

Spring 2021 Time Series Final Exam

Study Guide / Preparation / Take Home Component

Data

The csv file SP21Final contains monthly non-seasonally adjusted data for Florida from January 1990 thru March 2021, obtained from the St. Louis Federal Reserve Economic Database. The included variables are:

- ✧ *date*: String date in YMD mask
- ✧ *Construct*: Number of construction employees in Florida
- ✧ *Leisure*: Number of leisure and hospitality employees in Florida
- ✧ *Manufacture*: Number of manufacturing employees in Florida
- ✧ *Total*: Total private employment in Florida

Background

Employees in the economic base produce output that is (potentially) for export out of the local area where it is produced. Service industries, for example retail sales, support the employees in the economic base. Construction, leisure and hospitality, and manufacturing are thought to be part of Florida's economic base. Construction because serving in-migrant retirees, who need places to live and shop and pay for it with money earned outside of Florida, is part of Florida's economic base. Leisure and hospitality because of the role of Tourism. Manufacturing because manufactures may be sold anywhere and are not inherently service related.

Relation to the In-Class Final Exam

The next two pages present questions requiring evaluating the properties of the data, building a model to measure the relationship of interest, performing hypothesis tests, preparing a one step ahead forecast, and preparing a forecast for each month of the next year, along with guidance about how to do this. If we had unlimited time, I would just give you this to do on your own during the exam.

But time is not unlimited. So, work through this to prepare, and I will write an exam that takes far less time to take provided you have done this and understood what you have done. To do so, I will prepare my own work to address these analyses and give it to you the output for the exam. Sometimes I will provide results for more than one possible approach and will have to choose which is better and defend your choice. Other times I may ask you to explain something in my analysis by pointing to a command or intermediate result and asking how it works, what its purpose is, or how to interpret it.

Submission

Include your work on this preparatory assignment in your class portfolio and as a complete pdf along with the submission of the in-class (online) exam (as a separate file with your Canvas submission). I will refer to it, as needed, in deciding how much partial credit to award on the exam.

Spring 2021 Time Series Final Exam

Study Guide / Preparation / Take Home Component

Part 1: Modeling and Hypothesis Testing

Both *Construct* and *Leisure* reflect service and base components, since workers in the base need housing, and places to shop, and leisure too. You want to whether they are as impactful as *Manufacture* on total private employment. To do so, you want to determine how many additional jobs are associated with one additional job in each sector and if they differ. Specifically, you want to estimate a distributed lag model relating employment in these sectors to total private employment, calculate the total effects (the sum of the coefficients on the lags for the sector), and test the hypothesis that the total effects are equal. Assume only lags 0-3, 12, and 24 might matter.

Estimate an appropriate model, calculate the total effects based on the best model you come up with, and test the hypothesis that the total effects are equal. Here is a list of things you should do/consider/know/think about to accomplish this. Think of this as hints and guidelines, not necessarily complete step by step instructions or an exhaustive list.

- ✧ Explain/defend/justify whatever choices you make as you go.
- ✧ Produce time series plots for the level and first difference of both total private employment and its log. Use these to get a general impression of how the dependent variable has behaved over time, in case it helps you make modeling decisions later.
- ✧ Since you want to measure the multiplier associated with one job in each sector in terms of the number of resultant total private jobs, you do not want to take logs for this question. Think about and explain why this makes sense.
- ✧ Is total private employment highly persistent? Should the data be differenced? Consider visual and statistical evidence.
- ✧ Is there remaining serial correlation in the residuals? Consider visual and statistical evidence. If so, what should you do about it?
- ✧ Do lags 12 and 24 matter or should they be dropped? The test or testparm commands will be useful here.
- ✧ The test or testparm commands will be necessary for the hypothesis tests you are ultimately aiming for. Something like:
$$\text{test } z + l1.z + l2.z + l3.z + l12.z + l24.z = x + l1.x + l2.x + l3.x + l12.x + l24.x$$
where z and x are standing in for the two variables you test. You may or may not have all those lags. You may or may not have differenced z and q . You will have to do three such tests, one for each combination of sectors.
- ✧ Interpret your findings. Ultimately, what is your conclusion?

Spring 2021 Time Series Final Exam

Study Guide / Preparation / Take Home Component

Part 2: Forecasting

Use the data available to produce a thoroughly considered one step forecast of total private employment for April 2021 and a quick forecast through March 2022. Assumptions and things to think about or do:

- ✧ Do not worry about collecting additional data.
- ✧ Assume you can ignore lags other than 1-3, 12, and 24.
- ✧ Should you difference, log transform, or include month dummies?
- ✧ Explain how content knowledge should shape model search.
- ✧ Explain how you would use something like GSREG in model search.
- ✧ Explain what is typically wrong with *Stepwise* for model selection.
- ✧ Rather than a huge search using GSREG or something like it, just use LOOCV, AIC, and BIC (what are they and why use them?) to choose the best 2 of these 4 models.
 - AR only with lags 1-3 and month dummies
 - AR only with lags 1-3, 12, 24 and month dummies
 - ARDL with lags 1-3 for all 4 variables and month dummies
 - ARDL with lags 1-3,12, 24 for all 4 variables and month dummies
- ✧ Which 2 do you think are best of the 4? Why?
- ✧ For the best 2, use the rolling window procedure to pick the best model and window width and to estimate the RMSE of the forecast.
- ✧ Forecast total employment for April 2021 using the best model and window width. If you difference or use the log, reconstruct underlying total employment.
- ✧ Make a forecast evaluation plot with the rolling window point forecast, the actual, and a plus or minus 2-sigma interval for all available rolling window forecasts. How often does the interval contain the actual?
- ✧ Calculate an empirical and normal estimate of the 95% confidence interval for April 2021. Think about whether either is appropriate in this case. Do a histogram or other plot of your residuals to help decide if they are normal.
- ✧ When there are problems with the empirical interval and the normal interval (if there are here, explain), you can always fall back on Chebyshev's inequality as a (very) conservative way to get an interval forecast. A quick read of the Wikipedia page will tell you all you need to know for this purpose if you don't remember it from Stats 1.
- ✧ For the best model use ARIMA and dynamic forecasting to produce a quick forecast of total private employment through March 2022 and a fan chart. Use 1, 2, and 3 sigma bands for the fan chart and explain its interpretation using Chebyshev's inequality as a worst case and the normal distribution as a best case.
- ✧ Explain why dynamic forecasting using ARIMA may be a good choice for producing lots of forecasts many periods in the future but is usually not a great choice for producing a small number of high stakes forecasts.