**Group #23**

Robbie deMuth

Kimberly Hall

Mac Styslinger

Assignment #7

April 01, 2015

*The lab investigates the effects of certain safety laws on the number of traffic deaths. The lab isn’t a definitive study, both because there is much more data that can be brought to bear on the question and because there are likely to be difficulties with bias resulting from omitted variables. However, the lab does succeed in shedding light on some important issues that arise in more serious studies of the subject*.

*The data set we will be using is traffic.dta. It is monthly time series data on traffic accidents in California between January 1981 (observation 1) and December 1989 (observation 108). The variables in the data set are:*

*Accidents = total (that is, fatal plus nonfatal) traffic accidents in that month*

*Fatalaccidents = total fatal accidents in that month*

*Unemp = the state unemployment rate in that month*

*Speedlaw = a dummy variable that equals one after speed limits were raised to 65 m.p.h. (They had previously been 55 m.p.h.) Zero otherwise.*

*Beltlaw = a dummy variable that equals one after a mandatory seat belt law was instituted. (Previously, there had been no such law.) Zero otherwise.*

*Weekend = the number of Fridays and Saturdays that fell in that month.*

*month = month of the year: 1 = January, 2=February, 3=March, etc.*

***Part I***

***Step I.***

*(3 points) This data set has been provided to you without a time index variable and without the data having been declared to be a time series. Create a time index variable (name it “time”) and declare the data to be a time series, making sure that STATA correctly realizes the data is monthly from January 1981 to December 1989. Define a new variable, pctfatal, equal to the proportion of all accidents that are fatal. Summarize the data and include the summary in your report.*

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

Accidents | 108 42831.28 4608.346 32699 52971

Fatalaccid~s | 108 377.9352 48.54678 266 500

Unemp | 108 7.200926 1.790134 4.3 11.9

Speedlaw | 108 .2962963 .4587521 0 1

Beltlaw | 108 .4444444 .4992206 0 1

-------------+--------------------------------------------------------

Weekend | 108 8.703704 .8232516 8 10

month | 108 6.5 3.468146 1 12

time | 108 305.5 31.32092 252 359

pctfatal | 108 .0088564 .0009978 .0070168 .0121683

***Step II.***

*(5 points) Hypothesize what the coefficient signs on the independent variables will be when explaining pctfatal with Unemp, Speedlaw, Beltlaw, Weekend and time, and briefly explain your reasoning. Then run the regression, including a set of monthly dummy variables to capture seasonal effects, and include it in your report.*

First, we would hypothesize Unemp to have a negative correlation with pctfatal because unemployed workers are less likely to commute during the day and contribute to accidents. We would expect Speedlaw to have a positive coefficient because higher speed limits likely increase the percentage of accidents that result in fatalities. On the other hand, we would expect Beltlaw to have a negative coefficient because seat belts should hopefully reduce the percentage of fatal accidents. The Weekend variable should have a negative coefficient because drivers should hopefully practice safer driving skills during the weekend when they don’t have to commute to work during rush hour traffic. Lastly, we would expect the time variable to have a negative coefficient because safety measures roads and vehicles have typically become safer over time due to new technologies.

Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 14.52

Model | .765456056 16 .047841004 Prob > F = 0.0000

Residual | .299795605 91 .003294457 R-squared = 0.7186

-------------+------------------------------ Adj R-squared = 0.6691

Total | 1.06525166 107 .009955623 Root MSE = .0574

------------------------------------------------------------------------------

pctfatal | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -.0156028 .0055385 -2.82 0.006 -.0266044 -.0046012

Speedlaw | .0675701 .0205236 3.29 0.001 .0268025 .1083376

Beltlaw | -.029801 .0231737 -1.29 0.202 -.0758327 .0162306

Weekend | -.0044741 .007253 -0.62 0.539 -.0188812 .009933

time | -.0022441 .0004202 -5.34 0.000 -.0030787 -.0014094

|

month |

February | -.005131 .0282304 -0.18 0.856 -.0612071 .0509452

March | -.0022931 .027332 -0.08 0.933 -.0565847 .0519986

April | .0556416 .0275431 2.02 0.046 .0009306 .1103527

May | .0703068 .0276092 2.55 0.013 .0154646 .1251491

June | .0975121 .0277805 3.51 0.001 .0423295 .1526948

July | .1759399 .0271782 6.47 0.000 .1219538 .229926

August | .1908538 .0274705 6.95 0.000 .1362871 .2454206

September | .1568048 .0279039 5.62 0.000 .1013772 .2122324

October | .1002426 .0275904 3.63 0.000 .0454376 .1550476

November | .0102796 .0279722 0.37 0.714 -.0452837 .0658429

December | .007779 .0277093 0.28 0.780 -.0472622 .0628203

|

\_cons | 1.644221 .1581406 10.40 0.000 1.330094 1.958348

------------------------------------------------------------------------------

***Step III.***

*(5 points) Using either one or two sided tests on the coefficients, the choice being based on the prior reasoning set forth in Step II, what are the correct p-values for each coefficient (except, that is, for the seasonal dummies).Which coefficients are statistically significant at the 5% level? Then do an F-test of the joint significance of the seasonal dummy variables. Are the seasonal dummies statistically significant?*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Expected Sign** | **P-Value** | **Significant** |
| Unemp | - | 0.003 | Yes |
| Speedlaw | + | 0.0005 | Yes |
| Beltlaw | - | 0.101 | No |
| Weekend | - | 0.2695 | No |
| Time | - | 0 | Yes |

( 1) 2.month = 0

( 2) 3.month = 0

( 3) 4.month = 0

( 4) 5.month = 0

( 5) 6.month = 0

( 6) 7.month = 0

( 7) 8.month = 0

( 8) 9.month = 0

( 9) 10.month = 0

(10) 11.month = 0

(11) 12.month = 0

F( 11, 91) = 14.21

Prob > F = 0.0000

As we can see, the seasonal dummy variables are jointly significant at any reasonable significant level because the p-value obtained from the F-test is extremely close to zero.

***Step IV.***

*(3 points) Save the residuals from the regression performed in Step II and have STATA plot them against time. Include the plot in your report. Do you see any clear evidence of serial correlation in the plot?*



As we can see, it appears that positive residuals seem to be almost always followed by positive residuals and vice versa. This could be indicative of serial correlation and suggests that observations of the random error term may not be independent.

***Step V.***

*(3 points) For the regression fit in Step II, generate lagged residuals and graph residuals against their lagged values. Include the plot in your report. Is there clear evidence of serial correlation in this graph?*



As we can see, the plot indicates a strong linear relationship between the current residual and the lagged residual. This strong pattern suggests that serial correlation may exist because true errors cannot possible be independent.

***Step VI.***

*(5 points) Using the residuals obtained from the regression done in Step II, follow the procedure outlined in Studenmund on page 414 to perform what he terms “the Lagrange Multiplier test” for serial correlation. What is the critical value of the Chi-square statistic (hint: see Table B-8) when testing using ? Do you accept or reject the null? Does this suggest there is or there isn’t autocorrelation?*

The test statistic we obtain for the Lagrange Multiplier test is . For the given test, the critical value of the Chi-square statistic at the 1% level is 6.63. Given that the value we obtained is larger than 6.63, we reject the null hypothesis and conclude that there is indeed serial correlation exhibited in the residuals.

***Step VII****.*

*(4 points) The test done in Step VI is also known as the Breusch-Godfrey test. Perform this test on the regression estimated in Step II using STATA. (Do not use the Davidson and Mackinnon approach.) Your answer in Step VII should differ with that in Step VI by no more than rounding error.*

Breusch-Godfrey LM test for autocorrelation

---------------------------------------------------------------------------

lags(p) | chi2 df Prob > chi2

-------------+-------------------------------------------------------------

1 | 8.628 1 0.0033

---------------------------------------------------------------------------

H0: no serial correlation

As we can see, the STATA output of the Breush-Godfrey LM test confirms our conclusions from the previous step. Since the p-value is less than 1%, we reject the null hypothesis and conclude that there is serial correlation.

***Step VIII.***

*(5 points) Using perform a Durbin-Watson test for positive serial correlation on the regression run in Step II and include the result in your report. What is the acceptance, rejection, and inconclusive region for the d-statistic? Because of the large number of explanatory variables, Studenmund’s tables will be inadequate, and one must consult a more detailed table.[[1]](#footnote-1) Do you accept or reject the null of no positive serial correlation using the Durbin-Watson or is the test inconclusive? What is approximate value of rho-hat implied by the d-statistic?*

Durbin-Watson d-statistic( 17, 108) = 1.422958

Using the table found through JSTOR, we find that the values of and for appropriate degrees of freedom are approximately 1.203 and 1.922, respectively. This means we reject the null hypothesis if d < 1.203 and fail to reject the null hypothesis if d > 1.922. This also implies that the result is inconclusive if the test statistic lies between 1.203 and 1.922. Based on the value output from STATA, we are inconclusive about whether or not there exists positive serial correlation. By computing we find that the approximate value of rho-hat implied by this d-statistic is 0.2885.

***Step IX.***

*(4 points) Estimate the same relationship as in Step II, but use the Prais-Winsten version of Generalized Least Squares. Include a copy of the result in your report. Take note of the estimated serial correlation coefficient produced by the procedure. Is it close to the value suggested by the d-statistic?*

Iteration 1: rho = 0.0472 , criterion = -.29252578

Iteration 2: rho = 0.3046 , criterion = -.27555561

Iteration 3: rho = 0.2975 , criterion = -.27551476

Iteration 4: rho = 0.2910 , criterion = -.27550221

Iteration 5: rho = 0.2910 , criterion = -.27550221

Iteration 6: rho = 0.2910 , criterion = -.27550221

Iteration 7: rho = 0.2910 , criterion = -.27550221

Iteration 8: rho = 0.2910 , criterion = -.27550221

Iteration 9: rho = 0.2910 , criterion = -.27550221

Iteration 10: rho = 0.2910 , criterion = -.27550221

Iteration 11: rho = 0.2910 , criterion = -.27550221

Iteration 12: rho = 0.2910 , criterion = -.27550221

Iteration 13: rho = 0.2910 , criterion = -.27550221

Iteration 14: rho = 0.2910 , criterion = -.27550221

Prais-Winsten AR(1) regression -- SSE search estimates

Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 10.22

Model | .49515405 16 .030947128 Prob > F = 0.0000

Residual | .275502206 91 .003027497 R-squared = 0.6425

-------------+------------------------------ Adj R-squared = 0.5797

Total | .770656256 107 .007202395 Root MSE = .05502

------------------------------------------------------------------------------

pctfatal | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -.0135477 .007117 -1.90 0.060 -.0276848 .0005895

Speedlaw | .0650128 .0267827 2.43 0.017 .0118123 .1182133

Beltlaw | -.0251251 .0300843 -0.84 0.406 -.0848839 .0346338

Weekend | -.0042326 .0058912 -0.72 0.474 -.0159347 .0074696

time | -.0021696 .0005486 -3.96 0.000 -.0032592 -.0010799

|

month |

February | -.0067243 .023751 -0.28 0.778 -.0539028 .0404541

March | -.0036593 .026181 -0.14 0.889 -.0556646 .0483461

April | .0547938 .0273566 2.00 0.048 .0004533 .1091344

May | .0701686 .0278219 2.52 0.013 .0149037 .1254334

June | .0968241 .0278026 3.48 0.001 .0415977 .1520505

July | .1739205 .0272039 6.39 0.000 .1198833 .2279577

August | .189976 .0275922 6.89 0.000 .1351675 .2447846

September | .1564955 .0280683 5.58 0.000 .1007413 .2122497

October | .0996916 .0276122 3.61 0.000 .0448435 .1545398

November | .0095345 .0271243 0.35 0.726 -.0443446 .0634136

December | .0069601 .024381 0.29 0.776 -.0414697 .0553899

|

\_cons | 1.604311 .1937244 8.28 0.000 1.219502 1.989121

-------------+----------------------------------------------------------------

rho | .2909854

------------------------------------------------------------------------------

Durbin-Watson statistic (original) 1.422958

Durbin-Watson statistic (transformed) 1.999357

As we can see, the estimated serial coefficient produced by this procedure of 0.2910 is close to that suggested by the d-statistic of 0.2885.

***Step X****.*

*(4 points) Another approach to the serial correlation problem is to use Newey-West standard errors. Re-estimate the relationship from Step II using Newey-West standard errors. Use a number of lags equal to the integer part of n^.25 in this and all further applications of Newey-West standard errors in this lab assignment. Include a copy of the result in your report.*

Regression with Newey-West standard errors Number of obs = 108

maximum lag: 3 F( 16, 91) = 16.58

Prob > F = 0.0000

------------------------------------------------------------------------------

| Newey-West

pctfatal | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -.0156028 .0065352 -2.39 0.019 -.0285842 -.0026214

Speedlaw | .0675701 .0254566 2.65 0.009 .0170037 .1181364

Beltlaw | -.029801 .0323137 -0.92 0.359 -.0939882 .0343861

Weekend | -.0044741 .0058125 -0.77 0.443 -.0160199 .0070717

time | -.0022441 .0005465 -4.11 0.000 -.0033296 -.0011585

|

month |

February | -.005131 .017239 -0.30 0.767 -.0393742 .0291122

March | -.0022931 .0227866 -0.10 0.920 -.0475559 .0429698

April | .0556416 .0262284 2.12 0.037 .0035421 .1077411

May | .0703068 .0277057 2.54 0.013 .0152729 .1253408

June | .0975121 .0315798 3.09 0.003 .0347827 .1602415

July | .1759399 .0345256 5.10 0.000 .107359 .2445208

August | .1908538 .0241474 7.90 0.000 .142888 .2388197

September | .1568048 .0285138 5.50 0.000 .1001656 .213444

October | .1002426 .0304941 3.29 0.001 .0396698 .1608153

November | .0102796 .0298593 0.34 0.731 -.0490322 .0695914

December | .007779 .0275847 0.28 0.779 -.0470145 .0625726

|

\_cons | 1.644221 .1812667 9.07 0.000 1.284157 2.004285

------------------------------------------------------------------------------

***Step XI.***

*(6 points) Compare the results from Step II with those obtained in Step IX and Step X. Compare the t-values on each coefficient. (For the seasonal dummy variables, compare only the F-statistic for the test of their joint significance.) When one corrects for serial correlation, which variables are statistically significant? Include this in your report. Compare the point estimates of the coefficients and the estimated standard errors of those coefficients obtained from the Newey-West procedure and from GLS with those obtained from OLS. (For brevity, you may skip comparing coefficients and standard errors for the monthly dummy variables.) What do you notice? Discuss how this illustrates the discussion in Studenmund, sections 9.2 and 9.4.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **T-Value or F-statistic** | | | **Significant** | | |
|  | **Step II** | **Step IX** | **Step X** | **Step II** | **Step IX** | **Step X** |
| Unemp | -2.82 | -1.90 | -2.39 | Yes | Yes | Yes |
| Speedlaw | 3.29 | 2.43 | 2.65 | Yes | Yes | Yes |
| Beltlaw | -1.29 | -0.84 | -0.92 | No | No | No |
| Weekend | -0.62 | -0.72 | -0.77 | No | No | No |
| Time | -5.34 | -3.96 | -4.11 | Yes | Yes | Yes |
| Dummy Variables | 14.21 | 9.66 | 15.99 | Yes | Yes | Yes |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Point Estimate** | | | **Estimated Standard Errors** | | |
|  | **OLS** | **GLS** | **Newey-West** | **OLS** | **GLS** | **Newey-West** |
| Unemp | -0.0156 | -0.0135 | -0.0156 | 0.0055 | 0.0071 | 0.0065 |
| Speedlaw | 0.0676 | 0.0650 | 0.0676 | 0.0205 | 0.0268 | 0.0255 |
| Beltlaw | -0.0298 | -0.0251 | -0.0298 | 0.0232 | 0.0301 | 0.0323 |
| Weekend | -0.0045 | -0.0042 | -0.0045 | 0.0073 | 0.0059 | 0.0058 |
| Time | -0.0022 | -0.0022 | -0.0022 | 0.0004 | 0.0005 | 0.0005 |

First, note that the point estimates of the coefficients produced by OLS and the Newey-West procedure are identical, while those produced by GLS are slightly different. The values produced by OLS and the Newey-West procedure are the same because the Newey-West procedure assumes that the serial correlation does cause bias in the estimated coefficients. As discussed in Studenmund, it is not surprising that the estimated coefficients produced by OLS and GLS differ because they are allowed to differ even though they have the same expected value. Next, notice that the estimated standard errors are all slightly different. This is not surprising since each method calculates the estimated errors slightly differently. For the most part, the estimated standard errors produced by the Newey-West procedure are larger than those produced by OLS just as Studenmund describes. This is to be expected because we have a relatively small sample.

***Step XII.***

*(6 points) There is a difference between being statistically significant and practically significant. It is possible the effects of the speed law and the belt law are large and practically important but are not measured with precision in our small data set. Have STATA find the average number of accidents in a month, and then use the OLS and GLS point estimates for the coefficients to obtain a point estimate of the number of lives per month saved or lost because seat belt or speed laws turned fatal accidents into nonfatal accidents, or vice versa. Include the answer and its derivation in your report. Note that this procedure abstracts from any change in the number of accidents that might possibly be caused by this safety regulation. Exploring that issue is the subject of Part II of the assignment.*

From the summary statistics, the average number of accidents in a month is approximately 42,831.28. From OLS, the coefficients that we obtain for the speed law and belt law variables are 0.0676 and -0.0298, respectively. This means that the speed law results in an additional lives lost each month, while the belt law saves approximately lives each month.

***Part II****.*

*The approach we have followed is open to an important objection. We have examined the likelihood that an accident will prove fatal and related that to speed and seat belts. But the number of accidents that occur may depend on speed and whether seat belts are worn. Economists have pointed out that safety features such as seat belts that protect the driver in a crash are likely to encourage people to drive in a riskier fashion. One economist famously (and probably facetiously) remarked that the best safety device he could imagine would be a dagger imbedded in the steering wheel pointed at the driver’s heart. Is there any evidence in the data that might support this objection?*

***Step I.***

*(5 points) Hypothesize what the signs on the independent variables will be when explaining the number of accidents with Unemp Speedlaw, Beltlaw, Weekend and time, and briefly explain your reasoning. Then, using OLS, regress the number of accidents on Unempl Speedlaw, Beltlaw, Weekend and time, including a set of monthly dummy variables to capture seasonal effects, and include the result in your report.*

First, we would hypothesize Unemp to have a negative correlation with Accidents because as a greater percentage of workers are unemployed, we would expect fewer drivers out on the road during peak traffic times. We would expect Speedlaw to have a negative coefficient because as the speed limit is increased, we would expect different drivers to exhibit similar speeds in high speed areas which typically results in a lower amount of collisions. According to the discussion above, we would expect Beltlaw to have a positive coefficient because drivers may exhibit greater amounts of reckless driving behavior due to the knowledge that they are protected by a seat belt. We would expect the Weekend variable to have a positive coefficient because there are a greater number of drivers on the road during the weekend when they aren’t working. Last, we would expect the number of accidents to be positively correlated with time since each year, the amount of people owning cars increases, which is likely to result in a greater number of collisions.

Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 60.16

Model | 2.0761e+09 16 129753639 Prob > F = 0.0000

Residual | 196285370 91 2156982.09 R-squared = 0.9136

-------------+------------------------------ Adj R-squared = 0.8984

Total | 2.2723e+09 107 21236856.1 Root MSE = 1468.7

------------------------------------------------------------------------------

Accidents | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -830.0377 141.7186 -5.86 0.000 -1111.544 -548.5312

Speedlaw | -2082.378 525.151 -3.97 0.000 -3125.526 -1039.23

Beltlaw | 4329.89 592.9608 7.30 0.000 3152.046 5507.734

Weekend | 431.5998 185.5869 2.33 0.022 62.95424 800.2453

time | 43.71978 10.75163 4.07 0.000 22.363 65.07657

|

month |

February | -1024.828 722.3506 -1.42 0.159 -2459.689 410.0326

March | 3380.146 699.3627 4.83 0.000 1990.948 4769.344

April | 577.8573 704.7653 0.82 0.414 -822.0724 1977.787

May | 1139.48 706.4555 1.61 0.110 -263.8072 2542.767

June | 1148.148 710.8405 1.62 0.110 -263.8491 2560.146

July | 2187.449 695.4269 3.15 0.002 806.069 3568.829

August | 2575.397 702.9069 3.66 0.000 1179.159 3971.635

September | 1995.466 713.9962 2.79 0.006 577.2004 3413.732

October | 3650.814 705.9755 5.17 0.000 2248.48 5053.148

November | 3522.32 715.7436 4.92 0.000 2100.583 4944.057

December | 4499.551 709.0187 6.35 0.000 3091.173 5907.929

|

\_cons | 28417.03 4046.454 7.02 0.000 20379.24 36454.81

------------------------------------------------------------------------------

***Step II.***

*(5 points) Using either one or two sided tests on the coefficients, the choice being based on the prior reasoning set forth in Step I, what are the correct p-values for each coefficient, and which coefficients are statistically significant at the 5% level? Also do an F-test of the joint significance of the seasonal dummy variables. Are the seasonal dummies statistically significant?*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Expected Sign** | **P-Value** | **Significant** |
| Unemp | - | 0 | Yes |
| Speedlaw | - | 0 | Yes |
| Beltlaw | + | 0 | Yes |
| Weekend | + | 0.011 | No |
| Time | + | 0 | Yes |

( 1) 2.month = 0

( 2) 3.month = 0

( 3) 4.month = 0

( 4) 5.month = 0

( 5) 6.month = 0

( 6) 7.month = 0

( 7) 8.month = 0

( 8) 9.month = 0

( 9) 10.month = 0

(10) 11.month = 0

(11) 12.month = 0

F( 11, 91) = 10.68

Prob > F = 0.0000

As we can see, the seasonal dummy variables are jointly significant at any reasonable significance level because the p-value obtained from the F-test is extremely close to zero.

***Step III.***

*(3 points) Save the residuals from the regression performed in Step I and have STATA plot them against time. Include the plot in your report. Do you see any clear evidence of autocorrelation in the plot?*



As we can see, it appears that positive residuals seem to be almost always followed by positive residuals and vice versa. This could be indicative of serial correlation and suggests that observations of the random error term may not be independent.

***Step IV****.*

*(3 points) For the regression fit in Step I, generate lagged residuals and graph residuals against their lagged values. Include the plot in your report. Is there clear evidence of autocorrelation in this graph?*



As we can see, the plot indicates a somewhat strong linear relationship between the current residual and the lagged residual. This pattern suggests that serial correlation may exist because true errors may not be independent.

***Step V.***

*(4 points) Using , perform a Breusch-Godfrey test on the regression estimated in Step I using STATA. (Do not use the Davidson and Mackinnon approach.) What do you conclude? Is there autocorrelation in the residuals?*

Breusch-Godfrey LM test for autocorrelation

---------------------------------------------------------------------------

lags(p) | chi2 df Prob > chi2

-------------+-------------------------------------------------------------

1 | 18.606 1 0.0000

---------------------------------------------------------------------------

H0: no serial correlation

Based on the extremely small p-value of the given test, we reject the null hypothesis and conclude that there is indeed serial correlation exhibited in the residuals.

***Step VI.***

*(5 points) Using perform a Durbin-Watson test for positive serial correlation and include the result in your report. What is the acceptance, rejection, and inconclusive region for the d-statistic?[[2]](#footnote-2) Do we accept or reject the null of no positive autocorrelation or is the test inconclusive? What is the value of rho-hat implied by the d-statistic?*

Durbin-Watson d-statistic( 17, 108) = 1.145603

Using the table found through JSTOR, we find that the values of and for appropriate degrees of freedom are approximately 1.203 and 1.922, respectively. This means we reject the null hypothesis if d < 1.203 and fail to reject the null hypothesis if d > 1.922. This also implies that the result is inconclusive if the test statistic lies between 1.203 and 1.922. Based on the value output from STATA, we reject the null hypothesis and conclude that there is positive serial correlation. By computing we find that the approximate value of rho-hat implied by this d-statistic is 0.4272.

***Step VII.***

*(5 points) Estimate the same relationship as in Step I, but use the Prais-Winsten version of Generalized Least Squares. Include a copy of the result in your report. Take note of the estimated autocorrelation coefficient produced by the procedure. Is it close to the value suggested by the d-statistic?*

Iteration 1: rho = 0.0472 , criterion = -1.900e+08

Iteration 2: rho = 0.4675 , criterion = -1.650e+08

Iteration 3: rho = 0.4672 , criterion = -1.650e+08

Iteration 4: rho = 0.4534 , criterion = -1.650e+08

Iteration 5: rho = 0.4534 , criterion = -1.650e+08

Iteration 6: rho = 0.4534 , criterion = -1.650e+08

Iteration 7: rho = 0.4534 , criterion = -1.650e+08

Iteration 8: rho = 0.4534 , criterion = -1.650e+08

Iteration 9: rho = 0.4534 , criterion = -1.650e+08

Iteration 10: rho = 0.4534 , criterion = -1.650e+08

Iteration 11: rho = 0.4534 , criterion = -1.650e+08

Iteration 12: rho = 0.4534 , criterion = -1.650e+08

Iteration 13: rho = 0.4534 , criterion = -1.650e+08

Iteration 14: rho = 0.4534 , criterion = -1.650e+08

Iteration 15: rho = 0.4534 , criterion = -1.650e+08

Prais-Winsten AR(1) regression -- SSE search estimates

Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 37.09

Model | 1.0762e+09 16 67260898.2 Prob > F = 0.0000

Residual | 165014068 91 1813341.41 R-squared = 0.8671

-------------+------------------------------ Adj R-squared = 0.8437

Total | 1.2412e+09 107 11599892 Root MSE = 1346.6

------------------------------------------------------------------------------

Accidents | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -744.9884 210.6562 -3.54 0.001 -1163.431 -326.5458

Speedlaw | -1631.084 807.6597 -2.02 0.046 -3235.401 -26.76765

Beltlaw | 4550.21 900.2713 5.05 0.000 2761.932 6338.489

Weekend | 427.4303 131.4068 3.25 0.002 166.4069 688.4537

time | 35.70953 16.64507 2.15 0.035 2.646134 68.77293

|

month |

February | -918.6737 545.8445 -1.68 0.096 -2002.927 165.5793

March | 3585.492 640.4407 5.60 0.000 2313.335 4857.648

April | 844.0966 694.8367 1.21 0.228 -536.111 2224.304

May | 1399.636 722.038 1.94 0.056 -34.60417 2833.875

June | 1396.254 722.4905 1.93 0.056 -38.88431 2831.393

July | 2403.43 707.7865 3.40 0.001 997.4992 3809.361

August | 2851.483 718.8255 3.97 0.000 1423.624 4279.342

September | 2312.187 728.5994 3.17 0.002 864.914 3759.46

October | 3994.563 707.1634 5.65 0.000 2589.87 5399.257

November | 3904.944 673.6093 5.80 0.000 2566.902 5242.986

December | 4995.165 583.6313 8.56 0.000 3835.853 6154.476

|

\_cons | 29791.74 5738.979 5.19 0.000 18391.96 41191.51

-------------+----------------------------------------------------------------

rho | .4533902

------------------------------------------------------------------------------

Durbin-Watson statistic (original) 1.145603

Durbin-Watson statistic (transformed) 2.111846

As we can see, the estimated serial coefficient produced by this procedure of 0.4534 is somewhat close to that suggested by the d-statistic of 0.4272.

***Step VIII.***

*(5 points) Another approach to the autocorrelation problem is to use Newey-West standard errors. Re-estimate the relationship from Step I using Newey-West standard errors. Include a copy of the result in your report.*

Regression with Newey-West standard errors Number of obs = 108

maximum lag: 3 F( 16, 91) = 38.03

Prob > F = 0.0000

------------------------------------------------------------------------------

| Newey-West

Accidents | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

Unemp | -830.0377 213.413 -3.89 0.000 -1253.956 -406.1191

Speedlaw | -2082.378 701.7672 -2.97 0.004 -3476.352 -688.4036

Beltlaw | 4329.89 773.6085 5.60 0.000 2793.212 5866.568

Weekend | 431.5998 167.5134 2.58 0.012 98.85495 764.3446

time | 43.71978 15.66864 2.79 0.006 12.59595 74.84362

|

month |

February | -1024.828 573.5832 -1.79 0.077 -2164.181 114.5243

March | 3380.146 743.3502 4.55 0.000 1903.572 4856.72

April | 577.8573 678.4015 0.85 0.397 -769.7039 1925.419

May | 1139.48 701.4247 1.62 0.108 -253.8142 2532.774

June | 1148.148 786.9674 1.46 0.148 -415.0658 2711.362

July | 2187.449 824.0763 2.65 0.009 550.5228 3824.375

August | 2575.397 744.5011 3.46 0.001 1096.537 4054.257

September | 1995.466 736.344 2.71 0.008 532.8093 3458.123

October | 3650.814 899.6718 4.06 0.000 1863.726 5437.902

November | 3522.32 811.1927 4.34 0.000 1910.985 5133.655

December | 4499.551 1039.872 4.33 0.000 2433.973 6565.129

|

\_cons | 28417.03 5558.031 5.11 0.000 17376.68 39457.37

------------------------------------------------------------------------------

***Step IX.***

*(6 points) Compare the results from Step I with those obtained in Step VII and Step VIII. Compare the t-values on each coefficient. (For the seasonal dummy variables, compare only the F statistic for the test of their joint significance.) When one corrects for serial correlation, which variables are statistically significant? Discuss this in your report. Compare the point estimates of the coefficients and the estimated standard errors of those coefficients obtained from the Newey-West procedure and from GLS with those obtained from OLS. In this instance, is there a material difference in coefficients, standard errors, or t-statistics when one corrects for autocorrelation?*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **T-Value or F-statistic** | | | **Significant** | | |
|  | **Step I** | **Step VII** | **Step VIII** | **Step I** | **Step VII** | **Step VIII** |
| Unemp | -5.86 | -3.54 | -3.89 | Yes | Yes | Yes |
| Speedlaw | -3.87 | -2.02 | -2.97 | Yes | Yes | Yes |
| Beltlaw | 7.30 | 5.05 | 5.60 | Yes | Yes | Yes |
| Weekend | 2.33 | 3.25 | 2.58 | No | No | No |
| Time | 4.07 | 2.15 | 2.79 | Yes | Yes | Yes |
| Dummy Variables | 10.68 | 14.86 | 8.64 | Yes | Yes | Yes |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Point Estimate** | | | **Estimated Standard Errors** | | |
|  | **OLS** | **GLS** | **Newey-West** | **OLS** | **GLS** | **Newey-West** |
| Unemp | -830.04 | -744.99 | -830.04 | 141.71 | 210.66 | 213.41 |
| Speedlaw | -2082.38 | -1631.08 | -2082.38 | 525.15 | 807.66 | 701.77 |
| Beltlaw | 4329.89 | 4550.21 | 4329.89 | 592.96 | 900.27 | 773.61 |
| Weekend | 431.60 | 427.43 | 431.60 | 185.59 | 131.41 | 167.51 |
| Time | 43.72 | 35.71 | 43.72 | 10.75 | 16.65 | 15.67 |

As we can see from the above tables, the point estimates of the coefficients produced by OLS and the Newey-West procedure are identical, while those produced by GLS are slightly different. Furthermore, since the estimated standard errors produced both by GLS and the Newey-West procedure differ from OLS, we obtain different values of the t-statistic. Therefore, in this instance, when one corrects for serial correlation we obtain different estimates and statistics.

***Step X.***

*(6 points) Have STATA compute the average proportion in our data of all accidents that are fatal, and use that figure along with the OLS point estimates of the Speedlaw and Beltlaw coefficients obtained in Step VII to produce a point estimate of deaths that might have occurred because these laws either increased or decreased the number of accidents. Repeat this calculation using the coefficient estimates obtained in Step VIII. (This abstracts from any change in the survivability of the accident -- the issue explored in Part I of this assignment.)*

From the summary statistics in STATA, we find that on average, 0.88% of accidents are fatal. Using OLS, we obtain estimated coefficients for the Speedlaw and Beltlaw variables of -2082.39 and 4329.89. From this data, we find that enforcement of the speeding law would have resulted in approximately fewer deaths while a belt law would have resulted in an approximate additional deaths.

1. N. E. Savin and Kenneth J. White, “The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors,” *Econometrica,* Vol. 45, No. 8 (Nov., 1977), pp. 1989-1996. You can access this online using the JSTOR database available through VIRGO. [↑](#footnote-ref-1)
2. You will once again have to look up the critical values in the article cited in footnote 2. [↑](#footnote-ref-2)