Estimating Annual Pharmaceutical Drug Expenditure in the United States

Of all first-world countries, the United States takes the lead on pharmaceutical drug spending each year. Expenditure only seems to increase each year, making the pharmaceutical industry one of the most high-valued industries in the country. To estimate annual expenditure and explore the variables that may be directly associated with this variable, I decided to use annual data in the United States pertaining to the response variable and possible predictors.

For this analysis, I only used a sample of 30 observations primarily because of how difficult it is to find considerable, consecutive data prior to the 90s for my selected variables. For my response variable, I used the **Per Capita Pharmaceutical Drug Expenditure** from 1990-2020 (data obtained from the Organization for Economic Co-operation and Development). To adjust for inflation, I used constant 2020 dollars. For each year of observation, I used the following predictor variables: (i) **Per Capita Out of Pocket Health Expenditure** (adjusted, constant 2020); (ii) **Percent Share of Old Age Population** (Americans aged 65+); and (ii) **Total Opioid Prescriptions Dispensed by U.S. Retail Pharmacies** (in millions). Per Capita Out of Pocket Health Expenditure data was acquired from the Centers for Medicare & Medicaid Service National Health Expenditure data; the Percent Share of Old Age Population data was acquired from the Organization for Economic Co-operation and Development; and the data for Total Opioid Prescriptions Dispensed by U.S. Retail Pharmacies was obtained from the CDC and the National Institute on Drug Abuse.

Per Capita Out of Pocket Health Expenditure refers to how much the average person is spending on health-related commodities that is not being covered for by insurance or other programs that may absorb some medical costs. The percentage of how much of prescription drugs are covered by insurance programs has increased over time, meaning that drugs have essentially become more affordable. However, out of pocket health expenditure have only increased over time. This can either mean that the average consumer is purchasing more [health] related products, including prescription drugs, or that the average price of these drugs has increased. Thus, it would be reasonable to believe that as the Per Capita Out of Pocket Health Expenditure increases, as will the Per Capita Pharmaceutical Drug Expenditure due what seems like an obvious association (although, there may be no association at all).

I used the percentage of Americans aged 65+ as a predictor because to me, it seemed to make sense that as the population gets older, so will the frequency of health-related illnesses/diseases. With an increase in the amount of people getting diagnosed with medical conditions, including chronic illnesses and other diseases that are more common in people of older age, there may also be an increase in how many people are purchasing pharmaceutical drugs. An increase in demand would either increase the price, or perhaps decrease the price and increase the quantity being sold; regardless of which is the case and depending on the magnitude of the effects, both outcomes would have the potential to increase pharmaceutical drug expenditure.

Lastly, I used the total umber of Opioid Prescriptions Dispensed by U.S. Retail Pharmacies as somewhat of a way to measure an increase in demand for pharmaceutical drugs; of course, depending on what effect an aging population has over the demand and/or supply of pharmaceutical drugs, there may be some multicollinear between these two variables. There is no way of being certain though, especially since the most common type of drugs prescribed to older adults (60+) are lipid-lowering drugs, with only 4-9% using prescribed opioid drugs for chronic pain relief. So, the total amount of opioid prescriptions can potentially explain an increase in the need for these drugs for a younger demographic, which may drive expenditure, making it possible for there to be an association between this variable and Per Capita Pharmaceutical Drug Expenditure.

Prior to running the regression, I looked to see if any of the data needed to be transformed to assume a more "physically" correct shape. However, everything looked fine. The regression yielded the following output:

## Regression Equation

```
Per Capita Pharm. Drug Spending = -1574 + 1.647 Out-of-Pocket Health Expend.
                                + 47.4 Old Age Population (65+)
                                + 0.987 Opioid Prescriptions Dispensed
```

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | -1574 | 118 | -13.31 | 0.000 | |
| Out-of-Pocket Health Expend. | 1.647 | 0.177 | 9.31 | 0.000 | 5.08 |
| Old Age Population (65+) | 47.4 | 11.1 | 4.26 | 0.000 | 1.88 |
| Opioid Prescriptions Dispensed | 0.987 | 0.329 | 3.00 | 0.006 | 3.57 |

## Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 60.4211 | 96.98% | 96.65% | 96.03% |

## Analysis of Variance

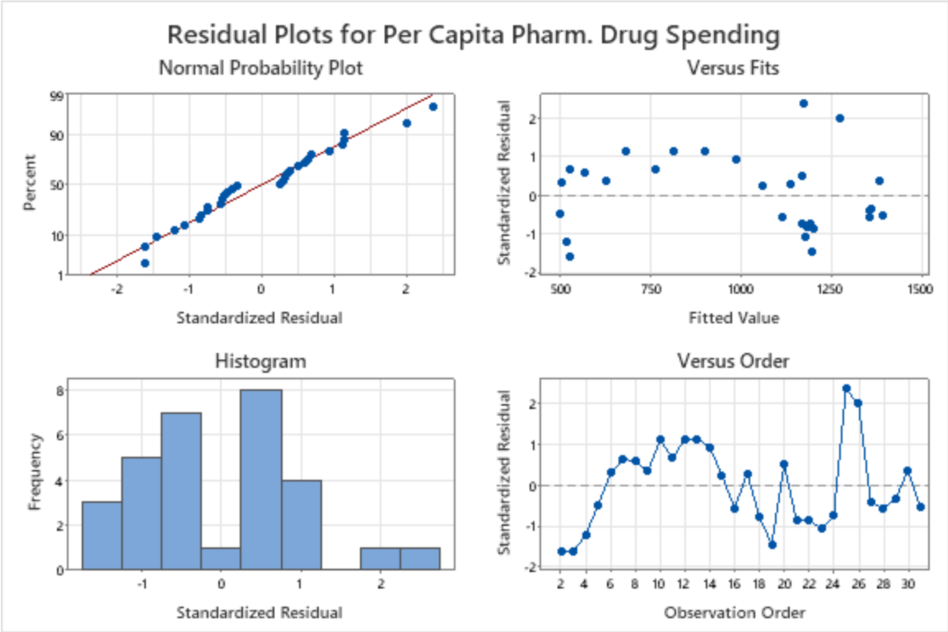
| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 3 | 3167315 | 1055772 | 289.20 | 0.000 |
| Out-of-Pocket Health Expend. | 1 | 316758 | 316758 | 86.77 | 0.000 |
| Old Age Population (65+) | 1 | 66387 | 66387 | 18.18 | 0.000 |
| Opioid Prescriptions Dispensed | 1 | 32878 | 32878 | 9.01 | 0.006 |
| Error | 27 | 98569 | 3651 | | |
| Total | 30 | 3265884 | | | |

Each predictor is very significant, with all P-values being less than $\alpha = 0.05$. The $R^2$ value for this model is also considerably high at 96.98%. While this model fits the data quite well, it can possibly be simplified, but also, it does not address the time element of this time-series model; furthermore, we can detect some autocorrelation.

## Durbin-Watson Statistic

```
Durbin-Watson Statistic = 0.460726
```

The Durbin-Watson statistics of 0.460726 is much less than its lower critical value of 1.006, indicating strong positive autocorrelation. We can also observe this autocorrelation when looking at the time series plot, which includes observation order vs standardized residuals.




At the $\alpha = 0.05$ level, we reject fail to reject the null hypothesis in favor of the alternative, which indicates that there is no autocorrelation in the data.

## Descriptive Statistics

| | | Number of Observations | |
|---|---|---|---|
| **N** | **K** | **≤ K** | **> K** |
| 30 | -0.0091723 | 15 | 15 |

*K = sample mean*

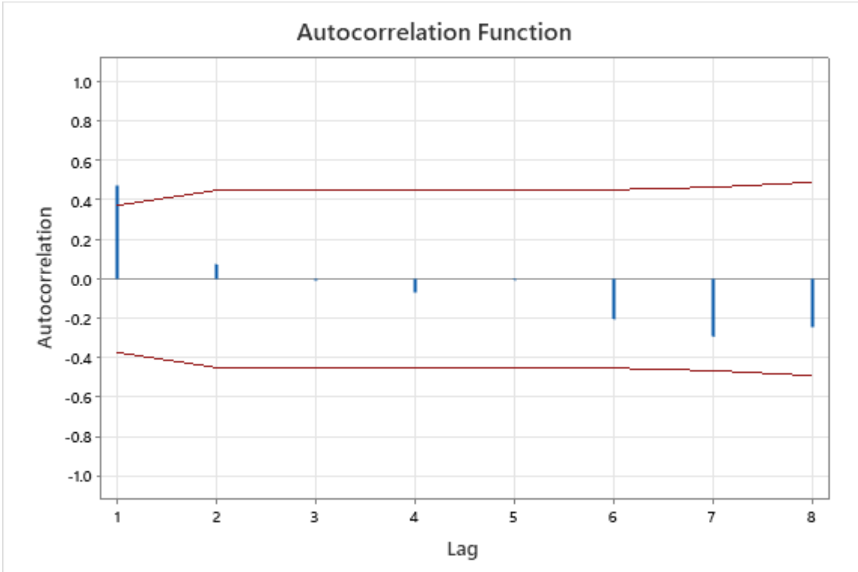
## Test

Null hypothesis $H_0$: The order of the data is random

Alternative hypothesis $H_1$: The order of the data is not random

| Number of Runs | | |
|---|---|---|
| **Observed** | **Expected** | **P-Value** |
| 11 | 16.00 | 0.063 |

However, when looking at the AFC plot, we observe some autocorrelation.




So, the present autocorrelation is certainly something that needs to be addressed. Thus, my first approach will be to lag the response variable and use that as a predictor; "**LaggedSpending**." I also decided to add in two more variables despite my initial mentioning of simplifying the model. It made some sense to me that perhaps the lagged version of Total Opioid Prescriptions Dispensed by U.S. Retail Pharmacies and Per Capita Out of Pocket Health Expenditure could have some influence on Per Capita Pharmaceutical Drug Expenditure. First, the number of opioid prescriptions written in a previous year could be some sort of indicator of the number of prescriptions that will be refilled in the following year, ultimately increasing expenditure. It may also provide some insight on the health status of the population. I called this variable "**LaggedPrescriptions**." I utilized the lag of Per Capita Out of Pocket Health Expenditure since if pharmaceutical drug prices are higher in previous years, which indicates a possible increasing demand/purchase of these drugs, it would make sense that that demand would only follow in and perhaps increase into the next month. I called this variable "**LaggedOutofPocket**."

The regression, Runs test, and AFC plot for this model look as follows.

## Regression Equation

```
Per Capita Pharm. Drug Spending = -163 + 0.930 Out-of-Pocket Health Expend.
                                  + 9.1 Old Age Population (65+)
                                  + 0.553 Opioid Prescriptions Dispensed
                                  + 0.897 LaggedSpending - 0.783 LaggedOutofPocket
                                  - 0.525 LaggedPrescriptions
```

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | -163 | 233 | -0.70 | 0.493 | |
| Out-of-Pocket Health Expend. | 0.930 | 0.263 | 3.54 | 0.002 | 42.70 |
| Old Age Population (65+) | 9.1 | 11.8 | 0.77 | 0.447 | 8.20 |
| Opioid Prescriptions Dispensed | 0.553 | 0.807 | 0.68 | 0.500 | 78.17 |
| LaggedSpending | 0.897 | 0.128 | 7.00 | 0.000 | 54.52 |
| LaggedOutofPocket | -0.783 | 0.223 | -3.50 | 0.002 | 30.98 |
| LaggedPrescriptions | -0.525 | 0.751 | -0.70 | 0.491 | 72.61 |

## Model Summary

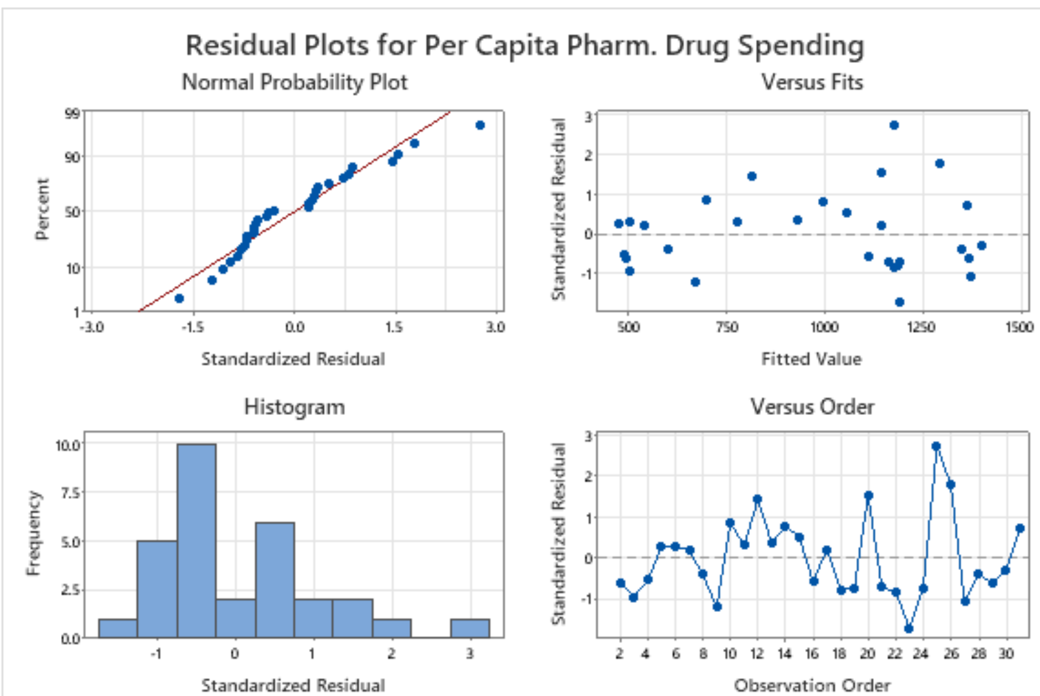
| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 30.5080 | 99.29% | 99.10% | 98.85% |

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 6 | 2991798 | 498633 | 535.74 | 0.000 |
| Out-of-Pocket Health Expend. | 1 | 11666 | 11666 | 12.53 | 0.002 |
| Old Age Population (65+) | 1 | 556 | 556 | 0.60 | 0.447 |
| Opioid Prescriptions Dispensed | 1 | 436 | 436 | 0.47 | 0.500 |
| LaggedSpending | 1 | 45580 | 45580 | 48.97 | 0.000 |
| LaggedOutofPocket | 1 | 11419 | 11419 | 12.27 | 0.002 |
| LaggedPrescriptions | 1 | 455 | 455 | 0.49 | 0.491 |
| Error | 23 | 21407 | 931 | | |
| Total | 29 | 3013205 | | | |

## Durbin-Watson Statistic

Durbin-Watson Statistic = 1.62214




## Descriptive Statistics

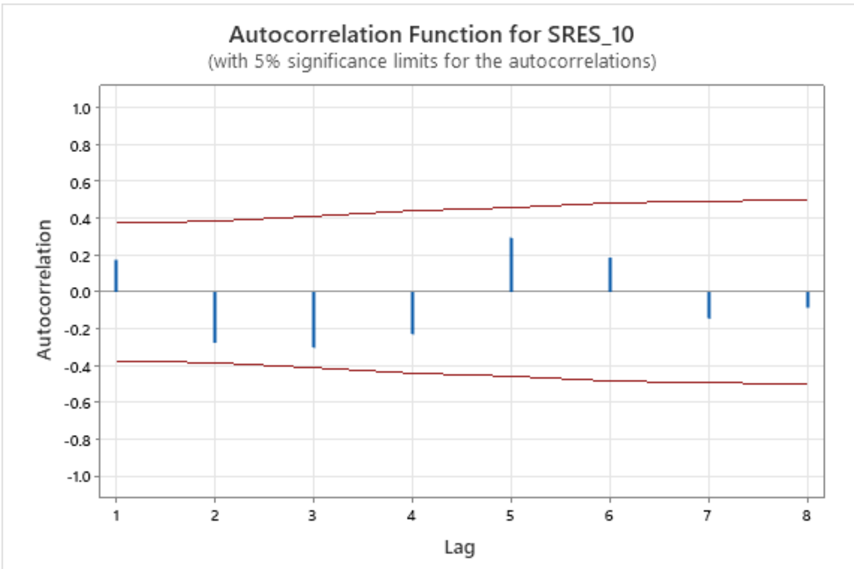
| | | Number of Observations | |
|---|---|---|---|
| N | K | ≤ K | > K |
| 30 | 0.0031289 | 16 | 14 |

*K = sample mean*

## Test

| | | |
|---|---|---|
| Null hypothesis | $H_0$: The order of the data is random | |
| Alternative hypothesis | $H_1$: The order of the data is not random | |

| Number of Runs | | |
|---|---|---|
| **Observed** | **Expected** | **P-Value** |
| 12 | 15.93 | 0.142 |




The time series plot using standardized residuals is a big improvement from the original plot, and we once again fail to reject to the null hypothesis for the Runs test, which further indicates no autocorrelation. Finally, the AFC plot looks very nice and does not depict any autocorrelation.

There are, however, some flaws with this model, and some indication that it is not the best fitted one. First, not all the predictor variables are statistically significant, with LaggedPrescriptions, Total Opioid Prescriptions Dispensed by U.S. Retail Pharmacies, and the Old Age Population Percent having P-values much greater than the alpha = 0.05 level. Some of the VIFs are also considerably high, butt the maximum of max {10, $1/1\text{-}R^2$} is 1/ (1-0.9929), which is equal to 140.85. None of the VIFs exceed this value, but they are high, nonetheless. There also seems to be two noticeably unusual observations in the time series plot—observations 23 and 25. However, before addressing these, I will first try to simplify our model.

## Best Subsets Regression: Per Capita Pharm. Drug Spending versus Out-of-Pocket Health Expend., Old Age Population (65+), Opioid Prescriptions Dispensed, LaggedOutofPocket, LaggedPrescriptions, LaggedSpending

| Vars | R-Sq | R-Sq (adj) | PRESS | R-Sq (pred) | Mallows Cp | S | AICc | BIC | Cond No |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 98.7 | 98.7 | 42740.8 | 98.6 | 14.5 | 36.690 | 306.140 | 309.420 | 1.000 |
| 1 | 95.7 | 95.6 | 151651.3 | 95.0 | 112.2 | 67.773 | 342.960 | 346.240 | 1.000 |
| 2 | 98.9 | 98.8 | 39594.2 | 98.7 | 11.4 | 34.919 | 304.757 | 308.762 | 81.847 |
| 2 | 98.8 | 98.7 | 41592.8 | 98.6 | 14.2 | 36.268 | 307.032 | 311.037 | 78.497 |
| 3 | 99.3 | 99.2 | 27327.5 | 99.1 | 1.8 | 29.170 | 295.732 | 300.238 | 140.005 |
| 3 | 98.9 | 98.8 | 42583.9 | 98.6 | 13.3 | 35.533 | 307.572 | 312.078 | 118.388 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | 99.3 | 99.2 | 28936.4 | 99.0 | 3.5 | 29.573 | 298.530 | 303.285 | 222.872 |
| 4 | 99.3 | 99.2 | 30402.8 | 99.0 | 3.6 | 29.653 | 298.692 | 303.447 | 177.176 |
| 5 | 99.3 | 99.1 | 34363.2 | 98.9 | 5.5 | 30.168 | 301.941 | 306.658 | 330.280 |
| 5 | 99.3 | 99.1 | 34707.1 | 98.8 | 5.5 | 30.182 | 301.967 | 306.684 | 340.519 |
| 6 | 99.3 | 99.1 | 34533.8 | 98.9 | 7.0 | 30.508 | 305.102 | 309.454 | 780.238 |

When we run best subsets, the best-looking model is one with just three variables. This three-variable model only uses LaggedSpending, LaggedOutofPocket, and Per Capita Out of Pocket Health Expenditure as predicting variables. It has the lowest Mallows' *Cp* of 1.8, minimizes the $AIC_C$ (295.732), and has an $R^2$-adjusted value of 99.2; two four variable models (Out of Pocket Health Expenditure, Old Age Population Share, LaggedOutofPocket, and LaggedSpending; Out of Pocket Health Expenditure, LaggedOutofPocket, LaggedPrescription, and LaggedSpending) also yield the same $R^2$-adjusted value, but their $AIC_C$ values and Mallows' *Cp* are lower, even though not by much. Worth noting, 10 out of 11 of these subsets included LaggedSpending as a predictor variable.

Simplifying our model to only include three variables, we run the regression and yield the following outcome.

## Regression Equation

```
Per Capita Pharm. Drug Spending = -26 + 0.846 Out-of-Pocket Health Expend.
                                  - 0.744 LaggedOutofPocket + 0.9385 LaggedSpending
```

## Coefficients

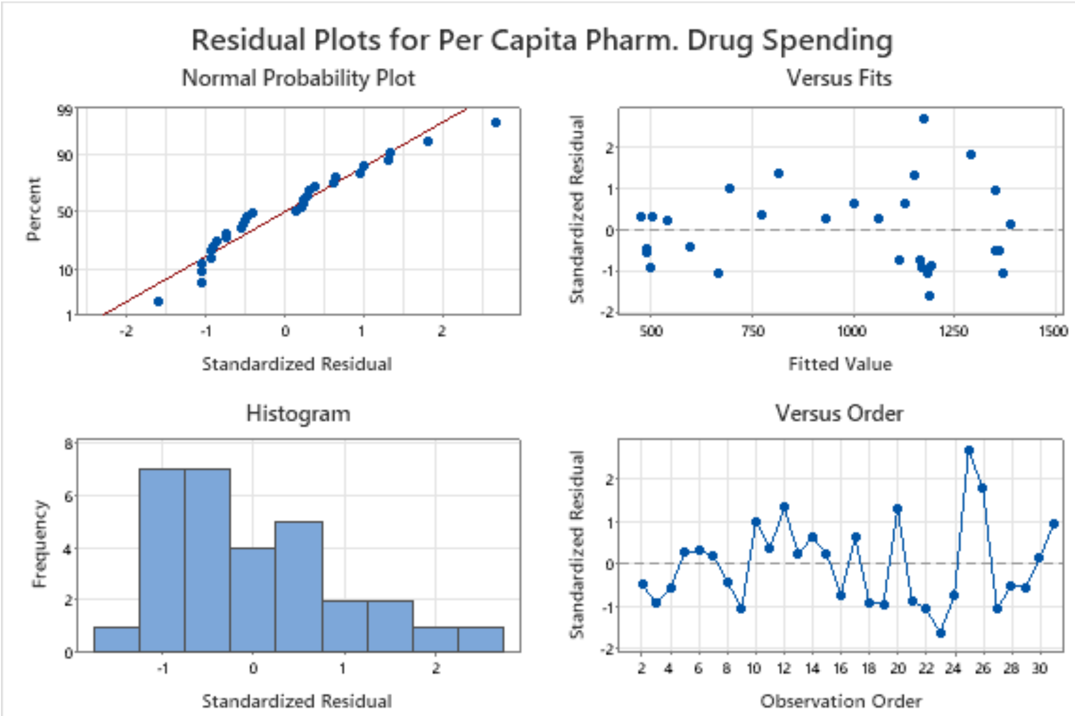
| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | -26 | 130 | -0.20 | 0.842 | |
| Out-of-Pocket Health Expend. | 0.846 | 0.213 | 3.97 | 0.001 | 30.82 |
| LaggedOutofPocket | -0.744 | 0.209 | -3.56 | 0.001 | 29.59 |
| LaggedSpending | 0.9385 | 0.0846 | 11.09 | 0.000 | 25.98 |

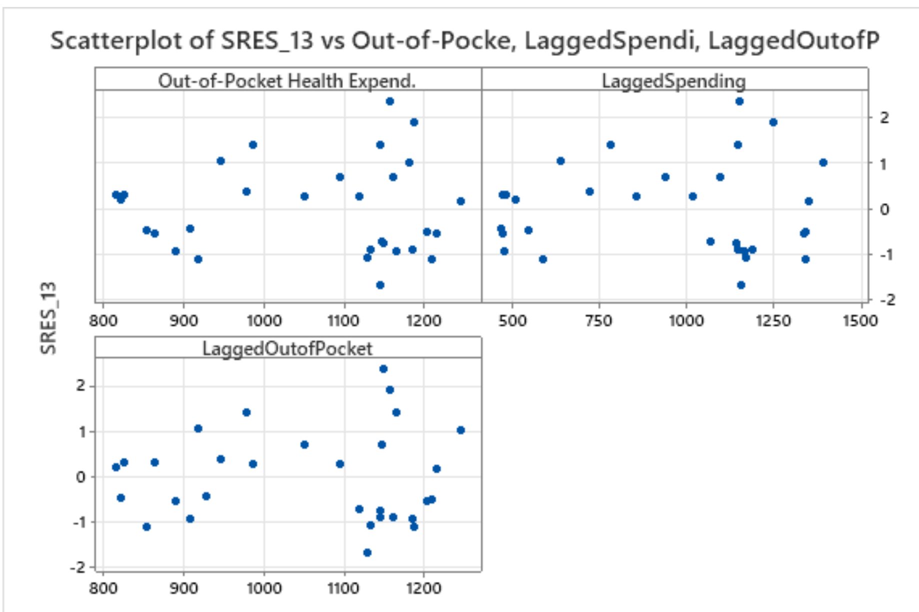
## Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 29.1702 | 99.27% | 99.18% | 99.09% |

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 3 | 2991081 | 997027 | 1171.73 | 0.000 |
| Out-of-Pocket Health Expend. | 1 | 13392 | 13392 | 15.74 | 0.001 |
| LaggedOutofPocket | 1 | 10799 | 10799 | 12.69 | 0.001 |
| LaggedSpending | 1 | 104676 | 104676 | 123.02 | 0.000 |
| Error | 26 | 22123 | 851 | | |
| Total | 29 | 3013205 | | | |




Each predictor variable is statistically significant, the VIF values are all lower than 1/ (1-0.9927), which is 136.99. The $R^2$ is 99.27% which indicates a very strong association between Per Capita Pharmaceutical Drug Expenditure and Per Capita Out of Pocket Health Expenditure, LaggedOutofPocket, and LaggedSpending. The p-value for each predictor variable is less than $\alpha = 0.05$, so all variables are statistically significant.

Here are the Runs Test and AFC Plot.

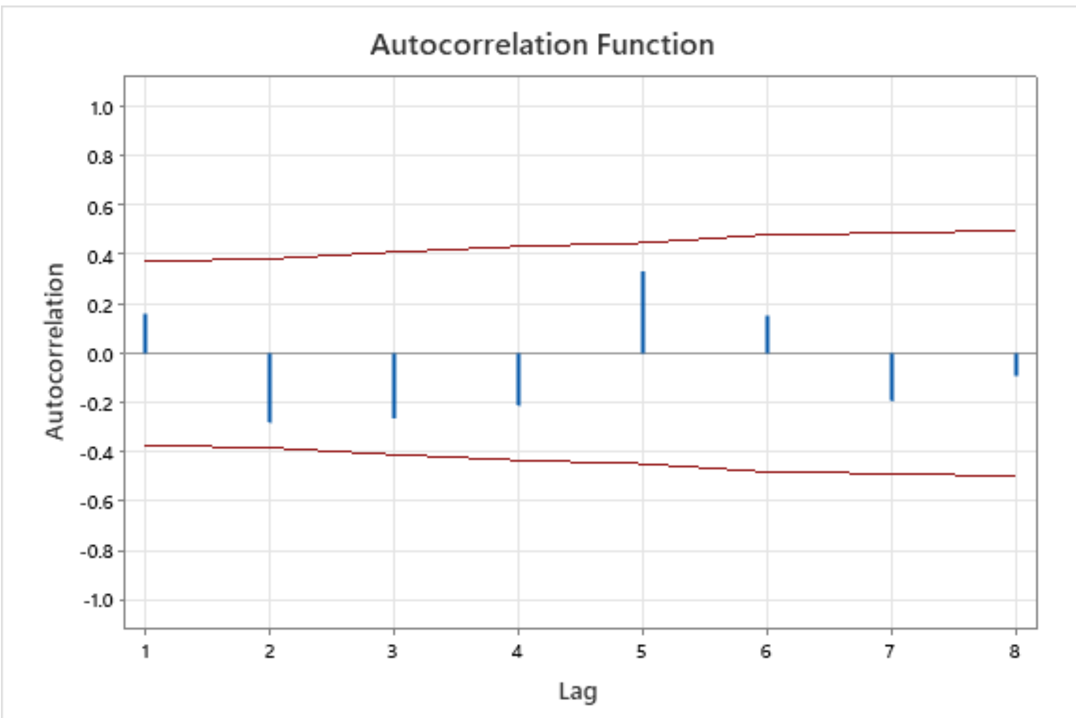
| | | Number of Observations | |
|---|---|---|---|
| N | K | ≤ K | > K |
| 30 | -0.0034058 | 15 | 15 |

*K = sample mean*

## Test

| | |
|---|---|
| Null hypothesis | $H_0$: The order of the data is random |
| Alternative hypothesis | $H_1$: The order of the data is not random |

| Number of Runs Observed | Expected | P-Value |
|---|---|---|
| 12 | 16.00 | 0.137 |




Based on the Runs Test and the AFC Plot, we do not have any autocorrelation that reappeared after taking the best subset, which is a positive thing. Using the lag of the predictor variable seemed to effectively resolve the autocorrelation presence, and because we kept that variable in this simplified model, this is likely why it has not reappeared.

This model tells us that, holding the previous year's Per Capita Out of Pocket Health Expenditure and Per Capita Out-of-Pocket Health Expenditure constant, a one dollar (constant 2020 dollars) increase in Per Capita Out-of-Pocket Health Expenditure is associated with a 0.846 increase in Per Capita Pharmaceutical Drug Expenditure (also constant 2020 dollars) for the current year. Holding all variables fixed, except for the previous year's Per Capita Out of Pocket Health Expenditure (LaggedOutofPocket), a one dollar increase (constant 2020 dollars) in the previous year's Per Capita Out of Pocket Health Expenditure is associated with a 0.744 decrease in Per Capita Pharmaceutical Drug Expenditure (constant 2020 dollars) for the current year. Holding all else fixed, except for the previous year's Per Capita Pharmaceutical Drug Expenditure (LaggedSpending), a one dollar (constant 2020 dollars) increase in the previous years' Per Capita Pharmaceutical Drug Expenditure is associated with a 0.9385 increase in Per Capita Pharmaceutical Drug Expenditure (constant 2020 dollars) for the current year.

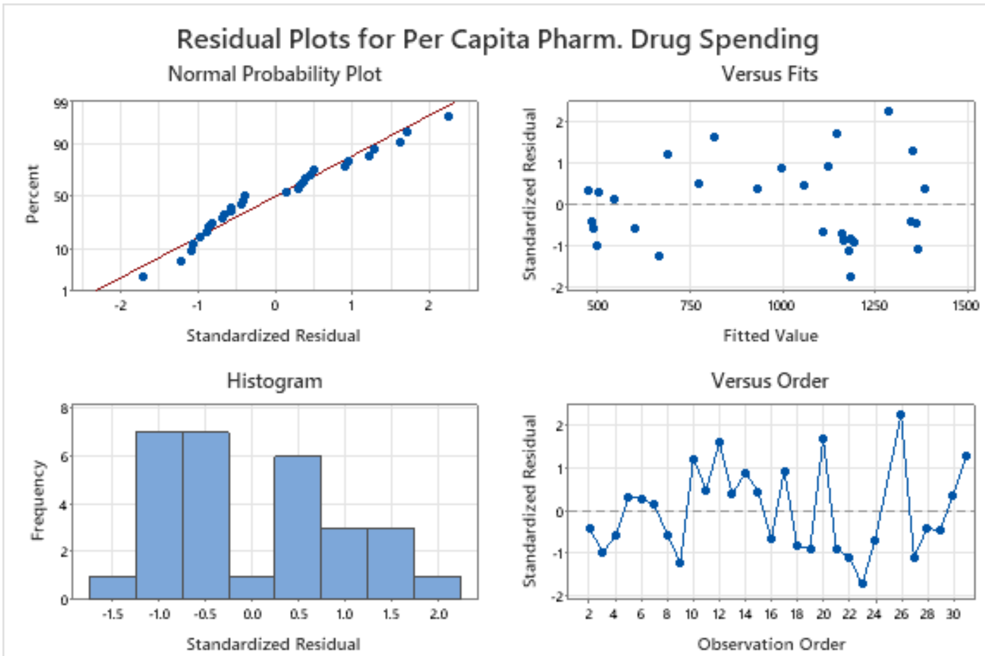
However, this model is still not perfect. Looking at the time series plot, there are some potentially unusual observations: 23, 25, and 26, which refer to the years of 2012, 2014, and 2015. When looking at the leverage values and Cook's Distances for these observations, there is nothing that unusual about the years of 2012 and 2015; these observations have values that are very consistent with the rest of the data. However, 2014 does raise some concern. The Cook's Distance for this observation is 0.108. This value is nowhere close to 1, but it is much higher than most of the other Cook's Distance values.

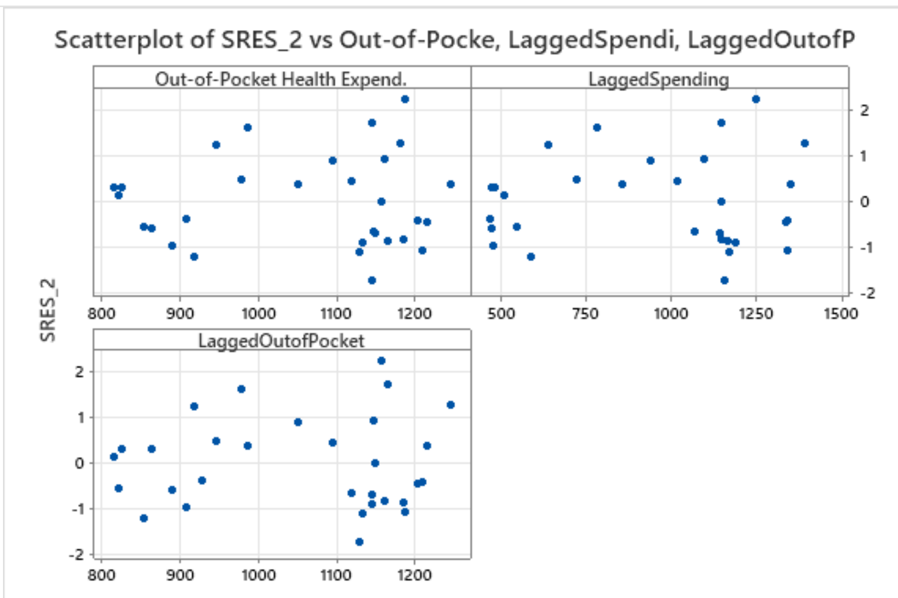
| Year | r Capita Pharm. Drug Spendi | it-of-Pocket Health Expend | LaggedSpending | LaggedOutofPocket | SRES | HI | COOK |
|---|---|---|---|---|---|---|---|
| 1990 | 466.4 | 928.07 | * | * | * | * | * |
| 1991 | 477.92 | 907.75 | 466.4 | 928.07 | -0.49292 | 0.323703 | 0.029074 |
| 1992 | 474.9 | 888.65 | 477.92 | 907.75 | -0.93695 | 0.203663 | 0.056129 |
| 1993 | 474.9 | 863.91 | 474.9 | 888.65 | -0.55641 | 0.16505 | 0.0153 |
| 1994 | 483.13 | 825.7 | 474.9 | 863.91 | 0.27868 | 0.20322 | 0.004952 |
| 1995 | 511.08 | 815.23 | 483.13 | 825.7 | 0.31597 | 0.193957 | 0.006006 |
| 1996 | 546.01 | 819.72 | 511.08 | 815.23 | 0.21946 | 0.234893 | 0.003697 |
| 1997 | 586.75 | 852.02 | 546.01 | 819.72 | -0.39549 | 0.227007 | 0.011483 |
| 1998 | 639.09 | 916.44 | 586.75 | 852.02 | -1.02198 | 0.220247 | 0.073753 |
| 1999 | 720.64 | 946.03 | 639.09 | 916.44 | 1.01055 | 0.083623 | 0.023298 |
| 2000 | 784.08 | 977.55 | 720.64 | 946.03 | 0.37719 | 0.071951 | 0.002758 |
| 2001 | 854.55 | 985.13 | 784.08 | 977.55 | 1.34038 | 0.043917 | 0.020632 |
| 2002 | 939.04 | 1051.18 | 854.55 | 985.13 | 0.26196 | 0.172529 | 0.003577 |
| 2003 | 1018.1 | 1095.59 | 939.04 | 1051.18 | 0.64975 | 0.106808 | 0.012621 |
| 2004 | 1068.45 | 1119.43 | 1018.1 | 1095.59 | 0.24199 | 0.065506 | 0.001026 |
| 2005 | 1094.77 | 1147.93 | 1068.45 | 1119.43 | -0.7286 | 0.086431 | 0.012556 |
| 2006 | 1148 | 1162 | 1094.77 | 1147.93 | 0.61835 | 0.101346 | 0.01078 |
| 2007 | 1164.67 | 1185.02 | 1148 | 1162 | -0.89951 | 0.103375 | 0.023321 |
| 2008 | 1144.88 | 1164.43 | 1164.67 | 1185.02 | -0.94778 | 0.11266 | 0.028512 |
| 2009 | 1187.07 | 1144.28 | 1144.88 | 1164.43 | 1.29654 | 0.08758 | 0.040339 |
| 2010 | 1170.77 | 1132.73 | 1187.07 | 1144.28 | -0.87337 | 0.086036 | 0.017951 |
| 2011 | 1155.66 | 1128.66 | 1170.77 | 1132.73 | -1.0415 | 0.075079 | 0.022013 |
| 2012 | 1142.88 | 1146.03 | 1155.66 | 1128.66 | -1.59708 | 0.053005 | 0.035691 |
| 2013 | 1145.23 | 1148.93 | 1142.88 | 1146.03 | -0.73454 | 0.04997 | 0.007095 |
| 2014 | 1248.82 | 1156.84 | 1145.23 | 1151.43 | 2.72241 | 0.055156 | 0.108162 |
| 2015 | 1340.61 | 1187.74 | 1248.82 | 1156.84 | 1.80108 | 0.089188 | 0.079412 |
| 2016 | 1342.85 | 1209.58 | 1340.61 | 1187.74 | -1.04936 | 0.116238 | 0.036208 |
| 2017 | 1338.48 | 1203.03 | 1342.85 | 1209.58 | -0.50126 | 0.10718 | 0.007541 |
| 2018 | 1348.88 | 1215.65 | 1338.48 | 1203.03 | -0.52952 | 0.090512 | 0.006976 |
| 2019 | 1392.96 | 1245.33 | 1348.88 | 1215.65 | 0.13727 | 0.105454 | 0.000555 |
| 2020 | 1376.27 | 1180.94 | 1392.96 | 1245.33 | 0.92989 | 0.364718 | 0.124105 |

Looking closely at the year of 2014, it seems that this is an outlier. It has a relatively high Per Capita Pharmaceutical Drug Expenditure of 1248.82 compared to all the years' prior Per Capita Pharmaceutical Drug Expenditure and given the predictor variable values. Basically, what looks to be the case is that there was a major leap in Per Capita Pharmaceutical Drug Expenditure during 2014 that does not seem to follow the general pattern of all the previous years. This situation is tough, because although the Per Capita Pharmaceutical Drug Expenditure increased substantially from all the previous years, it only continued to increase even more. The difference between 2014 and the following years, though, is that even though expenditure jumped again in 2015, it continued to increase much more steadily up until 2020.

The increase in Per Capita Pharmaceutical Drug Expenditure comes as no surprise. Upon further investigation, it turns out that in 2014 there was a surge in the demand for new, expensive specialty drugs and treatments. Specifically, there was an increased demand for cancer and multiple sclerosis treatments, and even a major price increase in insulin products and other major products. One existing article even explains that because of Medicaid expansion, there was an increase in pharmaceutical prescriptions being filled since fractions of the cost were being covered by Medicaid programs.

I decided to use the accommodation approach for removing this outlier. I created the indicator value "Year2014," and assigned 2014 a "1" while all other observations were assigned "0." I ran the regression again.




Truthfully, removing this point does not seem to resolve the issue of the sudden spike in Per Capita Pharmaceutical Drug Expenditure, especially given the fact that the effects of this spike seemed to continue. Therefore observation 26, which is 2015, now seems to be unusual. Regardless of whether I choose to keep 2014 or remove it, I seem to encounter a problem regarding a leap in the expenditure no matter. This observation is also a bit more recent, so it is hard to tell if in the future, the Per Capita Pharmaceutical Drug

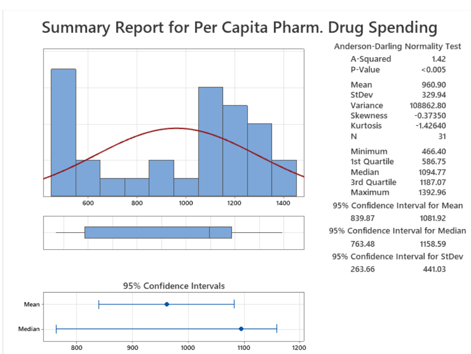
Expenditure will level off and once again resemble what it was before 2014, thus potentially making the years of 2014-2020 a pool of unusual observations.

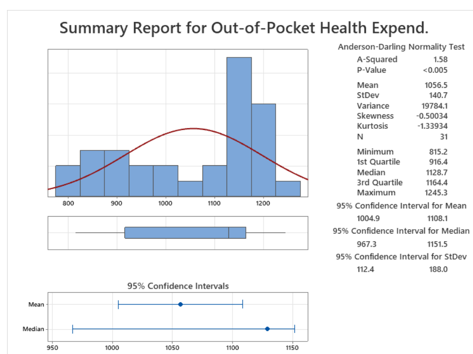
I decided to keep the indicator variable “Year2014” since the normal probability plot and standardized residuals plot looked a little bit better with the omission of this observation, although they certainly do no look perfect. The normal probability plot does not look too great, but it does not look to have any tails nor strong right skew. Additionally, the histogram does not look skewed, although it does have an awkward shape. The fitted values plot looks reasonable, with the plots bouncing randomly. The same can be said for the plots with residuals versus each predictor variable; this lack of pattern is a positive sign.

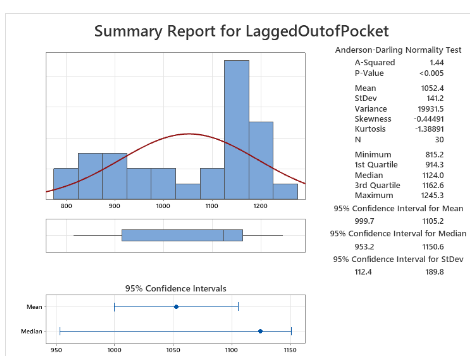
I ran best subsets and chose to continue using all the variables I had been using since this model yields the lowest Mallows’ *Cp* and $AIC_C$, and highest $R^2$-adjusted.

| Total | Vars | R-Sq | R-Sq (adj) | PRESS | R-Sq (pred) | Mallows Cp | S | AICc | BIC |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 99.0 | 98.9 | * | * | 25.1 | 34.154 | 303.428 | 307.432 |
| | 2 | 95.8 | 95.5 | * | * | 171.5 | 68.108 | 344.841 | 348.846 |
| | 3 | 99.1 | 99.0 | * | * | 20.9 | 32.524 | 302.262 | 306.768 |
| | 3 | 99.0 | 98.9 | * | * | 22.8 | 33.215 | 303.523 | 308.029 |
| | 4 | 99.5 | 99.4 | * | * | 5.0 | 25.314 | 289.201 | 293.956 |

Each variable in this model had a physically appropriate shape, with no alarming right skewedness and so forth. So, I did not transform the data and decided that this would be the best model I use for the analysis (unless autocorrelation once again appeared).



## Regression Equation

Per Capita Pharm. Drug Spending = 7 - 0.768 LaggedOutofPocket + 0.9530 LaggedSpending
+ 0.823 Out-of-Pocket Health Expend. + 82.2 Year2014

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 7 | 114 | 0.06 | 0.954 | |
| LaggedOutofPocket | -0.768 | 0.181 | -4.23 | 0.000 | 29.67 |
| LaggedSpending | 0.9530 | 0.0736 | 12.95 | 0.000 | 26.09 |
| Out-of-Pocket Health Expend. | 0.823 | 0.185 | 4.45 | 0.000 | 30.87 |
| Year2014 | 82.2 | 26.0 | 3.16 | 0.004 | 1.02 |

## Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 25.3141 | 99.47% | 99.38% | * |

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 4 | 2997185 | 749296 | 1169.31 | 0.000 |
| LaggedOutofPocket | 1 | 11483 | 11483 | 17.92 | 0.000 |
| LaggedSpending | 1 | 107493 | 107493 | 167.75 | 0.000 |
| Out-of-Pocket Health Expend. | 1 | 12664 | 12664 | 19.76 | 0.000 |
| Year2014 | 1 | 6387 | 6387 | 9.97 | 0.004 |
| Error | 25 | 16020 | 641 | | |
| Total | 29 | 3013205 | | | |

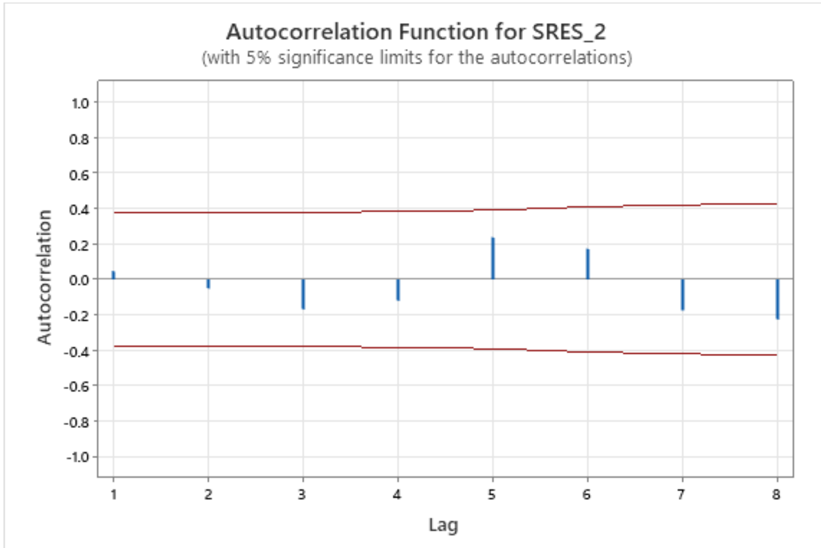
| | | Number of Observations | |
|---|---|---|---|
| N | K | ≤ K | > K |
| 30 | -0.0004559 | 15 | 15 |

*K = sample mean*

## Test

Null hypothesis $H_0$: The order of the data is random
Alternative hypothesis $H_1$: The order of the data is not random

| Number of Runs | | |
|---|---|---|
| Observed | Expected | P-Value |
| 12 | 16.00 | 0.137 |




Neither the Runs Test nor AFC Plot hint towards the presence of autocorrelation.
Each predictor variable has a VIF less than 188.68 which is a good sign regarding multicollinearity, and they are all statistically significant $\alpha$ alpha = 0.05 level.

This model has a very high $R^2$ of 99.47%, meaning there is a strong relationship; 99.47% of the variability in this model can be explained by the predictor variables. The F-value of 749296 is very large, and unsurprisingly greater than the critical value of 6.19 at the $\alpha$ = 0.05 significance level. There is convincing evidence of a relationship between Per Capita Pharmaceutical Drug Expenditure and the variables Per Capita Out of Pocket Health Expenditure, LaggedOutofPocket, LaggedSpending, and Year2014. After adjusting for variable selection (and using the approach that makes use of a separate worksheet and entering the original model's fitted values), a 95% hypothesis interval for Per Capita Pharmaceutical Drug Expenditure is ±117.2 (constant 2020 dollars).

The regression equation tells us that, holding the previous year's Per Capita Out of Pocket Health Expenditure and Per Capita Out-of-Pocket Health Expenditure constant, a one dollar (constant 2020 dollars) increase in Per Capita Out-of-Pocket Health Expenditure is associated with a 0.823 increase in Per Capita Pharmaceutical Drug Expenditure (also constant 2020 dollars) for the current year. Holding all variables fixed, except for the previous year's Per Capita Out of Pocket Health Expenditure (LaggedOutofPocket), a one dollar increase (constant 2020 dollars) in the previous year's Per Capita Out of Pocket Health Expenditure is associated with a 0.768 decrease in Per Capita Pharmaceutical Drug Expenditure (constant 2020 dollars) for the current year. Holding all else fixed, except for the previous year's Per Capita Pharmaceutical Drug Expenditure (LaggedSpending), a one dollar (constant 2020 dollars) increase in the previous years' Per Capita Pharmaceutical Drug Expenditure is associated with a 0.9530 increase in Per Capita Pharmaceutical Drug Expenditure (constant 2020 dollars) for the current year.

Given the previous years' Per Capita Pharmaceutical Drug Expenditure, previous years' Per Capita Out of Pocket Health Expenditure, and the current year's Per Capita Out of Pocket Health Expenditure, the current year's observed Per Capita Pharmaceutical Drug Expenditure for the year of 2014 was $82.2 (constant 2020 dollars) higher than expected, and this amount is significantly different from 0 ($p < 0.05$).

It seems reasonable for Per Capita Out-of-Pocket Health Expenditure for both the current year and previous year to drive Per Capita Pharmaceutical Drug Expenditure since they seem to reflect an increase in demand for health-related pharmaceutical, health products. As medicaid begins to cover the costs of more pharmaceutical drugs, it seems reasonable to attribute the growing Per Capita Out-of-Pocket Health expenditure to an increase in consumption; this could be a result of a variety of different forces. The Per Capita Pharmaceutical Drug Expenditure for a previous year also seems to be a strong variable since it reflects a pattern in consumer behavior. If demand is steadily increasing from year to year, it makes sense that it will continue to increase and could be used to predict by how much the expenditure will increase by.