hw06.R

Brian Kang

Mon Mar 11 20:41:43 2019

# hw6  
# ch.10  
# C1  
library("wooldridge")

## Warning: package 'wooldridge' was built under R version 3.5.2

data(intdef)  
dat1 <- intdef  
y79 <- ifelse(dat1$year>1979,1,0) # if true then 1, if false then 0  
cbind(dat1, y79) # finished making and adding dummy variable

## year i3 inf rec out def i3\_1 inf\_1 def\_1 ci3  
## 1 1948 1.04 8.1 16.2 11.6 -4.6000004 NA NA NA NA  
## 2 1949 1.10 -1.2 14.5 14.3 -0.1999998 1.04 8.1 -4.6000004 0.06000006  
## 3 1950 1.22 1.3 14.4 15.6 1.2000008 1.10 -1.2 -0.1999998 0.12000000  
## 4 1951 1.55 7.9 16.1 14.2 -1.9000006 1.22 1.3 1.2000008 0.32999992  
## 5 1952 1.77 1.9 19.0 19.4 0.3999996 1.55 7.9 -1.9000006 0.22000003  
## 6 1953 1.93 0.8 18.7 20.4 1.6999989 1.77 1.9 0.3999996 0.15999997  
## 7 1954 0.95 0.7 18.5 18.8 0.2999992 1.93 0.8 1.6999989 -0.97999996  
## 8 1955 1.75 -0.4 16.5 17.3 0.7999992 0.95 0.7 0.2999992 0.80000001  
## 9 1956 2.66 1.5 17.5 16.5 -1.0000000 1.75 -0.4 0.7999992 0.91000009  
## 10 1957 3.27 3.3 17.7 17.0 -0.7000008 2.66 1.5 -1.0000000 0.60999990  
## 11 1958 1.84 2.8 17.3 17.9 0.6000004 3.27 3.3 -0.7000008 -1.42999995  
## 12 1959 3.41 0.7 16.2 18.8 2.5999985 1.84 2.8 0.6000004 1.57000005  
## 13 1960 2.93 1.7 17.8 17.8 0.0000000 3.41 0.7 2.5999985 -0.48000002  
## 14 1961 2.38 1.0 17.8 18.4 0.6000004 2.93 1.7 0.0000000 -0.54999995  
## 15 1962 2.78 1.0 17.6 18.8 1.1999989 2.38 1.0 0.6000004 0.39999986  
## 16 1963 3.16 1.3 17.8 18.6 0.8000011 2.78 1.0 1.1999989 0.38000011  
## 17 1964 3.55 1.3 17.6 18.5 0.8999996 3.16 1.3 0.8000011 0.38999987  
## 18 1965 3.95 1.6 17.0 17.2 0.2000008 3.55 1.3 0.8999996 0.40000010  
## 19 1966 4.88 2.9 17.3 17.8 0.5000000 3.95 1.6 0.2000008 0.93000007  
## 20 1967 4.32 3.1 18.4 19.4 1.0000000 4.88 2.9 0.5000000 -0.55999994  
## 21 1968 5.34 4.2 17.6 20.5 2.8999996 4.32 3.1 1.0000000 1.01999998  
## 22 1969 6.68 5.5 19.7 19.4 -0.3000011 5.34 4.2 2.8999996 1.33999968  
## 23 1970 6.46 5.7 19.0 19.3 0.2999992 6.68 5.5 -0.3000011 -0.21999979  
## 24 1971 4.35 4.4 17.3 19.5 2.2000008 6.46 5.7 0.2999992 -2.11000013  
## 25 1972 4.07 3.2 17.6 19.6 2.0000000 4.35 4.4 2.2000008 -0.27999973  
## 26 1973 7.04 6.2 17.6 18.7 1.1000004 4.07 3.2 2.0000000 2.96999979  
## 27 1974 7.89 11.0 18.3 18.7 0.4000015 7.04 6.2 1.1000004 0.84999990  
## 28 1975 5.84 9.1 17.9 21.3 3.3999996 7.89 11.0 0.4000015 -2.04999971  
## 29 1976 4.99 5.8 17.1 21.4 4.2999992 5.84 9.1 3.3999996 -0.85000038  
## 30 1977 5.27 6.5 18.0 20.7 2.7000008 4.99 5.8 4.2999992 0.28000021  
## 31 1978 7.22 7.6 18.0 20.7 2.7000008 5.27 6.5 2.7000008 1.94999981  
## 32 1979 10.04 11.3 18.5 20.1 1.6000004 7.22 7.6 2.7000008 2.82000017  
## 33 1980 11.51 13.5 19.0 21.7 2.7000008 10.04 11.3 1.6000004 1.47000027  
## 34 1981 14.03 10.3 19.6 22.2 2.6000004 11.51 13.5 2.7000008 2.51999950  
## 35 1982 10.69 6.2 19.2 23.1 3.8999996 14.03 10.3 2.6000004 -3.34000015  
## 36 1983 8.63 3.2 17.4 23.5 6.1000004 10.69 6.2 3.8999996 -2.05999947  
## 37 1984 9.58 4.3 17.3 22.1 4.8000011 8.63 3.2 6.1000004 0.94999981  
## 38 1985 7.48 3.6 17.7 22.8 5.0999985 9.58 4.3 4.8000011 -2.09999990  
## 39 1986 5.98 1.9 17.5 22.5 5.0000000 7.48 3.6 5.0999985 -1.50000000  
## 40 1987 5.82 3.6 18.4 21.6 3.2000008 5.98 1.9 5.0000000 -0.15999985  
## 41 1988 6.69 4.1 18.1 21.2 3.1000004 5.82 3.6 3.2000008 0.86999989  
## 42 1989 8.12 4.8 18.3 21.2 2.9000015 6.69 4.1 3.1000004 1.42999983  
## 43 1990 7.51 5.4 18.0 21.8 3.7999992 8.12 4.8 2.9000015 -0.60999966  
## 44 1991 5.42 4.2 17.8 22.3 4.5000000 7.51 5.4 3.7999992 -2.09000015  
## 45 1992 3.45 3.0 17.5 22.1 4.6000004 5.42 4.2 4.5000000 -1.97000003  
## 46 1993 3.02 3.0 17.5 21.4 3.8999996 3.45 3.0 4.6000004 -0.43000007  
## 47 1994 4.29 2.6 18.1 21.0 2.8999996 3.02 3.0 3.8999996 1.26999998  
## 48 1995 5.51 2.8 18.5 20.7 2.2000008 4.29 2.6 2.8999996 1.22000027  
## 49 1996 5.02 3.0 18.9 20.3 1.3999996 5.51 2.8 2.2000008 -0.49000025  
## 50 1997 5.07 2.3 19.3 19.6 0.3000011 5.02 3.0 1.3999996 0.05000019  
## 51 1998 4.81 1.6 20.0 19.2 -0.7999992 5.07 2.3 0.3000011 -0.26000023  
## 52 1999 4.66 2.2 20.0 18.6 -1.3999996 4.81 1.6 -0.7999992 -0.15000010  
## 53 2000 5.85 3.4 20.9 18.4 -2.5000000 4.66 2.2 -1.3999996 1.19000006  
## 54 2001 3.45 2.8 19.8 18.6 -1.1999989 5.85 3.4 -2.5000000 -2.39999986  
## 55 2002 1.62 1.6 17.9 19.4 1.5000000 3.45 2.8 -1.1999989 -1.83000004  
## 56 2003 1.02 2.3 16.5 19.9 3.3999996 1.62 1.6 1.5000000 -0.60000002  
## cinf cdef y77 y79  
## 1 NA NA 0 0  
## 2 -9.3000002 4.40000057 0 0  
## 3 2.5000000 1.40000057 0 0  
## 4 6.6000004 -3.10000134 0 0  
## 5 -6.0000000 2.30000019 0 0  
## 6 -1.0999999 1.29999924 0 0  
## 7 -0.1000000 -1.39999962 0 0  
## 8 -1.1000000 0.50000000 0 0  
## 9 1.9000000 -1.79999924 0 0  
## 10 1.8000000 0.29999924 0 0  
## 11 -0.5000000 1.30000114 0 0  
## 12 -2.0999999 1.99999809 0 0  
## 13 1.0000000 -2.59999847 0 0  
## 14 -0.7000000 0.60000038 0 0  
## 15 0.0000000 0.59999847 0 0  
## 16 0.3000000 -0.39999771 0 0  
## 17 0.0000000 0.09999847 0 0  
## 18 0.3000001 -0.69999886 0 0  
## 19 1.3000001 0.29999924 0 0  
## 20 0.1999998 0.50000000 0 0  
## 21 1.0999999 1.89999962 0 0  
## 22 1.3000002 -3.20000076 0 0  
## 23 0.1999998 0.60000038 0 0  
## 24 -1.2999997 1.90000153 0 0  
## 25 -1.2000000 -0.20000076 0 0  
## 26 2.9999998 -0.89999962 0 0  
## 27 4.8000002 -0.69999886 0 0  
## 28 -1.8999996 2.99999809 0 0  
## 29 -3.3000002 0.89999962 0 0  
## 30 0.6999998 -1.59999847 1 0  
## 31 1.0999999 0.00000000 1 0  
## 32 3.7000003 -1.10000038 1 0  
## 33 2.1999998 1.10000038 1 1  
## 34 -3.1999998 -0.10000038 1 1  
## 35 -4.1000004 1.29999924 1 1  
## 36 -2.9999998 2.20000076 1 1  
## 37 1.1000001 -1.29999924 1 1  
## 38 -0.7000003 0.29999733 1 1  
## 39 -1.6999999 -0.09999847 1 1  
## 40 1.6999999 -1.79999924 1 1  
## 41 0.5000000 -0.10000038 1 1  
## 42 0.7000003 -0.19999886 1 1  
## 43 0.5999999 0.89999771 1 1  
## 44 -1.2000003 0.70000076 1 1  
## 45 -1.1999998 0.10000038 1 1  
## 46 0.0000000 -0.70000076 1 1  
## 47 -0.4000001 -1.00000000 1 1  
## 48 0.2000000 -0.69999886 1 1  
## 49 0.2000000 -0.80000114 1 1  
## 50 -0.7000000 -1.09999847 1 1  
## 51 -0.6999999 -1.10000038 1 1  
## 52 0.6000000 -0.60000038 1 1  
## 53 1.2000000 -1.10000038 1 1  
## 54 -0.6000001 1.30000114 1 1  
## 55 -1.1999999 2.69999886 1 1  
## 56 0.6999999 1.89999962 1 1

lm.1 <- lm(i3 ~ inf+def+y79, data = dat1)  
summary(lm.1)

##   
## Call:  
## lm(formula = i3 ~ inf + def + y79, data = dat1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.4674 -0.8407 0.2388 1.0148 3.9654   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.29623 0.42535 3.047 0.00362 \*\*   
## inf 0.60842 0.07625 7.979 1.37e-10 \*\*\*  
## def 0.36266 0.12025 3.016 0.00396 \*\*   
## y79 1.55877 0.50577 3.082 0.00329 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.711 on 52 degrees of freedom  
## Multiple R-squared: 0.6635, Adjusted R-squared: 0.6441   
## F-statistic: 34.18 on 3 and 52 DF, p-value: 2.408e-12

# I say that there is a shift in the interest rate equation  
# after 1979. We see a decrease in intercept from 1.73 to 1.2962,  
# decrease in coefficient of def from 0.513 to 0.3626. Also,  
# interpreting the coefficient of our dummy, we expect that the interest  
# rate i3 to increase by 1.5587 per increase in years after 1979 on avg.,  
# holding all else constant.  
  
# C7  
# i)  
data(consump)  
dat2 <- consump  
lm.2 <- lm(gc ~ gy, data = dat2)  
summary(lm.2)

##   
## Call:  
## lm(formula = gc ~ gy, data = dat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0140496 -0.0035407 -0.0005813 0.0044080 0.0116890   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.008079 0.001899 4.254 0.000155 \*\*\*  
## gy 0.570781 0.067354 8.474 6.75e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.007268 on 34 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.6787, Adjusted R-squared: 0.6692   
## F-statistic: 71.81 on 1 and 34 DF, p-value: 6.754e-10

# Form: gc = 0.008079 + 0.570781\*gy  
# Interpretation:  
# I: We estimate that when gy=0, gc will equal 0.008079, all else equal.  
# Slope: On average, we estimate that gc will increase by 0.570781 per increase  
# in gy, hold all else equal.  
# Both intercept and slope are statistically significant even at the 0.001 level  
  
# ii)  
tsdat1 <- ts(dat2) # create time series data  
# linear regression of gc on lag of gy  
#install.packages("dynlm")  
library(dynlm)

## Warning: package 'dynlm' was built under R version 3.5.2

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

# gc ~ gy + L(gy)  
res1 <- dynlm(gc ~ gy + L(gy), data = tsdat1)  
summary(res1)

##   
## Time series regression with "ts" data:  
## Start = 3, End = 37  
##   
## Call:  
## dynlm(formula = gc ~ gy + L(gy), data = tsdat1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0129446 -0.0041097 -0.0002394 0.0049605 0.0138004   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.006357 0.002262 2.811 0.00837 \*\*   
## gy 0.552251 0.069651 7.929 4.76e-09 \*\*\*  
## L(gy) 0.096213 0.069019 1.394 0.17292   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.007267 on 32 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.6953, Adjusted R-squared: 0.6762   
## F-statistic: 36.51 on 2 and 32 DF, p-value: 5.522e-09

# We see that the t-stat for the lag is 1.394, insignificant even at the 10%  
# confidence level. Also, the lag's slope is not large and the slopes of other  
# variables have not altered much too. We cannot say that adjustment lags  
# are really affecting consumptions.  
  
# iii)  
lm.3 <- lm(gc ~ gy+r3, data = dat2)  
summary(lm.3)

##   
## Call:  
## lm(formula = gc ~ gy + r3, data = dat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0135088 -0.0034771 -0.0005361 0.0036980 0.0125572   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0082181 0.0019665 4.179 0.000202 \*\*\*  
## gy 0.5781105 0.0715164 8.084 2.5e-09 \*\*\*  
## r3 -0.0002148 0.0006265 -0.343 0.733901   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.007364 on 33 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.6798, Adjusted R-squared: 0.6604   
## F-statistic: 35.03 on 2 and 33 DF, p-value: 6.902e-09

# The t-stat for the coefficient of r3 is -0.343 and the p-value is  
# 0.733901. It is insignificant at the 5% confidence level.  
# Also, the slope itself is very small -0.0002148. Therefore, we cannot  
# say that real interest rate affects consumption growth.  
  
# C11  
data(traffic2)  
dat3 <- traffic2  
# i)  
temp <- (subset(dat3, dat3[,"spdlaw"]==1)$t)[1] # the time trend  
dat3[dat3$t==temp,]

## year totacc fatacc injacc pdoacc ntotacc nfatacc ninjacc npdoacc  
## 77 1987 47072 421 20406 26245 43381 368 18967 24046  
## rtotacc rfatacc rinjacc rpdoacc ushigh cntyrds strtes t tsq unem  
## 77 375 18 174 183 1248 7306 3704 77 5929 5.6  
## spdlaw beltlaw wkends feb mar apr may jun jul aug sep oct nov dec  
## 77 1 1 15 0 0 0 1 0 0 0 0 0 0 0  
## ltotacc lfatacc prcfat prcrfat lrtotacc lrfatacc lntotacc lnfatacc  
## 77 10.75943 6.042633 0.8943746 4.8 5.926926 2.890372 10.67778 5.908083  
## prcnfat lushigh lcntyrds lstrtes spdt beltt prcfat\_1  
## 77 0.8482977 7.129298 8.896451 8.217169 77 77 0.8773477

# 1987, May the Speed Law was passed  
temp <- (subset(dat3, dat3[,"beltlaw"]==1)$t)[1]  
dat3[dat3$t==temp,]

## year totacc fatacc injacc pdoacc ntotacc nfatacc ninjacc npdoacc  
## 61 1986 43522 339 16895 26288 39900 305 15577 24018  
## rtotacc rfatacc rinjacc rpdoacc ushigh cntyrds strtes t tsq unem  
## 61 246 8 102 136 1060 6398 3258 61 3721 6.6  
## spdlaw beltlaw wkends feb mar apr may jun jul aug sep oct nov dec  
## 61 0 1 13 0 0 0 0 0 0 0 0 0 0 0  
## ltotacc lfatacc prcfat prcrfat lrtotacc lrfatacc lntotacc lnfatacc  
## 61 10.68102 5.826 0.7789164 3.252033 5.505332 2.079442 10.59413 5.720312  
## prcnfat lushigh lcntyrds lstrtes spdt beltt prcfat\_1  
## 61 0.764411 6.966024 8.763741 8.088869 0 61 0.7834257

# 1986, January the Seatbelt Law was passed  
  
# ii)  
tsdat2 <- ts(dat3, start = 1981, frequency = 12)  
#months <- paste(names(dat3)[c(23:33)],sep = "",collapse = "+")  
res2 <- dynlm(log(totacc) ~ t+feb+mar+apr+may+jun+jul+aug+sep+oct+nov+dec,data = tsdat2)  
summary(res2)

##   
## Time series regression with "ts" data:  
## Start = 1981(1), End = 1989(12)  
##   
## Call:  
## dynlm(formula = log(totacc) ~ t + feb + mar + apr + may + jun +   
## jul + aug + sep + oct + nov + dec, data = tsdat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.100216 -0.039778 -0.002915 0.038747 0.138015   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.4685664 0.0190028 550.895 < 2e-16 \*\*\*  
## t 0.0027471 0.0001611 17.057 < 2e-16 \*\*\*  
## feb -0.0426864 0.0244475 -1.746 0.084036 .   
## mar 0.0798245 0.0244491 3.265 0.001523 \*\*   
## apr 0.0184850 0.0244517 0.756 0.451532   
## may 0.0320983 0.0244554 1.313 0.192506   
## jun 0.0201919 0.0244602 0.825 0.411156   
## jul 0.0375825 0.0244660 1.536 0.127835   
## aug 0.0539828 0.0244729 2.206 0.029808 \*   
## sep 0.0423611 0.0244809 1.730 0.086810 .   
## oct 0.0821134 0.0244899 3.353 0.001149 \*\*   
## nov 0.0712786 0.0244999 2.909 0.004511 \*\*   
## dec 0.0961571 0.0245110 3.923 0.000165 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05186 on 95 degrees of freedom  
## Multiple R-squared: 0.7969, Adjusted R-squared: 0.7712   
## F-statistic: 31.06 on 12 and 95 DF, p-value: < 2.2e-16

# F-test  
library(car)

## Loading required package: carData

null = c("feb","mar","apr","may","jun","jul","aug","sep","oct","nov","dec")  
linearHypothesis(res2,null)

## Linear hypothesis test  
##   
## Hypothesis:  
## feb = 0  
## mar = 0  
## apr = 0  
## may = 0  
## jun = 0  
## jul = 0  
## aug = 0  
## sep = 0  
## oct = 0  
## nov = 0  
## dec = 0  
##   
## Model 1: restricted model  
## Model 2: log(totacc) ~ t + feb + mar + apr + may + jun + jul + aug + sep +   
## oct + nov + dec  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 106 0.40786   
## 2 95 0.25550 11 0.15236 5.1501 2.712e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# We can see that, basing on Jan, only Feb has a lower totacc (only it has neg. coeff).  
# Also, Dec is the peak, with 9.61571% more totacc than Jan.  
# From the F-test we get a F-stat of 5.1501 with DF 11,95. So, the monthly dummies are  
# jointly statistically significant even at a 0.1% confidence level.  
# Individually, feb,mar,aug - dev were statistically significant at the 5% level.  
# Also, the coefficient of trend variable t shows that, ignoring seasonality,  
# totacc increases by 0.27471% per month over the time series we have.  
# Therefore, we conclude that there is evidence of seasonality.  
  
# iii)  
res3 <- dynlm(log(totacc) ~ wkends+unem+spdlaw+beltlaw+t+feb+mar+apr+may+jun+jul+aug+sep+oct+nov+dec,data = tsdat2)  
summary(res3)

##   
## Time series regression with "ts" data:  
## Start = 1981(1), End = 1989(12)  
##   
## Call:  
## dynlm(formula = log(totacc) ~ wkends + unem + spdlaw + beltlaw +   
## t + feb + mar + apr + may + jun + jul + aug + sep + oct +   
## nov + dec, data = tsdat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.084489 -0.023998 -0.002176 0.024586 0.089315   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.6398628 0.0630859 168.657 < 2e-16 \*\*\*  
## wkends 0.0033333 0.0037761 0.883 0.37970   
## unem -0.0212173 0.0033974 -6.245 1.33e-08 \*\*\*  
## spdlaw -0.0537593 0.0126036 -4.265 4.87e-05 \*\*\*  
## beltlaw 0.0954526 0.0142350 6.705 1.65e-09 \*\*\*  
## t 0.0011011 0.0002579 4.270 4.79e-05 \*\*\*  
## feb -0.0338346 0.0177683 -1.904 0.06004 .   
## mar 0.0769530 0.0167941 4.582 1.46e-05 \*\*\*  
## apr 0.0104562 0.0170469 0.613 0.54115   
## may 0.0237075 0.0169389 1.400 0.16504   
## jun 0.0219334 0.0172149 1.274 0.20587   
## jul 0.0499292 0.0167036 2.989 0.00360 \*\*   
## aug 0.0559523 0.0168173 3.327 0.00127 \*\*   
## sep 0.0420694 0.0172819 2.434 0.01687 \*   
## oct 0.0817170 0.0169554 4.820 5.73e-06 \*\*\*  
## nov 0.0768721 0.0172455 4.458 2.36e-05 \*\*\*  
## dec 0.0990862 0.0170705 5.805 9.32e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.03524 on 91 degrees of freedom  
## Multiple R-squared: 0.9101, Adjusted R-squared: 0.8943   
## F-statistic: 57.61 on 16 and 91 DF, p-value: < 2.2e-16

# Slope of unem: -0.0212173  
# Yes this makes sense. Higher unemployment means a weaker and more unstable economy,  
# so it is likly that less cars will be on the street and so less car accidents will  
# occure. Also, 1% increase in unem causing a decrease in car accidents by 2.12173%  
# could seem plausible due to the decrease in economic activity.  
  
# iv)  
# spdlaw: We estimate that, after the spdlaw passes, log(totacc) will  
# decrease by 5.37593% on average, holding all else constant.  
# beltlaw: We estimate that, after the beltlaw passes, log(totacc) will  
# increase by 9.54526% on average, holding all else constant.  
# The effects of spdlaw make sense. With the spdlaw, cars will likely slow down,  
# and the liklihood that car accidents happening will decrease.  
# The effect of beltlaw could make sense, but not likely. With the belt law  
# requirement, more people put on seatbelts and this could cause a false sense of  
# security, driving people to drive more dangerously and cause accidents.  
# This would be contrary to the intention: mandating seatbels for people's safety.  
  
# v)  
mean(dat3$prcfat)

## [1] 0.8856363

# Avg. percentage of fatal accidents is 0.8856363. Less that 1% of accidents  
# turns out to be fatal apparently. Highest value is 1.217. I think this makes sense  
# because fatal would be defined "life threatening or impaling".  
# Although many accidents are very dangerous, I think the small pertencage of  
# "fatal" accidents occuring makes logical sense.  
  
# vi)  
res4 <- dynlm(prcfat ~ wkends+unem+spdlaw+beltlaw+t+feb+mar+apr+may+jun+jul+aug+sep+oct+nov+dec,data = tsdat2)  
summary(res4)

##   
## Time series regression with "ts" data:  
## Start = 1981(1), End = 1989(12)  
##   
## Call:  
## dynlm(formula = prcfat ~ wkends + unem + spdlaw + beltlaw + t +   
## feb + mar + apr + may + jun + jul + aug + sep + oct + nov +   
## dec, data = tsdat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.132085 -0.032573 0.000621 0.038226 0.132080   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.030e+00 1.030e-01 10.003 2.52e-16 \*\*\*  
## wkends 6.259e-04 6.162e-03 0.102 0.919328   
## unem -1.543e-02 5.544e-03 -2.782 0.006563 \*\*   
## spdlaw 6.709e-02 2.057e-02 3.262 0.001560 \*\*   
## beltlaw -2.951e-02 2.323e-02 -1.270 0.207288   
## t -2.235e-03 4.208e-04 -5.312 7.66e-07 \*\*\*  
## feb 8.607e-04 2.900e-02 0.030 0.976384   
## mar 9.226e-05 2.741e-02 0.003 0.997321   
## apr 5.822e-02 2.782e-02 2.093 0.039152 \*   
## may 7.164e-02 2.764e-02 2.592 0.011129 \*   
## jun 1.013e-01 2.809e-02 3.604 0.000510 \*\*\*  
## jul 1.766e-01 2.726e-02 6.479 4.64e-09 \*\*\*  
## aug 1.926e-01 2.744e-02 7.018 3.91e-10 \*\*\*  
## sep 1.600e-01 2.820e-02 5.674 1.64e-07 \*\*\*  
## oct 1.010e-01 2.767e-02 3.651 0.000435 \*\*\*  
## nov 1.395e-02 2.814e-02 0.496 0.621345   
## dec 9.200e-03 2.786e-02 0.330 0.741960   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05751 on 91 degrees of freedom  
## Multiple R-squared: 0.7174, Adjusted R-squared: 0.6677   
## F-statistic: 14.44 on 16 and 91 DF, p-value: < 2.2e-16

# F-test  
library(car)  
null = c("spdlaw","beltlaw")  
linearHypothesis(res4,null)

## Linear hypothesis test  
##   
## Hypothesis:  
## spdlaw = 0  
## beltlaw = 0  
##   
## Model 1: restricted model  
## Model 2: prcfat ~ wkends + unem + spdlaw + beltlaw + t + feb + mar + apr +   
## may + jun + jul + aug + sep + oct + nov + dec  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 93 0.33929   
## 2 91 0.30102 2 0.038271 5.7847 0.004316 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

null = c("spdlaw=1","beltlaw=1")  
linearHypothesis(res4,null)

## Linear hypothesis test  
##   
## Hypothesis:  
## spdlaw = 1  
## beltlaw = 1  
##   
## Model 1: restricted model  
## Model 2: prcfat ~ wkends + unem + spdlaw + beltlaw + t + feb + mar + apr +   
## may + jun + jul + aug + sep + oct + nov + dec  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 93 14.998   
## 2 91 0.301 2 14.697 2221.5 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# spdlaw: 6.709e-02; beltlaw: -2.951e-02  
# With the speed law, interestingly, the percentage of fatal accidents  
# increase, and with the belt law it decreases, as it should.  
# The spdlaw is individually significant at the 1% level, but  
# beltlaw is insignificant.  
# Jointly they are significant.  
  
  
# ch.11  
# C9  
# i)  
tsdat3 <- ts(dat3)  
acf(tsdat3[,"prcfat"],plot = F,lag.max = 1) # 1st Order autocorr coeff

##   
## Autocorrelations of series 'tsdat3[, "prcfat"]', by lag  
##   
## 0 1   
## 1.000 0.708

# The 1st order autocorr coeff is high but it should not be a huge  
# concern because it is does not contain a unit root.  
acf(tsdat3[,"unem"],plot = F,lag.max = 1)

##   
## Autocorrelations of series 'tsdat3[, "unem"]', by lag  
##   
## 0 1   
## 1.000 0.941

# But for unem, the 1st order autocorr coeff is very high (close to 1)  
# so we should be concerned about it possibily containing a unit root.  
  
# ii)  
res5 <- dynlm(d(prcfat) ~ wkends+d(unem)+spdlaw+beltlaw+t+feb+mar+apr+may+jun+jul+aug+sep+oct+nov+dec,data = tsdat3)  
summary(res5)

##   
## Time series regression with "ts" data:  
## Start = 2, End = 108  
##   
## Call:  
## dynlm(formula = d(prcfat) ~ wkends + d(unem) + spdlaw + beltlaw +   
## t + feb + mar + apr + may + jun + jul + aug + sep + oct +   
## nov + dec, data = tsdat3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.130088 -0.044411 0.002195 0.032828 0.166416   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.1268680 0.1048114 -1.210 0.22928   
## wkends 0.0068097 0.0072276 0.942 0.34862   
## d(unem) 0.0125342 0.0161094 0.778 0.43857   
## spdlaw -0.0071825 0.0237979 -0.302 0.76349   
## beltlaw 0.0008251 0.0265048 0.031 0.97523   
## t 0.0001433 0.0004849 0.296 0.76823   
## feb 0.0346228 0.0370460 0.935 0.35250   
## mar 0.0419346 0.0389248 1.077 0.28422   
## apr 0.0985703 0.0382988 2.574 0.01170 \*   
## may 0.0568102 0.0374416 1.517 0.13270   
## jun 0.0540339 0.0347738 1.554 0.12373   
## jul 0.0878394 0.0331103 2.653 0.00943 \*\*  
## aug 0.0589255 0.0396686 1.485 0.14092   
## sep 0.0065431 0.0379741 0.172 0.86358   
## oct -0.0323897 0.0352025 -0.920 0.35998   
## nov -0.0591083 0.0354151 -1.669 0.09859 .   
## dec 0.0272794 0.0363245 0.751 0.45462   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06724 on 90 degrees of freedom  
## Multiple R-squared: 0.3436, Adjusted R-squared: 0.2269   
## F-statistic: 2.945 on 16 and 90 DF, p-value: 0.0006167

summary(res5)$r.squared

## [1] 0.3436279

# We see some seasonality with only Oct and Nov being lower than  
# Jan. The slope of d(unem) is small and individually insignificant  
# so we cannot say that the difference in prcfat can be explained  
# by the difference in the difference of unemployment. Also,  
# most of the variables are considered insignificant. Even the trend  
# is showing insignificance and the slope is small. Note that  
# the high R2 value is affected by seasonality.  
  
# iii)  
# The reasoning isn't entirely correct. It is not necessarily the safest   
# strategy and we end up not getting similar results using the levels,   
# we mostly lose the interesting interpretations in our model. However,  
# this shouldn't be the reason for not using the different levels in  
# our models. It is difficult to say when to take the 1st order  
# differences.  
  
# C13  
# i)  
##############################################  
#CHAPTER 11 HOMEWORK PROBLEMS NOT DUE ANYMORE#  
##############################################