Paper 5: Analyzing the Factors Influencing Media Trust Over Time

Abby Sim

2023-05-05

## Theory

The American public’s trust in media is at a record low: according to Gallup, trust in media has decreased from 72% less than 50 years ago to 34% today. This paper explores how Americans’ trust in the news media has changed over time in response to trends in political polarization, social media usage, media freedom, and government trust.

Diagram

Description automatically generated

As depicted in the causal flow chart above, this paper proposes that each exogeneous variable has a direct causal impact on the public’s trust in news media.

* Political polarization: In line with Robert Vallone’s hostile media effect — a social psychology theory that refers to the tendency for individuals with strong preexisting attitudes on an issue to perceive media coverage as biased against their side — this paper theorizes that higher levels of polarization in American politics leads to lower levels of trust in the media. In addition, because the news media industry caters to consumers and individuals are more likely to seek out information that aligns with their pre-existing worldviews, the media industry might harness partisan bias to appeal to viewers during times of increased political polarization, causing increased distrust in media.
* Social Media Usage: With increased social media usage comes increased misinformation and increased skepticism towards the news media. As more and more Americans gain access to alternative forms of news coverage that might contradict or undermine the traditional news media and circulate misinformation, public trust in media declines.
* Freedom of the press: Freedom of the press refers to the ability of print, broadcast, and digital media to operate freely and without threat of repercussions. It encapsulates the editorial independence of both state-owned and privately owned outlets; access to information and sources; official censorship and self-censorship; the diversity of news available; and the transparency and concentration of media ownership. The freer a country’s press is, the more likely it is that the public will trust the news media they consume to be transparent and unbiased.
* Trust in government: During times where public trust in government increases, trust in the media is also likely to increase. When Americans trust the political process, the opinions of experts, and official government accounts of events, they are also more likely to trust the news media institution.

## Data

To operationalize this theory, this paper uses data aggregated from Pew Research Center’s Core Trends surveys conducted annually on national samples of approximately 1500 adults via telephone interviews; the Gallup Poll Social Surveys administered annually on national samples of at least 1000 adults via telephone interviews; Voteview, a dataset on every congressional role call in American history; and Freedom House’s Freedom of the Press report, an annual report on media independence around the world. This paper focuses its analyses on the years 1985 to 2022, providing a set of 38 data points, and utilizes the United States over time as its unit of analysis.

The variables used by this paper are:

1. Trust, representing the percentage of the American public reporting “a great deal/fair amount of trust in the media.”
2. Polar, representing the difference between the mean DW-NOMINATE scores of Senate Republicans and Senate Democrats. DW-NOMINATE scores are used widely to describe the political ideology of political actors, political parties and political institutions: a score closer to -1 is considered liberal, while a score closer to 1 is considered conservative. As a result, this variable, which measures the difference between the mean scores of Senate Republicans and Senate Democrats, theoretically ranges from 0 to 2. The bigger the Polar value is, the higher the level of political polarization.
3. Social, representing the percentage of the American public using any form of social media.
4. Free, representing the United States’ media independence using Freedom House’s Freedom of the Press index. This index assigns countries a total score from 0 (most free) to 100 (least free) based on a set of 23 methodology questions and 109 indicators divided into three broad categories covering the legal, political, and economic environment. The bigger Free is, the less independent the media is.
5. Gov.App, representing the percentage of the American public reporting that they “trust the government to do what is right always/most of the time.”

According to this operationalization, I hypothesize that the effect of Polar on Trust is negative (i.e. time periods with higher levels of political polarization see lower media trust); the effect of Social on Trust is negative (i.e. time periods with higher social media usage see lower media trust); the effect of Free on Trust is negative (i.e. time periods where press freedom is more inhibited see lower media trust); and the effect of Gov.App on Trust is positive (i.e. time periods with higher levels of governmental trust see lower media trust).

media <- na.omit(read.csv("Data/media44.csv")[6:43,1:7])  
media

## Year Time Trust Polar Social Freedom Gov.App  
## 6 1985 6 64 0.6195957 0 14 42  
## 7 1986 7 61 0.6195957 0 13 44  
## 8 1987 8 63 0.6082332 0 13 43  
## 9 1988 9 63 0.6082332 0 12 41  
## 10 1989 10 64 0.6162547 0 14 40  
## 11 1990 11 56 0.6162547 0 13 35  
## 12 1991 12 59 0.6187860 0 12 41  
## 13 1992 13 52 0.6187860 0 12 26  
## 14 1993 14 63 0.6337041 0 12 25  
## 15 1994 15 61 0.6337041 0 12 21  
## 16 1995 16 53 0.6524845 0 14 22  
## 17 1996 17 54 0.6524845 0 14 29  
## 18 1997 18 53 0.6839434 0 12 29  
## 19 1998 19 55 0.6839434 0 13 31  
## 20 1999 20 55 0.6583835 1 13 34  
## 21 2000 21 51 0.6583835 1 15 38  
## 22 2001 22 53 0.6631800 1 16 49  
## 23 2002 23 54 0.6631800 2 17 44  
## 24 2003 24 54 0.6498015 3 13 40  
## 25 2004 25 44 0.6498015 5 17 39  
## 26 2005 26 50 0.6858404 8 16 32  
## 27 2006 27 48 0.6858404 11 16 31  
## 28 2007 28 47 0.6871482 21 17 26  
## 29 2008 29 43 0.6871482 26 18 24  
## 30 2009 30 45 0.7063485 37 18 22  
## 31 2010 31 43 0.7063485 46 17 22  
## 32 2011 32 44 0.7436179 50 18 19  
## 33 2012 33 40 0.7436179 54 18 19  
## 34 2013 34 44 0.7968635 61 21 21  
## 35 2014 35 40 0.7968635 62 22 19  
## 36 2015 36 40 0.8130109 65 21 18  
## 37 2016 37 32 0.8130109 69 23 18  
## 38 2017 38 41 0.8260322 69 22 19  
## 39 2018 39 45 0.8260322 69 22 18  
## 40 2019 40 41 0.8407891 72 23 17  
## 41 2020 41 40 0.8407891 72 23 23  
## 42 2021 42 36 0.8733165 72 24 21  
## 43 2022 43 34 0.8733165 73 23 20

cat("The dataset has", paste(dim(media)[1]), "observations, and", paste(dim(media)[2]-2), " variables for each one of them", '\n')

## The dataset has 38 observations, and 5 variables for each one of them

descr::freq(ordered(media$Trust), plot = F)

## ordered(media$Trust)   
## Frequency Percent Cum Percent  
## 32 1 2.632 2.632  
## 34 1 2.632 5.263  
## 36 1 2.632 7.895  
## 40 4 10.526 18.421  
## 41 2 5.263 23.684  
## 43 2 5.263 28.947  
## 44 3 7.895 36.842  
## 45 2 5.263 42.105  
## 47 1 2.632 44.737  
## 48 1 2.632 47.368  
## 50 1 2.632 50.000  
## 51 1 2.632 52.632  
## 52 1 2.632 55.263  
## 53 3 7.895 63.158  
## 54 3 7.895 71.053  
## 55 2 5.263 76.316  
## 56 1 2.632 78.947  
## 59 1 2.632 81.579  
## 61 2 5.263 86.842  
## 63 3 7.895 94.737  
## 64 2 5.263 100.000  
## Total 38 100.000

descr::freq(ordered(media$Polar), plot = F)

## ordered(media$Polar)   
## Frequency Percent Cum Percent  
## 0.608233202 2 5.263 5.263  
## 0.616254658 2 5.263 10.526  
## 0.61878605 2 5.263 15.789  
## 0.619595745 2 5.263 21.053  
## 0.633704094 2 5.263 26.316  
## 0.649801471 2 5.263 31.579  
## 0.65248447 2 5.263 36.842  
## 0.65838354 2 5.263 42.105  
## 0.66318 2 5.263 47.368  
## 0.683943434 2 5.263 52.632  
## 0.685840404 2 5.263 57.895  
## 0.687148235 2 5.263 63.158  
## 0.706348485 2 5.263 68.421  
## 0.743617925 2 5.263 73.684  
## 0.796863463 2 5.263 78.947  
## 0.813010943 2 5.263 84.211  
## 0.826032197 2 5.263 89.474  
## 0.84078905 2 5.263 94.737  
## 0.873316527 2 5.263 100.000  
## Total 38 100.000

descr::freq(ordered(media$Social), plot = F)

## ordered(media$Social)   
## Frequency Percent Cum Percent  
## 0 14 36.842 36.84  
## 1 3 7.895 44.74  
## 2 1 2.632 47.37  
## 3 1 2.632 50.00  
## 5 1 2.632 52.63  
## 8 1 2.632 55.26  
## 11 1 2.632 57.89  
## 21 1 2.632 60.53  
## 26 1 2.632 63.16  
## 37 1 2.632 65.79  
## 46 1 2.632 68.42  
## 50 1 2.632 71.05  
## 54 1 2.632 73.68  
## 61 1 2.632 76.32  
## 62 1 2.632 78.95  
## 65 1 2.632 81.58  
## 69 3 7.895 89.47  
## 72 3 7.895 97.37  
## 73 1 2.632 100.00  
## Total 38 100.000

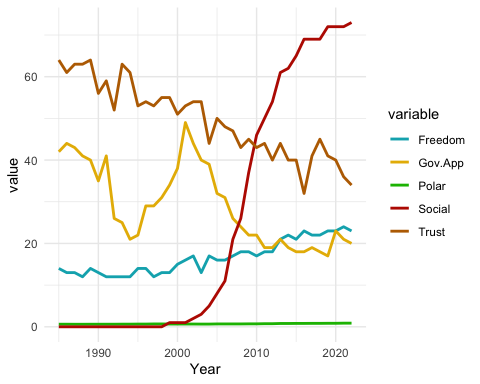
descr::freq(ordered(media$Freedom), plot = F)

## ordered(media$Freedom)   
## Frequency Percent Cum Percent  
## 12 6 15.789 15.79  
## 13 6 15.789 31.58  
## 14 4 10.526 42.11  
## 15 1 2.632 44.74  
## 16 3 7.895 52.63  
## 17 4 10.526 63.16  
## 18 4 10.526 73.68  
## 21 2 5.263 78.95  
## 22 3 7.895 86.84  
## 23 4 10.526 97.37  
## 24 1 2.632 100.00  
## Total 38 100.000

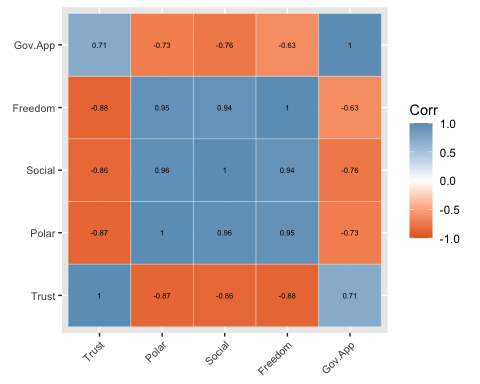
descr::freq(ordered(media$Gov.App), plot = F)

## ordered(media$Gov.App)   
## Frequency Percent Cum Percent  
## 17 1 2.632 2.632  
## 18 3 7.895 10.526  
## 19 4 10.526 21.053  
## 20 1 2.632 23.684  
## 21 3 7.895 31.579  
## 22 3 7.895 39.474  
## 23 1 2.632 42.105  
## 24 1 2.632 44.737  
## 25 1 2.632 47.368  
## 26 2 5.263 52.632  
## 29 2 5.263 57.895  
## 31 2 5.263 63.158  
## 32 1 2.632 65.789  
## 34 1 2.632 68.421  
## 35 1 2.632 71.053  
## 38 1 2.632 73.684  
## 39 1 2.632 76.316  
## 40 2 5.263 81.579  
## 41 2 5.263 86.842  
## 42 1 2.632 89.474  
## 43 1 2.632 92.105  
## 44 2 5.263 97.368  
## 49 1 2.632 100.000  
## Total 38 100.000

plotmedia <- media %>% dplyr::select(Year, Trust, Polar, Social, Freedom, Gov.App) %>% gather(key = "variable", value = "value", -Year)  
  
ggplot(plotmedia, aes(x = Year, y = value)) +   
 geom\_line(aes(color = variable), size = 1) +  
 scale\_color\_manual(values = c("#00AFBB", "#E7B800", "#10bb00", "#bb1c00", "#bb6d00")) +  
 theme\_minimal()



media2 <- media[c("Trust", "Polar", "Social", "Freedom", "Gov.App")]  
corr <- round(cor(media2, use = "complete.obs"), 2)  
  
ggcorrplot(corr, type = "full", lab = TRUE,  
 outline.col = "white",  
 ggtheme = ggplot2::theme\_gray,  
 colors = c("#E46726", "white", "#6D9EC1"),   
 lab\_col = "black", lab\_size = 2,   
 tl.cex = 8, tl.col = "black")



The correlation matrix supports the paper’s theory regarding the direction of the relationships between Trust and the other variables: Polar, Social, and Freedom and negatively correlated with Trust, while Gov.App is positively correlated with it. However, this correlation matrix demonstrates extremely high levels of multicollinearity, especially between the Freedom, Social, and Polar variables. Multicollinearity could potentially make it difficult to test individual regression coefficients due to inflated standard errors and unstable parameter estimates.

## Statistical analysis

### OLS Regression

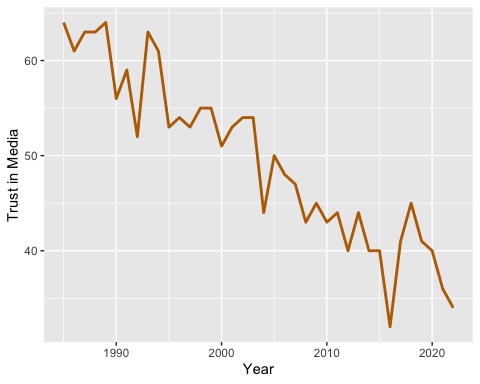
## OLS Regression Model  
mod1 <- lm(Trust ~ Polar+Social+Freedom+Gov.App, data = media)  
summary(mod1)

##   
## Call:  
## lm(formula = Trust ~ Polar + Social + Freedom + Gov.App, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9402 -2.8460 -0.5633 2.1621 7.2490   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 77.86341 18.01992 4.321 0.000134 \*\*\*  
## Polar -10.21386 31.17708 -0.328 0.745277   
## Social 0.06663 0.09414 0.708 0.484039   
## Freedom -1.84232 0.61683 -2.987 0.005284 \*\*   
## Gov.App 0.27436 0.11914 2.303 0.027725 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.017 on 33 degrees of freedom  
## Multiple R-squared: 0.8219, Adjusted R-squared: 0.8003   
## F-statistic: 38.07 on 4 and 33 DF, p-value: 6.305e-12

### Testing for non-stationarity: Time plots

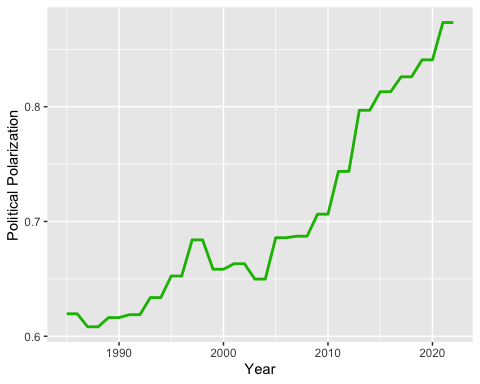
The first step in analyzing a time series models entails testing for non-stationarity, a potential indication of spurious regression. If a series is non-stationary, it is important to evaluate whether common patterns of trending lead to artificially enhanced correlations and, as a result, spurious inferences within the model.

ggplot(media, aes(x = Year, y = Trust)) + geom\_line(color="#bb6d00", size = 1) + ylab("Trust in Media")



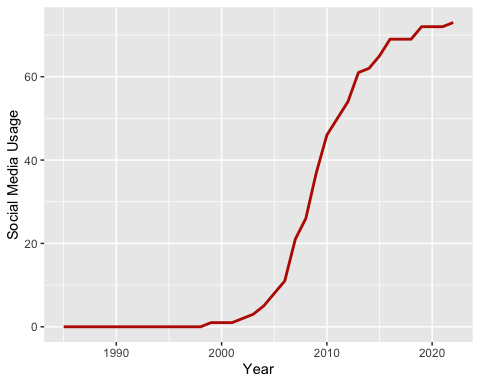
As an informal test of non-stationarity, we examine the time plots of each variable. This plot of Trust in Media over time suggests that Trust in Media is non-stationary (i.e. its mean and variance are not constant). However, its trend appears generally linear/constant, and as a result, there is a possibility for stationarity once time is controlled for.

ggplot(media, aes(x = Year, y = Polar)) + geom\_line(color = "#10bb00", size = 1) + ylab("Political Polarization")



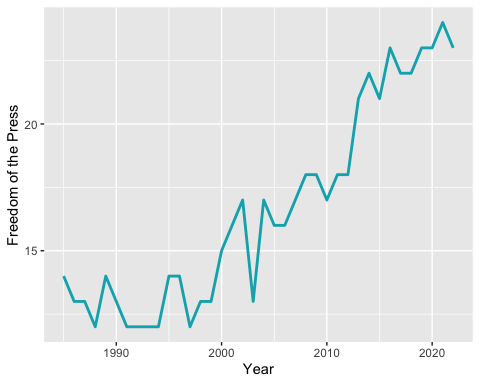
ggplot(media, aes(x = Year, y = Social)) + geom\_line(color = "#bb1c00", size = 1) + ylab("Social Media Usage")

This plot of Political Polarization over time suggests that this variable is non-stationary (i.e. its mean and variance are not constant).



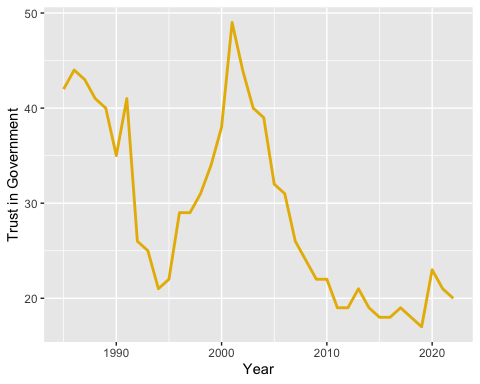
ggplot(media, aes(x = Year, y = Freedom)) + geom\_line(color = "#00AFBB", size = 1) + ylab("Freedom of the Press")

This plot of Social Media Usage over time suggests that this variable is non-stationary (i.e. its mean and variance are not constant).



ggplot(media, aes(x = Year, y = Gov.App)) + geom\_line(color = "#E7B800", size = 1) + ylab("Trust in Government")

This plot of Freedom of Press over time suggests that this variable is non-stationary (i.e. its mean and variance are not constant).

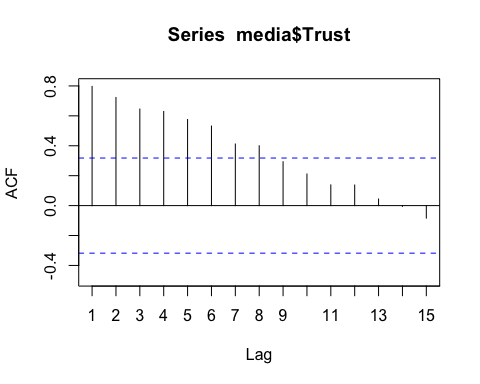


This plot of Trust in Government over time suggests that this variable is non-stationary (i.e. its mean and variance are not constant).

### 

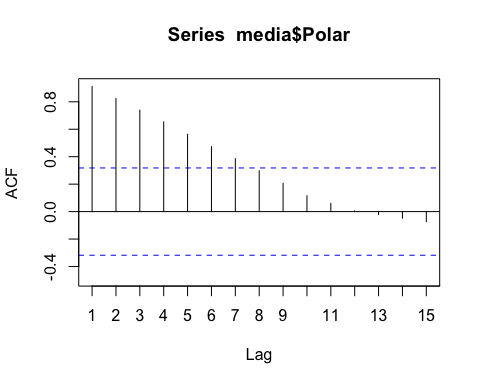
### Testing for non-stationarity: Auto-correlation function

Acf(media$Trust)



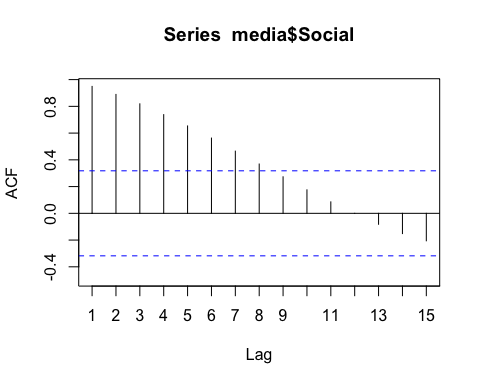
Autocorrelation plots indicating stationarity drop off to 0 extremely quickly and stay at 0. The ACF plot of the Trust in Media variable does not drop off quickly enough and as a result indicates non-stationarity.

Acf(media$Polar)



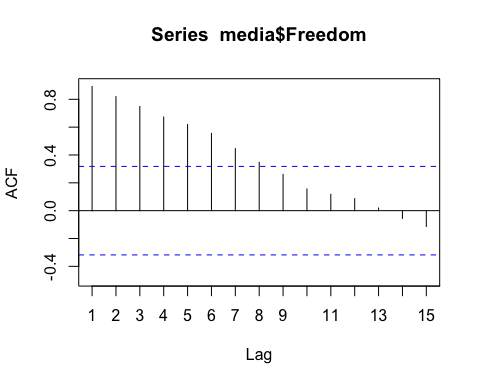
The ACF plot of the Political Polarization variable does not drop off quickly enough and as a result indicates non-stationarity.

Acf(media$Social)



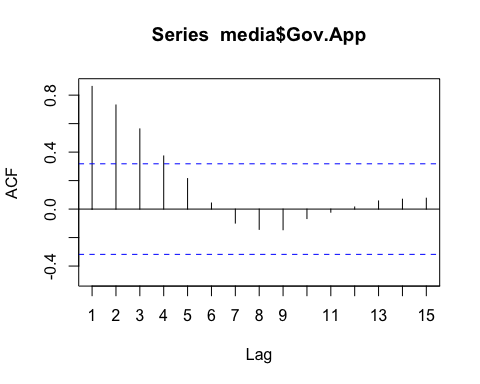
The ACF plot of the Social Media Usage variable does not drop off quickly enough and as a result indicates non-stationarity.

Acf(media$Freedom)



The ACF plot of the Prses Freedom variable does not drop off quickly enough and as a result indicates non-stationarity.

Acf(media$Gov.App)



The ACF plot of the Government Trust variable does not drop off quickly enough and as a result indicates non-stationarity.

### Testing for non-stationarity: Dickey-Fuller tests

#### Dickey-Fuller test for trust in media

media$Trust.Lag <- sapply(1:nrow(media), function(x) media$Trust[x-1])  
media$Trust.Lag <- car::recode(as.numeric(media$Trust.Lag),"numeric(0)=NA")  
  
media$Diff.Trust <- media$Trust-media$Trust.Lag  
  
# Regress first difference on the lagged variable  
dftrust1 <- lm(Diff.Trust ~ Trust.Lag, data = media)  
summary(dftrust1)

##   
## Call:  
## lm(formula = Diff.Trust ~ Trust.Lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.6798 -3.5773 0.3202 2.9356 12.0638   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 5.60362 4.32104 1.297 0.203  
## Trust.Lag -0.12822 0.08512 -1.506 0.141  
##   
## Residual standard error: 4.455 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.06088, Adjusted R-squared: 0.03404   
## F-statistic: 2.269 on 1 and 35 DF, p-value: 0.141

durbinWatsonTest(dftrust1)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.3560376 2.698997 0.032  
## Alternative hypothesis: rho != 0

Since the tau-value generated from regressing the first difference on the lagged trust variable is -1.506 and the Dickey-Fuller critical value for a sample size of 37 is ≈-3.00, we cannot reject the null hypothesis that there is a unit root (i.e. non-stationarity) for the trust variable.

However, since dL=1.36 (or equivalently 2.64) and dU=1.59 (or equivalently 2.41) for a model with k=2,n=37 and this model’s Durbin-Watson statistic is 2.698, there is evidence of serial correlation. As a result, we cannot rely on this version of the Dickey-Fuller equation.

# Regress first difference on the lagged variable and the time variable  
dftrust2 <- lm(Diff.Trust ~ Trust.Lag + Year, data = media)  
summary(dftrust2)

##   
## Call:  
## lm(formula = Diff.Trust ~ Trust.Lag + Year, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.1659 -1.6070 0.0239 1.7035 6.4315   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1618.1042 285.8469 5.661 2.38e-06 \*\*\*  
## Trust.Lag -1.0333 0.1720 -6.007 8.44e-07 \*\*\*  
## Year -0.7820 0.1386 -5.641 2.52e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.248 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.5149, Adjusted R-squared: 0.4864   
## F-statistic: 18.05 on 2 and 34 DF, p-value: 4.557e-06

durbinWatsonTest(dftrust2)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.002471299 1.988439 0.818  
## Alternative hypothesis: rho != 0

The next Dickey-Fuller equation regresses the first difference on the lagged trust variable and the time variable. Running a Durbin-Watson test on this version of the equation generates a statistic of 1.98. Since there is evidence of no serial correlation, we can use this equation to evaluate the model’s non-stationarity

Since the tau-value generated by running the Dickey-Fuller test on this model is -6.007 and the critical value for a sample size of 37 is ≈-3.60, we reject the null hypothesis that there is a unit root and conclude that the Trust variable is stationary when time is controlled for.

#### Dickey-Fuller test for political polarization

media$Polar.Lag <- sapply(1:nrow(media), function(x) media$Polar[x-1])  
media$Polar.Lag <- car::recode(as.numeric(media$Polar.Lag),"numeric(0)=NA")  
  
media$Diff.Polar <- media$Polar-media$Polar.Lag  
  
# Regress first difference on the lagged variable  
dfpolar1 <- lm(Diff.Polar ~ Polar.Lag, data = media)  
summary(dfpolar1)

##   
## Call:  
## lm(formula = Diff.Polar ~ Polar.Lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.031962 -0.007058 -0.004931 0.004194 0.045096   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.01363 0.02246 -0.607 0.548  
## Polar.Lag 0.02928 0.03191 0.918 0.365  
##   
## Residual standard error: 0.01538 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.0235, Adjusted R-squared: -0.0044   
## F-statistic: 0.8423 on 1 and 35 DF, p-value: 0.365

durbinWatsonTest(dfpolar1)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.2807042 2.541706 0.11  
## Alternative hypothesis: rho != 0

Since dL=1.36 (or equivalently 2.64) and dU=1.59 (or equivalently 2.41) for k=2 and n=37 and the Durbin-Watson statistic is 2.54, there is evidence of potential serial correlation. As a result, we cannot use this version of the Dickey-Fuller equation.

# Regress first difference on the lagged variable and the time variable  
dfpolar2 <- lm(Diff.Polar ~ Polar.Lag + Year, data = media)  
summary(dfpolar2)

##   
## Call:  
## lm(formula = Diff.Polar ~ Polar.Lag + Year, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.027760 -0.008570 -0.002542 0.006464 0.040329   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.6575302 1.1950068 -2.224 0.0329 \*  
## Polar.Lag -0.1441310 0.0840095 -1.716 0.0953 .  
## Year 0.0013798 0.0006236 2.213 0.0337 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01459 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.1464, Adjusted R-squared: 0.09622   
## F-statistic: 2.916 on 2 and 34 DF, p-value: 0.06778

durbinWatsonTest(dfpolar2)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.2267853 2.441704 0.288  
## Alternative hypothesis: rho != 0

For k=3 and n=36 (since we lose an observation due to the addition of another lag), dL=1.30 (or equivalently 2.70) and dU=1.65 (or equivalently 2.35). Since the Durbin-Watson statistic is 2.44, there is still evidence of potential serial correlation and we cannot use this version of the Dickey-Fuller equation.

# Augmented Dickey-Fuller Test   
media$Diff.Polar.Lag <- sapply(1:nrow(media), function(x) media$Diff.Polar[x-1])  
media$Diff.Polar.Lag <- car::recode(as.numeric(media$Diff.Polar.Lag), "numeric(0)=NA")  
  
# Regress the difference variable on the lagged variable, the time/trend variable and the lagged difference variable  
dfpolar3 <- lm(Diff.Polar ~ Polar.Lag + Year + Diff.Polar.Lag, data = media)  
summary(dfpolar3)

##   
## Call:  
## lm(formula = Diff.Polar ~ Polar.Lag + Year + Diff.Polar.Lag,   
## data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.029149 -0.008852 -0.000081 0.004804 0.037938   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.5962232 1.2820960 -2.025 0.0513 .  
## Polar.Lag -0.1182353 0.0902529 -1.310 0.1995   
## Year 0.0013410 0.0006688 2.005 0.0535 .  
## Diff.Polar.Lag -0.2576006 0.1724049 -1.494 0.1449   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01449 on 32 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.2031, Adjusted R-squared: 0.1284   
## F-statistic: 2.719 on 3 and 32 DF, p-value: 0.06086

durbinWatsonTest(dfpolar3)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.05745068 1.876876 0.476  
## Alternative hypothesis: rho != 0

For k=4, n=35: dL=1.22 (or equivalently 2.78) and dU=1.73 (or equivalently 2.27). Since the Durbin-Watson statistic is 1.87, there is evidence of no serial correlation and we can use this augmented Dickey-Fuller equation.

Since the tau-value generated from regressing the first difference on the lagged trust variable and the time variable is -1.310 and the Dickey-Fuller critical value for a sample size of 37 is ≈-3.60, we cannot reject the null hypothesis that there is a unit root and must conclude that the Political Polarization variable is non-stationary.

#### Dickey-Fuller test for Social Media Use

media$Social.Lag <- sapply(1:nrow(media), function(x) media$Social[x-1])  
media$Social.Lag <- car::recode(as.numeric(media$Social.Lag),"numeric(0)=NA")  
  
media$Diff.Social <- media$Social-media$Social.Lag  
  
# Regress first difference on the lagged variable  
dfsocial1 <- lm(Diff.Social ~ Social.Lag, data = media)  
summary(dfsocial1)

##   
## Call:  
## lm(formula = Diff.Social ~ Social.Lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.060 -1.439 -1.439 1.097 8.975   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.43938 0.63125 2.280 0.0288 \*  
## Social.Lag 0.02251 0.01686 1.335 0.1904   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.972 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.04847, Adjusted R-squared: 0.02128   
## F-statistic: 1.783 on 1 and 35 DF, p-value: 0.1904

durbinWatsonTest(dfsocial1)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.6108191 0.7579352 0  
## Alternative hypothesis: rho != 0

For k=2, n=37, dL=1.36 (or equivalently 2.64) and dU=1.59 (or equivalently 2.41). Since the Durbin-Watson statistic is 0.757, there is evidence of strong serial correlation and we cannot use this version of the Dickey-Fuller equation.

# Regress first difference on the lagged variable and the time variable  
dfsocial2 <- lm(Diff.Social ~ Social.Lag + Year, data = media)  
summary(dfsocial2)

##   
## Call:  
## lm(formula = Diff.Social ~ Social.Lag + Year, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3075 -1.9052 -0.4985 0.9355 7.8725   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -518.18366 195.43057 -2.651 0.0121 \*  
## Social.Lag -0.06400 0.03607 -1.774 0.0850 .  
## Year 0.26032 0.09790 2.659 0.0119 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.744 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2123, Adjusted R-squared: 0.1659   
## F-statistic: 4.581 on 2 and 34 DF, p-value: 0.01732

durbinWatsonTest(dfsocial2)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.562743 0.8431702 0  
## Alternative hypothesis: rho != 0

For Durbin-Watson statistic for k=3, n=36: dL=1.30 (or equivalently 2.70) and dU=1.65 (or equivalently 2.35). Since the Durbin-Watson statistic is 0.843, there is still strong evidence of serial correlation and we cannot use this version of the Dickey-Fuller equation.

# Augmented Dickey-Fuller Test   
media$Diff.Social.Lag <- sapply(1:nrow(media), function(x) media$Diff.Social[x-1])  
media$Diff.Social.Lag <- car::recode(as.numeric(media$Diff.Social.Lag), "numeric(0)=NA")  
  
# Regress the difference variable on the lagged variable, the time/trend variable and the lagged difference variable  
dfsocial3 <- lm(Diff.Social ~ Social.Lag + Year + Diff.Social.Lag, data = media)  
summary(dfsocial3)

##   
## Call:  
## lm(formula = Diff.Social ~ Social.Lag + Year + Diff.Social.Lag,   
## data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5021 -1.1151 -0.3374 0.8110 6.4972   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -374.05164 179.42010 -2.085 0.04515 \*   
## Social.Lag -0.05976 0.03119 -1.916 0.06433 .   
## Year 0.18777 0.08988 2.089 0.04474 \*   
## Diff.Social.Lag 0.57602 0.13833 4.164 0.00022 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.27 on 32 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 0.4862, Adjusted R-squared: 0.4381   
## F-statistic: 10.1 on 3 and 32 DF, p-value: 7.816e-05

durbinWatsonTest(dfsocial3)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.2417014 2.477279 0.304  
## Alternative hypothesis: rho != 0

For k=4, n=35, dL=1.22 (or equivalently 2.78) and dU=1.73 (or equivalently 2.27). Since the Durbin-Watson statistic is 2.47, there is evidence of potential serial correlation and we cannot use this version of the Dickey-Fuller equation.

media$Diff.Social.Lag2 <- sapply(1:nrow(media), function(x) media$Diff.Social.Lag[x-1])  
media$Diff.Social.Lag2 <- car::recode(as.numeric(media$Diff.Social.Lag2), "numeric(0)=NA")  
  
dfsocial4 <- lm(Diff.Social ~ Social.Lag + Year + Diff.Social.Lag + Diff.Social.Lag2, data = media)  
summary(dfsocial4)

##   
## Call:  
## lm(formula = Diff.Social ~ Social.Lag + Year + Diff.Social.Lag +   
## Diff.Social.Lag2, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7718 -1.1489 -0.3764 0.5753 5.7518   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -404.10255 181.87226 -2.222 0.0340 \*  
## Social.Lag -0.07450 0.03097 -2.406 0.0225 \*  
## Year 0.20275 0.09109 2.226 0.0337 \*  
## Diff.Social.Lag 0.35106 0.15909 2.207 0.0351 \*  
## Diff.Social.Lag2 0.40171 0.16468 2.439 0.0208 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.134 on 30 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.5685, Adjusted R-squared: 0.511   
## F-statistic: 9.881 on 4 and 30 DF, p-value: 3.19e-05

durbinWatsonTest(dfsocial4)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.1923961 2.374908 0.5  
## Alternative hypothesis: rho != 0

I added another lag to the augmented Dickey-Fuller equation since the model continues to report serial correlation. For k=5, n=34, dL=1.14 (or equivalently 2.86) and dU=1.81 (or equivalently 2.19). Since the Durbin-Watson statistic generated from this further lagged model is 2.37, there is still evidence of potential serial correlation and we cannot use this version of the Dickey-Fuller equation.

media$Diff.Social.Lag3 <- sapply(1:nrow(media), function(x) media$Diff.Social.Lag2[x-1])  
media$Diff.Social.Lag3 <- car::recode(as.numeric(media$Diff.Social.Lag3), "numeric(0)=NA")  
  
dfsocial5 <- lm(Diff.Social ~ Social.Lag + Year + Diff.Social.Lag + Diff.Social.Lag2 + Diff.Social.Lag3, data = media)  
summary(dfsocial5)

##   
## Call:  
## lm(formula = Diff.Social ~ Social.Lag + Year + Diff.Social.Lag +   
## Diff.Social.Lag2 + Diff.Social.Lag3, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6053 -0.9259 -0.2365 0.7158 5.5890   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -503.64412 180.71769 -2.787 0.00945 \*\*  
## Social.Lag -0.10400 0.03090 -3.366 0.00223 \*\*  
## Year 0.25257 0.09049 2.791 0.00935 \*\*  
## Diff.Social.Lag 0.18415 0.15686 1.174 0.25030   
## Diff.Social.Lag2 0.27050 0.15769 1.715 0.09733 .   
## Diff.Social.Lag3 0.46447 0.16749 2.773 0.00977 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.947 on 28 degrees of freedom  
## (4 observations deleted due to missingness)  
## Multiple R-squared: 0.66, Adjusted R-squared: 0.5993   
## F-statistic: 10.87 on 5 and 28 DF, p-value: 6.973e-06

durbinWatsonTest(dfsocial5)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.03379765 1.916856 0.504  
## Alternative hypothesis: rho != 0

After adding a third lag to the Dickey-Fuller equation, the Durbin-Watson test returns a statistic of 1.91, providing evidence of no serial correlation and allowing us to rely on this version of the equation.

Since the tau-value generated is -3.366 and the Dickey-Fuller critical value for a sample size of 33 is ≈-3.60, we cannot reject the null hypothesis that there is a unit root and conclude that the trend of the Social Media Usage variable is non-stationary.

#### Dickey-Fuller test for Freedom of Press

media$Free.Lag <- sapply(1:nrow(media), function(x) media$Freedom[x-1])  
media$Free.Lag <- car::recode(as.numeric(media$Free.Lag),"numeric(0)=NA")  
  
media$Diff.Free <- media$Freedom-media$Free.Lag  
  
# Regress first difference on the lagged variable  
dffree1 <- lm(Diff.Free ~ Free.Lag, data = media)  
summary(dffree1)

##   
## Call:  
## lm(formula = Diff.Free ~ Free.Lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.2232 -0.9888 -0.1841 0.7768 3.6206   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.88730 1.06622 0.832 0.411  
## Free.Lag -0.03907 0.06298 -0.620 0.539  
##   
## Residual standard error: 1.473 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.01087, Adjusted R-squared: -0.01739   
## F-statistic: 0.3847 on 1 and 35 DF, p-value: 0.5391

durbinWatsonTest(dffree1)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.371871 2.708223 0.04  
## Alternative hypothesis: rho != 0

For k=2,n=37, dL=1.36 (or equivalently 2.64) and dU=1.59 (or equivalently 2.41). Since the Durbin-Watson statistic generated by regressing the lagged freedom variable on the first difference is 2.708, there is evidence of potential serial correlation and we cannot use this version of the Dickey-Fuller equation.

# Regress first difference on the lagged variable and the time variable  
dffree2 <- lm(Diff.Free ~ Free.Lag + Year, data = media)  
summary(dffree2)

##   
## Call:  
## lm(formula = Diff.Free ~ Free.Lag + Year, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7815 -0.5958 -0.0245 0.8827 2.2594   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -374.11313 95.19367 -3.930 0.000396 \*\*\*  
## Free.Lag -0.52701 0.13470 -3.912 0.000416 \*\*\*  
## Year 0.19114 0.04852 3.940 0.000385 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.239 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.3209, Adjusted R-squared: 0.2809   
## F-statistic: 8.032 on 2 and 34 DF, p-value: 0.001391

durbinWatsonTest(dffree2)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.1837122 2.342295 0.418  
## Alternative hypothesis: rho != 0

The Durbin-Watson statistic for this version of the Dickey-Fuller equation is 2.34. For k=3 and n=36, dL=1.30 (or equivalently 2.70) and dU=1.65 (or equivalently 2.35). As a result, there is evidence of no serial correlation and we can use this version of the Dickey-Fuller equation.

Regressing the first difference on the lagged freedom variable and the time variable generates a tau-value of -3.912. Since the Dickey-Fuller critical value for a sample size of 36 is ≈-3.60, we can reject the null hypothesis that there is a unit root and conclude that the trend of the Freedom of Press variable is stationary.

#### Dickey-Fuller test for Government Approval

media$Gov.App.Lag <- sapply(1:nrow(media), function(x) media$Gov.App[x-1])  
media$Gov.App.Lag <- car::recode(as.numeric(media$Gov.App.Lag),"numeric(0)=NA")  
  
media$Diff.Gov.App <- media$Polar-media$Gov.App.Lag  
  
# Regress first difference on the lagged variable  
dfgov1 <- lm(Diff.Gov.App ~ Gov.App.Lag, data = media)  
summary(dfgov1)

##   
## Call:  
## lm(formula = Diff.Gov.App ~ Gov.App.Lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.108258 -0.030321 -0.007352 0.039072 0.125770   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.899302 0.029673 30.31 <2e-16 \*\*\*  
## Gov.App.Lag -1.006598 0.000965 -1043.09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05579 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.088e+06 on 1 and 35 DF, p-value: < 2.2e-16

durbinWatsonTest(dfgov1)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.7566466 0.370332 0  
## Alternative hypothesis: rho != 0

Since the Durbin-Watson statistic for this version of the Dickey-Fuller equation is 0.370, there is evidence of strong serial correlation and we must add a t-value to the equation.

# Regress first difference on the lagged variable and the time variable  
dfgov2 <- lm(Diff.Gov.App ~ Gov.App.Lag + Year, data = media)  
summary(dfgov2)

##   
## Call:  
## lm(formula = Diff.Gov.App ~ Gov.App.Lag + Year, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.049549 -0.016078 0.003579 0.019964 0.049778   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.191e+01 1.208e+00 -9.862 1.66e-11 \*\*\*  
## Gov.App.Lag -1.002e+00 6.693e-04 -1496.533 < 2e-16 \*\*\*  
## Year 6.320e-03 5.958e-04 10.607 2.52e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02727 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 2.277e+06 on 2 and 34 DF, p-value: < 2.2e-16

durbinWatsonTest(dfgov2)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.7166828 0.4152208 0  
## Alternative hypothesis: rho != 0

Since the Durbin-Watson statistic is still very low (0.415) even after adding the time variable, we cannot use this version of the equation and must proceed onto the augmented Dickey-Fuller test.

# Augmented Dickey-Fuller Test   
media$Diff.Gov.App.Lag <- sapply(1:nrow(media), function(x) media$Diff.Gov.App[x-1])  
media$Diff.Gov.App.Lag <- car::recode(as.numeric(media$Diff.Gov.App.Lag), "numeric(0)=NA")  
  
# Regress the difference variable on the lagged variable, the time/trend variable and the lagged difference variable  
dfgov3 <- lm(Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag, data = media)  
summary(dfgov3)

##   
## Call:  
## lm(formula = Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag,   
## data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.047344 -0.018619 0.000627 0.017348 0.043362   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.156e+01 1.158e+00 -9.981 2.37e-11 \*\*\*  
## Gov.App.Lag -9.998e-01 9.957e-04 -1004.101 < 2e-16 \*\*\*  
## Year 6.149e-03 5.710e-04 10.769 3.59e-12 \*\*\*  
## Diff.Gov.App.Lag 2.311e-03 9.967e-04 2.319 0.027 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02471 on 32 degrees of freedom  
## (2 observations deleted due to missingness)  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.756e+06 on 3 and 32 DF, p-value: < 2.2e-16

durbinWatsonTest(dfgov3)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.6680926 0.5124578 0  
## Alternative hypothesis: rho != 0

The augmented Dickey-Fuller test still shows signs of strong serial correlation (since dw=0.512). As a result, I continued to add further lags.

media$Diff.Gov.App.Lag2 <- sapply(1:nrow(media), function(x) media$Diff.Gov.App.Lag[x-1])  
media$Diff.Gov.App.Lag2 <- car::recode(as.numeric(media$Diff.Gov.App.Lag2), "numeric(0)=NA")  
  
dfgov4 <- lm(Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag + Diff.Gov.App.Lag2, data = media)  
summary(dfgov4)

##   
## Call:  
## lm(formula = Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag +   
## Diff.Gov.App.Lag2, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.045220 -0.019070 0.004309 0.015211 0.036777   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.110e+01 1.133e+00 -9.797 7.36e-11 \*\*\*  
## Gov.App.Lag -1.000e+00 9.312e-04 -1073.863 < 2e-16 \*\*\*  
## Year 5.928e-03 5.584e-04 10.615 1.12e-11 \*\*\*  
## Diff.Gov.App.Lag 5.205e-04 1.189e-03 0.438 0.6648   
## Diff.Gov.App.Lag2 2.071e-03 9.298e-04 2.227 0.0336 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02289 on 30 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.422e+06 on 4 and 30 DF, p-value: < 2.2e-16

durbinWatsonTest(dfgov4)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.7150106 0.4472563 0  
## Alternative hypothesis: rho != 0

Since the dw=0.447, there is evidence of potential serial correlation and we cannot use this version of the Dickey-Fuller equation.

media$Diff.Gov.App.Lag3 <- sapply(1:nrow(media), function(x) media$Diff.Gov.App.Lag2[x-1])  
media$Diff.Gov.App.Lag3 <- car::recode(as.numeric(media$Diff.Gov.App.Lag3), "numeric(0)=NA")  
  
dfgov5 <- lm(Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag + Diff.Gov.App.Lag2 + Diff.Gov.App.Lag2, data = media)  
summary(dfgov5)

##   
## Call:  
## lm(formula = Diff.Gov.App ~ Gov.App.Lag + Year + Diff.Gov.App.Lag +   
## Diff.Gov.App.Lag2 + Diff.Gov.App.Lag2, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.045220 -0.019070 0.004309 0.015211 0.036777   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.110e+01 1.133e+00 -9.797 7.36e-11 \*\*\*  
## Gov.App.Lag -1.000e+00 9.312e-04 -1073.863 < 2e-16 \*\*\*  
## Year 5.928e-03 5.584e-04 10.615 1.12e-11 \*\*\*  
## Diff.Gov.App.Lag 5.205e-04 1.189e-03 0.438 0.6648   
## Diff.Gov.App.Lag2 2.071e-03 9.298e-04 2.227 0.0336 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02289 on 30 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 1.422e+06 on 4 and 30 DF, p-value: < 2.2e-16

durbinWatsonTest(dfgov5)

## lag Autocorrelation D-W Statistic p-value  
## 1 0.7150106 0.4472563 0  
## Alternative hypothesis: rho != 0

Despite adding several lags to the model, the serial correlation persits. At this point, adding further lags is inadvisable considering the data set is very small and each lag requires the deletion of another observation. Nevertheless, all the tau-values generated by the Dickey-Fuller equations are less than the Dickey-Fuller critical value of 3.60, so we can conclude that the Government Trust variable is non-stationary.

### Adjusting for non-stationarity: Cointegration

# Residuals of original regression  
mod4 <- lm(Trust ~ Polar+Social+Freedom+Gov.App, data = media)  
summary(mod4)

##   
## Call:  
## lm(formula = Trust ~ Polar + Social + Freedom + Gov.App, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9402 -2.8460 -0.5633 2.1621 7.2490   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 77.86341 18.01992 4.321 0.000134 \*\*\*  
## Polar -10.21386 31.17708 -0.328 0.745277   
## Social 0.06663 0.09414 0.708 0.484039   
## Freedom -1.84232 0.61683 -2.987 0.005284 \*\*   
## Gov.App 0.27436 0.11914 2.303 0.027725 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.017 on 33 degrees of freedom  
## Multiple R-squared: 0.8219, Adjusted R-squared: 0.8003   
## F-statistic: 38.07 on 4 and 33 DF, p-value: 6.305e-12

media$res2 <- residuals(mod4)  
  
# Create first difference of residuals  
media$res2lag <- sapply(1:nrow(media), function(x) media$res2[x-1])  
media$res2lag <- car::recode(as.numeric(media$res2lag), "numeric(0)=NA")  
media$difres2 <- media$res2-media$res2lag  
  
# Dickey-Fuller test  
mod5 <- lm(difres2 ~ res2lag, data = media)  
summary(mod5)

##   
## Call:  
## lm(formula = difres2 ~ res2lag, data = media)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.4029 -2.4781 -0.0188 1.5884 8.4306   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.2056 0.5823 -0.353 0.726   
## res2lag -0.7008 0.1548 -4.528 6.61e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.541 on 35 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.3694, Adjusted R-squared: 0.3514   
## F-statistic: 20.51 on 1 and 35 DF, p-value: 6.61e-05

durbinWatsonTest(mod5)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.002778049 1.985513 0.962  
## Alternative hypothesis: rho != 0

Performing these Dickey-Fuller tests revealed that the trends of the Trust in Media and Freedom of the Press variable are stationary, while Political Polarization, Social Media Usage, and Government Trust (and their trends) are all non-statioinary. Since we can only be confident that spurious regression is not interfering with our analysis if ALL variables are stationary, we test for co-integration to determine if the nature of the non-stationary present in Polar,Social, and Gov.App is similar.

Since the tau-value generated from regressing the difference in residuals on the lagged residuals is -4.528 and the critical Dickey-Fuller value for 38 observations is ≈-3.00, we reject the null hypothesis that there is a unit root (i.e. that there is non-stationarity). Since the effect of the residual lag term on the difference is residuals is stationary, we can conclude that the model is cointegrated. In other words, the original OLS model is non-spurious.

### Testing for serial correlation

Once the model has been proven to be cointegrated, it is important to test for another common issue in time series analysis: serial correlation. Serially correlated models can lead to underestimated standard errors and exaggerated estimations of goodness of fit.

durbinWatsonTest(mod1)

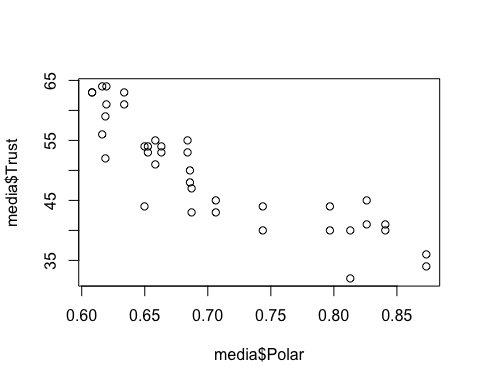
## lag Autocorrelation D-W Statistic p-value  
## 1 0.2932861 1.312211 0.004  
## Alternative hypothesis: rho != 0

Using the Durbin-Watson test to test for serial correlation in the OLS model generates a statistic of 1.31. Since the dL = 1.261 and dU = 1.722 for a model with n=38 and k=4, there is evidence of potential serial correlation.

### Adjusting for serial correlation: Prais-Winsten standard errors

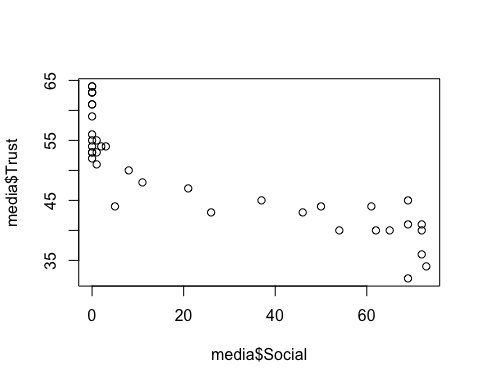
To determine if the serial correlation occurs as a consequence of a misspecified functional form, I use scatter plots comparing each predictor variable against the response variable Trust in Media to informally test whether the relationships are non-linear.

## Scatter plots  
plot(media$Polar, media$Trust)



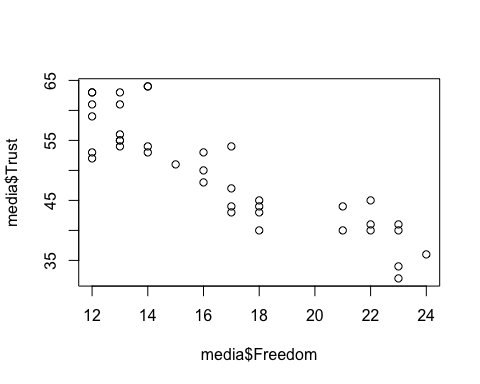
The relationship between Media Trust and Political Polarization appears to be negative and generally linear.

plot(media$Social, media$Trust)



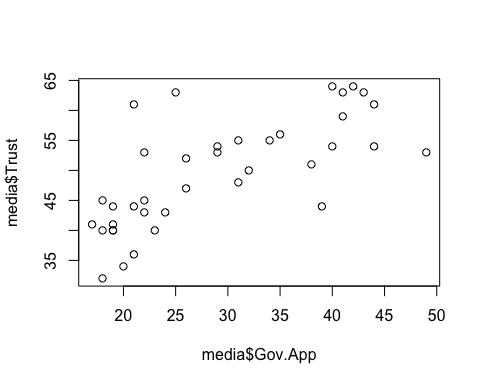
The relationship between Media Trust and Social Media Usage appears to be negative and generally linear.

plot(media$Freedom, media$Trust)



The relationship between Media Trust and Media Freedom appears to be negative and generally linear.

plot(media$Gov.App, media$Trust)



The relationship between Media Trust and Government Trust appears to be positive and generally linear. Since none of the relationships seem to be misspecified, we assume the model’s serial correlation is pure.

pw <- prais\_winsten(Trust ~ Polar+Social+Freedom+Gov.App, data = media, index="Year")

## Iteration 0: rho = 0  
## Iteration 1: rho = 0.2981  
## Iteration 2: rho = 0.3503  
## Iteration 3: rho = 0.3652  
## Iteration 4: rho = 0.3699  
## Iteration 5: rho = 0.3715  
## Iteration 6: rho = 0.372  
## Iteration 7: rho = 0.3722  
## Iteration 8: rho = 0.3722  
## Iteration 9: rho = 0.3722  
## Iteration 10: rho = 0.3722  
## Iteration 11: rho = 0.3722  
## Iteration 12: rho = 0.3722

summary(pw)

##   
## Call:  
## prais\_winsten(formula = Trust ~ Polar + Social + Freedom + Gov.App,   
## data = media, index = "Year")  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.9805 -3.4017 -0.5109 2.3010 7.2713   
##   
## AR(1) coefficient rho after 12 iterations: 0.3722  
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 63.45805 22.34034 2.841 0.00766 \*\*  
## Polar 5.96144 36.40877 0.164 0.87094   
## Social -0.02222 0.11277 -0.197 0.84497   
## Freedom -1.52155 0.57261 -2.657 0.01205 \*   
## Gov.App 0.27247 0.13901 1.960 0.05848 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.776 on 33 degrees of freedom  
## Multiple R-squared: 0.7848, Adjusted R-squared: 0.7587   
## F-statistic: 30.09 on 4 and 33 DF, p-value: 1.37e-10  
##   
## Durbin-Watson statistic (original): 1.312   
## Durbin-Watson statistic (transformed): 1.93

## 

## Conclusion: Comparison

compare <- na.omit(tibble(coefficients = c( "Intercept", "Political Polarization", "Social Media Usage", "Freedom of Press", "Government Trust"),OLS = c(summary(mod1)$coef[,1]), PW = c(summary(pw)$coef[,1]),OLS.Std.Err = c(summary(mod1)$coef[,2]),PW.Std.Err = c(summary(pw)$coef[,2]),OLS.Sig = c(summary(mod1)$coef[,4]),PW.Sig = c(summary(pw)$coef[,4])))  
data.frame(compare)

## coefficients OLS PW OLS.Std.Err PW.Std.Err  
## 1 Intercept 77.86340742 63.45804976 18.01991948 22.3403444  
## 2 Political Polarization -10.21386389 5.96143837 31.17708005 36.4087703  
## 3 Social Media Usage 0.06663359 -0.02222437 0.09414189 0.1127668  
## 4 Freedom of Press -1.84232162 -1.52155131 0.61682765 0.5726128  
## 5 Government Trust 0.27436183 0.27246549 0.11913580 0.1390124  
## OLS.Sig PW.Sig  
## 1 0.0001341512 0.007658265  
## 2 0.7452773814 0.870937427  
## 3 0.4840391822 0.844971747  
## 4 0.0052836093 0.012048471  
## 5 0.0277248534 0.058484073

The above table compares the OLS model’s coefficient estimates, standard errors, and p-values with the adjusted versions generated using the Prais-Winsten procedure. Comparing the two models, it appears that adjusting for serial correlation causes the magnitude of nearly every variable’s effect to decrease (the effect of Polar decreases from |-10.21| to |5.96|, the effect of Social Media Usage decreases from |0.066| to |-0.022|, and the effect of Press Freedom decreases from |-1.842| to |-1.521|). In addition, the Prais-Winsten adjustments increase all of the OLS model’s standard errors and p-values, suggesting that serial correlation made the model’s parameter estimates appear more precise than they actually are. After adjusting for serial correlation, the variables for Freedom of Press and Government Trust lose their statistical significance at the 0.05 level.

Notably, the Prais-Winsten model switches the sign/direction of effect for both Political Polarization and Social Media Usage ­– a potential indicator of multicollinearity, which aligns with the findings of the initial correlation matrix.

Since the Prais-Winsten version of the model adjusts for serial correlation, it is the preferred model. Interpreting the adjusted coefficients, a 1%-increase in political polarization leads to a 5.96%-increase in media trust; a 1%-increase in social media usage leads to a 0.02%-decrease in media trust; a 1-unit increase in the score of media independence (indicating a decline in press freedom) leads to a 1.84%-decrease in media trust, and a 1%-increase in government trust leads to a 0.27%-increase in media trust. The direction of the Social, Freedom, and Gov.App variables within the model support the paper’s causal theory; however, the direction of the Political Polarization effect is positive, which suggests that increased political polarization leads to increased media trust over time and contradicts the initial theory.