

Evaluation of CPI Predictions for Meats, Dairy Products, and Fruits and Vegetables from 2004 - 2025

Introduction:

Our group set out to examine how the Consumer Price Index (CPI) for commonly purchased household foods has changed over the past 20 years and how forecasted trends help explain shifts in food prices over time. To conduct this analysis, we used CPI forecast data published by the United States Department of Agriculture (USDA), which regularly produces inflation forecasts for various food categories as part of its economic research and outlook reports.

These forecasts are widely relied upon by economists, policymakers, and researchers to assess future inflationary pressures in food markets, making them especially valuable for understanding long-term pricing trends. The data set includes key attributes such as CPI item categories, forecast years, and projected percentage changes. To represent a single, central estimate for each forecast, we used the mean forecasted percentage change, which allowed for a clear and consistent comparison across food categories and years.

Because the USDA is a reputable federal agency that follows standardized data collection and reporting practices, this data set can be considered both reliable and authoritative for examining changes in food prices and their broader economic implications.

Exploratory Data Analysis:

Research Question

How have USDA forecasted CPI percent changes for meats, dairy products, and fruits and vegetables differed over time, and what do these forecasts reveal about relative price stability and volatility across food categories?

Data Provenance:

The data set used in this analysis, CPI Historical Forecast, originates from the United States Department of Agriculture - USDA. The USDA regularly publishes CPI forecasts for various food categories as part of its economic research and outlook reports. These forecasts are widely used by policymakers, economists, and researchers to anticipate inflationary pressures in food markets. The data includes multiple forecast attributes, such as year of forecast and year being forecast. Year of forecast is the year the predicted CPI percent change was created, and year being forecast is when they predict the change to occur. Since the USDA is a reputable federal agency that follows standardized data collection, the data are considered authoritative, reliable, and suitable for economic analysis.

This dataset is particularly well-suited for addressing our research question: how do USDA forecasted CPI percent changes differ across meats, dairy products, and fruits and vegetables, and what do these differences reveal about relative price stability and volatility over time?

FAIR and CARE Principles: The dataset largely satisfies the FAIR principles:

1. Findable: This dataset is publicly available through USDA publications and online repositories, with clear titles and documentation.
2. Accessible: This dataset can be downloaded without restrictions and read using standard tools (ex: R or Excel).
3. Interoperable: This dataset is provided in structured tabular formats that integrate easily with common data analysis software and libraries.
4. Reusable: This dataset has clear definitions of the variables, consistent formatting, and public licensing allows the data to be reused for multiple analytical purposes.

Overall, this dataset strongly adheres to the FAIR principles, making it suitable for transparent and reproducible research, and a trustworthy data set for our project.

However, the CARE principles - Collective Benefit, Authority to Control, Responsibility, and Ethics - are not as applicable to the context of this data set. This data set does not involve indigenous data or personally identifiable information, and therefore issues surrounding the idea of community authority or consent are minimal. However, this data set does support collective benefit as the data set supports public understanding of inflation and food prices, benefiting society broadly. Additionally, the data set supports responsibility and ethics as the USDA adheres to ethical standards in data reporting and dissemination. While CARE principles are not central to this dataset, the data are used responsibly and ethically in a public-interest context.

Attributes: The original attributes from the data set include:

- 1. Consumer Price Index Item which identifies the CPI category for the given items in the data set.
- 2. Year of Forecast: This is the year in which the forecast was made.
- 3. Year Being Forecast: This is the year for which the CPI percent change is being predicted to be in.
- 4. Forecast Percent Change: This is the predicted percent change in CPI for the specified category and year.

Distinguishing between the year of forecast and the year being forecast is critical for interpreting trends, as it allows us to evaluate how expectations of price changes evolve over time.

EDA Focus and Analytical Approach:

For our analysis, and due to the magnitude of possible items within the Consumer Price Index Item attribute, we focus on the categories meats, dairy products, and fruits and vegetables and their associated forecast percent changes. These categories were selected because they represent commonly purchased household foods and differ in supply dynamics, making them appropriate for comparing forecasted price volatility and stability.

To address our research question, we examine how forecasted CPI behavior varies across these three categories using both summary statistics and visualizations. Because multiple forecasts exist for a given year, we use the mean forecasted percent change to represent a central tendency, allowing for consistent comparison across categories and time.

For this project, we create three data visualizations and three tables to evaluate the USDA’s CPI predictions for each category:

We create line graphs to illustrate how forecasted CPI percent changes fluctuate over time for each CPI item from 2004 to 2025, highlighting periods of increased volatility and spread across categories.

We create box-and-whisker plots for each CPI item to evaluate the five descriptive statistics (minimum, Q1, median, Q3, and maximum), which allows us to compare distributional spread and identify which food categories exhibit greater forecast variability.

Finally, we created a scatterplot of the mean forecasted percent change for each CPI item, directly linking the EDA to the research question by comparing long-term average inflation expectations across food categories.

To support our visual analyses, we pair each graph with a corresponding data table to provide quantitative context and strengthen interpretation. Together, these exploratory steps establish early evidence that forecasted CPI behavior differs meaningfully across food categories, setting the foundation for the results and conclusions that follow.

Visual Analysis:

Figure 1:

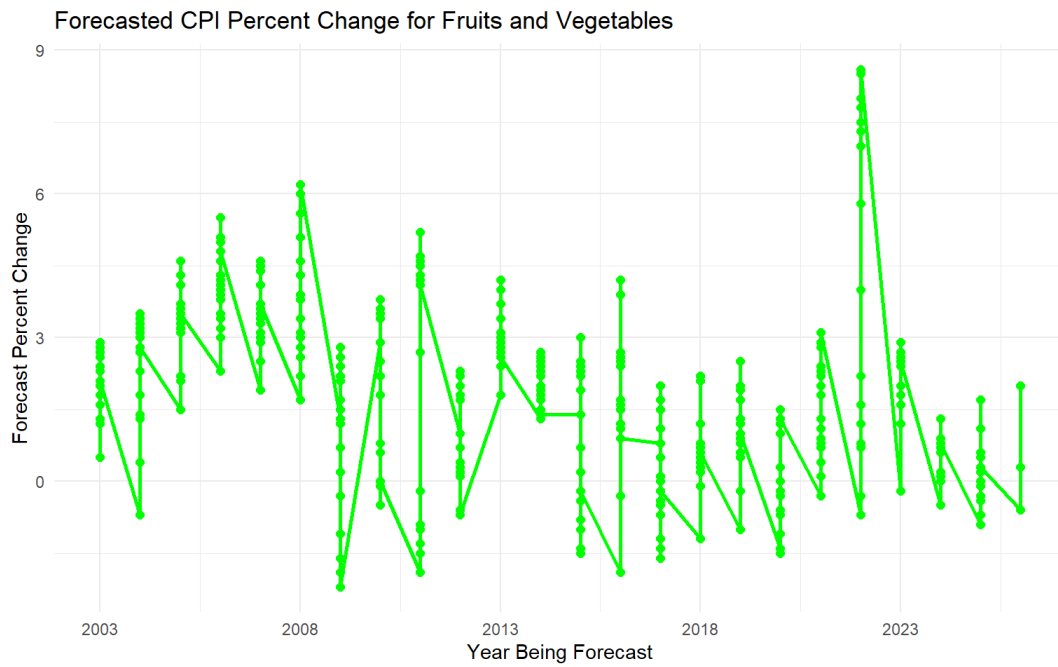
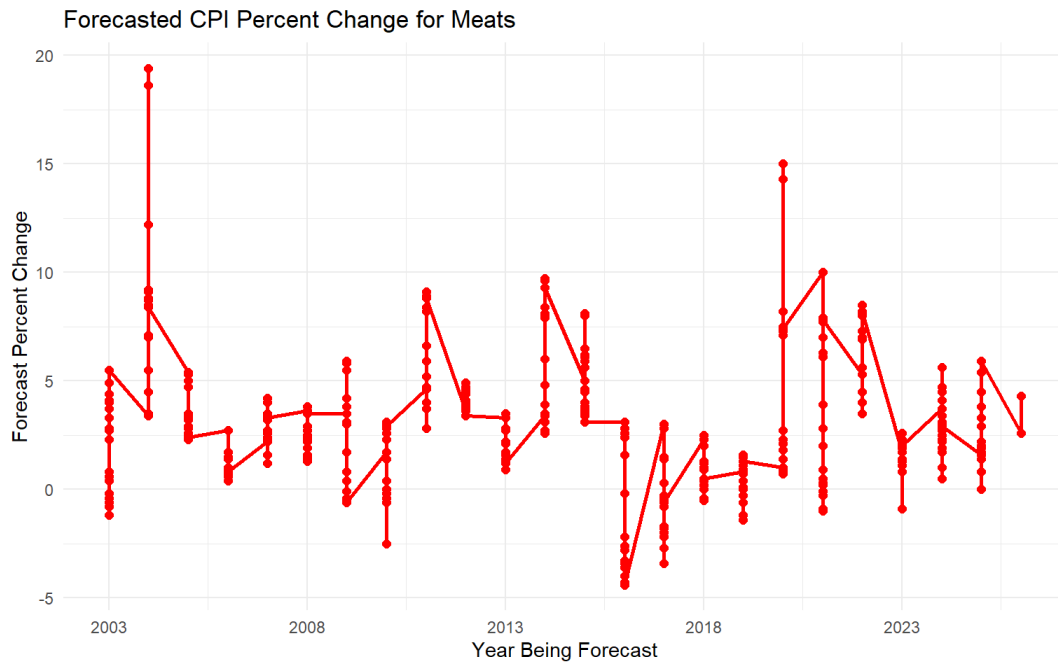
Year vs Forecast Percent Change for Each CPI Item			
consumer_price_index_item	year_of_forecast	year_being_forecast	forecast_percent_change
Meats	2002	2003	0.6
Meats	2002	2003	-0.2
Meats	2002	2003	0.8
Meats	2002	2003	-0.4
Meats	2002	2003	-1.2
Meats	2002	2003	-0.6
Meats	2003	2003	0.4
Meats	2003	2003	-0.8
Meats	2003	2003	3.3
Meats	2003	2003	2.8
Meats	2003	2003	2.7
Meats	2003	2003	2.3
Meats	2003	2003	3.7
Meats	2003	2003	4.0
Meats	2003	2003	4.1

consumer_price_index_item	year_of_forecast	year_being_forecast	forecast_percent_change
Meats	2003	2003	4.4
Meats	2003	2003	4.9
Meats	2003	2003	5.5
Dairy products	2002	2003	-1.2
Dairy products	2002	2003	-0.7
Dairy products	2002	2003	-0.2
Dairy products	2002	2003	-1.3
Dairy products	2002	2003	-0.8
Dairy products	2002	2003	-0.3
Dairy products	2003	2003	-0.4
Dairy products	2003	2003	-1.3
Dairy products	2003	2003	-0.3
Dairy products	2003	2003	-0.6
Dairy products	2003	2003	-1.6
Dairy products	2003	2003	-1.4
Dairy products	2003	2003	-2.2
Dairy products	2003	2003	-1.5
Dairy products	2003	2003	-0.6
Dairy products	2003	2003	-0.1
Dairy products	2003	2003	-0.1
Dairy products	2003	2003	-0.2
Fruits and vegetables	2002	2003	1.6
Fruits and vegetables	2002	2003	2.4
Fruits and vegetables	2002	2003	2.7
Fruits and vegetables	2002	2003	2.8
Fruits and vegetables	2002	2003	2.9
Fruits and vegetables	2002	2003	2.8
Fruits and vegetables	2003	2003	1.8
Fruits and vegetables	2003	2003	2.8
Fruits and vegetables	2003	2003	1.2
Fruits and vegetables	2003	2003	1.3
Fruits and vegetables	2003	2003	0.5
Fruits and vegetables	2003	2003	2.0
Fruits and vegetables	2003	2003	2.3
Fruits and vegetables	2003	2003	2.6

Figure 1 Analysis:

Figure 1 shows a data table of Year vs Forecast Percent Change for Each CPI Item. The plot includes attributes such as the consumer price index item which includes foods such as meats, dairy products, and fruits and vegetables, year of forecast, year being forecast, and forecast percent change, which shows how much the predicted value is expected to increase or decrease compared to the starting point.

Figure 2:



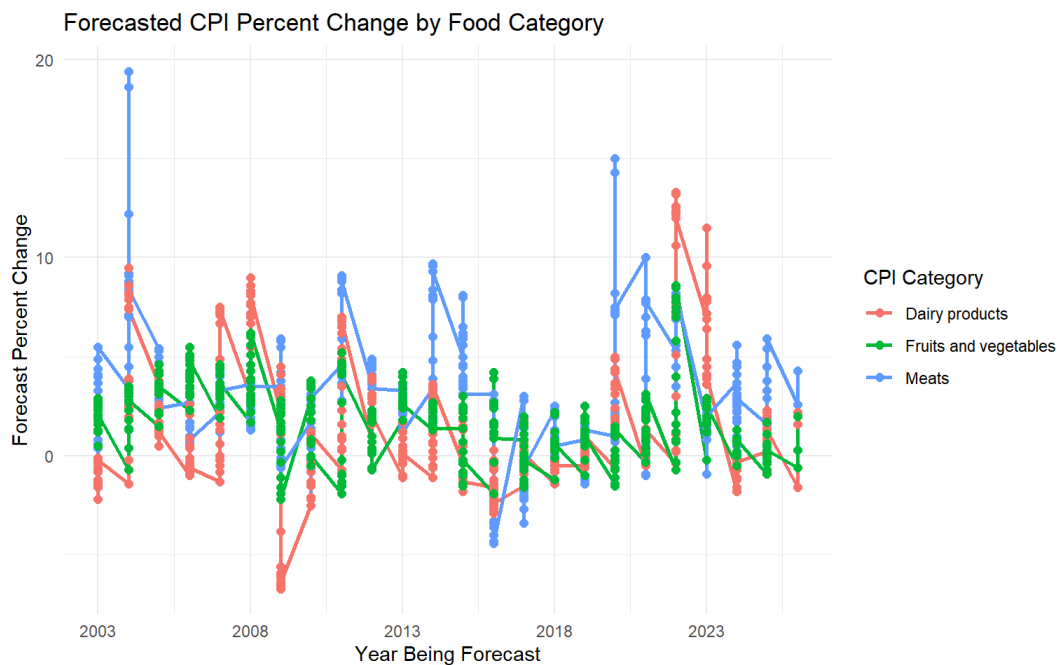
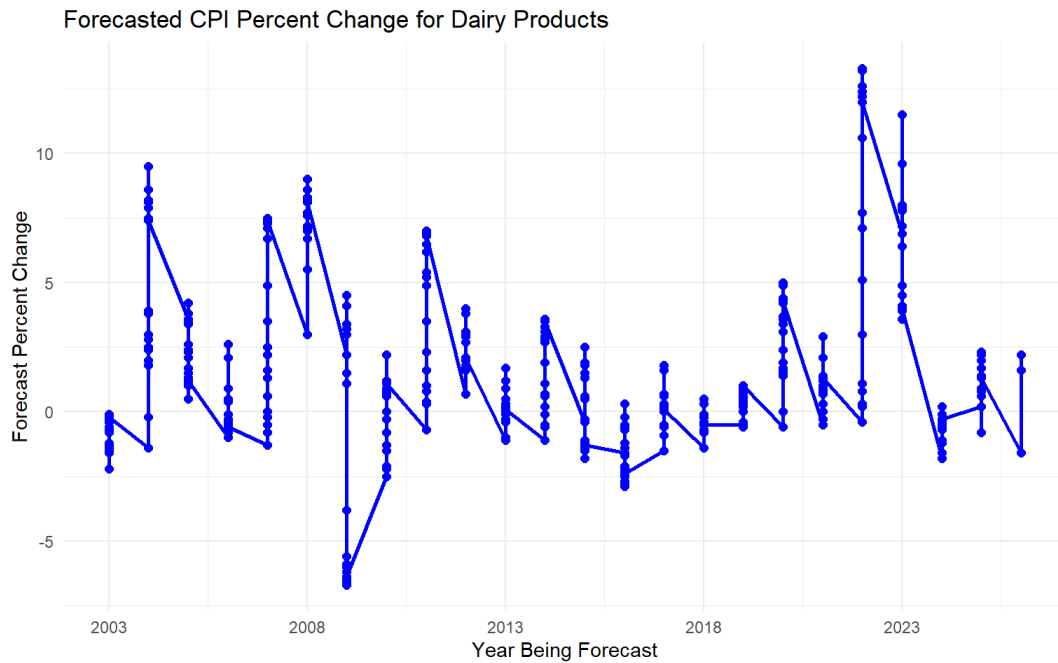


Figure 2 Analysis:

Figure 2 shows the forecasted CPI percent changes for meats, fruits and vegetables and dairy products. As you can see, there's big spikes for meats around 2004 and 2019, and a spike increase for fruits, vegetables, and dairy products in 2022. This shows how the forecasted percent change affected the price of these products over time. The last graph shows that forecasted CPI percent changes vary substantially across food categories and over time. Meats display the greatest volatility, with several sharp spikes and wider fluctuations compared to dairy products and fruits and vegetables. In contrast, fruits and vegetables exhibit more stable and tightly clustered forecasts, while dairy products generally fall between the two, reflecting moderate variability.

Figure 3:

Mean and Standard Deviation of CPI Forecasted Percent Change by Food Category and Year

consumer_price_index_item	year_being_forecast	mean_forecast_pct_change	sd_forecast_pct_change
Dairy products	2003	-0.8222222	0.6207733
Dairy products	2004	4.4611111	3.3099029
Dairy products	2005	2.2055556	1.1106290

consumer_price_index_item	year_being_forecast	mean_forecast_pct_change	sd_forecast_pct_change
Dairy products	2006	0.0611111	0.9989046
Dairy products	2007	2.7555556	3.2316855
Dairy products	2008	7.3777778	1.3575738
Dairy products	2009	-2.0777778	4.6655881
Dairy products	2010	-0.0444444	1.3699387
Dairy products	2011	3.4055556	2.6853573
Dairy products	2012	2.3833333	0.8361326
Dairy products	2013	0.1611111	0.7022141
Dairy products	2014	1.7166667	1.6607050
Dairy products	2015	-0.0833333	1.3750936
Dairy products	2016	-1.5388889	1.0169961
Dairy products	2017	0.1500000	0.7763035
Dairy products	2018	-0.3722222	0.4169999
Dairy products	2019	0.3888889	0.4945454
Dairy products	2020	2.8611111	1.7040225
Dairy products	2021	0.9333333	0.7977911
Dairy products	2022	6.8888889	5.4563389

Figure 3 Analysis:

Although the table does not display all 3 categories (would be a very long table if it did), Figure 3 shows the Mean and Standard Deviation of CPI Forecasted Percent Change by Food Category and Year. This data table includes attributes such as consumer price index item which includes products such as meats, fruits and vegetables, and dairy products, year being forecast, mean forecast percentage change, and the standard deviation forecast percentage change. Both the mean and standard deviations shows the fluctuation in forecast percentage change which can affect the price of the products mentioned.

Figure 4:

Range of Mid-Point CPI Forecasted Percent Change by Food Category

consumer_price_index_item	min_forecast_pct_change	max_forecast_pct_change
Dairy products	-6.7	13.3
Fruits and vegetables	-2.2	8.6
Meats	-4.4	19.4

Figure 4 Analysis:

From this table, we observe the minimum and maximum forecasted percent changes for each of the three categories. Meats exhibit the largest positive maximum forecasted percent change, while fruits and vegetables show the smallest maximum change. In contrast, dairy products have the largest minimum forecasted percent change, whereas fruits and vegetables again have the smallest minimum value. This shows that fruits and vegetables had the smallest variability in forecasted percent change among the 3 groups

Figure 5:

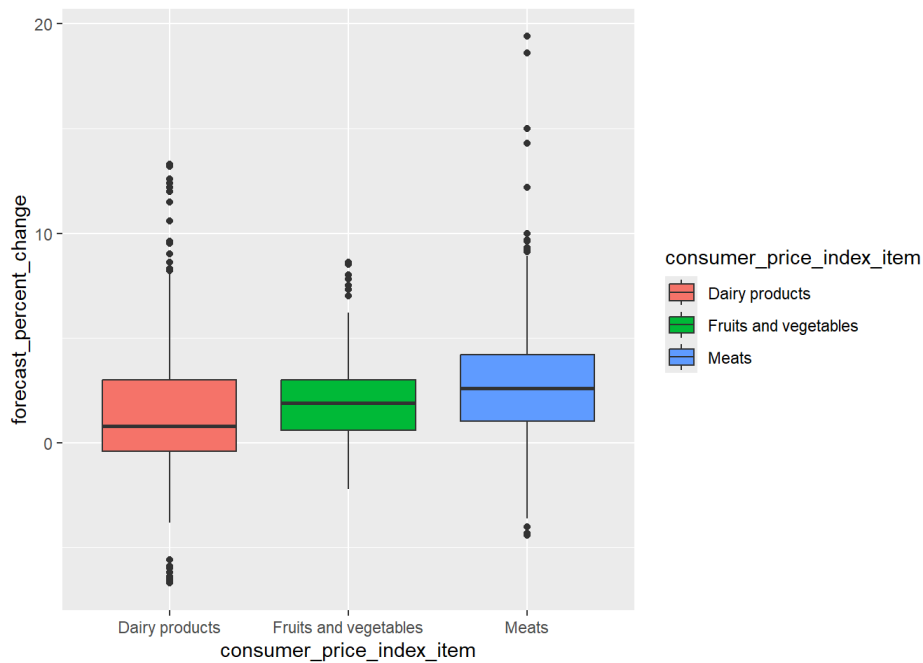


Figure 5 Analysis:

Figure 5 compares the distribution of forecasted percent changes across dairy products, fruits and vegetables, and meats. Meats show the highest median forecast percent change and the greatest variability, with several large positive outliers, indicating more volatile and potentially higher price growth expectations. Fruits and vegetables have a relatively higher median than dairy but a narrower spread, suggesting more moderate and stable forecasts. Dairy products display the lowest median forecasted change and include several negative outliers, reflecting weaker or more uncertain price growth compared to the other categories.

Figure 6:

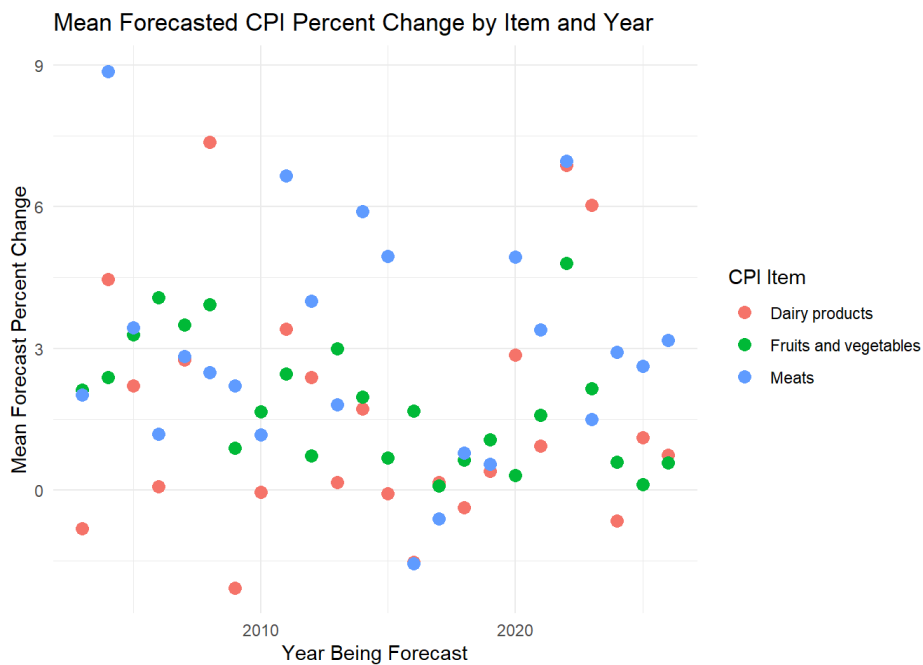


Figure 6 Analysis:

Figure 6 illustrates the mean forecasted CPI percent change over time for dairy products, fruits and vegetables, and meats. The mean forecasts for meats are generally higher and more volatile across years, with several periods showing higher average percent changes relative to the other categories. Fruits and vegetables showcase lower and more stable mean forecasted changes, typically clustered around small positive growth. Dairy products fall between the two, with mean

forecasts that fluctuate over time and occasionally approach zero or become negative, indicating greater uncertainty than fruits and vegetables but less volatility than meats.

Conclusion:

This study looked to answer the research question: how have USDA forecasted CPI percent changes for meats, dairy products, and fruits and vegetables differed over time, and what do these forecasts reveal about relative price stability and volatility across food categories? Overall, our analysis demonstrates that while food prices have generally trended upward from 2004 to 2025, the magnitude and consistency of forecasted inflation varies meaningfully by category.

Using USDA CPI forecast data and focusing on the mean forecasted percent change as a central measure, we found that meats consistently exhibit higher average forecasted CPI increases and greater variability, indicating stronger inflationary pressure and less price stability over time. In contrast, fruits and vegetables show lower and more stable mean forecasted percent changes, suggesting comparatively predictable price behavior. Dairy products fall between these two extremes, with moderate mean forecasts and occasional periods of negative or near-zero growth, reflecting intermittent volatility rather than sustained inflation.

These findings highlight that not all food categories respond equally to broader economic conditions, and that CPI forecasts can provide valuable insight into which household food items are more susceptible to sharp price changes. Because this analysis relies on forecasts from a reputable federal source, the results offer a credible framework for interpreting historical trends and comparing inflation risk across food categories.

One limitation of this study is the absence of observed CPI outcomes for years that have already occurred. If an additional dataset containing actual realized CPI values for meats, dairy products, and fruits and vegetables were available, it would allow for a direct comparison between predicted and actual CPI changes. Combining forecasted CPI data with observed CPI data would enable an evaluation of forecast accuracy, bias, and error over time, providing deeper insight into how well USDA predictions align with real-world price movements.

Based on the observed patterns, we anticipate that CPI for meats, dairy products, and fruits and vegetables will continue to increase beyond 2025, but at a slower and more stable pace than during recent high-inflation periods. Although food prices are expected to remain on the higher end, the data suggests that inflation will cool down over time rather than increase or decrease sharply.

Author Contribution:

The authors of this report are Eli Idiculla, Ethan Agaon, and Emily Shehan.

Code Appendix

```
# -----
# Code Header
# Primary Author: Emily Shehan
# Reviewer: Eli Idiculla
# -----

library(readr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)
library(readxl)
library(janitor)
library(knitr)

# -----
# Code Header
# Primary Author: Emily Shehan
# Reviewer: Eli Idiculla
# -----

CPIHistoricalForecast <- read.csv("CPIHistoricalForecast.csv")

# -----
# Code Header
# Primary Author: Eli Idiculla
```



```
# Reviewer: Ethan Agaon
# -----

cpi_tidy <- CPIHistoricalForecast %>% # Tidying the dataset to show Consumer Price Index Item (only showing "Meat
  clean_names() %>%
  # use only mid-point forecasts
  filter(attribute == "Mid point of prediction interval") %>%
  # keep only the desired CPI items
  filter(
    consumer_price_index_item %in% c(
      "Meats",
      "Fruits and vegetables",
      "Dairy products"
    )
  ) %>%
  select(
    consumer_price_index_item,
    year_of_forecast,
    year_being_forecast,
    forecast_percent_change
  )

kable(cpi_tidy[1:50,],caption = "Year vs Forecast Percent Change for Each CPI Item") # Viewing tidied dataset
```

```
# -----
# Code Header
# Primary Author: Eli Idiculla
# Reviewer: Ethan Agaon
# -----

ggplot(
  cpi_tidy %>% filter(consumer_price_index_item == "Meats"),
  aes(x = year_being_forecast,
      y = forecast_percent_change)
) +
  geom_line(color = "red", linewidth = 1) +
  geom_point(color = "red", size = 2) +
  scale_x_continuous(breaks = seq(
    min(cpi_tidy$year_being_forecast),
    max(cpi_tidy$year_being_forecast),
    by = 5
  )) +
  labs(
    title = "Forecasted CPI Percent Change for Meats",
    x = "Year Being Forecast",
    y = "Forecast Percent Change"
  ) +
  theme_minimal()
```

```
ggplot(
  cpi_tidy %>% filter(consumer_price_index_item == "Fruits and vegetables"),
  aes(x = year_being_forecast,
      y = forecast_percent_change)
) +
  geom_line(color = "green", linewidth = 1) +
  geom_point(color = "green", size = 2) +
  scale_x_continuous(breaks = seq(
    min(cpi_tidy$year_being_forecast),
    max(cpi_tidy$year_being_forecast),
    by = 5
  )) +
  labs(
    title = "Forecasted CPI Percent Change for Fruits and Vegetables",
    x = "Year Being Forecast",
    y = "Forecast Percent Change"
  ) +
  theme_minimal()
```

```

ggplot(
  cpi_tidy %>% filter(consumer_price_index_item == "Dairy products"),
  aes(x = year_being_forecast,
      y = forecast_percent_change)
) +
  geom_line(color = "blue", linewidth = 1) +
  geom_point(color = "blue", size = 2) +
  scale_x_continuous(breaks = seq(
    min(cpi_tidy$year_being_forecast),
    max(cpi_tidy$year_being_forecast),
    by = 5
  )) +
  labs(
    title = "Forecasted CPI Percent Change for Dairy Products",
    x = "Year Being Forecast",
    y = "Forecast Percent Change"
  ) +
  theme_minimal()

```

```

ggplot(
  cpi_tidy %>%
    filter(consumer_price_index_item %in% c(
      "Meats",
      "Dairy products",
      "Fruits and vegetables"
    )),
  aes(
    x = year_being_forecast,
    y = forecast_percent_change,
    color = consumer_price_index_item
  )
) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_x_continuous(
    breaks = seq(
      min(cpi_tidy$year_being_forecast),
      max(cpi_tidy$year_being_forecast),
      by = 5
    )
  ) +
  labs(
    title = "Forecasted CPI Percent Change by Food Category",
    x = "Year Being Forecast",
    y = "Forecast Percent Change",
    color = "CPI Category"
  ) +
  theme_minimal()

```

```

# -----
# Code Header
# Primary Author: Ethan Agaon
# Reviewer: Eli Idiculla
# -----

cpi_summary <- cpi_tidy %>%
  group_by(consumer_price_index_item, year_being_forecast) %>%
  summarise(
    mean_forecast_pct_change = mean(forecast_percent_change, na.rm = TRUE),
    sd_forecast_pct_change = sd(forecast_percent_change, na.rm = TRUE),
    .groups = "drop"
  )

kable(cpi_summary[1:20,],
      caption = "Mean and Standard Deviation of CPI Forecasted Percent Change by Food Category and Year")

```

```

# -----
# Code Header
# Primary Author: Emily Shehan
# Reviewer: Eli Idiculla

```

```
# -----

cpi_range_table <- cpi_tidy %>%
  group_by(consumer_price_index_item) %>%
  summarise(
    min_forecast_pct_change = min(forecast_percent_change, na.rm = TRUE),
    max_forecast_pct_change = max(forecast_percent_change, na.rm = TRUE),
    .groups = "drop"
  )

kable(
  cpi_range_table,
  digits = 2,
  caption = "Range of Mid-Point CPI Forecasted Percent Change by Food Category"
)
```

```
# -----
# Code Header
# Primary Author: Ethan Agaon
# Reviewer: Eli Idiculla
# -----

ggplot(cpi_tidy, aes(x = consumer_price_index_item,
                     y = forecast_percent_change,
                     fill = consumer_price_index_item)) +
  geom_boxplot()
```

```
# -----
# Code Header
# Primary Author: Emily Shehan
# Reviewer: Eli Idiculla
# -----

cpi_mean <- cpi_tidy %>%
  group_by(consumer_price_index_item, year_being_forecast) %>%
  summarise(
    mean_forecast_percent_change = mean(forecast_percent_change, na.rm = TRUE),
    .groups = "drop"
  )

ggplot(cpi_mean,
  aes(x = year_being_forecast,
      y = mean_forecast_percent_change,
      color = consumer_price_index_item)) +
  geom_point(size = 3) +
  labs(
    title = "Mean Forecasted CPI Percent Change by Item and Year",
    x = "Year Being Forecast",
    y = "Mean Forecast Percent Change",
    color = "CPI Item"
  ) +
  theme_minimal()
```