

Walmart Sale
ECON 5300- Self Project Report
Tien V Nguyen

Table of Contents

Contents

1.	Abbreviations	1
2.	Background and Motivation:	1
3.	Data	2
4.	Models.....	3
5.	Results.....	3
6.	Implications.....	5
7.	Limitations	5
8.	Sources	6
9.	AI Declaration.	6
10.	Technical Appendix.....	6

1. Abbreviations

Store	The store number
Date	The week of the sale in day-month-year format.
Weekly_Sales	The sales for the given store in the given week.
Holiday_Flag	Whether the week is a special holiday week (1) or not
Temperature	The average temperature in the region (in Fahrenheit)
Fuel_Price	The cost of fuel in the region (in dollars per gallon)
CPI	The consumer price index
Unemployment:	The unemployment rate in the region

Table 1: Data terminology explanation.

2. Background and Motivation

Understanding the impact of temperature variation on Walmart's weekly sales is an important empirical question. To investigate this relationship, data were collected from 45 Walmart stores covering the period from February 2010 through October 2012.

In addition to weekly sales and average temperature, several other variables were included to better control for external factors that may influence sales. These variables include the regional unemployment rate, the consumer price index (CPI), the week of sales (recorded in dd-mm-yyyy format), and the store identification number.

3. Data

The dataset consists of weekly sales from 45 Walmart stores spanning February 2010 through October 2012, totaling 6,435 observations.

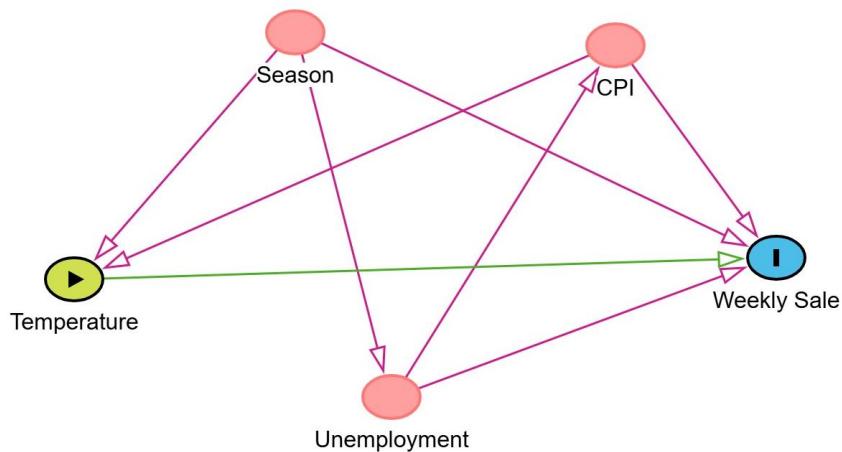


Figure 1: Model visualization.

Based on the collection dates, the data were grouped by season. According to U.S. meteorological classifications, January to March is defined as Winter, April to June as Spring, July to September as Summer, and October to December as Fall.

Seasonality was expected to influence both holiday frequency and unemployment. Since different seasons contain varying numbers of holidays that may affect sales patterns, the Holiday variable was excluded under the assumption that seasonal indicators could capture its effect. Although unemployment may fluctuate across seasons ⁽¹⁾, particularly during spring and summer hiring periods, it was retained in the model because it is expected to have independent effects on sales and broader economic conditions, including its association with CPI.

Fuel Price was also excluded due to its strong positive relationship with the Consumer Price Index (CPI) ⁽²⁾. As higher inflation is typically associated with increased fuel prices, including CPI in the model may sufficiently capture the variation explained by fuel prices, thereby reducing concerns of multicollinearity.

The relationships among the variables were structured under the following assumptions. CPI may influence weekly sales through its impact on consumer purchasing power, as higher inflation can alter household spending behavior, including grocery expenditures. CPI is also closely related to fuel prices. Increases in fuel prices, whether driven by market conditions or energy-related policies, may reduce fuel consumption and transportation activity. In the long run, reduced fuel

consumption could contribute to lower greenhouse gas emissions, which are associated with temperature patterns. While this channel is likely to operate gradually rather than immediately, it provides a theoretical basis for considering the indirect linkage between fuel prices, broader economic conditions, and temperature variation.

Additionally, seasonal factors directly affect temperature and labor market conditions, influencing unemployment rates. Higher unemployment may reduce aggregate consumer spending, thereby lowering weekly sales.

Lastly, the correlation matrix was run between all the chosen variables to ensure there was no multicollinearity among the independent variables. (Table 3)

4. Models

This report estimates two regression models.

In Model 1, a fixed-effects regression including store effects is estimated to examine the relationship between temperature and weekly sales without additional control variables. The objective is to assess the direct effect of temperature on weekly sales, independent of other store characteristics. Based on multiple graphical analyses exploring different functional forms, the logarithmic transformation of weekly sales appears to provide a better fit than the alternative specifications (Figure 2).

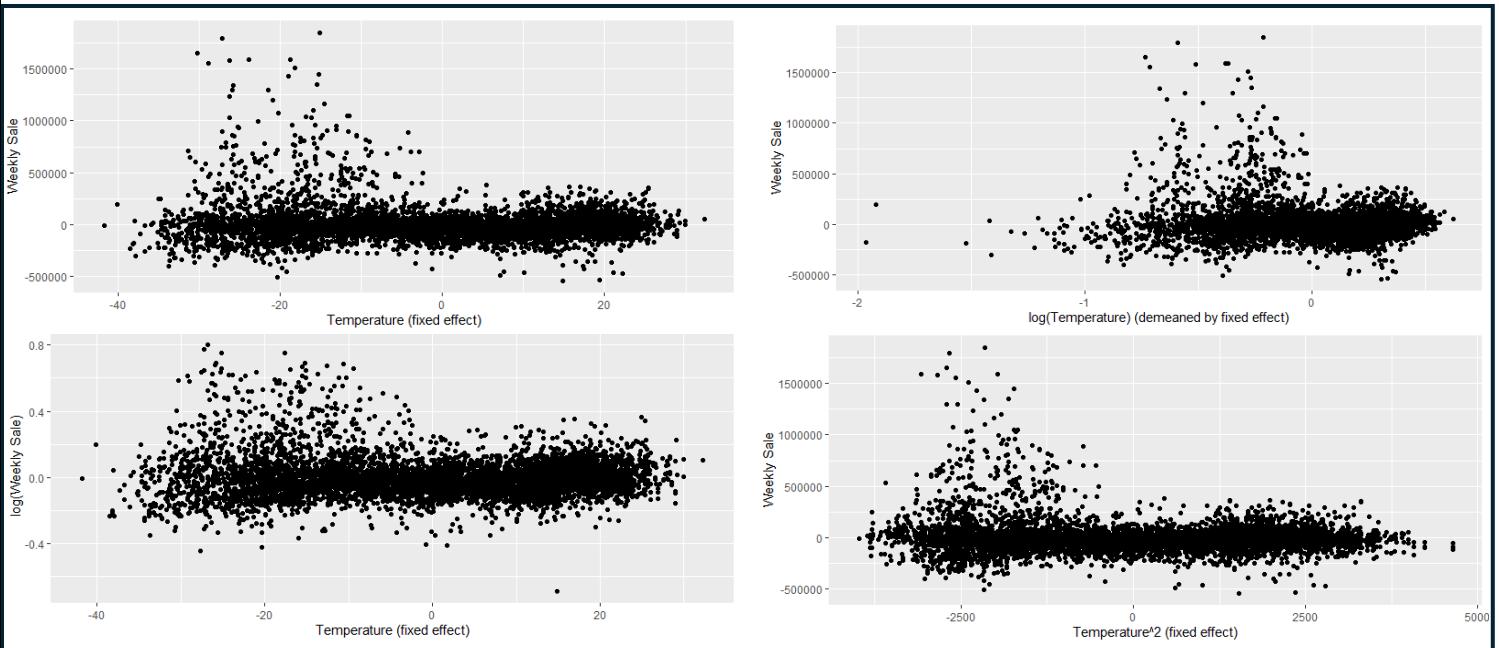


Figure 2: Four graphical representations are shown to compare alternative functional forms and determine which specification best fits the relationship between temperature and weekly sales.

In Model 2, the logarithmic transformation of weekly sales is retained. In addition to temperature, other control variables are included to examine how these characteristics contribute to differences in sales within each store (Figure 3).

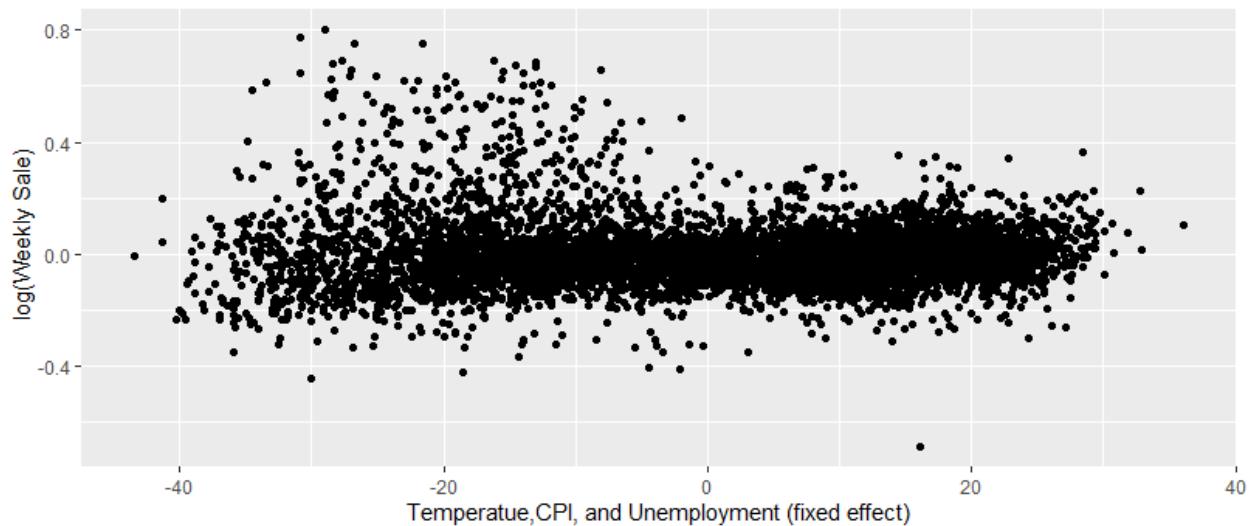


Figure 3: Scatter plots showing the within-store relationship between weekly sales and each additional explanatory variable included in the model.

5. Results

Parameter	Model 1	Model 2
Dependent Variable	Log (Weekly Sale)	
Temperature	-0.0005* (0.0002)	-0.0011*** (0.0002)
Season - Spring		-0.0397*** (0.0074)
Season - Summer		-0.0435*** (0.0055)
Season - Winter		-0.0977*** (0.0113)
CPI		-0.0005 (0.0022)
Unemployment		-0.0166 (0.0103)
S.E: Clustered	By: Store	
R ²	0.95726	0.96043

Within R ²	0.00375	0.07760
Observation		6435
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1		

Table 2: Regression Model Results.

6. Implications

The results from both models consistently indicate that higher temperatures are associated with lower weekly sales across the 6,435 observations.

In Model 1, since the coefficient is small in magnitude, it can be interpreted as an approximate percentage change. Within a store, a unit increase in temperature is associated with an average decrease of 0.05% in weekly sales. This relationship is statistically significant at the 5% level. (Otherwise, it can be interpreted as 0.01 unit increase in temperature is associated with 0.05% decrease in Weekly Sales)

In Model 2, the coefficient in temperature remains negative and statistically significant at the 1% level. Since the dependent variable is in logarithmic form and the coefficient is small in magnitude, the estimate can be interpreted approximately as a percentage change. So, within a store, a one-unit increase in temperature is associated with a 0.11% decrease in weekly sales, holding seasonal indicators, CPI, unemployment constant (or within a store, 0.01 unit increase in temperature is associated with 0.0011% decrease in weekly sales).

Seasonal indicators also show significant differences relative to the omitted category (Fall). Within a store, average weekly sales are lower by roughly 3.97% in Spring, 4.35% in Summer, and 9.77% in Winter compared to Fall. This relationship is statistically significant at the 1% level.

CPI and unemployment both have negative estimated coefficients in Model 2; however, neither coefficient is statistically significant at conventional levels. This suggests that while the estimated direction aligns with theoretical expectations, there is insufficient statistical evidence to confirm their independent effects on weekly sales in this specification.

Lastly, the R² and within R² values increase from 0.9572 to 0.9604 and from 0.00375 to 0.0776, respectively, from Model 1 to Model 2. This indicates that adding additional control variables improves the explanatory power of the model, particularly in terms of within-store variation in weekly sales.

Overall, the findings support the hypothesis that temperature negatively affects weekly sales. As temperature increases, weekly sales decline, and this result remains statistically significant after controlling for additional covariates.

7. Limitations

The survey covers only a two-year period and includes a limited number of explanatory variables. Important factors such as store size, local weather conditions, and population characteristics were not available, which may limit the explanatory power of the model.

8. Sources

1. Fiveable. (2024, July 31). Seasonal Unemployment – AP Macroeconomics.
<https://fiveable.me/key-terms/ap-macro/seasonal-unemployment>.
2. Kpodar, K., & Liu, B. (2022). *The distributional implications of the impact of fuel price increases on inflation*. *Energy Economics*, 108, 105909.
<https://doi.org/10.1016/j.eneco.2022.105909>

9. AI Declaration

ChatGPT assisted but no others than with refining the written content and troubleshooting coding errors encountered during the analysis.

10. Technical Appendix

Table 3: Correlation Matrix:

	Weekly Sales	Temperature	CPI	Unemployment
Weekly Sales	1.00	-0.06	-0.07	-0.11
Temperature	-0.06	1.00	0.18	0.10
CPI	-0.07	0.18	1.00	-0.30
Unemployment	-0.11	0.10	-0.30	1.00

Model 1: Information

OLS estimation, Dep. Var.: log(Weekly_Sales)

Observations: 6,435

Fixed-effects: Store: 45

Standard-errors: Clustered (Store)

	Estimate	Std. Error	t value	Pr(> t)
Temperature	-0.000475	0.000216	-2.20041	0.033073 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
RMSE: 0.121612 Adj. R2: 0.95696		Within R2: 0.003748		

Model 2: Information

OLS estimation, Dep. Var.: log(Weekly_Sales)

Observations: 6,435

Fixed-effects: Store: 45

Standard-errors: Clustered (Store)

	Estimate	Std. Error	t value	Pr(> t)
seasonSpring	-0.0397	0.0074	-5.343	3.08e-06 ***
seasonSummer	-0.043	0.005	-7.95	4.77e-10 ***
seasonWinter	-0.098	0.011	-8.668	4.518e-11 ***
Temperature	-0.001	0.0002	-4.707	2.518e-05 ***
CPI	-0.0005	0.002	-0.218	8.28e-01
Unemployment	-0.017	0.01	-1.606	1.155e-01
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
RMSE: 0.117017 Adj. R2: 0.96012		Within R2: 0.077604		

