Part A: Time Series Basics and Static Models

1) The first model in the Stata output regresses the unemployment rate in Florida on the US unemployment rate. The second adds month indicators and a linear time trend.

a) What is the purpose of adding the month indicators and the time trend?

The monthly indicators control for predictable seasonal fluctuations. When the dependent variable is in log form, the represent predictable proportional impacts of different months, for example due to holidays. In this case, it is a consistent addition to the underlying rate, not a proportional change. The time trend allows for growth at a constant rate over time.

b) What do you make of the change in the coefficient on US unemployment and its standard error that occurred when month indicators and a time trend were added?

The coefficient became larger and its standard error smaller. This would suggest that seasonal effects masked, or confused, part of the relationship between Florida and US unemployment, and controlling for them thus revealed a stronger relationship.

2) Nonstationarity

2a) What is a non-stationary process and why do we need to be wary of them?

The distribution of a non-stationary process changes in unpredictable ways over time. Here are two problems. First, to estimate a model from which to make predictions, we need to be able to expect the process to be the same in the future as in the data being used in the model. Second, two non-stationary series that are completely independent are almost certain to show strong statistically significant correlation with one another due only to the non-stationarity, so relationships found between non-stationary series are likely to be sufficiently spurious as to mean nothing for predictive purposes.

2b) Interpret the partial autocorrelogram and Dickey-Fuller test results for the Florida unemployment rate that follow the second regression model.

The first order partial autocorrelation coefficient is essentially 1. The null hypothesis it is one can be rejected at conventional levels, but it is so close to one that we are still better off differencing the data before modeling, because the impact of shocks will take a very long time to die out.

3) The third regression model in the output regresses the first difference of the unemployment rate in Florida on the first difference of the US unemployment rate, including month indicators and a time trend.

a) The time trend was much smaller and less statistically significant in the third model than in the second model. What do you make of that?

The original time trend reflected a constant growth rate, not an accelerating one, so when we took the first difference, the time trend is captured in the constant.

b) Compare the coefficients on US unemployment in the second and third models. What likely explains the difference?

The coefficient falls from 1.33 to 0.99 after first differencing the model. This likely reflects that a portion of the correlation picked up before differencing was purely spurious, which we should always suspect with series that are, or are nearly, I(1).

c) Interpret the Breusch-Godfrey test results that follow estimation of the third model.

After differencing and controlling for seasonal effects and changes in the US unemployment rate, the residuals do not exhibit first order correlation, but through 12 and 24 lags, there is still autocorrelation. The model is not dynamically complete.

d) In November 2017, the Florida unemployment rate was 3.8 and the US unemployment rate was 3.9. In December 2017, date was 695 and the US unemployment rate was 3.9. Using the third regression, what are the predicted values of the change in Florida unemployment rate and the Florida unemployment rate in December 2017? Show your work.

$$\widehat{\Delta u_{FL,t}} = -0.397 + 0.986\Delta u_{US,t} + 0.19December_t + 0.000002t$$

$$= -0.397 + 0.986(0) + 0.19(1) + 0.000002(695)$$

$$= -0.21$$

$$\widehat{u_{FL,695}} = -0.21 + 3.8 = 3.59$$

Part B: Model Selection

The Stata output for Part B contains four ARDL models relating the unemployment rate in Florida to the US unemployment rate and Florida building permits (in hundreds of thousands). Thoroughly make the case that Model 4 is the best of them.

The table below presents the usual measures of fit (those corrected for over-fitting), the number of parameters estimated, and the p-values for Breusch-Godfrey tests of serial correlation (the null hypothesis is no autocorrelation) for 1 month, a full year, and two full years.

Model		1	2	3	4
Fit Statistics	10-Fold CV RMSE	0.1314	0.1233	0.1203	0.1214
	LOOCV RMSE	0.1300	0.1202	0.1197	0.1187
	AIC	-418.3478	-437.8662	-439.2790	-443.8431
	BIC	-92.9858	-310.6041	-360.6759	-376.4690
# of Parameters (K+1)		87	34	21	18
Breusch-Godfrey p-values	1 lag	0.9023	0.1181	0.0886	0.7647
	12 lags	0.8229	0.5106	0.2739	0.4210
	24 lags	0.0846	0.6734	0.6385	0.7658

All models include lags of the Florida and US unemployment rate and Florida building permits, and also month indicators. Model 1 includes every lag through 24. Model 4 includes the fewest, only those estimated with relatively narrow confidence intervals, and so is the simplest. The other models are in between.

On fit, there is not a large difference between the models, though model 4 is best on AIC, BIC, and LOOCV and a close second on 10-fold CV. Model 4 is also the only one of the four that shows no evidence of serial correlation at 1 month, through 12 months, or through 24 months. This is consistent with visual inspection of the PACs of the model residuals as well.

Regarding the particular lags used for model 4, it is understandable that building permits pulled more than 2 months in the past might not matter once we control for US unemployment and lagged Florida unemployment and difference the data. It is also understandable that perhaps only current US unemployment and unemployment one month ago matter. But, if there is a weakness of model 4, it is that it is hard to conceptualize why the first lag of Florida unemployment would not matter if the second does.