

Avaliação de programas de formação profissional*

MÉTODOS E
ORIENTAÇÕES GERAIS

António F. M. G. Saraiva**

Dezembro de 1993

*Trabalho originariamente concebido como elemento de avaliação no âmbito da disciplina de **Economia do Trabalho e Gestão de Recursos Humanos** do **Curso de Mestrado**, orientada pelo Dr. Pedro Portugal, a quem muito agradeço a ajuda na revisão final do texto.

**Assistente do Instituto Superior de Contabilidade e Administração do Porto

PROPÓSITO

A produção deste texto subordina-se ao intuito de, sistematizada e compreensivamente, expor as principais soluções propostas para, com base em dados não-experimentais, determinar em que medida os programas de formação profissional produzem os efeitos para que foram criados, designadamente, o crescimento dos salários dos formandos induzido pelo acréscimo de produtividade potenciado pela formação.

Dadas as suas características, este texto poderá servir como uma elementar referência para quem pretenda fazer uma primeira abordagem à temática da avaliação de programas de formação profissional. O seu interesse deverá mesmo ser avaliado pela medida em que, sem o defraudar, conseguir convencer o leitor da insuficiência da sua leitura, e da necessidade do seu abandono em favor das referências bibliográficas para que remete.

INTRODUÇÃO

Nos E.U.A., o interesse por este assunto foi suscitado pela necessidade de avaliar os programas sociais¹ ali implementados a partir da década de sessenta e destinados a ajudar os estratos populacionais de baixos rendimentos.

Os primeiros estudos devem-se a Goldberger [1972] e Cain [1975] e, ainda, a Ashenfelter [1978]. Mas o avanço fundamental na abordagem desta temática deu-se aquando o desenvolvimento de métodos capazes de dar conta do enviesamento por selectividade (selectivity bias). Estes métodos propostos, entre outros, por James Heckman, foram pela primeira vez aplicados na avaliação de programas por Barnow, Cain e Goldberger [1980]. Posteriormente, os trabalhos de Heckman² (de parceria com outros autores) vêm a firmar-se como referências importantes para a concepção de modelos destinados a avaliar os programas de formação profissional com base em dados não-experimentais.³ Implícita em cada um destes modelos, está a ideia de que a formação profissional, enquanto investimento em capital humano, proporcionaria aumentos de produtividade os quais, por sua vez, explicariam o crescimento dos rendimentos.

Este encadeamento de efeitos é preconizado pela teoria do capital humano pelo que, de algum modo, pode pretender-se que, ao avaliar, nestes termos, os programas de formação

¹ Estes programas sucederam-se à cadência de um por década: Manpower development and training act (MDTA), na década de sessenta; Comprehensive employment and training act (CETA), na década de setenta; Job partnership and training act (JPTA), na década de oitenta.

² J. Heckman e R. Robb [1985] e J. Heckman e V. Hotz [1989].

³ Assim, não pode espantar que este texto revele, (porventura em demasia), traços característicos da abordagem feita por este autor.

profissional, se está, concomitantemente, a avaliar a relevância daquela teoria. Esta tem mesmo sido encarada como a forma possível de comprovar empiricamente os resultados da teoria do capital humano, dada a dificuldade em recolher e tratar dados relativos à evolução da produtividade.⁴

Uma forma de contornar esta dificuldade consiste em usar os dados relativos a uma única empresa. É que, em princípio, é possível exercer um controlo mais apurado sobre estes dados pois será grande a probabilidade de terem sido recolhidos em obediência a critérios consistentemente aplicados de período para período e de indivíduo para indivíduo. E se, para além disto, existirem registos sobre a performance dos trabalhadores, é virtualmente possível investigar a relação entre a evolução da produtividade e a evolução dos rendimentos. Pode, então, aproveitar-se o ensejo para, mais rigorosamente, testar a teoria do capital humano.⁵

No contexto desta teoria, distinguem-se, basicamente, dois tipos de formação profissional:

- "on-the-job training", i.e., a formação profissional proporcionada pelas empresas aos seus trabalhadores;
- "off-the-job training", i.e., a formação profissional ministrada por entidades, públicas ou privadas, para tal expressamente vocacionadas.⁶

⁴ Não se deixa, contudo, de ficar desarmado face ao argumento de que outros factores, que não a evolução da produtividade, possam explicar o perfil de rendimentos empiricamente observado.

⁵ Ver Ann P. Bartel [1992].

⁶ A distinção entre "on-the-job training" e "off-the-job training" é especialmente relevante quando o que está em causa é não apenas avaliar o impacto da formação sobre os rendimentos, mas, por exemplo, explicar o fenómeno da mobilidade no emprego. A este propósito, ver, por exemplo, Lisa Lynch [1992].

A "on-the-job training" apresenta-se como uma forma de investimento específico em capital humano, cujos custos e benefícios serão partilhados por empregadores e empregados. Os primeiros começam por pagar salários superiores ao valor do produto marginal mas, em compensação, passam a dispor de trabalhadores mais habilitados ao desempenho das suas tarefas, pelo que pretenderão mantê-los ao seu serviço; os últimos, porque terão incorrido num custo de oportunidade equivalente à diferença entre o salário que efectivamente recebem e o salário que, na melhor alternativa, receberiam, têm interesse em permanecer na empresa até conseguirem recuperar o seu investimento graças ao crescimento dos salários potenciado pela formação.

A "off-the-job training" apresenta-se como uma forma de investimento genérico em capital humano, cujos custos⁷ serão suportados pelos trabalhadores interessados em obter ou melhorar as suas qualificações profissionais e, assim, aumentar a probabilidade de verem crescer os seus rendimentos.

A influência da formação (específica ou genérica) sobre os rendimentos será traduzida através duma expressão onde estejam presentes, como variáveis explicativas, entre outras, o número de anos ao serviço da empresa actual e o número de anos de experiência profissional. A primeira dará conta do efeito da

⁷ Genericamente, este custo será dado pela expressão: salário na melhor alternativa + subsídio concedido aquando a formação - despesa com a formação - salário efectivo.

formação específica, enquanto que a segunda dará conta do efeito da formação genérica.⁸

Mas, atendendo a que o que está em causa é a avaliação de um determinado programa de formação profissional, impõe-se redefinir esta expressão de molde a possibilitar discernir o impacto sobre os rendimentos daquele programa em particular. Isto, como se verá, conseguir-se-á fundamentalmente através da introdução de uma variável dummy representativa da situação de cada indivíduo relativamente ao programa.

ESTIMAÇÃO DO IMPACTO DOS PROGRAMAS DE FORMAÇÃO PROFISSIONAL SOBRE OS RENDIMENTOS

O problema em questão, o qual consiste em saber como avaliar o impacto sobre os rendimentos (via produtividade) dos programas de formação profissional, pode ser encarado como um caso particular do problema mais geral que consiste em avaliar o impacto, sobre certos resultados, de programas de natureza diversa. No entanto, dado o âmbito definido para este trabalho, tomar-se-á o rendimento como o resultado que importa analisar.

⁸ Contrariando a visão convencional que atribuía à antiguidade uma importância significativa na explicação do crescimento dos rendimentos, Katherine Abraham e Henry Farber [1987] chamam a atenção para o facto de que especificações omissas quanto à influência sistemática de aspectos como a qualidade do trabalhador, do emprego e da entrosagem trabalhador/empregador (job-matching) sobre o crescimento dos rendimentos, implicarem estimadores de mínimos quadrados positivamente enviesados. Uma vez explicitamente considerados aqueles aspectos, conclui-se não ser, afinal, muito significativa a influência da antiguidade sobre o crescimento dos rendimentos, nem sequer muito perceptível o sentido do nexo de causalidade que os liga. Nesta perspectiva, o recurso à antiguidade para descortinar o efeito da formação específica só se explica por uma necessidade de simplificação, como afirmam Solomon Polachek e Stanley Siebert [1993].

Assim, adoptar-se-ão, de raiz, as seguintes definições:

$d_i = 0 \equiv$ o indivíduo i não recebeu formação.

$d_i = 1 \equiv$ o indivíduo i recebeu formação.

$y_{it} \equiv$ rendimento efectivo do indivíduo i , no período t .

$y^*_{it} \equiv$ rendimento, no período t , do indivíduo i :

-que não recebeu formação no período k ($d_i=0$)

-**se** não tivesse recebido formação no período k ($d_i=1$)

$\alpha_{it} \equiv$ impacto da formação sobre o rendimento do indivíduo i
relativo ao período t .

$$\alpha_{it} = y_{it} - y^*_{it} \Big|_{d_i=1} \quad ^9$$

Por simplificação, considere-se $\alpha_{it} = \alpha_t$, $\forall i$, o que equivale a aceitar i como o indivíduo representativo e permite escrever $\bar{\alpha}_t = E(\alpha_{it}) = \alpha_t$.

Consequentemente, a expressão

$$\bar{\alpha}_t = E(\alpha_{it} | d_i=1) = E(y_{it} - y^*_{it} | d_i=1) = E(y_{it} | d_i=1) - E(y^*_{it} | d_i=1)$$

apresenta-se como a "ideal" para traduzir o impacto do programa sobre o rendimento do indivíduo i , no período t , α_t .

A obtenção de estimativas fica, porém, comprometida pelo facto de, como já foi referido, o "rendimento presumido", y^*_{it} , não ser observável e, portanto, não se poder estimar o seu valor

⁹ Obviamente que, se o "rendimento presumido", y^*_{it} , daqueles que participaram efectivamente no programa fosse observável, o impacto do programa sobre o rendimento do indivíduo i , no período t , seria perfeitamente quantificado por esta diferença.

esperado. Oferecem-se, basicamente, duas maneiras de contornar esta dificuldade.

Uma passa pela constituição de um grupo de controlo constituído por indivíduos que, embora tendo sido seleccionados para participar no programa, acabaram por ser, aleatoriamente, excluídos dessa participação enquanto formandos. São os rendimentos destes indivíduos, no período t , que irão ser considerados relevantes para estimar $E(y^*_{it}|d_i=1)$ e, assim, tornar possível a avaliação do impacto do programa, no período t , α_t , — trata-se de um **método de avaliação experimental**.

A outra consiste em recorrer a valores observados de y_{it} , relativos a indivíduos que não tenham participado no programa, para estimar o impacto em causa, agora, congruentemente, traduzido pela expressão: $\tilde{\alpha}_t = E(y_{it}|d_i = 1) - E(y^*_{it}|d_i = 0)$.

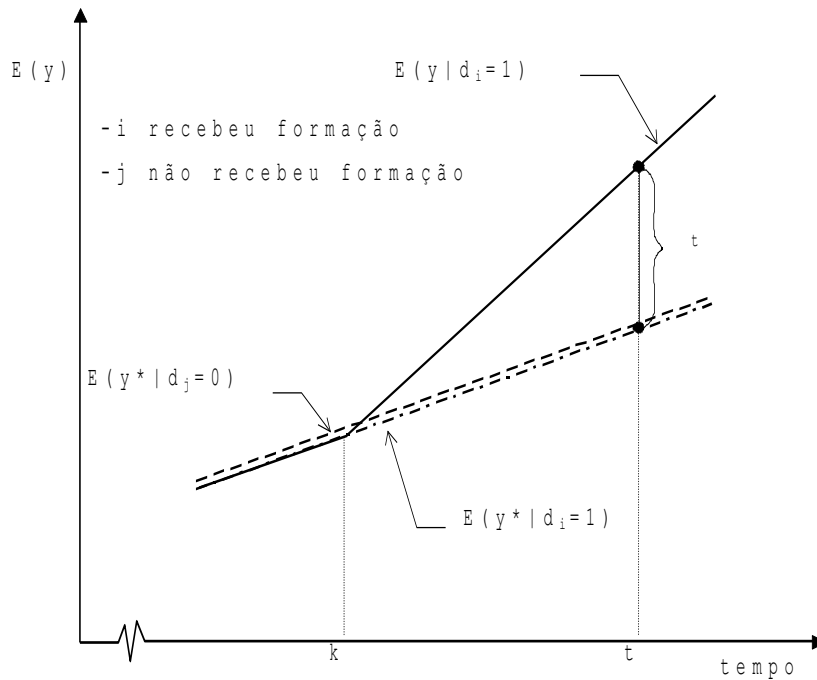


Figura 1

Na figura 1, pretende dar-se uma visão gráfica, necessariamente esquemática, da especificação agora considerada.

Neste caso, o **método de avaliação** diz-se **não-experimental** pois não requer a constituição de um grupo de controlo expressamente criado, mas apenas a disponibilidade de dados relativos a um grupo de comparação¹⁰ composto por indivíduos supostamente¹¹ não envolvidos no programa.

No entanto, as estimativas de $\tilde{\alpha}_t$ apenas serão tão fiáveis, para avaliar o impacto da formação, quanto o seriam as estimativas de $\bar{\alpha}_t$, se se verificar $E(y^*_{it}|d_i=1)=E(y^*_{it}|d_i=0)$ ¹², isto é, se o valor esperado do rendimento que presumivelmente um indivíduo obteria, no período t, caso não tivesse participado no programa, seja igual ao valor esperado do rendimento que, nesse mesmo período, será obtido por um indivíduo que não tenha, efectivamente, participado no programa.

Fala-se em enviesamento por selectividade quando $E(y^*_{it}|d_i=1) \neq E(y^*_{it}|d_i=0)$ pois, na origem desta desigualdade, estará a forma como os indivíduos são seleccionados, ou se auto-seleccionam, para participar no programa.

¹⁰ Os grupos de comparação costumam ser formados por:
 - indivíduos que pretenderam frequentar o programa e foram, à partida, rejeitados com base em certos critérios e não aleatoriamente;
 - indivíduos que não se candidataram;
 - indivíduos que integram amostras constituídas por certas instituições.

¹¹ Desde que não se ponha o problema da contaminação da amostra referente ao grupo de comparação por observações relativas a indivíduos que tenham recebido formação.

¹²

~

$$\tilde{\alpha}_t = \bar{\alpha}_t$$

$$E(y_{it}|d_i = 1) - E(y^*_{it}|d_i = 0) = E(y_{it}|d_i = 1) - E(y^*_{it}|d_i = 1)$$

$$E(y^*_{it}|d_i = 1) = E(y^*_{it}|d_i = 0)$$

É óbvio que este enviesamento torna incorrecto o recurso à formulação $\tilde{\alpha}_t$ para estimar o impacto do programa sobre os rendimentos do período t , α_t .

Subjacente a uma metodologia de avaliação não-experimental está uma determinada especificação¹³ de uma função de rendimentos.

Genericamente, ter-se-á $y_{it} = y^*_{it} + d_i\alpha_{it}$ ou atendendo à simplificação já introduzida, $y_{it} = y^*_{it} + d_i\alpha_t$.

Esta função rendimentos é genérica no sentido de que mede o rendimento de um certo período t quer o indivíduo i tenha, ou não, participado no programa de formação (ou quer t identifique um período posterior, ou anterior, ao período em que decorreu a formação).

Se não participou, $d_i=0$ e portanto $y_{it} = y^*_{it}$, o que é congruente com a definição de y_{it} e y^*_{it} ; se participou, $d_i=1$, pelo que $y_{it} = y^*_{it} + \alpha_t$.

Mas, se, na primeira hipótese, se desemboca numa trivial igualdade tautológica, na segunda é-se conduzido a uma relação entre os níveis de rendimento efectivo e "presumido" que apenas se tornará operacional se admitirmos que este último depende linearmente de um conjunto de características¹⁴ observáveis

¹³ Em geral, trata-se de uma especificação linear.

¹⁴ Estas características podem ser traduzidas por variáveis como sejam a idade, os anos de experiência profissional, o grau académico, os rendimentos obtidos em períodos anteriores, o sexo, ...

(X) e não observáveis (U), i.e., se admitirmos¹⁵

$$y^*_{it} = \mathbf{x}_{it}\beta + U_{it} \quad \forall i,$$

com $\mathbf{x}_{it} = [X_{1it} \ X_{2it} \dots X_{cit}] = [X_{jit}]$ e $\beta = [\beta_j] = [\beta_1 \ \beta_2 \dots \beta_c]'$
 $t=1, \dots, T$, $j=1, \dots, c$ e $c=n^\circ$ de variáveis que traduzem a influência de certas características sobre os rendimentos obtidos.

Pode, então, dar-se o seguinte aspecto à função rendimentos, já na forma estocástica potencialmente estimável¹⁶:

$$y_{it} = \mathbf{x}_{it}\beta + d_i\alpha_t + U_{it} \quad \{1\} \quad \forall i.$$

Inerente a esta especificação está a hipótese de que, abstraindo do impacto da formação, os rendimentos de formados e não formados são explicados pelas mesmas c variáveis que integram o vector \mathbf{x}_i .

No entanto, para que uma estimativa fiável de α_t possa ser obtida (nomeadamente por recurso a um método de mínimos quadrados) têm que verificar-se as hipóteses que garantem a centralidade dos estimadores a usar.

Designadamente, deverá acontecer $E(U_{it}|\mathbf{x}_i, d_i) = 0$ ou, equivalentemente, $E(y_{it}|\mathbf{x}_i, d_i) = \mathbf{x}_{it}\beta + d_i\alpha_t$ garantindo-se,

¹⁵ Este termo de perturbação, U_{it} , pode apresentar-se como um agregado de três componentes:

- um termo de perturbação específico do indivíduo i o qual pode traduzir não só um efeito fixo sobre o rendimento, mas também uma taxa de crescimento específica desse mesmo rendimento (tal será adiante considerado);
- um termo de perturbação específico do período t ;
- um termo de perturbação residual.

¹⁶ Para que esta equação reproduza o perfil (côncavo) de rendimentos habitualmente observado (rendimentos crescentes ao longo da vida activa do trabalhador, mas a taxas decrescentes) é, muitas vezes, conveniente dar-lhe uma forma polinomial quanto à idade e/ou autoregressiva quanto ao rendimento.

assim, que nenhuma variável explicativa do rendimento foi omitida na especificação {1}. É que, se o tivesse sido, estaria comprometida a centralidade dos estimadores de β e α_t na medida em que os valores assumidos por essa variável omitida estivessem correlacionados com os valores amostrais dos regressores X_{ji} ou d_i .

Tendo-se chegado à conclusão de que é virtualmente possível avaliar o impacto médio da formação, α_t , estimando a equação {1}, importa, agora, perceber como se traduz o enviesamento por selectividade quando uma especificação deste tipo é adoptada.

Nesse sentido, é conveniente ter presente que

$$E(y_{it} | \mathbf{x}_i, d_i=1) = \mathbf{x}_{it}\beta + \alpha_t + E(U_{it} | \mathbf{x}_i, d_i=1) \text{ e}$$

$$E(y_{it} | \mathbf{x}_i, d_i=0) = \mathbf{x}_{it}\beta + E(U_{it} | \mathbf{x}_i, d_i=0).$$

Refira-se, ainda, que se se aceita a hipótese de que na ausência de formação, as variáveis observáveis (X_{ji}) condicionam de igual modo e cabalmente o rendimento de todo e qualquer indivíduo independentemente de ter, ou não, recebido formação, as expressões acima podem ser rescritas de uma forma mais simplificada:

$$E(y_{it} | d_i=1) = \mathbf{x}_{it}\beta + \alpha_t + E(U_{it} | d_i=1) \text{ e}$$

$$E(y_{it} | d_i=0) = \mathbf{x}_{it}\beta + E(U_{it} | d_i=0).$$

Recordando que $\tilde{\alpha}_t = E(y_{it} | d_i = 1) - E(y_{it}^* | d_i = 0)$ vem
 $\tilde{\alpha}_t = \alpha_t + E(U_{it} | d_i = 1) - E(U_{it} | d_i = 0).$

Neste contexto, a igualdade entre $\tilde{\alpha}_t$ e $\bar{\alpha}_t (= \alpha_t)$ requer que se tenha $E(U_{it}|d_i = 1) = E(U_{it}|d_i = 0)$, enquanto que, anteriormente (ver nota 12 na pg. 9), se tinha concluído que aquela igualdade exigia que se verificasse $E(y^*_{it}|d_i = 1) = E(y^*_{it}|d_i = 0)$. Torna-se, assim evidente, desta perspectiva, a equivalência entre estas duas proposições.

Conseqüentemente, se anteriormente se traduziu o enviesamento por selectividade pela desigualdade $E(y^*_{it}|d_i = 1) \neq E(y^*_{it}|d_i = 0)$, pode-se, agora, fazê-lo através desta outra desigualdade $E(U_{it}|d_i = 1) \neq E(U_{it}|d_i = 0)$, a qual significa a existência de uma relação de dependência entre d_i e U_{it} .

Quer isto dizer que, quando se trata com a função rendimentos, o enviesamento por selectividade manifesta-se pela dependência entre d_i e U_{it} .

Compreende-se, assim, que este tipo de enviesamento é pernicioso porque inviabiliza a estimação do modelo na medida em que, se acontece $E(U_{it}|d_i = 1) \neq E(U_{it}|d_i = 0)$, não se pode assegurar que se verifique $E(U_{it}|d_i) = 0$, o que seria necessário para que se obtivessem estimadores cêntricos de β e de α_t .

A dependência entre d_i e U_{it} radica na forma como se processa a selecção (ou auto-selecção) dos candidatos ao programa de formação.

Independentemente dos contornos específicos do processo de selecção, é, em princípio, possível traduzi-lo através de uma função caracterizadora da(s) regra(s) de selecção adoptada(s).

Tomando uma função índice para esse efeito, virá $IN_i = \mathbf{z}_i \boldsymbol{\gamma} + V_i$ sendo que se passará a ter

$$d_i = 1 \Leftrightarrow IN_i > 0 \quad \text{≡ o indivíduo } i \text{ é seleccionado;}$$

$$d_i = 0 \Leftrightarrow IN_i \leq 0 \quad \text{≡ o indivíduo } i \text{ não é seleccionado.}^{17}$$

As variáveis observáveis Z_i que integram o vector \mathbf{z}_i , bem como as variáveis não observáveis, V_i , veiculam a influência dos diversos factores condicionantes da selecção.¹⁸

Considerando a função distribuição de V_i — $F(v_i) = P(V_i < v_i)$ — e supondo que V_i se distribui independentemente de \mathbf{z}_i , tem-se:

$$\begin{aligned} P(d_i = 1 | \mathbf{z}_i) &= E(d_i | \mathbf{z}_i) = 1 - F(-\mathbf{z}_i \boldsymbol{\gamma}) \\ &= 1 - P(V_i < -\mathbf{z}_i \boldsymbol{\gamma}) \\ &= 1 - P(\mathbf{z}_i \boldsymbol{\gamma} + V_i < 0) \\ &= 1 - P(IN_i < 0) \end{aligned}$$

¹⁷ Aceitando como válida a regra de selecção a posteriori, deve ter-se em conta que $d_i = 1$ não significa, obrigatoriamente, que o indivíduo i tenha, efectivamente, recebido formação.

¹⁸ Pode acontecer que \mathbf{z}_i inclua variáveis X_i . Muitas vezes, justifica-se que o rendimento do indivíduo i , no período da selecção (ou em períodos anteriores) integre o vector \mathbf{z}_i .

Estando, deste modo, representada a regra de selecção, é pertinente admitir que a dependência entre U_{it} e d_i pode resultar:

A — da dependência entre \mathbf{z}_i e U_{it} e/ou

B — da dependência entre V_i e U_{it} .

No primeiro caso, fala-se de selecção condicionada por variáveis observáveis; no segundo, de selecção condicionada por variáveis não-observáveis.

A distinção é relevante pois, consoante se trate de um caso ou de outro, diferente metodologia deve ser seguida para "neutralizar" o enviesamento por selectividade.

Sabe-se já que há enviesamento por selectividade quando

$E(U_{it}|\mathbf{x}_i, d_i) \neq 0$ (não porque não se verifique $E(U_{it}|\mathbf{x}_i) = 0$, mas porque $E(U_{it}|d_i) \neq 0$, ou seja, porque U_{it} e d_i estão correlacionados).

A. Quando a selecção é condicionada por variáveis observáveis tem-se, evidentemente, $E(U_{it}|\mathbf{x}_i, \mathbf{z}_i, d_i) \neq 0$. mas também $E(U_{it}|\mathbf{x}_i, \mathbf{z}_i, d_i) = E(U_{it}|\mathbf{x}_i, \mathbf{z}_i) [\neq 0]$.

Esta igualdade significa que o facto de o valor esperado condicional de U_{it} diferir de zero é, em última instância e cabalmente, explicado pela influência sobre o processo de selecção das variáveis observáveis reunidas em \mathbf{z}_i .

Por conseguinte, a consideração destas variáveis é suficiente para dar conta do enviesamento por selectividade, ou seja, para desfazer a dependência entre U_{it} e d_i .

Começando por notar que $E(y_{it}|\mathbf{x}_i, \mathbf{z}_i, d_i) =$
 $= \mathbf{x}_{it}\beta + d_i\alpha_t + E(U_{it}|\mathbf{x}_i, \mathbf{z}_i, d_i) = \mathbf{x}_{it}\beta + d_i\alpha_t + E(U_{it}|\mathbf{x}_i, \mathbf{z}_i)$, percebe-se que o que está em causa é estabelecer a forma funcional de $E(U_{it}|\mathbf{x}_i, \mathbf{z}_i)$.¹⁹

Considerando estas funções como funções de controlo a integrar na equação de regressão {1}, torna-se possível obter estimativas consistentes do impacto da formação sobre o rendimento desde que, evidentemente, aquelas funções de controlo sejam adequadamente definidas.²⁰

A equação {1}, apresentada na página 11, transfigurar-se-á em²¹

$$y_{it} = \mathbf{c}_{it}\delta + d_i\alpha_t + \tilde{U}_{it} \quad \text{[A]} \quad \text{com} \quad \mathbf{c}_{it} = \mathbf{x}_{it} \cup \mathbf{z}_{it}$$

$$\text{e} \quad \tilde{U}_{it} = U_{it} - E(U_{it}|\mathbf{c}_{it}, d_i) = U_{it} - E(U_{it}|\mathbf{c}_{it})$$

$$\text{e, portanto, } E(\tilde{U}_{it}|\mathbf{c}_{it}) = E(U_{it}|\mathbf{c}_{it}) - E(U_{it}|\mathbf{c}_{it}) = 0.$$

É, no entanto, possível criar duas variantes consoante se considere $E(\alpha_{it}|\mathbf{c}_{it}, d_i=1) = \alpha_t$, ou $E(\alpha_{it}|d_i=1) = \mathbf{c}_{it}\theta_t$.

¹⁹ Por simplificação, admite-se a sua linearidade.

²⁰ As estimativas assim obtidas resultam, portanto, de um processo bietápico.

²¹ Refira-se que a inclusão das variáveis Z_i como regressores nesta equação, se bem que não afecte a centralidade dos estimadores de δ e α_t , diminui-lhes, em alguma medida, a eficiência. Também acontece que aumenta a probabilidade de se estar a usar amostras afectadas por multicolinearidade, o que se repercutirá negativamente na relevância dos testes de inferência estatística que, eventualmente, se pretendam fazer.

Aceitar $E(\alpha_{it} | \mathbf{C}_{it}, d_i=1) = \alpha_t$ [variante 1] corresponde a admitir que o impacto da formação sobre o rendimento do indivíduo i , no período t , é independente do conjunto das suas características específicas (o indivíduo é tomado como representativo).

Aceitar $E(\alpha_{it} | d_i=1) = \mathbf{C}_{it} \boldsymbol{\theta}_t$ [variante 2]²² é reconhecer que o impacto da formação sobre o rendimento de certo indivíduo, no período t , é função dos factores específicos condicionantes directos do rendimento ou do processo de selecção.

Adoptar a primeira variante implica trabalhar com a equação

$$y_{it} = \mathbf{C}_{it} \boldsymbol{\delta} + d_i \alpha_t + \tilde{U}_{it} \quad [\text{A.1}],$$

enquanto que a opção pela segunda variante obriga a considerar a equação

$$y_{it} = \mathbf{C}_{it} \boldsymbol{\delta} + d_i \mathbf{C}_{it} \boldsymbol{\theta}_t + \tilde{U}_{it} \quad [\text{A.2}],$$

na estimação dos vectores de parâmetros $\boldsymbol{\delta}$ e $\boldsymbol{\theta}_t$.

Neste último caso, obtém-se uma estimativa do impacto da formação sobre o rendimento de um determinado indivíduo, num certo período t , calculando o produto $\mathbf{C}_{it} \hat{\boldsymbol{\theta}}_t$, onde $\hat{\boldsymbol{\theta}}_t$ é o vector de estimativas de $\boldsymbol{\theta}_t$.

B. Quando acontece $0 \neq E(U_{it} | \mathbf{X}_i, \mathbf{Z}_i, d_i) \neq E(U_{it} | \mathbf{X}_i, \mathbf{Z}_i) \neq 0$, não se pode imputar o enviesamento apenas à influência sobre o processo de selecção das variáveis observáveis presentes em \mathbf{Z}_i . Serão, pois, as variáveis não observáveis, cuja influência sobre o processo de selecção é veiculada por V_i , aquelas que permitirão explicar o enviesamento em questão.

²² Esta variante significa o abandono da simplificação, anteriormente introduzida, segundo a qual $\alpha_{it} = \alpha_t, \forall i$.

É razoável admitir-se que V_i e U_i estejam negativamente correlacionados: para um valor elevado de V_i é maior a probabilidade de se ter $d_i=1$; por outro lado, para um valor reduzido de U_i verificar-se-á que a estimativa do rendimento pré-formação daqueles que virão a receber formação é inferior à estimativa do rendimento daqueles que a não irão receber, mesmo que tenham sido consideradas na equação de regressão as variáveis Z_i .

Desde logo dois tipos de estimadores serão, em princípio, capazes de, nestas circunstâncias, permitir a obtenção de estimativas relevantes.

São eles: B.1. — o estimador de efeitos fixos²³ e

B.2. — o estimador de crescimento aleatório²⁴

B.1. Admita-se que $U_{it}=\phi_{1i}+v_{it}$ onde ϕ_{1i} consiste num termo de perturbação específico do indivíduo i ($E(\phi_{1i})=0$) devido ao qual, supostamente, se estabelece a dependência entre U_{it} e d_i . Relativamente a v_{it} , é necessário admitir a verificação das hipóteses habituais: $E(v_{it}|\phi_{1i})=0$ e $E(v_{it}v_{it'})=0 \quad \forall t, t' (t \neq t')$ (os termos aleatórios residuais v_{it} nem estão autocorrelacionados, nem dependem do termo de perturbação específico de i , ϕ_{1i}).

²³ Também designado estimador "primeiras diferenças".

²⁴ Também designado estimador "diferença das diferenças".

Nestas condições, verifica-se que

$$\begin{aligned}
 E(U_{it} - U_{it'} | \mathbf{X}_i, d_i) &= E(\phi_{1i} + v_{it} - \phi_{1i} + v_{it'} | \mathbf{X}_i, d_i) \\
 &= E(v_{it} - v_{it'} | \mathbf{X}_i, d_i) \\
 &= E(v_{it} | \mathbf{X}_i, d_i) - E(v_{it'} | \mathbf{X}_i, d_i) \\
 &= 0 - 0 \\
 &= 0
 \end{aligned}$$

Transformando, adequadamente, a equação de regressão é, então, possível garantir a centricidade dos estimadores inerentes à especificação:

$$y_{it} = \mathbf{x}_{it}\beta + d_i\alpha_t + U_{it}$$

$$y_{it'} = \mathbf{x}_{it'}\beta + U_{it'} \quad \text{pois } t' < k$$

$$\begin{aligned}
 y_{it} - y_{it'} &= (\mathbf{x}_{it} - \mathbf{x}_{it'})\beta + d_i\alpha_t + (U_{it} - U_{it'}) \\
 &= (\mathbf{x}_{it} - \mathbf{x}_{it'})\beta + d_i\alpha_t + (v_{it} - v_{it'})
 \end{aligned}$$

Desta equação²⁵ derivam estimadores cêntricos pois, por hipótese, $E(v_{it} - v_{it'} | \mathbf{X}_i, d_i) = 0$, como já se viu.

Distinguindo as mesmas duas variantes já consideradas vem:

[variante 1]

$$y_{it} - y_{it'} = (\mathbf{x}_{it} - \mathbf{x}_{it'})\beta + d_i\alpha_t + (v_{it} - v_{it'}) \quad \text{[B.1.1]}$$

[variante 2]

$$y_{it} - y_{it'} = (\mathbf{x}_{it} - \mathbf{x}_{it'})\beta + d_i\mathbf{c}_{it}\theta + (v_{it} - v_{it'}) \quad \text{[B.1.2]}$$

$$t' < k < t$$

²⁵ Repare-se que, desta forma, as perturbações individuais (ϕ_{1i}), assim como as variáveis X_i que assumem valores inalterados ao longo do tempo, foram removidas da equação de regressão, ou seja, não é necessário considerar qualquer das variáveis cujo nível permanece fixo de período para período e que se sabe condicionarem os rendimentos.

Mas, porque é que, tendo-se $U_{it} = \phi_{1i} + v_{it}$, se verifica enviesamento por selectividade ?

Se o indivíduo i participou no programa e o indivíduo j não participou, tem-se:

$$E(y_{it} | d_i=1) = E(\mathbf{x}_{it}\beta + d_i\alpha_{it} + U_{it} | d_i=1) = \mathbf{x}_{it}\beta + \alpha_{it} + \phi_{1i} \quad e$$

$$E(y_{jt} | d_j=0) = E(\mathbf{x}_{jt}\beta + d_j\alpha_{jt} + U_{jt} | d_j=0) = \mathbf{x}_{jt}\beta + \phi_{1j}$$

Portanto, mesmo que i e j tenham as mesmas características ($\mathbf{x}_{it}\beta = \mathbf{x}_{jt}\beta$), a diferença entre $E(y_{it} | d_i=1)$ e $E(y_{jt} | d_j=0)$, dada por $\alpha_{it} + \phi_{1i} - \phi_{1j}$, não permitirá obter uma estimativa fiável do impacto da formação, α_{it} , na medida em que ϕ_{1i} diferir de ϕ_{1j} .²⁶

Significa isto que o enviesamento por selectividade se traduz, desta vez, pela não verificação da igualdade

$$E(y^*_{it} - y^*_{it'} | d_i=1) = E(y^*_{it} - y^*_{it'} | d_i=0).$$

Torna-se, assim, evidente que especificar o impacto médio da formação como $\tilde{\alpha}_t$ se apresenta, agora, desapropriado. Nestas condições, aquele impacto deverá ser definido pela expressão

$$\tilde{\alpha}_t = E(y_{it} - y^*_{it'} | d_i = 1) - E(y^*_{it} - y^*_{it'} | d_i = 0) \quad \text{pois}$$

$$\tilde{\alpha}_t = E(\mathbf{x}_{it}\beta + \alpha_{it} + \phi_{1i} - \mathbf{x}_{it'}\beta - \phi_{1i}) - E(\mathbf{x}_{it}\beta + \phi_{1i} - \mathbf{x}_{it'}\beta - \phi_{1i}) = E(\alpha_{it}) = \alpha_{it} \quad t' < k < t$$

(isto porque $E(v_{it}) = 0 \quad \forall t$).

É claro que esta é apenas uma forma alternativa de caracterizar o estimador de efeitos fixos, a qual tem a vantagem de, mais intuitivamente, permitir compreender porque o impacto da

²⁶ Esta divergência só é relevante se o valor de ϕ_{1i} influenciar a (auto)selecção.

formação deve ser equacionado como $\tilde{\alpha}_t^D$ e não como $\tilde{\alpha}_t$. Isto está patente na seguinte ilustração gráfica (figura 2).

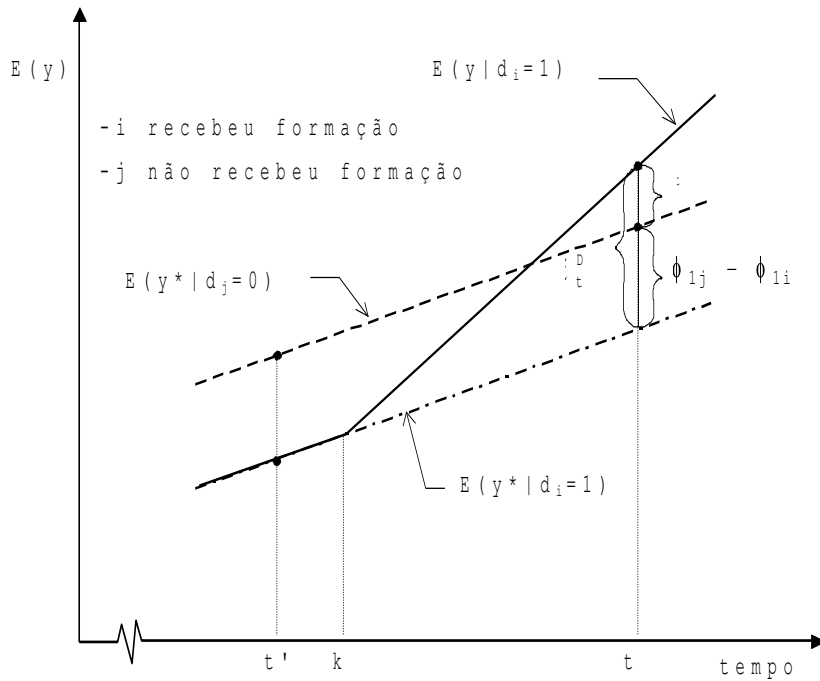


Figura 2

$$\tilde{\alpha}_t^D = E(y_{it} - y_{it}^* | d_i = 1) - E(y_{it}^* - y_{it}^* | d_i = 0)$$

B.2. Admita-se que $U_{it} = \phi_{1i} + t\phi_{2i} + v_{it}$ onde ϕ_{2i} representa a taxa de crescimento do rendimento do indivíduo i , e ϕ_{1i} e v_{it} têm o significado que já lhes foi atribuído.²⁷

Desta vez, a dependência entre U_{it} e d_i é explicada pela divergência entre os valores de ϕ_1 e ϕ_2 relativos a diferentes indivíduos.²⁸

Nestas condições, verifica-se que²⁹

$$E[(U_{it} - U_{it'}) - (t - t')(U_{it'} - U_{it'-m}) | \mathbf{x}_i, d_i] =$$

²⁷ $E(\phi_{2i}) = 0$

²⁸ É claro que esta diferença apenas é relevante se a (auto)selecção for, de algum modo, condicionada pela taxa de crescimento do rendimento dos indivíduos.

²⁹ Ter-se-á $m=1$, se as observações pré-formação disserem respeito a dois períodos consecutivos.

$$\begin{aligned}
&=E[(t\phi_{2i}-t'\phi_{2i})+(v_{it}-v_{it'})-(t-t')(t'\phi_{2i}-(t'-m)\phi_{2i})+ \\
&+(v_{it'}-v_{it'-m})|\mathbf{x}_i,d_i]= \\
&=E[(t-t')\phi_{2i}+(v_{it}-v_{it'})-(t-t')\phi_{2i}-(t-t')(v_{it'}-v_{it'-m})|\mathbf{x}_i,d_i]= \\
&=E[(v_{it}-v_{it'})-(t-t')(v_{it'}-v_{it'-m})|\mathbf{x}_i,d_i]= \\
&=0 \quad \text{pois, por hipótese, } E(v_{it})=0 \quad \forall t.
\end{aligned}$$

Assim, e mantendo as duas variantes já conhecidas, são as equações seguintes aquelas que originam estimativas consistentes, nomeadamente do impacto da formação:

[variante 1]

$$\begin{aligned}
&(y_{it} - y_{it'}) - (t-t')(y_{it'} - y_{it'-m}) = \\
&= [(\mathbf{x}_{it} - \mathbf{x}_{it'}) - (t-t')(\mathbf{x}_{it'} - \mathbf{x}_{it'-m})] \beta + d_i \alpha_t + [(v_{it} - v_{it'}) - \\
&-(t-t')(v_{it'} - v_{it'-m})] \quad t' < k < t \quad \mathbf{[B.2.1]}
\end{aligned}$$

[variante 2]

$$\begin{aligned}
&(y_{it} - y_{it'}) - (t-t')(y_{it'} - y_{it'-m}) = \\
&= [(\mathbf{x}_{it} - \mathbf{x}_{it'}) - (t-t')(\mathbf{x}_{it'} - \mathbf{x}_{it'-m})] \beta + d_i \mathbf{c}_{it} \theta + [(v_{it} - v_{it'}) - \\
&-(t-t')(v_{it'} - v_{it'-m})] \quad t' < k < t \quad \mathbf{[B.2.2]}
\end{aligned}$$

Sob a hipótese de que se verifica $U_{it} = \phi_{1i} + t\phi_{2i} + v_{it}$ a forma apropriada de conceber o impacto da formação será:

$$\tilde{\alpha}_t^{DD} = E((y_{it} - y_{it'}) - (y_{it'} - y_{it'-m}) | d_i = 1) - E((y_{it} - y_{it'}) - (y_{it'} - y_{it'-m}) | d_i = 0)$$

pois

$$\begin{aligned}
\tilde{\alpha}_t^{DD} &= E((\mathbf{x}_{it}\beta + \alpha_{it} + \phi_{1i} + t\phi_{2i}) - 2(\mathbf{x}_{it'}\beta - \phi_{1i} + t'\phi_{2i}) + (\mathbf{x}_{it'-m}\beta + \phi_{1i} + (t'-m)\phi_{2i})) - \\
&- E((\mathbf{x}_{it}\beta + \phi_{1i} + t\phi_{2i}) - 2(\mathbf{x}_{it'}\beta - \phi_{1i} + t'\phi_{2i}) + (\mathbf{x}_{it'-m}\beta + \phi_{1i} + (t'-m)\phi_{2i})) = E(\alpha_{it}) = \alpha_t
\end{aligned}$$

$$t'-m < t' < k < t \quad E(v_{it}) = 0 \quad \forall t.$$

Esta especificação revela-se apropriada desde que possa ser considerada como válida a igualdade

$$E((y_{it} - y_{it'}) - (y_{it'} - y_{it'-m}) | d_i = 1) = E((y_{it} - y_{it'}) - (y_{it'} - y_{it'-m}) | d_i = 0)$$

Caso contrário, subsistirá um enviesamento por selectividade que

$$\tilde{\alpha}_t^{DD} \text{ não controla.}$$

Para mostrar que $\tilde{\alpha}_t^D$ é, neste caso, uma formulação deficiente de α_t , considere-se que o indivíduo i participou no programa mas o indivíduo j não, pelo que se tem:

$$E(y_{it} - y_{it'} | d_i = 1) = E(\mathbf{x}_{it}\beta + \alpha_{it} + \phi_{1i} + t\phi_{2i} - \mathbf{x}_{it'}\beta - \phi_{1i} - t'\phi_{2i})$$

$$= (\mathbf{x}_{it} - \mathbf{x}_{it'})\beta + \alpha_{it} + (t - t')\phi_{2i}$$

$$E(y_{jt} - y_{jt'} | d_j = 1) = E(\mathbf{x}_{jt}\beta + \phi_{1j} + t\phi_{2j} - \mathbf{x}_{jt'}\beta - \phi_{1j} - t'\phi_{2j})$$

$$= (\mathbf{x}_{jt} - \mathbf{x}_{jt'})\beta + (t - t')\phi_{2j}$$

Portanto, mesmo que $\mathbf{x}_{it} - \mathbf{x}_{it'} = \mathbf{x}_{jt} - \mathbf{x}_{jt'}$, $\tilde{\alpha}_t^D$ não será uma boa formulação de α_t na medida em que ϕ_{2i} diferir de ϕ_{2j} , ou seja, na medida em que a taxa de crescimento do rendimento dos participantes no programa divirja da taxa de crescimento do rendimento dos não participantes relativa a períodos anteriores ao da formação.³⁰

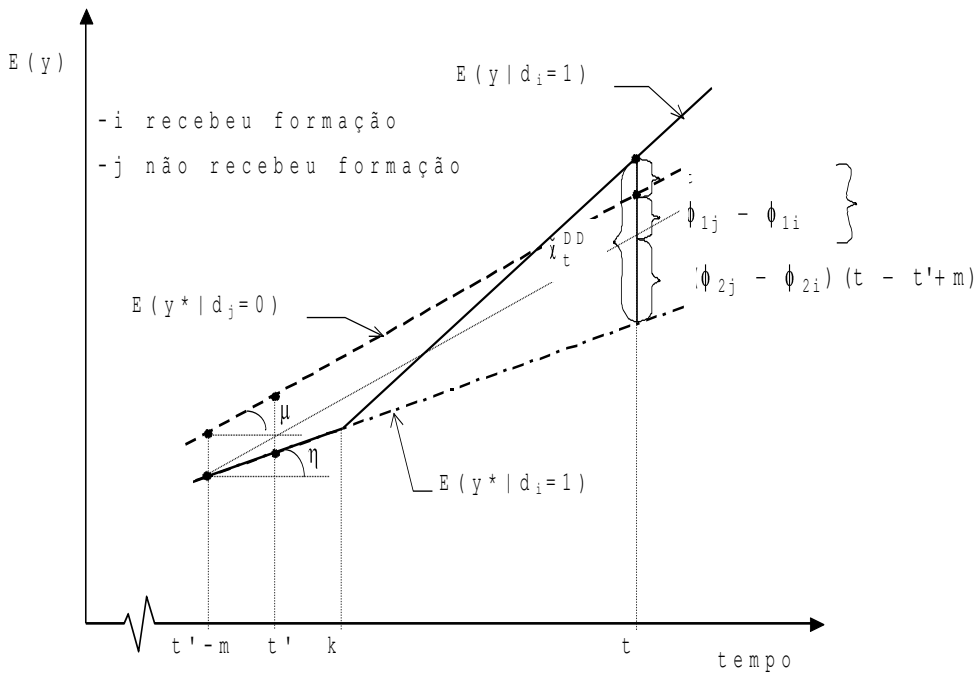


Figura 3³¹

³⁰ Obviamente que, tendo-se $U_{it} = \phi_{1i} + t\phi_{2i} + v_{it}$, considerar $\tilde{\alpha}_t^D$ é ainda menos acertado do que considerar α_t pois não se estará a ter em consideração o enviesamento devido a eventuais diferenças entre os valores de ϕ_{1i} relativos a diferentes indivíduos.

³¹ Para comprovar a relação entre $\tilde{\alpha}_t^D$ e $\tilde{\alpha}_t^{DD}$, aqui esquematizada, atenda-se a que se tem $\phi_{2i} = \text{tg}(\eta)$ e $\phi_{2j} = \text{tg}(\mu)$.

Na figura 3, pretende evidenciar-se o tipo de situação em que as formulações $\tilde{\alpha}_t$ e $\tilde{\alpha}_t^D$ se apresentam inadequados para avaliar o impacto da formação sobre os rendimentos.

A quebra nos rendimentos dos futuros formandos habitualmente³² registada nos s períodos imediatamente anteriores ao do arranque do programa de formação, explica-se pelos critérios adoptados para a selecção dos indivíduos a formar. Neste caso (ver Figura 4), considerar $\tilde{\alpha}_t^{DD}$ implica sobrestimar o impacto da formação sobre o rendimento de i , do período t , na medida em que uma parcela (\overline{AB}) do aumento do rendimento dos formados, relativamente ao rendimento dos não-formados, subsequente à formação, se explica pela retoma de um perfil de rendimentos apenas momentaneamente perturbado por factores conjunturais. Esta quebra poderá, pois, dificultar a tarefa de discernir o verdadeiro impacto da formação sobre o rendimento.

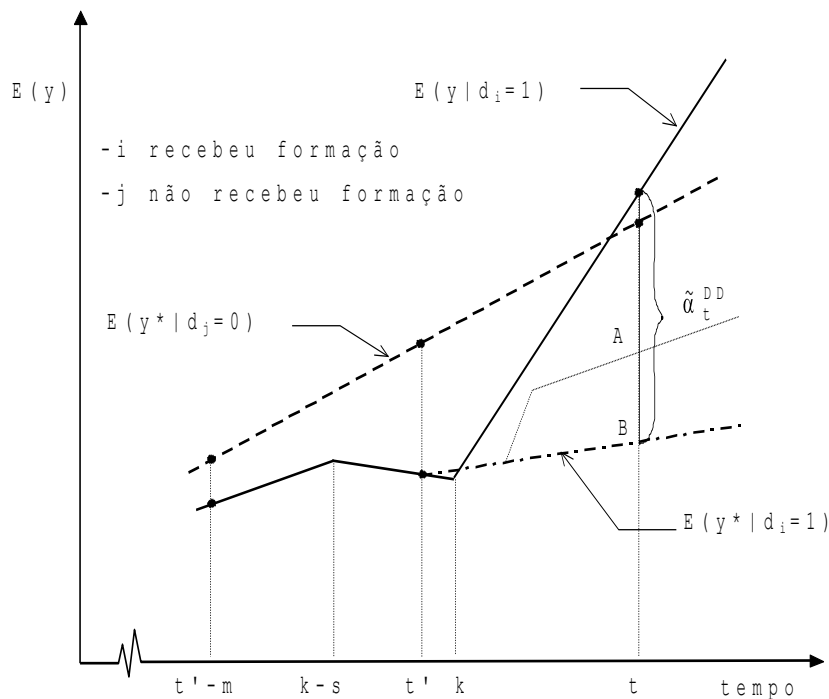


Figura 4

³² Ver O. Ashenfelter [1978].

Repare-se que as especificações até aqui apresentadas, dada a sua forma linear, não reflectem a concavidade do perfil de rendimentos empiricamente comprovada e teoricamente analisada no âmbito da economia do trabalho, onde coexistem diversas teorias que concorrem na sua explicação.

Para que, de alguma forma, a especificação adoptada traduza aquela concavidade, justificar-se-ia introduzir uma terceira componente, ϕ_3 , com o coeficiente $(t-t')^2$, na expressão representativa de U_i .

Ter-se-ia, então, $U_{it} = \phi_{1i} + t\phi_{2i} + (t-t')^2\phi_{3i} + v_{it}$, o que obrigaria a redefinir a equação de regressão e a dispor de dados relativos a mais um período anterior à formação.

Do que foi dito, infere-se que o estimador de α_t será tanto melhor quanto maior for o número de variáveis desfasadas que integrar.

Estendendo este raciocínio, conclui-se que a expressão $\tilde{\alpha}_t^A = E(y_{it}^* | d_i = 1, y_{ik}^*, \dots, y_{i,k-p}^*) - E(y_{it}^* | d_i = 0, y_{ik}^*, \dots, y_{i,k-p}^*)$

(com $p = n^\circ$ de períodos pré-formação relativamente aos quais existem dados disponíveis) será ajustada para traduzir α_t desde que $E(y_{it}^* | d_i = 1, y_{ik}^*, \dots, y_{i,k-p}^*) = E(y_{it}^* | d_i = 0, y_{ik}^*, \dots, y_{i,k-p}^*)$.

Significa isto que o estimador correspondente à formulação $\tilde{\alpha}_t^A$ apenas será cêntrico se dois quaisquer indivíduos, com idênticos

perfis de rendimento pré- formação, tiverem o mesmo rendimento, y^*_t , quer tenham, ou não, recebido formação.

Quando é verosímil assumir certas hipóteses sobre a distribuição das variáveis implicadas no modelo, designadamente, sobre a distribuição do "rendimento presumido", y^*_{it} , condicional em d_i , pode aferir-se o impacto da formação sobre os rendimentos através dos desvios verificados nos momentos³³ daquela distribuição, quando considerada a amostra composta pelos indivíduos que receberam formação.

No entanto, esta via de estimação do impacto raramente estará aberta, dada a impossibilidade de fundamentar as hipóteses sobre a distribuição de y^*_{it} com base em informações relativas a outros períodos. Constata-se que as únicas populações relativamente às quais é, em princípio, possível caracterizar a distribuição de y^*_{it} , são as relativas a um grupo de controlo expressamente criado, ou a um grupo definido por uma variável Z_i relevante para traduzir o processo de (auto)selecção. Mas, se, no primeiro caso, o problema da estimação em causa fica totalmente descaracterizado³⁴, no segundo caso, é-se reconduzido a um modelo já apresentado, revelando-se, pois, supérflua qualquer hipótese sobre a distribuição de y^*_{it} .

³³ Não, forçosamente, apenas os desvios registados na média, mas também em outros momentos de ordem superior.

³⁴ Se existe um grupo de controlo, o impacto da formação poderá ser estimado por recurso a dados experimentais.

TESTES DE ESPECIFICAÇÃO A MODELOS NÃO-EXPERIMENTAIS DE AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL

Sabe-se já que se podem obter estimativas do impacto dos programas de formação profissional recorrendo, essencialmente, a dois tipos de métodos: o método experimental e o método não-experimental.

O recurso a este último apresenta-se, à partida, como uma boa alternativa ao primeiro na medida em que é menos exigente em termos de meios necessários à sua aplicação. O método experimental, para além de, em geral, implicar a montagem de toda uma dispendiosa estrutura de controlo envolvendo volumes significativos de meios materiais e humanos, encontra resistências dada a sua intrínseca discricionariedade ser encarada como socialmente injusta, eticamente reprovável e, portanto, politicamente inconveniente.

Por outro lado, a validade das estimativas do impacto dos programas que fornece depende da possibilidade de se concretizar o controlo e acompanhamento previamente projectados.³⁵

Mas, se se admitir que, apesar das dificuldades, se consegue exercer o controlo de forma minimamente satisfatória, então as estimativas experimentais do impacto dos programas revelam-se particularmente precisas, podendo ser usadas como referência para avaliar a operacionalidade de modelos não-experimentais de avaliação do referido impacto.

³⁵ Muitas vezes torna-se difícil fazer cumprir o esquema delineado para a recolha de informações necessárias ao cálculo acurado das estimativas, o que se ficará a dever a uma multiplicidade de vicissitudes.

Porém, esta via para testar modelos econométricos alternativos não pode ser sistematicamente adoptada pois, muitas vezes, não estão disponíveis estimativas experimentais.

Pode-se, isso sim, aproveitar a disponibilidade dos dados e estimativas experimentais para comparar a fiabilidade dos testes econométricos destinados a avaliar diferentes especificações.

Uma vez comprovada a sua operacionalidade relativa, fica reforçada a legitimidade para os usar para escolher o modelo não-experimental gerador de estimativas virtualmente equivalentes às estimativas experimentais. Identificada, por este processo, a especificação correcta, obter-se-ão estimativas que podem ser consideradas fiáveis mesmo quando não seja possível cotejá-las com estimativas baseadas em dados experimentais correctamente recolhidos e controlados.

É conveniente distinguir, basicamente, dois tipos de testes de especificação consoante a hipótese que se pretenda verificar:

I. Hipótese HV1: $\alpha_t=0$ no caso de se querer testar a variante 1 da especificação adoptada;

Hipótese HV2: $\theta_t=0$ no caso de se querer testar a variante 2 da especificação adoptada;

II. Hipótese: $\gamma_t=0$ sendo γ_t o coeficiente da variável y_{it} com t identificando um período diferente daqueles que aparecem considerados na especificação a testar.

I. A aplicação deste teste à especificação que estiver em causa consiste em averiguar se o coeficiente de d_i não se apresenta significativamente diferente de zero³⁶ quando tal seja de esperar dada a definição da variável dummy, d_i , e a natureza dos dados utilizados.

Consoante se disponha apenas de dados não-experimentais ou, para além destes, estejam igualmente disponíveis dados experimentais, interessa considerar-se, respectivamente:

I.1.³⁷ $d_i=1$ se o indivíduo i recebeu formação;
 $d_i=0$ se o indivíduo i pertence ao grupo de comparação.

I.2.

I.2.1.³⁸ $d_i=1$ se o indivíduo i participa no programa (seja como formando, seja como elemento do grupo de controlo);
 $d_i=0$ se o indivíduo i pertence ao grupo de comparação.

I.2.2.³⁹ $d_i=1$ se o indivíduo i pertence ao grupo de controlo;
 $d_i=0$ se o indivíduo i pertence ao grupo de comparação.

³⁶ Poder-se-á tomar, como indicador da correcção da especificação em causa, a probabilidade de se verificar a hipótese nula relevante HV1 (ou HV2).

³⁷ Neste caso, o teste só poderá realizar-se com base em dados pré-formação, desde que disponíveis.

³⁸ Neste caso, o teste só poderá realizar-se com base em dados pré-formação, desde que disponíveis.

³⁹ Esta definição de d_i visa permitir a realização de um teste à especificação que estiver em causa com base em dados pós-formação, quando disponíveis.

II. A aplicação deste teste a cada uma das especificações [B.1.1],[B.1.2],[B.2.1] e [B.2.2]⁴⁰, consiste em averiguar se especificações alternativas que integrem como regressores variáveis y_i referidas a outro(s) período(s) que não os considerados naquelas equações, fornecem estimativas dos respectivos coeficientes, γ_t , não significativamente diferentes de zero. É que, se tal não suceder, deve reconhecer-se a deficiência das especificações originárias e, conseqüentemente, a futilidade das correspondentes estimativas.

Naturalmente que a realização deste tipo de testes requer a disponibilidade de dados relativos a, pelo menos, outro período para além dos inicialmente considerados, independentemente de se tratar de dados pré ou pós-formação.

Sob a hipótese HV1 (HV2), e para cada uma das definições relevantes de d_i consideradas em I.1 e I.2, pode testar-se a especificação [A.1]([A.2]), (se se admitiu que a selecção se faz sobre observáveis), ou as especificações [B.1.1],[B.1.2] ([B.2.1],[B.2.2])⁴¹, (se se admitiu que a selecção se faz sobre não-observáveis).

Testar a especificação [A.1]([A.2]) equivale a testar a validade da função (linear) de controlo correspondente. A função (linear)

⁴⁰ A aplicação deste teste às especificações [A.1] ou [A.2] redundaria numa prática de "stepwise regression", metodologicamente reprovável.

⁴¹ Como já se viu, outras especificações poderão ser consideradas conforme a forma que se admita para U_{it} .

de controlo adoptada será válida desde que, como é óbvio⁴², se possa mostrar que $\alpha_t(\theta_t)$ não se apresenta significativamente diferente de 0(0) quando se usam observações relativas a um período, t, anterior à formação ($t < k$) para estimar a equação dos rendimentos.

A estimativa de $\alpha_t(\theta_t)$ decorrente da estimação da equação [B.1.1]([B.1.2]) com base em dados relativos aos períodos t' e t anteriores ao período, k, da formação ($t' < t < k$), não deve apresentar-se significativamente diferente de zero (0), para que a especificação [B.1.1]([B.1.2]) possa ser considerada adequada, isto é, se o enviesamento por selectividade se ficar a dever apenas à discrepância dos efeitos fixos.

De modo análogo, quando se estima a equação [B.2.1]([B.2.2]) com base em dados relativos aos períodos t'-m, t' e t anteriores ao período, k, da formação ($t' < t < k$), não deve obter-se uma estimativa de $\alpha_t(\theta_t)$ significativamente diferente de zero (0), se se pretende usar a especificação [B.2.1]([B.2.2]) para avaliar o impacto do programa de formação sobre o rendimento.

Repare-se que a realização deste tipo de teste de especificação, no caso de se adoptar a especificação compatível com $U_{it} = \phi_{1i} + t\phi_{2i} + (t-t')^2\phi_{3i}$, exigiria, já, a disponibilidade de dados relativos a quatro períodos anteriores ao da formação.

⁴² Seria disparatado usar uma especificação que desse conta de um impacto da formação sobre o rendimento de períodos anteriores ao da própria formação.

Com d_i definida como em II.2, o teste conduzirá à rejeição da equação em causa se a estimativa de $\alpha_t(\theta_t)$ for significativamente diferente de zero, uma vez que tal resultado não se coaduna com o facto de apenas se terem usado observações relativas a indivíduos que não receberam formação e cujo rendimento, portanto, não pode ter sofrido qualquer impacto dela decorrente.

MÉTODO EXPERIMENTAL *VERSUS* MÉTODO NÃO-EXPERIMENTAL

Na sequência desta tipificação dos testes destinados a apreciar modelos não-experimentais de avaliação do impacto dos programas de formação profissional, vem a propósito referir o confronto entre aqueles que, como Robert Lalonde [1986], não confiam nas estimativas daquele impacto obtidas a partir de modelos econométricos, e aqueles que, como James Heckman [1989], acreditam ser possível garantir a fiabilidade daquelas estimativas desde que o modelo adoptado possa ser sancionado por testes de especificação.

Os primeiros argumentam que a grande discrepância entre as estimativas não-experimentais oriundas de diferentes modelos, e entre estas e as estimativas experimentais (quando disponíveis), retira qualquer credibilidade aos estimadores não-experimentais. Face a uma forte suspeição quanto ao recurso a modelos não-experimentais de avaliação, Heckman e Hotz contrapõem que a disparidade entre as estimativas não-experimentais geradas por modelos alternativos se deve ao facto de se verificarem diferenças sistemáticas entre formados e não-formados quanto às variáveis (observáveis e não-observáveis) que condicionam os

rendimentos e a selecção. Na sua perspectiva, aquela discrepância apenas é sintomática de tais diferenças sistemáticas, na medida em que estas são desigualmente consideradas pelos diversos modelos alternativos, não sendo legítimo invocá-la para demonstrar o alegado fracasso daqueles modelos. Contra-argumentam, ainda, que é possível avaliar objectivamente modelos não-experimentais alternativos recorrendo aos testes de especificação que propõem e cuja operacionalidade se comprovaria pelo facto de conduzirem:

- à rejeição de modelos não-experimentais que fornecem estimativas incongruentes com a evidência experimental e
- à não rejeição de modelos que fornecem estimativas congruentes com a evidência experimental.

PRINCIPAIS ASPECTOS A CONSIDERAR PARA PROCEDER À AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL

Nesta altura, deve ser evidente que o investigador interessado em (ou condenado a) servir-se de dados não-experimentais para avaliar o impacto sobre os rendimentos de programas de formação profissional, terá forçosamente que ponderar certas questões primordiais:

- como constituir o grupo de comparação;
- que especificação adoptar para explicar, adequadamente, os rendimentos;
- como traduzir o processo de selecção.

A primeira questão remete, desde logo, para a necessidade de caracterização, tão detalhada quanto possível, do grupo de formandos em análise. A composição do grupo de comparação será

depois ditada pela preocupação em fazer com que os dois grupos sejam, virtualmente, indistintos do ponto de vista estatístico.⁴³ Só assim se poderá pretender que a especificação a adoptar explique os rendimentos dos indivíduos independentemente de pertencerem a um, ou outro, grupo, o que constitui uma hipótese crucial do método de avaliação não-experimental.

Quanto à segunda questão, deve ter-se bem presente que a definição do modelo a adoptar para a estimação do impacto da formação não deve assentar no aproveitamento exaustivo de dados relativos ao rendimento e a potenciais variáveis Z_i . Pelo contrário, a recolha de dados deverá processar-se selectivamente, subordinando-se à especificação radicada no conhecimento da especificidade do processo de selecção, sob pena de se estar a comprometer, inapelavelmente, a validade dos resultados da investigação.

A propósito da terceira questão, é importante salientar que a pesquisa e escolha de variáveis Z_i assenta no estudo e conhecimento dos aspectos específicos do processo de selecção.

Designadamente, importa investigar as razões porque a acessibilidade aos programas de formação varia, seja no espaço geográfico, seja na estrutura social. Entre os factores que condicionam a acessibilidade destes programas, conta-se a heterogeneidade na divulgação e na disponibilidade dos elementos de informação pertinentes.

⁴³ Ver F. Thomas, R. Maynard e N. Lyle [1984].

LIMITAÇÕES QUANTO À DISPONIBILIDADE DE DADOS NÃO-EXPERIMENTAIS

Pouco se sabe, com segurança, sobre o verdadeiro impacto da formação sobre os rendimentos, o que se ficará a dever, sobretudo, à grande dimensão das amostras necessárias para o avaliar e aos elevados custos a suportar para acompanhar a evolução dos rendimentos dos formados.

Dado o peso destes condicionalismos, há que, pragmaticamente, definir estratégias de pesquisa e recolha de dados com nível de exigência gradual.

A recolha de dados deverá ser, sistematicamente, orientada no sentido de permitir testar aquelas hipóteses que anteriormente o não puderam ter sido, precisamente por insuficiência de dados. Assim, o conhecimento condensado em estudos já realizados pode aproveitar a cada novo estudo ou etapa da investigação anteriormente encetada.

Se estiverem disponíveis dados experimentais oferece-se a possibilidade de, mais seguramente, se identificar o modelo (função rendimentos e regra de selecção) que melhores estimativas fornece, dada a especificidade da investigação em curso. Subsequentemente, quando se pretender fazer análises similares, pode cingir-se a recolha de dados àqueles que são estritamente necessários à estimação do modelo previamente adoptado.

ÂMBITO DA AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL

Subjacente à abordagem aqui feita sobre a avaliação de programas de formação profissional, esteve a ideia básica de que esta avaliação passa exclusivamente pela mensuração do impacto dos programas sobre os rendimentos (porque sobre o valor do produto marginal) dos formandos.

No entanto, não será despropositado admitir-se que a referida avaliação se possa fazer numa perspectiva mais alargada. Por exemplo, poder-se-ia tentar isolar o efeito dos programas de formação sobre os níveis de emprego através da elaboração de um modelo de âmbito macroeconómico.

Um outro critério de avaliação passaria pela análise da influência dos programas de formação profissional sobre a competitividade das empresas face ao exterior. Sabendo-se que o impulso dado a estes programas decorreu do processo de integração na comunidade europeia, será incorrecto pretender avaliá-los medindo tão só o seu impacto sobre os rendimentos, e ignorando o seu efeito sobre outros aspectos, nomeadamente os traduzidos por indicadores de convergência.

Contudo, este interesse em proceder a uma avaliação mais lata dos programas de formação profissional, apenas se poderá concretizar desde que possam ser encontradas soluções ao nível da elaboração de modelos válidos, apesar das inevitáveis limitações quanto à disponibilidade e natureza dos dados a usar.

Abre-se, assim, um vasto campo à capacidade inventiva e ao esforço pragmático dos investigadores que nele se queiram aventurar.

REFERÊNCIAS BIBLIOGRÁFICAS

Abraham, Katherine G. e Henry Farber, [1987], "Job duration, seniority and earnings", *The American Economic Review*, 77 n° 3, 278-297. (*)

Ashenfelter, Orley, [1978], "Estimating the effect of training programs on earnings", *Review of Economics and Statistics*, 60, Fevereiro, 47-57. (*)

Ashenfelter, Orley e David Card, [1985], "Using the longitudinal structure of earnings to estimate the effect of training programs", *Review of Economics and Statistics*, 67, Novembro, 648-60. (*)

Barnow, B., G. Cain, e A. Goldberger, [1980], Issues in the analysis of selectivity bias. In *Evaluation Studies Review Annual*, E. Stornsdorfer e G. Farks vol. 5. Beverly Hills, CA: Sage, 42-59.

Bartel, Ann P., [1992], Training, wage growth and job performance: evidence from a company database, Working paper n° 4027, National Bureau of Economic Research, Cambridge, Março. (*)

Cain, G., [1975], "Regression and selection models to improve nonexperimental comparisons". In *Evaluating and Experiment*, C. A. Bennett e A. A. Lumsdaine. New York: Academic press, 297-317.

Goldberger, A., [1972], Selection bias in evaluating treatment effects: some formal illustrations. Discussion paper 123-72. Madison, W.I.: Institut for Research on Poverty.

Heckman, James e V. Joseph Hotz, [1989], "Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training", *Journal of the American Statistical Association*, 84, Dezembro, 862-74. (*)

- Lalonde, Robert, [1986]**, "Evaluating the econometric evaluations of training programs with experimental data", *American Economic Review*, 76, 604-620. (*)
- Linch, Lisa M., [1992]**, "Diferential effects of post-school training on early career mobility", Working paper n° 4034, National Bureau of Economic Research, Cambridge, Março. (*)
- Moffit, Robert, [1991]**, "Program evaluation with nonexperimental data", *Evaluation Review*, vol. 15, n° 13, Junho, 291-314. (*)
- Polachek, Soloman e Stanley Siebert, [1993]**, The economics of earnings, Cambridge University Press. (*)
- Rosen, Sherwin, [1976]**, "A theory of life earnings", *Journal of Political Economy*, vol.84, n° 4, University of Chicago. (*)
- Thomas, F., R. Maynard e N. Lyle, [1984]**, An assessment of alternative comparison groups methodologies for evaluating employment and training programs. Princeton: Mathematica Policy Research. Inc..

(*) Textos consultados.

ÍNDICE

| | |
|--------------------------------------------------------------------------------------------------------|----|
| PROPÓSITO | 2 |
| INTRODUÇÃO | 3 |
| ESTIMAÇÃO DO IMPACTO DOS PROGRAMAS DE FORMAÇÃO PROFISSIONAL SOBRE OS RENDIMENTOS | 6 |
| TESTES DE ESPECIFICAÇÃO A MODELOS NÃO-EXPERIMENTAIS DE AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL | 27 |
| MÉTODO EXPERIMENTAL <i>VERSUS</i> MÉTODO NÃO-EXPERIMENTAL | 32 |
| PRINCIPAIS ASPECTOS A CONSIDERAR PARA PROCEDER À AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL | 33 |
| LIMITAÇÕES QUANTO À DISPONIBILIDADE DE DADOS NÃO-EXPERIMENTAIS | 35 |
| ÂMBITO DA AVALIAÇÃO DE PROGRAMAS DE FORMAÇÃO PROFISSIONAL | 36 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 38 |