Universiteit Leiden
The Netherlands

Opleiding Informatica

Evaluating Causal Inference Methods
in Natural Experiment Setup

Tolunay Ahmedov

Supervisors:
Saber Salehkaleybar
Ana Letícia Garcez Vicente

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

07/08/2026

Abstract

This thesis examines the robustness of Difference-in-Differences (DiD) estimates, a widely used method for causal inference in applied economics. Because DiD relies on assumptions such as parallel trends, which cannot always be verified, we compare its results with those from a newer Cross-Moment approach that uses higher-order moments and proxy variables to address latent confounding. Using five publicly available datasets, we replicate published DiD estimates and compare them with cross-moment estimates. Across seven treatment effects, the two methods match in sign in five cases, while two estimates differ. These results suggest that the main qualitative conclusions-especially the direction of the effect-are fairly consistent across different identification strategies, while the exact size of the estimates appears much more sensitive to the underlying assumptions and model choices.

Contents

1 Introduction **1**

2 Preliminaries and Notation **4**
2.1 Difference-in-Differences . . . 4
2.1.1 Setup . . . 4
2.1.2 Estimation . . . 4
2.1.3 Assumptions . . . 4
2.1.4 Identification . . . 5
2.2 Cross-Moment Model . . . 5
2.2.1 Setup . . . 5
2.2.2 Identification Assumptions . . . 6
2.2.3 Identification Result . . . 7
2.2.4 Estimation . . . 7
2.2.5 Extension . . . 8

3 Methodology **9**
3.1 Design . . . 9
3.2 Datasets . . . 9
3.3 Software and Implementation . . . 12
3.4 Sample Restrictions and Time Windows . . . 12
3.5 Comparison Strategy . . . 13

4 Related Work **14**

5 Experiments **16**
5.1 Summary of Experimental Findings . . . 21

6 Conclusion and Further Research **22**

References **25**

1 Introduction

Causal inference aims to estimate how an intervention or treatment affects an outcome using observational data [IR15, Cun21]. Unlike randomized control trials, where treatment assignment is controlled, observational datasets come from complex, uncontrolled processes. In these datasets, treatment often correlates with both observed and unobserved variables [AP09, SW20]. This creates an identification problem. The observed data-generating process does not directly reflect causal relationships; instead, it reflects a mix of causal effects and confounding factors [Pea09, IR15]. Therefore, estimating causal effects requires adding more structure or assumptions to the data. It also involves designing algorithms that can identify parameters under these conditions [Pea09, Cun21]. This challenge is vital in modern data-driven applications, such as policy evaluation, recommendation systems, and machine learning pipelines that depend on observational logs [IR15, Pea09].

A widely used approach to this problem is the Difference-in-Differences (DiD) method [AP09, Cun21]. DiD exploits panel or repeated cross-sectional data structures, in which observations are indexed by both units and time [SW20]. By computing differences across time and between groups, the method effectively removes certain classes of unobserved heterogeneity, such as time-invariant unit-specific effects and common shocks [AP09]. This can be interpreted as a data transformation that simplifies the underlying statistical model, making causal parameters identifiable under some specific conditions [Woo10]. From an implementation standpoint, DiD is attractive due to its simplicity, scalability, and compatibility with standard regression frameworks, which have contributed to its widespread adoption in both economics and applied data science [Cun21].

However, the validity of DiD depends critically on the parallel trends assumption, which requires that, in the absence of treatment, the treated and control groups would evolve according to the same underlying trend [AP09, IR15]. In practice, this assumption imposes a strong constraint on the latent data-generating process and cannot be fully verified from observed data [MR12]. While diagnostic tools such as pre-treatment trend analysis are commonly used, they provide only partial evidence and do not guarantee that the assumption holds in counterfactual scenarios [Cun21]. From an identification perspective, violations of this assumption correspond to model misspecification, where the transformation applied by DiD fails to eliminate confounding structure. This can lead to biased estimators and unstable results, particularly in high-dimensional or noisy datasets [Pea09].

These limitations have driven the development of alternative identification strategies that rely on different structural assumptions or use additional information in the data. Recent work has focused on methods that go beyond time-based variation and instead exploit statistical properties of the joint distribution of observed variables. One such approach is the Cross-Moment method, which uses higher-order moment conditions to identify causal effects in the presence of unobserved confounding [KSK23]. This framework utilizes proxy variables that provide information about latent confounders and estimates causal effects through moment-based restrictions [KSK23]. However, the Cross-Moment approach also relies on strong assumptions, including a linear structural causal model, the presence of a single latent confounder, and access to valid proxy information. Importantly, identification requires non-Gaussianity of the latent confounding structure; when the observed variables are jointly Gaussian, the causal effect is not identifiable because higher-order moments do not provide additional information. This shifts the problem from differencing over time to solving

systems of equations derived from empirical moments, often requiring more advanced estimation procedures and careful handling of finite-sample properties.

The central research question of this thesis is:

> *How does the Cross-Moment approach perform compared to traditional Difference-in-Differences estimators when applied to real-world datasets when validity of the parallel trends assumption is uncertain?*

This thesis conducts an empirical comparison between traditional DiD estimators and a Cross-Moment-based approach using several real-world datasets. For each application, DiD results from the original studies are replicated and function as benchmarks for comparison with the corresponding Cross-Moment estimates. The primary focus is on whether both methods yield treatment effect estimates with the same sign, thereby signifying agreement in the qualitative direction of the effect. Sign agreement is particularly important in this setting because it indicates whether different identification strategies lead to the same substantive conclusion about the impact of a treatment. In applied econometric work, causal estimates are often used to inform whether an intervention increases or decreases an outcome, making the direction of the effect central to interpretation and policy relevance [AP09, IR15, Cun21]. Even when estimated effect sizes differ across methods, consistent signs suggest robustness in the underlying causal conclusion, whereas conflicting signs may indicate sensitivity to identification assumptions and lead to different substantive interpretations [R+23, Lec11, SRTT16].

The central hypothesis is that the Cross-Moment approach will generate treatment effect estimates with signs consistent with those from established DiD studies, only when the parallel trend assumption is not violated. This hypothesis is based on the expectation that both methods aim to recover the same underlying causal effect under different identifying assumptions [AP09, IR15]. If the assumptions for each method are approximately satisfied in practice, the two approaches are expected to recover a similar directional relationship between treatment and outcome, even if they differ in magnitude. As such, agreement in sign can be interpreted as evidence that the qualitative causal conclusion is robust to the choice of identification strategy rather than being driven by method-specific assumptions [R+23, Cun21].

This enables a direct assessment of directional consistency between a widely used econometric baseline and a more recent identification strategy. Although differences in magnitude are also reported, they are not the main focus of the analysis. Consistency across methods is ensured by replicating the initial DiD specifications and implementing the Cross-Moment estimator within the same computing platform. The objective is to compare the Cross-Moment approach in reproducing the directional conclusions of established DiD studies and to evaluate the sensitivity of causal sign inference to the chosen identification strategy. Agreement in sign is informative because it suggests that the qualitative causal conclusion is robust across methods that rely on different identifying assumptions. Conversely, disagreement in sign does not necessarily imply that one estimator is incorrect, but may instead reflect differences in identifying assumptions, violations of model conditions, or the fact that the two approaches recover different weighted causal effects. As a result, comparing directional consistency provides a diagnostic tool for assessing how sensitive

substantive conclusions are to the choice of identification framework rather than a definitive test of correctness.

The remainder of this thesis is structured as follows. Section 2 introduces the theoretical background and key definitions, including the potential outcomes framework, DiD estimation, and the Cross-Moment approach. Section 3 describes the methodology, including the empirical design, data, implementation, and comparison strategy. Section 4 reviews the relevant literature and positions the two approaches within causal inference. Section 5 presents the empirical results and compares the methods. Finally, Section 6 concludes the thesis and discusses future research directions.

2 Preliminaries and Notation

This section introduces the notation and causal framework used throughout the thesis and proceeds in a step-by-step manner through the main methodological components. First, we present the potential outcomes framework, which formalizes the definition of causal effects and serves as the basis for all subsequent analyses [IR15]. Next, we introduce the DiD estimator and its identifying assumptions, which represent the primary benchmark method used in the empirical applications. We then introduce the Cross-Moment and outline its corresponding identification assumptions.

Using the structural causal model formulation of Cross-Moment, we derive the main identification result and show how the causal parameter can be expressed in terms of higher-order moments. We further discuss the role of non-Gaussianity, which is essential for identification in this setting. Finally, we present the Cross-Moment estimation algorithm. Together, these components provide a comparison of the DiD and Cross-Moment estimators in terms of both their identifying assumptions and their implied causal effects.

2.1 Difference-in-Differences

Throughout this section, Y denotes the observed outcome variable. The binary variable $D \in \{0, 1\}$ is the treatment indicator, where $D = 1$ identifies treated units and $D = 0$ identifies untreated (control) units.

2.1.1 Setup

Suppose that we observe treated and control groups over two time periods $t \in \{0, 1\}$ for simplicity of exposition. This two-period setup is the standard framework for introducing DiD and highlighting the identification intuition. In the empirical section, we generalize this approach to multiple time periods using an extended specification that exploits variation across both groups and time.
The average outcome in the treated group is:

$$E[Y_t \mid D]. \tag{1}$$

2.1.2 Estimation

The DiD estimator is given by [AP09, Cun21]:

$$\delta^{DiD} = (E[Y_1 \mid D = 1] - E[Y_0 \mid D = 1]) - (E[Y_1 \mid D = 0] - E[Y_0 \mid D = 0]). \tag{2}$$

2.1.3 Assumptions

For the DiD framework to be valid, we assume the Stable Unit Treatment Value Assumption (SUTVA), which consists of two components. First, there is no interference between units, meaning that one unit's outcome is not affected by other units' treatment status. Second, the treatment is well defined with no hidden variations, so that all treated units receive the same version of the treatment.

Under SUTVA, the observed outcome is linked to potential outcomes through the consistency relation:

$$Y = DY(1) + (1 - D)Y(0). \tag{3}$$

Under the assumption of parallel trends [AP09, CK94], the treated and control groups would have experienced the same average change in outcomes over time in the absence of treatment. This implies that any deviation in their post-treatment evolution can be attributed to the treatment effect.

$$E[Y_1(0) - Y_0(0) \mid D = 1] = E[Y_1(0) - Y_0(0) \mid D = 0]. \tag{4}$$

2.1.4 Identification

The causal parameter of interest is the Average Treatment Effect on the Treated (ATT), defined as

$$ATT = E[Y_1(1) - Y_1(0) \mid D = 1]. \tag{5}$$

Under SUTVA and the parallel trends assumption, the DiD estimator identifies the ATT. Therefore,

$$\delta^{DiD} = ATT = E[Y_1(1) - Y_1(0) \mid D = 1]. \tag{6}$$

2.2 Cross-Moment Model

2.2.1 Setup

So far, the analysis has been framed in the potential outcomes framework. We now introduce an equivalent structural representation that is particularly convenient for the Cross-Moment identification strategy, which is naturally formulated using structural causal models (SCMs) [Pea09].

Structural causal models (SCMs) formalize causal relationships through a system of structural equations. In this framework, each variable is expressed as a deterministic function of its direct causes and an exogenous error term. We impose a linear and additive noise structure, where each equation is linear in its arguments and the error terms enter additively.
This implies that (i) effects are linear and proportional, (ii) unobserved heterogeneity enters through additive disturbances, and (iii) the exogenous noise terms capture all sources of variation not explicitly modeled in the system and are not further explained within the model. These exogenous terms are typically assumed to be mutually independent across equations.

We assume the following linear structural causal model (SCM) [Pea09]:

$$U = \varepsilon_u, \tag{7}$$
$$Z = \alpha_z U + \varepsilon_z, \tag{8}$$
$$D = \alpha_d U + \varepsilon_d, \tag{9}$$
$$Y = \beta D + \gamma U + \varepsilon_y, \tag{10}$$

where:

- U is an unobserved confounder,
- Z is an observed proxy variable,
- D is the treatment,
- Y is the outcome,
- β is the causal effect of interest.

The unobserved confounder U affects both the treatment D and the outcome Y, generating endogeneity. The variable Z serves as a noisy proxy for U, providing partial information about the unobserved heterogeneity.

We observe only (D, Y, Z), while U is unobserved. This induces confounding, so that the naive regression of Y on D generally does not recover β [AP09, IR15].
The key question is whether β can be identified from the observational distribution of (D, Y, Z).

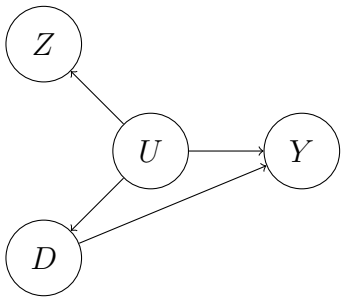

Figure 1: Linear structural causal model (SCM). The unobserved confounder U affects both treatment D and outcome Y, while Z acts as a noisy proxy for U.

2.2.2 Identification Assumptions

Identification holds under the following assumptions [KSK23]:

1. $\alpha_z \neq 0$: the proxy Z carries information about the latent confounder U.
2. $\mathrm{Var}(\varepsilon_d) > 0$: the treatment contains variation not fully determined by U.

In addition to these baseline conditions, identification of α_d/α_z relies on higher-order moment restrictions. Assuming that ε_u admits finite higher-order moments, for $n \geq 3$ define:

$$\begin{aligned} \text{num} &= \mathbb{E}[D^{n-1}Z] - (n-1)\mathbb{E}[D^{n-2}]\mathbb{E}[DZ], \\ \text{den} &= \mathbb{E}[Z^{n-1}D] - (n-1)\mathbb{E}[Z^{n-2}]\mathbb{E}[DZ]. \end{aligned}$$

Under these moment conditions, the magnitude of the ratio is identified as:

$$\left|\frac{\alpha_d}{\alpha_z}\right| = \left(\frac{\text{num}}{\text{den}}\right)^{\frac{1}{n-2}}. \tag{11}$$

with sign determined by $\text{sign}(\mathbb{E}[DZ])$.

Finally, identification crucially depends on the distributional properties of the confounder. If ε_u is Gaussian, higher-order moments do not provide additional identifying information, and the model is not identifiable. If ε_u is non-Gaussian, higher-order moments contain identifying information that allows recovery of α_d/α_z and thus β [KSK23].

2.2.3 Identification Result

Using moment conditions, the causal effect can be written as:

$$\beta^{CM} = \frac{\text{Cov}(D,Y) - \frac{\alpha_d}{\alpha_z}\text{Cov}(Y,Z)}{\text{Var}(D) - \frac{\alpha_d}{\alpha_z}\text{Cov}(D,Z)}. \tag{12}$$

Thus, β is identified from the joint distribution of (D, Y, Z) [KSK23].

2.2.4 Estimation

The estimator is computed using the algorithm presented in Algorithm 1 [KSK23].

Algorithm 1 Cross-Moment Algorithm

function GetBeta(D, Z, Y)
    ratio $\leftarrow$ GetRatio(D, Z)
    $\beta \leftarrow \dfrac{\mathbb{E}[DY] - \text{ratio} \cdot \mathbb{E}[YZ]}{\mathbb{E}[D^2] - \text{ratio} \cdot \mathbb{E}[DZ]}$
    return β
end function

function GetRatio(D, Z)
    found $\leftarrow$ False
    $n \leftarrow 2$
    while found = False **do**
        $n \leftarrow n + 1$
        num $\leftarrow \mathbb{E}[D^{n-1}Z] - (n-1)\mathbb{E}[D^{n-2}]\mathbb{E}[DZ]$
        den $\leftarrow \mathbb{E}[Z^{n-1}D] - (n-1)\mathbb{E}[Z^{n-2}]\mathbb{E}[DZ]$
        if den $\neq 0$ **then**
            ratio $\leftarrow \text{sign}(\mathbb{E}[DZ]) \cdot \left(\dfrac{\text{num}}{\text{den}}\right)^{\frac{1}{n-2}}$
            found $\leftarrow$ True
        end if
    end while
    return ratio
end function

2.2.5 Extension

The framework extends to multiple latent confounders under standard proxy assumptions [KSK23]:

- each confounder has a valid proxy,
- proxies are not descendants of the outcome,
- confounders are mutually independent in structure,
- each confounder-proxy pair satisfies the identification conditions.

Under these conditions, β remains identifiable from observational data.

3 Methodology

This thesis compares the performance of the DiD and Cross-Moment estimators in empirical applications with potential unobserved confounders. Both methods are applied to previously studied real-world datasets, selected based on their public availability and relevance as documented in [MR12]. Only datasets that were accessible for independent replication were included in the analysis, ensuring consistency in empirical design and comparability of the estimated treatment effects across the two identification strategies.

3.1 Design

The objective is to compare the estimated causal effects of a treatment variable D on an outcome Y obtained using two alternative approaches and to assess whether they produce similar results:

1. Difference-in-Differences
2. Cross-Moment approach

DiD relies on the parallel trends assumption and is one of the most widely used methods for causal inference featuring observational panel data, as established in the empirical economics and applied statistics literature [AP09, Cun21]. This popularity is due to its relative simplicity and interpretability under well-understood assumptions. In contrast, the Cross-Moment approach, recently introduced in the literature, relies on linear structural assumptions, the existence of a valid proxy variable, and non-Gaussian confounding [KSK23]. Its selection is justified by recent advances showing that these conditions enable the identification of causal effects even when traditional parallel trends may be violated, particularly in the presence of latent confounders. By applying both estimators to the same datasets, this thesis evaluates how sensitive empirical conclusions are to the underlying identifying assumptions in terms of both the sign and magnitude of the estimated effects. The Cross-Moment estimator is thus employed as an alternative identification strategy, motivated by evidence from the literature showing its strengths in situations where unobserved confounding may undermine the validity of DiD estimation [KSK23].

3.2 Datasets

These studies analyze treatment effects in different economic and policy contexts and were originally estimated by using quasi-experimental methods such as DiD. Prior to estimation, each dataset was cleaned and pre-processed according to the methods and requirements described in the respective original papers.

The following datasets were publicly available:

- Compulsory Licensing: Evidence from the Trading with the Enemy Act
  Moser et al. (2012)[MV09]
- Evidence on the Impact of Sulfa Drugs
  Jayachandran et al. (2010)[JLMS10]

- Traffic Congestion and Infant Health: Evidence from E-ZPass
  Currie and Walker (2011)[CW11]
- Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research
  Furman and Stern (2011)[FS11]
- Marrying Up: The Role of Sex Ratio in Assortative Matching
  Abramitzky et al. (2011)[ADV11]

[MV09]: The dataset includes patent grants by U.S. inventors across 7,248 chemical classes from 1875 to 1939. Each entry corresponds to a class year. We define a treatment indicator, `treat`, which is 1 if a class has at least one enemy-owned patent licensed to a U.S. firm. We also define a post-treatment indicator, `postTWEA`, which is 1 for years after 1918. The outcome variable measures the number of patents granted to U.S. inventors in a class-year (`count_usa`).

For the DID estimation, we use PanelOLS. The interaction `treat` × `postTWEA` serves as the main explanatory variable. We include class and year fixed effects to account for time-invariant class differences and shared shocks. We cluster standard errors at the class level.
In the Cross-Moment estimation, Z represents the pre-treatment `treat` values, D represents the post-treatment `treat` values, and Y represents the post-treatment number of U.S. patents (`count_usa`). Pre-treatment variables are used as proxy variables because they are unaffected by the treatment while potentially containing information about unobserved confounders that influence both treatment assignment and outcomes, making them informative signals for latent structure in Cross-Moment identification approaches [MGTT18, TT+20, KSK23]. All variables are centered by subtracting their mean. We align pre- and post-period observations to ensure equal length, which allows for the calculation of the Cross-Moment estimator.

***[JLMS10]*:** The dataset includes national mortality counts for measles (MMR), influenza/pneumonia, and scarlet fever from 1925 to 1943. Each observation relates to a disease-year. The treatment indicator, `treated`, is 1 for MMR, influenza/pneumonia, or scarlet fever, and 0 for control diseases (diseases not treatable by sulfa drugs). The post-treatment indicator, `Post37`, is 1 for years after 1936. The outcome variable is the log-transformed mortality count, noted as `lnMortality`.

The analysis focuses on all 3 mentioned diseases, using a DiD model. This model incorporates the interactions `treated × Post37` and `treated × Year` to reflect the immediate post-treatment effect and the trends before and after treatment. For the Cross-Moment approach Z represents pre-treatment `treated × Year × Post37`, D indicates the combined pre- and post-treatment `treated × Post37`, and Y refers to the stacked log mortality counts. All variables are centered at their mean, and pre- and post-treatment observations are arranged for the Cross-Moment calculation.

***[CW11]*:** In this paper daily air pollution data from monitoring stations near New Jersey highway toll plazas are analyzed. The outcome variable is $Y =$ `log_NO2`, defined as the natural logarithm of `dailyMeanno2`, which measures the daily mean nitrogen dioxide concentration. The key treatment indicator is $D =$ `near_toll`, a binary variable equal to 1 if the monitoring station is located within 2 km of a toll plaza based on `minDistanceMonitor`, and 0 otherwise. The post-treatment indicator `post` equals 1 for dates after the introduction of EZ-Pass (`ezpass_date`) and 0 otherwise. The

interaction `did` = `near_toll` × `post` captures the DiD treatment exposure. Additional control variables include `distanceMonitor`, time fixed effects (`year`, `month`), toll plaza fixed effects (`toll_id`), and monitor-specific linear time trends (`monitor_trend`).

In the DiD estimation, we estimate a linear model using OLS with clustered standard errors, where Y = `log_NO2` and the treatment effect is captured by D = `did`. Controls include the variables above.
In the Cross-Moment approach, Y = `log_NO2`, D = `near_toll`, and Z is the residualized instrument constructed from the outcome after removing the effects of `distanceMonitor`, `year`, `month`, `dow`, `toll_id`, and `monitor_trend` estimated from the pre-treatment period. All variables are mean-centered.

***[FS11]*:** The analysis uses the dataset `brc_core.dta`, which contains information on scientific articles and their citation outcomes over time. The outcome variable is Y = `cites9`, measuring the number of citations an article receives within nine years. The treatment variable is D = `brc`, a binary indicator equal to 1 if the article was deposited in the Biological Resource Center (BRC). The observation year (`year`) and publication year (`pub_year`) are used to construct two time indicators: `window`, equal to 1 for years within [`pub_year`-1, `pub_year`+1], and `post_brc`, equal to 1 for years $\geq$ `pub_year`+2. Observations are restricted to `sample3`=1 and missing values are removed. The regressions control for article age using indicators `age1` to `age30`, and standard errors are clustered at the pair level using `pair_num`.

The treatment effect is estimated using a DiD specification with OLS. Interaction terms `brc_window` = `brc` × `window` and `brc_post` = `brc` × `post_brc` capture the treatment effect.
For the Cross-Moment method, Y = `cites9`, D = `brc`, and Z is constructed from residualized pre-treatment citation outcomes after removing age controls (`age1`–`age30`). The centered variables are then used to compute the estimator.

***[ADV11]*:** This study utilizes a dataset that provides marriage-level details from various French departments and years following the war. The main outcome variable is Y = `classdiff`, defined as the difference between the bride's and groom's social class (`clbr - clgr`). The treatment variable is D = `post_mortality`. Additional controls include `rural`, `agebrd`, `agegrd`, and dummy variables derived from `clgr`.

For the DiD method, the specification is estimated using PanelOLS with observations indexed by `depc` and year. The Cross-Moment estimator uses Z = pre-treatment average of `classdiff`, Y = post-treatment average outcome, and D = `post_mortality`. All variables are centered before estimation.

3.3 Software and Implementation

All empirical analyses are implemented in Python using Jupyter notebooks, ensuring an interactive and reproducible workflow. The main libraries used include `numpy`, `pandas`, and `statsmodels`.

The complete code and datasets utilized in the analysis can be accessed at:
https://github.com/tahmedov/Causal-Inference-Methods

For each dataset, the following implementation strategy is used:

- Under SUTVA and the parallel trends assumption, the DiD estimator identifies the Average Treatment Effect on the Treated (ATT) defined in Section 2.
- Cross-Moment estimation is implemented using the linear structural model and the moment-based procedure described in Section 2.
- When required by the Cross-Moment procedure, additional sample construction is applied to define suitable pre- and post-treatment periods. These choices may differ from the DiD benchmark and depend on the empirical setting.

All computations are fully reproducible and executed within Jupyter notebook environments.

3.4 Sample Restrictions and Time Windows

The choice of sample period differs across applications depending on the requirements of each estimation strategy. The DiD estimator is typically less sensitive to the length of the observation window, as identification relies on the parallel trends assumption between treated and control groups. Therefore, longer time horizons can be used as long as the underlying trends remain comparable. In contrast, the Cross-Moment approach relies on structural assumptions regarding latent confounding, proxy validity, and the stability of higher-order moment relationships. Long time periods may introduce structural breaks, changes in treatment assignment, or shifts in the relationship between variables, potentially violating these assumptions. Therefore, in some applications, narrower time windows are used to obtain a more stable empirical setting for Cross-Moment identification.
For example, in the Jayachandran et al. (2010) application, both DiD and Cross-Moment estimates are obtained using the restricted period 1925–1943 surrounding the introduction of sulfa drugs. In contrast, for Abramitzky et al. (2011), the Cross-Moment estimator is constructed using pre- and post-war windows (1910–1918 and 1919–1925), while the DiD specification uses the full available panel. These differences reflect the methodological requirements of the Cross-Moment estimator rather than changes in the underlying research question.
This approach is consistent with latent variable and independent component analysis literature, where stability of the structural relationships over the estimation window is important for valid identification [HP01].

3.5 Comparison Strategy

To compare Difference-in-Difference and Cross-Moment results:

- For each dataset, the results are reported by the estimated effects of the treatment and, where available, the standard errors.
- A comparison mainly focused on the sign of the estimates is made, with magnitude and statistical significance being less relevant.
- Differences between β^{DiD} and β^{CM} may indicate sensitivity to identifying assumptions.

4 Related Work

Estimating causal effects from observational data is a fundamental problem in both economics and computer science, especially in situations where controlled experiments are not feasible [IR15, Pea09]. One of the most commonly used methods for this purpose is the DiD approach [AP09, Cun21]. DiD estimates treatment effects by comparing changes in outcomes over time between a treated group and a control group. However, its validity depends on the parallel trends assumption [AP09, CK94], which states that, without treatment, both groups would have followed the same trend over time. In practice, this assumption is difficult to verify and may not always hold, which has led to a growing body of research focused on relaxing or replacing it [Lec11, R+23].

A notable contribution to this literature is the work of Mora and Reggio [MR12]. Rather than relying solely on the standard parallel trends assumption, they develop a framework that characterizes treatment effect identification under a set of alternative assumptions about the evolution of untreated potential outcomes. Their approach shows that DiD estimates are generally not invariant to the choice of identifying assumptions, as different assumptions about counterfactual trends can lead to different estimands. A key insight from their analysis is that causal conclusions in DiD settings can be sensitive to the identifying assumptions imposed on the data, even within the DiD framework itself. In particular, they show that relaxing or modifying the parallel trends restriction can lead to meaningful changes in estimated effects, including differences in magnitude and, in some cases, statistical significance, while the sign of the effect is often more stable. Overall, their findings highlight that causal inference in DiD designs is sensitive to identifying assumptions, and that reporting a single specification may mask this uncertainty.

At the same time, a broader literature has studied causal inference under unobserved confounding using proxy variables. A central idea in this literature is the use of negative control outcome and exposure variables [LTTC10, SRTT16]. These are variables that are assumed not to be directly affected by the treatment but can still carry information about latent confounders, allowing for bias detection and adjustment under additional assumptions. For example, DiD can be interpreted as a special case of a negative control design in panel data settings [SRTT16]. More generally, proxy-based identification strategies exploit auxiliary variables to recover information about unobserved confounders. This includes single- and multiple-proxy models, where identification relies on conditions such as completeness or invertibility of the mapping between latent confounders and observed proxies [MGTT18]. In linear structural models, these approaches often depend on additional assumptions about functional form or conditional distributions, and in some cases multiple proxies are required for identification [KZCS14]. More broadly, recent work has developed general frameworks for causal inference with negative controls and proxy variables under unobserved confounding [TT+20].

Further advances in this area involve proximal causal inference, which formalizes the use of proxy variables within the potential outcomes framework and offers broader identification results under less strict assumptions [TT+20, MGTT18]. Newer studies build on these ideas for situations with latent confounding, taking into account linearity and non-Gaussianity assumptions [KSK23]. They demonstrate that causal effects can be retrieved even from partially observed systems. Meanwhile, methods based on independent component analysis (ICA) and non-Gaussian latent variable models

have been applied to identify causal structures in linear SCMs, often relying on higher-order statistical moments or cumulants [HP01, SHHK06].

At the same time, recent research has developed alternative methods for estimating causal effects without relying on DiD. One such approach is proposed by Kivva et al. in [KSK23], which we refer to as the Cross-Moment (CM) approach. The method considers settings with an unobserved confounder that jointly affects both treatment assignment and outcomes. Instead of exploiting time variation, identification is achieved through algebraic relationships between higher-order moments of the observed variables and an additional proxy variable that contains information about the latent structure. Under certain conditions-most importantly, the presence of non-Gaussianity in the data-generating process-these moment restrictions help distinguish causal effects from confounding effects [KSK23]. In particular, identification relies on higher-order moment information that is not available under Gaussian distributions, where second-order moments fully characterize the joint distribution.

Compared to traditional econometric methods, this framework changes the identification strategy. It moves from drawing on time or group variation to using the distributional properties of the data. Identification is therefore achieved through moment conditions instead of quasi-experimental variation, making these methods especially useful in cases where panel data is not available or where the assumptions of parallel trends are likely not met [IR15, AP09].

Under this framework, causal effects are typically uncovered in two steps. First, the ratio between treatment and proxy loadings is identified using higher-order Cross-Moments. Second, this ratio is inserted into the structural outcome equation to define the causal effect. This two-step process relies on linearity, the presence of valid proxy variables, and non-Gaussian confounding [KSK23, HP01].

These two papers [MR12] and [KSK23] represent different approaches to the same core problem: estimating causal effects when standard assumptions may not hold. While Ricardo Mora and Iliana Reggio [MR12] extend the DiD framework by making the parallel-trends assumption more flexible, Kivva et al. [KSK23] propose a completely different approach based on the properties of the data. However, there is limited work directly comparing these methods in practice, especially regarding their sensitivity to underlying assumptions [IR15].

This thesis addresses this gap by systematically comparing identification strategies. It replicates the baseline DiD estimates from the empirical applications in Mora and Reggio under the standard parallel trends assumption. Next, it applies the Cross-Moment method proposed by Kivva et al. using similar datasets. By directly comparing the results from both approaches, the analysis shows how their underlying assumptions affect the estimated treatment effects and the reliability of empirical conclusions.

5 Experiments

For each empirical study, we first summarize the original findings reported in the corresponding publication. We then present our DiD replication and assess its consistency with the published results. Next, we report the corresponding Cross-Moment estimate and compare it with the replicated DiD estimate. Finally, we discuss possible explanations for similarities and differences between the two approaches, with particular attention to their underlying identifying assumptions. Table 1 provides a compact overview of all results.

Moser et al. (2012):

Original findings

Moser and Voena [MV09] examine the effect of compulsory licensing introduced during World War I on domestic innovation in the United States. Using a DiD framework, the authors compare chemical patent subclasses affected by compulsory licensing with unaffected subclasses before and after the policy change in 1919. Their results indicate that compulsory licensing significantly increased domestic patenting activity. Depending on the specification, treated subclasses produced between 0.151 and 0.255 additional patents per year after 1919, corresponding to a 24–40% increase relative to the baseline level of 0.619 patents per subclass-year. These effects remain statistically significant when controlling for foreign patenting and when alternative treatment definitions are considered. The authors interpret these findings as evidence that compulsory licensing reduced barriers to knowledge access and stimulated domestic innovation by enabling firms to utilize previously restricted technologies [MV09].

Replication results

Replicating the DiD specification from Moser and Voena [MV09], we obtain an estimated treatment effect of 0.1505 (SE = 0.0359). This estimate is consistent with the original findings, preserving both the positive sign and similar magnitude of the reported effects. Therefore, the replication confirms the original qualitative conclusion that compulsory licensing increased domestic patenting activity.

Cross-Moment results

Applying the Cross-Moment estimator to the same empirical setting yields an estimated treatment effect of 0.1102. The Cross-Moment estimate has the same sign as both the original DiD estimate and the replicated DiD result, indicating agreement in the qualitative direction of the treatment effect. Although the magnitude differs from the DiD estimate, the primary conclusion remains unchanged: both approaches suggest that compulsory licensing increased domestic innovation.

Discussion

The agreement in sign between the DiD and Cross-Moment estimates suggests that the qualitative conclusion of the original study is robust across two identification strategies based on different assumptions. However, the difference in magnitude highlights that the estimators do not necessarily recover identical numerical effects. The DiD estimator identifies the treatment effect through temporal variation and relies on the parallel trends assumption, whereas the Cross-Moment approach relies on moment conditions and assumptions regarding latent confounding and proxy validity. Consequently, differences in effect size may arise from the distinct sources of identifying variation

used by each method. In this application, where both approaches recover a positive treatment effect, the Cross-Moment estimate supports the original conclusion that compulsory licensing increased domestic patenting. The consistency in sign indicates that the qualitative causal interpretation is not sensitive to the choice of estimator in this setting, even though the estimated magnitudes differ.

Jayachandran et al. (2010):

Original findings

Jayachandran et al. [JLMS10] investigate the effect of the introduction of sulfa drugs on mortality rates in the United States. Using a DiD framework, the authors compare mortality trends for diseases affected by sulfa drugs with diseases that were not affected by the treatment. Their preferred specification (Model 5.3) reports significant reductions in mortality after the introduction of sulfa drugs in 1937. The estimated effects are negative and statistically significant for measles, mumps, and rubella (MMR) (-0.304–0.148), pneumonia/influenza (-0.163–0.037), and scarlet fever (-0.862–0.495), corresponding to reductions in mortality of approximately 26–38%, 15–26%, and 58–72%, respectively. In contrast, the control disease tuberculosis does not show a significant break after 1937, supporting the validity of the original identification strategy. The authors interpret these findings as evidence that the availability of sulfa drugs generated substantial reductions in mortality by providing an effective treatment for previously difficult-to-treat infectious diseases [JLMS10].

Replication results

Replicating Model 5.3 from Jayachandran et al. [JLMS10], we obtain estimates of -0.304 (SE = 0.128) for MMR, -0.163 (SE = 0.112) for pneumonia/influenza, and -0.862 (SE = 0.334) for scarlet fever. These estimates closely match the original results, preserving the sign, approximate magnitude, and ordering of treatment effects across diseases. Therefore, the replication confirms the original qualitative conclusion that the introduction of sulfa drugs substantially reduced mortality.

Cross-Moment results

Applying the Cross-Moment estimator produces treatment effect estimates of 0.1881 for MMR, -1.4576 for pneumonia/influenza, and -4.8411 for scarlet fever. For pneumonia/influenza and scarlet fever, the Cross-Moment estimates have the same negative sign as the DiD estimates, indicating agreement in the qualitative direction of the treatment effect. For MMR, however, the Cross-Moment estimate is positive, which differs from the negative DiD estimate obtained in both the original study and the replication.

Discussion

The comparison between DiD and Cross-Moment provides mixed evidence in this application. While two of the three outcomes show sign agreement, the MMR estimate demonstrates that the qualitative conclusion can be sensitive to the choice of identification strategy. The disagreement for MMR does not necessarily imply that one estimator is incorrect, but may reflect differences in the identifying assumptions underlying the two approaches. The DiD estimator relies on the assumption that treated and control diseases would have followed similar trends in the absence of sulfa drugs, whereas the Cross-Moment approach relies on distributional assumptions and proxy validity conditions. Overall, the results indicate that Cross-Moment partially reproduces the directional conclusions of the original DiD analysis in this historical setting.

Currie and Walker (2011):

Original findings

Currie and Walker [CW11] question the effect of electronic toll collection through E-ZPass adoption on local air pollution and infant health outcomes. Using a DiD framework, the authors estimate the impact of E-ZPass adoption on pollutant concentrations near toll plazas. Their results show that E-ZPass adoption significantly reduced nitrogen dioxide (NO_2) concentrations near toll plazas. In their preferred specification, daily mean NO_2 levels decreased by approximately 10.8% at monitors located within 0.15 km of toll plazas. Additional specifications using alternative NO_2 control monitors also report significant declines, ranging from 6.5% to 20.8%. In contrast, the authors find no measurable effect on sulfur dioxide (SO_2) concentrations. They interpret these findings as evidence that electronic toll collection reduces vehicle congestion and idling time, thereby lowering traffic-related emissions and improving local air quality [CW11].

Replication results

Replicating the DiD specification using all available control monitors, we obtain an estimated treatment effect of -0.126 (SE = 0.020) for NO_2. This corresponds to an approximate 12% reduction in NO_2 concentrations near toll plazas. The replication therefore closely matches the original findings, preserving the negative sign and similar magnitude of the estimated effect. These results confirm the original qualitative conclusion that E-ZPass adoption reduced local NO_2 pollution.

Cross-Moment results

Applying the Cross-Moment estimator to the same empirical setting yields an estimated treatment effect of 0.9108. In contrast to the DiD estimate, the Cross-Moment result is positive, indicating an increase in NO_2 concentrations following E-ZPass adoption. Therefore, this application does not exhibit sign agreement between the two approaches.

Discussion

The disagreement between the DiD and Cross-Moment estimates illustrates a case where the qualitative conclusion depends on the identification strategy. The DiD results suggest that E-ZPass adoption reduced NO_2 pollution, consistent with the mechanism proposed by Currie and Walker [CW11]. The Cross-Moment estimate instead implies the opposite directional effect. The DiD approach relies on the parallel trends assumption and temporal changes associated with E-ZPass adoption, whereas the Cross-Moment estimator relies on moment conditions and assumptions regarding latent confounding and proxy variables. This divergence does not by itself establish that either estimator is incorrect. Instead, it demonstrates that when identifying assumptions differ substantially, causal conclusions may not always be invariant to the estimation strategy.

Furman and Stern (2011):

Original findings

Furman and Stern [FS11] analyze whether linking biological materials to Biological Resource Centers (BRCs) affects the subsequent scientific impact of research articles. Using DiD specifications, the authors compare citation outcomes for articles linked to BRCs with comparable articles that are not linked. Their results show that BRC linkage is associated with substantial increases in citations. In their OLS specifications, articles linked to BRCs experience a baseline citation increase of 46.5% (SE = 0.143) when controlling only for age fixed effects and 51.3% (SE = 0.122) when additionally including article family and year fixed effects. The authors also find positive effects for different timing windows, with the window-period effect increasing citations by approximately 35.1–35.5% and the post-deposit effect increasing citations by approximately 57.4–64.4%. They interpret these findings as evidence that BRCs facilitate knowledge diffusion by reducing barriers to accessing biological materials and enabling further reuse of scientific resources, thereby increasing the cumulative impact of research discoveries [FS11].

Replication results

Replicating the OLS DiD specification with age fixed effects, we obtain an estimated window-period effect of approximately 50% (SE = 0.020). This estimate is consistent with the original findings, preserving the positive sign and similar magnitude of the reported citation effect. Therefore, the replication supports the original qualitative conclusion that BRC linkage increases subsequent citation impact.

Cross-Moment results

Applying the Cross-Moment estimator yields an estimated treatment effect of approximately 150%. The Cross-Moment estimate is positive and therefore agrees with the DiD estimate in terms of qualitative direction. However, the estimated magnitude is substantially larger than the replicated DiD effect.

Discussion

The agreement in sign between the DiD and Cross-Moment estimates indicates that both approaches support the same qualitative conclusion: association with a Biological Resource Center increases citation impact. However, the difference in magnitude highlights the sensitivity of quantitative estimates to the choice of identification strategy. The DiD approach estimates the treatment effect using temporal variation and relies on assumptions regarding comparable trends between treated and control articles. The Cross-Moment approach instead relies on moment conditions and assumptions about latent confounding and proxy variables. Differences in these identifying assumptions may explain why the Cross-Moment estimator produces a larger effect size. In this application, the consistency in sign suggests that the main qualitative conclusion is robust across estimation strategies, while the difference in magnitude indicates that the size of the estimated effect is more sensitive to methodological choices.

Abramitzky et al. (2011):

Original findings

Abramitzky et al. [ADV11] review how changes in the local sex ratio caused by World War I military mortality affected marriage-market outcomes in France. Using a DiD framework, the authors exploit variation in regional male mortality before and after the war to estimate the effect of increased male scarcity on assortative matching. Their results show that higher male mortality improved marriage outcomes for men. In the full sample, the interaction between postwar period and regional mortality reduces the class difference between spouses by 0.020 (SE = 0.010), decreases the probability of men marrying down by 0.010 (SE = 0.004), and reduces the probability of marrying a low-class bride by 0.017 (SE = 0.005). These effects are generally larger when excluding higher-class grooms, with the estimated effect on class difference reaching -0.035 (SE = 0.013). The authors interpret these findings as evidence that reductions in the local sex ratio altered marriage-market equilibrium by increasing male bargaining power and improving assortative matching outcomes [ADV11].

Replication results

Replicating the DiD specification for the class difference outcome, we obtain an estimated treatment effect of -0.0021 (SE = 0.0126). This estimate preserves the negative sign of the original result and suggests that increased male mortality is associated with improved assortative matching outcomes. Although the magnitude is smaller than the baseline estimate reported in the original study (-0.020) and statistical significance is weaker (p = 0.0696), the replication remains directionally consistent with the published findings.

Cross-Moment results

The Cross-Moment estimator produces an estimated treatment effect of -0.0101. This estimate has the same negative sign as both the original DiD estimate and the replication result. Therefore, both estimation strategies agree that increased male scarcity is associated with improved marriage outcomes. Although the Cross-Moment estimate differs in magnitude from the DiD results, it supports the same qualitative conclusion.

Discussion

The agreement in sign across the original DiD results, the replication, and the Cross-Moment estimate suggests that the qualitative conclusion is robust to the choice of identification strategy in this application. Both approaches indicate that increased male scarcity following World War I improved assortative matching outcomes for men. The difference in magnitude between the estimators may result from their different identifying assumptions. The DiD approach relies on comparing changes over time between regions with different levels of mortality exposure, while the Cross-Moment approach identifies the treatment effect through moment conditions and assumptions about latent variables. Therefore, while the direction of the estimated effect is consistent across methods, the exact magnitude remains sensitive to the estimation framework. This result supports the central hypothesis of this thesis that Cross-Moment can recover the same qualitative causal conclusion as traditional DiD even when the underlying identification strategies differ.

Table 1: Difference-in-Differences and Cross-Moment Results

Article	DiD Estimate	Cross-Moment Estimate	Sign Agreement
Moser et al. (2012)	0.1505 (0.0359)	0.1102	✓
Jayachandran et al. (2010) – MMR	-0.3039 (0.1277)	0.1881	×
Jayachandran et al. (2010) – Pneumonia/Flu	-0.1631 (0.1122)	-1.4576	✓
Jayachandran et al. (2010) – Scarlet fever	-0.8618 (0.3343)	-4.8411	✓
Currie and Walker (2011)	-0.1260 (0.0200)	0.9108	×
Furman and Stern (2011)	0.4034 (0.0841)	0.9149	✓
Abramitzky et al. (2011)	-0.0021 (0.0126)	-0.0101	✓

Notes: Standard errors are reported in parentheses for DiD estimates. No standard errors are reported for Cross-Moment estimates, as the method is identification-based and does not provide a standard variance estimator in its baseline formulation. Sign agreement indicates whether the DiD and Cross-Moment estimates have the same qualitative direction.

5.1 Summary of Experimental Findings

The experiments compare the Cross-Moment approach with traditional DiD estimators across seven empirical applications from five studies. The main focus of the comparison is whether both methods recover treatment effects with the same qualitative direction, rather than whether their estimated magnitudes are identical. Table 1 summarizes the results and indicates sign agreement between the two approaches.

Across the seven applications, the Cross-Moment estimator produces estimates with the same sign as the DiD benchmark in five cases. Sign agreement is observed for the Moser et al. [MV09], two Jayachandran et al. [JLMS10] mortality outcomes, Furman and Stern [FS11], and Abramitzky et al. [ADV11] applications. Disagreement occurs for the MMR outcome in Jayachandran et al. [JLMS10] and the E-ZPass pollution application of Currie and Walker [CW11]. The latter represents the largest disagreement, as the two methods imply opposite conclusions regarding the effect of E-ZPass on NO_2 levels.

The empirical evidence presented in this chapter provides partial support for the central hypothesis of this thesis. In most applications, Cross-Moment recovers the same qualitative conclusion as the DiD benchmark, suggesting robustness of causal direction across different identification strategies. However, the cases of sign disagreement demonstrate that treatment effect conclusions may remain sensitive to the assumptions underlying each estimation approach.

6 Conclusion and Further Research

This thesis analyzed the robustness of causal effect estimation under different identifying assumptions by comparing the DiD framework with a Cross-Moment-based approach across several empirical applications. The replication results show that the DiD specifications closely match the original studies in both sign and magnitude, supporting the correctness of the implementation. For the Cross-Moment approach, agreement in sign with the DiD estimates is observed in most applications, while differences in magnitude suggest sensitivity to the identification strategy. In some applications, sign differences are also observed, highlighting that the choice of identification strategy can affect substantive conclusions. Despite these differences, the main qualitative conclusions are largely consistent across methods.

These findings are consistent with the central hypothesis stated in the introduction. In particular, while the qualitative direction of effects is often robust across methods, quantitative estimates are sensitive to the underlying identifying assumptions. DiD relies on the parallel trends assumption, while the Cross-Moment approach depends on structural assumptions regarding latent confounders and proxy validity. These distinct foundations naturally lead to differences in estimated magnitudes, and in some cases also in sign.

For example, in the Jayachandran et al. (2010) application [JLMS10], the DiD and Cross-Moment estimators produce opposite signs for the MMR mortality outcome. Similarly, in the Currie and Walker (2011) application [CW11], the two approaches imply opposite directions for the effect of E-ZPass adoption on NO_2 levels. These cases demonstrate that, although the methods often agree on the qualitative direction of the treatment effect, such agreement is not guaranteed. Differences in identifying assumptions, proxy construction, or the validity of the required conditions may lead to different substantive conclusions across estimation strategies.
Overall, the evidence suggests that while the two approaches often agree on the qualitative direction of effects, this agreement is not universal, and notable exceptions exist. At the same time, even in cases of agreement in sign, quantitative estimates can differ substantially across methods.

This study has several limitations. First, although Mora and Reggio [MR12] evaluate ten empirical applications, only five of these datasets were publicly available and could therefore be included in the analysis. As a result, the comparison does not cover the full range of studies considered in the original paper.
Second, the analysis focuses on the baseline DiD specifications reported in the selected studies. Several of the original papers also present alternative estimators, robustness checks, and extended model specifications-such as sensitivity analyses, placebo tests, alternative control groups, and varying time windows-that were not included in the present comparison. The omission of these robustness checks limits the ability to assess how sensitive the results are to variations in model specification or to alternative identification assumptions. Consequently, the results should be interpreted as a comparison between the Cross-Moment approach and the primary DiD estimates rather than a comprehensive assessment across all robustness and identification analyses considered in the original studies.
Third, the evaluation primarily focuses on whether the methods agree on the direction of the estimated treatment effect. While this provides useful information about the robustness of qualitative

conclusions, it does not fully capture differences in effect magnitude or statistical precision. Finally, the Cross-Moment identification assumptions cannot be directly verified in the empirical applications, which limits the ability to assess the validity of the structural conditions underlying the estimator. In addition, proxy variables are constructed differently across datasets, which may introduce heterogeneity in the strength and quality of the information used for identification.

Further research could extend this work in several directions. A simulation study could help identify the conditions under which each method performs well and investigate how violations of their respective assumptions affect estimation accuracy. Future work could also compare the Cross-Moment approach with additional identification strategies, such as synthetic control methods or modern staggered DiD estimators. Finally, applying these methods to a larger and more diverse collection of empirical studies would provide a broader assessment of their robustness and generalizability.

References

[ADV11] Ran Abramitzky, Adeline Delavande, and Luis Vasconcelos. Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, 3(3):124–57, July 2011.

[AP09] Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

[CK94] David Card and Alan B. Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, 84(4):772–793, 1994.

[Cun21] Scott Cunningham. *Causal Inference: The Mixtape*. Yale University Press, 2021.

[CW11] Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, 3(1):65–90, January 2011.

[FS11] Jeffrey L. Furman and Scott Stern. Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, 101(5):1933–63, August 2011.

[HP01] Aapo Hyvärinen and Petteri Pajunen. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2001.

[IR15] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

[JLMS10] Seema Jayachandran, Adriana Lleras-Muney, and Kimberly V. Smith. Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs. *American Economic Journal: Applied Economics*, 2(2):118–46, April 2010.

[KSK23] Yaroslav Kivva, Saber Salehkaleybar, and Negar Kiyavash. A cross-moment approach for causal effect estimation, 2023.

[KZCS14] Hyunseung Kang, Anru Zhang, Tony T Cai, and Dylan S Small. Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American Statistical Association*, 111(513):132–144, 2014.

[Lec11] Michael Lechner. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3):165–224, 2011.

[LTTC10] Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388, 2010.

[MGTT18] Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

[MR12] Ricardo Pérez Mora and Iliana Reggio. Treatment effect identification using alternative parallel assumptions. 2012.

[MV09] Petra Moser and Alessandra Voena. Compulsory licensing - evidence from the trading with the enemy act. Working Paper 15598, National Bureau of Economic Research, December 2009.

[Pea09] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

[R+23] Jonathan Roth et al. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Economic Literature*, 2023. Working paper / survey literature on modern DiD.

[SHHK06] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

[SRTT16] Tamar Sofer, David B Richardson, and Eric J Tchetgen Tchetgen. A negative control outcome approach to difference-in-differences estimation. *Epidemiology*, 2016.

[SW20] James H. Stock and Mark W. Watson. *Introduction to Econometrics*. Pearson, 4 edition, 2020.

[TT+20] Eric J Tchetgen Tchetgen et al. An introduction to proximal causal learning. *Statistical Science*, 2020.

[Woo10] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2 edition, 2010.