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), soja (CME\_S1), milho (CME\_C1) 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 (ainda não feito nesta versão)**

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

**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 AR, I, MA. Fazer diagnósticos do modelo e checar os resíduos, os quais não podem possuir tendências e devem ser distribuídos de forma normal.

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

Utilizar as bibliotecas para fazer previsão para os próximos períodos. As utilizadas são:

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

**1 Examinar os dados**

Ler arquivo CSV com os dados da commodity. Utilizando a função summary(), temos um resumo dos dados a serem analisados. Temos o valor mínimo, o valor no primeiro quadrante, a mediana, a média, o valor no terceiro quadrante, o valor máximo. Estes dados estão sumarizados na tabela abaixo.

Tabela Função summary

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| café | 86.65 | 97.54 | 104.33 | 105.61 | 112.66 | 135.90 |
| soja | 791.0 | 858.6 | 885.9 | 888.8 | 908.9 | 1043.2 |
| milho | 302.8 | 359.8 | 372.4 | 374.1 | 386.1 | 454.8 |
| ptax | 3.481 | 3.776 | 3.926 | 4.027 | 4.136 | 5.651 |

**2 Decompor os dados**

Aprendendo a utilizar as bibliotecas stl()

* stl café

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* stl soja

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

* stl milho

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

* stl ptax

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

**3 Estacionariedade**

* ACF - *Auto Correlation Function*Tela de computador com texto preto sobre fundo branco

  Descrição gerada automaticamente

Figura Função acf

Visualmente vemos que a função ACF, conforme observado na Figura 1, demonstra que os resultados dos lag ultrapassam o intervalo de confiança. Podemos observar que estas séries não são estacionarias, pois seus respectivos lags excedem o intervalo de confiança. Sendo assim, com as séries em seu formato original, não conseguimos determinar o valor do MA.

* PACF - *Partially Auto Correlation Function*

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

Figura Função pacf(commodity)

Já para o PACF, conforme observado na Figura 2, verificamos que para todos os dados analisados apenas o lag 1 excede a linha tracejada que indica o intervalo de confiança. Sendo assim, podemos determinar o valor do AR para as séries como 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 Tabela 2.

Tabela Teste Ljung-Box

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| X-squared | 486.98 | 473.17 | 479.38 | 486.13 |
| p-value | < 2.2e-16 | < 2.2e-16 | < 2.2e-16 | < 2.2e-16 |

* ADF - *Augmented Dickey–Fuller*

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). Os valores do teste podem ser observados na Tabela 3.

Tabela 3 Teste Augmented Dickey–Fuller

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| Dickey-Fuller | -2,6718 | -3,4708 | -1,7084 | -0,066756 |
| p-value | 0.294 | 0.04511 | 0.7017 | 0.99 |

**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. Sendo assim, vamos tirar a primeira diferença da série e rodar novamente os testes ACF, Ljung-Box e ADF.

Tabela Função summary aplicado aos dados com 1 (uma) diferença

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| café | -8.8 | -1.5 | -0.05 | -0.03131 | 1.325 | 7 |
| soja | -41.75 | -6.25 | 0.0000 | -0.3713 | 4.75 | 38.5 |
| milho | -25.00 | -3.00 | -0.25 | -0.1604 | 2.5 | 21.75 |
| ptax | -0.147100 | -0.016250 | 0.003600 | 0.003869 | 0.022000 | 0.210900 |

* Plot das séries originais

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

Figura Plot das séries originais

* Plot das séries após a primeira diferença

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

Figura Plot das séries com uma diferenciação

* ACF - *Auto Correlation Function*

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, conforme pode ser observado na Figura 5.

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

Figura Função acf aplicada à series com uma diferenciação

Para a PACF, já havíamos visto, sem que houvesse a aplicação da função diff() para a diferença, que encontramos o valor de AR em 1, como pode ser observado na Figura 6, os lags não ultrapassam a linha de intervalo de confiança.

* PACF - *Partially Auto Correlation Function*

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Figura Função pacf aplicada à series com uma diferenciação

* Ljung-Box

Uma série estacionária terá o valor de p (*p-value*) alto, conforme pode ser observado na Tabela 5.

Tabela Teste Ljung-Box para séries com uma diferenciação

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| X-squared | 1.2306 | 5.0046E-07 | 1.202 | 0.00001721 |
| p-value | 0.2673 | 0.9994 | 0.2729 | 0.9967 |

* ADF - *Augmented Dickey–Fuller*

Tabela Teste Augmented Dickey-Fuller para séries com uma diferenciação

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| Dickey-Fuller | -7.6159 | -8.0381 | -8.1507 | -6.7436 |
| p-value | 0.01 | 0.01 | 0.01 | 0.01 |

**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), de “drift” (termo constante em modelo que tem d > 0 devido à não estacionariedade e de média móvel (MA). Os modelos gerados por esta função estão resumidos na Tabela 7, junto com as métricas para avaliação.

Tabela Modelos e métricas geradas pela função auto.arima

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| modelo | ARIMA(0,1,0) | ARIMA(1,0,0) with non-zero mean | ARIMA(0,1,0) | ARIMA(0,1,0) with drift |
| sigma^2 | 5.334 | 91.46 | 30.09 | 0.001481 |
| log likelihood | -1152.79 | -1861.26 | -1576.14 | 925.38 |
| AIC | 2307.58 | 3728.52 | 3154.28 | -1846.76 |
| AICc | 2307.59 | 3728.57 | 3154.28 | -1846.73 |
| BIC | 2311.81 | 3741.2 | 3158.5 | -1838.32 |

Com estes indicadores, vamos extrair os resíduos do modelo através da função residuals(MODELO). Os resíduos foram então plotados para avaliarmos se estes são estacionários.

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Figura Residuos dos modelos gerados

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. Os valores de ACF e PACF podem ser observados nas Figuras 7 e 8 respectivamente.

* ACF - *Auto Correlation Function*

**Tela de celular com publicação numa rede social

Descrição gerada automaticamente**

Figura Função acf aplicada aos resíduos

* PACF - *Partially Auto Correlation Function*

**Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente**

Figura Função pacf aplicada aos resíduos

* Ljung-Box

Tabela Teste Ljung-Box para o resíduo das séries

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| modelo | ARIMA(0,1,0) | ARIMA(1,0,0) with non-zero mean | ARIMA(0,1,0) | ARIMA(0,1,0) |
| Q\* | 14,025 | 8,2778 | 7,4357 | 9,9893 |
| df | 10 | 8 | 10 | 10 |
| p-value | 0,1719 | 0,4068 | 0,6838 | 0,4414 |
| Model df | 0 | 2 | 0 | 0 |

* ADF - *Augmented Dickey–Fuller*

Tabela Teste Augmented Dickey-Fuller para o resíduo das séries

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| Dickey-Fuller | -7,6368 | -8 | -8 | -67.448 |
| p-value | 0.01 | 0.01 | 0.01 | 0.01 |

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

* CaféTela de celular com publicação numa rede social

  Descrição gerada automaticamente

Figura Função tsdialog() aplicada ao café

* SojaTela de celular com publicação numa rede social

  Descrição gerada automaticamente

Figura Função tsdialog() aplicada à soja

* MilhoTela de celular com publicação numa rede social

  Descrição gerada automaticamente

Figura Função tsdialog() aplicada ao milho

* PtaxTela de celular com publicação numa rede social

  Descrição gerada automaticamente

Figura Função tsdialog() aplicada ao dólar ptax

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

**6.1 Forecast**

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(). Na Figura 13 vemos o resultado dos modelos aplicados aos seus respectivos conjuntos de dados.

Uma imagem contendo mapa, screenshot, texto

Descrição gerada automaticamente

Figura 14 Modelos forecast aplicado ao café, soja, milho, ptax

Para verificarmos a qualidade da previsão usando o modelo gerado pela função arima(), vamos utilizar a função accuracy(), avaliando as métricas ME, RMSE, MAE, MAPE, MASE.

Tabela 10 Função accuracy e suas métricas

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| café | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
| Training set | -0.0310143 | 2.307198 | 1.77504 | -0.050152 | 1.661449 | 0.9981794 | -0.04881 |
| Test set | 15.625 | 15.877093 | 15.625 | 12.942386 | 12.942386 | 8.7865907 | NA |
| soja |  |  |  |  |  |  |  |
| Training set | -0.4809201 | 9.544657 | 7.212441 | -0.061893 | 0.8140671 | 1.000215 | 0.001775 |
| Test set | 161.3338 | 162.19792 | 161.33383 | 15.866435 | 15.866435 | 22.373633 | NA |
| milho |  |  |  |  |  |  |  |
| Training set | -0.1593 | 5.480207 | 3.939511 | -0.05601 | 1.051303 | 0.99822 | 0.048802 |
| Test set | 81.6 | 81.63149 | 81.6 | 20.75545 | 20.75545 | 20.67637 | NA |
| ptax |  |  |  |  |  |  |  |
| Training set | 0.003868 | 0.038596 | 0.026988 | 0.084224 | 0.657353 | 0.9982714 | -0.00021 |
| Test set | 4.94073 | 494.08135 | 494.073 | 98.913478 | 98.913478 | 1.827572+e04 | NA |

Vamos avaliar o resíduo deste modelo utilizando a função checkresiduals(), que de acordo com a própria biblioteca, avalia se os resíduos de um modelo de serie temporal é um ruído branco.

Tabela Função checkresiduals para avaliar os resíduos dos modelos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | café | soja | milho | ptax |
| modelo | ARIMA(0,1,0) | ARIMA(1,0,0) with non-zero mean | ARIMA(0,1,0) | ARIMA(0,1,0) |
| Q\* | 14.025 | 8.2778 | 7.4357 | 9.9893 |
| df | 10 | 8 | 10 | 10 |
| p-value | 0.1719 | 0.4068 | 0.6838 | 0.4414 |
| Model df | 0 | 2 | 0 | 0 |

* Café**Tela de celular com texto preto sobre fundo branco

  Descrição gerada automaticamente**

Figura Função tsdiag para avaliar os resíduos para o café

* Soja**Tela de celular com publicação numa rede social

  Descrição gerada automaticamente**

Figura Função tsdiag para avaliar os resíduos para a soja

* Milho**Tela de celular com texto preto sobre fundo branco

  Descrição gerada automaticamente**

Figura Função tsdiag para avaliar os resíduos para o milho

* Ptax**Tela de celular com publicação numa rede social

  Descrição gerada automaticamente**

Figura Função tsdiag para avaliar os resíduos para o dólar ptax

**6.2 Predict**

A função predict é uma função genérica para fornecer previsões de acordo com o modelo gerado.

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

Figura Valores previstos com a função predict()

Tabela Accuracy dos modelos criados com a função predict()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ME | RMSE | MAE | MPE | MAPE |
| café | 15.625 | 15.87709 | 15.625 | 12.94239 | 12.94239 |
| soja | 161.3338 | 162.1979 | 161.3338 | 15.86644 | 15.86644 |
| milho | 81.6 | 81.63149 | 81.6 | 20.75545 | 20.75545 |
| ptax | 494.073 | 494.0813 | 494.073 | 98.91348 | 98.91348 |

**6.3 Naive**

**Uma imagem contendo mapa, screenshot

Descrição gerada automaticamente**

Figura Valores previstos pela função naive()

Tabela Summary dos modelos criados com a função naive()

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
| café | -0.031311 | 2.309448 | 1.778278 | -0.05045 | 1.664504 | 1 | -0.04893 |
| soja | -0.371287 | 9.568797 | 7.210891 | -0.04534 | 0.814348 | 1 | -3.14E-05 |
| milho | -0.160396 | 5.485602 | 3.946535 | -0.05632 | 1.053187 | 1 | 0.048642 |
| ptax | 0.003869 | 0.038634 | 0.027034 | 0.084193 | 0.658461 | 1 | -0.00018 |

**6.4 SES**

Texto preto sobre fundo branco

Descrição gerada automaticamente

Figura Valores previstos com a função ses()

Tabela Summary dos modelos criados com a função ses()

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
| café | -0.03301 | 2.304622 | 1.767481 | -0.05288 | 1.654172 | 0.993928 | -0.00162937 |
| soja | -0.3706 | 9.559339 | 7.196634 | -0.04525 | 0.812738 | 0.998023 | 1.23E-04 |
| milho | -0.16009 | 5.480206 | 3.938749 | -0.05622 | 1.051109 | 0.998027 | 0.0487879 |
| ptax | 0.003861 | 0.038596 | 0.026981 | 0.084021 | 0.657172 | 0.99804 | -0.0003927 |

* Resíduos da função ses() café

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função ses() soja

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função ses() milho

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função ses() ptax

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

**6.5 Holt’s**

**Uma imagem contendo mapa

Descrição gerada automaticamente**

Figura Valores previstos com a função holt()

Tabela Summary dos modelos criados com a função holt()

|  |  |  |  |
| --- | --- | --- | --- |
|  | AIC | AICc | BIC |
| café | 4060.304 | 4060.423 | 4081.496 |
| soja | 5445.419 | 5445.539 | 5466.551 |
| milho | 4882.58 | 4882.7 | 4903.713 |
| ptax | -139.888 | -139.767 | -118.7746 |

* Resíduos da função holt () café

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função holt () soja

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função holt () milho

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

* Resíduos da função holt() ptax

Tela de celular com publicação numa rede social

Descrição gerada automaticamente