PSC 405 – Final Exam

1. Consider the data file newspaper.dta, which can be accessed using the link http://www.rochester.edu/College/faculty/mperess/lm2014/newspaper.dta. This data file contains a monthly time series with the following variables,

frac_pos - the fraction of positive mentions of the economy during the current month, a measure of the sentiment on the economy reported by a centrist newspaper in a large industrialized country

gr - the growth rate

un - the unemployment rate

inf - the inflation rate

rightpm - whether the current government has a right-wing prime minister

monthid - a time series index

(i) Estimate a model with *frac_pos* as the dependent variable and *gr*, *un*, and *inf* as independent variables. Interpret the effect of each independent variable and the constant.

(ii) Estimate an appropriate parametric model for the time series above (and justify your particular model). Does anything important change?

(iii) Correct for time-series dependence in the data using a non-parametric model and report the results. Does anything important change?

(iv) What are the advantages of applying the parametric approach, as in (ii), and what are the advantages of applying the nonparametric approach, as in (iii).

(v) Estimate the same model as in (i), but with a lagged-dependent variable this time. Interpret the results in this regression. Is a lagged dependent variable theoretically justified in this regression?

(vi) Estimate a parametric time-series model that corrects for time-series dependence in (v). Did the presence of a lagged dependent variable affect the model you chose in (vi) vs. (ii)? If so, why do think this happened?

(viii) We can use the variable *rightpm* to test if the newspaper is biased in favor of right-wing or left-wing prime ministers. Estimate the same model as in (i), this time including *rightpm* as a variable as well, and use either a parametric or nonparametric approach for correcting for time-series dependence. Interpret the independent variables in the model.

(ix) Would a lagged dependent variable be justified for the model you estimated in (viii)?

(x) Estimate the model from (viii), this time also including interactions between *rightpm* and the three economic variables, again correcting for time-series dependence using either a parametric or nonparametric approach. Interpret the effects of *rightpm*, *gr*, *un*, and *inf* on the dependent variable.

(xi) Based on the results of the model you estimated in (x), is the newspaper is "biased"?

(xii) Is a unit-root present in the dependent variable? What approach would you apply if a unit root were present in the data?

(xiii) If we use the *tsline* command to look at the pattern of growth (*tsline gr*), we see that this country experienced a major economic collapse around month 778. Test for a structural break at

month 778 using a chow test, using the simple regression you ran in (i). You don't have to correct for time-series dependence here. Interpret the results of the test.

2. Consider the data file gov_elec_2.dta, which can be accessed using the link http://www.rochester.edu/College/faculty/mperess/lm2014/gov_elec_2.dta. This data file contains election results from gubernatorial races with the following variables,

rep2pshare - the two party vote share of the Republican candidate

reppos - a measure of the position of the Republican candidate

dempos - a measure of the position of the Democratic candidate

ideo - a measure of the median voter's ideology, on a different scale than *dempos*

inc2pshare - the two party vote share of the incumbent candidate

incspend - spending by the incumbent candidate

chspend - spending by the challenger

statepop - the population of the state

chselfspend - the amount of money the challenger spent on his own campaign

year - the year of the election

timeidx - a time index for each state, that can be used to generate lagged variables

(i) We will start by estimating a model of proximity voting. Specifically, we will consider $y_n = \alpha_0 - \rho_0(r_n - m_n)^2 + \rho_0(d_n - m_n)^2 + \varepsilon_n$ where y_n is the Republican's two party vote share, r_n is the position of the Republican candidate, d_n is the position of the Democratic candidate, and m_n is the position of the median voter. The coefficient ρ_0 captures the effect of relative proximity—if the median voter is closer to the Republican candidate or further from the Democratic candidate, the Republican's vote share increases. We however do not observe the

positions of the candidates on the same scale as the position of the median voter. We observe $\tilde{m}_n$ where $m_n = a + b\tilde{m}_n$ (i.e. we observe the median voter's ideal point on a different scale). We can plug this into the model to obtain, $y_n = \alpha_0 - \rho_0(r_n - a_0 - b_0\tilde{m}_n)^2 + \rho_0(d_n - a_0 - b_0\tilde{m}_n)^2 + \varepsilon_n$ which we can manipulate to obtain, $y_n = \alpha_0 + 2\rho_0 a_0(r_n - d_n) - \rho_0(r_n^2 - d_n^2) + 2b_0\rho_0(r_n - d_n)\tilde{m}_n + \varepsilon_n$. We can write, $y_n = \beta_{10} + \beta_{20}(r_n - d_n) + \beta_{30}(r_n^2 - d_n^2) + \beta_{40}(r_n - d_n)\tilde{m}_n + \varepsilon_n$ where $\beta_{20} = 2\rho_0 a_0$, $\beta_{30} = -\rho_0$, and $\beta_{40} = 2b_0\rho_0$. This implies that we can run a regression with y_n as dependent variable and $(r_n - d_n)$, $(r_n^2 - d_n^2)$, and $(r_n - d_n)\tilde{m}_n$ as independent variables, and recover ρ_0, a_0, b_0 using $\hat{\rho} = -\hat{\beta}_3$, $\hat{a} = -\frac{\hat{\beta}_2}{2\hat{\beta}_3}$, and $\hat{b} = -\frac{\hat{\beta}_4}{2\hat{\beta}_3}$. Run such a regression and report the estimates of ρ_0, a_0, and b_0. (Warning: Note that stata always presents the constant term last).

(ii) Use the delta method to compute standard errors for $\hat{\rho}$, $\hat{a}$, and $\hat{b}$. Hint: remember that we can write $\hat{\rho}$, $\hat{a}$, and $\hat{b}$ as nonlinear functions of $\hat{\beta}$, i.e. $\hat{a} = f(\hat{\beta})$ where $f(x) = -\frac{x_2}{2x_3}$, so computing the standard errors will require developing a formula by taking derivatives of $f(x)$ for the three parameters.

3. Consider again the Gubernatorial elections data set. In this question, we will consider the effects of incumbent and challenger spending on the incumbent's vote share.

(i) First, compute spending per capita by dividing *incspend* and *chspend* by *statepop*, and call these *incspend_cap* and *chspend_ cap*. Then run a regression with *incspend_ cap* and *chspend_cap* as independent variables and *inc2pshare* as the dependent variable. Interpret the coefficients on *incspend_cap* and *chspend_cap*.

(ii) Alternatively, generate the variables *log_incspend_cap = log(1+incspend_cap)* and *log_chspend_cap =log(1+chspend_cap)*. Then run a regression with *log_incspend_cap* and *log_chspend_cap* as the independent variables. Interpret the coefficients on *log_incspend_cap* and *log_chspend_cap*.

(iii) Which specification between (i) and (ii) is more theoretically justified?

(iv) Which specification between (i) and (ii) is more empirically justified (or is it impossible to say?). Include a formal test if possible.

(v) The data set has a panel structure with *timeidx* being a time variable and *stateid* being an individual variable. Would it be appropriate to cluster the standard errors by state here? Why or why not? Report the results of a linear regression with clustered standard errors.

(vi) It can be argued that incumbent spending and challenger spending are endogenous. We could potentially correct for endogeneity using a fixed effects model. Explain why or why not fixed effects my help correct for endogeneity.

(vii) Estimate a model with state fixed effects.

(viii) Estimate a model with state random effects.

(ix) Will the model estimated in (viii) potentially correct for endogeneity? Why or why not?

(x) Would it be helpful to cluster the standard errors in the models estimated in (vii) and (viii)? Why or why not?

(xi) An alternative approach for correcting for endogeneity would be to use instrumental variables. Two potential instruments are included in the data set—*statepop* and a dummy variable for whether the challenger spent at least 500k of his own money in the campaign (*chselfspend*). Comment on whether these are potentially good instruments.

(xii) Run a first stage regression for each instrument. What do the first stage regressions tell us about the appropriateness of the instruments?

(xiii) Use IV to correct for endogeneity using the suggested instruments. Comment on the results, referencing both the observed results and the arguments you made in (xi) and (xii).

(xiv) Returning to the model estimated in (ii), would it make sense to include a lagged dependent variable in the specification? Why or why not?

(xv) If you were to estimate a model which included both fixed effects and a lagged dependent variable, which of these estimators would be appropriate to use—OLS, Anderson-Hsiao, and Arellano-Bond? Why?

(xvi) Estimate the model in (ii), this time interacting *year* with incumbent and challenger log spending per capita. Interpret the results of the incumbent and challenger spending variables.