.110   -0.9738  negative
103     0.103    0.9907  positive
104     0.148    0.9981  positive

[105 rows x 10 columns]
```

[19]:
```
df['PublishedDate'] = pd.to_datetime(df['PublishedDate'], format = '%Y-%m-%d')
```

[20]:
```
df.sort_values(by='PublishedDate', inplace = True)
df
```

[20]:
```
                                                 Title PublishedDate  \
104  Evergrande pain spreads to wealthy investors a…    2021-09-27
103  China asking state-backed firms to pick up Eve…    2021-09-28
102  Commentary: China's leaders risk a dangerous m…    2021-09-29
101  Commentary: Evergrande woes show China's overr…    2021-09-30
100  Commentary: China needs bolder moves for growt…    2021-10-27
..                                                 …           …
62   Will China's property headaches have broader e…    2023-10-05
61   Analysis:Evergrande crisis tests Beijing&#039;…    2023-10-05
60   Wall St brokerages raise China&#039;s economic…    2023-10-05
59   Chinese developer Country Garden faces fresh o…    2023-10-09
58   Country Garden says can&#039;t meet all offsho…    2023-10-10

                                                  Link  \
104  https://www.channelnewsasia.com/business/everg…
103  https://www.channelnewsasia.com/business/china…
102  https://www.channelnewsasia.com/commentary/eve…
101  https://www.channelnewsasia.com/commentary/eve…
100  https://www.channelnewsasia.com/commentary/chi…
..                                                 …
62   https://www.channelnewsasia.com/asia/china-pro…
61   https://www.channelnewsasia.com/business/analy…
60   https://www.channelnewsasia.com/business/wall-…
59   https://www.channelnewsasia.com/business/china…
58   https://www.channelnewsasia.com/business/china…
```

```
                                                  Text  \
104  evergrande pain spreads wealthy investors inte…
103  china asking state-backed firms pick evergrand…
102  commentary: china's leaders risk dangerous mis…
101  commentary: evergrande woes show china's overr…
100  commentary: china needs bolder moves growth en…
..                                                 …
62   china's property headaches broader economic ef…
61   analysis:evergrande crisis tests beijing's fal…
60   wall st brokerages raise china's economic grow…
59   chinese developer country garden faces fresh o…
58   country garden says can't meet offshore debt p…

                                                 score  negative  neutral  \
104  {'neg': 0.087, 'neu': 0.765, 'pos': 0.148, 'co…     0.087    0.765
103  {'neg': 0.046, 'neu': 0.852, 'pos': 0.103, 'co…     0.046    0.852
102  {'neg': 0.125, 'neu': 0.765, 'pos': 0.11, 'com…     0.125    0.765
101  {'neg': 0.082, 'neu': 0.798, 'pos': 0.121, 'co…     0.082    0.798
100  {'neg': 0.14, 'neu': 0.723, 'pos': 0.138, 'com…     0.140    0.723
..                                                 …       …       …
62   {'neg': 0.071, 'neu': 0.831, 'pos': 0.098, 'co…     0.071    0.831
61   {'neg': 0.11, 'neu': 0.796, 'pos': 0.093, 'com…     0.110    0.796
60   {'neg': 0.059, 'neu': 0.8, 'pos': 0.141, 'comp…     0.059    0.800
59   {'neg': 0.068, 'neu': 0.819, 'pos': 0.113, 'co…     0.068    0.819
58   {'neg': 0.063, 'neu': 0.796, 'pos': 0.141, 'co…     0.063    0.796

     positive  compound sentiment
104     0.148    0.9981  positive
103     0.103    0.9907  positive
102     0.110   -0.9738  negative
101     0.121    0.9867  positive
100     0.138   -0.9780  negative
..        …       …        …
62      0.098    0.9707  positive
61      0.093   -0.9677  negative
60      0.141    0.9918  positive
59      0.113    0.9841  positive
58      0.141    0.9970  positive

[105 rows x 10 columns]
```

[21]:
```
sentiment_index = df[['PublishedDate', 'compound']]
sentiment_index
```

[21]:
```
    PublishedDate  compound
104    2021-09-27    0.9981
```

```
103    2021-09-28    0.9907
102    2021-09-29   -0.9738
101    2021-09-30    0.9867
100    2021-10-27   -0.9780
..            …       …
62     2023-10-05    0.9707
61     2023-10-05   -0.9677
60     2023-10-05    0.9918
59     2023-10-09    0.9841
58     2023-10-10    0.9970

[105 rows x 2 columns]
```

[22]:
```
si_group = sentiment_index.groupby('PublishedDate')['PublishedDate'].count()
print(si_group)
```

```
PublishedDate
2021-09-27    1
2021-09-28    1
2021-09-29    1
2021-09-30    1
2021-10-27    1
             ..
2023-10-03    1
2023-10-04    1
2023-10-05    3
2023-10-09    1
2023-10-10    1
Name: PublishedDate, Length: 81, dtype: int64
```

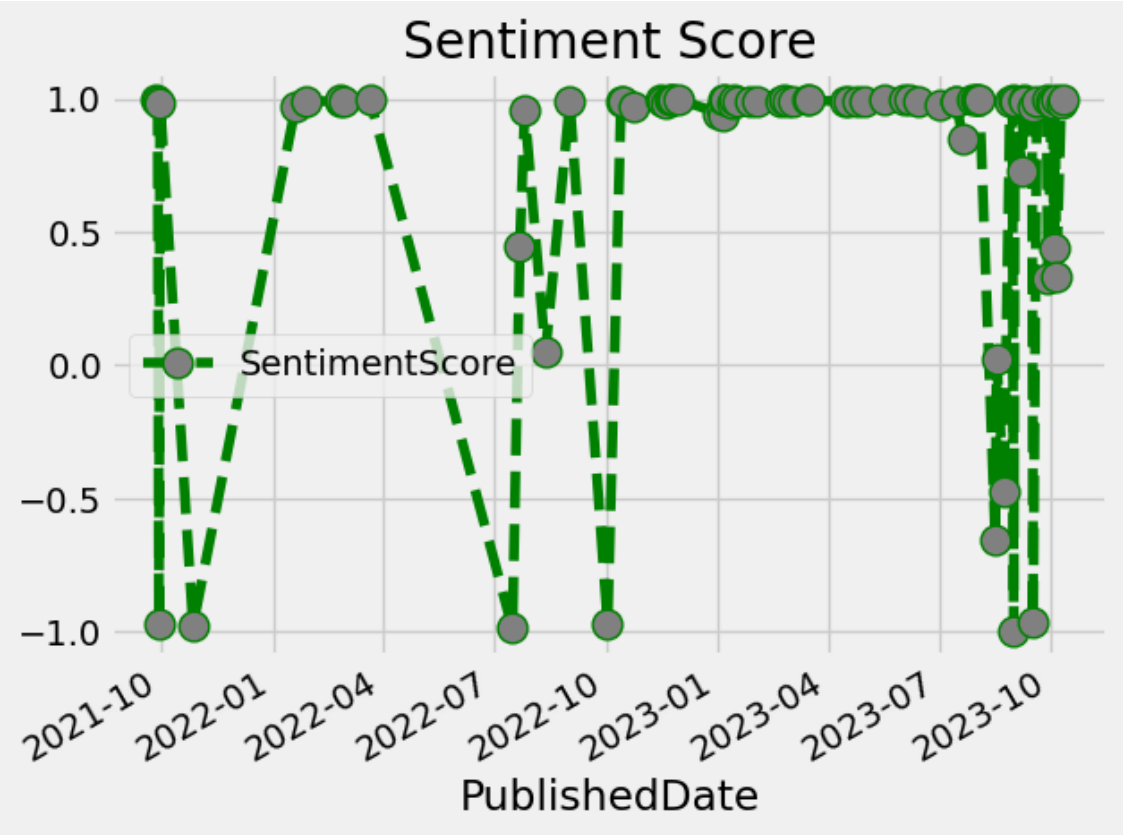
[23]:
```
si_mean = sentiment_index.groupby('PublishedDate').compound.agg(['mean']).
  ↪reset_index()
si_mean = si_mean.rename(columns={'mean':'SentimentScore'})
si_mean
```

[23]:
```
   PublishedDate  SentimentScore
0     2021-09-27          0.9981
1     2021-09-28          0.9907
2     2021-09-29         -0.9738
3     2021-09-30          0.9867
4     2021-10-27         -0.9780
..           …              …
76    2023-10-03          0.4381
77    2023-10-04          0.9946
78    2023-10-05          0.3316
79    2023-10-09          0.9841
80    2023-10-10          0.9970
```

[81 rows x 2 columns]

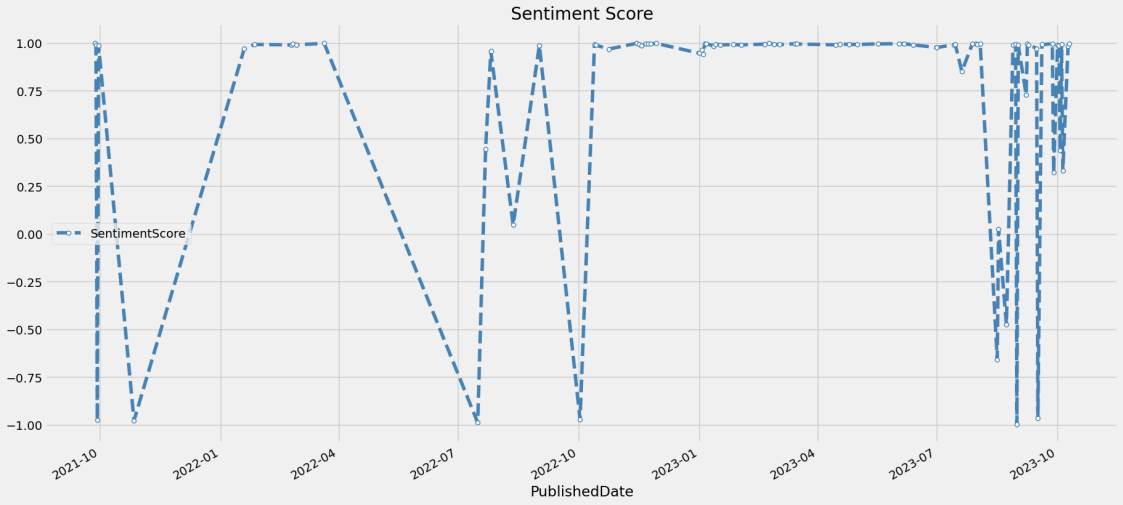
[24]:
```
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

plt.figure(figsize=(16, 4))
si_mean.plot(x='PublishedDate', y='SentimentScore', color='green', linestyle =␣
 ↪'dashed', title='Sentiment Score', marker='o', markerfacecolor='gray',␣
 ↪markersize=12)
plt.show()
```

<Figure size 1600x400 with 0 Axes>



[25]:
```
fig, ax = plt.subplots(figsize=(20,10))
si_mean.plot(x='PublishedDate', y='SentimentScore', color='steelblue',␣
 ↪linestyle = 'dashed', title='Sentiment Score', marker='o',␣
 ↪markerfacecolor='white', markersize=5, ax=ax)
plt.legend(loc = 'best', frameon= True)
```

[25]: <matplotlib.legend.Legend at 0x7fd56c2fa310>



---

2 Housing Price Index

[26]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
import warnings
warnings.filterwarnings('ignore')
```

[27]:
```
# import data
def load_data(file_name):
    # load from XLSX file
    dataset = pd.read_excel(io=file_name)

    # unpivot from wide to long format
    dataset = dataset.melt(id_vars='Year', var_name='Month', value_name='Rate')

    from pandas.tseries.offsets import MonthEnd

    # assign last day of month
    dataset['Date'] = pd.to_datetime(dataset[['Year', 'Month']].assign(DAY=1))␣
 ↪+ MonthEnd(1)

    # order ascending data values
    dataset = dataset.sort_values(by='Date', ascending=True)

    # drop unnecessary columns
```

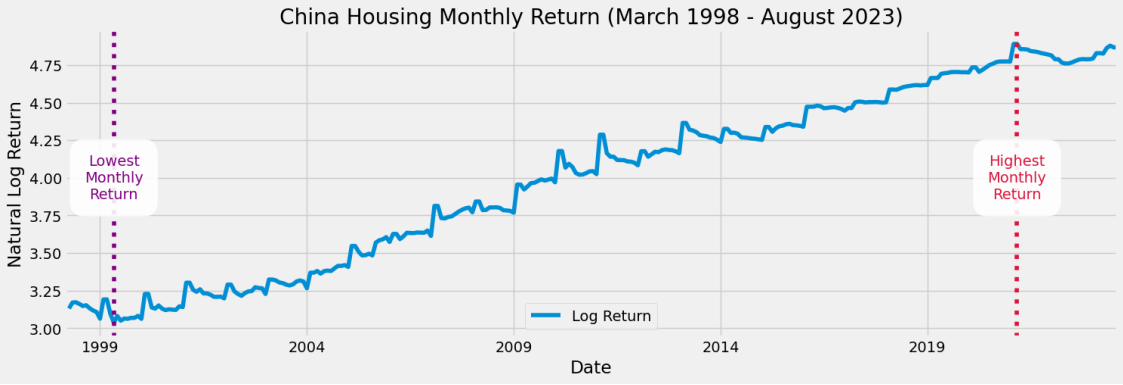
```
    dataset = dataset.drop(['Year', 'Month'], axis=1)

    # set date column as index
    dataset.set_index('Date', inplace=True)

    # drop NaN rows
    dataset.dropna(subset=['Rate'], inplace=True)

    return dataset
```

3 Find the return in HPI

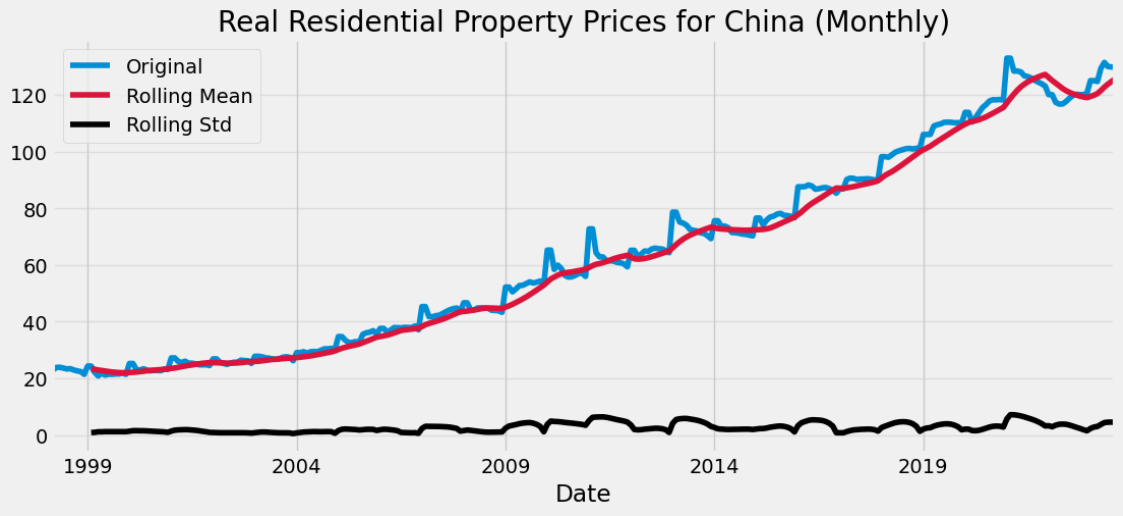
[28]:
```
rpi_data = load_data(file_name='ChinaResidentialPropertyIndex_CEIC.xlsx')
```

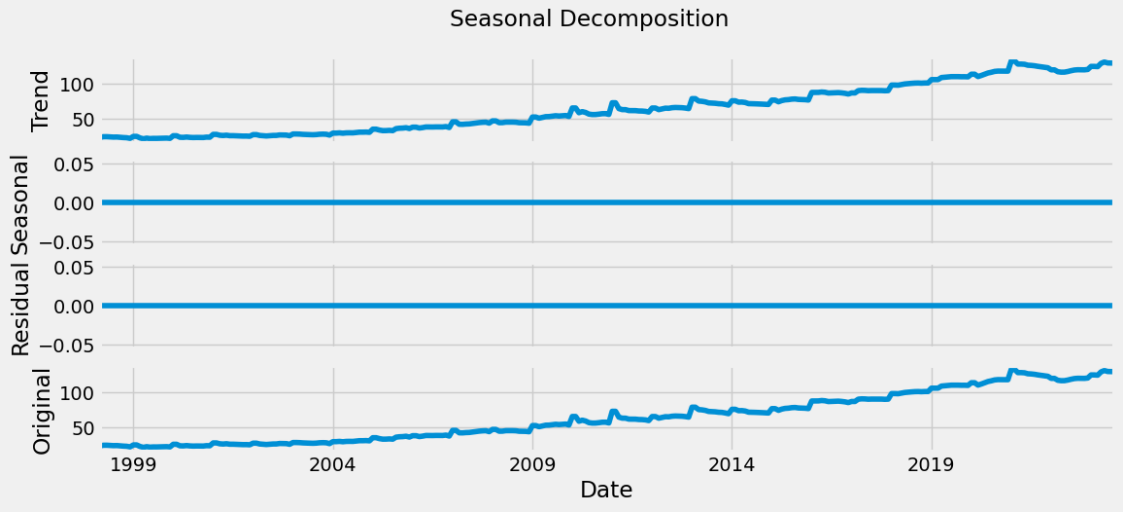
[29]:
```
rpi_data['Log Return'] = np.log(rpi_data['Rate'].shift(1))
```

[30]:
```
rpi_data['Log Return'].plot(figsize=(16,5))
plt.ylabel("Natural Log Return")
plt.xlabel("Date")
plt.title("China Housing Monthly Return (March 1998 - August 2023)")
plt.legend(loc="best")

event_list = [(pd.to_datetime('1999-05-31'), 'Lowest\nMonthly\nReturn',␣
 ↪'purple'),
              (pd.to_datetime('2021-03-31'), 'Highest\nMonthly\nReturn',␣
 ↪'crimson')]
for date_point, label, clr in event_list:
        plt.axvline(x=date_point, color=clr, linestyle=':')
        plt.text(x=date_point, y=4, s=label, horizontalalignment='center',␣
 ↪verticalalignment='center',
                 color=clr, bbox=dict(facecolor='white', alpha=0.9,␣
 ↪boxstyle='round, pad=1', linewidth=0.2))

plt.show()
```



[31]:
```
# visualize target data
def plot_time_series(series):
    mean_rolling = series.rolling(window=12).mean()
    std_rolling = series.rolling(window=12).std()

    # plot inflation rates
    series.plot(figsize=(12, 5), label='Original')
    mean_rolling.plot(color='crimson', label='Rolling Mean')
    std_rolling.plot(color='black', label='Rolling Std')
    plt.title('Real Residential Property Prices for China (Monthly)')
    plt.grid(axis='y', alpha=0.5)
    plt.legend(loc='best')
    plt.show()

    from statsmodels.tsa.seasonal import seasonal_decompose

    # plot decomposition components
    decomp = seasonal_decompose(series, model='additive', period = 1)
    fig, axes = plt.subplots(ncols=1, nrows=4, sharex=True, figsize=(12, 5))
    fig.suptitle('Seasonal Decomposition')

    decomp.trend.plot(ax=axes[0], legend=False)
    axes[0].set_ylabel('Trend')

    decomp.seasonal.plot(ax=axes[1], legend=False)
    axes[1].set_ylabel('Seasonal')

    decomp.resid.plot(ax=axes[2], legend=False)
    axes[2].set_ylabel('Residual')

    decomp.observed.plot(ax=axes[3], legend=False)
    axes[3].set_ylabel('Original')
    plt.show()

rpi_data = load_data(file_name='ChinaResidentialPropertyIndex_CEIC.xlsx')
plot_time_series(rpi_data['Rate'])
```





[32]:
```
# ADF statistical test
def adf_test(series):
    from statsmodels.tsa.stattools import adfuller

    result = adfuller(series, regression='c', autolag='AIC')
    print('======= Augmented Dickey-Fuller Test Results =======\n')
    print('1. ADF Test Statistic: {:.6f}'.format(result[0]))
    print('2. P-value: {:.6f}'.format(result[1]))
    print('3. Used Lags: {}'.format(result[2]))
    print('4. Used Observations: {}'.format(result[3]))
    print('5. Critical Values:')
    for key, value in result[4].items():
```

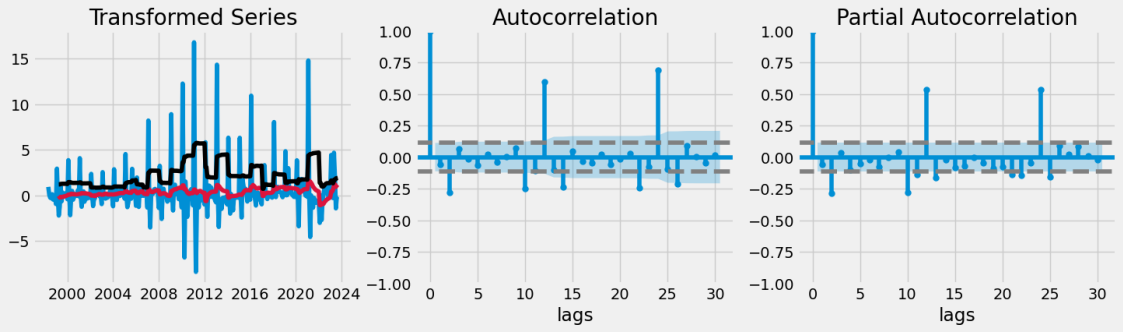
```
        print('\t{}: {:.6f}'.format(key, value))

    critical_value = result[4]['5%']
    if (result[1] <= 0.05) and (result[0] < critical_value):
        print('\nStrong evidence against the null hypothesis (H0), reject the␣
 ↪null hypothesis.\
        Data has no unit root and is stationary.')
    else:
        print('\nWeak evidence against null hypothesis, time series has a unit␣
 ↪root, indicating it is non-stationary.')
    return

# run function
adf_test(rpi_data['Rate'])
```

```
======= Augmented Dickey-Fuller Test Results =======

1. ADF Test Statistic: 1.004735
2. P-value: 0.994316
3. Used Lags: 13
4. Used Observations: 292
5. Critical Values:
        1%: -3.452945
        5%: -2.871490
        10%: -2.572071

Weak evidence against null hypothesis, time series has a unit root, indicating
it is non-stationary.
```

[33]:
```
# perform data transformation
# series: must be a pandas dataframe
def series_transformation(series):
    from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
    from statsmodels.tsa.stattools import adfuller

    # 1st plot - data after transformation
    fig = plt.figure(figsize=(16, 4))
    ax1 = fig.add_subplot(1, 3, 1)
    ax1.set_title('Transformed Series')
    ax1.plot(series)
    ax1.plot(series.rolling(window=12).mean(), color='crimson')
    ax1.plot(series.rolling(window=12).std(), color='black')

    # 2nd plot - partial autocorrelation plot
    ax2 = fig.add_subplot(1, 3, 2)
    plot_acf(series.dropna(), ax=ax2, lags=30, title='Autocorrelation')
    # plot 95% confidence intervals
```

```
    plt.axhline(y=-1.96/np.sqrt(len(series)), linestyle= '--', color= 'gray')
    plt.axhline(y=1.96/np.sqrt(len(series)), linestyle= '--', color= 'gray')
    plt.xlabel('lags')

    # 3rd plot - partial autocorrelation plot
    ax3 = fig.add_subplot(1, 3, 3)
    plot_pacf(series.dropna(), ax=ax3, lags=30, title='Partial Autocorrelation')
    plt.axhline(y=-1.96/np.sqrt(len(series)), linestyle= '--', color= 'gray')
    plt.axhline(y=1.96/np.sqrt(len(series)), linestyle= '--', color= 'gray')
    plt.xlabel('lags')
    plt.show()

    # ADF test
    result = adfuller(series.dropna(), regression='c', autolag='AIC')
    critical_value = result[4]['5%']
    if (result[1] <= 0.05) and (result[0] < critical_value):
        print('P-value = {:.6f}, the series is likely stationary.'.
↪format(result[1]))
    else:
        print('P-value = {:.6f}, the series is likely non-stationary.'.
↪format(result[1]))
    return
```

[34]:
```
# first difference
series_transformation(rpi_data['Rate'].diff())
```

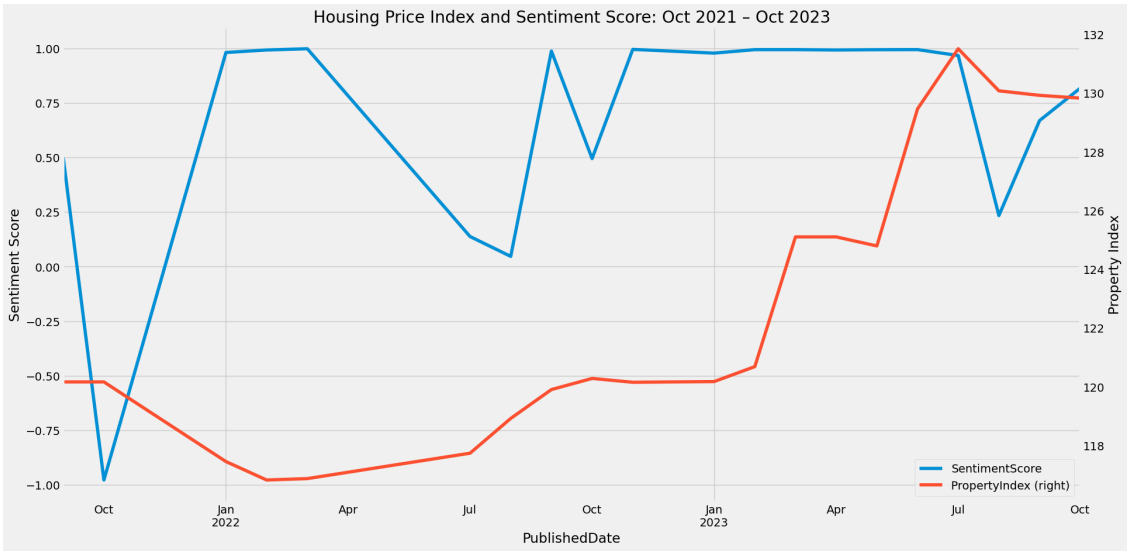



```
P-value = 0.000077, the series is likely stationary.
```

4 Finding the relationship between Sentiment Index and Housing Prices

[35]:
```
si_mean = sentiment_index.groupby('PublishedDate').compound.agg(['mean']).
 ↪reset_index()
si_mean = si_mean.rename(columns={'mean':'SentimentScore'})
si_mean_monthly = si_mean.groupby(pd.PeriodIndex(si_mean['PublishedDate'],␣
 ↪freq="M"))['SentimentScore'].mean().reset_index()
selected_month = rpi_data.loc['2022-01-01':'2023-09-01']

###

selected_month2 = selected_month.values.tolist()
# print(selected_month2)

rpi_list = []
for value in selected_month2:
    for item in value:
        rpi_list.append(item)

###

si_mean_monthly['PropertyIndex'] = rpi_list

###

fig, ax = plt.subplots(figsize=(20,10))
si_mean_monthly.plot(x='PublishedDate', y='SentimentScore', ax=ax, ylabel =␣
 ↪"Sentiment Score")
si_mean_monthly.plot(x='PublishedDate', y='PropertyIndex', ylabel = "Property␣
 ↪Index", ax=ax, secondary_y=True)
plt.title('Housing Price Index and Sentiment Score: Oct 2021 - Oct 2023')
```

[35]:
```
Text(0.5, 1.0, 'Housing Price Index and Sentiment Score: Oct 2021 - Oct 2023')
```



[36]:
```
corr = si_mean_monthly["SentimentScore"].corr(si_mean_monthly["PropertyIndex"],␣
 ↪method='pearson')
print("Correlation between Sentiment Score and Property Index is: ",␣
 ↪round(corr, 5))
```

```
Correlation between Sentiment Score and Property Index is:  0.13072
```