ISA 444: Business Forecasting

19 - ARIMA Models (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Spring 2021

Outline

Recap of What we Have Covered Last Two Weeks

ARIMA Models: Models we considered may have three components, an autoregressive component (AR), integrated (I for differencing) and a moving average component (MA).

Main Learning Outcomes from Last 3 Classes

- **Describe the behavior of the ACF and PACF of an ARMA (p,q) process.**
- **Fit an ARMA model to a time series, evaluate the residuals of a fitted ARMA model to assess goodness of fit, use the Ljung-Box test for correlation among the residuals of an ARIMA model.**
- **Use nonseasonal differencing to attain stationarity for a time series.**
- **Fit an ARIMA model to a time series, evaluate the residuals of a fitted ARMA model to assess goodness of fit, use the Ljung-Box test for correlation among the residuals of an ARIMA model.**
- **Show that you can fit reasonable ARIMA models based on both simulated and actual data cases.**

Learning Outcomes for Today's Class

Main Learning Outcomes

- **Describe AIC, AICc, and BIC and how they are used to measure model fit.**
- **Describe the algorithm used within the auto.arima() function to fit an ARIMA model.**
- **Describe the results of the auto.arima() function.**

Outline

Additional Measures of Goodness of Fit [1]

Akaike's Information Criterion: Akaike suggests measuring the goodness of fit of a model by balancing the error of the fit against the number of parameters in the model.[1]

$AIC = 2k - 2\ln(\hat{L})$,

where k is the number of parameters in the model and $\hat{L}$ is the sample likelihood function. The value of k that gives a minimum AIC gives the best model. This is simply a penalty imposed on the error variance for the number of parameters in the model.

Bias Corrected AIC: $AICc = AIC + \frac{2k^2+2k}{n-k-1}$.

The AICc is usually preferred over the AIC.

Bayesian Information Criterion: $BIC = \ln(n)k - 2\ln(\hat{L})$.

BIC is also known as the Schwarz Information Criterion (SIC). The BIC has a larger penalty for model size and tends to choose smaller models.

Additional Measures of Goodness of Fit [2]

Studies have shown:

- BIC does well at getting correct model in large samples.
- AICc tends to get correct models in smaller samples with a large number of parameters.

Why did we discuss these metrics today?

- They were printed with some of the models that we have examined in class.
- They are used with the `auto.arima()`, which comes from the forecast package (loaded with `fpp2`).

[1]Slides are based on Dr. Allison Jones-Farmer's lecture notes, Miami University, Spring 2020.

Outline

The `auto.arima()` Function [1]

The auto.arima() function can be used to automatically fit ARIMA models to a time series. It is a useful function, but it should be used with caution.[2]

The function

- Uses “brute force” to fit many models and then selects the “best” based on a certain model criterion
- Works best when the data are stationary, but can be used with nonstationary data
- Tends to overfit the data
- Should always be used as a starting point for selecting a model and all models derived from the `auto.arima()` function should be properly vetted and evaluated.

The `auto.arima()` function combines
- Unit root tests (KPSS by default)
- Minimization of AICc to obtain an $ARIMA(p, d, q)$ model using the following algorithm:

The `auto.arima()` Function [2]

1. Determine the number of differences, d, using a sequence of KPSS tests.
2. Determine p and q by minimizing AICc after differencing the data d times. Rather than considering all possible p and q combinations, a stepwise approach is taken.
    - The best initial model with lowest AICc is selected from the following four:
        - ARIMA (2,d,2),
        - ARIMA (0,d,0),
        - ARIMA (1,d,0), and
        - ARIMA (0,d,1).
        - *If d=0, then a constant, c, is included. If* $d \geq 1$*, then the constant is set to 0. The results of this step is called the current model.*
    - Variations on the current model are considered by
        - Vary p and/or q from current model by ± 1
        - Include/exclude c from current model.

The `auto.arima()` Function [3]

    - The best model considered so far (either current or one of variations) becomes the *new current model*.
- Repeat previous step until no lower AICc can be found.

[2] Slides are based on Dr. Allison Jones-Farmer's lecture notes, Miami University, Spring 2020.

Live Coding: Example 1

In class, we will use a different snapshot of the GNP data that we have explored in class so far. The purpose of this different snapshot is two-fold:

- We are not sure whether the model we fit last class would be appropriate (so this is somewhat of a revision of what to do in order to fit the "best" ARIMA model by hand).
- Walk you through the process of finding the model selected from the `auto.arima()`

```
pacman::p_load(astsa)
gnpData = gnp # will be loaded from the astsa package until 2002

# We will build on this example in class
```

Live Coding: Example 2 (Many Datasets/Models)

Similar to what we did in class 11 (Slide 5), but using the `auto.arima()` instead of `holt()`.

```
crypto = tq_get(c('BTC-USD', 'ETH-USD', 'LTC-USD', 'ADA-USD', 'LINK-USD',
                  'ZIL-USD'),
                from = '2020-10-15', to = '2021-03-03')

crypto %<>% select(c(symbol, date, adjusted))
is_grouped_df(crypto) # answer was FALSE (so we will group it)
```

```
## [1] FALSE
```

```
crypto %<>% group_by(symbol) %>% mutate(adjustedLog = log(adjusted))
```

Outline

Summary of Main Points

Main Learning Outcomes

- **Describe AIC, AICc, and BIC and how they are used to measure model fit.**
- **Describe the algorithm used within the auto.arima() function to fit an ARIMA model.**
- **Describe the results of the auto.arima() function.**

Things to Do to Prepare for Next Class

- Thoroughly read Chapters 6.2 – 6.8 of our textbook.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- Go through the posted assignment (see next slide)

Graded Assignment 17: Evaluating your Understanding

Please go to Canvas (click here) and answer the questions. **The assignment is due April 5, 2021 [11:40 AM, Ohio local time].**

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of ARIMA modeling. To reinforce your understanding of the covered material, I also suggest reading Chapter 6.1 – 6.8 of the book.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment.
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the Remote Desktop)

ISA 444: Business Forecasting

19 - ARIMA Models (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: Click here to schedule an appointment

Spring 2021