Problem Set 3: Time Series Properties and Forecasting

So far the estimated models of the Japanese Kuznets curve have not accounted for potential time dependence or lagged effects. The following exercises ask you to investigate the time series properties and use your models to forecast.

1) Time series Properties of the Data

Now we turn to the time series properties of the data:

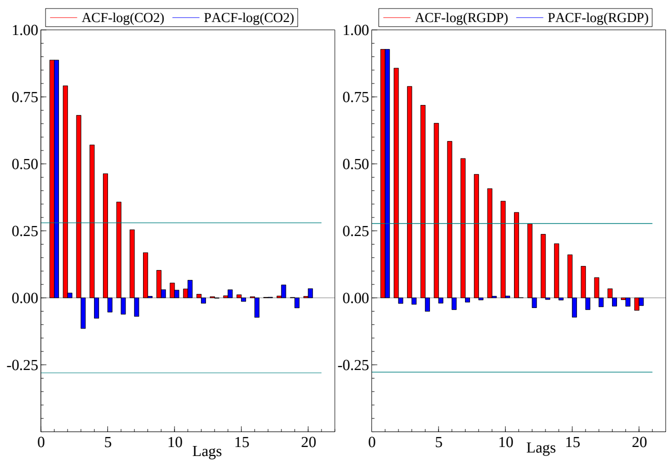
1. Plot the Auto-correlation and Partial auto-correlation functions of log(CO2) and log(RGDP) and comment on the plots. What lag length would you suggest is optimal?

2. Estimate an auto-regressive model with 1-lag (AR-1 models) for log(CO2), report the results in equation format and comment on the output.

$$\log(\text{CO2})_t = \alpha_0 + \alpha_1 \log(\text{CO2})_{t-1} + \epsilon_t \tag{1}$$

Solution:

1. Auto-correlation and partial auto-correlation functions

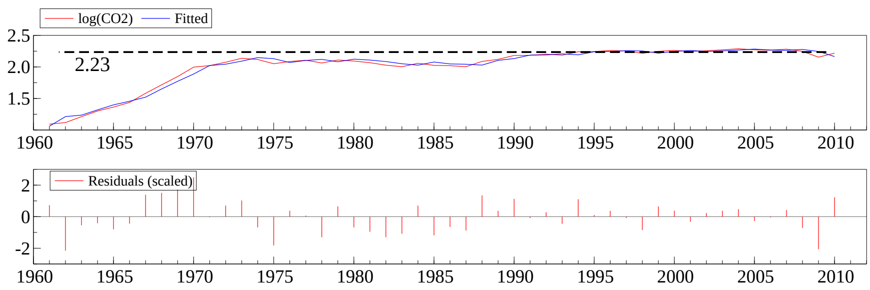


The auto-correlation for both time series decreases with the lag length, consistent with an auto-regressive process for both time series. The partial auto-correlation function for both series shows that, when controlling for all possible lags up to 20, only the first lag is significant. This suggests that an auto-regressive model of order one including one lag, (an AR(1) model), is appropriate. Failure to account for the apparent auto-correlation likely results in residual auto-correlation which leads to a violation of the assumption of independence of the error terms.

2. The estimated AR(1) model for log(CO2 per capita) is:

$$\log(CO2) = \underset{(0.018)}{0.895} \log(CO2)_{t-1} + \underset{(0.036)}{0.234}$$

The first lag of the dependent variable, the auto-regressive variable is highly significant in the regression model. This suggests that per capita CO2 emissions are relatively persistent, emissions this year are highly dependent on emissions of the previous year. The model satisfies most diagnostic tests at the 1% level of significance. The model residuals show little to no pattern confirming this result. The model long-run equilibrium $\mu = \frac{\alpha_0}{1-\alpha_1}$ is estimated to be $\frac{0.234}{1-0.895} = 2.23$ log(CO2 emmissions per capita), which corresponds to 9.29kg CO2 per capita when inverting the log transformation.



2) Auto-regressive Model Forecasts

now you will investigate the forecasting performance of the models you have estimated so far. For the current scenario, imagine it is the year 1990 and you are interested in forecasts of GPD per capita and CO2 emissions per capita.

1. Estimate AR(1) models for log(RGDP) and log(CO2) up until 1990 and compute the static long-run solution (you can set the sample length when being asked to choose the estimation sample in PcGive).

   - Use a 1-step forecast to generate forecasts from 1991-2010, plot the forecast graphs, comment on the performance. What is the root-mean-squared forecast error?

- Use a dynamic forecast to generate forecasts from 1991-2040 (50 years), comment on the outcome. How do the results relate to the long-run solution? Do the results seem sensible to you?

Solution:

1. AR(1) Models and long-run solution up until 1990:

   - log(CO2):

   $$\log(\text{CO2}) = \underset{(0.0252)}{0.894} \log(\text{CO2})_{t-1} + \underset{(0.047)}{0.235}$$

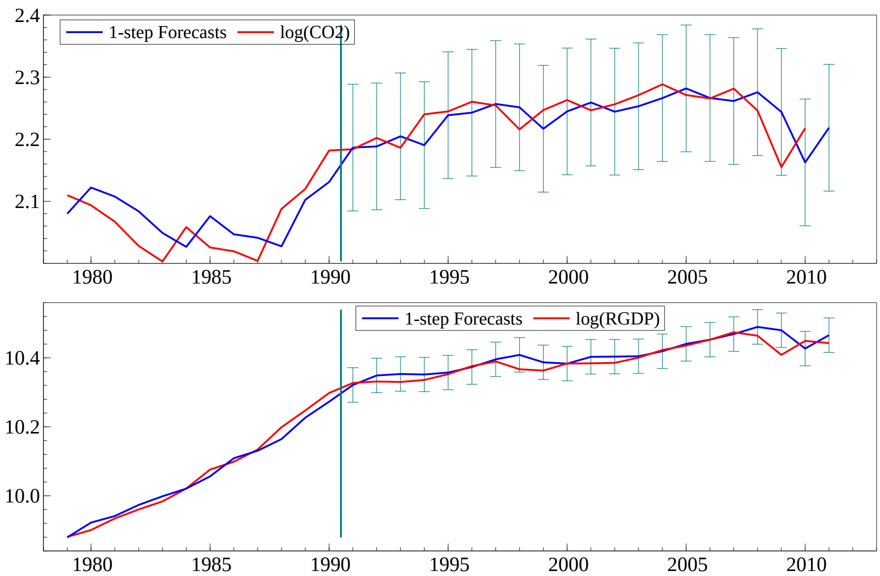
   with the long-run solution being $\hat{\mu} = \frac{\hat{\alpha_0}}{1-\hat{\alpha_1}} = 2.22$

   - log(RGDP):

   $$\log(\text{RGDP}) = \underset{(0.0106)}{0.956} \log(\text{RGDP})_{t-1} + \underset{(0.102)}{0.477}$$

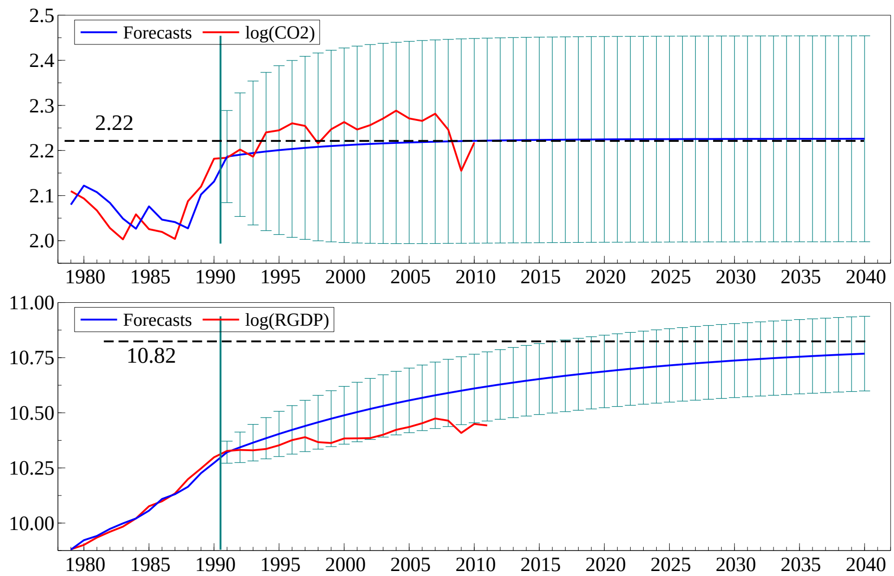
   with the long-run solution being estimated as $\hat{\mu} = 10.82$.

2. 1-step (1-year) ahead forecasts for log(CO2) and log(RGDP) from 1991-2010



Both auto-regressive forecasts closely track realized observations. Due to the high persistence in both time series (as can be seen through the high auto-regressive coefficient), this year's value provides a relatively good forecast for next year's value. Given the short forecast horizon (1-year) and high-persistence, the good forecast performance is unsurprising. However, the forecast fails (the observed value lies outside the confidence range of the forecast) when there is sudden dramatic change, this is particularly notable in 2009 when both real GDP and emissions are over-predicted. The root mean-squared forecast errors (RMSE) are 0.03 and 0.02 for the CO2 and RGDP forecasts respectively.

3. Dynamic forecasts for log(CO2) and log(RGDP) from 1991-2040



Dynamic forecasts use the previous period's predicted value to create forecasts for the next period, this allows out-of-sample forecasts to be made. Given the repeated use of predicted values, the forecast uncertainty increases quickly - for both log(CO2) and log(RGDP) forecasts the confidence bands span a wide range. The dynamic forecasts converge to the long-run mean estimated in sample. Due to the higher persistence (higher estimated auto-regressive coefficient), the log(RGDP) series has not converged to the long-run equilibrium even after 50 years. In contrast, the log(CO2) forecasts converge quickly. The models assume that the series are stationary and the estimated long-run mean accurately reflects the "true" long-run mean. This is not a sensible assumption in the context of CO2 emissions and real GDP per capita as these series are driven by economic activity and other stochastic (random) processes which change over time. Therefore very little confidence can be placed in these long-run forecasts.