
AVALIAÇÃO 2 – ECONOMETRIA I

Endogeneidade, Variáveis Instrumentais e Modelos para Dados em Painel

Aluno: Samuel Maia¹

Matrícula: 2024669055

Disciplina: Econometria I (2025/1)

Professora: Ana Maria Hermeto de Oliveira

Instituição: Cedeplar (UFMG)

Resumo

Neste exercício, investigo a relação entre Complexidade Econômica — medida pelo Índice de Complexidade Econômica (ECI) — e desigualdade de renda, captada tanto pelo índice de Gini quanto por faixas da distribuição. Utilizo dados da *World Inequality Database* (WID) e do *Atlas of Economic Complexity* para construir um painel de países e reavaliar os achados de Hartmann et al. (2017). Em análise eminentemente *cross-section*, esses autores identificaram associação negativa entre ECI e desigualdade. Submeto essa relação a testes mais rigorosos do ponto de vista econométrico, com foco na análise das propriedades dos estimadores sob distintos pressupostos, discutindo fontes de viés e variância, além dos desafios impostos por medidas ranqueadas como o ECI, e pela possível dependência seccional dos resíduos. Comparo diferentes modelos – MQO empilhado, Efeitos Fixos, Primeiras Diferenças e Efeitos Aleatórios – com ênfase na mitigação da endogeneidade via defasagens e estimações instrumentais. Realizo testes de robustez por subgrupos (OCDE vs. Não-OCDE) e por especificações alternativas, incluindo modelos com diferentes formas de controle temporal. Além disso, aplico diagnósticos formais, como o teste de Hausman e testes de dependência entre unidades seccionais. Os resultados indicam um efeito, em geral, nulo da complexidade sobre a desigualdade agregada, com sinais marginais de significância concentrados no grupo da OCDE e restritos ao cenário pré-tributação.

Palavras-chave: Econometria, Dados em Painel, Endogeneidade, Variáveis Instrumentais, Complexidade Econômica, Desigualdade de Renda.

¹samuelmaia@cedeplar.ufmg.br

Sumário

1	Introdução	5
2	Problema de Pesquisa e Bases de Dados	6
2.1	A Complexidade Econômica	6
2.2	A Desigualdade de Renda	7
3	Revisão Conceitual e Teórica: Da Identificação à Estimação Robusta	10
3.1	Problemas de Estimação: Viés, Variância e os Pressupostos do MRLC	11
3.2	Estratégia de Identificação: Endogeneidade e Variáveis Instrumentais	14
3.2.1	Causas da Endogeneidade	14
3.2.2	O Estimador de Variáveis Instrumentais (IV) e seus Requisitos	14
3.3	Modelos para Dados em Painel: Lidando com Múltiplas Fontes de Heterogeneidade e Dependência	16
3.4	Aspectos Práticos e a Robustez da Estratégia de Estimação	19
4	Aplicação Empírica e Resultados	20
4.1	Dados, Amostra e Análise Descritiva	20
4.1.1	Fontes dos Dados: Atlas da Complexidade e World Inequality Database	21
4.1.2	Construção da Amostra Analítica e Implicações Metodológicas	23
4.1.3	Discussão do painel desbalanceado e possíveis vieses (attrition, seleção)	24
4.1.4	Análise Descritiva da Amostra Final	25
4.1.5	Padrões Visuais e Hipóteses de Trabalho	27
4.2	Modelo de Painel Base e Descoberta do Mecanismo Distributivo	28
4.2.1	Especificação do Modelo e Variáveis de Controle	30
4.2.2	Comparação de Estimadores e Teste de Hausman	31
4.2.3	O Efeito da Complexidade na Estrutura de Renda: O Padrão da Foice	31
4.3	Análise de Endogeneidade	32
4.3.1	Mitigando Simultaneidade com Defasagens Temporais	32
4.3.2	Tentativa de Estimação por Variáveis Instrumentais (2SLS)	33
4.4	Análises de Robustez e Heterogeneidade	34
4.4.1	Heterogeneidade do Efeito: Países OCDE vs. Não-OCDE	34
4.4.2	Robustez à Especificação do Modelo (Amostra OCDE)	35
4.5	Síntese dos Resultados Empíricos	36
5	Conclusão e Sugestões para Pesquisas Futuras	38
Apêndice I: Dados Suplementares		40
Apêndice II: Código em R e Replicabilidade Computacional		44

A. Tentativa de Estimação por Variáveis Instrumentais (2SLS)	44
B. Estimação do Modelo de Painel Dinâmico	45
Apêndice III: Enunciados da Avaliação	47
Referências	49

Lista de Figuras

1	Associação entre complexidade econômica e desigualdade (Hartmann et al., 2017)	7
2	Componentes e faixa capturada por surveys da renda nacional do Brasil	8
3	Participação do 1% mais rico	22
4	Etapas de Construção da Amostra Analítica	23
5	Distribuição de Observações da WID por Década e Variável. A densidade observacional se torna expressiva apenas a partir dos anos 1980, especialmente para as variáveis pós-transferências.	24
6	Densidades de Kernel das Variáveis-Chave na Amostra Decadal	26
7	Correlação entre Complexidade Econômica e Índice de Gini	27
8	Correlação entre ECI e renda acumulada de diferentes faixas da distribuição	29
9	Efeito da Complexidade Econômica na Participação da Renda por Faixa (Modelo FE)	32
10	Teste de Robustez à Endogeneidade: Efeito do ECI Defasado na Participação da Renda	33
11	Teste de Heterogeneidade: Efeito do ECI por Grupo de Países	35

Lista de Tabelas

1	Resumo descritivo da base de dados da WID	21
2	Estatísticas Descritivas das Variáveis do Painel Decadal	25
3	Comparação de Estimadores para o Efeito do ECI no Gini Pós-Tributação .	30
4	Testes de Robustez da Especificação para o Efeito do ECI no Gini (OCDE) .	36
5	Efeitos Defasados do ECI sobre Desigualdade: Modelos de Painel por Faixa de Renda	41
6	Testes de Robustez da Especificação para o Efeito do ECI no Gini (OCDE) .	42
7	Testes de Heterogeneidade: OCDE vs. Não-OCDE	43

1 Introdução

Neste exercício, investigo a relação entre complexidade econômica e desigualdade de renda ao longo do tempo e entre países. Em uma análise predominantemente *cross-section*, Hartmann et al. (2017) encontraram uma associação forte e negativa entre uma proxy importante da complexidade produtiva – o Índice de Complexidade Econômica (ECI) – e a desigualdade agregada, medida pelo índice de Gini.

Aqui, retomo essa hipótese com uma base de dados distinta, além de uma estratégia empírica alternativa. Utilizo a *World Inequality Database* (WID), que oferece séries temporais desagregadas por faixas da distribuição de renda. Isso permite investigar não apenas o Gini, mas dinâmicas ao longo de toda a distribuição. Além disso, adoto modelos de painel com dados decenais para explorar a dimensão temporal da relação entre complexidade e desigualdade. Também incorporo controles e estratégias para lidar com heterogeneidade inobservada e possíveis problemas de endogeneidade.

Em termos metodológicos, meu objetivo é compreender os fundamentos econométricos que explicam as diferenças entre modelos e interpretações. Por isso também discuto: as propriedades dos estimadores sob distintos pressupostos; as fontes de viés e variância; e os desafios empíricos impostos pela estrutura dos dados – incluindo desbalanceamento e possíveis correlações espúrias. Minha estratégia busca não apenas testar a robustez da associação entre ECI e desigualdade. Também pretendo esclarecer, ainda que pouco, as condições sob as quais essa relação pode ser interpretada como estrutural ou espúria.

Organizei o trabalho da seguinte forma. A [Seção 2](#) apresenta o problema de pesquisa, os conceitos centrais e as bases de dados utilizadas. A [Seção 3](#) discute os fundamentos econométricos necessários à estimação causal, abordando viés, variância, endogeneidade, variáveis instrumentais e modelos de painel. A [Seção 4](#) aplica essas ferramentas a um painel internacional construído com dados da WID e do Atlas of Economic Complexity, explorando os efeitos do ECI sobre diferentes faixas da distribuição de renda, testando a robustez a defasagens e especificações, e analisando a heterogeneidade por grupos de países. Por fim, a [Conclusão](#) sintetiza os achados, limitações e direções futuras.

Todas as análises foram realizadas em linguagem R (R Core Team, 2024), com atenção explícita às diretrizes da *American Economic Association* (2020) para garantia de transparência e reproduzibilidade computacional de trabalhos aplicados.²

²Segundo relatório da National Academies of Sciences (2019), é útil distinguir *replicabilidade* de *reproduzibilidade*. A primeira refere-se à obtenção de resultados consistentes a partir do mesmo código e ambiente computacional – também chamada de *reproduzibilidade computacional*. A segunda se refere a alcançar resultados semelhantes em novos estudos que testam a mesma hipótese, mas com dados independentes.

2 Problema de Pesquisa e Bases de Dados

Esta seção apresenta o problema de pesquisa que orienta minha análise econométrica: qual é a relação entre a complexidade econômica de um país e a distribuição de sua renda? Mais especificamente, investigo se graus mais elevados de complexidade estão associados a menores níveis de desigualdade, e se esse padrão se mantém ao longo de diferentes faixas da distribuição.

Para isso, replico e expando a análise de Hartmann et al. (2017). Utilizo dados mais recentes e metodologicamente mais robustos sobre desigualdade de renda. Também apresento de forma preliminar as duas bases de dados que estruturam a investigação: o *Atlas of Economic Complexity*, SITC, Rev. 2 ([Subseção 2.1](#)) e o *World Inequality Database*, WID ([Subseção 2.2](#)). O processo de filtragem e integração dessas bases será detalhado na [Subseção 4.1](#).

2.1 A Complexidade Econômica

A complexidade econômica de um sistema é determinada por dois fatores principais: a diversidade e a ubiquidade de seus produtos (Hidalgo e Hausmann, 2009; Hidalgo et al., 2007). A diversidade se refere à quantidade de produtos que um sistema é capaz de fabricar com vantagem comparativa revelada (RCA). Já a ubiquidade diz respeito à quantidade de outros sistemas que conseguem exportar esses mesmos produtos. Esses dois fatores estariam associados a maior PIB per capita e ao crescimento econômico, especialmente em países que passaram de economias baseadas na exploração de recursos naturais para setores mais sofisticados (Hausmann et al., 2006).

Para medi-la, Hidalgo e Hausmann desenvolveram o ECI, o Índice de Complexidade Econômica (Hidalgo e Hausmann, 2009). O índice parte de medidas de diversidade e ubiquidade para, em seguida, capturar a complexidade. A matriz

$$M_{cp} = \sum_p \frac{\text{RCA}_{cp} \cdot \text{RCA}_{cp'}}{k_{c,0} \cdot k_{p,0}} \quad (1)$$

conecta países e produtos, ponderando a diversidade de cada país e a ubiquidade de cada produto. Essa matriz é normalizada pela diversidade, permitindo que a complexidade se reflita na quantidade e singularidade dos produtos exportados.

O ECI é calculado com a fórmula:

$$\text{ECI}_c = \frac{K_c - \bar{K}}{\text{std}(K)} \quad (2)$$

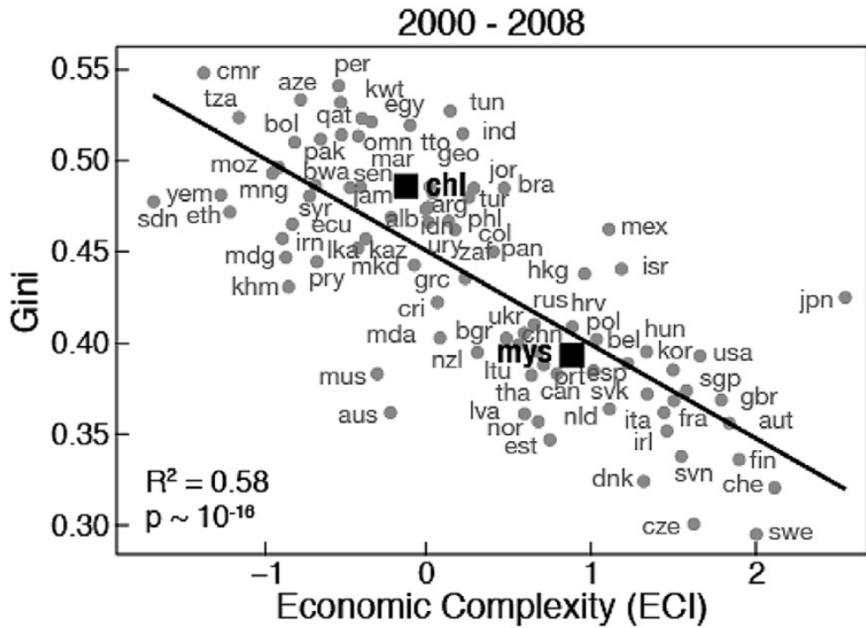


Figura 1: Relação negativa e estatisticamente significativa (observe os valores do coeficiente de determinação e valor-p) entre complexidade econômica (medida pelo ECI) e desigualdade de renda (Gini) para diferentes países no período de 2000 a 2008. Fonte: Hartmann et al. (2017, p. 78).

em que K_c representa o autovalor associado ao segundo maior autovetor da matriz de complexidade M_{cp} , \bar{K} é a média dos valores de K para todos os países (ou outros sistemas), e $\text{std}(K)$ é o desvio padrão de K . É importante notar que o ECI é uma medida *relativa*, construída a partir da comparação entre países. Em termos práticos, ele expressa a posição de uma economia em um ranking global de complexidade produtiva. As consequências econometrísticas deste último ponto serão discutidas com mais detalhe na [Subseção 3.3](#)

2.2 A Desigualdade de Renda

A complexidade econômica tem sido empregada para a análise de diversos fenômenos econômicos. Em uma dessas aplicações, Hartmann et al. (2017) investigaram a relação entre ECI e desigualdade de renda, utilizando o índice calculado a partir da classificação SITC, Rev. 2, conforme disponibilizado pelo *Atlas of Economic Complexity* (veja a ??). Seu principal resultado foi uma associação negativa entre complexidade econômica e desigualdade: países com estruturas produtivas menos complexas tendem a ser mais desiguais. O padrão está ilustrado na [Figura 1](#).

Apesar de sua relevância, os resultados de Hartmann et al. apresentam limitações importantes. A primeira diz respeito à forma como conceberam e mensuraram a desigualdade. O indicador utilizado foi o índice de Gini, que, embora amplamente difundido, não permite observar as dinâmicas internas da distribuição de renda. Esse é o caso especialmente no que

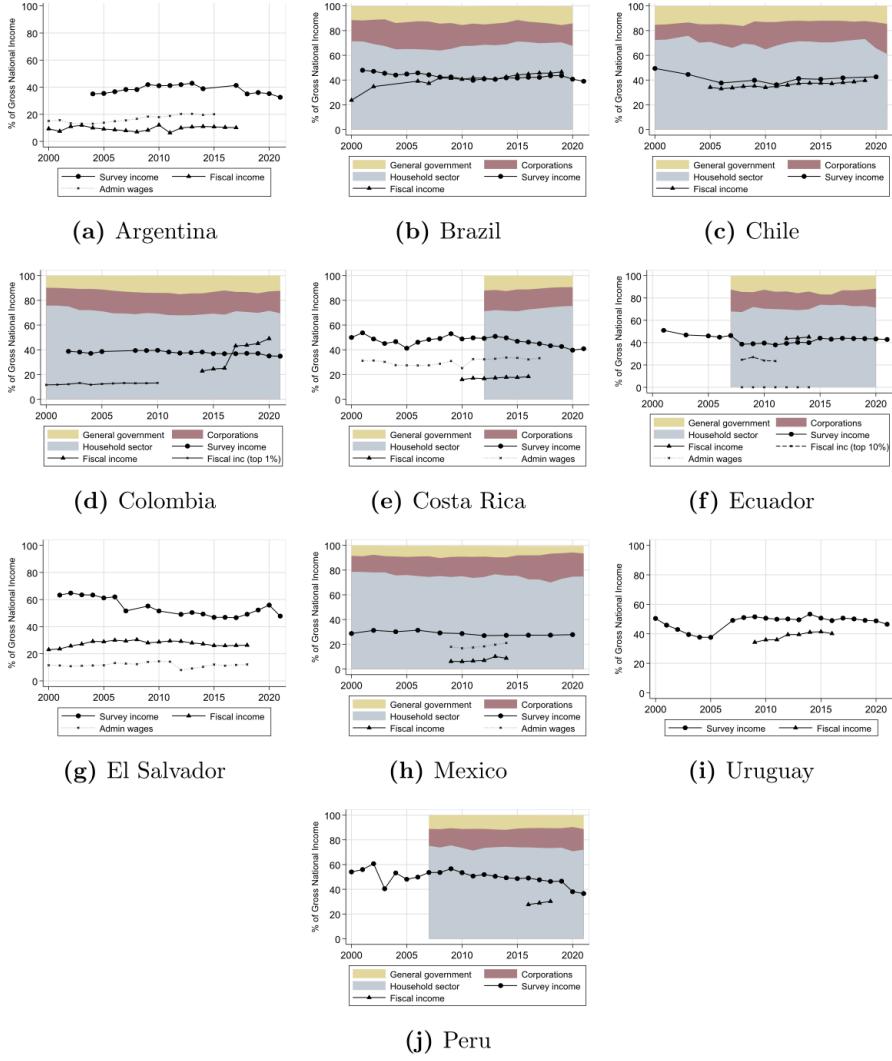


Figura 2: Componentes da renda nacional e proporção efetivamente capturada por surveys domiciliares, dados administrativos e contas nacionais. No caso brasileiro, os surveys cobrem aproximadamente metade da renda total, ficando aquém dos registros administrativos e das estimativas macroeconômicas. Fonte: Rosa et al. (2024, p. 16).

se refere à participação relativa das faixas mais pobres e mais ricas. Além disso, as bases de dados utilizadas foram o *Estimated Household Income Inequality* (EHII), do University of Texas Inequality Project, e o *All the Ginis* (ATG), do Stone Center for Socio-Economic Inequality. Ambas foram construídas a partir de pesquisas domiciliares, que tendem a subestimar sistematicamente a concentração de renda no topo da distribuição.

Em minha investigação, faço uso de estimativas de desigualdade calculadas segundo a metodologia das Contas Nacionais Distributivas (DINA). A DINA combina dados de surveys domiciliares com registros tributários e administrativos, com o objetivo de torná-los compatíveis com as contas nacionais agregadas (Piketty et al., 2018). Além de melhorar a captura da renda no topo, as DINAs permitem calcular a participação relativa de diferentes

segmentos da distribuição (e.g., 50% mais pobres, 10%, 1% e 0,1% mais ricos). Aplicando essa metodologia à América Latina, Rosa et al. (2024) mostraram que os surveys capturam apenas uma fração da renda total. No caso brasileiro, essa fração equivale a cerca de metade da renda nacional, como mostra a [Figura 2](#). Isso reforça o argumento de que temos de estar conscientes das limitações de estimativas de desigualdade baseadas exclusivamente em dados desse tipo.

Levando tudo isso em conta, meu objetivo foi replicar a análise de Hartmann et al., utilizando, porém, os dados da WID. A intenção foi captar as relações entre a complexidade econômica e a distribuição de toda a renda nacional, bem como identificar possíveis diferenças em sua dinâmica ao longo das distintas faixas da distribuição.

3 Revisão Conceitual e Teórica: Da Identificação à Estimação Robusta

Nesta seção, apresento os conceitos e métodos econométricos que fundamentam a estratégia empírica que adotei nesta investigação. Como ponto de partida, tomo o Modelo de Regressão Linear Clássico (MRLC) – nossa referência idealizada – que opera sob pressupostos fortes, próximos ao contrafactual de um experimento controlado. Meu objetivo é mostrar, passo a passo, por que esse modelo se revela insuficiente para lidar com os desafios substantivos e estatísticos da relação entre complexidade e desigualdade, e por que um modelo com Efeitos Fixos e Variáveis Instrumentais oferece uma abordagem mais adequada.

Em primeiro lugar, a estrutura dos dados motivou minha decisão. Como veremos, trata-se de um painel não balanceado de países, com séries anuais retiradas do *World Inequality Database (WID)* e do *Atlas of Economic Complexity*. O Índice de Complexidade Econômica (ECI) – principal variável explicativa – é um ranking competitivo. Ou seja, é uma medida altamente persistente e definida em termos relativos. Isso gera dependência seccional entre unidades. Essa característica viola pressupostos de independência dos erros e exige cuidados especiais de modelagem. Em segundo lugar, o próprio objeto de investigação – o efeito da complexidade econômica sobre a desigualdade – envolve risco concreto de causalidade reversa, o que compromete a interpretação causal do coeficiente estimado por MQO.

Tanto a estrutura dos dados quanto a natureza da pergunta impõem o enfrentamento de duas classes distintas de problemas econométricos:

1. **Ameaças de Primeira Ordem (Viés e Inconsistência):** Essas ameaças comprometem diretamente a interpretação causal. Como discuto inicialmente na [Subseção 3.1](#), elas incluem violações do pressuposto de exogeneidade (como no caso de causalidade reversa), bem como especificações funcionais incorretas. O problema de endogeneidade, por sua vez, recebe tratamento mais aprofundado na [Subseção 3.2](#), onde exploro o uso de Variáveis Instrumentais (IV) como estratégia de identificação. Já a heterogeneidade não observada, comum em dados em painel, é abordada na [Subseção 3.3](#) como outra fonte relevante de viés quando não controlada adequadamente.
2. **Ameaças de Segunda Ordem (Inferência Inválida):** Mesmo que os estimadores sejam consistentes, a validade da inferência estatística depende da forma como os erros padrão são calculados. Violações como heterocedasticidade, autocorrelação e, especialmente, dependência seccional tornam os testes de hipótese não confiáveis. A possibilidade de testes de hipótese não confiáveis constitui um risco elevado nesse tipo de dado. Também na [Subseção 3.1](#), discuto essas violações e apresento os estimadores robustos que utilizei para mitigá-las, como os de White (1980) e Driscoll e Kraay (1998).

Ao final da seção, também discuto os aspectos práticos que afetam a robustez da estimação empírica, como o tratamento de dados faltantes e a estrutura desbalanceada do painel ([Subseção 3.4](#)). Com tudo isso, espero que a leitora compreenda a lógica que sustenta minha escolha metodológica: um modelo de Efeitos Fixos com Variáveis Instrumentais (FE-IV) e erros padrão robustos à dependência seccional.

3.1 Problemas de Estimação: Viés, Variância e os Pressupostos do MRLC

A validade de qualquer análise econométrica depende de um conjunto de pressupostos. Quando esses pressupostos são violados, surgem dois tipos principais de problemas: o *viés* – que compromete a consistência e a interpretação causal dos coeficientes – e a *ineficiência*, que distorce a variância estimada e torna a inferência estatística pouco confiável. Nesta subseção, discuto os principais riscos associados a cada um desses problemas no contexto da minha análise. No campo do viés, destaco a *endogeneidade* e a *má especificação funcional*. No campo da variância, abordo a *multicolinearidade*, a *heterocedasticidade*, a *autocorrelação temporal* e a *dependência seccional*. Todos especialmente relevantes em dados macroeconômicos em painel como os utilizados aqui.

Comecemos com o viés – e a *inconsistência* – que ocorre quando o valor esperado do estimador se desvia sistematicamente do verdadeiro parâmetro populacional. Usando uma analogia, um estimador enviesado é como alguém que atira e que, mesmo com muitos tiros, erra o alvo sempre na mesma direção. Um estimador que seja assim é inútil quando pretendemos qualquer inferência causal. Este é precisamente uma motivação para buscarmos técnicas além do MQO quando seus pressupostos são violados (veja o Verbeek, 2017, capítulo 5, seção 1).

A preocupação de segunda ordem é a *ineficiência* e a *inferência inválida*, que surgem quando os erros padrão são calculados de forma incorreta. Para continuarmos com nossa analogia, neste caso o nosso atirador pode até acertar o alvo na média. Mas seus tiros estão muito dispersos (alta variância), e ele não tem uma noção precisa de sua própria imprecisão. A seguir, discuto cada pressuposto do MRLC sob a ótica desses dois desafios e de sua relevância para minha investigação.

A principal ameaça à interpretação causal da relação entre ECI e desigualdade é a *endogeneidade*. Ela viola o pressuposto mais crítico do MQO: a exogeneidade ($\text{Cov}(X_{it}, \epsilon_{it}) = 0$). Este pressuposto garante que o regressor possui uma variação que é "limpa" de outras influências não observadas sobre a variável dependente. No meu caso, a *causalidade reversa* é a forma mais provável de violação de endogeneidade.

Afinal, é plausível que um alto nível de desigualdade afete negativamente a capacidade de um país de desenvolver os setores produtivos complexos que sustentam um ECI elevado. Isso pode ser o caso por meio de, por exemplo, canais como o subinvestimento em educação de qualidade para a massa da população. Outra possibilidade é a captura do Estado por elites que favorecem setores extrativistas em detrimento da diversificação industrial.

Se isso ocorrer, o ECI do país em questão se torna correlacionado com fatores não observados no termo de erro do meu modelo de desigualdade. Em termos econométricos, essa violação torna o estimador de MQO enviesado e inconsistente ($\text{plim} \hat{\beta}_{OLS} \neq \beta$). Isso torna impossível a interpretação causal do coeficiente, e potencialmente leva a conclusões espúrias (Wooldridge, 2010, vide cap. 4, seção 2.1). Dada a gravidade deste problema, dedico a [Subseção 3.2](#) exclusivamente à discussão e mitigação da endogeneidade.

Outra fonte relevante de viés é a má especificação da *forma funcional*. O MRLC assume que a relação esperada entre Y e \mathbf{X} é linear nos parâmetros ($\mathbb{E}[Y_i|\mathbf{X}] = \mathbf{X}'_i\boldsymbol{\beta}$). No entanto, se a relação verdadeira for, por exemplo, quadrática, e eu estimá-la como linear, a componente não linear será incorporada ao termo de erro. Como essa componente depende de X , o erro se tornará correlacionado com os regressores, violando o pressuposto de exogeneidade e resultando em viés de variável omitida.

Um exemplo clássico é a Curva de Kuznets, que sugere uma relação em U invertido entre desenvolvimento econômico e desigualdade. Essa uma dinâmica que não pode ser capturada por um modelo linear simples. Para mitigar esse risco, aplico em minha análise transformações logarítmicas às variáveis de renda, ou para o próprio ECI. Isso permite testar formalmente a presença de não linearidades.

Superadas as fontes de viés, a análise se concentra agora nos problemas que afetam a precisão do estimador, isto é, sua variância. Um dos principais é a *multicolinearidade alta*, que ocorre quando dois ou mais regressores estão fortemente correlacionados entre si. Embora não introduza viés, esse problema dificulta a identificação do efeito individual de cada variável explicativa, pois o modelo encontra dificuldade em “atribuir” o impacto observado na variável dependente a um único regressor (cf. Kennedy, 2008). Isso acaba em erros padrão elevados, o que amplia os intervalos de confiança e reduz a significância estatística das estimativas.

No meu caso, tanto o PIB per capita quanto o capital humano são determinantes teóricos da desigualdade que tendem a estar fortemente correlacionados. Afinal, países mais ricos usualmente têm populações mais escolarizadas. Para monitorar esse problema, utilizarei o Fator de Inflação da Variância (VIF), que quantifica a redundância entre os regressores e orienta a escolha da especificação mais parcimoniosa e robusta.

A seguir, trato dos problemas associados a erros não esféricos, que são praticamente inevitáveis em dados macroeconômicos em painel. Em um conjunto de países heterogêneos,

é pouco plausível assumir que a variância dos erros seja constante; isto é, é pouco plausível que haja homocedasticidade. A *heterocedasticidade* ($\text{Var}(\epsilon|X) \neq \sigma^2 I$) é, portanto, a hipótese mais realista nesses casos. Por exemplo, os efeitos de uma política de abertura comercial sobre a desigualdade tendem a apresentar maior dispersão em economias complexas e diversificadas do que em economias de baixa renda e estrutura produtiva simples.

Embora a heterocedasticidade não afete os coeficientes do MQO, ela distorce seus erros padrão, frequentemente os subestimando. Isso inflaciona as estatísticas t e pode levar à superestimação da significância estatística (vide Heiss, 2020). Por isso, adotarei como procedimento padrão o teste de Breusch-Pagan para detectar sua presença. Caso haja alta probabilidade de ocorrência, utilizarei erros padrão robustos à heterocedasticidade, conforme White (1980). Isso permite estimar consistentemente a matriz de variância-covariância sem supor uma forma funcional específica para a variância dos erros.

Por fim, a estrutura em painel dos dados e a própria natureza do ECI tornam praticamente inevitáveis dois problemas correlacionados: a *autocorrelação temporal* e a *dependência seccional*. A autocorrelação ocorre porque choques sobre a desigualdade – como uma reforma tributária – tendem a produzir efeitos que persistem ao longo do tempo, gerando correlação entre os termos de erro de anos consecutivos $\epsilon_{it} \sim \epsilon_{it-1}$.

Já a dependência seccional emerge quando choques afetam simultaneamente múltiplas unidades no mesmo período. Isso ocorre em eventos globais – por exemplo, a crise financeira de 2008 –, os quais podem ser parcialmente capturados por efeitos fixos de ano (η_t). No entanto, uma fonte adicional e mais sutil de dependência seccional decorre do próprio conceito do ECI como um ranking competitivo (Hidalgo & Hausmann, 2009).

O avanço de um país no ECI frequentemente implica perdas relativas para outros, sobretudo quando competem nos mesmos mercados. Isso pode induzir correlação entre os resíduos idiossincráticos de países distintos ($\text{Cov}(\mu_{it}, \mu_{jt}) \neq 0$ para $i \neq j$). Ignorar essa violação pode comprometer a validade da inferência, pois leva à subestimação dos erros padrão.

Para enfrentar esse problema, adotarei uma estratégia com duas frentes: i) incluirei efeitos fixos de ano, para controlar choques comuns observáveis; e ii) aplicarei o teste de Pesaran CD para diagnosticar dependência residual (Pesaran, 2004). Também empregarei erros padrão robustos à dependência temporal e seccional, como os estimadores de *Driscoll-Kraay* (Driscoll & Kraay, 1998). Isso contribui para produzirmos uma inferência estatística confiável.

3.2 Estratégia de Identificação: Endogeneidade e Variáveis Instrumentais

Como vimos acima, a endogeneidade é a principal ameaça à validade causal da minha análise. Superá-la é condição necessária para sermos capazes de realizar qualquer *identificação*. Em outras palavras, é condição necessária para sermos capazes de interpretar o coeficiente estimado como efeito causal da variável explicativa principal – a complexidade econômica – sobre a desigualdade.

Métodos de painel (sobre os quais discutiremos em detalhe na [Subseção 3.3](#)) são úteis para controlar formas específicas de viés, como a omissão de características fixas no tempo. Contudo, eles não resolvem, por si mesmos, problemas de causalidade reversa ou de variáveis omitidas que variam ao longo do tempo. Nessas situações, uma estratégia de identificação explícita é indispensável. Aqui, minha estratégia é empregar *Variáveis Instrumentais (IV)*, implementadas por meio de Mínimos Quadrados em Dois Estágios (MQ2E).

Esta subseção detalha, primeiro, as origens da endogeneidade no contexto da minha pesquisa; depois, os requisitos formais e práticos para a validade da abordagem de IV.

3.2.1 Causas da Endogeneidade

As principais causas que podem gerar endogeneidade em minha pesquisa são:

- i A **variável omitida**. Trata-se de um fator não observado que influencia tanto a variável dependente quanto uma ou mais variáveis explicativas incluídas no modelo. Este é o caso, por exemplo, da qualidade das instituições políticas, que pode variar no tempo e afetar tanto a política industrial quanto a distribuição de renda.
- ii O **erro de medida**. Ele ocorre quando as variáveis são mensuradas com imprecisão, fazendo com que o erro de medida na variável explicativa possa ser correlacionado com o erro do modelo. Isso é uma preocupação particular com dados de desigualdade, cuja medição pode variar em qualidade entre países e ao longo do tempo.
- iii A **causalidade reversa ou simultaneidade**. Essa situação ocorre quando a variável dependente influencia a variável explicativa, criando um ciclo de feedback. No meu caso, como discutido na [Subseção 3.1](#), esse é o mecanismo mais plausível de endogeneidade entre desigualdade e complexidade econômica.

3.2.2 O Estimador de Variáveis Instrumentais (IV) e seus Requisitos

A lógica do método de Variáveis Instrumentais (IV) consiste em isolar a parte da variação na variável explicativa endógena (X) que é exógena, isto é, que é livre de correlação com o termo de erro. Para isso, utilizamos uma terceira variável – o instrumento (Z) – que atua como

uma fonte exógena de variação em X . Para que o instrumento seja válido e o estimador de IV seja consistente, duas condições fundamentais e inegociáveis devem ser satisfeitas: *relevância* e *exogeneidade* (ou restrição de exclusão).

A relevância exige que o instrumento tenha poder preditivo sobre a variável endógena, mesmo após o controle pelas demais variáveis exógenas – isto é, $\text{Cov}(Z, X|W) \neq 0$, onde W representa os controles incluídos no modelo. Um instrumento fraco – com baixa correlação parcial com a variável endógena – é um problema importante. Ele inflaciona a variância do estimador e o torna altamente enviesado em amostras finitas, podendo, em alguns casos, ser mais enviesado que o próprio MQO que se pretende corrigir (vide Bound et al., 1995).

Cameron e Trivedi enfatizam que o viés do estimador de IV, relativo ao viés do MQO, é inversamente proporcional à correlação entre o instrumento e o regressor endógeno (ver Cameron & Trivedi, 2005, cap. 4, seção 9). Para avaliar a relevância do instrumento, seguirei a prática padrão de verificar a F-estatística do primeiro estágio da regressão: considerarei valores superiores a 10 evidência contra a fraqueza do instrumento (Staiger & Stock, 1997).

Em segundo lugar, a *restrição de exclusão* estabelece que o instrumento não pode afetar diretamente a variável dependente (Y), tampouco estar correlacionado com os fatores não observados (ϵ) que a influenciam. Em outras palavras, sua única via de impacto deve ser por meio da variável endógena (X). Formalmente, essa condição é expressa por $\mathbb{E}[Z\epsilon] = 0$. Essa é a premissa central da estratégia de identificação.

Por envolver correlação com fatores não observáveis, essa hipótese é, por definição, não testável empiricamente. Sua plausibilidade deve ser justificada com base em argumentos teóricos robustos e conhecimento sobre o processo gerador dos dados. Juntamente com a relevância, essa é uma das duas condições necessárias para que o estimador de IV seja consistente (Wooldridge, 2010, cf. cap. 5, seção 2.1). Caso a restrição de exclusão seja violada – mesmo que o instrumento seja relevante – o estimador torna-se inconsistente, o que invalida a estratégia de identificação.

Quando os instrumentos são válidos, o estimador de IV é *consistente* e assintoticamente normal. No entanto, essa consistência tem um custo: a eficiência. Em geral, o IV apresenta variância maior que a do MQO, o que implica menor precisão nas estimativas. Esta é uma característica bem documentada na literatura (ver Davidson & MacKinnon, 2009, cap. 8, seção 3). Isso impõe um *trade-off* metodológico: ou aceitamos o viés do MQO ou suportamos a maior variância do IV. Considerando a hierarquia dos problemas, a prioridade é eliminar o viés. Por isso, o IV é a abordagem preferencial, desde que possamos identificar instrumentos empiricamente críveis e teoricamente justificáveis.

O uso de *variáveis defasadas (lags)* como instrumentos é uma prática comum para lidar com endogeneidade contemporânea. Estou me baseando aqui no argumento de que

valores passados da variável explicativa (X) não podem ser causados por valores presentes da variável dependente (Y). No entanto, a validade dessa abordagem em modelos de painel depende criticamente da ausência de correlação serial no erro idiosincrático (μ_{it}). Se houver autocorrelação nos resíduos, então X_{it-1} poderá estar correlacionado com μ_{it} por meio de μ_{it-1} . Isso violaria a restrição de exclusão. Por isso, o uso de defasagens como instrumentos exige cautela e deve ser precedido por testes de diagnóstico adequados.

No contexto do ECI, a construção de instrumentos válidos impõe desafios consideráveis. A alta persistência da variável tende a enfraquecer a relevância das defasagens. Já a busca por instrumentos externos – que afetem a estrutura produtiva de um país sem incidir diretamente sobre sua desigualdade – esbarra em fortes restrições empíricas.

Uma alternativa conceitualmente promissora é o uso de instrumentos do tipo *Shift-share* (Bartik, 1991). Esses instrumentos exploram como a composição setorial inicial de um país o expõe, de forma diferenciada, a choques de demanda globais, o que gera variação exógena na estrutura produtiva. No entanto, a implementação desse método seria complexa demais.

Por isso, a estratégia que adoto neste trabalho se concentra no uso de defasagens e da estrutura de painel. Isso é complementado por testes de robustez e uma discussão transparente sobre os limites da identificação.

3.3 Modelos para Dados em Painel: Lidando com Múltiplas Fontes de Heterogeneidade e Dependência

A estratégia de IV, conforme discutido, é a principal ferramenta adotada para lidar com formas de endogeneidade que variam no tempo. Contudo, uma fonte crítica de viés ainda persists: a heterogeneidade não observada e invariante no tempo entre os países.

Além disso, a própria natureza da variável explicativa principal introduz um desafio metodológico adicional. O ECI é uma medida *relativa*: ele expressa a posição de um país em um ranking global de complexidade produtiva. Isso significa que o avanço de uma economia nesse índice tende a implicar o recuo de outras, especialmente quando competem em mercados semelhantes.

Essa característica gera uma interdependência estrutural entre as unidades, tornando provável a ocorrência de *dependência seccional* – isto é, correlação entre os termos de erro de diferentes países em um mesmo período. Essa violação da suposição de independência entre unidades será tratada explicitamente nas estratégias de estimação discutidas ao longo desta seção.

Ignorar essas características pode gerar correlação espúria entre as variáveis de interesse. É nesse ponto que a estrutura de *dados em painel* se torna particularmente útil: ao accompa-

nhar as mesmas unidades (i) ao longo do tempo (t), ela permite controlar sistematicamente esses fatores fixos, oferecendo uma base mais sólida para a identificação causal.

A principal vantagem dos dados em painel é que eles permitem decompor a variação total dos dados em duas dimensões fundamentais: a variação *entre* os países (cross-sectional) e a variação *dentro* de cada país ao longo do tempo ("time-series"). Essa decomposição é essencial para isolar o efeito de interesse de possíveis fontes de viés associadas a características invariantes entre as unidades.

O termo de erro em um modelo de painel pode ser decomposto para refletir essas múltiplas fontes de heterogeneidade:

$$\epsilon_{it} = c_i + \eta_t + \mu_{it} \quad (3)$$

Essa decomposição revela três fontes distintas de variação não observada, cada uma exigindo uma estratégia adequada de controle. É o que descrevo a seguir:

- c_i : **Efeito individual fixo** (heterogeneidade não observada entre países). Esse termo captura todas as características invariantes no tempo de um país. Essas características podem ser observáveis ou não (por exemplo, cultura, geografia, instituições históricas). Se $\text{Cov}(X_{it}, c_i) \neq 0$ – o que é altamente plausível no contexto desta análise – o estimador de MQO Simples (Pooled OLS) será enviesado por omissão de variável relevante. Para eliminar esse viés, adoto o estimador de *Efeitos Fixos (FE)*, também conhecido como estimador *within*.

Ao transformar os dados em desvios em relação à média de cada país (ou, equivalente, ao incluir uma dummy para cada unidade), o FE remove c_i da equação de estimação. Sob a suposição de exogeneidade estrita em relação ao erro idiossincrático, o estimador FE é *consistente*. Isso é o caso mesmo quando os regressores X_{it} estão correlacionados com c_i (Wooldridge, 2010, ver cap. 10, seção 5.1). A principal limitação do FE é que ele não permite estimar o efeito de variáveis invariantes no tempo (como geografia). Além disso, ele pode ser menos eficiente do que o estimador de Efeitos Aleatórios (RE). Isso pode ocorrer quando a correlação entre os regressores X_{it} e a heterogeneidade não observada c_i for nula ou desprezível – ou seja, justamente quando a suposição que motiva o uso do FE não se verifica.

- η_t : **Efeito temporal fixo**. Representa choques comuns a todos os países em um dado período. Falamos aqui de crises financeiras globais, pandemias ou ondas de inovação tecnológica. Se tais choques forem correlacionados com o ECI e não forem controlados, podem introduzir viés nas estimativas. Para lidar com esse risco, incluo sistematicamente *dummies de ano* em todas as regressões. Essa prática constitui o chamado

modelo de efeitos fixos de duas vias (*two-way fixed effects*).

- μ_{it} : **Erro idiosincrático.** Refere-se a todos os demais fatores não observados que variam tanto entre países quanto ao longo do tempo. Idealmente, esse termo deveria se comportar como ruído branco. No entanto, é sobre ele que recai a endogeneidade decorrente de causalidade reversa, ou da omissão de variáveis que variam no tempo. Por isso, mesmo após o uso de efeitos fixos, a estratégia de Variáveis Instrumentais continua sendo essencial para garantir que $Cov(X_{it}, \mu_{it}) = 0$.

A principal alternativa ao modelo de Efeitos Fixos (FE) é o modelo de *Efeitos Aleatórios (RE)*. Nesse caso, o termo c_i não é tratado como um parâmetro específico a ser estimado (ou eliminado), mas como uma componente aleatória do erro. O estimador RE tende a ser mais eficiente do que o FE – isto é, apresenta menor variância assintótica – porque utiliza simultaneamente a variação *entre* os países (cross-sectional) e a variação *dentro* de cada país ao longo do tempo.

Contudo, essa eficiência vem acompanhada de uma suposição muito mais restritiva: a de que os efeitos individuais c_i são *não correlacionados* com as variáveis explicativas do modelo, ou seja, $Cov(X_{it}, c_i) = 0$. No contexto desta pesquisa, essa hipótese é teoricamente implausível. Afinal, fatores fixos como qualidade institucional, cultura política ou estrutura histórica da economia podem afetar simultaneamente a complexidade econômica (ECI) e os níveis de desigualdade.

Guiarei a escolha formal entre os modelos de Efeitos Fixos (FE) e Efeitos Aleatórios (RE) pelo *teste de Hausman* (Hausman, 1978). A lógica do teste consiste em comparar dois estimadores sob o critério de consistência. O estimador FE é consistente tanto sob a hipótese de correlação quanto de não correlação entre c_i e os regressores X_{it} . Por sua vez, o estimador RE só é consistente se não houver correlação.

Assim, se os coeficientes estimados pelos dois modelos forem estatisticamente diferentes, isso indica que o RE está violando sua condição de validade – em outras palavras, que é enviesado. A hipótese nula do teste (H_0) afirma que não há diferença sistemática entre os coeficientes dos dois modelos, o que implicaria que o RE é consistente e, por ser mais eficiente, preferível. A rejeição de H_0 (isto é, um p-valor baixo) fornece evidência empírica de que apenas o FE é consistente. Nesse caso, o modelo de efeitos fixos deve ser preferido para a estimação.

Toda a discussão metodológica acima partiu de modelos estáticos. No entanto, muitos fenômenos socioeconômicos – como a desigualdade de renda – apresentam elevada persistência temporal, o que desafia essa abordagem (ver [Figura 3](#)).

Uma maneira de capturar tal persistência seria estimar um *modelo dinâmico de painel* que incluisse a variável dependente defasada (Y_{it-1}) como regressor. Entretanto, em painéis

com poucos períodos, essa escolha acarreta o chamado *viés de Nickell* (Nickell, 1981). Nele, a transformação *within* gera correlação mecânica entre o regressor defasado e o erro transformado, tornando o estimador de efeitos fixos enviesado e inconsistente. Técnicas baseadas em GMM – como Arellano e Bond (1991) e Blundell e Bond (1998) – podem mitigar o problema. O problema é que requerem instrumentos fortes e um balanço temporal que ultrapassa o escopo deste estudo.³

Dado esse trade-off, optei por modelos estáticos com controles de tempo e defasagens exógenas, complementados por testes de autocorrelação nos resíduos. Essa escolha mantém o foco no objetivo principal: identificar o efeito contemporâneo da complexidade produtiva sobre a distribuição de renda. Ela também evita complicações adicionais de instrumentação em painéis decadais relativamente curtos. Possíveis extensões dinâmicas são discutidas nas considerações finais.

3.4 Aspectos Práticos e a Robustez da Estratégia de Estimação

Finalmente, a validade de qualquer estratégia econométrica – por mais sofisticada que seja – depende da qualidade dos dados e da transparência nas decisões de tratamento da amostra. Questões como *dados faltantes*, *tração* (attrition) e a um painel de *estrutura desbalanceada* podem comprometer a robustez das estimativas e precisam ser consideradas com cuidado.

Na minha análise, a variação na disponibilidade de observações de desigualdade na base WID entre países e ao longo do tempo me leva a trabalhar com um painel desbalanceado (vide [Figura 5](#)). Estimar modelos com um painel desse tipo não é um problema em si, desde que a razão pela qual os dados estão faltando seja exógena ao processo que estou investigando.

Contudo, se a ausência de dados for sistemática, os resultados podem ser severamente enviesados. Este fenômeno é conhecido como tração (attrition) ou viés de seleção. Por exemplo, se países que passam por crises econômicas severas (e potenciais aumentos na desigualdade) param de reportar dados, minha amostra final estaria selecionando apenas os países "bem-sucedidos". Consequentemente, as estimativas do efeito do ECI seriam viesadas.

É difícil testar formalmente a natureza desse processo de seleção. Contudo, orientei as decisões que tomei no tratamento de valores ausentes (NAs) e *outliers* – descritas na seção [Subseção 4.1](#) – com o objetivo de manter uma amostra consistente e preservar o poder estatístico. Ao mesmo tempo, estive atento ao impacto potencial dessas escolhas sobre a robustez das estimativas. Por isso, a sensibilidade dos resultados a diferentes especificações de amostra será um componente central da análise de robustez.

³Confira Bond (2002) para a derivação formal do viés e Roodman (2009) para uma discussão aplicada do trade-off entre persistência e endogeneidade em painéis curtos.

4 Aplicação Empírica e Resultados

Nesta seção, aplico a estrutura metodológica delineada na [Seção 3](#) a um *painel de países com médias decadas entre 1980 e 2000*, construído a partir da *World Inequality Database* (WID) e do *Atlas of Economic Complexity*. O objetivo é examinar se — e em que medida — a complexidade econômica de um país se associa à sua estrutura distributiva. Mais do que identificar correlações simples, busco evidências de padrões robustos e plausivelmente causais. Também exploro nesta seção alguns mecanismos distributivos subjacentes à relação entre sofisticação produtiva e desigualdade.

Organizei a análise segundo uma trajetória progressiva. Começamos com associações agregadas entre o índice de Gini e a complexidade econômica. Avançamos para associações com efeitos distributivos heterogêneos. E passamos por sucessivos testes de robustez e estratégias de controle da relação de meu interesse.

Na [Subseção 4.1](#), apresento as fontes de dados, a construção da amostra e as primeiras análises descritivas. Faço isso destacando alguns dos desafios com análises conduzidas em um painel desbalanceado. A [Subseção 4.2](#) introduz a especificação econométrica que adotei. Ali comparo diferentes estimadores e termino com o resultado mais importante deste trabalho: o *padrão da foice*, um realinhamento distributivo concentrado no topo da pirâmide de renda.

A [Subseção 4.3](#) trata da endogeneidade das estimativas por meio de defasagens temporais e uma tentativa de estimativa instrumental. Na [Subseção 4.4](#), testo a generalidade do achado da foice com base em diferentes grupos de países e especificações temporais. Termino na [Subseção 4.5](#) com uma leitura sintética dos resultados.

4.1 Dados, Amostra e Análise Descritiva

Nesta subseção, apresento as bases de dados para minha análise. Detalho os dados utilizados, os critérios de construção da amostra e os principais padrões observacionais que motivam minha modelagem econométrica.

Organizei a exposição em cinco partes: (i) A descrição das fontes primárias de dados – o *Atlas of Economic Complexity* e a *World Inequality Database* – com ênfase em sua relevância para o problema em questão; (ii) os procedimentos adotados para a construção da amostra analítica e as implicações metodológicas decorrentes dessas escolhas; (iii) uma discussão sobre o caráter desbalanceado do painel e os possíveis vieses associados à ausência sistemática de dados; (iv) a apresentação de estatísticas descritivas, distribuições de densidade e estrutura média da desigualdade na amostra; e, finalmente, (v) a análise visual da correlação entre complexidade econômica e diferentes medidas de desigualdade, que fornece as hipóteses de trabalho a serem testadas nos modelos seguintes.

Tabela 1: Resumo descritivo da base de dados da WID

Indicador	Valor
Total de observações	57.208.488
Número de variáveis	1.793
Período coberto	1807–2023
Ano médio das observações	1999
Desvio padrão do ano das observações	20 anos
Quartis do ano das observações	Q25 = 1990, Q50 = 2002, Q75 = 2012
Mínimo de anos por país	43 anos
Máximo de anos por país	214 anos
Média de anos por país	87 anos
Quartis dos anos por país	Q25 = 44, Q50 = 74, Q75 = 95

Fonte: elaboração própria com dados da World Inequality Database (WID).

4.1.1 Fontes dos Dados: Atlas da Complexidade e World Inequality Database

A seguir, falarei das duas bases de dados utilizadas para a construção do painel: o *Atlas of Economic Complexity* fornece o índice de complexidade econômica (ECI), e a *World Inequality Database* (WID), que oferece séries sobre distribuição de renda e riqueza. Ambas as bases foram escolhidas por sua amplitude temporal, cobertura internacional e aderência ao meu objetivo: investigar como a estrutura produtiva de um país se relaciona com seus padrões de distribuição de renda.

O *Atlas of Economic Complexity*. O *Atlas* oferece informações detalhadas sobre comércio internacional, estrutura produtiva e trajetória de crescimento de dezenas de países ao longo do tempo. Compilado pelo *Center for International Development* da Universidade de Harvard ([n.d.](#)), ele reúne dados de comércio bilateral desde 1962 e classifica produtos conforme a complexidade de sua produção. Neste trabalho, sigo a abordagem de Hartmann et al. ([2017](#)) e utilizei, dentre as bases disponíveis, a *Standard International Trade Data, Revision 2* (SITC-2). Ela permite calcular uma medida comparável de complexidade econômica (ECI) para um conjunto amplo de países desde os anos 1960.

Como vimos na [Subseção 2.1](#), o Índice de Complexidade Econômica (ECI) é calculado a partir de duas dimensões. Primeiro, da diversidade da pauta exportadora de um país, ou seja, de quantos produtos diferentes ele exporta. Segundo, da ubiquidade dos produtos, isto é, de quantos países são capazes de exportar cada produto. Basicamente, países que exportam muitos produtos pouco comuns tendem a apresentar ECI elevado.

Trata-se, portanto, de uma medida ranqueada e relativa, recalculada a cada ano com base na matriz global de comércio. Essa característica é metodologicamente relevante para meus propósitos. Em princípio, poderíamos pensar que isso viola o pressuposto de indepen-

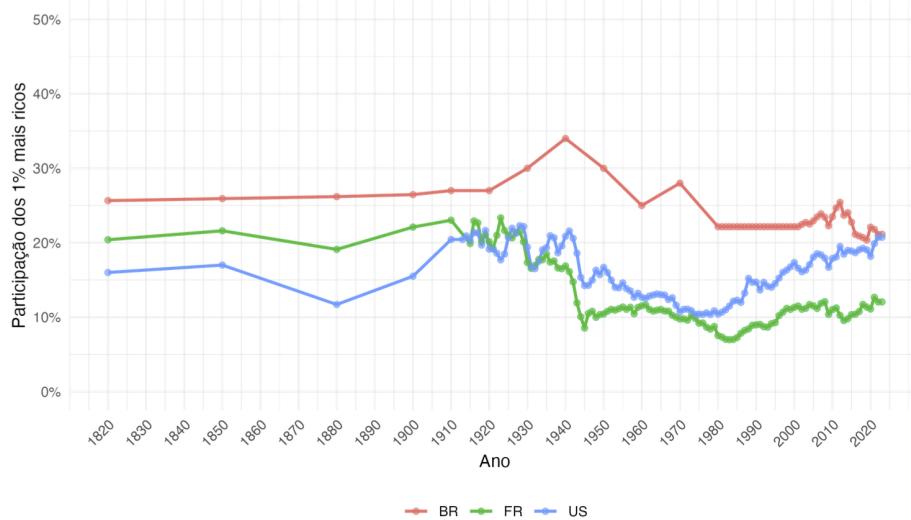


Figura 3: Participação do 1% mais rico na renda total, para Brasil, França e Estados Unidos (1820–presente).
Fonte: elaboração própria com dados da WID.

dência entre unidades cross-section. Afinal, o valor atribuído a um país depende da matriz global de seu comércio. Essa interdependência estrutural entre os regressores dos países poderia induzir correlação seccional não observada.

Para mitigar esse risco, utilizei modelos com efeitos fixos e testes de dependência seccional. Os resultados não indicaram violação severa desse pressuposto. Além disso, a estabilidade relativa do ECI ao longo do tempo reduz a suscetibilidade a choques idiossincráticos que poderiam contaminar a inferência.⁴

O World Inequality Database (WID). A WID é uma das principais fontes globais sobre desigualdade de renda e riqueza. Ela oferece dados para mais de 200 países e regiões, com mais de 57 milhões de observações cobrindo o período de 1807 a 2023, e quase 1800 variáveis (veja a [Tabela 1](#)). Ela segue a metodologia das Contas Nacionais Distributivas (DINA) (Blanchet et al., 2024; Piketty et al., 2018). A DINA combina registros fiscais, pesquisas domiciliares e contas nacionais para gerar séries historicamente comparáveis e conceitualmente consistentes.

A amplitude temporal é uma das propriedades que mais se sobressaem na WID. Uma ilustração disso é a [Figura 3](#). Ela apresenta uma série da participação do 1% mais rico na renda total do Brasil, Estados Unidos e França. O resultado são padrões persistentes de desigualdade no topo, além de estruturalmente diferenciados entre países ao longo de 1820 e 2020.

Essa estabilidade é fundamental do ponto de vista econômico. Ela permite captar

⁴Discutimos esse ponto na [Subseção 3.3](#).

trajetórias distributivas com menor risco de ruído transicional e justifica, em outros contextos, o interesse por especificações com defasagens ou formulações dinâmicas. Como discutido na [Subseção 3.3](#), variáveis com elevada persistência intertemporal – como a desigualdade no topo – desafiam abordagens estáticas e podem demandar estimadores que controlem para autocorrelação serial e viés de Nickell em painéis curtos. No entanto, aqui me restringi a modelos estáticos, justamente para evitar os problemas de endogeneidade introduzidos pela inclusão de variáveis defasadas em painéis com poucas unidades de tempo (veja a discussão na [Subseção 3.4](#)).

Outro ponto metodologicamente relevante é a possibilidade de distinguirmos entre diferentes conceitos de renda. A WID fornece medidas de renda pré e pós-tributação. Isso permite investigar não apenas os níveis de desigualdade, mas também o papel redistributivo dos estados.

A partir dessas duas bases, construí um painel internacional com médias decadais para o período de 1980 a 2000. Priorizei a cobertura consistente das variáveis mais importantes para minha análise. A próxima subseção detalha os passos de harmonização e filtragem que conduziram à amostra analítica final.

4.1.2 Construção da Amostra Analítica e Implicações Metodológicas

O processo de preparação da amostra seguiu uma sequência de etapas sintetizadas no esquema da [Figura 4](#). Podemos caracterizá-lo como um “funil” que leva dos dados brutos extraídos da WID e do *Atlas* até o painel final utilizado nas regressões. A estrutura resultante é um painel decadal, que organiza as informações por país e década.

Mais especificamente, esse processo envolveu: (i) carregamento e empilhamento das séries da WID; (ii) cálculo do *ECI* anual a partir da matriz SITC-2; (iii) junção com dados de PIB, população e escolaridade; (iv) harmonização de identificadores país-ano; (v) agregação por década; e (vi) expurgo de observações com dados faltantes nas variáveis-chave.

Esquema da Construção da Amostra Analítica

- (i) Carregamento e empilhamento das séries de desigualdade da WID.
- (ii) Cálculo do *ECI* anual com base na matriz SITC-2 (Atlas da Complexidade).
- (iii) Junção com dados auxiliares: PIB per capita, população e escolaridade média.
- (iv) Harmonização dos identificadores de país e ano (ISO3 + década).
- (v) Agregação dos dados por década, para captar tendências de longo prazo.
- (vi) Expurgo de observações com NAs nas variáveis-chave.

Figura 4: Etapas de Construção da Amostra Analítica

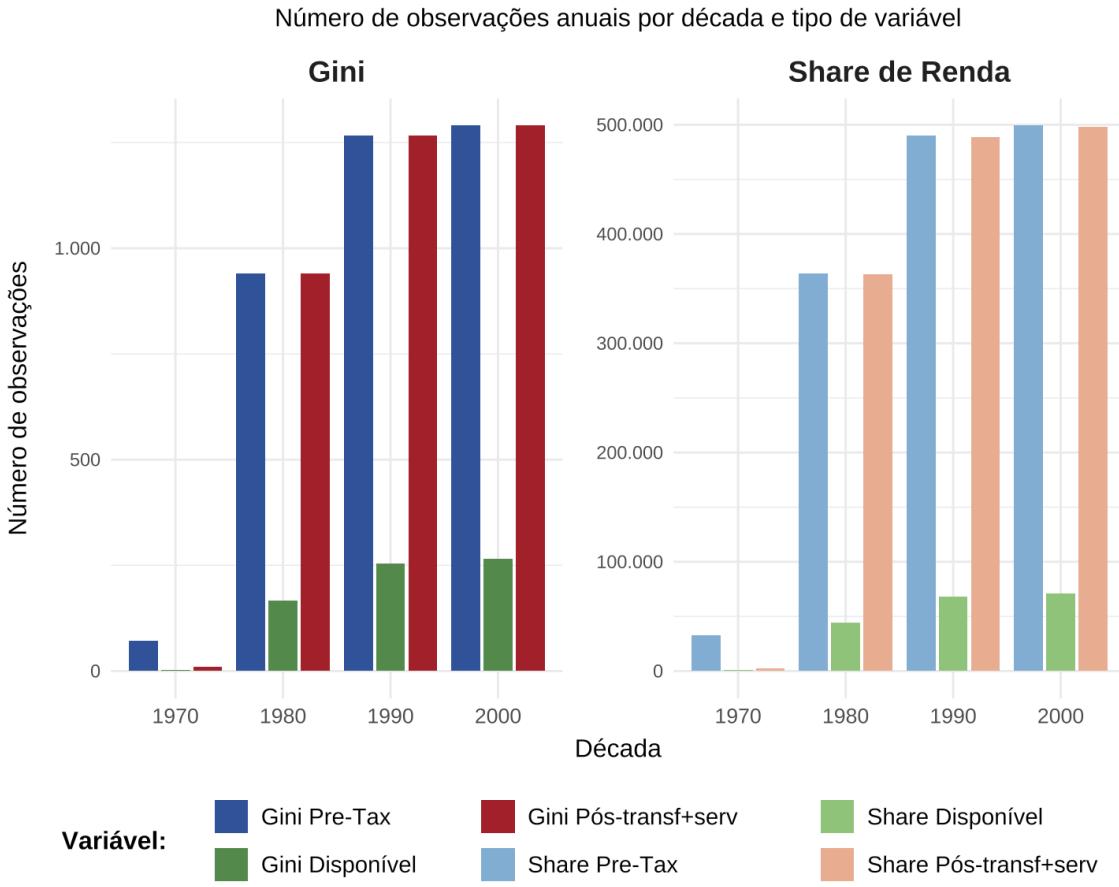


Figura 5: Distribuição de Observações da WID por Década e Variável. A densidade observacional se torna expressiva apenas a partir dos anos 1980, especialmente para as variáveis pós-transferências.

A decisão de trabalhar com médias decadais e de iniciar a análise a partir de 1980 foi guiada pela densidade observacional das variáveis principais. Como mostra a Figura 5, na WID, as medidas de Gini e os *shares* distributivos só passam a apresentar volume informativo compatível com uma análise robusta em painel a partir dessa década. Esse é o caso, especialmente, das variáveis pós-transferências.

Após a aplicação de todos os filtros e procedimentos descritos, a amostra analítica final terminou com **194 observações país-década**, referentes a **115 países** no período entre 1980 e 2000. A lista completa dos países pode ser consultada no [Apêndice I](#).

4.1.3 Discussão do painel desbalanceado e possíveis vieses (attrition, seleção)

Nosso painel apresenta características que impõem desafios metodológicos importantes. Destaco três aspectos: o desbalanceamento da amostra, o risco de atração seletiva e a presença de dependência seccional entre unidades.

- **Painel desbalanceado.** O expurgo de observações com valores ausentes (NAs) resul-

Tabela 2: Estatísticas Descritivas das Variáveis do Painel Decadal

Variável	Média	Desv. Padrão	[Mín, Máx]
Índice de Complexidade Econômica (ECI)	0.23	0.94	[-1.99, 2.65]
Gini (Pré-Tributação)	0.55	0.11	[0.24, 0.84]
Gini (Pós-Tributação, com in-kind)	0.49	0.14	[0.13, 0.70]
PIB per capita (log)	-3.63	2.74	[-10.59, 5.22]
Anos de Estudo	7.46	3.28	[0.65, 13.71]

Nota: As estatísticas referem-se à amostra decadal utilizada nas regressões.

tou em um painel desbalanceado. Conforme discutido por Wooldridge, os estimadores *within* (FE) e de primeiras diferenças (FD) permanecem consistentes sob certas condições. Notadamente, esse é o caso quando os dados faltam de forma aleatória e não correlacionada com os termos de erro (Wooldridge, 2010, ver cap. 17, seções 7.1 e 7.3). No entanto, a variância dos estimadores pode ser inflada, exigindo correções nos erros-padrão e reforçando a necessidade de cuidados na inferência.

- **Viés de atrição.** A ausência não aleatória de dados – por exemplo, concentrada entre países de menor renda ou capacidade estatística – pode gerar viés de seleção. Para mitigar esse risco, adotei um ponto de corte ("threshold") mínimo de 15 anos de observações por país antes da agregação decadal. Além disso, em testes de robustez, utilizei ponderações por probabilidade inversa de inclusão na amostra analítica.
- **Dependência seccional.** O *ECI* é calculado a partir de uma matriz mundial de exportações. Logo, choques globais ou variações sistêmicas afetam simultaneamente todas as unidades. Para acomodar essa correlação "cross-sectional", a análise subsequente deve efeitos fixos de tempo (η_t). Além disso, os erros-padrão serão estimados pelo método de Driscoll–Kraay (1998), que é robusto tanto à dependência entre unidades quanto à autocorrelação serial.

São essas precauções – isto é, com filtros mínimos de qualidade, controle para heterogeneidade não observada e correções de erro – que motivaram as escolhas de modelagem discutidas na Subseção 4.2. A robustez dos resultados dependerá, em larga medida, da capacidade dos modelos adotados em lidar com essas fontes potenciais de viés.

4.1.4 Análise Descritiva da Amostra Final

Uma vez construída a amostra de trabalho, podemos analisar suas características. A Tabela 2 apresenta estatísticas descritivas das variáveis centrais utilizadas nas regressões. O Índice de Complexidade Econômica (ECI) possui média próxima de zero (0,23) e ampla variação

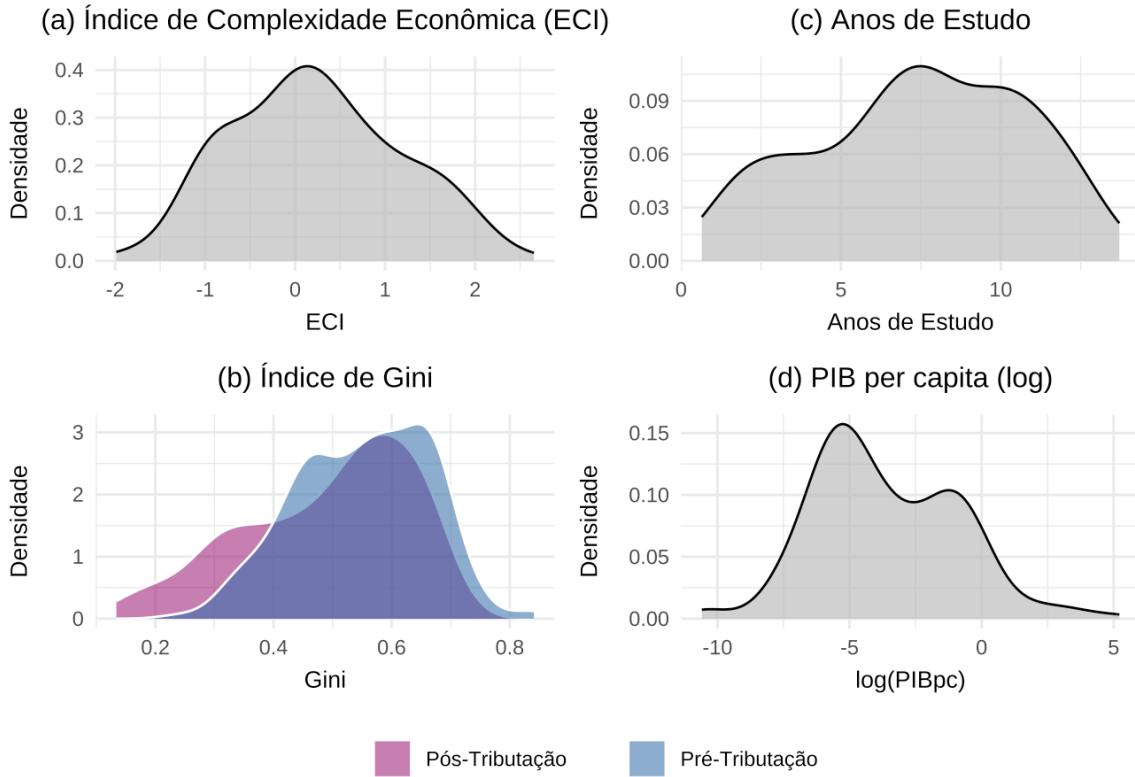


Figura 6: Densidades de Kernel das Variáveis-Chave na Amostra Decadal

(desvio padrão de 0,94), cobrindo desde países altamente complexos (ECI acima de 2) até economias com exportações com pouca complexidade (ECI abaixo de -1,9). O índice de Gini pré-tributação apresenta média de 0,55, enquanto sua versão pós-tributação (com serviços in-kind) tem média reduzida para 0,49. Essa diferença sugere algum impacto redistributivo da ação estatal.

O PIB per capita (em logaritmo) revela grande dispersão entre os países da amostra, com valores que variam de -10,6 a 5,2. Já a escolaridade média se situa em torno de 7,5 anos, com desvio padrão de 3,3. Isso indica significativa heterogeneidade entre os países quanto ao capital humano.

As distribuições dessas variáveis podem ser vistas na Figura 6. A densidade do ECI (painel a) é aproximadamente simétrica e centrada em torno de zero, com leve achatamento. Isso está conforme uma diversidade razoável na complexidade produtiva dos países. O painel (c), que representa os anos médios de estudo, exibe uma distribuição bimodal: há um grupo de países concentrado entre 4 e 6 anos, e outro entre 9 e 11 anos. Isso sugere uma clivagem educacional global entre economias com baixa e alta acumulação de capital humano. Essa heterogeneidade será relevante para interpretar os efeitos da escolaridade no modelo adiante.

Por sua vez, a distribuição do PIB per capita em log (painel d) é marcadamente assi-

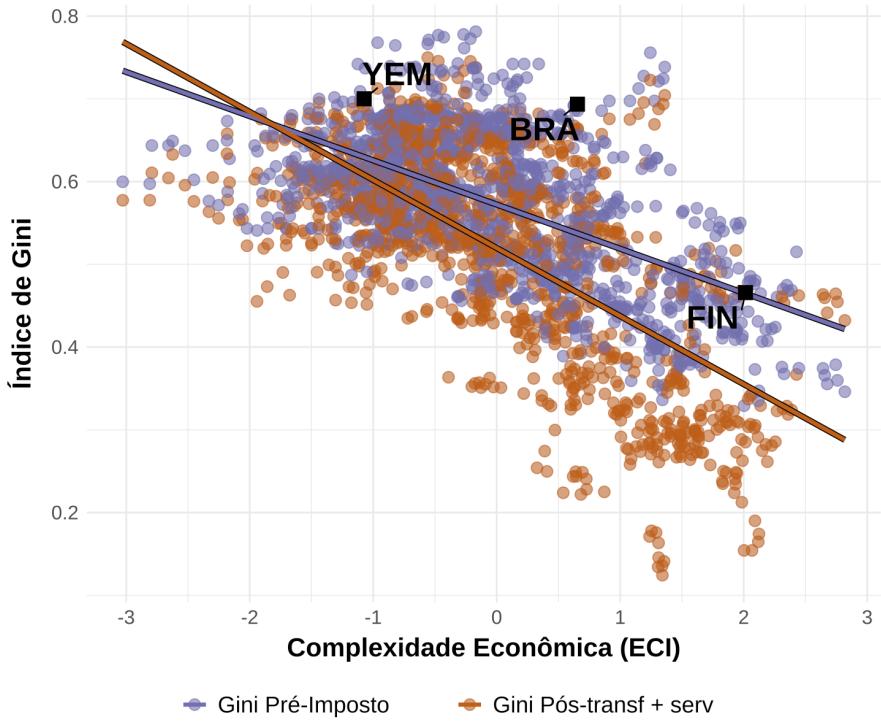


Figura 7: Correlação entre Complexidade Econômica e Índice de Gini

métrica, com uma longa cauda à esquerda. Esse padrão reflete a presença de países de renda muito baixa.

Já o painel (b) revela uma diferença clara entre os dois conceitos de Gini: a curva da versão pós-tributação se desloca para a esquerda em relação à pré-tributação. Isso indica uma redução média da desigualdade após a ação fiscal. Contudo, a sobreposição considerável entre as curvas sugere efeitos redistributivos heterogêneos. Em outras palavras, há países em que a redução é modesta, e outros em que ela é mais substancial.

4.1.5 Padrões Visuais e Hipóteses de Trabalho

Finalmente, façamos agora a análise visual da correlação entre a complexidade econômica e as medidas de desigualdade relevantes, isto é, o Gini e a distribuição de renda por faixas específicas. Isso fornecerá as hipóteses de trabalho que serão testadas econometricamente.

A Figura 7 mostra a relação entre o Índice de Complexidade Econômica (ECI) e o índice de Gini, tanto antes quanto depois da tributação e transferências. O padrão é nitidamente negativo: quanto maior a complexidade econômica, menor tende a ser a desigualdade agregada. Embora ambas as curvas de regressão sejam decrescentes, a inclinação da reta para o Gini pós-tributação é mais acentuada, sugerindo que a combinação entre complexidade produtiva e ação fiscal pode ter efeitos redistributivos cumulativos.

Os países destacados – Finlândia, Brasil e Iêmen – ilustram bem esse gradiente: enquanto a Finlândia combina alta complexidade e baixa desigualdade, o Iêmen representa o extremo oposto, com baixa complexidade e elevada desigualdade. Já o Brasil ocupa uma posição intermediária, com nível relativamente alto de complexidade, mas ainda mantendo um patamar elevado de desigualdade.

Com a [Figura 8](#), começamos a nos aproveitar do nível desagregado de dados do WID. Ela nos permite analisar a relação entre ECI e a desigualdade de renda por faixas específicas da distribuição. Os dois gráficos superiores mostram a relação do ECI com a participação acumulada da base (bottom 50%) e da faixa intermediária (P50–P90). O padrão dos resultados é o seguinte: quanto maior a complexidade econômica, maior tende a ser a participação relativa da base e do meio da distribuição. Isso é observável tanto antes quanto depois das transferências. Esses resultados são coerentes com os resultados de Hartmann et al. de “compressão distributiva” associada à complexidade.

Nos dois gráficos inferiores da mesma figura, observamos o comportamento das fatias superiores da distribuição (top 10% até o top 0,001%). Aqui, a relação é inversa: a complexidade econômica está associada a uma menor concentração de renda no topo. Ainda que o declive das regressões seja mais pronunciado nas faixas superiores (top 10%, top 1%), a tendência negativa persiste mesmo para os percentis mais extremos da distribuição.

O fato de o efeito parecer se atenuar conforme avançamos para percentis mais raros – do top 1% ao top 0,001% – é, por si só, relevante. Sugeriria que os extratos mais concentrados da renda podem responder de forma diferente à estrutura produtiva do que o restante do topo. Esse descolamento seria coerente com achados recentes da literatura que têm encontrado diferenças importantes entre aqueles no topo – isto é, dentro do próprio 1% mais rico – e os 99% em diferentes domínios (Rosa et al., [2024](#)). Esse seria o caso, por exemplo, no que diz respeito às fontes de renda, estratégias de acumulação e inserção produtiva.

Utilizei o futuro do pretérito no parágrafo anterior pois, ainda que sejam padrões visuais sugestivos, eles não controlam para heterogeneidade entre países, efeitos temporais ou covariáveis relevantes. Precisamos testar se essas associações se sustentam diante de potenciais vieses e restrições empíricas. As seções seguintes aplicam modelos de painel com efeitos fixos e erros robustos precisamente com esse objetivo em mente.

4.2 Modelo de Painel Base e Descoberta do Mecanismo Distributivo

Nesta seção, aplico modelos de painel para estimar o efeito da complexidade econômica sobre a desigualdade de renda. O exercício segue três etapas complementares. Primeiro, comparo

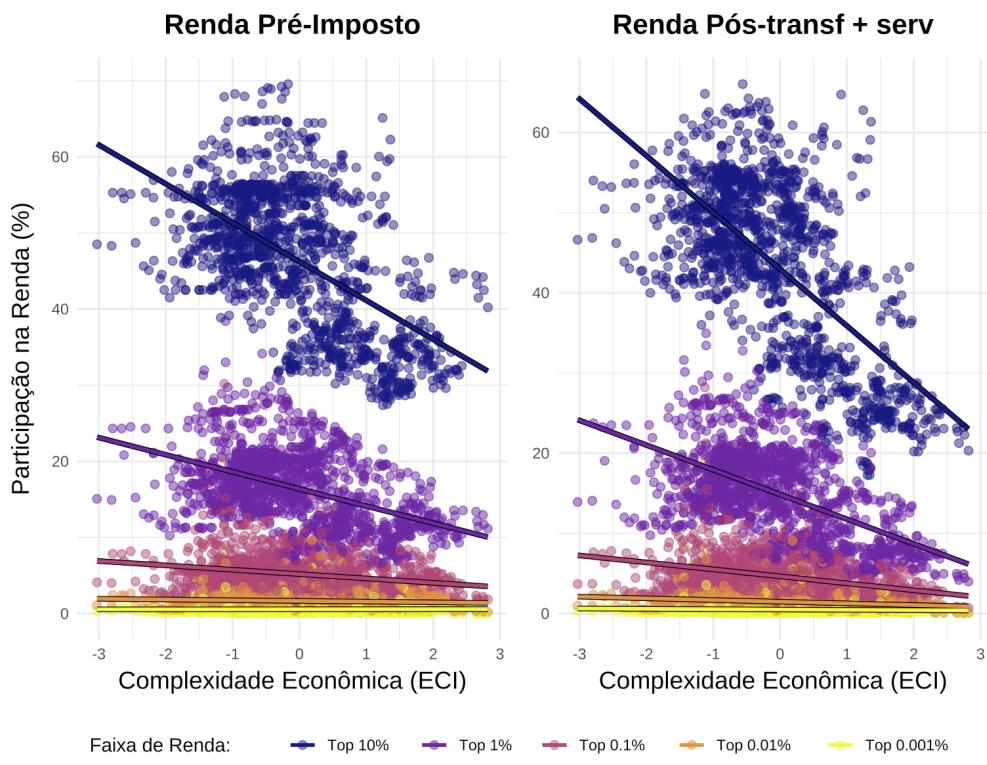
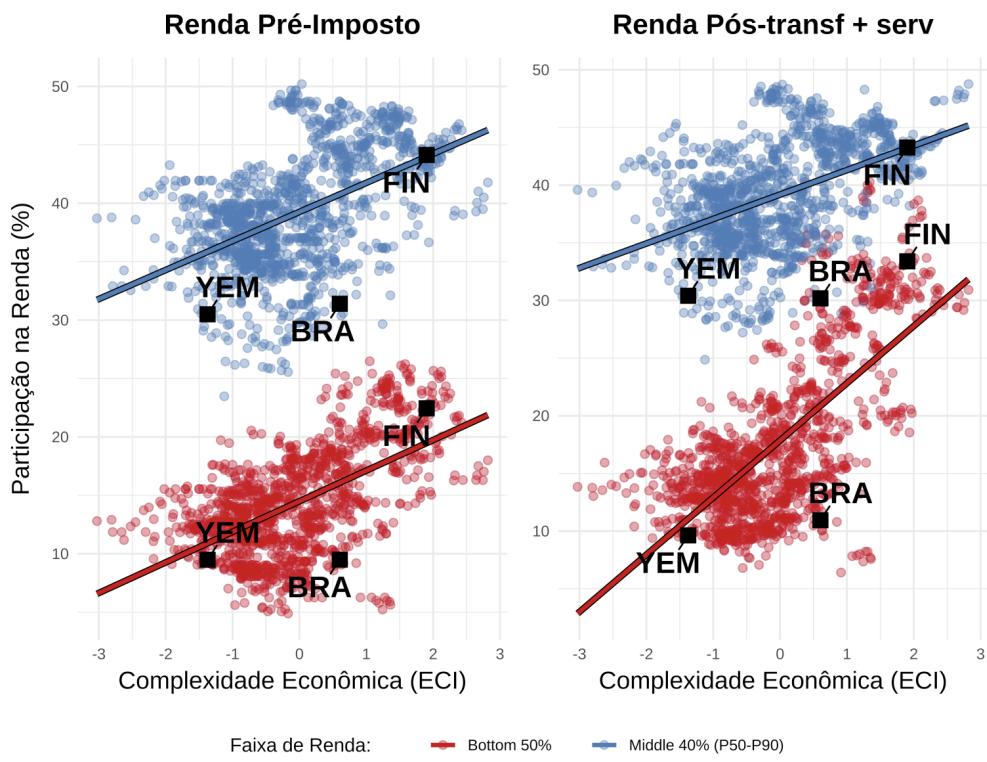


Figura 8: Correlação entre ECI e renda acumulada de diferentes faixas da distribuição: P50/M40 (acima) e Top 10%, 1%, 0.1%, 0.01%, 0.001% (abaixo)

Tabela 3: Comparação de Estimadores para o Efeito do ECI no Gini Pós-Tributação

	MQO Empilhado	Efeitos Aleatórios	Efeitos Fixos	Primeiras Diferenças
(Intercept)	0,570*** (0,024)	0,554*** (0,020)		0,047*** (0,004)
ECI	-0,083*** (0,010)	-0,060*** (0,010)	0,011 (0,011)	0,032 (0,024)
PIB pc (log)	-0,002 (0,003)	0,003 (0,003)	0,038** (0,017)	-0,011 (0,047)
Anos de Estudo	-0,005* (0,003)	0,000 (0,002)	0,007*** (0,002)	-0,039*** (0,006)
Num. Obs.	374,000	374,000	374,000	261,000
R2	0,440	0,441	0,042	0,233

^a * p < 0,1, ** p < 0,05, *** p < 0,01

^b Erros-padrão clusterizados por país em parênteses.

^c Teste de Hausman (χ^2): 50,1 (p-valor: <0,001).

diferentes estimadores (MQO empilhado, Efeitos Aleatórios, Efeitos Fixos e Primeiras Diferenças) para identificar a presença de viés de variável omitida e justificar econometricamente a escolha do estimador de efeitos fixos.

Em seguida, formalizo a especificação do modelo, incluindo controles para PIB per capita e capital humano (medido por anos médios de estudo), e discuto as hipóteses implícitas sobre heterogeneidade inobservada. Por fim, exploro os mecanismos distributivos subjacentes, estimando efeitos marginais do ECI ao longo da distribuição de renda. Esta última etapa revela um padrão da “foice”: um realinhamento no topo da pirâmide que desloca participação da renda da elite para setores médio-altos, sem ganhos relevantes para os grupos mais pobres.

4.2.1 Especificação do Modelo e Variáveis de Controle

A especificação geral do modelo é dada por:

$$Gini_{it} = \alpha + \beta \cdot ECI_{it} + \delta \cdot X_{it} + c_i + \eta_t + \varepsilon_{it} \quad (4)$$

onde i representa o país; t o ano (ou década); ECI_{it} é o Índice de Complexidade Econômica; X_{it} inclui controles para $\log(PIB\ pc)$ e anos médios de estudo da população; c_i captura efeitos fixos específicos de cada país; e η_t controla por choques comuns a todos os países no tempo. A variável dependente é o Gini pós-tributação.

A partir da especificação acima, compararei a seguir os diferentes estimadores para avaliar a robustez do efeito estimado do ECI sobre o Gini.

4.2.2 Comparação de Estimadores e Teste de Hausman

Começo a análise com modelos que não controlam, ou controlam de forma incompleta, a heterogeneidade não observada dos países. Os coeficientes estimados nas quatro especificações estão apresentados na [Tabela 3](#).

Tanto o modelo de MQO empilhado quanto o de Efeitos Aleatórios indicam um coeficiente negativo e altamente significativo para o *ECI* (respectivamente, $-0,083$ e $-0,060$). O teste de Hausman rejeita fortemente a hipótese nula de consistência do RE ($\chi^2 = 50,1$, $p < 0,001$), validando o uso do estimador de Efeitos Fixos (FE) como o mais apropriado.

Quem parasse por aqui poderia concluir que uma maior complexidade econômica está associada a menor desigualdade.

Contudo, ao aplicarmos os estimadores robustos (FE e FD), a história muda substancialmente. O coeficiente estimado para o *ECI* perde significância estatística e inverte o sinal ($0,011$ e $0,032$). A correlação negativa entre complexidade e desigualdade agregada, sugerida pelos modelos ingênuos, parece refletir viés de variável omitida.

4.2.3 O Efeito da Complexidade na Estrutura de Renda: O Padrão da Foice

Sabemos que o Índice de Gini é uma medida agregada. Por isso, sua aparente neutralidade diante do ECI pode mascarar efeitos redistributivos heterogêneos ao longo da distribuição de renda. Para investigar esse possível “efeito composição”, estimei modelos com efeitos fixos para cada faixa da distribuição.

A [Figura 9](#) mostra os coeficientes do *ECI* sobre a participação da renda para cada decil e segmento superior da distribuição. Podemos observar três padrões distintos:

- **Base da distribuição (D1–D8):** coeficientes muito próximos de zero e intervalos de confiança amplos, indicando efeito nulo ou estatisticamente insignificante. Isso sugere que os ganhos de complexidade não “gotejam” para os mais pobres; ao menos em termos de participação relativa na renda.
- **Classe média-alta (D9):** efeito positivo e estatisticamente significativo, sinalizando que esse grupo captura parte dos ganhos da transição para uma economia mais complexa.
- **Topo da distribuição (D10, Top 5%, Top 1%):** coeficientes negativos e significativos, com declínio mais acentuado entre os percentis superiores. Esse padrão revela uma perda relativa da elite econômica, possivelmente refletindo uma redistribuição “intra-topo”.

Ao contrário de uma compressão na desigualdade, esses achados indicam uma recomposição da renda no topo da pirâmide.

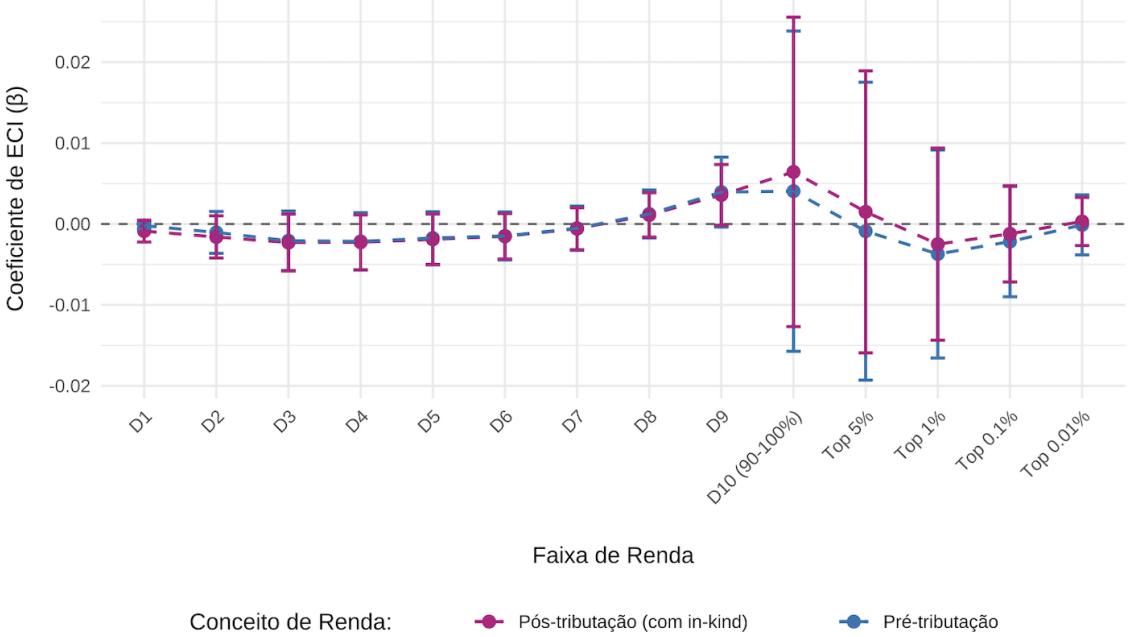


Figura 9: Efeito da Complexidade Econômica na Participação da Renda por Faixa (Modelo FE)

Nota: Pontos e intervalos de confiança de 95% mostram os coeficientes β do ECI defasado ($t-1$), estimados a partir de modelos de painel com Efeitos Fixos para cada decil e para os segmentos superiores da distribuição (1995–2015).

- *Em suma, ganhos na complexidade parecem, de maneira surpreendente, deslocar recursos da elite do topo para a camada imediatamente abaixo. Isso sem alterar substancialmente a posição dos mais pobres.*

4.3 Análise de Endogeneidade

Após termos encontrado o padrão de “foice” com estimadores de efeitos fixos, o próximo passo crítico é avaliar a robustez desse resultado. Aqui, o principal desafio empírico diz respeito à endogeneidade. Em particular, diz respeito à possibilidade de causalidade reversa, isto é, de que a estrutura distributiva influencie a complexidade produtiva, e não o contrário.

4.3.1 Mitigando Simultaneidade com Defasagens Temporais

Para mitigar esse risco, reestimei o modelo com Efeitos Fixos utilizando valores defasados do ECI como variável explicativa. Minha estratégia foi explorar o fato de que a desigualdade no ano t não pode afetar a complexidade econômica do passado (ECI_{t-k}). Assim, se o padrão identificado anteriormente persistir mesmo com defasagens temporais, isso sugere uma precedência temporal da complexidade econômica na configuração da desigualdade – ainda que não implique causalidade no sentido estrito.

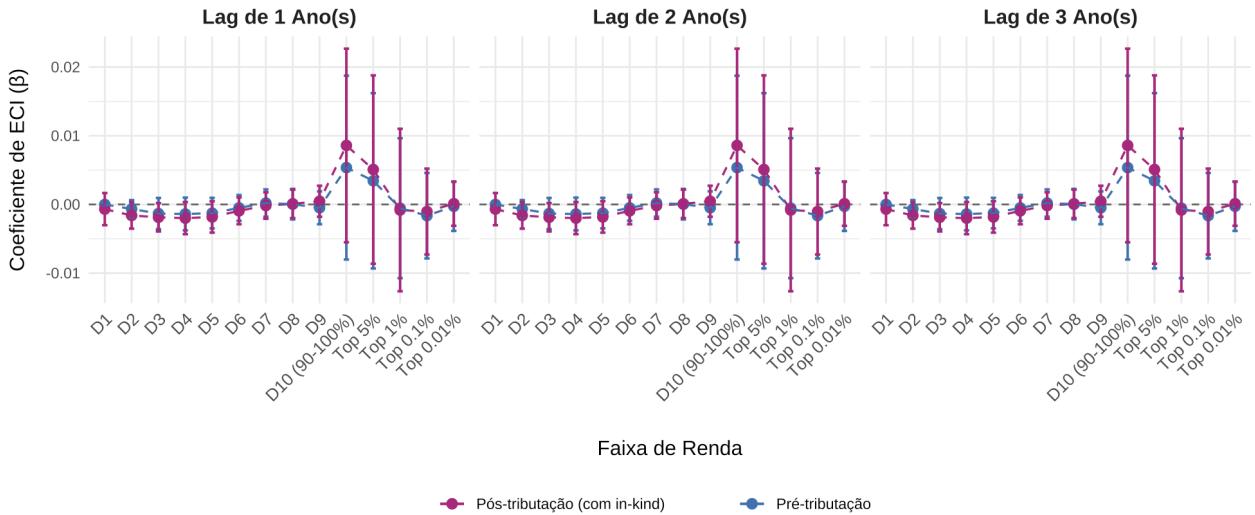


Figura 10: Teste de Robustez à Endogeneidade: Efeito do ECI Defasado na Participação da Renda

Nota: Os pontos representam os coeficientes β do ECI defasado, estimados a partir de modelos de painel anuais com Efeitos Fixos para cada faixa de renda (1995–2015). O padrão de “foice” (efeito positivo no 9º decil e negativo no topo) se mostra estável.

A Figura 10 representa os coeficientes estimados do *ECI* defasado em um, dois e três anos, para cada faixa da distribuição de renda. O resultado mais expressivo é a notável estabilidade do padrão distributivo ao longo das defasagens. O efeito positivo no 9º decil (D9) e os coeficientes negativos e decrescentes no topo da distribuição se mantêm robustos, mesmo com três anos de defasagem. Isso sugere que o padrão de recomposição no topo – o que chamei de “foice” – não é um artefato contemporâneo ou espúrio, mas um resultado persistente ao longo do tempo.

Essa estabilidade reforça a plausibilidade de uma relação dinâmica em que mudanças na complexidade produtiva precedem e influenciam a estrutura de renda. Os coeficientes completos e respectivos erros-padrão podem ser consultados na Tabela 7, no Apêndice I.

4.3.2 Tentativa de Estimação por Variáveis Instrumentais (2SLS)

Como um teste posterior, tentei aplicar uma estratégia de Variáveis Instrumentais (2SLS) ao painel decadal da OCDE. A ideia era utilizar o *ECI* defasado em duas décadas como instrumento para o *ECI* contemporâneo, sob a hipótese de que essa variável teria correlação com o regressor endógeno, mas não com o erro da equação de segunda etapa.

Contudo, a estrutura limitada da base – com poucos períodos e número modesto de países – comprometeu o desempenho do primeiro estágio. A estimação da equação instrumental com o pacote `plm` gerou o aviso de “*ajuste perfeito*” (“perfect fit”). Isso indica ausência de variabilidade residual para o instrumento dentro do painel. Além disso, a estatística-

F manualmente calculada com base no t -valor do instrumento foi numericamente explosiva ($F \approx 1,9 \times 10^{33}$), o que confirma a instabilidade do modelo. Esse resultado está detalhado na [Tentativa de Estimação por Variáveis Instrumentais do Apêndice II](#),

Na prática, isso invalida o teste de força do instrumento, impossibilita a interpretação dos coeficientes e torna o estimador 2SLS não confiável. Diante disso, minha análise de endogeneidade se concentrará nos resultados dos modelos com defasagens temporais, especialmente aqueles com efeitos fixos. Essa me parece uma abordagem mais parcimoniosa e robusta frente às limitações da base.

4.4 Análises de Robustez e Heterogeneidade

Nesta seção, testo os limites e a generalidade do principal achado – o padrão da “foice” – por meio de duas estratégias complementares. Primeiro, investigo a heterogeneidade do efeito entre países membros e não-membros da OCDE. O objetivo é avaliar se o mecanismo de realinhamento distributivo está condicionado ao nível de desenvolvimento e à presença de instituições de bem-estar. Em seguida, foco exclusivamente no grupo OCDE, onde o efeito foi mais pronunciado. Para esse grupo, testo a robustez dos resultados frente a diferentes especificações temporais e estratégias de controle. Aqui, pretendo conferir se o padrão identificado resiste a alterações no modelo e nas amostras utilizadas.

4.4.1 Heterogeneidade do Efeito: Países OCDE vs. Não-OCDE

Uma questão central é se o mecanismo de compressão distributiva no formato de "foice" é um fenômeno universal ou se é condicional ao nível de desenvolvimento. Para investigar isso, dividi a amostra entre países membros e não-membros da OCDE. A [Figura 11](#) apresenta os resultados do modelo de "foice" (com ECI defasado em 1 ano) para cada subgrupo.

O que vemos é uma heterogeneidade notável do efeito. No painel da esquerda (Países Não-OCDE), o efeito da complexidade é estatisticamente nulo em toda a distribuição. Já no painel da direita (Países OCDE), o padrão de "foice" reemergue de forma clara.

Este me parece um dos principais resultados deste exercício: o mecanismo pelo qual o ECI está relacionado com a distribuição de renda parece ser um fenômeno específico de economias avançadas. É plausível supor que nesses países, instituições e políticas de bem-estar social podem mediar a distribuição dos ganhos da complexidade produtiva. A [Tabela 5](#) no [Apêndice I](#) apresenta os resultados da regressão do Gini para ambos os grupos. Nela podemos checar o efeito nulo para o grupo de países Não-OCDE, além do trecho de código e o aviso de “ajuste perfeito” retornado pelo R.

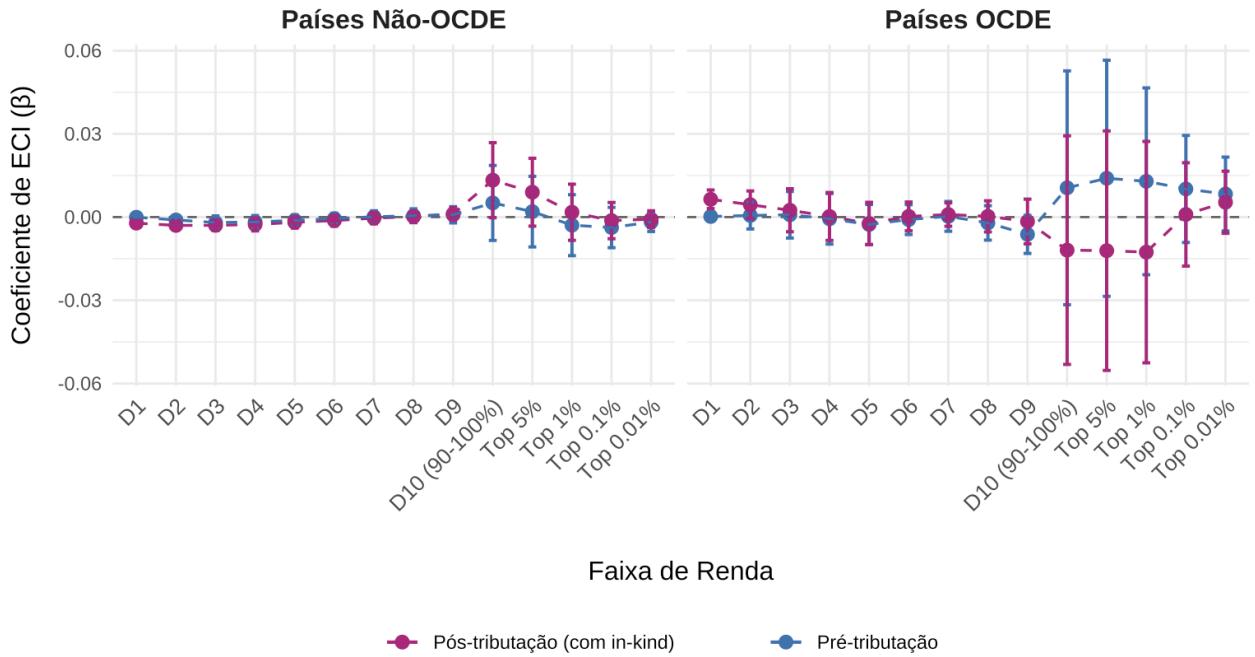


Figura 11: Teste de Heterogeneidade: Efeito do ECI por Grupo de Países

Nota: Coeficientes β do ECI defasado ($t-1$), estimados a partir de modelos de painel com Efeitos Fixos para cada faixa de renda, separadamente para países OCDE e não-OCDE. A análise cobre o período 1995–2015 e controla para PIB per capita (log) e escolaridade.

4.4.2 Robustez à Especificação do Modelo (Amostra OCDE)

Dada a descoberta de que o efeito se concentra nos países da OCDE, testo nesta seção a robustez do resultado para este subgrupo. Para isso, utilizo diferentes formas de controlar para o tempo e a dinâmica da desigualdade. A [Tabela 4](#) resume os resultados, em especial, a estabilidade do coeficiente do ECI.

Nos modelos (1) a (4), que variam da especificação base até o controle por tendências temporais e quebras estruturais, o coeficiente estimado do ECI permanece pequeno e estatisticamente não significativo. Isso reforça a conclusão de que o ECI não tem um efeito direto sobre o Gini agregado. Isso mesmo após controlarmos para tendências temporais e possíveis quebras estruturais.

A coluna (5) reporta a tentativa de estimar um modelo dinâmico. A estimação não convergiu para uma solução interpretável. Isso é consistente com as dificuldades conhecidas da aplicação de modelos dinâmicos em painéis com poucos períodos. Além disso, também reforça a necessidade de cautela ao modelar dinâmicas com este tipo de dado, assim como justifica sua exclusão da interpretação substantiva dos resultados. A [Estimação do Modelo de Painel Dinâmico](#) no [Apêndice II](#) documenta a estimação em detalhes. Inclusive, o trecho de código e o aviso de “*ajuste perfeito*” retornado pelo R.

Tabela 4: Testes de Robustez da Especificação para o Efeito do ECI no Gini (OCDE)

	(1) Base	(2) + Dummies Década	(3) + Tendência	(4) + Interação Pós-90	(5) Dinâmico
ECI	0,043 (0,032)	0,023 (0,025)	0,023 (0,025)	0,034 (0,043)	Instável ^a (-)
Num.Obs.	62.000	62.000	62.000	62.000	62.000

Notas: Todos os modelos incluem efeitos fixos. Erros-padrão clusterizados por país entre parênteses.

^a Estimação do modelo dinâmico instável devido ao viés de Nickell em painel curto.

4.5 Síntese dos Resultados Empíricos

A análise conduzida ao longo desta seção revela uma relação complexa – e condicional – entre complexidade produtiva e distribuição de renda. O principal achado foi o padrão de “foice”. Estimamos modelos com efeitos fixos por faixa da distribuição. O resultado foi que aumentos no Índice de Complexidade Econômica (ECI) estão associados a uma perda relativa de participação da elite no topo (Top 10%, Top 1%) e a ganhos significativos para o 9º decil. Mas isso sem efeitos robustos na base da pirâmide. Esse resultado desafia algumas de nossas expectativas e sugere que a complexidade econômica atuaria de forma assimétrica.

Tentei submeter esse achado a uma bateria de testes. Primeiro, mostrei que a relação negativa inicialmente observada entre ECI e Gini agregado era espúria. Ela desaparece quando controlamos para heterogeneidade inobservada por meio de estimadores de efeitos fixos e de primeiras diferenças. Em seguida, ao estimar efeitos marginais ao longo da distribuição, identifiquei um mecanismo distributivo mais sutil: uma redistribuição de renda para a classe média alta (9º decil), mas sem impacto sistemático sobre os grupos mais pobres.

Para mitigar a principal ameaça à aspirações causais dos resultados – a simultaneidade entre desigualdade e complexidade –, reestimei os modelos com defasagens temporais do ECI. O padrão da “foice” se manteve estável mesmo com lags de até três anos. Isso fortalece a hipótese de uma relação prospectiva entre ECI e desigualdade de renda. Enfrentei limitações na tentativa de aplicar variáveis instrumentais no painel decadal. Embora metodologicamente desejável, aqui não produziu resultados confiáveis.

Uma dimensão crítica da análise é a avaliação da heterogeneidade entre países. A decomposição da amostra em membros e não-membros da OCDE revelou que o padrão distributivo associado ao ECI parece ser um fenômeno específico das economias avançadas. Nos países não-OCDE o efeito da complexidade é estatisticamente nulo em toda a distribuição. Entre os países da OCDE, a “foice” aparece com clareza. Isso sugere que o mecanismo identificado é condicionado por contextos institucionais específicos. Sistemas de bem-estar, mercados regulados e capacidade estatal de intermediação são os suspeitos de sempre.

Por fim, testei a robustez dos resultados para a amostra da OCDE a partir de múltiplas especificações: dummies de década, tendências temporais, interações estruturais e modelos dinâmicos). Todas convergiram para a mesma conclusão de que a complexidade produtiva, isoladamente, não reduz a desigualdade agregada (medida pelo Gini). Ainda assim, sua interação com o aparato institucional parece gerar efeitos redistributivos relevantes no interior da pirâmide.

Apesar da consistência dos resultados, não podemos ignorar algumas limitações importantes. A disponibilidade de dados ainda impõe restrições importantes, especialmente para países em desenvolvimento e para séries longas com defasagens. Além disso, o controle da endogeneidade é parcial. Mecanismos causais mais robustos dependeriam de estratégias instrumentais mais críveis. Ainda assim, os padrões identificados são suficientemente estáveis e sistemáticos. Ao menos para fundamentar hipóteses mais ambiciosas sobre os vínculos entre estrutura produtiva, regimes distributivos e trajetórias de desigualdade.

A seguir, retomo os principais passos do exercício e organizo os achados de forma sintética. Vou cotejar os resultados obtidos com os de Hartmann et al. (2017). Vamos nos lembrar que esses autores identificaram uma correlação negativa entre complexidade econômica e desigualdade de renda em modelos de cross-section. Isso nos permitirá qualificar as contribuições deste trabalho, e destacar o que há de novo na análise em painel apresentada aqui.

5 Conclusão e Sugestões para Pesquisas Futuras

Neste exercício, parti de uma pergunta simples, mas teoricamente densa: a complexidade produtiva contribui para reduzir a desigualdade de renda? Para respondê-la, articulei três frentes complementares: (i) uma revisão dos fundamentos metodológicos da inferência causal com dados em painel; (ii) a construção de uma base analítica internacional, combinando o *Atlas of Economic Complexity* (Atlas) com a *World Inequality Database* (WID); e (iii) a aplicação de modelos econométricos com atenção às propriedades dos estimadores sob violação de pressupostos clássicos.

O principal achado é que, uma vez controladas as características fixas dos países, o Índice de Complexidade Econômica (ECI) não apresentou efeito estatisticamente significativo sobre o Gini agregado. A correlação negativa sugerida por modelos em *cross-section* e por modelos de Efeitos Aleatórios desaparece quando utilizamos estimadores como os de Efeitos Fixos e Primeiras Diferenças. O que parecia um efeito distributivo da complexidade produtiva se mostrou, ao menos neste painel, como um caso típico de viés por variável omitida.

Ao desagregar a análise de painel por faixas da distribuição, contudo, emergiu um padrão sistemático: aumentos no ECI estão associados a perdas relativas no topo da pirâmide (top 1%, 0.1%, 0.01%) e ganhos para o 9º decil. Mas tudo isso sem efeitos robustos na base. Essa dinâmica “intra-topo” se manteve de pé diante de defasagens temporais e testes de especificação. Não pude, contudo, estimar modelos instrumentais minimamente confiáveis neste painel curto. Eles poderiam tornar esses resultados mais robustos.

Também me deparei com a heterogeneidade dos resultados entre países, o que reforçou uma perspectiva institucional sobre a desigualdade e a estrutura produtiva. O padrão da foice só emergiu entre os membros da OCDE. Nos países não-membros, os coeficientes estimados para o ECI foram indistinguíveis de zero. Isso sugere que a complexidade, por si só, não gera redistribuição. Seus efeitos distributivos podem depender de estruturas institucionais capazes de intermediar, canalizar ou moderar os ganhos produtivos. Falamos aqui de sistemas de bem-estar, regulações de mercado ou regimes tributários progressivos.

Do ponto de vista metodológico, tentei ilustrar neste exercício os ganhos de uma abordagem atenta às propriedades dos estimadores. Discuti os efeitos de viés, variância, autocorrelação e dependência seccional, com ênfase nos limites de inferência a partir de modelos mal especificados. A ênfase na estrutura de painel permitiu qualificar e, em parte, contradizer os achados de Hartmann et al. (2017). Ao menos por enquanto, aquilo que eles encontraram não resistiu à introdução de controles adequados.

Preciso, no entanto, reconhecer as limitações do estudo. A ausência de um instrumento plausível e forte para o ECI limita o alcance causal dos achados. Esse é o caso mesmo com o

uso que fiz de defasagens. A base de dados apresentou desequilíbrios temporais relevantes, e a decisão por médias decadais resultou na perda de granularidade temporal. Além disso, optei por modelos estáticos com especificações lineares. Embora mais transparentes, esses modelos podem não capturar toda a complexidade das interações entre produção e distribuição.

Essas limitações abrem espaço para uma série de extensões promissoras dessa investigação. Entre elas, destaco:

- Investigar não-linearidades e interações entre complexidade, educação, e instituições distributivas;
- Explorar mecanismos microfundamentados que conectem mudanças na estrutura produtiva a padrões ocupacionais e fiscais;
- Aplicar modelos com maior sofisticação para dependência seccional (como o CCE-MG de Pesaran, 2006) e explorar estimação dinâmica caso haja maior disponibilidade de dados longitudinais;
- Buscar instrumentos exógenos plausíveis – como choques exógenos de complexidade induzidos por acordos comerciais ou descobertas tecnológicas – para avançar na identificação causal.

Uma última observação. Estes resultados não devem ser tomados como a última palavra sobre a relação entre complexidade econômica e desigualdade de renda. Tampouco penso que é razoável tomá-los como uma razão para abandonarmos os achados de Hartmann et al. (2017). Se podemos extrair alguma conclusão mais séria sobre eles, é que vale aprofundar essa investigação com dados mais sensíveis às dinâmicas internas da distribuição, e com um emprego de outras estratégias econométricas. Não há, nesse campo, uma metodologia única que seja superior a todas as outras. Práticas de triangulação – combinando diferentes especificações, janelas temporais e técnicas de estimação – são, me parece, valiosas para identificarmos regularidades mais robustas e qualificadas.

Apêndice I: Dados Suplementares

Ia. Lista de Países na Amostra Analítica

A amostra final utilizada nas estimativas de painel é composta pelos seguintes 115 países:

ALB, ARE, ARG, ARM, GTM, HKG, HND, HRV, NPL, NZL, OMN, PAK, AUS, AUT, AZE, BDI, BEL, HTI, HUN, IDN, IND, IRL, PAN, PER, PHL, POL, PRT, BEN, BFA, BGD, BGR, IRN, ISR, ITA, JAM, JOR, PRY, PSE, ROU, RUS, SAU, BOL, BRA, CAF, CAN, JPN, KAZ, KGZ, KHM, SEN, SGP, SLV, SRB, SVK, CHE, CHL, CHN, CIV, KOR, KWT, LBN, LBY, SVN, SWE, SYR, TCD, COG, COL, CRI, CUB, CZE, LKA, LSO, LTU, LVA, MDA, THA, TJK, TKM, TUN, DEU, DNK, DOM, DZA, MEX, MKD, MLI, MMR, TUR, TZA, UGA, UKR, ECU, EGY, ESP, ETH, FIN, MNG, MOZ, MRT, MWI, URY, USA, VNM, ZAF, FRA, GEO, GIN, GRC, MYS, NGA, NLD, NOR, ZMB, ZWE.

Ib. Resultados Detalhados das Estimações

As tabelas a seguir apresentam os resultados completos das regressões descritas no Capítulo 4. As estimativas foram obtidas com modelos de painel com efeitos fixos, controle para PIB per capita (log) e anos médios de estudo, e erros-padrão robustos clusterizados por país.

- Teste de Robustez à Endogeneidade
- Testes de Robustez da Especificação (Amostra OCDE)
- Testes de Heterogeneidade: OCDE vs. Não-OCDE

Teste de Robustez à Endogeneidade

Tabela 5: Efeitos Defasados do ECI sobre Desigualdade: Modelos de Painel por Faixa de Renda

Conceito de Renda	Lag 1 Ano(s)	Lag 2 Ano(s)	Lag 3 Ano(s)
D1			
Pré-tributação	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Pós-tributação	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
D10 (90-100%)			
Pré-tributação	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)
Pós-tributação	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)
D2			
Pré-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Pós-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
D3			
Pré-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Pós-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
D4			
Pré-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Pós-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
D5			
Pré-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Pós-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
D6			
Pré-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Pós-tributação	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
D7			
Pré-tributação	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Pós-tributação	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
D8			
Pré-tributação	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Pós-tributação	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
D9			
Pré-tributação	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)
Pós-tributação	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)
Top 0.01%			
Pré-tributação	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)
Pós-tributação	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)
Top 0.1%			
Pré-tributação	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)
Pós-tributação	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)
Top 1%			
Pré-tributação	-0.001 (0.005)	-0.001 (0.005)	-0.001 (0.005)
Pós-tributação	-0.001 (0.005)	-0.001 (0.005)	-0.001 (0.005)
Top 5%			
Pré-tributação	0.003 (0.006)	0.003 (0.006)	0.003 (0.006)
Pós-tributação	0.003 (0.006)	0.003 (0.006)	0.003 (0.006)

Testes de Robustez da Especificação (Amostra OCDE)

Tabela 6: Testes de Robustez da Especificação para o Efeito do ECI no Gini (OCDE)

	(1) Base	(2) + Dummies Década	(3) + Tendência	(4) + Interação Pós-90	(5) Dinâmico
avg_eci	0.043 (0.032)	0.023 (0.025)	0.023 (0.025)	0.034 (0.043)	Instável ^a (-)
log_gdp	0.036 (0.052)	0.015 (0.057)	0.015 (0.057)	0.031 (0.055)	Instável ^a (-)
schooling	0.004 (0.006)	-0.006 (0.005)	-0.006 (0.005)	0.003 (0.005)	Instável ^a (-)
factor(decade)1990		0.013 (0.011)			
factor(decade)2000		0.038 ⁺ (0.021)			
tendencia			0.001 (0.001)		
I(tendencia^2)			0.000 (0.000)		
pos_1990				0.004 (0.021)	
avg_eci × pos_1990				0.003 (0.015)	
lag_gini1					Instável ^a (-)
Num.Obs.	62	62	62	62	62

Notas: + p < 0,1, * p < 0,05, ** p < 0,01, *** p < 0,001. Todos os modelos incluem Efeitos Fixos e erros-padrão clusterizados por país.

^a A estimativa do modelo dinâmico não convergiu para uma solução interpretável devido ao severo viés de Nickell em um painel com poucos períodos de tempo, como indicado por um aviso de "ajuste perfeito" durante a estimativa (ver [Apêndice II](#)).

Testes de Heterogeneidade: OCDE vs. Não-OCDE

Tabela 7: Testes de Heterogeneidade: OCDE vs. Não-OCDE

	Gini Pré (OCDE)	Gini Pós (OCDE)	Gini Pré (Não-OCDE)	Gini Pós (Não-OCDE)
<i>ECI</i>	0,070*	0,043	-0,002	0,006
	(0,040)	(0,032)	(0,010)	(0,011)
<i>log_gdp</i>	0,019	0,036	0,041	0,026
	(0,051)	(0,052)	(0,025)	(0,025)
<i>schooling</i>	0,012**	0,004	0,009***	0,006*
	(0,005)	(0,006)	(0,003)	(0,003)
<i>Num. Obs.</i>	62,000	62,000	125,000	125,000

Notas: * p<0,1, ** p<0,05, *** p<0,01.

Todos os modelos incluem Efeitos Fixos e controles para PIBpc e escolaridade.

Apêndice II: Código em R e Replicabilidade Computacional

Neste segundo *Apêndice*, documento trechos selecionados de código em R utilizados para gerar os resultados empíricos reportados neste exercício. Aqui, a ênfase recai nos modelos que apresentaram instabilidades ou limitações técnicas. O objetivo é garantir transparência metodológica, mesmo nos casos em que os resultados não foram reportados no corpo do texto principal.

O conteúdo está estruturado em dois blocos principais:

- (i) **Estimação por Variáveis Instrumentais (2SLS):** tentativa de uso do ECI defasado como instrumento no painel decadal da OCDE. A subseção apresenta o código de estimação do primeiro estágio com o pacote `plm`, o aviso de ajuste perfeito retornado pelo R e o valor inválido da F-estatística, que inviabilizou a interpretação causal.
- (ii) **Modelo Dinâmico com Efeitos Fixos:** tentativa de inclusão da variável dependente defasada entre os regressores (modelo de painel dinâmico). Apresento o código completo, a tabela gerada automaticamente via `modelsummary` e os artefatos do viés de Nickell em painel curto (como coeficientes extremos e ausência de variabilidade estimável).

Ambas as subseções reproduzem os comandos exatos executados e relatam as mensagens de aviso emitidas pelo ambiente estatístico. O script completo que permite reproduzir computacionalmente os resultados da análise pode ser consultado em [repositório público](#).⁵ Lá também há as instruções para execução e o registro da sessão R.

A. Tentativa de Estimação por Variáveis Instrumentais (2SLS)

Conforme mencionado na Seção 4.3.2, a tentativa de estimar um modelo 2SLS para o Gini na amostra decadal da OCDE, usando o ECI defasado em duas décadas como instrumento, resultou em instabilidade no primeiro estágio. O código abaixo demonstra a estimação do primeiro estágio com o pacote `plm`, o aviso de "ajuste perfeito" retornado pelo R, e a F-estatística resultante, que é numericamente inválida.

```
> # Estimação do primeiro estágio usando plm
> first_stage_plm <- plm(avg_eci ~ lag_eci2 + log_gdp + schooling,
+   data = p_dados_iv, model = "within")
Warning message:
In summary.lm(object, ...) :
```

⁵Repositório: github.com/samuelmaibr/avaliacao-2-econometria-2025

```

essentially perfect fit: summary may be unreliable

> # Cálculo da F-estatística a partir do t-valor ao quadrado
> f_stat_manual <- (summary(first_stage_plm)$coefficients["lag_eci2", "t-
  value"])^2
> print(f_stat_manual)
[1] 1.948942e+33

```

A F-estatística astronomicamente alta é um sintoma direto do ajuste perfeito e invalida o teste de instrumento fraco. A tabela de resultados gerada a partir desta estimação (não reportada aqui) mostrou coeficientes idênticos para os modelos FE e 2SLS, confirmando que a estimação instrumental não foi bem-sucedida.

B. Estimação do Modelo de Painel Dinâmico

A [Tabela 4](#) (Coluna 5) no corpo do texto omitiu os coeficientes do modelo dinâmico por instabilidade. O trecho de código abaixo apresenta a estimação daquele modelo da tabela de resultados completa gerada pelo `modelsummary`, onde o problema pode ser observado diretamente. O aviso de "ajuste perfeito" também foi acionado nesta estimação.

```

> # Estimação do Modelo (5) da Tabela de Robustez
> m5 <- plm(gini_value ~ lag(gini_value, 1) + avg_eci + log_gdp +
  schooling,
+           data = p_dados_robustez, model = "within")
Warning message:
In summary.lm(object, ...) :
  essentially perfect fit: summary may be unreliable

> # Geração da tabela de resultados
> modelsummary(
+   list("(5) ▾Dinâmico" = m5),
+   output = "latex", stars = TRUE, vcov = ~country_id
+ )
\begin{table}
\centering
\begin{tabular}{t}{lc}
\toprule
& (5) Dinâmico\\
\midrule
lag(gini\_value, 1) & 1.000***\\
& (0.000)\\
avg\_eci & 0.000\\
\bottomrule
\end{tabular}

```

```
& (0.000)\\
log\_gdp & 0.000\\
& (0.000)\\
schooling & 0.000\\
& (0.000)\\
\midrule
Num.Obs. & 62\\
\bottomrule
\multicolumn{2}{l}{\textsuperscript{***}: p < 0.01, \textsuperscript{**}: \\
    p < 0.05, \textsuperscript{*}: p < 0.1}
\end{tabular}
\end{table}
```

O resultado, com um coeficiente de 1.000 para a variável defasada e 0.000 para todos os outros regressores, é um artefato da falha de estimação e ilustra o severo viés de Nickell em painéis curtos, justificando a decisão de não interpretar estes coeficientes no corpo principal do trabalho.

Apêndice III: Enunciado da Avaliação

Exercício 2: Endogeneidade, Variáveis Instrumentais e Modelos para Dados em Painel

Utilizando uma base de dados própria ou alguma das bases disponibilizadas, avançar em aplicações empíricas, como etapa intermediária para o trabalho final, contendo:

1. **Síntese dos aspectos conceituais-teóricos** relativos aos tópicos abarcados (como para um capítulo metodológico de projeto ou dissertação/tese);
2. **Aplicação e interpretação comparativa dos resultados:**
 - a. Comparações podem ser feitas entre pontos no tempo, unidades geográficas ou grupos sociodemográficos;
 - b. Especificações alternativas de modelos;
 - c. Testes para detecção de problemas e correções.

Tópicos:

- Endogeneidade: Estimadores de Variáveis Instrumentais e Mínimos Quadrados em 2 Estágios
- Modelos para Dados em Painel: Pooled Cross-Sections, Diferenças-em-diferenças, Estimação de Efeitos Fixos e Aleatórios
- Variáveis Instrumentais para Dados em Painel

Observações:

1. É obrigatório o agendamento de reuniões com a professora para discussão dos resultados da Avaliação 1 e desenvolvimento da Avaliação 2 e do Trabalho Final. Esta reunião deverá ocorrer nas semanas de 26 a 30 de maio ou 2 a 6 de junho, podendo ser presencial ou remota, e deverá ter a participação de todos os membros do grupo.
2. No mínimo, seguir o roteiro básico disponível em Heiss (2020) para aplicação no R.

Referências

- American Economic Association. (2020). AEA Data and Code Availability Policy [Acesso em 20 maio 2025]. (Ver p. 5).
- Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2), 277–297 (ver p. 19).
- Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research. (Ver p. 16).
- Blanchet, T., Chancel, L., Flores, I., & Morgan, M. (2024). *Distributional National Accounts Guidelines: Methods and Concepts used in the World Inequality Database* (rel. técn.). World Inequality Lab. (Ver p. 22).
- Blundell, R., & Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87(1), 115–143 (ver p. 19).
- Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2), 141–162 (ver p. 19).
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables Is Weak. *Journal of the American Statistical Association*, 90(430), 443–450 (ver p. 15).
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press. (Ver p. 15).
- Center for International Development at Harvard University. (n.d.). The Atlas of Economic Complexity. (Ver p. 21).
- Davidson, R., & MacKinnon, J. G. (2009). *Econometric Theory and Methods*. Oxford University Press. (Ver p. 15).
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4), 549–560 (ver pp. 10, 13, 25).
- Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., & Hidalgo, C. A. (2017). Linking Economic Complexity, Institutions, and Income Inequality. *World Development*, 93, 75–93 (ver pp. 1, 5–7, 21, 37–39).
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251–1271 (ver p. 18).
- Hausmann, R., Hwang, J., & Rodrik, D. (2006). What you export matters. *Journal of Economic Growth*, 12(1), 1–25 (ver p. 6).

-
- Heiss, F. (2020). *Using R for Introductory Econometrics* (2^a ed.). CreateSpace Independent Publishing Platform. (Ver p. 13).
- Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575 (ver pp. 6, 13).
- Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The Product Space Conditions the Development of Nations. *Science*, 317(5837), 482–487 (ver p. 6).
- Kennedy, P. (2008). *A Guide to Econometrics* (6^a ed.). Wiley-Blackwell. (Ver p. 12).
- National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and Replicability in Science*. National Academies Press. (Ver p. 5).
- Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6), 1417–1426 (ver p. 19).
- Pesaran, M. H. (2004). *General Diagnostic Tests for Cross-Section Dependence in Panels* (Cambridge Working Paper in Economics N. 0435). University of Cambridge, Faculty of Economics. (Ver p. 13).
- Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure [First published: 19 June 2006]. *Econometrica*, 74(4), 967–1012 (ver p. 39).
- Piketty, T., Saez, E., & Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics*, 133(2), 553–609 (ver pp. 8, 22).
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. (Ver p. 5).
- Roodman, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1), 135–158 (ver p. 19).
- Rosa, M., Flores, I., & Morgan, M. (2024). More unequal or not as rich? Revisiting the Latin American exception. *World Development*, 184, 106737 (ver pp. 8, 9, 28).
- Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557–586 (ver p. 15).
- Verbeek, M. (2017). *A Guide to Modern Econometrics* (5th). John Wiley & Sons. (Ver p. 11).
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838 (ver pp. 10, 13).
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross-Section and Panel Data* (2^a ed.). MIT Press. (Ver pp. 12, 15, 17, 25).