4
Combinando Imputação e Reponderação

Vimos, no capítulo anterior, que as abordagens de Imputação e Reponderação permitem a estimação consistente e assintoticamente eficiente do Efeito Médio de Tratamento. Embora estas propriedades forneçam um importante argumento a favor do uso desses métodos, algumas ressalvas devem ser consideradas.

Por se tratarem de propriedades assintóticas, consistência e eficiência não permitem conclusões imediatas sobre o desempenho para uma quantidade determinada de observações. Além disto, como vimos na discussão do limite de eficiência, a classe dos estimadores regulares não exige uniformidade na convergência em distribuição. Logo, a princípio, o conceito de 'amostra suficientemente grande' depende de valores desconhecidos e, portanto, não se pode verificar se uma determinada amostra satisfaz tal requerimento. Confirmando esta razão teórica contra a orientação exclusiva pela otimalidade assintótica, estudos de simulação (a serem discutidos no próximo capítulo) apresentam resultados distintos sobre o desempenho dos diversos métodos 'eficientes'.

Desta forma, no atual estágio de desenvolvimento da literatura, o uso das implementações tradicionais da Imputação e da Reponderação não é recomendado na prática, sendo preferível o recurso a alguma combinação desses métodos, tipicamente interpretada como forma de reduzir o viés (Imbens e Wooldridge, 2009). Uma forma de fazê-lo é sugerida pelas técnicas de estimação duplamente robusta, associada a James Robins e seus coautores, discutida a seguir, neste capítulo.

4.1
Estimação Duplamente Robusta

Métodos que combinam as estimações da probabilidade de seleção e das funções de regressão foram inicialmente propostos para o caso de modelos paramétricos para esses objetos, nos trabalhos de Robins e Rotnitzky (1995) e Robins, Rotnitzky e Zhao (1995) sobre inferência com dados faltantes. Sua finalidade, nesse contexto, é proteger contra má especificação desses modelos. Um exemplo é o estimador de reponderação 'aumentado', discutido por Scharfstein, Robins e Rotnitzky (1999):

$$\hat{\beta}_{DR} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{p(X_i;\hat{\gamma})} - \frac{(1-T_i)Y_i}{1-p(X_i;\hat{\gamma})} - (T_i - p(X_i;\hat{\gamma}))\left(\frac{m_1(X_i;\hat{\delta}_1)}{p(X_i;\hat{\gamma})} + \frac{m_0(X_i;\hat{\delta}_0)}{1-p(X_i;\hat{\gamma})}\right) \qquad (4\text{-}1)$$

onde $p(.;\hat{\gamma})$ e $m_t(.;\hat{\delta}_t)$ são, respectivamente, estimativas para $p(.)$ e $m_t(.)$ baseadas nos modelos paramétricos $\{p(.;\gamma);\gamma \in \Gamma\}$ e $\{m_t(.;\delta_t);\delta \in \Delta\}$, que devem satisfazer condições de regularidade.

É visível a analogia entre (4-1) e a função influência eficiente apresentada no Capítulo 2. De fato, trata-se de uma versão amostral de $E[\psi^*] = 0$. O termo adicional é similar à correção devida à estimação de $p(.)$. Seu efeito é tornar $\hat{\beta}_{DR}$ consistente desde que ao menos um dos modelos paramétricos esteja bem especificado, i.e., contenha o valor verdadeiro. Para verificar isto, observamos, primeiramente, que a equação populacional:

$$\beta = E\left[\frac{TY}{p(X)} - \frac{(1-T)Y}{1-p(X)}\right] - E\left[(T-p(X))(\frac{\bar{m}_1(X)}{p(X)} + \frac{\bar{m}_0(X)}{1-p(X)})\right], \qquad (4\text{-}2)$$

análoga a (4-1), é satisfeita para os valores verdadeiros de β e $p(.)$, independente das funções $\bar{m}_t(.)$. Neste caso, a segunda parcela do lado direito é nula, pois $T - p(X) = T - E[T|X]$ é o resíduo da projeção ortogonal de T sobre o espaço das funções de X. Portanto, (4-2) é idêntica a (2-5) quando $p(.)$ assume seu valor populacional. A consistência de $\hat{\beta}_{DR}$ segue, sob condições de regularidade e argumentos usuais da teoria de estimação por método dos momentos.

Para avaliar a consistência quando $m_0(.)$ e $m_1(.)$ são corretamente estimadas, é útil reescrever (4-2) da seguinte forma:

$$\beta = E\left[m_1(X) - m_0(X)\right] - E\left[\frac{T}{\bar{p}(X)}(Y - m_1(X)) + \frac{1-T}{1-\bar{p}(X)}(Y - m_0(X))\right]. \qquad (4\text{-}3)$$

Se m_0 e m_1 forem as verdadeiras funções de regressão, o segundo termo do lado direito de (4-3) é nulo para qualquer função $\bar{p}(.)$, pois

$$E\left[\frac{T}{\bar{p}(X)}(Y - m_1(X))|X\right] = \frac{1}{\bar{p}(X)}E\left[E\left[T(Y-m_1(X))|X,T\right]|X\right]$$
$$= \frac{1}{\bar{p}(X)}E[E[Y_1 - m_1(X)|X,T=1]Pr(T=1|X)|X] = 0$$

e, analogamente,

$$E\left[\frac{1-T}{1-\hat{p}(X)}(Y - m_0(X))|X\right] = 0.$$

Logo (4-3) equivale a (2-3), e recorrendo, da mesma forma que no caso anterior, a resultados bem conhecidos, é possível estabelecer a consistência de $\hat{\beta}_{DR}$. Estimadores como este, que satisfazem a consistência mesmo quando um entre dois modelos paramétricos falha, são chamados Duplamente Robustos. Além de $\hat{\beta}_{DR}$, outros estimadores com esta propriedade são conhecidos, como os apresentados por Robins et al. (2007) e Egel, Graham e Pinto (2008).

4.2
Relevância para Estimação Semiparamétrica

Claramente, um procedimento duplamente robusto é preferível a outro que dependa do funcionamento de um único modelo paramétrico pouco confiável. Isto não mostra, porém, como regressão e imputação podem ser combinadas para produzir estimadores com propriedades desejáveis quando não há modelos paramétricos críveis. Além disso, os métodos de Imputação e Reponderação semiparamétricos apresentados no capítulo anterior permitem a estimação consistente e eficiente sem impor restrições além da hipótese de identificação e condições de regularidade.

Uma motivação para a combinação de abordagens é o fato de que, dado o tamanho da amostra disponível, o estimador não paramétrico é geralmente viesado. Em particular, se o estimador em questão é do tipo *sieve*, trata-se efetivamente de um modelo paramétrico mal especificado. Então, a discussão sobre dupla robustez sugere que acrescentar uma versão amostral do termo de correção para o primeiro passo pode reduzir o viés. Esta idéia é apontada no estudo de simulações de Bang e Robins (2005), que sugerem a validade de uma versão aproximada do conceito de robustez dupla: se ao menos um dos modelos é aproximadamente correto, o viés da estimação é pequeno. Também é interessante observar, neste sentido, que as recentes discussões sobre implementação ótima dos estimadores de Imputação e Reponderação (Imbens, Newey e Ridder, 2005, Ichimura e Linton, 2005) utilizam tanto o *propensity score* quanto as funções de regressão para estimar os termos de viés e variância a serem minimizados.

Formalmente, para o caso de dados faltantes com Y binária, Robins e Ritov (1997) apontam uma razão para utilizar um estimador análogo a $\hat{\beta}_{DR}$ sob uma especificação semiparamétrica. Seguindo a argumentação dos autores, um problema dos resultados assintóticos dos estimadores semiparamétricos é a dependência de hipóteses de suavidade sobre a probabilidade de observação/*propensity score* e função de regressão. Robins e Ritov mostram que, supondo apenas mensurabilidade dessas funções, não existem estimadores convergentes a taxas algébricas, i.e, da forma $n^{-\alpha}$, $\alpha > 0$. Na demonstração deste fato são considerados modelos alternativos distintos do verdadeiro por um grande número de irregularidades em $p(X)$ e $m(X)$ introduzidas dentro de pequenos intervalos da distribuição de X. A construção destas regiões é feita de modo a limitar a probabilidade de cada uma delas conter mais de um ponto na amostra.

Esta idéia é comparada aos problemas de desempenho em amostra finita. À medida que a dimensão de X cresce, a probabilidade de observações terem valores próximos diminui rapidamente. Logo, a detecção de comportamentos locais importantes dos modelos de seleção e/ou regressão torna-se difícil. Assim, a partir do estudo das fragilidades das propriedades assintóticas, Robins e Ritov apontam para a necessidade de cautela quando a etapa não paramétrica está sujeita a esta 'maldição da dimensionalidade'. Desenvolvem então a 'Teoria Assintótica Apropriada à Maldição da Dimensionalidade' que requer uma delimitação subjetiva dos parâmetros de ruído potencialmente mal estimados, devido a alguma suspeita de baixa suavidade ou à alta dimensão. A estimação do parâmetro de interesse satisfaz esta teoria quando se baseia em momentos não viesados sob possível má especificação dos parâmetros de ruído suspeitos.

Para o modelo de dados faltantes, a equação que iguala a função influência a zero é adequada tanto para o caso de 'seleção bem estimada'/'regressão mal estimada' quanto para o oposto. Robins e Ritov mostram ser então possível construir um estimador RAL em ambos os casos. Adicionalmente, no primeiro, o estimador também é Uniformemente Assintoticamente Normal Não Viesado (UANU). Esta última propriedade é definida pela existência de uma sequência $s_n(F)$, tal que

$$\sup_F |Pr_F\left[N^{1/2}(\hat{\beta} - \beta(F))/s_N(F) < t\right] - \phi(t)| \to 0,$$

onde F indexa as distribuições admissíveis no modelo. É uma condição mais fraca do que o estimador ser Uniformemente Regular Gaussiano, mas suficiente para construir intervalos de confiança uniformemente assintoticamente válidos, com s_N estimado por *bootstrap*. O estimador corresponde à expressão (4-1), para $Y(0) \equiv 0$, mas com estimadores (não paramétricos) de histograma para $p(.)$ e $m(.) = m_1(.)$, em que cada uma das estimações preliminares e o passo final são feitos a partir de sub-amostras independentes.

4.3 Estimadores Semiparamétricos Duplamente Robustos

Em vista da discussão acima, propomos combinações das técnicas de Imputação e Reponderação em estimadores que: (i) sejam baseados na condição de momento $E[\psi^*] = 0$, que como vimos na seção 4.1, se verifica quando ou $m_1(.)$ e $m_0(.)$, ou $p(.)$ assumem seus valores verdadeiros; e, ao mesmo tempo, (ii) envolvam a estimação não paramétrica dessas funções. Desta forma, pretendemos obter, além de uma inferência consistente no modelo semiparamétrico em que apenas se supõe Ignorabilidade, também um melhor desempenho em amostras finitas, dada a menor vulnerabilidade a parâmetros de ruído pouco suaves.

A primeira sugestão que analisaremos é utilizar estimadores não paramétricos num procedimento de reponderação aumentado como em $\hat{\beta}_{DR}$:

$$\hat{\beta}_{FI} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} - (T_i - \hat{p}(X_i))\left(\frac{\hat{m}_1(X_i)}{\hat{p}(X_i)} + \frac{\hat{m}_0(X_i)}{1-\hat{p}(X_i)}\right), \tag{4-4}$$

onde $\hat{p}(.)$ e $\hat{m}_t(.)$ são estimativas por *sieve*/logit e *sieve*/mínimos quadrados, respectivamente. Embora sugerido em alguns estudos, este procedimento não foi explicitamente estudado até recentemente. Cattaneo (2007) analisa as propriedades de um estimador análogo a $\hat{\beta}_{FI}$ no contexto da estimação do efeito médio de um tratamento de múltiplos níveis.

Um fato interessante sobre $\hat{\beta}_{FI}$ é que ele pode ser representado como combinação linear de três estimadores bem conhecidos. Reescrevendo (4-4) de forma conveniente, obtemos:

$$\begin{aligned}\hat{\beta}_{FI} = &\frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)}\\ &+\frac{1}{N}\sum_{i=1}^{N}\hat{m}_1(X_i) - \hat{m}_0(X_i)\\ &-\frac{1}{N}\sum_{i=1}^{N}\frac{T_i\hat{m}_1(X_i)}{\hat{p}(X_i)} - \frac{(1-T_i)\hat{m}_0(X_i)}{1-\hat{p}(X_i)},\end{aligned}$$

o que correponde exatamente à soma dos estimadores de Imputação e Reponderação, subtraída do estimador 'modificado' de Imbens, Newey e Ridder (2005). Esta representação permite estabelecer, imediatamente, condições para consistência e eficiência assintótica. Imbens, Newey e Ridder (2005) demonstram, sob hipóteses de diferenciabilidade, que os três estimadores são assintoticamente lineares com função influência eficiente. Nestas condições, portanto:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_{FI} - \beta) &= \sqrt{n}(\hat{\beta}_{rep} + \hat{\beta}_{imp} - \hat{\beta}_m - \beta) \\
&= \sqrt{n}(\hat{\beta}_{rep} - \beta) + \sqrt{n}(\hat{\beta}_{imp} - \beta) - \sqrt{n}(\hat{\beta}_m - \beta) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \psi^*(Z_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \psi^*(Z_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \psi^*(Z_i) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \psi^*(Z_i) + o_p(1),
\end{aligned}
$$

e, portanto, $\hat{\beta}_{FI}$ também é assintoticamente eficiente. Desta forma, concluímos que, no contexto semiparamétrico considerado neste trabalho, estimadores duplamente robustos não são necessariamente ineficientes, ao contrário do que ocorre em modelos paramétricos onde a robustez contra a violação das hipóteses de especificação é obtida ao custo da perda de eficiência sob a validade das mesmas[1]. Cattaneo (2007) apresenta outra demonstração de eficiência para este estimador, e observa que as condições necessárias são mais fracas, pois a semelhança entre a condição de momento estimada e a função influência eficiente implica que a estimação dos parâmetros de ruído não precisa deixar os 'restos' de ordem $\frac{1}{\sqrt{N}}$ vistos na seção 3.3. É interessante observar, portanto, que procedimentos duplamente robustos podem oferecer uma maneira simples de se aproveitar informações adicionais sobre o *propensity score*.

Outra forma de se obter um procedimento duplamente robusto é através da escolha adequada do estimador preliminar em um procedimento de Imputação, como apontado por Robins et al. (2007). Neste sentido, observamos que um estimador derivado da condição $E[\psi^*] = 0$ tem a forma

$$
\begin{aligned}
\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} & \hat{m}_1(X_i) - \hat{m}_0(X_i) \\
& - \left(\frac{T_i(Y_i - \hat{m}_1(X_i))}{\hat{p}(X_i)} - \frac{(1 - T_i)(Y_i - \hat{m}_0(X_i))}{1 - \hat{p}(X_i)} \right).
\end{aligned}
$$

Verifica-se que $\hat{\beta}$ se distingüe de um estimador de Regressão pelo termo entre parênteses, correspondente à correção devida ao uso dos parâmetros de ruído $\hat{m}_t$ estimados. Nota-se ainda que este é composto por produtos dos pesos da Reponderação com os resíduos da estimação das funções de regressão. Logo, se

[1]Deve ficar claro que nos referimos a limites de eficiência distintos. Afirmamos que um estimador duplamente robusto baseado em dadas especificações paramétricas para $m_t(.)$ e $p(.)$ não atinge o limite de eficiência do modelo em que estas são válidas. Entretanto, existem estimadores desse tipo, como o de Egel, Graham e Pinto (2008) que atingem o limite de eficiência semiparamétrico quando verificam-se ambas as restrições funcionais. Estes são ditos localmente eficientes.

o método de estimação dessas funções produzir resíduos ortogonais aos pesos, $\hat{\beta}$ será algebricamente equivalente a um estimador de Regressão.

Isto pode ser realizado pela inclusão dos pesos como regressores. Neste caso, as equações normais implicam diretamente na ortogonalidade requerida. Este procedimento, no entanto, não é recomendável, pois a distribuição dos pesos, por construção, se concentra em pontos diferentes entre os grupos. Por exemplo, $\hat{p}(X_i)$ assumirá valores mais altos no grupo de tratamento, e possivelmente muito baixos no grupo de controle. Logo, incluir o regressor $\frac{1}{\hat{p}(X_i)}$ pode produzir um caso grave de extrapolação.

O termo de correção também pode ser anulado se substituirmos $\hat{m}_1(.)$ e $\hat{m}_0(.)$, pelos estimadores *sieve*/mínimos quadrados ponderados $\tilde{m}_1(.;\hat{p}(.))$ e $\tilde{m}_0(.;\hat{p}(.))$, que utilizam, respectivamente, $\frac{1}{\sqrt{\hat{p}(X_i)}}$ e $\frac{1}{\sqrt{1-\hat{p}(X_i)}}$ como pesos. Se a base de funções escolhida contém uma constante, a equação normal para o coeficiente correspondente é, no cálculo de $\tilde{m}_1(.;\hat{p}(.))$:

$$-\frac{2}{N_1}\sum_{i=1}^{N_1}\frac{1}{\hat{p}(X_i)}(Y_i - \tilde{m}_1(X_i;\hat{p}(.))) = 0$$

ou

$$\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}(Y_i - \tilde{m}_1(X_i;\hat{p}(.))) = 0,$$

valendo uma equação análoga para $\tilde{m}_0(.;\hat{p}(.))$. Neste caso, definimos o estimador de Regressão Ponderada:

$$\hat{\beta}_{RP} = \frac{1}{N}\sum_{i=1}^{N}\tilde{m}_1(X_i;\hat{p}(.)) - \tilde{m}_0(X_i;\hat{p}(.)), \tag{4-5}$$

onde $\tilde{m}_t(.;\hat{p}(.))$ são obtidos pelos coeficientes da projeção de mínimos quadrados ponderados de Y em funções de X nas sub-amostras, com os pesos mencionados acima.

Este estimador foi implementado por Hirano e Imbens (2001)[2], que, além disso, propuseram um algoritmo para seleção dos estimadores preliminares. O procedimento consiste em estabelecer um conjunto grande de variáveis (possivelmente incluindo transformações e interações) e um par de valores reais positivos (t_{prop}, t_{reg}). Cada variável é então utilizada em regressões logísticas simples do indicador de tratamento, e são observadas as estatísticas-t do teste com a hipótese nula de que o coeficiente associado é zero. São utilizadas, na estimação de $p(.)$ as variáveis cujo valor absoluto da estatística-t

[2]Anteriormente, Robins, Rotnitzky e Zhao (1995), haviam considerado um estimador próximo, envolvendo a estimação não paramétrica do *propensity score*, mas uma especificação paramétrica para as funções de regressão.

supera t_{prop}. De maneira semelhante, cada variável do conjunto mais amplo, é utilizada numa regressão de Y_i juntamente com T_i. São incluídas nas estimações de $m_1(.)$ e $m_0(.)$ aquelas com estatísticas-t (do teste da nulidade de seu coeficiente) de valor absoluto maior que t_{reg}. Os autores observam ainda que este procedimento inclui como casos particulares os estimadores de Imputação e Reponderação, quando, respectivamente, na estimação de $p(.)$ e $m_t(.)$ são utilizadas apenas constantes.

O estimador $\hat{\beta}_{RP}$ também permite uma nova interpretação para a propriedade de robustez dupla, baseada na idéia de omissão de variáveis numa regressão (conforme notado por Imbens e Wooldridge, 2009). Enquanto a Imputação corresponde a controlar diretamente para a variável X, a Reponderação, ao equilibrar a distribuição de X na amostra, elimina sua correlação com T, de modo que sua omissão não mais causa viés. Quando não se sabe exatamente a correta especificação do efeito do controle X, o estimador duplamente robusto, recorrendo às duas formas de correção, aumenta as chances de uma inferência com baixo viés.