Chapter 2:

Counterfactual Thinking and What-If Analysis

2.1 What Are Counterfactuals

2.1.1 Definition and Intuition

Counterfactuals refer to statements about what would have happened under alternative circumstances that did not actually occur. In causal analysis, they answer questions of the form: "What would the outcome have been if a different action, treatment, or policy had been chosen?" This idea is fundamentally different from prediction, which only concerns what is likely to happen given observed patterns. Instead, counterfactual reasoning compares the observed reality with a hypothetical alternative reality in which some condition is changed.

Formally, counterfactuals describe outcomes under unrealized scenarios, outcomes that are not observed but are conceptually well-defined. To illustrate, in the case of a person who was given a policy intervention, the counterfactual outcome is a situation of what would have occurred to the same person had the intervention not been implemented. It is then given a causal effect that is with a comparison made of these two possible states of the world. This mode of thinking is at the core of the contemporary causal inference in which the causal effects are determined not by the association of data, but rather by comparison of actual and hypothetical outcome.

Recent research stresses that counterfactuals can be used to offer an intuitive connection between the human cognition and the formal causal study. As Wang et al. (2024) [22] emphasize, counterfactuals are inherently stated in terms of what-if options that can assist analysts and decision-makers to reason the cause-effect relationships, as opposed to their correlations. In line with this, Imbens (2024) [23] emphasizes that causal questions are counterfactual in the social sciences, as such always involve the comparison of what actually occurred to what it would have occurred had a different intervention or choice been made. Causal claims cannot be assigned any specific meaning without this hypothetical comparison.

Accordingly, the intuition of counterfactuals is quite straightforward and persuasive, causality concerns contrast between realities, the world as it is and the world as it would be with a different action or policy. This difference is what separates the causal analysis and the mere prediction or descriptive statistical analysis.

2.1.2 Historical Background in Economics

Counterfactual reasoning has a long history in economics, especially program evaluation, policy analysis, and welfare economics. Even when only one policy is actually implemented, thought experiments have been important in the assessment of the implications of alternative policies like tax changes, subsidy changes, or regulation changes, by economists. In this respect the use of counterfactuals in economic reasoning has always been implicit: an evaluation of a policy must involve asking how things would have turned out under an alternative policy choice.

According to Heckman and Pinto (2024) [21], thought experiments are at the core of econometric causality and therefore, they are essentially counterfactual. Their work demonstrates that the fundamental activity of causal econometrics is to build decent comparisons between the actual world and well-defined hypothetical alternatives. In this view, counterfactuals are not of recent origin, instead, it is a formalization of an old tradition in the field of economics of thinking about alternative phenomena to evaluate the effect of a policy and social outcomes.

This tradition has been formalized in the contemporary empirical economics in the form of causal inference structures that render counterfactuals explicit and central. According to Imbens (2024) [23], the contemporary causal analysis in the social sciences defines the causal effects as the possibilities of the outcomes, which are explicitly counterfactual objects: in regard to each unit or individual, the outcomes are conceptualized under various possible interventions, although it is only possible to observe one of them in reality. Such formalization is a significant move to make economic reasoning about causality more specific, transparent and testable.

The general history of counterfactual thinking in economics entails a change in the informal thought experiments of policy to formalized statistical and econometric models that can put counterfactuals at the center of the causal analysis. Counterfactuals are now known as essential instruments of responding to causal questions, assessing policies, and explaining economic processes since without them, the concept of a causal impact would be definitionally unclear.

2.2 Counterfactual Reasoning Framework

The counterfactual reasoning system offers a formal model of the causal relations. Essentially, it poses a question that is very basic yet profound at the same time; What would have happened had I done or treated it differently? To provide an answer to this, one has to compare the actual results with the hypothetical results in different circumstances and this is what constitutes causal inference in economics and social sciences.

2.2.1 Potential Outcomes Model

The most important concept in this type of reasoning is the Potential Outcomes Model. It theorizes that every person or unit in research may have various results in case they are provided with a treatment or not. Although only one of the outcomes is witnessed in real life, the other outcome is a counterfactual. This model offers a strict methodology of definitional causal effects and causal questions are clearly identified as being related to hypothetical alternatives as opposed to mere associations.

Recent studies state that this framework is flexible. To illustrate, Han et al. (2024) [24] demonstrate that the causal effects can be estimated using potential outcomes even in a situation where experimental and observational data are not entirely overlapping. On the same note, Sert et al. (2025) [27] show a Bayesian semiparametric model whereby the treatment effects can be robustly estimated in complex data structures. Combined with these studies indicates the value of the potential outcomes model in the basis of contemporary causal inference as it offers a rational method of formalizing otherwise possibly intuitive what-if reasoning.

2.2.2 Treatment and Control Groups

In practice, the estimation of causal effects is frequently done in terms of comparing control and treatment groups. In experiments, the participants would be randomly assigned to the treatment and the control group and this assists in isolating the causal effect of the treatment. In observational studies though, the assignment of the treatment is not regulated, making it prone to biases. The modern techniques, like doubly robust estimation presented by Du et al. (2024) [25], correct such differences by considering the observed covariates, which result in more causal inferences.

Interestingly, there are instances when there will be a lack of a classic control group. In a study focusing on how to assess the effectiveness of policies that do not include a conventional control group, Cerqua et al. (2024) [26] employ well-crafted statistical models to create plausible counterfactual comparison. This shows how the counterfactual reasoning framework can be applied to the real-life problems, and make the causal analysis possible even in non-standard or complicated study design.

2.2.3 Average Treatment Effect (ATE)

One of the concepts of the potential outcomes model is the Average Treatment Effect (ATE). It is the mean effect of treatment in a population, which summarizes the causal effect in a manner that is understandable as well as policy-relevant. The ATE is estimated by comparing the outcomes of the treated and the control groups and must be done with a serious consideration of the assumptions, including but not limited to, the absence of unmeasured confounders and enough overlap of groups.

Han et al. (2024) [24] give the approaches to estimate ATE in the situation when observational and experimental data are different, whereas Du et al. (2024) [25] [27] Pin A robust estimation strategies that can be used to combine both model-based and design-based methods. Such developments demonstrate that the counterfactual reasoning framework is not only conceptually clarifying but also informative of practical methods of the estimation of causal effects in various situations.

In general, the counterfactual reasoning model transforms the intuitive thinking processes of what-if into a systematic and strict method of causal inference. Through this, researchers can strictly address the causal questions and offer evidence-based information that can be used in the evaluation and decision-making of a policy by clearly defining the possible outcomes, distinguishing between the treatment and control groups, and calculating population-level outcomes, including the ATE.

2.3 What-If Analysis in Economics

What-if analysis is a useful economic instrument, which implements the counterfactual reasoning framework to practical decision-making. Although the earlier parts of this paper have addressed the conceptual and formal basis of counterfactuals, this section will show how economists and policymakers apply the ideas in examining other scenarios, weighing policy choices, and making evidence-based decisions.

2.3.1 Policy Evaluation Examples

Counterfactual analysis enables economists to approximate the impact of those policies that have already been implemented or simulate the impact of those policies that have not been experimented. As an example, in the paper by Chen, Phillips and Shi (2025) [28], the authors research into housing market interventions in New Zealand by using what-if scenarios to determine the effect of changes in interest rates on property bubbles. They can measure the causal impacts of policy interventions by creating a counterfactual situation as to what would otherwise have happened without the implementation of the policy, which can be used to give evidence of the effectiveness of alternative strategies. Equally, in macroeconomics, structural counterfactual analysis is employed by Wang (2024) [29] to analyze the manner in which alternative policies, including alternative fiscal or monetary intervention policies, would have impacted macroeconomic outcomes. The examples of these studies demonstrate that counterfactual thinking helps fill the gap between theory and practice through enabling policymakers to conduct systematic and data-driven analysis of the notion of what-if.

2.3.2 Tax Reforms

Another important area where what-if analysis is important is in tax policies. Babilla (2023) [30] discusses the possible outcomes of tax change and universal basic income in a currency union and assesses how each situation can impact on growth in the long run, income disparity, and welfare. With the study, by modeling alternative tax structures and redistribution policies of incomes, the study can give an insight into which kind of policy choices will yield desired results. On the same note, Peprah, Ocansey, and Asirifi (2025) [31] discuss the tax reforms in the developing economies and predict the effect of various reforms on economic growth, revenue collection, and income inequality using a counterfactual analysis. These studies explain the practical significance of counterfactual situations in determining the extension of the implications of fiscal policies before implementation.

2.3.3 Subsidies and Interest Rate Changes

Although the tax reforms have a direct impact on the household income and the incentives to firms, the other policy tools that are commonly assessed with the help of what-if analysis are subsidies and interest rates. Through simulating other allocation strategies, it is possible to evaluate subsidy programs, e.g., agricultural subsidies or social welfare subsidies, to learn about their probable impact on the production level, the consumption level, and the equity level. On the same note, interest rate policy is also analyzed on a regular basis using counterfactual models to determine the effect on investment, consumption and macroeconomic stability. Indicatively, the research introducing interest rate changes to understand the possible developments in the housing market, such as Chen et al. (2025) [28], offers the policymaker with quantitative data on how effective monetary interventions are likely to be.

In general, what-if analysis illustrates the translation of counterfactual reasoning into the practical economic analysis. Economists can attempt to estimate the effect of interventions they may have by simulating different policy environment scenarios, that is, vary in taxes, subsidies, or interest rates, and inform the policy design and assist decision-makers in choosing those strategies that can maximize welfare or meet certain goals. This part makes it explicit that counterfactual thinking is not only a theoretically constructed concept but an indispensable instrument of evidence-based economic policy.

2.4 Statistical Representation of Counterfactuals

Counterfactual understanding involves having a clear statistical framework that makes explicit the meaning of what would have happened in different hypothetical conditions. This is given the Neyman Rubin Causal Model (NRCM), also referred to as the potential outcomes framework. The model assumes that, in this model, there is a number of possible outcomes of each unit (e.g., individual, firm, or region) with respect to a variety of treatment or intervention. The inherent problem is that observed outcomes of any unit are only one, and the other outcomes, the counterfactual outcomes are not observed. This gap is termed the basic issue of causal inference, and it is what prompts the necessity of the use of statistical assumptions to determine the causal impact (Imbens et al. (2025) [32]; Keller and Branson (2024) [35]).

Neyman-Rubin model describes the causal effects as variation between the possible outcomes. As an illustration, the difference between what one unit should have experienced the influence of a policy intervention would be formalized as the difference between the outcome with treatment and without treatment. Since the two outcomes cannot be measured at the same time using the same unit, researchers use identification assumptions to measure these effects using observed data. Key assumptions include:

- **Ignorability (or unconfoundedness):** It is assumed that the treatment assignment is not a deterministic factor of the possibly occurring outcomes, conditioned on observed covariates. This gives the researcher a chance to consider the observational information as random after taking those covariates into consideration (Keller and Branson (2024) [35]).
- **Overlap (or positivity):** This is the condition that every unit is likely to have a non-zero probability of obtaining any treatment, and makes comparisons between treated and untreated groups significant (Imbens et al. (2025) [32]).
- **Consistency (SUTVA – Stable Unit Treatment Value Assumption):** The observed outcome for a unit under a given treatment corresponds exactly to the potential outcome defined for that treatment, and no interference occurs between units (Imbens et al. (2025) [32]; Keller & Branson (2024) [35]).

These are the assumptions that are vital in the translation of the NRCM into practice. In case they are held, the observed data can be interpreted and concluded on the causal effects e.g. the average treatment effect (ATE) at a population level. Some of the recent researches note the use of these principles in various areas: utilizing multiple sources of data to explore persistent confounding in longitudinal studies (Imbens et al. (2025) [32]) and modeling causal behavior using trajectory data in transportation (Wu et al. (2025) [33]) and the combination of causal inference and network theory to understand delay propagation in bus networks (Zhang et al. (2025) [34]). These assumptions play a vital role in research on education, whereby the soundness of evaluations of interventions is based on plausible statistical arguments (Keller and Branson (2024) [35]).

Average Treatment Effect (ATE):

$\tau\_ATE = E[Y_i(1) - Y_i(0)]$

Where:

- $Y_i(1)$ = potential outcome for unit i under treatment
- $Y_i(0)$ = potential outcome for unit i under control
- $E[\cdot]$ = expected value (average across the population)
- τ_ATE = average treatment effect

This single formula encapsulates the core concept of counterfactual reasoning in causal inference. It defines the causal effect as the difference between what would happen if everyone received treatment versus if no one received treatment. The fundamental problem is that we can never observe both $Y_i(1)$ and $Y_i(0)$ for the same unit simultaneously, which is why we need the three key assumptions to estimate this quantity from observed data.

The Neyman- Rubin framework is a formal statistical model of counterfactuals by providing an explicit statement of the assumptions and relating them to observable data. It fills the void between idealized knowledge of theoretical situations and practical forecasting in applied studies, and thus is a fundamental part of contemporary causal forecasting in economics, political policy, transport, and education.

2.5 Challenges in Counterfactual Analysis

The causal inference has some counterfactual reasoning as its basis but its practical application has some major challenges. The key challenges, such as the absence of data, selection bias, and unobserved confounders, are discussed in this section. These challenges are necessary to understand how to interpret causal estimates and how to create reliable conclusions in a study.

2.5.1 The Fundamental Missing Data Problem

The main problem in counterfactual analysis is that of every unit, we are only able to observe a single outcome, the outcome that would be obtained using the treatment. The result would not have been counterfactual with alternate treatments and this presents an underlying missing data issue (Benz et al. (2025) [37]).

This weakness explains why causal effects would not be measurable and have to be statistically deduced. Missing outcomes are dealt with through a number of different methods, such as:

- Imputation-based approaches, which approximate the unobserved possible outcomes with observed patterns of data.
- Randomization-based or design-assisted methods, that is, they utilize study design to minimize missing data bias (Heng et al. (2025) [38]).
- Covariate adjustment and modeling, in which covariates that have been observed are modelled to give counterfactual results, on the assumption that all included confounding variables are measured.

In spite of these measures, the problem of missing data creates some uncertainty in itself, highlighting that the derivation of causal conclusions is frequently assumption-based.

2.5.2 Selection Bias in Observational Studies

Selection bias is a situation in which the assignment to treatment is not random but it is related to unobserved variables that also influence outcomes. In education or healthcare research, an example of such a difference is a systematic difference between the students or patients who have been treated and those who have not been, not completely accounted by the observed variables.

Estimated causal effects can be misleading and confounded without taking into consideration the selection bias. Some of the methods of reducing selection bias are:

- Propensity score matching: forming similar groups with regard to observed covariates.
- Weighting methods: correcting the chances of the assignment to treatments.
- Instrumental variable approaches: employing external variables that affect the assignment of treatment, but have no effect in outcomes.

Heng et al. (2025) [38] point out that despite the advanced approaches, residual selection bias might still persist in the case of missing or poorly measured covariates of interest.

2.5.3 Unobserved Confounders

Unobserved confounders refer to the variables, which influence the treatment and the outcome but cannot be measured in the dataset. Their attendance goes against the principle assumption of ignorability, which is the essential element of causal inference without biases (Byrnes and Dee (2025) [36]).

The treatment effect estimates can be biased in the presence of unobserved confounding even when the potential outcomes framework is used. Researchers often use:

- Sensitivity analysis to approximate the stability of causal inferences to unobserved confounding factors.
- The methods include: advanced modeling methods in which an effort is made to integrate several data sources in order to decrease confounding bias.
- Despite the fact that random assignments are not approximated, the design-based methods, including randomized encouragement designs or quasi-experiments, can be used.

The absence of data, selection bias, and unmeasured confounders together show that application of causal inference is not practical in many cases and it takes some very strong assumptions to interpret the estimates in a way that should be believed.

2.5.4 Implications for Practice

These issues have a number of implications on researchers and policymakers:

- Transparency of assumptions: It is of great importance to make known the assumptions of any causal analysis, such as that of ignorability and that of overlap.
- Methodological rigor: Causal estimates can be made more trustworthy by relying on strong statistical methods and sensitivity analysis.
- Emphasis on data quality: To lessen the combined effects of missing data, selection bias, and unobserved confounders, it is important to measure covariates accurately and design studies carefully..

By confronting these limitations, researchers can produce more credible and interpretable causal findings, while also understanding the bounds of inference in applied settings (Benz et al. (2025) [37]; Byrnes & Dee (2025) [36]; Heng et al. (2025) [38]).

2.6 Case Study: Effect of Education on Income

One example of counterfactual reasoning involves the classic economics question, which is the role of education in income, and how do changes in both impact each other. The presented case study will implement the ideas and theories of the preceding sections to a real-life empirical example, where both the possibilities of outcomes framework and the actual difficulties of causal inference will be observed.

2.6.1 Defining the Treatment and Outcome

The level of education attained is the treatment, and the outcome, individual income is the outcome in this research. More specifically:

- Treatment (Education): This may involve; years in school, highest level of education or parental education where intergenerational effects are involved.
- Outcome (Income): Typically measured as annual earnings, lifetime income, or wage rate.

Lee, Roys, and Seshadri (2024) [39] examine the causal impact of parent education on the earnings of children, which can be implemented in such a way that treatment and outcome are formulated in counterfactual terms. In this case, the counterfactual question is:

What would have been the income of a child had his or her parents achieved another level of education?

This framing of the problem makes us clearly see that we are unable to observe both possible outcomes of an individual which is the fundamental missing data issue in counterfactual analysis.

2.6.2 Counterfactual Interpretation

The counterfactual here is the hypothetical income that a person would receive in a different scenario of education. This allows researchers to:

- Approximate the Average Treatment Effect (ATE) of education on income.
- Understand heterogeneity in treatment effects, e.g., whether effects differ by socioeconomic background or region.
- Look at indirect or inter-generational impact, e.g. the impact of parental education on the economic outcome of children (Lee et al., 2024) [39].

Notably, naive techniques, like comparing simple change scores of observational data, do not give causal estimates, warns Tennant et al. (2022) [40]. Any differences that are observed are likely to be biased without the appropriate consideration of confounding variables.

2.6.3 Sources of Bias

Causal estimates can be influenced by a number of sources of bias, even with a careful definition of treatment and outcome:

- Ability Bias: Those with greater innate ability can further their education and also earn higher, regardless of education.
- Family Background: Income of the families, educational level of parents, and social surroundings may confound between education and income.
- Selection Bias: Professionals who continue their education more might be different in a systematic way compared to those who do not.

To mitigate these biases, it is common to use statistical adjustment which could include regression with covariates or instrumental variables or quasi-experimental design (Lee et al., 2024) [39]. Counterfactual reasoning offers a conceptual approach to define the comparisons that should be made when coming to a causal conclusion.

2.6.4 Applying the Counterfactual Framework in Practice

In practice, potential outcomes framework is used in the following way:

- Define potential outcomes: Income at each possible level of education of each person.
- Identify treatment and control groups: For example, children whose parents had higher education vs. lower education.

- Estimate causal effects: With statistical techniques, controlling confounding, selection and missing counterfactuals.

Lee et al. (2024) [39] apply a similar framework to this study but based on observational data and that takes into account confounders to conclude on the impact of parental education on the earnings of children. As noted by Tennant et al. (2022) [40], unofficial observational comparisons may be deceptive unless conducted properly.

2.6.5 Key Insights from the Case Study

- Counterfactual reasoning enables researchers to create causal questions in a specific way even in cases where the two possible results cannot be observed.
- It is an end-to-end instance of the counterfactual framework application that the case study illustrates: the definition of treatment, conceptualizing counterfactuals, determining biases, and estimating causal effects.

2.7 Summary: Linking to Econometric Foundations

The chapter has addressed the concept of counterfactual reasoning as a tool to be used to study the cause and effect in the field of economics. It has already shifted away to conceptual definitions to formal structures, operationalizations, statistical models and problems -and ends in a case study of the impact of education on income.

2.7.1 Recap of Core Concepts

Counterfactuals were identified as the definition of what would have happened were (if) given the case, which enabled economists to formulate questions beyond correlations and predictions (Heckman and Pinto, 2024 [21]; Wang et al., 2024 [22]; Imbens, 2024 [23]). Researchers are in a position to explain causal processes by taking into consideration other possible results which will determine the impact of interventions like education, policy changes or economic interventions.

The counterfactual reasoning model operationalizes those concepts through potential outcomes, treatment and control groups and the primary causal estimand of the Average Treatment Effect (ATE) (Han et al., 2024 [24]; Du et al., 2024 [25]; Cerqua et al., 2024 [26]; Sert et al., 2025 [27]). Such instruments enable economists to pose specific causal questions and assess interventions, either through individual-level interventions or macroeconomic policies (Chen et al., 2025 [28]; Wang, 2024 [29]; Babilla, 2023 [30]; Peprah et al., 2025 [31]).

2.7.2 Statistical Foundations and Identification

The article identified the Neyman-Rubin potential outcomes framework as a statistical model of counterfactuals (Imbens et al., 2025 [32]; Wu et al., 2025 [33]; Zhang et al., 2025 [34]; Keller and Branson, 2024 [35]). According to this framework, the causal effects are impossible to observe directly and identification assumptions like, ignoring, overlap and consistency have to be made in order to draw valid inferences.

The issues of the counterfactual analysis were also addressed such as missing data, selection bias and unobserved confounders (Byrnes and Dee, 2025 [36]; Benz et al., 2025 [37]; Heng et al., 2025 [38]). These limitations are important in identifying empirical strategies to be used and credible estimates of causation.

2.7.3 Insights from the Case Study

The example of the impact of education on income case study (Lee et al., 2024 [39]; Tennant et al., 2022 [40]) presented an entire instance of the use of counterfactual reasoning:

- The causal question is clear by defining treatment and outcome variables.
- Counterfactual logic is useful in explaining what a person would have earned in other educational conditions.
- The fact that ability, family background, or selection may harm awareness of bias in the model points to the relevance of econometric rigor.
- There are statistical methods such as regression adjustment, instrumental variables, or quasi-experimental designs that implement the framework mentioned in the previous sections.

2.7.4 Linking Back to Econometric Foundations

The chapter is a bridge the gap between conceptual counterfactual reasoning and the actual tools of formal econometrics to estimating causal effects. The chapter prepares the subsequent chapters in that it introduces the potential outcomes, treatment effects, and identification assumptions that will be used in subsequent chapters and will revolve around:

- Causal inference based on advanced estimation,
- Dealing with confounding and missing data,
- Policy evaluation of an observational and experimental data,
- Contemporary usage of Bayesian, doubly robust, and machine-learned causal procedures (Du et al., 2024 [25]; Sert et al., 2025 [27]).

Essentially, counterfactual thought offers the conceptual perspective as well as the statistical structure to interpret causality in economics preparing the reader to applied econometric analysis in the later chapters.