

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO INSTITUTO DE ECONOMIA PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

CAROLINA TAVARES RESENDE

ECONOMIC EFFICIENCY AND POLICYMAKING: FROM THEORY TO PRACTICE

RIO DE JANEIRO 2025 Carolina Tavares Resende

ECONOMIC EFFICIENCY AND POLICYMAKING: FROM THEORY TO PRACTICE

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Rio de Janeiro, como requisito à obtenção do título de Doutora em Ciências Econômicas.

Orientadora: Prof^a. Dr^a. Esther Dweck Co-orientador: Prof. Dr. Carlos Pinkusfeld Monteiro Bastos

Rio de Janeiro 2025

FICHA CATALOGRÁFICA

R433e	Resende, Carolina Tavares. Economic efficiency and policymaking: from theory to practice / Carolina Tavares Resende. – 2025. 154 f.
	Orientadora: Esther Dweck. Coorientador: Carlos Pinkusfeld Monteiro Bastos. Tese (doutorado) – Universidade Federal do Rio de Janeiro, Instituto de Economia, Programa de Pós-Graduação em Economia da Indústria e da Tecnologia, 2025. Bibliografía: f. 147 – 154.
	 Políticas públicas. 2. Eficiência. 3. Bem-estar social. I. Dweck, Esther, orient. II. Bastos, Carlos Pinkusfeld Monteiro, coorient. III. Universidade Federal do Rio de Janeiro. Instituto de Economia. IV. Título.
	CDD 320.6

Ficha catalográfica elaborada pela bibliotecária: Luiza Hiromi Arao CRB 7 – 6787 Biblioteca Eugênio Gudin/CCJE/UFRJ Carolina Tavares Resende

ECONOMIC EFFICIENCY AND POLICYMAKING: FROM THEORY TO PRACTICE

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Rio de Janeiro, como requisito à obtenção do título de Doutora em Ciências Econômicas.

Rio de Janeiro, 25 de abril de 2025.

Prof^a. Dr^a. Esther Dweck – Presidente Universidade Federal do Rio de Janeiro

Prof. Dr. Carlor Pinkusfeld Monteiro Bastos – Co-orientador Universidade Federal do Rio de Janeiro

Prof. Dr. Fábio Neves Perácio de Freitas – Membro interno Universidade Federal do Rio de Janeiro

Prof. Dr. Norberto Montani Martins – Membro interno Universidade Federal do Rio de Janeiro Prof. Dr. Francisco Luiz Cazeiro Lopreato – Membro externo Universidade Estadual de Campinas

Dr. Gabriel Ferraz Aidar – Membro externo Banco Nacional de Desenvolvimento Econômico e Social (BNDES)

AGRADECIMENTOS

Fazer o doutorado na UFRJ foi uma decisão acertada. A instituição reúne um corpo docente excelente, com quem tive a oportunidade de aprender muito. Para tanto, o apoio da Câmara dos Deputados, por meio da concessão de uma licença, foi fundamental. Fazer esta pósgraduação em dedicação exclusiva certamente fez toda diferença.

Como não realizamos nada sozinhos, esse é um ótimo espaço para agradecer a tantas pessoas queridas que fizeram parte da minha jornada. Primeiramente, gostaria de agradecer à minha orientadora, Prof^a Esther Dweck, por sua parceria durante todo o processo. Meu desejo desde antes do ingresso no doutorado era ser orientada por ela, que, apesar de ter aceito novos e intensos desafios, seguiu comigo nesta jornada. De suas aulas às reuniões de orientação, Esther sempre nos presenteia com sua generosidade intelectual, paciência e tratamento atencioso. Não menos importante foi a presença do meu co-orientador, Prof. Carlos Pinkusfeld, que esteve muito presente e com quem tive muitas trocas valiosas, sempre acompanhadas de bom-humor.

Agradeço o apoio incondicional da minha família, em especial minha mãe, meu pai e minha irmã, que sempre foram suporte em tudo que já realizei. Ao meu companheiro, com quem divido a vida, os sonhos e as lutas diárias. Ao seu lado, tudo parece possível. Aos meus bebês amados. Otto que me acompanhou no ventre na qualificação e ao segundo que está por vir e me acompanhará na defesa. Aos meus sogros e minha cunhada, que formam uma rede de apoio tão importante pra nós.

Aos amigos queridos pelas trocas que trazem leveza à vida. Aos colegas de curso e do GESP pelos valiosos debates. Em especial, ao Bruno, um presente que esse doutorado me deu. Obrigada, meu amigo, pelas infinitas conversas, botecos e doses diárias de gargalhadas. Não teria sido o mesmo sem você. À minha amiga Nádia, grande incentivadora de tudo que me proponho a fazer.

RESUMO

Esta tese discute a relação entre eficiência econômica e políticas públicas da teoria à prática. No primeiro capítulo são discutidos os suportes filosófico e teórico para a definição da eficiência como critério de bem-estar social. Primeiramente, trata-se da influência do positivismo na determinação do que é conhecimento científico na ciência econômica, evidenciando os efeitos desse paradigma nos processo de avaliação de política. Também no primeiro capítulo discute-se a evolução da teoria neoclássica que culminou na determinação do conceito de eficiência como critério central de bem-estar. Por fim, o primeiro capítulo trata das implicações de ter a eficiência como benchmark das políticas públicas. O segundo capítulo faz a conexão entre o campo teórico e a economia aplicada, discutindo as principais metodologias utilizadas para estimar a eficiência das políticas públicas. Nesse sentido, o capítulo faz um debate sobre os elementos internos dos modelos utilizados para quantificar eficiência técnica e alocativa, evidenciando a falta de transparência de hipóteses dos modelos, a imprecisão das estimativas e os juízos de valor implícitos, muito embora essas metodologias produzam respostas binárias e supostamente estritamente técnicas. O terceiro capítulo faz um estudo de caso com aplicações da vida real desses modelos discutidos anteriormente. Para tanto, foram selecionados estudos que ilustram os pontos discutidos no capítulo anterior, deixando ainda mais claro que o processo de avaliação de políticas públicas não pode prescindir de juízos políticos e é influenciado por visões de mundo particulares. Por fim, o terceiro capítulo traz exemplos de metodologias alternativas, que não se comprometem com o princípio da eficiência como critério central de bem-estar social. Com isso, mostra-se que, em uma perspectiva econômica distinta, outros benchmark de políticas públicas são não só possíveis como desejáveis, além de responderem primordialmente a interesses coletivos.

Palavras-chave: bem-estar social, eficiência alocativa, eficiência técnica, políticas públicas, gasto público.

ABSTRACT

This thesis discusses the relationship between economic efficiency and public policies, from theory to practice. The first chapter addresses the philosophical and theoretical foundations for defining efficiency as a social welfare criterion. First, it explores the influence of positivism in determining what constitutes scientific knowledge in economics, highlighting the effects of this paradigm on policy evaluation processes. The chapter also discusses the evolution of neoclassical theory, which led to establishing efficiency as a central welfare criterion. Finally, the first chapter examines the implications of using efficiency as a benchmark for policymaking. The second chapter bridges the gap between theoretical frameworks and applied economics by discussing the most relevant methodologies used to estimate the efficiency of public policies. In this context, the chapter debates the internal elements of models employed to quantify technical and allocative efficiency, exposing the lack of transparency in the models' assumptions, the inaccuracy of estimates, and the implicit value judgments, even though these methodologies produce binary and supposedly strictly technical answers. The final chapter presents a case study with real-life applications of previously discussed models. To this end, studies were selected to illustrate the points discussed in the prior chapter, making it even clearer that the process of evaluating public policies is inseparable from political judgments and is influenced by particular worldviews. Lastly, the third chapter introduces examples of alternative methodologies that do not commit to the principle of efficiency as the central criterion of social welfare. Consequently, from a different economic perspective, other benchmarks for public policies are not only possible but desirable, as they primarily respond to collective interests.

Key words: social welfare, allocative efficiency, technical efficiency, policymaking, public spending.

LIST OF TABLES

Table 1 - Various Positions in the Philosophy of Science	18
Table 2 - Principles that guide positivist research	20
Table 3 - A Summary of Market Failures and Their Implications for Efficiency	
Table 4 - Summary of analytical framework for spending reviews	51
Table 5 - The major steps in CBA	60
Table 6 - Link between property rights and WTP and WTA	69
Table 7 - Selected overview of discounting guidance across countries	78
Table 8 - A taxonomy of frontier methods	92
Table 9 - DEA model assumptions	94
Table 10 - International Income-Adjusted Estimates of the VSL - selected countries	102
Table 11 - Examples of valuation studies using differing valuation techniques	108
Table 12 - Overview of different discounting approaches	120
Table 13 - Discounted health outcomes of HPV vaccination using the different discoun	ting
approaches	121
Table 14 - Methodological steps from ROA	140

LIST OF FIGURES

Figure 1: Pareto x Potential Pareto Efficiency
Figure 2 - The Scitovsky Paradox
Figure 3 - A Classification of Goods: Private and Public
Figure 4 - Main objectives of Spending Reviews over the past 10 years
Figure 5 - Valuation methods for non-market goods
Figure 6 - Demand curve for a hypothetical greening project
Figure 7 - Social Discount Rates (in %) by country and approach
Figure 8 - Size of Government and Performance Indicators in Industrial Countries, 1990
(Percentage of GDP, unless otherwise indicated)
Figure 9 - Summary of monetary values for each service by biome - International dollars
per hectare per year, 2007 price level 106
E : 10 D 1 $(1 + 1 + 1 + 1)$ 1 $(1 + 1 + 1)$ 1 $(1 + 1)$ 1 $(1 + 1)$
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices 107 Figure 11 - The VSL by Age Group (in 2019 millions of dollars) 111 Figure 12 - Effect of Alternative Approaches on Analytic Results 112 Figure 13 - Top 10 countries with VSL savings 112 Figure 14 - Temporal and generational timing of benefits from four vaccines 119 Figure 15 - Input and Output categories for hospitals
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices 107 Figure 11 - The VSL by Age Group (in 2019 millions of dollars) 111 Figure 12 - Effect of Alternative Approaches on Analytic Results 112 Figure 13 - Top 10 countries with VSL savings 112 Figure 14 - Temporal and generational timing of benefits from four vaccines 119 Figure 15 - Input and Output categories for hospitals 125 Figure 16 - Distribution, by income quintile, of deaths averted and cases of medical
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices 107 Figure 11 - The VSL by Age Group (in 2019 millions of dollars) Figure 12 - Effect of Alternative Approaches on Analytic Results Figure 13 - Top 10 countries with VSL savings Figure 14 - Temporal and generational timing of benefits from four vaccines Figure 15 - Input and Output categories for hospitals Figure 16 - Distribution, by income quintile, of deaths averted and cases of medical impoverishment averted by vaccines to be administered in 41 low- and middle-income
Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices 107 Figure 11 - The VSL by Age Group (in 2019 millions of dollars) Figure 12 - Effect of Alternative Approaches on Analytic Results 112 Figure 13 - Top 10 countries with VSL savings 112 Figure 14 - Temporal and generational timing of benefits from four vaccines 119 Figure 15 - Input and Output categories for hospitals 125 Figure 16 - Distribution, by income quintile, of deaths averted and cases of medical impoverishment averted by vaccines to be administered in 41 low- and middle-income countries, 2016–30.

CONTENTS

INTRODUCTION	12
1. EFFICIENCY AND POLICYMAKING: GROUNDS AND IMPLICATIONS	14
1.1 Positivism: replacing human judgment	17
1.2 Positivism and policymaking	22
1.3 The changing nature of neoclassical efficiency: Theoretical foundations	25
1.3.1 The marginalist and the ordinalist revolutions: the rise of Pareto's efficiency as welfare criterion	a 25
1.3.2 Kaldor-Hicks efficiency and hypothetical compensations: a flexible concept	29
1.3.3 Challenges to neoclassical efficiency	34
1.4 Market and efficiency: Benchmarks for Policymaking	37
1.4.1 Limits on circumstances and instruments for policymaking	37
1.4.2 Enforcement of a way of thinking	44
1.4.3 Potential Pareto efficiency as a sufficient condition for welfare improvement	46
1.4.4 Policy financing: efficiency and Spending Reviews	48
Final remarks	53
2. EFFICIENCY AND METHODOLOGY: FROM THEORY TO PRACTICE	55
2.1 Allocative Efficiency: The Cost-Benefit of Public Funding	56
2.1.1 The basic framework	58
2.1.2 Valuation: assigning monetary value to non-market goods	63
2.1.3 Dealing with time: intertemporal discount	72
2.1.4 Distributive concerns	80
2.2 Technical efficiency: seeking value for money	85
2.2.1 Performance Indicators	88
2.2.2 Data Envelopment Analysis (DEA) and Stochastic Frontier Models (SFA)	91
Final remarks	97
3. APPLIED EFFICIENCY: IMPLICATIONS TO POLICYMAKING	98
3.1 Allocative efficiency in practice	99
3.1.1 Is WTP a proxy of welfare?	100
3.1.2 The Valuation Process	104
3.1.3 Discounting: discussing inter-generational effects	117
3.2 Technical efficiency in practice	122

REFERENCES	147
CONCLUSION	142
Final remarks	141
3.4 Alternative methodologies	135
3.3 Political and Collective Interests	131
3.2.2 The relevance of context and interpretation	125
3.2.1 How Internal Methodological Choices Affect Results	123

INTRODUCTION

Neoclassical welfare economics establishes economic efficiency as the benchmark for resource allocation. In other words, the theory defines efficiency as the welfare criterion. Consequently, to enhance social welfare, government actions must be deemed efficient. Embracing this perspective has significant effects, which are often unclear to society.

Acknowledging the neoclassical approach raises several questions that encompass interesting debates. What implicit distributive judgement does defining allocative efficiency as the welfare criterion entail? What political space is opened for policymaking? Which tools can be used to assess whether a policy is efficient? What do the internal elements of the methodologies reveal about the expected performance of government actions? What are the practical implications of adopting economic efficiency as a guiding principle for policymaking? Are there alternative methodologies that engage with a different approach to economic theory?

This work addresses these questions through a comprehensive critical review of the relationship between efficiency and policymaking. The revision covers the theoretical and philosophical supports for electing efficiency as a guiding principle, going through the most common methodologies employed to estimate it. Finally, it assesses applied cases to discuss practical implications for society.

In this sense, Chapter One begins by emphasising the relevance of the positivist paradigm in defining what science means to economics due to the direct impact this discussion has on both theory and the methodological framework that connects efficiency and policymaking. It also critically reviews economic theory literature, focusing on the evolution that culminated in determining economic efficiency as the welfare criterion to guide policymaking and changes that had to be made to accommodate theory with applied economics. Moreover, Chapter One assesses the implications for policymaking derived from the centrality of markets and the definition of economic efficiency as the benchmark for government actions.

Chapter Two connects efficiency to the most commonly employed allocative and technical efficiency methodologies. It links theory to practice by critically discussing these methods' assumptions, implicit value judgments, and controversial internal elements in detail. This chapter aims to uncover the specific worldviews considered technical components of the analysis, demonstrating that every evaluation carries political valuations, even if some are made implicitly.

More specifically, Chapter Two reviews the internal methodological elements that compose the Cost-Benefit Analysis (CBA), which is the most popular methodology for assessing allocative efficiency, and the Data Envelopment Analysis (DEA) and Stochastic Frontier Models (SFA), which are commonly employed to estimate technical efficiency.

Concerning the CBA, it presents a) the basic framework; b) the controversial valuation process, where monetary values are assigned to non-market goods (externalities); c) the social discount rate applied to evaluate projects with different maturities; and d) the distributive concerns related to the methodology. Regarding technical efficiency, it evaluates how internal choices strongly influence results and the relevance of context and interpretation in analysing public policies from the perspective of achieving value for money.

Chapter Three presents a case study on applied efficiency, demonstrating the practical implications of each methodology discussed. It provides an illustrative, though not exhaustive, list of examples to address selected internal elements of the methods covered in Chapter Two. More specifically, regarding the CBA, it a) exhibits the real-life effects of using the principle of willingness to pay as a proxy for welfare; b) highlights inaccuracies and controversies surrounding practical examples in the valuation process; c) reveals how CBA overlooks crucial intergenerational concerns through real cases.

Concerning technical efficiency, it includes studies that demonstrate how sensitive DEA, the most widely used method, is to variable selection. Additionally, it emphasises the relevance of context in interpreting results, challenging the common misconception that every inefficiency presents an opportunity for cutting spending.

In both allocative and technical efficiency, since methodologies work on a binary rationale, they should be founded on widely accepted assumptions, present accurate estimates, and provide transparent outcomes. Instead, Chapter Three reveals something quite different through applied studies. In fact, the case study illustrates how politics and collective interests cannot be separated from policy analysis, casting a shadow over the notion of impartial evaluations.

Finally, Chapter Three offers a preview of alternative methodologies for policy analysis that are not committed to any specific benchmark. Impact Analysis and Risk-Opportunity Analysis support decision-making without restricting themselves to neoclassical efficiency as a welfare criterion. Instead, these methods can be connected to any collectively or politically decided benchmark, allowing for broader definitions of social welfare. With this discussion, the chapter briefly presents some possible alternative paths to policy analysis.

This work aims to contribute to and enrich the academic literature with originality as it is not restricted to the literature of the economic field but also resorts to philosophy, anthropology, and public management. This feature is significant because debates over social welfare must be interdisciplinary.

Even though it does not exhaust the debate over the implications of electing efficiency as the welfare criterion, this work provides a thorough discussion that ranges from philosophical and theoretical support, spans methodologies connecting theory to practice, and concludes with a discussion of the practical implications of determining efficiency as a guiding principle for policymaking by presenting a case study of applications.

Lastly, it is important to highlight the relevance of this debate within the context of widespread agreement on the need to reduce spending. The use of efficiency as a synonym for the quality of public spending has bolstered arguments for cuts, which explains why a thorough understanding of the implicit assumptions, value judgments, internal elements, and implications of employing efficiency as a welfare criterion is essential.

1. EFFICIENCY AND POLICYMAKING: GROUNDS AND IMPLICATIONS

Efficiency is a well-accepted concept that guides policymaking in contemporary society. However, far from having a single, specific definition, efficiency encompasses many meanings. One reason this principle is so popular as a guideline for public actions is its association with avoiding waste. Although this idea may closely align with the interpretation of technical efficiency, it does not encapsulate the concept of efficiency in economic theory. In economics, efficiency pertains to the allocation of resources within society. The rationale is straightforward: the market is primarily responsible for all social transactions, while the State provides goods and services only when there is an economic justification or when, for some reason, such a market is unable to coordinate or establish prices for all transactions.

Understanding the nature of economic efficiency, the foundations of this reasoning, and the implications for collective action are essential for comprehending the political space available for policymaking. In light of this reflection, this chapter discusses three dimensions of how efficiency and policymaking are correlated. The first element of this Chapter addresses the philosophical foundations that connect positivism, efficiency, and policymaking. More specifically, it discusses how the positivist paradigm remains robust in shaping what Science entails. This recognition is notably reflected in the assessments and evaluations of public policies, resulting in the dominance of quantitative methods that aim to provide objective and unbiased results, supplanting human judgment. Analytical models offer binary answers – true or false, good or bad, efficient or inefficient – a crucial aspect of the validation challenge within the positivist paradigm.

Conversely, a postpositivist perspective on science acknowledges the impossibility of entirely objective evaluations, which must consider historical, contextual, and subjective elements of scientific knowledge. By recognising these factors, the postpositivist paradigm paves the way for diverse conclusions, benchmarks, principles, and methods in policymaking.

Secondly, it reviews the theoretical foundations of efficiency in neoclassical economics, exploring how alterations to the original concept of Pareto's efficiency were essential for enabling public interventions in applied economics. The Kaldor-Hicks efficiency introduced a concept of hypothetical compensations that permitted distributional changes in policymaking. Furthermore, a transition from utility to wealth in welfare analysis became crucial for applied economics. Prior to this, the theory struggled to support efficient state interventions that did not adversely affect at least one individual, which was the essence of Pareto's distributional judgment.

Adaptations of neoclassical economic efficiency served various purposes in applied economics. Firstly, they enabled policymakers to rely on the widely accepted principle of efficiency to justify public actions. Furthermore, these adaptations transformed efficiency into a more adaptable concept that functioned as a welfare criterion in practice. In doing so, they facilitated a more active State without accommodating other guiding principles in policymaking, such as equity, as prioritising these principles risked efficiency losses. Consequently, the second section of Chapter One explores the theoretical developments that established the connection between economic theory and applied economics despite the persistent inconsistencies observed.

The third section discusses how having the market and the concept of efficiency as benchmarks affects the political space of policymaking. First, it debates the theoretical limits imposed on government interventions, from circumstances to instruments allowed. Two movements were perceived. The first, associated with the changing notion of Pareto's efficiency in applied economics, expanded the range of possibilities for government interventions. So, the previous corrective market failure perspective was replaced by a 'hands-on' State with a broader space for action. Conversely, the second limited the way of making such interventions because policymaking became more attached to market mechanisms.

Secondly, the third section argues that the blossoming of what became known as the 'economic way of thinking' enforced a specific line of reasoning for public action that disregards other perspectives to address social problems. When analysing public policies, economic reasoning replaces principles such as universalism, equity and rights for competition, incentives, and choices. With this, the scope of analysis and the treatment of social problems have changed, framing public action through a market lens. For instance, pollution has become an externality to be priced rather than a condemnable external effect to be addressed in environmental policymaking. Another example is research on antipoverty measures that examines how income transfers influence an individual's work or family decisions instead of evaluating the impacts of such transfers on achieving a particular standard of living.

Third, section three discusses how defining potential Pareto efficiency as a sufficient and greater condition for social improvement also significantly impacts the political space for policymaking. First, it equates this specific and controversial concept of efficiency with social welfare. Moreover, it reduces society's measure of well-being to income.

However, a growing body of literature provides evidence that raises serious doubt about the adequacy of efficiency as a guiding principle for assessing welfare. In fact, evidence shows that income is an essential variable in welfare analysis. Nevertheless, other elements such as access to social security net protection, quality jobs and marital status are equally relevant. Recognising this evidence is vital as it refutes the widespread notion that efficiency is either a sufficient or more important guiding principle for public policy.

Lastly, the third section examines the link between evaluations and the budgeting process, highlighting how selecting efficiency as the standard for policymaking influences the funding of public policies in contemporary society. The Spending Reviews framework merges efficiency analysis with the government's budget, serving as a crucial tool for implementing efficiency. This framework directly impacts the financing of public policy.

Therefore, Spending Reviews reflect the end of welfare economics regarding the role of the government in market economies. They are the last step from theory to practice in how neoclassical economics determines the appropriate state actions. Understanding the connection between the elements of this Chapter is of great importance since it affects how society decides to dispose of resources.

1.1 Positivism: replacing human judgment

Economics is a scientific field that remains closely tied to the positivist paradigm. Consequently, quantitative methods are often deemed indispensable and superior to alternative approaches for generating valuable information. This belief is linked to the claimed objectivity afforded by analytical models and the widely accepted notion that scientific knowledge is produced independently of the researcher.

This line of reasoning often seeks to delegate the decision-making process, privileging the binary answers provided by analytical models: true or false, efficient or inefficient, and so on. This approach to determining scientific knowledge and, consequently, valuable information links positivism with policymaking, as the paradigm significantly influences assessments and evaluations. A relevant consequence of this process is the substitution of human judgment in evaluating the merits of a public policy.

Therefore, understanding the philosophical debate that underpins the paradigm dispute in the Philosophy of Science is important. Without this, one is limited to critiquing specific methodologies from within. Conversely, embracing a postpositivist perspective of science allows for the acceptance of diverse sources of scientific information, which is valuable when considering policymaking, assessments, and evaluations.

The validation problem concerns the approval of the methodological paths taken by a researcher. It constitutes the socially accepted standards that provide proof to some arguments. Epistemically, analytical models are instruments to support scientific theories. That is, they represent methodological choices to produce scientific evidence. However, as Kleindorfer, O'Neill, and Ganeshan (1998) pointed out, it is not possible to follow the validation problem in estimates without connecting it with the underlying features of the Philosophy of Science. The main reason is that analytical models are methodological decisions that embody personal or shared beliefs about epistemology. In other words, what people believe about the nature of scientific knowledge is a determinant of the validation problem.

Therefore, confronting the subsistent paradigms associated with a particular method is relevant since there is no value-free analysis. The methodology represents just one aspect of a broader framework within the Philosophy of Science, linking the general epistemological focus with validation approaches based on specific criteria. Recognising this framework is crucial for transcending the notion that there is only one pathway to obtaining valuable knowledge. As Caldwell (1980) highlighted, "*the philosophy of science can be a useful tool for clarifying one's understanding of methodological issues in economics*" (p. 64).

This perspective holds significant relevance for public policy evaluation. As Guba and Lincoln (1989) noted, paradigms shape our worldview and greatly influence how we approach political matters, define ethics, and ensure social justice. Therefore, it is impossible to separate methodology, epistemology, and public policy appraisals and evaluations.

Ramacciotti and Bernardino (2020) discuss the most significant differences between positivist (rationalist) and postpositivist (constructivist, interpretivist, argumentative, poststructuralist, etc.) perspectives, which arise from the epistemology of science. That is, from the nature of scientific knowledge. They argue that comprehending this epistemological debate and the underlying assumptions of each branch enables us to select our methodology more clearly. Furthermore, the evolution of this debate should assist us in overcoming closed and dogmatic positions, acknowledging that methodological pluralism and often complementary research strategies are essential for achieving effective results.

In discussing the validation problem, Kleindorfer, O'Neill, and Ganeshan (1998) summarise various positions in the Philosophy of Science, which we shall associate with their underlying paradigms to better understand the nature of scientific knowledge according to each position.

Positions in the Philosophy of Science	General Epistemological Focus	Criterion of the Philosophy	Representative Philosophers	Validation Approaches
Rationalism		Logical reduction	Descartes	Derived from rational foundation
Classical Empiricism	Logical justification of knowledge claims	Inductive generalization	J. S. Mill J. N. Keynes	Induced from empirical data
Logical Positivism		Empirical verification	Carnap, Russell Wittgenstein	Derived from empirical foundation

Table 1 - Various Positions in the Philosophy of Science

	Instrumentalism	Theories as frameworks for prediction and testing	Predictive success, simplicity, or other aesthetic value	Pierce Friedman	Shown by predictive accuracy, simplicity, or other value
	Dogmatic Falsificationism		"theory-free" observations to test theories	Popper	Continued testing to eliminate faulty models
	Methodological Falsificationism		Survival of testing and criticism	Lakatos' version of Popper	Shown by testing and criticism
	Bayesianism	Consistent treatment of probabilistic induction	Increase subjective probability	Howson Urbach	Empirical success increasing belief
		Progressive			
	Kuhnianism	Progressive	Growth of knowledge through Paradigm shifts	Kuhn, Polyani Bohm Weimer	Accordance with expert opinion, professional acceptance
1	Kuhnianism Lakatos' MSRP	Progressive historical growth of knowledge	Growth of knowledge through Paradigm shifts Growth of knowledge through Research Programmes	Kuhn, Polyani Bohm Weimer Popper, Lakatos Bartley, Agassi	Accordance with expert opinion, professional acceptance Increase empirical and theoretical content without ad hoc adjustment

Source: Kleindorfer, O'Neill and Ganeshan (1998), p. 1090.

Essentially, much of the literature on analytical models consists of the objective/foundationalist and the relativist/anti-foundationalist approaches to the nature of science, with some less strict positions in between. The authors emphasise that the term 'validation' itself carries a philosophical commitment, with the result granting some form of truth. Thus, this is an objective perspective of analytical models. The primary difference between the two approaches lies in the capacity to make inferences from model estimates, given our limited set of perspectives and experiences. Is it possible for anyone to simulate capturing the essential parameters and structure?

The foundationalists believe that if models are validated, they can significantly represent reality. Theories or models can achieve an ultimate and neutral foundation if they are 'resolvable'. For most dogmatic empiricists and rationalists, human judgment is excluded from the process. Rationalism, Classical Empiricism, and Logical Positivism fall into this category. The underlying paradigm of objectivists is positivism, which supposedly provides unequivocal scientific knowledge through selected methodologies.

For positivists, knowledge holds value to the extent that it can be verified, which implies testability. This criterion has evolved over time, with the initial dependence on complete verification through observational evidence becoming more flexible. Nonetheless, the fundamental emphasis on empirical evidence has endured (Caldwell, 1980). Positivism is

fundamentally a hypothetical-deductive model that involves verifying assumptions through the manipulation of variables. Methodologies grounded in the positivist paradigm concentrate on identifying correlations or inferring causality between variables (Park, Konge, and Artino, 2020). The authors outline positivist principles that continue to guide much contemporary research.

	· · · · · · · · · · · · · · · · · · ·
Goals of Science	Social and natural sciences should focus on the discovery of laws that facilitate explanation and prediction.
Methodology	Social and natural sciences should use the same methodology based on the hypothetico-deductive model of Science (theory, hypothesis, operationalization, experimentation).
Sampling and inference	Larger samples are favourable over smaller, idiosyncratic samples; larger samples reveal generalizable tendencies, causes, and the nature of reality.

Table 2 - Principles that guide positivist research

Source: Park, Konge, and Artino (2020). Adapted.

Caldwell (1980) highlights a relevant consequence of the positivist paradigm: "*A* corollary of that view, which further ensures the legitimacy of our explanations, is that explanation and prediction are logically symmetrical, the only difference between them being temporal" (p. 60). In this sense, analytical models can objectively explain reality and possess predictive power. Consequently, they ultimately shape the nature of truth and guide us towards the future.

Another relevant feature of objectivism, and thus positivism, is the idea that one can separate the subject from the object in the sense that reality is independent of us (Bernstein, 1983). This element reflects the belief that the production and verification of scientific knowledge are independent of the researcher's background.

According to Barlas and Carpenter (1990), Thomas Kuhn's work was essential in challenging the traditional concept of objectivity in the philosophy of science. By presenting the historical context of scientific progress, he argues that the evolution of scientific knowledge is accompanied by new paradigms. And so "*The perspective, the methods and rules to be followed, and even the norms of rationality are restated*" (p. 156). But not only the objective perspective is denied, as Bernstein (1983) points out:

In its strongest form, relativism is the basic conviction that when we turn to the examination of those concepts that philosophers have taken to be the most fundamental -- whether it is the concept of rationality, truth, reality, right, the good, or norms -- we are forced to recognize that in the final analysis all such concepts must

be understood as relative to a specific conceptual scheme, theoretical framework, paradigm, form of life, society, or culture (p. 8).

Following the 1960s, the ahistorical and acultural Analysis became incompatible with scientific inquiry, paving the way for the emergence of relativist philosophies. The validation problem within the relativistic approach is semiformal and gradual. It relies on discussions and social acceptance, as estimates represent merely one method of analysing a situation. Also, "*No particular representation is superior to all others in any absolute sense, although one could prove to be more effective*" (Barlas and Carpenter, 1990, p. 157). Moreover, the impossibility of neutral observation ensures the connection between the scientist and the theory. Nevertheless, this connection could render any model equally valid or invalid, which is subjective (Kleindorfer, O'Neill and Ganeshan, 1998).

According to Kleindorfer, O'Neill, and Ganeshan (1998), Instrumentalism, Falsificationism, and Kuhnianism occupy a grey area, fluctuating between objectivist and relativist approaches. For instrumentalists like Milton Friedman, positive economics play a role in organising empirical evidence. The general structure of models includes a conventionalist element, with assumptions governing empirical observations. The aim is to harness the predictive power of analytical models, even if the hypothesis contradicts reality. In summary, the conventionalist theoretical component, combined with the objectivist empirical domain where theory resides, illustrates, as the authors suggest, the fluctuating philosophical positions of instrumentalists. Bernstein (1983) shares a similar perspective, viewing Kuhnianism as occupying a grey area between objectivism and relativism, as Kuhn employed positivist arguments, a paradigm associated with objectivism, along with logical empiricism, which he aimed to critique.

Finally, Kleindorfer, O'Neill, and Ganeshan (1998) argue that the Bayesianist approach is a particular case in dealing with the problem of induction since, instead of resorting to deterministic theories, they estimate models probabilistically. As for their position in the Philosophy of Science, the authors emphasise that although the Bayesianists would not consider themselves foundationalists, many may see them "*as the last holdover of foundationalism*" (p. 1095).

This duality in the philosophy of science is questioned by Bernstein (1983), who argues for the necessity of overcoming the polarity between objectivism and relativism. He references the work of Hans Georg Gadamer to explore the epistemology of hermeneutics, which relies on a new paradigm: constructivism. For him, we are urged to transcend the 'Cartesian Anxiety', thinking beyond these two fields in the philosophy of science. As he explains, one cannot separate history from the concept of reason, which defines rationality as we perceive it. Rather than being a limitation of reason, historical context is the essence of reason.

This conclusion leads to a new understanding of knowledge and truth. It is a perspective that rejects the excessive reduction of human plurality, as objectivists tend to, but does not succumb to the complete irreducibility of subjective interests as relativists do. Instead, it strives to find common ground when divergences arise. Conflict and negotiation are key concepts that guide the chosen path. As Bernstein (1983) highlights:

Nevertheless, the problem of our society is that the longing of the citizenry for orientation and normative patterns invests the expert with an exaggerated authority. Modern society expects him to provide a substitute for past moral and political orientations. [...] I think, then, that the chief task of philosophy is to justify this way of reason and to defend practical and political reason against the domination of technology based on Science. That is the point of philosophical hermeneutic. It corrects the peculiar falsehood of modern consciousness: the idolatry of scientific method and of the anonymous authority of the sciences and it vindicates again the noblest task of the citizen -- decision-making according to one's own responsibility -- instead of conceding that task to the expert. In this respect, hermeneutic philosophy is the heir of the older tradition of practical philosophy (Bernstein, 1983, p. 39).

1.2 Positivism and policymaking

But what does this discussion have to do with methodology and public policy analysis? While the Philosophy of Science has advanced in overcoming the aforementioned Cartesian duality, Economics is still heavily influenced by positivist principles in research. Consequently, there exists a shared belief that scientific knowledge is produced somewhat independently of an individual's subjective experiences and values. Furthermore, results vest themselves with an aura of objective truth, not a snapshot of reality based on predefined assumptions. This asserted objectivity elevates analytical models to a revered status, undermining social debates and alternative perspectives.

Regarding the connection between positivism and policymaking, Verdung (2010) argues that public policy analysis evolved in "waves", shaping today's framework. These waves are part of a broader political process determining public sector governance doctrines. Therefore, public policy appraisals and evaluations result from a wider movement of shared societal beliefs anchored in paradigms.

The first wave was Science-Driven, which consolidated around the 1960s the evaluation as an element of rationalism that made public policy analysis 'scientific'. By Science, they referred to experimental processes from which 'truths' would emerge. In these processes, researchers tested instruments that allowed previously set goals to be achieved more efficiently. Thus, to become more rational and less dependent on politics, government decisions should rely on feedback from scientific evaluation.

In its more radical form, the 'engineering model' determined that evaluation results must be used instrumentally. That is, the findings about efficient means ought to become binding elements of the government's decision-making process. Here, positivism is the fundamental paradigm that guides evaluations. A handful of contemporary methods, including zero-based budgeting, multi-annual planning and cost-benefit Analysis, are considered a legacy of the Science-Driven wave.

In the mid-1970s, the Dialogue-Oriented wave emerged, questioning society's role in evaluation proceedings. According to Verdung (2010), calls for a more democratic process involved a pluralistic perspective. It was during this movement, he argues, that Guba and Lincoln (1989) introduced the alternative constructivist paradigm for policy analysis. This new paradigm, also referred to as naturalistic, hermeneutic, or interpretative, challenged the conventional positivist basis of assessments. The constructivist paradigm denies the existence of an objective reality, asserting that reality is socially constructed. In this sense, policy appraisals and evaluations do not describe the "true state of affairs" but are constructions that reflect the views of those involved in the process. Furthermore, these constructions are shaped by a shared value system aimed at achieving consensus. Therefore, the key dynamic of the evaluation process is negotiation.

The direct consequence of this point of view concerning policymaking is the impossibility of reaching a single path. There is no convergence to the truth of facts. As Verdung (2010) argues, "Instead of producing truths, dialogical evaluation would generate broad agreements, consensus, political acceptability and democratic legitimacy" (p. 270). As Guba and Lincoln (1989) stressed, "First and foremost, evaluation is a sociopolitical process" (p. 253).

Another direct consequence of the constructivist paradigm in public policy appraisals and evaluations is the lack of a superior methodology, or family of methods, to yield valuable information. Because constructivism is a dialectical process that entails the management of conflicting ideas, constructivists reject the approximation to reality as a criterion of superiority among different methodologies (Verdung, 2010).

Accordingly, the blossoming of the postpositivist perspective of science paved the way for including new values in public policy analysis. Instead of reducing evaluation to an alleged objective single reality measured in terms of statistical and empirical evidence through efficiency, efficacy and effectiveness indicators, values such as ethics, social justice, solidarity and equity became considered objectives in public debate (Ramacciotti and Bernardino, p. 2020).

However, in the 1980s, the New Public Management represented what Verdung (2010) termed the Neo-Liberal Wave. With the slogan of market efficiency, citizens became customers, and rather than solving problems, the Public Sector was perceived as the problem. From this viewpoint, the marketisation of government was expected to yield greater declared values such as efficiency, effectiveness, and productivity. From a budgetary perspective, for instance, independent agents—bureaucrats—were to select the methods to achieve the predetermined goals. There was no opportunity for dialogue, and evaluation prioritised value-for-money analysis. Here, we clearly define superior values and a benchmark for allegedly impartial assessments. The market serves as the benchmark, and market relations are to be pursued.

The fourth and final wave emerged in the late 1990s as part of a nostalgic movement from the Science-Driven wave. The Evidence wave advocates for government guidance derived from lessons learned, governed by the principle that 'What matters is what works'. Verdung (2010) notes that contemporary evaluation is partly a legacy of the doctrine that defines Science as an experimental process where researchers act as distanced and neutral observers. The Evidence wave employs a 'systematic review' to guide the decision-making process, as this collection of available evidence provides solid, independent, objective, and unbiased information (Verdung, 2010).

Howlett, Ramesh, and Perl (2013) argue in the same direction that public policy analysis is mainly seen as a neutral and technical procedure to verify the government's success in dealing with society's demands, which is aligned with the positivist spirit of research. As such, from a Evidence Wave's perspective, the evaluation process must be objective, systematic and empirical.

On the other hand, from a postpositivist perspective, public policy evaluation is a political process with a technical component. Intrinsically, there is no absolute answer to the correct methodology to be applied or an ultimate truth to reach. Evaluation will rely on a common ground among multiple actors. This does not imply that evaluation is irrational or merely a political process. Instead, it highlights the need to understand the limits of human rationality as well as the political forces and historical context that shape procedures. (Howlett, Ramesh, and Perl, 2013).

In this sense, applied models can be a powerful tool for decision-making, but results must be put into perspective in line with the comprehension that they are tied to assumptions, beliefs, historical context and particular worldviews.

In the opposite direction, this chapter addresses the connection between positivism and economics, highlighting how it influences the dominant economic theory. In Chapter 2, by discussing the current methodologies applied to estimate policy efficiency, the discussion aims to link these methods with the literature of the Philosophy of Science to illuminate the fact that the approaches are not only tied to a specific branch of economics—the one in which the market economy serves as the benchmark—but also that the methodologies are linked to the positivist paradigm.

1.3 The changing nature of neoclassical efficiency: Theoretical foundations

Economic efficiency in neoclassical economics arises from Pareto's concept of resource allocation within society, which provides the rationale for interventions. To reach this principle, developments in economic theory were made because the marginalist revolution resulted in the conclusion that there would be no trade-off between efficiency and equity. Central to his idea were discussions surrounding the principle of diminishing marginal utility and the feasibility of interpersonal utility comparisons, which collectively supported the case for distributive public policies.

1.3.1 The marginalist and the ordinalist revolutions: the rise of Pareto's efficiency as a welfare criterion

The marginalist revolution entailed that if utility was cardinally measured and the marginal utility of income diminished, redistribution of wealth efficiently increased total welfare. Thus, the extra dollar of income to a rich man provided less utility than the same extra dollar to a poor man. Therefore, if the marginal utility theory was correct, government policy that redistributed wealth in society was efficient. Consequently, there would be no tradeoff between efficiency and equity. In Pigou's (1920) words:

It is evident that any transference of income from a relatively rich man to a relatively poor man of similar temperament, since it enables more intense wants to be satisfied at the expense of less intense wants, must increase the aggregate sum of satisfaction. The old 'law of diminishing utility' thus leads securely to a second main proposition, which may be stated as follows: '*Any cause which increases the proportion of the national dividend received by poor persons, provided that it does not lead to a contraction of the dividend and does not injuriously affect its variability, will, in general, increase economic welfare'. This conclusion is further fortified by the fact that, of the satisfaction yielded by the incomes of rich people, a specially large proportion comes from their <i>relative* rather than their *absolute* amount, and, therefore, will not be destroyed if the incomes of all rich people are diminished together (Pigou, 1920, p. 52-53).

This idea was unpopular among economists as it contradicted the market's role in efficiently allocating resources. Consequently, advancements were made in the measurement and interpretation of utility. According to Cooter and Rappoport (1984), it was a "*dramatic change in the conceptual framework of economics*" (p. 508). The authors contrast the views of Cannan, Pigou, and Marshall, which they referred to as the material welfare school, with the work of Robbins and Pareto¹, presenting the significant changes in the interpretation of utility and the impacts on policy prescriptions.

The material welfare school distinguished goods by the type of satisfaction they granted. The scale from "material" or "purely economic" to "nonmaterial" or "noneconomic" provided a sense of hierarchy that economists ought to consider. The first accounted for essential human life and health goods; the latter came as comforts and luxuries. Pigou positively linked increases in income with an enhancement in welfare related to the consumption of essential goods. However, this could not be strictly inferred for non-material goods. Therefore, assuming that the rich can afford all the material goods they require, but the poor cannot, a redistribution of wealth to the poor would enhance overall welfare, given their increase in satisfaction (Cooter and Rappoport, 1984).

Moreover, Pigou (1920) discussed the relationship between industrial wages and efficiency and advocated a rise in the working class's income. He examined the connection

¹ The authors refer to Pigou's *Economics of Welfare* (1920), Cannan's *Wealth* (1914), Marshall's *Principles of Economics* (1890), Pareto's *Cours d'economie Politique* (1896-1897), and *Manuel d'economie Politique* (1906), and Robbins' An essay on the nature and significance of economic science (1932), Live and dead issues in the methodology of economics (1938a), and Interpersonal comparisons of utility: a comment (1938b).

between industrial wages and efficiency. The author argued that improved living conditions and access to essential goods would enhance efficiency and, thus, the national dividend. He recognised that material deprivation hindered full human development, adversely affecting productivity.

The objectivity brought by material welfare economists made utility comparisons possible. Furthermore, it enabled the debate on the benefits of the redistribution of wealth in society. Therefore, the interpersonal comparison of utility according to this hierarchy of human needs had substantial policymaking implications since this framework encouraged a strong role of governments in promoting equity, especially since no tradeoff between efficiency and equity was verified. On the contrary, equity could enforce efficiency. They were complementary concepts.

The controversial conclusions drawn by the material school caused significant discomfort among liberal economists, given that the superiority of the market in allocating resources was fundamental to economic theory. Two responses to such arguments emerged. The first, inspired by Robbins's work, contended that interpersonal comparisons were unscientific. The other, based on Pareto's efficiency, asserted the impossibility of such comparisons (Cook, 2022).

According to Robbins, utility comparisons based solely on objective material well-being were problematic since there may not be a consensus. Conversely, utility should reflect subjective preferences, regardless of the goods' ends and the neoclassical marginal diminishing principle. In this sense, a fact that influenced Robbins to disregard the material school as scientific was the inherent value judgments contained in cardinal utility—the idea of a hierarchy of goods that addressed fundamental human needs required some ethical reasoning.

Indeed, making interpersonal utility comparisons and establishing a hierarchy of basic human needs require some consensus, which does not render them unscientific. The context of his discussion, which remains remarkably relevant today, is the influence of positivism in the philosophy of science.

In the same direction of ordinal preferences, one of Pareto's contributions to neoclassical theory was separating the theory's roots from hedonism and utilitarianism by associating utility with desire. From this consumer's desire perspective, the ability of the good to give real satisfaction or to increase welfare was irrelevant (Dobb, 1975). Conversely, the utilitarian tradition was associated with the material welfare school. It has strongly influenced

policymaking by suggesting methods for analysing policies, linking them to the capacity to fulfil individuals' needs (Cooter and Rappoport, 1984).

Indifference curves allowed the theory to drop the assumption that utility could be cardinally quantified. So, marginal utility maximization still relied on consumer utility maximisation, but there was no such thing as a hierarchy of needs. Since desires and preferences were not directly observable, and the individual was the only judge of his will, interpersonal comparisons were impossible.

The association between utility and preferences, as well as desire, has directly influenced policymaking. For example, if utility were linked to meeting various human needs, an antipoverty policy aimed at fulfilling this need would be efficient. Cooter and Rappoport (1984) exemplify this idea:

Departures from industrial efficiency may be documented by pointing to the inadequacy of diets, frequency of illness, high mortality rates and so on. Thus, an important implication of confining utility to the material end of the hierarchy of goods or satisfactions, is that it made the production of utility observable and verifiable. On the other hand, the generation of ophelimity for a particular individual is not answerable to any standard of verification external to that individual because each person is the best judge of his or her own preferences. (Cooter and Rappoport, 1984, p. 516).

Therefore, a relevant consequence of this movement was resetting the trade-off between efficiency and equity, which also greatly impacted policymaking. In the context of Pareto's concept of efficiency rising as a broadly accepted welfare criterion, a sufficient condition for social improvement, equity concerns come at the cost of efficiency losses.

Agafonow (2007) highlights that Pareto's own distributive judgment was endorsed by neoclassical economics. Similarly, Ng (2003) points out that this acceptance relates, explicitly or implicitly, to the endorsement of a set of value judgments. In other words, the Pareto principle is not free from value judgments; it merely depends on less socially contentious ones. But what does it actually mean to assert that resources are efficiently allocated in a Pareto sense?

Pareto-optimality indicates that any further reallocation will reduce the welfare of at least one individual. A Pareto-superior allocation means that another arrangement can be made, improving the welfare of at least one person without making anyone worse off. In this sense, policies grounded in Pareto's standards produce winners but no losers. For this, there is no need to estimate if welfare is enhanced (Coleman, 1980).

Nevertheless, the impossibility of interpersonal utility comparisons and the definition of Pareto's efficiency as the welfare criterion created a significant problem in neoclassical economics: how could policymaking occur without harming at least one individual? There was little scope for efficient interventions. Cook (2022) argues that neoclassical economists challenged their own capacity to provide policy advice to avoid the distributive consequences of their marginal utility theory.

Atkinson and Stiglitz (2015) emphasise that if it were not possible for policies to exist that do not worsen the situation of at least one individual, then the natural procedure would be to evaluate the effects of programmes on different groups. However, assessing the impacts on various groups to determine the merits of a policy violates Pareto's standard as a welfare criterion, as it does not permit losers or any value judgments.

Consequently, a more flexible concept of economic efficiency was essential to applied economics. Then, theoretical developments were necessary to address policymaking without explicitly resorting to interpersonal comparisons. The New Welfare Economics emerged in this context to allow policy prescriptions. According to Drakopoulos (1989), Kaldor-Hicks' efficiency and Bergson-Samuelson's social welfare functions attempted to address this issue in line with the positivist philosophy of science.

1.3.2 Kaldor-Hicks efficiency and hypothetical compensations: a flexible concept

Kaldor (1939) argued that if utility incomparability is "strictly pressed," no policy prescriptions could be made, and "the economist as an adviser is completely stultified" (p. 449). To illustrate this, he discusses what came to be known as the hypothetical compensation principle. According to this concept, if an intervention yields a net benefit over the costs, there is potential for compensating those who are harmed. Thus, the policy is deemed efficient, even if the compensation does not take place, because aggregate welfare will increase. In his own words:

In all cases, therefore, where a certain policy leads to an increase in physical productivity, and thus of aggregate real income, the economist's case for the policy is quite unaffected by the question of the comparability of individual satisfactions; since in all such cases it is *possible* to make everybody better off than before, or at any rate to make some people better off without making anybody worse off. There is no need for the economist to prove- as indeed he never could prove- that as a result of the adoption of a certain measure nobody in the community is going to suffer. In order to establish his case, it is quite sufficient for him to show that even if all those who suffer as a result are fully compensated for their loss, the rest of the community will still be better off than before. Whether the landlords, in the free-trade case, should in fact be

given compensation or not, is a political question on which the economist, *qua* economist, could hardly pronounce an opinion (p. 550).

According to Kaldor (1939), hypothetical compensations allow policy recommendations that increase aggregate welfare without resorting to value judgments. Hicks (1940) makes a similar argument by stating that welfare should be analysed in general - in aggregate. However, as Ng (2003) observed, compensations are not implemented in practice; otherwise, this welfare criterion would be unnecessary since the Pareto principle would be satisfied: no one would incur losses. Because no compensation is required to achieve efficiency, some individuals end up worse off. Thus, "a Kaldor-Hicks efficient allocation need neither be Pareto superior nor Pareto optimal though it may be either or both" (Coleman, 1980, p. 514). It is also known as a potential Pareto improvement.

Most arguments for disregarding full compensation revolve around informational costs involved in repaying everyone harmed by the policy. In addition, concerns about side transfers negatively impacting investments and the labour market are also mentioned. Lastly, full compensation would be an incentive to overstate costs and receive more (Boardman et al., 2018). In the same direction, Campbell and Brown (2023) conclude that compensation transfers may distort economic behaviour and impose economic costs. Therefore, the costs of redistribution may be excessively high.

Boardman et al. (2018) illustrate the distinction between the Pareto and potential Pareto frontiers by allocating a fixed amount of \$100 between two persons. They must agree on how to split the money; otherwise, each will get \$25. The shaded triangle is the Pareto frontier, where any decision beneath \$25 for each person is impossible since it would make at least one person worse. Consequently, even if they choose to allocate the money unevenly, there is a limit referred to by the authors as the 'status quo point'. According to the Pareto Frontier, no one can receive less than \$25.

On the other hand, the potential Pareto frontier allows for a broader range of distributions, extending to the point where one person receives the entire \$100. This is feasible due to the hypothetical compensation principle designed to offset losses. Theoretically, the individual who receives the \$100 could compensate the one who receives nothing. Whether compensation actually occurs or not would not be an economic discussion but rather a matter of value judgment. What is crucial is that, in aggregate, the full \$100 would be distributed.

Figure 1: Pareto x Potential Pareto Efficiency



Source: Boardman et al. (2018, p. 29)

The logic behind the potential Pareto improvement is that the total sum of benefits will be higher, thus enhancing welfare. In contrast, the pure Pareto principle would not permit such distribution because the initial minimum amount of \$25 is mandatory, even if it results in distributing less than \$100.

Initial endowments must be respected from a pure Pareto perspective. In real life, this means that the potential Pareto frontier allows policy analysis to conclude that some intervention is efficient if the net benefit exceeds the present situation, regardless of who benefits from it.

Besides the hypothetical compensations, another relevant aspect of Kaldor-Hicks efficiency is the substitution of utility with money in assessing aggregate well-being. Indeed, Kaldor (1939) directly refers to individuals' satisfaction and their comparisons of gains and losses in monetary terms. Furthermore, Coleman (1980) points out that a net gain in utility does not necessarily follow from satisfying the Kaldor-Hicks test, as this efficiency criterion is not an index of utility. One can meet the Kaldor-Hicks efficiency test without relying on any standard interpersonal comparisons, which was one of the goals of the criteria in the first place. Nevertheless, the presence of winners and losers renders the net result on utility uncertain. Interpersonal-cardinal comparability is essential to evaluate the net result on utility.

In fact, reliance on preferences and utility makes the Kaldor-Hicks an inconsistent test over social states because two states may be Kaldor-Hicks efficient to one another, which is known as the Scitovsky Paradox (Coleman, 1980). In this sense, Gowdy (2003) explains that the potential Pareto principle ran into many theoretical difficulties, one of which resulted in this paradox. The author points out that moving from X to X'' can be potential Pareto improving, which can also be the case when moving back from X'' to X. From X to X''' both people are better off compared to X''. "*This is called "cycling" and it is only one of a number of paradoxes in neoclassical welfare theory*" (Gowdy, 2003, p. 3).





Utility of Person 2

Source: Gowdy, 2003, p. 3.

The potential compensation principle makes the reversal from the later social state to the previous one possible, and eliminating the paradox depends on being unable to compensate losers in one of the social states. However, this challenges the notion of preference transitivity, a vital axiom of consumer theory (Coleman, 1980).

However, Coleman (1980) emphasises that the Scitovsky Paradox only emerges when Kaldor-Hicks is applied in terms of utility. Wealth maximisation, in contrast, ranks social states in monetary units, which is comparable. However, as the author points out, wealth maximization is "*neither more defensible than utilitarianism nor is it an alternative efficiency criterion. Indeed it is not an efficiency criterion at all*" (p. 521).

Cook (2022) argues that this shift from utility to wealth was essential to allowing policy prescriptions without going down a redistributive road: maximizing utility was only possible with interpersonal comparisons, leading to redistribution recommendations. Conversely, maximising aggregate wealth paved the way for an objective measure without considering how

wealth would be distributed. Coleman (1980) adds that the Kaldor-Hicks rule aims to "*replace free, but inefficient, market exchanges*" (p. 543) by mimicking market relations.

In conclusion, the Kaldor-Hicks hypothetical compensation principle, along with the shift from utility to wealth maximisation, sidestepped the trap set by neoclassical Pareto efficiency to the State's intervention, becoming central to applied economics. Thus, the Potential Pareto principle has paved the way for government actions to adopt the discourse of efficient allocation.

This acknowledgement brings us to a critical feature of Kaldor-Hicks' efficiency: it treats everyone equally, irrespective of initial endowments. The Pareto principle also treats everyone equally, but initial endowments must be respected. On one hand, Kaldor-Hicks' efficiency could theoretically promote a better distribution. In practice, however, methodologies that estimate efficiency based on this principle tend to favour the wealthier, as discussed in detail in Chapter 2, making this efficiency concept a tool for justifying controversial interventions.

Conversely, Pigou's analysis within the material welfare school placed a higher value on a dollar given to the poor compared to the wealthy, favouring interventions that benefit materially deprived individuals. Cooter and Rappoport (1984) stress that "*Kaldor made a different conventional judgment from Pigou, rather than no judgment*" (p. 526). They argue that this acknowledgement is particularly significant because it influenced a generation of economists to believe that treating everyone equally is scientific, while placing greater value on the most deprived is seen as a nonscientific approach.

The background of this discussion is whether there exists a trade-off between efficiency and equity. As previously discussed, a consequence of the material welfare school's framework was that these principles were complementary. Therefore, there was no contradiction in policymaking. Conversely, the ordinalist revolution re-established the trade-off, viewing efficiency and equity as distinct and opposing issues. For this reason, a general guideline must prioritise one of these objectives in policymaking.

A consensus has been reached over efficiency being the primary welfare criterion. With this, equity can be addressed, but it comes at the cost of efficiency losses, strengthening arguments against distributive policies and portraying them as burdensome. Therefore, other principles and objectives were overlooked as central to welfare economics.

1.3.3 Challenges to neoclassical efficiency

The theoretical developments have raised numerous questions directly connecting the theory and policy implications, even within the neoclassical tradition. From this perspective, criticism revolves around the trade-off between efficiency and equity. Klasen (2008) examines four interrelated strands of the literature that challenge the notion of treating efficiency and equity as distinct elements: the experimental literature on equity and inequality-adjusted wellbeing, as well as the theoretical and empirical literature on inequality-growth linkages. Conversely, this empirical literature provides evidence that equity is essential for achieving efficiency.

The first strand pertains to the findings regarding the significance of equity and fairness, as revealed by the experimental literature that assesses deviations from standard selfish utility maximisation behaviour. This literature offers evidence for inequality aversion, demonstrating that fairness constitutes a relevant motivation for economic agents. The second strand encompasses the importance of relative incomes and inequality on subjective well-being. The emphasis here is on how inequalities influence an individual's perceived welfare. The author argues that this literature addresses a similar issue to the first strand but employs a different methodological approach that evaluates responses concerning life satisfaction and happiness. The third strand includes studies that merge measures of mean incomes with income inequality, highlighting the substantial impact of incorporating inequality on well-being. Lastly, Klasen (2008) presents literature examining the relationship between income, gender inequality, and economic growth, showcasing studies within the neoclassical tradition that suggest inequality hinders economic growth.

The author concludes that the alleged trade-off between efficiency and equity does not seem to hold in practice and that "*reluctance to explicitly consider equity issues in theoretical and applied economics is distorting analysis and policy advice and leads to erroneous and sub-optimal policy recommendations*" (p. 271). Drakopoulos (2024) argues in the same direction, advocating for the incorporation of comparability to establish solid foundation for policymaking and evaluations in applied economics, such as "*the justification and the level of minimum wages, environmental protection policies, desirability of provision of public goods, and the social impact of large-scale projects*" (p. 4).

Indeed, inconsistencies between theory and practice have been largely noticed. Applied economics often resorts to comparisons and distributive judgments to address welfare, moving beyond the traditional theoretical efficiency criterion, as real-life politics and collective demands play a significant role in policymaking. Drakopoulos (2024) cites common examples of welfare concepts that rely on interpersonal comparisons. For instance, many countries' national social security systems are designed to enhance the well-being of certain segments of society based on these comparisons. Another example is the prevalent progressive taxation system, which operates on the assumption that comparability is achievable.

From an alternative economic perspective, the criticism lies in the core components that underpin the neoclassical tradition. Garegnani (2007) discusses relevant policy implications derived from the classical revival, specifically concerning content and method. Regarding content, the author stresses three related aspects with remarkable effects on policy analysis: a) the distributive rather than the competitive role of prices, b) the fact that labour unemployment and capacity underutilisation is the rule, not the exception, in capitalist economies, deriving from the inexistence of the neoclassical symmetry between factor demand and supply, c) the relevance of aggregate demand for the level and growth of social product.

When comparing the differences between the classical and the marginal approaches to income distribution, Petri (2021) addresses the fundamental aspects of the points made by Garegnani (2007). The central difference is the explanation of how the social product is divided between wages and profits. While in the marginal approach, the return on capital and real wages are determined by a symmetrical and simultaneous mechanism that leads toward an equilibrium between supply and demand, the classical approach claims that capitalists are able to appropriate more or less of the social product according to their bargaining power. Petri (2021) clarifies that the most relevant divergence is not the existence of a competitive labour market but rather the presence or absence of a cooperation mechanism of the production factors that facilitates the substitution process. In the marginal approach, the substitution mechanism engenders a decreasing demand curve for each factor, which entails that real wages and the level of employment are endogenously determined. Conversely, for the classical approach, there is no necessary relationship between real wages and the level of employment, which depends on relative bargaining power.

The change in the explanation of the distribution of the social product between wages and profits alters the interpretation of factor prices, directly impacting the sense of Pareto optimality. If, from a marginal perspective, prices reflect the scarcity of production factors, in the classical approach, sociopolitical elements are essential to determine factor prices. Garegnani (2007) clarifies the implication of this argument for policymaking: The denial of a determination of the distributive variables by an equilibrium between demand and supply of 'factors', i.e. by the 'full employment' of them, breaks open the magic circle within which policy analysis tends to be confined by the neoclassical preoccupation with distortions of the Pareto-optimal allocation of resources allegedly brought about by competition (p. 234).

That is, if the economy does not tend to a full employment equilibrium, there are no efficiency costs related to allocative changes of factor transfers. Furthermore, there is no necessary maximization of the social product brought by competitive markets. In fact, this acknowledges what Garegnani (2007) provocatively names 'the principle of underutilization of resources in a market economy'. This principle, the author stresses, leads to the most critical implication of classical theory to policy analysis: the relevance of aggregate demand for capital accumulation and growth. "*The critique of the neoclassical theory of distribution entails therefore the critique of the idea of a long-period irrelevance of aggregate demand for the level and growth of the social product*" Garegnani (2007, p. 235).

In this sense, acknowledging the contributions derived from the work of Keynes and Kalecki, which bring a different perspective on the dynamics in capitalist economies, is essential to the debate about the role of governments in society. Petri (2021) outlines that the Keynesian-Kaleckian approach highlights production's tendency to adapt to demand, which requires a degree of production flexibility. Furthermore, the propensity to underutilise production factors in capitalist economies is evident, contrary to expectations of neoclassical full employment equilibrium brought by competitive markets. Thus, a direct consequence of the classical perspective is a different role for the State's action, one not tied to efficiency as the guiding welfare criterion.

As for methods in policy analysis, Garegnani (2007) highlights that, in the classical revival, central elements are defined outside the core theory, which leaves little space for routine predictions about policy effects. In this sense, the author argues that there are few grounds left for the idea that policymaking can be guided by experts who can objectively translate collective interests. In fact, the classic revival opens the room for accepting that policy decisions will have different impacts on social groups and will also respond to power relations, which aligns with a postpositivist perspective of science and paves the way to the understanding of policymaking decisions as a sociopolitical process supported by appraisals and evaluations instead of a purely technical series of action to achieve maximum social welfare.
1.4 Market and Efficiency: Benchmarks for Policymaking

Efficiency as a welfare criterion originates from the fundamental theorems of Neoclassical welfare economics. According to the First Theorem, competitive markets ensure the efficient allocation of resources in society. In this framework, economic efficiency is the principle that leads to an allocation yielding the maximum total surplus. Consequently, the market economy serves as the benchmark for social interactions, with efficiency as the guiding principle of any economic analysis. Recognising this framework carries significant implications for policymaking.

The first implication of using the market as a benchmark is that it theoretically restricts government actions to specific circumstances and permissible instruments. Secondly, it promotes a mindset that dismisses alternative perspectives on social issues. This approach establishes methodologies for policy analysis rooted in market principles, transforming the citizen into a mere consumer of services and, thus, perpetuating market-related inequalities in the public sphere. Thirdly, it establishes a potential Pareto improvement as a sufficient condition for social progress. Finally, it directly impacts policy financing since contemporary societies have been using evaluations based on neoclassical efficiency to guide the budgeting process.

Transitioning from theory to practice involves a significant journey that goes beyond the mere application of theoretical foundations. Consequently, exploring the implications mentioned above is crucial for understanding contemporary policymaking.

1.4.1 Limits on circumstances and instruments for policymaking

In a traditional approach, as Stiglitz (1991) discusses, the Fundamental Theorems of Welfare Economics assign a prominent role to the competitive market and restrict the circumstances and instruments available to the State for ensuring Pareto efficiency. Understanding the conditions outlined in the First Theorem that enable the economy to achieve efficiency, along with the policies that can restore it, forms the Market Failure approach to Modern Economics. Furthermore, the Second Theorem delineates the instruments through which the State can act to rectify the undesirable distribution arising from competitive markets.

Therefore, the Fundamental Theorems of neoclassical economics suggest fields open to government interventions. Market failure gives the government the legitimacy to act (Tresch, 2015). In addition, addressing equity concerns is possible through a narrow set of tools. This

approach implies that the government's actions should primarily correct the elements that prevent the proper functioning of markets and eventually deal with undesirable distributions stemming from economic interaction.

Weimer and Vining (2017) summarise the contemporary, commonly accepted market failures and their implications for economic efficiency. Traditional market failures refer to situations where public intervention can yield a higher social surplus than the outcome produced under market equilibrium. Thus, the market serves as the reference point for all analyses. The fundamental question to address regarding market failures is: Why doesn't the market allocate this good efficiently? The answer to this question justifies public interventions.

Traditional Market Failures				
Public Goods	Pure public goods (undersupply) Open access/common property (overconsumption, underinvestment) Toll goods (undersupply)			
Externalities (Missing Markets)	Positive externalities (undersupply) Negative externalities (oversupply)			
Natural Monopoly	Declining average cost (undersupply) With costly monitoring (undersupply, X-inefficiency)			
Information Asymmetry	Asymmetry Quality overestimation of experience, post-experience goods (overconsumption) Quality underestimation of experience, post-experience goods (underconsumption)			
Macroeconomic Dynamics	Business cycles (underemployed resources)			
Other Limitations of the Competitive Framework				
Thin Markets	Cartelization (undersupply)			
Preference Problems	Endogenous preferences (typically overconsumption) Utility interdependence (distributional inefficiency) Unacceptable preferences (overconsumption)			
Uncertainty Problems	ainty Problems Moral hazard, adverse selection, unique assets (incomplete insurance) Misperception of risk (violation of expected utility hypothesis)			
Intertemporal Problems	Nontraded assets, bankruptcy (incomplete capital markets)			
Adjustment Costs	Adjustment Costs Sticky prices (underemployed resources)			

Table 3 - A Summary of Market Failures and Their Implications for Efficiency

Source: Weimer and Vining (2017, p. 238), adapted.

For instance, public goods are examples of goods not efficiently provided by private markets. The common characteristics attributed to these goods are non-rivalry in consumption and nonexcludability in ownership. However, these two features alone do not sufficiently address the variety of public goods required to understand the nature of the failure and, thus, the expected public response. A third element – congestion or crowding – in demand is essential for constructing a complete framework. Congestion pertains to demand levels, which can vary and potentially change the characteristics of a good. For example, non-rivalry may be a feature of a good only to a certain extent. Specifically, consumers can impose costs on each other over some range of use. Hiking in wilderness areas exemplifies non-rivalry that could become rivalrous in the case of excessive hikers, thereby interfering with each other's enjoyment. Thus, congestion refers to the scenario where the marginal social cost of consumption exceeds the marginal private cost of consumption (Weimer and Vining, 2017).

Weimer and Vining (2017) present an insightful board that classifies the nature of goods based on these characteristics. For example, 'Toll Goods', such as roads and bridges, are considered non-rivalrous in consumption and excludable in ownership. They are likely to be inefficiently provided by private markets due to their non-rival nature. Regarding demand, uncongested toll goods are only efficiently supplied by the market at zero price. However, because exclusion is possible, private supply may occur. In this case, private supply at any positive price would lead to underconsumption, which is inefficient. Conversely, if the toll good is congested, the pricing process becomes more complex, as efficiency would require variable prices: a price of zero during off-peak times and positive prices during rush hours, which is unlikely to occur in reality. Therefore, privately provided toll goods typically result in deadweight losses due to inefficient pricing.

The rest of the figure crosses the three characteristics—rivalry, excludability, and congestion—to explain why the market does not allocate a certain good efficiently. This explanation is essential to justifying any government action.

Figure 3 - A Classification of Goods: Private and Public



Source: Weimer and Vining, 2017, p. 157

Nevertheless, the framework for public goods does not explain why rival and excludable goods, such as health services and education, are so commonly publicly provided in contemporary society.

A common response to this question involves the presence of externalities, which appear to encompass all publicly provided private goods that the theory does not explain. In Weimer and Vining's (2017) words: "*There has been considerable debate about the existence, and magnitude, of positive externalities associated with the consumption of housing, education, health services, and food. These are empirical issues that can only be considered on a case-by-case basis*" (p. 401). The idea of an externality relates primarily to consequences for third parties not directly involved in a voluntary exchange. Thus, from this perspective, publicly providing rival, excludable and uncongested goods is justifiable to the extent that it indirectly benefits society. Another more plausible explanation Stiglitz and Rosengard (2015, chap. 5)

pointed out is that public education provision, for example, relates to common distributive concerns: a large part of society believes this opportunity should not depend on parents' wealth. Therefore, justification relies on equity concerns, not efficiency.

However, to theoretically justify this intervention on efficiency grounds, economists often resort to the distribution of vouchers. In this sense, Weimer and Vining (2017) argue that in-kind grants administrated through vouchers and distributed in fixed quantities at price zero are equivalent to lump-sum transfers, meaning that efficiency would not be compromised by publicly providing private goods. "*Vouchers are used for subsidizing a wide range of goods, such as primary and secondary education, day care, food and nutrition, environmental protection, and housing*" (p. 403).

Once more, the argument fails to explain why the market did not efficiently allocate rival, excludable, and uncongested goods initially. The voucher system resorts to clarifying the undesirable distribution outcomes that arose from market interactions to justify public involvement. Furthermore, this represents a specific scenario that does not pertain to the majority of publicly provided private goods.

Another justification lies in the realm of 'merit goods', which Musgrave (1957) defined as public expenditure not derived from preferences. Sometimes, choices within the marketplace do not result in welfare maximisation. For this reason, interventions are necessary to rectify limitations arising from individual preferences, thereby enhancing social welfare. In such cases, separating the discussion of public provision and distribution is impossible. "*I refer to as merit wants may be thought of as provided for in a separate branch. Here a strict separation from the Distribution problem does, indeed, become untenable*" (p. 341).

Besides market failure, the concept of government failure appears as another reason for government intervention (Stiglitz and Rosengard, 2015, chap. 1). Weimer and Vining recognise that "*the social sciences have yet to produce a theory of government failure as comprehensive or widely accepted as the theory of market failure*" (p. 277). However, the authors argue that governments can ultimately fail to promote social good. Government failures are characterised as situations with likely opportunities for improvement. For example, in representative governments, the issue of self-interested representatives may lead to inefficient policies. They emphasise a tendency to respond to their constituencies rather than focusing on the interests of society as a whole. With this, geographically inefficient allocations may occur. Also, the influence of interest groups in a world of imperfect information and costly monitoring causes inefficiency through rent-seeking.

Finally, electoral cycles stimulate socially excessive discount rates, stimulating public spending in the present.

However, market failure, government failure, or distributive concerns alone do not justify public action within this framework. According to Stiglitz and Rosengard (2015, chap. 4), the second essential element for government action is the analysis that demonstrates the intervention will enhance social welfare. This is where methodologies that estimate efficiency come into play². Consequently, the state's actions will be supported only after a proper justification of why private suppliers cannot fully provide a good or service, alongside an estimation that public intervention will yield positive net benefits.

At the beginning of this section, it was highlighted that the traditional neoclassical welfare approach perceives market failure through a corrective lens. From this viewpoint, public interventions seek to amend the factors that hinder the proper functioning of the market economy, ideally employing non-distortionary instruments, such as lump-sum taxes and transfers, which are considered the best solutions. Nevertheless, costs and the availability of information pose significant barriers to the utilisation of these instruments. The second-best solutions can be regarded as indirect control solutions since governments cannot directly control the intended element. For instance, governments might aim to reduce the consumption level of a particular good and, therefore, resort to instruments that influence individual behaviour. Consequently, the equilibrium achieved will be different (Atkinson and Stiglitz, 2015, chap. 11).

Nevertheless, in addition to informational costs, political feasibility is another crucial element when transitioning from theory to applied economics (Atkinson and Stiglitz, 2015, chap. 11). Furthermore, recognising that market failures are not exceptions to a general framework but rather the most common scenario also alters the perspective regarding the role of government in society. Thus, the simplistic corrective policy prescriptions viewpoint that theory presents is inadequate to explain the observed policymaking in contemporary society. As Stiglitz (1991) points out:

There was a simple prescription: government activity should be limited to lump sum redistributions and correcting a well defined and limited set of market failures. But now that we see that market failures (in the sense of constrained Pareto efficiency) are pervasive, that they arise in all aspects of economic life, and that issues of efficiency and equity cannot be neatly separated, these issues of political economy cannot be ignored (p.40 e 41).

² These methods will be discussed in detail in the next chapter.

Additionally, Atkinson and Stiglitz (2015, chap. 11) emphasise that government actions have a more modest objective than the theory suggests. Rather than social welfare maximisation, welfare improvements are more suitable for understanding policymaking in applied economics.

This reasoning, combined with the Kaldor-Hicks concept of efficiency, paved the way for a more active State. Thus, the notion of market correction was supplanted by a marketoriented active State. As discussed in the previous section, the potential Pareto improvement was a theoretical development that enabled more 'efficient' policymaking. Without the shift in the understanding of efficient interventions, governments were unlikely to act without making at least one person worse off. In this sense:

The use of Kaldor-Hicks efficiency in cost-benefit analyses is a perfect example of this crucial intellectual shift as it was invented explicitly to replace classical liberal, 'handsoff' policy approaches and legitimise a far more aggressive, activist and neo-liberal form of technocratic – yet market-oriented – economic governance and management in which the economy is run not by the free market or democratically elected politicians but rather market-minded expert-economists (Cook, 2022, p. 5).

In conclusion, the speech on market failure, government failure, and, to a lesser extent, distributive concerns related to estimating whether an intervention will broadly improve social welfare, justifies contemporary policymaking. However, although the space for state intervention has expanded through theoretical development, which has made the concept of efficiency more flexible, substantial limitations still exist within this policymaking framework.

The most significant aspect is the enforcement of a market-oriented perspective on public actions, which will become clearer when discussing the methodologies that evaluate efficiency. Policymaking has increasingly aligned with market mechanisms, using this rationale not only to justify intervention but also to integrate it into the methods that assess the merits and scope of a policy.

For example, from a philosophical perspective, Cost-Benefit Analysis (CBA) has emerged as a fundamental tool in the public decision-making process. Indeed, it determines whether a policy has merits, with those merits directly contingent on its perceived efficiency. CBA assesses a policy through a net measure of benefits and costs, which relies on a market rationale for the estimations, subordinating collective interest to this questionable efficiency measure, as discussed in detail in the next chapter. In summary, by employing CBA, one delegates the decision regarding the merits of addressing a social problem.

In addition to delegating the decision regarding the merits, contemporary methods also define the scope of policies. In this regard, another commonly employed method is Cost-Effectiveness³, which acknowledges a policy's merits but often limits its applicability. Rather than adhering to the principles of universalism, equality, and rights, this methodology embraces the notions of competition, choice, and incentives to justify targeted policies over universal ones. Indeed, as efficiency became increasingly central to policymaking, so too did the barriers to universal programs (Berman, 2022).

1.4.2 Enforcement of a way of thinking

The second implication of having the market as a benchmark is that it enforces a way of thinking that disregards different perspectives. Berman (2022) discusses the emergence of this distinctive way of thinking about policy, which she calls the 'economic style of reasoning' – a microeconomic approach. The central elements of this reasoning are simplifying complex situations, resorting to quantitative methods, weighing costs and benefits, and thinking at the margin. They become constituent elements of policymaking.

The author devotes considerable effort to explaining how this way of thinking became pivotal to policy analysis. By 'economic reasoning,' she means how market values, which focus on choice, competition, and incentives, replaced previously praised principles, such as universalism, equity, and rights, in conducting public policies. Therefore, noneconomic values that justified broader government actions were substituted for the neoclassical economic efficiency that vested policymaking with the market lens.

Consequently, the economic reasoning changed the political space for policymaking and the terms of the debate. The commitment to efficiency led policymakers to a particular type of solution to social problems. For example, Berman (2022) explores the shift in environmental policymaking from condemning pollution to accepting it as an externality to be priced. The market lens turned the 'one-size-fits-all' regulatory approach to a cost-benefit logic that views the environment as a service and, thus, a priceable asset. The author argues that this example

³ Cost-effectiveness is not extensively debated because this work focuses on discussing the most commonly applied methods to assess allocative and technical efficiency.

only illustrates the wide range of changes in policy analysis between the 1960s and 1980s toward this 'economic reasoning'.

This acknowledgement aligns with the previously discussed shift from a 'hands-off' State, associated with the notion of a market correction, to an active, market-oriented 'handson' State. This economic reasoning perspective gained popularity to varying degrees across all social areas, including health services, education, housing, and poverty alleviation. Consequently, even in policy domains where the link to rights was more firmly established, the market perspective became predominant.

Berman (2022) argues that this style of reasoning was institutionalised through various channels, including organisational changes, legal frameworks, and administrative rules, which created a positive feedback loop. Indeed, some elements of this perspective were formally integrated into decision-making processes, making it harder to contest. The author cites a series of United States Supreme Court decisions that rendered allocative efficiency the only legitimate goal of antitrust policy. However, there are many other examples.

The World Bank explicitly recommends that project appraisals include a cost-benefit analysis to determine whether the project should be financed (World Bank, 2010). The same is true for the guidelines of the Asian Development Bank (ADB, 2013). Cook (2022) demonstrated that Regulatory Impact Assessments within the OECD began to depend on cost-benefit analysis in all member countries. Furthermore, this practice has also spread to many developing nations such as Uganda, Kenya, Brazil, and Vietnam.

As the economic style of reasoning became institutionalised, conflicting values such as health and education as civil rights began to be regarded as unreasonable and irrational. Consequently, benchmarking efficiency came at the expense of undervaluing conflicting principles. A significant factor in this process was that elevating efficiency to the highest standing is viewed as scientific and politically neutral.

In this sense, Berman (2022) stresses the connection between the growth of evaluation research and economic reasoning, arguing that massive financing to support organisations and studies that privileged quantitative and econometric methods and, thus, reproduced the economic style was a relevant element of this process. Over time, it became established as the scientific method among experts and the wider community. Claims of neutrality and rationality

empowered these researchers to centre efficiency and cost-effectiveness in their analysis, delegitimising 'value-based judgements' that underpin competing worldviews.

Onwards, this line of reasoning set the tone for policy evaluations, even in social areas. For instance, economic anti-poverty research predominantly focuses on how income affects an individual's work or family decisions to prevent extensive policy measures rather than examining how such policies influence society from a fundamental human rights perspective. Another example is research on health services coverage policy, which concentrates on moral hazards to justify cost-sharing and avoid universal full coverage, further illustrating the marketdriven approach in policymaking. Indeed, many conclusions highlight how a healthcare market could effectively stimulate competition and lower prices. Access would be provided through 'choice'.

Therefore, the economic reasoning also shaped the content and the way relevant aspects of policy analysis were rationalised. Efficiency became the channel through which the change from a collective perspective of society to an individualistic microeconomic view occurred. This is what Berman (2022) argued when eliciting the passage mentioned earlier, from universalism, equity and rights as founding values of policymaking to competition, choice and incentives. This line of reasoning shaped applied economics with significant effects on appraisals and evaluations.

In sum, this specific way of thinking has gradually spread and has come to prevail in policymaking. Once economic reasoning started ruling the decision-making process, it became harder to challenge it with different principles. That is, naturalising efficiency measures as a central goal of policy analysis hinders other ways of thinking from being considered viable alternatives when discussing and deciding on public policies (Berman, 2022). As Fontaine (2023) highlighted, the economic reasoning dresses up their solutions "*with an aura of inescapability*" (p. 3).

1.4.3 Potential Pareto efficiency as a sufficient condition for welfare improvement

Efficiency takes many forms, as discussed in detail in the next chapter. It is typically linked to waste avoidance, which is one of the reasons it is so widely accepted in society. However, when the principle is examined from a theoretical perspective, it commonly pertains to allocative efficiency, which is also a synonym for economic efficiency. In this context,

allocative efficiency involves making choices between policy alternatives, basing the decision on a strict notion of welfare improvement: a Pareto or potential Pareto improvement.

This perspective presents a limited view of the role of governments in society, as it simplifies the nature of social relations to a mere exchange system governed by preferences, which serves as a basic framework for policy prescriptions. As previously discussed, the potential Pareto efficiency as a decision-making rule disregards distributional concerns and any other guiding principle, making this specific concept of efficiency the equivalent of social well-being.

Another pertinent concern regarding the potential Pareto efficiency as a sufficient condition for social improvement is the reduction of society's well-being to income. That is, the welfare criterion is confined to a single variable: money. Of course, income is an important factor in assessing individual well-being, but limiting policy analysis to this one variable diminishes the decision-making process.

In this sense, a growing body of literature provides evidence that ultimately challenges the ability of economic efficiency to translate into welfare. For instance, the literature on happiness emerges by questioning the centrality of income in determining subjective wellbeing. Richard Easterlin pioneered this area of economics, offering valuable insights into social welfare discussions. His well-known paradox posits that the rich are happier than the poor within the same country, yet this is not corroborated across different countries or over time. This finding does not imply that income is irrelevant to happiness but rather strongly indicates the insufficiency of this variable alone to explain welfare (Graham, 2005).

Easterlin (2003) explains that the inference derived from the revealed preference that "more is better" is not absolute. Survey evidence supports a positive correlation between income and happiness in cross-sectional data but not in time series data. That is, over a life cycle, as income increases, happiness remains unchanged. He argues that adaptation and social comparison are relevant aspects of one's well-being, along with non-economic factors such as health, marital status, and education.

Graham (2005) examines the discrepancies between income measures and reported well-being in the literature on happiness to offer insights into welfare discussions. The author emphasises how the earliest economists, such as Jeremy Bentham, endeavoured to comprehend individual happiness. However, with the evolution of welfare economics, definitions of wellbeing became increasingly narrow, confined to utility and preferences.

Graham (2005) identifies several elements often mentioned in surveys about welfare that are overlooked in traditional measures of well-being. These factors include employment, public services such as health, education, and crime prevention, as well as norms surrounding fairness and justice. Moreover, even dimensions related to income, like poverty and inequality, have broader implications that traditional estimates fail to capture. For example, well-known poverty lines facilitate international comparisons but seldom reflect individual perceptions of poverty within countries.

The findings indicate that the effects of age, income, education, health, employment, and marriage are surprisingly comparable in both developed and developing countries, highlighting common elements that substantially influence social welfare. (Graham, 2005).

In the same vein, Petri (2021) emphasises that perceived well-being is not entirely subjective but is strongly correlated with the aforementioned indicators. In fact, he notes that research on happiness suggests that once basic needs for survival and good working conditions are met, material factors cease to influence long-term self-perceived happiness. From this point on, well-being becomes associated with positive affective relationships and secure employment.

Discussions and new insights on well-being are highly relevant because the narrow perspective of the traditional neoclassical framework restricts welfare criteria to achieving efficiency, thereby shaping government policies. Conversely, recognising that other elements—such as access to social security protection, a universal health system, and quality education or jobs—strongly influence the definition of social welfare and thus undermine the capacity of efficiency as a welfare criterion to guide policymaking.

1.4.4 Policy financing: efficiency and Spending Reviews

The final implication of using efficiency as a benchmark for policymaking concerns the potential impact on the financing of various policies. Efficiency is linked to the budgeting process in many countries through the Spending Reviews (SR) framework. "A spending review is the process of identifying and weighing saving options, based on the systematic scrutiny of

49

baseline expenditure" (Bova, Ercoli and Bosch, 2020, p. 8). Therefore, SRs serve as instruments for analysing public policies, taking into account their burden on public budgets.

Generally, the SR is a flexible framework that can accommodate any benchmark, as no single criterion has been defined previously. As Bova, Ercoli and Bosch (2020) argue, it could target policies that enhance growth or promote long-term environmental sustainability, for instance. However, in the context of rising public debts and calls for fiscal consolidation, the SR emerges as a strategic tool to reduce aggregate spending by primarily focusing on efficiency analysis. Therefore, although the SR can have other objectives, it has primarily been used to cut spending.

The historical antecedents of SR also include frameworks that aimed to integrate budgeting and evaluations from a performance perspective to justify expenditure. Examples include Performance Budgeting in the 1950s, the Planning-Programming-Budgeting System, Zero-Based Budgeting in the 1970s, and the Expenditure Management System in the 1980s (Catalano and Erbacci, 2018).

Actually, Schick (2014) argues that these initiatives are very similar to Performance Budgeting (PB), albeit with new labels. He states: "*PB has become an elastic label that is stretched to encompass a wide range of processes that purport to expand fiscal space*" (p. 3). Indeed, Doherty and Sayegh (2022) outline the primary objectives of SR: a) fiscal consolidation, b) creation of fiscal space, c) reprioritisation of existing expenditures, and d) achieving better value for money. These objectives are realised by selecting efficiency as the benchmark for the evaluations.

As discussed in this section, efficiency is a concept closely tied to market principles, as understood by marginalist theory. Not surprisingly, an expanding body of literature identifies a theoretical link between SR and cutback management, which are essentially reactive strategies for addressing declining organisations. The cutback management literature emphasises the idea of strategic–selective cuts, in contrast to the linear approach, which would uniformly reduce budgeting across numerous actions. The linear strategy is often employed in public organisations due to a lack of performance information. For this reason, SR is deemed indispensable as it identifies opportunities for non-linear savings. Because of this selective approach, efficiency emerges as the dominant principle: cuts would concentrate on less efficient programmes/units to avoid undermining overall productivity (Catalano and Erbacci, 2018). This perspective makes sense for a production plant but does not necessarily apply to public policies, given the remarkably different nature and objectives of firms and the State. Of course, the selective approach can be a good strategy, rather than indiscriminately and linearly cutting public spending. However, resorting to efficiency as the supreme benchmark is reasonable from a profit viewpoint, not from a multiple objective perspective, which is the case for government actions. Therefore, even if cutting or limiting spending is broadly agreed upon as a social priority, it does not necessarily mean that it should be based on efficiency grounds.

By imposing efficiency as the primary principle guiding evaluations, SR vests itself with a rhetoric of technicality that supposedly guarantees an impartial framework to favour policies worth financing. It is no coincidence that SR is frequently associated with the idea of ensuring the 'quality' of public spending. This direct association between efficiency and quality is often seen when claims for cutting or limiting spending are present. As Bova, Ercoli and Bosch (2020) point out about SR: "*They are a key tool to enhance the quality of public finance, as they can promote allocative efficiency*" (p.8).

Thus, the narrative involves the belief that enforcing reduced spending encourages waste reduction by prioritising efficient policies. However, in practice, there is no automatic correlation; this narrative assumes evaluation is a neutral process, separate from the natural political conflicts and disputes in a democratic society. Furthermore, it confines the adjective 'quality' to a specific principle – efficiency – as if prioritising other benchmarks in policy evaluations would result in less relevant choices, as discussed in the previous section. Finally, it makes the association of quality in public spending with a particular worldview about the dynamics of capitalist economies, with a specific distributive perspective and a singular explanation about the behaviour of families in society, as also previously discussed.

Nevertheless, SR has gained popularity among OECD countries. In 2023, 34 out of 35 OECD countries reported using SR at least once. Also, 20 countries claimed to use it regularly and 9 countries periodically⁴.

In a summary of the analytical framework for spending reviews, Fallov and Georgieva-Andonovska (2018) discuss the four problem areas that spending reviews address: a) effectiveness; b) allocative efficiency; c) technical efficiency; and d) functional coherence.

⁴ https://www.oecd.org/en/topics/sub-issues/spending-reviews.html

Each area requires specific methodologies, yet the recommendations can be summed up with one goal: to terminate or reduce funding.

Problem area	Examples of possible scenarios	Types of Analysis	Examples of areas of recommendations
Effectiveness	A government program or activity is not achieving its intended objectives	 Desk reviews of documents Analysis of quantitative and qualitative performance data Impact evaluation 	Terminate activities or reduce funding and reallocate to more effective activities
Allocative efficiency	 Activities not aligned with government strategies or policies; Low socioeconomic returns of public funds 	 Profile analysis Trend analysis Comparing/benchmarking budget allocation/prioritization Budget composition analysis Estimating marginal returns Cost benefit analysis Cost-effectiveness analysis Cost-utility analysis 	Terminate activities or reduce funding and reallocate to higher priority activities
Technical efficiency	 Constant or increasing funding despite decreases in underlying cost drivers, e.g.: decline in number of beneficiaries simplifications in the regulatory or international requirements available new technology new organizational models, processes or modes of service delivery The Budget for a specific budget heading is higher than comparable budget headings in other entities with no 	 Data Envelopment Analysis (DEA) Budget deviation analysis Business processes review IT systems gap analysis Desk review of legal, regulatory or international requirements 	- Reduced funding based on identified efficiency potential - Recommendations on how to implement new technology or business processes
Functional coherence	Indications of duplicative or inefficient functions carried out by an entity	 Review of mandates, goals and objectives; Benchmarking with other countries Gap analysis of systems 	Legal and organizational changes to streamline/eliminate/outsource/(de)centralize functions and generate corresponding savings

Table 4 - Summary of analytical framework for spending reviews

Source: Fallov and Georgieva-Andonovska (2018) p.24. Adapted

The possibility that a policy justifies its existence for reasons other than efficiency and, to a lesser extent, effectiveness does not exist. Furthermore, even if effectiveness appears as a guiding principle of SR, it has not been prioritised in practice. It is important to note that effectiveness is the sole principle listed in Fallov and Georgieva-Andonovska (2018) that relates directly to a previously established public goal, such as reducing poverty or universalising health services. However, as Bova, Ercoli, and Bosch (2020) demonstrate, allocative and technical efficiency dominate SR, serving as the primary guideline of this framework. The distinctions between allocative and technical efficiency are discussed in detail in Chapter 2.

Figure 4 - Main objectives of Spending Reviews over the past 10 years



Source: Bova, Ercoli and Bosch (2020), p. 39.

The idea of constantly evaluating public spending to identify possibilities for improvements is very appealing. Problems arise when collective social objectives are reduced to neoclassical efficiency performance analysis to justify public expenditure. The pursuit of efficiency as a primary goal of evaluations is not a natural aspect of the public policy cycle. Instead, one central element of this choice is to grant the evaluation process a 'technical' stamp, as if one could preclude context, political interests, societal values, and individual judgments.

However, how can one justify the merits of preserving cultural heritage or promoting gender or race equality on the grounds of efficiency? Even if one can justify them, for instance, by linking these policies to better performance in the labour market, is this the only (or even the best) reason to fund such policies? Should the budgeting process be guided by a framework solely focused on current performance without considering the future impacts of these policies? Should terminating or reducing spending be the primary recommendation of a framework designed to dictate public funding? Should public policy evaluations adhere to the logic of a private firm?

Reflecting on these questions is fundamental to determining the best way to integrate evaluations and the budgeting process. In fact, acknowledging the abovementioned challenges would either end current practices or at least make ongoing choices more transparent.

Final remarks

All of the above questions are crucial reflections to assess the existing political space in policymaking. This space has limitations that are fundamentally philosophical and theoretical. From a philosophical perspective, we still observe a close tie between economics and positivism, which affects the understanding of scientific knowledge. Consequently, quantitative methods gain prominence as sources of information and decision-making rules for policymaking. For instance, CBA determines the merits of a policy by deciding whether it is efficient. Moreover, these analytical models are considered impartial and value-free, even though the research object, the reasoning applied, the chosen method, and the internal elements of these methods exemplify how the subjectivity of the researcher strongly influences the results.

This acknowledgement does not advocate for the abandonment of quantitative research methods but rather for their improved utilisation. From a postpositivist perspective, policymaking and its appraisals and evaluations are political processes with a technical component. In this sense, estimation models are relevant but ancillary in the decision-making of government actions, as the most pertinent aspect of this process is political. Within this framework, efficiency is merely one of many elements, not necessarily the most relevant.

Secondly, the limitations on the political space for policymaking are grounded in theory. The dominant neoclassical approach has developed a theory that centralises market economy relations as defined within its framework. Consequently, a narrow space has been created for government interventions. Generally, market failure justifies the actions of the State, which must adhere to an essential allocative judgement: efficiency. This principle became a guideline for policymaking. However, the understanding of what efficiency entails has evolved over time. In a Pareto sense, it involved a distributive perspective that upheld the status quo. Conversely, it restricted policymaking because worsening the situation of at least one person could challenge policy recommendations.

In this context, Kaldor-Hicks' efficiency emerged as a concept in applied economics, suggesting that efficient interventions might produce losers even if the net result were positive. That is, overall welfare improved, irrespective of who benefited. Consequently, another approach to distributive justice was introduced. Furthermore, the potential for Pareto improvement shifted from utility to wealth, enabling comparability without necessitating interpersonal comparisons that could favour redistribution policies.

Adapting the concept of efficiency was essential in connecting theory and applied economics with broad implications for policymaking. The more flexible principle of efficiency facilitated a shift from a hands-off approach to an active, market-oriented State. As efficiency could be measured, methodologies like CBA were developed to evaluate the net outcomes of an intervention and perspectives on market relations were incorporated into these methodologies, which form the subject of the next chapter.

The implications of using the market as a benchmark and efficiency as a guideline for policymaking are considerable. It influences the political space available for policymaking in numerous ways. Firstly, it imposes a particular line of reasoning to address social issues that equates the government with private firms. Naturally, the discourse around minimizing waste is quite appealing, but placing efficiency at the forefront of policymaking does not solely pertain to this. In fact, it suggests that every aspect of human relations will be subjected to a market exchange rationale, even concerning social policies.

In this sense, defining efficiency as a sufficient and sole condition for social improvement impoverishes policy analysis and reduces our notion of well-being to one variable: money. Indeed, income is fundamental to welfare assessments, particularly in areas with weak social safety nets and limited access to universal systems. However, emerging research increasingly demonstrates how various other factors significantly influence our sense of well-being. This evidence cannot be overlooked in discussions of policymaking. Furthermore, the debate surrounding the meaning of welfare in contemporary society must take into account other fields of expertise.

Understanding the challenges to neoclassical economics can significantly explain why pursuing efficiency does not make sense. By not recognizing the neoclassical explanation of the distribution between the factors of production and the tendency to fully employ them in competitive markets, allocative efficiency becomes meaningless since allocative changes would not necessarily involve real costs. Moreover, recognizing that families do not act like private firms, in the sense of presenting a constant maximizing behaviour in every decision, strongly defies the theoretical construction of the methodologies applied to estimate efficiency, notably the Cost-Benefit Analysis. The acknowledgement of this argument makes the extension of the symmetrical behaviour of private firms imposed on families to policymaking arbitrary.

Lastly, the discussion regarding efficiency as a guiding principle of policymaking is not limited to theoretical differences in the functioning of capitalist economies. It has practical implications for the financing of public policies, necessitating a full understanding of what prioritising efficiency truly entails. The Spending Reviews framework is increasingly popular in shaping public budget allocations, particularly in the context of pressure to reduce expenditures. Although this framework is not rigid and can incorporate multiple principles and objectives, it has been employed to cut spending based on controversial methodologies, imbuing the analysis with an air of unquestionable technicality.

2. EFFICIENCY AND METHODOLOGY: FROM THEORY TO PRACTICE

The connection between evaluations and budgeting is not straightforward. Even though there is a broad consensus over the merits of analysing the results of public interventions and, thus, whether they are worth financing, many elements behind this process remain ignored. The benchmark, the methodology, the belief about the nature of scientific knowledge, the context and the interpretation of results are crucial and far from impartial dimensions of this process.

Chapter One has discussed the philosophical and theoretical foundations of the broadly accepted idea that efficiency should be a leading principle of policymaking. More specifically, it debated the emergence of the flexible Kaldor-Hicks concept of efficiency, enabling this principle to integrate the decision-making process in practice. It also responded to claims for making policymaking more objective and neutral, in line with positivism, which made quantitative methods the most reliable source of information.

Chapter One also discussed the implications of having the market and, thus, efficiency as a benchmark for policymaking. First and foremost, it highlighted the effects this line of reasoning has on the political space opened to policymaking. It restricts the situations in which the State is allowed to act, the rationale governments must apply to deal with social problems, the definition of what society's well-being means, and ultimately, how it impacts the financing of public policies.

Economic efficiency has two main perspectives: allocative and technical. The allocative aspect of efficiency defines the value of a policy—whether a program is worth financing. The technical element relates to the performance of the units administering the policy, embracing the idea of value for money.

Discussions regarding allocative and technical efficiency methodologies aim to clarify why reliance on this principle is insufficient to direct the budgeting process. Chapter Two seeks to link the foundations of this rationale with the methodologies used to estimate efficiency. In doing so, this chapter intends to reveal the effects of applying market-oriented methods in policy analysis.

Furthermore, even if one agrees that efficiency ought to be the central element of evaluations, the methodological shortcomings significantly undermine the method's ability to produce reasonable results without depending on human judgment. Consequently, there is no value-neutral analysis, and the political dimension of evaluations becomes apparent.

Understanding the connection between the elements of this chapter and the previous one is crucial, as it influences how society chooses to manage resources. Despite assertions that efficiency is merely a technical guideline for setting priorities and minimising waste, it reflects a system of beliefs regarding how modern societies ought to function.

2.1 Allocative Efficiency: The Cost-Benefit of Public Funding

Cost-benefit analysis is the most common method for determining allocative efficiency. This methodology equalizes the State's to a firm's decision-making process by adopting a maximizing market-oriented rationale as a guideline. Social benefits replace revenue, and opportunity costs substitute costs to create a pros and cons list that drives decisions. It serves as a tool for "informing decisions about the most efficient allocation of resources" (Bamberger, Clark and Sartorius, 2004, p. 20). It "assesses the impact of different options on social welfare"

(United Kingdom, 2022, p. 40). Furthermore, it provides "a framework for assessing the relative efficiency of policy alternatives" (Boardman et al., 2018, p. 28).

Thus, CBA is the methodology for assessing the neoclassical perspective of efficiency in applied economics. It connects neoclassical welfare economics with public policy analysis from an allocative standpoint. It addresses neoclassical allocation in the real world, where efficiency is a sufficient condition for welfare improvement.

According to OECD (2006), the fusion of CBA and decision-making responded to increasing social pressure for government efficiency, aligning with the previously discussed notion of efficiency as avoiding waste.

The concepts of consumer surplus and externality are the underlying elements that distinguish CBA from traditional profit-and-loss accounting. The difference is CBA's capacity to identify and quantify factors previously overlooked in economic analysis, such as externalities. The main idea is to estimate the net benefit of projects, incorporating social assessments rather than focusing solely on financial aspects (Mishan and Quah, 2020).

As Pigou (1920) pointed out, externality refers to the divergence between private and social costs. CBA addresses this issue by estimating non-market elements (externalities) to capture all relevant aspects that influence social welfare. That is, CBA encompasses all relevant costs and benefits involved, even those not traded in the market. It assigns monetary values to each element and, consequently, provides a sort of "pros and cons" list with a net result. The essence of CBA is that benefits signify increases in human well-being, while costs denote losses in well-being. Therefore, the net result indicates whether the policy represents an improvement in general welfare.

The foundation of this reasoning lies in the hypothetical compensation principle, which is explored in Chapter One, to address real-life situations in applied economics. When benefits surpass costs, the aggregate outcome is positive, which underpins the Kaldor-Hicks efficiency principle. Consequently, the positive net results of CBA signify efficiency in allocating public resources to the assessed policy and suggest welfare enhancement for society.

Formally, the cost-benefit criterion can be expressed in this simple notation $\sum V_i > 0$; where V represents the net present value of the project. Positive values of V indicate that benefits exceed costs, while negative values of V suggest the opposite. The positive values of V are potential Pareto improvements that will make the project efficient (Mishan and Quah, 2020).

The general idea of weighing the benefits and costs to evaluate the merits of something is quite appealing. In this sense, CBA aligns with common sense, which is that every decisionmaking process must involve assessing the pros and cons of a change. However, understanding this general concept does not elucidate what adopting CBA truly entails. This section tackles the internal methodological controversies that involve selecting allocative efficiency as a guiding principle for policymaking.

2.1.1 The basic framework

CBA is a general framework consisting of basic principles that lead to an approach rather than a specific method. In other words, there is no singular way to apply CBA. Instead, a common framework exists with various options that the analyst makes throughout the process (Sen, 2002). Mishan and Quah (2020) noted that a CBA analyst must address seven key questions: a) Who should be the reference target group?; b) What are the benefits and costs?; c) How can these benefits and costs be measured?; d) What should the discount rate be?; e) Are there equity concerns?; f) How do we manage uncertainties?; g) What investment decision criteria do we adopt?

The first observation is that the analyst makes choices that significantly influence results by addressing these underlying questions. The same policy can be regarded as efficient or inefficient depending on each CBA analyst's personal decision. In fact, every policy appraisal or evaluation involves individual choices, such as those regarding the scope, the data, the targeted group, and so on. Issues occur when one attempts to classify a methodology as neutral and objective, as this interpretation assumes the absence of potential conflicts or divergences involved. Furthermore, it is concerning when society determines that this purported objective methodology ought to dictate what is worthy of funding.

Mishan and Quah (2020) illustrate how individual value judgments significantly affect CBA results. They delve into the history of the U.S. Executive Orders from various presidents to demonstrate the arbitrariness in determining an efficient allocation of resources according to this methodology. For instance, President Reagan explicitly required the efficiency criterion in the Regulatory Impact Analysis. In a similar vein, Clinton mandated that all costs and benefits of available regulatory alternatives be taken into account. On the other hand, Bush prioritised

costs by asserting that regulatory measures affecting economic growth should be eliminated. President Obama revised an Executive Order highlighting the necessity to consider equity and distributional impacts on CBA analysis. Trump again shifted the government's orientation towards costs, establishing controversial calculations in cost-benefit analyses. For instance, the authors reference environmental estimates concerning greenhouse gas emissions that overlook global impacts, focusing solely on the effects within the U.S. In doing so, he secured the costbenefit analysis's efficiency stamp by disregarding the consequences of greenhouse gas emissions in other countries.

With this brief history of general guidance of CBA analysis within the U.S. government, the authors end the frequently claimed objectivity of the CBA methodology in determining efficient allocation and social welfare. The changes in the Executive Orders demonstrate how political preferences influence CBA, casting a shadow over the notion of an objective methodology based solely on technical criteria elements.

Boardman et al. (2018) acknowledge how individual viewpoints influence CBA, but they classify the effects based on the bureaucrat's role within the government. For them, the perception of costs and benefits differs depending on whether the bureaucrat is an 'analyst', 'guardian', or 'spender'. Guardians and spenders possess biases intrinsic to their professional functions. Guardians are identified as primarily operating in budget offices, where concerns regarding public expenditure shape CBA analysis. Consequently, social benefits outside the budget are frequently overlooked.

Conversely, Spenders typically work in the service department and maintain a more positive perspective on public spending. According to the authors, they incorporate this "bias" into their CBA analysis. Spenders are linked to proponents of economic impact analysis, which assesses the effects of policies on economic activity, often utilizing an input-output framework and estimating multiplier effects.

For Boardman et al. (2018), impact analysis fails to address social welfare. Indeed, economic impact analysts do not seek to measure neoclassical economic efficiency, as many do not regard the potential Pareto efficiency principle as a reasonable estimate for social wellbeing or as an appropriate concept to define what an efficient allocation entails.

It is crucial to state that the impact analysis approach is more aligned with a postpositivist science perspective, which breaks the binary logic of analysis. In this sense, an

impact analysis is uncommitted to efficiency as a policymaking benchmark. Other benchmarks can be pursued, such as reducing income inequality, fighting unemployment, stimulating economic activity, etc.

Furthermore, there is no definitive answer arising from the methodology's estimation. In other words, the method itself does not yield an answer in a binary fashion as CBA does efficient or inefficient. Rather, multiple effects result from social choices. These should be measured, and their impacts discussed in order to arrive at a conclusion regarding the merits of a project or policy. Thus, any technique serves as a tool to support the analysis rather than being the central figure in the decision-making process.

Finally, Boardman et al. (2018) classify the 'analysts' as professionals aligned with the 'correct' CBA path, supposedly devoid of biases. Although there is a lack of further explanation regarding this bureaucrat profile, implicit in this statement is the distanced and neutral researcher discussed in Chapter One. Therefore, for the authors, there exists a specific, albeit unidentified, type of public agent capable of making 'the correct' CBA analysis.

It is not uncommon to overlook the difficulty of distinguishing the evaluator from an ordinary person shaped by thoughts, beliefs, and socially constructed intrinsic values. Nevertheless, this reality is surprising in the context of CBA analysis because of the numerous personal choices the evaluator must make to achieve the final binary result and ascertain whether the policy is efficient or not.

The fundamental CBA framework consists of a series of commonly shared steps. Each step necessitates decisions involving methods and personal choices. Numerous sources outline and explore these key steps. (Mishan and Quah, 2020; Campbell and Brown, 2023; United Kingdom, 2022; Brasil 2018a; Brasil, 2018b).

Table 5 - The major steps in CBA

1. Explain the purpose of the CBA		
2. Specify the set of alternative projects		
3. Decide whose benefits and costs count (specify standing)		
4. Identify the impact categories, catalogue them, and select		
metrics		
5. Predict the impacts quantitatively over the life of the project		
6. Monetize (attach dollar values to) all impacts		
7. Discount benefits and costs to obtain present values		

8. Compute the net present value of each alternative		
9. Perform sensitivity analysis		
10. Make a recommendation		
Source: Boardman et al. (2018, p. 5)		

The first step concerns the 'rationale' behind public intervention—that is, why changes to the status quo may enhance welfare. The starting point is that markets and individual selfinterest efficiently guide the allocation of resources in society. Interventions are thus welcome in cases of market or government failure, the latter being identified as a clear opportunity for enhancing current policies, as discussed in Chapter One. This initial step serves as the rationale for the CBA study. Here, we clearly delineate the defined role of government action, limiting options to those deemed 'efficient' policies.

The next step is to identify alternative projects that offer the counterfactual for the analysis. Typically, this is the status quo or the absence of changes. A critical aspect of the CBA is highlighted here: the methodology's incremental or marginal approach. This aspect is particularly problematic, especially when addressing structural changes in society, as the CBA tends to exhibit a status quo bias.

The Economics of Energy Innovation and System Transition (EEIST) consortium discussed this status quo bias in a report that addresses the transition in energy technologies (Grubb, M. et al., 2021). Previously characterised as inefficient, costly, and limited, wind energy increased from 1% to 10-15% of electricity supply in Brazil and Europe over the past decade due to declining costs, transitioning from being 'the most expensive way to reduce carbon emissions' to offering 'the cheapest electricity in history' (p. 10). They further assert that "our key finding is that these have been achieved despite, not because of, traditional approaches to policy assessment" (p. 10). By traditional approaches, they specify public research and development (R&D) efforts and cost-benefit analysis (CBA). The marginal feature of CBA renders the methodology incapable of capturing innovative benefits. Consequently, CBA analyses are "systematically biased towards the status quo in the context of transitions" (p. 10).

Another relevant aspect of comparing alternative projects is the common notion that CBA will provide a ranking with all policy alternatives. However, this rarely occurs. Analysts do not compare projects in different fields, such as health care and national defence, because of the varying objectives of these policies. How could one classify a national defence policy as superior to a health care one on efficiency grounds? This comparison is unreasonable. Society can, indeed, prefer one over the other. It simply won't be based on efficiency, as that does not make sense. Therefore, CBA is not an appropriate methodology for selecting policies with diverse objectives.

The third step of CBA involves identifying the costs and benefits to be included. This stage is subject to various individual judgments that significantly influence the outcome. Typically, the literature highlights the need to include 'all relevant' costs and benefits without specifying what those might be. There is no common ground, even in fields such as health or education. It seems reasonable that this is the case, as policies may encompass several particularities, including regional differences. The problem is trying to classify the methodology with an objective label when it clearly contains subjective elements.

Another key step involves identifying the impact categories, cataloguing them, and selecting the relevant metrics. The final output of this stage will be an impact matrix that summarises the effects of each policy alternative. An important observation made by Boardman et al. (2018) is that impacts are only acknowledged if they positively or negatively affect human beings. Flora and fauna, for instance, do not possess legal standing. Consequently, projects that lead to the extinction of entire species should only have this impact assessed if some individuals perceive this as a cost. This observation raises significant ethical concerns that should not be overlooked.

The fifth procedure concerns the temporal dimension of the project in question. The impacts should be assessed during the discount period, which is the project's life. Therefore, this step involves quantitatively predicting the impacts over the project's duration. Naturally, the longer the project lasts, the less accurate those estimations tend to be. Consequently, CBA predictions exhibit greater sensitivity to extended time horizons.

Attributing monetary value to costs and benefits is a highly controversial seventh mission in CBA. Generally speaking, relying on willingness to pay to measure benefits presents the primary point of disagreement from a broader perspective. Additionally, estimation techniques – the so-called valuations - have numerous shortcomings. The following section explores this particular step in greater depth when discussing the valuation techniques.

The eighth stage involves calculating the net present value of each alternative. The primary objective here is to summarise all impacts in the present, enabling comparability. Therefore, future costs and benefits are discounted to derive their present value. This step is

also subject to criticism, primarily due to the analyst's choice of discount rate. The temporal dimension is closely linked with intergenerational concerns, as future impacts tend to be underestimated. This intertemporal feature of CBA is also the focus of a more detailed discussion in this chapter.

Performing sensitivity analysis is the subsequent step in addressing uncertainties arising from the various assumptions made during the CBA process. Different scenarios and parameter estimations that could alter the overall evaluation from 'efficient' to 'inefficient' are frequently employed.

Finally, CBA makes recommendations. The most common decision-making rule considers society as risk-neutral and endorses proposals with a higher net present value (NPV) because, in this framework, the NPV leads to a more efficient allocation of resources. Thus, CBA presents a normative perspective on how resources ought to be appropriated. From this standpoint, the argument for economic efficiency underlies the decision-making process.

2.1.2 Valuation: assigning monetary value to non-market goods

Assigning non-existent values or adjusting existing ones is crucial in the CBA. The valuation process includes the notion that trade-offs in the public sector must be measured accurately to facilitate the decision-making process, thereby achieving an efficient allocation of resources.

Campbell and Brown (2023) observe that public goods lack a market price because they are non-rival and non-excludable. Consequently, they are referred to as 'externalities' in the sense that they exist outside the market system. Nevertheless, since CBA must encompass all pertinent impacts on welfare, the methodology must estimate their values. Thus, they should be integrated into the shadow-pricing process⁵ of the efficiency analysis. For instance, the OECD (2006) provides examples of early CBAs where shadow prices were utilised to evaluate the value of accident risks or time savings associated with certain policies.

Furthermore, as discussed in Chapter One, the neoclassical framework also encompasses certain rival and excludable goods when market failures are identified. For instance, the positive externality resulting from the provision of health services and education

⁵ Shadow prices are the estimates given by economists to benefits or costs that are either unpriced or unsatisfactorily priced.

justifies public intervention. In such cases, these positive effects must be quantified to legitimise policymaking.

Nevertheless, assigning monetary values to non-market goods remains a highly contentious issue. Firstly, it predominantly depends on Willingness to Pay (WTP), a principle that raises significant equity concerns by favouring the wealthier. Furthermore, the valuation process has methodological deficiencies that undermine the purported objectivity of CBA and the precision of the estimates, thereby challenging this method's capacity to determine the allocation of resources within society and, ultimately, social welfare.

CBA must encompass all impacts valued in monetary terms. The conceptual foundation for these valuations is WTP and opportunity cost. Boardman et al. (2018) emphasise that there is no singular method for undertaking this monetisation process. For instance, when impacts alter quantities consumed in an undistorted market, prices can be utilised to assess policy effects. Conversely, when markets are distorted, shadow prices may be employed⁶. Furthermore, there are valuation techniques available to estimate the effects of impacts external to the market system. "*More often, analysts must piece together evidence to support predictions from a variety of sources and use shadow prices with varying degrees of provenance*" (Boardman et al., 2018, p. 193).

Quantifying the intangible impacts of public policies in monetary terms is an increasingly relevant aspect of appraisals and evaluations. Numerous important effects of policymaking cannot be directly observed through traditional market mechanisms such as price and consumption, leading to the emergence of various methods for estimating the value of these impacts. (OECD, 2006).

Mishan and Quah (2020) present a chart summarising the valuation methods for nonmarket goods (externalities). According to the authors, the three most common ways to measure the economic impacts of non-market goods are by deriving demand curves to assign prices. However, there are also approaches that do not rely on demand curves.

Concerning the demand curve methods, which are more frequently applied, the Contingent Valuation Method (CVM) depends on expressed preference techniques to assign

⁶ Boardman et al. (2018) clarify that many impacts have already estimations that analysist resort to. "researchers have estimated a number of shadow prices for commonly encountered impacts. Some shadow prices, such as the willingness to pay for reductions in mortality risk (a basis for estimating the value of a statistical life), the social cost of noise, or the opportunity cost of commuting time, have been estimated in a sufficiently large number of studies to make meta-analyses possible" (p. 192).

monetary values to non-market goods. This technique resorts to surveys to assess individual preferences. On the other hand, the Travel Cost and Hedonic Pricing methods are founded in revealed preference approaches. In these cases, the valuation is estimated through the observation of actual behaviour in the market system.

Figure 5 - Valuation methods for non-market goods



Source: Mishan and Quah (2020, p. 212)

The first of the most commonly applied techniques is the CVM, which estimates WTP (Willingness to Pay) and WTA (Willingness to Accept) to derive the demand curve. The monetary values arise from surveys based on hypothetical situations. Surveys directly question individuals about their WTP for a policy change. This is why CVM is considered an expressed preference method. The WTP and WTA of the sample are projected onto the entire population with adjustments where necessary. For instance, information regarding income levels and personal characteristics may be included to establish a 'typical' respondent. (Mishan and Quah, 2020; Campbell and Brown, 2023).

Mishan and Quah (2020) illustrate the CVM through a hypothetical example of a citywide greening project. Respondents are asked about their willingness to pay for this project across a range of costs (\$5, \$10, \$20, \$50, or \$100, with each amount presented to 20% of the participants). In this hypothetical example, the city has one million citizens. The results below display the demand curve derived from the responses and the aggregate WTP of \$34,750,000⁷.

Figure 6 - Demand curve for a hypothetical greening project

⁷ Total WTP are estimated by summing the areas of the four trapezoids.



Source: Mishan and Quah (2020, p. 212)

Four types of biases are commonly discussed concerning CVM. The hypothetical market bias pertains to the lack of reality in the survey. Respondents may provide misleading answers as they are not faced with actual market situations. Additionally, the strategic bias relates to the possibility of underestimating or overestimating their WTP if respondents believe they can influence the results. The design bias is associated with survey characteristics and how they are presented, namely, the ability of the survey administrator to influence results. Finally, the 'part-whole' bias refers to the observed inconsistency of respondents assigning the same WTP for a single component as for the entire good. For instance, the same WTP value is given to recreational fishing in one river and to the whole river system (Campbell and Brown, 2023).

The authors acknowledge that improvements to mitigate biases do not address a fundamental concern: the ability of humans to express their preferences for non-market goods, such as environmental or health services, in the same way they would in a supermarket or shopping centre. This poses a challenge to the CVM as an accurate method for assessing costs and benefits.

The second valuation technique for non-market goods is the Hedonic Pricing Method (HPM). This method attributes value by considering differences in one element of existing markets as proxies for prices. For instance, the price related to noise would be the difference between the prices of two properties with similar characteristics, apart from their ambient noise

levels (Mishan and Quah, 2020). When all other possible factors that could influence prices are controlled, any price variation is ascribed to the specific element being valued. Fluctuations in WTP would reflect an implicit measure of the unit cost of the additional noise in this instance (Campbell and Brown, 2023). "Then, the slope of the hedonic price function with respect to each characteristic is equal to the implicit price" (OECD, 2006, p. 94).

The OECD (2006) highlights that the HPM has been utilised to assess various nonmarket factors in the property market, including traffic or aircraft noise, air pollution, water quality, geographical location, and so on. Furthermore, this technique has been employed in the labour market to quantify the value of mitigating the risk of death or injury. In these instances, estimates are based on wage differentials among jobs with varying risk exposures.

The most common criticism of hedonic pricing is the reliance of the valuation on perfect competition, as any imperfection in market conditions would imply that price differentials do not necessarily reflect the attribute being valued (Mishan and Quah, 2020). In this regard, the OECD (2006) emphasises that the HPM depends on perfect information. For instance, in the case of the wage-risk premium, workers may not be fully aware of the risks they encounter in their workplace. Consequently, the wage differentials may not accurately represent their valuation of risks.

Campbell and Brown (2023) identify three weaknesses of this valuation method. Firstly, it relies on a 'high degree of statistical knowledge'. Furthermore, it suggests a linear relationship among the attributes, wherein the sum of individual elements constitutes the price. Similarly, the OECD (2006) highlights the multicollinearity issue, indicating that non-market characteristics often occur together. For example, properties near roads experience increased noise and air pollution, complicating the task of isolating their effects. Lastly, Campbell and Brown (2023) emphasise that the methodology assumes a continuous range of choices encompassing every possible combination of attributes for the specific object under analysis, which appears to be an unreasonable assumption.

Recently, the travel cost method (TCM) has been employed to estimate the value of recreational areas, such as mountain resorts and parks. According to the OECD (2006), the method is based on the idea that travel and recreational areas are complementary. To achieve this, a surrogate market is established using travel information to value services related to environmental assets. The primary objective is to derive the demand curve by using travel costs stratified by distance and population as a proxy for price. Changes in visitation costs are linked

to the pricing process. This method connects the individual's demand for environmental goods with the costs incurred during the trip. Thus, the demand curve indicates that an increase of one dollar in travel costs is akin to a rise of one dollar in prices (Mishan and Quah, 2020; Campbell and Brown, 2023). For instance, Mishan and Quah (2020) utilise data on the number of trips per million and the travel costs associated with departures from various cities. They perform linear extrapolation to estimate the relationship between trips and costs.

The authors discuss several shortcomings of this valuation technique. Firstly, by relying on actual trips, the travel cost method overlooks option demand, which refers to what individuals would be prepared to pay to preserve the site, even if they do not utilise it. Given the importance of environmental protection, particularly in uncharted areas, neglecting option demand is a significant issue. Additionally, by focusing solely on trip expenditures, the method tends to underestimate the true value of the good, as the analysis fails to take other factors into account. Lastly, the travel cost method assumes uniform preferences across entire populations, which is unreasonable (Mishan and Quah, 2020).

Campbell and Brown (2023) raise three additional concerns regarding TCM. They emphasise the method's limited capacity to estimate the value of individual site attributes. Furthermore, the complexity of valuing a trip with multiple destinations or motives represents another significant concern. The third issue pertains to the omission of potential substitute sites, which could lead to a biased analysis.

In summary, the three most common valuation techniques used to estimate the costs and benefits of non-market goods in the CBA framework face significant criticism. Given that the valuation process is likely the most crucial aspect of CBA, these comments raise important questions about the methodology's capacity to yield consistent and reliable results, particularly since these findings are intended to inform the decision-making process in policymaking and, ultimately, the funding of public policies.

In fact, Flyvbjerg and Dirk (2022) analysed a large dataset on infrastructure investments that employed CBA analysis to examine how accurate and unbiased the estimations were. They collected data from 2,062 projects in 104 countries across six continents concerning infrastructure investments in eight areas: bridges, buildings, bus rapid transit, dams, power plants, rail, roads, and tunnels. The data spans from 1927 to 2013. They found that forecasters tend to underestimate costs and overestimate benefits across every type of investment, on average, revealing a "strong and consistent bias" (p. 178). With 'overwhelming statistical

significance', the estimations were indeed biased and inaccurate. As a result, they argue that this systematic and significant bias is likely to lead to a misallocation of resources: "*Then, the assumption that cost-benefit analysis is a rational way to improve resource allocation would be a fallacy*" (p. 174).

Another essential aspect of the valuation process is the reliance on WTP. Two pertinent debates surround this element. The first concerns the correct choice between WTP and WTA when estimating costs and benefits. The second pertains to unequal weighting in the valuation process, as WTP is influenced by the ability to pay, likely favouring the wealthier.

In general, WTP refers to the amount that an individual is willing to pay to have the policy or project approved. Willingness to Accept (WTA), on the other hand, is the amount that an individual is willing to accept as compensation for personal losses resulting from the approval of the policy or project. However, implicit in this interpretation is the notion that individuals have the right to their initial situation.

The OECD (2006) emphasises that CBA typically assumes the right to the status quo. Therefore, a policy that benefits individuals should generally be evaluated based on their WTP. Conversely, if an individual possesses the right to the new situation, then the WTA to relinquish the benefit is the appropriate reference for the valuation. In this regard, they assert that the right to a new situation is indeed "*at the heart of environmental debates*" (p. 158), such as the right to clean air or water.

Property Rights	Policy $\uparrow Q$ or $\downarrow P$ (better)	Policy $\downarrow Q$ or $\uparrow P$ (worse)
Right to the status quo	WTP for the change	WTA to tolerate loss
Right to the new situation	WTA to forego the benefit	WTP to avoid the loss

Table 6 - Link between property rights and WTP and WTA.

Source: OECD (2006), p. 159. Adapted.

This debate over the appropriate basis for estimating costs and benefits—specifically, the discussion surrounding property rights—would be irrelevant if WTP and WTA were empirically similar. However, an increasing number of studies provide empirical evidence that WTP and WTA differ, often by considerable amounts. One reason for this discrepancy is the

income effect; the greater the income elasticity for a good, the larger the disparity between WTP and WTA⁸.

In light of this finding, it seems problematic that CBA often relies too heavily on WTP without considering the relevant debate regarding individuals' rights. This can alter the reference point for property rights and, consequently, the accurate estimation measure. Furthermore, the OECD (2006) highlights that estimates in CBA frequently equate the financial and opportunity costs of projects, which is "*at best an approximation of true opportunity cost*" (p. 156).

Another relevant aspect of CBA concerns the intrinsic reliance of WTP on the ability to pay, which results in overweighing the individual preferences of the wealthier segment of society, thereby granting this group more influence in collective decisions. In fact, WTP reflects the rationale of market discipline in applied economics for measuring the benefit estimates required by CBA. In this sense, it not only reduces welfare to a single metric—money—but also transposes market inequalities into policymaking. With citizens regarded as customers in an uneven society, the disparity in the ability to pay highlights the likelihood of wealthier individuals exercising more influence than the poor in pursuing their interests.

As Heinzerling and Ackerman (2007) noted, "*It is hard to see what WTP could contribute to an objective approach to welfare. If WTP is meaningfully connected to ability to pay, then the rich will have greater WTP for trifles than the poor will have for necessities*" (p. 368). In the same direction, the OECD (2006) points out that CBA analysis does not consider how the ability to pay constrains WTP sums.

More interestingly, the organisation stress an implicit feature that shows how income inequality might influence results: "*In this system, mean WTP is preferred to median WTP as a more accurate reflection of the variance in preferences across the mass of individuals whose aggregation is considered to represent society's preference*" (OECD, 2006, p.118). This implies that the mean overvalues a minority with 'stronger preferences'. On the other hand, by using the median, the methodology would capture the value that most people are willing to pay.

Frank (2020) argues that WTP reflects how CBA addresses the intensity of preferences, an essential attribute for addressing welfare. Implicit in his argument regarding intensity is the

⁸ Discussing all the reasons is beyond the scope of this work. For more details see OECD (2006, chap. 11).

notion that it is reasonable for some people to hold more influence than others in public decisions. WTP would exhibit this difference. He acknowledges that the principle underrepresents the interests of low-income individuals, favouring the wealthiest. However, he asserts that the Pareto criterion is only applicable to initial allocation. By justifying that the methodology should be applied to marginal analysis, equity concerns would be relegated to further decisions beyond the economic framework. He states that distributive issues rely on more profound changes in the tax system, the labour market, and so on. Therefore, it would be "counterproductive to interfere with Pareto-improving results that can be obtained through cost-benefit analysis and associated stakeholder negotiations" (p. 562).

Cook (2022) clarifies how WTP, by relying on an individual's initial endowments rather than utility, ensured a neoclassical efficient allocation that favoured the wealthier, despite the higher marginal utility of an extra dollar for the poor:

Yet in turning to these subjective notions of willingness to pay or accept, Hicks was actually ensuring that the Kaldor-Hicks criterion would often favour the wealthy. The reason for has been noted by numerous critics: Willingness to pay is extremely dependent on ability to pay and often reflects the initial endowments of the individual and not the level of his utility from the proposed policy change or consumer good. On the flip side, a poor man who is constrained by a meagre income will accept a far lower level of compensation than a rich man since – ironically – the marginal utility of a dollar for him is much higher since he is poor. In the discarded 'Benthamite' welfare economics of Pigou, measuring welfare or efficiency in units of marginal utility of income inherently favoured the poor since, as we have seen, they would enjoy that final dollar more than the rich. With Kaldor-Hicks and the New Welfare School the exact opposite was true since now welfare and efficiency were measured by willingness to pay which favours the rich (p.11).

Therefore, reliance on WTP reinforces inequalities (Heinzerling and Ackerman, 2007; Petri, 2021; and Cook, 2022). Even the potential for weighting WTP does not address equity concerns due to the inherent arbitrariness. Furthermore, as discussed in Chapter One, the compensation principle remains hypothetical. Thus, combining greater power for the wealthier in determining what constitutes efficient allocation in practice with the absence of actual side transfers to compensate the disadvantaged renders CBA a highly controversial methodology in relation to equity concerns in policymaking.

Moreover, the valuation techniques used to assign monetary values to non-market goods possess several methodological shortcomings that undermine their capacity to accurately estimate values for these goods, which are essential to the results of CBA. Consequently, even if we accept that market relations should serve as the benchmark for public policy, placing a price on non-market goods is not without its unreasonable aspects.

In light of this finding, it seems difficult to accept the CBA's valuation process as a reliable estimate of the social impacts arising from policymaking. Moreover, it appears even more challenging to accept that these estimations can inform the decision on what allocative efficiency entails in real-life situations.

2.1.3 Dealing with time: intertemporal discount

Intertemporal discounting refers to the valuation of time. Regarding CBA, the Social Discount Rate (SDR) deals with projects with different maturities. SDR integrates timedifferentiated costs and benefits to allow for the evaluation of the long-term impact of policies. Two arguments underpin this concept. First, from a consumer perspective, the assertion is that society prefers goods and services now rather than later. This viewpoint accounts for pure time preference, expressing impatience and the risk of not being alive in the future, as well as the expectation of increasing future consumption; hence, the marginal utility of consumption would diminish. In summary, this perspective addresses how much society is willing to postpone consumption, as defined by the Social Rate of Time Preference (STP).

The second argument that supports discounting is a producer's perspective, which asserts that investing in a project incurs an opportunity cost. Consequently, costs and benefits should be estimated in such a way that the expected return at least equals the opportunity cost. The Social Opportunity Cost of Capital (SOC) illustrates this producer's viewpoint. In a context of perfect competition, the market interest rate reflects the opportunity cost of investment. Thus, it would serve as the applicable Social Rate of Discount (SRD) to allocate resources efficiently, equating the STP and the SOC (ADB, 2013).

However, in the real world, the rates differ, and the choice of the SDR is one of the most significant controversies surrounding CBA. We might conclude from the theoretical foundations supporting the STP and SOC that identifying the appropriate SDR equates to finding the proper substitute for the market interest rate in the absence of perfect competition. Here, once again, we observe that the CBA approach limits government actions to a market benchmark. It equates the justification for the state's and firms' intertemporal investment decisions.
Much discussion has occurred in the literature concerning which SDR is suitable for CBA evaluation. The fundamental aspect distinguishing the approaches is the assessment of how public projects influence consumption, private investment, and the cost of international borrowing. There are four approaches: a) the STP, b) the SOC, c) the weighted average approach, and d) the shadow price of capital (SPC) approach (ADB, 2013). Essentially, the STP overlooks the cost of funding, addressing only the demand side, while the SOC focuses solely on the production side, neglecting the consumption trajectory. The other methods attempt to reconcile both approaches, addressing the potential blind spots of each (Groom et al., 2022).

The STP reflects the rate at which society is willing to trade present consumption for future consumption. Two empirical methods are used to estimate the STP. The first is the aftertax rate of return on government bonds or other low-risk assets. The second method involves Ramsey's formula, which arises from a growth model and comprises two components. The first component indicates pure time preference through a utility discount parameter. The second component is the product of the elasticity of marginal utility of consumption and the annual growth rate of per capita real consumption (ADB, 2013).

STP =
$$\rho + g\varepsilon$$
, where:

 ρ = is the rate at which future utility is discounted;

g = is the percentage change in per-capita consumption;

 ϵ = is the absolute value of the elasticity of the marginal utility of consumption with respect to changes in consumption.

In other words, the first component signifies impatience, while the second reflects society's preference for smoothing consumption over time. Thus, STP denotes the rate that maximises the present value of current and future per capita consumption. Concerning government projects, the purpose of employing this rate is to achieve an optimal consumption growth rate (Boardman et al., 2018).

Once again, as observed in the valuation process of costs and benefits in CBA, two market mechanisms are evident: reducing welfare to consumption and limiting well-being to choice. Moreover, the notion that individuals smooth consumption over a lifetime overlooks various socioeconomic backgrounds and standardises distinct realities within and across countries. For instance, unmet demand in developing nations would argue against the smoothing hypothesis. An essential assumption of STP is that government funding for public projects originates from reduced consumption by individuals. This hypothesis has faced criticism for considering the opportunity cost solely in terms of foregone consumption, neglecting the widely accepted notion in mainstream economics that public spending crowds out private investments. As ADB (2013) highlights, if public investment occurs at the expense of displacing private investments, then the opportunity cost should also account for the displaced private investment, which the SOC can address. Consequently, STP is generally lower than the SOC.

Implicit in the choice to use the SOC as the SDR is the theoretical assumption that public projects crowd out private investments. For this reason, public investment should yield results at least as favourable as those of private investment. Otherwise, total welfare could be enhanced by reallocating resources to the private sector, thereby making public investment inefficient.

ADB (2013) points out that the pre-tax rate of return on risk-free assets has been proposed as a method for estimating the SOC, such as Moody's AAA bonds. However, the use of SOC would only be acceptable in two-period models with a fixed amount of capital available for investment. In this scenario, public investment crowds out private investment, and the marginal rate of return on private investment reflects the SOC. However, if either assumption is abandoned, the argument becomes unsound since if public funding is at least partially derived from postponing current consumption, the rate of return should be lower. Therefore, the SDR would be less than the SOC.

Boardman et al. (2018) also emphasise that using market rates only accounts for currently living individuals, disregarding future generations, which is particularly problematic in projects with long-term impacts. In fact, the sensitivity of discount rates to long time horizons is one of the most significant issues in cost-benefit analysis, as it raises serious intergenerational concerns.

Boardman et al. (2018) discuss another method for estimating the SDR that builds on Arnold Harberger's work, incorporating further theoretical developments suited for an open economy. Groom et al. (2022) refer to it as "*the SOC approach in the second-best world*" (p. 471). ADB (2013) categorises it as the weighted average approach, asserting that the SDR should represent the weighted average of SOC, STP, and the cost of foreign borrowing. The weights reflect the proportion of resources derived from each term. In this approach, SOC is:

$$SOC = a*CRI + b*ROI + c*CFF$$

CRI = trade-off consumption today for consumption in the future is called the consumption rate of interest;

ROI = return on investment;

CFF = the marginal cost of foreign funds.

Boardman et al. (2018) summarise numerous criticisms related to the estimation of SOC and its respective parameters. They initially emphasise the potential for error in the weights applied. More specifically, they highlight the possibility of overestimating ROI, which is the most significant parameter of the method. Another pertinent concern is that SOC may vary from project to project, as they can be funded in various ways. Therefore, applying the same rate across different projects would be inappropriate. Thirdly, they note the conventional approach of estimating the average ROI instead of the marginal ROI, which fails to reflect the theoretical concept of SOC. Additionally, market estimates often include risk premiums that should not be considered as part of ROI. Finally, the authors emphasise that ROI is frequently biased because it estimates actual returns rather than social ones, which take into account market failures and negative externalities. For these reasons, they conclude that SOC results in an overestimation of the SDR.

ABD (2013) notes that the weighted average approach assumes all benefits will be consumed immediately, neglecting the potential for reinvestment. Consequently, this methodology acknowledges the high social cost of public funding while overlooking the social value of reinvested benefits. The longer the project lasts, the greater the bias in the weighted average approach.

Another method that aims to reconcile STP and SOC is the shadow price of capital (SPC), which addresses the STP shortcoming of disregarding the mainstream consensus of the crowding-out effect. The SPC approach seeks to demonstrate how public investment influences private investments regarding lost future consumption. In CBA, the shadow price is employed to adjust costs that may displace private investments (Groom et al., 2022). Additionally, SPC aims to address the failure of the weighted average approach to account for potential benefits of reinvestments that create the aforementioned bias. Boardman et al. (2018) present the equation:

$$SPC = \frac{(1-f)ROI}{STP - f.ROI}$$
, where:

f = the fraction of the return that is reinvested each period.

Both ADB (2013) and Boardman et al. (2018) emphasise that although the concept of the SPC is quite appealing, its implementation is challenging. SPC estimation is highly sensitive to STP, SOC, the assumptions regarding depreciation and reinvestment, and the duration of the project under analysis.

Regarding the application of the SDR in CBA analysis in countries and international organisations, there is no consensus on which approach better reflects how public projects impact private investments and consumption. Groom et al. (2022) provide a selected overview of SDRs across countries and institutions. Their survey shows that more countries follow SOC (n=14) than STP (n=8), with two countries using both. On average, STP is set 3.7% lower than SOC.

It is important to remember that an excessively high SDR places little value on future effects, thereby undervaluing the benefits and costs for future generations. Consequently, it may prevent worthwhile public projects from being undertaken.

Groom et al. (2022) also present notable differences, even when the same type of SDR is selected. In contrast, the SOC in European countries is significantly lower than that in Latin American countries. For instance, the Netherlands employs a 2.25% SOC, whereas Mexico utilises a 10% SOC.



Figure 7 - Social Discount Rates (in %) by country and approach

Source: Groom et al. (2022). Note: Some countries, e.g. Brazil, Columbia, Mexico and Peru, follow the weighted SOC approach outlined in Harberger (1969); Abbreviations: ADB: Asian Development Bank; IDB: Inter-American Development Bank; SOC: Social Opportunity Cost of capital; STP: Social Time Preference.

According to ADB (2013), the differences between developing countries that use SDRs of 8-15% and developed countries that apply rates of 3-7% relate to the "*perceived social opportunity cost of public funds across countries in the extent to which the issue of intergenerational equity is taken into consideration in setting the SDR*" (p. 42). In this regard, there seems to be an understanding that developing countries experience a significantly higher opportunity cost of capital or demonstrate a much lower concern for future generations. The same reasoning extends to multilateral institutions such as the Asian Development Bank, Inter-American Development Bank, World Bank, African Development Bank, and European Bank for Reconstruction and Development, with rates varying from 9-12% according to Campos, Serebrisky, and Suárez-Alemán (2015) and Groom et al. (2022)⁹.

In relation to uncertainty, Groom et al. (2022) identify five approaches to account for unclear future events. Firstly, some countries, such as Australia and Denmark, utilise sensitivity

⁹ Recently, the Asian Development Bank switched to using the STP method and reduced its recommended rate to 9 per cent (Boardman et al., 2018).

analysis with alternative SDRs. The second approach incorporates a risk premium atop a riskfree rate, as seen in the Netherlands and New Zealand. Thirdly, nations such as Australia, Norway, and Peru implement declining discount rates (DDR) to address uncertainty.

DDR is an important subject, particularly when considering long-term projects. The assumption of a constant SDR is inappropriate when dealing with intergenerational projects—that is, investments with significant effects across generations. For instance, discussions about environmental and health policies fall into this category and illustrate the severe impacts a narrow-minded perspective of public spending can have on society. To exemplify this notion, we might refer to Boardman et al. (2018) for an example of investing in an environmental policy:

With a constant SDR the social discount factors decline geometrically. Even using a modest SDR, costs and benefits that occur sufficiently far in the future have a negligible value. The use of a constant discount rate much in excess of 1.0 or 2.0 percent implies that it is not allocatively efficient for society to spend even a small amount today in order to avert a very costly environmental disaster, provided that the disaster occurs sufficiently far in the future. For example, if greenhouse gas buildup imposes a huge cost of, say, \$1 trillion in 400 years' time, this has an NPV of less than \$336 million today at a constant discount rate of 2 percent and an NPV of less than \$113,000 at a discount rate of 4 percent. Thus, if CBA used a discount rate of more than 4 percent, we would conclude that it is not worth spending \$113,000 today to avert a major disaster with a cost of \$1 trillion in 400 years (p. 257).

The fourth approach identified by Groom et al. (2022) for addressing risk and uncertainty concerns the cases of Sweden and the UK, which incorporate catastrophic extinction risk as an additive component to the rate of pure time preference. Lastly, some countries consider risk separately from the SDR. Furthermore, some nations apply lower rates for non-market effects, such as environmental and health projects, whereas others adjust relative prices for such policies.

Groom et al. (2022) emphasise that none of the selected guidelines indicates SDR adjustments to address intra-temporal inequalities. The Canadian guideline, for instance, describes equity concerns as "*too controversial*." Other countries, such as Germany and the UK, adopt a distinct approach to addressing inequality in specific contexts.

Table 7 - Selected overview of discounting guidance across countries

Country Year SDR SOC or STP	Uncertainty	Inequality	Non-market effects
--------------------------------	-------------	------------	-----------------------

ADB	2017	9%	STP			6% SDR for env. or soc. benefits
Australia	2020	7%	SOC	Sensitivity with 3% & 10%, DDR to 3.7% (>300yrs), lower risk premium for nonmarket benefits		4% SDR for non-market benefits
Brazil	2020	8,5%	SOC			
Canada	2007	8%	SOC			3% SDR may be used for health & env. goods
Chile	2021	6%	SOC			
Columbia	2019	9%	SOC			
Denmark	2021	3,5%	Both	1.5% systematic risk premium, falling to 0% (>70yrs). DDR to 1.5% (>70yrs)		
EU	2021	3%	STP	DDR for >50yrs		
France	2013	4,5%	SOC	Risk-free SDR from 2.5% to 1.5% & risk premium from 2% to 3% (>2070)		RPA for env. goods
Germany	2018	1%	STP		Equity weights for SCC	
IDB	2021	12%	SOC			
Ireland	2019	4%	STP	DDR to 1.5% >275yrs		
Mexico	2014	10%	SOC			
Netherlands	2020	2,3%	SOC	3.25% risk premium; SDR sensitivity with 1.85% to 2.65%		1% RPA for env. goods
New Zeland	2020	5%	SOC	7% risk-premium		
Norway	2021	4%	SOC	DDR from 4% to 2% (>75yrs). 1.5% risk premium		
Peru	2021	8%	SOC	DDR from 8% to 1% (>200yrs)		
Sweden	2014	3,5%	STP	1% catastrophic risk premium		
Uk	2018	3,5%	STP	1% catastrophic risk premium; DDR to 2.14% (>75yrs)	Separate equity weights	1.5% SDR for health; 2% RPA for air pollution damages
USA	2003	7% & 3%	Both	Lower SDRs may be used for important intergen. effects		

Source: Groom et al. (2022) adapted.

Abbreviations: ADB: Asian Development Bank; IDB: Inter-American Development Bank; SOC: Social opportunity cost of capital; STP: Social Time Preference; DDR: Declining discount rate; SCC: Social cost of carbon; RPA: Relative price adjustment. Brazil, Columbia, Mexico and Peru follow the SOC approach outlined in Harberger (1969).

Thus, discounting encompasses several relevant elements that significantly influence the CBA results. From a theoretical standpoint, differences regarding the effects of public investment on the private sector determine the nature of the SDR. In all instances, however, there exists an implicit consensus that the government should regard the market economy as the standard. Citizens' well-being is reduced to consumption, and investments are justifiable only insofar as they do not disrupt private ventures.

Concerns regarding future generations remain in the very early stages when selecting the SDR in CBA, as few countries have adopted discount rates that encourage projects with long-term benefits, such as environmental and health policies. In a world facing significant climate threats, for instance, the CBA is a conservative approach for addressing structural changes or addressing other critical intergenerational challenges. Moreover, accounting for intra-generational inequalities through SDR adjustments is still irrelevant in real-world CBA analysis.

Differences in SDR rates also reveal that developing countries encounter even more resistance to public funding, particularly for long-term projects, as discounting at higher interest rates renders projects with future benefits less appealing from a market perspective view.

Recognising these aspects of the SDR is very important, as managing time in policies or projects with varying maturities is one of the most crucial elements of CBA, which can significantly affect the outcomes.

2.1.4 Distributive concerns

The purpose of the CBA is to address efficient resource allocation. To achieve this, it relies on a framework that disregards any kind of distributive concerns: individual, regional, sectoral, gender, and race. However, when it comes to policymaking, the method should not abandon the responsibility of assessing distributive justice in its various forms, as determining the impacts of a project on different social groups is essential to the decision-making process. Furthermore, as previously argued, given that public and private enterprises have entirely different objectives, it is unreasonable to constrain public spending as if the State were a private firm.

Implied in this discussion is the trade-off between efficiency and equity that arises directly from neoclassical welfare economics, as explored in Chapter One. By introducing the Kaldor-Hicks compensation principle, mainstream economics re-established 'impartiality' since efficiency and equity were to be considered separately. Furthermore, by following the logic of the market as a benchmark and citizens as customers, individuals' ability to pay for goods or services determines the value of things and expands inequalities, as addressed in section 2.1.2.

Discussions about equity are often framed as society's willingness to 'sacrifice efficiency' for social concerns. Generally, the intergenerational distribution dimension is addressed through the discount rate, as discussed in the previous section. Regarding intragenerational distribution concerns, some of the literature has evolved to incorporate distributional weights within CBA. The greater the difference between the weights, the higher the analysis's sensitivity to equity concerns.

Boardman et al. (2018) discuss three arguments frequently advanced in favour of addressing equity concerns within the CBA framework. The first asserts that higher-income groups experience diminishing marginal utility; in other words, the effect of a dollar benefit or cost is less significant for wealthier individuals. This rationale was examined in Chapter One within the context of the material welfare school, which suggests that there is no trade-off between efficiency and equity.

Campbell and Brown (2023) explain that the weights are derived under the assumption of diminishing marginal utility. Consequently, the weight attributed to additional consumption is based on the marginal utility it provides at that level of income in comparison with some base level. As marginal utility declines with increasing income, higher income results in a lower weight being attached. The weights would be:

 $d_i = \left(\frac{\bar{Y}}{Y_i}\right)^n$, where d_i = the distribution weight for group i \bar{Y} = the average level of income for the economy Y_i = the average level of income for group i

n = the elasticity of the marginal utility regarding income increases

Problems with this perspective arise when estimating this responsiveness 'n'. Campbell and Brown (2023) demonstrate how sensitive the weight distribution is to variations in the parameter 'n', concluding that the elasticity of marginal utility in relation to income increases is merely a judgment on the value of an additional dollar to the wealthy or the impoverished.

The second argument discussed by Boardman et al. (2018) for addressing equity concerns within the CBA framework is the perception that inequality is excessively high in our society and that welfare would improve if income distribution were more equitable. Several claims support this assertion. First, extreme inequality can provoke civil disorders and increase crime rates, worsening general welfare. Secondly, society agrees on a minimum standard of living, below which no human being should fall. Moreover, not every individual's sole concern is their own well-being. Therefore, improvements in the welfare of low-income individuals would also benefit society. These arguments suggest the necessity to reject projects considered efficient if they fail to address equality concerns.

The 'One person, one vote' principle is the third argument and addresses the consequences of the methodology's reliance on WTP, as discussed in section 2.1.2. Since high-income individuals can dispose of more money than low-income individuals, their influence on decision-making is more significant. Therefore, the fundamental idea is that everyone should have equal influence over public policy decisions, irrespective of their purchasing power.

The OECD (2006) highlights that the essence of applying distributional weights within the CBA framework is that public interventions should not only be based on individual or household WTP. Instead, the methodology should be more sensitive to equity considerations. They present a simple case to illustrate how conventional CBA implicitly addresses the distributive effects of the projects under analysis. Suppose a project only affects two individuals:

$$NPV = a_1 NPV + a_2 NPV$$
, where

a = the weight assigned to each individual regarding the project's net present value.

Conventional CBA assumes that $a_1 = a_2 = 1$. In this sense, equal weight is assigned to each individual, irrespective of who benefits from the project and who bears the burden. If "a" is considered society's choice regarding the distribution of benefits, then any distributional

impact arising from CBA is acceptable within the conventional framework. It is essential to remember that the benefits are not confined to income but also include non-market goods.

Boardman et al. (2018) illustrate how the conventional framework of CBA impacts environmental agreements that favour developed nations. They emphasise that without distributional weights, the negative emission effects tend to be greater in developing countries. One reason is the smaller valuation of the statistical life in these regions due to lower WTP¹⁰. That is, simply because the ability to pay in lower-income countries is reduced, their lives hold less value.

The 1991 memorandum by Larry Summers, then chief economist at the World Bank, also illustrates how the lack of distributional weights in CBA analysis favours wealthy countries due to its theoretical basis in WTP. The internal document was leaked to a Brazilian newspaper, which published the justification for relocating polluting industries to less developed countries. The CBA rationale formed the basis of his arguments (Petri, 2021).

First, Summers argued that health-impairing pollution should be situated in the country with the lowest costs, specifically the one with lower wages. He stated: "*I think the economic logic behind dumping a load of toxic waste in the lowest-wage county is impeccable and we should face up to that*" (Petri, 2021, p. 1288). The logic underpinning this economic argument is that the valuation of human life is based on its potential contribution to society, with wages reflecting this capacity. The theoretical foundation for this argument is the neoclassical theory of income distribution, which posits that wages reflect a worker's productivity. Therefore, any illnesses or injuries that affect an individual's productivity should be valued as potential losses. The outcome is that the lives of high-income individuals are deemed more valuable.

Secondly, Summers asserted that the costs of pollution are likely to be quite low in their initial increments, which would render less developed countries 'under polluted'. He affirmed that "*their air quality is probably vastly inefficiently low compared to Los Angeles or Mexico City*" (Petri, 2021, p. 1288). According to his regional unweighted analysis based on the WTP CBA approach, the presence of good air quality in countries that have not yet explored into many dirty industries is inefficient, and the economic rationale supports the migration of pollution. Furthermore, the former chief economist of the World Bank argues that the demand for a clean and healthy environment likely has a high income elasticity. Both arguments rely on

¹⁰ The Value of Statistical Life (VSL) is an estimation that quantifies the value that society is willing to pay to avoid mortality risks.

the vulnerability of less developed countries, which might be willing to accept imported pollution if adequately compensated (Petri, 2021).

It is crucial to note that the reliance of the CBA on WTP, combined with the usual practice of ignoring distributional weights, results in analyses that favour the wealthier, whether they be individuals, countries, regions, or otherwise. This stems directly from the Kaldor-Hicks potential compensation principle discussed in Chapter One, which disregards who benefits from the project as long as the net gain is positive. Furthermore, this framework does not address the impacts of policymaking on other inequalities, such as gender or race, illustrating how welfare is only related to income and consumption. For these reasons, a growing body of literature is advocating for the inclusion of distributional weights within the CBA framework.

However, the concept of assigning weights is not universally accepted. According to Mishan and Quah (2020), CBA does not incorporate implicit weights, as no interpersonal comparison is presumed. Moreover, by recognising distributive concerns, CBA deviates from the principles of allocative efficiency. However, the decision not to apply any distributional weight constitutes a value judgement in itself, whether made explicitly or implicitly (Campbell and Brown, 2023).

Official guidelines for CBA appraisals do not typically address equity concerns directly. Instead, they assign inequality issues to separate analyses. Consequently, equity concerns are often excluded from CBA appraisals and evaluations, as demonstrated by the works of Groom et al. (2022) and Boardman et al. (2018)—only the UK advocates for distributional weighting in CBA within official guidelines. Therefore, in practice, as the OECD (2006) concluded earlier, equity concerns have little or no role in social decisions regarding the selection and design of policies.

Mishan and Quah (2020) discuss two reasons commonly cited to justify the absence of distributional concerns within the CBA framework. Firstly, there is a lack of consensus on distributional weights, particularly as they may fluctuate with changes in the political scenario. Political shifts, they argue, might support political projects at the expense of economic rationale. Conversely, if economists are tasked with choosing the weights, arbitrariness becomes the main concern.

Implicit in this argument is the notion that CBA is a value-neutral methodology and that the application of distributional weights would undermine the method's objectivity. Furthermore, there is a perception that policy appraisals and evaluations can disregard political considerations, neglecting the fact that no method can replace collective decisions and that politics is an inseparable aspect of life within society.

The second claim for dispensing distributional weights is that even assuming an accurate estimate of weights, the goal of improving income distribution is not guaranteed. For any set of weights, CBA analysis might still result in distributionally regressive projects if the rich affected by the project are numerous or made much richer.

In conclusion, despite the numerous issues surrounding the valuation process and the social discount rate, CBA remains unsuitable for public analysis because it overlooks distributive justice, thereby granting greater influence to the preferences of the affluent. The assignment of distributional weights, as discussed in the literature, does not adequately address this problem. In addition to facing practical challenges, efforts to incorporate distributive weights presume that economists can prescribe a notion of collective fairness.

Nevertheless, not only should the distributive consequences of policymaking be subject to public debate but also all relevant impacts derived from government policies. Society cannot delegate the decision over the merits of public actions to a flawed methodology. In fact, as Rose-Ackerman (2011) pointed out, policies with significant social impacts on current or future generations require a different normative framework. Thus, the CBA framework is insufficient to guide the allocation of resources in society, even if distributional weights are applied.

2.2 Technical efficiency: seeking value for money

Another concept of efficiency is linked to productive activity. Technical efficiency pertains to the idea of maximisation, where firms aim to achieve the highest profit under given conditions. In the realm of policymaking, the primary concern is to ensure a better utilisation of resources, as scarcity is the central issue of economic theory. Consequently, 'value for money' is applied to public policies, highlighting the symmetry between the government and firms yet again.

Bogetoft and Otto (2011) discuss another, more practical reason for seeking technical efficiency in policymaking. The authors argue that the concept of effectiveness involves comparing actual performance with an ideal one, as the goal is explicit. However, when the

goal is not clear-cut or the ideal performance cannot be accurately estimated, evaluations depend on proxy objectives, and efficiency becomes the focus.

The authors argue that technical efficiency is a rational method of assessment, as it identifies minimum cost as the social preference and evaluates the unit's performance against potential peers. This rational evaluation would be relevant in the real world, as the scarcity of information on ideal performances is quite common. This is because addressing ideal performances implies comprehensive knowledge of utility and technology. Consequently, evaluations shift from effectiveness to efficiency, with the primary goal being to estimate a frontier technology in order to account for relative efficiency. In this sense, "*Benchmarking is a way to overcome these fundamental practical problems by moving from effectiveness to relative efficiency*" (p. 8). They specify:

$$Efficiency = \frac{Minimal\ cost}{Actual\ cost} \quad \mathbf{x} \quad Effectiveness = \frac{Actual\ performance}{Ideal\ performance} = \frac{U\ (A)}{U\ (ideal)}$$

However, shifting from effectiveness to efficiency often leads to a disregard for policy objectives. The evaluation process ceases to address the underlying reasons for the policy's initial implementation and instead focuses on minimising costs. In this regard, Bogetoft and Otto (2011) acknowledge that efficiency is not a sufficient criterion and can, indeed, be misleading:

It is worth remembering, however, that this logic also means that while efficiency is a necessary condition for effectiveness, it is not a sufficient one. In fact, in terms of a particular technology, an inefficient firm may well be better than a fully efficient one. We could rephrase this by saying that it is not sufficient to run fast; it is also important to run in the correct direction—and it may be better to run at a moderate speed in the right direction than at full speed off-course (p. 25).

Fundamentally, the technical aspect of efficiency concerns the idea of minimising waste. This aim can be realised by producing as much output as possible with the available inputs and technology, or by decreasing inputs to achieve the same outputs. In the input orientation, the focus is on reducing inputs (radial reduction) proportionally while maintaining the production level. In contrast, the output orientation aims for the maximum radial expansion of outputs with the given inputs and technology (Fried, Lovell, and Schmidt, 2008). The combination of inputs and outputs is a crucial element in measuring technical efficiency and reveals the main characteristic of this approach: it relies on estimations of frontier functions to compare the relative performance of the entities analysed. As Murillo-Zamorano (2004) pointed out, the notion of defining an efficient frontier to assess the current performance of productive units has been popular for a long time. It originates from neoclassical production theory, which emphasises optimising a firm's performance subject to constraints.

Thus, technical efficiency involves comparing observed performances with optimal ones, where optimal is limited to the best verified productions. This concept encompasses benchmarking, with comparisons aimed at enhancing productivity. Bogetoft and Otto (2011) emphasize that benchmarking pertains to one or more objectives: learning, coordination, and motivation. They assert that in scientific studies, learning is undoubtedly the primary objective of benchmarking. In this regard, assessing the relative efficiency of units can yield significant insights for enhancing performance.

However, Fried, Lovell, and Schmidt (2008) highlight three common problems with the methods used to estimate technical efficiency. First, which inputs and outputs should be selected? Additionally, how should they be weighted in the comparison? Lastly, how should the technology or potential production be determined? The first problem relates to the analyst's failure to select all relevant variables, specify the economic objective, and establish the proper constraints. If these elements are not correctly addressed, they can lead to inefficiency. The second problem pertains to estimating efficiency while considering multiple inputs and outputs. In this case, prices often serve as reference weights for the variables. Nonetheless, price fluctuations or the presence of monopolies can distort these weights. Moreover, acknowledging numerous non-market goods related to the analysis presents another challenge in technical efficiency estimation. The final problem involves determining the possible frontier since no one knows the 'true' potential of the units being analysed. Thus, the efficient frontier is seen as the 'best-practice' frontier, which could indeed serve as a poor reference.

The three common problems identified by Fried, Lovell, and Schmidt (2008) are easily observed when discussing the methods for estimating technical efficiency. Analysing each technique allows for a comprehension of its strengths and weaknesses in relation to these potential issues, ensuring that the results are used with the necessary precautions. Furthermore, contextualising the results and seeking possible causes are essential elements of any analysis. Otherwise, conclusions may be misleading.

2.2.1 Performance Indicators

A popular method for analysing performance is through Key Performance Indicators (KPIs). The premise is that they somehow reflect the essence of the entity's purpose. Most industries have specific indicators to provide comparable accomplishments. For instance, productivity, the ratio of inputs to outputs, is a well-known indicator for assessing performance. (Bogetoft and Otto, 2011).

In the context of increasing KPIs in the private sector, the public sector has also developed numerous indicators to facilitate performance comparisons. Shore and Wright (2015) discuss this phenomenon, illustrating how this process oversimplifies complex situations and operates within a ranking logic to define qualitative features. The authors argue that this 'audit culture'—the traditionally technical analysis of private firms—has permeated other domains, such as public management—an application that bears no resemblance to finance or bookkeeping, as the auditing represents an independent evaluation for internal control.

Indeed, comparing gross margin (gross profit/net sales) among private firms in a specific industry makes sense for evaluating relative performance, particularly as private firms share the same objective: profit. Conversely, comparing public achievements is considerably more complex, especially when engaging with peers who possess entirely different backgrounds and goals.

Shore and Wright (2015) point out:

'Governing by numbers' – reducing complex processes to simple numerical indicators and rankings for purposes of management and control – has become a defining feature of our times. At the heart of this process is an increasing fetishisation of statistical measurement and competitive ranking as robust and reliable instruments for calculating (and enhancing) what are largely qualitative features such as 'excellence', 'quality', 'value' and 'effectiveness' (p. 22).

Therefore, in addition to standardising complex realities, the 'audit culture' fosters subjectivities that link efficiency with quality, enabling the acceptance of symmetrical evaluations between firms and the government. It is no coincidence that debates over policymaking frequently centre on the 'quality of public spending', a characteristic that directly connects expenditures and efficiency performances, often without regard for policy goals.

The most relevant aspect of this process is replacing social and political debate with 'technical expertise'. As Merry (2011) argued: "*A key dimension of the power of indicators is*

their capacity to convert complicated contextually variable phenomena into unambiguous, clear, and impersonal measures" (p.84). It is this feature that gives them an aura of objective truth.

Public indicators are generally multidimensional because they aim to capture society's objectives with some precision. Each dimension features a sub-indicator that measures results through a chosen proxy. For instance, if the public objective is to enhance education, the multidimensional indicator may include sub-indicators encompassing students' performances in mathematics and reading, average years of schooling, and the percentage of the population with higher education, among others. (Boueri, Rocha, and Rodopoulos, 2015).

However, Bogetoft and Otto (2011) discuss how these micro KPIs' have implicit assumptions that might compromise results. First, comparing entities with different sizes assumes that inputs and outputs are scaled linearly, thus assuming constant returns to scale. Second, KPI's evaluations are frequently only partial, implying that they may not correctly reflect the entity's entire objective. Consequently, partial contrasting may provide misleading conclusions or unfeasible comparisons. The third limitation is known as the Fox Paradox and shows how an entity can be less productive in sub-processes, being the most productive in overall evaluations.

Performance indicators can also address macro measures that capture a general sense of how the State contributes to society's welfare (Boueri, Rocha, and Rodopoulos, 2015). For instance, it is trendy to compare general welfare correlating the level of public spending (the 'size' of the State) with respect to the country's per capita Gross Domestic Product (GDP). Typically, results show that public spending does not correlate with increased well-being.

For example, Tanzi and Schuknecht's renowned work, Reconsidering the Fiscal Role of Government: The International Perspective (1997), finds no evidence that public spending enhances welfare, based on comparisons between the ratio of public expenditure to GDP and selected socioeconomic indicators used as proxies for welfare. The authors argue that the increasing share of public spending relative to GDP is significant and responds to pressures to expand the government's role in areas such as education, health, the provision of public pensions, and assistance to the unemployed, among others. Although the authors recognise that these policies have contributed to higher welfare through improvements in literacy and human capital, reduced mortality, and increased life expectancy, they raise the question of whether there is a limit to government contributions to general welfare, that is, whether diminishing returns to welfare gains exist. (Tanzi and Schuknecht, 1997).

To address this question, they present a static and ahistorical table that divides selected countries into three groups based on the percentage of spending relative to GDP, classifying them as big, medium, or small governments while comparing selected indicators among them. For instance, the authors identify the lowest unemployment rate in countries with small governments and the lowest infant mortality rate.

Figure 8 - Size of Government and Performance Indicators in Industrial Count	ries, 1	1990
(Percentage of GDP, unless otherwise indicated)		

	Siz	e of governi	nent ^a
Indicator	Big	Medium	Small
Total public expenditure:	55.1	44.9	34.6
Public consumption	18.9	17.4	15.5
Subsidies and transfers	30.6	21.5	14.0
Economic indicators:			
Real GDP growth ^b	2.0	2.6	2.5
Standard deviation of			
GDP growth	1.6	2.1	1.9
Gross fixed capital			
formation	20.5	21.3	20.7
Inflation ^b	3.9	3.7	3.7
Unemployment rate	8.5	11.9	6.6
Public debt	79.0	59.9	53.3
Social indicators:			
Life expectancy (years)	77	77	77
Infant mortality/1,000			
births	6.7	7.1	6.4
Secondary school			
enrollment	92.8	99.1	89.0
Income share of poorest			
40 percent	24.1	21.6	20.8

Sources: Information on the specific sources is available from the authors upon request.

^a Big governments (meaning public expenditure is more than 50 percent of GDP) include Belgium, Italy, the Netherlands, Norway, and Sweden. Medium-sized governments (public expenditure 40–50 percent of GDP) include Austria, Canada, France, Germany, Ireland, New Zealand, and Spain. Small governments (public expenditure less than 40 percent of GDP) include Australia, Japan, Switzerland, the United Kingdom, and the United States. ^b 1986–1994 period.

Source: Tanzi and Schuknecht, 1997, p. 167.

Regarding these performance indicators, recognising that they are static and a-historical significantly undermines the capacity of these KPIs to offer reliable information. Comparing remarkably diverse realities and linking them to a single metric without considering historical contexts is unreasonable. For instance, when analysing countries, one cannot overlook the importance of addressing the role of history. However, it is precisely because these indicators decontextualise results that comparisons become feasible.

Aside from implicit assumptions and neglecting context and history, interpreting the results is another crucial element of comparing efficiency. Similar to what occurs in private firms, differences in performance in policymaking have often been interpreted as opportunities for reducing expenditure, irrespective of the underlying reasons for such differences. For instance, in the previously mentioned example of a public indicator in education, if two cities yield the same results but one utilises more inputs, the typical response is to advocate for budget cuts, suggesting that the same outcomes can be achieved with fewer resources. However, municipalities are not private firms, and numerous factors may account for the need for greater input. For example, a greater distance between students and the school or an increased requirement for teachers due to the challenges faced by students from disadvantaged backgrounds. Additionally, some reasons are not directly related to performance. Variations in teachers' salaries and workloads, which are standard inputs in education, are not necessarily indicative of productivity but may instead raise significant discussions about the minimum conditions for workers in the labour market. In this context, the public sector may have valid social and economic justifications for increasing wages.

As Shore and Wright (2015) discussed, complex situations cannot be reduced to a simple numerical indicator. Of course, indicators can support the decision-making process by providing information for debate. However, social and political considerations cannot be substituted with the simplistic solution of cutting expenditure whenever there are differences in performance. Furthermore, society should not adopt corporate thinking, associating it with quality, rationality, and evidence-based decisions simply because the goals are fundamentally different. Nobody would advocate for wasting public resources. Nevertheless, debates about what constitutes waste in methodologies that assess technical efficiency are vital for policymaking. To achieve this, it is essential to understand the internal elements of the methods employed to evaluate technical efficiency.

2.2.2 Data Envelopment Analysis (DEA) and Stochastic Frontier Models (SFA)

In addition to KPIs, numerous models exist for estimating performance. Each category of models features numerous variations that address different assumptions regarding technology, noise distribution, and so on. Bogetoft and Otto (2011) categorise the methods as deterministic or stochastic, as well as parametric or non-parametric. Typically, they share common foundations in estimating efficient frontier functions against which a unit's performance is assessed.

	Deterministic	Stochastic
Denserveterie	Corrected Ordinary Least Squares (COLS)	Stochastic Frontier Analysis (SFA)
Parametric	Aigner and Chu (1968), Lovell (1993), Greene (1990, 2008)	Aigner et al (1977), Battese and Coelli (1992), Coelli et al (1998a)
Non nonomotorio	Data Envelopment Analysis (DEA)	Stochastic Data Envelopment Analysis (SDEA)
Non-parametric	Charnes et al (1978), Deprins et al (1984)	Land et al (1993), Olesen and Petersen (1995), Fethi et al (2001)

Table 8 - A taxonomy of frontier methods

Source: Bogetoft and Otto, 2011, p.18.

Data Envelopment Analysis (DEA) and Stochastic Frontier Models (SFA) are the most commonly used techniques for measuring technical efficiency (Lampe and Hilgers, 2015). Consequently, this section examines these two approaches, emphasising their fundamental differences, limitations, and recent developments.

DEA is a non-parametric method, meaning there is no pre-established functional form. It is primarily atheoretical and based on the available data regarding selected inputs and outputs to estimate the production frontier and the efficiency scores of each unit considered in the data. In contrast, SFA is a parametric approach with a pre-defined functional form against which the unit's performance is measured. The necessity of specifying a functional form for technology and the error term presents the main disadvantage of SFA, due to potential specification and estimation problems. (Murillo-Zamorano, 2004; Fried, Lovell, and Schmidt, 2008).

Another classification that distinguishes DEA from SFA is the tools used to solve the problem. Most DEA applications rely on mathematical programming, highlighting its deterministic nature. Deterministic specifications assume that all deviations from the efficient frontier are due to inefficiency, an element within the agent's control. However, many factors outside the agent's control influence results—for instance, regulations, socio-economic factors, and demographic characteristics. This is the main disadvantage noted in the literature on DEA: the deterministic nature of the models. In contrast, SFA models are stochastic, allowing them to differentiate between noise and inefficiency. (Murillo-Zamorano, 2004).

On the advantages side, for being non-parametric, the DEA framework enables the model to avoid the effects of confounding misspecification of the functional form with inefficiency. Additionally, flexibility is a common benefit of DEA models discussed in the literature. Apart from the absence of previous specifications, a good example of the adaptability of DEA models is the lack of pre-determined weights on inputs and outputs. In fact, these will be calculated to position the Decision Making Unit (DMU) in the best possible light. This feature is particularly relevant for estimating efficiency when no a priori weights are known. Conversely, the stochastic parametric perspective accounts for statistical noise, recognising that individual observations may be influenced by random noise. Therefore, it attempts to distinguish the structure from the noise, providing a basis for inferences. (Murillo-Zamorano, 2004; Fried, Lovell, and Schmidt, 2008).

According to Bogetoft and Otto (2011), choosing between DEA and SFA involves selecting between a flexible structure and precision in the noise. Ideally, one would opt for Stochastic Data Envelopment Analysis (SDEA), which combines the advantage of relying on data rather than excessive theoretical assumptions, with reduced model sensitivity to random variations in the data. However, they argue that this decision is not costless, as "*the estimation task becomes bigger, the data need larger, and still, we cannot avoid a series of strong assumptions about the distributions of the noise terms*" (p. 19). They conclude that utilising either DEA or SFA models is adequate for addressing technical efficiency, aligning with the applied studies identified in the literature. Both methods necessitate minimal or no assumptions regarding preferences or prices and can accommodate a variety of inputs and outputs.

Developments have been made to address shortcomings in each approach. For instance, DEA models primarily differ in the adoption of four hypotheses. First is the reliance on constant returns to scale. Constant returns to scale assume that each DMU operates at optimal scale, which is unrealistic. To tackle this issue, a variable return to scale model was designed to ensure that each DMU is only compared to one of similar size, thus avoiding the impact of scale efficiency on the technical efficiency scores. Second is the assumption of free disposability, which states that all inputs and unwanted outputs can be easily discarded; this is a strong hypothesis. Most common DEA models still depend on this premise. Third, the convexity assumption posits that any weighted average (convex combination) of feasible production is also possible, which is regarded as a 'convenient' hypothesis. Some models operate without convexity. Lastly, the additive principle also considers the sum of feasible production plans as possible. Only one model makes this assumption (Bogetoft and Otto, 2011). The table below

provides a summary of the types of DEA models and the underlying assumptions associated with them.

Model	A1 - Free disposability	A2 - Convexity	A3 - Scale*	A4 - Additivity
FDH - Free disposability hull	V		k = 1	
VRS - Varying return to scale	V	V	k = 1	
DRS - Decreasing return to scale	V	V	$k \leq 1$	
IRS - Increasing return to scale	V	V	$k \ge 1$	
CRS - Constant return to scale	V	V	$k \ge 0$	
FRH - Free replicability hull	V		k = 1	V

Table 9 - DEA model assumptions

Source: Bogetoft and Otto, 2011, p.88. Adapted.

* A3: γ -returns to scale. Production can be scaled with any of the given factors: $(x,y) \in T$, $k \in I'(\gamma) \rightarrow k.(x,y) \in T$ (p.85).

Another relevant development in DEA models was the treatment of variables over which DMUs have no control – the discretionary variables. The aim is to isolate their effect from the final performance of the sample units. Progress in DEA models is also observed in the treatment of complete flexibility, which can grant extremely low or high multipliers that are incompatible with economic theory and support specification. The Assurance Region and the Cone-Ratio methods are examples (Murillo-Zamorano, 2004).

More importantly, developments in bootstrapping methods to account for noise or measurement error in DEA models are promising (Fried, Lovell, and Schmidt, 2008). The bootstrap enables sensitivity analysis of efficiency scores relative to sampling variation, thus providing a better estimate of technical efficiency (Murillo-Zamorano, 2004).

However, according to Souza, Scatena, and Kehrig (2016), two elements continue to significantly influence DEA results: a) determining whether the model assumes constant returns to scale – the CCR model¹¹ - or variable returns to scale – the BCC model¹²; b) the orientation of the DEA model, specifically whether it is input or output oriented¹³. Furthermore, variable selection can also remarkably influence results. For example, a model with many variables

¹¹ Following the work of Charnes, Cooper e Rhodes (1978) and further developments.

¹² Following the work of Banker, Charnes e Cooper (1984) and further developments.

¹³ An input-oriented DEA model estimates the minimum resources (inputs) needed to keep the same results, while the output-oriented DEA model assumes inputs are fixed and estimates the maximum output according to the determined resources.

tends to be overly generous, resulting in numerous DMUs achieving a score of 1, which indicates performance at the efficiency frontier¹⁴.

Concerning SFA developments, models have transitioned to more flexible functional forms that accommodate multiple technologies. The fundamental cross-section Cobb-Douglas case is:

 $Ln Y_i = \beta_o + \sum_{n=1}^{N} \beta_n Ln X_i + v_i - u_i, \text{ where:}$ $v_i = \text{statistical noise (randomness)}$ $u_i = \text{technical inefficiency}$

In this case, strong distribution assumptions are made, which is regarded as one of the issues with SFA. The statistical noise is considered to be independent and identically distributed. Regarding the technical component, several distributions have been assumed, including half-normal and exponential. Another primary issue with cross-sectional SFA is that their estimates are inconsistent despite being unbiased. Panel Data Stochastic Frontier Models (PDMs) were developed to remedy these two shortcomings. The first advantage is that PDMs do not require the inefficiency term and the input levels to be independent. Furthermore, by incorporating a time dimension in the same unit, PDMs achieve consistent estimates of the inefficiency term. Lastly, no assumptions regarding the distribution of efficiency terms are necessary in PDMs (Murillo-Zamorano, 2004; Boueri, Rocha, and Rodopoulos, 2015).

The PDMs can be either time-invariant or time-variant. As the name suggests, timeinvariant models assume that efficiency remains constant over time. However, as the time dimension expands or in sectors where technological innovations are common, it seems reasonable to allow efficiency to fluctuate over time (Boueri, Rocha, and Rodopoulos, 2015).

Duality theory is a further step in improving flexibility in SFA models by allowing multiple equation estimation procedures. Dual representations of the production technology enable many outputs, quase-fixed inputs, alternative objectives, etc. The duality consists of choosing between the direct estimation of the production structure or the indirect estimation of a cost function. However, even though duality tools have contributed to a more accurate estimation of technical efficiency, many issues remain to be addressed. The Bayesian approach is a new popular perspective to overcome the necessity of imposing a priori restrictions on the

¹⁴ For this, there is a rule of thumb which indicates the number os DMUs must be at least the double of the total inputs and outputs considered.

efficiency term. The Bayesian techniques also allow the model to deal with multiple or undesired outputs (Murillo-Zamorano, 2004).

In terms of the application of technical efficiency in the real world, Lampe and Hilgers (2015) surveyed DEA and SFA applications from 1987 to 2011, highlighting the increasing significance of both methods, particularly in recent years. The number of publications rose from fewer than 20 per year for both methods combined to over 800 in 2010. DEA has consistently been more popular, with data from 2010 showing nearly 700 publications compared to fewer than 150 SFA studies.

A survey conducted by Liu et al. (2013) on DEA applications from 1978 to 2010 revealed that this method is primarily used in five areas: banking, healthcare, agriculture, transportation, and education, which collectively account for 41% of all applications. Furthermore, they conclude that the trend is to apply newly developed models as soon as they become available, suggesting that the most accurate model is invariably the one yet to be published. This finding is also echoed in Lampe and Hilgers (2015), who identified the same trend for SFA models.

Studies indicate a growing utilisation of DEA models in specific areas of expertise, which aligns with the previously mentioned finding of Liu et al. (2013). For instance, Mardani et al. (2017) provide a comprehensive review of DEA applications for assessing energy efficiency. This review identified 144 publications in 45 prestigious journals between 2006 and 2015. According to the authors, employing DEA is crucial due to the challenges of estimating a production function, which can be difficult or even unfeasible. Consequently, DEA has become increasingly prevalent for evaluating efficiency in the energy sector. A similar advantage of not needing to estimate a frontier function is noted by Toma et al. (2015) when discussing agricultural efficiency, as well as by Kohl et al. (2018) in relation to healthcare efficiency.

Conversely, particularly in Latin America, a survey conducted by Drei and Angulo-Meza (2023) on DEA applications in health services, one of the most prevalent areas of DEA analysis, revealed how minimal research has been carried out using the tool. They identified only 65 studies in the Scopus, Web of Science, and Scielo databases up to September 2021, which indicated no preference for a specific model. Furthermore, they observed that many studies did not acknowledge their limitations or propose future developments. Moreover, when suggestions were offered, no additional studies were found. In conclusion, significant advancements in DEA and SFA have addressed the most prevalent criticism concerning internal methodological shortcomings. Nevertheless, numerous challenges persist within these models, undermining their capacity to yield conclusive results. Furthermore, in addition to the discussions surrounding the internal elements of technical efficiency models presented in this Chapter, the necessity of carefully contextualising and interpreting results becomes particularly prominent when analysing studies of technical efficiency.

There remains minimal debate regarding the context and causes of performance variations, prompting policymakers to view all projected inefficiencies as opportunities for budget reductions. The next chapter discusses contextualising and interpreting results, accompanied by practical examples.

Final remarks

Chapter Two linked the theoretical foundations of neoclassical economic thought with the tools developed to assess whether a policy conforms to efficiency—the benchmark for policymaking. Indeed, as previously mentioned, determining resource allocation based on efficiency involves imposing behaviour on governments as if they were private firms—an analogy already observed in the symmetrical behaviour of the market's equilibrium imposed on consumers.

To clarify the argument about the symmetry between firms and the State, this chapter discusses the internal elements of the methodologies commonly applied to assess both allocative and technical efficiency, the two dimensions that should guide policymaking from a neoclassical perspective. In fact, uncovering the internal elements of the methodologies used to analyse government actions elucidates the implicit market principles governing these methods, thereby connecting theory to practice. Furthermore, it dispels the notion that seeking efficiency is a rational and impartial approach to guiding policymaking.

Regarding allocative efficiency, the CBA places collective interests beneath a net measure that evaluates whether a policy is efficient. In other words, the methodology assesses the merit of a public policy. Furthermore, the internal components of a CBA demonstrate how this method depends on market principles to achieve a conclusion. The first component discussed was the reliance on WTP to estimate the benefits of a policy – a principle that directs

market dynamics and determines prices. The second component debated was the contentious methods used to assess the values ascribed to non-market elements (externalities), which pose a challenge to the valuation process in achieving accurate outcomes, a crucial aspect for a methodology that delivers binary conclusions. Thirdly, defining a social discount rate to evaluate projects with long-term impacts was another internal aspect of the CBA that made the market centrality apparent. All the approaches reflect the mainstream view of how public financing influences consumption and private investments. On the one hand, it is the cost of foregone consumption, while on the other, it is the cost of displacing private investments. Furthermore, discussions regarding the SDR highlighted the ethical conflict surrounding intergenerational value judgments. Lastly, the section examined the implicit distributional consequences of applying CBA to inform policymaking, which assigns equal weight to each individual, irrespective of who benefits from the project and who bears the burden.

Regarding technical efficiency, the debate centred on how establishing symmetry between private firms and the government often results in losing sight of policy goals, as pursuing minimal costs becomes the primary objective to achieve good performance. The concept of seeking value for money often decontextualises central elements of the analysis and leads to the misinterpretation of results. In fact, despite the debated internal inconsistencies of the DEA and SFA methodologies, differences in performance arising from these methods are frequently interpreted as opportunities for reducing public expenditure.

Thus, examining the internal elements of the most widely used methodologies to assess allocative and technical efficiency reveals the implicit assumptions, value judgments, and principles underlying those methods, which challenge common assertions, such as impartiality and the existence of a purely technical decision-making process. In this sense, the following chapter illustrates the discussions presented in this chapter, showcasing real-life applications of these methodologies. Furthermore, it highlights alternative approaches that do not focus on benchmarking efficiency.

3. APPLIED EFFICIENCY: IMPLICATIONS TO POLICYMAKING

While Chapter 1 discussed the theoretical foundations of efficiency as a guiding principle for policymaking and the broad implications for government actions, Chapter 2 explored the most common methodologies for assessing allocative and technical efficiency,

presenting internal methodological elements that prevent the methods from achieving accurate results. Furthermore, the discussion showed how internalising the market rationale in appraisals and evaluations elicits ethical dilemmas besides separating the analysis from policy goals.

Chapter 3 moves to practice, presenting applications of such methods and stressing the typical consequences of adopting market-based analysis in the public sphere. Additionally, analysing the real-life employment of these estimations can uncover implicit ethical choices within these methods, which are not always straightforward.

This chapter illustrates how the internal elements discussed in Chapter 2 can significantly alter the conclusions of appraisals and evaluations. In other words, the applications highlight the sensitivity of the methodologies assessing efficiency to slight parameter variations or individual value judgements. This raises doubts about the ability of such methods to serve as the primary tool for budgeting decisions without any collective and political debate.

The first section discusses allocative efficiency, precisely three key elements of CBA: the willingness to pay as a proxy for welfare, the valuation process of non-market goods, and the strong inter-generational effects influenced by discounting decisions. The second section explores technical efficiency applications, emphasising how internal methodological choices—such as the selection of inputs and outputs and model orientation—affect results, as well as the significance of context and interpretation when addressing the estimations. The third section asserts the inevitability of politics in decision-making, providing examples of the World Bank's financing decisions, political resolutions that led to significant structural changes, and the evaluation process of public policies in Brazil. The fourth section illustrates alternative methodologies that can support decision-making without using efficiency as the benchmark for policymaking. Finally, it concludes.

3.1 Allocative efficiency in practice

Allocative efficiency is supposedly the best way to decide on different projects and policies, making the decision-making more rational and objective. As discussed in Chapter 2, CBA is the most commonly applied method to estimate allocative efficiency.

This section presents cases that illustrate inconsistencies and ethical dilemmas concerning three elements discussed in Chapter 2 of the CBA, which challenge the wellaccepted notion of neutrality. First, there is the reliance on WTP as a proxy for welfare. Additionally, examples of the benefits valuation process and the discounting methods used for projects with varying maturities reveal the lack of objectivity in numerous analyses. Lastly, this section briefly demonstrates how policymaking cannot ignore political and collective interests, which are fundamental to any social analysis.

3.1.1 Is WTP a proxy of welfare?

In neoclassical economics, preferences underpin the definition of welfare, and the traditional method for assessing whether a policy enhances social well-being is through people's willingness to pay (WTP) to implement that change. This is because, through WTP, policymakers can estimate allocative efficiency. Therefore, within this framework, decision-making in the public sphere should adhere to the same principles that guide purchases in the marketplace. In this context, WTP is viewed as an effective proxy for valuing individual welfare.

Sustein (2007) highlights that the WTP principle is "*an administrable way of ascertaining the welfare consequences of one or another approach*" (p. 306). That is, it appears as a convenient tool to guide allocative efficiency. However, the author stresses that the disconnection between the WTP for a good and the well-being obtained is often verified due to "affective forecasting"¹⁵, excessive optimism, myopia, and related phenomena.

For instance, he argues that people often underestimate their likelihood of being involved in a car accident, losing their job, or developing heart disease. Consequently, their willingness to pay for policies that mitigate these risks tends to be lower than the actual welfare these policies provide. Another example pertains to myopia, a common tendency to prioritise immediate gratification over future well-being despite the potential negative impacts on longterm welfare. He exemplifies this with issues such as obesity, lack of exercise, smoking, and drinking. Therefore, in the context of myopia, the WTP for preventive health policies that enhance long-term well-being and decrease future treatment costs is significantly less than the corresponding welfare.

In fact, defining WTP as a proxy for welfare has remarkable implications for policymaking that are much deeper than the abovementioned individual incorrect valuations.

¹⁵ People's current emotional states often have a large impact on their affective forecasts in a way that can produce significant errors. A simple demonstration, involving consumption choices, is that shoppers at grocery stores are greatly influenced by how hungry they are at the time that they shop (Sustein, 2007, p. 323).

101

For example, Sunstein (2007) questions whether WTP should govern international agreements on climate change or whether society should guide decisions regarding water pollution and protection against risks based on how much people are willing to pay to prevent adverse consequences.

Empirical studies on environmental policies reveal that the WTP for ecological protection is higher for the rich¹⁶, which suggests that developing countries might undervalue, from a CBA perspective, the importance of preserving the environment. The reason for this finding lies in the connection between WTP and the ability to pay, which serves as another critical explanation for why WTP is not a reliable welfare metric.

In fact, the connection between WTP and the ability to pay clarifies that using WTP as a proxy for welfare transposes market inequalities into policy evaluations. In a straightforward explanation, Sunstein (2007) argues that if individuals are poor, they will have limited ability to pay, which is directly reflected in their WTP for a policy, distorting any analysis. Therefore, when low-income individuals exhibit a small WTP for a policy, it does not indicate low gains in terms of welfare; it actually reflects their limited ability to pay for something.

Numerous practical examples can illustrate how applying the rationale of market dynamics, as revealed by the principle of WTP, raises significant social justice concerns. For instance, in the context of environmental preservation, policies that relocate polluting industries to low-income countries may be deemed efficient, as wealthier nations typically exhibit a higher WTP to safeguard their national environment and minimise pollution. In this sense, Ackerman and Heinzerling (2002) argue that resorting strictly to CBA and WTP to guide environmental policymaking implies that "*most environmental burdens will end up being imposed on the countries, communities, and individuals with the least resources*" (p. 1575).

Another relevant finding presented by Shao, Tian, and Fan (2018) is that the willingness to pay for environmental protection in China is more closely related to the degree of pollution experienced by the local population than to income alone. This evidence aligns with the myopia discussed above by Sustein (2007), which reflects a general tendency to place greater value on immediate well-being over future well-being.

However, more importantly, these examples emphasise the inadequacy of the WTP as a fundamental principle for assessing the benefits of policymaking. Exploring the Value of

¹⁶ For a literature review, see Shao, Tian and Fan (2018).

Statistical Life (VSL)¹⁷ can also illustrate this argument, showing how resorting to WTP transposes market inequalities to policymaking.

The VSL is a widely used tool in policy assessments to estimate how much individuals would be willing to pay to avoid certain risks. This statistic is applied in CBA and various methodologies across many areas, including health, transport, environmental, and regulatory analysis. Ozawa et al. (2011) highlight the employment of VSL in U.S. CBAs conducted by the Department of Transportation and the Environmental Protection Agency, as well as in analyses from Canadian and European countries. They also reference its application in cancer risk prevention programmes in China, tuberculosis control in Sub-Saharan Africa, and HIV/AIDS counselling and testing in Tanzania.

Since WTP is linked to the ability to pay, it is unsurprising that the VSL varies across countries. Viscusi and Masterman (2017) estimate an international ranking for nearly 200 countries' VSL based on the USA's \$9.6 million VSL. According to the authors, due to unreliable data, the calculated values represent the current best option for assessing VSL, particularly in low and middle-income countries. In addition, simply using the VSL of the USA would be inappropriate because differences in "*income levels, life expectancies, and social norms regarding risk and death may influence a particular nation's VSL*" (p. 227). Considering these reasons, the authors estimate, for instance, a VSL of \$ 102 thousand for Ethiopia or Mozambique and over \$ 16 million for Norway.

Country	VSL (\$ millions)
Argentina	2,144
Brazil	1,695
Cambodia	0,184
Canada	8,179

Table 10 - International Income-Adjusted Estimates of the VSL¹⁸ - selected countries

¹⁷ Ozawa et al. (2011) explain that the VSL embraces the idea of a trade-off between death risks and income. "Values are derived from both wage risk studies, which use labor-market data, and stated population preference studies, which asks individuals how much they are willing to pay to avoid certain risks of death" (p. 1012). For more details of the methods see Viscusi and Aldy (2003).

¹⁸ "Our approach is able to derive a VSL for almost 200 countries using a base U.S. VSL of \$9.6 million, a U.S. income of \$55,980, and a VSL elasticity of 1.0 to calculate a VSL for each country. We also estimate that low income, lower-middle income, upper-middle income, and upper income countries should use average VSLs of \$107,000, \$420,000, \$1.2 million, and \$6.4 million, respectively, based on the World Bank income group criteria. The international differences in the VSL will have ramifications for efficient regulatory policies, which will tend to be more stringent in more affluent countries. Because many countries will continue to lack detailed data on employment and workplace fatalities, transferring VSLs from the full sample of VSL estimates remains the most reliable way to generate a revealed preference VSL estimate for low and middle income countries. This benefit transfer approach will remain desirable until researchers develop more credible countryspecific VSL estimates" (Viscusi and Masterman, 2017, pg. 248).

China	1,364
Denmark	10,073
Ethiopia	0,102
France	6,975
Germany	7,904
India	0,275
Luxembourg	13,247
Mozambique	0,102
Norway	16,127
United States	9,631

Source: Based on Viscusi and Masterman, 2017, pgs. 245-247.

However, the difference between the VSLs does not reflect the difference in welfare obtained by avoiding risks. In fact, by relying on the principle of WTP, a market mechanism that governs private exchanges, policymakers inadvertently introduce inequalities in the decision-making process and implicitly assign greater value to high-income lives.

The same rationale can be applied to policy analysis within a country. For example, Becerra-Pérez et al. (2024) assess VSLs at sub-national levels in Mexico, obtaining the VSL for 32 states, ranging from \$400 thousand in Chiapas to \$3.3 million in Mexico City. The authors argue that the differences reflect the distinct perceived value of life and risk reduction in Mexico and that "*they provide valuable insights for policymakers and stakeholders involved in decision-making processes related to investment projects, resource allocation, and risk management strategies*" (p. 9).

In practice, a higher VSL indicates a priority in policymaking. For instance, from a CBA perspective, it is more efficient to provide vaccines to wealthier countries rather than poorer ones because the wealthier willingness to pay (WTP) is greater, implying more gains in welfare. Another relevant aspect concerns the decision of whether to implement a specific policy. In this context, Odon (2024) presents a hypothetical situation in Brazil during the pandemic, where the judiciary must decide on a lockdown. Agencies estimate a VSL of R\$ 5.67 million and outline two possible scenarios. The first predicts losses of R\$ 600 million and 220 lives, while the second predicts losses of R\$ 900 million and 150 lives. Thus, the question is: Are 50 lives worth R\$ 300 million? Since the VSL is R\$ 5.67 million, the answer is no. The author argues that total welfare is increased and there is no waste of resources. The answer would be yes if the same scenario were presented for any country with a VSL exceeding R\$ 6 million.

Besides the evident ethical divergences that might arise from those results, two other elements are worth noticing. First, the \$6 million per life does not even represent the value of what a government could pay to promote a lockdown and save an extra 50 lives. It is actually more connected to the lower or higher ability of the population to pay for something. The second is that WTP is not a good proxy for welfare. If someone is willing to pay \$20 for a vaccine and a richer person would pay \$200, it does not follow a difference in welfare provided but is more likely to mean a difference in income and wealth. In this sense, even though Sunstein (2007) disagrees with the differences in the VSL advocate against using the WTP principle, the author recognizes that "*if welfare is our guide, serious risks faced by people in poor nations deserve special priority and the WTP of those people is a poor proxy for welfare*" (p. 314).

3.1.2 The Valuation Process

As CBA provides binary answers, estimations must be accurate and based on transparent and widely accepted assumptions to ensure the results serve as a reliable metric for policymaking. However, this is seldom the case. For instance, a World Bank report on CBA in the organisation's projects highlights the obvious when justifying the low rates of CBA applications: many project analyses indicate that the benefits are not quantifiable, particularly in four areas: education, health, technical assistance, and the environment. The lack of data is another reason cited to justify the absence of a CBA in many projects, despite it being mandatory (World Bank, 2010).

On the other hand, when valuations are present, it is common to rely on previously measured generic values; that is, lists of estimated impacts are broadly applied irrespective of specificities and context. In this regard, Rowell (2021) discusses how CBA analysts make numerous choices when determining internal methodological elements, particularly concerning valuation. The general approach of assigning a monetary value based on aggregate individual preferences and their willingness to pay has given rise to what the author refers to as an 'industry' that has developed an extensive list of monetised impacts for non-market goods. Some examples include the Value of Statistical Life (VSL), unemployment, ecosystem services, recreation, and non-lethal incident illnesses.

For instance, the OECD (2018) summarises the monetary value of ecosystem services, divided by biome¹⁹. In addition, it classifies the services to connect the notion of ecosystem services from the natural sciences with economic requirements. For example, the cultural ecosystem services of recreation have an estimated value of \$96.302 per hectare per year in coral reefs, while this valuation is \$7 in woodlands. These numbers can be used as reference values to estimate benefits in a cost-benefit analysis to determine whether a specific policy is worth financing. For regulating services, climate regulation in tropical forests is valued at \$2.044 per hectare per year and \$65 in coastal wetlands.

¹⁹ Estimations are missing from the table for reasons such as a lack of data or the irrelevance of a particular ecosystem service for some biome.

		Marine	Coral reefs	Coastal systems	Coastal wetlands	Inland wetlands	Rivers and lakes	Tropical forests	Temperate forests	Woodlands	Grasslands
	Provisioning services total	102	55 724	2 396	2 998	1 659	1 914	1 828	671	253	1 305
1	Food	93	667	2 384	1 111	614	106	200	299	52	1 192
2	Water				1 217	408	1 808	27	191		60
3	Raw materials	8	21 528	12	358	425		84	181	170	53
4	Genetic resources		33 048		10			13			
5	Medicinal resources				301	99	1 504				1
6	Ornamental resources		472			114				32	
	Regulating services total	65	171 478	25 847	171 515	17 364	187	2 529	491	51	159
7	Air quality regulation							12			
8	Climate regulation	65	1 188	479	65	488		2 044	152	7	40
9	Disturbance moderation		16 991		5 351	2 986		66			
10	Water flow regulation					5 606		342			
11	Waste treatment		85		162 125	3 015	187	6	7		75
12	Erosion prevention		153 214	25 368	3 929	2 607		15	5	13	44
13	Nutrient recycling				45	1 713		3	93		
14	Pollination							30		31	
15	Biological control					948		11	235		
	Habitat services total	5	16 210	375	17 138	2 455		39	862	1 277	1 214
16	Nursery services			194	10 648	1 287		16		1 273	
17	Genetic diversity	5	16 210	180	6 490	1 168		23	862	3	1 214
	Cultural services total	319	108 837	300	2 193	4 203	2 166	867	989	7	26
18	Aesthetic information		11 390			1 292					167
19	Recreation	319	96 302	256	2 193	2 211	2 166	867	989	7	26
20	Inspiration					700					
21	Spiritual experience			21							
22	Cognitive development		1 145	22					1		
	Total economic value	491	352 249	28 918	193 844	25 681	4 267	5 263	3 014	1 588	2 871

Figure 9 - Summary of monetary values for each service by biome - International dollars per hectare per year, 2007 price level

Source: OECD, 2018, p. 319.

In the case of ecosystem services, the OECD (2018) stresses that using valuations to guide social decision-making is very challenging for many reasons. One of them is that it fails to reflect the importance of biodiversity. For example, because the valuation techniques rely on preferences, "charismatic" animals would have a greater chance of surviving extinction threats if CBA is the allocation method applied. The report illustrates this point by stating that the WTP to conserve lions, even when they are not at risk of extinction, is higher than the WTP to save frogs "*even when it is on the brink of extinction*" (OECD, 2018, p. 327).

From a cost perspective in a CBA, the valuation of the burdens associated with the environment includes estimations of morbidity, specifically non-fatal illnesses that represent social losses and, therefore, should be considered when evaluating public policies. For instance, when examining the potential negative impacts of increased air pollution, the OECD (2018) provides empirical evidence for the valuation of a) hospital admissions for both respiratory and cardiovascular issues related to ambient ozone or particulate matter (PM); b) restricted activity days (RADs) or lost work days attributable to ambient ozone or PM; c) chronic bronchitis in adults related to PM; d) acute bronchitis in children (aged 6 to 18 years) linked to PM; e) acute lower respiratory illness (ALRI) in very young children (under 5 years old) related to PM. The central value for hospital admissions is estimated at \$2,000 per case, while the valuation for a case of chronic bronchitis is \$334,750. Conversely, work losses are country-specific owing to wage differences. These estimations serve as reference values utilised to measure the costs (or part of the costs) when conducting a CBA to ascertain the merits of a policy.

Morbidity end-point	Central value	Range
Chronic bronchitis (per case)	334 750	41 700 to 889 000
Hospital admission (per case)	2 000	600 to 3 300
Work loss (per day)	Coun	try specific
RAD (per day)	170	41 to 268
Minor RADS (per day)	62	53 to 70
Acute bronchitis in children (per case)	464	301 to 511
ALRI in very young children (per case)	464	301 to 511

Figure 10 - Proposed unit values for selected morbidity end-points - USD, 2010 prices

Source: OECD, 2018, p. 319.

Mishan and Quah (2020) also provide many valuation examples according to the methodologies discussed in Chapter 2: the contingent valuation method, hedonic pricing, and the travel cost method. In common, the lack of accuracy stands out. For instance, from the

CVM, the authors cite the study of Hammitt and Zhou (2006), which discusses the health risks

associated with air pollution in China and estimates the value of preventing a cold at \$3 to \$6 per episode and the value of avoiding chronic bronchitis at \$500 to \$1000 per case. The values proposed for chronic bronchitis differ remarkably from the OECD (2018) abovementioned, reflecting, among other elements, that valuation techniques rely on WTP.

Another example brought by Mishan and Quah (2020) is the hedonic pricing method used by Day et al. (2007), which estimates the WTP for noise avoidance in Birmingham (UK). The authors estimate a value from £31.49-201.16 per annum (1997 value) for road noise reduction (1dB) and £83.61-1,488.88 for rail noise reduction (1dB). The values depend on the original noise level and reference the estimation of the benefits provided by implementing some policies to improve the acoustic conditions. This study argues that it regresses the observed noise exposure levels "*on calculated implicit prices in order to estimate a demand function for peace and quiet*" (Day et al., 2007, p. 213). It is important to recall that the hedonic pricing method equals the marginal WTP for each attribute of an object study to price. The authors recognize many caveats in their research. For example, the data refers only to preferences in property purchasing, not including the property-rental market, which excludes a significant part of the population under analysis.

Finally, Mishan and Quah (2020) provide examples of valuation using the travel cost method. For instance, Gürlük and Rehber (2008) estimate the value of recreational bird-watching at Lake Manyas (Turkey) at US\$ 103.23 million per annum, which would be a reference for benefit estimations in a CBA. Another unusual research conducted by Jeuland et al. (2010) uses the travel cost method to assess the willingness to pay for the cholera vaccine in Beira (Mozambique), which is only US\$ 0.85 per complete treatment (of two doses)²⁰. The remarkably low value assigned as a reference to estimate the CBA of the vaccination treatment against cholera again suggests a strong connection between WTP and the ability to pay, rather than indicating that the WTP principle serves as a good proxy for welfare.

Table 11 - Examples of valuation studies using differing valuation techniques

Valuation method Author (year)	Valuation item	Estimated item value
-----------------------------------	----------------	----------------------

²⁰ In 2005, Beira participated in a free-of-charge campaign for cholera vaccination, which resulted in 30.000 people from outside the targeted zone joining the trial. Therefore, the authors estimated travel cost methods to explore their WTP for those vaccines.
	Quah and Tan (1996)	Scenic View of East Coast Park (Singapore)	Present value of Singapore dollars S\$ 2.1-7.2 billion
	Amirnejah et al. (2006)	Existence Value of North Forests (Iran)	US\$ 30.12 per household per annum
Contingent	Aabø (2005)	Public Libraries (Norway)	400-2,000Kr per household per annum
Valuation Method (CVM)	Hammitt and Zhou (2006)	Air-pollution-related health risk (China)	Prevention of a cold episode: US\$ 3-6 per episode Prevention of chronic bronchitis: US\$ 500-1000 per case
	Yu and Abler (2010)	Air pollution in Beijing (China)	120.15 to 128.6 Yuan for blue skies
	Xie and Zhao (2018)	Green electricity in Tianjing (China)	32.63 Yuan per month per household
	Dewenter et al. (2007)	Mobile phone brand name Premiums (Germany)	Brand premium in the range \$57 -172
	Day et al. (2007)	Noise avoidance in Birmingham (UK)	Road noise reduction (1dB)*: £31.49- 201.16 per annum (1997 value) Rail noise reduction (1dB): £83.61- 1,488.88 per annum (1997 value)
Hedonic Pricing	Kong et al. (2007)	Percentage of urban green landscape within 0.3km radius in Jinan City (China)	63.55 yuan per percentage point increase
	Jiao and Liu (2010)	Recreational spaces of Changjian River and the East Lake in Wuhan, and city-level parks (China)	Up to 4109.2 Yuan/m2
	Gibbs et al. (2017)	Airbnb price listings based on physical and host characteristics, and location (Canada)	Varying for different characteristics and locations
Travel Cost Method	Shrestha et al. (2007)	Nature-based recreation in public natural areas of Apalachicola River, Florida (US)	US\$ 74.18 per visit day US\$ 484.56 million per annum
	Fleming and Cook (2008)	Lake McKenzie (Australia)	AU\$ 13.7 - 31.8 million per annum AU\$ 104.30 - 242.84 per person per visit
	Gürlük and Rehber (2008)	Recreational value for bird- watching at Lake Manyas (Turkey)	US\$ 103.23 million per annum
	Jeuland et al. (2010)	Private benefits of 'free' cholera vaccine in Beira (Mozambique)	US\$ 0.85 per complete treatment (of two doses)
	Mayer and Woltering (2018)	Recreational ecosystem services of national parks (Germany)	€385.3 million to €2.751 billion

* Value depends on the original noise level Source: Mishan and Quah, 2020, p. 211. Examples of the VSL applications can also demonstrate how the valuation process of a CBA may yield controversial or conflicting results. Furthermore, certain examples elucidate how internal methodological choices significantly influence results, which is problematic given that CBA offers binary outcomes and serves as an important reference for budgetary decisions.

For instance, Robinson, Sullivan, and Shogren (2020) explore three distinct results of VSL derived from various approaches to the impact of age on statistics, presenting implications for COVID-19 cost-benefit analysis. They argue that understanding the trade-offs between reducing death risks and increasing economic costs is fundamental to guiding policy responses to such crises. For them, the relationship between age and the VSL cannot be overlooked, as the virus's greater effects on older adults were soon observed. Additionally, the conceptual framework of CBA comprises the fundamental element that the sum of the effects across individuals represents the overall impact of a policy on welfare. "*This focus on individual preferences means that it is essential to supplement conventional benefit–cost analysis with assessment of the distribution of the effects*" (Robinson, Sullivan, and Shogren, 2020, p. 2).

In light of this observation, the authors present three approaches to age-adjusting the VSL. The first is an invariant VSL, which considers deaths irrespective of age. The second introduces a constant VSLY (value per statistical life year), which reflects the rate at which a person exchanges money for gains in life expectancy²¹. The final approach is the VSL that exhibits an inverse-U pattern, peaking in middle age. This viewpoint reflects the belief held by some that the relationship between age and the VSL should align with the consumption pattern throughout the lifecycle (Robinson, Sullivan, and Shogren, 2020).

Results show very different VSLs according to the relationship assumed between age and statistics, representing "*unit values per death reported*" (Robinson, Sullivan, and Shogren, 2020, p. 6). The invariant VSL presents a constant measure of \$10.63 million. The values of the VSLY decrease with age, ranging from \$13.88 million under 1 year old to \$2.03 million for adults over 85. Finally, the VSLs for the inverse-U approach increase until the ages of 45-54, when they start to decrease.

²¹ The VSL, on the other hand, is the rate at which a person would trade money for mortality risks. To estimate the constant VSLY, the authors divide the VSL by the average (discounted) remaining life expectancy of the population under analysis.

Age Group	Invariant VSL	Constant VSLY	Inverse-U Relationship
Under 1 year	\$10.63	\$13.88	\$5.38
1–4 years	\$10.63	\$13.74	\$5.38
5–14 years	\$10.63	\$13.37	\$5.38
15–24 years	\$10.63	\$12.64	\$5.38
25–34 years	\$10.63	\$11.76	\$8.50
35–44 years	\$10.63	\$10.63	\$10.63
45–54 years	\$10.63	\$9.19	\$10.72
55–64 years	\$10.63	\$7.54	\$8.15
65–74 years	\$10.63	\$5.68	\$8.15
75–84 years	\$10.63	\$3.72	\$8.15
85 years and over	\$10.63	\$2.03	\$8.15

Figure 11 - The VSL by Age Group (in 2019, millions of dollars)

Source: Robinson, Sullivan, and Shogren, 2020, pg. 6.

The different impacts of the three approaches become evident when applied to analyses of the effects of social distancing policies, which would subsequently guide policymaking. For instance, Robinson, Sullivan, and Shogren (2020) utilise the VSLs estimated above concerning three notable studies on the effects of the COVID-19 pandemic²². The authors compare the reduction in mortality risk with the scenario of no mitigation in each study by applying the estimated age-varying value of statistical life (VSL).

The results illustrate how different approaches to age-adjusting assumptions can significantly alter the conclusion regarding whether a policy is cost-beneficial. In general, Robinson, Sullivan, and Shogren (2020) demonstrate the tendency of the invariant VSL to yield higher benefits, and the differences are substantial enough to produce divergent policy orientations. For example, the positive cost-benefit result of the social distance policy of Thunström et al. (2020) depends on the valuation perspective. The original approach reaches a net benefit of \$5.2 trillion, indicating that the policy is cost-beneficial. On the other hand, if the constant VSLY age-adjusting is applied, results shift to a net cost of \$1,66, suggesting the inefficiency of the social distancing policy.

Regarding the study of Greenstone and Nigam (2020), although the authors do not estimate the costs that would make a CBA possible, differences in the effects of the three ageadjusting approaches are significant in the benefits valuation. These variations are even more prominent when applied to the baseline of Acemoglu et al. (2020).

²² Thunström et al. (2020), Greenstone and Nigam (2020), and Acemoglu et al. (2020).

	Costs	Lives Saved	Original Approach	Invariant VSL	Constant VSLY	Inverse-U Relationship
Thunström et al. (2020) Greenstone and Nigam (2020) Acemoglu et al. (2020)	\$7.2 trillion N/A \$2.15 trillion	1.24 million 1.76 million 8.7 million	\$12.4 trillion \$7.94 trillion N/A	\$13.16 trillion \$18.72 trillion \$92.44 trillion	\$5.54 trillion \$7.88 trillion \$38.93 trillion	\$10.30 trillion \$14.64 trillion \$72.31 trillion

Figure 12 - Effect of Alternative Approaches on Analytic Results

Source: Robinson, Sullivan, and Shogren, 2020, pg. 7.

Considering that CBA results determine whether a policy is efficient or not and that this outcome should be the guiding principle for the allocation of resources according to the mainstream perspective, it is disturbing that one single element of the estimation (there are many others) can provide such different outcomes, which can even result in opposite policy orientations.

Ozawa et al. (2011) also present a controversial study estimating the benefits of expanding vaccine coverage for children under 5 years old in 72 low- and middle-income countries. The results indicate that it would prevent the deaths of 6.4 million children, amounting to \$231 billion in the value of statistical lives saved between 2011 and 2020. The authors explain that this \$231 billion represents the value that these countries' populations would be willing to pay to prevent the deaths of their children.

However, the results vary significantly across countries, illustrating how income differences strongly influence outcomes. For instance, when comparing Angola and Nigeria, it is more efficient, from a CBA perspective, to save the lives of 12,700 children in Angola than to save 91,100 children in Nigeria because the VSL per capita in Angola is \$326,300, whereas the VSL per capita in Nigeria is \$39,300. The same reasoning can be observed when comparing other countries below.

Total VSL saved (millions of US\$)

Rank	Country	GDP per capita	VSL per capita	Vaccine-preventable child deaths	Estimate	Range
1	India	1,860	41,100	181,300	7,451	(3,544–20,646)
2	Angola	7,390	326,300	12,700	4,139	(2,272–8,484)
3	Nigeria	1,800	39,300	91,100	3,578	(1,696–9,985)
4	Indonesia	4,440	152,000	15,900	2,411	(1,260–5,473)
5	Pakistan	1,350	25,600	26,700	683	(321–1,999)
6	Sudan	2,440	62,100	8,900	552	(276–1,414)
7	Bhutan	2,810	76,500	6,100	467	(237–1,164)
8	Afghanistan	780	11,300	28,400	320	(140–1,065)
9	Kenya	1,480	29,300	9,300	274	(130–785)
10	Cameroon	1,260	23,000	11,200	258	(121–767)

Figure 13 - Top 10 countries with VSL savings

Since income and, therefore, willingness to pay (WTP) strongly influence results, the lives of those in poorer circumstances are undervalued both across and within countries. For policymakers, this implies favouring investments and policies that benefit those who are able to pay more. In fact, certain criteria ought to be applied to prioritise actions. However, these must be transparent and accountable to society. When someone argues that it is more efficient to select one option over another in health services, for instance, it is typically linked to an option that will save more lives or is deemed more cost-effective, which contradicts the rationale of the VSL.

Another relevant element to consider is the association of this rationale with welfare. Reducing welfare to income, as discussed previously, impoverishes policymaking because many other factors are essential to social well-being. More than just acknowledging this, applying the principle of WTP as a proxy for welfare introduces inequality as a guide for decision-making while presenting a misleading facade of technicality and neutrality.

Petri (2021) also cites two examples of controversial valuations, which raises doubts about whether the valuation process depends on widely accepted assumptions. The first concerns the valuation method used in US court decisions to assess the damages caused by illness or death for compensation purposes. The author emphasises that the standard method relies on the human capital approach, which bases calculations on investment theory. According to this method, people are viewed as capital investments with the sole objective of production. In this context, a worker's value is determined by their earnings, aligning with the neoclassical theory of income distribution, which posits that wages reflect the marginal productivity of labour. A direct consequence of this valuation method is that the illness or death of low-income individuals is considered to be worth less than that of high-income individuals, resulting in unequal compensation for the same damages. Petri (2021) also argues that this rationale is fundamentally the same when a country decides that health treatments are private services, as access to additional lives is only afforded to those who are willing and able to pay for them.

Another example cited by the author is the 'shocking use of cost-benefit analysis' in the 1971 Ford Pinto case. The car's design flaws caused it to catch fire in rear-impact collisions. Apparently, this problem was identified before sales commenced, but the company chose not to rectify the issue. According to them, a cost-benefit analysis was conducted, and estimates indicated that issuing a recall would be inefficient from a welfare perspective. They calculated the cost of wrongful death lawsuit settlements to be \$49.5 million compared to \$137.5 million in recall costs. However, subsequent information revealed that preventive costs would not exceed \$10 per car (no more than 2.5 million Ford Pintos were sold). Petri (2021) argues that "the episode is illustrative of the possibility to bend cost-benefit analysis to aims that general morality strongly rejects" (p. 1292).

In fact, morality and other value judgments are inherent to internal decisions on a CBA evaluation. However, Rowell (2021) discusses how CBA estimations are taken as technical and seldom investigated.

In this sense, just as some questions presented to agencies are "trans-scientific"—what Alvin Weinberg has termed "questions which can be asked of science and yet which cannot be answered by science" —so there are questions that are "trans-economic:" questions that can be posed to economics, but not answered by it. The trans-economic character of cost-benefit methodologies can obscure the ethical choices underlying existing practices (Rowell, 2021, p. 6).

The author highlights that valuations in CBA are based on ethical choices that are rarely uncovered. In fact, because few people are aware of such options, CBA analysts make these 'trans-economic' decisions without the support of other social scientists, philosophers, or legal academics. To illustrate these ethical choices, Rowell (2021) discusses three types of value judgment embedded in CBA valuation decisions. The first concerns the scoping of decisions. The second relates to the treatment of benefit transfer practices, specifically the valuation of mortality risks. Lastly, she examines the treatment of altruistic preferences, using the COVID pandemic as an example.

Ethical decisions regarding the scope of CBA involve various choices, such as who has standing, how impacts should be valued over time, whether national borders serve as the geographical delimitation, and whether future lives should be considered, among others. These choices are often opaque and, despite referencing non-economic elements, are made without drawing on insights from other fields of knowledge.

Differences in estimating the Social Cost of Carbon (SCC) exemplify how variations in the scope of CBA influence its outcomes. The monetisation of the damage caused by carbon dioxide emissions is a valuation process that has played a central role in US federal climate change policy. A single methodological choice regarding the scope of impacts was responsible for altering the results of the CBA. By determining the global scope of the Clean Power Plan and accounting for the effects of emissions on climate change worldwide, President Obama's team estimated a net benefit of \$26-\$45 billion for his climate policy aimed at reducing emissions. In contrast, Trump's administration narrowed the scope to national, overlooking the global impacts of carbon emissions, which resulted in a net cost of \$28 billion. Consequently, the climate policy for emission reduction was deemed inefficient. The key difference lay in the value assigned to benefits, which was \$48 per ton when impacts were considered globally, compared to \$1-\$7 per ton when effects were assessed solely on a national basis. (Rowell, 2021).

The point here is not to condemn choices but to reveal the often opaque individual ethical decisions that analysts must navigate within the CBA methodology. The discussion is not merely about scope but primarily about a political worldview that distinguishes both presidents. This is why the author emphasises the importance of understanding how CBA calculations are executed.

Secondly, Rowell (2021) discusses the ethical considerations surrounding benefit transfer practices, specifically, the decision of whether a more general estimation should be applied in a specific valuation process. As mentioned previously, there are lists of monetary impacts for many areas commonly utilised in CBA. A notable example is the VSL estimated for an entire population, irrespective of the sociodemographic characteristics of those exposed to the mortality risk in question.

The author stresses that if policymakers and regulators used income-specific valuations, CBA would consider inequality as a relevant element of the analysis "*in resource distribution and justifying the expenditure of more resources to protect those who are well-off than to protect those who are impoverished*" (Rowell, 2021, p. 9). In this sense, the author cites the example of the USA Science Advisory Board (SAB) to conduct different VSL approaches according to the impacts' timing.

Indeed, the SAB report recommend wealth-blind VSLs for current populations, even as it recommends wealth-sensitive VSLs for valuing risks to future people. Theoretically, such a distinction might be justified on ethical grounds, for example if the SAB believed that risks to future people are ethically distinct from risks to present persons. But the SAB—despite having previously recognized the moral, ethical, and political issues related to demographic-sensitive VSLs—did not cite moral, ethical, or political reasons for its decision. Instead, it frames its recommendations as having been based in the economics literature (Rowell, 2021, p. 10). In fact, not only income but also other factors, such as race and gender, could be included in the valuation process when discussing mortality risks. However, when these elements are introduced, the discourse surrounding ethical conflicts arises, as if the current practice fails to embrace its own ethical choices.

Third, Rowell (2021) discusses the treatment of altruistic preferences, regarded as "other preferences". These preferences are disregarded and excluded from CBA. However, people might have strong ethical concerns about how they affect others, especially when there is a close association between the possible harm inflicted and the author and a close relationship between the two parties. To illustrate the argument, the author highlights the valuation process of COVID-19 for many agencies, which applied generic VSL, the metric that estimates people's WTP to avoid mortality risks. In doing so, analysts implicitly disregarded the dangers of close contagion, that is, the risk of infecting a loved one. In this context, people would probably be willing to pay more to avoid mortality risks, resulting in a higher VSL. But by not considering "other preferences", analysts implicitly reduce the estimated benefits of a government's social distancing policy, undervaluing actions aimed at fighting COVID-19.

Ultimately, the examples presented in this section share a common point: the lack of transparency regarding the assumptions made in the valuation process. Key hypotheses and internal elements are frequently undisclosed and treated as technical choices, which complicates social oversight. In this context, Grubb et al. (2021) advocate for greater transparency in such valuations.

Any approach for valuation – intentionally or not – determines the relative weighting given to different interests and outcomes. Political decisions become at risk of being made implicitly by analysts. Analysts may become risk of politicizing their analysis through their choice of valuation methods. It may be preferable to ensure that the weighting of interests and outcomes be made independently by legitimate decisionmakers who can be held accountable for their choices or challenged in public debate (Grubb et al., 2021, p. 39).

In light of all the examples discussed in this section, it seems unreasonable to assume CBA can provide accurate valuations to support the binary conclusions elicited by the methodology. Moreover, the ethical and internal choices made in the valuation process are far from purely technical, embodying personal and collective value judgments that are seldom disclosed. These acknowledgements are especially relevant, considering that CBA is often

regarded as a 'technical' solution for the best allocation of public funding, which becomes vital to the budgeting process.

3.1.3 Discounting: discussing inter-generational effects

As discussed in Chapter 2, the choice of a social discount rate has a significant impact on CBA results, particularly when the analysis encompasses long-term effects. Therefore, this section presents practical applications that clearly demonstrate the influence of the discounting process on appraisal conclusions.

In analysing environmental policies, discounting is particularly relevant as it typically compares current costs with future benefits. This raises discussions not only regarding the appropriate rate but also an inter-generational debate concerning rights and collective interests. In this regard, the famous Stern Review²³ highlights how, beyond fostering an open public debate about policymaking to tackle climate change, internal methodological choices for determining the appropriate discount rate reflect worldviews that can significantly influence outcomes.

The Review supported strong measures against carbon emissions, which would require 1% of global GDP to avoid significant temperature increases. Stern's CBA estimations concluded that the benefits of early actions against climate change outweighed the costs, pointing to possible irreversible effects of a status quo scenario. However, cost estimations and the discount rate applied were the subject of much criticism. Stern used Ramsey's formula, which is discussed in Chapter 2, to assess the discount rate with the following parameters: $\rho = 0.1\%$, $\eta = 1$, and g = 1.3% to yield $r = 1.4\%^{24}$ (Cole, 2008).

The choice of the parameters was broadly criticised because values deviated from the literature and "best practices". The values were considered too low, violating the premise of reflecting somehow how people behave in a market-oriented context. In fact, Cole (2008) highlights that Stern's Review actually disregarded the market-based observations and behavioural inferences, placing his own ethical judgment into the analysis and, thus, challenging the common practice of economic analysis on climate change. For example, by

²³ Stern, N. (2006).

²⁴ Where: $r = \rho + \eta g$, $\rho =$ pure time preference, $\eta =$ coefficient of risk aversion and g = assumed growth rate of per capita consumption.

setting the time preference parameter " ρ " at 0.1, Stern ignored how people behave in markets, deviating significantly from the conventional range.

However, as Cole (2008) observed, not only Stern's but all CBAs are, to some extent, political and ethical documents given the inherently subjective elements that comprise the methodology. In the author's words:

From the valuations of nonmarket goods (including human lives) to the choice of value parameters (including discount rates) -and given the influence of those subjective elements on outcomes, each and every BCA inevitably is informed by the ethical, political, and/or ideological predilections of its author(s)²⁵ (Cole, 2008, p. 80).

Therefore, the assertion that Stern's Review is mainly political is true, but it is also true for all CBAs. The merit of this work lies in questioning whether market dynamics should indeed serve as the guiding parameters for analysing climate change policies. Furthermore, economic analyses cannot evade the political debate, particularly when policies address long-term effects and the impacts of diverse geographical contexts²⁶. In this sense, it is impressive that anyone would classify the CBA allocative tool as an efficient way to assign resources in society due to its technical and neutral characteristics.

As in environmental debates, discussions about discounting are particularly pertinent in healthcare decisions due to the time disparities between current costs and future health benefits. According to Jit and Mibei (2015), who conduct a critical review of discounting in the evaluation of vaccine programmes, time preference serves as the foundation for the theoretical debate in health economics, with the rate selected for discounting reflecting society's willingness to trade present consumption for future consumption.

The authors discuss how discounting health benefits significantly impacts vaccine programmes due to pertinent time differences between current costs and future benefits. Furthermore, the interaction of these time differences can be quite complex. To illustrate their argument, Jit and Mibei (2015) provide a table detailing the temporal and generational timing of health benefits from four vaccines: smallpox, human papillomavirus, varicella, and paediatric influenza. For instance, the human papillomavirus vaccine offers relevant benefits solely for future generations and only in the future. In contrast, the impact of the varicella vaccine varies considerably when viewed from intra- and inter-generational perspectives.

 25 BCA = CBA

²⁶ It is widely known that climate change effects are likely disproportionately affecting developing countries.

Moreover, benefits could even be negative for adults in the future due to the potential increase in zoster incidence among vaccinated adults.

Vaccine	Benefits	Benefits		
	Present generation (present adults)	Future generations (present children)		
Smallpox Present (1970) Future	Very small Very small	Large Very large		
Human papillomavir u Present (2010) Future	s Small Small	Small Large		
Varicella Present (2010) Future	Very small Possibly negative	Medium Large		
Paediatric influenza Present (2010) Future	Large Very small	Small Very small		

Figure 14 - Temporal and generational timing of benefits from four vaccines

Source: Jit and Mibei, 2015, p. 3790.

Balancing these time differences is a challenging process, made more complex by other elements involved, such as determining whether the rates for discounting are equivalent to the benefits and costs. Jit and Mibei (2015) clarify that numerous value judgments underpin the choice between constant and differential discounting. The argument for equal discounting is based on the premise that health costs and benefits hold equal priority when the value of health remains constant over time²⁷. In addition, supporters of equal discounting claim "time neutrality" by giving similar treatment to beneficiaries. On the other hand, counter-arguments include acknowledging that beneficiaries live in societies with different income levels and available health technologies and, thus, health valuations. Constant discounting is more accepted for intra-generational impacts.

However, inter-generational effects necessarily involve a wider reasoning about fairness, which becomes central to determining differential discounting. In the authors' words:

The validity of differential discounting depends on whether the decision maker is seeking to maximize welfare or health itself, whether the budget for health care is fixed, and whether the value of health changes over time. They show that the differential between the discount rate for costs and health can be informed by growth in either the value of health, or the cost-effectiveness threshold (Jit and Mibei, 2015, p. 3791).

²⁷ A smaller discount rate for health benefits implies an increasing value of health.

In their study, Jit and Mibei (2015) analysed 84 publications on economic evaluations of vaccine programmes. They discovered that most vaccine evaluations depend on equal, positive, and constant discount rates, which range from 3% to 6%, despite the fact that differential discounting would promote a sense of equity between generations. The authors conclude that decisions regarding discounting must be approached with caution due to the sensitivity of the results to this factor.

To illustrate the argument, Westra et al. (2012) present a study highlighting the varying health benefits from Human Papillomavirus (HPV) vaccination based on the method used to estimate discounting, which refers to weighing future health effects.

Results showed a substantial variation in health gains when different discount rates and approaches were estimated²⁸. The authors examined five discounting approaches, concentrating on different discount rates and methods to evaluate health benefits²⁹. The first was the constant discounting, which devalues future costs and benefits at a constant rate. The authors estimated four discount rates for health outcomes (0%, 1.5%, 3% and 4%)³⁰. The second approach was stepwise discounting, in which a constant rate is applied for a period and then lowered in subsequent periods³¹. The hyperbolic and proportional discounting methods, where rates gradually decline over time, followed the work of Asenso-Boadi, Peters, and Coast (2008) to estimate the parameters. Lastly, the time-shifted approach seeks to account for demand in health valuations for preventive interventions by discounting at the moment of risk reduction, rather than when health is affected "gained"³². The authors applied two discount rates to estimate the time-shifted approach (4% and 1,5%) (Westra et al., 2012).

Table 12 - Overview of different discounting approaches

Approach Description

²⁸ According to the authors, "All five discounting approaches were applied to the health outcomes of a Dutch HPV model. This model predicts the incidence of cervical intraepithelial neoplasia and cervical cancer incidence with and without HPV vaccination, reflecting the current Dutch situation". In addition, the model is static and does not account for the dynamic transmission of HPV, which would be more appropriate for infectious diseases.
²⁹ The authors did not vary discounting for costs, applying a constant level of 4%.

³⁰ The difference between the discount rate for costs and health effects would be the expected growth rate in health value.

³¹ The authors applied the rates recommended by the UK treasury (i.e., 3.5% for years 0–30, 3% for years 31–75, 2.5% for years 76–125, 2% for years 126–200, 1.5% for years 201–300, and 1% thereafter).

³² With this, the discounting happens at the moment of risk reduction.

Constant	Discount rate is constant over time, can be uniform or differential
Stepwise	Discount rate declines stepwise after specific time intervals
Hyperbolic	Discount rate declines gradually over time
Proportional	Discount rate declines gradually over time
Time-shifted	Time period between vaccination and the prevention of infection is discounted rather than full period up to actual quality-adjusted life-year gains

Source: Westra et al., 2012, p. 565.

Results showed the impacts of the different approaches on the estimated Quality-Adjusted Life Year (QALY)³³, which is reflected in the measure of Incremental Cost-Effectiveness Ratios (ICER)³⁴. The undiscounted perspective presents a higher QALY and, consequently, the lowest ICER, that is, the most cost-effective estimate. As for the discounted approaches, the proportional and hyperbolic ones are the most cost-ineffective, bringing the lowest estimation on QALY gained. The constant discounting, which is the most applied method for health-economic evaluations according to the authors, also presents relevant differences according to the specific rate chosen, ranging from an ICER of €18,400 to €59,100 (Westra et al., 2012).

Approach	QALYs gained	ICER (€/QALY)
Undiscounted	3.462	7.600
Constant 1.5%	1.423	18.400
Constant 3%	715	37.000
Constant 4%	438	59.100
Proportional	164	165.400
Hyperbolic	160	164.500
Stepwise	718	36.800
Shifted 4%	2.117	13.200
Shifted 1.5%	2.811	9.400

Table 13 - Discounted health outcomes of HPV vaccination using the different discounting approaches

³³ The QALY is a common metric applied to assess the value of medical interventions and compare different health options. It is basically the product of the life expectancy (or the years of life gained) with quality of life, which scales from 0 to 1.

³⁴ The ICER represents the incremental costs per QALY gained.

Therefore, Westra et al. (2012) conclude that choosing the approach and the specific rate within this approach can remarkably influence results. In fact, the authors argue that "*extremely large and relevant differences in the ICER were found between the various approaches investigated, moving from extremely cost-effective up to extremely cost-ineffective (when compared with commonly cited thresholds)*" (Westra et al., 2012, p. 565). This finding illustrates how internal methodological choices, which reflect the researcher's belief about how future benefits should be valued, are decisive for results.

3.2 Technical efficiency in practice

Technical efficiency embraces the notion of value for money, bringing the idea of producing more with fewer resources from private firms to the public sector. As discussed in Chapter 2, problems arise when the evaluation process loses sight of policy goals; that is, when the pursuit of minimal costs obscures policy objectives or fails to meet shared social values. This acknowledgment is pertinent, given that the government does not share the same objective as private firms: profits. In fact, policymaking encompasses multiple goals depending on the social issue at hand, necessitating various guiding principles and evaluation strategies to achieve valuable outcomes.

As discussed in Chapter 2, the methodologies for estimating technical efficiency include internal elements that can significantly influence the results. They are highly sensitive to the selection and weighting of inputs and outputs, technology choices, and potential production. Additionally, results are often not put in perspective, which requires thoughts about context that shed light on possible causes and solutions to the estimated inefficiency.

In light of these reflections, this section presents practical applications of technical efficiency analysis, discussing how internal methodological choices may affect results, thereby undermining the method's capacity to provide definitive rather than intermediate answers. Furthermore, it highlights the importance of context and interpretation in assessing real-life cases.

3.2.1 How Internal Methodological Choices Affect Results

The relationship between inputs and outputs forms the core of technical efficiency assessments, as it encapsulates the vital concept of minimising waste by achieving more with fewer resources. Nevertheless, selecting the variables, determining the model's orientation, and recognising that results represent merely one of the potential outcomes are critical aspects for understanding how this tool can facilitate decision-making.

In contrast to this reasoning, the World Bank's report on Brazilian public spending—A Fair Adjustment: Efficiency and Equity of Public Spending in Brazil—presents a conclusive analysis despite the obvious fact that internal choices are decisive for the results. One example is the selection of the DEA orientation.

The report indicates notable inefficiencies in health spending when Brazil is compared to other countries. Employing a DEA output-oriented approach for health spending in 2012, Brazil achieved a score of 0.91, suggesting that the health sector could only improve outputs by 9% with the same amount of funding. Conversely, in the input-oriented DEA, Brazil obtained a score of 0.41 in 2012, indicating that the country could spend less than half of its current resources to attain the same outcomes. The expressive differences depending on the approach can be explained by private sector providers, "*which spend large amounts on a small share of the population*" (World Bank, 2017, p. 107). That is, the private sector serves a small part of the population only in places where high profits are compatible with a firm's objective, thereby influencing the efficiency frontier.

On the other hand, the Brazilian government, which is responding to a constitutional mandate to universalise health access, provides services in all municipalities, irrespective of profit provisions. This mandate necessitates several fixed and potentially idle resources that clearly influence costs, which explains why the input-orientation DEA model for Brazil shows a low score when compared to the health systems of other countries that are largely private.

This is why efficiency comparisons must be made cautiously, and results should be subjected to careful contextual analysis. By comparing the Brazilian health system to those of other countries, the report failed to acknowledge essential differences between them, which largely explains what is presented as inefficiency in the report. It actually reflects diverse social agreements regarding public health provisions. Additionally, it makes evident how a single methodological choice – the model orientation – can significantly influence results.

Souza, Scatena, and Kehrig (2016) examine hospital efficiency in the Brazilian state of Mato Grosso, seeking to determine whether private or public hospitals are more efficient through a DEA. Given the importance of context, the authors advocate for the output-oriented DEA model, as public hospitals face high fixed costs and limited control over human resources due to the hiring framework in the public sector in Brazil. Their findings reveal that private hospitals are more efficient than public ones, with average scores of 0.93 compared to 0.81. However, the authors underscore certain caveats that must not be overlooked.

Firstly, they discovered that during the analysis period, the two largest public hospitals experienced underfunding, which may have compromised an important output proxy of the study – quality. Secondly, the authors emphasise the difference in the objectives of private and public hospitals, a significant factor that strongly influences the results. Furthermore, they emphasise that scores approaching one do not render private hospitals centres of excellence, serving merely as an indicator of consistency between inputs and outputs. Additionally, an important caveat is that the limited number of units under analysis complicates the model's ability to differentiate efficiency among them. Finally, Souza, Scatena, and Kehrig (2016) make a pertinent observation: results could vary if different methods were applied or if the selection of variables (inputs and outputs) was altered.

To illustrate this point, the work of Kohl et al. (2018), which reviews 262 papers on DEA applications in healthcare, particularly concerning hospitals, summarises the usual inputs and outputs selected, highlighting the vast number of variables and possible combinations for estimating a DEA model. Additionally, each category presents similar properties but may encompass many different variables. For instance, the category "nurses" below represents a group that includes nurse hours, assistant nurses, nurses' salaries, registered nurses, and licensed nurses. Consequently, each category includes numerous potential variables, and selecting one over another could influence the outcome. Furthermore, if the chosen variable is nurse hours, for instance, variations in legislation between the private and public sectors might introduce a significant bias.

Used at least once	totally used	Output category	used at least once	totally used
184	196	Outpatients	120	146
149	189	Other / total cases	118	226
114	129	Inpatients	103	120
93	124	Surgery	62	81
65	69	Services	48	100
60	70	Performance / quality	41	79
40	54	Others	37	47
37	38	Revenue	13	30
29	46	Case mix	5	10
27	35			
16	33			
5	17			
	Used at least once 184 149 114 93 65 60 40 37 29 27 16 5	Used at leasttotally used184196149189114129931246569607040543738294627351633517	Used at least oncetotally usedOutput category184196Outpatients149189Other / total cases114129Inpatients93124Surgery6569Services6070Performance / quality4054Others3738Revenue2946Case mix273516517	Used at least oncetotally usedOutput categoryused at least once184196Outpatients120149189Other / total cases118114129Inpatients10393124Surgery626569Services486070Performance / quality414054Others373738Revenue132946Case mix5273517

Figure 15 - Input and Output categories for hospitals

Source: (Kohl et al., 2019, p. 252).

Many combinations are possible, potentially leading to varied outcomes, as each combination establishes a unique efficiency frontier. Consequently, scores and rankings in performance may fluctuate, highlighting the sensitivity of DEA models to variable selection.

Another interesting observation made by Mardani et al. (2017) when discussing energy efficiency is that the concept may have different definitions depending on who is defining it. There are variations in this concept, for instance, among environmentalists, engineers, or economists. Consequently, they would choose different variables to assess energy efficiency. That is, the selection of inputs and outputs would most likely differ completely, substantially influencing the results.

3.2.2 The relevance of context and interpretation

Another essential feature of technical efficiency is the need to put estimation outcomes into perspective. This acknowledgment highlights a crucial aspect of any evaluation: the necessity of interpreting its results, given that any analysis of technical efficiency is merely an intermediate outcome that demands reflection. Consequently, the common conclusion that any identified inefficiency represents an opportunity for expenditure reductions is unjustifiable. Analysing the conclusions of numerous DEA applications can exemplify this argument. For instance, in the agriculture sector, Toma et al. (2015) utilised the DEA approach to estimate the regional production efficiency of 36 countries, which were categorised into three geographical areas: plain (20 countries), hill (8 countries) or mountain areas (8 countries). The authors explain this geographical separation by citing differences in the production patterns of these areas, which would make them incomparable.

Their results indicate that only 14 countries operate on an efficient scale (5 in plain, 5 in hill and 4 in mountain areas). According to them, other countries must adjust their input combinations, particularly concerning excessive working hours. In contrast, others must increase output levels by adapting their use of fixed capital (Toma et al., 2015). However, no consideration has been given, for example, to climate variations or any contextual factors, which makes all potential production differences a consequence of inefficient practices.

Nevertheless, certain countries may have, for example, stringent legislation against the use of agricultural pesticides and fertilisers, which could impact productivity. Another possible explanation might be variances in regulations regarding cultivation areas to protect the soil and prevent degradation. In such cases, the decisions are political rather than stemming from technical inefficiency and ought to be considered within the evaluation process.

In another example of entirely disregarding context, the previously mentioned World Bank report provided DEA estimates for Brazilian public spending in certain social areas. In addition to overlooking local realities, legislation, and established government goals, the report failed to analyse results within the appropriate policy context, rendering the evaluation a rhetorical tool to legitimise their position on the necessity to reduce public spending.

An excellent example is the conclusion regarding the significant potential savings in health expenditure by optimising the inefficient scale of service provision. According to the report's findings, if all municipalities operated like the DEA's efficient frontier, the government could save R\$22 billion. The most pertinent argument is that Brazil has many inefficient small hospitals. Therefore, the government should 'rationalise' the service delivery network, a euphemism for advocating the closure of public health assistance in small municipalities (World Bank, 2017).

However, as previously mentioned, universalising healthcare is a constitutional mandate, and deciding whether small municipalities should have access to this service is not a

technical decision but a political one. Santos (2023) adds that the isolated recommendation to reduce the number of small hospitals is problematic since the number of beds per inhabitant in Brazil is 2.2 per thousand, indicating low availability. Additionally, the author presents a strong argument regarding comparability in technical efficiency models. Inputs and outputs must be comparable between DMUs. Threats to comparisons include service heterogeneity, analysis of aggregate services, and distinct risk profiles of patients. None of these elements were discussed in the report, highlighting a poor interpretation of results and a lack of consideration for context.

This acknowledgement does not imply that municipalities ought not to adopt good practices from the most efficient DMUs, nor does it suggest that no adjustments should be made to enhance resource utilisation. It signifies that merely concluding that any estimated inefficiency indicates a fiscal space for reducing the budget without compromising the healthcare system is an absurd conclusion.

When discussing DEA applications in hospitals, Kohl et al. (2019) arrive at a similar conclusion, asserting that further exploration of DEA results is a crucial component of the analysis. In their own words:

Many studies end by reporting the efficiency values and illustrate how DMUs need to reduce the inputs in order to become efficient. In theory this might be true, in practice these results are almost never viable as discussed by Chen et al. [61]. Following theoretical results, a hospital might have to cut its inputs by 40% and cover the same amount of patients in order to be considered efficient. However, no hospital in the world will be able to implement this result straight away. Many inefficient units will more likely collapse than become efficient if they try to cut down their personnel in the way DEA results suggest (Kohl, 2019, p. 259).

The World Bank's report analysis on education spending serves as another example of how crucial it is to contextualise and carefully interpret results for the evaluation of technical efficiency. For instance, they argue that public spending on higher education in Brazil is inefficient, "*such that almost 50 per cent of resources could be saved*" (World Bank, 2017, p. 177) because students in public institutions cost two or three times more than students in private ones.

They also argue that the 'value added' by universities does not vary significantly between public and private institutions. To support this conclusion, they develop an indicator of 'value added' that supposedly reflects the difference between performance in ENADE – a compulsory examination taken by students at the end of their degree – and ENEM – an assessment that evaluates students before they enter university. The rationale for estimating the value added is

that students from public universities generally perform better in ENADE, but they attribute this to their prior education (primary and secondary). Therefore, in order to evaluate the efficiency of higher education institutions, they argue that the appropriate measure would be the value added for them. (World Bank, 2017).

The numbers and conclusions presented for evaluating the technical efficiency of public spending in higher education were controversial and subject to much criticism. Many argued that the report disregarded local context and distorted data to achieve results. First, critics stressed that private and public universities have many different systems and, thus, are not so easily comparable. For instance, while private universities primarily focus on graduation, public universities also invest in research and extension programmes, which require organisation, time, and additional human resources. None of these elements were included in the 'value added' estimation. The same reasoning applies to cultural activities, university hospitals, research laboratories, and so on.

Furthermore, the calculation of the cost per student in public universities included pensions, unlike private universities, which are subject to different legislation. Another controversy relates to the indicator created to estimate the value added by the institutions. Some argue that it is precisely because students in public universities have better backgrounds that their progress is lower. By achieving a certain level, there would be no space for improvement. Also, the student profile in public universities is changing, with lower-income students gaining more access, a fact not yet captured in this report. Finally, some argue that ENADE involves very specific content only learned in universities, making the exam alone the appropriate tool for measuring students' performance.

The report's conclusion regarding higher education is that public universities should charge for teaching, given that spending on tertiary education is regressive. Regarding this statement, it is notable that the agenda of charging for education is presented as a solution without any discussion of its political implications. Additionally, it is essential to highlight that the justification for implementing fees relies on a different benchmark: the regressiveness of public spending rather than efficiency. Interestingly, the use of alternative benchmarks for policymaking occurs only at convenient times.

The progressivity of any public spending is welcome; however, like efficiency, it cannot be applied out of context. There is a significant effort to expand access to public universities in Brazil, reducing the gap between higher-income and lower-income students. Furthermore, universities should serve as spaces of integration, and charging for higher education may push wealthier students towards private institutions, creating a specific excellence market that segregates students by income. Thus, an alternative approach to addressing regressivity in higher education involves increasing public spending, providing more university places, and reducing the entry gap, which aligns with the constitutional mandate that education is a social right that the state must ensure.

In conclusion, neither argument, for or against charging for higher education, can escape a discussion that transcends mere technicalities but involves worldviews and projects for civil society, independently of which benchmark is chosen.

The examples demonstrate that efficiency is an intermediate result, despite the methodology being deemed adequate. Technical efficiency, which focuses on outputs, should not overshadow outcomes, which represent the policy objectives (Santos, 2023). The author emphasises that the relationship between outputs and outcomes is not always straightforward, citing examples related to healthcare.

For instance, governments pursue facilitating access to healthcare and improving overall health conditions, not specifically making a determined number of heart surgeries. Another example is the lower correlation between the number of medical appointments or hospital discharges (inputs) and quality years of life (outcomes) in OECD studies. For example, Sweden and Mexico have the same average of three medical appointments per person yearly, but very different quality of life expectancy: 74 in Sweden and 67 in Mexico (Santos, 2023). These findings strengthen the assertion that technical efficiency is a pertinent yet intermediate aspect of evaluations.

Additionally, Santos (2023) summarises how the simple examples of disconnection between outputs and outcomes draw attention to three reflections. First, models that estimate technical efficiency cannot capture all relevant dimensions involving health systems, which include many variable inputs according to different contexts. Furthermore, technical efficiency does not necessarily address effectiveness. For instance, an excessive number of exams can actually be harmful to health due to exposure to iatrogenesis. This point was made in Chapter 2 when discussing the ideas of Bogetoft and Otto (2011), which highlight that moving from effectiveness to efficiency—that is, from policy goals to minimal costs—might result in losing sight of relevant objectives. Third, determinants external to the health sector and the quality of care are undoubtedly significant aspects of outcomes that technical efficiency models do not capture.

Therefore, conclusions over the estimations cannot preclude context, social and political debate and connection with public goals. Technical efficiency is undoubtedly a relevant aspect of policy evaluations, especially concerning the common idea of avoiding waste. In addition, good practices compatible with a set of shared social values are also definitely welcome. However, if better performances mean constraining human rights, society might decide not to seek this pattern. For example, targets might be more efficient due to lower payrolls or excessive pressure on the workforce.

More importantly, technical efficiency cannot be separated from public objectives and become an end in itself. Unlike private firms that aim to maximise profits, public actions pursue multiple goals that cannot be reduced to a single guiding principle. As recognized by Boueri, Rocha, and Rodopoulos (2015), "*Technical efficiency of a given good in isolation does not provide relevant information*" (p. 225 in free translation).

In conclusion, it is indisputable that agents must act with care and due diligence when dealing with public spending. It seems hard to believe that anyone would argue against that. Thus, technical efficiency is a relevant dimension of evaluations. Analysts should pursue best practices that can improve the goods and services provided to society. The learning dimension of benchmarking is welcome.

However, as some advocates say, efficiency cannot dictate public funding. Decisions about public funding should embrace collective choices about society's goals, which relate to other benchmarks. For instance, if reducing inequality is a crucial objective, then efficiency is not a proper guiding principle. With the caution that methodological shortcomings have shown, technical efficiency analyses should be a tool that supports the decision-making process, but not in an absolute role. Instead, it should guide analysts to adopt good practices aligned with broader social goals.

This acknowledgement is especially pertinent considering that efficiency is a guiding principle for policymaking and is frequently associated with the quality of public spending. This makes the commonly accepted idea of an evaluation providing binary answers a real problem, especially because technical efficiency analyses are often made without context or reservations, making every result less than the score one related to the idea of waste, that is, with avoidable inefficiency. As a result, recommendations usually involve cutting spending to the same extent as the inefficiency estimated.

3.3 Political and Collective Interests

Policymaking is a political process that should be supported by both technical quantitative and qualitative analysis. However, these analyses cannot replace the fundamental sociopolitical elements that precede any decision-making. Relevance, targets, objectives, and goals form the foundation of public policy and lie outside the technical domain. Therefore, decision-making cannot ignore the collective priorities established by society.

It is not surprising that, despite the assertion that policymakers should pursue allocative efficiency, decision-making remains within the political sphere, even when the rhetoric surrounding the supremacy of efficiency is vigorous. For instance, an independent evaluation group presented a report analysing four decades of cost-benefit analysis (CBA) applications in the World Bank's appraisals. The report finds that the percentage of projects with such analysis fell from 70% to 25% between 1970 and 2008, despite the Bank's explicit policy that all projects must include a CBA³⁵. Even in projects that present CBA, the lack of transparency on data is often a problem (World Bank, 2010).

According to the report, the most relevant finding is that CBA "*is usually conducted after the decision has been reached to pursue a project*" (World Bank, 2010, p. 32). Furthermore, 82% of project leaders stated that CBA was not a primary criterion for deciding to fund a project. In light of these findings, the report concludes that the Bank's policy requires revision to enforce the use of CBA because "*other than quantitative cost-benefit analysis, no methodology can answer the basic question of whether benefits exceed costs*" (p. 48). With this, the Bank would protect the decision-making process from political and sectoral influences.

However, this conclusion does not make sense because politics is an inherent aspect of any decision-making, as it has been shown in the report's analysis over four decades of the

³⁵ Only in exceptional cases when monetary values cannot be attributed to benefits can a Cost-Effectiveness analysis be applied. Even so, Cost-Effectiveness has not been a decision-making tool for the World Bank. According to the report: "A classic cost-effectiveness analysis starts by stating a specific goal, such as reducing the incidence of a disease in a town by 50 percent in four years, presents data on the expected cost of two or more methods of achieving this goal, and then selects the least-cost alternative. The 24 projects that invoke cost-effectiveness analysis, however, do not mention a specific alternative to the project chosen. Second, project documents usually examine the costs of doing the project rather than those of achieving a meaningful goal such as disease reduction" (World Bank, 2010, p. 15).

World Bank's funding. In addition, if the decision to pursue a project happens before any CBA estimations, it actually raises even more doubt about the quality of the CBAs performed, which are tied to a previously defined result. In this case, the supposedly technical CBA is probably applied to eliminate the political debate over the Bank's funding priorities. Indeed, this may be one of the reasons why the lack of transparency is one of the relevant findings of the report.

Moreover, even if CBA is deemed a good estimate for allocative efficiency, the Bank cannot preclude decisions of a political nature because deciding to fund projects in different areas requires some previous guidance. For instance, a project yielding a positive net outcome for health in Brazil cannot be compared to one with a net positive result for transport in India, even if one has a higher net present value, simply because they are not equivalent. The goals are entirely different, and if funds are limited, the Bank would need to prioritise funding, whether that pertains to areas, regions, population profiles, etc.

Another example of how policymaking cannot preclude political decisions is the discussion of successful policies in the low-carbon transition. The Economics of Energy Innovation and System Transition (EEIST) Consortium concludes that CBA is inappropriate for dealing with the risks and opportunities associated with structural changes and, thus, is not an adequate methodology for dealing with climate change policies. Firstly, the framework does not account for systemic uncertainties, such as structural changes. Secondly, the results are consistently biased towards the status quo due to their marginal nature, which conflicts with substantial transformations. In the author's words:

Prices, demand, productivity or other macro variables are assumed in CBA to be unchanged by the application of the policy. Yet deep decarbonization will be entwined with innovation and huge changes in energy markets, invalidating the original assumptions, and hence conclusions, of classical CBA based on implicit assumptions of marginal change (Grubb et al., 2021, p. 4).

The authors present three case studies to illustrate how CBA can be misleading. They challenge the common argument that estimating the costs and benefits of policies is preferable to not doing so. In all the cases discussed, transformations in key energy technologies—wind energy, solar photovoltaic, and efficient lighting—resulted from strategic and political choices and would not have survived a CBA decision-making process.

For instance, in Brazil, the development of onshore wind energy was underpinned by a policy framework endorsed by numerous institutions, a long-term strategy, and the crucial role

of accessing funding with lower interest rates from the Brazilian Development Bank (BNDES). The strategic plan aimed to diversify the energy mix, which in 2001 comprised 75% hydropower and was impacted by an extended period of drought, thereby ensuring supply security and access to affordable energy. The set of measures involved contracts for long-term purchase agreements at a fixed price, funding tied to a 60% nationalisation index, and infrastructure investment programmes. Consequently, investment costs significantly decreased, and a relevant supply chain industry was developed, employing 150,000 people. As a result, between 2010 and 2020, the supply of wind energy increased from negligible levels to almost 10% of all electricity in Brazil (Grubb et al., 2021).

Another example is solar photovoltaics (PV) in Germany, which was driven by oil shocks, acid rain, and the Chernobyl nuclear disaster. In the 1990s, the country introduced Feedin Tariffs (FiTs) for renewable energy, contrary to the consensus in mainstream economics regarding the need for neutrality. The Renewable Energy Act of 2000 established a target for renewables to achieve 20% of electrical production by 2010. Once more, the emphasis on solar PV was both strategic and political, "*with little sign of CBA*" (Grubb et al., 2021, p. 15), especially because it was an expensive policy. Between 2004 and 2010, Germany accounted for over half of the world's solar PV capacity. During this same period, costs did not decrease, which again argues against the policy. Instead, the government involved citizens in the energy transformation, turning local communities into energy producers that benefited from high tariffs.

On the other hand, costs were socialised. Later, costs dropped with the expansion of the supply chain. The authors stress that all forecasts, for instance, from the International Energy Agency (IEA), were expressively wrong and that "*had they been fed into CBA, they would undoubtedly have indicated the programmes should be cancelled*" (Grubb et al., 2021, p. 17). In the author's words:

The decisions which led to these developments were driven primarily by strategic and political factors. To judge from the estimates, projections and common pronouncements on solar PV and wind a decade earlier, formalized CBA based on trying to monetize the costs and benefits on the basis of marginal, 'equilibrium' welfare economics would probably have killed these programmes (Grubb et al., 2021, p. 21).

Once again, the importance of not limiting a strategic sociopolitical decision to a costbenefit analysis becomes clear. If CBA had been applied in the abovementioned studies, how much would individuals be willing to pay for those policies? How would all the intangible benefits be assessed? Given the substantial costs associated with implementing those policies, what would the conclusion be?

Lastly, this section discusses the evaluation process of public policies in Brazil, demonstrating that despite the common rhetoric surrounding the necessity to prioritise technical and allocative efficiency, the assessments reveal other principles based on the objectives previously defined. In this regard, this work analysed 65 reports identified by the Brazilian government as exhibiting allocative or technical efficiency analysis³⁶.

The 65 reports included evaluations from various sectors, with 55% covering health, social security, social assistance, education, science and technology, along with a category of undefined areas that involve transversal policies. Despite the assertion that all 65 reports included efficiency analyses, only six provided any estimation of efficiency. The remaining 59 made generic references or assessments that do not relate to the discussed concepts of allocative or technical efficiency.

For example, many reports brought cost estimations without relating them to any measure of benefits, which would make them converge to the concept of a cost-benefit analysis, nor did these cost estimations make connections between inputs and outputs to allow a technical efficiency analysis, making room for the idea of value for money. Furthermore, the six reports that included some efficiency analysis did not highlight efficiency as the most significant feature to justify the relevance of the policy. In other words, even when the principle of efficiency was present, it did not emerge as the central element of the assessment.

The reports provided a thorough analysis of compliance, accountability, effectiveness, impacts on various dimensions, distributive effects, and opportunities for improvement. These assessments were comprehensive and did not limit themselves to the narrow concept of efficiency in evaluating the merits of the policy under review. The reports serve as a valuable tool for decision-making and support the political choices made during the budgeting process.

Therefore, the Brazilian government does not resort to allocative efficiency to dictate public spending, making the evaluation process a relevant but ancillary tool. Of course, there is always room for improvement, including increasing estimates of savings opportunities that are

³⁶ The reports were requested by an accountability tool guaranteed by a Brazilian law on access to public information. In December of 2024, it was required all evaluations conducted by the federal government that contained allocative or technical efficiency analysis. The Brazilian federal government formally answered the request with the list of the 65 reports.

compatible with other public guiding principles. Additionally, these reports could be more integrated into budgeting decisions. However, as emphasised earlier in this work, evaluations cannot dictate choices by presenting binary answers to assess the merits of a policy. They must serve as a comprehensive support mechanism, incorporating various elements, principles, and estimations to provide high-quality information to policymakers, politicians responsible for budgeting, and society as a whole.

3.4 Alternative methodologies

Despite attempts to reduce decision-making in the public sphere to a market-oriented perspective—equating the rationale of policymaking with firms' guidelines—there exist alternative approaches that do not adhere to such a viewpoint. Firstly, these approaches do not rely on binary reasoning; that is to say, estimations do not dictate the merits of a policy. Secondly, they do not commit to efficiency as the sole benchmark for policymaking. In fact, any other benchmark, such as reducing poverty, pursuing inclusive growth, or enhancing labour market equality, can inform decision-making. The benchmark aligns with the policy's overarching objectives.

For instance, impact analysis offers another perspective on policymaking. According to Gertler et al. (2016), this approach falls within the evidence-based agenda, aligning with the global trend to focus on outcomes or results. The most significant characteristic of impact analysis is its assessment of whether a particular policy intervention affects the well-being of the population or a specific targeted group. As Khandker, Koolwal, and Samad (2009) clarify, impact analysis specifically seeks to establish a connection between the programme and its effects, clarifying what can be directly attributed to the policy in question. Therefore, a key element of this perspective is the causal relationship between the policy under consideration and the outcomes.

Three remarkable elements distinguish the traditional market-oriented perspectives from impact analysis. First, unlike CBA analysis, in which welfare is a synonym of neoclassical allocative efficiency, the impact analysis approach does not previously define what welfare means. The idea of a policy that increases welfare is closely tied to the defined policy objectives, which have a sociopolitical nature. Therefore, impact analysis can have various benchmarks. For instance, enhancing welfare could include income distribution, poverty reduction, economic growth, combating violence, preserving historical heritage, and so forth. Another relevant element of impact analysis is the direct link between evaluations and public policy objectives, which is not present in efficiency evaluations. In these, the performance is associated with the idea of maximization, which loses sight of the policy's objective. On the other hand, in impact assessments, the relationship is directly with the desired objective, with costs being a relevant but accessory dimension of the process. This difference reflects the lack of commitment of the impact analysis approach to making a symmetry between governments' and firms' behaviour.

Finally, impact analysis does not aim to quantify in monetary terms every possible effect of a policy to provide a conclusion. In fact, impact assessments can be quantitative or qualitative. A qualitative impact evaluation is essential for understanding the sociocultural and institutional context, programme details, and so on, serving as a fundamental tool for informing the mechanisms through which the programme benefits the beneficiaries. On the other hand, quantitative impact assessments aim to confront policy effects with a counterfactual, which is the outcome in the absence of the intervention, directly connecting the policy to results. They are complementary and vital for providing quality information for decision-making.

The ex ante quantitative impact analysis typically employs structural models to estimate the effects of future programmes and policies. The ex post quantitative impact assessment evaluates the effects of ongoing or terminated programmes to determine the impacts on beneficiaries attributed to the programme. Various approaches can be utilised: Randomised evaluations; Matching methods, specifically propensity score matching (PSM); Doubledifference (DD) methods; Instrumental variable (IV) methods; Regression discontinuity (RD) design and pipeline methods; Distributional impacts; Structural and other modelling approaches (Khandker, Koolwal and Samad, 2009).

A practical example of the difference between an efficiency-guided approach and an impact perspective concerns the assessment of a vaccine programme. As discussed earlier in this chapter, the CBA employs the VSL as a parameter to evaluate a programme's efficiency. In other words, it refers to the principle of WTP, a controversial market-oriented element, to determine the merits of a vaccine programme. Alternatively, it is possible to use other benchmarks to evaluate the same programme.

For instance, Chang et al. (2018) examine the health and economic impacts of vaccines on ten antigens across 41 low- and middle-income countries from 2016 to 2030. The authors concentrate on the distributional health and economic effects that these vaccines could entail, particularly concerning the lowest quintile of the population, which directly aligns with the goal of poverty alleviation.

According to the authors, in 2010, the World Health Organization reported that healthcare costs push 150 million people into poverty each year. Therefore, reducing out-of-pocket health spending could help prevent extreme poverty. For this, besides the health benefits, measured as the number of deaths averted, they estimate the household economic impact of the vaccine program. The financial effect is examined in terms of the number of instances of medical impoverishment; specifically, the reduction in the number of households falling below the World Bank's poverty line (\$1.90 a day) due to medical spending (Chang et al., 2018).

The total number of deaths averted is estimated to be nearly 36 million across the 41 countries. In comparison to the counterfactual scenario, which estimated the number of potential deaths that would have occurred in each quintile of the population without the vaccine programme, the authors find that the poorest quintile benefits the most from immunisation. The lowest quintile experienced the highest percentage of deaths averted (23-34%), and the two lowest quintiles together accounted for over half of the deaths averted (Chang et al., 2018).

The estimated number of cases of medical impoverishment averted totals 24 million cases for the period 2016-2030, which, considering 2013 data, represents almost 9% of individuals in low-income countries living below the \$1.90 poverty line. Thus, the vaccine programme demonstrates significant results in preventing medical impoverishment. Moreover, the quintile that stands to benefit the most is the poorest one, highlighting not only the distributional health impacts but also the economic ones (Chang et al., 2018).

Figure 16 - Distribution, by income quintile, of deaths averted and cases of medical impoverishment averted by vaccines to be administered in 41 low- and middle-income countries, 2016–30.





In conclusion, the immunisation programme would likely provide greater health and economic benefits for the poorest quintile of low- and middle-income countries, enhancing equity and reducing extreme poverty. Furthermore, Chang et al. (2018) emphasise the pertinent policy implications of their analysis, which suggest that prioritising the poorest quintile would yield the most substantial advantages in health and economic terms. This finding aligns with the established benchmark: improving equity and alleviating poverty.

Finally, the authors highlight the need for a more active role in reaching the lowest quintile:

Merely ensuring equal access to vaccines will not reduce the health and economic outcome gaps that exist across income quintiles. The poor face higher baseline risks, which are tied to social determinants of health, and they have lower access to treatment. Additional steps may be needed to address those factors (Chang et al., 2018, p. 322).

Concerning this point, it is essential to note that decision-making may not necessarily be the most cost-effective solution because, as stressed by the authors, "*reaching the poorest quintiles could be substantially more expensive than reaching the richest*" (Chang et al., 2018, p. 319). A policy might enhance equity and alleviate poverty without necessarily benefiting the

Source: Chang et al., 2018, p. 320.

poorest quintile. Consequently, the programme may incur higher costs to achieve this particular outcome, which involves a political decision.

This example of a vaccine programme illustrates the differences between an efficiency approach and an impact analysis. The market-oriented cost-benefit analysis (CBA) approach values the programme based on people's willingness to pay for it, irrespective of social objectives or their ability to pay. The merit of the vaccine programme is determined by comparing these benefit estimates with the associated costs, reflecting the idea of maximising behaviour. Conversely, the impact perspective links the selected benchmarks with a causal simulation that aims to uncover the effects of the vaccine programme. The programme's value is defined by the chosen social objectives, rather than by the notion that the government is acting like a private firm.

Another example of an alternative methodology to support policymaking concerns decisions involving structural changes. As discussed in Chapter 2, one of the criticisms of the CBA is its marginalist nature, which biases the method towards the status quo. To address this, Grubb et al. (2021) propose a new appraisal approach. The Risk-Opportunity Analysis (ROA) is a framework for supporting decision-making that focuses on assessing the risks and opportunities associated with transformational changes.

The authors argue that the rationale for CBA is correcting market failure because marginal changes do not entail new types of economic resources. Conversely, non-marginal policies require a different perspective due to the continuous creation of new structures and resources. "*Without an equilibrium, an optimal allocation of resources cannot be identified*". With this: "*The focus is on dynamic effectiveness instead of static allocative efficiency*" (Grubb et al., 2021, p. 38).

The authors discuss the case of low-carbon transition, arguing that traditional CBA fails to capture many essential and intangible benefits, such as the positive impacts of developing new technologies, supply chains, and jobs, as well as feedback loops. Unlike CBA, the ROA considers all opportunities, even when assigning a numerical value to this impact is not feasible. Thus, instead of artificially reducing all effects to a single metric—money—like the CBA does, the ROA identifies the various possible outcomes to assess the pros and cons, making diverse interests transparent and accountable to society. They argue that quantifying every impact in monetary terms undermines transparency, as numerous decisions are made implicitly. Consequently, ROA transitions from CBA's one-dimensional approach to a multidimensional assessment (Grubb et al., 2021).

Another pertinent aspect of the ROA is the dynamic nature of the analysis. The authors emphasise that in complex systems, the relationship between components conveys more significance than the individual components themselves. Therefore, policy analysis that involves transformational changes should not be evaluated in isolation, but rather in terms of the relationships among the fundamental elements. "*These may include innovation, diffusion, growth, contraction, reorganization, or replacement of one or more sets of economic resources, assets, or structures, with another*" (Grubb et al., 2021, p. 39).

	Where the aim or expectation is marginal change	Where the aim or expectation is non-marginal change	Reason for difference (in non-marginal case)
Purpose of the policy intervention	Allocative / static efficiency	Dynamic effectiveness	Primary concern is not how efficiently resources are allocated (optimisation), but how effectively economic structures are changed or created (steering)
Rationale for policy	Market failure	Market shaping	Over periods or scales of concern, existing markets are changing, or new ones emerge, so that optimal states cannot be reliably identified
Appropriate analysis	СВА	ROA	Fundamental uncertainty makes precise expected future costs and benefits unknowable
Appropriate models	Equilibrium / optimising	Disequilibrium / simulating	Need to assess effect of policy on processes of change, not just on destination
Theoretical basis	Equilibrium / welfare economics	Complexity economics	Need theory that can explain non-marginal, irreversible and transformational change where relevant

Figure 17 - Key differences between CBA and ROA

Source: (Grubb et al., 2021, p. 41).

Mercure et al. (2021) summarise the methodological steps in ROA, which are presented in the table below. One notable aspect of ROA is that both quantitative and qualitative evidence are equally relevant and fully considered in decision-making. Additionally, it is important to note that the method does not monetise every impact arising from policies, rendering a binary result inconsistent with ROA. Consequently, unlike CBA, ROA does not offer a definitive answer. Rather, the method evaluates both quantitative and qualitative impacts derived from the system under analysis to aid the decision-making process. Ultimately, this approach incorporates the effects of various policies, acknowledging the interdependence of policymaking, thereby making this dynamic perspective significantly more engaging in the realm of public policy.

Table 14 - Methodological steps from ROA

Step	Description
1	Identify the boundaries of the system and map the relationship between the components. Make dynamical quantitative and/or qualitative analysis models and datasets accordingly.
2	Estimate median (not mean) outcomes and impacts on the process associated with each policy and make possible scenarios with ranges of uncertainty.
3	Make risk assessments for each policy to predict worst-case scenarios along with their likelihood.
4	Make an opportunity assessment by varying the scenarios and using other methods to identify the potential of each policy.
5	Report to decision-makers median impacts, direction of system change feedbacks, risks and opportunities, in all dimensions considered, along with uncertainty ranges and/or confidence levels. Report both qualitative and quantitative evidence, against current regulatory norms and risk tolerances. The normative weighting or valuation of outcomes is not considered part of ROA.

Source: Based on Mercure et al. (2021), p. 5.

In conclusion, both impact analysis and ROA do not adhere to the market-oriented rationale that creates a false symmetry between government and firms, imposing market dynamics on policy analysis. In contrast, these alternative methods pursue estimations that link policymaking with social and political objectives, allowing for qualitative assessments and other benchmarks for public policies.

Final remarks

If methodologies that produce binary results, such as the CBA allocative approach or the most common application for technical efficiency—the DEA—cannot provide accurate, objective, and rational outcomes, how can policymakers follow a neutral guideline? The answer is straightforward and has long been indicated: they cannot, simply because there is no objective and technically neutral analysis. Every methodology involves a set of hypotheses that reveal, at least partially, a worldview. Moreover, assuming that some method can replace human judgment merely highlights an outdated concept of what science is.

Acknowledging this thought paves the way for recognising that policy evaluation is a complex process that involves weighing collective interests and defining guiding principles and ancillary estimation methods that better align with broader social goals. Furthermore, the imperative of efficiency as a benchmark for policymaking ceases to exist. In fact, any

142

benchmark can guide public actions, and choosing between them relates more to defined policy goals and collective interests than to any principles previously established by economists.

This reflection prompts us to disassociate efficiency estimations from the definition of quality in public spending. This term relates more to the link between spending and the attainment of social objectives than to the association of government performance with market-oriented behaviour, like that of firms. In other words, quality spending can be defined as significantly reducing poverty or protecting the environment.

Consequently, as an intermediate principle in policymaking, efficiency should not be the sole factor in budgeting, as resource allocation within society must address collective interests rather than market-driven outcomes. This entails acknowledging that budgeting decisions ought to remain what they have always been: political decisions.

CONCLUSION

Critics of neoclassical economic theory have long noted the false symmetry between firms and families when discussing the firms' maximising behaviour imposed on families by neoclassical consumer theory in welfare economics.

The theory has also established economic efficiency as the welfare criterion to guide the state's actions, implicitly defining a specific distributive judgment. Consequently, if government policies adhere to efficiency as a guiding principle, total welfare is enhanced. In this context, certain methodologies have been developed to assess the efficiency of a policy. In this transition from theory to practice, it has been agreed that the methods used to analyse policymaking follow the logic of those applied to private firms, imposing upon the state the performance expectations of a private firm. Therefore, once again, a false symmetry is observed, but now between governments and private firms.

Allocative and technical efficiency provide a rationale for government actions that aim to transpose market dynamics into the public sector. Chapter One discussed how this analogy integrates the theoretical framework of neoclassical economics, thereby determining the political space available for policymaking. This framework defines the concept of economic efficiency, which dictates what social welfare entails. Determining what welfare encompasses is crucial for shaping public policies as governments seek to enhance welfare. In this regard, benchmarking neoclassical efficiency holds many implications, as outlined in Chapter One. It enforces a way of thinking about social issues, restricts the circumstances and tools permitted in policymaking, establishes methodologies rooted in market principles, and ultimately impacts the budgeting process in contemporary societies. This occurs despite the State and private firms having remarkably different objectives. As Wildavsky (1966) articulated:

Public works projects have a multitude of objectives and consequences. Projects may generate economic growth, alleviate poverty among some people, provide aesthetic enjoyment and opportunities for recreation, improve public health, reduce the risks of natural disaster, alter travel patterns, affect church attendance, change educational opportunities, and more. No single welfare criterion can encompass these diverse objectives (p. 294).

Indeed, restricting governmental actions to satisfy a single welfare criterion is illogical, as no principle can encompass the intricacies of various objectives in policymaking. This is especially true for a principle that does not directly consider the complexity of human wellbeing but instead defines welfare based on a particular and implicit judgement regarding the distribution of resources within society.

However, justifications for benchmarking efficiency are supported by some wellaccepted elements among economists. First, the dominant theory underpins efficiency as a guiding principle. Second, by resorting to analytical models that can be validated and that present binary results, benchmarking efficiency engages with the popular branch of positivism in the philosophy of science, which supposedly provides true answers, making outcomes indisputable and sidelining them in public debates. Lastly, efficiency is often associated with rationality, which is particularly appealing.

Rationality has various interpretations within this debate. Firstly, it pertains to the perception of what constitutes valid information to support decision-making, specifically the understanding of the nature of scientific knowledge, as discussed in Chapter One. The purported objectivity and impartiality of efficiency estimations would substantiate the results and provide adequate evidence to assess the merits of public policies.

Political and social debates regarding impacts, priorities, and other benchmarks for policymaking become secondary considerations in this approach. However, separating politics from public management disconnects essential elements that justify the adoption or persistence of a policy. As Jannuzzi (2022) argues, evaluations must not only address objectives, targets,

and design but also discuss the underlying institutional framework that supports the policy, the main actors involved, and the most relevant ideas and ideals that legitimise the intervention. The author adds that contextualising the policy in question helps to clarify values, interests, and disputes rather than implicitly covering them with a seemingly technical argument. This is particularly significant in linking the evaluation with other principles that should guide the analysis, such as relevance, coherence, effectiveness, impact, and sustainability. Therefore, the disconnection between politics and evaluation limits collective decisions based on a predetermined and supposedly purely technical benchmark.

The second meaning that integrates the concept of rationality and benchmarking efficiency is the general idea of avoiding waste. Allegedly, electing efficiency as a guiding principle for policymaking signifies acting with due diligence concerning the public budget. However, prioritising efficiency in policymaking effectively entails institutionalising the symmetry between the State and private firms in practical terms.

In the case of allocative efficiency, as discussed in Chapter Two, the CBA methodology resorts to the appealing yet generic concept of weighing benefits and costs to determine the merits of a policy. However, similar to any market transaction where goods are traded based on the WTP of consumers, the adoption of public projects or policies will depend on the WTP to implement the change. Also, estimating the costs and benefits involves a valuation process that assigns monetary values to every social impact arising from the policy under analysis.

For example, instead of debating the harms of pollution and legislating to prohibit it, pollution will come with a cost for those who can afford it. However, beyond the controversial idea that any policy effect can be measured in monetary terms, estimation techniques are often inaccurate, as discussed in Chapter Two. Thus, it is unreasonable for these estimations to serve as a foundation for conclusions about the merits of a policy or project, particularly regarding their financial viability, especially since they yield binary outcomes.

Furthermore, CBA raises significant intergenerational concerns owing to its sensitivity to long time horizons, specifically regarding impacts in the distant future. This debate is particularly pertinent for policies and projects that affect the environment, as relevant negative effects in the future will have minimal influence on current decisions. Lastly, completely ignoring distributive impacts when deciding on public actions is misleading, especially in highly unequal societies. Of course, policymakers may opt for a policy or project with negative distributive impacts because policymaking has multiple objectives. However, rendering the
distributive element irrelevant to the analysis is inexcusable, especially when the methodology carries an implicit distributional judgment of its own.

Concerning technical efficiency, the concept of rationality related to minimising waste becomes clearer. The rationale for value for money equates factories and schools by determining the minimal resources required to achieve equivalent results. Naturally, when one considers saving physical resources or ensuring an adequate number of employees, the notion of technical efficiency is logical. However, the analysis is not entirely straightforward. Savings can stem from reducing wages, pressuring for longer working hours, and so forth.

Moreover, interpreting results is not as clear-cut as often suggested. The most common approach concludes that any inefficiency represents an opportunity to cut costs. However, a programme may perform poorly not because it wastes resources, but due to insufficient funding. This explanation may not arise from the methods used, which is why every methodology can only act as an ancillary to the decision-making process and cannot replace social and political debates about the causes, priorities, and effects of policymaking.

Finally, the most worrisome connection between rationality and benchmarking efficiency permeates debates about budgeting. The association vests the expression 'quality of public spending'. Thus, the quality of expenditure derives from the results of methodologies grounded on market principles, which makes the State and private firms alike, even though their goals are entirely different.

In this sense, how can policymakers justify the merits of preserving cultural heritage or promoting gender and race equality on efficiency grounds? How can analysts measure the monetary value of preserving historical assets? Is it reasonable to argue that the justification for promoting gender or race equality stems from productivity in the labour market rather than from humanitarian reasons? Does financing these policies render the Budget a poor-quality spending?

The answers lie outside neoclassical economics. In fact, when discussing the 'classical revival' of economics for policy analysis, Garegnani (2007) makes two relevant observations. The first concerns content, highlighting three interconnected features that directly oppose the neoclassical approach and have enormous impacts on policy analysis: a) the distributive rather than allocative role of prices; b) the rule (not exception) of labor unemployment and capacity

underutilization; c) the relevance of aggregate demand. The distinct perspective of these three elements would render the State a remarkably different political space for policymaking.

The second observation concerns methods. Garegnani (2007) argues that the notable differences between classical and neoclassical approaches also affect policy analysis. In the authors' words, "*it can be said that there is bound to be less space for 'routine predictions' about the effects of policy, whether from theoretical analysis or econometric models*" (p. 236). Policy decisions, he argues, will inevitably favor some groups and damage others. "*Even a policy of full labour employment, which might seem to be in the obvious interest of the whole community, meets and has in fact met obstacles*" (p. 236).

This acknowledgment aligns with the discussion made throughout the three chapters, which concludes that there is no space for a single benchmark or a pre-defined welfare criterion. The merits of a policy must be collectively determined based on context, priorities, policy goals, and social preferences. Moreover, this acknowledgment also aligns with a post-positivist approach to science, wherein negotiation plays a crucial role in policy analysis, as discussed in Chapter One.

In this direction, Chapter Three has presented two examples of alternative methodologies that do not commit to efficiency or any specific single benchmark. Impact Analysis and Risk-Opportunity Analysis are examples of policy analysis that connect directly to the defined social objectives, thereby assuming different benchmarks according to policy goals.

The alternative approaches indicate that quantifying policymaking effects is a relevant aspect of the evaluation process. However, applying any methodology requires critical thinking and recognition that estimations can only provide intermediate results, serving as a necessary but insufficient tool for the decision-making process. Evaluating policy performance demands caution because history, context, and multiple objectives are essential elements of any analysis. Additionally, it is crucial to recognise that not every effect can be quantified, which makes qualitative analysis, social debates, and collective decisions equally important components of the decision-making process.

More importantly, recognising a different perspective on economic theory enables various welfare criteria, creating opportunities for other scientific fields to contribute to discussions about what enhances social welfare. Consequently, it alters the political space opened for policymaking, clearly indicating that governments do not have to operate like private firms.

REFERENCES

Acemoglu, D., Chernozhukov, V., Werning, I., Whinston, M. D. A multi-risk SIR model with optimally targeted lockdown. National Bureau of Economic Research, Working Paper 27102, 2020. Available at: < <u>https://www.nber.org/papers/w27102</u>>. Accessed on: Feb. 22, 2023.

Ackerman, F.; Heinzerling, L. **Pricing the Priceless: Cost-Benefit Analysis of Environmental Protection**. University of Pennsylvania Law Review, Rev. 150(5), 1553, 2002.

ADB – Asian Development Bank.Cost-benefit Analysis for Development: A PracticalGuide.Filipinas:ADB,2013.Availableat:<https://www.adb.org/sites/default/files/institutional-document/33788/files/cost-benefit</td>analysis-development.pdf>.Accessed on: Jan. 7, 2024.

Agafonow, A. Los límites de la eficiencia económica en una sociedad democrática. Revista de economía institucional, 9(16), 89-119, 2007.

Atkinson, A. B.; Stiglitz, J. E. Lectures on public economics: Updated edition. Princeton University Press, 2015.

Bamberger, J. M.; Clark, M.; Sartorius, R. Monitoring and Evaluation: Some Tools, Methods, and Approaches: Monitoring and evaluation: some tools, methods, and approaches (English). Evaluation Capacity Development working paper series, ECD Washington, D.C.: World Bank Group, 2004. Available at: < <u>http://documents.worldbank.org/curated/en/829171468180901329/Monitoring-and-evaluation-some-tools-methods-and-approaches</u> > Accessed on: Jan. 20, 2024.

Barlas Y, Carpenter S. **Philosophical roots of model validation: two paradigms**. System Dynamics Review 6(2): 148–166, 1990.

Becerra-Pérez, L. A.; Ramos-Alvarez, R. A.; DelaCruz, J. J.; García-Páez, B. Value per Statistical Life at the Sub-National Level as a Tool for Assessing Public Health and Environmental Problems. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *61*, 00469580241246476, 2024.

Berman, E. P. Thinking like an Economist: How Efficiency Replaced Equality in U.S. Public Policy. Princeton University Press, 2022.

Bernstein, R. J. **Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis**. Philadelphia: University of Pennsylvania Press, 1984.

Boardman, A. E.; Greenberg, D. H.; Vining A. R.; Weimer, D. L. Cost–Benefit Analysis: Concepts and Practice. 5th Edition. Cambridge Univ. Press, 2018.

Bogetoft, P.; Otto, L. Benchmarking with DEA, SFA and R. Springer, New York, 2011.

Boueri, R.; Rocha, F.; Rodopoulos, F. Avaliação da qualidade do gasto público e mensuração da eficiência. Brasília: Secretaria do Tesouro Nacional, 2015.

Bova, E.; Ercoli, R.; Bosch, X. V. **Spending reviews: some insights from practitioners.** Discussion Paper N. 135. Directorate General of Economic and Financial Affairs. European Commission, 2020.

Branch, K.; Khalkhal, I. Critical Analysis of Kaldor-Hicks efficiency criterion, with respect to moral values, social policy making and incoherence. Advances in Environmental Biology, 6(7), 2032-2038, 2012.

BRASIL. Casa Civil da Presidência da República. **Avaliação de políticas públicas: guia prático de análise ex ante**. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea), 2018a.

BRASIL. Casa Civil da Presidência da República. Avaliação de políticas públicas: guia prático de análise ex post. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea), 2018b.

Caldwell, B. **Positivist Philosophy of Science and the Methodology of Economics.** Journal of Economic Issues, 14(1), 53–76, 1980.

Campbell, H. F.; Brown, R. P. **Benefit-cost Analysis: financial and economic appraisal using spreadsheets**. Third Edition. Cambridge University Press, 2023.

Campos, J.; Serebrisky, T.; Suárez-Alemán, A. **Time goes by: recent developments on the theory and practice of the discount rate**. Note IDB-TN-861 Inter-Am. Dev. Bank Washington, DC, 2015. Available at: < <u>https://publications.iadb.org/publications/english/document/Time-Goes-By-Recent-Developments-on-the-Theory-and-Practice-of-the-Discount-Rate.pdf</u> > Accessed on: Oct. 15, 2024.

Catalano, G.; Erbacci, A. A theoretical framework for spending review policies at a time of widespread recession. OECD Journal on Budgeting 17(2), 9–24, 2018.

Chang, A. Y. et al. The equity impact vaccines may have on averting deaths and medical impoverishment in developing countries. Health Affairs, 37(2), 316-324, 2018.

Clark, A. E.; Frijters, P.; Shields. M. A. **Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles**. Journal of Economic Literature, 46 (1): 95-144, 2008.

Cole, D. H. The Stern Review and its critics: Implications for the theory and practice of benefit-cost analysis. Nat. Resources J., 48, 53, 2008.

Cook, E. Efficiently unequal: The global rise of Kaldor-Hicks neoliberalism. Global Intellectual History, 9(1-2), 247-269, 2024.

Cooter, R.; Rappoport, P. Were the ordinalists wrong about welfare economics? Journal of Economic Literature, 22(2), 507-530, 1984.

Day, B.; Bateman, I.; Lake, I. **Beyond implicit prices: recovering theoretically consistent and transferable values for noise avoidance from a hedonic property price model.** Environmental and Resource Economics, 37(1), 211–232, 2007.

Dobb, M. Theories of Value and Distribution since Adam Smith: Ideology and Economic Theory. Cambridge: Cambridge University Press, 1975.

Drakopoulos, S. A. The Historical Perspective of the Problem of Interpersonal Comparisons of Utility. Journal of Economic Studies, 16(4), 1989.

Drakopoulos, S. A. Value judgements, positivism and utility comparisons in economics. Journal of Business Ethics, 189(3), 423-437, 2024.

Drei, S. M.; Angulo-Meza, L. Data Envelopment Analysis (DEA) in health: a literary review of Latin America. Multidisciplinary Reviews, 6(2), 2023011-2023011., 2023.

Doherty, L.; Sayegh, A. **How to design and institutionalize spending reviews.** International Monetary Fund, 2022. Available at: < <u>https://www.imf.org/en/Publications/Fiscal-Affairs-Department-How-To-Notes/Issues/2022/09/20/How-to-Design-and-Institutionalize-Spending-Reviews-523364</u> > Accessed on: Apr. 17, 2024.

Easterlin, R. A. **Explaining happiness**. Proceedings of the National Academy of Sciences, 100(19), 11176-11183, 2003.

Fallov, J. A.; Georgieva-Andonovska, E. G. **Bulgaria - Spending review manual.** Washington, D.C.: World Bank Group, 2018. Available at: < <u>http://documents.worldbank.org/curated/en/940921532366222268/Bulgaria-spending-review-manual</u>> Accessed on: Apr. 17, 2024.

Ferreira, E. C. **Um ajuste justo ou mais alguns passos atrás para a educação básica pública no Brasil?** Educação e Pesquisa, [S. l.], v. 46, p. 1–18, 2020. Available at: < <u>https://www.revistas.usp.br/ep/article/view/187149</u>> Accessed on: Jun. 12, 2024.

Flyvbjerg, B. B.; Dirk W. How (In)Accurate Is Cost-Benefit Analysis? Data, Explanations, and Suggestions for Reform. In José A. Gómez-Ibañez and Zhi Liu, eds., Infrastructure Economics and Policy: International Perspectives. Cambridge, MA: Lincoln Institute of Land Policy, pp. 174-196, 2022.

Fontaine, P. Review of "Thinking like an Economist: How Efficiency Replaced Equality in U.S. Public Policy" by Elizabeth Popp Berman. Journal of the History of Economic Thought (forthcoming). Preprint at SocArXiv, osf.io/preprints/socarxiv, 2023.

Frank, R. H. Why is Cost-Benefit Analysis so Controversial? The Journal of Legal Studies, 29(S2), 913–930, 2000.

Frank, R. H. Microeconomics and behavior. 10th Edition. New York: McGraw-Hill, 2020.

Fried, H.; Lovell, C.; Schmidt, S. The measurement of productive efficiency and productivity growth. New York: Oxford University Press, 2008.

Garegnani, P. **Professor Foley and classical policy analysis**. Review of Political Economy, 19(2), 221-242, 2007.

Gertler, P. J. et. al. **Impact evaluation in practice.** World Bank Publications, 2016. Available at: <<u>https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice</u>> Accessed on: Dec. 10, 2024.

Gowdy, J. Contemporary welfare economics and ecological economics valuation and policy. International Society for Economics Internet Encyclopaedia of Economical Economics, 2003.

Graham, C. **Insights on development from the economics of happiness.** The World Bank Research Observer, 20(2), 201-231, 2005.

Greenstone, M.; Nigam, V. **Does social distancing matter?** Covid Economics. Mar 27;7(1):23, 2020.

Groom, B. et. al. **The future, now: A review of social discounting.** Annual Review of Resource Economics, 14, 467-491, 2022.

Grubb, M. et al. **The New Economics of Innovation and Transition: Evaluating Opportunities and Risks**. EEIST Consortium Report, 2021. Available at: < https://eeist.co.uk/eeist-reports/the-new-economics-of-innovation-and-transition-evaluating-opportunities-and-risks/ > Accessed on: Oct. 23, 2023.

Helliwell, J. F. et. al. **World Happiness Report 2024**. University of Oxford: Well-being Research Centre, 2024. Available at: < <u>https://worldhappiness.report/ed/2024/</u>> Accessed on: Oct. 18, 2024.

Heinzerling, L.; Ackerman, F. **Wasting Away in Paretoville: A Reply to Cass Sunstein.** Harvard Law & Policy Review, Vol. 1, Georgetown Law and Economics Research Paper No. 966457, Georgetown Public Law Research Paper No. 966457, 2007. Available at: < <u>https://ssrn.com/abstract=966457</u>> Accessed on: Sep. 18, 2024.

Hollingsworth, B. **The measurement of efficiency and productivity of health care delivery.** Health Economics, 17(10), 1107–1128, 2008.

Howlett, M.; Ramesh, M; Perl A. Política Pública: seus ciclos e subsistemas: uma abordagem integral. Rio de Janeiro: Elsevier, 2013.

Jannuzzi, P. de M. A importância do contexto institucional, político e ideacional na avaliação de políticas públicas. Revista Brasileira de Avaliação, 11(2), e113722, 2022.

Jit, M.; Mibei, W. Discounting in the evaluation of the cost-effectiveness of a vaccination programme: a critical review. Vaccine, 33(32), 3788-3794, 2015.

Kaldor, N. Welfare propositions of economics and interpersonal comparisons of utility. Economic Journal, 49: 549–52, 1939.

Khandker, S. R.; Koolwal, G. B.; Samad, H. A. **Handbook on impact evaluation: quantitative methods and practices.** World Bank Publications, 2009.

Klasen, S. The Efficiency of Equity. Review of Political Economy, 20:2, 257-274, 2008.

Mankiw, N. G. **Principles of microeconomics**. 10th ed. CENGAGE Learning Custom Publishing, 2024.

Kleindorfer, G. B.; O'Neill, L.; Ganeshan, R. Validation in Simulation: Various Positions in the Philosophy of Science. Management Science 44(8):1087-1099, 1998.

Kohl S.; Schoenfelder J.; Fügener A., Brunner J.O. The use of Data Envelopment Analysis (DEA) in healthcare with a focus on hospitals. Health Care Manag Sci. Jun;22(2):245-286, 2019.

Hicks, J. R. The valuation of social income. Economica, 7: 105–24, 1940.

Lampe, H. W.; Hilgers, D. **Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA.** European journal of operational research, 240(1), 1-21, 2015.

Liu, John S. et. al. A survey of DEA applications. Omega, Elsevier, vol. 41(5), pages 893-902, 2013.

Mardani, A. et. al. A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency. Renewable and Sustainable Energy Reviews, 70, 1298–1322, 2017.

Martins Drei, S.; Angulo-Meza, L. **Data Envelopment Analysis (DEA) in health: a literary review of Latin America.** Multidisciplinary Reviews, 6(2), 2023011, 2023.

Mercure, J. F. et. al. **Risk-opportunity analysis for transformative policy design and appraisal.** Global Environmental Change, 70, 102359, 2021.

Merry, S. E. Measuring the world: Indicators, human rights, and global governance. Current anthropology, 52(S3), S83-S95, 2011.

Minken, H. **The Pareto criterion and the Kaldor Hicks criterion.** International Encyclopedia of Transportation; Elsevier: Amsterdam, The Netherlands, 1, 7, 2021.

Mishan, E.J.; Quah, E. Cost-Benefit Analysis. 6th ed. Routledge, 2020.

Murillo-Zamorano, L. R. Economic Efficiency and Frontier Techniques. Journal of Economic Surveys, Vol. 18, No. 1, pp. 33-77, 2004.

Musgrave R. A. A multiple theory of budget determination. Finanzarchiv 17(3):333–343, 1957.

Nordhaus, W. D. A Review of the Stern Review on the Economics of Climate Change. Journal of Economic Literature, 45 (3): 686-702, 2007.

Ng, Y. K. Welfare economics: towards a more complete analysis. Springer, 2003.

Ocké-Reis, C. O. O. et. al. **SUS: avaliação da eficiência do gasto público em saúde**. 2023. Available at: < <u>https://repositorio.ipea.gov.br/handle/11058/12029</u>> Accessed on: Mar. 6, 2024.

Odon, T. I. Risco, valor de uma vida estatística e análise de custo-benefício: novos caminhos para comissões parlamentares de inquérito. Brasília: Núcleo de Estudos e Pesquisas/CONLEG/Senado, Fevereiro 2024 (Texto para Discussão n ° 326). 2024. Available at: < <u>www.senado.leg.br/estudos</u> >. Accessed on: Dec. 6, 2024.

OECD. **Equity and Cost-benefit Analysis**, in Cost-Benefit Analysis and the Environment: Recent Developments, OECD Publishing, Paris, 2006. Available at: < <u>https://doi.org/10.1787/9789264010055-16-en</u>> Accessed on: Dec. 6, 2024.

OECD. Cost-Benefit Analysis and the Environment: Further Developments and Policy Use, OECD Publishing, Paris, 2018. Available at: < <u>https://doi.org/10.1787/9789264085169-en</u>> Accessed on: Dec. 6, 2024.

OECD. **Applying Evaluation Criteria Thoughtfully**, OECD Publishing, Paris, 2021. Available at: < <u>https://doi.org/10.1787/543e84ed-en</u>> Accessed on: Dec. 12, 2023.

Ozawa, S. et. al. During the 'decade of vaccines,' the lives of 6.4 million children valued at **\$231 billion could be saved.** Health affairs, 30(6), 1010-1020, 2011.

Park, Y.; Konge, L.; Artino, A. R. **The Positivism Paradigm of Research.** Academic medicine: Journal of the Association of American Medical Colleges, 95 (5), 2020.

Petri, F. Microeconomics for the Critical Mind: Mainstream and Heterodox Analyses. Switzerland: Springer, 2021.

Pigou, A. C. The Economics of Welfare. London: Macmillan. 1920.

Ramacciotti, B. L.; Bernardino, M. C. da C. Vertentes epistemológicas das políticas públicas: entre as análises descritiva e prescritiva. Cadernos de Pesquisa, São Paulo, v. 50, n. 175, p. 56-76, jan./mar. 2020.

Rose-Ackerman, S. Putting Cost-benefit Analysis In Its Place: Rethinking Regulatory Review, 65 U. Miami L. Rev. 335, 2011. Available at: https://repository.law.miami.edu/umlr/vol65/iss2/3 Accessed on: Dec. 12, 2023.

Robinson, L. A.; Sullivan, R.; Shogren, J. F. Do the Benefits of COVID-19 Policies Exceed the Costs? Exploring Uncertainties in the Age–VSL Relationship. Risk Analysis, 2020. Available at: < https://pmc.ncbi.nlm.nih.gov/articles/PMC7405126/ > Accessed on: Sep. 12, 2024.

Rowell, A. **Ethical Choices Inside Regulatory Cost-Benefit Analysis.** Georgetown Journal of Law & Public Policy, Vol. 19, University of Illinois College of Law Legal Studies Research Paper No. 22-08, 2021. Available at: https://srn.com/abstract=3944299 Accessed on: Sep. 12, 2024.

Santos, M. A. B. D. Eficiência e ineficiência nos sistemas de saúde: a perspectiva internacional do debate. In: SUS: avaliação da eficiência do gasto público em saúde. Org. Ocké-Reis, C. O. O. 2023.

Schick, A. **The metamorphoses of performance budgeting.** OECD Journal on Budgeting, vol. 13/2, 2014. Available at: <<u>https://doi.org/10.1787/budget-13-5jz2jw9szgs8</u>> Accessed on: Sep. 12, 2024.

Sen, A. The discipline of cost-benefit Analysis. Journal of Legal Studies, 29 (2 Part 2), 931-952, 2000.

Shao, S.; Tian, Z.; Fan, M. Do the rich have stronger willingness to pay for environmental protection? New evidence from a survey in China. World Development, 105, 83-94, 2018.

Shore, C.; Wright, S. Governing by numbers: audit culture, rankings and the new world order. Social Anthropology, 23(1), 22–28, 2015.

Souza, p. C. D.; Scatena, j. H. G.; Kehrig, R. T. Aplicação da análise envoltória de dados para avaliar a eficiência de hospitais do sus em mato grosso. Physis: Revista de Saúde Coletiva, v. 26, p. 289-308, 2016.

Stiglitz, J. E. The invisible hand and modern welfare economics. 1991.

Stiglitz, J. E.; Rosengard, J. K. Economics of the public sector: Fourth international student edition. WW Norton & Company. 2015.

Sunstein, C. R. Willingness to pay vs. welfare. Harv. L. & Pol'y Rev., 1, 303, 2007.

Toma, E. et. al. **DEA Applicability in Assessment of Agriculture Efficiency on Areas with Similar Geographically Patterns.** Agriculture and Agricultural Science Procedia, 6, 704–711, 2015.

Thunström, L. et. al. **The benefits and costs of flattening the curve for COVID-19.** Journal of Benefit-Cost Analysis, FirstView. 2020.

UNITED KINGDOM. H.M. Treasury. The Green Book: Central Government Guidance on Appraisal and Evaluation. 2022. Available at: < <u>https://assets.publishing.service.gov.uk/media/623d99f5e90e075f14254676/Green_Book_202</u> 2.pdf>Accessed on: Sep. 7, 2024.

Vedung, E. Four Waves of Evaluation Diffusion. Evaluation, 16(3), 263–277, 2010.

Viscusi, W.K.; Aldy, J.E. The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World. Journal of Risk and Uncertainty 27, 5–76, 2003.

Viscusi, W. K.; Masterman, C. J. **Income Elasticities and Global Values of a Statistical Life.** Journal of Benefit-Cost Analysis, 8(02), 226–250, 2017.

Weimer, D.; Vining, A. Policy analysis: Concepts and practice. Routledge. 2017.

Westra, T. A. et. al. On discounting of health gains from human papillomavirus vaccination: effects of different approaches. Value in Health, 15(3), 562-567, 2012.

Wildavsky, A. The political economy of efficiency: cost-benefit analysis, systems analysis, and program budgeting. Public Administration Review, 292-310, 1966.

World Bank. Cost-benefit Analysis in World Bank projects (English). IEG Fast Track Brief Washington, D.C.: World Bank Group. 2010. Available at: <<u>http://documents.worldbank.org/curated/en/216151468335940108/Cost-benefit-analysis-in-World-Bank-projects</u>> Accessed on: Sep. 7, 2024.

World Bank. A fair adjustment: efficiency and equity of public spending in Brazil: Volume 1 - Overview (English). Washington, D.C: World Bank Group. 2017. Available at: http://documents.worldbank.org/curated/en/641520429223428/Volume-1-Overview Accessed on: Sep. 10, 2023.