STA 642: Fall 2018 – Homework #3 Exercises

1. Continue to work through the AR(p) notes and connect with relevant sections of the P&W text. Then, read ahead on material in the introductory sections of P&W Chapter 4 on general DLMs, and the support slides on the course web page. We will move there next class after spending more time on AR(p) models and review of this homework with its broader connections.

2. **Course mini-project proposal and development**. Re-read the expectations and process on the course assessment pdf linked to first class on the web site. See also the advance information on the *minproject-interim.pdf* file linked with this homework. You must be progressing on this. Before Fall break, you will be required to submit a 1-2 page outline of project ideas. Per the start of semester information and comments on this, that later mid-term checkpoint will be assessed (as part of the mid-term take-home exam) based on your identification of a topic, initial outline/sketch of the problem area, comments on data available/sources, and comments on initial goals for modelling, time series analysis and/or forecasting.

   Make sure to discuss with TA and/or myself– contact via email to discuss your initial thoughts, questions, and bounce-off project ideas. Nothing to hand-in here this week .. but, be sure to get focused on project planning!

3. The exploratory AR(p) Matlab code has, at the end, some basic exploration of quarterly US macro-economic data. See data in Figure 1. That example just reads in data, selects quarterly inflation rates and converts to *quarterly changes (i.e., difference values of quarterly actual inflation levels)*, and fits one or two AR(p) models. Earlier in the Matlab code there are examples of simulating the reference posterior in any AR(p) model, and generating synthetic futures representing the posterior distribution via Monte Carlo samples. The former can be used to explore (by Monte Carlo– as in the example using the SOI time series worked through in detail in that example code, and from class) the posterior for moduli and wavelengths of any quasi-periodic components suggested by the model. You may/should find this code useful in exploring this data further.

   Fit the reference analysis of an AR(8) model to this data, and address the following.

   (a) If the AR(8) is accepted as a good model for this data, do you think the data-model match supports stationarity? Give full numerical support for this based on the reference posterior.

   As in the example code, we can trivially generate a direct Monte Carlo sample from the posterior for the characteristic roots (a.k.a. eigenvalues) of the AR(8) model, simply by direct transformation of samples of the AR vector ϕ. Transforming to the roots and saving the $p = 8$ moduli $r_{1:8}$ gives us a sample from the posterior for these moduli, to be summarized. The proportion of samples with at least one r_j value lying outside $(-1, 1)$ can be trivially calculated– the Monte Carlo estimate of the posterior probability of non-stationarity.

   Figure 2 shows a Monte Carlo sample of posterior draws for the 8 moduli $r_{1:8}$, in terms of boxplots; it seems clear they all lie below $r = 1$, and it is trivially checked that in

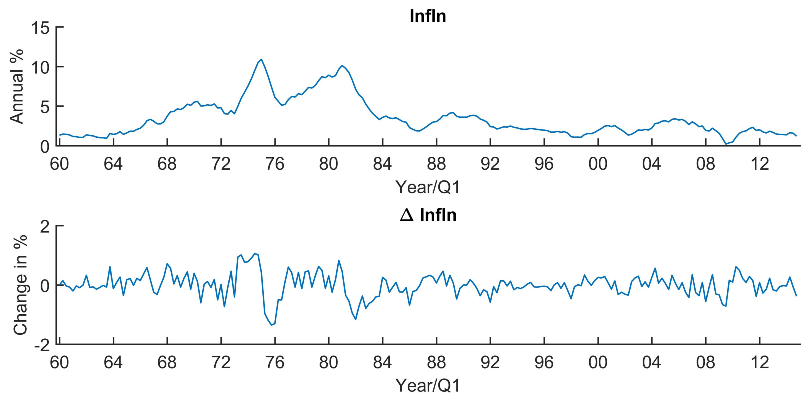

Figure 1: Quarterly US inflation and ΔInfln

fact they all do. Repeating this with larger samples maintains this, i.e., the posterior probability of non-stationarity is zero, up to Monte Carlo error.

Additional comments: The sampling of posterior for eigenvalues inherently suffers from a basic identification question. The simulation here orders Monte Carlo sampled values down by moduli, and saves the corresponding wavelengths (which are 0 or ∞ in cases of real eigenvalues). There are a number of details to then address in summarizing inferences. First, the boxplots of moduli reflect the ordering; it also shows that most of the samples give 4 pairs of complex conjugate eigenvalues, while a (very) few give 3 pairs plus 2 real eigenvalues of low moduli. Second, boxplots here just match wavelengths with moduli based on the initial ordering of the moduli. Then, to infer the ordered wavelengths reordering is needed. You can see that, in cases of two or more components with similar moduli, ordering by these will confound inference on wavelengths. This does not arise with this example, but does and may with others.

(b) Assuming that there is some indication of quasi-periodic behaviour under this posterior, summarise inferences on the *maximum wavelength (a.k.a. period)* of (quasi-)periodic components.

For each posterior sample of the roots computed above, we need to identify any complex conjugate roots and save the resulting values of the wavelength. It turns out that the posterior for this analysis very strongly supports at least one quasi-periodic component, so that repeat Monte Carlo draws almost surely include at least one sampled wavelength. Matlab code to do this simply adds this check on each sampled ϕ vector.

Figure 2 shows the histogram of the samples of the posterior for the maximum wavelength (and also the second largest). In this (and multiple repeat simulations) *all* Monte Carlo samples have at least one complex conjugate root, so that the histogram is of the

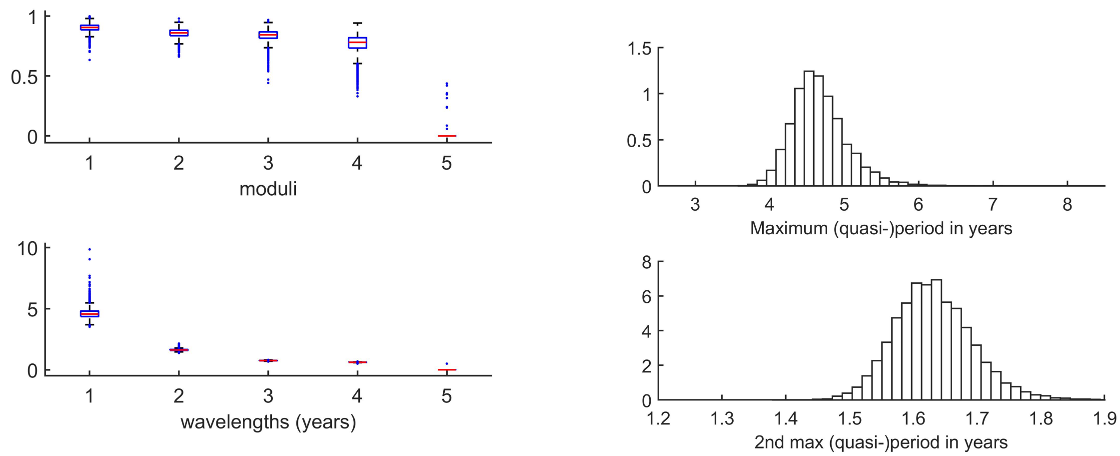

Figure 2: Boxplots of posteriors of moduli $r_{1:8}$ and histogram of the posterior for the largest two wavelengths of quasi-periodic components, in the reference AR(8) analysis of ΔInfln. Note that component are ordered down by moduli, and the wavelengths are then summarized by reordering.

full number simulated. This shows, by the way, the the posterior probability of at least one quasi-periodic component in the series (assuming we like the model) is basically one.

Then, some summaries of the posterior for the maximum wavelength (up to one decimal place) are as follows: mean and median are 4.6 years, interquartile range 4.3-4.8 years, 95% equal-tails interval 4.0-5.4 years.

(c) Explore and discuss aspects of inference on the implied decomposition of the series into underlying components implied by the eigenstructure of the AR model.

Posterior inferences for the sampled latent components are shown in Figure 3. The dominant component is a quasi-cyclical "business cycle" while the higher-order components refine that.

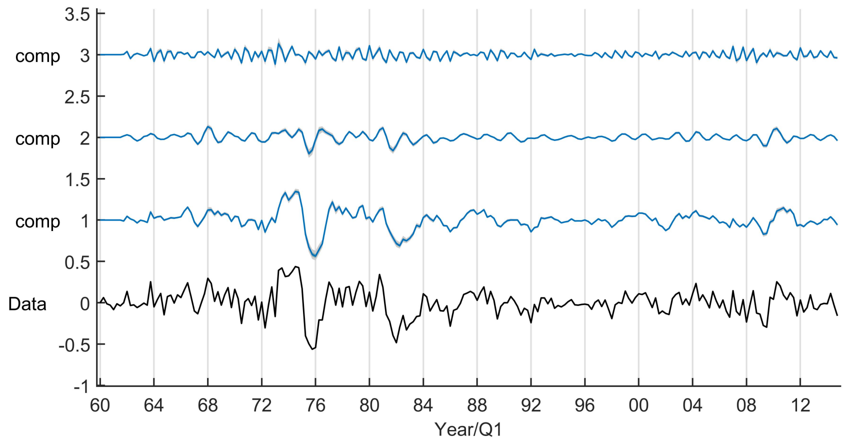

Figure 3: Posterior inferences on latent components in ΔInflnbased on AR(8) analysis.

(d) Produce and display graphical summaries– in terms of (Monte Carlo based) posterior medians, upper and lower quartiles, and 10% and 90% points of the predictive distributions of *actual inflation* over the 12 quarters following the end of the data series.

Each Monte Carlo sample of ϕ can be used to generate a path of future ΔInfln over the next 12 quarters, just recursively simulating the model into the future. Figure 4 shows the requested summaries, quarter-by-quarter, for ΔInfln. Now, to map to predictions of actual inflation we need to cumulatively sum over the coming quarters. Write y_t for actual % inflation in quarter t and x_t for ΔInfln, so that $x_t = y_t - y_{t-1}$ or $y_t = y_{t-1} + x_t$. Hence, from any "current" quarter $t = 0$, with a known value of current inflation y_0, we project over the next $h = 12$ quarters recursively; this implies the cumulative form $y_t = y_0 + \sum_{r=1:t} x_t$. From the above Monte Carlo prediction of ΔInfln with 20,000 samples, the posterior predictive samples of x_t over the next 12 time steps from the end of the data can be trivially mapped to actual inflation. Notice that the cumulation effect expands uncertainty as we go further ahead; see Figure 4.

As a bonus, note also the impact of the quasi-cyclicity in the forecasts, i.e., the Monte Carlo estimate of the forecast function.

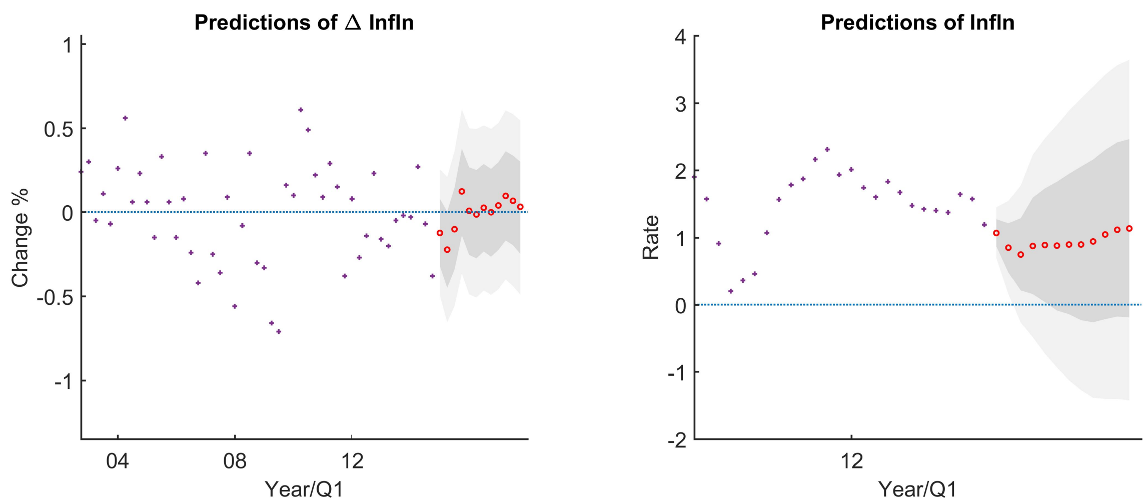

Figure 4: Synthetic futures of ΔInflnand actual Inflation.

(e) Assuming an AR(p) model is agreeable, do you think $p = 8$ makes sense for the differenced inflation series? Consider, for example, features of the fitted residuals from the model. In addition to exploratory and graphical analysis– and all you know about applied evaluation of linear regression model "fits"– the `arpcompare.m` function (noted and advised to review in Homework #2) may be of interest.

Some exploration of ACF and PACF summaries in Figure 5 indicate some quasi-periodic behaviour consistent with $p > 1$ for sure, while are strongly suggestive of $p > 5$, less strongly but evidently $p \approx 8$.

AIC, BIC and reference marginal likelihood for model order in Figure 5 also suggest $p = 8$ is interesting and relevant.

The qqplot and acf/pacf summaries of the fitted residuals from the AR(p) model, in Figure 6 indicate little evidence of non-normality which supports the normal innovation assumption. The residual correlations and autocorrelations suggest much of the dependence structure has been explained by the model, but there are perhaps some questions raised about the apparent "annual" pattern in PACF values, even though they are not obviously significant. The little bumps in both ACF and PACF at exactly 2 years might also raise a question– maybe a higher order model? or, maybe add one longer-term lag but not the intervening ones?

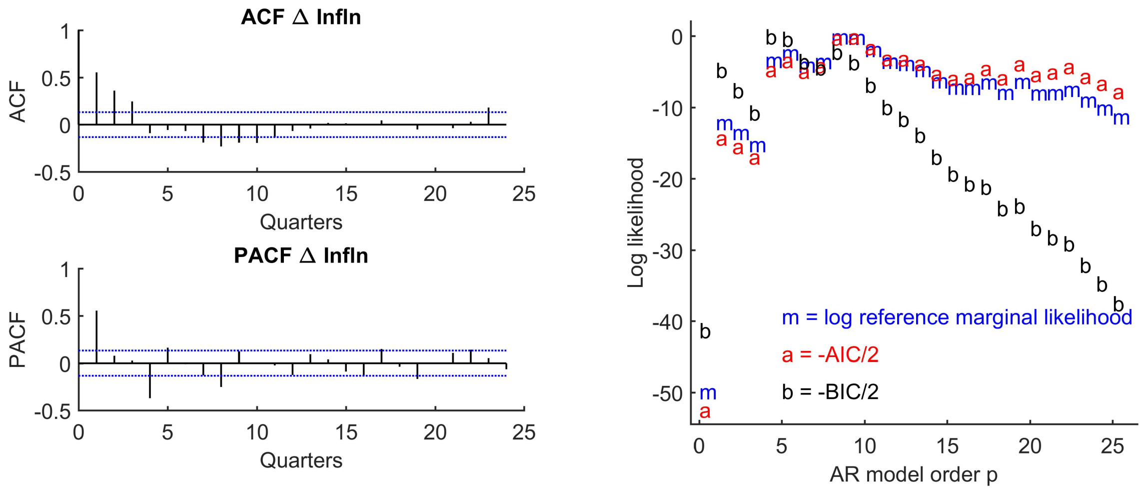

Figure 5: Sample correlations and model order assessments for ΔInfln.

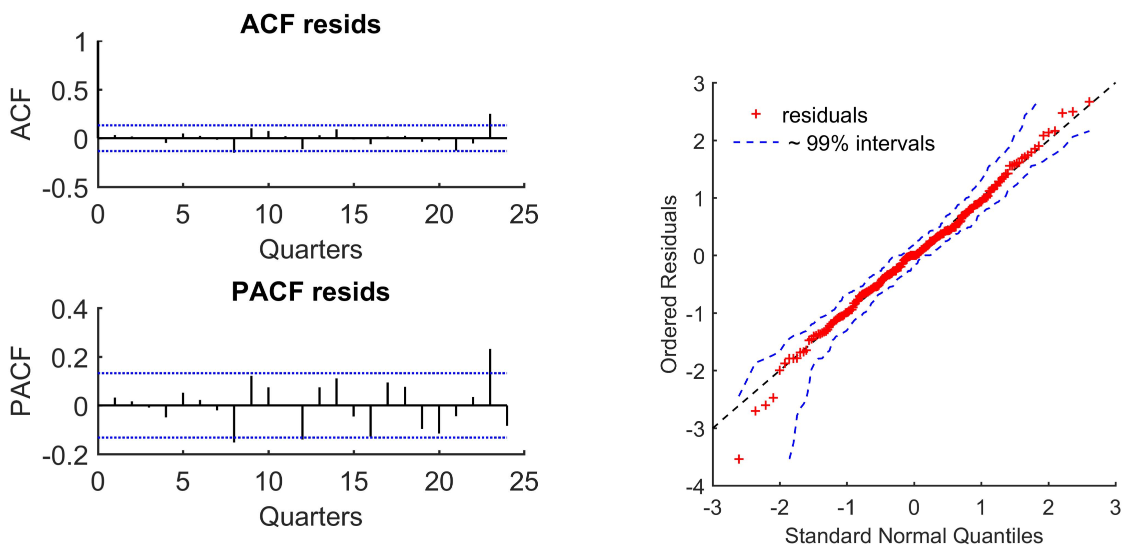

Figure 6: Sample correlations and normal qq plot for AR(8) residuals .

Bonus points for discussion of any additional analyses of other models with different values of p, comparisons, general discussion of model selection questions, etc. For example:

- Fit a higher-order model, such as with $p = 12$ or $p = 16$. Exploration of this can/may find some indication of longer wavelength quasi-periodic structure, but of course accompanied by a high level of uncertainty. Reflective and maybe rationalizing some of why economists argue so much about what the business cycle is.

- Perhaps the data plots in Figure 1 suggest that there may be different levels of variation at different time periods, i.e., changes in volatility. How might this impact these results? How to explore?