Time Series Modeling and Forecasting **Final Exam—Solution**

Part A: Theory and Specification Questions

1) Derive the autoregressive representation of the MA(1) process $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ where $0<\theta<1$ and ε is a typical mean zero disturbance. Explain why for modeling a limited number of AR terms can represent this process well even though there are an arbitrarily large number of AR terms in the exact representation (one for each period since the first no matter how long).

$$
\begin{aligned}
u_t &= \theta\left(u_{t-1} - \theta\varepsilon_{t-2}\right) + \varepsilon_t \\
&= \theta u_{t-1} - \theta^2\varepsilon_{t-2} + \varepsilon_t \\
&= \theta u_{t-1} - \theta^2\left(u_{t-2} - \theta\varepsilon_{t-3}\right) + \varepsilon_t \\
&= \theta u_{t-1} - \theta^2 u_{t-2} + \theta^3\varepsilon_{t-3} + \varepsilon_t \\
&= \theta u_{t-1} - \theta^2 u_{t-2} + \theta^3\left(u_{t-3} - \theta\varepsilon_{t-4}\right) + \varepsilon_t \\
&= \theta u_{t-1} - \theta^2 u_{t-2} + \theta^3 u_{t-3} - \theta^4\varepsilon_{t-4} + \varepsilon_t \\
&\vdots \\
&= \sum_{h=1}^{t}(-1)^{h-1}\theta^h u_{t-h} + \varepsilon_t
\end{aligned}
$$

Note $\varepsilon_t = u_t - \theta\varepsilon_{t-1}$ for all $t>0$. Looking at the final step of the algebra to the right, since $0<\theta<1$, θ^h will approach 0 rapidly as long as θ is not very close to 1, meaning a limited number of AR terms can achieve a relatively accurate approximation.

2) What is the Wold representation theorem, why is it important for time series modeling and forecasting, and how does this relate to your answer to (1) above?

It states any stationary time series can be expressed as a deterministic portion plus a stable MA process in the forecast errors (shocks/innovations) from time t=0. We showed a stable MA(1) process could be expressed as an AR process. This is true for higher order MA processes. Together with the Wold theorem this implies any covariance stationary process can be well approximated by an AR(q) process without q being very large.

3) Consider the model $y_t = y_{t-1} + \delta J_t + \lambda t + \beta x_t + \varepsilon_t$ where t is the time index, J is 1 in January and 0 otherwise, x is a predictor variable, and ε is a mean zero disturbance. Write out the first difference. Explain why the first difference is stationary and not trending while the undifferenced process is trending and not trend stationary (will not tend to return quickly to trend after a large shock). In the differenced model, what is δ multiplied by in February? Interpret the meaning of this—it makes sense in context of the difference given the underlying model, explain why.

- The first difference is: $\Delta y_t = \Delta y_{t-1} + \delta\Delta J_t + \lambda + \beta\Delta x_t + \Delta\varepsilon_t$.
- y increased with t at λ per period, but the Δy is centered around $\delta\Delta J_t + \lambda + \beta\Delta x_t$, not trending given x (t is not a variable in the expression, just an index).
- Before differening, the effect of a shock at time t, ε_t, was permanent because it is transmitted forever through the lagged y values.
- After differencing, ε_t increases Δy_t but at time t+1 will increase Δy_t by exactly the amount that it decreases $\Delta\varepsilon_{t+1}$, so its effect is gone after one period.
- Δy_t is 1 in January, -1 in February, and 0 otherwise. What to make of that? There is a bump of δ in January that increased Δy in January. In February that bump is gone, so Δy falls by δ. In the differenced version we undo the January bump in February.

Part B: Time Series Modeling (Not Forecasting)

1) Refer to the part of the Stata output for Part B Question 1:

i) Explain why model B.1.2 is preferable to model B.1.1.

The partial autocorrelogram indicates nonfarm employment should be differenced to make it stationary, and this is supported by the dickey fuller test. Therefore, the second model, estimated on first differences, is better.

ii) Interpret the results of model B.1.2. What does the model say about the structural relationship between total employment and the two components?

There is a positive and statistically significant relationship between these two components and nonfarm employment that is not fully felt for a number of periods. This does not prove causation, but it is entirely consistent with the notion that these two sectors are important in increasing total employment, and don't simply displace other things.

2) Refer to the part of the Stata output for Part B Question 2:

i) Why are the results of model B.2.2 preferable to those of model B.2.1?

The Bruesh-Godfrey test indicates autocorrelation in the residuals, which should have been suspected anyway. The second model uses Newey-West standard errors to correct for this, and so is better.

ii) Explain how adding lags 1 and 12 of (the log difference of) nonfarm employment makes sense in a dynamic structural model of this type. (Think in terms of modeling the adjustment process.) Interpret these two coefficients in terms of what they imply for the adjustment process of nonfarm employment over time.

It takes a time for the economy to adjust, so where we move to today will depebd in pat on where we were before today, regardless of what our ultimate destination might be. Including these lags gets at that kind of thinking about the adjustment process. Both are statistically significant. The one period lag suggests that if last month was strong in employment growth, this month will see a small correction—that is some of the most recent growth should be expected to disappear in short order. The one year lag though suggests that in the longer term, if growth this month was high a year ago, it is likely to be high again this year.

Part C: Forecasting

1) Refer to the output for Part C Question 1

i) Argue that model C.1.2 is better than model C.1.1 or model C.1.3 for its intended purpose.

Both the LOOCV RMSE and the AIC are smallest for model 2, though the difference is to slight to worry much about. Further, model 2 retains the lag of amspk as a variable, which earlier analysis and theory suggests is sensible, while it is not present in model 3.

ii) What is the purpose of model C.1.2 and why should it differ from B.2.2 to that end?

The purpose is purely forecasting, so it cannot contain any contemporaneous variables (lag 0) that are needed for structural time series modeling and hypothesis testing about that structure (part B) but not available for pure forecasting.

***The caveat to that last statement, not required for full credit but I'll give a bonus if someone gets it, is that to predict an outcome where a decision regarding a variable that influences that outcome will be made based on seeing the results of the model, you must have uncovered the underlying structure by appropriate statistical inference, not only have estimated the best preditvie model which, ironically, is by definition not of use predicting if the action to be based on seeing the prediction has an important impact on what is being predicted.

2) Refer to the output for Part C Question 2:

i) Argue that model C.2.3 is better than model C.2.1 or model C.2.2 for its intended purpose.

Both the LOOCV RMSE is smallest for model 3. The AIC is smaller for model 3 than for model 1, but the smallest value is for Model 2. On these criteria, one could argue for model 2 or model 3. , though the difference is to slight to worry much about. Further, model 2 retains the lag of amspk as a variable, which earlier analysis and theory suggests is sensible, while it is not present in model 3. However, nothing we have seen yet has suggested the second lagged difference of nonfarm employment matters, and the first lag can't be used to forecast two periods ahead. The third model is more parsimonious, having dropped it. Also nothing we have seen makes us very confident the 2 and 3 year lags of the log difference of nonfarm employment matter independently, and the third specification drops them. So, since the two are split on performance measures, on grounds of simplicity, I would go with model 3.

ii) What is the purpose of model C.2.3 and why ***MUST*** it differ from C.1.2 to that end?

The purpose is to forecast two months ahead. Therefore, the dependent variable must be the change over two months, not one, and no lags that are not at least two months old can be used.

3) Refer to the output for Part C Question 2:

i) Refer to lines 500-523 and the figure that follows. What is that code doing and what is illustrated by the figure?

It is generating predicted values for the one month difference and the standard forecast error, then translating that back to the undifferenced values and then back to the unlogged values, correcting for the change of the in the shape of the distribution from normal to log normal. Then, it plots the point and approximate 95% interval forecast one month ahead against the history using the model to give a visual idea of the model's historical performance.

ii) Refer to lines 524-547 and the figure that follows. What is that code doing and what is illustrated by the figure?

Same as the previous one, but for the two month difference.

iii) Refer to lines 548-566 and the figure that follows. What is that code doing and what is illustrated by the figure?

It is using the one period results for the point and interval forecast for January 2018 and the two month model results for the point and interval forecast for February 2018, then preparing an illustration of the cone of uncertainty, or a fan chart, to give a visual representation of both the forecast the growing uncertainty as we move ahead in time.

iv) Briefly discuss the forecast for January and February 2018 nonfarm employment from the output for Part C, particularly in terms of the accuracy of the model historically (the first two figures and the LOOCV RMSE) and the width of the forecast interval (in the last figure).

The series was made stationary by differencing. The models use a full year of monthly nonfarm lags, and potentially lags from 24 and 36 months ago, and lags of bldmt and potentially of amspk, all of which are reasonable, and produce low values for the RMSE in leave one out cross validation. In the first two figures, upward and downward changes are usually predicted, the interval is relatively narrow interval of movements of nonfarm, and the interval almost always contains the actual. So, the model appears reasonably sound. The last figure predicts a January downturn, which is common, with a relatively tight interval, with a February upturn, also common, but the prediction has a much wider interval.

v) Why is the method used in Part C for forecasting two periods out conceptually better than a dynamic autoregressive or vector autoregressive approach?

If either of the dynamic versions achieved the best performance, the version estimating a separate model for each forecast horizon would simply replicate those results. But if this method achieves the best result, neither of those two dynamic methods can replicate it because they are not looking at the data in a way that subsumes this methodology. The downside is it may be more work and work that can't be done by canned routines in many programs. It also makes it easier to produce the fan chart, or cone of uncertainty, to give a useful visual presentation of uncertainty of the untransformed and undifferenced variable.