

<https://github.com/SaarthShah/BigMacIndex/blob/main/notebook.ipynb>

COGS9: Introduction to Data Science

Final Project

Group Member Information:

Assignment will be completed in a group of 4-5 people. These will be the same people you work with for the final project. This assignment will help get you started on the final project.

First Name	Last Name	PID
Saarth	Shah	A16963610
Anoushka	Sahai	A17047217
Arjun	Varshney	A17048280
Aryamun	Das	A17050602

Question

Is the change in Big Mac Prices a true indicator of rising inflation or just a strategy by McDonald's to increase profit margins?

Hypothesis

The change in Big Mac prices is a true indicator of inflation trends in the United States, trying to showcase how the percentage change in ingredients prices of the Big Mac Burger is exactly equivalent to the percentage change in the prices of the Big Mac in the McDonald's menu.

Justification

Rising Inflation is a hot topic these days, with the Federal Reserve calling it transitory. We came

up with this hypothesis after reading about the Big Mac Index from the Economist— a reading which compares the purchasing power of countries on the basis of the prices of the Big Mac burger by McDonald's. Provided that the Economist's Big Mac Index is accurate, the price of Big Mac burgers should grow at a rate that strictly follows inflation patterns. With our study, we want to see if the Big Mac Prices actually have a correlation with the change in prices of ingredients and the labor needed to make the sandwich.

Background Information

The Big Mac Index, originally created by the Economist as a light-hearted way of measuring the value between different currencies, has stood the test of time. Even today, the index is used to gauge the purchasing power between different countries and to find how under or overvalued one currency is to another. While the Economist article assures us that biases have been taken care of by adding a separate Gross Domestic Product Index there still remain some biases in the index as noted by the Economist itself in the end.

The Big Mac is supposed to represent a representative basket of goods and services in comparison to the goods and services consumed by an average American family as posited, in the Article by [The Economist](#). However, with recent [third-party analysis](#) of the data, some discrepancies have been found and the index has proven to not align with the consumer price index in the way it was supposed to. Interestingly, the actual price followed closely with the Consumer Price Index calculated price up till 2005, and then there was a steady increasing difference between the two leading to a gap of 1.5\$ between the two. This 43% difference highlights a problem with the Big Mac Index which makes it an unreliable measure to find the purchasing power parities between countries. Moreover, reverse engineering the inflation using the prices of the big Mac reveals “a 1% higher inflation as stated by the CPI”(What Do Big Mac Prices Tell Us About True Inflation?, 2021). Moreover, this trend is not replicated by the Mac prices in the United Kingdom as the CPI calculated price and annual price follow along until 2010, and then the two start to separate leading to a 34% difference. These irregularities further motivate us to find the rise in individual prices of the ingredients and compare those to the inflation rates.

Data

The perfect dataset for our study would contain the annual prices of individual ingredients that are required to make a big mac, prices of big macs in the United States, the minimum wage in the United States, and the annual Inflation rate for the United States. As the Economist Dataset of Big Mac Prices is from 1986 to 2021, we would ideally require observations for every year for all the commodities. For our variables, we would ideally need either the change in prices (per a specific quantity eg. per pound) or their actual prices (we can calculate the annual change) for ingredients, minimum wage, Big Mac prices, and inflation.

Dataset of the Big Mac Index:

<https://github.com/TheEconomist/big-mac-data/blob/master/source-data/big-mac-historical-source-data.csv>

Dataset for ingredients:

Lettuce cost per pound: <https://fred.stlouisfed.org/series/APU0000FL2101>

<https://fred.stlouisfed.org/series/APU0000712211>

Cheddar Cheese per pound: <https://fred.stlouisfed.org/series/APU0000710212>

Cucumbers/Pickles: <https://www.producepay.com/current-us-market-for-fresh-cucumber/>

Beef per pound: <https://fred.stlouisfed.org/series/PBEEFUSDQ>

Bread per pound: <https://fred.stlouisfed.org/series/APU0000702111>

Onions per pound: https://ycharts.com/indicators/us_onions_dry_price_received

Dataset for minimum hourly wage:

<https://fred.stlouisfed.org/series/FEDMINNFRWG>

Dataset for Average Inflation in the United States:

<https://fred.stlouisfed.org/series/FPCPITOTLZGUSA>

Weighted index in big mac:

<https://everythingwhat.com/how-many-grams-is-a-mcdonalds-big-mac>

The ingredients data is highly inconsistent with the time frame, for example, the lettuce data is from 2006 to 2022 and the bread data is from 1980 to 2022. Some of the data is in cents and some in dollars which would require some data wrangling. We will use these datasets to collect the approximate cost for replicating a big mac and calculating the profit margins. We would require pandas to combine all the data together into a single data frame, then we should be able to find the change in prices (**our expected Big Mac inflation**) and compare that to the dataset for the annual average inflation rate and the actual Big Mac Inflation.

Ethical Considerations

Data science projects must always consider the ethical consideration checklist to ensure the prevention of biases and ensure proper representation of information collected in the project.

Example Data Science Ethics Checklist

- ☒ Is your team biased?
- ☒ Data collection avoid sampling bias?
- ☒ Are the data biased?
- ☒ Informed consent obtained for all participants?
- ☒ Can one's data be removed on request?
- ☒ Does our analysis discriminate?
- ☒ Is our analysis transparent?
- ☒ What are the negative or unintended consequences of our findings?
- ☒ Do we have a plan to check for biases going forward and update our algorithm?

Team Bias:

The team analyzing the data is composed of international students and none of us have domain expertise in economics. While some members of the team do have financial analysis experience, we might still miss out on some important factors/parameters that would have heavily influenced our study.

Sampling Bias:

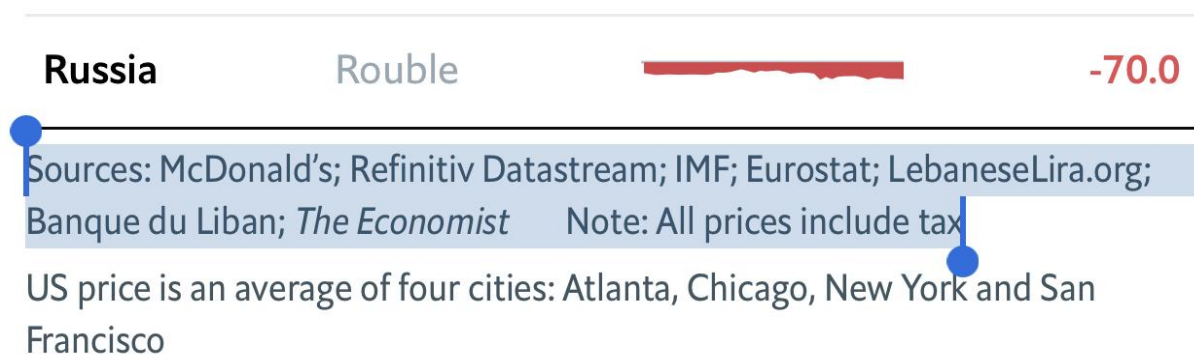
The source of the collection first is The Economist and Federal Reserve Economic Data. A [survey](#) of The Economist finds that it is a reliable source, very slightly leaning towards the left-liberal side of economics. However, The Big Mac data is collected from big cities like Atlanta, Chicago, New York, and San Francisco so there might be sampling bias as it does not accurately represent the smaller cities of the United States.

Data Bias:

The data is biased because there were no accurate datasets for very specific ingredients, for example, the dataset was available for Romaine lettuce while McDonald's uses shredded iceberg lettuce which might have a different change in price at a given time. Another example of data bias would be that a lot of data is not consistent over the years and while doing the analysis we might face issues when actually creating the algorithm.

Consent:

The data is secondarily corrected, directly by the source, The Economist. Appropriate credit is given to the providers by stating the different sources of data collection by the article writers. As the datasets, planned to be used in a study, are available for public use under the MIT license, we are not required to collect individual consent.



Algorithmic Bias / Discrimination:

We would be weighing each component of the big mac with an approximate weightage rather than accurately accounting for each ingredient. For example, an average cheese slice weighs 1 oz while the one used by McDonald's might weigh 1.5oz. While the McDonald's [website](#) accurately showcases which ingredients are present, it does not accurately explain the weight of each ingredient. Also, we were unable to account for the Thousand Island Dressing and Dill Pickles that McDonald's uses because there was no dataset available.

Data Privacy / Ownership:

The data used by us is mostly open source. The Big Mac dataset generated by The Economist is open-source and available under the [Creative Commons Attribution 4.0 International License](#).

Federal Reserve Economic Data(FRED) claims that its data can be used for anything, ["As long](#)

as you don't engage in a prohibited/restricted use and do adhere to any applicable copyright restrictions, you are free to use FRED for your own non-commercial, educational, and personal uses."

As we are strictly following the rules given by FRED, Economist, and Kaggle while not collecting any data by ourselves, we do not have any privacy or ownership considerations.

Transparency:

Our study would be 100% transparent with the data being open-sourced with our code and documentation and findings are available openly on Deepnote and GitHub for everyone to see and replicate.

Unintended Consequences:

The unintended consequences of our study could be that we might not take into account the rise in the cost of running a company and might create a negative image of the McDonald's corporation. Another unintended consequence could be that due to some flaws in our algorithm, our study might raise false doubts against the original Big Mac index study.

Continued Monitoring / Accountability:

As the public datasets continue to grow annually, we plan on validating our study with new data to remain accountable to our claims with time. As the analysis will be personally done by us, we have justifiable answers for every element of our algorithm.

Analysis Proposal

Data Collection:

Our analysis started with a **collection of data** of the average cost of the ingredients from Federal Reserve Economic Data. The datasets were recorded from different time ranges, hence to **obtain consistent data**, we decided to filter our dataset from 2000 onwards. We have obtained the data for the Big Mac Index from The Economist itself. We weren't able to find consistent data for cucumbers so we created our own dataset, consisting of yearly average prices starting from 2000 to 2020 by **scraping data** from various reputed sources and checking our average prices using **open-source data** from various websites. For our SPDR S&P 500 or SPY and \$MCD (Mc

Donalds' stock) dataset, we used pandas DataReader to import data directly from Yahoo Finance. To calculate the cost of all of the ingredients in a Big Mac, we found the proportion of ingredients used to create a Big Mac in grams and converted it to pounds. A weighted formula then helped us calculate the approximate ingredient cost of a big mac - summation of weighted prices of ingredients.

Our weighted formula was as follows:

Bun - 0.163 pounds, Beef patty - 0.099 pounds x 2 = 0.198 pounds, Cheese - 0.03 pounds, Lettuce - 0.062 pounds, Pickle - 2 x 0.0075 pounds = 0.015 pounds, Onion - 0.015 pounds, = Total Weight - 0.598 pounds.

Data Wrangling:

In the process of data wrangling, the first and most important step is to provide uniformity and a common time period for analyzing big mac prices. Many of the data were provided in scattered periods. For example, 1986 to 2022 for prices of cucumbers or lettuce from 2000-2022. We have followed the strict timeline from 2000-2020 for the analysis of inflation in big mac prices, as it gives an idea of the latest trends and the data can be verified from sources published more recently, which are in abundance. Moreover, keeping a uniform period for the data helps us to avoid confusion and errors during data analysis.

Furthermore, we have considered only the US local market big mac prices, averaged over a year, to get an idea of the annual trends in inflation, rather than taking the quarterly prices which fluctuate. Upon plotting a line graph (Figure 1) we find a steep and constant upward incline trend, which follows a trend very close to a straight line.

For our analysis, we decided to do it on Deepnote as it allowed real-time collaboration between our team. After downloading the datasets and reading the CSV files using Pandas, we found the data to be very inconsistent. In wrangling the datasets for the ingredients, we initially had to uniformize the format of prices of ingredients to a dollar per pound, as many of the prices were in dollar per kg or thousands of dollars per cwt. Moreover, we filtered out null values by making all the null values as 0 and adding a condition ($x > 0$) to each ingredient dataset, thereby getting rid of them. Null values were causing a discontinuity in our graphs, which in turn would have impacted our data analysis. After this, we took the annual mean of the prices of the ingredients in the market similar to the earlier process with big mac prices, so as to average out the quarterly fluctuations.

As for inflation and minimum wage, we followed the same process as the above, grouping the values by date, removing null values, and getting an average of the values for each year.

Similarly, we cleaned the SPY index dataset and GDP dataset, by fitting them in the annual time period from 2000 to 2020. For the GDP, we took the maximum reported value at 10-01 of each year, consistent with our trend of analyzing data on an annual criterion. To check the overall trend of the stock market, we decided to use the SPY Index. The SPDR S&P 500 is an exchange-traded fund (ETF) that was created to provide an investment vehicle that produces returns roughly in line with the S&P 500 Index before expenses. For SPY, we followed a little elaborate process, getting highs, lows, market opening, and closing prices, and finding the low on the 1st financial day and high on the last financial day. We then used this to calculate the change SPY index, which formed the basis of the analysis.

Finally, after merging and cleaning our datasets, **Exploratory Data Analysis** was performed starting with a plot of ingredient cost of a Big Mac vs actual prices (Big Mac Index).

Descriptive & Exploratory Data Analysis:

Once we had calculated the annual cost of all the ingredients of a big mac combined, we started comparing our study with popular economic indicators such as the Annual Inflation in the United States, SPDR S&P 500 (SPY), Change in the Gross Domestic Product.

First, to find the difference between the actual price of a big mac and our calculated price, we plotted a graph comparing these two values. We found that the change in actual prices is much steeper than the change in the total cost of the ingredients but annual percentage change follows the trends very closely (with similar peaks and dips).

As our data does not consider some of the smaller financial contributors in creating a Big Mac and emphasizes more on the actual ingredients, after a few rounds of analysis, it became evident that we needed to compare the change in the price of the two burgers, as it was more relevant in comparing to factors like Inflation and change in GDP. Thus, change in percentage is an accurate indicator to test our hypothesis of whether ingredients prices of the Big Mac Burger equal the prices given by the Big Mac index, thus following inflation trends in the United States.

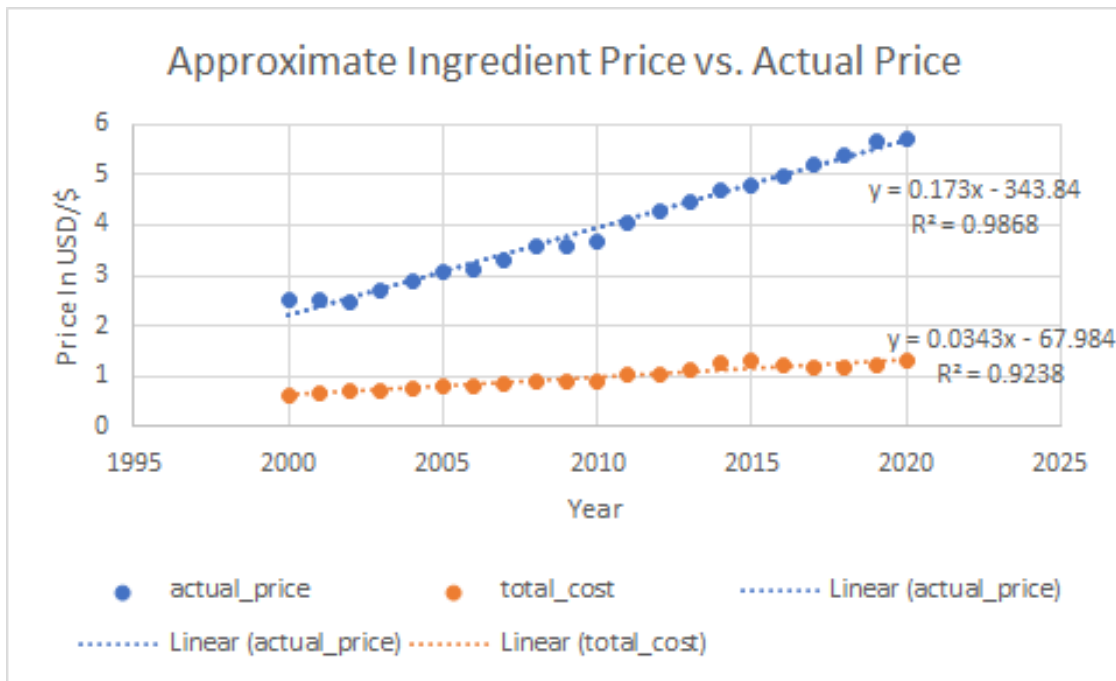


Figure 1. Approximate Ingredient price vs Actual price (Big Mac Index)

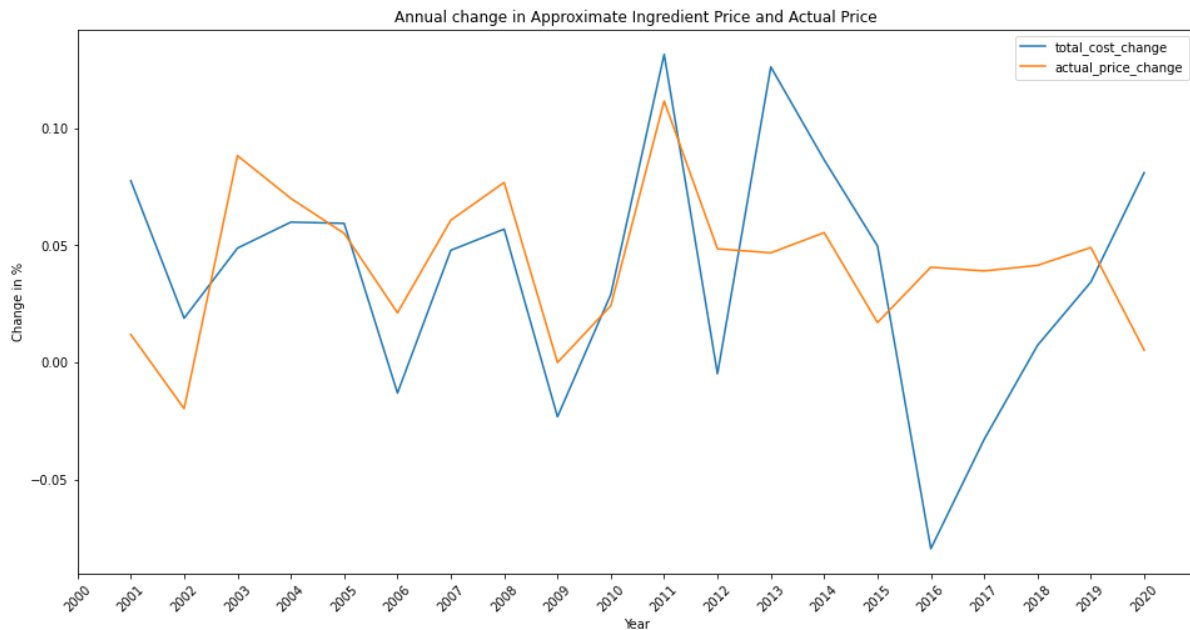


Figure 2. Annual change in Calculated Total Ingredient price and Actual price(Big Mac Index)

Data Visualization:

To visualize the results of our Exploratory Data Analysis, we made extensive use of line charts. These are very useful for displaying specific quantitative data over a certain time interval. For financial analysis, these charts appeared to be most suitable as they gauge the change of price over a certain time period. Furthermore, line charts easily help us identify the general trend of the prices of the ingredients versus the increase in prices of Big Macs and prove our hypothesis.

We also utilized Bar Charts to display the difference between the long-term percentage change in Total Price, SPY, and Actual Price. Bar charts are much better for visualizing the overall difference between values, hence we used them in this particular case.

We kept our visualizations simple and apt in order to quickly convey the necessary information without causing any confusion. Most of our charts were made through Python's Matplotlib package, with a few built on Microsoft excel.

Statistical Analysis (Inference, A/B testing, etc.):

We found that in 20 years, the actual price of a Big Mac had risen by 126% and our calculated cost had risen by around 106%. To compare this overall 20-year % increase in actual Big Mac prices and our total calculated prices, we plotted a bar chart. In the same time frame, S&P 500 provided a return of around 160%.

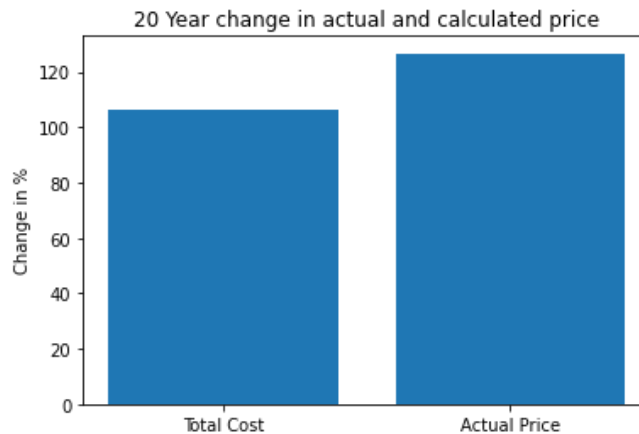


Figure 3. 20-year change in Calculated Total Ingredient price and Actual price(Big Mac Index)

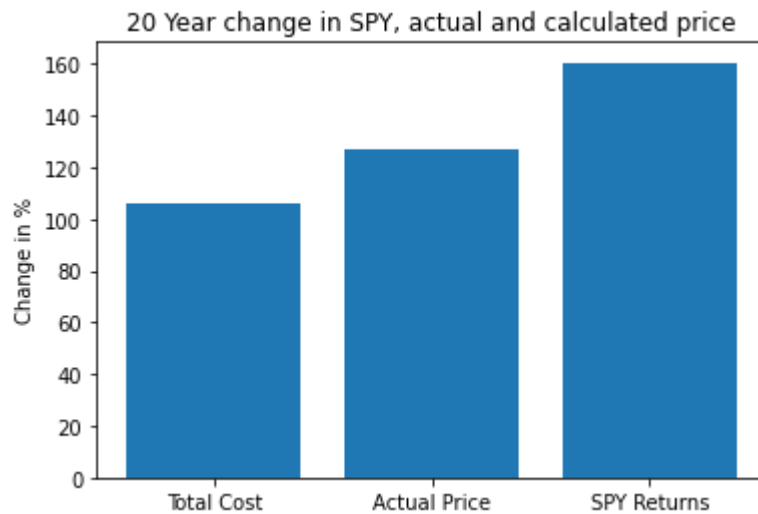


Figure 4. 20-year change in Calculated Ingredient price vs Actual price vs SPY returns

To gauge the overall pattern in which ingredient prices were increasing in comparison to our calculated total and actual prices, we plotted a line chart of prices of all the ingredients (in USD) with time. Since the prices of most of the individual ingredients are much lower than the actual price, we can relate this to the graphs in Figure 1 where the weighted sum of all ingredient prices is a line much lower than that of the actual price.

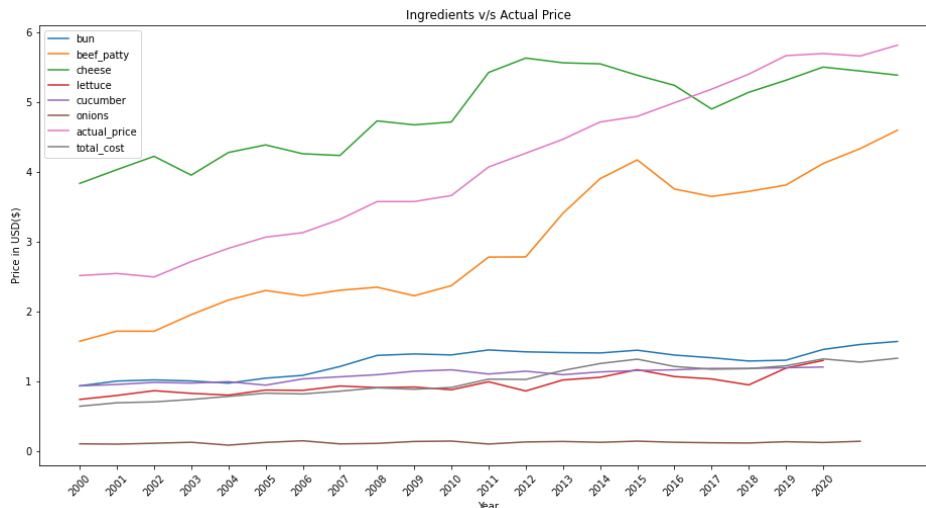


Figure 5. 20-year prices of each ingredient vs the Calculated Ingredient price vs Actual price.

Comparing the percentage change in our calculated prices and actual prices with the annual

inflation rate, we discovered that all 3 of them followed a very similar trend in peaks and dips but overall inflation was more consistent at around 4%.

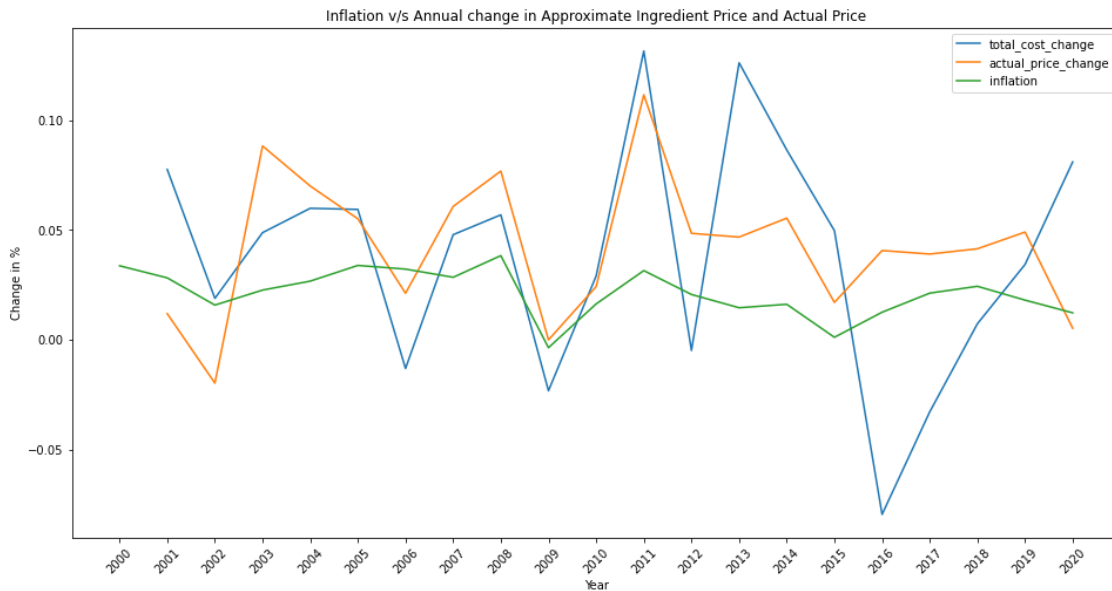


Figure 6. Inflation trend vs Change in Calculated Ingredient price and Actual price

Comparing the percentage change in our calculated prices and actual prices with the change in Gross Domestic Product, we discovered that all 3 of them followed a very similar trend in peaks and dips.

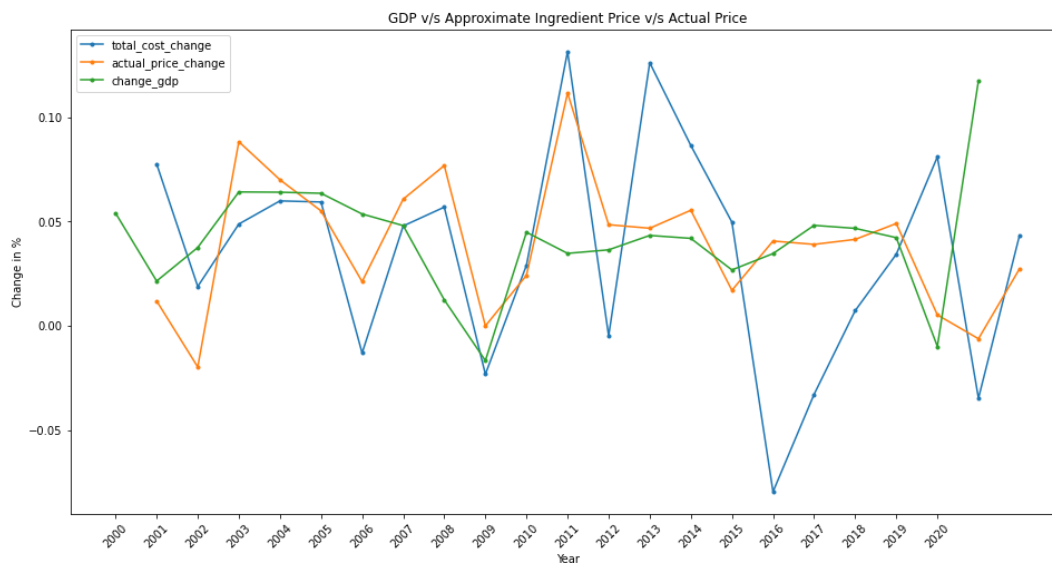


Figure 7. 20-year change in Calculated Ingredient price vs Actual price(Big Mac Index) vs GDP

Comparing the percentage change in our calculated prices and actual prices with the change in S&P 500, we discovered that the stock market had almost no correlation with Big Mac prices.

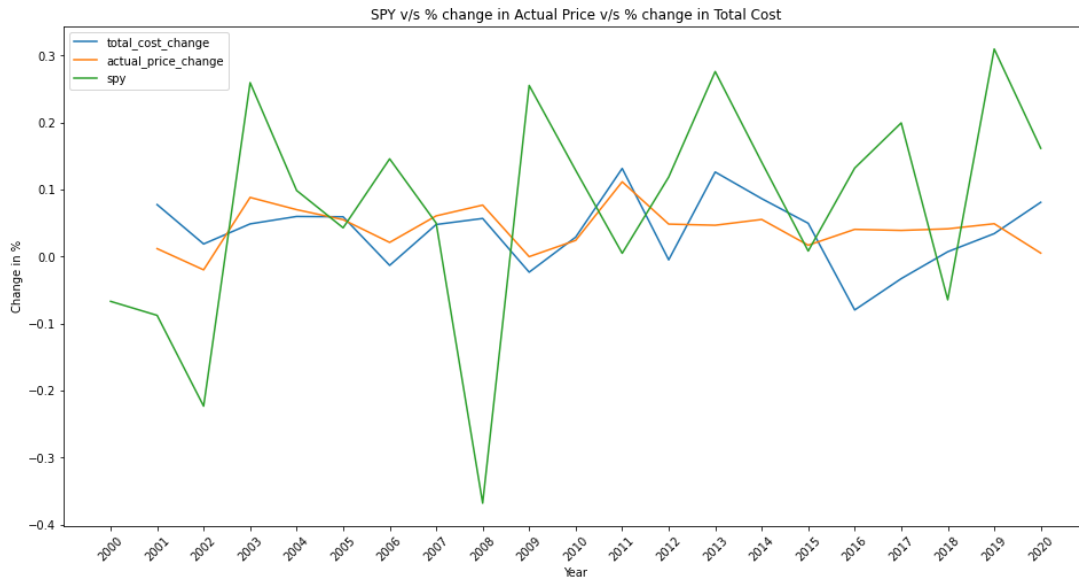


Figure 8. Percentage change in Calculated Ingredient price
(Total Cost) vs Actual Price(Big Mac Index) vs SPY over 20 Years.

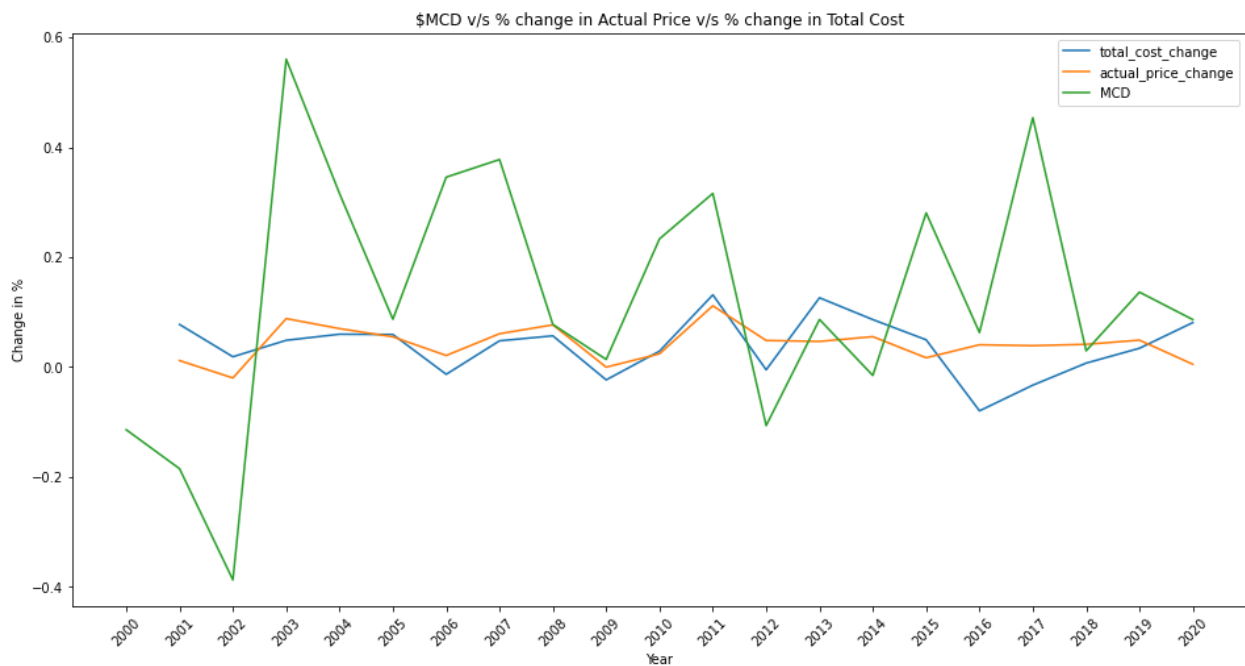


Figure 9. Percentage change in Calculated Ingredient price
(total cost) vs Actual price(Big Mac Index) vs MCD.

Discussion

From the various graphs drawn, we can see a clear trend of rising in the cost of prices of the ingredients and the Big Mac respectively; however, whenever there are falls in the prices of ingredients, the prices of the Big Mac doesn't decrease proportionally. We can see this in Figure 2, where we have plotted the cost of all ingredients (in pounds) and compared them to the actual price of a Big Mac. This was facilitated by the fact that the prices of almost all of the ingredients increased slowly over 20 years except for the prices of cheese and beef, which seem to be growing at a much faster rate. Even when the prices of cheese and beef reduced between 2015 and 2016, the price of Big Macs gradually increased. The cost was reduced by more than 7% in 2016 due to a first in a 50-year drop in the cost of beef and dairy products causing a sudden fall in prices, however, McDonald increased prices by almost 4%. Moreover, figure 6 further illustrates the relationship between inflation trends, Actual price, and calculated prices such that the trend for the actual price was similar to that of inflation. This means that the rise in daily goods prices in the United States does have a correlation with the increase in the price of a Big Mac sandwich.

Furthermore, we can see that there is a clear similarity between the rate of increase of the calculated ingredient price and the Big Mac up till 2012 and then a staunch difference as the prices of the ingredients have changed but the price of the Big Mac hasn't. This correlation, in the beginning, is further strengthened by the R-value of 0.96 found between our calculated price and the actual price of the Big Mac. Moreover, the rise in prices of Big Mac is steeper than the rise in prices of ingredients as the starting base of total ingredient price is around 0.64 dollars while the starting price of an actual Big Mac is 2.51 dollars, thus the actual increase in prices varies greatly and compounds over 20 years leading to a wide difference between the two. This can be observed in the gradients of the two lines in the graph, Actual Prices versus Actual Cost, as the gradient of actual prices is 0.173 while the gradient of the total cost is 0.034. This can be further seen in their annual income statement as they have earned record-breaking revenue of 23 billion dollars in 2021.

There are multiple pitfalls in our methodology due to the lack of accountability of other fixed costs paid by McDonald's such as the advertisement costs, labor costs, land costs as well as costs associated with producing the burger itself such as the cost of the spices, oil, and sauces which

remain unknown to the public. We also assume that McDonald's sources all of its ingredients from the USA, rather than importing ingredients at a lower price from other countries. Other factors such as the daily maintenance of equipment, purchasing of new equipment which reduces costs in the long run, as well as electricity, water, and other basic utilities, are not considered. To further incorporate these values in our analysis we would require official data from franchisees of McDonald's which would represent a good sample of all franchisees in the USA. We will need to collect various important data on all of the franchisees and then randomly sample them and convince them to share their day-to-day costs data and other yearly costs. Revealing such data as part of a public study could lead to financial secrets of McDonald's being revealed and the firm would not be comfortable allowing students to have such a close look at their financial documents.

From our analysis, we can come to the conclusion that the Big Mac Index is an accurate indicator when comparing the percentage change in Big Mac prices with inflation however it is not a good indicator of the actual prices of a basket of common goods and services as the change in ingredients prices of the Big Mac Burger are not equivalent to prices indicated by the values of the Big Mac index. This can be witnessed in Figure 6 where we compare the inflation with actual price change as the shapes of the two lines were quite similar however the R-value is 0.56 as the actual percentage price change is more volatile than inflation.

Therefore, in our study, we come to the conclusion that the Big Mac Index approximately represents the actual inflationary trends in the United States.

Our Python Notebook is available open-source on:

<https://github.com/SaarthShah/BigMacIndex/blob/main/notebook.ipynb>

Group Participation

Saarth performed the data wrangling operations, exploratory and statistical analysis, formulated the question, hypothesis, and justifications. Anoushka contributed to grammatical edits, proofreading, ethics checklist, and ethical considerations, verifying biases at each step of the project. Arjun contributed to the collection and wrangling of datasets, discussions, statistical analysis, and the overall organization of the document. Aryamun contributed to the collection of datasets, proofreading, and checking the final and initial stages of the project.

Works Cited:

- <https://www.economist.com/big-mac-index>

- <https://medium.datadriveninvestor.com/what-do-big-mac-prices-tell-us-about-true-inflation-27f44912500f#:~:text=Between%201986%20and%202021%2C%20the,than%20recorded%20CPI%20per%20annum.>