Instrumentos Fracos

Victor Gomes

31 de julho de 2022

1 Validade dos Instrumentos

Instrumentos precisam ser validos para produzir estimações confiáveis. Sem instrumentos com as características adequadas as estimativas não possuem validade. A primeira característica de um bom instrumento é a sua *validade*. Validade do instrumento está relacionado com sua capacidade de explicar a variável endógena.

Suponha o seguinte modelo (seguindo Andrews, Stock e Sun, 2019):

$$Y_i = X_i'\beta + W_i'\kappa + e_i \tag{1}$$

$$X_i = Z_i\pi + W_i'\gamma + V_i, \tag{2}$$

substituindo X em (1) se tem:

$$Y_i = Z_i'\delta + W_i'\tau + U_i \tag{3}$$

O interesse do pesquisador é estimar β, mas X_i é potencialmente endógena. Então utilizamos o instrumento Z.

Neste setup o estimador 2SLS pode ser escrito na seguinte forma:

$$\hat{\beta}_{2SLS} = (\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi})^{-1}\hat{\pi}\hat{Q}_{ZZ}\hat{\delta} \tag{4}$$

para $\hat{Q}_{ZZ} = \frac{1}{n}Z_iZ_i'$. E com 2-step GMM:

$$\hat{\beta}_{2SGMM} = (\hat{\pi}'\hat{\Omega}_{ZZ}\hat{\pi})^{-1}\hat{\pi}\hat{\Omega}_{ZZ}\hat{\delta} \tag{5}$$

$\hat{\Omega}_{ZZ}$ é o estimador para a variância de $\hat{\delta} - \hat{\pi}\beta$.

Sob condições de regularidade o sampling error é:

$$\sqrt{n}\begin{pmatrix} \hat{\delta} - \delta \\ \hat{\pi} - \pi \end{pmatrix} \rightarrow_d N(0, \sigma^*) \tag{6}$$

Pela aproximação assintótica de (6),

$$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma\right) \tag{7}$$

Modelo IV implica em

$$\delta = \pi\beta.$$

Então

$$\beta = \delta/\pi.$$

β é altamente não-linear quando π é próximo de zero.

Instrumentos fracos ocorre quando π é próximo de zero.

1.1 Testes com erros homocedásticos

Stock e Yogo (2005) consideram o problema de testes para instrumentos fracos no caso com erros homocedásticos. Primeiro eles realizam uma discussão criteriosa sobre a definição de instrumentos fracos. Aqui neste texto, vamos assumir que instrumentos fracos possuem estimativa pontual que não será estatisticamente diferente de zero. Stock e Yogo fazem testes para o caso de um único regressor endógeno utilizando estatística F do primeiro estágio. O valor crítico depende do número de instrumentos. A tabela com os valores é disponibilizado no artigo de Stock e Yogo.[1]

Se os instrumentos fracos são definidos como o pior caso de um teste t avaliado em 5% do estimador de mínimos quadrados de dois estágios exceda 15%, então o valor crítico depende fortemente do número de instrumentos e é igual a 8.96 em casos com um único instrumento mas aumenta para 44.78 em casos com 30 instrumentos. A estatística exceder 15% na estimativa 2SLS é função direta do poder do teste a 5% na estimativa do primeiro estágio. Os números para os casos mais parcimoniosos são próximos da regra de bolso de F acima de 10 de Staiger e Stock (1997) – veja Stock e Watson (2015), cap. 12.

Stock e Yogo também consideram desenhos com múltiplos regressores endógenos. Para estes casos eles desenvolvem valores críticos para o uso com a estatística Cragg-Donald (1993) para testar a hipótese de que π possui posto reduzido.[2]

[1] Veja também a Tabela 12.4 em Hansen (2022, cap. 12).

[2] A partir destes resultados de Stock e Yogo, Sanderson e Windmeijer (2016) consideram testes para quando os instrumentos são fracos para o propósito de estimação e inferência sobre um de múltiplas variáveis endógenas.

1.2 Testes com erros não-homocedásticos

Os resultados de Stock e Yogo (2005) dependem fortemente da hipótese de homocedasticidade. No caso homocedástico a matriz de variância Σ para $(\hat{\delta}, \hat{\pi})$ pode ser escrita como produto de Kronecker de uma matriz 2×2 com uma matriz $k \times k$. Portanto, a análise de Stock e Yogo (2005) não se aplica aqui.

A despeito da inaplicabilidade da regra de Stock e Yogo, estatística F são frequentemente reportadas em situações de não-homocedasticidade com múltiplos instrumentos. De acordo com Andrews, Stock e Sun (2019), em alguns casos os autores reportam estatísticas F que são robustas a não-homocedasticidade (F^R):

$$F^R = \frac{1}{k}\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}, \tag{8}$$

enquanto outros relatam estatísticas F tradicional que *não* é robusta à não-homocedasticidade.

$$F^N = \frac{1}{k}\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi} = \frac{n}{k\hat{\sigma}_V^2}\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}, \tag{9}$$

tal que

$$\hat{\Sigma}_{\pi\pi} = \frac{\hat{\sigma}_V^2}{n}\hat{Q}_{ZZ}^{-1}$$

e $\hat{\sigma}_V^2$ é um estimador para $\mathrm{E}[V_i^2]$.

O uso de estatística-F em setup não-homocedástico é comum em pacotes estatísticos. Por exemplo, quando se estima um modelo 2SLS utilizando o comando `ivreg2` no programa estatístico Stata, automaticamente é gerado a estatística Wald de Kleibergen-Paap (2007) de que π tem posto reduzido em relação aos valores críticos de Stock-Yogo. Cabe observar que o usuário é avisado de que os valores críticos de Stock e Yogo são para o caso homocedástico.[3]

Em configurações com uma única variável endógena, a estatística Wald de Kleibergen-Paap é equivalente a estatística F robusta à não-homocedasticidade F^R para testar $\pi = 0$. Quando se tem múltiplos regressores endógenos ela é análoga a estatística robusta Cragg-Donald.

Uma estatística F alternativa foi proposta por Montiel Olea e Pffueger (2013). A estatística F efetiva do primeiro estágio é

$$F^{Eff} = \frac{\hat{\pi}'\hat{\Sigma}_{N,\pi\pi}^{-1}\hat{\pi}}{traco(\hat{\Sigma}_{\pi\pi}\hat{Q}_{ZZ})} = \frac{traco(\hat{\Sigma}_{\pi\pi,N}\hat{Q}_{ZZ})}{traco(\hat{\Sigma}_{\pi\pi}\hat{Q}_{ZZ})}F^N = \frac{k\hat{\sigma}^2}{traco(\hat{\Sigma}_{\pi\pi}^{*}\hat{Q}_{ZZ})}F^N \tag{10}$$

[3] Ver Baum et al. 2007. Veja também a tabela de valores críticos de Kleibergen-Paap na Tabela 12.5 de Hansen (2022, cap. 12).

Em casos com erros homocedásticos, a estatística F^{Eff} se reduz na F^N, enquanto que com erros não-homocedásticos se adiciona uma correção multiplicativa da estimativa robusta da variância. De forma similar, no caso exatamente identificado F^{Eff} se reduz na estatística F^R,[4] enquanto que no caso não-homocedástico $\hat{\pi}$ é ponderado por $\hat{Q}_{ZZ}$ ao invés de $\hat{\Sigma}_{\pi\pi}^{-1}$.

As equações do estimador 2SLS (4) e as dos testes F^N, F^R e F^{Eff} fornecem a intuição porque F^{Eff} é a estatística apropriada para testar a força do instrumento em 2SLS no caso não-homocedástico. O estimador 2SLS não se comporta bem quando o seu denominador, $\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}$, é próximo de zero. A estatística F^N mede o mesmo objeto, mas calcula o erro-padrão equivocado e não tem uma distribuição F não-central como em Stock-Yogo. O problema aqui é F^N pode ser muito grande mesmo quando $\hat{\pi}'\hat{Q}_{ZZ}\hat{\pi}$ é muito pequeno. Por outro lado, a estatística F^R mede o objeto populacional errado, $\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}$, ao invés de $\hat{\pi}'Q_{ZZ}^{-1}\hat{\pi}$, embora ela tenha uma distribuição não-central não é correspondente ao estimador 2SLS.

Por último, a estatística F^{Eff} mede o objeto correto e consegue capturar o erro-padrão correto na média. F^{Eff} é distribuída como uma média ponderada de variáveis χ^2 não-central tal que os pesos são positivos e somam um. Os pesos para formar F^{Eff} são dados pelos autovalores de $\hat{\Sigma}_{\pi\pi}^{1/2}\hat{Q}_{ZZ}\hat{\Sigma}_{\pi\pi}^{1/2}/traco(\hat{\Sigma}_{\pi\pi}\hat{Q}_{ZZ})$. Montiel Olea e Pflueger (2013) mostram que a distribuição de E^{Eff} pode ser aproximada por uma distribuição não-central χ^2, mas os resultados somente se aplicam ao estimadores 2SLS e LIML. Os testes deles rejeitam quando a estatística-F efetiva excede um valor crítico.

Quando $k = 1$, i.e. com apenas um instrumento, $\Sigma_{\pi\pi}$, $\Sigma_{\pi\pi,N}$, e Q_{ZZ} são todos escalares e, portanto, $F^R = F^{Eff}$. Ambas as estatísticas possuem distribuição F não-central com o mesmo parâmetro de não-centralidade que governa a distribuição do estimador de variáveis instrumentais. Portanto, nesta situação se pode usar os valores críticos de Stock e Yogo (2005).

Andrews, Stock e Sun (2019), consideram que para o caso com $k > 1$ o melhor é utilizar a estatística F^{Eff} aplicando a regra de corte de 10. A principal conclusão deles é de que F^{Eff} é a melhor estatística para detectar instrumentos fracos nos casos sobre-identificados não-homocedásticos. Em suma, F^{Eff} pode ser comparada com os valores críticos de Stock e Yogo (2005) quando $k = 1$, e com os valores críticos de Montiel Olea e Pflueger (2013) para k > 1, ou ainda, com a regra de bolso de valor de corte de 10.

Lee, McCrary, Moreira, e Porter (2022) mostram que, em modelos de uma única variável instrumental, que o verdadeiro teste t a 5% requer estatística F do primeiro estágio superior a 104.7 para o estimador 2SLS. Para usar a regra de corte da F como 10, requer trocar o valor crítico de 1.96 por 3.43.

[4] Que é igual a estatística Kleibergen-Paap.

Uma importante limitação de adotar um único valor de corte para F é a perda de informação dos dados quando o primeiro estágio está abaixo do valor de corte (Andrews, Stock, Sun, 2019). Para resolver este problema, Lee, McCrary, Moreira, e Porter propõem o chamado tF teste.

Para acomodar ocorrências da F abaixo de 104.7, a proposta é utilizar o teste tF. Dada a definição para $c(F)$,[5] "F-dependent critical value for t^2",

$$\mathrm{E}[t^2 > c(F)] \leq 0.05$$

Por exemplo, se uma pessoa calcula uma estimativa pontual de 3.2 com erro-padrão (convencional) de 1.5, e a estatística F do primeiro estágio como 9, então a entrada na Tabela de LMMP para $\sqrt{F}$ é 3.65, que significa que o erro-padrão tradicional deve ser inflado pelo fator $\frac{3.65}{1.96} = 1.862$, resultando no *erro-padrão tF 0.05* de $1.5 \times 1.862 \approx 2.79$.

Uma alternativa ao teste F do primeiro estágio seria a estatística Anderson-Rubin (1949). O teste é composto por dois passos: (i) calcule uma nova variável $Y_i^* = Y_i - \beta_{1,0}x_i$, tal que $\beta_{1,0}$ é a estimativa do parâmetro β_1 usando instrumentos fracos. (ii) No segundo passo, faça a regressão de Y_i^* contra os regressores exógenos incluídos (x) e os instrumentos (z). A estatística Anderson-Rubin é a F testando a hipótese de o coeficiente dos z's serem todos zero. Sob a hipótese nula de que $\beta_1 = \beta_{1,0}$, se os instrumentos satisfazem a condição de exogeneidade, eles serão não-correlacionados com o termo de erro nessa regressão, e a hipótese nula será rejeitada em 5% de todas as amostras.

Um conjunto de confiança pode ser construído a partir do conjunto dos parâmetros que não são rejeitados pelo teste de hipótese. O conjunto de valores de β_1 que não são rejeitados pelo teste Anderson-Rubin de 5% compõe um intervalo de confiança de 95% para β_1. Quando a estatística é calculada usando a fórmula de homocedasticidade, o intervalo de confiança Anderson-Rubin pode ser construído solucionando uma equação quadrática. Este teste nunca assume relevância dos instrumentos. O intervalo de confiança do teste terá probabilidade cobertura de 95% em amostras grandes, não importando se os instrumentos são fortes, fracos ou mesmo irrelevantes.[6]

Detalhes (Andrews, Stock, e Sun (2019), seção 5.1). Especificamente o modelo implica que $\beta = \frac{\delta}{\gamma}$. Portanto, sob a hipótese nula $H_0 : \beta = \beta_0$

[5] LMMP: $c(F) = 1[F < 104.7].\tilde{c}(\sqrt{F}) + 1[F \geq 104.7].1.96^2$. $\tilde{c}(\sqrt{F})$ é um polinômio descrito no apêndice de LMMP.

[6] Quando os instrumentos são fortes, com a regressão 2SLS valendo, e os coeficientes são sobre-identificados, o teste Anderson-Rubin é ineficiente no sentido de ser menos poderoso do que o teste t do 2SLS. Um teste alternativo é o teste de razão de verossimilhança condicional de Moreira (2003).

sabemos que $\delta - \gamma\beta_0 = 0$ e por conseguinte

$$g(\beta_0) = \hat{\delta} - \hat{\gamma}\beta_0 \sim N(0, \Omega(\beta_0)), \tag{11}$$

para

$$\Omega(\beta_0) = \Sigma_{\delta\delta} - \beta(\Sigma_{\delta\gamma} + \Sigma_{\gamma\delta}) + \beta^2\Sigma_{\gamma\gamma}.$$

As matrizes que compõem $\Omega(\beta_0)$ são:

- $\Sigma_{\delta\delta}$ é a variância de $\hat{\delta}$,
- $\Sigma_{\gamma\gamma}$ é a variância de $\hat{\gamma}$, e
- $\Sigma_{\delta\gamma}$ é a covariância entre $\hat{\delta}$ e $\hat{\gamma}$.

A estatística Anderson-Rubin AR é definida como

$$AR(\beta) = g(\beta)'\Omega(\beta)^{-1}g(\beta), \tag{12}$$

e ela segue uma distribuição χ^2_k sob $H_0 : \beta = \beta_0$ não importando o valor de γ. Andrews, Stock e Sun (2019) observam que a Anderson e Rubin (1949) consideram o caso com erro normal homocedásticos. A versão apresentada acima (12) é formalmente uma uma generalização do teste AR quer permite erros não-homocedásticos.

Usando a estatística AR (12) podemos formar um teste AR de $H_0 : \beta = \beta_0$ como

$$\phi_{AR} = 1\left\{AR(\beta_0) > \chi^2_{k,1-\alpha}\right\}, \tag{13}$$

para a quantil $1-\alpha$ de uma distribuição χ^2_k. Isto resulta um teste tamanho-α que é robusto a instrumentos fracos (Andrews, Stock e Sun, 2019, Staiger e Stock, 1997). Nesta fórmula $1\{\cdot\}$ representa uma variável indicadora que será igual a 1 em função da regra definida entre as chaves.

Podemos formar um conjunto de confiança robusto a instrumentos fracos, CS_{AR}, coletando os valores não rejeitados por (13). É mais direta a formação do conjunto de confiança no caso homecedástico.[7]

[7] Uma vez que os conjunto de confiança AR possui a cobertura correta independente da força dos instrumentos, sabemos que ele tem tamanho infinito com probabilidade positiva. Especificamente, o CS_{AR} pode assumir uma de três formas em especificações com um único instrumento: (i) um intervalo limitado $[a, b]$, (ii) uma linha real $(-\infty, \infty)$, e (iii) a linha real excluindo o intervalo $(-\infty, a] \cup [b, \infty)$. Em especificações com mais de um instrumento mas com erros homocedásticos, o conjunto de confiança AR pode assumir uma das três formas ou ainda pode ser vazio.

Estimação com instrumentos fracos. Para trabalhar na presença de instrumentos fracos, o economista possui algumas alternativas. Se o modelo for exatamente identificado, então a saída extrema é encontrar instrumentos melhores (fortes). Se o problema empírico tiver muitos instrumentos, tente excluir as variáveis instrumentais mais fracas e observar se o erro-padrão vai crescendo (ele cresce a medida que se exclui os instrumentos mais fracos).

No caso não em que não se pode excluir instrumentos fracos é aconselhável não utilizar o modelo de mínimos quadrados de dois estágios (2SLS). Uma alternativa na presença de instrumentos fracos é o modelo LIML (*limited information maximum likelihood*) – que é contrapartida usando máxima verossimilhança ao 2SLS. O estimador LIML também é o valor de $\beta_{1,0}$ que minimiza a estatística Anderson-Rubin homocedástica. Portanto, se o intervalo de confiança Anderson-Rubin não é vazio, ele irá conter o estimador LIML. Quando os instrumentos são fracos, o estimador LIML é adequado pois ele é mais centrado no valor verdadeiro β_1 do que o 2SLS. Para uma introdução ao estimador LIML veja o capítulo 12 de Hansen (2020) e 7 de Hayashi (2000).

2 Testes para Modelos Sobre-Identificados

Para o caso de modelos sobre-identificados não existe ainda um procedimento claro. Um grande número de alternativas estão disponíveis, mas não existe um consenso sobre a melhor.

O teste AR é robusto mas ineficiente sob identificação forte em modelos sobre-identificados. Observe que no modelo (de distribuição) normal dos parâmetros, a estatística AR para testar $H_0 : \beta = \beta_0$ depende dos dados apenas por meio de $g(\beta_0) = \hat{\delta} - \hat{\pi}\beta_0$. De acordo com Andrews, Stock e Su (2019), para construir procedimentos que funcionem bem como o teste t no caso fortemente identificado, é importante incorporar informações sobre $\hat{\pi}$, que é informativo sobre que desvios de $\delta - \pi\beta_0$ de zero correspondem a violações das restrições paramétricas do modelo. Especificamente, sob valor de parâmetro alternativo β, $\hat{\delta} - \hat{\pi}\beta_0 \sim N\left(\pi(\beta - \beta_0), \Omega(\beta_0)\right)$. Para construir procedimentos que funcionem bem como o teste t nos casos sobre-identificados e bem-identificados, diversos autores tem considerado testar estatísticas que dependem de $(\hat{\pi}, \hat{\delta})$ mais do que $\delta - \pi\beta_0$.

O primeiro problema enfrentado neste setup é que mesmo sob a hipótese nula $H_0 : \beta = \beta_0$, a distribuição de $(\hat{\pi}, \hat{\delta})$ dependem dos parâmetros desconhecidos do primeiro estágio, π. Portanto, para uma estatística genérica $s(\beta_0)$ que dependa de $(\hat{\pi}, \hat{\delta})$, a distribuição de $s(\beta_0)$ sob a hipótese nula irá depender de π.

Um exemplo é tomar $s(\beta_0)$ como a estatística t absoluta dada por $|\hat{\beta} - \beta_0|/\hat{\sigma}_{\hat{\beta}}$, sabemos que a distribuição da estatística t sob a hipótese nula depende da força dos instrumentos. Pode-se a princípio encontrar o maior quantil possível $1-\alpha$ para $s(\beta_0)$ sob a hipótese nula consistente com algum conjunto de valores para π. Por exemplo, para um conjunto de confiança inicial como na abordagem Bonferroni de Staiger e Stock (1997). Entretanto, para muito das estatísticas $s(\beta_0)$ isto requer simulação extensiva e se torna computacionalmente intratável, e além disso tipicamente se tem perda de poder do teste.

Uma abordagem alternativa pode eliminar a dependência de π por meio de condicionamentos. Especificamente, sob a hipótese nula $H_0 : \beta = \beta_0$,

$$\begin{pmatrix} g(\beta_0) \\ \hat{\pi} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ \pi \end{pmatrix}, \begin{pmatrix} \Omega(\beta_0) & \Sigma_{\delta\pi} - \Sigma_{\pi\pi}\beta_0 \\ \Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta_0 & \Sigma_{\pi\pi} \end{pmatrix} \right). \tag{14}$$

Portanto, se for definido

$$D(\beta) = \hat{\pi} - (\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta)\,\Omega\,(\beta)^{-1}\,g\,(\beta)\,, \tag{15}$$

se tem que $(g(\beta), D(\beta))$ é uma transformação 1:1 de $(\hat{\pi}, \hat{\delta})$, e sob $H_0 : \beta = \beta_0$

$$\begin{pmatrix} g(\beta_0) \\ D(\beta_0) \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ \pi \end{pmatrix}, \begin{pmatrix} \Omega(\beta_0) & 0 \\ 0 & \Psi(\beta_0) \end{pmatrix} \right), \tag{16}$$

tal que

$$\Psi(\beta_0) = \hat{\pi} - (\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta)\,\Omega\,(\beta)^{-1}\,(\Sigma_{\delta\pi} - \Sigma_{\pi\pi}\beta)\,.$$

Portanto, sob a hipótese nula H_0 o parâmetro de distúrbio π entra na distribuição dos dados apenas por meio da estatística $D(\beta_0)$, ao mesmo tempo que $g(\beta_0)$ é independente de $D(\beta_0)$ e tem distribuição conhecida. Assim sendo, a distribuição condicional de $g(\beta_0)$ (e consequentemente $(\hat{\pi}, \hat{\delta})$) dado $D(\beta_0)$ não depende de π.

Esta abordagem condicional foi introduzida na literatura de instrumentos fracos primeiro por Marcelo Moreira (2003) que estudou o caso homocedástico. Na configuração com erros homocedásticos, $g(\beta_0)$ e $D(\beta_0)$ são transformações das estatísticas S e T definidas pelo autor.

A distribuição condicional de qualquer estatística $s(\beta_0)$ dado $D(\beta_0)$ pode ser simulada. Esta simulação pode ser realizada sob a hipótese nula H_0 pela retirada de elementos de $g(\beta_0)^* \sim N(0, \Omega(\beta_0))$, construindo $(\hat{\delta}^*, \hat{\pi}^*)$ como

$$\begin{pmatrix} \hat{\delta}^* \\ \hat{\pi}^* \end{pmatrix} = \begin{pmatrix} I + \beta_0(\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta_0)\Omega(\beta_0)^{-1} & \beta_0 I \\ (\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta_0)\Omega(\beta_0)^{-1} & I \end{pmatrix} \begin{pmatrix} g(\beta_0)^* \\ D(\beta_0) \end{pmatrix}, \tag{17}$$

para dado $D(\beta_0)$, e tabulando a distribuição resultante de $s^*(\beta_0)$ com cálculo baseado em $(\hat{\delta}^*, \hat{\pi}^*)$. Se representarmos o quantil condicional $1-\alpha$ como $c_\alpha(D(\beta_0))$, podemos construir um teste condicional baseado em s como $\phi_s = 1\{s(\beta_0) > c_\alpha(D(\beta_0))\}$, e dado que $s(\beta_0)$ é continuamente condicionalmente distribuído sobre $D(\beta_0)$ este teste tem probabilidade de rejeição exatamente α sob a hipótese nula, $\mathrm{E}_{\beta_0,\pi}[\phi_s(\beta_0)] = \alpha$ para todo ϕ, enquanto se a distribuição condicional de $s(\beta_0)$ possui ponto de massa, o teste tem tamanho menor do que ou igual a α. Como observado por Moreira (2003) para o caso homocedástico, isto nos permite construir um teste de tamanho α baseado em qualquer estatística de teste $s(\cdot)$.[8]

Andrews, Stock e Su (2019) dizem que testes com probabilidade de rejeição exatamente α para todos os valores de parâmetros consistente com com a hipótese nula são considerados similares. Para o caso homecedástico, a recomendação é utilizar o teste CLR de Moreira (2003). No caso não homecedástico não existe uma recomendação consensual.

Moreira (2003) propôs testes LR e Wald condicionais, baseado em comparar as estatísticas Wald e LR a um valor crítico condicional. Diferente da estatística AR, tanto a estatística LR como a score dependem de $(\hat{\delta}, \hat{\pi})$, e testes condicionais baseado nestas estatísticas são eficientes no caso bem-identificado. O teste CLR de Moreira (2003) tem propriedades muito boas. Entre elas estão a estrutura baseado em um produto de Kronecker da matriz de variância Σ, que significa que nesse caso o problema é imutável em transformações lineares nos instrumentos.

Entretanto, Andrews, Moreira, e Stock (2006) mostraram que o poder invariante do teste de Moreira depende apenas da correlação entre os erros (U, V), do tamanho do π no primeiro estágio (normalizado pela variância), e do valor verdadeiro do parâmetro β. Impondo uma forma adicional de invariância para limitar a atenção aos testes de dois lados, esses autores mostraram numericamente que o teste CLR possui poder perto do limite superior de qualquer teste similarmente invariante sobre um grande intervalo de valor dos parâmetros, tal que o cálculo se torna possível pela baixa dimensão do espaço dos parâmetros invariante.

Caso não-homocedástico. As simplificações obtidas usando a estrutura de produto de Kronecker de Σ não são disponíveis não caso não-homocedástico e isto complica a forma de testar.

As contribuições para testes sem homoncedasticidade são generalizações ou inovações a partir do teste CLR de Moreira. Todas as alternativas são extensões deste teste CLR e são eficientes sob identificação forte, exceto o

[8] Andrews, Stock e Su (2019) sugerem Andrews e Mikusheva (2016) para mais detalhes.

proposto por Andrews (2016). Entretanto, enquanto estas generalizações objetivam o caso não-homocedástico, evidência desta performance para casos fracamente identificados tem sido obtidos apenas com resultados de simulação.

As modificações foram propostas por Andrews et al (2004), que propuseram uma versão do teste CLR, enquanto Kleibergen (2015) introduziu uma estatística condicionante $D(\beta_0)$ para o caso não-homocedástico e e desenvolveu estatísticas score e quasi-CLR aplicáveis para este caso. Andrews e Guggenberger (2015) introduziram dois testes alternativos quasi-CLR para casos não-homocedásticos que permite o uso de uma matriz de covariância Σ singular. Moreira e Moreira (2015) e Andrews e Mikusheva (2016) introduzem uma generalização do teste CLR em configurações com erros não-homocedásticos, que novamente comparam a estatística razão de verossimilhança a um valor crítico condicional.[9]

Para derivar testes com probabilidade de otimalidade que possa ser provável para o caso não-homocedástico fracamente identificado, uma literatura recente tem focado em otimizar o poder médio ponderado. Isto significa que o poder do teste é integrado com respeito aos pesos (β, π). Especificamente, um teste similar que maximiza o poder médio ponderado com respeito aos pesos ν,

$$\int \mathrm{E}_{\beta,\pi}[\phi] d\nu(\beta, \pi)$$

rejeita quando

$$s(\beta_0) = \int f(\hat{\delta}, \hat{\pi}; \beta, \pi) d\nu(\beta, \pi) / f(\hat{\delta}, \hat{\pi} \mid D(\beta_0); \beta_0) \tag{18}$$

excede o valor crítico condicional.

Intuitivamente, este teste ótimo do poder médio ponderado rejeita quando o dado observado é suficientemente mais provável aumentar sob a alternativa ponderada $H_1 : \beta \neq \beta_0$, ponderado por ν, do que sob a hipótese nula $H_0 : \beta = \beta_0$. Como a descrição sugere, a escolha do peso ν tem um papel importante em determinar o poder e outras propriedades do teste resultante, embora o uso de valores críticos condicionais garantam controle de tamanho para todas as escolhas de ν.

A escolha dos pesos é importante neste setup. Testes similares ao descrito acima – poder ótimo de média ponderada – podem atingir essencialmente qualquer função poder admissível pela escolha de pesos apropriados. Alguns artigos discutem em mais detalhes as escolhas dos pesos.

[9] Andrews (2016) mostrou que testes baseados em combinações lineares de estatísticas AR e score são similares ao teste CLR.

Referências

Anderson, T. W., and Herman Rubin. “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations.” *Annals of Mathematical Statistics,* 20, 1949.

Andrews, Donald, Marcelo J. Moreira, e James H. Stock. “.” *Econometrica,* 2006.

Andrews, Isaiah, James H. Stock, e Liyang Sun. “Weak Instruments in IV Regression: Theory and Practice.” *Annual Review of Economics,* 2019.

Hansen, Bruce. *Econometrics.* Princeton, Princeton University Press, 2022.

Hayashi, Fumio. *Econometrics.* Princeton, Princeton University Press, 2000.

Lee, David S., Justin McCrary, Marcelo J. Moreira, Jack Porter. “Valid t-ratio Inference for IV.” *American Economic Review*, a ser publicado, 2022.

Montiel Olea, José L., e Carolin Pflueger. “A Robust Test for Weak Instruments.” *Journal of Business & Economic Statistics,* 31, 2013.

Moreira, Marcelo J. “A Conditional Likelihood Ratio Test for Structural Models.” *Econometrica,* 71, 2003.

Stock and Watson. *Introduction to Econometrics.* 3rd ed. 2015.