

**Hillary de Oliveira dos Santos**

**Atribuição de pesos em indicadores  
compostos: metodologias baseadas em dados  
e uma aplicação ao Brasil**

Niterói - RJ, Brasil

14 de junho de 2023

**Hillary de Oliveira dos Santos**

**Atribuição de pesos em indicadores  
compostos: metodologias baseadas  
em dados e uma aplicação ao Brasil**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Wilson Calmon Almeida dos Santos

Niterói - RJ, Brasil

14 de junho de 2023

**Hillary de Oliveira dos Santos**

**Atribuição de pesos em indicadores  
compostos: metodologias baseadas em dados  
e uma aplicação ao Brasil**

Monografia de Projeto Final de Graduação sob o título "*Atribuição de pesos em indicadores compostos: metodologias baseadas em dados e uma aplicação ao Brasil*", defendida por Hillary de Oliveira dos Santos e aprovada em 14 de junho de 2023, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Wilson Calmon Almeida dos Santos**  
Departamento de Estatística – UFF

---

**Profa. Márcia Marques de Carvalho Brasil**  
Departamento de Estatística – UFF

---

**Prof. Jorge Nogueira de Paiva Britto**  
Departamento de Economia – UFF

Niterói, 14 de junho de 2023

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

S237a Santos, Hillary de Oliveira dos  
Atribuição de pesos em indicadores compostos: metodologias baseadas em dados e uma aplicação ao Brasil / Hillary de Oliveira dos Santos. - 2023.  
51 f.

Orientador: Wilson Calmon Almeida Dos Santos.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2023.

1. Indicadores compostos. 2. Análise de Componentes Principais. 3. Análise Envoltória de Dados. 4. Avaliação de governos. 5. Produção intelectual. I. Calmon Almeida Dos Santos, Wilson, orientador. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD - XXX

# Resumo

O uso de indicadores compostos vem sendo cada vez mais comum para avaliar e orientar a condução de políticas governamentais. Tais indicadores reconhecidamente sintetizam algum fenômeno complexo, muitas vezes descritos por uma coleção de indicadores simples que representam diferentes dimensões de tal fenômeno. Este é o caso, por exemplo, do Índice de Desenvolvimento Humano (IDH). Há várias etapas envolvidas na criação de um indicador composto, sendo uma das mais relevantes a escolha da importância relativa (pesos) dos diferentes indicadores a serem combinados. O presente trabalho compara duas metodologias de atribuição de pesos baseadas em dados: Análise de Componentes Principais e Análise Envoltória de Dados. Com base nestes métodos, dois indicadores compostos serão criados e comparados com as propostas de governo dos mandatos entre 1997 a 2022 afim de avaliar a qualidade das políticas econômicas adotadas.

Palavras-chave: Indicadores compostos. Análise de Componentes Principais. Análise Envoltória de Dados. Avaliação de governos.

# Dedicatória

Dedico esta monografia à minha mãe, que sem dúvidas está olhando por mim com muito orgulho. Obrigada por ter sido exemplo de luta e coragem. Celebro mais essa vitória contigo no meu coração

# Agradecimentos

Primeiro, sou imensamente grata a Deus por todas as bênçãos em minha jornada acadêmica. Pai, obrigada por acreditar mais do que eu mesma, estar perto e me apoiar sempre. Tio Rogério, obrigada por mobiliar minha cozinha e ajudar meu pai no trabalho. Sem o apoio de vocês nada seria possível. Su, Tia Marilda, Vó Raquel e Tia Vera obrigada por todas as orações e incentivos. Maristela, obrigada por cada conselho, colo e cházinho. Mila por escolher ser minha irmã. Iago, pela companhia durante quase toda jornada. Aos amigos que se tornaram família em Niterói (Bia, Brendha, Caio, Érica, Priscila e Paola). A Ingrid por ser minha parceira da infância a graduação. Ao meu grupinho (Bia, Karol e Vitória) por tentarem constantemente me lembrar que existia vida além da faculdade. A equipe do IBRE - FGV (Taíse, Julio, Luciano, João e Pedro) por compartilharem suas experiências comigo. Ao Matheus Teles por abrir portas e acreditar no meu potencial. Aos professores do GET que são minha inspiração: Ana Maria Farias, Jessica Kubrusly, Rafael Erbisti, Eduardo Ferioli e Yasmin Cavaliere. Ao meu orientador Wilson, que tornou a criação desse trabalho leve e me deu todo apoio necessário. A minha banca e professora Márcia pela sua doçura e amor pelos indicadores. Ao professor Jorge Britto pela disponibilidade em contribuir com o trabalho. E a todos que se interessarem por este estudo.

# Sumário

## Lista de Figuras

Lista de Tabelas	p. 1
<b>1 Introdução</b>	p. 2
1.1 Motivação . . . . .	p. 2
1.2 Revisão Bibliográfica . . . . .	p. 3
1.3 Objetivos . . . . .	p. 4
1.4 Organização . . . . .	p. 4
<b>2 Materiais e Métodos</b>	p. 5
2.1 OCDE e a construção de indicadores compostos . . . . .	p. 5
2.2 Análise de Componentes Principais (PCA) . . . . .	p. 8
2.2.1 Definição PCA . . . . .	p. 8
2.2.2 Padronização e PCA . . . . .	p. 10
2.2.3 PCA amostral ou empírica . . . . .	p. 11
2.3 Análise Envoltória de Dados (DEA) . . . . .	p. 12
2.3.1 DEA com retornos constantes de escala (CRS) . . . . .	p. 14
2.4 Ponderação nos Indicadores Compostos . . . . .	p. 17
2.4.1 Pesos via PCA . . . . .	p. 17
2.4.2 Pesos via DEA: Abordagem BoD . . . . .	p. 18
2.5 Banco de Dados . . . . .	p. 19

*Sumário*

<b>3 Resultados</b>	p. 22
3.1 Dois Indicadores Compostos para Economia . . . . .	p. 22
3.2 Desempenho Econômico no Brasil: 1997-2022 . . . . .	p. 30
<b>4 Conclusão</b>	p. 39
<b>Referências</b>	p. 40

# Lista de Figuras

1	scree plot ilustrativo feito pela autora . . . . .	p. 12
2	Ilustração retirada do livro do Coeoper et al, 2007, p.3 . . . . .	p. 13
3	Ilustração retirada do livro do Coeoper et al, 2007, pp.8-9 . . . . .	p. 16
4	Evolução dos indicadores simples (%) . . . . .	p. 23
5	Evolução dos indicadores simples transformados . . . . .	p. 25
6	Scree plot das variâncias estimadas . . . . .	p. 26
7	Importância relativa das componentes principais iniciais . . . . .	p. 26
8	ICs obtidos via PCA e via média aritmética simples . . . . .	p. 28
9	IC obtido via DEA e via média aritmética simples . . . . .	p. 28
10	Evolução anual dos ICs (apenas IC via DEA projetado no eixo da direita)	p. 31
11	Indicador composto por mandado (médias ao longo dos anos) . . . . .	p. 31
12	Autocorrelação entre os ICs . . . . .	p. 32
13	Crescimento percentual em relação à mandato anterior . . . . .	p. 33
14	Evolução dos Pesos DEA (por ano e mandato) . . . . .	p. 33

# Lista de Tabelas

1	Coefficientes das componentes principais . . . . .	p. 27
2	Correlação entre ICs . . . . .	p. 29
3	ICs por mandato . . . . .	p. 29
4	Correlação entre ICs por mandato . . . . .	p. 29

# 1 Introdução

## 1.1 Motivação

O Índice de Desenvolvimento Humano (IDH) tem sido empregado para quantificar o desenvolvimento dos países, sendo um dos principais exemplos de indicador composto (IC), pois consiste em um resumo univariado de três indicadores simples: a expectativa de vida ao nascer, a escolaridade e a renda per capita.

A construção de um IC deve ser baseada em uma análise teórica que permita, dentre outras coisas, selecionar indicadores simples relevantes e atribuir importâncias ou pesos aos mesmos. Neste sentido, uma metodologia útil é a análise de componentes principais, que gera combinações lineares não correlacionadas dos indicadores simples; a análise en-voltória de dados, outro método a ser considerado, permite que pesos variáveis sejam definidos para cada observação de modo que o valor do IC reflète em uma distância relativa para a melhor combinação possível (dos indicadores simples). Este último método, aliás, tem se mostrado alternativa metodológica interessante em tempos mais recentes pois, em especial, não exige normalização prévia dos indicadores simples.

O IC construído neste trabalho deseja medir a performance econômica do Brasil de 1997 a 2022. Para isso é preciso entender como governos com propostas tão diferentes: FHC Cardoso (1994), Lula Silva (2002) e Silva (2006) e Bolsonaro Bolsonaro (2018) contribuíram para o país. Ou seja, entender suas performances e quais foram os mandatos mais eficientes. Cabe ressaltar que não foi encontrado nem um outro trabalho ou indicador com a mesma abordagem.

## 1.2 Revisão Bibliográfica

Em decorrência da popularidade que os indicadores compostos - como o Índice de Desenvolvimento Humano (IDH) - alcançaram nos últimos tempos, surge o interesse em discutir as melhores abordagens metodológicas para sua construção. Nesse sentido, Boysen (2002) é um dos trabalhos pioneiros no campo, explicando métodos baseados em percepções subjetivas (consulta de especialistas e percepção do analista) ou em abordagens empíricas (análise de componentes principais). Em Union e Commission (2008), também são listados alguns métodos que resumem informações de opiniões de especialistas ou do público em geral (como, por exemplo: processo de alocação de orçamento e análise conjunta), bem como métodos empíricos - dentre os quais destacamos análise de componentes principais, análise fatorial, análise envoltória de dados e modelo de componentes não observáveis. Listas e classificações similares são encontradas em Greco et al. (2019) ou (DECANCQ; LUGO, 2013). Este último, por sua vez, subdivide os tipos de ponderação em três categorias: Orientada a dados (Pesos baseados em frequência, Pesos estatísticos e Pesos mais favoráveis), Normativos (Pesos iguais ou arbitrários, Pesos da opinião de especialistas, Pesos com base no preço) e Híbridos (Pesos preferenciais declarados e Pesos hedônicos).

De acordo com Ram (1982), em virtude de sua a fácil aplicabilidade, a análise de componentes principais pode ser empregada para lidar com a eventual sobreposição de alguns dos indicadores simples e que, por vezes, apresentam elevada correlação. Slottje (1991), porém, questiona, mediante análise de sensibilidade, a variabilidade que tal método proporciona quando variamos o número de componentes escolhidas - em especial, isto promove uma alta variabilidade dos rankings associados aos diferentes indicadores compostos obtidos. Mazziotta e Pareto (2019) destaca ainda que ao resumir dados a determinado número de componentes, pode haver uma perda de informação considerável.

Os métodos anteriores usualmente pressupõe um sistema fixo de pesos e uma normalização prévia dos indicadores simples (CHERCHYE et al., 2007). Como discutido em Rogge (2018), a análise envoltória de dados, por sua vez, não possui tais limitações. No estudo conduzido em tal trabalho, as unidades observacionais correspondem a países e atribui-se um sistema de pesos diferentes para cada país que podem ser compreendidos como pesos ideais para atender a demandas/objetivos específicos destes. Porém, Fusco (2015) ressalta que, em tal abordagem, a atribuição dos pesos é bastante sensível à presença de outliers.

## 1.3 Objetivos

O presente trabalho tem como objetivo geral discutir o uso de importantes metodologias de atribuição de pesos (ponderação) em indicadores compostos, sendo os objetivos específicos:

1. Apresentar as principais diretrizes para criação de um indicador composto e dois dos principais métodos baseados em dados: análise de componentes principais e análise envoltória de dados.

2. Ilustrar os dois procedimentos propostos, criando um indicador composto associado a cada um deles, a partir de dados reais dos indicadores: taxa de crescimento do **PIB**, taxa de **juros**, taxa de **inflação**, taxa de **desemprego**, **dívida pública** em relação ao PIB, saldo da **balança comercial** em relação ao PIB, saldo da **conta corrente** em relação ao PIB e taxa de **risco país**.

3. Avaliar a performance dos mandatos comparando os programas de governo com os resultados obtidos pelos pesos da análise envoltória de Dados.

## 1.4 Organização

O Capítulo 2 apresenta conceitos relacionados à construção de indicadores compostos, baseando-se, para isso, no manual da OCDE. Na sequência, são descritas duas técnicas que viabilizam a atribuição de pesos: análise de componentes principais e análise envoltória de dados. Dito isso, é feita uma discussão de como os pesos são definidos a partir dos dois métodos. Encerrando esse capítulo tem-se o banco de dados a ser estudado. No Capítulo 3, são apresentados os resultados da aplicação dos dois métodos na construção de um indicador composto. Em seguida, são discutidas as propostas de governo e feitas as considerações para saber a qualidade do Governo Federal do Brasil ao longo do tempo. O Capítulo 4 contém as considerações finais.

## 2 Materiais e Métodos

### 2.1 OCDE e a construção de indicadores compostos

O popular Índice de Desenvolvimento Humano (IDH) tem sido empregado para medir o desenvolvimento dos países, apresentando-se como uma medida resumo de três indicadores simples (expectativa de vida ao nascer, escolaridade e renda per capita) que medem, cada um, três diferentes dimensões vistas como fundamentais para uma sociedade (saúde, educação e padrão de vida, respectivamente).<sup>1</sup>

O IDH é um dos exemplos mais importantes de Indicador Composto (IC). Em essência, um IC é um índice unidimensional que se propõe a medir conceitos multidimensionais (no caso do IDH, por exemplo, saúde, educação e padrão de vida) e, para isso, combina informações de diferentes indicadores simples (cada qual associado a uma dimensão individual). Usualmente, a construção do IC demanda uma estruturação e análise teórica que permita selecionar e atribuir pesos aos diferentes indicadores simples.<sup>2</sup>

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) consiste em um fórum único onde os governos de diferentes países trabalham juntos para enfrentar os problemas sociais, ambientais e econômicos do mundo. Um debate foi promovido pela OCDE sobre as metodologias de construção de ICs. A OCDE defende a transparência na proposição metodológica e, com base no debate realizado, elaborou um passo-a-passo que pode embasar a formulação de novos ICs – conforme (UNION; COMMISSION, 2008). Os passos identificados pela OCDE são discutidos na sequência.

---

<sup>1</sup>Maiores detalhes em <https://hdr.undp.org/data-center/human-development-indexhdr.undp.org/data-center/human-development-index>.

<sup>2</sup>Ver <https://stats.oecd.org/glossary/detail.asp?ID=6278stats.oecd.org/glossary/detail.asp?ID=6278>.

**1. Delimitar marco teórico:** o fenômeno de interesse a ser medido deve ser definido com clareza, deixando evidente o objetivo final do IC a ser construído. A visão de onde se pretende chegar vai guiar todas as próximas etapas da metodologia de construção do IC, auxiliando na seleção de indicadores individuais, definição de pesos e avaliação final, por exemplo.

**2. Seleção dos indicadores simples e tratamento de valores ausentes:** a qualidade do IC depende, em grande parte, da qualidade dos indicadores simples escolhidos. Idealmente, deve-se considerar a relevância dos diferentes indicadores e das dimensões a eles associadas, a qualidade metodológica do próprio índice individual, bem como a acessibilidade aos dados. Há de se considerar a disponibilidade dos dados, uma vez que dimensões relevantes para os objetivos globais podem não ser bem medidas por indicadores simples. Não raro, outro problema a ser resolvido é a falta de dados, ou a presença de valores ausentes e que, por vezes, exige: (i) exclusão de unidades observacionais particulares (regiões, datas, etc) (ii) ou algum tipo de imputação.

**3. Análise multivariada:** é conveniente avaliar o conjunto de dados associados aos diversos indicadores individuais previamente selecionados e entender como eles se relacionam. A análise multivariada ajudará a entender se os indicadores individuais podem ser substituídos por uma quantidade menor de indicadores ou se carregam informações que, de algum modo, se sobrepõem.

**4. Normalização:** não raro, é necessário normalizar as unidades de medida. Diversas estratégias podem ser adotadas como, por exemplo: (i) o uso de z-scores (padronização); (ii) a adoção de uma transformação Min-Max (normaliza os indicadores para que pertençam ao intervalo  $[0, 1]$  subtraindo do dado original o valor mínimo e dividindo a diferença pela amplitude observada nos dados originais); (iii) a utilização de distâncias (ou seja, medindo a posição relativa de um determinado indicador em relação a um ponto de referência ou fronteira).

**5. Ponderação e agregação:** a ponderação consiste, tipicamente, na avaliação objetiva da contribuição que cada índice individual deve apresentar para o IC. Em muitos casos, o IC é apresentado como uma média ponderada dos indicadores simples e, sendo assim, é necessário atribuir um peso a cada índice, o que pode ser feito com a ajuda de métodos multivariados como Análise de Componentes Principais (PCA), Análise Fatorial (FA) ou Análise de Envoltória de Dados (DEA), dentre outros. Os pesos podem também ser definidos pelas opiniões de especialistas ou por meio de consultas públicas. Finalmente, apesar de boa parte dos ICs mais conhecidos corresponderem a médias pon-

deradas (agregação linear) de indicadores simples, outras estratégias de agregação podem ser consideradas, dentre as quais destacamos a agregação geométrica (associada ao uso de médias geométricas).

**6. Análise de sensibilidade:** uma etapa importante da construção do IC é a apreciação da incerteza associada ao IC que, em parte, é associada à incerteza oriunda dos indicadores simples considerados e dos pesos a eles atribuídos. A análise de sensibilidade de um IC compreende a avaliação da contribuição individual de cada índice para a incerteza global do IC. Não raro, resultados da análise de sensibilidade relatam o impacto provocado pela remoção ou exclusão de um índice individual.

**7. Desagregação:** é desejável observar as contribuições das subcomponentes de um IC (eventualmente, de forma sub-agrupada). Muitas vezes, destas análises são identificados sub-aspectos relevantes (geralmente associados a conceitos previamente conhecidos e mensuráveis) que ajudam a entender o fenômeno global estudado. Análises de correlação (e similares) podem auxiliar na quantificação da importância relativa de um índice individual ou de um sub-grupo deles para o IC.

**8. Visualização dos resultados:** a construção de um IC deve proporcionar a melhor visualização possível, de modo que os usuários de tal indicador devem conseguir entender e capturar a maior informação possível. Neste sentido, tabelas, gráficos de barras, gráficos de linhas e diagramas, podem ser convenientes para permitir uma exposição adequada dos resultados.

Na sequência (Seções 2.2-2.3) descrevemos sucintamente três métodos de análise de dados que podem ser úteis para derivação dos pesos (etapa de ponderação) dos indicadores simples na construção de um IC.

## 2.2 Análise de Componentes Principais (PCA)

### 2.2.1 Definição PCA

Em uma Análise de Componentes Principais (PCA, de *Principal Component Analysis*), o objetivo é explicar a estrutura de segundo momento (variância-covariância) de um conjunto de  $p$  variáveis  $X_1, \dots, X_p$  por meio  $k$  combinações lineares dessas variáveis, onde  $k$  é desejavelmente bem menor que  $p$ . De modo geral, a PCA é encarada como um método de “redução de dados” que, eventualmente, pode ajudar na interpretação dos mesmos.<sup>3</sup>

Seja  $\Sigma$  a matriz de variância-covariância ou, mais simplesmente, variância do vetor aleatório  $\mathbf{X} = (X_1, \dots, X_p)$ . A matriz de correlação é denotada por  $\rho$ . Dados dois vetores  $\mathbf{a} = (a_1, \dots, a_p)$  e  $\mathbf{b} = (b_1, \dots, b_p)$ , em  $\mathbb{R}^p$  – de agora em diante considerados sempre uma matriz coluna –, temos que  $\mathbf{a}^\top \mathbf{X} = a_1 X_1 + \dots + a_p X_p$  e  $\mathbf{b}^\top \mathbf{X} = b_1 X_1 + \dots + b_p X_p$  são combinações lineares das coordenadas de  $\mathbf{X}$ . Segue, das propriedades usuais da covariância, que

$$\text{Cov}(\mathbf{a}^\top \mathbf{X}, \mathbf{b}^\top \mathbf{X}) = \mathbf{a}^\top \Sigma \mathbf{b}$$

$$\text{e, em particular, vale: } \text{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \Sigma \mathbf{a}.$$

Na sequência (Definição 2.2.1), apresentamos a definição de componentes principais baseadas em (JOHNSON; WICHERN, 2007; MORRISON, 2005).

As  $p$  componentes principais, denotadas  $Y_1, Y_2, \dots, Y_p$ , são definidas de forma recursiva, via:

- A **primeira componente principal**, é a combinação linear:

$$Y_1 = a_{11}X_1 + \dots + a_{1p}X_p = \mathbf{a}_1^\top \mathbf{X}, \text{ onde}$$

$$\mathbf{a}_1 \equiv \underset{\mathbf{a} \in \mathbb{R}^k; \mathbf{a}^\top \mathbf{a} = 1}{\text{argmax}} \text{Var}(\mathbf{a}^\top \mathbf{X}).$$

---

<sup>3</sup>Quando observamos dados de um número grande de variáveis é, usualmente, difícil de interpretá-las simultaneamente. Neste sentido, é possível que se obtenha um maior entendimento dos dados efetivamente obtidos ou transformados a partir da PCA, desde que as componentes encontradas em si sejam interpretáveis (JOHNSON; WICHERN, 2007).

- Para  $i = 2, 3, \dots, p$ , a  **$i$ -ésima componente principal** é a combinação linear

$$Y_i = a_{i1}X_1 + \dots + a_{ip}X_p = \mathbf{a}_i^\top \mathbf{X}, \text{ onde}$$

$$\mathbf{a}_i \equiv \underset{\mathbf{a} \in \mathbb{R}^k; \mathbf{a}^\top \mathbf{a} = 1; \text{Cov}(\mathbf{a}^\top \mathbf{X}, \mathbf{a}_j^\top \mathbf{X}) = 0, j=1, \dots, i-1}{\text{argmax}} \text{Var}(\mathbf{a}^\top \mathbf{X}).$$

A proposição 2.2.1 apresenta uma definição das componentes principais do ponto-de-vista da álgebra linear, assim como algumas de suas propriedades. As demonstrações podem ser encontradas em Johnson, Navarro.

Digamos que  $\Sigma$  possua autovalores  $\lambda_1, \dots, \lambda_p$ , onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p-1} \geq \lambda_p \geq 0$  e que  $\mathbf{e}_i = (e_{i1}, \dots, e_{ip})$  seja o autovetor associado ao  $i$ -ésimo autovalor. Então,

- Para  $i = 1, \dots, p$ , a  **$i$ -ésima componente principal** satisfaz:

$$Y_i = e_{i1}X_1 + \dots + e_{ip}X_p = \mathbf{e}_i^\top \mathbf{X}; \quad (2.1)$$

- Vale ainda, que:

$$\text{Var}(Y_i) = \mathbf{e}_i^\top \Sigma \mathbf{e}_i = \lambda_i \text{ e } \text{Cov}(Y_i, Y_j) = \mathbf{e}_i^\top \Sigma \mathbf{e}_j = 0, \text{ se } i \neq j;$$

- A chamada **variância populacional total** é dada pelas somas das variâncias de  $X_1, \dots, X_p$  e satisfaz:

$$\sum_{i=1}^p \text{Var}(X_i) = \text{traço}(\Sigma) = \sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \lambda_i.$$

Com base no último resultado apresentado no Teorema 2.2.1, podemos utilizar, tal como feito em Morrison (2005), uma medida de poder explicativo da  $k$ -ésima componente principal (ou das primeiras  $i$  componentes principais) que apresentamos na Definição 2.2.1.

A **proporção da variação total explicada pela  $k$ -ésima componente** é dada por:

$$\frac{\lambda_k}{\lambda_1 + \dots + \lambda_p}.$$

Em especial, a **proporção da variação total explicada pelas primeiras  $i$  componentes principais** assume a forma:

$$\frac{\lambda_1}{\lambda_1 + \dots + \lambda_p} + \dots + \frac{\lambda_i}{\lambda_1 + \dots + \lambda_p} = \frac{\lambda_1 + \dots + \lambda_i}{\lambda_1 + \dots + \lambda_p}.$$

Obviamente, o poder de ajuste da  $k$ -ésima componente diminui conforme  $k$  aumenta. Mais ainda, o poder explicativo das primeiras  $i$  componentes principais está sempre entre 0 e 1, sendo que valores mais próximos de 1 implicam maior poder explicativo. Em Härdle e Simar (2008), é apontado que se uma parcela expressiva (como por exemplo, mais de 80% ou 90%) da variação total é atribuída às  $i$  componentes iniciais, então, estas  $i$  componentes poderiam substituir as  $p$  variáveis originais sem muita perda de informação.

## 2.2.2 Padronização e PCA

É válido ressaltar que as componentes principais são vulneráveis a variações de escala<sup>4</sup> e, portanto, de acordo com Johnson e Wichern (2007), ao invés de trabalhar com as variáveis originais  $X_1, \dots, X_p$ , não raro se utiliza suas versões padronizadas  $Z_1, \dots, Z_p$ , em que  $Z_i \equiv (X_i - E[X_i]) / \sqrt{Var[X_i]}$ , para  $i = 1, \dots, p$ . O objetivo, de modo geral, é atenuar o efeito de eventuais diferenças entre as unidades de medida das diferentes variáveis. Fazendo  $\mathbf{Z} = (Z_1, \dots, Z_p)$ , segue-se que sua matriz de variância é  $\rho$ , que é exatamente a matriz de correlação de  $\mathbf{X} = (X_1, \dots, X_p)$ . Podemos definir as componentes principais baseadas em  $Z_1, \dots, Z_p$ , a serem denotadas  $\tilde{Y}_1, \dots, \tilde{Y}_p$ , de modo similar ao que foi feito com as variáveis originais, (HÄRDLE; SIMAR, 2008).

Propriedades análogas às apresentadas na seção 2.2.1 são válidas para as componentes na versão padronizada. Em especial, se  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_p$  denotam os autovalores (não negativos e dispostos em ordem decrescente) de  $\rho$ , com autovetores associados  $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_p$ , então, vale que  $\tilde{Y}_i = \tilde{\mathbf{e}}_i^\top \mathbf{Z}$ . De

$$\sum_{i=1}^p Var(Z_i) = \sum_{i=1}^p Var(\tilde{Y}_i) = \sum_{i=1}^p \tilde{\lambda}_i = p,$$

segue que  $\tilde{\lambda}_k/p$  e  $(\tilde{\lambda}_1 + \dots + \tilde{\lambda}_i)/p$  medem o ajuste da  $k$ -ésima componente e das  $i$  primeiras componentes principais, respectivamente, após a padronização. As componentes principais obtidas a partir das variáveis padronizadas não guardam uma relação tão simples em relação àquelas que aparecem associadas às variáveis originais. Recomenda-se, na prática, utilizar as versões padronizadas quando as originais são medidas em escalas com amplitudes muito diferentes.

---

<sup>4</sup>A técnica de PCA é sensível a mudanças de escala. Se multiplicarmos uma variável por um escalar, obtemos diferentes autovalores e autovetores. Isso se deve ao fato de que a decomposição singular é realizada na matriz de covariância e não na matriz de correlação” (HÄRDLE; SIMAR, 2008).

### 2.2.3 PCA amostral ou empírica

Na prática, a matriz de variância  $\Sigma$  é desconhecida e, assim, versões amostrais das componentes principais são obtidas. Considere que os  $n$  vetores  $p$ -dimensionais  $\mathbf{x}_1, \dots, \mathbf{x}_n$  são realizações independentes e identicamente distribuídas (iid) do vetor aleatório  $p$ -dimensional  $\mathbf{X}$ . Sejam  $\bar{\mathbf{x}}$  o vetor média amostral e  $\mathbf{S}$  a matriz de variância amostral (ambos, estimadores de  $\mu = E[\mathbf{X}]$  e  $\Sigma = Var[\mathbf{X}]$ , respectivamente). Sejam  $\mathbf{a} = (a_1, \dots, a_p)$  e  $\mathbf{b} = (b_1, \dots, b_p)$  dois vetores em  $\mathbb{R}^p$ . Considere que os vetores aleatórios  $p$ -dimensionais  $\mathbf{X}_S$  e  $\mathbf{X}'_S$  possuem distribuição empírica baseada nos vetores  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . O valor esperado associado, correspondente à média amostral, é da forma  $\frac{1}{n} \sum_{j=1}^n \mathbf{a}^\top \mathbf{x}_j = \mathbf{a}^\top \bar{\mathbf{x}}$ . Por outro lado, a covariância entre  $\mathbf{a}^\top \mathbf{X}_S$  e  $\mathbf{b}^\top \mathbf{X}'_S$  é dada por  $\mathbf{a}^\top \mathbf{S} \mathbf{b}$  (covariância amostral) e, em especial, a variância de  $\mathbf{a}^\top \mathbf{X}_S$  é dada por  $\mathbf{a}^\top \mathbf{S} \mathbf{a}$  (variância amostral).

Denotaremos por  $\hat{y}_1, \dots, \hat{y}_p$  as chamadas **componentes principais amostrais** que são definidas do mesmo modo que na seção 2.2.1, mas com base na  $\mathbf{S}$  ao invés de  $\Sigma$ . Sejam,  $\hat{\lambda}_1, \dots, \hat{\lambda}_p$ , autovalores não negativos de  $\mathbf{S}$  cujos autovetores associados são, respectivamente,  $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_p$ . Então, a  $i$ -ésima **componente principal amostral** satisfaz  $\hat{y}_i = \mathbf{e}_i^\top \mathbf{X}_S$ , cuja variância amostral é  $\hat{\lambda}_i$ . Em especial, a covariância amostral entre  $\hat{y}_i$  e  $\hat{y}_j$  é nula. A qualidade da  $k$ -ésima componente amostral é dada por  $\hat{\lambda}_k / (\hat{\lambda}_1 + \dots + \hat{\lambda}_p)$ , enquanto o ajuste das primeiras  $i$  componentes amostrais é dado por  $(\hat{\lambda}_1 + \dots + \hat{\lambda}_i) / (\hat{\lambda}_1 + \dots + \hat{\lambda}_p)$ . Cabe destacar ainda que, não raro, trabalha-se com a versão padronizada, na qual basicamente as componentes principais são definidas com base na matriz de correlação amostral ( $\mathbf{R}$ , estimador de  $\rho$ ).

Nas aplicações, como discutido em Morrison (2005), para se definir o número de componentes principais (relevantes) utiliza-se comumente um gráfico, chamado **scree plot**, que exibe os tamanhos relativos (medidas de ajuste) dos autovalores  $\hat{\lambda}_1, \dots, \hat{\lambda}_p$  em suas respectivas ordens. O total ideal de componentes relevantes é, assim, identificado informalmente como o ponto de máxima curvatura (joelho/cotovelo/elbow), que é exatamente o ponto a partir do qual a inclusão de uma componente adicional gera poucos ganhos em termos acumulados (como ocorre a partir da terceira componente no gráfico ilustrativo exibido na Figura 1).

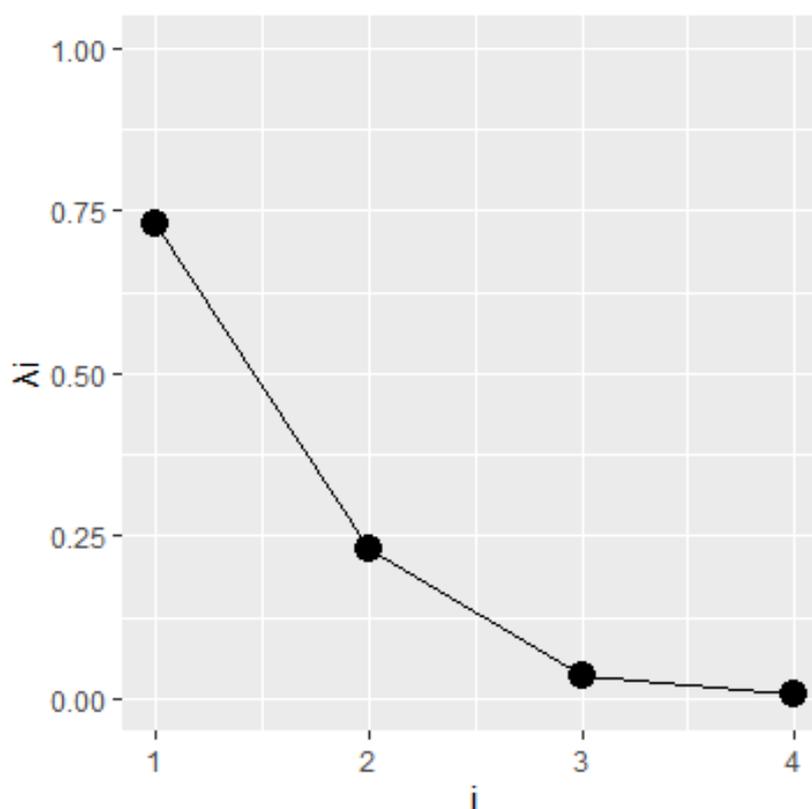


Figura 1: scree plot ilustrativo feito pela autora

## 2.3 Análise Envoltória de Dados (DEA)

A Análise Envoltória de Dados ou DEA (de *Data Envelopment Analysis*) é um método que permite avaliar a performance de uma determinada Unidade Decisória ou DMU (de *Decision Making Unit*) em relação a outras, como discutido em ((COOPER; SEIFORD; TONE, 2006)). Usualmente, uma DMU corresponde a uma firma, no entanto, o método é bastante geral e pode ser aplicado para comparar agências governamentais, países, pessoas, dentre outros tipos de entidades.

A DEA provê uma estratégia simples de avaliar a eficiência na qual uma certa DMU utiliza um ou mais insumos para produzir um ou vários produtos. A estratégia se baseia fundamentalmente na ideia de que, tipicamente, eficiências (produtividades) são baseadas em razões da forma:

$$\frac{\text{produto}}{\text{insumo}},$$

que são medidas naturais de performance. Alguns exemplos são o lucro por unidade de custo, a produção por trabalhador e a renda per capita.

Store	A	B	C	D	E	F	G	H
Employee	2	3	3	4	5	5	6	8
Sale	1	3	2	3	4	2	3	5
Sale/Employee	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

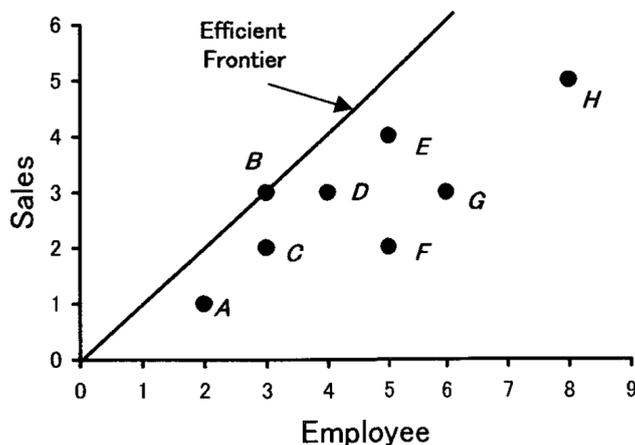


Figura 2: Ilustração retirada do livro do Cooeper et al, 2007, p.3

No caso específico em que um único insumo e um único produto são observados, a comparação das performances é feita de modo trivial a partir da razão **produto/insumo**. A título de ilustração, considere a tabela de dados e o gráfico da Figura 2. Oito lojas (*stores*), nomeadas de “A” a “H”, são comparadas quanto a performance de vendas (*sales*) de um produto único a partir da utilização de quantidades diferentes do insumo único correspondente ao número de empregados (*employee*). Repare que a tabela de dados também apresenta a razão **produto/insumo** de cada loja, sendo a loja “B” a de melhor performance, cuja razão (a maior observada) é igual a 1. A linha apresentada no gráfico correspondente à reta que passa pela origem e pelo ponto B (ou, de outro modo, pelo par ordenado correspondente à loja “B”, a mais eficiente), cujo coeficiente angular é exatamente 1 (a razão **produto/insumo** da loja “B”). Como discutido em Coelli et al. (2005), tal reta representa a **fronteira eficiente**. Observe que os pares ordenados de todas as demais lojas, que são menos eficientes ou produtivas que “B”, situam-se abaixo de tal fronteira.

### 2.3.1 DEA com retornos constantes de escala (CRS)

O problema de medir a performance se torna relativamente mais complexo quando há mais do que um insumo ou mais do que um produto. Digamos, de modo mais geral, que as DMUs “ $o$ ” são associados um vetor de insumos  $(x_{1o}, \dots, x_{mo})$  e um vetor de produtos  $(y_{1o}, \dots, y_{so})$  (COELLI et al., 2005). Se existirem um certo vetor de pesos para os insumos  $(v_{1o}, \dots, v_{mo})$  e outro para os produtos  $(u_{1o}, \dots, u_{so})$ , então, a eficiência ou produtividade da DMU “ $o$ ” poderia ser avaliada através da razão:

$$\frac{u_{1o}y_{1o} + \dots + u_{so}y_{so}}{v_{1o}x_{1o} + \dots + v_{mo}x_{mo}} = \frac{\text{produto virtual}}{\text{insumo virtual}}. \quad (2.2)$$

Na DEA, os “pesos” de cada DMU “ $o$ ” são escolhidos de modo a maximizar a performance da DMU (COELLI et al., 2005). De outra maneira, eles não precisam ser especificados *a priori* e podem ser obtidos a partir do conjunto de dados.

Digamos que sejam avaliadas  $n$  DMUs. Para cada  $j = 1, \dots, n$  devem ser observados um vetor numérico de insumos  $(x_{1j}, \dots, x_{mj})$  e outro de produtos  $(y_{1j}, \dots, y_{sj})$ . Em princípio, quanto maiores os valores dos produtos e menores os dos insumos, maior deveria ser a eficiência da DMU.

Na discussão que se segue, nos basearemos em um dos modelos DEA mais conhecidos, conhecido como CCR (Charnes, Cooper and Rhodes). Nele, a partir do conjunto de dados define-se um sistema de pesos ótimo para cada DMU “ $o$ ” ( $\mathbf{v}_o = (v_{1o}^*, \dots, v_{mo}^*)$  e  $\mathbf{u}_o = (u_{1o}^*, \dots, u_{so}^*)$ ) como solução do problema de otimização (COELLI et al., 2005):

$$\begin{aligned} (FP_o) \quad & \max_{\mathbf{v}, \mathbf{u}} \frac{u_{1o}y_{1o} + \dots + u_{so}y_{so}}{v_{1o}x_{1o} + \dots + v_{mo}x_{mo}} \\ \text{sujeito a} \quad & \frac{u_{1j}y_{1j} + \dots + u_{sj}y_{sj}}{v_{1j}x_{1j} + \dots + v_{mj}x_{mj}} \leq 1 (j = 1, \dots, n) \\ & v_{1o}, v_{2o}, \dots, v_{mo} \geq 0 \\ & u_{1o}, u_{2o}, \dots, u_{so} \geq 0 \end{aligned} \quad (2.3)$$

que, por sua vez, equivale ao seguinte problema de Programação Linear:

$$\begin{aligned}
(LP_o) \quad & \max_{\mathbf{v}, \mathbf{u}} u_{1o}y_{1o} + \cdots + u_{so}y_{so} \\
\text{sujeito a} \quad & v_{1o}x_{1o} + \cdots + v_{mo}x_{mo} = 1 \\
& u_{1j}y_{1j} + \cdots + u_{sj}y_{sj} \leq v_{1j}x_{1j} + \cdots + v_{mj}x_{mj} \quad (j = 1, \dots, n) \\
& v_{1o}, v_{2o}, \dots, v_{mo} \geq 0 \\
& u_{1o}, u_{2o}, \dots, u_{so} \geq 0.
\end{aligned} \tag{2.4}$$

Coelli et al. (2005) argumentam que os valores ótimos não dependem das unidades de medida adotadas para cada insumo ou produto, desde que as unidades sejam as mesmas para todas as DMUs. O problema de programação linear  $LP_o$  (Equação 2.4) pode ser resolvido através do método **simplex**. Se  $\mathbf{v}^* = (v_{1o}^*, \dots, v_{mo}^*)$  e  $\mathbf{u}^* = \mathbf{u}_o = (u_{1o}^*, \dots, u_{so}^*)$  são os pesos ótimos obtidos, então, a eficiência da DMU “ $o$ ” pode ser escrita como

$$\theta_o^* = \frac{u_{1o}^*y_{1o} + \cdots + u_{so}^*y_{so}}{v_{1o}^*x_{1o} + \cdots + v_{mo}^*x_{mo}}.$$

Diz-se que a DMU “ $o$ ” é CCR-eficiente se  $\theta_o^* = 1$  e existem vetores ótimos  $\mathbf{v}^* > \mathbf{0}$  e  $\mathbf{u}^* > \mathbf{0}$ . Caso contrário, a DMU em questão é dita ineficiente.

É interessante destacar que se  $\theta_o^* < 1$ , então, existe  $j$  tal que  $1 \leq j \leq n$  tal que  $u_{1o}^*y_{1j} + \cdots + u_{so}^*y_{sj} = v_{1o}^*x_{1j} + \cdots + v_{mo}^*x_{mj}$  pois, do contrário, poderíamos aumentar o valor da função objetivo aumentando todos os pesos do produto em alguma proporção comum (COOPER; SEIFORD; TONE, 2006). Seja, então, o conjunto:

$$E'_o = \{j; u_{1o}^*y_{1j} + \cdots + u_{so}^*y_{sj} = v_{1o}^*x_{1j} + \cdots + v_{mo}^*x_{mj}\}.$$

O conjunto de referência para a DMU “ $o$ ” é o subconjunto  $E_o \subset E'_o$  composto das DMUs eficientes. O conjunto gerado por  $E_o$  (as combinações lineares de seus elementos) é a fronteira eficiente para a DMU “ $o$ ”.

Consideremos um importante caso particular em que há um único insumo e de valor constante para todas as DMUs ( $m = 1$  e  $x_{11} = \cdots = x_{1n} = 1$ ). Nele, o vetor de pesos associados ao insumo é definido por um único escalar  $v_{1o}$  cujo valor ótimo é (trivialmente)  $v_{1o}^* = 1$ . Finalmente, o problema descrito na Equação 2.4 é simplificado para:

$$\begin{aligned}
(LP_o^*) \quad & \max_{\mathbf{u}} u_{1o}y_{1o} + \cdots + u_{so}y_{so} \\
\text{sujeito a} \quad & u_{1j}y_{1j} + \cdots + u_{sj}y_{sj} \leq 1 \quad (j = 1, \dots, n) \\
& u_{1o}, u_{2o}, \dots, u_{so} \geq 0.
\end{aligned} \tag{2.5}$$

Para ilustrar este caso mais simples, considere a Figura 3, que apresenta dados de oito lojas (nomeadas de “A” a “H”) que são comparadas quanto a performance na produção de dois produtos – clientes (*customers*) e vendas (*sale*) –, mantendo o insumo único (empregados) constante e igual a 1.

Store		<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>
Employees	$x$	1	1	1	1	1	1	1
Customers	$y_1$	1	2	3	4	4	5	6
Sales	$y_2$	5	7	4	3	6	5	2
EFICIÊNCIA		0.714	1	0.7	0.75	1	1	1

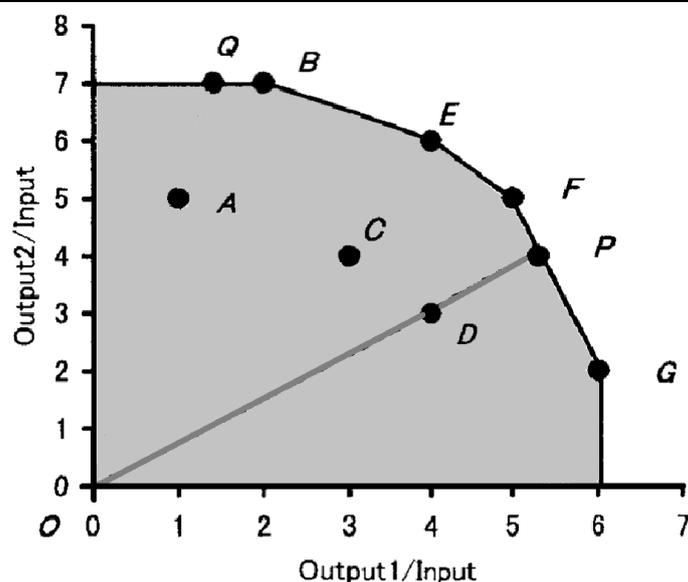


Figura 3: Ilustração retirada do livro do Coeper et al, 2007, pp.8-9

A Figura 3 exibe uma tabela de dados, em que são apresentadas as eficiências das lojas obtidas pela resolução do problema descrito na Equação 2.5. No gráfico exibido são apresentados os pares ordenados de produtos de cada loja e uma fronteira eficiente é formada pelos pontos associados às lojas mais eficientes (que são aquelas com eficiência unitária) e pelos segmentos de reta que ligam tais pontos. Apenas três lojas não são eficientes (“A”, “C” e “D”), cujos pontos estão abaixo da fronteira. Os valores de suas eficiências, menores que 1, possuem uma interpretação interessante. Considere, por exemplo, a loja “D”, que possui eficiência de 0.75. Se multiplicarmos cada produto desta loja pelo inverso da eficiência (ou seja, por  $1/0.75 = 4/3$ ), obtemos o ponto “P”, que está destacado sobre a fronteira. O ponto “P” (uma espécie de projeção do ponto “D” na fronteira) pode ser interpretado como a combinação de produtos que a loja “D” deveria produzir para ter eficiência máxima igual a 1. Portanto, se os pesos ideais associados à loja “D” são utilizados, a eficiência (0.75), por sua vez, é igual à razão entre o comprimento do segmento que

liga a origem ao ponto D e o comprimento do segmento que liga a origem ao ponto “P”. Dessa maneira, quanto mais próximo estiver o par ordenado de uma DMU em relação à fronteira, maior será sua eficiência.

## 2.4 Ponderação nos Indicadores Compostos

Nesta seção serão apresentados os dois métodos utilizados no presente trabalho para construir os Indicadores Compostos: um baseado na PCA (cf. Seção 2.3) e o outro na DEA (cf. Seção 2.3). A partir daqui, assumiremos que são observados dados de  $p$  variáveis (indicadores simples) para  $n$  unidades observacionais (nas nossas aplicações, anos; mas, em outros contextos poderiam representar diferentes países, por exemplo). O conjunto de dados pode ser representado através de uma matriz  $\mathbf{X}$  (não confundir com o vetor aleatório  $\mathbf{X}$ ) de dimensão  $n \times p$ , cujas linhas, em ordem, são representadas por  $\mathbf{x}_1, \dots, \mathbf{x}_n$  e, para  $i = 1, \dots, n$ ,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  é o vetor  $p$ -dimensional que contém dados de cada um dos  $p$  indicadores individuais associados à  $i$ -ésima observação ( $x_{ij}$  representa o valor do  $j$ -ésimo indicador simples observado para a  $i$ -ésima unidade observacional) – em nossas aplicações, consideraremos que  $x_{ij}$  já foi *previamente padronizado* mediante uma transformação do tipo “*min-max*” – ver Seção 3.1 .

### 2.4.1 Pesos via PCA

Usando a PCA, podemos identificar um conjunto de  $q$  componentes principais (amostrais) mais relevantes, onde  $1 \leq q \leq p$ , como discutido na Seção 2.2.3. Se  $\mathbf{S}$  representa a matriz de variância amostral (essencialmente,  $S$  é uma matriz simétrica  $p \times p$ , cujo elemento da linha  $l$  e coluna  $j$  é definido como a covariância amostral entre o  $j$ -ésimo e o  $l$ -ésimo indicadores simples), então, as componentes principais podem ser definidas a partir dos autovetores de  $\mathbf{S}$ , sendo que as variâncias das componentes principais correspondem aos autovalores que aparecem, por construção, em ordem decrescente:  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \hat{\lambda}_p$ . O vetor de coeficientes da  $l$ -ésima componente principal ( $l = 1, \dots, p$ ) corresponde ao autovetor  $\hat{\mathbf{e}}_l = (\hat{e}_{l1}, \dots, \hat{e}_{lp})$ , que é exatamente aquele associado ao  $l$ -ésimo autovalor  $\hat{\lambda}_l$ .

Tendo escolhido um número  $q$  de primeiras componentes principais mais relevantes<sup>5</sup>, podemos obter um Indicador Composto via PCA seguindo os mesmos passos que (BOUDT et al., 2022). Essencialmente, Boudt propõe que o peso de cada variável

<sup>5</sup>Por exemplo, através do *scree plot* ou da medida de qualidade ajustada das  $q$  componentes principais iniciais, definida como  $(\hat{\lambda}_1 + \dots + \hat{\lambda}_q)/(\hat{\lambda}_1 + \dots + \hat{\lambda}_p)$ , de acordo com o que discutimos na Seção 2.2.

$j = 1, \dots, p$  seja obtido como uma combinação linear dos pesos de tal variável nas  $q$  componentes principais escolhidas  $\hat{e}_{j1}, \hat{e}_{j2}, \dots, \hat{e}_{jq}$ , com coeficientes respectivamente dados por  $\hat{\lambda}_1/p, \hat{\lambda}_2/p, \dots, \hat{\lambda}_q/p$ . Formalmente, o Indicador Composto associará a  $i$ -ésima observação o valor:

$$IC_i = w_1x_{i1} + w_2x_{i2} + \dots + w_px_{ip},$$

onde o peso  $w_j$  (associado ao  $j$ -ésimo indicador simples e que não depende ou varia com as unidades observacionais) é dado por:

$$w_j = \hat{e}_{j1} \frac{\hat{\lambda}_1}{p} + \hat{e}_{j2} \frac{\hat{\lambda}_2}{p} + \dots + \hat{e}_{jq} \frac{\hat{\lambda}_q}{p}.$$

## 2.4.2 Pesos via DEA: Abordagem BoD

Segundo Cherchye et al. (2007), o método do Benefício da Dúvida (BoD) utiliza a DEA (cf. Seção 2.3) para construir o indicador composto. Diferente do método baseado na PCA, na análise via BoD obtém-se um vetor de pesos diferente para cada unidade observacional  $i = 1, \dots, n$  e define-se, então, formalmente, o valor do indicador composto para a  $i$ -ésima observação via:

$$IC_i = w_{1i}x_{i1} + w_{2i}x_{i2} + \dots + w_{pi}x_{pi},$$

sendo  $w_{ji}$  o peso escolhido para o  $j$ -ésimo indicador Ssimples na obtenção do indicador composto para a  $i$ -ésima unidade observacional.

Como discutido em Fusco (2015), para se encontrar os pesos  $w_{1i}, \dots, w_{pi}$  associados a uma unidade observacional  $i$  particular (para  $i = 1, \dots, n$ ), basta resolver um **problema de Programação Linear** equivalente ao descrito na Equação 2.5, associado à DEA com múltiplos ( $p$ , mais precisamente aqui) produtos e um único insumo. Especificamente,  $w_{1i} = u_{1i}^*, \dots, w_{pi} = u_{pi}^*$  são os valores dos argumentos que resolvem o seguinte problema:

$$\begin{aligned} (LP_o^*) \quad & \max_{u_{1i}, \dots, u_{pi}} u_{1i}x_{i1} + \dots + u_{pi}x_{ip} \\ \text{sujeito a} \quad & u_{1i}x_{i\ell} + \dots + u_{pi}x_{i\ell} \leq 1 \quad (\ell = 1, \dots, n) \\ & u_{1i}, u_{2i}, \dots, u_{pi} \geq 0. \end{aligned} \tag{2.6}$$

Repare que o método BoD aplica a DEA em um contexto com um único insumo ou *input* e utiliza como produtos ou *outputs* os dados dos  $p$  indicadores simples. Cada  $i$ -ésima unidade observacional corresponde a uma DMU. Destaca-se, como discutido na Seção 2.3, que são resolvidos  $n$  problemas de programação linear como o descrito na Equação 2.6,

sendo um para cada  $i = 1, \dots, n$ . Assim, para cada unidade observacional se obtém um sistema de pesos particular e é isso que justifica a nomenclatura “Benefício da Dúvida”, uma vez que o sistema de pesos obtido para a observação  $i$  é aquela que torna a sua performance a melhor possível. Como o BoD é uma aplicação da DEA, herda todas suas vantagens. Em especial, não é afetado por mudanças na escala de um indicador simples. Mais detalhes sobre o método em (CHERCHYE et al., 2007; FUSCO, 2015).

## 2.5 Banco de Dados

O portal Trading Economics<sup>6</sup> disponibiliza dados históricos oficiais e previsões de indicadores econômicos e financeiros para 196 países, tendo seu site mais de 1 bilhão de visualizações desde 2011. No presente trabalho, iremos apresentar indicadores econômicos compostos construídos a partir dos **oito indicadores simples** apontados como sendo aqueles mais relevantes pelo portal. Na sequência, descreveremos sucintamente cada um destes indicadores, informando a fonte na qual obtivemos dados a eles associados e o período de cobertura disponibilizado na respectiva fonte.<sup>7</sup>

**1. PIB (Taxa de crescimento do PIB):** o Instituto Brasileiro de Geografia e Estatística (IBGE) define o Produto Interno Bruto como sendo a “*soma dos valores de todos os bens e serviços finais produzidos por um país, estado ou cidade, geralmente em um ano*”.<sup>8</sup> Será utilizada aqui a taxa de crescimento real anual do Produto Interno Bruto (doravante, PIB), que corresponde à variação percentual do PIB real entre dois anos consecutivos – grosso modo, o PIB real corrige o PIB usual, chamado PIB nominal, no caso fixando os preços dos produtos nos valores observados para 2015. Dados das taxas de crescimento do PIB real podem ser encontrados no site do Banco Mundial<sup>9</sup> no período de 1961 a 2021. Para o ano de 2022, utilizamos uma projeção apresentada no portal G1, da Globo.<sup>10</sup>

---

<sup>6</sup>[www.tradingeconomics.com](http://www.tradingeconomics.com)

<sup>7</sup>Trabalhamos apenas com dados anuais.

<sup>8</sup>[ibge.gov.br](http://ibge.gov.br)

<sup>9</sup><https://data.worldbank.org/indicator>

<sup>10</sup>[g1.globo.com/economia/noticia/2023/03/02/edit-pib-do-brasil-avanca-29percent-em-2022.ghtml](https://g1.globo.com/economia/noticia/2023/03/02/edit-pib-do-brasil-avanca-29percent-em-2022.ghtml)

**2. Juros (Taxa básica de juros, Selic):** segundo o Banco Central do Brasil (Bacen)<sup>11</sup>, a taxa Selic (taxa básica de juros) é uma ferramenta importante de política monetária e cujo valor afeta positivamente as demais taxas de juros da economia – por exemplo, taxas de juros em empréstimos ou financiamentos, rendimentos de aplicações. No portal Ipeadata<sup>12</sup>, é possível encontrar dados da taxa Selic para o período de 1975 a 2022. Consideramos aqui a média anal das taxas de juros - Selic - diárias fixadas pelo Comitê de Política Monetária.

**3. Inflação (Taxa de inflação):** de acordo com o IBGE<sup>13</sup>, o Índice de Preços ao Consumidor Amplo (IPCA) “*mede a variação de um conjunto de produtos e serviços comercializados no varejo, referentes ao consumo pessoal das famílias, cujo rendimento varia entre 1 e 40 salários mínimos*”, sendo um dos indicadores oficiais mais importantes de inflação no Brasil. Aqui, será utilizada a taxa de inflação anual (correspondente à variação percentual do índice IPCA), cujos dados podem ser encontrados no site do IBGE<sup>14</sup> para o período de 1995 a 2022.

**4. Desemprego (Taxa de desemprego):** a taxa de desemprego é definida pelo IBGE<sup>15</sup> como a proporção de pessoas desempregadas (que não estão trabalhando, mas estão disponíveis e tentam encontrar trabalho) na força de trabalho (pessoas com 14 anos ou mais que trabalham ou estão desempregadas). Dados da taxa de desemprego podem ser encontrados no portal Ipeadata<sup>16</sup> para o período de 1991 a 2021 – para o ano de 2022, utilizamos a taxa divulgada pelo IBGE<sup>17</sup>.

**5. Dívida Pública (Dívida líquida do setor público em relação ao PIB):** o portal Infomoney<sup>18</sup> define Dívida Pública como o “*resultado dos empréstimos feitos por entes estatais com o setor privado e com órgãos públicos financeiros*”. Aqui, trabalharemos com a razão entre a a Dívida Líquida do Setor Público (total - governo federal e Banco Central) e o PIB nominal, ambas medidas em milhões de reais e cujos dados estão disponíveis no portal Ipeadata<sup>19</sup>, para o período de 1996 a 2022.

---

<sup>11</sup> <<https://www.bcb.gov.br/controleinflacao/taxaselic>>

<sup>12</sup> <<http://www.ipeadata.gov.br>>

<sup>13</sup> <<https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos>>

<sup>14</sup> <<https://www.ibge.gov.br/estatisticas>>

<sup>15</sup> <<https://ibge.gov.br/explica/desemprego.php>>

<sup>16</sup> <<http://www.ipeadata.gov.br>>

<sup>17</sup> <<https://ibge.gov.br>>

<sup>18</sup> <<https://www.infomoney.com.br>>

<sup>19</sup> <[www.ipeadata.gov.br](http://www.ipeadata.gov.br)>

**6. Balança comercial (Saldo da balança comercial em relação ao PIB):** como reportado no portal Infomoney<sup>20</sup>, o saldo da balança comercial corresponde à diferença entre os valores das exportações e importações de um determinado país. Serão utilizados aqui, os dados da razão entre o saldo da balança comercial brasileira (em valores FOB ou *Free on Board*) e o PIB a Câmbio Médio<sup>21</sup> – ambos medidos em milhões de dólares –, obtidos no portal Ipeadata<sup>22</sup> e disponíveis conjuntamente para o período de 1995 a 2022.

**7. Conta Corrente (Saldo da conta corrente em relação ao PIB):** segundo o Banco Mundial<sup>23</sup>, o saldo da conta corrente denota a soma das exportações líquidas de bens e serviços, com a renda primária líquida e a renda secundária líquida, refletindo, deste modo, um resumo das transações econômicas (de bens, serviços, renda e ativos financeiros) dos residentes de um país com residentes dos demais países. No presente trabalho são considerados os dados da razão entre o saldo da conta corrente e o PIB, obtidos no portal Ipeadata<sup>24</sup> e disponíveis de 1980 a 2021 – para o ano de 2022 será considerado o valor reportado no portal Money Times<sup>25</sup>.

**8. Risco país (Taxa de risco país):** Mede o risco a que os investidores estrangeiros estão sujeitas ao investir no país refletindo, assim, a confiabilidade ou atratividade para se investir em um dado país, sob o ponto de vista dos investidores estrangeiros, como explicado no Bacen<sup>26</sup>. Neste trabalho, utilizaremos dados do risco país do Brasil segundo o indicador EMBI+ (*Emerging Markets Bond Index Plus*), disponibilizado no portal Ipeadata<sup>27</sup> para o período de 1995 a 2022.

---

<sup>20</sup><https://www.infomoney.com.br>

<sup>21</sup>Ou seja, o PIB anual mensurado em dólares correntes utilizando, para isso, a média entre as taxas de câmbio de compra e venda ao longo do ano.

<sup>22</sup><http://www.ipeadata.gov.br/>

<sup>23</sup><https://data.worldbank.org>

<sup>24</sup><http://www.ipeadata.gov.br/>

<sup>25</sup>[www.moneytimes.com.br](http://www.moneytimes.com.br)

<sup>26</sup><https://www.bcb.gov.br/conteudo>

<sup>27</sup><http://www.ipeadata.gov.br/Default.aspx>

## 3 Resultados

### 3.1 Dois Indicadores Compostos para Economia

Na Seção 2.4 apresentamos dois métodos – baseados em dados – que viabilizam a construção de indicadores compostos, um associado à análise de componentes principais (Seção 2.2) e outro na análise envoltória de dados (Seção 2.3). Neste capítulo serão apresentados dois Indicadores Compostos (ICs) – um para cada método acima mencionado – a fim de medir a performance do Governo Federal em relação à condução da política econômica. Tais ICs são construídos a partir de oito indicadores simples, a saber: (Taxa de Crescimento do) **PIB**, (Taxa de) **Juros**, (Taxa de) **Inflação**, (Taxa de) **Desemprego**, **Dívida Pública** (em relação ao PIB), (Saldo da) **Balança comercial** (em relação ao PIB), (Saldo da) **Conta corrente** (em relação ao PIB) e (Taxa de) **Risco País**. Dados anuais destes Indicadores Simples foram obtidos, de fontes oficiais principalmente, para o maior período comum disponível: de 1997 até 2022. Definições sucintas dos indicadores e as fontes dos dados podem ser encontradas na Seção 2.5. O período analisado contempla os seguintes mandatos presidenciais:

- (i) Dois últimos anos do primeiro mandato de Fernando Henrique Cardoso, de 1997 a 1998 (**FHCI**)
- (ii) Segundo mandato de Fernando Henrique Cardoso, de 1999 a 2002 (**FHCII**)
- (iii) Primeiro mandato de Luiz Inácio Lula da Silva, de 2003 a 2006 (**Lula1**)
- (iv) Segundo mandato de Luiz Inácio Lula da Silva, de 2007 a 2010 (**Lula2**)
- (v) Primeiro mandato de Dilma Vana Rousseff, de 2011 a 2014 (**Dilma1**)
- (vi) Primeiro ano do segundo mandato de Dilma Vana Rousseff, em 2015 (**Dilma2**)
- (vii) Transição dos governos DilmaII e Temer, em 2016 (**Transição**)

- (vii) Dois últimos anos do mandato de Michel Temer (Temer), de 2017 a 2018 (**Temer**)
- (viii) Mandato de Jair Messias Bolsonaro de 2019 a 2022 (**Bolsonaro**)

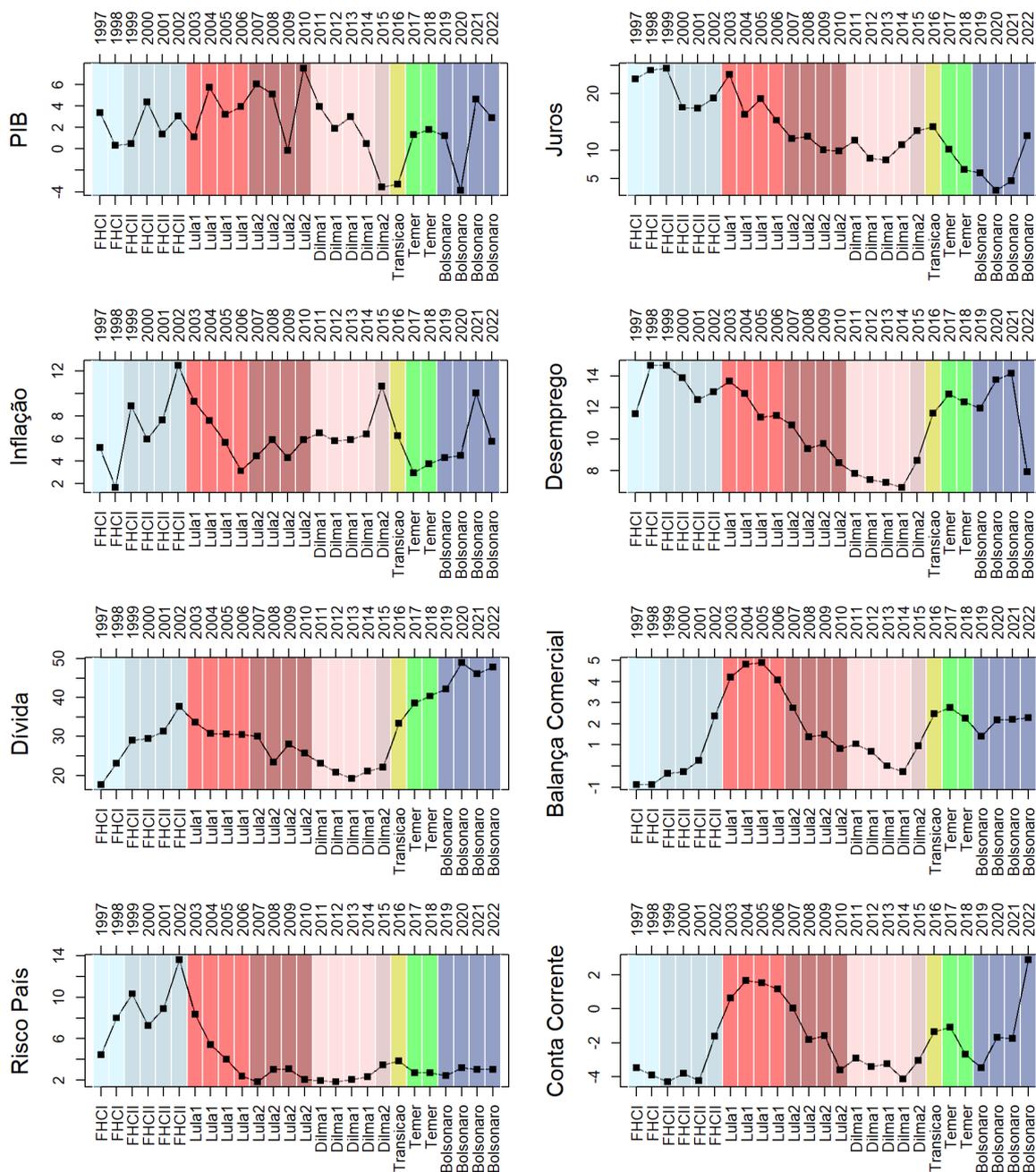


Figura 4: Evolução dos indicadores simples (%)

Na Figura 4 exibimos os gráficos (de linha) de cada um dos indicadores simples no período de 1997 a 2022. É possível notar que a taxa de crescimento do PIB evoluiu de modo bastante irregular, tendo os maiores valores sido observados nos dois mandatos do

Lula e os menores nos mandatos da Dilma e no governo Bolsonaro. A taxa de juros, por sua vez, aparentemente segue uma tendência de decrescimento no período analisado – com uma potencial reversão no último governo. A taxa de inflação, que teve seu menor valor em 1998 (último ano do primeiro mandato de FHC) atinge seu pico às vésperas da primeira eleição do Lula e volta a apresentar picos secundários após a reeleição de Dilma e no penúltimo ano do governo Bolsonaro. A taxa de desemprego apresenta um padrão de queda desde o segundo mandato de FHC, atravessando todo o governo Lula (dois mandatos) e o primeiro mandato de Dilma. Há, porém, uma reversão da tendência a partir daí que culmina em um pico secundário no penúltimo ano do governo de Bolsonaro. A dívida pública, como proporção da riqueza produzida no país, cresce durante os dois mandatos de FHC, é reduzida ao longo dos governos de Lula e Dilma e volta a crescer, de forma ainda mais acentuada a partir daí. O risco país tem uma queda acentuada a partir do primeiro mandato de Lula e muda de patamar – antes, nos governos de FHC, oscilava em torno de 10% – do segundo mandato Lula em diante é sempre inferior a 5%. Os saldos da balança comercial e da conta corrente apresentam comportamentos similares: crescem substancialmente no governo FHC até atingir o pico no primeiro mandato de Lula. A partir daí, decrescem até o final do período em que Dilma esteve na presidência, passando a oscilar nos últimos anos. Uma das principais diferenças entre tais indicadores é que, enquanto a balança comercial se estabiliza no governo Bolsonaro, a conta corrente aumenta drasticamente, atingindo o pico em todo o período investigado.

Alguns indicadores simples considerados em nossa análise (como a taxa de desemprego, por exemplo) são tais que, quanto maior seus valores, pior (**interpretação invertida**). Além do indicador mencionado, vale o mesmo para a taxa de juros, a taxa de inflação, a dívida pública em relação a PIB e o risco país. Em adição ao problema reportado, há um outro: as escalas dos diferentes indicadores não são necessariamente comparáveis. Tais problemas foram solucionadas da seguinte maneira: (i) no primeiro caso, os dados dos indicadores simples com interpretação invertida foram multiplicados por  $-1$  (revertendo a interpretação); (ii) os dados de todos os Indicadores Simples foram padronizados através da transformação min-max, tal como discutido em OCDE:

$$IST = \frac{IS_t - \min IS}{\max IS - \min IS}, \quad (3.1)$$

onde  $\min IS$  representa o menor valor,  $\max IS$  representa o maior valor e  $IS_t$  o valor observado no ano  $t$  para o correspondente indicador simples – ou eventualmente o indicador simples multiplicado por  $-1$ , quando necessário, conforme (i). Deste modo, os indicadores simples transformados devidamente: (a) assumem valores entre 0 e 1 e (b) quanto

maiores seu valores, melhor. A Figura 5 apresenta os gráficos dos dados associados aos indicadores simples já transformados. Destacamos que os indicadores compostos serão construídos como combinações lineares dos indicadores simples já transformados.

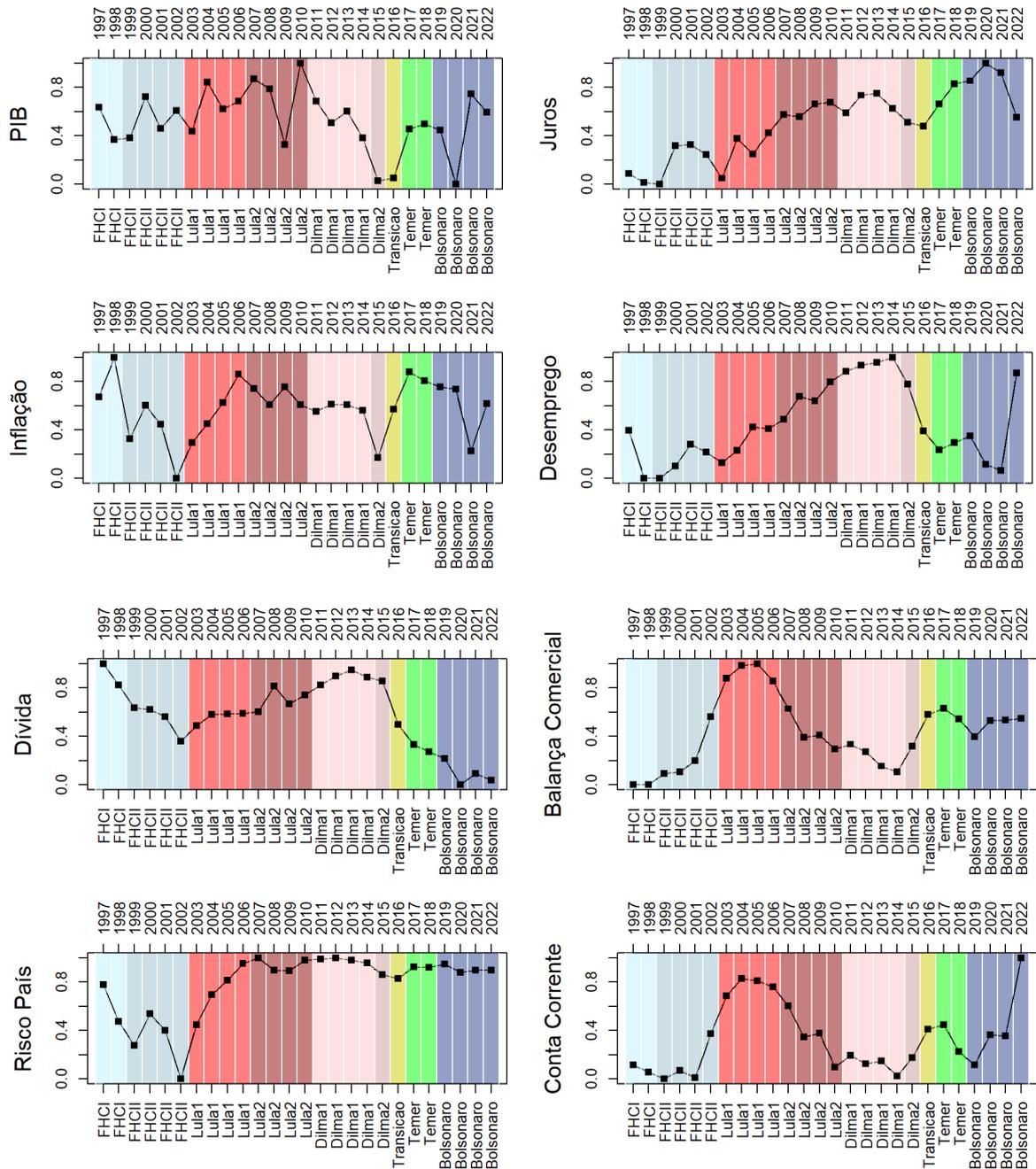


Figura 5: Evolução dos indicadores simples transformados

Apresentaremos na sequência os dois indicadores compostos produzidos no presente trabalho: um baseado na PCA e outro na DEA (BoD). Também consideraremos um terceiro indicador, mais simples, baseado na média aritmética dos oito indicadores simples – que servirá de referência.

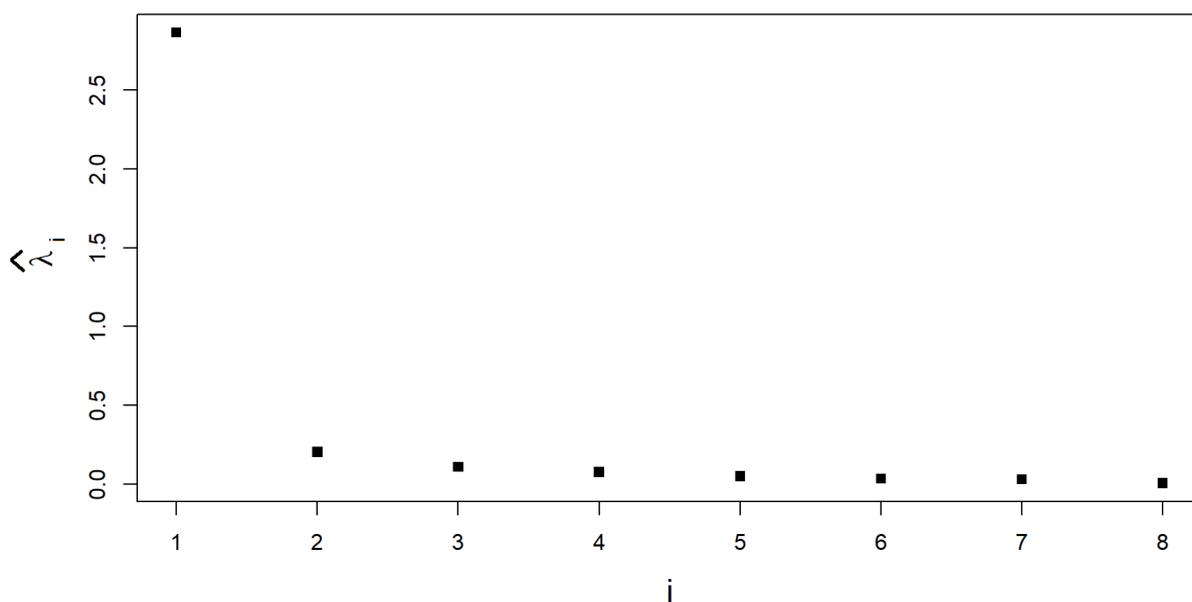


Figura 6: Scree plot das variâncias estimadas

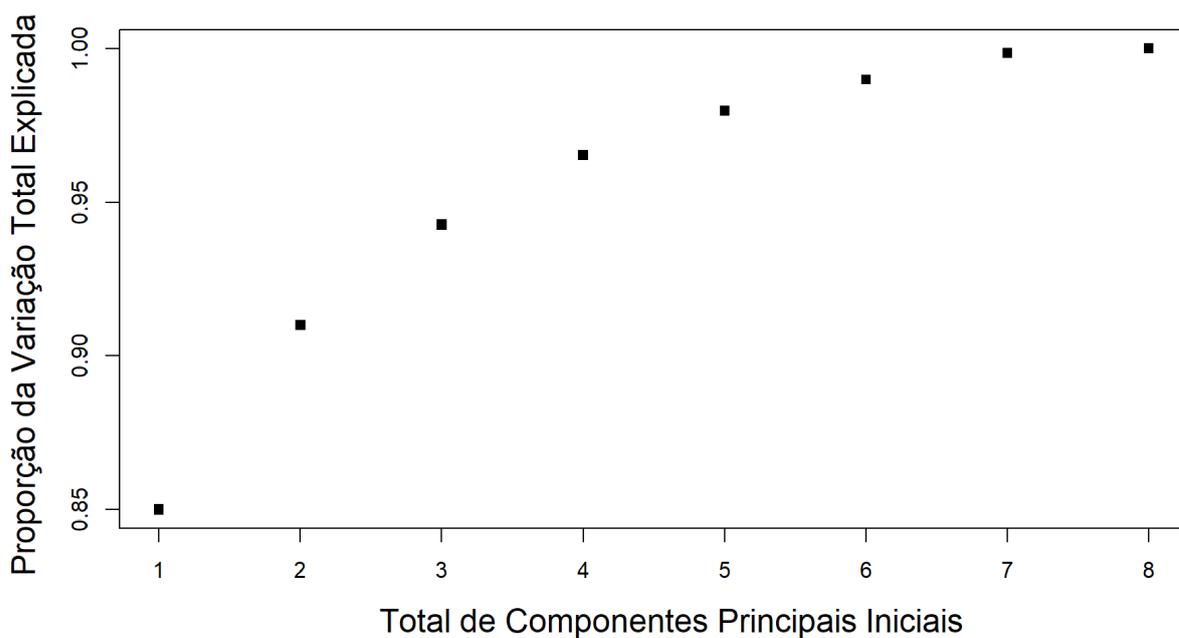


Figura 7: Importância relativa das componentes principais iniciais

Para o IC baseado na PCA, pontuamos que há inicialmente  $p = 8$  variáveis e, assim, precisamos escolher o número de componentes principais a utilizar. O objetivo é obter um total de componentes principais mais relevantes bem menor que este número, sem perder o poder explicativo. Como vimos na (Seção 2.2), podemos analisar o scree plot (Figura 6), assim como o poder de ajuste, medido pela proporção da variância total explicada pelas primeiras componentes principais (Figura 7). O scree plot indica que o ponto de máxima curvatura está associado ao uso das duas primeiras componentes

principais: ou seja, incluir uma componente adicional a partir da terceira parece pouco relevante. A situação é corroborada pela Figura 7, na qual visualizamos que as duas primeiras componentes principais juntas respondem por mais de 90% da variabilidade total dos Indicadores Simples Transformados. Destacamos, aliás, que a primeira componente simples explica cerca de 85% da variabilidade total. Assim, escolhemos trabalhar apenas com a primeira componente principal – como veremos, mais adiante, a inclusão da segunda componente principal tem pouco efeito sobre o IC final.

Tabela 1: Coeficientes das componentes principais

	<b>CP1</b>	<b>CP2</b>	<b>CP3</b>	<b>CP4</b>	<b>CP5</b>	<b>CP6</b>	<b>CP7</b>	<b>CP8</b>
PIB	<b>-0.3</b>	-0.010	-0.080	-0.532	-0.292	-0.638	0.309	-0.112
Juros	<b>-0.5</b>	-0.220	0.310	0.324	0.03871	0.209	0.400	-0.582
Inflação	<b>-0.4</b>	0.009	0.039	0.552	-0.259	-0.439	-0.552	0.048
Desemprego	<b>-0.3</b>	-0.116	-0.924	0.139	0.101	0.143	0.082	-0.044
Dívida	<b>-0.4</b>	-0.520	0.187	-0.199	0.450	0.032	-0.105	0.497
Balança Com.	<b>-0.3</b>	0.643	0.072	0.253	0.128	-0.008	0.442	0.471
Risco País	<b>-0.4</b>	0.188	0.050	-0.332	-0.529	0.575	-0.269	0.073
Conta Corrente	<b>-0.2</b>	0.468	0.004	-0.272	0.580	-0.055	-0.392	-0.412

A Tabela 1 exibe a matriz de carregamento, contendo os coeficientes das oito componentes principais. Observe que os coeficientes associados à primeira componente principal, destacados em negrito, são todos negativos. Assim, precisamos multiplicá-los por  $-1$ , uma vez que cada Indicador Simples (devidamente Transformado) tem a interpretação “quanto maior, melhor”.

A Figura 8 exibe o IC construído usando diferentes métodos, sendo eles: a PCAI (baseado apenas na primeira componente principal), a PCAII (usando as duas primeiras componentes principais) e o construído via média aritmética (Média). É notável que as três curvas estão próximas, logo o método mais eficiente é a média aritmética, já que exige menor esforço em cálculos e retorna resultados semelhantes. A figura também sugere que os diferentes Indicadores Simples Transformados têm quase que a mesma relevância, uma vez que na média aritmética seus pesos são iguais. Outra conclusão é que ao adicionar a segunda componente principal não há grande aumento na quantidade de informação substancial, ou seja, a PCAI mais eficiente que a PCAII.

No método da DEA, obtemos, para cada ano, o conjunto de pesos que maximiza a performance daquele ano. Todavia, como há poucas observações (apenas 26 anos) em relação ao total de variáveis (8 Indicadores Simples), a fronteira não é bem estimada e, assim, a eficiência máxima (1) é alcançada em inúmeros anos – como pode ser visto na Figura 9. Mesmo com tal problema, podemos obter – como será feito na Seção 3.2 –

o melhor sistema de pesos de cada ano e interpretá-los. Ainda observando a Figura 9, destacamos que ao comparar a DEA com a média aritmética simples, verificamos que em anos nos quais a eficiência não é máxima (ou seja, o IC construído via DEA tem valor inferior a 1), o IC via média aritmética tende a apresentar valores inferiores aos dos demais anos.

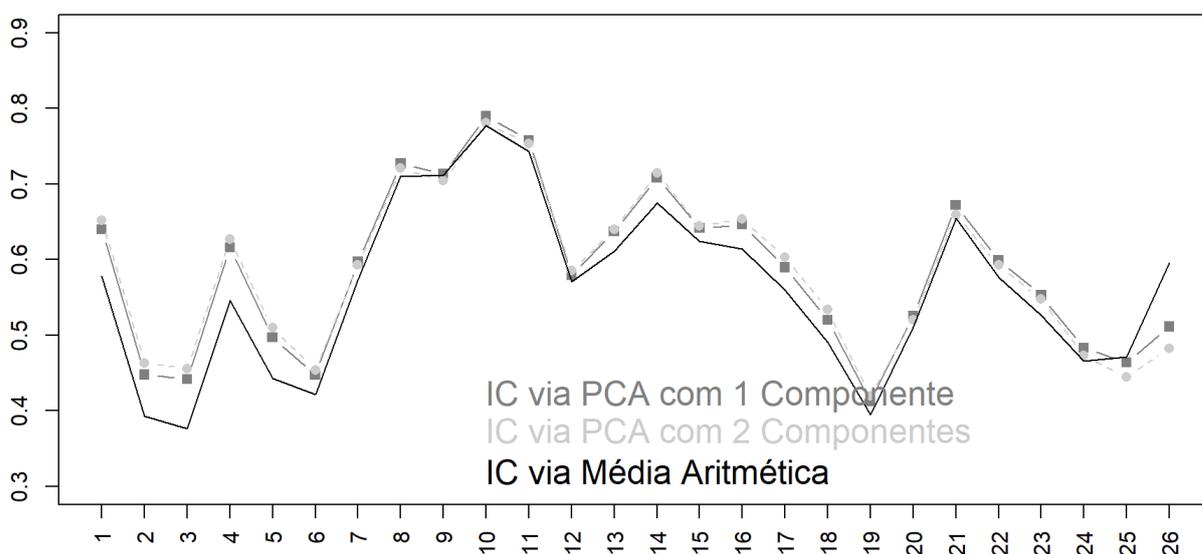


Figura 8: ICs obtidos via PCA e via média aritmética simples

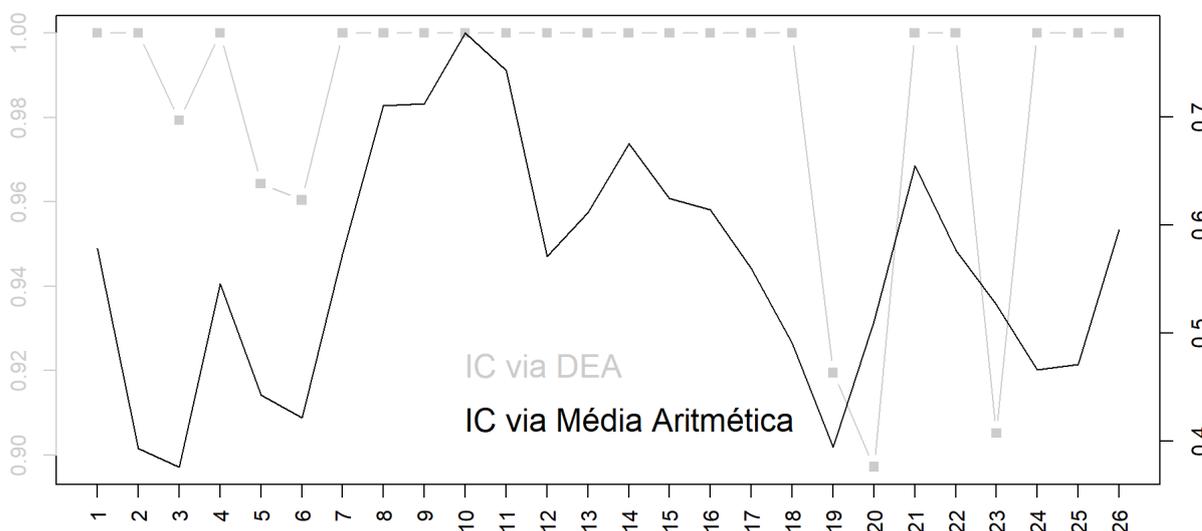


Figura 9: IC obtido via DEA e via média aritmética simples

A matriz de correlação (de Pearson) obtida para os diferentes ICs é exibida na Tabela 2 e sugere uma relação de dependência linear positiva forte entre o indicador via PCA (que usa uma única componente principal) e aquele associado à média aritmética simples. Nos demais casos, observa-se uma dependência linear positiva moderada. Os resultados

são positivos no sentido de que apontam que os diferentes ICs levam a comparações semelhantes dos diferentes anos.

Tabela 2: Correlação entre ICs

	<b>IC - média aritmética</b>	<b>IC - DEA</b>	<b>IC - PCA</b>
IC - média aritmética	1,000	0,418	0,965
IC - DEA	0,418	1,000	0,424
IC - PCA	0,965	0,424	1,000

Uma forma de avaliar a qualidade dos governos seria considerando ICs por mandato. Neste caso, podemos calcular a média dos ICs ao longo dos anos de cada mandato (cf. Tabela 3).

Tabela 3: ICs por mandato

	<b>IC - média aritmética</b>	<b>IC - DEA</b>	<b>IC - PCA</b>
FHCI	0.4855571	1.0000000	0.5436658
FHCII	0.4462970	0.9759943	0.5005339
Lula1	0.6931250	1.0000000	0.7068148
Lula2	0.6500547	1.0000000	0.6708878
Dilma1	0.5721961	1.0000000	0.5994390
Dilma2	0.3944007	0.9194480	0.4121965
Transicao	0.5107946	0.8971659	0.5258144
Temer	0.6154178	1.0000000	0.6359745
Bolsonaro	0.5147129	0.9763026	0.5026952

As correlações de Pearson (ver Tabela 4) sugerem que os ICs por mandato associados aos diferentes métodos são consistentes entre si: um governo com melhor performance de acordo com um método tende, em geral, a apresentar boa performance de acordo com os demais métodos.

Tabela 4: Correlação entre ICs por mandato

	<b>IC - média aritmética</b>	<b>IC - DEA</b>	<b>IC - PCA</b>
ICMA	1.000	0.592	0.976
ICDEA	0.592	1.000	0.668
ICPCA	0.976	0.668	1.000

## 3.2 Desempenho Econômico no Brasil: 1997-2022

Nesta seção, avaliaremos a evolução do desempenho econômico do Brasil ao longo dos anos no período 1997-2022. Utilizaremos, para isso, os dois ICs propostos e apresentados no presente trabalho (cf. Seção 3.1). Iniciamos com a comparação entre os indicadores compostos via média aritmética, DEA e PCA (Figura 10). É possível notar que, apesar da diferença de escala do IC obtido via DEA, há certas semelhanças entre os dois ICs aqui produzido (e aquele obtido de acordo com a média aritmética dos indicadores simples). De fato, os três indicam uma piora da condução da política econômica durante o governo FHC, primeiro do penúltimo ano de seu primeiro mandato até o primeiro ano de seu segundo mandato (1999) e, depois, de 2000 até o final de seu segundo mandato (2022). Lula herda uma economia em baixa (pior nível até então observado pelo IC obtido via DEA ou um dos três piores de acordo com os outros ICs) e consegue uma boa performance ao longo de seu mandato – obtendo resultados que se equiparam ou superam os melhores anos de FHC. No segundo mandato de Lula a performance é mantida (IC via DEA) ou piora ligeiramente (outros dois ICs) – mas, ainda assim, o desempenho é melhor do que no governo de FHC. Há de se destacar que a tendência de crescimento observada no IC via PCA (e média aritmética) que se inicia no primeiro mandato de Lula é revertida em uma tendência de queda que se inicia em 2007, o primeiro ano da crise financeira internacional (oriunda da crise do subprime) e que se prolonga até o final do governo Dilma, pelo menos. Não captado pelo IC via DEA, o governo Dilma, aliás, tem performance pior, ano após ano. Na transição e, principalmente, no governo Temer, a economia volta a se recuperar, porém, tal recuperação é interrompida com o governo Bolsonaro. É interessante observar que o IC via DEA aponta que, em alguns anos onde os outros dois ICs sugerem uma performance ruim da economia, talvez os governos não tenham feito um mal trabalho: eventualmente, só mudaram o foco de prioridade da condução de política econômica, associados estes aos diferentes indicadores simples.

Analisamos, em sequência, a condução da política econômica por mandatos. Para isso, calculamos a média aritmética simples dos ICs nos anos associados a cada mandato. A Figura 11 apresenta os ICs baseados na DEA e na PCA (os resultados observados para a média aritmética foram parecidos com os obtidos para a PCA). É possível observar que, independentemente do método utilizado, as melhores performances aparecem associadas ao governo Lula (primeiro e segundo mandatos) e ao governo Temer, assim como as piores performances associadas ao segundo mandato de Dilma. Há também uma piora do primeiro para o segundo mandato de FHC. Como previamente apontado, no IC obtido

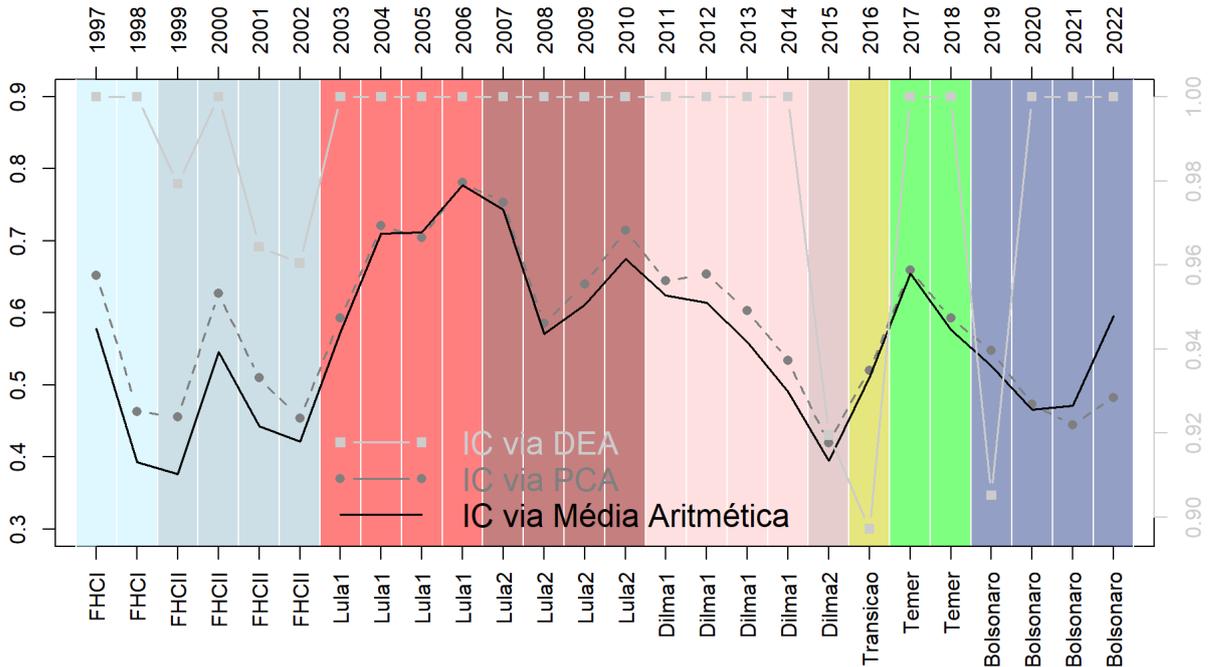


Figura 10: Evolução anual dos ICs (apenas IC via DEA projetado no eixo da direita)

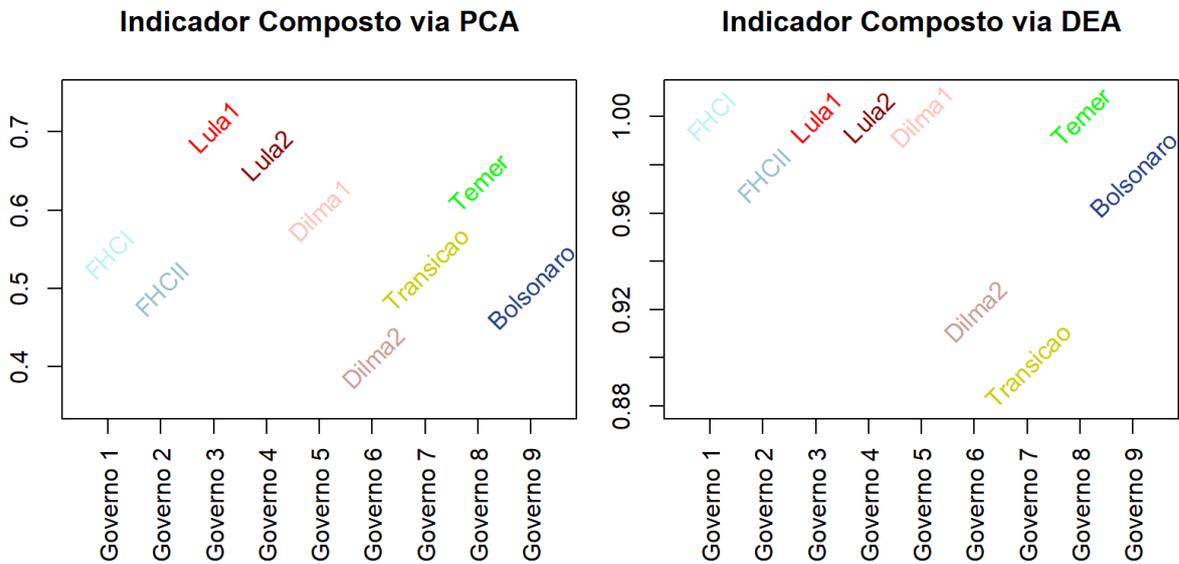


Figura 11: Indicador composto por mandato (médias ao longo dos anos)

via DEA os pesos tendem a favorecer cada governo. De um lado, isto explica a razão de vários mandatos apresentarem performances próximas da máxima possível de acordo com tal método. Por outro lado, ainda assim é possível observar que vários governos deixaram a desejar, obtendo avaliações abaixo da máxima possível (estes são os casos dos mandatos Dilma2 e Transição e, em uma proporção menor, de FHCII e Bolsonaro).

É razoável imaginar que um governo que herda um cenário econômico desfavorável

(deixado por seu antecessor) tenha um desafio maior do que outro que sucede um mandato bem sucedido quanto ao desempenho econômico. Apresentamos na Figura 12 as função de autocorrelação (ACF ou total; e PACF ou parcial) dos ICs obtidos via PCA e DEA. Enquanto no IC baseado na DEA não observamos qualquer relação dependência temporal, encontramos indícios de dependência de primeira ordem e positiva no IC construído a partir da PCA. Ao menos de acordo com este último IC, confirmamos que é mais fácil observar uma boa performance econômica de um governo cujo antecessor obteve êxito.

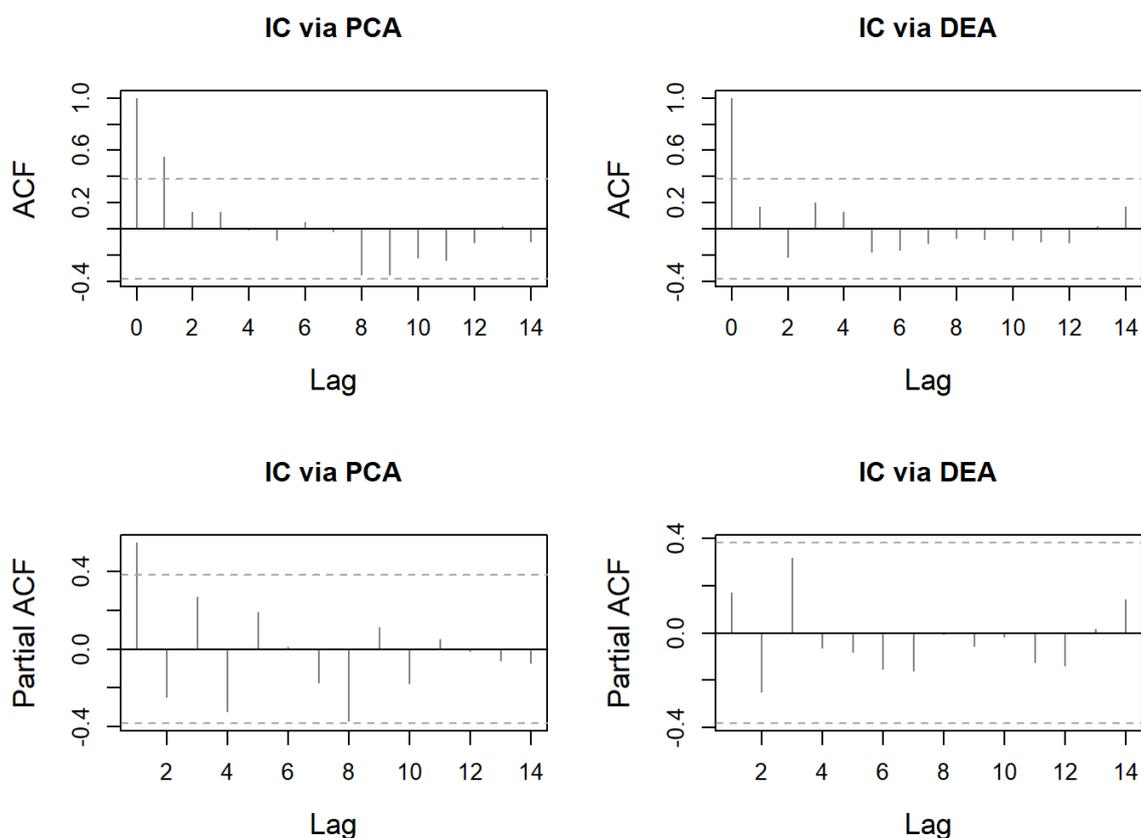


Figura 12: Autocorrelação entre os ICs

Levando em conta a inércia associada ao desempenho econômico, avaliamos a evolução do desempenho de cada governo em relação ao seu antecessor. Para isso, avaliamos a taxa de variação percentual do último ano de cada mandato em relação ao último ano do mandato imediatamente anterior. Tais taxas são apresentadas na Figura 13. Destacamos que no IC obtido via PCA, a melhor evolução foi obtida no primeiro mandato de Lula (crescimento de cerca de 60

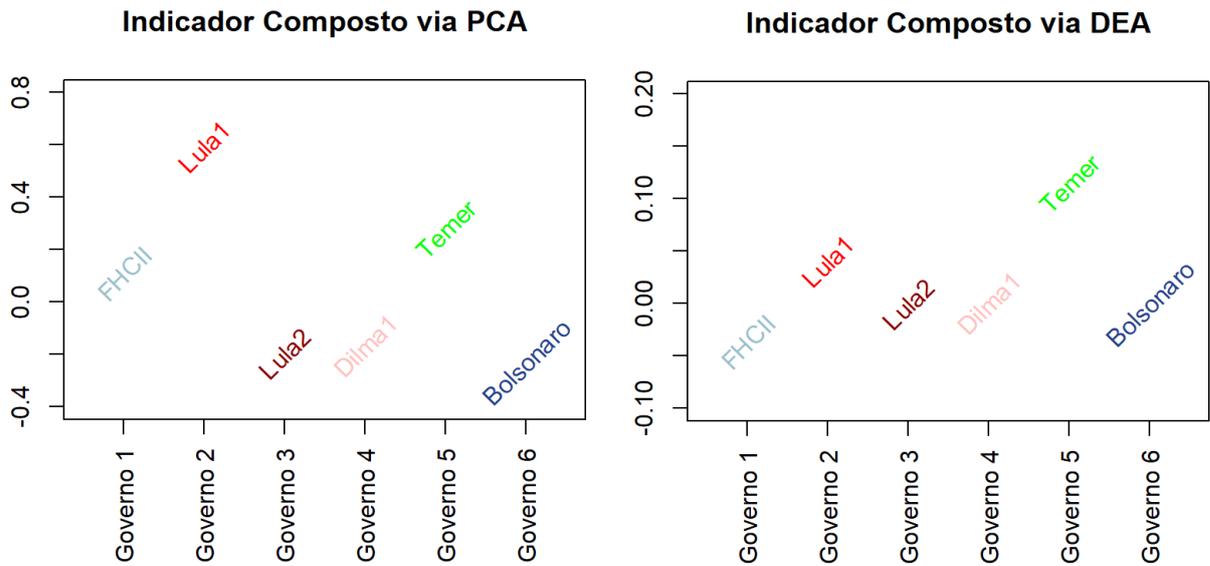


Figura 13: Crescimento percentual em relação à mandato anterior

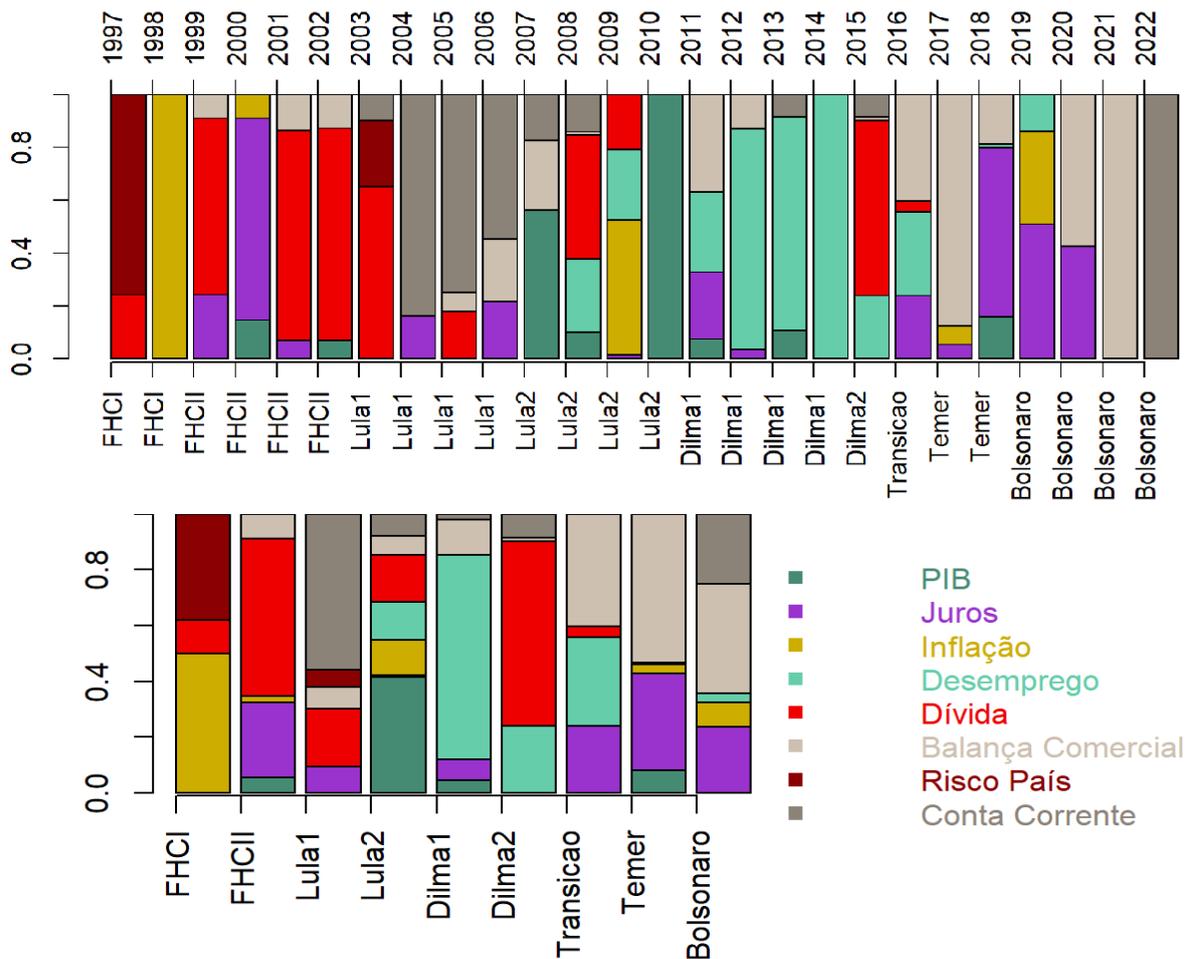


Figura 14: Evolução dos Pesos DEA (por ano e mandato)

Levando em conta ainda a Figura 13, observamos que no IC construído a partir da DEA, as evoluções estão um pouco mais próximas – o destaque positivo fica por conta do governo de Temer (crescimento de pouco mais de 10% em relação ao seu sucessor), ao passo que o negativo é o do segundo mandato de FHC, que apresentou a maior queda em termos relativos.

Como visto anteriormente, os pesos da DEA podem acabar gerando distorções e beneficiando todos os mandatos. Na Figura 14 tem-se os pesos DEA relativizados – para que a soma seja igual a 1 – tanto por ano, como por mandato (considerando, neste último caso, a média dos pesos entre os diferentes anos de cada mandato). A ideia é, inicialmente, identificar quais foram os indicadores simples mais relevantes em cada ano ou mandato e que poderiam refletir as prioridades de cada governo. Em um segundo momento, podemos contrastar tais prioridades reveladas pela DEA com os discursos políticos no momento em que cada governante assumiu a presidência.

A Figura 14 indica que FHC, em seu primeiro mandato, priorizou o risco país e a inflação. O risco país, aliás, só foi prioridade também para o primeiro mandato de Lula. É interessante observar que as ditas “prioridades” também variaram bastante em cada ano de um mandato. No caso de FHC, por exemplo, destacamos que no ano de 1998 o foco esteve quase que exclusivamente no combate à inflação. Uma vez que uma das estratégias usualmente adotada para combater a inflação é a elevação da taxa básica de juros, não nos surpreende que no ano seguinte (1999) o controle dos juros tenha se tornado tão relevante, assim como da dívida pública que é, de certo modo, afetado também por este. O controle da dívida pública foi prioridade nos três anos iniciais do primeiro mandato de Lula que também priorizou o equilíbrio das contas correntes. É interessante, inclusive, observar que na transição de seu primeiro mandato o governo Lula aumentou a importância do controle sobre o saldo da balança comercial que, não por acaso, afeta diretamente a conta corrente. O crescimento econômico (assim como o combate ao desemprego) aparece como prioridades no segundo mandato de Lula. Nenhum outro mandato deu a mesma importância relativa ao crescimento econômico. Na transição de Lula para Dilma, destacamos a inversão das prioridades em relação ao crescimento econômico e o emprego, se tornando este último marca do primeiro mandato de Dilma. A análise de Dilma<sup>2</sup> feita aqui envolve apenas o primeiro ano deste mandato e, portanto, é um tanto quanto insipiente. Contudo, ao analisar a evolução dos pesos entre os governos de Dilma e Temer, percebemos uma mudança brusca de foco: diminuiu consideravelmente a importância relativa do emprego e aumentou substancialmente a importância dos juros e da balança comercial. As novas prioridades foram mantidas durante o governo Bolsonaro.

Todavia, em tal governo o controle dos juros foi prioridade apenas nos dois anos iniciais de seu governo. Da pandemia em diante, o foco nos saldos da balança comercial e, posteriormente, da conta corrente foi concentrando o esforço de suas ações.

A partir daqui, cabe comparar as prioridades reveladas pelos pesos da DEA com o discurso de cada presidente eleito (mais precisamente, descritas em seus programas de governo). Desconsideraremos, todavia, o segundo mandato Dilma (que contempla apenas um ano completo) e o período de transição (de DilmaII para Temer).

O governo de FHC Cardoso (1994) iniciou tendo como objetivo superar o problema da inflação. Para isso, seu plano de governo elegeu cinco metas – emprego, educação, saúde, agricultura e segurança. Antes de tudo era preciso combater a miséria e melhorar a distribuição de renda, nesse sentido haviam metas sobre a educação, desenvolvimento científico e tecnológico, aliado a expansão e modernização da agricultura. O Brasil precisava também garantir a estabilidade do ambiente econômico e implementar um novo modelo de desenvolvimento, para que pudesse ampliar sua capacidade de negociação com os países desenvolvidos e principalmente competir com os mesmos. No âmbito da economia, desejava-se aumentar a capacidade de poupança interna e, a mais audaciosa das metas, executar a terceira fase do Programa de Estabilização Econômica - Plano Real - a moeda que prenuncia o fim da inflação. Retomando os resultados da DEA, é possível observar que de fato FHC cumpriu com seus objetivos iniciais e priorizou a inflação, o risco país e a dívida pública.

A reeleição de FHC Cardoso (1998) estava comprometida em dar continuidade aos projetos dos últimos quatro anos. Além disso, extirpar a miséria, fome e desigualdade social, se tornaram novas prioridades do governo. Essas eram sintetizadas por meio do progresso material e social, cujas metas eram fazer o país crescer de forma sustentada, elevar a escolaridade média do trabalhador brasileiro e gerar oportunidades de emprego. Outro ponto abordado era o desejo de enraizar a democracia e promoção dos Direitos Humanos, nas pautas do país. No âmbito da economia, o Plano Real já havia sido cumprido e estava gerando resultados, principalmente por promover a estabilização econômica e a preservação do poder aquisitivo. Portanto os objetivos passavam a ser o fim da dívida pública em relação ao produto interno, redução no peso dos impostos, reconstruir o sistema de crédito nacional, aumentar as exportações e consolidar o espaço econômico do Mercosul. Pela DEA, as prioridades do governo foram: dívida pública, juros, PIB, balança comercial e Risco País, correspondendo ao projeto inicial da campanha.

O primeiro mandato de Lula Silva (2002) carregava os problemas enfrentados pelo governo anterior tais como fome e miséria. Além disso, a implementação do Plano Real desencadeou grandes dívidas externas que precisavam ser sanadas. Sendo assim, as pautas do governo estavam voltadas para: Reduzir as desigualdades sociais; criação de políticas públicas nas áreas de saúde, educação, previdência social, habitação e distribuição de renda; promover o desenvolvimento tecnológico do país; diminuir a dependência externa; crescimento econômico - focado na agricultura; desprivatização do Estado; estabelecimento de segurança e modernização das Forças Armadas (FFAA). De acordo com a DEA, Lula priorizou: conta corrente, dívida pública, juros, balança comercial e risco país. Sendo assim, o governo foi bom para solucionar pendências com o exterior. Em contrapartida, não obteve resultados satisfatórios na promoção de empregos.

Na reeleição de Lula Silva (2006), os principais objetivos eram: distribuição de renda, desenvolvimento e educação de qualidade. O mandato continuaria se desenvolvendo rumo a erradicação da fome - com os programas Fome Zero e Bolsa Família; avançaria na consolidação do Sistema Único de Saúde (SUS); e permaneceria determinado em garantir luz para todos. Melhorias relacionadas a qualidade nas redes de transporte, no turismo, pesca, apoio à agricultura familiar, o incentivo ao agro-negócio e implantação de uma Reforma Agrária também estavam nas pautas. As propostas relacionadas a economia consistiam em: reduzir as taxas de juros, exterminar as dívidas públicas, ampliar o crédito às empresas produtivas e aos trabalhadores. Nesse sentido, comparando com a Figura - que identifica PIB, dívida pública, inflação, desemprego, conta corrente e balança comercial como prioridades. Logo, o governo foi satisfatório em relação ao seu discurso.

O Governo de Dilma Rousseff estava pautado em 13 pilares, que podem ser vistos em Rousseff (2010), dentre eles é válido destacar: "Crescer mais, com expansão do emprego e da renda, com equilíbrio macroeconômico, sem vulnerabilidade externa e desigualdades regionais- Ampliando os investimentos, a poupança e as conquistas sociais; "Dar seguimento a um projeto nacional de desenvolvimento que assegure grande e sustentável transformação produtiva do Brasil- Com a política industrial (fortemente apoiada na inovação), política agrícola (fortalecendo a agricultura familiar e o agronegócio) e a internacionalização da política energética (sobretudo relacionada com países pobres e em desenvolvimento); "Erradicar a pobreza absoluta e prosseguir reduzindo as desigualdades- mantendo programas como o Bolsa Família; "Promover a igualdade, com garantia de futuro para os setores discriminados na sociedade"; "O Governo Dilma será de todos os brasileiros e brasileiras e dará atenção especial aos trabalhadores"; "Garantir educação para a igualdade social, a cidadania e o desenvolvimento- Expandindo recursos para Pes-

quisas de Desenvolvimento e fornecendo Banda larga nas escolas; "Transformar o Brasil em potência científica e tecnológica"; "Defender a soberania nacional. Por uma presença ativa e altiva do Brasil no mundo". O mandato priorizou - de acordo com a DEA - majoritariamente o desemprego, porém ainda é considerável a balança comercial, juros, PIB e conta corrente. Sendo assim, Dilma foi eficiente na proposta de empregar os brasileiros, estabelecer relações exteriores e controlar os Juros.

O governo de Michel Temer (2015) inicia devido ao Impeachment de Dilma em seu segundo mandato. Tal informação, por si só, carrega o retrato de um país movido por crises e grave recessão econômica. Sendo assim, os objetivos do mandato consistiam em: Combater os déficits elevados e a tendência do endividamento do Estado a fim de retomar o crescimento econômico; Exterminar o desequilíbrio fiscal controlando o aumento da inflação, juros altos, incerteza sobre a evolução da economia, pressão cambial e retração do investimento privado. Reduzir as Taxas de Impostos. Reformar a previdência social. Por um fim em todas as indexações. Interromper o crescimento da Dívida Pública. Criar as condições para o crescimento sustentado da economia brasileira. Viabilizar a participação mais efetiva e predominante do setor privado na construção e operação de infraestrutura. Cooperar junto ao setor privado na abertura dos mercados externos, maior número de alianças ou parcerias - por exemplo com a Ásia e no Atlântico Norte e assim extimular o agronegócio. Pela Figura 14 Temer, tem como principais pesos balança comercial, juros, PIB e inflação. Logo, com excessão das dívidas, o governo priorizou o que havia descrito em sua proposta.

Bolsonaro (2018) possuía como desafios urgentes o fim da criminalidade, recuperar o país na economia e na Democracia. Em relação a camada mais pobre, desejava estabelecer o Programa da Renda Mínima - com valores acima do Bolsa Família. No âmbito do emprego, a Modernização da Legislação Trabalhista, criação da carteira de trabalho verde e amarela e ideias de Reforma na Previdência seriam fundamentais para o futuro. O governo propusera também o aumento da produtividade transformando as relações de trabalho, estimulando a criação de “startups” e a Quarta revolução industrial. O governante tinha como meta equilibrar as contas públicas, buscando um superávit primário que estabilize a relação dívida / PIB. Quebrar o círculo vicioso do crescimento da dívida, adotando práticas auxiliares como o “Orçamento Base Zero” e o corte de privilégios. Por fim, implantar o liberalismo econômico como forma de reduzir a inflação, baixar os juros, elevar a confiança, os investimentos e gerar crescimento. A DEA concluiu que os pesos bolsonaristas foram: balança comercial, juros, conta corrente, inflação e desemprego. Logo, o governo que antes indicava preocupação com as dívidas públicas e desemprego não adotou bom desempenho. Vale ressaltar que a inflação e os juros, que aparecem como prioridades, possivelmente foram motivados pela pandemia do Covid 19.

Em suma, os governos observados no presente trabalho obtiveram pesos relacionados às propostas iniciais de seus mandatos, com exceção de Bolsonaro - que possuía muitas metas relacionadas ao desemprego e ao fim das dívidas públicas porém o maior peso foi em balança comercial e juros . E Dilma, que focou excessivamente no desemprego e não obteve pesos tão consideráveis nas demais variáveis. Cabe ressaltar que a análise feita é superficial e simplória - não sendo avaliado a qualidade dos empregos gerados, ou o quanto os demais indicadores evoluíram. Outra observação é que dimensões paralelas importantes como distribuição de renda, educação e meio-ambiente foram deixadas de fora propositamente nas análises por não fazerem parte dos oito indicadores adotados pelo Main Indicators. <sup>1</sup>

---

<sup>1</sup><https://tradingeconomics.com/indicators>

## 4 Conclusão

A presente monografia buscou explorar de maneira aprofundada a criação dos indicadores compostos, tendo como objetivo avaliar os métodos de ponderação baseados em dados, aplicá-los a um banco sobre a economia e relacionar os resultados obtidos com as propostas dos governos do Brasil de 1997 a 2022.

Durante essa jornada, foi empregado o passo a passo de como criar um indicador composto de acordo com a OCDE. Para isso os métodos: Análise de Componentes Principais e Análise Envoltória de Dados foram escolhidos, definidos e aplicados, gerando dois indicadores compostos que viabilizaram a avaliação e comparação dos governos quanto à condução da política econômica em um sentido mais global.

Os resultados obtidos revelaram importantes insights sobre a relação entre a PCA e a média aritmética, trazendo a ideia que a média, mesmo sendo um método simples, é quase tão eficiente quanto os pesos adotados pela PCA. Já nos resultados obtidos pela DEA foi possível analisar a coerência entre o discurso político no início do mandato e a condução efetiva da política econômica. Além disso, foi descoberto que o método possui limitações quanto ao tamanho do banco de dados - como a base era pequena, a fronteira obtida não foi eficiente para as análises.

Destaca-se que este estudo não se esgota em si mesmo. Pelo contrário, abre portas para possíveis investigações futuras sobre quaisquer indicadores, como por exemplo ambientais ou sociais. Novos estudos focados em indicadores econômicos poderiam complementar este trabalho e proporcionar uma visão ainda mais completa incluindo o uso de variáveis exógenas como o crescimento da economia mundial e o preço das commodities.

Em suma, a presente monografia reafirma a importância de se aprofundar nos estudos sobre indicadores compostos. A partir dos resultados apresentados e das reflexões realizadas, espera-se que este trabalho possa servir como um estímulo para novas pesquisas e debates, fortalecendo assim o conhecimento do tema.

# Referências

- BOLSONARO, J. *O Caminho da Prosperidade*. 2018.
- BOOYSEN, F. An overview and evaluation of composite indices of development. *Social indicators research*, Springer, v. 59, p. 115–151, 2002.
- BOUDT, K. et al. Interpretability of composite indicators based on principal components. *Journal of Probability and Statistics*, Hindawi, v. 2022, 2022.
- CARDOSO, F. H. *Mãos à obra, Brasil: Proposta de governo*. 1994.
- CARDOSO, F. H. *Avança Brasil: Proposta de governo*. 1998.
- CHERCHYE, L. et al. An introduction to ‘benefit of the doubt’ composite indicators. *Social indicators research*, Springer, v. 82, p. 111–145, 2007.
- COELLI, T. et al. *An Introduction to Efficiency and Productivity Analysis*. Springer US, 2005. ISBN 9780387242651. Disponível em: (<https://books.google.com.br/books?id=NMYB0Mh8ljcC>).
- COOPER, W.; SEIFORD, L.; TONE, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer US, 2006. ISBN 9780387452814. Disponível em: (<https://books.google.com.br/books?id=PxyFswEACAAJ>).
- DECANCQ, K.; LUGO, M. A. Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, Taylor & Francis, v. 32, n. 1, p. 7–34, 2013.
- FUSCO, E. Enhancing non-compensatory composite indicators: A directional proposal. *European journal of operational research*, Elsevier, v. 242, n. 2, p. 620–630, 2015.
- GRECO, S. et al. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social indicators research*, Springer, v. 141, p. 61–94, 2019.
- HÄRDLE, W.; SIMAR, L. *Applied Multivariate Statistical Analysis*. [S.l.]: Springer International Publishing, 2008.
- JOHNSON, R.; WICHERN, D. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 2007. (Applied Multivariate Statistical Analysis). ISBN 9780131877153. Disponível em: (<https://books.google.com.br/books?id=gFWcQgAACAAJ>).
- MAZZIOTTA, M.; PARETO, A. Use and misuse of pca for measuring well-being. *Social Indicators Research*, Springer, v. 142, p. 451–476, 2019.

- MORRISON, D. *Multivariate statistical methods*. Thomson/Brooks/Cole, 2005. (Duxbury Advanced Series). Disponível em: <<https://books.google.com.br/books?id=0-TMuQEACAAJ>>.
- RAM, R. Composite indices of physical quality of life, basic needs fulfilment, and income: A ‘principal component’ representation. *Journal of Development Economics*, Elsevier, v. 11, n. 2, p. 227–247, 1982.
- ROGGE, N. Composite indicators as generalized benefit-of-the-doubt weighted averages. *European Journal of Operational Research*, Elsevier, v. 267, n. 1, p. 381–392, 2018.
- ROUSSEFF, D. *Os 13 compromissos programáticos de Dilma Rousseff para debate na sociedade brasileira*. 2010.
- SILVA, L. I. L. da. *Programa de Governo 2002*. 2002.
- SILVA, L. I. L. da. *Programa de Governo 2006*. 2006.
- SLOTTJE, D. J. Measuring the quality of life across countries. *The Review of economics and statistics*, JSTOR, p. 684–693, 1991.
- TEMER, M. *Uma ponte para o futuro Fundação Ulysses Guimarães*. 2015.
- UNION, E.; COMMISSION, J. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, 2008. ISBN 9789264043466. Disponível em: <<https://books.google.com.br/books?id=N-jVAgAAQBAJ>>.