

A.2. O PROBLEMA DA MULTICOLINEARIDADE ENTRE AS VARIÁVEIS DOS MODELOS ESTIMADOS

No presente trabalho foi estimada uma série de funções de demanda para diversos alimentos, onde as variáveis independentes captam alguns fatores ligados às condições sócio-econômicas das famílias, como renda, tamanho da família e escolaridade. Sabe-se a priori que esses fatores são intensamente relacionados entre si, uma vez que cada fator pode ser causa e efeito dos demais.

Quando se está estimando funções onde comparecem variáveis desse tipo, o que importa é saber se a correlação entre essas variáveis é estatisticamente significativa, pois se a multicolinearidade derivada desse fato por muito forte, pode afetar a eficiência dos estimados.

A existência da multicolinearidade pode ser verificada pela observação da tabela de correlação simples entre cada duas variáveis, e pela comparação entre o R^2 obtido na regressão múltipla e a soma de todos os R^2 das regressões simples entre a variável dependente e cada uma das independentes separadamente.

A tabela de correlação simples obtida neste estudo é apresentada a seguir, onde se nota que os coeficientes de correlação simples são em geral bastante baixos, com exceção da correlação entre as variáveis dentro do grupo de "dummies" para educação.

TABELA 1
COEFICIENTES DE CORRELAÇÃO SIMPLES

	ln Y	ln DY ₂	ln DY ₃	ln N	ln DL	ln DEY	ln DEN
ln Y	1,0						
ln DY ₂	0,09525	1,0					
ln DY ₃	0,77133	-0,45907	1,0				
ln N	-0,32354	-0,15163	-0,17515	1,0			
ln DE	0,42619	-0,02157	0,39136	-0,07342	1,0		
ln DEY	0,47877	-0,05363	0,44919	-0,08101	0,99112	1,0	
ln DEN	0,38071	-0,02164	0,36037	0,06760	0,94996	0,93567	1,0

Pelos resultados da tabela se é levado a concluir pela existência de pequena multicolinearidade entre cada uma das variáveis renda, tamanho da família e educação e as duas outras. No entanto, a

consideração apenas da tabela de correlação simples, para tirar conclusões sobre este problema, não é suficiente, principalmente quando se utilizou variáveis "dummies". Assim, por exemplo, o valor obtido para a correlação entre a variável $\ln Y$ (log da renda da 1a. classe) e a variável DY_3 ("dummy" para log da renda da 2a. classe) de 0,77 não implica na introdução de multicolinearidade nas regressões, pois a própria forma de especificação das funções, utilizando "dummies" para separar as classes de renda, evita esse fenômeno.

Assim, além da tabela de correlação simples, estimaram-se regressões da variável dependente contra as variáveis renda, tamanho da família e educação separadamente, para comparar as somas dos R^2 "simples" com o R^2 múltiplo. Se o R^2 obtido na regressão múltipla for significativamente maior que a soma dos R^2 simples, ficaria caracterizada a multicolinearidade e, portanto, a eficiência dos estimadores seria afetada.

As estimativas foram feitas para três produtos (carne bovina, leite natural e ovos) escolhidos ao acaso, e para cada um deles as especificações funcionais adotadas foram as seguintes: (1)

$$1. \ln Q_i = \ln \alpha_0 + \alpha_1 (\ln Y_i) + \alpha_2 \ln (DY_2) + \alpha_3 \ln (DY_3) + u_i$$

para a variável renda

$$2. \ln Q_i = \ln \beta_0 + \beta_1 \ln N_i + v_i$$

para o tamanho da família

$$3. \ln Q_i = \ln \gamma_0 + \gamma_1 \ln Y_i + \gamma_2 \ln (DY_2) + \gamma_3 \ln (DY_3) + \gamma_4 \ln (DE) + \gamma_5 \ln (DEY) + \gamma_6 \ln (DEN) + w_i$$

para se obter o poder explicativo da variável educação.

Nota-se que a função (1) não se trata propriamente de uma regressão simples, pois tem três variáveis independentes. Como, porém, duas das variáveis são "dummies", não seria possível estimar funções com cada uma das dummies separadamente. Algo semelhante ocorre com as variáveis para educação, todas especificadas na forma de "dummies", o que obrigou a estimar uma função do tipo (3), para que se obtivesse o poder explicativo de educação pela diferença entre o

(1) A denominação das variáveis é a mesma do item 6 do capítulo III.

R^2 da função (2) e o de (1). Portanto, os tipos de especificação adotadas não satisfazem efetivamente às necessidades do teste de multicolinearidade, que exige regressões simples contra cada uma das variáveis independentes, o que não foi possível devido à presença de "dummies", mas permite que se tenha uma idéia do problema.

Obtiveram-se, então, os seguintes resultados:

COEFICIENTES DE EXPLICAÇÃO - R^2					
ESPECIFICAÇÃO (1)	ESPECIFICAÇÃO (2)	ESPECIFICAÇÃO (3)	(3)-(1)= (4)	$\Sigma(1)+(2)+(4)$	MÚLTIPLO
0,2598	0,0807	0,2604	0,0060	0,3411	0,2751
0,1197	0,1215	0,1251	0,0054	0,2467	0,1855
0,1481	0,1454	0,1591	0,0110	0,3045	0,2273

Como mostram os resultados do quadro, para cada um dos três produtos testados o R^2 múltiplo (última coluna) não diferiu significativamente da soma dos R^2 simples (coluna 5).

Portanto, se é levado a concluir que, ainda que haja uma pequena multicolinearidade implícita nas funções estimadas, esta é pequena, o que não deve prejudicar a eficiência dos estimadores obtidos nas regressões múltiplas.