

TÓPICOS PARA O AJUSTE DISCUTIDOS NA ÚLTIMA REUNIÃO

- Ver o impacto do frete de retorno no frete da ida. As rotas que tem frete de retorno tem frete mais barato do que as que não tem? Usa fertilizante só para isso.

Ok! Variável inserida.

- Os fretes rodoviários tem comportamento de preço por faixas por quilômetros frete ate 500 é mais caro do que frete para distancia superiores a 500 criar binária para isso.

Variável não inserida. Foi criada a binária para verificar essa condição, no entanto ela não se mostrou significativa a 5%. Além disso, essa variável apresentou sinal contrário do esperado. A hipótese do do preço do frete ser menor em distâncias menores foi, de certa forma, testada nas regressões quantílicas, onde os resultados mostraram que o impacto da distância percorrida no preço do frete tende a ser menor nos valores de frete mais elevados (quantis 50% e 75%) do que que nos valores de frete menores (quantil 25%).

- Período safra e entre safra das culturas por UF. Vou usar definição do Ministério da Agricultura.

Ok! Variável inserida.

- preços das cotações internacionais ou preço pago ao produtor.

Varável não inserida. Como temos apenas a base de dados para o ano de 2014 e os valores dos preços seriam iguais para todos os indivíduos do modelo (rotas) acabei não inserindo. Caso a gente decida inserir, mesmo que a variável não traga variabilidade ao modelo, precisamos pensar como inserir. Inserimos apenas o valor médio das cotações para o ano de 2014? Ou inserimos a taxa de crescimento do valor médio das cotações nos anos 2013 e 2014?

- Qualidade das vias segundo índice criado por Jamile e Andrea

OK! Variável inserida.

- Preço do diesel

OK! Variável inserida.

VARIÁVEIS UTILIZADAS

As variáveis utilizadas no trabalho foram:

- logfrete = logaritmo natural do valor do frete medido em reais por tonelada (R\$/ton)
- xretorno = binária criada para verificar o impacto do frete de retorno no logfrete. Recebeu valor 1 caso a rota tivesse frete de retorno e 0 caso a rota não tivesse.
- XSAFRA = binária criada para verificar o impacto do período de safra no logfrete. Foi ajustada considerando o período de safra de cada cultura (milho e soja) em cada unidade da federação presente na base (MT, MS, SP, PR, GO). Recebeu valor 1 caso a data associada a rota estivesse vinculada ao período safra e 0 caso contrário.
- Xlogdistkm = logaritmo da distância da rota medida em quilômetros
- xlogqualirota = logaritmo do índice de qualidade das rota. O índice de qualidade da rota foi obtido por uma média simples entre o índice de qualidade das rodovias da UF de origem e o índice de qualidade das rodovias da UF de destino.
- Xlogdiesel = logaritmo do preço médio do litro do diesel na rota. O preço médio do litro do diesel na rota foi obtido por uma média simples entre o preço médio do diesel na UF de origem e o preço médio do litro do diesel na UF de destino.

A variável dependente (y) foi logfrete e as outras foram inseridas nos ajustes como variáveis independentes (xi). Para avaliar o efeito das variáveis independentes no valor médio de logfrete foi ajustado uma modelo de regressão linear múltipla. Já para avaliar o impacto das variáveis independentes em diferentes faixas de valores de frete foram realizados ajustes quantílicos. Esses ajustes foram efetuados para os percentis ou quantis 25%, 50% (mediana) e 75% do logfrete. Em outras palavras, procurou-se verificar os diferentes efeitos das variáveis independentes nos valores

de frete mais baratos (primeiro quartil), valores de frete intermediários (segundo quartil) e nos valores de frete mais caros (terceiro quartil).

REGRESSÃO LINEAR MÚLTIPLA

O modelo de regressão linear mostrou-se significativo a 1% (p-value: $< 2.2e-16$). Todas as variáveis inseridas também foram significativas a 1%, com exceção de xretorno que foi significativa a 5%. O ajuste possui um bom poder explicativo, com um valor de R quadrado ajustado de aproximadamente 87% (Adjusted R-squared: 0.8754).

A variável contínua com maior impacto em logfrete foi xlogdiesel. Cada 1% de aumento no preço do óleo diesel leva a um aumento de aproximadamente 2,4% no valor do frete. As variáveis xlogdistkm e xlogiqualirota também se mostraram importantes. Para cada 1% de aumento na distância da rota e no índice de qualidade da rota, espera-se respectivamente um aumento de 0,7% e 0,5% no valor do frete. O impacto positivo da qualidade da rota no valor do frete pode ser atribuído ao fato de rodovias melhores cobrarem pedágios mais elevados.

Call:

```
lm(formula = y ~ x, data = frete)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.11665	-0.13105	0.00118	0.12672	1.10854

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.578658	0.589698	-6.069	1.49e-09	***
xretorno	0.022534	0.011205	2.011	0.0444	*
xSAFRA	0.125531	0.009142	13.732	$< 2e-16$	***
xlogdistkm	0.706071	0.005888	119.925	$< 2e-16$	***
xlogiqualirota	0.580919	0.108249	5.366	8.77e-08	***
xlogdiesel	2.403254	0.418184	5.747	1.02e-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2234 on 2502 degrees of freedom
(5 observations deleted due to missingness)

Multiple R-squared: 0.8757, Adjusted R-squared: 0.8754

F-statistic: 3525 on 5 and 2502 DF, p-value: $< 2.2e-16$

As binárias XSAFRA e xretorno também apresentaram efeito positivo em logfrete. O fato do transporte de grão ocorrer em período de safra torna o valor do frete em média 12,5% mais elevado. Além disso, se a rota possuir frete de retorno com carga de fertilizantes pode-se esperar um valor de frete de grãos 2,25% mais elevado. Isso pode ser explicado pelo fato de que as rotas que possuem frete de retorno também são as mais longas, que geralmente tem como partida o local de origem da

carga de grãos e como destino o porto de Santos ou de Paranaguá, o que também torna essas rotas as mais caras. Isso faz com que o frete de retorno esteja associado as rotas mais caras e longas da base de dados utilizada no trabalho.

REGRESSÕES QUANTÍLICAS

Nas regressões quantílicas todas as variáveis foram significativas para explicar o comportamento de logfrete, com exceção de xretorno, que foi significativa apenas no ajuste referente ao quantil 25%. A variável xlogdiesel e a binária xSAFRA possuem impacto crescente a medida que se avança para as distribuições maiores de valor de frete. Na categoria mais elevada de valor de frete (quantil 75%) um aumento de 1% no preço do diesel gera um aumento de 2,91% no valor do frete, enquanto que na categoria inferior (quantil 25%), um aumento de 1% no valor do diesel gera uma elevação de 2% no valor do frete.

A qualidade da rota também se mostrou mais importante para valores de frete maiores. Uma elevação de 1% no índice de qualidade da rota (xlogiqualirota) nos percentis 50% e 75%, produz respectivamente um aumento de 0,66% e 0,65% no valor do frete. Já uma elevação similar no índice de qualidade da rota para o percentil 25% gera um aumento no valor do frete (logfrete) de 0,62%.

Já a distância da rota apresentou impacto menor nas distribuições mais elevadas de valor de frete. Em outras palavras, quanto maior o valor do frete menor o impacto da distância da rota em sua composição. Nos percentis 75% e 50%, um aumento de 1% na distância percorrida (xlogdistkm) gera respectivamente uma elevação de 0,71% e 0,73% no valor do frete (logfrete). Já na categoria de valores de fretes menores (percentil 25%) uma elevação de 1% na distância percorrida produz uma elevação no valor do frete de 0,76%. Isso sugere que quanto menor o valor do frete, maior é o espaço para que os ajustes de preço em função da distância sejam elevados. Ou então, sendo o valor do frete proporcional a distância percorrida, rotas com distâncias menores possuem preços de frete por quilômetro maiores do que as rotas mais longas.

Percentil 25%

Call: rq(formula = y ~ x, tau = 0.25)

tau: [1] 0.25

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-3.78878	0.71645	-5.28830	0.00000
xretorno	0.04856	0.01220	3.97943	0.00007
xSAFRA	0.10612	0.01082	9.81066	0.00000
xlogdistkm	0.76500	0.00733	104.41881	0.00000
xlogiqualirota	0.62117	0.13046	4.76148	0.00000
xlogdiesel	2.00842	0.50743	3.95802	0.00008

Percentil 50%

Call: rq(formula = y ~ x, tau = 0.5)

tau: [1] 0.5

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-3.91365	0.60225	-6.49833	0.00000
xretorno	0.01783	0.00957	1.86201	0.06272
xSAFRA	0.11977	0.00861	13.91024	0.00000
xlogdistkm	0.73482	0.00688	106.81841	0.00000
xlogiqualirota	0.66670	0.11198	5.95395	0.00000
xlogdiesel	2.38334	0.42863	5.56039	0.00000

Percentil 75%

Call: rq(formula = y ~ x, tau = 0.75)

tau: [1] 0.75

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-4.08040	0.64381	-6.33789	0.00000
xretorno	-0.00584	0.01015	-0.57536	0.56510
xSAFRA	0.12491	0.00870	14.35294	0.00000
xlogdistkm	0.70899	0.00733	96.66490	0.00000
xlogiqualirota	0.65201	0.11993	5.43674	0.00000
xlogdiesel	2.91309	0.45833	6.35581	0.00000