**Week 4 Lab Handout- Forecasting Models**

**PA 5033 – Multivariate Techniques**

April 2nd, 2021

Wenchen Wang

**PART A: UNCONDITIONAL FORECAST ~15 min**

**PART B: CONDITIONAL FORECAST ~15 min**

**PART C: ARIMA MODEL ~20 min**

We wish to predict the value of the Puerto Rican employment-to-population ratio for the year 1987 using three different forecasting models: an unconditional forecast, a conditional forecast, and an ARIMA model. Open prtime.dta in Canvas.

**Part A. Unconditional Forecast**: a forecast using known values of the independent variables to develop predictions for the dependent variable.



1. **Run the regression for all years prior to the year to be forecasted.**

Run the regression on all years prior to 1986 (since you are making a prediction for 1986 and 1987).



1. **Determine the regression equation.**

Write the equation for the regression of **prepop** as a function of **prgnp** and **kaitz:**

1. **Find the forecasted value of the dependent variable.**

To formulate a prediction for 1986 **prepop**, substitute in the 1986 values for **prgnp** and **kaitz** and solve the equation:

To formulate a prediction for 1987 **prepop**, substitute in the 1987 values for **prgnp** and **kaitz** and solve the equation:

1. **Calculate the forecasting error.**

To calculate forecasting error, compare predicted to actual **prepop**:

1986

1987

Since the sign on the forecasting error is positive, this implies that our predictions overestimated the actual value. We could calculate the Mean Absolute Percentage Error (MAPE) or Root Mean Square Error (RMSE) for these.

**Part B. Conditional Forecast**: a forecast using predicted values of the independent variables to predict the value of the dependent variable. Suppose you wanted to predict **prepop** using the same model as above, but you did not have the 1987 value of **prgnp** to insert into the equation. In this case, you need to first predict 1987 **prgnp** before you could determine 1987 **prepop**.



The regression equation can be determined as below:

1. **Estimate *prgnp* as a function of *usgnp*.**

Before running the regression, make sure that only the years less than 1987 are still selected.



Determine the regression equation:

1. **Find the predicted value of 1987 *prgnp*:**

To formulate a prediction for 1987 **prgnp**, substitute in 1987’s value of **usgnp** and solve the equation:

1. **Find the forecasted value of 1987 *prepop*:**

Substitute the predicted 1987 **prgnp** into the **prepop** equation from page 1 and the actual value of **kaitz** in 1987 to find the forecasted 1987 **prepop**.

1. **Calculate the Forecasting Error:**

The forecasting error in this case is:

.

Note that, as expected, the error is larger using conditional forecasting than unconditional forecasting to predict **prepop** in 1987.

## Part C. ARIMA Model: ARIMA forecasting models use prior values of the dependent variable and prior values of the error term to predicted current values of the dependent variable. The model is essentially implying that last year’s value plus a “fudge factor” is all you need to determine this year’s value. We will ignore the issue of stationarity for the sake of simplicity.

1. Create a new variable that is a one-period lag of the dependent variable. In our case, **prepop**. (This variable is already created in our dataset)
2. **Estimate the regression of the dependent variable against the lagged dependent variable and save the residuals:**

Run a regression with **prepop** as a function of **prepop\_1**, and save the unstandardized value residuals (name them ***residuals*)**. (Because we are forecasting prepop for 1987, make sure that only yr<87 are used in the regression)

Run the regression:



Determine the residuals:



1. **Lag the saved residuals:**

The residuals from the previous time period will be used to help predict the value of the current time period’s ***prepop***. Create a new variable, which will be ***residuals*** lagged one period (***residuals\_1*)**.



1. **Regress the forecasted dependent variable against the lagged dependent variable and the lagged residuals*.*** *This is a ARIMA (1,0,1) model*

Now run a regression of **prepop** as a function of **prepop\_1** and **residuals\_1**. The autoregressive term is **prepop\_1**, and the moving average term is **residuals\_1**.



Determine the regression equation:

1. **Use the regression results to find the forecasted value.**

To find the 1987 value of **prepop**, substitute in for **prepop\_1** and **residuals\_1** (i.e., **prepop** and **residuals** in 1986) and solve the equation:

1. **Calculate the forecasting error.**

The forecasting error is:

The ARIMA model underestimated the true 1987 value of **prepop** as seen by the negative forecasting error.