MBA 547 Case Report, Homework 4

Topic: Consumer Spending and Income

Due Date: November 22, 2024

Submitted by: Ofuka Abung, Sadaf Vora, and Samyak Shah

Team Lead - Ofuka Abung, Data Analysts and Script writers - Sadaf Vora and Samyak Shah

Consumer Spending and Income

Executive Summary

This paper used data from the previous ten-year time series to investigate the relationship between U.S. Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI). We sought to determine how variations in people's available income influence their expenditure of money. We employed several statistical models to accomplish this: autocorrelation tests, finite distributed lag models, AR (1) models, cointegration tests, and error correction models (ECM).

We started with a rudimentary view of the data. The PCE and DPI values were somewhat dissimilar, with a mean PCE value of 4,905.65 and a mean DPI value of 5,452.48. The statistics also revealed significant variations between the lowest and highest figures, therefore indicating considerable fluctuations in the state of the economy. While both elements were rising, PCE has risen faster during the past few years than DPI. Consumer expenditure has expanded more quickly than pay rise. People may thus depend more on loans or savings during periods of economic growth.

With regard to autocorrelation, the Durbin-Watson test results for DPI and PCE respectively were 2.16 and 2.82, respectively, indicating no appreciable autocorrelation either. Consequently, the data does not violate any necessary presumptions for accurate time series modeling. The finite distributed lag (FDL) model did, however, suffer certain flaws. Given some points when PCE was all one in the data, it was difficult to determine how lagged DPI affected PCE. We sought to address this issue using an AR (1) model. A coefficient of 0.1434 demonstrated that the current value is unaffected by the past value of PCE.

This one fits better than the FDL model since it helps to prevent multicollinearity issues. PCE and DPI are cointegrated; they have been linked for a long time based on the Augmented Dickey-Fuller (ADF) test for cointegration. The Error Correction Model (ECM) supported this by demonstrating that the next period fixes around 9% of any PCE departure from equilibrium. This validated the theory that PCE and DPI would show a steady long-term correlation. Though the ECM was crucial, DPI had little effect on PCE over the near run. This implies that, over time, DPI affects PCE more than it does in the near term.

The study reveals that whereas the short-term changes in DPI do not affect PCE, the two are strongly correlated over time. The results reveal that any deviations from long-term equilibrium are corrected; changes in consumer spending depending on changes in income occur gradually and over time. These results should cause companies and legislators to consider long-term patterns while forecasting consumer expenditure. More studies, including other elements influencing behavior, can help models be more accurate and better at explaining the complexity of consumer behavior.

Introduction

Using U.S. Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI), we investigate the relationship between consumer expenditure and income levels in our study report. In our work, we apply models including autocorrelation testing, cointegration analysis, finite distributed lag models, AR (1) models, and an error correction model (ECM). We want to know how variations in income affect consumer purchasing behavior by looking at these factors over ten years. Our findings result from methodologies of research and modeling that provide useful insights for guiding company strategies and trend prediction in the economy.

Data

Two gauges—U.S. personal consumption expenditures (PCE) and Disposable Personal Income (DPI) obtained from FRED (Federal Reserve Economic Data)—constitute the basis of the analysis. With updates for both PCE and DPI included in the dataset, these measurements help explain how consumer expenditures and income are entwined in the United States over a period of ten years.

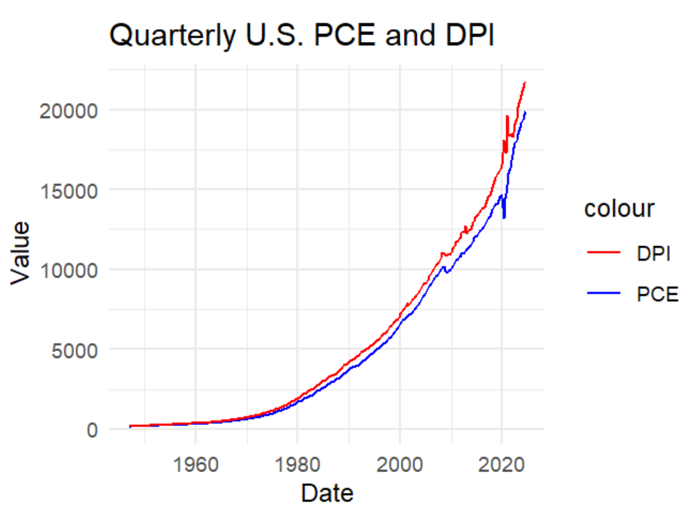
Descriptive Statistics

Summary Statistics for PCE and DPI

Statistic	N	Mean	St. Dev.	Min	Max
Mean_PCE	1	4,905.65		4,905.65	4,905.65
SD_PCE	1	5,232.50		5,232.50	5,232.50
Min_PCE	1	156.16		156.16	156.16
Max_PCE	1	19,935.21		19,935.21	19,935.21
Mean_DPI	1	5,452.48		5,452.48	5,452.48
SD_DPI	1	5,782.41		5,782.41	5,782.41
Min_DPI	1	170.24		170.24	170.24
Max_DPI	1	21,804.90		21,804.90	21,804.90

We first computed some fundamental summary statistics to help us to grasp the dataset better. With a standard deviation of 5,232.50, PCE on average for the period is 4,905.65. This suggests that the figures vary significantly even if the average consumer expenditure is about 4,105.65. With a minimum measured PCE value of 156.16 and a maximum of 19,935.21, the PCE can change with time. For DPI, the mean is 5,452.48, with a 5,782.41 standard deviation. DPI also has a significant variation; the least value is 170.24, and the greatest is 21,804.90. These broad variations in PCE and DPI mirror the historical economic fluctuations encompassing periods of income disparity, inflation, and economic expansion as well as periods of recession.

Graphical Illustration



Other than the summary statistics, a graphical representation was developed to understand the link between PCE and DPI. With notable increases in the most recent decades, the plot that illustrates the patterns of both variables over time showing PCE and DPI indicates a steady upward tendency. This trend shows a period of economic expansion in which disposable income and consumer expenditure have grown consistently. Fascinatingly, in the latter years, the PCE line (shown in blue) seems to have risen faster than the DPI line (shown in red), suggesting that personal consumption has expanded faster than disposable income. This implies that consumers have been spending more during times of economic growth, depending on loans or savings to support their consumption.

The summary statistics and graphical representation clearly show rising income over time as well as consumption. This prepares the ground for more research to investigate the elements driving these changes and the relationships between PCE and DPI and more general economic circumstances.

Analysis

1. Test for Autocorrelation

```
Durbin-Watson test

data:  PCEC ~ PCEC_lag1
DW = 2.16, p-value = 0.9124
alternative hypothesis: true autocorrelation is greater than 0

> # Test for autocorrelation on the DPI variable
> dwtest(DPI ~ DPI_lag1, data = data_clean)

        Durbin-Watson test

data:  DPI ~ DPI_lag1
DW = 2.8216, p-value = 1
alternative hypothesis: true autocorrelation is greater than 0
```

The Durbin- Watson (DW) test was used to probe autocorrelation in both PCE and DPI data. The data demonstrate no noticeable autocorrelation, given a p-value of 0.9124 and a DW statistic of 2.16 for PCE. In DPI, the DW statistic is 2.82 with a p-value of 1, suggesting similar no autocorrelation. The DW test looks at whether the residuals of a regression model are linked with one another; in both cases, the result shows that autocorrelation has no effect on the PCE and DPI series.

2. Finite distributed lag model

```
# FDL Model: PCE explained by current and lagged DPI
> fdl_model <- dynlm(PCEC ~ DPI + lag(DPI, 1) + lag(DPI, 2), data = data)
> summary(fdl_model)

Time series regression with "zoo" data:
Start = 1970-01-01, End = 1970-01-01

Call:
dynlm(formula = PCEC ~ DPI + lag(DPI, 1) + lag(DPI, 2), data = data)

Residuals:
ALL 1 residuals are 0: no residual degrees of freedom!

Coefficients: (3 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    631.7        NaN     NaN      NaN
DPI               NA         NA      NA       NA
lag(DPI, 1)       NA         NA      NA       NA
lag(DPI, 2)       NA         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
```

We aimed to build a Finite Distributed Lag (FDL) model using current and lagged values of DPI as predictors for PCE in order to evaluate the link between Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI). The FDL model thus generated poor model estimate results since the coefficients for the present and lagged values of DPI were not defined. Singularities produce this; hence, the lagged and present DPI values showed great correlation, and separating their respective effects on PCE proved difficult.

Neither pertinent t-values nor p-values could be found; the singularity issue generated NA (unavailable) values for the coefficient estimates. This suggests that the model could not distinguish between prior periods' influence on PCE and the instantaneous DPI impact. It implies that the relationship between PCE and DPI may be unduly tightly coupled or that the lagged variables lacked more insights not already reflected by the present value of DPI.

These results indicate that more model adjustments are necessary. One next action may be addressing the multicollinearity between DPI and its lags or modifying the count of the lags in the model. The FDL model can provide appreciable insight into the link between PCE and DPI by resolving these issues. This underlines the need for more careful data preparation or different modeling approaches to capture the variables' interactions correctly and directs us to the following analytical section.

3. Fitting an AR (1) Model

We examined the AR (1) model on the differenced Personal Consumption Expenditures (PCE) data. The AR (1) model is a time series model with a data current value dependent on its past value. We then computed the PCE data difference, eliminating trends and increasing the series' stability, guaranteeing that the data was stationary, which is required for time series analysis.

AR (1) Model Results and Interpretation

```
# Fit an AR(1) model on the differenced PCE series
> ar1_model <- arima(dPCE, order = c(1, 0, 0))
> summary(ar1_model)

Call:
arima(x = dPCE, order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.1434    63.8463
s.e.  0.0563     9.0767

sigma^2 estimated as 18760:  log likelihood = -1965,  aic = 3936.01

Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE     MASE
ACF1
Training set 0.05244746 136.9686 59.21419 -94.23874 818.2996 1.374378 -
0.03391336
```

The output of the AR (1) model exposes a coefficient of 0.1434. As such, the present value is expected to vary roughly 0.1434 units for every unit change in the prior value of the differenced PCE. The intercept, the baseline level of the differenced series, is 63.85 having a zero-prior value.

The model's Akaike Information Criteria (AIC) is 3936.01, which helps one compare its fit with other models: lower AIC values indicate a better fit. Based on a Root Mean Squared Error (RMSE) of 136.97, the model estimates errors of about 137 units. Reflecting rather high data volatility, the Mean Absolute Percentage Error (MAPE) is 818.3%.

How This Helps with the FDL Model Problem

Our previous FDL model could have run better since several correlated variables prevented the model from estimating some coefficients. Focusing on the link between the present and the prior value of the PCE, the AR(1) model streamlines the analysis and facilitates time-dependent data collecting.

We eliminate difficulties resulting from utilizing too many lags in the FDL model by applying the AR (1) model. This method guarantees that we avoid encountering problems like singularities or models that are too complicated and clarifies how PCE increases over time.

4. ADF Based Cointegration Testing

We conducted an Augmented Dickey-Fuller (ADF) test on the residuals of the cointegration regression (where PCE was regressed on DPI) to investigate the cointegration between Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI). The ADF test searches the residuals for a unit root, determining whether the series are cointegrated.

ADF Test Results and Interpretation

```
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-2922.54   -14.29     1.80    12.07   855.67

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.31411    0.04904  -6.405 5.63e-10 ***
z.diff.lag -0.16774    0.05634  -2.977  0.00314 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 218.3 on 307 degrees of freedom
Multiple R-squared:  0.211,    Adjusted R-squared:  0.2059
F-statistic: 41.05 on 2 and 307 DF,  p-value: < 2.2e-16


Value of test-statistic is: -6.4053

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

The test statistics are well below the critical values for the 1%, 5%, and 10% significance levels, -6.4053 (the critical values are -2.58, -1.95, and -1.62, respectively). We reject the null hypothesis of a unit root since the test statistic is smaller than the crucial values at all levels. This implies that PCE and DPI are cointegrated since the cointegration regression residuals seem stationary.

The tiny p-value for the test of 5.63e-10 helps further justify the null hypothesis's rejection. Understanding their economic link depends critically on the substantial statistical evidence that PCE and DPI move together over time and are in a long-run equilibrium relationship.

With an Adjusted R-squared of 0.2059, the lagged values of the differenced series help to explain roughly 20.6% of the variance in the residuals. Although not particularly strong, this indicates the model catches some important data trends.

The ADF test finds that PCE and DPI are, in fact, cointegrated. This relationship helps to support the theory that the two variables have a long-term equilibrium; deviations from this equilibrium will usually be corrected over time. This is a significant outcome since it suggests that, over time, changes in PCE and DPI are linked, and their motions reflect an underlying relationship rather than being entirely random.

5. Approximate the Error Correction Model (ECM).

While it accounts for their long-term relationship, the Error Correction Model (ECM) clarifies the short-term link between Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI). The results reveal that DPI and the error correction mechanism bringing PCE back to equilibrium following any deviations affect variations in PCE.

ECM Test Results and Interpretation

```
# Define and fit the Error Correction Model (ECM)
+   ecm_model <- dynlm(dPCE ~ dDPI + ect, data = as.data.frame(data_ecm))
+ print(summary(ecm_model))
+ } else {
+   stop("No valid data available after removing NA values. Check your data
preparation steps.")
+ }

Time series regression with "numeric" data:
Start = 1, End = 310

Call:
dynlm(formula = dPCE ~ dDPI + ect, data = as.data.frame(data_ecm))

Residuals:
    Min      1Q  Median      3Q     Max
-999.84  -59.65  -20.74   23.35 1325.68

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 68.50978    8.22515   8.329 2.73e-15 ***
dDPI        -0.06734    0.04270  -1.577  0.11577
ect          0.09021    0.03007   3.000  0.00292 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 135 on 307 degrees of freedom
Multiple R-squared:  0.05787,	Adjusted R-squared:  0.05173 
F-statistic: 9.428 on 2 and 307 DF,  p-value: 0.0001062
```

With a coefficient of 68.51, everything else held constant and a low p-value; the intercept in the model is significant. This suggests a considerable baseline effect, a continuing influence on PCE, even without changes in other parameters. Although it is not statistically significant, meaning that short-term fluctuations in DPI have little impact on PCE, the differenced DPI (dDPI) coefficient is -0.06734, maintaining everything else. This suggests that DPI may have long-term impacts on intake, even though not immediately.

Given that everything else is constant, the Error Correction Term (ECT) has a positive and significant coefficient of 0.09021. This means that about 9% of the variance will be adjusted in the next quarter should PCE stray from its long-term link with DPI. This is important since it shows that PCE evolves with time to match DPI, hence maintaining long-term equilibrium.

The residual standard error of 135 indicates a modest fit of the model, given an average error of about 135 units. The model with an R-squared value of 0.05787 shows that other factors influencing PCE are not sufficiently captured and explains just around 5.8% of the variation in PCE. Still, the F-statistic of 9.424 with a quite low p-value shows that the model is statistically significant generally.

The error correcting procedure is crucial in returning PCE to its long-term relationship with DPI even if variations in DPI have little effect on short-term PCE. Though the model is useful, it also suggests that other factors are involved in variances in PCE.

Conclusion and Recommendation

The analysis of the relationship between Personal Consumption Expenditures (PCE) and Disposable Personal Income (DPI) provides an interesting examination of how these economic variables interact over time. We verified using the Augmented Dickey-Fuller test that the residuals from the cointegral model were stationary, therefore establishing a long-term relationship between PCE and DPI. The Error

Correction Model (ECM) enhanced this even further by proving that deviations from equilibrium are rectified over time, with over 9% of the deviation in PCE corrected in the next period. The AR (1) model indicated that although PCE's historical values influence short-term changes in PCE, changes in DPI did not instantly affect consumption, suggesting that DPI's effects may be more long-term than short-term.

Regarding the finite distributed lag model (FDL), the results exposed issues with model identification whereby some coefficients were undefined, primarily due to data singularities. Notwithstanding this, the overall trend showed that lags in DPI can affect PCE; even with this link, additional research is needed to obtain closer knowledge. Autocorrelation studies for PCE and DPI confirmed no appreciable autocorrelation, as shown by the Durbin-Watson test results for both variables with values near 2. This suggests that the residuals of our models do not show trends that would challenge the assumptions of our time series models.

These findings lead us to urge businesses and legislators stressing consumer expenditure to realize that changes in DPI can affect consumption over the long run instead of immediately. The Error Correction Model suggests that should PCE stray from its long-term trend, it usually corrects itself gradually. Given the low explanatory ability of the models, more research involving other elements could be necessary to improve the fit and better represent the complexity of consumer expenditure behavior. Examining other outside factors influencing consumption will also enable one to understand the dynamics between income and expenditure.