São Paulo School of ECONOMICS

Working Paper

3/2013

Jan 2013

CENTER FOR APPLIED MICROECONOMICS

Combinando Estratégias para Estimação de Efeitos de Tratamento

Sergio Firpo
Rafael de Carvalho Cayres Pinto

Combinando Estratégias para Estimação de Efeitos de Tratamento

Sergio Firpo [Sao Paulo School of Economics, FGV]
Rafael de Carvalho Cayres Pinto [PUC-Rio, Depto. de Economia]

Janeiro 2013

Abstract

Uma ferramenta importante na avaliação de políticas econômicas é a estimação do efeito médio de um programa ou tratamento sobre uma variável de interesse. Em geral, a atribuição do tratamento aos potenciais participantes não é aleatória, o que pode causar viés de seleção quando desconsiderada. Uma maneira de resolver esse problema é supor que o econometrista observa um conjunto de características determinantes, a menos de um componente estritamente aleatório, da participação. Sob esta hipótese, existem na literatura estimadores semiparamétricos do efeito médio do tratamento que são consistentes e capazes de atingir, assintoticamente, o limite de eficiência semiparamétrico. Entretanto, nas amostras freqüentemente disponíveis, o desempenho desses métodos nem sempre é satisfatório. O objetivo deste trabalho é estudar como a combinação das duas estratégias pode produzir estimadores com melhores propriedades em amostras pequenas. Para isto, consideramos duas formas de integrar essas abordagens, tendo como referencial teórico a literatura de estimação duplamente robusta desenvolvida por James Robins e co-autores. Analisamos suas propriedades e discutimos por que podem superar o uso isolado de cada uma das técnicas que os compõem. Finalmente, comparamos, num exercício de Monte Carlo, o desempenho desses estimadores com os de imputação e reponderação. Os resultados mostram que a combinação de estratégias pode reduzir o viés e a variância, mas isso depende da forma como é implementada. Concluímos que a escolha dos parâmetros de suavização é decisiva para o desempenho da estimação em amostras de tamanho moderado.

Keywords: Efeito Médio de Tratamento; Ignorabilidade; Imputação; Reponderação; Estimação Duplamente Robusta; Método de Monte Carlo.

1 Introdução

A identificação do efeito causal de um tratamento ou programa sobre uma variável de interesse é um dos temas principais da literatura econométrica, sendo parte essencial da avaliação de políticas públicas como intervenções ativas no mercado de trabalho (ver, por exemplo, os trabalhos de Heckman, Ichimura e Todd (1997), e Heckman et al. (1998)). A preocupação central no estudo deste problema é a relação entre os componentes não observados que determinam os resultados e aqueles que afetam a participação no programa. Assim, na impossibilidade de realização de experimentos aleatórios, deve haver auto-seleção para o tratamento, que em geral segue padrões desconhecidos pelo econometrista, ocasionando o fenômeno conhecido como viés de seleção, devido ao efeito de tratamento individual ser diferente entre as unidades afetadas e as não afetadas.

Nestas condições, alguma suposição adicional deve ser feita para identificar um parâmetro de interesse. A hipótese de Ignorabilidade (Rosenbaum e Rubin, 1983) afirma que toda a informação relevante sobre a heterogeneidade pode ser captada por variáveis auxiliares observadas para todas as unidades. Em outras palavras, não há viés de seleção sistemático quando comparam-se indivíduos semelhantes quanto a determinadas características. Considerada isoladamente, esta hipótese define um modelo semiparamétrico para a população, ou seja, impõe que a distribuição de probabilidade que a descreve pertença a uma dada classe a qual, embora seja um subconjunto próprio do universo de todas as medidas de probabilidade, é ampla a ponto de não poder ser indexada por um parâmetro de dimensão finita. Quando há interesse na inferência do efeito de tratamento sobre toda a população, com dados não experimentais, este é o principal modelo considerado na literatura que permite identificação exata de um parâmetro.

A teoria sobre estimação semiparamétrica do Efeito de Tratamento sob Ignorabilidade em grandes amostras encontra-se num estágio de desenvolvimento avançado. Uma variedade de métodos bastante distintos entre si foi profundamente estudada, e, para cada um, foram desenvolvidas condições sob as quais a estimação assintoticamente eficiente é assegurada. Dois dos mais importantes métodos são a Imputação (ou Regressão) e a Reponderação. No primeiro, os dados de cada grupo são usados para estimar a relação entre os valores potenciais e as variáveis auxiliares, a chamada função de regressão; em seguida estas estimativas são usadas como substitutas do valor potencial não observado no outro grupo. O segundo método envolve estimar a relação entre as variáveis auxiliares e a seleção para o tratamento, dada pela probabilidade desta condicionada àquelas, conhecida como *propensity score*. Esta informação descreve a representação relativa dos grupos para cada valor das características auxiliares, permitindo uma reponderação que torna a amostra representativa de uma população na qual a participação foi atribuída aleatoriamente.

Uma peculiaridade da teoria assintótica para alguns problemas semiparamétricos, inclusive este, é que os procedimentos são equivalentes entre si, desde que condições de regularidade relativamente fracas sejam satisfeitas (Newey, 1994). Questões práticas, entretanto, motivam o interesse no desempenho da inferência em amostras pequenas. A fragilidade da relação entre a teoria, predominantemente assintótica, e as propriedades em amostras finitas motivou uma série de estudos empregando simulações. Estes conseguem apontar a importância da forma de implementação dos métodos, além dos méritos relativos de cada um. Em particular, uma questão apresentada recentemente é o potencial benefício em se integrar diferentes técnicas em procedimentos que compartilhem o bom desempenho de cada uma delas em certas circunstâncias.

Este trabalho visa a um melhor entendimento das possibilidades oferecidas pela combinação das técnicas de Imputação e Reponderação. Para atingir este objetivo, foram realizados dois tipos de esforço. Por um lado, razões teóricas para a superioridade da abordagem mista, ligadas à literatura de estimação duplamente robusta (Robins e Rotnitzky, 1995, Robins, Rotnitzky e Zhao , 1995), são discutidas. É importante ressaltar, entretanto, que a motivação principal para a combinação em nosso contexto se distingue daquela considerada em grande parte dessa produção científica. De fato, a inferência duplamente robusta tem sido aplicada, na maioria dos estudos, para combinar estimadores paramétricos de Imputação e Reponderação de modo a assegurar consistência, desde que a especificação usada em ao menos um deles descreva corretamente o modelo populacional. No presente estudo, pelo contrário, partimos do modelo semiparamétrico onde apenas se supõe Ignorabilidade, e cada abordagem, isoladamente, produz estimadores consistentes sob condições gerais. Procuramos então melhorar o desempenho em amostras finitas pela combinação de Imputação e Reponderação semiparamétricas, possibilidade que, embora apoiada pelo trabalho de Robins e Ritov (1997), permanece pouco explorada na literatura. Com base nessa discussão, propomos dois procedimentos Duplamente Robustos. O primeiro é uma generalização direta do estimador de Scharfstein, Robins e Rotnitsky (1999), na qual usamos estimativas preliminares do tipo *sieve* no lugar das paramétricas. Este método, que até recentemente não havia sido considerado de forma explícita na literatura, coincide com o estudado por Cattaneo (2007). O outro consiste num procedimento semiparamétrico de Imputação no qual estimativas das funções de regressão são obtidas por mínimos quadrados ponderados pelo inverso do *propensity score* estimado (ou de seu complemento). Este estimador foi implementado por Hirano e Imbens (2001), mas carece de um estudo detalhado de suas propriedades.

Pelo outro lado, de maneira complementar à análise teórica, são realizadas simulações de Monte Carlo comparando diferentes implementações de métodos de Imputação, Reponderação, e combinações destes na forma dos estimadores Duplamente Robustos propostos. Como forma de avaliar a relevância destes últimos, bem como testar previsões e sugestões da teoria, este exercício é reproduzido sob diversos modelos populacionais. As especificações consideradas se distinguem quanto à forma funcional do *propensity score* e das funções de regressão, a heterocedasticidade dos valores potenciais e a dimensão do conjunto de variáveis auxiliares. Quanto ao primeiro atributo, buscamos variar a suavidade do modelo de uma maneira sistemática, apoiada em um conceito 'função suave' relevante para a teoria, evitando assim manipulações *ad hoc*, freqüentes em estudos de simulação. Da mesma forma, a heterocedasticidade foi introduzida de modo a gerar cenários extremamente favoráveis e desfavoráveis para o estimador Duplamente Robusto baseado em Imputação, conforme se alinha ou se opõe a ponderação empregada neste método àquela que otimiza a estimação das funções de regressão.

Os resultados mostram que a combinação de Imputação e Reponderação em procedimentos duplamente robustos permite a redução do erro quadrático médio mesmo para modelos desfavoráveis. Adicionalmente, corroborando a análise teórica, o ganho de eficiência possibilitado pelos estimadores Duplamente Robustos é maior em modelos menos suaves e quando as variáveis auxiliares são multidimensionais. Mudanças na heterocedasticidade, por sua vez, não produzem efeito sobre a vantagem da integração de técnicas, diferentemente do esperado. Por fim, a combinação da estimação de funções de regressão com a ponderação por funções do *propensity score* exato mostrou-se um meio efetivo, na maior parte dos modelos, de aproveitamento do conhecimento prévio desta função.

2 Identificação do Efeito de Tratamento sob Ignorabilidade

Nesta seção, 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 nesta seção 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

[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.

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 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

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

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]:

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

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

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

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

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:

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

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:

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

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, 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

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

onde, a primeira expressão existe, por (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,

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

uma representação do parâmetro de interesse como diferença de médias ponderadas de valores observados. A equação 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.

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 paramétri-

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

[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.

cos. 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 determinada 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.

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 $(\Theta, \{f(z;\theta); \theta \in \Theta\})$:

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

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

[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.

parâmetro diferenciável, com derivada

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

enquanto o espaço tangente é constituído por funções do tipo

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

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 (7) pertence ao espaço tangente. Logo, coincide com sua projeção, e determina a função influência eficiente

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

e o limite de eficiência

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

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 na próxima seção. Chen, Hong e Tarozzi (2008) discutem a estimação de soluções de condições de momento bastante gerais, não necessariamente suaves.

3 Estimação por Imputação e Reponderação

Nesta seção serão discutidos dois dos métodos que têm recebido maior atenção da literatura de estimação de efeito de tratamento sob ignorabilidade[8].

O primeiro, consiste na imputação do valor estimado da função de regressão no lugar dos valores potenciais, viabilizando a estimação do Efeito Médio de Tratamento por uma média de diferenças. O outro envolve a comparação dos valores médios da variável de interesse entre os grupos, sendo estes devidamente reponderados de forma a representar a população.

Apesar das inspirações distintas, as duas abordagens levam a estimadores que podem ser analisados através de uma teoria geral como a desenvolvida em Newey (1994).

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

[8]Na resenha de Imbens e Wooldridge (2009), os principais métodos são classificados em três grupos: regressão, métodos baseados no *propensity score*, e *matching*. Neste trabalho são abordadas a técnica de regressão e a de reponderação, que se enquadra no segundo grupo. Os métodos de *matching*, ineficientes em geral, não serão discutidos aqui por não possuirem ligação próxima com a proposta deste estudo.

Esta teoria garante que ambas possuem boas propriedades assintóticas, sob hipóteses adicionais de suavidade das funções envolvidas e regularidade do modelo probabilístico subjacente.

3.1 Regressão/Imputação

Uma abordagem para a estimação do Efeito Médio de Tratamento é motivada pela equação (3), a qual pode ser interpretada como uma condição de momento associada a funções que podem ser estimadas com os dados disponíveis. Isto sugere que estimemos β por:

$$\hat{\beta}_{imp} = \frac{1}{N} \sum_{i=1}^{N} \hat{m}_1(X_i) - \hat{m}_0(X_i), \tag{11}$$

onde $\hat{m}_t(.)$ é uma estimativa para $m_t(.)$. O procedimento envolve, portanto, dois passos: no primeiro, estimar as funções de regressão; no segundo, integrar a diferença entre as estimativas sobre a distribuição empírica de X_i. Intuitivamente, este estimador substitui a diferença entre valores potenciais da variável de interesse, que requer o uso de dados indisponíveis, pela diferença entre valores das funções de regressão, que podem ser estimados e imputados.

O método de imputação foi implementado por Rubin (1977), num contexto em que $m_1(.)$ e $m_0(.)$ seguem um modelo paramétrico previamente conhecido. Mais recentemente, foi estudada a implementação no caso de interesse para este trabalho, em que as formas funcionais das funções de regressão são desconhecidas. Heckman, Ichimura e Todd (1997), e Heckman, Ichimura, Smith e Todd (1998) empregaram métodos de *kernel* para a primeira etapa.

Outra possibilidade é obter $\hat{m}_t$ por um método de *sieves* (Grenander (1981), Geman e Hwang (1982)), que consiste em estimar modelos paramétricos, porém sucessivamente mais abrangentes quanto maior a amostra. Um exemplo é a estimação por *sieve*/mínimos quadrados, em que se considera uma seqüência crescente de bases $B_K = \{q_k(X), k = 1, 2, ..., K\}$ de funções de X, e, para cada tamanho de amostra n, $\hat{m}_t(X_i) = q^{k(n)}(X_i)'\hat{\gamma}_{k(n),t}$, onde $\hat{\gamma}_{k,t}$ são os coeficientes da projeção ortogonal de Y sobre $q_1(X), ..., q_{k(n)}(X)$ na sub-amostra $\{(Y_i, X_i, T_i); T_i = t\}$. Ou seja, introduzindo a notação $(X_i^1, Y_i^1)_{i=1}^{N_1}$, para a subamostra das unidades tratadas, e $(X_i^0, Y_i^0)_{i=1}^{N_0}$, das de controle,

$$
\begin{aligned}
q^k(X) &= (q_1(X), ..., q_k(X)) \\
\hat{\gamma}_{k,t} &= \left(Q^{k,t\prime k,t}\right)^{-1} Q^{k,t\prime} Y^t \\
Q^{k,t} &= (q^k(X_1^t)^{\prime k}(X_{n_t}^t)^\prime).
\end{aligned}
$$

Como requisito mínimo para a convergência de $\hat{m}_t$, $m_t(.)$ deve ser aproximada arbitrariamente bem no espaço gerado por B_K, para K suficientemente grande, e a dimensão $k(n)$ do espaço de projeção deve crescer arbitrariamente com, porém menos que proporcionalmente a, o tamanho da amostra.

Esta técnica foi adotada por Chen, Hong e Tamer (2005), num problema de estimação similar ao nosso[9]. Imbens, Newey e Ridder (2005) estudaram a versão do mesmo estimador na qual as B_K são bases de polinômios, no contexto de estimação de Efeito Médio de Tratamento sob Ignorabilidade.

3.2 Reponderação

Um grupo de métodos alternativos à Regressão/Imputação se desenvolveu ligado ao trabalho de Rosenbaum e Rubin (1983, 1984), que mostraram que o *propensity score* contém a informação necessária sobre a diferença entre os grupos de tratamento e controle. Formalmente, Rosenbaum e Rubin demonstram que, sob Ignorabilidade:

$$T \perp (Y(0), Y(1)) \mid p(X), \tag{12}$$

e que, portanto, não há viés em se comparar unidades com o mesmo *propensity score*. Esta observação motivou a busca por estimadores baseados nessa função, evitando métodos de imputação que, por envolverem uma estimação de dimensão alta na primeira etapa, freqüentemente apresentavam dificuldades computacionais e mau desempenho em amostras pequenas.

De particular interesse para este trabalho é o método que utiliza o *propensity score* numa reponderação das unidades, como na segunda fórmula de identificação, (5). A reponderação pode ser interpretada como uma forma de equilibrar a distribuição de X

[9] Chen, Hong e Tamer (2005) consideram o problema de estimar a solução de uma condição de momento envolvendo variáveis latentes, X^*, observadas apenas numa amostra auxiliar. A hipótese de identificação, de que a distribuição de X^* condicional às variáveis sempre observadas X independe de a observação estar na amostra auxiliar, é análoga à Ignorabilidade, com $X^* = Y(t)$ e as unidades que recebem $T = t$ correspondendo à amostra auxiliar na estimação de $E[Y(t)]$.

entre os grupos na amostra.

Procedimentos de reponderação foram inicialmente empregados em contextos nos quais o *propensity score* era conhecido. Horwitz e Tompson (1952), freqüentemente apontados como precursores da técnica, utilizaram a ponderação pelo inverso da probalidade de seleção em amostras estratificadas.

No caso da estimação do Efeito Médio de Tratamento com dados observacionais, o *propensity score* geralmente é desconhecido, e portanto deve ser estimado. Por analogia à equação (5), um estimador de Reponderação tem a forma:

$$\hat{\beta}_{rep} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)}, \tag{13}$$

ou

$$\tilde{\beta}_{rep} = \frac{\sum_{i=1}^{N} T_iY_i/\hat{p}(X_i)}{\sum_{i=1}^{N} T_i/\hat{p}(X_i)} - \frac{\sum_{i=1}^{N}(1-T_i)Y_i/(1-\hat{p}(X_i))}{\sum_{i=1}^{N}(1-T_i)/(1-\hat{p}(X_i))}, \tag{14}$$

onde $\hat{p}(.)$ é uma estimativa do *propensity score*. A diferença entre os dois estimadores é que $\tilde{\beta}_{rep}$ impõe que a soma dos pesos valha um. Se uma normalização dos pesos for incluída no primeiro passo, pela forma como $\hat{p}(.)$ é estimado, temos $\hat{\beta}_{rep} = \tilde{\beta}_{rep}$.

Como o *propensity score* corresponde a uma média condicional, pode ser estimado por métodos de *sieve*/mínimos quadrados, da mesma forma que as funções de regressão no estimador de Imputação discutido acima. As propriedades de um estimador deste tipo foram estudadas por Hirano, Imbens e Ridder (2000), com bases polinomiais e, mais tarde, por Chen, Hong e Tarozzi (2008), de forma mais geral.

No entanto, o emprego do método de mínimos quadrados permite que os valores estimados $\hat{p}(X_i)$ não pertençam ao intervalo unitário, e portanto se produzam pesos negativos[10]. Devido a esta desvantagem, Hirano, Imbens e Ridder (2003) propõem estimar $p(.)$ por um método *sieve* alternativo. No lugar da projeção por mínimos quadrados, utilizam a estimação por máxima verossimilhança da regressão logística de T nas funções da base $B_{k(n)}$, ou seja, $\hat{p}(X_i) = L(q^{k(n)}(X_i)'\gamma_{k(n)})$, onde $L(.)$ é a função logística e

$$\hat{\gamma}_k = \arg\max_{\gamma\in\mathbb{R}^k}\sum_{i=1}^{N} T_i\log(L(q^k(X_i)'\gamma)) + (1-T_i)(1-\log(L(q^k(X_i)'\gamma)).$$

Um fato interessante, observado em diversos trabalhos sobre o assunto (ver, por ex-

[10]Uma crítica semelhante se aplica ao estimador do primeiro passo da Imputação quando Y é variável limitada.

emplo, Hahn (1998), Hirano, Imbens e Ridder (2003)), é que a ponderação utilizando o verdadeiro $p(.)$ não permite atingir o limite de eficiência, enquanto isto é possível empregando valores estimados. Por outro lado, a necessidade de obter estas estimativas torna a Reponderação praticamente tão complexa quanto a Imputação, ao contrário do sugerido pela motivação inicial para o estudo de métodos baseados no *propensity score*.

3.3 Propriedades Assintóticas

Ambas as abordagens apresentadas nas seções anteriores envolvem procedimentos em duas etapas: uma estimação não-paramétrica e uma resolução de condição de momento utilizando a estimativa obtida preliminarmente. Estimadores deste tipo são objeto de particular interesse entre os métodos semiparamétricos, e diversos estudos propõem estruturas gerais para analisá-los, como Bickel et al. (1993), Newey (1994), Newey e McFadden (1994) e Chen, Linton e Van Keilegom (2003).

Na terminologia dessa literatura, a função estimada no primeiro passo é denominada parâmetro de ruído (*nuisance parameter*), em contraste com o parâmetro de interesse. Uma questão fundamental é a forma como a estimação deste é afetada pelo desconhecimento do primeiro, tomando como referência o estimador de método dos momentos obtido caso aquela função fosse conhecida.

Por resultados bem conhecidos da teoria sobre o método dos momentos, a consistência dos estimadores é implicada pela convergência da condição de momento amostral, uniformemente com respeito ao parâmetro de interesse, para seu valor populacional. Isto é, no caso da Imputação, $\hat{\beta} \xrightarrow{p} \beta$ se

$$\begin{aligned} \sup_{\beta} & \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{m}_1(X_i) - \hat{m}_0(X_i)) - \beta - (E\left[m_1(X) - m_0(X)\right] - \beta) \right] \\ & = \frac{1}{N} \sum_{i=1}^{N} (\hat{m}_1(X_i) - \hat{m}_0(X_i)) - E\left[m_1(X) - m_0(X)\right] \xrightarrow{p} 0 \end{aligned} \tag{15}$$

e, no caso da Reponderação, se

$$\frac{1}{N} \sum_{i=1}^{N} \frac{T_i Y_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} - E\left[\frac{TY}{p(X)} - \frac{(1-T)Y}{1-p(X)}\right] \xrightarrow{p} 0. \tag{16}$$

Uma forma de demonstrar (15) e (16) é observar que as parcelas individuais $\hat{m}_1(X_i) - \hat{m}_0(X_i)$ e $\frac{T_i Y_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)}$ são contínuas com respeito a $(\hat{m}_0(.), \hat{m}_1(.))$ e $\hat{p}(.)$, se considerarmos para estes últimos a norma do supremo. Portanto, sob hipóteses que garantam a convergência, nessa norma, dos parâmetros de ruído a seus valores verdadeiros, um argumento de Lei Uniforme dos Grandes Números permite verificar esses limites. Newey (1994, Lema 5.2) sugere esta estratégia, que é utilizada por Hirano, Imbens e Ridder (2000) e Chen, Hong e Tamer (2005), respectivamente para Reponderação e Imputação.

O estudo das propriedades relacionadas à $\sqrt{N}$-consistência é facilitado pela observação de Newey (1994), de que, quando o modelo semiparamétrico é suficientemente amplo, qualquer estimador regular assintoticamente linear (RAL) é eficiente. Este resultado é conseqüência de outro (Newey, 1990) que diz que a função influência de um estimador RAL, ψ, satisfaz, no lugar de d, a equação de diferenciabilidade por caminhos (6). Logo, usando a notação da seção 2.5, para qualquer submodelo paramétrico regular,

$$E[(\psi - \delta)S'_\theta] = \frac{\partial\beta(\theta_0)}{\partial\theta} - \frac{\partial\beta(\theta_0)}{\partial\theta} = 0, \tag{17}$$

i.e., a diferença $\psi - \delta$ é ortogonal a qualquer escore. A condição de abrangência do modelo semiparamétrico é que o espaço tangente contenha todas as funções de média zero dos dados. Quando válida, temos que $\psi - \delta$ é simultaneamente pertencente e ortogonal ao espaço tangente; portanto, deve ser idênticamente nula.

De fato, o resultado de Scharfstein, Robins e Rotnitzky (1999) apresentado na seção 2.4, conforme os próprios autores apontam[11], mostra que o modelo em que apenas se impõe Ignorabilidade é amplo como requerido por Newey. Logo, independente da abordagem (se por Reponderação ou Imputação), sob condições de regularidade, um estimador assintoticamente linear de β terá a função influência eficiente. Em vista disso, duas formas de escrever esta função são interessantes:

$$\begin{aligned}\psi^* &= m_1(X) - m_0(X) - \beta + \frac{T}{p(X)}(Y - m_1(X)) - \frac{1-T}{1-p(X)}(Y - m_0(X)) \\ &= \frac{T}{p(X)}Y - \frac{1-T}{1-p(X)}Y - \beta - (T - p(X))(\frac{m_1(X)}{p(X)} + \frac{m_0(X)}{1-p(X)}).\end{aligned}$$

Em cada uma das formas, o primeiro termo é a função influência dos estimadores de Imputação e Reponderação que seriam obtidos usando-se os valores verdadeiros das funções de regressão e *propensity score*. O segundo termo pode ser visto como uma correção para a estimação dessas funções. Para o caso da estimação por reponderação, o termo de correção é negativamente correlacionado com o primeiro, daí o ganho de eficiência em

[11] Op. cit., p. 1118.

se estimar o *propensity score*, mesmo quando se conhece esta função.

As propriedades sobre a distribuição assintótica dependem, portanto, da convergência dos 'resíduos':

$$\frac{1}{N}\sum_{i=1}^{N}\hat{m}_1(X_i) - m_1(X_i) - (\hat{m}_0(X_i) - m_0(X_i)),$$

da estimação de $m_t(.)$ na Regressão, e

$$\frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} - \left(\frac{T_iY_i}{p(X_i)} - \frac{(1-T_i)Y_i}{1-p(X_i)}\right),$$

da estimação de $p(.)$ na Reponderação, para os respectivos termos de correção a uma taxa mais rápida que $\frac{1}{\sqrt{N}}$. Esta condição impõe restrições adicionais sobre a taxa à qual os estimadores dos parâmetros de ruído devem convergir. No caso da estimação por Reponderação, um desdobramento da ineficiência do uso do *propensity score* verdadeiro é que a convergência da estimativa deste deve, por outro lado, ser lenta o suficiente para deixar o 'resto' de ordem $\frac{1}{\sqrt{N}}$ mostrado acima. Isto torna não-trivial a incorporação de informações adicionais (como restrições de forma funcional ou de suavidade de $p(.)$) que possam auxiliar o primeiro passo desta abordagem[12].

A velocidade de convergência que um método não paramétrico de estimação de uma função pode atingir depende da suavidade do estimando. Para *sieve*/mínimos quadrados, por exemplo, Newey (1997) mostra que as taxas máximas de convergência uniforme e em média quadrática são funções decrescentes da dimensão do espaço dos regressores e crescentes do número de derivadas contínuas da função estimada. Assim, quanto maior o conjunto de variáveis auxiliares, maior a suavidade a ser exigida das funções desconhecidas.

Para seu estimador de Reponderação, Hirano, Imbens e Ridder (2003) mostram a normalidade e eficiência assintótica supondo que o *propensity score* é $7r$ vezes diferenciável, onde r é a dimensão do espaço das variáveis auxiliares. Chen, Hong e Tamer (2005) demonstram essas propriedades para o estimador de Imputação, desde que $m_t(.)$ pertençam a determinados espaços de Hölder[13] e $\frac{p(.)}{1-p(.)}$ seja suave o suficiente para admitir determinadas aproximações nos espaços *sieve*.

[12] Chen, Hong e Tarozzi (2008) mostram uma forma de fazê-lo

[13] Uma função $f(.)$ pertence ao espaço de Hölder (γ, p) se tem derivadas até a ordem $\lfloor\gamma\rfloor$ (maior inteiro menor ou igual a γ, sua parte inteira), sendo as derivadas de maior ordem funções Hölder-contínuas de expoente $\alpha = (\gamma - \lfloor\gamma\rfloor)$, i.e., com $\frac{\|f(x)-f(y)\|}{\|x-y\|^{\alpha}}$ limitada.

4 Combinando Imputação e Reponderação

Vimos, na seção 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 na próxima seção) 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, nesta seção.

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):

$$\begin{aligned}\hat{\beta}_{DR} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{p(X_i;\hat{\gamma})} - \frac{(1-T_i)Y_i}{1-p(X_i;\hat{\gamma})} \\ - (T_i - p(X_i;\hat{\gamma}))\left(\frac{m_1(X_i;\hat{\delta}_1)}{p(X_i;\hat{\gamma})} + \frac{m_0(X_i;\hat{\delta}_0)}{1-p(X_i;\hat{\gamma})}\right) \qquad (18)\end{aligned}$$

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 (18) e a função influência eficiente apresentada na Seção 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:

$$\beta = E\left[\frac{TY}{p(X)} - \frac{(1-T)Y}{1-p(X)}\right] - E\left[(T-p(X))(\frac{\bar{m}_1(X)}{p(X)} + \frac{\bar{m}_0(X)}{1-p(X)})\right], \tag{19}$$

análoga a (18), 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, (19) é idêntica a (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 (19) da seguinte forma:

$$\beta = E\left[m_1(X) - m_0(X)\right] - E\left[\frac{T}{\bar{p}(X)}(Y - m_1(X)) + \frac{1-T}{1-\bar{p}(X)}(Y - m_0(X))\right]. \tag{20}$$

Se m_0 e m_1 forem as verdadeiras funções de regressão, o segundo termo do lado direito de (20) é nulo para qualquer função $\bar{p}(.)$, pois

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

e, analogamente,

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

Logo (20) equivale a (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 na seção 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 (18), 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 $\hat{\beta}_{DR}$:

$$\hat{\beta}_{FI} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\hat{p}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{p}(X_i)} \\ - (T_i - \hat{p}(X_i))\left(\frac{\hat{m}_1(X_i)}{\hat{p}(X_i)} + \frac{\hat{m}_0(X_i)}{1-\hat{p}(X_i)}\right), \tag{21}$$

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 (21) 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[14]. 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,

[14] 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.

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 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:

$$\hat{\beta}_{RP} = \frac{1}{N}\sum_{i=1}^{N}\tilde{m}_1(X_i;\hat{p}(.)) - \tilde{m}_0(X_i;\hat{p}(.)), \tag{22}$$

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)[15], 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

[15] 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.

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 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.

5 Simulação Monte Carlo

A variedade de metodologias para a estimação de Efeito de Tratamento sob Ignorabilidade, bem como de possíveis implementações de cada uma, motivou, nos últimos anos, uma série de estudos de simulação comparando o desempenho das alternativas em amostras finitas. Nesta seção, revisaremos a literatura de simulações relacionadas à questão da combinação de métodos proposta ao longo deste trabalho. Em seguida, realizaremos um novo estudo de simulações com o objetivo de avaliar a relevância dos estimadores propostos na seção 4.3.

5.1 Resultados da Literatura de Simulações sobre Combinação de Métodos

Estudos de simulação considerando métodos combinados foram motivados pelos estudos teóricos sobre esses estimadores e, devido a isto, são recentes. Uma parte considerável dessa literatura foi desenvolvida pelos próprios proponentes dos procedimentos duplamente robustos, (Bang e Robins (2005), Robins et al. (2007), Neugebauer e Van der

Laan (2002)), como forma de divulgá-los. Os resultados publicados até o presente são nitidamente divergentes entre si.

Lunceford e Davidian (2004) estudam o comportamento em amostras finitas de diversos estimadores baseados em *propensity score*. O foco do trabalho é a comparação entre métodos de Reponderação e de estratificação da amostra, com vantagem para a primeira abordagem, nas simulações apresentadas. Um ponto interessante para nossa discussão é que tanto Reponderação quanto estratificação apresentam melhor desempenho quando ajustadas por uma regressão. A recomendação final é o uso de um estimador duplamente robusto na forma de Reponderação aumentada como proposto por Scharfstein, Robins e Rotnitzky (1999).

Bang e Robins (2005) apresentam extensões da estimação duplamente robusta para diferentes contextos envolvendo dados faltantes. Nas simulações, são destacadas as situações em que os modelos usados para estimar a probabilidade de observação e os valores potenciais estão mal especificados. O principal resultado é o bom desempenho dos estimadores duplamente robustos, mesmo quando ambas as especificações estão incorretas.

Num trabalho semelhante, Kang e Schaeffer (2007) discutem métodos duplamente robustos e os comparam com métodos baseados em Regressão e Reponderação. Também é dada ênfase o comportamento dos diferentes procedimentos quando as estimações preliminares falham, de forma limitada, em captar os modelos populacionais de seleção e regressão. Os resultados obtidos mostram sensibilidade dos métodos que usam *propensity score* à má especificação dos modelos para este objeto. Diferentemente do que ocorre em Bang e Robins (2005), os melhores desempenho são dos estimadores simples de Regressão, enquanto os duplamente robustos representam melhoria em relação aos de Reponderação em alguns casos. Num comentário sobre este trabalho, Robins et al. (2007) questionam os resultados, argumentando que peculiaridades das simulações favorecem particularmente os estimadores de Regressão. A fragilidade dos resultados de Kang e Schaeffer é sugerida pela consideração do modelo modificado pela inversão do indicador T, que apresenta resultados divergentes em relação aos do modelo original.

Busso, DiNardo e McCrary (2009) comparam estimadores de Reponderação, *Matching* e o estimador duplamente robusto em forma de Regressão Ponderada. Uma característica peculiar dos modelos adotados é que as funções de regressão dependem de X apenas via $p(X)$. Em suas simulações, os estimadores de Reponderação apresentam os melhores resultados, exceto quando a hipótese de Sobreposição é violada, ou em situações próximas, em que o *propensity score* atinge valores próximos de zero ou um. Neste caso, todos os métodos apresentam considerável viés e pouca semelhança com as previsões da teoria assintótica. Os autores ressaltam a importância desta hipótese, e apontam a sua falha como razão de resultados inusitados em outros estudos de simulação.

5.2 Um Novo Estudo de Simulações

O estudo de simulações a seguir tem o objetivo de avaliar a interação entre elementos do modelo populacional e o desempenho da combinação de Reponderação e Imputação em relação ao uso de apenas uma das técnicas, além de suas diferentes formas de implementação. Neste sentido, quatro elementos principais serão analisados.

O primeiro aspecto que buscamos ressaltar é o efeito da forma funcional das funções de regressão e do *propensity score*, particularmente quanto à sua suavidade. Pela discussão da seção anterior, a motivação para utilizar o inverso do *propensity score* como peso numa regressão é eliminar a correlação entre os regressores e variáveis omitidas. A má especificação pode ser incluída neste contexto, se interpretada como a omissão de uma parte de $m_t(X)$. Portanto, funções de regressão pouco suaves, que projetadas numa pequena base das variáveis auxiliares deixam um resíduo considerável, devem implicar ganhos importantes em reponderar previamente a amostra.

Em segundo lugar, são estudados os efeitos da heterocedasticidade sobre a conveniência de realizar Regressões Ponderadas. Uma vez que a reponderação da regressão em $\hat{\beta}_{RP}$ segue um critério completamente independente da eficiência dessa estimação, o efeito sobre a precisão do estimador pode ser adverso.

Terceiro, será considerado o efeito da quantidade de dimensões do espaço das variáveis auxiliares. A discussão do artigo de Robins e Ritov (1997) sugere que a 'maldição da dimensionalidade' tem mais chances de comprometer o desempenho de estimadores não robustos. Por outro lado, vimos que a hipótese de Ignorabilidade exige do econometrista um esforço em obter toda a informação possível para controlar a heterogeneidade, o que deve refletir no número de variáveis auxiliares. Apesar disto, a maioria das simulações disponíveis na literatura emprega X univariada. Dos estudos mencionados anteriormente, apenas o de Lunceford e Davidian (2004) expressam esta preocupação.

Finalmente, avaliaremos o uso do *propensity score* verdadeiro como elemento do cálculo dos pesos na Regressão Ponderada e do termo de correção na Reponderação Aumentada. Conforme apontamos na seção anterior, é possível que um estimador Duplamente Robusto que empregue o valor populacional de $p(.)$, no lugar de uma estimativa, atinja o Limite de Eficiência Semiparamétrico. Isto indicaria que a incorporação do conhecimento disponível sobre essa função pode ser feita de maneira simples, reduzindo a dificuldade computacional do procedimento sem prejuízo à variância assintótica.

5.3 Estimadores e Modelos

Consideraremos estimadores de Regressão ($\hat{\beta}_{rep}$), Imputação ($\hat{\beta}_{imp}$) e duplamente robustos dos tipos $\hat{\beta}_{FI}$ (Reponderação Aumentada ou Função Influência) e $\hat{\beta}_{RP}$ (Regressão Ponderada). As estimativas preliminares de $p(.)$ serão realizadas por *sieve*/logit, para todos os estimadores. As funções de regressão $m_t(.)$, $t = 0, 1$ serão estimadas por *sieve*/mínimos quadrados ordinários, para Regressão e $\hat{\beta}_{FI}$, e por *sieve*/mínimos quadrados ponderados (com os pesos descritos na seção 4.3), para o uso em $\hat{\beta}_{RP}$. Além disso, calculamos também os estimadores $\tilde{\beta}_{FI}$ e $\tilde{\beta}_{RP}$, análogos a $\hat{\beta}_{FI}$ e $\hat{\beta}_{RP}$, porém combinando o *propensity score* verdadeiro e as funções de regressão estimadas.

Como forma de estudar a seleção do número de termos das bases dos estimadores *sieve*, consideraremos todo o intervalo de valores $L = 1, 2, 3, ...\bar{L}$ para a estimação de $m_t(.)$ e $K = 1, 2, 3, ...\bar{K}$ para a estimação de $p(.)$. As bases consistem de polinômios das variáveis auxiliares, em ordem (i) crescente no grau, (ii) decrescente no maior expoente, e (iii) decrescente no expoente da variável de menor número de ordem. Como observado no seção anterior, os estimadores de Regressão e Imputação são casos particulares dos estimadores duplamente robustos, respectivamente, quando *propensity score* e funções de regressão são estimados por uma constante. Logo o cálculo $\hat{\beta}_{FI}$ e $\hat{\beta}_{RP}$ inclui o de $\hat{\beta}_{rep}$, quando $L = 1$ e $\hat{\beta}_{imp}$, quando $K = 1$. Além disso, para $L = 1$, $\tilde{\beta}_{FI}$ e $\tilde{\beta}_{RP}$ são iguais ao estimador (ineficiente) de Reponderação pelo *propensity score* populacional.

Para destacar o efeito da suavidade de $p(.)$ e $m_t(.)$, consideramos inicialmente um modelo com uma única variável auxiliar X_i, distribuída uniformemente entre -1 e 1, e

$$\begin{aligned}
Y_i(1) &= m_1(X_i) + u_{1,i} \\
Y_i(0) &= m_0(X_i) + u_{0,i} \\
Y_i &= T_i(m_1(X_i) + u_{1,i}) + (1 - T_i)(m_0(X_i) + u_{0,i})
\end{aligned}$$

com Y_i homocedástica, ou seja, $u_i = (u_{0,i}, u_{1,i})$ independente e identicamente distribuída conforme uma distribuição normal bivariada de média zero e variância σ^2. A variação das especificações de $p(.)$ e $m_t(.)$ é feita pela manipulação dos coeficientes das seguintes representações:

$$
\begin{aligned}
log\left(\frac{p(x)}{1-p(x)}\right) &= \sum_{k=0}^{\infty} a_k \lambda_k(x) \\
m_1(x) - m_0(x) &= \sum_{k=0}^{\infty} b_k \lambda_k(x) \\
m_0(x) &= \sum_{k=0}^{\infty} c_k \lambda_k(x),
\end{aligned}
$$

onde λ_k é o polinômio de Legendre de grau k. Os polinômios de Legendre são construídos de forma a serem ortogonais entre si, em relação ao produto interno $< f, g >= \int fgdx$; portanto definem funções não correlacionadas entre si de uma variável com distribuição uniforme. O primeiro polinômio é constante ($\lambda_0(x) \equiv 1$), logo, por ortogonalidade, os demais têm média zero, e portanto o Efeito Médio de Tratamento é dado por

$$
\beta = E[Y_i(1) - Y_i(0)] = \sum_{k=0}^{\infty} b_k E[\lambda_k(x)] = b_0.
$$

A velocidade de decaimento dos coeficientes das séries simula a suavidade das funções correspondentes. De fato, o papel das hipóteses de suavidade na teoria assintótica dos estimadores estudados é assegurar determinadas taxas de convergência do estimador não paramétrico do primeiro passo. No caso da estimação por *sieve*, a conseqüência relevante da suavidade é garantir a existência de uma seqüência de aproximações (não aleatórias) $\sum_{k=0}^{K} \tilde{a}_k^K \lambda_k(x)$ convergindo a determinada taxa (em termos de K) para a função estimada (Newey, 1997). Este é justamente o efeito simulado através da manipulação de especificações realizada nos modelos 1 a 3.

No modelo 1, utilizamos $\sigma^2 = 9,5114 I_2$ (I_2 é a matriz identidade 2 por 2); $a_0 = 0,5$; $a_k = \frac{k^{-1}}{2}$, se $1 \leq k \leq 10$; $a_k = 0$, se $k > 10$; $b_0 = 1$; $b_k = 20a_k$, se $k \geq 1$; $c_k \equiv 1$. Com esta escolha de parâmetros, o *propensity score* é uma transformação monotônica da função de regressão, o que intuitivamente parece tornar redundantes Reponderação e Imputação, e inócua a combinação entre as duas. O modelo é homocedástico, o que tornaria atrativo o estimador de regressão corretamente especificado.

Nos modelos 2 e 3, são alteradas as especificações de modo que os coeficientes se distribuem diferentemente. No primeiro caso, o decaimento é mais rápido: $a_k = 1,0922\frac{k^{-2}}{2}$, se $1 \leq k \leq 10$; no segundo, não há decaimento: $a_k = \frac{0,5926}{2}$, se $1 \leq k \leq 10$. É feito o correspondente ajuste para manter $b_k = 20a_k$, se $k \geq 1$, em cada caso. As constantes multiplicadas na redefinição de a_k mantêm o limite de eficiência semiparamétrico igual ao do modelo 1.

Os modelos 4 e 5 introduzem heterocedasticidade. São consideradas as especificações do modelo 1 para a_k, b_k e c_k, mas u_i são independentemente distribuídas com média zero e variância $\sigma^2(X_i)$. No modelo 4, definimos $\sigma^2(X_i) = 20,7767 diag(1-p(X_i), p(X_i))$ ($diag(d_1, d_2)$ é a matriz que tem d_1 e d_2 na diagonal principal e zero nas demais entradas), uma especificação favorável à Regressão Ponderada, pois, por exemplo, observações tratadas com *propensity score* alto, que receberão pesos baixos, têm variância maior. Se os pesos são corretamente especificados, correspondem àqueles que otimizam as regressões de $Y(t)$ em X por mínimos quadrados ponderados. No modelo 5, a situação é oposta, com $\sigma^2(X_i) = 17,5421 diag(p(X_i), 1-p(X_i))$. As constantes nas expressões para $\sigma^2(X_i)$ são introduzidas para preservar o mesmo limite de eficiência semiparamétrico, como fizemos nos modelos 2 e 3.

Para a análise do efeito da dimensão de X_i, o modelo 6 considera três componentes para esta variável, todas identicamente distribuídas, uniformemente no intervalo $[-1, 1]$, e independentes entre si. Os modelos de seleção para o tratamento e de média condicional são descritos abaixo:

$$
\begin{aligned}
&p(x) = 0,8x_3 L(x_1) + (1 - 0,8x_3) L(x_2) \\
&L(.) = \frac{\exp(.)}{1 - \exp(.)} \\
&m_1(x) - m_0(x) = 5\exp(2x_3)(1.5 + \sin(\frac{\pi}{4}(x_1 + x_2)) - (x_1 - x_2)^2) \\
&m_0(x) \equiv 0.
\end{aligned}
$$

O par de valores potenciais da variável de interesse tem o componente aleatório u_i independente e identicamente distribuído de acordo com uma distribuição normal bivariada de média zero e variância $10I_2$.

5.4 Resultados

Para as simulações com X unidimensional, foram consideradas amostras com $N = 100$ e $N = 200$ observações. Acima destes valores, os estimadores apresentaram desempenhos indistingüíveis. As tabelas 1 a 4 reportam os erros quadráticos médios simulados dos estimadores, para as duas formas de combinar métodos, com a especificação do modelo 1.

[inserir Tabela 1 aqui]
[inserir Tabela 2 aqui]
[inserir Tabela 3 aqui]
[inserir Tabela 4 aqui]

É notável que o melhor estimador, em todos os casos, utiliza mais de um termo tanto para estimar o *propensity score* ($K > 1$) quanto as funções de regressão ($L > 1$). Contrastado com a observação na seção anterior, de que o modelo 1 é particularmente desfavorável, este fato é um forte indício a favor da combinação. O ganho em utilizar a melhor combinação, em relação a Reponderação ou Imputação, é da ordem de 5%, para $N = 100$, e 1% , para $N = 200$, em termos de erro quadrático médio.

Outro ponto interessante é a proximidade entre $\hat{\beta}_{FI}$ e $\hat{\beta}_{RP}$, quando utilizados o mesmos números de termos K e L. Curiosamente, enquanto, na maioria dos casos, o estimador ótimo é $\hat{\beta}_{RP}$, para valores muito altos de K e L, $\hat{\beta}_{FI}$ é mais estável.

A tabela 5 mostra os resultados da simulação com o modelo 2, mais suave, com $N = 200$. Pela similaridade observada, para este modelo e os próximos, são reportados apenas os valores de $\hat{\beta}_{RP}$. Verifica-se que, quando o modelo é muito suave, a diferença entre os todos os estimadores desaparece rapidamente. Nesta especificação, embora o melhor estimador empregue estritamente uma combinação das técnicas, percebe-se que há tanto estimadores de Reponderação quanto de Imputação com precisão distante em menos de 3% do Limite de Eficiência.

[inserir Tabela 5 aqui]

A especificação menos suave deteriora a performance de todos os estimadores, conforme mostrado na tabela 6. Percebe-se que os melhores estimadores empregam um grande número de termos na estimação de $p(.)$, e que a introdução de pelo menos um termo além da constante, na estimação de $m_t(.)$, geralmente reduz o erro quadrático. Também é notada uma mudança no desempenho relativo entre Imputação pura e Reponderação pura, favorável à última.

[inserir Tabela 6 aqui]

Uma possível explicação para esta observação é que $p(.)$ é uma transformação de $m_1(.) - m_0(.)$ pela função logística, $L(.)$. Isto pode fazer com que a perda de suavidade tenha sido desigual, pois a composição com $L(.)$ atenua variações.

Os modelos utilizados para avaliar o efeito da heterocedasticidade, embora aparentemente representando casos extremos, não forneceram evidência favorável à sugestão de que esta poderia afetar os benefícos da Regressão Ponderada. As tabelas 7 e 8 reportam os resultados, respectivamente, para os casos supostamente favorável e desfavorável. Os estimadores ótimos, em ambos os casos, utilizam o mesmo número de termos, contrariando a expectativa de que, no segundo, seria recomendada uma reponderação mais precisa.

[**inserir Tabela 7 aqui**]
[**inserir Tabela 8 aqui**]

A introdução de múltiplas variáveis auxiliares reforça o ganho da combinação de estratégias, quando utilizadas pelo estimador de Regressão Ponderada, como mostrado na tabela 9. O melhor estimador Duplamente Robusto, neste caso, tem erro quadrático médio 10% e 8% menor que os melhores estimadores de, respectivamente, Imputação e Reponderação. Diferentemente das especificações anteriores, porém, a forma de combinar os métodos influencia o resultado. Na tabela 10, que considera o estimador de Reponderação Aumentado, ocorrem perdas consideráveis em algumas combinações de K e L, e o melhor estimador utiliza apenas Reponderação. No entanto, o estimador $\hat{\beta}_{FI}$ que utiliza a combinação ótima para $\hat{\beta}_{RP}$ ainda se encontra entre os melhores.

[**inserir Tabela 9 aqui**]
[**inserir Tabela 10 aqui**]

Por fim, conforme esperado, a reponderação pelo inverso do *propensity score* verdadeiro resultou em desvios muito maiores que o Limite de Eficiência Semiparamétrico, como podemos observar na tabela 11. Por outro lado, o uso de $p(.)$ combinado com procedimentos de imputação permitiu atingir desempenho semelhante ao das implementações ótimas de $\hat{\beta}_{FI}$ e $\hat{\beta}_{RP}$, exceto nos modelos 3 e 6. Sinteticamente isto é um indício de que o uso de informação sobre o *propensity score*, embora piore o desempenho assintótico em estimadores de Reponderação simples, permite atingir eficiência assintótica quando se combina com a estimação das funções de regressão. Em modelos menos afetados pela 'maldição da dimensionalidade', o recurso a esta informação foi particularmente efetivo em facilitar a obtenção de um estimador ótimo, uma vez que eliminou a necessidade de se escolher o número de termos a utilizar em $\hat{p}(.)$, ao mesmo tempo em que possibilitou alcançar um erro quadrático médio praticamente idêntico ao de $\hat{\beta}_{FI}$ e $\hat{\beta}_{RP}$.

[**inserir Tabela 11 aqui**]

6 Conclusão

Neste trabalho, discutimos a combinação das estratégias de Regressão/Imputação e Reponderação para a estimação de Efeitos de Tratamento sob a hipótese de Ignorabilidade. Argumentamos que a mistura de métodos oferece uma oportunidade para a melhoria da inferência em amostras finitas, tomando como referencial teórico a literatura de estimação duplamente robusta.

No contexto da inferência paramétrica, a teoria sobre estimação duplamente robusta sugere o uso de um ajuste, baseado em Regressão, do estimador de Reponderação, quando se considera o risco de um dos modelos estar mal especificado. A versão semiparamétrica dessa literatura apresenta um procedimento similar para lidar com a potencial falha das hipóteses de suavidade subjacentes às propriedades assintóticas. A partir do resultado formal de robustez desse estimador semiparamétrico, é sugerido seu bom desempenho em amostras pequenas, relacionado ao problema da 'maldição da dimensionalidade'.

Sob esta motivação, dois estimadores semelhantes foram apresentados e testados num exercício de Monte Carlo. A realização das simulações confirmou a possibilidade de estimar mais precisamente o parâmetro de interesse através de um mistura dos métodos tradicionais. Foi verificada, ainda, a relação entre a 'maldição da dimensionalidade' e a conveniência de misturar métodos, conforme proposto por Robins e Ritov (1997). Um maior número de dimensões do espaço das variáveis auxiliares, e/ou a menor suavidade das funções de regressão e *propensity score*, foram acompanhados por um hiato maior no desempenho dos estimadores. Por outro lado, contrariamente ao que se esperava, a heterocedasticidade não apresentou efeito sobre o ganho de ponderar as regressões. Outra vantagem dos estimadores Duplamente Robustos observada no Monte Carlo foi a capacidade de incorporação de conhecimentos sobre o *propensity score* de forma trivial e sem perda de eficiência.

Uma importante questão que não foi abordada é a metodologia para seleção do número de termos da base dos estimadores *sieve*, ou, de forma mais geral, do parâmetro de suavidade das estimações semiparamétricas da primeira etapa. O estudo das simulações mostrou simplesmente a possibilidade de, realizando uma escolha suficientemente boa, superar o desempenho das técnicas de Imputação e Reponderação. Porém, enquanto estas exigem que se defina apenas um parâmetro de suavidade, a combinação impõe duas decisões. A capacidade de propor uma regra, dependente dos dados, que selecione bem estes parâmetros é, por conseguinte, crítica para o sucesso prático desta estratégia de estimação. Analisar este problema é o próximo passo lógico na linha de estudo considerada aqui.

Outra extensão possível para este trabalho é a análise do problema de inferência sobre parâmetros do tipo Efeito de Tratamento Sobre os Tratados. Embora tenha recebido, em geral, um tratamento paralelo ao do Efeito Médio de Tratamento na literatura, não se trata de uma teoria equivalente. Em particular a literatura sobre estimação duplamente robusta é, neste caso, incipiente. Dos trabalhos investigados neste trabalho, apenas Egel, Graham e Pinto (2009), que também apontam esta deficiência, explicitam um estimador para esse contexto.

Referências Bibliográficas

BANG, H.; ROBINS, J. Doubly Robust Estimation in Missing Data and Causal Inference Models. Biometrics, v.61, p.962-972, 2005.

BICKEL, P.; KLASSEN, C.; RITOV, Y. ; WELLNER, J. Efficient and Adaptive Estimation for Semiparametric Models. Nova Iorque: Springer-Verlag, 1993.

BUSSO, M.; DINARDO, J. ; MCCRARY, J. Finite Sample Properties of Semiparametric Estimators of Average Treatment Effects, 2009. Disponível em: <http://www.econ.berkeley.

CATTANEO, M. Efficient Semiparametric Estimation of Multivalued Treatment Effects. 2007. Tese (PhD. em Economia) - Department of Economics, University of California, Berkeley.

CHEN, X.; HONG, H. ; TAMER, E. Measurement Error Models With Auxiliary Data. Review of Economic Studies, v.72, p.343-366, 2005.

CHEN, X.; HONG, H. ; TAROZZI, A. Semiparametric efficiency in GMM models with auxiliary data. Annals of Statistics, v.36, n.2, p.808-843,2008.

CHEN, X.; LINTON, O. ; KEILEGOM, I. V. Estimation of semiparametric models when the criterion function is not smooth. Econometrica, v.71, p.1591-1608, 2003.

CRUMP, R.; HOTZ, V.; IMBENS, G. ; MITNIK, O. Dealing with Limited Overlap in Estimation of Average Treatment Effects. Biometrika, v.96, n.2, p.1-13, 2009.

EGEL, D.; GRAHAM, B. ; PINTO, C. Inverse Probability Tilting and Missing Data Problems NBER Working Paper Series, n.W13981, 2008. Disponível em: <http://ssrn.com/abstract=1131634>.

FIRPO, S. Efficient Semiparametric Estimation of Quantile Treatment Effects. Econometrica, v.75, n.1, p.259-276, 2007.

FRÖLICH, M. Finite Sample Properties of Propensity-Score Matching and Weighting Estimators. Review of Economics and Statistics, v.86, n.1, p.77-90, 2004.

GEMAM, S.; HUANG, C. Nonparametric Maximum Likelihood Estimation by the Method of Sieves. The Annals of Statistics, v.10, n.2, p.401-414, 1982.

GRENANDER, U. Abstract Inference. Nova Iorque: Wiley, 1981.

HAHN, J. On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. Econometrica, v.66, n.2, p.315-331, 1998.

HAHN, J.; TODD, P. ; VAN DER KLAAUW, W. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica, v.69, n.1, p.201-209, 2000.

HECKMAN, J.; ICHIMURA, H.; SMITH, J. ; TODD, P. Characterizing Selection Bias Using Experimental Data. Econometrica, v.66, n.5, p.1017-1098, 1998.

HECKMAN, J.; ICHIMURA, H. ; TODD, P. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. Review of Economic Studies, v.64, p.605-654, 1997.

HIRANO, K.; IMBENS, G. Estimation of Causal Effects Using Propensity Score Weighting: An Application of Data on Right Hear Catherization. Health Services and Outcomes Research Methodology , v.2, p.259-278, 2001.

HIRANO, K.; IMBENS, G. ; RIDDER, G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. NBER Working Paper Series, n.T0251, 2000. Disponível em: <http://ssrn.com/abstract=228061>.

HIRANO, K.; IMBENS, G. ; RIDDER, G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica, v.71, n.4, p.1161-1189, 2003.

HORVITZ, D.; THOMPSON, D. A Generalization of Sampling Without Replacement from a Finite Universe. Journal of the American Statistical Association, v.47, p.663-685, 1952.

ICHIMURA, H.; LINTON, O. Asymptotic Expansions for some Semiparametric Program Evaluation Estimators, In: ANDREWS, D.; STOCK, J. (Ed.). Identification and Inference for Econometric Models. Cambridge: Cambridge University Press, 2005. cap. 8, p. 149-170.

IMBENS, G. The Role of the Propensity Score in Estimating Dose Response Functions. Biometrika, v.87, n.3, p.706-710, 2000.

IMBENS, G.; ANGRIST, J. Identification and Estimation of Local Average Treatment Effects. Econometrica, v.61, n.2, p.467-476, 1994.

IMBENS, G.; NEWEY, W. ; RIDDER, G. Mean-squared-error Calculations for Average Treatment Effects. IEPR Working Paper, n.05.34, 2005. Disponível em: <http://ssrn.com/abstract=820205>.

IMBENS, G.;WOOLDRIDGE, J. Recent Developments in the Econometrics of Program Evaluation. Journal of Economic Literature, v.47, n.1, p.5-86, 2009.

KANG, J.; SCHAEFFER, J. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. Statistical Science, v.22, n.4, p.523-539, 2007.

LUNCEFORD, J.; DAVIDIAN, M. Stratification and Weighting via the Propensity Score in estimation of Causal Treatment Effects: A Comparative Study. Statistics in Medicine, v.23, p.2937-2960, 2004.

MANSKI, C. Nonparametric Bounds on Treatment Effects. American Economic Review Papers and Proceedings, v.80, p.319-323, 1990.

MANSKI, C. Partial Identi¯cation of Probabality Distributions. Nova Iorque: Springer-Verlag, 2003.

NEUGEBAUER, R.; DER LAAN, M. V. Why Prefer Double Robust Estimates? Journal of Statistical Planning and Inference, v.129, p.405-426, 2005.

NEWEY, W. Semiparametric Efficiency Bounds. Journal of Applied Econometrics, v.5, p.99-135, 1990.

NEWEY, W. The Asymptotic Variance of Semiparametric Estimators. Econometrica, v.62, n.6, p.1349-1382, 1994.

NEWEY, W. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, v.79, p.147-168, 1997.

NEWEY, W.; MCFADDEN, D. Large Sample Estimation and Hypothesis Testing, In: ENGLE, R.; MCFADDEN, D. (Ed.). Handbook of Econometrics, v.4. Amsterdã: North-Holland, 1994. cap. 38, p. 2111-2242.

ROBINS, J.; RITOV, Y. Towards a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-parametric Models. Statistics in Medicine, v.16, p.285-319, 1997.

ROBINS, J.; ROTNITZKY, A. Semiparametric Efficiency in Multivariate Regression Models with Missing Data. Journal of the American Statistical Association, v.90, n.429, p.122-129, 1995.

ROBINS, J.; ROTNITZKY, A. ; ZHAO, L. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. Journal of the American Statistical Association, v.90, n.429, p.106-121, 1995.

ROBINS, J.; SUED, M.; LEI-GOMEZ, O. ; ROTNITZKY, A. Comment: Performance of Double-Robust Estimators When 'Inverse Probability' Weights Are Highly Variable. Statistical Science, v.22, n.4, p.544-559, 2007.

ROSENBAUM, P.; RUBIN, D. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, v.70, p.41-55, 1983.

RUBIN, D. Matching to Remove Bias in Observational Studies. Biometrics, v.29, p.159-183, 1973.

RUBIN, D. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational Statistics, v.2, n.1, p.1-26, 1977.

RUBIN, D. Bayesian inference for causal effects: The Role of Randomization. Annals of Statistics, v.6, n.1, p.34-58, 1978

SCHARFSTEIN, D.; ROBINS, J. ; ROTNITZKY,A. Adjusting for Non-ignorable Drop-Out Using Semiparametric Nonresponse Models. Journal of the American Statistical Association, v.94, n.448, p.1096-1146, 1999.

STEIN, C. Efficient nonparametric testing and estimation. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 3; 1956, Berkeley. Proceedings. Berkeley: University of California Press, 1956. p.187-196.

Tabela 1: Modelo 1, Regressão Ponderada, N=100

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	3,578	1,004	0,9	0,869	0,914	1,015	1,969	3,767	>10
2	0,957	1,047	0,9	0,864	0,899	1,003	1,952	4,42	>10
3	0,89	0,88	0,9	0,867	0,902	0,998	1,961	4,541	>10
4	0,889	0,862	0,9	0,87	0,902	1,006	1,914	4,52	>10
5	0,9	0,855	0,9	0,857	0,905	1,012	1,922	4,219	>10
6	0,911	0,855	0,9	**0,855**	0,901	1,012	1,952	4,745	>10
7	0,93	0,863	0,9	0,86	0,906	1,015	1,928	4,67	>10
8	0,949	0,868	0,9	0,865	0,908	1,015	1,984	4,639	>10
9	0,991	0,882	0,9	0,876	0,923	1,023	2,046	4,624	>10
10	1,018	0,93	9,3	>10	>10	>10	>10	>10	>10
11	1,07	0,917	0,9	3,951	>10	>10	>10	>10	>10
12	1,12	1,114	>10	>10	>10	>10	>10	>10	>10

N=100, Limite de Eficiência Semiparamétrico = 0,831, 5000 replicações

Tabela 2: Modelo 1, Reponderação Aumentada, N=100

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	3,578	1,004	0,9	0,869	0,914	1,015	1,969	3,767	>10
2	0,957	1,062	0,9	0,869	0,914	1,015	1,969	3,767	>10
3	0,89	0,881	0,9	0,871	0,913	1,015	1,969	3,767	>10
4	0,889	0,867	0,9	0,875	0,915	1,016	1,969	3,767	>10
5	0,9	0,86	0,9	0,859	0,918	1,018	1,969	3,767	>10
6	0,911	0,859	0,9	**0,858**	0,911	1,017	1,968	3,767	>10
7	0,93	0,867	0,9	0,864	0,917	1,017	1,965	3,766	>10
8	0,949	0,873	0,9	0,87	0,923	1,02	1,968	3,767	>10
9	0,991	0,887	0,9	0,883	0,941	1,032	1,985	3,773	>10
10	1,018	0,903	0,9	0,891	0,947	1,039	1,98	3,765	>10
11	1,07	0,928	0,9	0,911	0,965	1,062	1,995	3,777	>10
12	1,12	0,936	0,9	0,918	0,977	1,066	2,009	3,782	>10

N=100, Limite de Eficiência Semiparamétrico = 0,831, 5000 replicações

Tabela 3: Modelo 1, Regressão Ponderada, N=200

Erros Quadráticos Médios									
K/L	1	2	3	4	5	6	7	8	9
1	2,56	0,512	0,458	0,439	0,439	0,436	0,445	0,45	0,481
2	0,49	0,542	0,457	0,438	0,438	0,437	0,444	0,45	0,476
3	0,45	0,449	0,459	0,439	0,438	0,436	0,444	0,451	0,476
4	0,44	0,437	0,438	0,44	0,438	0,437	0,444	0,45	0,475
5	0,44	**0,433**	0,435	0,434	0,438	0,437	0,444	0,45	0,475
6	0,44	0,435	0,436	0,434	0,436	0,437	0,443	0,45	0,476
7	0,44	0,436	0,437	0,435	0,437	0,436	0,444	0,45	0,478
8	0,45	0,435	0,437	0,434	0,436	0,436	0,443	0,451	0,478
9	0,45	0,437	0,439	0,436	0,438	0,437	0,445	0,452	0,478
10	0,45	0,437	0,438	0,435	0,437	0,436	0,445	0,451	0,477
11	0,46	0,439	0,442	0,437	0,44	0,439	0,447	0,453	0,48
12	0,47	0,442	0,444	0,438	0,442	0,44	0,448	0,455	0,483

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 4: Modelo 1, Reponderação Aumentada, N=200

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	2,56	0,512	0,458	0,439	0,439	0,436	0,445	0,45	0,481
2	0,49	0,548	0,457	0,439	0,439	0,436	0,445	0,45	0,481
3	0,45	0,449	0,461	0,439	0,439	0,436	0,445	0,45	0,481
4	0,44	0,438	0,438	0,441	0,439	0,437	0,445	0,45	0,481
5	0,44	**0,433**	0,436	0,433	0,439	0,437	0,445	0,451	0,481
6	0,44	0,435	0,437	0,434	0,437	0,437	0,445	0,45	0,481
7	0,44	0,436	0,438	0,435	0,438	0,436	0,445	0,451	0,481
8	0,45	0,436	0,438	0,434	0,437	0,436	0,444	0,451	0,482
9	0,45	0,438	0,44	0,436	0,439	0,437	0,446	0,452	0,481
10	0,45	0,437	0,44	0,434	0,438	0,436	0,445	0,451	0,48
11	0,46	0,44	0,443	0,437	0,441	0,438	0,448	0,453	0,482
12	0,47	0,443	0,447	0,439	0,445	0,44	0,45	0,455	0,484

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 5: Modelo 2, Regressão Ponderada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	2,42	0,442	0,428	0,428	0,43	0,432	0,437	0,443	0,466
2	0,43	0,446	0,428	0,427	0,43	0,432	0,437	0,443	0,467
3	0,43	0,427	0,428	0,428	0,43	0,432	0,437	0,443	0,467
4	0,43	0,427	**0,427**	0,428	0,429	0,432	0,437	0,443	0,468
5	0,43	0,429	0,429	0,429	0,429	0,433	0,437	0,443	0,468
6	0,44	0,43	0,431	0,43	0,431	0,433	0,437	0,444	0,467
7	0,45	0,431	0,432	0,432	0,433	0,434	0,438	0,444	0,467
8	0,45	0,433	0,433	0,433	0,433	0,435	0,438	0,444	0,467
9	0,45	0,434	0,434	0,433	0,434	0,436	0,439	0,445	0,468
10	0,46	0,436	0,437	0,436	0,437	0,438	0,442	0,447	0,469
11	0,48	0,44	0,44	0,439	0,44	0,441	0,445	0,449	0,472
12	0,47	0,442	0,443	0,442	0,443	0,443	0,447	0,45	0,474

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 6: Modelo 3, Regressão Ponderada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	2,683	1,346	0,892	0,674	0,598	0,533	0,532	0,503	0,548
2	1,357	1,435	0,908	0,678	0,595	0,53	0,524	0,495	0,524
3	0,879	0,9	0,95	0,703	0,6	0,531	0,524	0,495	0,524
4	0,679	0,685	0,702	0,734	0,61	0,536	0,52	0,492	0,52
5	0,578	0,576	0,586	0,595	0,627	0,545	0,525	0,496	0,521
6	0,523	0,52	0,523	0,526	0,544	0,556	0,532	0,497	0,525
7	0,491	0,486	0,491	0,489	0,506	0,505	0,538	0,501	0,527
8	0,479	0,47	0,473	0,471	0,482	0,479	0,504	0,504	0,533
9	0,458	0,453	0,462	0,457	0,469	0,461	0,486	0,479	0,537
10	0,449	0,448	0,453	0,449	0,457	0,452	0,469	0,468	0,512
11	0,446	0,442	0,446	**0,44**	0,448	0,441	0,459	0,456	0,499
12	0,451	0,445	0,45	0,442	0,452	0,444	0,464	0,459	0,504

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 7: Modelo 4, Regressão Ponderada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	2,62	0,531	0,457	0,445	0,44	0,438	0,45	0,448	0,477
2	0,504	0,562	0,458	0,443	0,439	0,438	0,448	0,448	0,471
3	0,45	0,451	0,461	0,445	0,439	0,438	0,448	0,448	0,47
4	0,444	0,441	0,442	0,445	0,439	0,438	0,447	0,448	0,47
5	0,44	**0,436**	0,437	0,437	0,44	0,438	0,447	0,448	0,471
6	0,443	0,436	0,437	0,436	0,438	0,438	0,447	0,448	0,473
7	0,446	0,438	0,44	0,438	0,44	0,44	0,447	0,448	0,473
8	0,448	0,438	0,44	0,437	0,44	0,439	0,446	0,449	0,476
9	0,451	0,44	0,442	0,439	0,441	0,44	0,447	0,45	0,477
10	0,457	0,443	0,443	0,44	0,443	0,442	0,449	0,451	0,48
11	0,462	0,445	0,445	0,442	0,445	0,443	0,451	0,453	0,482
12	0,466	0,446	0,446	0,443	0,445	0,444	0,451	0,453	0,483

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 8: Modelo 5, Regressão Ponderada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	2,564	0,521	0,453	0,432	0,431	0,431	0,447	0,453	0,483
2	0,494	0,55	0,454	0,432	0,431	0,432	0,446	0,452	0,472
3	0,443	0,446	0,457	0,433	0,431	0,432	0,446	0,455	0,471
4	0,432	0,432	0,432	0,435	0,431	0,432	0,446	0,455	0,473
5	0,432	**0,43**	0,43	0,43	0,432	0,433	0,447	0,455	0,473
6	0,434	0,43	0,431	0,43	0,431	0,434	0,447	0,455	0,474
7	0,441	0,433	0,434	0,432	0,434	0,435	0,447	0,456	0,474
8	0,443	0,435	0,436	0,434	0,436	0,437	0,449	0,458	0,474
9	0,45	0,437	0,439	0,436	0,438	0,439	0,452	0,46	0,475
10	0,457	0,442	0,443	0,44	0,442	0,443	0,455	0,463	0,477
11	0,459	0,442	0,445	0,441	0,444	0,444	0,457	0,464	0,477
12	0,465	0,444	0,447	0,443	0,445	0,445	0,459	0,465	0,479

N=200, Limite de Eficiência Semiparamétrico = 0,4155, 5000 replicações

Tabela 9: Modelo 6, Regressão Ponderada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	6,313	5,336	4,547	4,25	4,201	4,463	4,519	3,826	3,906
2	5,319	5,28	4,52	4,184	4,093	4,349	4,393	3,791	3,892
3	4,443	4,505	4,478	4,16	4,089	4,183	4,234	3,738	3,803
4	3,995	4,068	4,026	4,178	4,094	4,164	4,189	3,677	3,76
5	3,878	3,967	3,925	4,078	4,1	4,172	4,195	3,671	3,763
6	3,805	3,892	3,928	4,093	4,09	4,189	4,208	3,695	3,783
7	3,825	3,899	3,94	4,089	4,075	4,154	4,197	3,668	3,768
8	3,746	3,725	3,63	3,785	3,762	3,836	3,887	3,673	3,779
9	3,755	3,727	3,622	3,75	3,728	3,796	3,868	3,636	3,758
10	4,124	3,91	3,696	3,669	3,64	3,72	3,779	**3,475**	3,6
11	4,006	3,844	3,657	3,663	3,636	3,726	3,791	3,513	3,625
12	3,851	3,847	3,738	3,744	3,712	3,796	3,862	3,547	3,669

N=100, 5000 replicações

Tabela 10: Modelo 6, Reponderação Aumentada

Erros Quadráticos Médios Simulados									
K/L	1	2	3	4	5	6	7	8	9
1	6,313	5,336	4,547	4,25	4,201	4,463	4,519	3,826	3,906
2	5,319	5,286	4,574	4,295	4,2	4,462	4,519	3,827	3,907
3	4,443	4,547	4,536	4,281	4,212	4,312	4,363	3,832	3,907
4	3,995	4,113	4,096	4,28	4,203	4,275	4,304	3,76	3,87
5	3,878	4,039	4,026	4,218	4,238	4,315	4,339	3,766	3,894
6	3,805	3,979	4,044	4,268	4,26	4,362	4,377	3,806	3,933
7	3,825	3,993	4,052	4,267	4,262	4,33	4,364	3,778	3,899
8	**3,746**	3,824	3,765	4,037	4,024	4,067	4,127	3,847	3,986
9	3,755	3,848	3,799	4,045	4,032	4,066	4,149	3,851	3,999
10	4,124	4,319	4,151	4,46	4,43	4,361	4,473	3,875	4,049
11	4,006	4,183	4,047	4,407	4,399	4,386	4,474	3,925	4,072
12	3,851	3,999	3,958	4,203	4,214	4,321	4,313	3,823	3,961

N=100, 5000 replicações

Tabela 11: Estimativas usando *propensity score* verdadeiro

Erros Quadráticos Médios Simulados										
Modelo	N\L	1	2	3	4	5	6	7	8	9
1,β~RP	100	1,348	0,906	0,87	**0,856**	0,898	0,992	1,851	3,899	>10
1,β~RP	200	0,661	0,459	0,442	**0,43**	0,434	0,434	0,446	0,448	0,502
2,β~RP	200	0,741	0,436	**0,428**	0,428	0,432	0,436	0,449	0,463	0,506
3,β~RP	200	0,51	0,494	0,501	0,491	0,499	**0,484**	0,519	0,484	0,789
4,β~RP	200	0,709	0,471	0,447	**0,437**	0,438	0,438	0,442	0,447	0,488
5,β~RP	200	0,656	0,452	0,436	**0,429**	0,432	0,432	0,442	0,452	0,484
6,β~RP	100	4,58	4,633	4,573	4,299	4,325	4,427	4,452	**3,853**	3,873
6,β~RP	100	4,58	4,687	4,649	4,408	4,451	4,587	4,602	3,946	**3,924**

WORKING PAPER 3/2013 • JAN, 2013 •