Transparência em Pesquisa Estrutural: Andrews, Gentzkow e Shapiro
Notas de Aula

Victor Gomes
Universidade de Brasília

27 de outubro de 2020

Estimadores de modelos não-lineares com múltiplas interações de agentes ou setores podem ser funções complicadas dos dados e portanto difíceis dos leitores entenderem. Estes modelos formam uma parte importante do arsenal dos economistas para analisarem problemas do mundo real. Um dos problemas apontados aos modelos estruturais é a falta de transparência. A falta de transparência pode fazer parte da natureza destes modelos, que muitas vezes usam métodos computacionais complexos necessários para implementação (veja Heckman, 2010).

Andrews, Gentzkow e Shapiro em vários artigos mostram a sensibilidade de modelos estruturais (não-lineares).

Pesquisa empírica estrutural é uma caixa-preta?

O valor da transparência em econometria estrutural: definição formal de *transparência*. Transparência é definida como uma propriedade das estatísticas que os pesquisadores relatam.

Transparência é distinta de identificação. Também é distinta de medidas tradicionais de robustez. Existem outras medidas de transparência, mas aqui o conceito se aplica a modelos estruturais.

Modelo Comunicação Científica

Andrews, Gentzkow e Shapiro propõem um modelo de comunicação científica. Conjuntamente com testes de robustez e eficiência, pesquisadores devem se preocupar com transparência quando estimando modelos estruturais.

No modelo, os pequisadores observam dados informativos sobre uma quantidade de interesse c. O pesquisador apresenta uma estimativa $\hat{c}$ junto com estatísticas auxiliares $\hat{t}$ para um cojunto de leitores indexados por r.

1 Um Modelo de Transparência

Um pesquisador observa dados $D \in \mathcal{D}$. O pesquisador assume um conjunto de hipóteses a_0 sob as quais $D \sim F(a_0, \eta)$ para $\eta \in \mathcal{H}$ ser um parâmetro desconhecido.

O pesquisador calcula uma estimativa pontual $\hat{c} = \hat{c}(D)$ de uma quantidade de interesse escalar $c(a_0, \eta)$, conjuntamente com um vetor de estatísticas auxiliares $\hat{t} = \hat{t}(D)$. Estas estatísticas auxiliares podem incluir evidência descritiva, análise de sensibilidade, e várias estatísticas auxiliares (vamos ver à frente quais).

O pesquisador apresenta $(\hat{c}, \hat{t})$ aos leitores $r \in \mathcal{R}$, que por sua vez não possuem acesso aos dados utilizados na pesquisa. Primariamente os autores consideram que a dimensão da estatística auxiliar é muito menor do que os dados utilizados pelo pesquisador: $\dim(\hat{t}) << \dim(D)$. Dado isto, se pergunta ao leitores (r) o que eles aprendem de $(\hat{c}, \hat{t})$.

Preocupação dos leitores: o modelo do pesquisador é mal especificado? Leitores consideram que hipóteses $a \in \mathcal{A}$ podem ser diferentes de a_0 (o conjunto de hipóteses originais do pesquisador). Sob as hipóteses $a \in \mathcal{A}$, $D \sim F(a_0, \eta)$ para η desconhecido, e a quantidade de interesse é $c(a, \eta)$.

Cada leitor r possuir uma *prior* π_r sobre o conjunto de hipóteses a e o parâmetro do modelo η, e deseja estimar $c(a, \eta)$, escolhendo a decisão $d_r \in \Re$ e incorrendo na função perda quadrática:

$$L(d_r, c(a, \eta)) = (d_r - c(a, \eta))^2.$$

Defina o *risco de comunicação* do leitor r de $(\hat{c}, \hat{t})$ como sua perda esperada ex-ante de tomar a ação ótima baseada em $(\hat{c}, \hat{t})$. Sob o quadrado do erro de perda esta ação ótima é simplesmente a média posterior do leitor r para c dado $(\hat{c}, \hat{t})$:

$$\mathrm{E}_r \left[\min_{d_r} \mathrm{E}_r \left[(d - c(\cdot))^2 \mid \hat{c}, \hat{t} \right] \right] = \mathrm{E}_r \left[\mathrm{Var}_r(c \mid \hat{c}, \hat{t}) \right]. \tag{1}$$

Aqui $\mathrm{E}_r[\cdot]$ e $\mathrm{Var}_r(\cdot)$ representam a expectativa e a variância sob π_r, respectivamente. c é na verdade $c(a, \eta)$.

O risco de leitor r de observar os dados completos é

$$\mathrm{E}_r \left[\mathrm{Var}_r(c \mid D) \right] \leq \mathrm{E}_r \left[\mathrm{Var}_r(c \mid \hat{c}, \hat{t}) \right] \leq \mathrm{Var}_r(c)$$

A primeira parte da desigualdade é satisfeita com igualdade apenas se a média posterior de r baseada em $(\hat{c}, \hat{t})$ é quase certamente a mesma se fosse baseada em todos os dados.

Andrews, Gentzkow e Shapiro definiem *transparência de* $(\hat{c}, \hat{t})$ *para* r como a redução no risco de comunicação de observar $(\hat{c}, \hat{t})$ relativo à redução de observar todos os dados:

$$T_r\left(\hat{c}(\cdot), \hat{t}(\cdot)\right) = \frac{\mathrm{Var}_r(c) - \mathrm{E}_r\left[\mathrm{Var}_r(c \mid \hat{c}, \hat{t})\right]}{\mathrm{Var}_r(c) - \mathrm{E}_r\left[\mathrm{Var}_r(c \mid D)\right]} \tag{2}$$

Transparência é definida como um quando o denominador for zero. Portanto

$$0 \leq T_r\left(\hat{c}(\cdot), \hat{t}(\cdot)\right) \leq 1$$

Transparência é igual a 1 quando observando $(\hat{c}, \hat{t})$ resulta no mesmo risco para r quando ele observa todos os dados. Transparência á igual a zero quando observando $(\hat{c}, \hat{t})$ não resulta em redução de risco, quando observando D resultaria em alguma redução.

Os autores argumentam que em alguns casos é "direto" construir relatórios totalmente transparentes, i.e. $T_r\left(\hat{c}(\cdot), \hat{t}(\cdot)\right) = 1$. Se $\hat{t}$ é suficiente para (a, η), então $(\hat{c}, \hat{t})$ é totalmente transparente para todos os leitores.

Quando não for possível relatar uma estatística suficiente, nós ainda podemos construir relatório totalmente transparente para o leitor r, utilizando o fato de que a média posterior do leitor é $\hat{t} = \mathrm{E}_r[c \mid D]$. Neste caso, $(\hat{c}, \hat{t})$ não precisa ser transparente para os leitores r' com $\pi_{r'} \neq \pi_r$. A heterogeneidade entre os leitores é fundamental para o estudo da transparência.

1.1 Exemplo: IV

Andrews, Gentzkow e Shapiro (2020) apresentam um exemplo de variáveis instrumentais. Suponha os seguintes dados: $D = \{Y_i, X_i, Z_i\}_{i=1}^n$; são observações de uma variável resultado Y_i, um regressor endógeno X_i e um candidato a instrumento Z_i, sendo todas escalares.

Os leitores acreditam que os dados seguem:

$$Y_i = X_i c + Z_i a + \varepsilon_i \tag{3}$$

$$X_i = Z_i \gamma + V_i \tag{4}$$

tal que os instrumentos Z_i são fixados. O erro forma-reduzida de se estimar Y_i como função de Z_i é

$$U_i = cV_i + \varepsilon_i$$

Se assume que os erros (U_i, V_i) são iid normal ao longo de i, $(U_i, V_i) \sim N(0, \Sigma)$, com Σ comummente conhecido, tal que o parâmetro $\eta = (c, \gamma) \in$

$\Re^2$. Suponha que $\mathcal{A} = \Re$, tal que hipóteses $a \in \mathcal{A}$ correspondem ao coeficiente em Z_i em (3), e que a hipótese do pesquisador é $a_0 = 0$. Sob hipóteses a_0, Z_i é um instrumento valido na regressão de Y_i sobre X_i, enquanto que sob $a \neq 0$, a exclusão de restrição falha.

Represente a estimativa IV usual como

$$\hat{c} = \frac{\sum Z_i Y_i}{\sum Z_i X_i}$$

e o coeficiente de primeiro estágio como

$$\hat{\gamma} = \frac{\sum Z_i X_i}{\sum Z_i^2}$$

O relatório sobre $\hat{c}$ pode não ser totalmente transparente. Por exemplo, considere um leitor r que possui uma intepretação a priori sobre $a \neq a_0$, mas uma prior conjunta contínua sobre (c, γ).[1] Como o leitor é incerto sobre o valor de γ, entrentanto, ele não pode inferir o valor de c da estimativa $\hat{c}$, mesmo com amostra grande. Por outro lado, com acesso aos dados completos ele poderia descobrir o valor de γ e assim ser capaz de inferir c.

Então, a medida que $n \to \infty$, a transparência de $\hat{c}$ para este leitor é limitada a partir de 1. Por outro lado, o relatório de $(\hat{c}, \hat{\gamma})$ possui transparência $T_r = 1$ para todos os leitores r, a medida que $(\hat{c}, \hat{\gamma})$ é suficiente para os parâmetros desconhecidos (c, γ). Neste exemplo, relatar a estatística auxiliar $\hat{\gamma}$ aumenta a transparência.

2 Análise Descritiva

O primeiro elemento que contribui para a transparência em pesquisa estrutural é a análise descritiva. Na abordagem de AGS, uma análise descritiva inclui estatísticas auxiliares $\hat{s}$ à $\hat{t}$ – estas estatísticas são diretamente informativas sobre c ou informativa sobre a plausibilidade das hipóteses a_0.

Exemplos de estatísticas incluem sumários, visualização dos dados ou correlações que ilustrem as principais relações causais. Algumas vezes estas evidências devem ser livres de modelos, no sentido de que a interpretação não precisa contar hipóteses do modelo estrutural. Pakes (2014) formaliza o papel da análise descritiva em fornecer um conjunto de fatos que o modelo estrutural deve racionalizar.

Os autores sugerem duas formas de análise descritiva para transparência:

[1] Observe que para n grande, $\hat{c}$ converge em probabilidade para $c + a/\gamma$, sob condições moderadas.

1. A estatística descritiva $\hat{s}$ pode fornecer evidência sobre c, que é informativa sob uma leque de informações de hipóteses além de a_0. Neste caso, a estatística deveria descrever o que se estima e o que está ocorrendo, por exemplo.

2. A estatística $\hat{s}$ pode fornecer evidência que ajuda ao leitor r julgar as hipóteses a_0.

Um exemplo da segunda estatística: Allcott et al (2019) estimam um modelo estrutural de demanda por grãos que permite a eles decompor as fontes de desigualdade nutricional nos EUA. Para estimar a sensitividade ao preço (elasticidade), os autores usam como instrumentos os preços de um produto em uma dada loja com o preço do mesmo produto em outras lojas da mesma cadeia (instrumentos de Hausman). A restrição de exclusão aqui é que a variação nos preços devido a composição das cadeias em um mercado particular é ortogonal as diferenças de preferências que não são observadas. Na análise descritiva, os autores buscam mostrar que as hipóteses são razoáveis mostrando que a variação nos preços é ortogonal as variáveis demográficas que são usadas para prever as escolhas.

3 Identificação

Discussão explícita sobre identificação é importante para aumentar a transparência. Uma quantidade c é identificada no modelo do pesquisador se

$$c(a_0, \eta) \neq c(a_0, \eta') \text{ implica em } F(a_0, \eta) \neq F(a_0, \eta'). \tag{5}$$

Valores distintos de c correspondem a distribuições distintas dos dados sob as hipóteses adotadas pelo pesquisador. Uma quantidade c é identificada por um vetor específico de estatísticas $\hat{s}$ se $c(a_0, \eta) \neq c(a_0, \eta')$ implica em distribuição distinta de $\hat{s}$ sob $F(a_0, \eta)$ e $F(a_0, \eta')$.

AGS acreditam que discussões claras e precisas sobre identificação tem um papel importante na transparência de pesquisa estrutural. Eles argumentam que grande parte do uso recente da idéia de identificação está mais relacionado a estratégia de uso do modelo empírico e a escolha dos dados, i.e. em estimar, do que com a definição clássica de identificação econométrica.

Nesta abordagem, as discussões de identificação podem ser vistas como uma forma de comunicar e "clarificar" as implicações primárias a_0 e o espaço das alternativas relevantes $a \neq a_0$, permitindo aos leitores formar *prior* π_r mais precisas. Discussões que deixam clara a identificação utilizada são importantes na comunicação com o leitor.

Discussões de identificação podem iluminar a forma que as hipóteses do modelo mapeiam a distribuição das variáveis observáveis para a quantidade c. Este tipo de estratégia é um complemento direto à inspeção das hipóteses em termos matemáticos.

3.1 Exemplo: Escolha Discreta

Suponha um modelo de escolha discreta que assume que a utilidade de um consumidor i pelo bem j contém um termo de erro aditivo ε_{ij}, que é iid valor extremo tipo 1. Como um leitor que não é familiar com tais modelos, poderia avaliar a hipótese distribucional sobre o termo de erro?

A hipótese é a de que a função de densidade cumulativa (FDC) de ε_{ij} é $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. O uso desta função pode trazer algumas duvidas, mas um gráfico da FDC mostraria que esta função não é muito diferente de uma normal (com caudas mais grossas). Para um leitor, pode parecer desafiador se esta hipótese é uma representação razoável da utilidade do consumidor, ou em quais circunstâncias isto seria uma aproximação pior ou melhor, apenas pela análise da fórmula ou do gráfico da FDC.

Estudar as implicações da hipótese de valor extremo para identificação é na verdade instrutivo. Impor esta forma aos erros pode significar que a participação dos consumidores escolhendo o bem j é suficiente (sozinha) para identificar:

1. a elasticidade-preço própria relativa e o markup para dois bens j e k;
2. como os consumidores realocam se qualquer bem é removido do conjunto de escolha;
3. bem-estar relativo do consumidor sob conjuntos de escolha diferentes.

Um leitor não familiarizado que soubesse destas implicações poderia entender que a hipótese é mais forte do que se imaginava e que valeria a pena maior análise do modelo. O leitor também poderia ser a capacidade de formar novas intuições sobre quais hipóteses alternativas a são as mais relevantes para se considerar – por exemplo, distribuições alternativas de erro que decompõem os padrões de substituição a partir dos market shares (BLP, 1995).

3.2 Recomendação

AGS sugerem dois princípios para transparência e identificação:

1. As discussões deveriam ser precisas – *identificação* usada apenas no sentido formal de econometria. Ex: declarações sobre identificação que

não são óbvias deveriam ser acompanhadas por uma prova formal ou apresentada explicitamente como conjectura.

A afirmação *c é identificado por* um vetor particular de estatísticas $\hat{s}$ significa que a distribuição de $\hat{s}$ é suficiente para inferior o valor de c utilizando o modelo. Se esta declaração se aplica apenas dado o conhecimento de algum outro parâmetro, então isto deve ser tornado explícito.

2. Discussão de identificação ser claramente distinta da estimação. Como c e $\hat{s}$ é relacionado no modelo é distinto de como $\hat{s}$ é relacionado com o estimador específico $\hat{c}$. Além disso, a afirmação "c é identificado por $\hat{s}_j$" não precisa implicar que $\hat{s}_j$ é um determinante importante de $\hat{c}$. A transparência é geralmente melhor quando a discussão sobre identificação elucida as mesmas relações que venham a ser importantes na estimação.

4 Estimação

Como tornar a estimação estrutural mais transparente.

Saber a forma do estimador $\hat{c}$ é importante para a transparência por duas razões:

1. Para mostrar a influência das estatísticas $\hat{s}$, que comandam o estimador $\hat{c}$, na violação de hipóteses (qual hipótese importa mais e qual a violação de hipótese tem maior efeito);

2. Saber como o estimador depende da estatística $\hat{s}$ pode ajudar ao leitor a julgar a chance de incorrer em viés devido a violações específicas.

Foco no valor das estatísticas $\hat{s}$ que determinam o estimador $\hat{c}$ – determinação exata ou aproximada – bem como na relação entre $\hat{s}$ e $\hat{c}$. Sem perda de generalidade podemos escrever

$$\hat{c} = h(\hat{s}) + v_h \tag{6}$$

tal que $h(\cdot)$ é alguma função e v_h é um resíduo cuja estrutura depende de $h(\cdot)$.

1) Se discute a escolha do estimador $\hat{c}$ tal que $v_h = 0$ e então a caracterização da função $h(\cdot)$.

2) Discutir é a escolha do estimador $\hat{c}$ tal que $v_h \neq 0$ e então demonstrar que v_h é pequeno e apropriado no sentido de que $\hat{c} \approx h(\hat{s})$

4.1 Estatísticas Descritivas em Estimação

Qual o papel de $\hat{s}$ na estimação: $\hat{c} = h(\hat{s})$. Isto é mais importante quando c é identificado por $\hat{s}$.

Grande transparência: quando c é identificado pelo valor da população s de $\hat{s}$ e a relação $c = \Gamma(s, a)$ pode não variar entre as alternativas de interesse a (i.e. $c = \Gamma(s)$ para todo $a \in \mathcal{A}$). Neste caso, o estimador $\hat{c} = h(\hat{s}) = \Gamma(\hat{s})$ pode ter grande transparência para todos os leitores.

Na prática, a estimação baseada em um vetor alvo de estatísticas descritivas $\hat{s}$ é geralmente impletando via alguma forma de estimador de distância mínima que escolhe os parâmetros para parear o $\hat{s}$ observado do valor previsto sob o modelo de interesse. A transparência fornece uma justificativa potencial para a escolha de tais estimadores mesmo quando estimadores mais eficientes são disponíveis (como MLE).

Exemplo: Modelo IV. Suponha que os dados consistem em n draws i.i.d. $\{Y_i, X_i, Z_{i,1}, ..., Z_{i,J}\}_{i=1}^n$ para $Z_i = (Z_{i,1}, ..., Z_{i,J})$ ser um vetor de J instrumentos mutuamente ortogonal e média-zero proposto pelo pesquisador.

Para $c \in \Re$ e $a \in \mathcal{A} = \Re^J$, o dados seguem:

$$Y_i = X_i c + Z_i' a + \varepsilon_i \tag{7}$$

onde agora tratamos os instrumentos Z_i como aleatórios e permitimos que o termo de erro ε_i seja não-normal. Faça G representar a distribuição conjunta de $(X_i, Z_i, \varepsilon_i)$ e assuma que todos os leitores acreditam que $G \in \mathcal{G}$ para alguma classe de distribuições com

$$\mathrm{E}_G[Z_i \varepsilon_i] = 0$$

para todo $G \in \mathcal{G}$.

Os instrumentos são válidos sob a hipótese do pesquisador, $a_0 = 0$, mas os leitores suspeitam que ela pode ser inválida.

Suponha que cada instrumento j possui um coeficiente de primeiro-estágio que é não-nulo: $\mathrm{E}_G[Z_{j,i} X_i] \neq 0$ para todo $G \in \mathcal{G}$. Sob a distribuição G, parâmetro verdadeiro c, e hipóteses a, a probabilidade limite do estimador de variáveis instrumentais baseado no j-ésimo instrumento sozinho $\hat{c}_j = \sum Z_{i,j} Y_i / \sum Z_{i,j} X_i$ é

$$\frac{\mathrm{E}_{(G,a,c)} Z_{i,j} Y_i}{\mathrm{E}_G \sum Z_{i,j} X_i} = c + \frac{\mathrm{E}_G[Z_{i,j}^2]}{\mathrm{E}_G[Z_{i,j} X_i]} a = c + \frac{a_j}{\gamma_j} \tag{8}$$

para

$$\gamma_j = \frac{\mathrm{E}_G[Z_{i,j}X_i]}{\mathrm{E}_G[Z_{i,j}^2]} \tag{9}$$

que é o primeiro estágio do instrumento j.

4.1.1 Regressão IV Sobre-identificada

AGS assumem que a variável descritiva $\hat{s}$ consiste nos primeiros m coeficientes dos instrumentos isolados (um de cada vez) $\hat{s} = (\hat{c}_1, ..., \hat{c}_m)$ para $m < J$. Eles assumem que os leitores tem fortes *priors* sobre o viés destas m estimativas: o leitor r acredita que os m primeiros elementos do vetor de viés $b = (a_1/\gamma_1, ..., a_J/\gamma_J)'$ igual a um vetor b_r com probabilidade um, $\mathrm{Prob}_r\{(a_1/\gamma_1, ..., a_J/\gamma_J)' = b_r\} = 1$. Portanto, todos os autores estão certos sobre o viés usando os primeiros m instrumentos, enquanto são incertos sobre os demais instrumentos.[2]

Neste caso, o estimador que tem como alvo a estatística descritiva $\hat{s}$ pode ser mais transparente do que estimador MLE sob as hipóteses $a_0 = 0$. Por exemplo, suponha que o tamanho da amostra seja grande o suficiente que $\hat{c}$ é aproximadamente normal e se negligenciamos o erro de aproximação para obter

$$\hat{c} = \iota c + b + \xi, \; \xi \sim N(0, \Omega), \tag{10}$$

para ι o vetor de 1s.

Neste modelo assintótico $\eta = (c, \gamma, \Omega)$. Suponha além disso que o pesquisador observa apenas $D = (\hat{c}, \Omega)$, que c e (b, Ω) são independentes de sob π_r para todo $r \in \mathcal{R}$, e que Ω é comumente conhecida.

Faça

$$\hat{c}_0 = \left(\iota'\Omega^{-1}\iota\right)^{-1} \iota'\Omega^{-1}\hat{c} \tag{11}$$

ser o estimador MLE sob a hipótese $a_0 = 0$, e faça $\hat{c}_s$ representar o estimador que eficientemente minimiza a distância entre S_c e $\hat{s} = S\hat{c}$ para S a matriz de seleção tal que $S\hat{c}$ seleciona os primeiros m elementos de $\hat{c}$. A variância de $\hat{c}_0$ dado c sob π_r é

$$\mathrm{Var}_r(\hat{c}_0 \mid c) = \underbrace{\left(\iota'\Omega^{-1}\iota\right)^{-1}}_{\textbf{(1)}} + \underbrace{\left(\iota'\Omega^{-1}\iota\right)^{-2} \iota'\Omega^{-1}\mathrm{Var}_r(b)\Omega^{-1}\iota}_{\textbf{(2)}} \tag{12}$$

tal que:

[2]Isto pode ocorrer em virtude do viés potencial dos m primeiros instrumentos. Isto porque estes instrumentos são possuem muita credibilidade e $b_r = 0$, ou porque o pesquisador não possui dúvidas sobre a respeitos de viés potencial no uso dos instrumentos – esse ponto pode ter ficado claro para o leitor através de análise descritiva ou pela discussão de identificação.

(1) é a variância amostral da estimativa MLE;

(2) reflete a invalidade do instrumento.

Por contraste, a variância de $\hat{c}_S$ dado c sob π_r, é simplesmente a variância amostral de $\hat{c}_S$. Quando o leitor r é muito incerto sobre a validade dos instrumentos (no sentido de que a variância dos últimos $J-m$ elementos de b é muito grande), $\hat{c}_S$ pode ser mais transparente do que $\hat{c}_0$. Isto parece ser intuitivo para o caso onde $b_r = 0$. tal que o leitor r acredita que os primeiros m instrumentos são válidos. Isso pode permanecer verdade mesmo quando $b_r \neq 0$. Portanto, o que é importante para a transparência neste *setup* é que os primeiros m instrumentos são válidos, mas que os leitores tem credos precisos sobre o viés que estes instrumentos induzem.

Relação de (12) com outros problemas: A equação (12) pode ser interpretada como como um modelo de regressão para $Y = \hat{c}$, com variável omitida b. Este modelo pode ser entendido como uma aproximação assintótica ao GMM com má-especificação local.

Comunicar aos leitores a geometria de $h(\cdot)$ aumenta transparência. Sabendo que o estimador tem a forma $\hat{c} = h(\hat{s})$ significa que os leitores sabem que o estimador depende dos dados apenas por meio de $\hat{s}$, mas eles podem não saber a natureza da dependência. Existem muitas funções diferentes $h(\cdot)$ que constituem estimadores validos.[3]

Algumas destas funções $h(\cdot)$ podem convencer um grande grupo de leitores, enquanto outras destas funções apenas convencem quem aceita $a_0 = 0$.

Na prática, os pesquisadores mostram uma definição formal do estimador usado, mas mesmo com uma definição pode não ser óbvia a geometria de $h(\cdot)$. Mesmo em modelos lineares esta caracterização pode ser difícil, e.g. papers com regressão em descontinuidade e estimador com dois efeitos fixos podem sofrer críticas por usarem dados que não são bem avaliados por muitos leitores.

Em modelos não-lineares a caracterização é mais difícil. Uma solução poderia ser uma descrição completa de $h(\cdot)$ por força bruta, embora isso não seja sempre possível.

Sensibilidade. AGS (2017) propõem focar na *sensibilidade local* do estimador em relação à estatísticas "alvo" na estimação. Sensibilidade corresponde às derivadas de $h(\cdot)$ quando $h(\cdot)$ é diferenciável. É possível aproximar

[3] Quais transformações de s ou subconjuntos diferentes são suficientes para identificar c.

esta derivada numericamente. Exemplo: calculando o estimador nas pertubações com a forma $\hat{s} + \varepsilon e_j$ para ε sendo um número pequeno e e_j o vetor base-padrão (com j linhas), e então computando a derivada númerica

$$\frac{h(\hat{s} + \varepsilon e_j) - h(\hat{s})}{\varepsilon}$$

A estimação repetida pode ser computavelmente demandante, mas AGS (2017) mostram que em muitas aplicações a estimação repetida não é necessária se o leitor está disposto a focar no valor assintótico da derivada.

Continuação do exemplo. Os primeiros m elementos de b são conhecidos sob π_r, mas agora suponha que Ω pode não ser comumente conhecido. O estimador $\hat{c}_S$ pode ser escrito como

$$\begin{aligned} \hat{c}_S &= \left(\iota' S'(S\Omega S')^{-1} S\iota\right)^{-1} \iota' S'(S\Omega S')^{-1}\hat{s} \\ &= \Lambda_S \hat{s} \end{aligned} \tag{13}$$

para $\hat{s} = S\hat{c}$, tal que Λ_S é a sensibilidade de $\hat{c}_S$ em relação a $\hat{s}$ (AGS, 2017).

Relatando $(\hat{c}_S, \sigma_S, \Lambda_S)$ – i.e. tomando $\hat{c} = \hat{c}_S$ e $\hat{t} = (\sigma_S, \Lambda_S)$, σ_S é o erro-padrão de $\hat{c}_S$ – é fracamente mais transparente para todos os leitores r do que relatar apenas $(\hat{c}_S, \sigma_S)$.

Argumento de transparência de AGS (2017). Suponha que o leitor r possui uma *prior*-normal sobre c, $c \sim N(0, \omega_r^2)$. A variância posterior média para o leitor r baseado no conjunto completo dos dados é portanto limitada inferiormente por $\mathrm{E}_r[\mathrm{Var}_r(c \mid D, b)] = \mathrm{E}_r\left[(\omega_r^{-2} + \sigma_0^{-2})^{-1}\right]$ para todo σ_0 o erro-padrão usual de $\hat{c}_0$, enquanto a variância posterior média baseada na observação de $(\hat{c}_S, \sigma_S, \Lambda_S)$ é $\mathrm{E}_r\left[(\omega_r^{-2} + \sigma_0^{-2})^{-1}\right]$. Isto limita a transparência de reportar $(\hat{c}_S, \sigma_S, \Lambda_S)$ por baixo.

Por outro lado, a transparência de $(\hat{c}_S, \sigma_S)$ vai a zero a medida que

$$\mathrm{Var}_r(\Lambda_S b_r) \to \infty.$$

Intuitivamente, mesmo se o leitor conhece o viés b_r das estimativas baseada nos primeiro m instrumentos são combinados para formar $\hat{c}_S$. Na falta de tal conhecimento, incerteza sobre como o viés em $\hat{s}$ se traduz em viés em $\hat{c}_S$ leva a $\hat{c}_S$ sem ser informativo quando b_r é grande.

4.2 Mostre Quanto o Estimador Depende da Estatística Descritiva

Baseando estimativas diretamente na estatística $\hat{s}$ pode não ser possível ou desejável. Quando isto for verdade, $\hat{c} = h(\hat{s}) + v_h$ para v_h não necessariamente igual a zero. Portanto, deixando claro para os leitores a magnitude de v_h bem como a forma de $h(\cdot)$ pode aumentar a transparência.

Um exemplo é onde pelo menos alguns leitores acreditam que $c = \Gamma(s, a)$, no caso em que eles podem encontrar $\hat{c}$ especialmente informativo quando $v_h \approx 0$ e $h(\cdot) \approx \Gamma(\cdot, a)$.

Caracterizar a relação amostra-finita entre $\hat{c}$ e $\hat{s}$, tanto analiticamente ou numericamente, pode ser difícil.[4]

AGS (2020b) mostram que, sob condições assintóticas relacionadas com as hipóteses de AGS (2017), muitos estimadores comuns podem ser representados na forma

$$\hat{c} \approx \text{constante} + \Lambda\hat{s} + v \tag{14}$$

para Λ um análogo a sensibilidade local (como definido em AGS-2017) e v assintoticamente não-correlacionado com $\hat{s}$.

AGS (2020b) propõem medir o tamanho de v pela *informatividade local* de $\hat{s}$ para $\hat{c}$, que é dado por

$$\Delta = \frac{\mathrm{Avar}(\Lambda\hat{\mathrm{s}})}{\mathrm{Avar}(\hat{\mathrm{c}})} = 1 - \frac{\mathrm{Avar}(\mathrm{v})}{\mathrm{Avar}(\hat{\mathrm{c}})} \tag{15}$$

para Avar(V) a variância assintótica da variável aleatória V.

Quando a informatividade local $\Delta = 1$, então $v = 0$ e retornamos as considerações da seção anterior. Quando $\Delta = 0$, $\hat{c}$ é assintoticamente independente de $\hat{s}$, no caso em que o leitor acreditaria que $c = \Lambda(s, a)$ pode *não* encontrar que $\hat{c}$ seja muito informativo sobre c.

AGS (2020b) mostram que é geralmente possível aproximar a sensibilidade local e a informatividade local sem ser necessário simulações ou estimações adicionais do modelo estrutural. Além disso, embora ambas sensibilidade local e informatividade local podem depender do processo gerador de dados, AGS mostram que as aproximações que eles usam funcionam sob violações locais das hipóteses do pesquisador, significando que estes objetos podem ser interpretados mesmo se o leitor não tem confiança completa no modelo do pesquisador.

AGS também estabelecem condições sob as quais maior informatividade Δ corresponde a maior redução no viés do pior caso do estimador $\hat{c}$ de aceitar a relação implicada pelo modelo entre c e o valor populacional da estatística descritiva $\hat{s}$.

[4]Exemplo: exploração numérica pela repetido sorteio de dados D de um ou mais processo gerador dos dados e então computar os valores "implicados" por $\hat{c}$ e $\hat{s}$ pode ser muito demandante em termos computacionais.

4.3 Implementação

Em um grande espectro de aplicações, estimativas convenientes de $\hat{\Sigma}$ de Σ são disponíveis a partir de resultados assintóticos padrão ou por meio de bootstrap. Dadas tais estimativas, se pode construir uma estimativa "plug-in"

$$\hat{\Delta} = \frac{\hat{\Sigma}_{c\gamma}\hat{\Sigma}_{\gamma\gamma}^{-1}\hat{\Sigma}_{\gamma c}}{\hat{\sigma^2}} \tag{16}$$

dado que $\hat{\Sigma}$ é consistente sob $S(0,0)$, consistência de $\hat{\Sigma}$ e $\hat{\Delta}$ sob as sequências de análise seguem imediatamente sob as hipóteses mantidas de AGS (2020b) que $\sigma_c^2 > 0$ e $\hat{\Sigma}_{\gamma\gamma}$ é posto completo. (16) é a contrapartida empírica da definição de informatividade de AGS (2020b).

Segue a hipótese adicional: $\hat{\Sigma} \underset{p}{\to} \Sigma$ sob $S(0,0)$. Sob a hipótese anterior e a hipótese 3 de AGS (2020b), $\hat{\Sigma} \underset{p}{\to} \Sigma$ e $\hat{\Delta} \underset{p}{\to} \Delta$ sob $S(h,z)$ para qualquer $h \in \mathcal{H}$.

4.3.1 Implementação: Estimação Distância Mínima

O estimador de distância mínima é um caso especial importante que é útil para um grande número de aplicações.

Formalmente suponha que podemos decompor $\eta = (\theta, \omega)$, tal que θ é de dimensão-finita e $c(\eta)$ depende de η apenas por meio de θ, tal que podemos escrever isto como $c(\theta)$. Assumimos que $c(\theta)$ é continuamente diferenciável em θ.

O pesquisador forma uma estimativa $\hat{c} = c(\hat{\theta})$, tal que $\hat{\theta}$ soluciona

$$\min_{\theta} \left\{ \hat{g}(\theta)'\hat{W}\hat{g}(\theta) \right\} \tag{17}$$

para $\hat{g}(\theta)$ ser um vetor de momentos de dimensão k_g e $\hat{W}$ a matriz de ponderação com dimensão $k_g \times k_g$.

O pesquisador também calcula $\hat{\gamma}$ solucionando:

$$\min_{\gamma} \left\{ \hat{m}(\gamma)'\hat{U}\hat{m}(\gamma) \right\} \tag{18}$$

para $\hat{m}(\gamma)$ ser um vetor de momentos com dimensão k_m e $\hat{U}$ a matriz de ponderação com dimensão $k_m \times k_m$.

Dado que $\hat{W}$ e $\hat{U}$ convergem em probabilidade para os limites W e U, enquanto que $\sqrt{n}\hat{g}(\theta(\eta_0))$ e $\sqrt{n}\hat{m}(\gamma(\eta_0))$ são conjuntamente assintoticamente normal sob $S(0,0)$,

$$\sqrt{n}\begin{pmatrix} \hat{g}(\theta(\eta_0)) \\ \hat{m}(\gamma(\eta_0)) \end{pmatrix} \rightarrow_d N\left(0, \begin{pmatrix} \Sigma_{gg} & \Sigma_{gm} \\ \Sigma_{mg} & \Sigma_{mm} \end{pmatrix}\right)$$

resultados existentes implicam que sob $S(0,0)$ e condições de regularidade padrão,

$$\sqrt{n}\begin{pmatrix} \hat{c} - c(\theta(\eta_0)) \\ \hat{\gamma} - \gamma(\eta_0) \end{pmatrix} \rightarrow_d N(0, \Sigma)$$

com

$$\Sigma = \begin{pmatrix} \Lambda_{cg} & 0 \\ 0 & \Lambda_{\gamma m} \end{pmatrix}\begin{pmatrix} \Sigma_{gg} & \Sigma_{gm} \\ \Sigma_{mg} & \Sigma_{mm} \end{pmatrix}\begin{pmatrix} \Lambda_{cg} & 0 \\ 0 & \Lambda_{\gamma m} \end{pmatrix}'$$

tal que $\Lambda_{cg} = -C(G'WG)^{-1}G'W$ e $\Lambda_{\gamma m} = -(M'UM)^{-1}M'U$ são sensibilidades de $\hat{c}$ com respeito a $\hat{g}(\theta(\eta_0))$ e de $\hat{\gamma}$ com respeito a $\hat{m}(\gamma(\eta_0))$ como definido em AGS (2017), e

$$C = \frac{\partial}{\partial \theta} c(\theta(\eta_0))$$

Por sua vez, C pode ser consistentemente estimado por

$$\hat{C} = \frac{\partial}{\partial \theta} c(\hat{\theta})$$

Se $\hat{g}(\theta)$ e $\hat{m}(\gamma)$ são continuamente diferenciáveis então sub certas condições de regularidade podemos estimar G consistentemente por $\hat{G} = \frac{\partial}{\partial \theta}\hat{g}(\hat{\theta})$ e M por $\hat{M} = \frac{\partial}{\partial \gamma}\hat{m}(\hat{\gamma})$. Portanto, dado as estimativas consistentes de $\hat{\Sigma}_{gg}$, $\hat{\Sigma}_{gm}$ e $\hat{\Sigma}_{mm}$, podemos estimar Σ por

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Lambda}_{cg} & 0 \\ 0 & \hat{\Lambda}_{\gamma m} \end{pmatrix}\begin{pmatrix} \hat{\Sigma}_{gg} & \hat{\Sigma}_{gm} \\ \hat{\Sigma}_{mg} & \hat{\Sigma}_{mm} \end{pmatrix}\begin{pmatrix} \hat{\Lambda_{cg}} & 0 \\ 0 & \hat{\Lambda}_{\gamma m} \end{pmatrix}'$$

para $\hat{\Lambda}_{cg} = -\hat{C}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}$ e $\hat{\Lambda}_{\gamma m} = -(\hat{M}'\hat{U}\hat{M})^{-1}\hat{M}'\hat{U}$.

O que resta é construir estimadores $\hat{\Sigma}_{gg}$, $\hat{\Sigma}_{gm}$ e $\hat{\Sigma}_{mm}$. Quando $\hat{\theta}$ e $\hat{\gamma}$ são estimadores GMM ou ML, podemos escrever

$$\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \phi_g(D_i; \theta) \tag{19}$$

$$\hat{m}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} \phi_m(D_i; \gamma), \tag{20}$$

para $(\phi_g(D_i;\theta), \phi_m(D_i;\gamma))$ as funções dos momentos para GMM ou funções score para ML. Se pode estimar Σ por

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \hat{\phi}_c(D_i)^2 & \hat{\phi}_c(D_i)\hat{\phi}_\gamma(D_i)' \\ \hat{\phi}_\gamma(D_i)\hat{\phi}_c(D_i)' & \hat{\phi}_\gamma(D_i)\hat{\phi}_\gamma(D_i)' \end{pmatrix}, \tag{21}$$

para

$$\hat{\phi}_c(D_i) = \hat{\Lambda}_{cg}\phi_g(D_i;\hat{\theta}) = -\hat{C}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\phi_g(D_i;\hat{\theta})$$

e

$$\hat{\phi}_\gamma(D_i) = \hat{\Lambda}_{\gamma m}\phi_m(D_i;\hat{\gamma}) = -(\hat{M}'\hat{U}\hat{M})^{-1}\hat{M}'\hat{U}\phi_m(D_i;\hat{\gamma})$$

No caso GMM, $\phi_g(D_i;\hat{\theta})$ e $\phi_m(D_i;\hat{\gamma}$ são disponíveis imediatamente do cálculo do objetivo final do "solver" para (17) e (18), respectivamente.

No caso do MLE, o *score* é geralmente computado como parte do gradiente numérico para a verossimilhança. Os elementos de $\hat{\Lambda}_{cg}$ e $\hat{\Lambda}_{\gamma m}$ também são comumente pré-computados. Os ponderadores $\hat{W}$ e $\hat{U}$ são diretamente envolvidos no cálculo das objetivos (17) e (18), respectivamente.

Quando $\hat{g}(\theta)$ e $\hat{m}(\gamma)$ são diferenciáveis, $\hat{G}$ e $\hat{M}$ são utilizados nas formulações tradicionais para inferência assintótica sobre θ e γ, e o gradiente $\hat{C}$ é utilizado nos cálculos do método delta para inferência assintótica sobre c.

Receita de AGS para calcular $\hat{\Delta}$ (GMM/ML com momentos diferenciáveis)

1. Estime $\hat{\theta}$ e $\hat{\gamma}$ seguindo (17) e (18), respectivamente, e calcule $\hat{c} = c(\hat{\theta})$.
2. Guarde $\left\{\phi_g(D_i;\hat{\theta})\right\}_{i=1}^{n}$ e $\{\phi_m(D_i;\hat{\gamma})\}_{i=1}^{n}$ dos cálculos da função objetivo em (17) e (18), respectivamente.
3. Guarde os gradientes numéricos $\hat{G} = \frac{\partial}{\partial\theta}\hat{g}(\hat{\theta})$ e M por $\hat{M} = \frac{\partial}{\partial\gamma}\hat{m}(\hat{\gamma})$, e $\hat{C} = \frac{\partial}{\partial\theta}c(\hat{\theta})$ dos cálculos dos erros-padrão assintóticos para $\hat{\theta}$, $\hat{\gamma}$ e $\hat{c}$.
4. Calcule $\hat{\Lambda}_{cg} = -\hat{C}(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}$ e $\hat{\Lambda}_{\gamma m} = -(\hat{M}'\hat{U}\hat{M})^{-1}\hat{M}'\hat{U}$ usando os pesos $\hat{W}$ e $\hat{U}$ das funções objetivo (17) e (18), respectivamente.
5. Calcule $\hat{\phi}_c(D_i) = \hat{\Lambda}_{cg}\phi_g(D_i;\hat{\theta})$ e $\hat{\phi}_\gamma(D_i) = \hat{\Lambda}_{\gamma m}\phi_m(D_i;\hat{\gamma})$ para cada i.
6. Calcule $\hat{\Sigma}$ como em (21).
7. Calcule $\hat{\Delta}$ como em (16).

Mercado para seguro de longo prazo de cuidados de saúde (Hendren, 2013). $\hat{c}$ é o estimador MLE para o *mínimo da razão de preço "pooled"*, uma quantidade que determina o intervalo de preferências para o qual o mercado de seguros não existe. Por sua vez, $\hat{s}$ é a estatística que resume a distribuição conjunta das crenças subjetivas dos indivíduos sobre a possibilidade de precisar de cuidados de longo prazo (LTC) e sua eventual necessidade para tal serviço.

Hendren explica que os parâmetros que determinam o *minimum pooled price ratio* são identificados da relação entre credos elicitados e a eventual realização de eventos de perda como cuidados de longo prazo. AGS definem quatro vetores de estatísticas descritivas:

1. *fractions in focal-point groups:* consiste na fração de respondentes que reportam exatamente 0, a fração dos que reportam exatamente 0.5, e a fração dos que reportam exatamente 1.
2. *fractions on non-focal-point groups:* consiste das frações de respondentes cujo relatórios estão em cada um dos intervalos:
    - (0.1, 0.2]
    - (0.2, 0.3]
    - (0.3, 0.4]
    - (0.4, 0.5]
    - $\vdots$
    - (0.9, 1]
3. *fraction in each group needing LTC:* consiste da fração dos respondentes de cada relatório que eventualmente precisam de cuidado de longo prazo (LTC).
4. Soma de todos os 3.

Hendren sugere que o terceiro vetor é especialmente informativo sobre a *minimum pooled price ratio.*

AGS (2020b) estimam que estas estatísticas descritivas tem informatividade de 0.68 para o estimador $\hat{c}$, implicando que estas estatísticas descritivas podem explicar 68%[5] da variação no estimador sob distribuição assintótica conjunta do estimador e estatística descritivas, e que (sob certas condições) aceitando a relação implicada pelo modelo entre a razão de preços mínima e o

[5] Interpretação similar ao R2.

Tabela 1: Informatividade Estimada Δ da Estatística Descritiva

Estatística descritiva $\hat{s}$	Informatividade Estimada Δ
1. *fractions in focal-point groups*	0.005
2. *fractions on non-focal-point groups*	0.018
3. *fraction in each group needing LTC*	0.678
4. Todos	0.700

A informatividade estimada é calculada de acordo com o código de AGS (2020b, seção 5.1) usando o programa e dados de replicagem de Hendren (2013), além de cálculos adicionais fornecidos pelo autor.

valor populacional da estatística descritiva reduz o viés do pior aso do estimador por um fator de $\sqrt{1-0.68} \approx 0.43$.

Exemplo IV. O resultado assintótico de AGS (2020b) vale com exatidão no exemplo de variáveis instrumentais linear. Para ilustrar o valor do cálculo da informatividade se supõe de novo que sob π_r, os primeiros m elementos de b são conhecidos e iguais a b_r, além disso $c \sim N(0, \omega_r^2)$. Além disso, AGS supõem que o leitor r pensa que o grau de má-especificação é limitado em relação a incerteza amostral, no sentido de que

$$\text{Prob}\left\{\sqrt{b'\Omega^{-1}b} < \mu^2\right\} = 1$$

para alguma constante μ.

Neste caso, se pode mostrar que para $\hat{c}_0$ de novo o estimador de máxima verossimilhança, o aumento no variância posterior média do leitor r de observar $(\hat{c}_0, \sigma_0, \Lambda_0)$ para σ_0 o erro-padrão de $\hat{c}_0$ e Λ_0 e a sensibilidade de $\hat{c}$ para $\hat{s}$, ao invés dos dados completos, é limitado superiormente por

$$\mathrm{E}_r = \left[\left(\frac{\omega_r^2}{\omega_r^2 + \sigma_0^2}\right)^2 \sigma_0^2 \mu^2(b_r)(1-\Delta)\right] \tag{22}$$

para

$$\mu(b_r) = \sqrt{\mu^2 - b_r'(S\Omega S')^{-1}b_r}$$

Portanto, quando o leitor r é confiante sobre o impacto da má-especificação sobre $\hat{s}$ e a informatividade Δ para $\hat{s}$ para $\hat{c}_0$ é alto, observando $\hat{c}_0$ é quase bom como observar os dados completos.

Tabela 2: BLP (1995), Resultados de Algumas Especificações Alternativas: Markups Preço-Custo Marginal

	Caso Base	Inclui ln(q) na f. custo	Substitui AT por AIR	Peso e HP no lugar de HP/Wt	Termos de interação f. custo	3 dummies de regiões e realibility	Interações na f. custo e weight
Mazda323	\$ 801	\$ 1616	\$ 1012	\$ 1073	\$ 828	\$1125	\$1389
Nissan Sentra	\$ 880	\$ 1769	\$1153	\$1271	\$ 912	\$1308	\$1487
Ford Escort	\$1077	\$2043	\$1326	\$1470	\$1111	\$2094	\$1690

5 Análise de Sensibilidade

Um pesquisador deveria reportar um estimador que é ótimo para cada (tipo) leitor. Como as conclusões mudam com pequenas mudanças no conjunto de hipóteses que são adotadas na análise principal?

5.1 Mostrar Conclusões sob Hipóteses Alternativas

Suponha que sob um conjunto de hipóteses $a \in \mathcal{A}$ existe um estimador natural $\hat{c}_a$ – MLE ou GMM eficiente. Se o pesquisador sabe que muitos dos leitores consideram apenas um conjunto limitado de hipóteses, no sentido que que cada prior π_r coloca massa em um único $a \in \mathcal{A}$ e o número de elementos é pequeno, então é natural relatar a estimativa $\hat{c}_a$ para cada elemento de $\mathcal{A}$.

Exemplo. No estudo de BLP (1995) de demanda por automóveis, os autores relatam como uma conclusão chave – a saber, o markup associado a cada modelo de veículo – muda sob 6 modelos alternativos, cada qual correspondendo a uma modificação na função utilidade ou função custo utilizada no modelo base. Então,um leitor que acredita em uma destas especificações a_j será capaz de saber o valor de c que é assintoticamente válido sob a_j.

É útil contrastar a análise de sensibilidade com análise de limites que relata o conjunto de estimativas $\{\hat{c}_a : a \in \mathcal{A}\}$ sem especificar que estimativa corresponde a qual hipótese. Em alguns casos, fazer um mapeamento de hipóteses a elementos do conjunto de estimativas pode parecer óbvio, pelo menos em pontos extremos – limites inferior e superior, por exemplo. Entretanto, em outros casos a análise de limites pode ser menor transparente do que uma análise de sensibilidade em relação ao mesmo conjunto de hipóteses.

5.2 Mostrar Como as Conclusões Dependem das Hipóteses

Se o conjunto de hipóteses $\mathcal{A}$ entendido pelos leitores é suficientemente rico, então relatando um estimador $\hat{c}_a$ associado com cada hipótese $a \in \mathcal{A}$ não é mais factível.

Uma alternativa possível é ofertar informação sobre a função (possivelmente aleatória) $u(a) = \hat{c}_a - \hat{c}_{a0}$ que relaciona o estimador sob a hipótese base do pesquisador, a_0, à hipótese naturalmente preferida do leitor, a. Se todos os leitores conhecerem $u(\cdot)$, então cada leitor pode ajustar a estimativa base $\hat{c}_{a0}$ para refletir a própria hipótese preferida do leitor, a.

OVBF. A fórmula do viés de variável omitida (OVBF) talvez seja a ferramenta mais famosa para intuir as propriedades de $u(\cdot)$. Dado os credos a sobre as propriedades da covariância de um regressor omitido, a OVBF permite ao leitor determinar o viés do estimador de um dado coeficiente resultante da exclusão daquele regressor – que pode corresponder à hipótese básica do pesquisador, a_0. Portanto, a OVBF evita a necessidade de enumerar o viés resultante de todos os possíveis credos sobre os regressores omitidos.[6]

AGS (2017) estudam o problema de traduzir estas ideias para modelos estruturais não-lineares, onde a OVBF não se aplica diretamente. Na ampla classe de modelos que pode ser estimado via distância mínima, as hipóteses de identificação podem ser representadas como restrições sobre o valor populacional a de uma condição de momento sob o valor verdadeiro dos parâmetros estruturais. Exemplo: em estimadores de variáveis instrumentais não-lineares, como o estudo de BLP (1995) sobre demanda por automóveis, a restrição é a de que o vetor de instrumentos observados é ortogonal em população ao vetor dos erros estruturais não-observados.

Nestas abordagens, violações específicas das hipóteses de identificação podem ser representadas como restrições específicas alternativas – por exemplo, que a covariância entre os instrumentos e o erro estrutural assume algum valor específico não-zero na população, ou que o modelo sistematicamente faz previsões ruins, em algum montante, do valor populacional de alguma estatística.

- Modelo linear: o OVBF e suas análogas dizem ao leitor como ajustar o estimador para acomodar tais perturbações às hipóteses identificadoras

[6] Conley, Hansen e Rossi (2012) generalizam a OVBF para a abordagem de variáveis instrumentais, mostrando como traduzir credos sobre violações de restrição de exclusão à credos sobre viés no estimador IV. Como o OVBF, a abordagem dos autores permite diferente leitores atingir diferentes conclusões considerando o ajuste apropriado no estimador relatado.

do pesquisador.

- Modelo não-linear: fórmula similar a OVBF é desconhecida, e checar exaustivamente as implicações de cada perturbação possível pode ser custosa ou impossível.

AGS (2017) mostram que se a perturbações são local então o viés assintótico do estimador é dado por

$$\Lambda(a - a_0) \tag{23}$$

tal que:

- Perturbação local: perturbação pequena no sentido assintótico apropriado;
- Λ é a sensibilidade local discutida na seção 4.1, mas agora trocando a estatística descritiva pelo vetor dos momentos estimados calculados no valor do parâmetro verdadeiro.

Portanto é prático aproximar e relatar os coeficientes da fórmula de viés assintótico em várias aplicações. Neste sentido, a sensibilidade local fornece um análogo ao OVBF para estimadores de distância mínima sob pequenas violações nas hipóteses identificadoras.

AGS (2017) relatam a sensibilidade local do valor médio do markup de veículos estimado em relação às hipóteses identificadoras de BLP (1995). Esta análise mostra que o estimador é especialmente sensível a violações nas hipóteses de que choques não-observados na utilidade de ou do custo de produzir um dado modelo de carro são ortogonais ao número de outros modelos oferecidos pela mesma firma ou ofertado por empresas rivais. BLP discutem a interpretação econômica destas e outras hipóteses de identificação. AGS mostram como um leitor pode usar informação sobre sensibilidade para aproximar os efeitos de violações nas hipóteses de identificação que tem sentido econômico interessante (que tem apelo). Um ponto importante é que a análise de sensibilidade não requer que o pesquisador conheça previamente as hipóteses de interesse alternativas a. Os leitores podem calcular as implicações de diferentes hipóteses usando informações sobre sensibilidade.

5.2.1 Implementação

O estimador de sensitividade é definido como (AGS, 2017):

$$\hat{\Lambda} = -\left(\hat{G}(\hat{\theta})'\hat{W}\hat{G}(\hat{\theta})\right)^{-1}\hat{G}(\hat{\theta})'\hat{W} \tag{24}$$

$\hat{W}$ é a matriz de pesos e $\hat{G}(\cdot)$ é o Jacobiano da função objetivo $g(\cdot)$.

5.3 Mostrar como Reverter a Conclusão

Algumas questões respondidas por modelos estruturais são qualitativas. Exemplos: uma dada política aumenta ou diminui o excedente do consumidor? Uma fusão irá aumentar ou diminuir a qualidade de um produto? A resposta empírica a estas questões depende da realização dos dados.

Para caracterizar esta dependência pode ser útil para os pesquisadores discutir as realizações dos dados que podem levar a conclusão oposta. Em alguns casos as propriedades dos dados para reverter a conclusão pode ser óbvio.

Por outro lado, para estimadores de modelos não-lineares algumas vezes não é óbvio que realizações dos dados possam levar a conclusões diferentes da apresentada pelo pesquisador (ou mesmo se tal realização existe). AGS pensam que mostrando tais realizações pode incrementar a transparência.

AGS chamam este exercício de *análise de sensibilidade reversa.* Ao invés de mudar os insumos (e.g., dados ou hipóteses) e investigar os efeitos sobre os resultados (conclusões), como em na análise de sensibilidade tradicional. Na análise de sensibilidade reversa se busca mudar o resultado com engenharia reversa mostrando a mudança nos insumos necessária para tanto. AGS (2020a) descrevem as mudanças necessárias nos dados ou parâmetros para sensibilidade reversa.

Exemplo. Conclusão binária.

AGS (2020a) seguem Abadie (2020) e consideram quanto um leitor r faz o update dos seus credos sobre c baseado no aprendizado de que $\hat{c} > 0$. A lei da probabilidade total implica que para qualquer conjunto de valores $\mathcal{C}$ para c,

$$\mid \text{Prob}_r\{\mathcal{C}\} - \text{Prob}_r\{\mathcal{C} \mid \hat{c} > 0\} \mid = \tag{25}$$

$$= \frac{\text{Prob}_r\{\hat{c} \leq 0\}}{\text{Prob}_r\{\hat{c} > 0\}} \mid \text{Prob}_r\{\mathcal{C}\} - \text{Prob}_r\{\mathcal{C} \mid \hat{c} > 0\} \mid \leq \frac{\text{Prob}_r\{\hat{c} \leq 0\}}{\text{Prob}_r\{\hat{c} > 0\}}.$$

Portanto, se o leitor r pensa que com muita probabilidade o modelo irá gerar uma estimativa positiva, $\text{Prob}_r\{\hat{c} > 0\} \approx 1$, ele praticamente não atualiza seus credos quando dito que $\hat{c} > 0$.

Por outro lado, se um pesquisador pode fornecer evidência de realizações plausíveis dos dados podem levar a uma conclusão diferente, saber que $\hat{c} > 0$ se torna muito mais informativo. Para formalizar isto no modelo de AGS, suponha que o relatório é

$$(1\{\hat{c} > 0\}, \hat{t}(X, \mathbf{\Omega})), \text{ tal que } \text{Prob}_r\{\hat{c} > 0 \mid \hat{t}\} \ll \text{Prob}_r\{\hat{c} > 0\}$$

Por exemplo, $t(X, \boldsymbol{\Omega})$ pode registrar um conjunto de valores $\mathcal{Y}$ para Y que pode levar a estimativas negativas. Mesmo assim ainda temos:

$$| \operatorname{Prob}_r\{\mathcal{C} \mid \hat{t}\} - \operatorname{Prob}_r\{\mathcal{C} \mid \hat{c} > 0, \hat{t}\} |= \tag{26}$$

$$= \frac{\operatorname{Prob}_r\{\hat{c} \leq 0 \mid \hat{t}\}}{\operatorname{Prob}_r\{\hat{c} > 0 \mid \hat{t}\}} \mid \operatorname{Prob}_r\{\mathcal{C} \mid \hat{t}\} - \operatorname{Prob}_r\{\mathcal{C} \mid \hat{c} \leq 0, \hat{t}\} \mid .$$

o leitor pode agora atualizar substancialmente seu credo apesar saber que $\hat{c} > 0$.

Referências

Abadie, Alberto, "Statistical Non-Significance in Empirical Economics." *American Economic Review: Insights.* 2 (2), 2020.

Andrews, Isiah, Matthew Gentzkow, e Jesse Shapiro, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics,* 132 (4), 2017.

Andrews, Isiah, Matthew Gentzkow, e Jesse Shapiro, "Transparency in Structural Research," *Journal of Business and Economic Statistics,* 38 (4), 2020a.

Andrews, Isiah, Matthew Gentzkow, e Jesse Shapiro, "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica,* 2020b.

Berry, Steve, James Levinsohn, e Ariel Pakes, "Automobiles Prices in Market Equilibrium," *Econometrica,* 1995.

Conley, Timothy, Christian Hansen, e Peter E. Rossi, "Plausibility Exogenous." *Review of Economics and Statistics,* 94 (1), 2012.

Heckman, James "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature,* 48, 2010.

Hendren, Nathaniel, "Private Information and Insurance Rejections." *Econometrica,* 81 (5), 2013.

Pakes, Ariel "Behavioral and Descriptive Forms of Choice Models." *International Economic Review,* 55 (3), 2014.