Gus Lipkin

Question 1

> What do you take away from Part 1 of the analysis that is important for modeling total employment? Explain how those conclusions relate to specific parts of the output.

The first thing we see is the summary statistics with each of the three components of total employment, construct, leisure, and manufacture. It stands out that manufacture is, on average, greater than that of construct and manufacture combined. This also holds for the minimum and maximum values of all three. It's also interesting that as high as leisure is, it caps out at around 1/6 of total employment.

The time series plots of private employment show general upwards trends with drop-offs after recessions in the 90s, 2000s, and for COVID. There is also a slight dip that begins right around 2001 that might have to do with the US beginning another war.

The autocorrelogram and partial autocorrelogram of Total together indicate that we should difference due to serial correlation. However, the Dickey-Fuller test disagrees with a p-value greater than .05. The same is true for lnTotal. In both cases, differencing the variable significantly reduced the serial correlation.

Question 2

Both *Construct* and *Leisure* reflect service and base components, since workers in the base consume housing, places to shop, and leisure. You want to determine whether they are as impactful as *Manufacture* on total private employment. Part 2 of the Stata output presents results for eight different finite distributed lag models. Following each model are tests of the hypotheses that the total effects are equal between each pair of sectors.

2a

Which model is best for conducting these hypothesis tests? Explain why.

As discussed in Part 1, the high serial correlation in the first lags suggests that we should difference, even if the Dicky-Fuller test disagrees. Thus, I am not going to consider models 1, 3, 5, and 7 because they do not difference total. Models 2 and 8 both only have lags zero through three while models 4 and 6 also have lags twelve and twenty-four. Models 6 and 8, however, also have lag limits to the twelfth lag. It doesn't make sense to me to include the twenty-fourth lag and not check it for autocorrelation and so I will not choose the model 6. I can then discard models 2 and 4 as I believe a Newey-West regression more helpful in dealing with any lingering serial correlation. This leaves me with model 8.

2b

For the model you chose, carefully and fully interpret the hypothesis test. What conclusion do you draw about the question of interest?

The three test operations all have p-values greater than 0.05. Therefore, we can accept the null hypothesis that the impacts of all three variables are equal.

Based on Part 1, I would think that manufacture would not have the same effect as construct and leisure, but rather than manufacture and construct would have the same effect as leisure. This is, of course, simply based off of the summary statistics.

Question 3

Part 3 of the Stata output presents results for eight potential models for forecasting total employment. Though numbering restarts at Model 1, these models differ from the eight models in Part 2. Do not be confused by the numbering.

3a

All eight of these models use log variables. Why is that most natural here?

When forecasting, proportional change is preferred as it allows you to describe year-over-year change as a percent change rather than an absolute change. It also prevents data from being forecasted as less than zero in situations where it doesn't make sense.

3b

Other than not being in log form, why are the models in Part 2 of the output unsuitable for forecasting?

While some of the models in Part 2 are not differenced, I think the bigger issue is including lag 0 in the models. I'm not entirely sure what lag 0 is, but if lag 1 is the data offset by one, then lag 0 would be the data point itself and you can't use the data to predict itself.

3c

At the end of Part 3 of the output, there is a table of model evaluation measures. Why does it include the number of variables includes, the LOORMSE, and the AIC, and not measures like in sample RMSE or R2?

In sample RMSE and R^2 are not reliable methods of measuring forecast model fit. R^2 can be arbitrary depending on the data and produce a low score when the fit is good or vice versa. I do not remember the exact reason why we should not use either, but I do know that it was drilled into us by Dr Dewey.

3d

Which of the eight models seems best for forecasting total employment? Explain why.

Before doing something silly like actually looking at the models and getting myself into trouble, I'm going to consider the AIC and LOORMSE first so I can reduce the number of models I'm considering. Models 1, 2, 3, and 6 have the lowest AIC. Models 1 through 4 have the lowest LOORMSE. I will discard models 4 and 6 as they are not in the top four for both tests.

This leaves me with models 1, 2, and 3. Model 3 has the lowest BIC and LOORMSE, model 2 the lowest AIC, and model 1 wins nothing. I discard model 1. Although model 3 wins two of the three tests, I'm going to choose model two because it is not autoregressive. I believe that construct, leisure, and manufacture do have separate properties that change with time and thus it is better to consider the three of them rather than total employment's own history.

Question 4

> Discuss the use of rolling windows estimation in forecasting. At minimum, address each of the following:
>
> - What is a rolling windows forecast?
> - Why is the RMSE from a rolling window estimation the most appropriate way to estimate forecast accuracy?
> - What is the trade-off implicit in choosing window width?
> - How does the rolling window estimation procedure determine the best window width?

A rolling window forecast cycles through many forecast window widths, the length of data used to forecast and does this as many times as possible as it progress from the start to the end of the data. It records the average root mean squared error for each window width across the whole dataset. This is known as the rolling window root mean squared error.

The rolling window root mean squared error is the most appropriate measure because it considers all of the individual measurements for that window width as an aggregate value rather than a simple value.

The implicit trade-off in choosing window width is that you may cut off data that is important to that period in time. If for example you are forecasting black Friday weekend sales at a Starbucks in a mall, it may be important to use a longer window size so that growth can be properly assessed over time and not just in the weeks leading up to black Friday on the year in question.

In STATA, users determine the best window width by generally choosing the lowest RWrmse value. They could, however, if they so chose, also consider other factors like I mentioned above and choose a window width that has a low RWrmse and includes data far enough back but does not have the lowest RWrmse.

Question 5

Suppose you have determined that the best model to forecast y one period ahead is:
$\Delta ln(y) = 0.01 + 0.4\Delta ln(y_{t-1}) - 0.2\Delta ln(y_{t-2})$. (Remember Δ means difference.) You estimate the rolling window RMSE is 0.05. The values of y at times (t) -2, -1, and 0 are given in the table below.

t	y	Forecast	2 Sigma Upper Bound	2 Sigma Lower Bound
-2	245			
-1	207			
0	238			
1		263	291	238
2		269	310	233
3		269	320	226

5a

Show that the forecast for time t=1, made at time t=0, is 263, and that the ± 2 sigma bounds are 291 and 238 (as in the table).

Forecast

$$\Delta ln(y(t)) = 0.01 + 0.4(ln(y_{t-1}) - ln(y_{t-2})) - .2(ln(y_{t-2}) - ln(y_{t-3})) \tag{1}$$
$$\Delta ln(y(1)) = 0.01 + 0.4(ln(238) - ln(207)) - 0.2(ln(207) - ln(245))$$
$$\Delta ln(y(1)) = 0.0995$$
$$\Delta y(1) = 1.104$$
$$y(1) = 262.906 = 263$$

Upper and Lower Bound

$$bounds = e^{\frac{(0.05^2)}{2}} * e^{ln(1)+ln(263)\pm(2*.05)} \tag{2}$$
$$bounds = 1.001 * e^{0+5.562\pm.1}$$
$$upper = 1.001 * e^{0+5.562+.1} = 291$$
$$lower = 1.001 * e^{0+5.562-.1} = 238$$

5b

> Show that the dynamic autoregressive forecast for time t=3, made at time t=0, is 269, and that the
> ± 2 sigma bounds are 320 and 226 (as in the table).

My brain is absolutely fried so I can't convert the stata commands back to normal math, but here's the
code for it.

```
1   scalar rmse_mod1 = .05
2   arima **regression here**
3   predict pnonfarm, dynamic(tm(0))
4   predict mse, mse dynamic(mofd(tm(3)))
5   gen totmse = mse if Date==tm(3)
6   replace totmse = l.totmse+mse if Date>tm(3)
7   gen pnonfarma = Total if Date==tm(0)
8   replace pnonfarma = l.pnonfarma*exp(pnonfarm+mse/2) if Date>tm(0)
9   gen ub1a = pnonfarma*exp(totmse^.5)
10  gen ub2a = pnonfarma*exp(2*totmse^.5)
11  gen ub3a = pnonfarma*exp(3*totmse^.5)
12  gen lb1a = pnonfarma/exp(totmse^.5)
13  gen lb2a = pnonfarma/exp(2*totmse^.5)
14  gen lb3a = pnonfarma/exp(3*totmse^.5)
15  replace ub1a=Total if Date == tm(2021m3)
16  replace ub2a=Total if Date == tm(2021m3)
17  replace ub3a=Total if Date == tm(2021m3)
18  replace lb1a=Total if Date == tm(2021m3)
19  replace lb2a=Total if Date == tm(2021m3)
20  replace lb3a=Total if Date == tm(2021m3)
```

5c

> Why is dynamic forecasting with an AR or ARMA model useful for forecasting several variables for
> many periods when the stakes are low, but generally not the best approach for forecasting a few
> variables at a few points in time when the stakes are high?

It's useful because you can run an AR model that forecasts out several periods ahead. Of course,
because it's an AR model, it does not take into account any variables other than itself. It's a quick and
dirty way to forecast but when you want all the variables properly considered, it is important to actually
include them in the forecast. However, when you include them in the forecast, the process of predicting
becomes much more arduous.