Fabrício Guimarães

Análise de Séries Temporais e Previsão de valores futuros aplicados a commodities agrícolas.

Os dados foram coletados do site <http://www.quandl.com>, sendo as commodities agrícolas café (ICE\_KC1), milho (CME\_C1), soja (CME\_S1) e o dólar ptax (BCB/1). Foram coletados dados de fechamento diário de 30 de abril de 2020 à 30 de abril de 2018.

**1 Examinar os dados**

Verificar a integridade dos dados e examinar padrões e irregularidades. Remover *outliers* e valores ausentes.

**2 Decompor os dados**

Quais são as características dos dados. Os dados possuem tendências? Sazonalidades? Ciclos?

Utilizar stl() para decompor a série em suas tendências, sazonalidades e ciclos.

**3 Estacionariedade**

A série original é estacionaria? Avaliar através dos testes:

* ACF - *Auto Correlation Function*
* PACF - *Partially Auto Correlation Function*
* Ljung-Box
* ADF - *Augmented Dickey–Fuller*

**4 Tornar a série estacionária através da diferenciação**

Utilizar a função diff() para calcular a diferenciação e tornar a série estacionária.

**5 Criar um modelo ARIMA**

Treinar um modelo ARIMA com os melhores parâmetros para as partes AR, I, MA. Fazer diagnósticos do modelo e checar os resíduos.

Utilizar a função residuals() para a avaliação dos resíduos.

**6 Utilizar o modelo ARIMA para fazer previsão**

Utilizar algumas bibliotecas para fazer previsão para períodos futuros e avaliar os erros. As bibliotecas utilizadas são:

* forecast()
* predict()
* naive()
* ses()
* holt()

**1 Examinar os dados**

Ler arquivo CSV com os dados da commodity.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

**2 Decompor os dados**

A biblioteca stl() decompõe a série temporal em suas tendências, sazionaidades e ciclos utilizando o método loess.Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**3 Estacionariedade**

* ACF - *Auto Correlation Function*
* PACF - *Partially Auto Correlation Function*

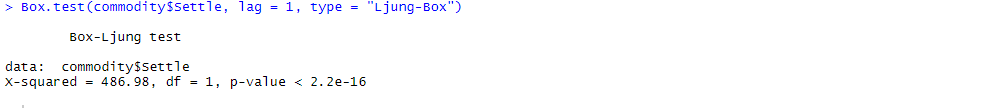
Tela de celular com publicação numa rede social

Descrição gerada automaticamente

Visualmente vemos que a função ACF demonstra que os resultados dos lag ultrapassam o intervalo de confiança. Podemos observar que esta é uma serie não estacionaria, pois os lags excedem o intervalo de confiança. Sendo assim, com a série original, não conseguimos determinar o valor do MA.

Já para o PACF, verificamos que apenas o lag 1 excede a linha tracejada que indica o intervalo de confiança. Sendo assim, podemos determinar o valor do AR para a série sendo 1.

* Ljung-Box



Examina se existem evidências de correlações diferentes de zero em determinadas defasagens. Uma série não estacionária terá o valor de p (p-value) muito baixo, conforme pode ser visto na imagem acima.

* ADF - *Augmented Dickey–Fuller*

*Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente*

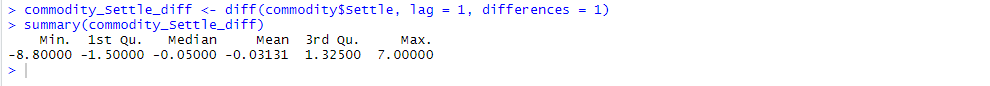
Outro teste que podemos realizar é o teste estatístico t de *Augmented Dickey – Fuller* (ADF) para descobrir se a série possui uma raiz unitária (uma série com uma linha de tendência terá uma raiz unitária e resultará em um grande valor p).

**4 Tornar a série estacionária através da diferenciação**

Geralmente, séries não estacionárias podem ser corrigidas por uma transformação simples, como a diferenciação. Diferenciar a série pode ajudar a remover sua tendência ou ciclos. A ideia por trás da diferenciação é que, se a série temporal de dados original não tiver propriedades, média, variância e autocovariância são constantes ao longo do tempo, a mudança de um período para outro pode ser. A diferença é calculada subtraindo os valores de um período dos valores do período anterior.

A montagem de um modelo ARIMA exige que a série seja estacionária.

Como verificamos nas análises de autocorrelação a série não é estacionaria. Vamos tirar a primeira diferença da série e verificar os testes ACF, Ljung-Box e ADF.



* ACF - *Auto Correlation Function*)
* PACF - *Partially Auto Correlation Function*

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

Agora podemos verificar que a função ACF demonstra que somente o lag 1 ultrapassa o intervalo de confiança. Sendo assim, podemos determinar o valor do MA para esta serie sendo 1.

Para a PACF, já havíamos visto, sem que houvesse a aplicação do filtro diff para a diferença, que encontramos o valor de AR em 1, como pode ser observado na figura, os lags não ultrapassam a linha de intervalo de confiança.

* Ljung-Box

Uma imagem contendo faca

Descrição gerada automaticamente

Uma série estacionária terá o valor de p (p-value) alto, conforme pode ser visto na imagem acima.

* ADF - *Augmented Dickey–Fuller*

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Plot da série original e da série após a primeira diferença

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

**5 Criar um modelo ARIMA**

A função **auto.arima** retorna o melhor modelo conforme os valores de p, d e q), os termos autorregressivos (AR) e de média móvel (MA), de “drift” (termo constante em modelo que tem d > 0 devido à não estacionariedade. O modelo é ARIMA(1,1,1), dizendo que houve um de autorregressão, uma derivação e um de médias móveis, respectivamente.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

**6 Utilizar o modelo ARIMA para fazer previsão**

Para a previsão, devemos criar um modelo de previsão. Mas antes de utilizarmos a série temporal para criar modelo, vamos analisar os resíduos. Estes resíduos devem ser estacionários, com média zero e variância constante no tempo.

* ACF - *Auto Correlation Function*
* PACF - *Partially Auto Correlation Function*

**Tela de celular com publicação numa rede social

Descrição gerada automaticamente**

* Ljung-Box

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* ADF - *Augmented Dickey–Fuller*

**Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente**

Vamos agora fazer um diagnóstico do modelo gerado, através da função tsdiag()

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Com o modelo com e os resíduos do modelo possuindo dados estacionários, podemos utilizar este modelo criado para fazer a previsão utilizando a biblioteca forecast().

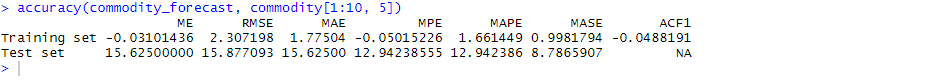
**6.1 Forecast**

Método genérico para previsão de séries temporais.

Uma imagem contendo ar, grupo

Descrição gerada automaticamente

Para verificarmos a qualidade do modelo gerado, vamos utilizar a função accuracy(), avaliando as métricas ME, RMSE, MAE, MAPE, MASE.



Vamos avaliar o resíduo deste modelo utilizando a função checkresiduals().

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**6.2 Predict**

Função genérica para predição.Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Accuracy do modelo utilizando a função predict().



**6.3 Naïve**

Para os métodos naïve, a previsão os valores previstos repetem o último valor. Adicionamos aos valores previstos limites inferiores e superiores na com intervalo de confiança em 80% e 95%. Este modelo equivale ao modelo ARIMA(0,1,0).

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Summary para o modelo gerado pela função naïve.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Verificando os resíduos do modelo gerado pela função naïve.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

**6.4 SES *Simple Exponential Smoothing***

Método de previsão para dados sem evidência de tendências ou sazionalidades. Utiliza um método naïve para previsão do próximo valor, isto é, repete o último valor da série. Contudo, o que a diferencia dos métodos naïve, é que essa metodologia é mais sensível aos últimos dados observados, atribuindo mais importância aos valores mais recentes do que aos valores mais distantes. Os métodos naïve atribuem pesos iguais a todos valores observados.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Summary para o modelo gerado pela função ses().

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Verificando os resíduos do modelo gerado pela função ses.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

**6.5 Holt’s**

É uma extensão do método *Simple Exponential Smoothing* que permite prever valores a partir de uma série de dados com tendência. Este método envolve uma equação para a previsão e duas equações para *smoothing*, um para a tendência e outro para o nível.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Summary para o modelo gerado pela função holt().

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Verificando os resíduos do modelo gerado pela função holt().

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente