Difference-in-Differences Models

Martin J. Conyon

November 2024

Abstract

This document introduces Difference-in-Differences (DiD), a foundational method in causal inference for evaluating treatment effects in observational studies. This lecture provides a rigorous yet accessible exploration of its theoretical underpinnings, assumptions, and practical applications. Particular attention is paid to the historical development of DiD, its technical implementation, and its critical role in identifying causal effects when randomized controlled trials are infeasible. These notes are based on Scott Cunningham's excellent *Causal Inference: The Mixtape* and are designed to support student learning.

While this document is a standalone resource, readers are encouraged to consult Cunningham's *Mixtape* for deeper insights. Visit the book's companion site at Scott Cunningham's Mixtape.

Contents

1 Introduction to Difference-in-Differences

Context and Purpose: Difference-in-Differences (DiD) is a widely used econometric technique designed to estimate causal effects in non-experimental settings. The core idea of DiD is straightforward: by observing how outcomes evolve over time for treated and untreated groups, and then taking the *difference of the differences* between these groups, we can estimate the causal impact of a treatment or intervention under specific assumptions. This approach is particularly effective in addressing time-invariant unobserved confounders that might otherwise bias causal inference.

DiD is a cornerstone of modern econometrics, offering a practical and interpretable framework for identifying causal effects when randomized controlled trials are impractical, unethical, or infeasible. Its applications span disciplines such as public policy, labor economics, health economics, and education, where natural experiments and policy variations often serve as quasi-experimental settings.

1.1 Historical Context and Evolution

The conceptual foundation of DiD can be traced to early natural experiments, most notably John Snow's investigation into cholera outbreaks in 19th-century London. Snow's study examined how differences in water supply systems, managed by competing companies, corresponded to cholera mortality rates. Although Snow's analysis predated the formalization of DiD, it exemplifies how comparing outcomes across groups and time could reveal causal relationships.

The modern formalization of DiD as an econometric tool emerged with advancements in policy evaluation and labor economics. A pivotal example is Card and Krueger's (1994) study on the employment effects of minimum wage increases. By comparing employment trends in New Jersey (where the minimum wage increased) and neighboring Pennsylvania (where it did not), the authors demonstrated how DiD could isolate the causal impact of

policy changes, even in observational data settings. Their work underscored the power of DiD to answer policy-relevant questions while accounting for unobservable confounding factors that remain constant over time.

1.2 Why Use Difference-in-Differences?

The primary strength of DiD lies in its ability to control for unobservable, time-invariant factors that might otherwise confound causal inference. By focusing on changes in outcomes over time rather than levels, DiD effectively subtracts away biases arising from fixed differences between treated and untreated groups. This makes it an essential tool for evaluating interventions where direct randomization is not feasible.

Key Advantages of DiD:

- **Simplicity and Transparency:** DiD relies on intuitive comparisons of before-and-after changes across groups.
- **Robustness to Unobserved Confounders:** By differencing, DiD eliminates biases from time-invariant unobservables.
- **Policy Relevance:** DiD is well-suited for evaluating natural experiments and real-world policy interventions, such as tax reforms, education programs, and health policies.
- **Scalability:** DiD can be extended to more complex settings, including multiple treatment periods, staggered interventions, and additional covariates.

1.3 Applications in Social Sciences

DiD has become a standard tool for causal analysis across a wide range of fields. Key examples include:

Public Policy: Evaluating the effects of taxation policies, welfare reforms, and public infrastructure investments. For instance, examining the impact of universal healthcare implementation on population health metrics.

Labor Economics: Analyzing the effects of labor market policies such as minimum wage laws, job training programs, and unionization on employment outcomes. Card and Krueger's seminal study is a hallmark application in this domain.

Health Economics: Studying the impact of hospital mergers, pharmaceutical pricing regulations, or Medicaid expansions on health outcomes and expenditures.

Education Economics: Assessing the impact of school reforms, voucher programs, or changes in teacher compensation on student performance.

1.4 Core Intuition: Why "Differences of Differences"?

The essence of DiD lies in isolating the causal effect of a treatment by leveraging two dimensions of variation:

1. **Time Dimension:** The difference in outcomes for the same group before and after the treatment captures the temporal change. However, this difference alone cannot account for other changes occurring over time that might also affect outcomes. 2. **Group Dimension:** The difference between treated and untreated groups at a given point in time controls for cross-sectional differences in levels. However, this comparison does not address potential trends over time.

By combining these two dimensions—taking the difference in differences—DiD eliminates biases from both time-invariant unobserved factors and group-invariant time trends, leaving an estimate of the causal treatment effect under specific assumptions.

1.5 Key Takeaways

- DiD is a powerful method for estimating causal effects in observational studies, particularly when randomized experiments are not feasible.
- Its intuitive framework relies on comparing changes in outcomes across treated and untreated groups over time.
- The method's validity hinges critically on the **parallel trends assumption**, which posits that in the absence of treatment, the treated and untreated groups would have followed the same trajectory over time.

2 TLDR: Key Insights, Results, and Roadmap

Context and Purpose: This lecture provides a comprehensive introduction to the Difference-in-Differences (DiD) method, a cornerstone of modern causal inference. DiD leverages temporal and group variation to estimate treatment effects in observational settings, addressing key challenges of unobserved confounding. These notes aim to equip graduate students with the theoretical foundations, practical applications, and limitations of DiD, ensuring they can confidently apply it in empirical research. While these notes are designed as a standalone resource, students are encouraged to explore Scott Cunningham's *Causal Inference: The Mixtape* for more in-depth discussions and examples.

What We Will Cover: This lecture systematically explores the following aspects of DiD:

- **Introduction to DiD:** The origins of the method, its key assumptions, and its role in causal inference.
- **Estimation Framework:** Detailed derivation of the DiD estimator, its link to the potential outcomes framework, and the interpretation of interaction terms in regression models.
- **Inference and Robustness:** Addressing challenges such as clustering, serial correlation, and the parallel trends assumption; testing assumptions and ensuring robust results.
- **Extensions and Limitations:** Expanding DiD to accommodate staggered adoption, multi-period settings, and heterogeneous treatment effects; addressing its inherent limitations and potential pitfalls.

Core Insights and Results

- **Difference-in-Differences Framework:** DiD estimates the causal effect of a treatment by comparing changes in outcomes over time between treated and untreated

groups. The key intuition is that, under the parallel trends assumption, any difference in pre-treatment trends can be used to infer the counterfactual post-treatment outcome for the treated group.

- **The DiD Estimator:** The basic DiD estimator can be derived as:

$$\hat{\delta} = (\overline{Y}_{\text{treated, post}} - \overline{Y}_{\text{treated, pre}}) - (\overline{Y}_{\text{control, post}} - \overline{Y}_{\text{control, pre}}),$$

where $\hat{\delta}$ represents the average treatment effect on the treated (ATT) under parallel trends.

- **Regression Formulation:** In a regression framework, DiD is estimated as the coefficient on the interaction term:

$$Y_{it} = \alpha + \beta_1 \text{treated}_i + \beta_2 \text{post}_t + \delta(\text{treated}_i \cdot \text{post}_t) + \epsilon_{it},$$

where δ is the DiD estimator.

- **Key Assumption: Parallel Trends:** The parallel trends assumption states that, in the absence of treatment, treated and untreated groups would have followed the same trajectory over time. This assumption is untestable but can be partially validated using pre-treatment data (e.g., placebo tests, event studies).

- **Inference Challenges:** Standard errors must account for clustering, serial correlation, and heteroskedasticity. Robust methods such as clustered standard errors, bootstrapping, and randomization inference are critical for valid statistical inference.

- **Extensions to DiD:** Modern extensions address staggered treatment adoption, multi-period settings, and heterogeneous treatment effects. Techniques like triple differences (DiDiD) and synthetic controls enhance its applicability in complex settings.

- **Limitations:** DiD relies heavily on the parallel trends assumption, which may not hold in practice. Additionally, spillover effects, non-stationary time trends, and small sample issues can bias estimates. Robustness checks and alternative methods are essential to mitigate these challenges.

Document Roadmap

To ensure a structured understanding, the lecture is organized as follows:

1. **Introduction to Difference-in-Differences:** An overview of the method, its historical context, and its foundational assumptions.

2. **Estimation Framework:** A rigorous derivation of the DiD estimator, including its connection to the potential outcomes framework.

3. **Inference and Robustness:** Practical considerations for inference, testing the parallel trends assumption, and addressing clustering and serial correlation.

4. **Extensions and Limitations:** Advanced methods like staggered adoption and triple differences, and a discussion of common pitfalls and solutions.

This lecture combines theoretical depth with practical relevance, preparing students to apply DiD rigorously in their own empirical research.

3 Core Assumptions of Difference-in-Differences

The validity of the Difference-in-Differences (DiD) estimator hinges on several key assumptions. These assumptions ensure that the estimated effect of treatment reflects the causal effect and is not confounded by other factors. This section provides a detailed discussion of these assumptions, their mathematical underpinnings, and the implications of potential violations.

3.1 The Parallel Trends Assumption

The cornerstone of the DiD framework is the **parallel trends assumption**. It posits that, in the absence of treatment, the average outcomes of the treated and untreated groups would have followed the same trajectory over time. Formally, let $Y_{it}(0)$ denote the potential outcome for unit i at time t under no treatment. The parallel trends assumption can be expressed as:

$$E[Y_{it}(0) \mid D_i = 1] - E[Y_{it-1}(0) \mid D_i = 1] = E[Y_{it}(0) \mid D_i = 0] - E[Y_{it-1}(0) \mid D_i = 0],$$

where D_i is an indicator for whether unit i is treated. This equality states that any changes in the untreated potential outcomes over time are the same for both the treated and untreated groups.

3.1.1 Intuition Behind Parallel Trends

The parallel trends assumption allows us to attribute differences in post-treatment changes between the treated and untreated groups to the treatment itself. In simpler terms, it posits that the only systematic difference between the treated and untreated groups after accounting for time is the causal effect of the treatment. This assumption is untestable in practice because $Y_{it}(0)$ for treated units is unobservable after treatment. However, researchers often

validate this assumption by examining pre-treatment trends in the data.

3.1.2 Graphical and Theoretical Intuition

Consider a scenario with two groups: one treated and one untreated. If the outcomes of both groups are plotted over time, the parallel trends assumption implies that, in the absence of treatment, the two groups would have exhibited the same slope in their trends over time. The treatment introduces a divergence in these slopes, which the DiD estimator captures as the causal effect.

3.1.3 Implications of Violating Parallel Trends

If the parallel trends assumption does not hold, the DiD estimator will be biased. For example, if the treated group was on a systematically different trajectory compared to the untreated group even before treatment, attributing the post-treatment difference solely to the treatment effect becomes invalid. Researchers often address this issue by:

- Including covariates to control for observed differences that might drive differential trends.
- Using pre-treatment data to test for violations of the parallel trends assumption.
- Adopting alternative methods, such as matching, to ensure that treated and untreated groups are more comparable.

3.2 Additive Separability of Errors

Another critical assumption in DiD is the **additive separability of errors**. This assumption posits that the outcome can be decomposed into components that are additively separable, as follows:

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + \epsilon_{it},$$

where:

- α_i represents unit-specific fixed effects, capturing time-invariant characteristics of unit i.
- λ_t represents time-specific fixed effects, capturing shocks or trends common to all units in period t.
- δ is the treatment effect to be estimated.
- ϵ_{it} is an idiosyncratic error term.

3.2.1 Implications of Additive Separability

The additive separability of errors ensures that unobserved factors affecting the outcome can be disentangled into time-invariant and time-varying components. This assumption is particularly crucial for identifying the treatment effect δ because it allows us to isolate the variation attributable to treatment from other confounding influences.

Violations of additive separability can occur if there are interactions between time and group-specific effects. For example, if certain unobserved characteristics of the treated group evolve differently over time compared to the untreated group, the DiD estimator may confound these interactions with the treatment effect.

3.3 Stable Unit Treatment Value Assumption (SUTVA)

The **Stable Unit Treatment Value Assumption (SUTVA)** has two key components:

1. **No Interference Between Units:** The treatment assigned to one unit does not affect the outcomes of other units. Formally, for any two units i and j, Y_i depends only on D_i and not on D_j.

2. **No Hidden Variants of Treatment:** The treatment is consistently applied across all treated units, without variations in implementation that might lead to differential effects.

3.3.1 Relevance of SUTVA to DiD

SUTVA is essential for ensuring that the treatment effect is well-defined and comparable across units. Violations of SUTVA, such as spillovers between treated and untreated units or heterogeneous treatment types, can bias the DiD estimates. For instance, if treated individuals influence untreated individuals' outcomes (e.g., through market competition or social interactions), the observed differences in outcomes may reflect these spillover effects rather than the treatment itself.

3.4 Exogeneity of Treatment Assignment

The **exogeneity of treatment assignment** requires that treatment is independent of potential outcomes after conditioning on group and time. Formally:

$$(Y_{it}(0), Y_{it}(1)) \perp D_i \mid X_{it},$$

where X_{it} represents covariates that may influence both treatment assignment and outcomes. Exogeneity ensures that any systematic differences in outcomes between treated and untreated groups are due to the treatment rather than pre-existing differences.

3.4.1 Challenges in Observational Studies

In observational settings, treatment assignment is often non-random, which complicates causal inference. Researchers typically address this challenge by:

- Including covariates to control for confounding factors.
- Using propensity score matching to balance treated and untreated groups.

- Employing instrumental variables to address endogeneity.

3.5 Consequences of Violating Assumptions

Violations of any of the above assumptions can lead to biased DiD estimates. For example:

- **Violations of Parallel Trends:** Differential trends in outcomes between treated and untreated groups introduce bias.
- **Violations of Additive Separability:** Non-additive interactions between unit and time effects confound the treatment effect.
- **Violations of SUTVA:** Spillovers or heterogeneous treatments distort the interpretation of causal effects.
- **Violations of Exogeneity:** Non-random treatment assignment leads to selection bias.

3.6 Summary of Assumptions

The assumptions underpinning DiD are critical for its validity and interpretability. Table 2 summarizes these assumptions and their implications.

Table 1: Core Assumptions in Difference-in-Differences

Assumption	**Implication**
Parallel Trends	Ensures that untreated potential outcomes evolve similarly for treated and untreated groups.
Additive Separability of Errors	Allows decomposition of outcomes into fixed and time-varying components.
Stable Unit Treatment Value Assumption (SUTVA)	Prevents interference and heterogeneity in treatment implementation.
Exogeneity of Treatment Assignment	Ensures differences in outcomes are attributable to the treatment.

4 Core Assumptions of Difference-in-Differences

The validity of the Difference-in-Differences (DiD) estimator hinges on several key assumptions. These assumptions ensure that the estimated effect of treatment reflects the causal effect and is not confounded by other factors. This section provides a detailed discussion of these assumptions, their mathematical underpinnings, and the implications of potential violations.

4.1 The Parallel Trends Assumption

The cornerstone of the DiD framework is the **parallel trends assumption**. It posits that, in the absence of treatment, the average outcomes of the treated and untreated groups would have followed the same trajectory over time. Formally, let $Y_{it}(0)$ denote the potential outcome for unit i at time t under no treatment. The parallel trends assumption can be expressed as:

$$E[Y_{it}(0) \mid D_i = 1] - E[Y_{it-1}(0) \mid D_i = 1] = E[Y_{it}(0) \mid D_i = 0] - E[Y_{it-1}(0) \mid D_i = 0],$$

where D_i is an indicator for whether unit i is treated. This equality states that any changes in the untreated potential outcomes over time are the same for both the treated and untreated groups.

4.1.1 Intuition Behind Parallel Trends

The parallel trends assumption allows us to attribute differences in post-treatment changes between the treated and untreated groups to the treatment itself. In simpler terms, it posits that the only systematic difference between the treated and untreated groups after accounting for time is the causal effect of the treatment. This assumption is untestable in practice because $Y_{it}(0)$ for treated units is unobservable after treatment. However, researchers often

validate this assumption by examining pre-treatment trends in the data.

4.1.2 Graphical and Theoretical Intuition

Consider a scenario with two groups: one treated and one untreated. If the outcomes of both groups are plotted over time, the parallel trends assumption implies that, in the absence of treatment, the two groups would have exhibited the same slope in their trends over time. The treatment introduces a divergence in these slopes, which the DiD estimator captures as the causal effect.

4.1.3 Implications of Violating Parallel Trends

If the parallel trends assumption does not hold, the DiD estimator will be biased. For example, if the treated group was on a systematically different trajectory compared to the untreated group even before treatment, attributing the post-treatment difference solely to the treatment effect becomes invalid. Researchers often address this issue by:

- Including covariates to control for observed differences that might drive differential trends.
- Using pre-treatment data to test for violations of the parallel trends assumption.
- Adopting alternative methods, such as matching, to ensure that treated and untreated groups are more comparable.

4.2 Additive Separability of Errors

Another critical assumption in DiD is the **additive separability of errors**. This assumption posits that the outcome can be decomposed into components that are additively separable, as follows:

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + \epsilon_{it},$$

where:

- α_i represents unit-specific fixed effects, capturing time-invariant characteristics of unit i.
- λ_t represents time-specific fixed effects, capturing shocks or trends common to all units in period t.
- δ is the treatment effect to be estimated.
- ϵ_{it} is an idiosyncratic error term.

4.2.1 Implications of Additive Separability

The additive separability of errors ensures that unobserved factors affecting the outcome can be disentangled into time-invariant and time-varying components. This assumption is particularly crucial for identifying the treatment effect δ because it allows us to isolate the variation attributable to treatment from other confounding influences.

Violations of additive separability can occur if there are interactions between time and group-specific effects. For example, if certain unobserved characteristics of the treated group evolve differently over time compared to the untreated group, the DiD estimator may confound these interactions with the treatment effect.

4.3 Stable Unit Treatment Value Assumption (SUTVA)

The **Stable Unit Treatment Value Assumption (SUTVA)** has two key components:

1. **No Interference Between Units:** The treatment assigned to one unit does not affect the outcomes of other units. Formally, for any two units i and j, Y_i depends only on D_i and not on D_j.

2. **No Hidden Variants of Treatment:** The treatment is consistently applied across all treated units, without variations in implementation that might lead to differential effects.

4.3.1 Relevance of SUTVA to DiD

SUTVA is essential for ensuring that the treatment effect is well-defined and comparable across units. Violations of SUTVA, such as spillovers between treated and untreated units or heterogeneous treatment types, can bias the DiD estimates. For instance, if treated individuals influence untreated individuals' outcomes (e.g., through market competition or social interactions), the observed differences in outcomes may reflect these spillover effects rather than the treatment itself.

4.4 Exogeneity of Treatment Assignment

The **exogeneity of treatment assignment** requires that treatment is independent of potential outcomes after conditioning on group and time. Formally:

$$(Y_{it}(0), Y_{it}(1)) \perp D_i \mid X_{it},$$

where X_{it} represents covariates that may influence both treatment assignment and outcomes. Exogeneity ensures that any systematic differences in outcomes between treated and untreated groups are due to the treatment rather than pre-existing differences.

4.4.1 Challenges in Observational Studies

In observational settings, treatment assignment is often non-random, which complicates causal inference. Researchers typically address this challenge by:

- Including covariates to control for confounding factors.
- Using propensity score matching to balance treated and untreated groups.

- Employing instrumental variables to address endogeneity.

4.5 Consequences of Violating Assumptions

Violations of any of the above assumptions can lead to biased DiD estimates. For example:

- **Violations of Parallel Trends:** Differential trends in outcomes between treated and untreated groups introduce bias.
- **Violations of Additive Separability:** Non-additive interactions between unit and time effects confound the treatment effect.
- **Violations of SUTVA:** Spillovers or heterogeneous treatments distort the interpretation of causal effects.
- **Violations of Exogeneity:** Non-random treatment assignment leads to selection bias.

4.6 Summary of Assumptions

The assumptions underpinning DiD are critical for its validity and interpretability. Table 2 summarizes these assumptions and their implications.

Table 2: Core Assumptions in Difference-in-Differences

Assumption	**Implication**
Parallel Trends	Ensures that untreated potential outcomes evolve similarly for treated and untreated groups.
Additive Separability of Errors	Allows decomposition of outcomes into fixed and time-varying components.
Stable Unit Treatment Value Assumption (SUTVA)	Prevents interference and heterogeneity in treatment implementation.
Exogeneity of Treatment Assignment	Ensures differences in outcomes are attributable to the treatment.

5 Estimation of Difference-in-Differences

Difference-in-Differences (DiD) is a method rooted in the Potential Outcomes Framework, allowing researchers to estimate causal effects in observational settings. By leveraging variation across groups and over time, DiD isolates the impact of a treatment or intervention while controlling for time-invariant unobserved heterogeneity. This section explores the theoretical foundation of DiD, its relationship to potential outcomes, and its practical implementation. Extensions to more complex scenarios are also discussed.

5.1 DiD and the Potential Outcomes Framework

In the Potential Outcomes Framework, each unit i at time t has two potential outcomes:

- $Y_{it}(1)$: The outcome if the unit is treated.
- $Y_{it}(0)$: The outcome if the unit is not treated.

The observed outcome for unit i at time t can then be written as:

$$Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0),$$

where D_{it} is an indicator equal to 1 if the unit is treated at time t and 0 otherwise. The goal of causal inference is to estimate the **treatment effect** for unit i at time t, defined as:

$$\delta_{it} = Y_{it}(1) - Y_{it}(0).$$

However, the fundamental problem of causal inference is that we observe either $Y_{it}(1)$ or $Y_{it}(0)$, but never both for the same unit at the same time. DiD addresses this problem by exploiting group-level and temporal variation to construct a counterfactual outcome.

5.1.1 Constructing the Counterfactual with DiD

DiD assumes the treated group's post-treatment outcomes can be compared to what their outcomes would have been in the absence of treatment, using the untreated group as a counterfactual. Under the parallel trends assumption (discussed in Section 2), the change in outcomes for the untreated group over time approximates the change that would have occurred for the treated group in the absence of treatment. This allows us to estimate the average treatment effect on the treated (ATT) as:

$$\delta = (E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]) - (E[Y_{it} \mid D_i = 0, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 0]),$$

where:

- $D_i = 1$ indicates the treated group.
- $D_i = 0$ indicates the untreated group.
- $T_t = 1$ indicates the post-treatment period.
- $T_t = 0$ indicates the pre-treatment period.

The terms in parentheses represent the changes in outcomes for the treated and untreated groups, respectively, over time.

5.2 Mathematical Framework and the DiD Estimator

The standard DiD model can be expressed as:

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \delta (D_i \cdot T_t) + \epsilon_{it},$$

where:

- α: Baseline outcome for the untreated group in the pre-treatment period.

- β: Difference in baseline outcomes between the treated and untreated groups.
- γ: Common change in outcomes over time for both groups, capturing time trends.
- δ: The DiD estimate, measuring the causal effect of the treatment.
- ϵ_{it}: Idiosyncratic error term.

The coefficient δ on the interaction term $D_i \cdot T_t$ captures the difference in differences and represents the treatment effect under the parallel trends assumption.

5.2.1 Deriving the DiD Estimator

Let us break down the derivation of δ using group-level means. Define:

- $\bar{Y}_{\text{treated, post}}$: Mean outcome for the treated group in the post-treatment period.
- $\bar{Y}_{\text{treated, pre}}$: Mean outcome for the treated group in the pre-treatment period.
- $\bar{Y}_{\text{untreated, post}}$: Mean outcome for the untreated group in the post-treatment period.
- $\bar{Y}_{\text{untreated, pre}}$: Mean outcome for the untreated group in the pre-treatment period.

The DiD estimator is:

$$\hat{\delta} = (\bar{Y}_{\text{treated, post}} - \bar{Y}_{\text{treated, pre}}) - (\bar{Y}_{\text{untreated, post}} - \bar{Y}_{\text{untreated, pre}}).$$

This formulation explicitly separates the treatment effect δ from other sources of variation, including baseline differences between groups (β) and common time trends (γ).

5.2.2 Connection to Potential Outcomes

In terms of potential outcomes, the DiD estimator approximates:

$$\delta = E[Y_{it}(1) - Y_{it}(0) \mid D_i = 1].$$

This corresponds to the average treatment effect on the treated (ATT), as the treated group is the focus of the causal estimation.

5.3 Practical Implementation of DiD

To implement DiD in practice, researchers follow these steps:

1. **Define Treated and Untreated Groups:** Identify units exposed to treatment and those not exposed. The untreated group serves as the counterfactual.

2. **Split Observations into Pre- and Post-Treatment Periods:** Define the time period before and after the treatment is applied.

3. **Compute Group-Level Averages:** Calculate average outcomes for treated and untreated groups in each time period.

4. **Apply the DiD Formula:** Use:

$$\hat{\delta} = (\bar{Y}_{\text{treated, post}} - \bar{Y}_{\text{treated, pre}}) - (\bar{Y}_{\text{untreated, post}} - \bar{Y}_{\text{untreated, pre}}).$$

5. **Estimate Using Regression:** To incorporate covariates or account for additional heterogeneity, estimate the regression model:

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \delta (D_i \cdot T_t) + \epsilon_{it}.$$

5.4 Extensions to Multi-Period and Staggered Treatments

5.4.1 Multi-Period DiD

When there are multiple time periods, the standard DiD model can be generalized to:

$$Y_{it} = \alpha_i + \lambda_t + \delta_t D_{it} + \epsilon_{it},$$

where:

- α_i: Unit-specific fixed effects, capturing time-invariant characteristics.
- λ_t: Time-specific fixed effects, capturing common shocks.
- δ_t: Treatment effect at time t.

This extension allows for dynamic treatment effects, capturing how the impact of treatment evolves over time.

5.4.2 Staggered Treatment Timing

In settings where treatment is introduced at different times for different units, a staggered DiD approach is required. Researchers often use event-study designs to estimate treatment effects relative to the timing of treatment, with the model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{K} \delta_k D_{i,t-k} + \epsilon_{it},$$

where δ_k measures the treatment effect k periods before or after treatment.

5.5 Limitations and Challenges

While DiD is a powerful method, it is subject to limitations:

- **Violation of Parallel Trends:** Differential trends between groups bias the estimate of δ.
- **Heterogeneous Treatment Effects:** If treatment effects vary across units or time, the interpretation of δ as a single effect becomes less meaningful.
- **Time-Varying Confounders:** Unobserved factors that evolve differently across groups can confound the results.

Researchers must rigorously test assumptions, often using robustness checks such as pre-treatment trend validation or placebo tests.

5.6 Conclusion

The DiD framework provides a theoretically grounded and practical approach to causal inference in observational studies. By leveraging differences across groups and time, DiD isolates treatment effects while addressing many sources of bias. However, its validity depends critically on assumptions such as parallel trends, which require careful evaluation in empirical applications.

6 Inference and Robustness in Difference-in-Differences

While the Difference-in-Differences (DiD) framework provides a powerful tool for estimating causal effects, valid statistical inference and robustness are critical for ensuring that the results are credible and reliable. This section addresses the challenges of inference in DiD settings, discusses methods for testing key assumptions, and outlines strategies for conducting robustness checks.

6.1 Statistical Inference in DiD

DiD estimates rely on sample data to approximate the true treatment effect δ. To make valid inferences, it is crucial to account for the specific structure of the data, particularly serial correlation, clustering, and heteroskedasticity.

6.1.1 Clustering and Serial Correlation

In most DiD applications, data are structured into groups (e.g., regions, firms) observed over time. This structure often introduces correlation within groups, both cross-sectionally and over time. Standard inference methods assuming independence across observations can severely underestimate standard errors, leading to overconfident (anti-conservative) statistical tests.

Clustering Standard Errors: Clustered standard errors account for within-group correlation by allowing the error term ϵ_{it} to be arbitrarily correlated across units within a group. For example, in a two-way fixed effects DiD model:

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + \epsilon_{it},$$

where α_i are unit fixed effects and λ_t are time fixed effects, the error term ϵ_{it} is typically clustered at the group level. The variance of $\hat{\delta}$ is computed as:

$$Var(\hat{\delta}) = \frac{\sigma^2}{n_{\text{treated}}} + \frac{\sigma^2}{n_{\text{untreated}}},$$

where σ^2 accounts for within-group correlation and $n_{\text{treated}}, n_{\text{untreated}}$ are the group sizes.

Serial Correlation: Serial correlation arises when the error term for a unit i at time t is correlated with its error term at time $t-1$. This is particularly problematic in settings with many time periods, as it can lead to underestimated standard errors. Bertrand, Duflo, and Mullainathan (2004) showed that failing to account for serial correlation can result in inflated t-statistics and spurious findings of statistical significance.

Solutions:

- **Clustered Standard Errors:** Cluster at the group level to account for both cross-sectional and serial correlation. This is the default approach in most modern econometric software.
- **Randomization Inference:** When the number of clusters is small, randomization inference can provide exact p-values under the null hypothesis.
- **Bootstrap Methods:** Non-parametric bootstrapping can be used to generate valid standard errors, particularly in the presence of heteroskedasticity.

6.1.2 Small Sample Issues

When the number of treated clusters is small, conventional clustered standard errors can be biased. Cameron, Gelbach, and Miller (2008) recommend using wild bootstrap methods to improve inference in such settings.

6.2 Testing the Parallel Trends Assumption

The validity of the DiD estimator depends on the parallel trends assumption, which states that in the absence of treatment, the treated and untreated groups would have followed the same trajectory over time. This assumption cannot be directly tested but can be assessed using pre-treatment data.

6.2.1 Pre-Treatment Placebo Tests

A common approach to validate parallel trends is to estimate DiD coefficients in pre-treatment periods, where no treatment effect should exist. Let:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{-1} \delta_k D_{i,t+k} + \epsilon_{it},$$

where δ_k captures the "placebo effect" k periods before treatment. If the parallel trends assumption holds, we expect $\delta_k \approx 0$ for all $k < 0$.

6.2.2 Event Studies

Event study designs allow researchers to visually and statistically examine trends before and after treatment. The model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{K} \delta_k D_{i,t+k} + \epsilon_{it},$$

estimates δ_k, the effect of treatment k periods relative to treatment initiation. If the parallel trends assumption holds, pre-treatment coefficients (δ_k for $k < 0$) should be statistically indistinguishable from zero.

6.3 Robustness Checks

Robustness checks evaluate the sensitivity of DiD estimates to key assumptions and model specifications. These include:

6.3.1 Placebo Tests

Placebo tests introduce a hypothetical treatment period prior to the actual treatment. If DiD estimates remain significant for this false period, the observed effects may be driven by spurious trends rather than the treatment.

6.3.2 Alternative Control Groups

Using different untreated groups as controls can help ensure the results are not driven by the specific choice of comparison group. For example, if DiD estimates vary significantly across control groups, the treated and untreated groups may not be comparable.

6.3.3 Covariate Adjustment

Including time-varying covariates in the regression model can improve robustness by accounting for confounding factors that vary over time. The extended model becomes:

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \delta(D_i \cdot T_t) + \mathbf{X}_{it}'\boldsymbol{\theta} + \epsilon_{it},$$

where $\mathbf{X}_{it}$ represents a vector of covariates.

6.4 Extensions for Improved Inference

6.4.1 Triple Differences (DiDiD)

The triple difference (DiDiD) method extends DiD by introducing a third dimension of variation, such as an additional control group or time period. The DiDiD estimator is:

$$\Delta^3 = (\Delta_{\text{treated}} - \Delta_{\text{control}})_{\text{post}} - (\Delta_{\text{treated}} - \Delta_{\text{control}})_{\text{pre}} ,$$

where Δ denotes a difference in means. This method is particularly useful for addressing heterogeneity in treatment effects or violations of the parallel trends assumption.

6.4.2 Synthetic Control Methods

When untreated units are not comparable to treated units, synthetic control methods construct a weighted combination of untreated units to create a more suitable counterfactual. This approach is particularly useful for cases with staggered treatment timing or limited data.

6.5 Conclusion

Inference and robustness are essential components of the DiD framework. By addressing challenges such as clustering, serial correlation, and violations of key assumptions, researchers can ensure their estimates are both statistically valid and substantively credible. Robustness checks and extensions, such as triple differences and synthetic controls, further enhance the reliability of DiD analyses, making them an indispensable tool in causal inference.

7 Extensions and Limitations of Difference-in-Differences

While the Difference-in-Differences (DiD) framework provides a powerful and intuitive approach to causal inference, it is not without its limitations. Over time, researchers have developed extensions to address its challenges and broaden its applicability. This section explores key extensions of the DiD method, particularly for multi-period and staggered treatment settings, and highlights its limitations.

7.1 Extensions of Difference-in-Differences

Traditional DiD assumes a binary treatment status, a single treatment period, and parallel trends between treated and untreated groups. However, many empirical settings involve more complex scenarios, necessitating extensions to the standard framework.

7.1.1 Multi-Treatment and Multi-Period DiD

In settings where treatment effects vary across time or units, the basic two-group, two-period framework becomes inadequate. The generalized DiD model introduces flexibility by allowing for multiple treatment groups and time periods. The model can be written as:

$$Y_{it} = \alpha_i + \lambda_t + \delta_t D_{it} + \epsilon_{it},$$

where:

- α_i are unit-specific fixed effects, capturing time-invariant characteristics of each unit.
- λ_t are time-specific fixed effects, capturing common shocks or trends affecting all units in period t.
- δ_t is a time-varying treatment effect, capturing how the impact of treatment evolves over time.

This framework is particularly useful for estimating dynamic treatment effects. For example, it can capture whether the impact of treatment increases, decreases, or remains constant as time progresses after treatment.

Event Study Designs: Event study designs are a popular application of multi-period DiD models. The event study specification can be written as:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{k=-K}^{K} \delta_k D_{i,t-k} + \epsilon_{it},$$

where δ_k measures the treatment effect k periods relative to the time of treatment. This approach allows researchers to visualize and test the timing and persistence of treatment effects. Pre-treatment coefficients (δ_k for $k < 0$) also serve as a test of the parallel trends assumption, as they should not deviate significantly from zero under the assumption.

7.1.2 Staggered Treatment Adoption

In many policy evaluations, treatment is implemented at different times for different groups. The standard DiD framework does not directly accommodate staggered treatment timing because untreated units at one time may eventually receive treatment. Recent advancements, such as those by Callaway and Sant'Anna (2021), propose methods to estimate group-time average treatment effects (ATTs) under staggered adoption. The staggered treatment model can be written as:

$$Y_{it} = \alpha_i + \lambda_t + \delta_{g,t} D_{g,t} + \epsilon_{it},$$

where $\delta_{g,t}$ represents the treatment effect for group g at time t.

Heterogeneity in Treatment Effects: Staggered adoption models account for heterogeneity in treatment effects by explicitly estimating effects for each group and time period. The overall treatment effect is then aggregated across groups and periods, providing a more nuanced understanding of the intervention.

7.1.3 Triple Differences (DiDiD)

Triple Differences (DiDiD) extends the DiD framework by adding a third dimension of variation, such as geography, demographics, or policy type. The DiDiD model can be expressed as:

$$Y_{ijt} = \alpha + \beta_1 D_i + \beta_2 S_j + \beta_3 T_t + \delta(D_i \cdot S_j \cdot T_t) + \epsilon_{ijt},$$

where:

- S_j is an indicator for the second level of variation (e.g., state or sector).
- $D_i \cdot S_j \cdot T_t$ is the triple interaction term, capturing the causal effect of treatment.

Triple differences are particularly useful in settings where parallel trends are likely violated in the original DiD framework. By introducing an additional comparison group, the DiDiD model adjusts for differential trends across groups or regions.

7.2 Limitations of Difference-in-Differences

Despite its versatility, the DiD framework has several important limitations that researchers must carefully consider.

7.2.1 Violation of the Parallel Trends Assumption

The validity of the DiD estimator hinges on the parallel trends assumption. If treated and untreated groups exhibit systematically different trends in the absence of treatment, the estimator will be biased. Potential violations can arise from:

- Pre-existing differences in growth rates between groups.
- Structural changes affecting one group disproportionately (e.g., economic shocks).

Addressing Violations: Researchers often address violations of parallel trends by:

- Including group-specific time trends in the regression model:

$$Y_{it} = \alpha_i + \lambda_t + \delta(D_i \cdot T_t) + \beta T_t \cdot D_i + \epsilon_{it}.$$

- Using synthetic control methods to construct a more suitable counterfactual.

7.2.2 Heterogeneous Treatment Effects

When treatment effects vary across units or over time, interpreting the DiD estimate as a single causal effect becomes problematic. The DiD framework assumes constant treatment effects across units and time, which may not hold in practice. For example, policy interventions may have larger effects in urban areas compared to rural areas or among different demographic groups.

Solutions: Extensions such as staggered adoption models and triple differences can explicitly account for treatment effect heterogeneity. Additionally, researchers can present disaggregated results to illustrate variation in effects across subgroups.

7.2.3 Non-Stationary Time Trends

If untreated units experience diverging trends due to structural changes unrelated to treatment, the parallel trends assumption is violated. For example, regions undergoing rapid economic development may exhibit upward trends in outcomes regardless of treatment.

Solutions: Including covariates to control for time-varying confounders or using synthetic control methods can mitigate biases from non-stationary trends.

7.2.4 Spillover Effects and Interference

The Stable Unit Treatment Value Assumption (SUTVA) assumes no interference between units. However, spillover effects are common in practice. For example, a policy implemented in one region may affect neighboring regions through labor market dynamics or trade. Such spillovers violate SUTVA and bias the DiD estimate.

Solutions: Spillover effects can be addressed using spatial econometric models or by explicitly modeling the spillover mechanism.

7.2.5 Small Sample Issues

When the number of treated or untreated clusters is small, standard inference methods for DiD may produce biased estimates. Clustered standard errors, in particular, can underperform in small samples.

Solutions: Wild bootstrap methods and randomization inference can improve inference in small-sample settings.

7.3 Future Directions and Advanced Techniques

Advancements in causal inference are addressing the limitations of traditional DiD. These include:

- **Synthetic Control Methods:** A flexible alternative to DiD, particularly when untreated units are not comparable to treated units.
- **Machine Learning Integration:** Combining DiD with machine learning to model complex heterogeneity and improve counterfactual predictions.
- **Doubly Robust DiD:** Methods that combine propensity score weighting and regression adjustments for more robust causal inference.

7.4 Conclusion

Difference-in-Differences is a foundational tool in applied econometrics, offering a robust framework for causal inference in observational settings. While its extensions enhance its versatility, researchers must remain vigilant about its assumptions and limitations, employing robustness checks and alternative methods where necessary. These advancements ensure that DiD remains a vital method for understanding the impacts of policies and interventions.

8 Implementing DiD in Stata 18: A Cookbook Approach

In this section, we will walk through a series of practical steps to implement the Difference-in-Differences (DiD) method using Stata 18. The examples and code are based on the Stata 18 manual, which provides comprehensive guidance on the syntax and usage of the `didregress` and `xtdidregress` commands for estimating causal treatment effects.

The following examples demonstrate how to estimate the Average Treatment Effect on the Treated (ATET) using DiD, including setting up the data, running the regressions, and performing diagnostics. All of the Stata commands provided are designed for immediate use, ensuring students can replicate these analyses in their own environments.

8.1 Basic Syntax of DiD Commands

The `didregress` command is used for repeated cross-sectional data, while the `xtdidregress` command is used for panel data. Here is the basic syntax for each:

```
didregress (outcome var) (treatment var), group(groupvar) time(timevar)
xtdidregress (outcome var) (treatment var), group(groupvar) time(timevar)
```

Where:

- `outcome var` is the dependent variable (e.g., satisfaction score, wages).
- `treatment var` is the binary treatment indicator (1 if the group is treated, 0 otherwise).
- `groupvar` indicates the group structure (e.g., individual, county, hospital).
- `timevar` is the time variable (e.g., pre- and post-treatment periods, year, month).

In the following examples, we will use the `didregress` command for repeated cross-sectional data.

8.2 Example 1: Health Provider Study

Let's consider a study on a health provider that wants to estimate the effect of a new admissions procedure on patient satisfaction. The data includes monthly observations from January to July, and the new procedure was implemented in April. We will treat hospitals that implemented the procedure in April as the treatment group.

To begin, we load the data and set up the environment in Stata:

```
use https://www.stata-press.com/data/r18/hospdd, clear
```

This dataset contains information on hospitals and patient satisfaction. The variable `satis` represents the satisfaction score, and the variable `procedure` indicates whether the new procedure was used (1 if used after April, 0 if used before). The time variable is `month`, which records the month of the observation.

Next, we estimate the Average Treatment Effect on the Treated (ATET) using the `didregress` command:

```
didregress (satis) (procedure), group(hospital) time(month)
```

This command estimates the ATET by comparing the changes in satisfaction between the treatment group (hospitals that implemented the procedure) and the control group (hospitals that did not). It also includes group and time fixed effects.

8.2.1 Interpreting the Results

The output will provide the coefficient for the treatment variable, which represents the estimated treatment effect. In this example, you might see something like:

```
Treatment and time information
Time variable: month
Control: procedure = 0
Treatment: procedure = 1
Control Treatment
Group
hospital 28 18
Time
Minimum 1 4
Maximum 1 4

Difference-in-differences regression Number of obs = 7,368
Data type: Repeated cross-sectional
(Std. err. adjusted for 46 clusters in hospital)
Robust

satis           Coefficient    Std. err.    t    P>|t|    [95% conf. interval]
ATET
procedure (New vs Old)   0.8479      0.0321    26.41   0.000    [0.7833, 0.9127]
```

The ATET estimate is 0.85, meaning the new procedure increased patient satisfaction by approximately 0.85 points on average, relative to the control group.

8.3 Graphical Diagnostics: Parallel Trends

Before interpreting the DiD results, it is crucial to verify the *parallel trends assumption*, which suggests that the treated and control groups would have followed similar trends over time if the treatment had not been applied. To check this assumption, we can plot the mean satisfaction scores for both groups before and after the treatment:

```
estat trendplots
```

This command produces a graph showing the observed means over time for both the treatment and control groups, allowing you to visually inspect whether the groups followed parallel trends before the treatment. If the trends are parallel, this suggests that the estimated treatment effect can be interpreted causally.

Additionally, you can run a formal test for parallel trends using the following command:

```
estat ptrends
```

This test checks whether the pre-treatment trends for both groups are parallel. If the test result indicates that the trends were parallel, you can proceed with interpreting the treatment effect confidently.

8.4 Example 2: DDD (Difference-in-Differences-in-Differences) Model

Sometimes, there are additional confounding factors that need to be addressed. For example, if the treatment effect is suspected to vary across subgroups (e.g., based on patient visit frequency), a Difference-in-Differences-in-Differences (DDD) model may be more appropriate. In this case, we modify the treatment definition to account for the frequency of hospital visits.

First, we create a new variable `hightrt`, which identifies patients with high or very high hospital visit frequency and who were treated after the new procedure was implemented:

```
generate hightrt = procedure == 1 & (frequency == 3 | frequency == 4)
label define trt 0 "Untreated" 1 "Treated"
label values hightrt trt
```

Now, we can estimate the treatment effect using a triple difference model by including both hospital and frequency effects:

```
didregress (satis) (hightrt), group(hospital frequency) time(month)
```

This model accounts for both hospital-level effects and the interaction with frequency of hospital visits. It adjusts for the possibility that the treatment effect might differ based on patient behavior or characteristics.

8.5 Conclusion

In this section, we have covered the basics of implementing Difference-in-Differences (DiD) using Stata 18. We demonstrated how to run DiD regressions, check the parallel trends assumption, and address potential confounders using a DDD model. The examples provided, taken directly from the Stata 18 manual, are designed to help students understand both the theory and practice of DiD analysis in Stata. By following these steps, you can replicate the analyses in your own data and apply DiD methods to estimate causal treatment effects.

9 Advanced Techniques in DiD: Aggregation, Wild Bootstrap, and Triple Differences (DDD)

In this section, we will explore some advanced techniques used to improve the robustness and flexibility of Difference-in-Differences (DiD) analyses. These techniques include data aggregation methods, wild bootstrap for robust standard errors, and Triple Differences (DDD) for addressing more complex confounding factors. All of these methods are crucial for ensuring accurate causal inference when working with DiD models in Stata 18.

9.1 Data Aggregation in DiD

When working with DiD, particularly with small groups or few observations within each group, it is often beneficial to aggregate the data to a higher level. This technique can reduce the dimensionality of the model and mitigate the risk of unreliable estimates due to small sample sizes.

The `didregress` command in Stata allows for data aggregation by specifying the `aggregate()` option. There are different aggregation methods available, such as the standard aggregation and the Donald and Lang (2007) method. Below we walk through the aggregation process.

9.1.1 Standard Aggregation

In the standard aggregation method, data is aggregated at the group and time level. The following command demonstrates how to perform standard aggregation using the `aggregate()` option in Stata:

```
didregress (satis) (procedure), group(hospital) time(month) aggregate(standard)
```

This command aggregates the data at the hospital and month levels. The standard aggregation method first estimates group-time level effects and then aggregates the remaining covariates. The final regression is estimated on the aggregated data, using cluster-robust standard errors.

9.1.2 Donald and Lang (2007) Aggregation

The Donald and Lang (2007) aggregation method provides an alternative to the standard method. It allows for bias-corrected standard errors in the presence of small group sizes. To apply this method, use the `aggregate(dlang)` option:

```
didregress (satis) (procedure), group(hospital) time(month) aggregate(dlang)
```

This aggregation technique is particularly useful when you have a small number of clusters and wish to account for potential heterogeneity across groups. By aggregating the data in this way, you obtain more reliable estimates of the treatment effect and its standard errors.

9.2 Wild Bootstrap for Robust Standard Errors

In situations where there are a small number of groups, standard errors can be biased, especially when the number of groups is less than 20. One solution is to use the wild bootstrap, which adjusts for the potential bias in cluster-robust standard errors. The wild bootstrap generates confidence intervals and p-values by resampling residuals from the regression model.

9.2.1 Applying Wild Bootstrap

To apply the wild bootstrap in Stata, you can use the `wildbootstrap` option with the `didregress` or `xtdidregress` commands. For example, the following command computes wild bootstrap p-values and confidence intervals:

```
didregress (satis) (procedure), group(hospital) time(month) wildbootstrap
```

This command performs the wild bootstrap by resampling residuals, adjusting the inference based on the resampled data. You can further specify the error weight type (e.g., `rademacher`, `mammen`, `webb`, etc.) using the `errorweight()` option:

```
didregress (satis) (procedure), group(hospital) time(month) wildbootstrap(errorweight(ra
```

The wild bootstrap is especially useful in cases where traditional cluster-robust standard errors may not be sufficient, particularly with a small number of clusters.

9.3 Triple Differences (DDD) Model

The Triple Difference (DDD) approach extends the traditional Difference-in-Differences (DiD) method by adding an additional dimension of difference. This can be useful when you suspect that the treatment effect differs across more than one dimension, such as time and group characteristics.

For instance, imagine a study where we not only want to compare pre- and post-treatment periods across treated and control groups but also want to account for a third variable (e.g., visit frequency) that might influence the treatment effect. In this case, we can estimate the Triple Difference (DDD) by including this third dimension.

9.3.1 Implementing a DDD Model in Stata

Let's extend our previous health provider study example to estimate the treatment effect on a subset of patients who have high or very high visit frequencies. We create a new variable `hightrt` to identify treated patients based on their visit frequency.

First, we generate the `hightrt` variable:

```
generate hightrt = procedure == 1 & (frequency == 3 | frequency == 4)
label define trt 0 "Untreated" 1 "Treated"
label values hightrt trt
```

Now, we estimate the Triple Difference (DDD) by including both the hospital and frequency of visits as group variables. This model accounts for both the group-level effects and the interaction with time effects:

```
didregress (satis) (hightrt), group(hospital frequency) time(month)
```

This command estimates the ATET by comparing the treatment and control groups with the additional layer of distinction based on visit frequency. This allows for a more nuanced understanding of the treatment effect, as it controls for potential heterogeneity in treatment effects across different subgroups.

9.4 Conclusion

In this section, we discussed advanced techniques for improving the robustness of DiD estimates, including data aggregation methods, the wild bootstrap, and the Triple Differences (DDD) approach. Each of these methods enhances the accuracy and reliability of DiD analysis, particularly in the presence of small groups or complex confounding factors.

Students can apply these techniques by simply modifying their Stata commands, as shown in the examples provided. The aggregation methods help to address small sample sizes, while the wild bootstrap ensures robust standard errors. Meanwhile, the DDD approach offers a more detailed treatment of heterogeneous effects, accounting for multiple dimensions of variation.

These advanced techniques are essential for robust causal inference in DiD models and should be considered whenever you face complex data or small group sizes.

10 Implementing DiD with Panel Data in Stata

10.1 Overview of Panel Data DiD

The ‘xtdidregress‘ command in Stata is designed for panel data analysis, where the data consist of repeated observations for the same units (e.g., individuals, companies, hospitals) over time. This command is particularly useful when we have longitudinal data and we want to estimate the treatment effect using the Difference-in-Differences (DiD) approach, while accounting for both individual and time effects.

The main difference between ‘xtdidregress‘ and the ‘didregress‘ command (used for repeated cross-sectional data) lies in the structure of the data. In panel data, the same units are observed over multiple periods, and we need to account for within-unit correlation over time, which ‘xtdidregress‘ does automatically by using fixed effects.

10.2 Step-by-Step Guide for Implementing DiD with Panel Data

10.2.1 Command Syntax

The basic syntax for the ‘xtdidregress‘ command is as follows:

xtdidregress (outcome) (treatment), group(groupvar) time(timevar)

Where: - `outcome` is the dependent variable (e.g., patient satisfaction, patents granted, etc.). - `treatment` is the binary treatment indicator (1 if the unit is treated after the intervention, 0 otherwise). - `groupvar` specifies the variable that indicates the group structure (e.g., individual, county, firm). - `timevar` is the time variable (e.g., year, month).

This command runs a Difference-in-Differences analysis for panel data with fixed effects for both individuals (or units) and time periods. The fixed effects account for unobserved heterogeneity in both dimensions.

10.2.2 Example 2: Patent Innovation Study

In this example, we will explore the effect of compulsory licensing during World War I on domestic inventions. We will use panel data that tracks the number of patents granted in different subclasses of the chemical industry over several years. The data is structured with patents granted by different subclasses and years.

We begin by loading the data and setting up the panel structure:

```
use https://www.stata-press.com/data/r18/patents, clear
xtset classid year
```

This command loads the dataset and sets the panel structure with `classid` as the panel variable (representing the subclass) and `year` as the time variable.

Next, we estimate the treatment effect of compulsory licensing on domestic patents granted after 1918. We define `gotpatent` as a binary treatment variable where 1 indicates a subclass received a TWEA patent after 1918, and 0 otherwise. The outcome variable is the number of patents granted to US inventors.

```
xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)
```

This command estimates the Average Treatment Effect on the Treated (ATET) for the number of US patents granted after the treatment (compulsory licensing).

The results from this command will provide the coefficient for the treatment variable (`gotpatent`), which tells us the impact of compulsory licensing on the number of US patents granted, adjusted for fixed effects.

10.3 Dealing with Time-Varying Treatment Effects

10.3.1 Event Study Approach

In some cases, treatment effects may vary over time. To explore this, we can use the event study approach, which examines how the treatment effect evolves before and after the intervention. The `estat grangerplot` command in Stata allows us to perform this event study and visualize dynamic treatment effects.

To perform an event study, we first run the DiD model, then use the `estat grangerplot` command to plot the leads and lags of the treatment effect:

```
xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)
estat grangerplot
```

This will generate a plot showing the treatment effect over time, including the pre-treatment period (leads) and post-treatment period (lags). The coefficients for each period indicate how the treatment effect changes over time.

10.4 Aggregation Methods and Bias-Correction

10.4.1 Data Aggregation

In some cases, especially with small sample sizes, it may be useful to aggregate data to improve inference. The `aggregate` option in `xtdidregress` allows us to aggregate the data

at the group and time levels, which can help mitigate issues with small group sizes or unbalanced data.

For example, we can aggregate the data using the standard method:

```
xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year) aggregate(stand
```

This command first aggregates the data at the group and time levels and then fits the DiD model. The aggregated results will provide more reliable standard errors in the case of small or unbalanced panels.

10.4.2 Bias-Corrected Standard Errors

When dealing with small samples or clusters, it may be necessary to compute bias-corrected standard errors. One method is to use the `vce(hc2)` option, which adjusts the standard errors for small-sample bias.

The following command applies the bias-correction using the `vce(hc2)` option:

```
xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year) vce(hc2)
```

This will calculate robust, bias-corrected standard errors for the treatment effect. The results will provide more accurate inference when working with small sample sizes or clusters in the data.

10.5 Conclusion

By using the ‘xtdidregress‘ command in Stata, researchers can efficiently estimate treatment effects in panel data settings, while accounting for both individual and time fixed effects. The ability to perform event studies and apply bias-corrected standard errors further strengthens the reliability of these estimates. These methods are crucial for causal inference in situations where treatment effects vary over time or where the data structure may introduce challenges to standard DiD estimation.

References

[1] Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press. https://doi.org/10.1515/9781400829828

[2] Angrist, J. D., & Pischke, J.-S. (2014). *Mastering 'Metrics: The Path from Cause to Effect.* Princeton University Press. https://doi.org/10.1515/9781400852383

[3] Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. https://doi.org/10.1162/003355304772839588

[4] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

[5] Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4), 772–793. Available at https://www.jstor.org/stable/2118030

[6] Cheng, C., & Hoekstra, M. (2013). Does Strengthening Self-Defense Law Deter Crime or Escalate Violence? Evidence from Castle Doctrine Laws. *Journal of Human Resources*, 48(3), 821–854. https://doi.org/10.3368/jhr.48.3.821

[7] Cunningham, S. (2021). *Causal Inference: The Mixtape.* Yale University Press. https://doi.org/10.2307/j.ctv15d8186

[8] Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

[9] Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics

of Program Evaluation. *Journal of Economic Literature*, 47(1), 5–86. https://doi.org/10.1257/jel.47.1.5

[10] Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Differences Methods. *Foundations and Trends in Econometrics*, 4(3), 165–224. https://doi.org/10.1561/0800000014

[11] Mora, R., & Reggio, I. (2015). Treatment Effect Identification Using Alternative Parallel Assumptions. *Econometrics Journal*, 18(3), 223–244. https://doi.org/10.1111/ectj.12052

[12] Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, 11(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

[13] Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

[Scott Cunningham's Mixtape Website] Cunningham, S. (2021). Causal Inference: The Mixtape. Access the companion website: https://mixtape.scunning.com

[Scott Cunningham's GitHub] Cunningham, S. (2021). Access datasets and Stata/R code on the companion GitHub repository: https://github.com/scunning1975/mixtape

[16] StataCorp. (2024). *Stata 18 Base Reference Manual*. https://www.stata.com/manuals/