PS 699 Problem Set 9: Heteroskedasticity and Serial Correlation

You have a data set called MES_699.dta. The data were given to me by Harold Clark, then of the University of North Texas. The data are part of those given to him by Mackuen, Erikson, or Stimson (MES) from their article:

MacKuen, Michael B., Erikson, Robert S., and Stimson, James A., "Peasants or bankers? the American electorate and the U.S. economy," The American Political Science Review 86:597-611 Sept.1992.

MES wish to ask whether presidential approval responds to the economy in either or some combination of four ways: are people retrospective and/or prospective (do they reward/punish a president for past economic performance and/or for expected future performance)? and are they personal and/or sociotropic in these economic evaluations (do they consider their own economic situation and/or their perception of the national economy)?

This assignment is not intended as a comment one way or another on their questions, their answers, or their research design. Nor should what we do here be viewed as sufficient to answer these questions. Nor should the sequence of questions or the questions themselves be taken as a guide to a research program (it has many extraneous and questionable steps as such). It's a series of exercises useful to practice, nothing more.

[Note: Stata-command hints given below relate to syntax previous versions' syntax, syntax which may have changed. Since those hints were written, Stata added a battery of useful time-series commands & procedures. You may want to 'time-series set: `tsset`' the data to capitalize on these. Try "`help time series`".]

(1) Taking the above questions as nestable ("putable" in one model), you consider first estimating the following model (the data are quarterly: one observation per quarter (three month period) from Eisenhower to Bush):

$$APP = \beta_0 + \beta_{pe}\, PEXP + \beta_{pa}\, PAGO + \beta_{be}\, BFUT + \beta_{ba}\, BAGO + \beta_v\, VIET + \beta_e\, EVENTS + \varepsilon$$

where APP is the % of people surveyed who approve of the president's handling of his job
PEXP is % of people who expect their financial situation to approve for them in the near future
PAGO is % of people who reported that their financial situation was good in the near past
BFUT is % of people who expect the economy to improve over the near term
BAGO is % of people who reported that the economy was doing well in the near past
VIET is the number of war dead in vietnam (in thousands that year-to-date)
EVENTS is an indicator variable coded 1 for the occurrence of a high-profile media-covered event involving the president (see the article for more details on all these variables).

(a) Estimate this model by OLS (in Stata, e.g., using "reg" or "fit" command). Print the results.
(b) Report on your coefficient estimates (i.e., interpret them (some substantively meaningful sort of "a such-&-such change in such-&-such is estimated to produce a such-and-such change in approval"), and be sure to give some indication of the certainty of those estimates (standard error, p-level, confidence interval...).

(2) It occurs to you now, that *Approval* may be more variant at some times than at others. You decide that perhaps you ought to test for heteroskedasticity. You decide to implementWhite's General Test.
(a) Indicate the auxiliary regression of that test, estimate it, and print the regression output.
(b) Report the results of the White test (indicate the test statistic, its p-level, your conclusion).

(3) Based on these results, you decide you had better "be safe" and use White's "robust" standard errors instead of the usual OLS ones.
(a) Re-estimate with White's heteroskedasticity-consistent variance-covariance estimates & print the results.
(b) Comment on any noticeable changes: indicate them and briefly consider what the change(s) may suggest(s) regarding the structure of your error variances relative to the regressors.
(c) Assume the true residuals in this data-generating process are drawn from independent distributions that have possibly different variances from observation to observation.
(i) What are the properties of these OLS coefficient estimates (bias? consistency? efficiency?)
(ii) What about the OLS standard errors? White's standard errors?

(4) Referring back to the auxiliary regression of White's test, suppose you decided to view it as constructive. In particular, suppose you take these results as possibly indicating some heteroskedasticity as a function of BAGO, PAGO, $BAGO^2$, $PAGO^2$ (in a real project, you'd want to have a theory which predicted such a thing).

(a) Report on a Glesjer's test regressing $ln(err^2)$ on these four variables (and a constant).

(b) Use the predictions from this regression as estimates of the natural log of the variance of each observation. Use those predictions to generate a weighting variable (call it *wt*) by which you could conduct FWLS (re-)analysis of the model. (Note: the weight involves an estimate of the variance of the observation, *not* the log of the variance. Be sure to transform back to the right scale using e^x.) Graph the weighting variable over time. (In Stata, *graph wt timeindx , s([date])* will do.)

(5) Re-estimate the model of (1) by FWLS using this weighting series from #4. (in Stata, after the regression model type *[aweight=wt]*). **[Be careful: software differs in what relation it expects your input weighting series, here *wt*, to be relative to the variance of epsilon. Some want *wt*=var., some =s.d., some inverses thereof, etc. Be sure to check how your software, including the version of Stata you are using, wants this info!]**

(a) Present and report the results of this FWLS estimation (interpret coefficients and give some statement about their estimated certainty).

(b) R^2 and M.S.E. (i.e., s^2_e) have changed. Why? What, if anything, should be made of this?

(c) Highlight and briefly discuss any other noticeable changes in the results.

(d) If your model of the error-variance structure is right...

(i) what are the properties of your FWLS coefficient estimates?

(ii) what about your FWLS standard errors?

(6) After all this finally occurs to you that, perhaps, just possibly, an observation in one quarter is not necessarily independent of the previous observation... What does the Durbin-Watson statistic from model 1 say about this? (i.e., what is the DW stat, and what conclusion does it suggest here? Consult a DW-distribution table or software command to get & report d^{upper} and d^{lower}. (In Stata, you used to have to estimate the model again, this time using *regdw*, to get the DW. That has likely changed.).

(7) Uh-Oh! You decide you'd better use Newey-West standard errors (and you consider any lag beyond 5 periods to be negligible) because you no longer trust your OLS or your White's standard errors. Print the results estimating the model by OLS with Newey-West's heteroskedasticity & autocorrelation (HAC) consistent estimated-coefficient variance-covariance estimates. (In Stata, the command was: `newey put model here , lag(5)`).

(a) Briefly discuss any noticeable changes from model 1. *E.g.*, have your *p*-levels changed much? What does this tell you about your previous assumption of independent observations?

(b) If all lags beyond 5 show no correlation, and the rest of your assumptions hold:

(i) what are the properties of your coefficient estimates?

(ii) what about your standard errors?

[In all of the following models, you should include the dummies (all of them) for the 1st quarter of an administration. This prevents any effects from previous administrations from "bleeding" into the next ones.]

(8) Next in your handy econometrics test regarding autocorrelation in the G(N)LRM, it recommends: estimate a model where $\varepsilon_t=\rho\varepsilon_{t-1}+\gamma_t$. Re-estimate model 1, only this time using a procedure that estimates $\varepsilon_t=\rho\varepsilon_{t-1}+\gamma_t$. Several could do (Cochrane-Orcutt, Prais-Winston, Hildreth-Lu, ...); use Cochrane-Orcutt here. (In stata, this was once the command `corc put model here , t(timeindx)`.

(a) What is the estimated coefficient on the previous period's residual in this model?

(b) Discuss any noticeable changes in the results from model 1. (Be sure to mention the change in DW statistics from model without to model with AR(1).)

(c) In particular, what do you think of your previously held independence assumption now? What about OLS's consistency and unbiasedness in the face of correlated residuals? Is that reassuring given the changes from question 1 to here? What about your Newey-West results? Do they seem adequate now?

(d) If this AR(1) process is the process that truly generated your residuals:

(i) what are the properties of your coefficient estimates from this (corc-estimated) FGLS?
(ii) what about its reported standard errors?

(e) Does the DW statistic give you much confidence that this is the right residual process? What other test(s) or statistics might you want to consider regarding residual autocorrelation from this model estimation?

(9) Generate a variable equal to lagged approval (once: `gen app1l=app[_n-1]`, easier now if data `tsset`). Using this variable, reconsider the temporal process generating your data. In particular, estimate a model where current approval depends on previous approval, and the rest of model 1.

(a) Estimate this model by OLS and print the results.

(b) Discuss any noticeable changes from model 1.

(c) Using an applicable test, consider whether the estimated residuals from this regression suggest appreciable first-order correlation remaining in residuals.

(d) Assuming the conclusions of this test are correct:

(i) what properties do the coefficient estimates in this lagged-dependent-variable LDV model have?
(ii) what about the standard errors?

(e) Which model seems better now, (8) or (9)?

(10) Re-estimate model 9, but this time include the set of dummy variables for each president (dde, jfk, *etc.*). Why do you suppose your coefficient on lagged approval changed the way it did?

(11) Using the estimates from the model in question (10) and a spreadsheet (or you could do it in other software if you prefer), graph the responses of approval to two hypotheticals: to a temporary, and to a permanent, 1-point increase in BFUT (that's two response lines). *(For challenging fun: generate and plot confidence intervals to go with these estimated responses. See lab exercises from Rob's summer courses.)*

(12) Use your statistics package to graph estimated residuals from the model estimated in question (10) against time (old syntax: `graph residuals timeindx, s([date])`.) Notice any outliers? Anything happen in that quarter that might explain this? (hint: a Bush is in White House, troops were involved, oil was involved...) (That should have been in the events series, don't you think?)