Answer Key 3 Time Series Data and Serial Correlation

1. (Adapted from C11.5, C12.1) Suppose we are interested in how personal tax exemptions (*pe*) affect the general fertility rate (*gfr*). Use the data in FERTIL3.RAW for this exercise.
2. Estimate the following equation:

*gfrt = β0 + β1 pet + β2ww2t + β3pillt + εt*

where *ww2* is a dummy variable for the years 1941-1945, and *pil*l is a dummy variable that is equal to one for the years 1963 on (after the pill was available). Discuss the significance of the coefficients and interpret their magnitudes.

reg gfr pe ww2 pill

 Source | SS df MS Number of obs = 72

-------------+------------------------------ F( 3, 68) = 20.38

 Model | 13183.6215 3 4394.54049 Prob > F = 0.0000

 Residual | 14664.2739 68 215.651087 R-squared = 0.4734

-------------+------------------------------ Adj R-squared = 0.4502

 Total | 27847.8954 71 392.223879 Root MSE = 14.685

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

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

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

 pe | .08254 .0296462 2.78 0.007 .0233819 .1416981

 ww2 | -24.2384 7.458253 -3.25 0.002 -39.12111 -9.355684

 pill | -31.59403 4.081068 -7.74 0.000 -39.73768 -23.45039

 \_cons | 98.68176 3.208129 30.76 0.000 92.28003 105.0835

A $1 increase in personal exemptions increases fertility by .08 children per 1000 women, which is statistically significant at the 1% level. Is it economically significant? A $1 increase is fairly trivial, so another way to interpret the effect is to look at a 1 standard deviation increase in pe ($65.88 according to the summary stats). This translates into reduction of 5 children per 1000 women, about a fourth of the standard deviation for gfr. The WWII and post-pill periods both had lower average fertility rates than the earlier periods: the average fertility rate in other periods (when pe=0) is 98.7, but is 74.5 during WWII and 67.1 after the introduction of the pill.

1. Fertility may react to personal exemptions with a lag. Reestimate your equation adding 2 lags of *pe* (*pet-1*and *pet-2*). Are these variables jointly significant? What are the degrees of freedom of your F-test and why?

tsset year

reg gfr pe ww2 pill l1.pe l2.pe

 Source | SS df MS Number of obs = 70

-------------+------------------------------ F( 5, 64) = 12.73

 Model | 12959.7886 5 2591.95772 Prob > F = 0.0000

 Residual | 13032.6443 64 203.635067 R-squared = 0.4986

-------------+------------------------------ Adj R-squared = 0.4594

 Total | 25992.4329 69 376.701926 Root MSE = 14.27

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

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

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

 pe | .0726718 .1255331 0.58 0.565 -.1781094 .323453

 ww2 | -22.1265 10.73197 -2.06 0.043 -43.56608 -.6869196

 pill | -31.30499 3.981559 -7.86 0.000 -39.25907 -23.35091

 pe |

 L1. | -.0057796 .1556629 -0.04 0.970 -.316752 .3051929

 L2. | .0338268 .1262574 0.27 0.790 -.2184013 .286055

 \_cons | 95.8705 3.281957 29.21 0.000 89.31403 102.427

Note that this is the same as results using the transformed variables included in the dataset:

reg gfr pe ww2 pill pe\_1 pe\_2

 Source | SS df MS Number of obs = 70

-------------+------------------------------ F( 5, 64) = 12.73

 Model | 12959.7886 5 2591.95772 Prob > F = 0.0000

 Residual | 13032.6443 64 203.635067 R-squared = 0.4986

-------------+------------------------------ Adj R-squared = 0.4594

 Total | 25992.4329 69 376.701926 Root MSE = 14.27

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

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

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

 pe | .0726718 .1255331 0.58 0.565 -.1781094 .323453

 ww2 | -22.1265 10.73197 -2.06 0.043 -43.56608 -.6869196

 pill | -31.30499 3.981559 -7.86 0.000 -39.25907 -23.35091

 pe\_1 | -.0057796 .1556629 -0.04 0.970 -.316752 .3051929

 pe\_2 | .0338268 .1262574 0.27 0.790 -.2184013 .286055

 \_cons | 95.8705 3.281957 29.21 0.000 89.31403 102.427

. test pe\_1 pe\_2

 ( 1) pe\_1 = 0

 ( 2) pe\_2 = 0

 F( 2, 64) = 0.05

 Prob > F = 0.9480

No, these variables are not jointly significant.

1. What are the first order autocorrelations () for *gfr* and *pe*? What do these suggest about possible unit root(s)? What does this suggest about your OLS results in (i)?

You can regress gfr on lagged gfr, pe on lagged pe. Here is an alternative approach:

corrgram pe

 -1 0 1 -1 0 1

 LAG AC PAC Q Prob>Q [Autocorrelation] [Partial Autocor]

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

1 0.9471 0.9479 67.314 0.0000 |------- |-------

2 0.8776 -0.2493 125.94 0.0000 |------- -|

3 0.8103 0.0594 176.64 0.0000 |------ |

4 0.7414 -0.0868 219.7 0.0000 |----- |

. corrgram gfr

 -1 0 1 -1 0 1

 LAG AC PAC Q Prob>Q [Autocorrelation] [Partial Autocor]

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

1 0.9447 0.9777 66.977 0.0000 |------- |-------

2 0.8730 -0.3048 124.99 0.0000 |------ --|

3 0.8072 0.1752 175.3 0.0000 |------ |-

4 0.7345 -0.2981 217.56 0.0000 |----- --|

The first order autocorrelation is very close to 1, suggesting that there may be a possible unit root. Unless gfr and pe are cointegrated, this implies that our results may be driven by spurious correlation.

1. Re-estimate (i) using first differences—that is, changes in *gft* and changes in *pe*. (Do not difference *ww2* and *pill*.) How does the effect of *pe* compare with your estimates in levels in (i)?

reg cgfr cpe ww2 pill

 Source | SS df MS Number of obs = 71

-------------+------------------------------ F( 3, 67) = 4.10

 Model | 197.055018 3 65.6850059 Prob > F = 0.0098

 Residual | 1072.52736 67 16.007871 R-squared = 0.1552

-------------+------------------------------ Adj R-squared = 0.1174

 Total | 1269.58238 70 18.1368911 Root MSE = 4.001

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

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

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

 cpe | -.0935102 .0325581 -2.87 0.005 -.1584965 -.0285239

 ww2 | 5.131562 2.249415 2.28 0.026 .6417119 9.621413

 pill | -1.987072 1.04927 -1.89 0.063 -4.081424 .1072791

 \_cons | -.4703762 .6034214 -0.78 0.438 -1.67481 .7340579

Note this is the same as

reg D.gfr D.pe ww2 pill

 Source | SS df MS Number of obs = 71

-------------+------------------------------ F( 3, 67) = 4.10

 Model | 197.055018 3 65.6850059 Prob > F = 0.0098

 Residual | 1072.52736 67 16.007871 R-squared = 0.1552

-------------+------------------------------ Adj R-squared = 0.1174

 Total | 1269.58238 70 18.1368911 Root MSE = 4.001

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

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

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

 pe |

 D1. | -.0935102 .0325581 -2.87 0.005 -.1584965 -.0285239

 ww2 | 5.131562 2.249415 2.28 0.026 .6417119 9.621413

 pill | -1.987072 1.04927 -1.89 0.063 -4.081424 .1072791

 \_cons | -.4703762 .6034214 -0.78 0.438 -1.67481 .7340579

Now we see that a 1 dollar change in personal exemptions leads to a .09 reduction in the growth rate of fertility, very different than the results from before. This suggests that some of the positive result from before may have been driven by spurious correlation--an overall reduction in both the birth rate and exemptions over time. Here are two graphs to give you a sense of what is going on:





In the first graph, the two series move more or less together. In the second graph, we can see that the annual changes often have an inverse relationship. (You can also see that differencing these series leads to stationarity).

1. Reestimate (ii) using first differences of *gft*, *pe*, and lagged *pe*. (Again, do not difference *ww2* and *pill*.) Interpret the coefficients and comment on their statistical significance.

reg cgfr cpe cpe\_1 cpe\_2 ww2 pill

 Source | SS df MS Number of obs = 69

-------------+------------------------------ F( 5, 63) = 5.29

 Model | 372.894284 5 74.5788568 Prob > F = 0.0004

 Residual | 888.565534 63 14.1042148 R-squared = 0.2956

-------------+------------------------------ Adj R-squared = 0.2397

 Total | 1261.45982 68 18.5508797 Root MSE = 3.7556

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

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

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

 cpe | -.0751636 .0323566 -2.32 0.023 -.1398232 -.010504

 cpe\_1 | -.0513865 .0331632 -1.55 0.126 -.1176579 .0148848

 cpe\_2 | .0882556 .0279766 3.15 0.002 .0323488 .1441624

 ww2 | 4.839225 2.831973 1.71 0.092 -.8200213 10.49847

 pill | -1.676145 1.004766 -1.67 0.100 -3.684009 .3317186

 \_cons | -.6502546 .5817652 -1.12 0.268 -1.81282 .5123105

Now we see that while change in personal exemptions are negatively related to changes in fertility in a given year, after two years, a change in fertility is associated with increased fertility, indicating that the effect may take some time to appear. This lagged model makes sense, given that outcomes fertility decisions take time to be realized.

1. Add a linear time trend to the model in (v). Is a time trend necessary in the first-difference equation?

It turns out that whether a time trend is significant depends on the specification. The first specification includes dummies for the post ww2 and pill periods. This essentially shifts the time trend in changes up after ww2 and down again after the pill:

. reg cgfr cpe cpe\_1 cpe\_2 ww2 pill t

 Source | SS df MS Number of obs = 69

-------------+------------------------------ F( 6, 62) = 5.79

 Model | 453.089851 6 75.5149751 Prob > F = 0.0001

 Residual | 808.369968 62 13.0382253 R-squared = 0.3592

-------------+------------------------------ Adj R-squared = 0.2972

 Total | 1261.45982 68 18.5508797 Root MSE = 3.6108

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

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

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

 cpe | -.0618201 .0315717 -1.96 0.055 -.124931 .0012908

 cpe\_1 | -.039124 .0322664 -1.21 0.230 -.1036236 .0253757

 cpe\_2 | .0951653 .0270425 3.52 0.001 .0411081 .1492225

 ww2 | 3.250812 2.797162 1.16 0.250 -2.340636 8.84226

 pill | -4.888069 1.615706 -3.03 0.004 -8.117819 -1.658319

 t | .0944006 .0380635 2.48 0.016 .0183127 .1704884

 \_cons | -3.141163 1.149618 -2.73 0.008 -5.439216 -.8431097

If we do not include those time period dummies, then t has no effect. If you look back and the graph of changes above, you will see that there does not appear to be a time trend overall, but that the WWII and post 63-pill periods do have some large outliers. In general, time trends are often not necessary to include after differencing the data, but that is not always the case: there may be a reason why CHANGES in the variable trend.

. reg cgfr cpe cpe\_1 cpe\_2 t

 Source | SS df MS Number of obs = 69

-------------+------------------------------ F( 4, 64) = 4.88

 Model | 294.856737 4 73.7141842 Prob > F = 0.0017

 Residual | 966.603082 64 15.1031732 R-squared = 0.2337

-------------+------------------------------ Adj R-squared = 0.1859

 Total | 1261.45982 68 18.5508797 Root MSE = 3.8863

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

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

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

 cpe | -.0348352 .0272856 -1.28 0.206 -.0893445 .019674

 cpe\_1 | -.0131442 .0278616 -0.47 0.639 -.0688042 .0425158

 cpe\_2 | .11109 .0272773 4.07 0.000 .0565974 .1655826

 t | .0078781 .0242282 0.33 0.746 -.0405233 .0562796

 \_cons | -1.267445 1.046219 -1.21 0.230 -3.357507 .822617

1. Using the model in (vi), test for whether there is AR(1) serial correlation in the errors.

. reg cgfr cpe cpe\_1 cpe\_2 ww2 pill

 Source | SS df MS Number of obs = 69

-------------+------------------------------ F( 5, 63) = 5.29

 Model | 372.894284 5 74.5788568 Prob > F = 0.0004

 Residual | 888.565534 63 14.1042148 R-squared = 0.2956

-------------+------------------------------ Adj R-squared = 0.2397

 Total | 1261.45982 68 18.5508797 Root MSE = 3.7556

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

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

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

 cpe | -.0751636 .0323566 -2.32 0.023 -.1398232 -.010504

 cpe\_1 | -.0513865 .0331632 -1.55 0.126 -.1176579 .0148848

 cpe\_2 | .0882556 .0279766 3.15 0.002 .0323488 .1441624

 ww2 | 4.839225 2.831973 1.71 0.092 -.8200213 10.49847

 pill | -1.676145 1.004766 -1.67 0.100 -3.684009 .3317186

 \_cons | -.6502546 .5817652 -1.12 0.268 -1.81282 .5123105

. estat dwatson

Durbin-Watson d-statistic( 6, 69) = 1.561962

. estat durbinalt

Durbin's alternative test for autocorrelation

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

 lags(p) | chi2 df Prob > chi2

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

 1 | 3.143 1 0.0763

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

 H0: no serial correlation

This is somewhat borderline. The p-value of .0763 indicates that there is a 7 percent chance that we would observe the autocorrelation that we do even if the truth were no serial correlation. Although this is not 5% or smaller, given the way the test is structured, we err on the side of accepting the null. We might want to do more to look at the model (including t, for example.)

1. Suppose we are interested in how laws and economic conditions might affect driving behavior. Use TRAFFIC2.RAW (monthly observations from CA from Jan 1981-Dec 1989) to answer these questions.
	1. The variable *prcfat* is the percentage of accidents resulting in at least on fatality. Note that this variable is a percentage, not a proportion. What is the average of this variable over this period?

sum prcfat

 Variable | Obs Mean Std. Dev. Min Max

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

 prcfat | 108 .8856363 .0997777 .7016841 1.216828

Note that the max is greater than one because there are a few periods with many fatalities per accident (i.e., more than one person died on average in accidents)

* 1. Run a regression of *prcfat* on a linear time trend, 11 monthly dummies (set January as your base month), *wkends*, *unem*, *spdlaw*, and *beltlaw*. Discuss the estimated effects of *unem*, *spdlaw*, and *beltlaw*. Do the signs and magnitudes make sense to you?

reg prcfat t wkends unem spdlaw beltlaw feb-dec

 Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 14.44

 Model | .764228387 16 .047764274 Prob > F = 0.0000

 Residual | .301019769 91 .00330791 R-squared = 0.7174

-------------+------------------------------ Adj R-squared = 0.6677

 Total | 1.06524816 107 .00995559 Root MSE = .05751

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

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

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

 t | -.0022352 .0004208 -5.31 0.000 -.0030711 -.0013993

 wkends | .0006259 .0061624 0.10 0.919 -.011615 .0128668

 unem | -.0154259 .0055444 -2.78 0.007 -.0264392 -.0044127

 spdlaw | .0670877 .0205683 3.26 0.002 .0262312 .1079441

 beltlaw | -.0295053 .0232307 -1.27 0.207 -.0756503 .0166397

 feb | .0008607 .0289967 0.03 0.976 -.0567377 .0584592

 mar | .0000923 .0274069 0.00 0.997 -.0543481 .0545327

 apr | .0582201 .0278195 2.09 0.039 .0029601 .11348

 may | .0716392 .0276432 2.59 0.011 .0167293 .1265492

 jun | .1012618 .0280937 3.60 0.001 .0454571 .1570665

 jul | .1766121 .0272592 6.48 0.000 .122465 .2307592

 aug | .1926117 .0274448 7.02 0.000 .1380959 .2471274

 sep | .1600164 .028203 5.67 0.000 .1039947 .2160381

 oct | .1010357 .0276702 3.65 0.000 .0460722 .1559991

 nov | .013949 .0281436 0.50 0.621 -.0419548 .0698528

 dec | .0092005 .027858 0.33 0.742 -.046136 .064537

 \_cons | 1.029799 .1029523 10.00 0.000 .8252964 1.234301

Higher speed limits are estimated to increase the percent of fatal accidents, by .067 percentage points. This is a statistically significant effect. The new seat belt law is estimated to decrease the percent of fatal accidents by about .03, but the two-sided *p*-value is about .21.

Interestingly, increased economic activity also increases the percent of fatal accidents. This may be because more commercial trucks are on the roads, and these probably increase the chance that an accident results in a fatality.

* 1. Test the errors for AR(1) serial correlation.

Here is how you could conduct the test using the Breusch-Godfrey method: After getting the OLS residuals, , run the regression  (Included an intercept, but that is unimportant.)

tsset t

predict error, resid

reg error l.error

 Source | SS df MS Number of obs = 107

-------------+------------------------------ F( 1, 105) = 8.91

 Model | .023534599 1 .023534599 Prob > F = 0.0035

 Residual | .277302016 105 .002640972 R-squared = 0.0782

-------------+------------------------------ Adj R-squared = 0.0695

 Total | .300836614 106 .002838081 Root MSE = .05139

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

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

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

 error |

 L1. | .2816415 .0943463 2.99 0.004 .0945701 .4687129

 \_cons | .0002981 .0049684 0.06 0.952 -.0095533 .0101496

The coefficient on  is .281 (se = .094). Thus, there is evidence of some positive serial correlation in the errors (*t ≈* 2.99). Note that this test is only valid if there are no concerns about the endogeneity of the regressors. A strong case can be made that all explanatory variables are strictly exogenous. Certainly there is no concern about the time trend, the seasonal dummy variables, or *wkends*, as these are determined by the calendar. It is seems safe to assume that unexplained changes in *prcfat* today do not cause future changes in the state-wide unemployment rate. Also, over this period, the policy changes were permanent once they occurred, so strict exogeneity seems reasonable for *spdlaw* and *beltlaw*. (Given legislative lags, it seems unlikely that the dates the policies went into effect had anything to do with recent, unexplained changes in *prcfat*.

* 1. Re-estimate the model accounting to serial correlation.

Couple of options:

. arima prcfat t wkends unem spdlaw beltlaw feb-dec, ar(1)

(setting optimization to BHHH)

Iteration 0: log likelihood = 168.775

Iteration 1: log likelihood = 168.81849

Iteration 2: log likelihood = 168.85351

Iteration 3: log likelihood = 168.8576

Iteration 4: log likelihood = 168.85947

(switching optimization to BFGS)

Iteration 5: log likelihood = 168.86032

Iteration 6: log likelihood = 168.8641

Iteration 7: log likelihood = 168.86419

Iteration 8: log likelihood = 168.86439

Iteration 9: log likelihood = 168.8644

Iteration 10: log likelihood = 168.8644

Iteration 11: log likelihood = 168.8644

ARIMA regression

Sample: 1 - 108 Number of obs = 108

 Wald chi2(17) = 191.09

Log likelihood = 168.8644 Prob > chi2 = 0.0000

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

 | OPG

 prcfat | Coef. Std. Err. z P>|z| [95% Conf. Interval]

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

prcfat |

 t | -.0021501 .0005349 -4.02 0.000 -.0031986 -.0011017

 wkends | .0006179 .0054809 0.11 0.910 -.0101244 .0113602

 unem | -.0132124 .0068961 -1.92 0.055 -.0267285 .0003037

 spdlaw | .0641854 .0298944 2.15 0.032 .0055935 .1227773

 beltlaw | -.0248763 .0338792 -0.73 0.463 -.0912783 .0415257

 feb | -.0008934 .0361499 -0.02 0.980 -.071746 .0699591

 mar | -.0011545 .0306679 -0.04 0.970 -.0612625 .0589534

 apr | .0575487 .0315385 1.82 0.068 -.0042656 .1193629

 may | .0718131 .0263242 2.73 0.006 .0202186 .1234076

 jun | .1007231 .0252368 3.99 0.000 .0512599 .1501863

 jul | .174789 .0225581 7.75 0.000 .1305759 .2190022

 aug | .1919537 .0308971 6.21 0.000 .1313964 .252511

 sep | .159901 .0287129 5.57 0.000 .1036247 .2161772

 oct | .1007692 .0243195 4.14 0.000 .0531039 .1484345

 nov | .0133081 .0275731 0.48 0.629 -.0407342 .0673504

 dec | .0085411 .0281464 0.30 0.762 -.0466249 .0637071

 \_cons | 1.009016 .1103693 9.14 0.000 .7926966 1.225336

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

ARMA |

 ar |

 L1. | .2859698 .1117903 2.56 0.011 .0668649 .5050747

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

 /sigma | .0506463 .0041883 12.09 0.000 .0424375 .0588551

Or this one:

. prais prcfat t wkends unem spdlaw beltlaw feb-dec

Iteration 0: rho = 0.0000

Iteration 1: rho = 0.2816

Iteration 2: rho = 0.2884

Iteration 3: rho = 0.2887

Iteration 4: rho = 0.2887

Iteration 5: rho = 0.2887

Prais-Winsten AR(1) regression -- iterated estimates

 Source | SS df MS Number of obs = 108

-------------+------------------------------ F( 16, 91) = 10.16

 Model | .495019683 16 .03093873 Prob > F = 0.0000

 Residual | .277020817 91 .003044185 R-squared = 0.6412

-------------+------------------------------ Adj R-squared = 0.5781

 Total | .7720405 107 .007215332 Root MSE = .05517

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

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

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

 t | -.0021487 .0005479 -3.92 0.000 -.003237 -.0010604

 wkends | .0006166 .0050041 0.12 0.902 -.0093234 .0105567

 unem | -.0131807 .0071065 -1.85 0.067 -.0272969 .0009354

 spdlaw | .0641361 .0267953 2.39 0.019 .0109105 .1173616

 beltlaw | -.024816 .0301099 -0.82 0.412 -.0846256 .0349936

 feb | -.0009093 .0244211 -0.04 0.970 -.0494188 .0476001

 mar | -.0011624 .0262733 -0.04 0.965 -.0533512 .0510264

 apr | .0575507 .0275861 2.09 0.040 .0027543 .112347

 may | .071829 .0278625 2.58 0.012 .0164836 .1271745

 jun | .1007237 .0280584 3.59 0.001 .0449891 .1564583

 jul | .1747688 .0272886 6.40 0.000 .1205634 .2289741

 aug | .1919517 .0275896 6.96 0.000 .1371484 .246755

 sep | .1599066 .0283165 5.65 0.000 .1036594 .2161538

 oct | .1007763 .0277006 3.64 0.000 .0457525 .1558002

 nov | .013306 .0273035 0.49 0.627 -.040929 .0675411

 dec | .0085443 .0245607 0.35 0.729 -.0402424 .0573311

 \_cons | 1.00872 .1016073 9.93 0.000 .8068894 1.21055

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

 rho | .2887052

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

Durbin-Watson statistic (original) 1.430031

Durbin-Watson statistic (transformed) 1.994739

.

* 1. Compute the first order autocorrelations () for *unem* and *prcfat*. What do these suggest about possible unit root(s)?

corrgram unem

 -1 0 1 -1 0 1

 LAG AC PAC Q Prob>Q [Autocorrelation] [Partial Autocor]

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

1 0.9410 0.9573 98.306 0.0000 |------- |-------

2 0.8928 0.0854 187.65 0.0000 |------- |

3 0.8601 0.1641 271.35 0.0000 |------ |-

4 0.8366 0.0972 351.3 0.0000 |------ |

. corrgram prcfat

 -1 0 1 -1 0 1

 LAG AC PAC Q Prob>Q [Autocorrelation] [Partial Autocor]

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

1 0.7077 0.7094 55.611 0.0000 |----- |-----

2 0.4439 -0.1127 77.696 0.0000 |--- |

3 0.1802 -0.1834 81.37 0.0000 |- -|

4 -0.0559 -0.1727 81.728 0.0000 | -|

The first order autocorrelation for *prcfat* is .709, which is high but not necessarily a cause for concern. For *unem*, , which is cause for concern in using *unem* as an explanatory variable in a regression.

* 1. Estimate the model in (ii) using first differences for unem and prcfat (Do not difference the month or policy variables.) Compare your results to those in (ii).

. reg D.prcfat t wkends D.unem spdlaw beltlaw feb-dec

 Source | SS df MS Number of obs = 107

-------------+------------------------------ F( 16, 90) = 2.94

 Model | .213030831 16 .013314427 Prob > F = 0.0006

 Residual | .406915412 90 .004521282 R-squared = 0.3436

-------------+------------------------------ Adj R-squared = 0.2269

 Total | .619946244 106 .005848549 Root MSE = .06724

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

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

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

 t | .0001433 .0004849 0.30 0.768 -.00082 .0011067

 wkends | .0068097 .0072276 0.94 0.349 -.0075492 .0211685

 unem |

 D1. | .0125342 .0161094 0.78 0.439 -.01947 .0445385

 spdlaw | -.0071825 .0237979 -0.30 0.763 -.0544612 .0400962

 beltlaw | .0008251 .0265048 0.03 0.975 -.0518312 .0534814

 feb | .0346228 .037046 0.93 0.352 -.0389755 .1082211

 mar | .0419346 .0389248 1.08 0.284 -.0353964 .1192656

 apr | .0985703 .0382988 2.57 0.012 .022483 .1746577

 may | .0568102 .0374416 1.52 0.133 -.0175742 .1311946

 jun | .0540339 .0347738 1.55 0.124 -.0150503 .1231182

 jul | .0878394 .0331103 2.65 0.009 .02206 .1536187

 aug | .0589255 .0396686 1.49 0.141 -.0198832 .1377342

 sep | .0065431 .0379741 0.17 0.864 -.068899 .0819852

 oct | -.0323897 .0352025 -0.92 0.360 -.1023255 .0375462

 nov | -.0591083 .0354151 -1.67 0.099 -.1294666 .01125

 dec | .0272794 .0363245 0.75 0.455 -.0448856 .0994445

 \_cons | -.126868 .1048114 -1.21 0.229 -.3350941 .0813581

This regression basically shows that the change in *prcfat* cannot be explained by the change in *unem* or any of the policy variables. It does have some seasonality, which is why the *R*-squared is .344.

1. Use the data in PHILIPS.RAW for this exercise. (This follows several of the examples in Wooldridge, but using the full set of the data, rather than only through 1996.)

The Phillips curve posits a relationship between unemployment and inflation:

$$inf\_{t}-inf\_{t}^{e}=β\_{1}\left(unem\_{t}-μ\_{0}\right)+e\_{t}$$

Here $inf\_{t}^{e}$ is the expected rate of inflation for year t that was formed in year t-1. The above formulation posits that there is a relationship between unanticipated inflation (deviations from expectations) and cyclical unemployment—deviations of unemployment in year t from the natural rate of unemployment, $μ\_{0}$. One assumption of this model is that the natural rate of unemployment is constant.

Under the adaptive expectations model, current expected values of inflation depend on recently observed inflation, resulting in the following:

$inf\_{t}-inf\_{t-1}=β\_{0}+β\_{1}\left(unem\_{t}\right)+e\_{t}$ = $∆inf\_{t}=β\_{0}+β\_{1}\left(unem\_{t}\right)+e\_{t}$

where $β\_{0}=-β\_{1}μ\_{0}$

1. **Estimate this equation. Interpret the coefficients. Using your estimates, calculate the natural rate of unemployment.**

. reg inf unem

 Source | SS df MS Number of obs = 56

-------------+------------------------------ F( 1, 54) = 3.58

 Model | 31.599858 1 31.599858 Prob > F = 0.0639

 Residual | 476.815691 54 8.8299202 R-squared = 0.0622

-------------+------------------------------ Adj R-squared = 0.0448

 Total | 508.415549 55 9.24391907 Root MSE = 2.9715

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

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

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

 unem | .5023782 .2655624 1.89 0.064 -.0300424 1.034799

 \_cons | 1.053566 1.547957 0.68 0.499 -2.049901 4.157033

Note that the coefficient here is positive—not what we’d expect.

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

1. **Obtain the residuals from this estimation. Is there evidence of serial correlation in these residuals?**

. dwstat

Durbin-Watson d-statistic( 2, 56) = .8014823

This is far below 2, and well below the lower bound—see the Appendix table D I handed out. So evidence of positive serial correlation.

Breusch Godfrey method:

. tsset year

 time variable: year, 1948 to 2003

 delta: 1 unit

. predict error, resid

. reg error l.error

 Source | SS df MS Number of obs = 55

-------------+------------------------------ F( 1, 53) = 27.91

 Model | 155.221317 1 155.221317 Prob > F = 0.0000

 Residual | 294.721637 53 5.5607856 R-squared = 0.3450

-------------+------------------------------ Adj R-squared = 0.3326

 Total | 449.942953 54 8.33227692 Root MSE = 2.3581

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

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

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

 error |

 L1. | .5724722 .1083545 5.28 0.000 .3551407 .7898038

 |

 \_cons | -.1118079 .3179895 -0.35 0.727 -.7496141 .5259983

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

The coefficient on  is .572 (se = .108). Again, there is evidence of positive serial correlation in the errors (*t ≈* 5.28).

1. Re-estimate this model accounting for serial correlation using the Prais-Winsten method of FGLS. Comment on the difference in coefficient estimates.

**prais inf unem**

**Iteration 0: rho = 0.0000**

**Iteration 1: rho = 0.5721**

**Iteration 2: rho = 0.7350**

**Iteration 3: rho = 0.7792**

**Iteration 4: rho = 0.7871**

**Iteration 5: rho = 0.7883**

**Iteration 6: rho = 0.7885**

**Iteration 7: rho = 0.7885**

**Iteration 8: rho = 0.7885**

**Iteration 9: rho = 0.7885**

**Prais-Winsten AR(1) regression -- iterated estimates**

 **Source | SS df MS Number of obs = 56**

**-------------+------------------------------ F( 1, 54) = 8.39**

 **Model | 38.377534 1 38.377534 Prob > F = 0.0054**

 **Residual | 246.917431 54 4.57254502 R-squared = 0.1345**

**-------------+------------------------------ Adj R-squared = 0.1185**

 **Total | 285.294965 55 5.18718118 Root MSE = 2.1384**

**------------------------------------------------------------------------------**

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

**-------------+----------------------------------------------------------------**

 **unem | -.7139659 .2897858 -2.46 0.017 -1.294951 -.1329804**

 **\_cons | 7.999443 2.048343 3.91 0.000 3.892762 12.10612**

**-------------+----------------------------------------------------------------**

 **rho | .7885234**

**------------------------------------------------------------------------------**

**Durbin-Watson statistic (original) 0.801482**

**Durbin-Watson statistic (transformed) 1.913928**

Note that the sign of the coefficient has flipped!

1. Then estimate the adaptive expectations model: $∆inf\_{t}=β\_{0}+β\_{1}\left(unem\_{t}\right)+e\_{t}$

Obtain the residuals from this estimation? Is there evidence of serial correltion in these residuals? If there is, re-estimate the model.

reg cinf unem

 Source | SS df MS Number of obs = 55

-------------+------------------------------ F( 1, 53) = 6.13

 Model | 32.6324798 1 32.6324798 Prob > F = 0.0165

 Residual | 282.055894 53 5.32180932 R-squared = 0.1037

-------------+------------------------------ Adj R-squared = 0.0868

 Total | 314.688374 54 5.82756247 Root MSE = 2.3069

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

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

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

 unem | -.5176487 .209045 -2.48 0.017 -.9369398 -.0983576

 \_cons | 2.828202 1.224871 2.31 0.025 .3714212 5.284982

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

. predict cerror, resid

(1 missing value generated)

. reg cerror l.cerror

 Source | SS df MS Number of obs = 54

-------------+------------------------------ F( 1, 52) = 0.08

 Model | .300505001 1 .300505001 Prob > F = 0.7798

 Residual | 197.891764 52 3.80561085 R-squared = 0.0015

-------------+------------------------------ Adj R-squared = -0.0177

 Total | 198.192269 53 3.73947677 Root MSE = 1.9508

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

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

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

 cerror |

 L1. | -.0326971 .1163578 -0.28 0.780 -.2661861 .2007919

 |

 \_cons | .1674464 .2654783 0.63 0.531 -.3652747 .7001676

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

Note that in this model, there is no evidence of serial correlation, the coefficient on unem is still negative (as in the AR adjusted model).

1. **An alternative model (the expectations augmented Phillips curve) allows the natural rate of unemployment to depend on past levels of unemployment. Reestimate the above model using *changes* in unemployment rather than levels as the independent variable. Comment on the difference between your results here and those in (a).**

The estimated equation in first differences is

  −.072 − .833 

(.306) (.290)

 *n* = 55, *R*2 = .135

The coefficient on Δ*unem* has the sign that implies an inflation-unemployment tradeoff, and the coefficient is quite large in magnitude. In fact, the estimated coefficient is not statistically different from –1, which would imply a one-for-one tradeoff.

1. **Compute a first order autocorrelation for unem. In your opinion, is the root close to one?**

The first order autocorrelation of *unem* is about .75. This is one of those tough cases: the correlation between *unemt* and *unemt*-1­ is large, but it is not especially close to one.

1. **Based on what you found from your various estimation results using different models and on the autocorrelations in errors and in the series, explain the pattern of your results. What would you conclude about which model is most appropriate?**

The levels regressions need to be adjusted for autocorrelation. However, the results for both the Phillips curve estimation (with the change in inflation as the dependent variable) and the augmented Phillips equation (with first differences of both variables) do not indicate autocorrelation in the errors. The main question is whether the correlations are driven by unit root processes in the variables of interest. Again, given an autocorrelation in unem of .75, this is sort of borderline.