

METODOLOGIA DE DIFERENÇAS-EM-DIFERENÇAS

Em economia, muitas pesquisas são feitas analisando os chamados experimentos naturais. Nas palavras de Wooldridge (2003), os experimentos naturais ocorrem quando algum evento exógeno, como, por exemplo, uma mudança de política do governo, muda o ambiente no qual indivíduos, famílias, firmas ou cidades operam. Para analisarmos um experimento natural sempre temos que ter um grupo de controle, isto é, um grupo que não foi afetado pela mudança, e um grupo de tratamento, que foi afetado pelo evento, ambos com características semelhantes. Ao contrário de um experimento real em que os grupos de tratamento e controle são escolhidos aleatoriamente para impedir viés nas estimativas, os grupos em um experimento natural emergem da forma com que a mudança é efetuada. Para estudarmos as diferenças entre os dois grupos precisamos de dados de antes do e de depois do evento para os dois grupos. Assim, nossa amostra é dividida em quatro grupos: o grupo de controle antes da mudança, o grupo de controle depois da mudança, o grupo de tratamento antes da mudança e o grupo de tratamento depois da mudança.

Esquemáticamente, podemos representar o procedimento a partir do seguinte quadro.

	Antes	Depois	<i>diferenças</i>
Controle	A	B	A-B
Tratamento	C	D	C-D
<i>Diferenças</i>	A-C	B-D	(C-D)-(A-B)

A-B e C-D representam em que medida o grupo de controle e o de tratamento se alteraram, respectivamente, entre o período anterior e posterior ao evento que está sendo examinado. Como por hipótese o grupo de controle não sofreu impacto do evento, essas mudanças se deveram a outros fatores, que também devem ter influenciado o grupo de tratamento. Já A-C e B-D representam as diferenças entre os grupos de controle e de tratamento antes e depois do evento, respectivamente.

Subtraindo então A-B de C-D, ou A-C de B-D, que é exatamente a mesma coisa, encontraremos a diferença da diferença verificada entre os grupos, entre os 2 períodos, ou

visto pelo outro lado, a diferença verificada entre a diferença entre os 2 períodos, entre cada um dos grupos. Daí a razão do nome diferenças-em-diferenças, ou dif-in-dif.

Matematicamente, podemos representar o método de diferenças em diferenças com a seguinte equação:

$$g3 = (y_{2,b} - y_{2,a}) - (y_{1,b} - y_{1,a})$$

onde cada Y representa a média da variável estudada para cada ano e grupo, com o número subscrito representando o período da amostra (1, para antes da mudança e 2, para depois da mudança) e a letra representando o grupo a qual o dado pertence (A, para o grupo de controle e B, para o grupo de tratamento). E g3 será nossa estimativa a partir da diferenças em diferenças. Obtendo g3 determinamos o impacto do experimento natural sobre a variável que gostaríamos de explicar.

Representando o método através de uma regressão e criando as variáveis indicadoras (ou dummies): dB, igual a um para os indivíduos do grupo de tratamento e zero para os indivíduos do grupo de controle; e d2, igual a um quando os dados se referem ao segundo período, pós-mudança, e zero caso os dados se refiram ao período pré-mudança, temos:

$$Y = g0 + g1*d2 + g2*dB + g3*d2*dB + \text{outros fatores}$$

onde Y representa a variável estudada, g1 o impacto de se estar no segundo período sobre a variável estudada, g2 o impacto de se estar no grupo de tratamento sobre a variável estudada, e g3 o impacto pós-evento do grupo de tratamento vis-à-vis do grupo de controle sobre a variável estudada (que é justamente o que se quer descobrir). Assim, g0 capta justamente o valor esperado da variável estudada quando se analisa o grupo de controle antes da mudança, o que nos dá, basicamente, o parâmetro de comparação.

No entanto, é preciso controlar por outros fatores relevantes na regressão, o que no jargão econométrico quer dizer que, antes de alegarmos que g3 nos dará o impacto da política exógena, temos que descobrir e isolar o efeito de todas as outras variáveis que podem estar causando mudanças na variável estudada. Isso é feito inserindo as variáveis de controle relevantes na regressão, como foi mostrado na segunda equação, evitando-se assim que efeitos de outras variáveis produzam viés na nossa estimação. Com esse procedimento determinamos, portanto, o efeito puro do experimento natural sobre a variável que gostaríamos de explicar.