

O emprêgo de deflatores inadequados e o problema de êrro comum nas variáveis em estudos econométricos ¹

KENNETH KING

A maioria das estatísticas econômicas é compilada em têrmos nominais enquanto os modelos são construídos em têrmos reais. Nas abordagens estatísticas aos modelos econômicos os valores nominais são deflacionados por algum índice de preços de maneira a aproximar o comportamento das variáveis reais. Se o índice de preços representa falsamente o seu comportamento, as variáveis deflacionadas poderão conter uma covariância sistemática que as verdadeiras variáveis reais não encerram. Êste problema tem sido praticamente ignorado na literatura teórica bem como nos estudos econométricos que dependem fundamentalmente de variáveis deflacionadas. No entanto, como se argumenta abaixo, em certas condições os vieses resultantes do emprêgo de um deflator inadequado podem revestir-se de extrema seriedade.

A natureza do problema é facilmente apreciada mediante o exame da especificação típica de um modêlo econométrico, como por exemplo,

$$\frac{M_t}{P_t} = a + b \frac{Y_t}{P_t} + c \dot{p}_t \quad (\text{eq. 1})$$

A equação 1 é uma versão simples da equação da demanda monetária onde M_t e Y_t representam demanda por moeda e renda em têrmos

¹ As críticas e sugestões de Arnold Herberger e do Corpo Editorial desta Revista foram muito úteis. Humberto de Lemos do IBRE-EPGE (Fundação Getúlio Vargas) escreveu os programas de computador usados nas simulações.

Nota da Redação — O autor pertence à Universidade de Oxford, Inglaterra e encontra-se atualmente escrevendo sua tese de doutorado na Escola de Pós-Graduação em Economia (EPGE) da FGV.

nominais no período t ; P_t é o valor de um índice de preços utilizado como deflator e \dot{p}_t representa a taxa de variação de P_t . A equação é uma aproximação de um modelo que relaciona saldos monetários reais, renda real e custo esperado de manutenção de saldos em caixa. Suponhamos que num período t a mudança no índice de preços exagere o aumento real dos preços; os valores deflacionados M/P e Y/P serão subestimativas da demanda real e da renda real nesse período, enquanto \dot{p}_t superestima a taxa de variação dos preços. Suponhamos que no período seguinte a variação no índice de preços subestima a verdadeira elevação do nível de preços. As variáveis deflacionadas serão, portanto, superestimativas dos valores reais e a taxa de inflação será mais alta do que a indicada por \dot{p}_t . Erros repetidos no índice de preços resultarão num viés ascendente do coeficiente b e descendente do coeficiente c na equação 1. Este exemplo contém os vieses que contribuem para o tipo de coeficientes sugeridos pela teoria, isto é, com saldos reais dependendo positivamente da renda real e negativamente do custo de conservar saldos.

Por outro lado, o viés não é reduzido pela compensação nos erros. No exemplo acima é possível que o índice de preços reflita acuradamente uma variação no nível de preços durante dois períodos tomados em conjunto, ao mesmo tempo em que subestima um período e superestima o outro. Obviamente, isto induzirá dois movimentos espúrios nas variáveis deflacionadas, e uma rápida reflexão sobre a definição dos parâmetros econométricos é suficiente para demonstrar que o viés introduzido pelos erros no deflator é cumulativo no tempo e não se cancela por um simples processo de promediamento.

Que importância poderá ter este viés? Em que condições ele afetará seriamente os estudos que utilizem dados deflacionados? A resposta a essas perguntas exige a especificação dos prováveis mecanismos de erro dos índices de preço e o estudo da variância do erro relativamente à variância das variáveis reais.

Natureza dos erros num deflator de preços

O deflator "verdadeiro" de preço é melhor definido pelo método utilizado para a construção de um deflator implícito do PNB. Dividindo o valor nominal do produto em qualquer ano pelo valor real

medido em unidades de moeda de um determinado ano-base, obtemos o valor do verdadeiro deflator de preços relativo àquela base. A definição é circular, entretanto como o deflator “verdadeiro” de preço não pretende desempenhar qualquer outra função além de relacionar um valor nominal a um valor real, a questão da circularidade torna-se irrelevante. Existindo informações perfeitas, o deflator “verdadeiro” de preços pode ser construído para qualquer período desejado. Na ausência das informações necessárias, algum outro índice “proxy” de preços terá que ser empregado, como por exemplo, um índice de custo de vida, ou um índice de preços por atacado. A diferença entre o deflator “verdadeiro” e o índice “proxy” é aqui denominada de erro na “proxy”. Pode-se, proveitosamente considerar o erro como contendo dois componentes, o primeiro atribuível a diferentes padrões sazonais, no deflator e na sua “proxy”, e o segundo como proveniente de erros de amostragem na “proxy”. Conforme se verá na discussão abaixo, somente os erros nos índices mensais ou trimestrais de preços precisam ser considerados uma vez que os erros numa “proxy” anual — caso fôsse necessário utilizá-la — seriam, ao que tudo indica, extremamente pequenos. Numa base apriorística, parece provável que um índice de custo de vida exiba um comportamento diferente, dentro de um ano, daquele do nível global de preços. As flutuações dos preços dos alimentos e as mudanças no custo das utilidades públicas por exemplo, afetarão o comportamento, a curto prazo, do primeiro em muito maior extensão do que do segundo. Considerações semelhantes possivelmente se aplicarão a qualquer outra “proxy” do deflator implícito. Em segundo lugar, embora teoricamente o índice de preços será uma amostra destinada a indicar o comportamento do preço de todos os bens abrangidos pelo mesmo, praticamente, apenas um número limitado é incluído, e a cada um é atribuída uma ponderação em si mesma sujeita a erro. Dêste modo, o índice por sua própria natureza, sofre de erros de amostragem. Poucas pessoas levariam a sério as variações de preços indicada pelo comportamento do índice mensal, ou mesmo trimestral, pelos motivos já indicados, uma vez que automaticamente tende-se a incluir certa margem de erro; no deflacionamento de uma série por um índice de preços normalmente este problema não é levado em consideração.

Existem diferentes maneiras de especificar o mecanismo de erro num índice de preços, mas três apenas servirão às nossas finalidades. A suposição mais extrema consiste em supor que o índice de preços contém erros (dos dois tipos mencionados acima), de tal magnitude que serve apenas para indicar tendência — *Especificação I* — e que portanto qualquer informação adicional será considerada espúria. O segundo método consiste em salientar a propriedade de amostragem do índice, de modo que

$$E_t = P_t^+ - P_t = \varepsilon_t P_t$$

onde E_t constitui o erro no índice “proxy” de preços — definido como a diferença no período t entre o mesmo e o deflator implícito; P_t^+ é o índice “proxy” e, P_t , o deflator implícito. ε_t é um número aleatório que recebe limites apropriados dependendo da confiabilidade do índice “proxy”, como, por exemplo

$$- 0,01 \leq \varepsilon_t \leq + 0,01$$

Neste caso o erro de amostragem é definido no intervalo de $\pm 1\%$ de modo que, no caso extremo, o índice “proxy” poderia indicar um aumento de preços de 2% enquanto na realidade o índice de preços permaneceria constante (ou que o “proxy” poderia indicar preços constantes enquanto, na verdade, subiria em 2%) entre quaisquer dois períodos. Dada a especificação do mecanismo de erro, tais extremos serão alcançados apenas raramente, de modo que a performance implícita do índice “proxy” de preços não é de todo irrealista. A especificação de ε_t (aleatório e centrado em zero) mostra que é impossível saber que confiança merece o índice “proxy”. No caso do índice de preços ser mensal, a soma de doze ε_t provavelmente se aproximará de zero e o índice funcionará bem em comparação com um deflator anual implícito. Esta especificação não prevê qualquer divergência sistemática a longo prazo entre a “proxy” e o verdadeiro índice e, por isso mesmo, o que de certo modo favorece a utilização de índices “proxy”.

A *Especificação II* encerra a propriedade de que o viés introduzido pelos erros na “proxy” é independente do comportamento tanto do deflator implícito como do índice “proxy” e depende apenas da

variância de ε_t em relação à variância das variáveis reais em estudo. Isto pode ser facilmente compreendido: suponhamos que R_t é o valor real da variável no período t ; por definição, seu valor nominal é igual ao seu valor real multiplicado pelo deflator implícito, isto é, $P_t \cdot R_t$. A *especificação II* mostra que o deflator realmente utilizado é $P_t + \varepsilon_t P_t$ de modo que a aproximação deflacionada da variável real será igual a,

$$\frac{P_t R_t}{(1 + \varepsilon_t) P_t} = \frac{R_t}{1 + \varepsilon_t}$$

A *Especificação II* é satisfatória para uma economia que experimenta uma taxa de inflação mais ou menos constante, embora seja muito menos satisfatória nos casos em que a taxa de inflação varia consideravelmente no tempo. Cabe pois esperar que o erro do índice seja proporcional à taxa de mudança dos preços e não uma proporção constante do nível absoluto deste índice. Seguindo estes fundamentos lógicos, a *Especificação III* será

$$\varepsilon'_t = \eta_t P_t$$

onde η_t é novamente uma série de números aleatórios que recebem limites apropriados tais como $-0,5 \leq \eta_t \leq +0,5$. A fim de manter erros da mesma ordem que os da *Especificação II* faz-se necessário normalmente ampliar os limites da especificação de erro (a magnitude absoluta do erro depende agora do comportamento do índice de preços, daí termos usado a palavra "normalmente").

A importância do viés

A força do impacto deste erro comum nas variáveis, depende da variância das variáveis reais em relação à variância dos erros. A forma dessa dependência é vista com maior facilidade no caso da *Especificação II*. Senão, vejamos,

$$\frac{X_t}{1 + \varepsilon_t} = \left(\frac{Y_t}{1 + \varepsilon_t} \right)^{\beta} \cdot u_t$$

onde foi omitido o termo constante, a fim de simplificar a notação uma vez que o mesmo não afeta o problema. Defina-se $\epsilon''_t = 1 + \epsilon'_t$ e tomemos logaritmos em ambos os lados, de modo que

$$X_t - \epsilon''_t = \hat{\beta} (Y_t - \epsilon''_t) + u'_t$$

Da definição de $\hat{\beta}$, obtém-se

$$\hat{\beta} = \frac{M_{xy} - M_{\epsilon''}y - M_x \epsilon'' + M_{\epsilon''} \epsilon''}{M_{yy} + M_{\epsilon''} \epsilon'' - 2M_y \epsilon''} \cong \frac{M_{xy} + M_{\epsilon''} \epsilon''}{M_{yy} + M_{\epsilon''} \epsilon''}$$

Os momentos omitidos tenderão a zero com o aumento do tamanho das amostras porquanto não há covariação entre os termos de erro e as variáveis reais. $\hat{\beta}$ apresenta um viés para a unidade e a força deste viés depende da relação entre os valores dos momentos. Quanto maiores M_{yy} e M_{xy} em relação a $M_{\epsilon''\epsilon''}$, menos importância terá o viés. Por outro lado a magnitude dos momentos das variáveis reais depende de dois fatores: primeiro, do tipo de relação entre os momentos e a variância das variáveis independentes, mas igualmente da maneira como se aborda a econometria. Considerando-se a linearidade e as variações sazonais como obstáculos à estimação eficiente dos dados usados nas investigações econométricas, as relações entre as variáveis serão pré-processadas — isto é, eliminadas a tendência e a sazonalidade — reduzindo assim os valores dos dois momentos reais, embora deixe imutável o valor do momento do erro, aumentando-lhe a importância relativa. Uma estimação que utilize dados brutos torna obscuro o problema de erro comum nas variáveis, sem entretanto solucioná-lo. Neste caso o erro fica mascarado pelos problemas adicionais causados pela existência de multicolinearidade.

As variáveis econômicas reais, especialmente quando eliminadas a tendência e a sazonalidade, provavelmente não variarão mais do que 1 ou 2% entre períodos, ou mais de 5 a 10% em relação à tendência em períodos mais longos. Admita-se então a possibilidade de erros de 2% ou mesmo de 1% no deflator entre quaisquer dois períodos. Dêste modo é possível apreciar a importância do viés em estimativas que utilizem dados mensais (e, também em menor extensão, no caso de dados trimestrais).

Simulação

Como exemplo da possível importância do problema, aplicou-se uma análise de regressão a uma série de números aleatórios sujeitos aos três tipos de especificação de erro. O método consiste em tomar duas séries de números aleatórios (que pela sua própria construção têm uma correlação nula), multiplicando por um deflator implícito “verdadeiro” e deflacionando-se as séries resultantes por um índice de preço “proxy” construído de acordo com uma das especificações de erro. Como “proxy” foi utilizado o índice geral de preços do Brasil (índice 2 da *Conjuntura Econômica*). Os resultados da regressão indicam a extensão do viés gerada pelos erros no deflator “proxy”.

Simulação usando a Especificação I

Foram gerados dois conjuntos de números aleatórios, com limites em torno de $1.000 \leq R_{it} \leq 5.000$ ($i = 1,2$; $t = 1,2, \dots, 228$). O índice geral de preços do Brasil toma o valor 1,3 no período 1 e o valor 208 no período 228, de modo que se multiplicando as duas séries de números aleatórios por uma variável $t = 1,2, \dots, 228$ (o deflator “real” de preços) e deflacionando-se a série resultante pelo índice geral de preços, obtém-se uma aproximação à *Especificação I*. Os resultados da regressão ² foram os seguintes,

$$\frac{tR_1}{P} = \text{constante} + 0,73 \frac{tR^2}{P} \quad \bar{R}^2 = 0,50$$

(15,3) d = 1,6

e, com a inclusão da variável de preço \dot{P} , obtém-se

$$\frac{tR_1}{P} = \text{constante} + 0,72 \frac{tR^2}{P} - 10650 \dot{P} \quad \bar{R}^2 = 0,51$$

(14,7) (1,2) d = 1,55

Observa-se que os coeficientes são impressionantes, considerando que R_1 e R_2 constituem dois conjuntos de números aleatórios, os quais, podem variar 200% em relação aos seus valores médios e

² Os números entre parênteses são estatísticas de t , e d é o coeficiente de Durbin-Watson.

possuem uma variação máxima entre períodos de 400%. Esta formulação do mecanismo de erro é interessante uma vez que admite autocorrelação nos erros gerados e padrões sazonais nos mesmos, note-se ainda que ambos os problemas estarão provavelmente presentes quando do uso de deflatores inadequados.

Simulação usando-se as Especificações II e III

Essas duas Especificações foram aplicadas a duas séries de números aleatórios para os quais permitiu-se inicialmente um intervalo da variação entre 1.000 e 5.000. Na *Especificação II* foram inicialmente usados erros ε''_t iguais a uma série de números aleatórios com os limites percentuais a $\pm 1\%$ conforme já mencionado. O resultado — um \bar{R}^2 praticamente nulo — não constitui surpresa em virtude da grande variância das variáveis reais em relação à variância do mecanismo de erro. A fim de avaliar a importância do viés causado por erros deste tipo, existem dois métodos possíveis: expandir os limites do mecanismo gerador de erro ou reduzir a variância das variáveis pseudo-reais. A expansão dos limites do gerador de erros produziu resultados significativos apenas quando $0,7 \leq \varepsilon''_t \leq 1,3$ ocasião em que se obtiveram as seguintes equações:

$$\text{Log } \frac{R_1}{\varepsilon''} = + 0,57 \text{ Log } \dot{E}\dot{P} \quad \bar{R}^2 = 0,08$$

(8,84) d = 2,17

$$\text{Log } \frac{R_1}{\varepsilon''} = + 0,55 \text{ Log } \dot{E}\dot{P} + 0,10 \text{ Log } \frac{R_2}{\varepsilon'}$$

(2,57) (1,19) $\bar{R}^2 = 0,08$

d = 2,17

$$\text{Log } \frac{R_1}{\varepsilon''} = 0,10 \frac{R_2}{\varepsilon''} \quad \bar{R}^2 = 0,03$$

(1,66) d = 2,18

Sendo $\dot{E}\dot{P}$ a mudança percentual no índice de preços com um mecanismo de erro correspondente a *Especificação II*. Conforme se pode ver, considerando que as variáveis reais variam amplamente no tempo, o viés causado por esta especificação provavelmente não será significativo para qualquer nível razoável de erro no índice "proxy".

Em Economia, as variáveis reais tendem a ser serialmente correlacionadas e não variam dentro dos amplos limites estabelecidos acima. Entretanto, tomando-se uma média móvel de três períodos de cada série (R_1 e R_2) consegue-se, até certo ponto, remover êstes problemas. Nas equações abaixo, \tilde{R}_1 e \tilde{R}_2 representam as séries que resultam das médias móveis R_1 e R_2 .

$$\text{Log } \frac{\tilde{R}_1}{\varepsilon''} = 0,25 \quad \text{Log } \frac{\tilde{R}_2}{\varepsilon''} + 0,41 \quad \text{Log } E\dot{P} \quad \bar{R}^2 = 0,20$$

$$(3,1) \quad d = 0,73$$

É evidente que a redução dos limites de êrro da "proxy" para níveis mais razoáveis e a restrição do intervalo de variação das variáveis pseudo-reais restabeleceriam a significância da *Especificação II*. Esta simulação, contudo, foi executada somente para a *Especificação III*.

Para η_t foi construída uma série de números aleatórios com os limites já mencionados ($\pm 0,5$), sendo o índice "proxy" construído através da multiplicação de η_t pelas taxas de variação do índice geral de preços no Brasil e adição do resultado a êste último. Foram utilizadas 100 observações do índice de preços, estendendo-se de janeiro de 1963 a abril de 1971 — período em que a taxa anual de inflação atingiu um máximo de quase 100% ao ano caindo, em seguida, para 20%. As duas séries pseudo-reais foram geradas com limites de 101.000 a 105.000. Multiplicando-se essas duas novas variáveis pseudo-reais, R_3 e R_4 , pelo índice geral de preços (que está sendo usado como se fôsse o deflator implícito) e dividindo o resultado pelo índice "proxy" construído segundo a *Especificação III*, obtêm-se os seguintes resultados:

$$\text{Log } \frac{R_3 \cdot P}{P + \eta \dot{P}} = + 0,87 \frac{R_4 \cdot P}{P + \eta \dot{P}} \quad \bar{R}^2 = 0,67$$

$$(13,97) \quad d = 2,05$$

$$\text{Log } \frac{R_3 \cdot P}{P + \eta \dot{P}} = + 0,87 \frac{R_4 \cdot P}{P + \eta \dot{P}} - 0,0003 E' P \quad \bar{R}^2 = 0,68$$

$$(13,92) \quad (1,1) \quad d = 2,07$$

Onde $E'P$ é a percentagem de mudança no índice "proxy" dos preços entre períodos. Dada a grande variância do termo de êrro relativa à variância das variáveis pseudo-reais, os resultados são, talvez,

esperados. Não obstante, este resultado torna bastante convincente o argumento segundo o qual o viés causado pelos erros no índice de preços usados como deflator constitui um elemento importante em estudos empíricos de economias inflacionárias utilizando dados mensais. Note-se que a variável preço não é significativa. Em princípio, parece que o erro em \hat{P} precisa ser excessivamente grande de modo a ocasionar uma covariância significativa entre o mesmo e a variável dependente. Argumenta-se abaixo que, pelo menos em um estudo empírico, esta condição foi atendida.

Exemplos de trabalho empírico afetado pelo problema de um erro comum nas variáveis

Qualquer estudo que utilize variáveis deflacionadas por um índice "proxy" está sujeito a este viés. Por outro lado, espera-se que, quanto mais alta a taxa de inflação (e, além disso, quanto mais ela flutua) maior será o erro e, como conseqüência, maior o viés. A definição da *Especificação I* contém ambos esses argumentos, a *Especificação III* produz erros que dependem da taxa de inflação, mas não de sua variabilidade, ao passo que a *Especificação II* produz erros independentes do comportamento do índice de preços — exceto na medida em que os limites de ε_t sejam estabelecidos como função do comportamento daquele índice. Parece plausível esperar que o problema tenha significação observável apenas em economias com uma taxa de inflação de cerca de 20% anualmente ou mais.

Dentre os estudos mais conhecidos onde o problema do erro comum afigura-se como um problema, provavelmente o mais importante foi o realizado por Philip Cagan.³ O estudo de Cagan consiste basicamente de uma regressão entre saldos monetários reais e uma variável que representa a taxa esperada de mudança de preços em várias economias que experimentaram hiperinflação antes e durante a II Guerra Mundial. As taxas de mudança de preços consideradas foram da ordem de 50% ao mês. A taxa esperada de mudança de preços é representada por uma soma exponencialmente

³ P. Cagan "The Monetary Dynamics of Hyperinflation" em *Studies in the Quantity Theory of Money* editado por M. Friedman (Chicago, University of Chicago Press, 1956).

ponderada da taxa de variação de preços mensais em períodos anteriores. Muito embora o promediamento das referidas taxas de inflação reduza a importância do viés no caso de erros gerados pelas *Especificações II* ou *III*, parece ainda provável que o viés seja considerado sério; no caso de erros serialmente correlacionados, um processo de promediamento não reduzirá de muito o viés.

Cagan caracteriza os dados de seu estudo como inexatos e pergunta:

“Se ... grande parte dos dados pode estar sujeita a grandes erros, porque os coeficientes seriam... tão altos? Dados medíocres tendem a aumentar os erros residuais de um ajustamento de mínimos quadrados. Os altos coeficientes de correlação sugerem que a maioria dos dados não estão sujeitos a grandes erros aleatórios.”⁴

O presente trabalho sugere que os grandes erros contidos no índice de preços melhoram o coeficiente de determinação, ao contrário do que afirma Cagan. Parece provável que existam grandes erros nos índices usados por Cagan como “proxies” para os deflatores.⁵

O deflator “proxy” para a Grécia consiste de um índice do custo de alimentação em Atenas; o índice polonês, representa os preços a varejo de alimentos; enquanto o russo, engloba os preços a varejo em todo o país. A relação entre a variável dependente “deflacionada” e a variável preço incluindo um erro, constitui a parte menos convincente das simulações acima; é provável entretanto, que as variâncias dos termos de erro nas regressões de Cagan sejam grandes em relação às variâncias dos saldos monetários reais e da taxa da inflação. Pelo menos, o valor dos coeficientes pode ser colocado em dúvida.

Os monetaristas encontram na América Latina um promissor campo para seus estudos econométricos. No Brasil, Afonso Celso Pastore tem utilizado a análise da regressão para investigar possíveis relações de renda monetária. No trabalho “Alguns Aspectos da Inflação Brasileira”,⁶ aquele autor utiliza o índice geral de preços do Brasil

⁴ Cagan, *op. cit.*, p. 47.

⁵ *Ibid.*, Apêndice B, pp. 97-117.

⁶ *Revista Brasileira de Economia* (janeiro-março, 1969).

como deflator para dados mensais nominais. Seguindo a mesma orientação, fizemos alguns trabalhos sobre dados brasileiros mensais e ficamos crescentemente impressionados pelo paradoxo mencionado por Cagan — quando a qualidade dos dados mensais é tão pobre, por que devem os resultados econométricos serem tão esmagadoramente bons? Os dados são reconhecidamente pobres, por exemplo, foi usada uma “proxy” mensal da renda e os coeficientes de determinação mostraram-se suspeitosamente altos, mesmo levando-se em conta problemas tradicionais como a multicolineariedade. Até certo ponto, o viés causado por um erro comum nas variáveis explica o paradoxo.

É possível simular especificações de erros em modelos nos quais as séries pseudo-reais sejam, serialmente correlacionadas e contenham componentes de tendência sazonal e o deflacionamento por um índice de preços com um mecanismo de erro mais complexo do que as *Especificações II e III* levaria a coeficientes de determinação extremamente altos e coeficientes de regressão fortemente significativos. O exercício não seria de todo inútil caso conseguisse despertar dúvidas em alguns entusiastas de estudos empíricos.