2
Identificação do Efeito de Tratamento sob Ignorabilidade

Neste capítulo, será definido o problema ao qual se aplicam os métodos discutidos no restante do trabalho. Consideramos uma população heterogênea de unidades (como indivíduos, domicílios ou firmas) que podem estar sujeitas a uma variedade de regimes alternativos, os tratamentos (por exemplo, benefícios sociais, situação geográfica, ou sistemas tributários). Pretendemos estimar o 'Efeito de Tratamento', o qual pode ser definido de forma abrangente como o efeito causal do tratamento sobre um certo atributo das unidades, a variável de interesse.

Seguindo a metodologia de Rubin (1973, 1977, 1978), o efeito causal é obtido pela comparação entre os possíveis valores da variável de interesse sob diferentes condições de tratamento. A dificuldade central decorrente desta abordagem é que, como cada unidade é observada sob apenas um dos tratamentos, a comparação necessária envolve valores não observados.

Diversos substitutos para a comparação (impossível) entre valores contrafactuais foram sugeridos e estudados pela literatura de avaliação de programas. O modelo a ser estudado supõe a disponibilidade de informações adicionais sobre as unidades, chamadas variáveis auxiliares, covariáveis ou variáveis pré-tratamento. Sob um determinado tipo de hipótese, denominada genericamente hipótese de Ignorabilidade, é legítimo comparar os valores da variável de interesse entre unidades de mesmas características auxiliares, mas sujeitas a diferentes tratamentos.

Estas condições sugerem a realização de experimentos onde grupos de indivíduos homogêneos quanto às variáveis auxiliares sejam submetidos, aleatoriamente, a diferentes tratamentos. Neste caso, diz-se que estão disponíveis dados experimentais. No entanto, freqüentemente o analista não tem a possibilidade de realizar tais experimentos, contando apenas com dados obtidos por amostragem de uma população onde atribuição do tratamento está fora de seu controle. O modelo apresentado neste capítulo representa este último caso, o problema de inferência, a partir de dados não-experimentais, ou observacionais, sob uma hipótese de Ignorabilidade.

2.1
Elementos Básicos

Os dados disponíveis resultam da observação de uma amostra aleatória de N unidades da população de interesse (Y, X, T), indexadas por $i = 1, 2, ..., N$. Para cada unidade i, o valor observado da variável de interesse é Y_i, que tomaremos, como é típico, como valores na reta real. O tratamento atribuído a i é indicado por T_i. Consideraremos, como na maior parte da literatura, um conjunto de tratamentos alternativos binário, representando a participação ou não no programa que se pretende avaliar[1]. Neste contexto, distinguem-se o grupo de tratamento, indicado por $T_i = 1$, constituido pelas unidades afetadas pela intervenção, e o grupo de controle, indicado por $T_i = 0$, das unidades não afetadas. Finalmente, X_i representa o conjunto de variáveis auxiliares da unidade i. A propriedade fundamental de X_i é o fato de que nem sua observação, nem seu valor dependem, no sentido causal, do tratamento T_i. Uma possibilidade em que isto se verifica logicamente é X_i ser observada antes da determinação do tratamento recebido. Deste caso particular segue a denominação de variáveis pré-tratamento.

A abordagem de valores potenciais de Rubin nos leva a postular as variáveis $Y_i(t)$, definidas para cada valor possível t de T_i, representando o valor de Y_i caso o indivíduo i estivesse sujeito à condição de tratamento $T_i = t$. Este artifício, importante na elaboração do modelo[2], traz implícita a hipótese de que cada unidade não é influenciada pelo tratamento recebido pelas outras. Esta hipótese, bastante plausível no contexto original de experimentos clínicos (como em Rubin, 1978), requer cuidadosa interpretação do modelo se aplicada a fenômenos sociais.

Outro conceito devido a Rubin é o *propensity score* (Rosenbaum e Rubin, 1983), definido como a probabilidade de seleção para o tratamento, condicional ao valor das variáveis auxiliares.

$$p(x) \equiv Pr[T_i = 1|X_i = x] = E[T_i|X_i = x]$$

Para a discussão ao longo deste trabalho, é conveniente também definir as

[1]As técnicas empregadas para o estudo do efeito de um tratamento que pode assumir um número finito de valores são similares às do caso binário. Uma análise de hipóteses de identificação análogas às que usamos nesta dissertação pode ser vista em Imbens (2000). Cattaneo (2007) desenvolve, nesse contexto, versões de dois dos estimadores que discutiremos. A recente resenha de Imbens e Wooldridge (2009) oferece uma discussão sobre métodos para tratamentos com valores contínuos.

[2]Imbens e Wooldridge (2009) discutem as dificuldades de se formular o problema utilizando apenas valores observados.

chamadas funções de regressão

$$m_t(x) \equiv E[Y_i(t)|X_i = x]$$

e de variância condicional

$$\sigma_t^2(x) \equiv V[Y_i(t)|X_i = x].$$

2.2 Parâmetro de Interesse

O parâmetro que se pretende estimar é

$$\beta = E\left[Y(1) - Y(0)\right] = E\left[Y(1)\right] - E\left[Y(0)\right],$$

a expectativa da diferença entre os valores potenciais do atributo de interesse, chamado Efeito Médio de Tratamento. O foco na inferência sobre o efeito médio não deve ser visto como uma limitação importante, pois generaliza-se imediatamente para transformações da variável de interesse. Em particular, tomando funções do tipo $g(Y) = 1(Y \leq y)$, e como os estimadores, na verdade, se decompõem em um termo para cada média de valor potencial, é possível estimar as distribuições marginais de $Y(t)$, $F_t(y) = Pr(Y(t) \leq y) = E[g(Y(t))]$, em cada ponto y. Outra possibilidade é discutida adiante, na seção 2.6.

Uma família de parâmetros de interesse relacionada que também tem recebido grande atenção da literatura é a dos Efeitos de Tratamento Sobre os Tratados. Estes consideram médias condicionais à atribuição do tratamento, medindo desta forma o efeito do programa sobre a população das unidades que efetivamente sofreram a intervenção. O exemplo principal deste tipo de parâmetro é o Efeito Médio de Tratamento Sobre os Tradados, $\beta_{EMTT} = E\left[Y(1) - Y(0)|T = 1\right]$. Por limitação de tempo, estimadores para β_{EMTT} não serão abordados neste trabalho.

2.3 Identificação

Se ambos os valores potenciais fossem observados para cada unidade, o problema de avaliar o Efeito Médio de Tratamento consistiria simplesmente na inferência sobre a média populacional de $Y_i(1) - Y_i(0)$. No entanto, como os indivíduos são observados apenas sob uma das condições de tratamento, este procedimento não é factível. Além disso, devido à heterogeneidade entre os

grupos, a estimação das médias populacionais de $Y(1)$ e $Y(0)$ pelas médias nas subpopulações em que são observadas sofre viés de seleção. A identificação do parâmetro de interesse depende, portanto, de hipóteses sobre o relacionamento entre as variáveis não observadas e as observadas.

Neste trabalho, serão estudados estimadores baseados em uma hipótese de Ignorabilidade, que corresponde a afirmar que a heterogeneidade relevante e/ou sistemática é captada pelas variáveis auxiliares. Mais especificamente, consideraremos a hipótese de Ignorabilidade Forte, que chamaremos simplesmente de Ignorabilidade, devida a Rosenbaum e Rubin (1983), definida pela conjunção das hipóteses de Independência Condicional[3]:

$$T_i \perp (Y_i(1), Y_i(0)) \mid X_i, \tag{2-1}$$

que determina a independência da atribuição do tratamento em relação aos valores potenciais, dadas as variáveis auxiliares, e de Sobreposição:

$$0 < \varepsilon < p(x) < 1 - \varepsilon < 1, \ \forall \, x, \tag{2-2}$$

que estabelece um limite uniforme, maior que zero, para a probabilidade de seleção e de não seleção.

Um caso análogo muito estudado é o de dados faltantes (*missing data*), no qual se procura fazer inferência sobre a distribuição marginal de uma variável Y, observada apenas em parte da amostra. A hipótese de Dados Faltando Aleatoriamente (*missing at random*) supõe que a probabilidade de observar Y é independente de seu valor, condicionalmente às variáveis auxiliares. Isto define um modelo idêntico ao de Efeito de Tratamento sob Ignorabilidade, quando se considera $Y(0)$ identicamente nulo. Inversamente, estimar o Efeito Médio de Tratamento equivale a resolver um problema de dados faltantes para obter estimativas da média de cada valor potencial, $E[Y(1)]$ e $E[Y(0)]$, e então tomar a diferença. Neste caso, a hipótese de Ignorabilidade na formulação original implica na de Dados Faltando Aleatoriamente nos dois problemas em que se decompõe.

Uma forma de mostrar a identificação de β sob Ignorabilidade é motivada pela seguinte aplicação da Lei das Expectativas Iteradas:

$$\begin{aligned} \beta = E\left[Y_i(1) - Y_i(0)\right] &= E\left[E\left[Y_i(1) - Y_i(0)|X_i\right]\right] \\ &= E\left[E\left[Y_i(1)|X_i\right] - E\left[Y_i(0)|X_i\right]\right] \\ &= E\left[m_1(X_i) - m_0(X_i)\right]. \end{aligned}$$

[3]Também conhecida como *Unconfounded Treatment Assignment* (Rosenbaum e Rubin, 1983).

Ou seja, o parâmetro de interesse pode ser representado como a média populacional de uma diferença entre funções das variáveis observadas X. Embora as funções de regressão sejam supostas desconhecidas e envolvam resultados potenciais, temos, como conseqüência imediata da Independência Condicional, as igualdades

$$E\left[Y_i(t)|X_i\right] = E\left[Y_i(t)|X_i, T_i = t\right] = E\left[Y_i|X_i, T_i = t\right],\ t = 0, 1.$$

Logo é possível reescrever $m_1(.)$ e $m_0(.)$, utilizando, respectivamente, os dados referentes às observações tratadas e de controle:

$$\beta = E\left[m_1(X_i) - m_0(X_i)\right] = E\left[E\left[Y_i|X_i, T_i = 1\right] - E\left[Y_i|X_i, T_i = 0\right]\right]. \quad (2\text{-}3)$$

A hipótese de Sobreposição permite, por exemplo, estimar $E\left[Y_i(1)|X_i\right] = E\left[Y_i|X_i, T_i = 1\right]$ e avaliá-la usando a distribuição empírica de X. De fato, por 2-2, em qualquer região à qual a distribição marginal de X associa probabilidade não nula, digamos q, a probabilidade condicionada a $T = 1$ é no mínimo $q\frac{\varepsilon}{1-\varepsilon}$, logo também positiva. Desta forma, está garantida (probabilisticamente) a observação do comportamento de $Y(1)$ (e $Y(0)$) para todos os valores de X.

Outra maneira de mostrar a identificação é observar que, valendo a hipótese de Ignorabilidade, temos

$$\begin{aligned} E\left[\frac{TY}{p(X)}\right] &= E\left[E\left[\frac{TY}{p(X)}|X\right]\right] = E\left[\frac{E\left[TY|X\right]}{p(X)}\right] \\ &= E\left[\frac{E\left[Y(1)|X\right]Pr(T=1|X)}{p(X)}\right] \\ &= E\left[E\left[Y(1)|X\right]\right] = E\left[Y(1)\right], \end{aligned} \quad (2\text{-}4)$$

onde, a primeira expressão existe, por (2-2) e, na passagem para a segunda linha, usa-se

$$\begin{aligned} E\left[TY|X\right] &= E\left[Y|X, T=1\right]Pr(T=1|X) + 0Pr(T=0|X) \\ &= E\left[Y(1)|X, T=1\right]Pr(T=1|X) \end{aligned}$$

e a hipótese de Independência Condicional, que implica $E\left[Y(1)|X, T=1\right] = E\left[Y(1)|X\right]$. Analogamente, temos $E\left[\frac{(1-T)Y}{1-p(X)}\right] = E\left[Y(0)\right]$ e, portanto,

$$\beta = E\left[Y_i(1) - Y_i(0)\right] = E\left[\frac{TY}{p(X)}\right] - E\left[\frac{(1-T)Y}{1-p(X)}\right], \quad (2\text{-}5)$$

uma representação do parâmetro de interesse como diferença de médias ponderadas de valores observados. A equação 2-4 mostra que pesos proporcionais

ao inverso do *propensity score* tornam a média da variável de interesse entre os indivíduos tratados representativa da média do valor potencial $Y(1)$ na população.

2.4 Crítica e Alternativas à hipótese de Ignorabilidade

Entre os componentes da Ignorabilidade, a hipótese de Sobreposição é a parte testável, e sua verificação é recomendada. Sua falha significa que determinadas unidades de um grupo têm poucos correspondentes (ou nenhum) no outro, e portanto, que a escolha dos grupos para comparação é inadequada. Busso, DiNardo e McCrary (2009) mostram que, quando isto ocorre, a precisão dos estimadores baseados em Ignorabilidade é ruim, e a teoria assintótica pouco informativa sobre o desempenho. Este problema pode ser contornado pela exclusão de observações com valores extremos do *propensity score* (estimado), como feito normalmente em trabalhos aplicados. Crump et al. (2009) discutem sistematicamente esta possibilidade, mostrando com minimizar a variância assintótica de estimadores de Efeito Médio de Tratamento através do descarte de observações fora de determinado subconjunto A^* de perfis das variáveis auxiliares. O mesmo estudo apresenta ainda condições sob as quais A^* é determinado unicamente pelo *propensity score* das observações, isto é, tem a forma $A^* = \{x \in supp(X) | p(x) \in [\alpha, 1-\alpha]\}$[4], para algum $\alpha \in (0,1)$. Observamos, no entanto, que esta metodologia envolve uma redefinição da população, ou uma alteração do parâmetro de interesse, pois em relação à população original, os estimadores passarão a estimar $E[Y(1) - Y(0) | X \in A^*]$ em vez do Efeito Médio de Tratamento $\beta = E[Y(1) - Y(0)]$. Assim, se é importante avaliar o efeito do tratamento sobre unidades pouco representadas em algum dos grupos, uma estratégia de comparação de unidades similares, como Ignorabilidade, claramente não pode funcionar.

O uso da hipótese de Ignorabilidade é questionado, de forma mais contundente, no que se refere à validade da Independência Condicional. A razão é que, em aplicações econométricas, as unidades são tipicamente agentes econômicos que se importam com o valor da variável Y. É plausível que estes disponham de mais informação sobre os valores potenciais que os dados disponíveis nas variáveis auxiliares, e que se auto-selecionem para o tratamento conforme suas expectativas. Logo, a Independência Condicional é ameaçada pela possibilidade de a informação adicional não ser independente da atribuição do tratamento, dado X.

[4] $supp(X)$ representa o conjunto dos possíveis valores de X

Este argumento sugere que Independência Condicional pode ser uma hipótese demasiadamente forte. No entanto, Scharfstein, Robins e Rotnitzky (1999) mostram que trata-se de um requisito mínimo para identificação, na ausência de hipóteses sobre o *propensity score* e a distribuição conjunta de $(Y(t), X)$. Mais precisamente, é considerado um modelo de dados faltantes, em que se observa (YT, X, T), isto é, o valor de Y é conhecido somente quando $T = 1$. A distribuição conjunta dos dados completos (Y, X), $F_{X,Y}(.,.)$, é irrestrita, e a probabilidade de observação, $P(T = 1|Y, X)$, é representada pelo produto entre um componente arbitrário mas dependente apenas de X, $\lambda(X)$, e uma função potencialmente dependente de Y, $r(Y, X; \alpha_0)$, contida numa família paramétrica $\{r(.,.;\alpha)|\alpha \in A\}$[5]. Demonstra-se que neste modelo nem a distribuição $F_{X,Y}(.,.)$, nem os componentes da probabilidade de observação $\lambda(.)$ e α_0 são identificados. Adicionalmente, fixo um valor arbitrário $\tilde{\alpha} \in A$ para o componente paramétrico do modelo, qualquer distribuição de (YT, X, T) (satisfazendo determinadas condições de regularidade) pode ser exatamente reproduzida mediante alguma combinação dos componentes não paramétricos. Conseqüentemente, (i) na ausência de hipóteses sobre os dados completos $(F_{X,Y}(.,.))$ ou sobre a relação entre variáveis observadas e a seleção $(\lambda(X))$, a identificação exige que se fixe uma hipótese sobre a relação entre a variável potencialmente não observada e a probalidade de observação, ou seja, que se fixe um α, e (ii) é impossível inferir α_0, com base nos dados observáveis, isto é, não se pode rejeitar qualquer valor possível para este em um teste estatístico.

A dificuldade em justificar a hipótese de Ignorabilidade deve, portanto, ser contraposta ao poder de identificação que ela proporciona. Neste sentido, é útil considerar algumas alternativas para identificação. Manski (1990, 2003) mostra que, com hipóteses mais fracas, é possível, a partir da distribuição dos dados observáveis, delimitar um intervalo ao qual o efeito de tratamento pode pertencer. Outra possibilidade, o uso de variáveis intrumentais (Imbens e Angrist, 1994), requer a observação de instrumentos, que induzem exogenamente a participação de determinadas unidades. A estratégia permite estimar o efeito médio de tratamento para esta subpopulação, sem restringir a relação entre valores potenciais e participação, mas a condição de exogeneidade imposta aos instrumentos é, em geral, forte. Algumas alternativas se voltam para a identificação do Efeito de Tratamento em subpopulações ainda mais específicas. A abordagem de 'Regression Discontinuity' (Hahn, Todd e Van der Klaauw, 2000), também permite identificação sob heterogeneidade arbitrária, mas apenas para observações próximas a um ponto, na distribuição de uma determi-

[5]O modelo considerado pelos autores é um pouco mais complicado, admitindo que as variáveis X e T evoluam ao longo de um intervalo de tempo, mas contém a versão reproduzida aqui como caso particular.

nada variável, em que há descontinuidade na probabilidade de seleção para o tratamento. Métodos envolvendo dados de mais de um período permitem controlar para fatores não observados afetando os valores potenciais, identificando o Efeito de Tratamento para o grupo das unidades que mudaram de situação de tratamento.

2.5 Limite de Eficiência Semiparamétrico

Isoladamente, a Ignorabilidade Forte define um modelo semiparamétrico para os dados, pois o conjunto de distribuições possíveis, embora limitado pela hipótese, não pode ser parametrizado por um conjunto de dimensão finita. Logo, ainda que o interesse seja a estimação do objeto de dimensão finita β, não se aplica o limite inferior de Cramer-Rao como tradicionalmente definido. No entanto, um conceito análogo, proposto inicialmente por Stein (1956) e desenvolvido, entre outros, por Bickel et al. (1993), permite estabelecer a máxima precisão que uma determinada classe de estimadores semiparamétricos pode atingir.

A idéia por trás do limite de eficiência semiparamétrico é considerar a estimação em certos submodelos paramétricos, i.e., famílias $(\Theta, \{f(z;\theta); \theta \in \Theta\})$ de distribuições dos dados, parametrizadas por Θ de dimensão finita, tais $f(z;\theta)$ satisfaz as restrições semiparamétricas para todo $\theta \in \Theta$, e, para algum $\theta_0 \in \Theta$, $f(z;\theta_0)$ é a distribição verdadeira. Para cada um destes θ_0 o limite de Cramer-Rao, se bem definido, deve ser menor que a variância de um estimador válido no modelo semiparamétrico. Formalmente, o limite de eficiência semiparamétrico é definido pelo supremo dos limites de eficiência dos submodelos paramétricos regulares (v. definição em Newey, 1990).

O limite de eficiência semiparamétrico se aplica aos estimadores regulares. Um estimador é dito regular se, para todo submodelo paramétrico regular $(\Theta, \{f(z;\theta); \theta \in \Theta\})$, e toda seqüência $(\theta_n) \subseteq \Theta$ tal que $\sqrt{n}(\theta_n - \theta_0)$ é limitado, a seqüência das distribuições de $\sqrt{n}(\hat{\beta} - \beta(\theta_n))$ sob θ_n converge para um mesmo limite. Esta classe exclui os chamados 'estimadores super-eficientes' e aqueles que utilizam mais informação que a contida no modelo semiparamétrico. Por outro lado, regularidade é um requerimento mais brando que convergência (mesmo localmente) uniforme em distribuição[6] . Logo, geralmente, as aproximações implicadas pelas propriedades assintóticas dos estimadores regulares dependem de tamanhos de amostra desconhecidos.

[6]Bickel et al. (1993) mostram que a existência de estimadores uniformemente convergentes impõe restrições fortes demais sobre os modelos, o que inviabiliza os semiparamétricos em geral.

Seguindo o estudo de Newey (1990), um parâmetro é dito diferenciável se (i) é diferenciável com respeito aos parâmetros de qualquer submodelo paramétrico suave, e (ii) existe uma função d, de variância finita, satisfazendo, para qualquer submodelo paramétrico regular $(\Theta, \{f(z;\theta); \theta \in \Theta\})$:

$$\frac{\partial \beta(\theta_0)}{\partial \theta} = E\left[dS'_\theta\right], \tag{2-6}$$

a chamada equação de diferenciabilidade por caminhos, onde S_θ é o escore (derivada da log-verossimilhança de uma única observação) do submodelo. Para um parâmetro diferenciável, Newey mostra que o limite de eficiência é dado por $E[\delta\delta']$, onde δ é a projeção de d (componente a componente) no espaço linear fechado gerado pelos escores, também chamado espaço tangente.

Como δ tem média zero, $E[\delta\delta']$ é a sua variância. Logo, um estimador asssintoticamente linear com função influência δ, é assintoticamente eficiente. Por esta razão, esta é chamada a função influência eficiente do modelo.

Hahn(1998) mostra que, sob Ignorabilidade Forte, o Efeito Médio de Tratamento é parâmetro diferenciável, com derivada

$$\frac{\partial \beta(\theta_0)}{\partial \theta} = E\Bigg[\Bigg(m_1(X) - m_0(X) - \beta + \\ \frac{T}{p(X)}(Y - m_1(X)) + \frac{1-T}{1-p(X)}(Y - m_0(X))\Bigg)S_\theta\Bigg], \tag{2-7}$$

enquanto o espaço tangente é constituído por funções do tipo

$$a(X)(T - p(X)) + b(X) + Ts_1(Y,X) + (1-T)s_0(Y,X), \tag{2-8}$$

onde $a(.)$ é uma função quadrado-integrável arbitrária, $b(X)$ tem média zero sobre a distribuição verdadeira de X, e $s_t(Y,X)$, $t = 0,1$, têm média zero sobre a distribuição de Y condicional a qualquer valor de X. Verifica-se que a expressão para d implicada por (2-7) pertence ao espaço tangente. Logo, coincide com sua projeção, e determina a função influência eficiente

$$\psi^* = m_1(X) - m_0(X) - \beta + \frac{T}{p(X)}(Y - m_1(X)) - \frac{1-T}{1-p(X)}(Y - m_0(X)). \tag{2-9}$$

e o limite de eficiência

$$E[\psi^{*2}] = E\left[(m_1(X) - m_0(X) - \beta)^2 + \frac{\sigma_1^2(X)}{p(X)} + \frac{\sigma_0^2(X)}{1-p(X)}\right]. \tag{2-10}$$

2.6
Generalização para outros parâmetros de interesse

Como observamos na seção 2.2, uma técnica de inferência para o Efeito Médio de Tratamento estende-se facilmente para a estimação da distribuição

de cada valor potencial num ponto qualquer. Isto permite aproximar arbitrariamente bem as distribuições marginais da variável de interesse sob as diferentes condições de tratamento e realizar quase qualquer comparação entre elas. Tais aproximações, contudo, dependerão das estimativas para um grande número de pontos, o que evidentemente torna pouco prática a avaliação das propriedades de tal procedimento.

Entretanto, se pretende-se avaliar como as distribuições diferem quanto a um parâmetro definido por uma condição de momento incondicional, outra generalização pode ser empregada. Seja $\mu^* = \mu_1 - \mu_0$ a quantidade que se pretende estimar, com μ_t, $t = 0, 1$, soluções das condições de momento

$$G(\mu_t) = E[g(X, Y(t); \mu_t)] = 0,$$

onde $g(.)$ é conhecida. Como, para dados X e μ, temos que $g(X, Y(1); \mu)$ e $g(X, Y(0); \mu)$ são funções conhecidas de $Y(1)$ e $Y(0)$, a hipótese de Independência Condicional

$$T_i \perp (Y_i(1), Y_i(0)) \mid X_i$$

implica em

$$T_i \perp (g(X, Y(1); \mu), g(X, Y(0); \mu)) \mid X_i.$$

Logo, se a hipótese de Ignorabilidade é satisfeita por $(Y(1), Y(0), X, T)$, então também vale para $(g(X, Y(1); \mu), g(X, Y(0); \mu), X, T)$[7]. Assim é possível estimar, por exemplo, o lado esquerdo da condição de momento $E[g(X, Y(1); \mu] = 0$ para cada μ, e obter, pelo valor que anula esta função (estimada), uma estimativa $\hat{mu}_1$ para μ_1. Além disso, a consistência e a eficiência assintótica desse tipo de estimador têm como condição essencial que a estimação da condição de momento apresente as mesmas propriedades.

Por exemplo, podemos observar como Firpo (2007) desenvolve, utilizando um procedimento como este, um estimador para o Efeito de Tratamento por Quantil, diferença dos quantis das distribuições potenciais, que compartilha as propriedades desejáveis dos estimadores do Efeito Médio considerados no próximo capítulo. Chen, Hong e Tarozzi (2008) discutem a estimação de soluções de condições de momento bastante gerais, não necessariamente suaves.

[7]Nota-se que, trivialmente, a hipótese de Sobreposição não se altera.