

Final Data Analysis

Ryan Zhou

May 8, 2016

1.Introduction

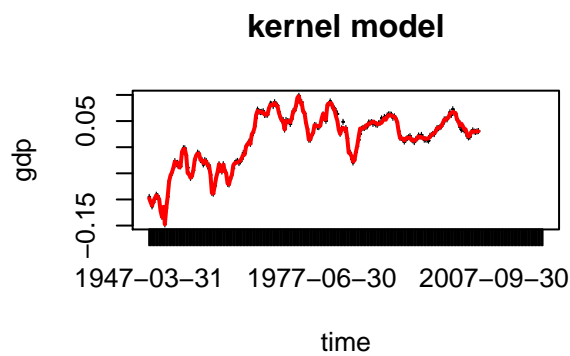
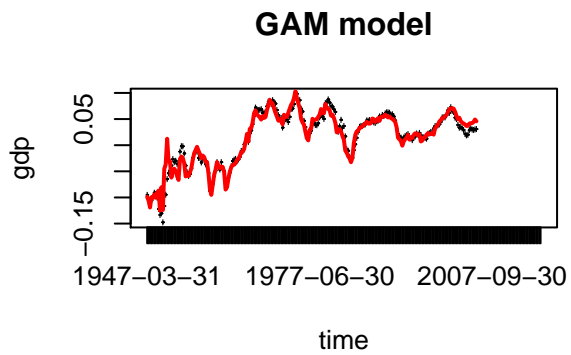
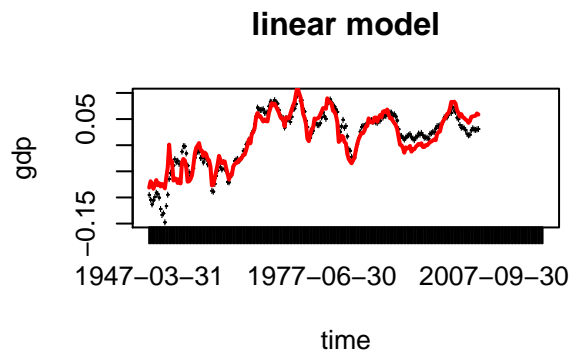
“Real business cycle” is an important and fundamental theory in modern economics. It measures how the economy reacts to the real change in society, which can be monetary, fiscal or other social causes. In this data analysis, we are particularly interested in people’s productivity. The theory believes that productivity is an exogenous variable, and all the fluctuations are driven by it. We will be using de-trended data to examine if such theory holds.

2.Specific Problems and Models

2.1

First of all, we are asked to build a regression model for GDP at time t as a function of other variables at time t . The predicting variable we will be using are consumption, investment, hours worked and productivity. Through the series courses, we have covered a few different regression models and methods and we will fit the data to each model and see how well the model is doing. And here, we will be using linear model, general additive model and kernel model.

There are two ways for us to calculate the uncertainty of a model, in-sample MSE and cross-validated MSE. First let’s calculate the in-sample MSE. Linear model gives MSE of 3.44×10^{-4} , general additive model gives MSE of 1.78×10^{-4} while kernel regression model gives MSE of 7.73×10^{-6} . Then, we will check the cv MSE for each model. Linear model gives MSE of 3.71×10^{-4} , general additive model gives MSE of 2.1×10^{-4} while kernel regression model gives MSE of 1.71×10^{-4} . We can see clearly kernel regression gives us better in-sample MSE and cross-validated MSE. Thus, for the first part of the question, we will be using kernel regression. We will also visualize it.

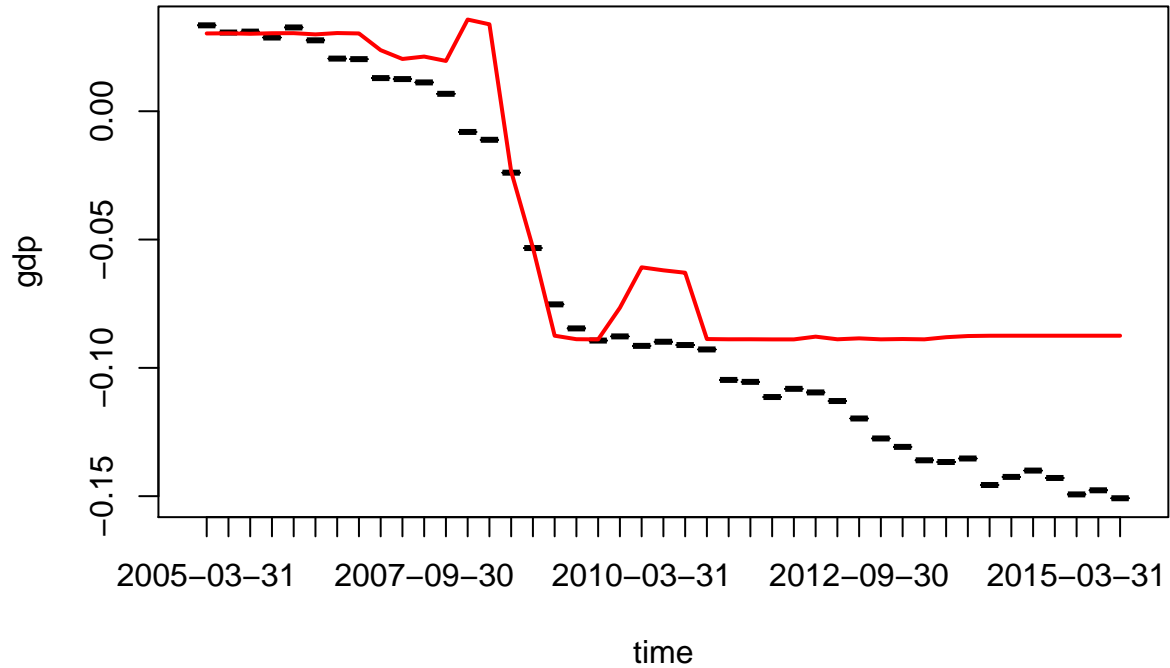


From the kernel model, we can see that investment has the widest bandwidth among the predictive variable, which implies that investment are one of the least important variable to consider for GDP and productivity has the smallest bandwidth, which means it might be more important to consider for GDP.

Then, we will try to use this model to predict the data post 2005. The way we measure the performance is by measuring the in-sample MSE by fitting the post 2005 data to our model.

As we can see, the prediction's in-sample mse is 0.00104, which is much larger the in-sample mse for the training data. Let's also visualize this on graph. As we can see from the graph, it has pretty poor performance of predicting the data especially after 2010.

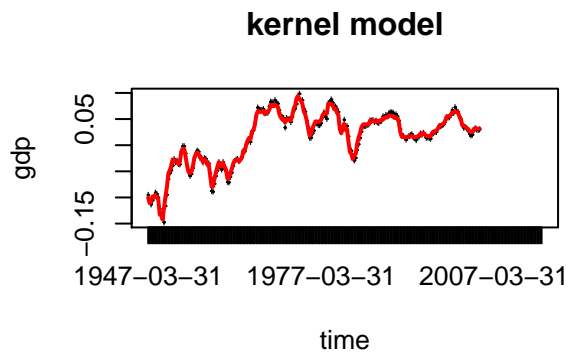
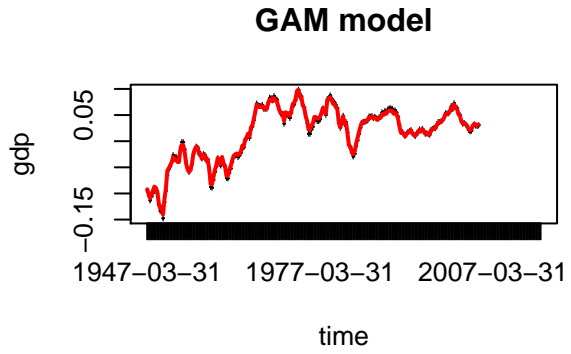
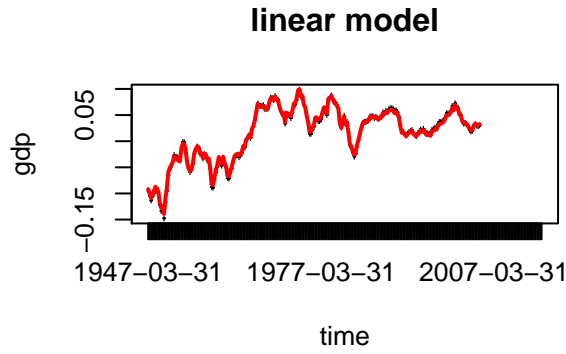
Prediction Performance



Such performance implies that using kernel regression results in extremely good performance to re-predict the data for knowing knowledge (repredicting the past) but pretty bad performance to predict the unknown knowledge (actual prediction towards the future).

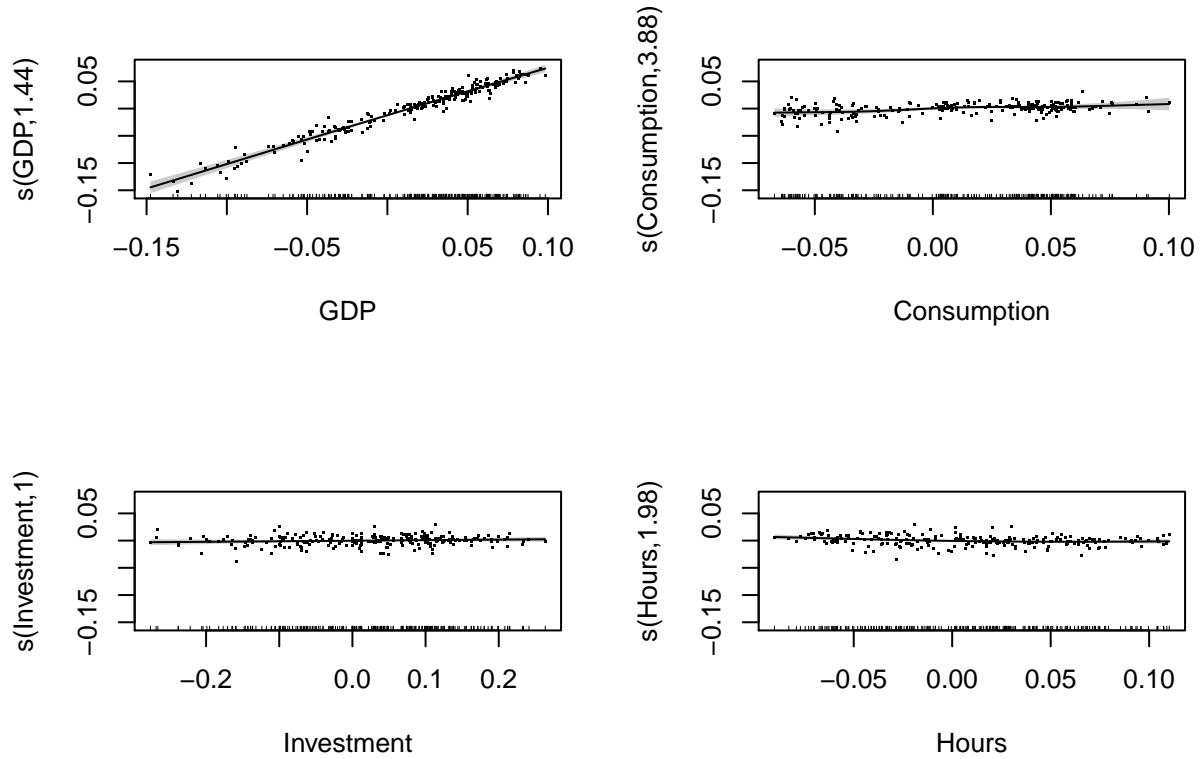
2.2

There are two ways for us to calculate the uncertainty of a model, in-sample MSE and cross-validated MSE. First let's calculate the in-sample MSE. Linear model gives MSE of 9.02×10^{-5} , general additive model gives MSE of 8.55×10^{-5} while kernel regression model gives MSE of 5.43×10^{-5} . Then, we will check the cv MSE for each model. Linear model gives MSE of 9.64×10^{-5} , general additive model gives MSE of 9.28×10^{-5} while kernel regression model gives MSE of 10^{-4} .



Although kernel has smallest in-sample MSE, its cross validated MSE is significantly larger than both GAM and linear model. Furthermore, the GAM model has smaller in-sample and cross-validated MSE than linear model. Thus, we will use GAM for this particular model.

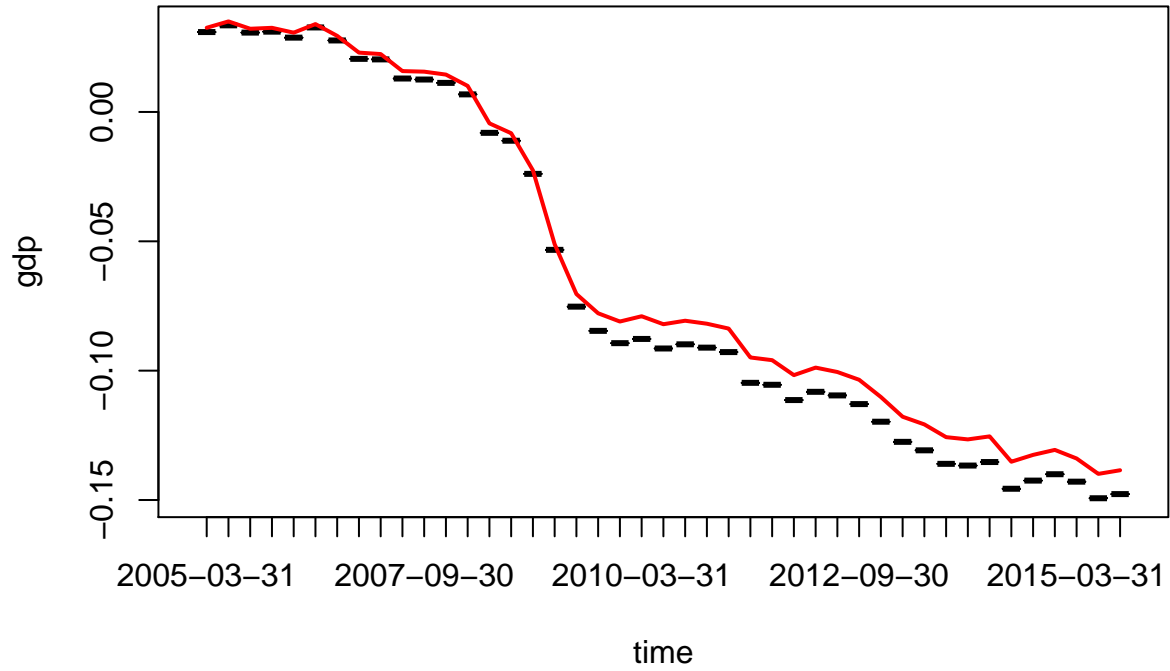
We will run a partial response for GAM here, which is visualized down here. All of the variables are linear without curvature. That is to say, next year's $\text{gdp}(gdp_t)$ grows linearly with last year's $\text{GDP}(GDP_t)$, $\text{consumption}(Consumption_t)$, $\text{investment}(Investment_t)$ and $\text{productivity}(Productivity_t)$.



Then, we will try to fit the data to our GAM to see how well GAM predicts.

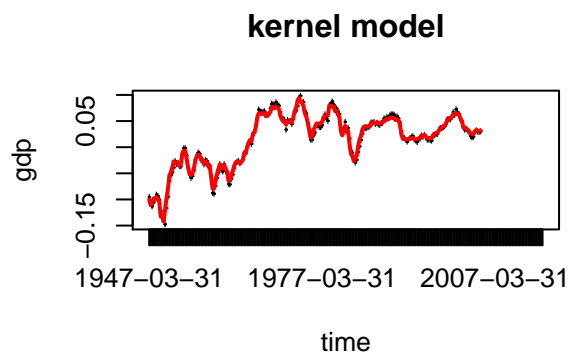
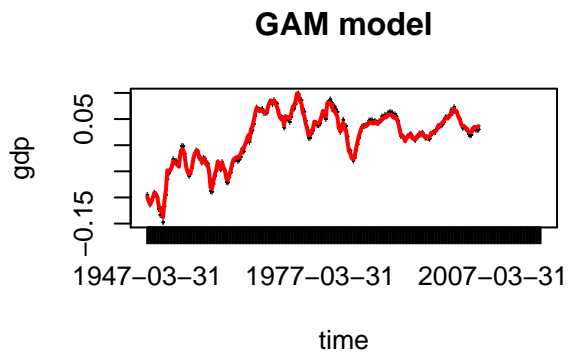
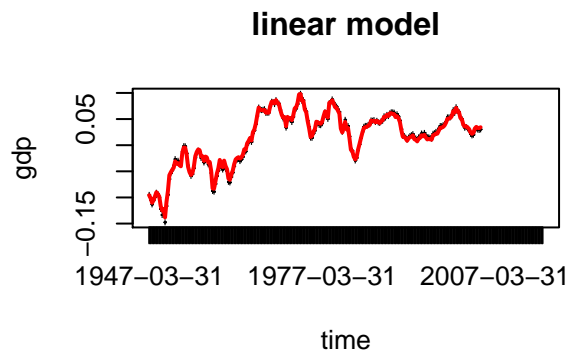
As we can see, the prediction's in-sample mse is 5.42×10^{-5} , which is a little bit smaller the in-sample mse for the training data. Let's also visualize this on graph. As we can see from the graph, it has pretty good performance for predicting the data after 2005, and it looks like a much better fit than the one in previous section.

Prediction Performance



2.3

There are two ways for us to calculate the uncertainty of a model, in-sample MSE and cross-validated MSE. First let's calculate the in-sample MSE. Linear model gives MSE of 8.81×10^{-5} , general additive model gives MSE of 7.51×10^{-5} while kernel regression model gives MSE of 5.43×10^{-5} . Then, we will check the cv MSE for each model. Linear model gives MSE of 9.45×10^{-5} , general additive model gives MSE of 8.7×10^{-5} while kernel regression model gives MSE of 9.21×10^{-5} .

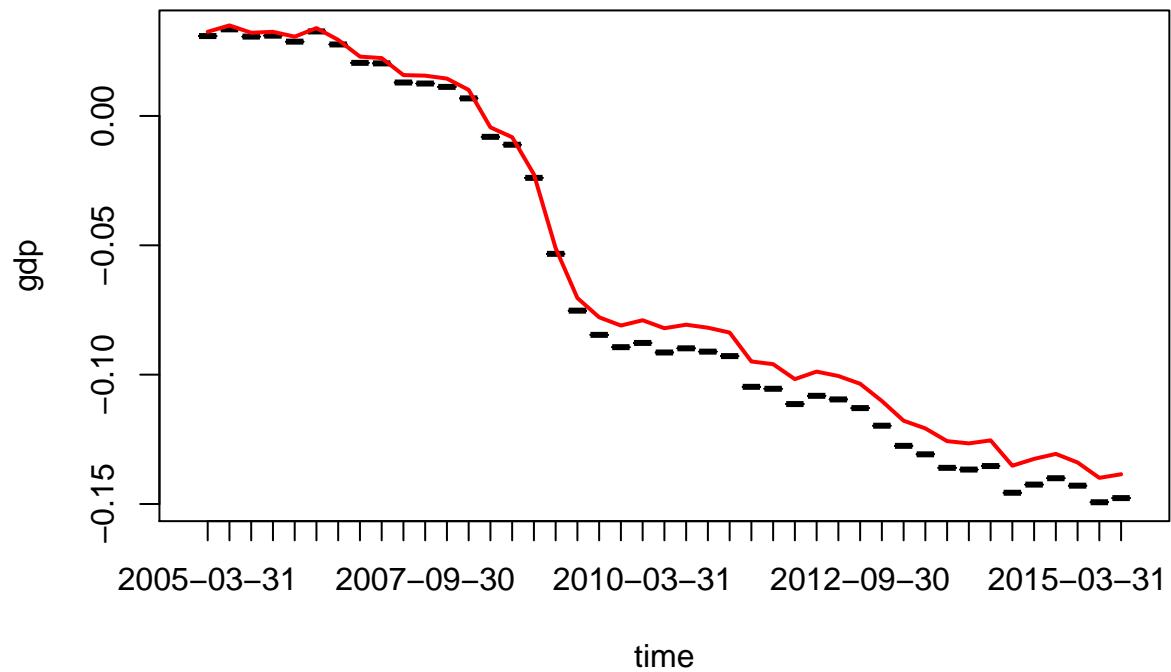


Kernel regression have smallest in-sample MSE and cross validated MSE. Thus, we will choose kernel regression for this particular model.

Then, we will try to fit the data to our kernel model to see how well it predicts.

As we can see, the prediction's in-sample mse is 1.32×10^{-4} , which is a little bit larger the in-sample mse for the training data. Let's also visualize this on graph. As we can see from the graph, it has pretty good performance for predicting the data after 2005(good fit).

Prediction Performance

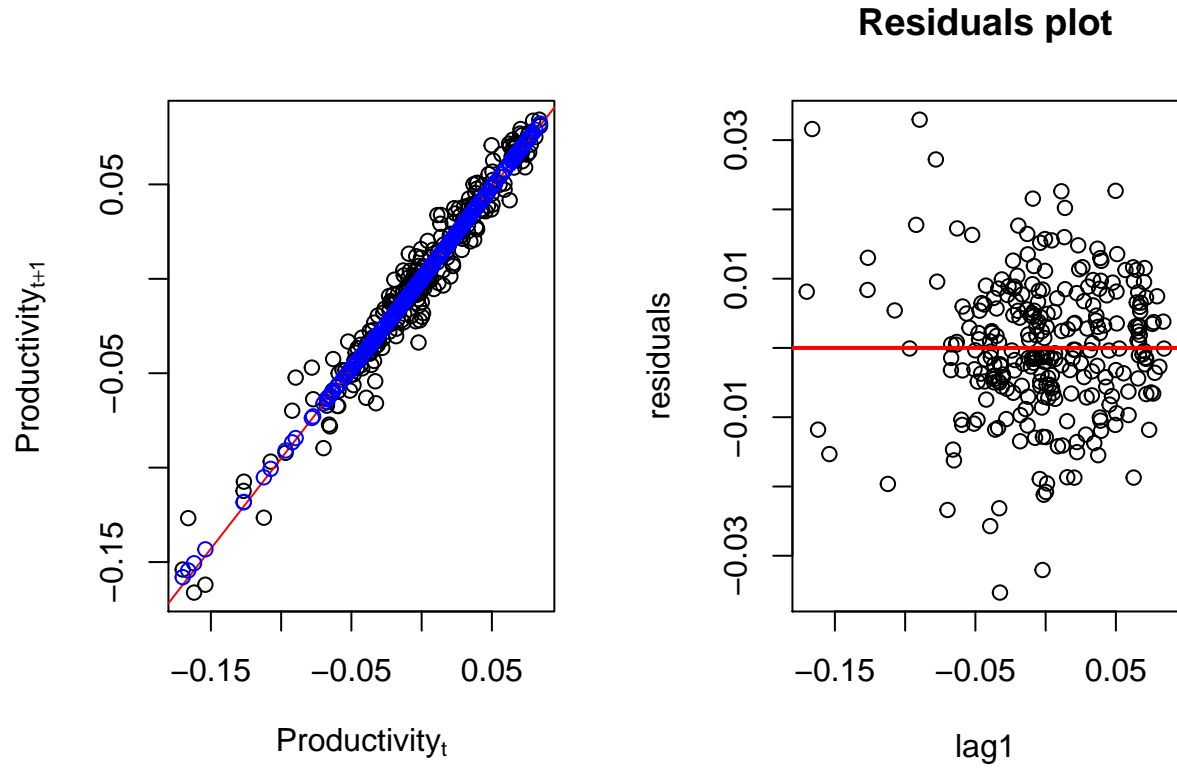


It suggests that productivity may be an important factor for driving GDP.

2.4

As we draw each model's partial response to productivity, we can see all of the plots look really similar, where the productivity shows a little curvature but around 0.

2.5



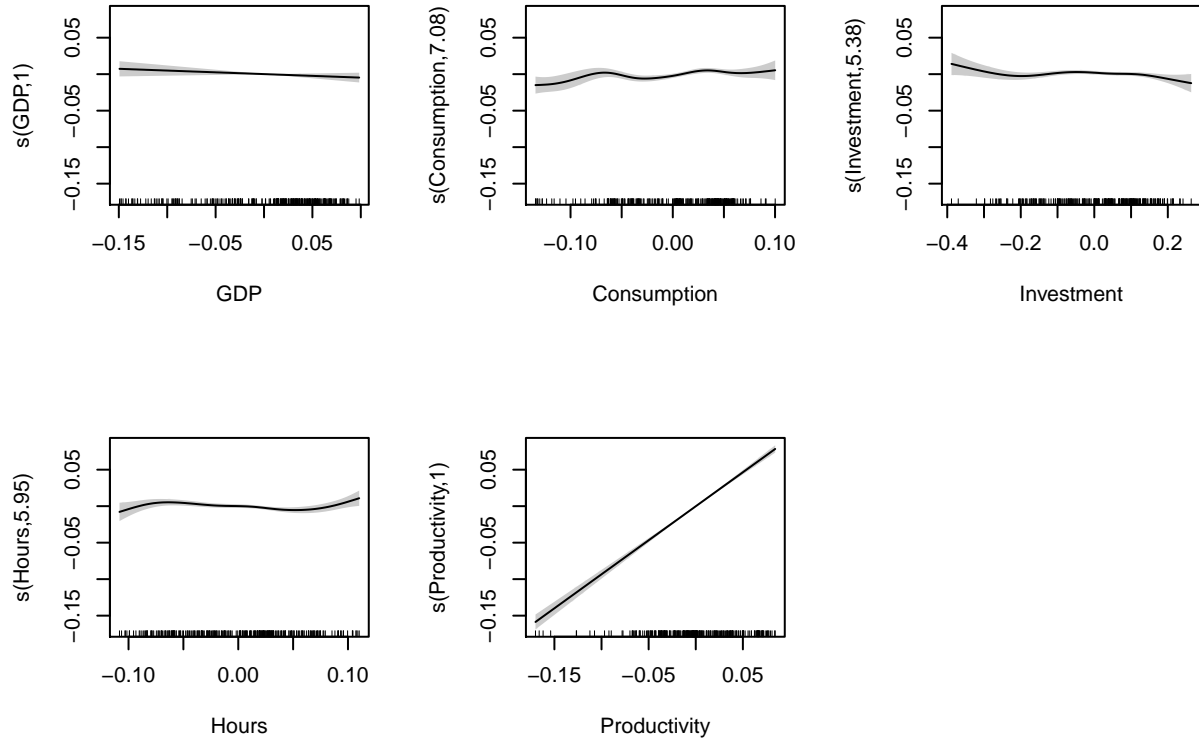
After running the autoregressive process, it shows that the process is a linear first-order autoregressive model with Gaussian noise.

2.6

There are two ways for us to calculate the uncertainty of a model, in-sample MSE and cross-validated MSE. First let's calculate the in-sample MSE. Linear model gives MSE of 9.36×10^{-5} , general additive model gives MSE of 7.25×10^{-5} while kernel regression model gives MSE of 3.05×10^{-5} . Then, we will check the cv MSE for each model. Linear model gives MSE of 1.02×10^{-4} , general additive model gives MSE of 8.53×10^{-5} while kernel regression model gives MSE of 8.48×10^{-5} .

We can see kernel regression gives both smaller in-sample MSE and cross-validated MSE. However, here we are more interested in determining the relationship between productivity and other variables, we would prefer gam so we can run the partial response to understand the relationship better.

Now, we will check the partial difference for the gam we constructed. Everything seems to be linear with productivity except last year's productivity, where there is a minor curvature.



Then, we can check how well this model performs. Similarly, we will find the mse by bootstrapping the whole data frame.

After bootstrapping, we found out that the mse for the gam model is $\text{signif}(\text{mse.prod}, \text{digits} = 3)$ and the mse for the autoregressive model is $\text{signif}(\text{mse.lag}, \text{digits} = 3)$. Thus, we can see the gam model is better. This implies that the productivity is not exogenous. If the productivity is exogenous, the autoregressive model should act better than the gam model. However, this is not the case here.

2.7 Conclusion and Suggestions

There are definitely some evidence that productivity is a little bit different from other factors to predict GDP (bandwidth in kernel regression and the partial responses). However, I don't think we are able to use analytic tool to evaluate the statement. People's productivity is a state depending on various factors and thus extremely hard to quantify. The theory is sound, but we need better analatic method to measure it. As for the future suggestions for the data analysis, I think we need to first have more detailed and broad way of describing productivity. Secondly, I think we should invest more time on testing if it is exogenous, and if so, which variables are productivity not depending on completely.