`Comentário de Conjuntura`

Estimando um modelo de correção de erros entre a Produção de Veículos e a Produção Industrial

Vítor Wilher, Mestre em Economia e Cientista de Dados

07 de janeiro de 2020

Abstract

Nesse comentário, fazemos uma análise de cointegração entre a produção de veículos da ANFAVEA e a produção industrial divulgada pelo IBGE.

Contents

1 Pacotes e atualizações

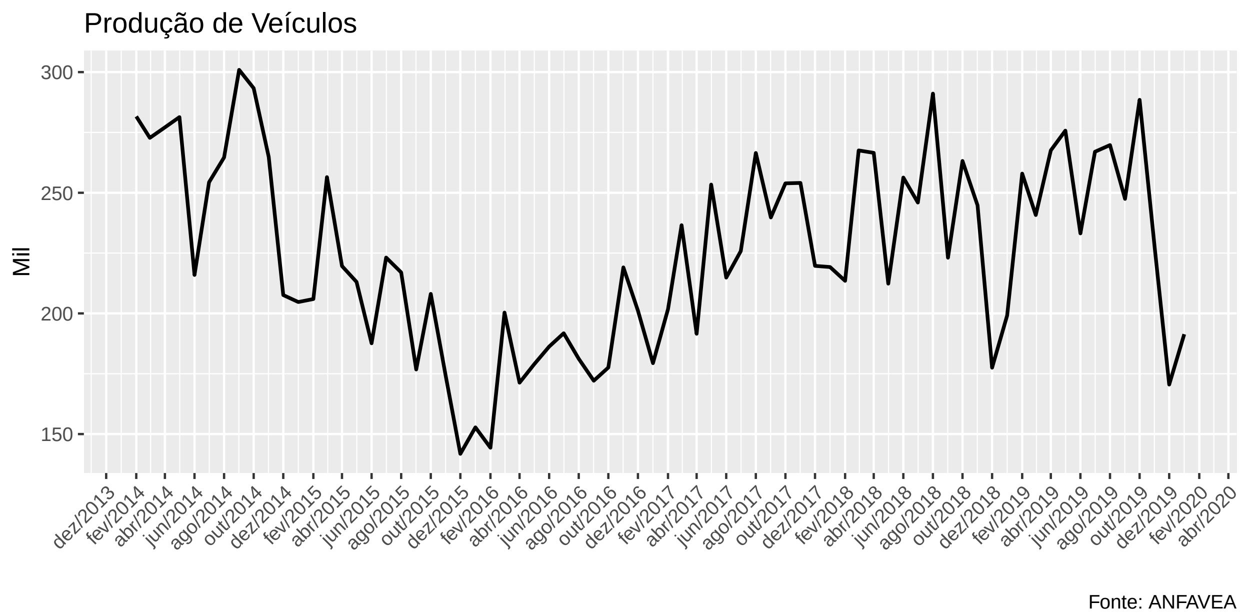
```
library(ggplot2)
library(xtable)
library(forecast)
library(gridExtra)
library(readxl)
library(dplyr)
library(magrittr)
library(scales)
library(sidrar)
library(vars)
library(dynlm)
library(stargazer)
```

2 Coleta de Dados

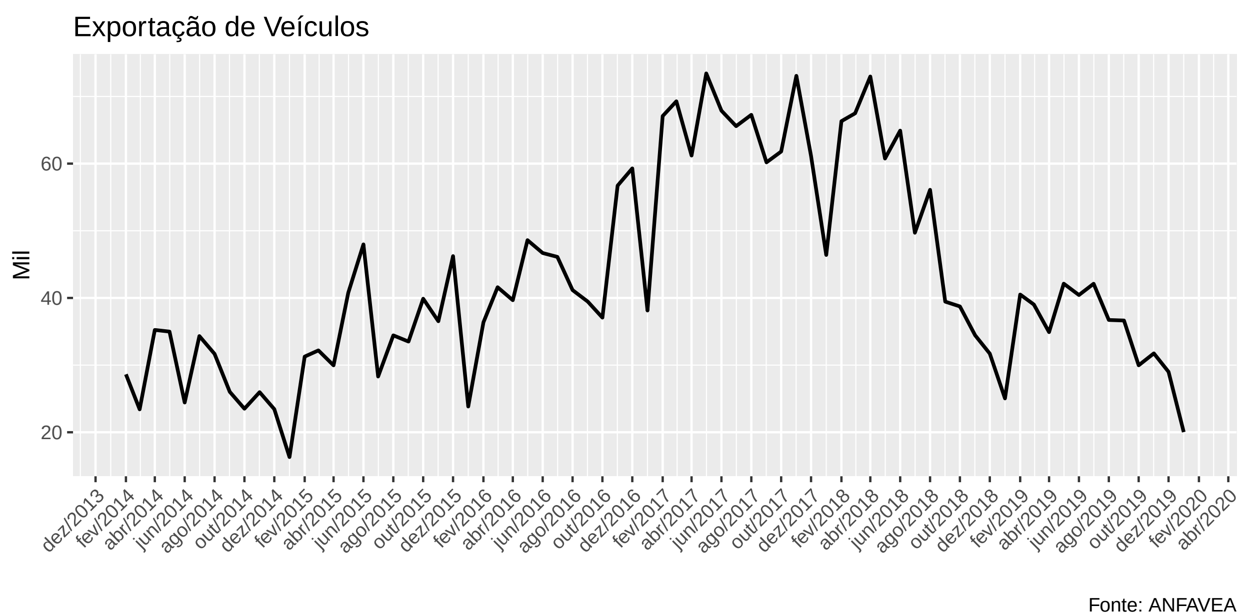
```
# Produção de Veículos
url = 'http://www.anfavea.com.br/docs/SeriesTemporais_Autoveiculos.xlsm'
download.file(url, destfile = 'veiculos.xlsm', mode='wb')
veiculos = read_excel('veiculos.xlsm', col_types = c('date',
                                                      rep('numeric', 25)),
                      skip=4)
veiculos$...1 = as.Date(veiculos$...1, format="%d/%m/%Y")
colnames(veiculos)[1] = 'dates'
veiculos = veiculos[!rowSums(veiculos[,-1])==0,]

## indústria
table1 = get_sidra(api='/t/3653/n1/all/v/3134,3135/p/all/c544/all/d/v3134%201,v3135%201')

geral_sa = table1$Valor[table1$`Variável (Código)`==3134 & table1$`Seções e atividades industriais (CNAE

geral = table1$Valor[table1$`Variável (Código)`==3135 & table1$`Seções e atividades industriais (CNAE 2

date = seq(as.Date('2002-01-01'), as.Date('2019-12-01'), by='1 month')
industria = tibble(dates=date, geral, geral_sa)
```

3 Visualização de Dados

```
filter(veiculos, dates > '2014-01-01') %>%
  ggplot(aes(x=dates, y=Produção...5/1000))+
  geom_line(size=.8)+
  scale_x_date(breaks = date_breaks("2 months"),
               labels = date_format("%b/%Y"))+
  theme(axis.text.x=element_text(angle=45, hjust=1))+
  labs(x='', y='Mil', title='Produção de Veículos',
       caption='Fonte: ANFAVEA')
```

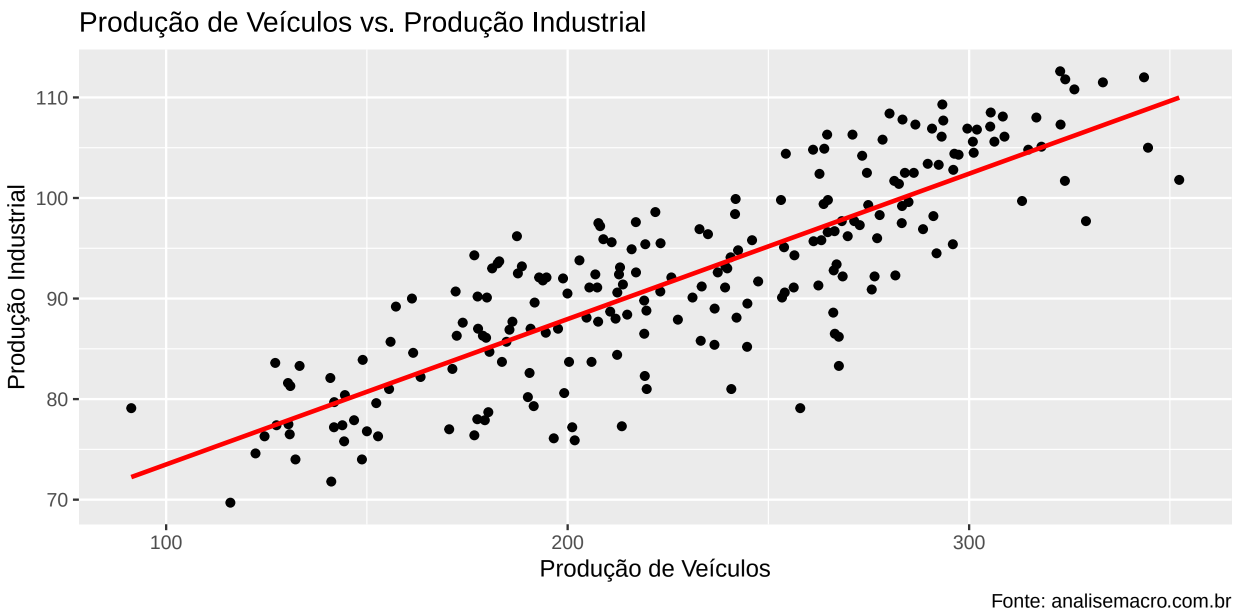



```r
filter(veiculos, dates > '2014-01-01') %>%
  ggplot(aes(x=dates, y=Exportação...6/1000))+
  geom_line(size=.8)+
  scale_x_date(breaks = date_breaks("2 months"),
               labels = date_format("%b/%Y"))+
  theme(axis.text.x=element_text(angle=45, hjust=1))+
  labs(x='', y='Mil', title='Exportação de Veículos',
       caption='Fonte: ANFAVEA')
```




```r
filter(veiculos, dates > '2002-01-01') %>%
  inner_join(industria, by='dates') %>%
  ggplot(aes(x=Produção...5/1000, y=geral))+
  geom_point()+
  geom_smooth(se=FALSE, method='lm', colour='red')+
```

```
  labs(x='Produção de Veículos', y='Produção Industrial',
       title='Produção de Veículos vs. Produção Industrial',
       caption='Fonte: analisemacro.com.br')
```



4 Cointegração

```
data = filter(veiculos, dates > '2002-01-01') %>%
  inner_join(industria, by='dates') %>%
  dplyr::select("Produção...5", geral)

colnames(data) = c('veiculos', 'industria')
```

```
### Passo 01
reg <- lm(industria~veiculos, data=data)
ur <- ur.df(resid(reg), type='trend')
ur@teststat
```

```
##                 tau3     phi2    phi3
## statistic -5.429977 9.874461 14.8087
```

```
unitrootTable(statistic='t', trend='ct')
```

```
##      0.010  0.025  0.050  0.100  0.900  0.950  0.975  0.990
##  25 -4.374 -3.945 -3.603 -3.238 -1.146 -0.820 -0.525 -0.173
##  50 -4.153 -3.795 -3.502 -3.181 -1.198 -0.882 -0.594 -0.251
## 100 -4.052 -3.727 -3.455 -3.153 -1.223 -0.911 -0.627 -0.289
## 250 -3.995 -3.687 -3.428 -3.137 -1.237 -0.929 -0.647 -0.311
## 500 -3.976 -3.673 -3.419 -3.132 -1.242 -0.935 -0.653 -0.318
## Inf -3.958 -3.660 -3.410 -3.127 -1.246 -0.940 -0.660 -0.326
## attr(,"control")
##      table      trend  statistic
## "unitroot"       "ct"        "t"
```

```
### Passo 02
data = ts(data, start=c(2002,01), freq=12)
resid <- ts(resid(reg), start=start(data), freq=12)
ecm <- dynlm(d(industria)~stats::lag(resid,-1)+d(veiculos), data=data)
```

```
stargazer(ecm, header=FALSE)
```

Table 1:

	Dependent variable:
	d(industria)
stats::lag(resid, -1)	-0.245^{***}
	(0.046)
d(veiculos)	0.0001^{***}
	(0.00001)
Constant	0.006
	(0.241)
Observations	214
R^2	0.653
Adjusted R^2	0.649
Residual Std. Error	3.530 (df = 211)
F Statistic	198.212^{***} (df = 2; 211)
Note:	*p<0.1; **p<0.05; ***p<0.01

5 Comentário

Uma exceção ao caso de regressão espúria visto anteriormente vem à tona quando dois processos aleatórios compartilham a mesma tendência estocástica.[1] Para ilustrar, considere, como Verbeek (2012), duas séries integradas de ordem 1, Y_t e X_t, e suponha que exista uma relação linear entre elas, dada por $Y_t = \beta X_t + \epsilon_{Yt}$. Isso implica no fato de existir algum valor de β tal que $Y_t - \beta X_t$ seja integrado de ordem zero, mesmo com as séries originais sendo ambas não estacionárias. Nesses casos, diz-se que as séries são **cointegradas** e as mesmas compartilham a mesma tendência.[2]

Sendo um pouco mais formal, com base em Pfaff (2008), a ideia por trás do conceito de cointegração é \textbf{encontrar uma combinação linear entre duas variáveis $I(d)$ de tal sorte que isso leve a uma variável de menor ordem de integração. Isto é,

Os elementos do vetor x_t são ditos cointegrados de ordem d, b, denominado por $x_t \sim CI(d, b)$, se todos os elementos de x_t são $I(d)$ e o vetor $\alpha(\neq 0)$ existe tal que $z_t = \alpha' x_t \sim I(d-b)$, onde $b > 0$. O vetor α é então chamado cointegrante.

[1] Para uma demostração detalhada, ver Enders (2009).

[2] Observe, por suposto, que a relação entre X_t e Y_t poderá ser caracterizada pelo vetor $[1, -\beta]'$.

Para os economistas, por exemplo, esse tipo de análise permite estabelecer relações de longo prazo entre variáveis não estacionárias. O problema, por suposto, passa a como estimar o vetor cointegrante e como modelar o comportamento dinâmico das variáveis $I(d)$. Para resolver, vamos ilustrar o **método de dois passos de Engle-Granger**, exposto em Pfaff (2008). No primeiro passo, estimamos o seguinte modelo contendo variáveis não estacionárias de mesma ordem de integração[3]

$$y_t = \alpha_1 x_{t,1} + \alpha_2 x_{t,2} + \alpha_K x_{t,K} + z_t \tag{1}$$

para $t = 1, ..., T$, onde z_t é um termo de erro. O vetor cointegrante $(K+1)$ $\hat{\alpha}$ estimado é dado por $\hat{\alpha} = [1, -\hat{\alpha^*}]'$, onde $\hat{\alpha^*} = (\hat{\alpha_1}, ..., \hat{\alpha_K})$. Assim, acaso exista uma relação de cointegração entre as variáveis, z_t nada mais é do que o erro em relação ao **equilíbrio de longo prazo** entre elas. Nesse caso, z_t será necessariamente estacionário.[4]

Se conseguirmos evidências de que z_t é de fato estacionário, podemos passar adiante. O passo seguinte é especificar um **modelo de correção de erros (ECM, no inglês)**. Para simplificar, vamos considerar, como em Pfaff (2008), o caso bivariado, onde y_t e x_t são cointegradas, sendo ambas $I(1)$. O ECM é então especificado, de forma geral, como segue

$$\Delta y_t = \psi_0 + \gamma_1 \hat{z_{t-1}} + \sum_{i=1}^{K} \psi_{1,i} \Delta x_{t-i} + \sum_{i=1}^{L} \psi_{2,i} \Delta y_{t-i} + \varepsilon_{1,t} \tag{2}$$

$$\Delta x_t = \xi_0 + \gamma_2 \hat{z_{t-1}} + \sum_{i=1}^{K} \xi_{1,i} \Delta y_{t-i} + \sum_{i=1}^{L} \xi_{2,i} \Delta x_{t-i} + \varepsilon_{2,t} \tag{3}$$

onde $\hat{z_t}$ é o erro do modelo estimado em 1 e $\varepsilon_{1,t}$ $\varepsilon_{2,t}$ são ruídos brancos. Nesses termos, o ECM na equação 2 implica que mudanças em y_t são explicadas pela sua própria estória, mudanças defasadas em x_t e pelos erros obtidos da relação de equilíbrio no passo 1. O valor do coeficiente γ_1 determina, por suposto, a velocidade de ajustamento e deveria ser sempre negativo. De outra forma, o sistema poderia divergir da sua trajetória de equilíbrio de longo prazo.

Referências

Enders, W. 2009. *Applied Econometric Times Series*. Wiley Series in Probability and Statistics. Wiley.

Pfaff, B. 2008. *Analysis of Integrated and Cointegrated Time Series with R*. Second. New York: Springer.

Verbeek, M. 2012. *A Guide to Modern Econometrics*. Editora Wiley.

[3]A suposição implícita é que as séries são integradas de primeira ordem.

[4]Pelo fato de z_t ser uma variável estimada, é preciso testar a presença de raiz unitária com outros valores críticos. No R, esses valores podem ser obtidos com a função `unitrootTable` do pacote **fUnitRoots**.