

Week 8: Difference-in-Differences

Minimum wages and E-ZPass · PP5001 · Martinmas 2026

[Class slides](#) · [Slides PDF](#) · [Notes PDF](#) · [Post-class exercise](#)

Learning Objectives

By the end of this topic, you should be able to:

- Calculate a difference-in-differences estimate from four group means and represent its counterfactual graphically.
- Explain parallel trends in terms of untreated potential outcomes and derive the average treatment effect on the treated under that assumption.
- Connect the four means to the coefficients of a regression with group, period, and interaction terms.
- Assess a policy comparison using evidence about timing, group composition, spillovers, and other changes affecting the outcome.
- Estimate a basic DiD in Python and interpret its magnitude and uncertainty.

Before class

Read these notes and **Card and Krueger (1994)** and **Currie and Walker (2011)**, available in Moodle. The first paper evaluates a minimum-wage increase using restaurants in two states. The second studies electronic toll collection and infant health using distance from toll plazas. Prepare to explain the policy, data, identification strategy, and results in each case.

Textbook references

Use the books to support your reading of the notes. You do not need to read all four accounts.

- **The Effect:** [Chapter 18, Difference-in-Differences](#), especially the basic design and regression interpretation.
- **Causal Inference: The Remix:** [Chapter 9, Difference-in-Differences Fundamentals](#).
- **Mastering 'Metrics:** Chapter 5, *Differences-in-Differences*. [Companion resources](#) and [Mastering Econometrics videos at MRU](#).
- **Mostly Harmless Econometrics — advanced reading:** Chapter 5, *Parallel Worlds: Fixed Effects, Differences-in-Differences, and Panel Data*.

The Evaluation Problem

A policy begins at a particular time in a particular place. We observe outcomes before and after its introduction.

The before–after change combines:

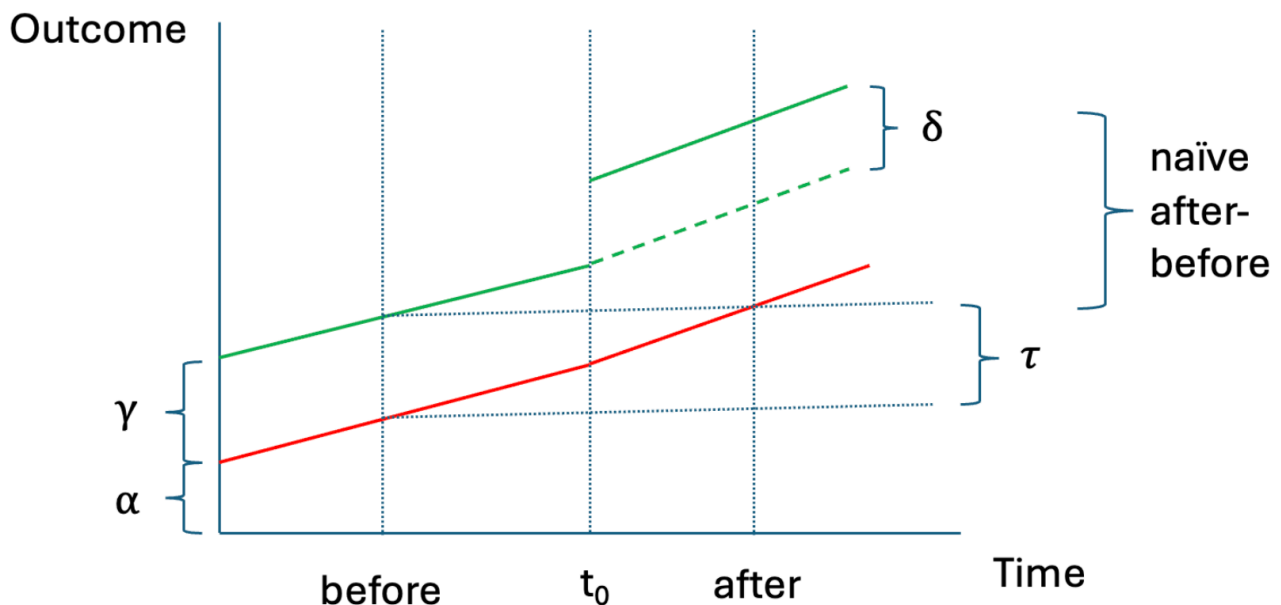
- The effect of the policy.
- The change that would have occurred even without the policy.

We need a comparison that tells us about that second component.

For example, employment might fall after a minimum-wage increase because the policy reduced labour demand, because the economy entered a recession, or because both happened. Equally, a beneficial policy may be followed by a deterioration in outcomes if other conditions worsen sufficiently. The direction of the observed change alone does not identify the effect.

This is the same missing-counterfactual problem we have encountered throughout the module. Here, we observe the treated group at an earlier date. Its earlier outcome is useful, but time itself brings changes that we must account for.

Before and After



The graph distinguishes the observed post-policy outcome from the outcome that would have occurred without the intervention. A comparison with the pre-policy level attributes the entire change to treatment. To isolate the policy effect, we must also allow for the underlying change over time.

The Basic DiD Idea

Use the change in the control group to estimate the change the treated group would have experienced without treatment.

$$\hat{\delta}_{DiD} = (\bar{Y}_{T,post} - \bar{Y}_{T,pre}) - (\bar{Y}_{C,post} - \bar{Y}_{C,pre}).$$

Taking each group's change removes its fixed level. Subtracting the control group's change removes the common change over time.

The subscripts T and C label the treatment and control groups. "Pre" and "post" label the two periods. Each bar denotes the sample mean in one of the four cells.

We can also subtract the treatment-control difference before the policy from the treatment-control difference after the policy. Rearranging the same four terms gives the same answer. Both calculations ask how the gap between the groups changed.

The 2×2 Parameterization

Group	Before	After	Change
Treatment	$\alpha + \gamma$	$\alpha + \gamma + \tau + \delta$	$\tau + \delta$
Control	α	$\alpha + \tau$	τ
Treatment minus control	γ	$\gamma + \delta$	δ

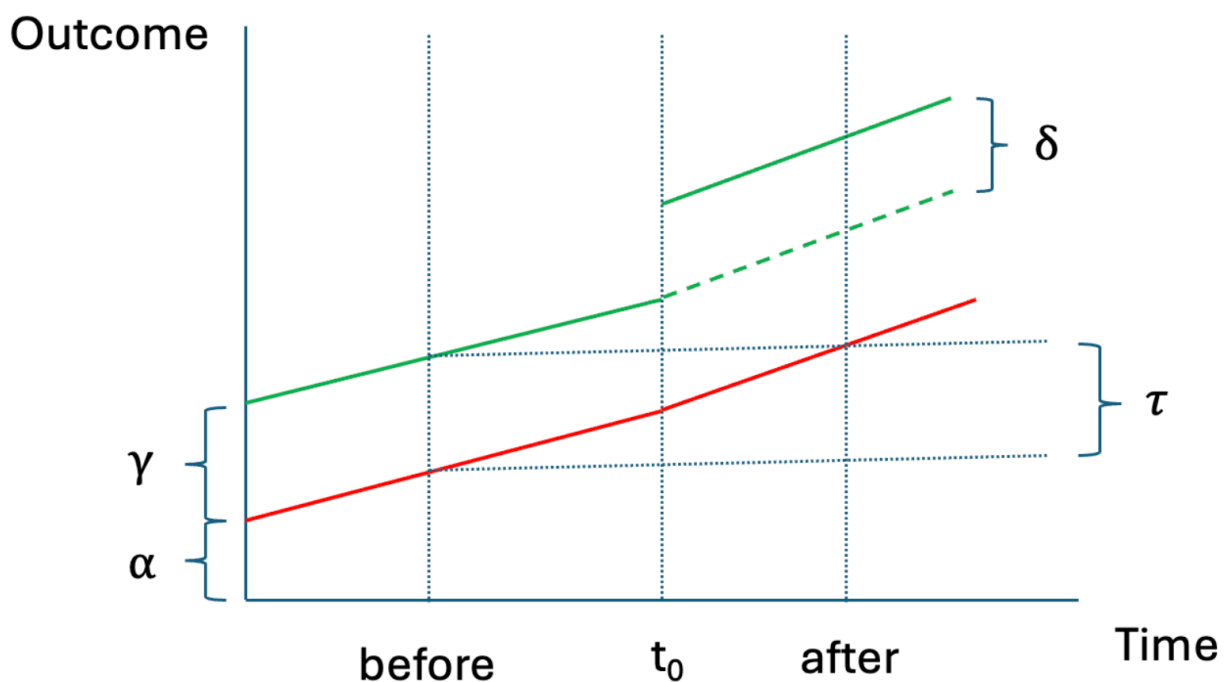
α : control-group baseline. γ : initial group difference.

τ : common time change. δ : the additional change in the treated group.

Start in the control group's pre-policy cell. Its mean is α . Moving to the treatment group adds γ . Moving to the post period adds τ . The treatment group in the post period has one further component, δ .

This parameterization can describe any four cell means. Interpreting δ as a causal effect requires an assumption about how the treated group's outcome would have changed without the policy. We will state that assumption explicitly below.

The Parameters in the Graph



The original graph labels the common change τ , the initial group difference γ , and the treatment effect δ . The dashed continuation of the treated group's path represents its counterfactual. The vertical gap between that path and the observed treated outcome after the policy is the effect of interest.

Regression Representation

Let $T_i = 1$ identify membership in the treatment group and $Post_t = 1$ identify the post-policy period.

$$Y_{it} = \alpha + \gamma T_i + \tau Post_t + \delta (T_i \times Post_t) + u_{it}.$$

- T_i allows the groups to have different initial levels.
- $Post_t$ allows for the common time change.
- $T_i \times Post_t$ equals one only for treated observations after the policy.

The individual or unit is indexed by i and time by t . In the restaurant example, i identifies a restaurant. In this basic design, group membership is fixed, while actual treatment exposure is $D_{it} = T_i \times Post_t$.

We use τ consistently with the original graph. In the exercise's generic coefficient notation, $\beta_0 = \alpha$, $\beta_1 = \gamma$, $\beta_2 = \tau$, and $\beta_3 = \delta$.

Why the Interaction Coefficient Is DiD

With an intercept and these three indicators, OLS fits all four cell means:

$$\begin{aligned}\hat{\alpha} &= \bar{Y}_{C,pre}, \\ \hat{\gamma} &= \bar{Y}_{T,pre} - \bar{Y}_{C,pre}, \\ \hat{\tau} &= \bar{Y}_{C,post} - \bar{Y}_{C,pre}, \\ \hat{\delta} &= \bar{Y}_{T,post} - \hat{\alpha} - \hat{\gamma} - \hat{\tau}.\end{aligned}$$

Substituting the first three expressions into the last gives the difference-in-differences estimator.

To see the last step explicitly,

$$\begin{aligned}\hat{\delta} &= \bar{Y}_{T,post} - \bar{Y}_{C,pre} - (\bar{Y}_{T,pre} - \bar{Y}_{C,pre}) - (\bar{Y}_{C,post} - \bar{Y}_{C,pre}) \\ &= (\bar{Y}_{T,post} - \bar{Y}_{T,pre}) - (\bar{Y}_{C,post} - \bar{Y}_{C,pre}).\end{aligned}$$

There are four regression parameters and four group-period cells, so the regression can fit each cell mean exactly. This equivalence applies to the unadjusted regression using the same observations and weights as the calculation of means. Adding covariates or changing the sample changes the comparison.

Potential Outcomes and the ATT

Let Y_{it}^1 and Y_{it}^0 denote outcomes with and without treatment.

The effect of interest is

$$ATT = E[Y_{i,post}^1 - Y_{i,post}^0 \mid T_i = 1].$$

We observe $Y_{i,post}^1$ for the treatment group. Its mean untreated outcome after the policy is the missing counterfactual.

Assume no anticipation and no spillovers to the control group.

Group membership T_i and treatment exposure D_{it} play different roles. Members of the treatment group are untreated in the pre period and treated in the post period. The control group remains untreated in both periods.

We assume no anticipation: the policy does not affect the treatment group's pre-policy outcome. We also need the policy to leave the control group's outcomes unaffected, including through spillovers. As in earlier weeks, outcomes must correspond to a well-defined treatment and be measured comparably.

Parallel Trends

In the absence of treatment, the two groups would have experienced the same average change:

$$E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 1] = E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 0].$$

The groups may start at different levels. The assumption concerns their **changes in untreated potential outcomes**.

This is an assumption about a counterfactual after treatment starts. We can observe the control group's untreated change. For the treatment group, we observe its pre-policy untreated outcome and its post-policy treated outcome. We cannot directly observe the untreated post-policy change on the left-hand side.

A convincing argument therefore needs knowledge of why the policy was adopted, why the comparison group is suitable, and what other events occurred. If several earlier periods are available, similar prior trends can support that argument. Their similarity cannot guarantee that the counterfactual trends would remain parallel after the policy starts.

Constructing the Counterfactual

Parallel trends implies

$$E[Y_{i,post}^0 | T_i = 1] = E[Y_{i,pre}^0 | T_i = 1] + E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 0].$$

Treated group's baseline + control group's change = treated group's counterfactual post-policy mean.

The assumption lets us fill in the missing cell. Begin at the treatment group's observed baseline and add the change observed in the control group. This preserves the initial difference between the groups while applying a common untreated trend.

The counterfactual may rise or fall. A positive treatment effect means that the observed outcome exceeds this counterfactual, even if the treatment group's observed outcome falls over time.

From the Counterfactual to the Effect

Substitute that counterfactual into the definition of the ATT:

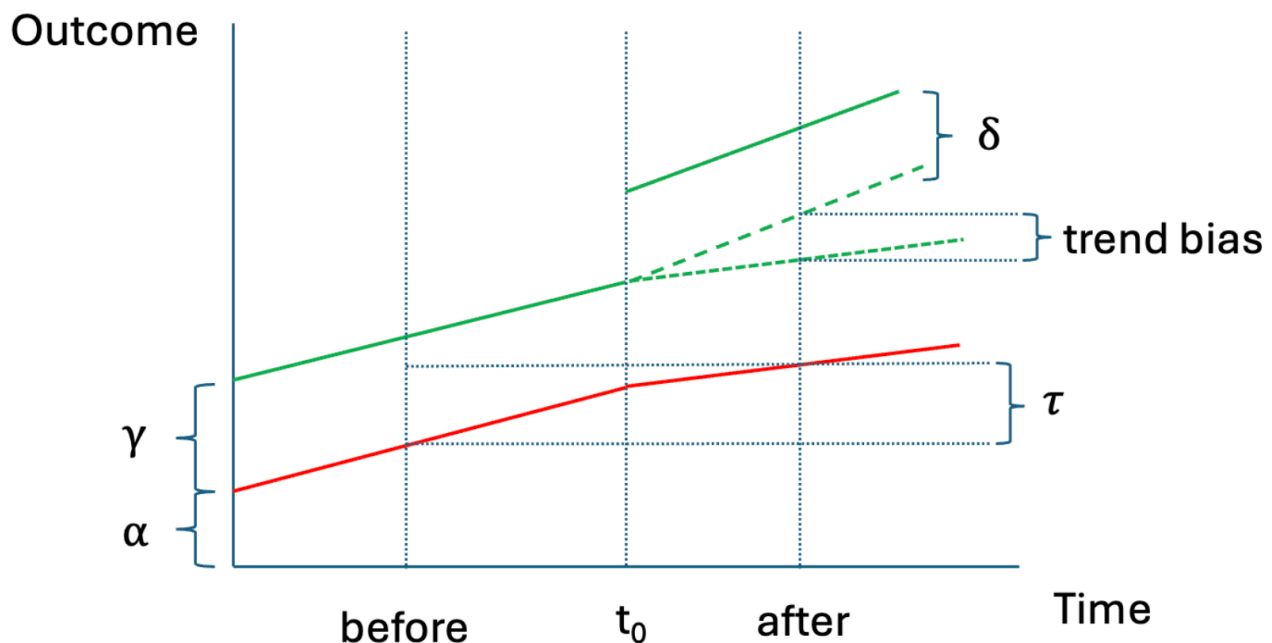
$$ATT = E[Y_{i,post} | T_i = 1] - E[Y_{i,pre} | T_i = 1] - (E[Y_{i,post} | T_i = 0] - E[Y_{i,pre} | T_i = 0]).$$

Use the analogy principle to replace the four population conditional means with their sample counterparts.

We can write these expectations using observed Y because of the treatment pattern and no-anticipation assumption: the treatment group reveals its treated outcome after the policy and its untreated outcome before it. The control group reveals untreated outcomes in both periods.

This derivation allows effects to differ across treated units. The basic DiD identifies their average post-policy effect under the stated assumptions. It does not require each restaurant or each mother to experience the same effect.

Different Underlying Trends



If the treatment group's untreated outcome would have grown faster, the control group's change understates its counterfactual growth. The DiD then attributes some of that extra growth to the policy. A slower untreated trend produces the opposite error.

More formally, the population DiD equals the ATT plus

$$E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 1] - E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 0] .$$

Parallel trends sets this second component to zero. Increasing the number of observations can make the estimate more precise without removing a systematic difference in those untreated trends.

Card and Krueger: Policy, Data, and Identification

For discussion: What changed in New Jersey? Who was surveyed, when, and why is eastern Pennsylvania a plausible control group? State the identifying assumption in this setting.

New Jersey raised its minimum wage from \$4.25 to \$5.05 per hour on 1 April 1992. Pennsylvania's minimum remained \$4.25. Card and Krueger surveyed fast-food restaurants in February–March and again in November–December. The initial sample contained 410 restaurants.

Fast-food restaurants employ many low-wage workers, making the policy particularly relevant to them. Nearby Pennsylvania restaurants provide a comparison exposed to some of the same regional economic conditions. The causal argument still requires that their employment changes represent what New Jersey restaurants would have experienced without the increase.

The study followed existing restaurants. Its employment measure includes changes at restaurants that closed, but the design does not directly measure employment generated by new restaurant openings. Keep that scope in mind when making a policy recommendation.

Card and Krueger: Figure 1

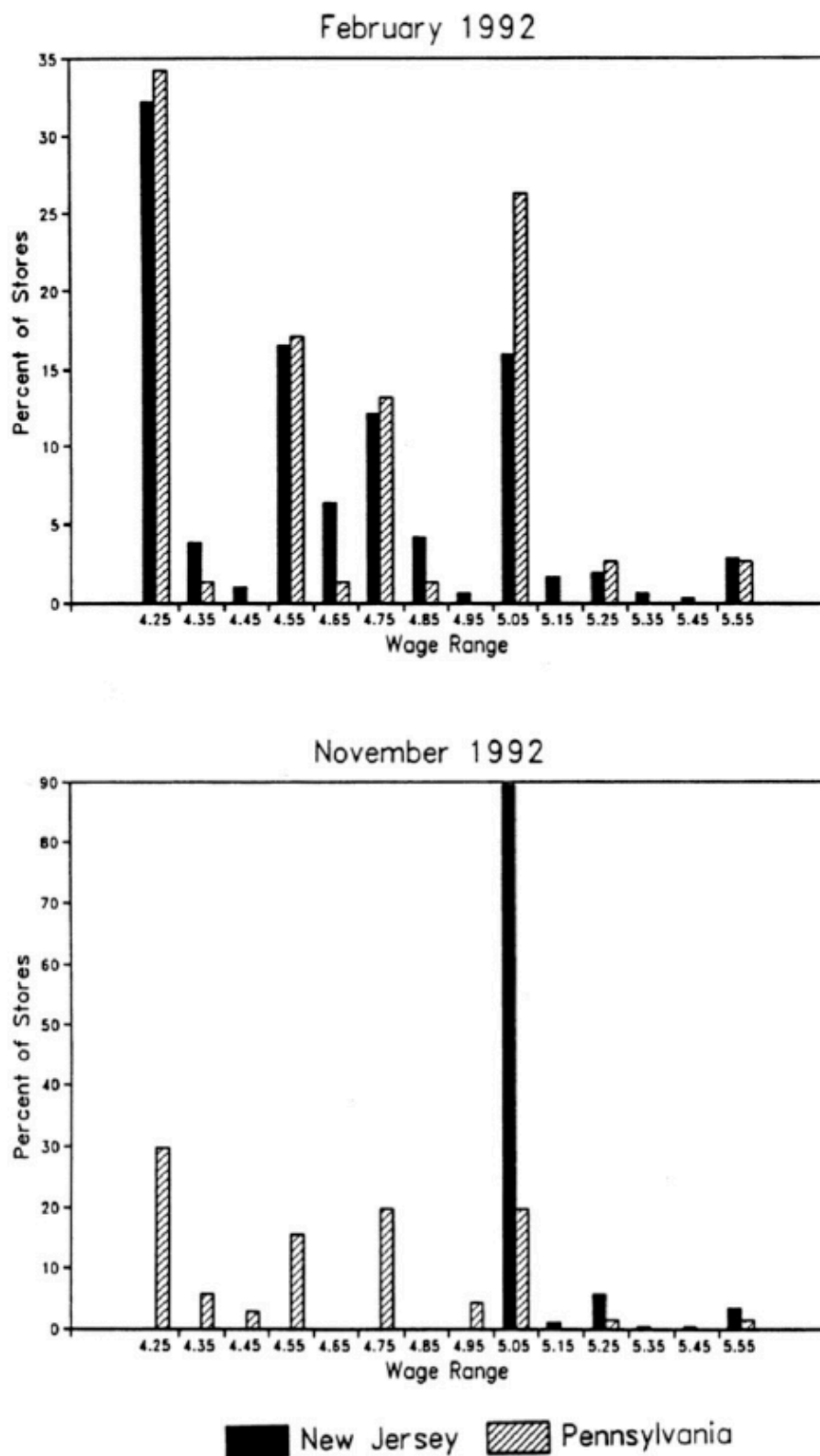


FIGURE 1. DISTRIBUTION OF STARTING WAGE RATES

Card and Krueger (1994), Figure 1.

The distributions of starting wages show how the increase changed the wage floor in New Jersey relative to Pennsylvania. This provides evidence that the policy affected the price of labour faced by the sampled restaurants. The employment comparison then asks how employment responded.

Card and Krueger: Table 3

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE
IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey ^a			Differences within NJ ^b	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low– high (vii)	Midrange– high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores ^c	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 ^d	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

^aStores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

^bDifference in employment between low-wage (\$4.25 per hour) and high-wage ($\geq \$5.00$ per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

^cSubset of stores with available employment data in wave 1 and wave 2.

^dIn this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

For discussion: Explain columns (i)–(iii), including the DiD. Why do rows 3 and 4 differ? What is a full-time-equivalent employee?

Full-time-equivalent employment counts managers and full-time employees at weight one and part-time employees at weight one-half. The unit is an equivalent full-time employee per restaurant.

Rows 1 and 2 report the available employment means in each wave. Row 3 differences those means. Row 4 uses restaurants with employment observed in both periods, holding the set of restaurants fixed. Missing employment information means these are slightly different comparisons. The distinction matters when translating a published table into code.

The post-class exercise reproduces these calculations. It then uses the balanced sample to show that the four-means calculation and the regression interaction coefficient agree. Permanently closed restaurants have zero post-period employment in the data, which is part of the policy-relevant outcome rather than a missing observation.

Covariates in the Levels Model

Let X_{ik} be pre-treatment characteristics held fixed across the two periods.

$$Y_{it} = \alpha + \gamma T_i + \tau Post_t + \delta (T_i \times Post_t) + \sum_{k=1}^K \beta_k X_{ik} + \sum_{k=1}^K \theta_k (X_{ik} \times Post_t) + u_{it}.$$

- β_k allows baseline outcomes to differ with X_{ik} .
- θ_k allows the change over time to differ with X_{ik} .
- δ is the additional change associated with treatment, conditional on these characteristics.

The original parameters keep their meanings: α is the control-group baseline when the covariates equal zero, γ is the initial group difference conditional on the covariates, τ is the control-group time change at those reference values, and δ is the treatment-group additional change. The new coefficients are β_k for baseline covariate differences and θ_k for covariate-specific changes.

For example, restaurants from different chains may start with different employment levels and may experience different employment changes. The terms involving β_k allow the first pattern. The interactions with $Post_t$, with coefficients θ_k , allow the second.

For a causal interpretation, we need a credible parallel-trends argument conditional on the pre-treatment characteristics, appropriate overlap, and a suitable regression specification. In this additive model, a common δ represents the conditional treatment effect. With heterogeneous effects, an adjusted OLS coefficient need not equal the overall ATT without further restrictions or appropriate averaging.

From Levels to Differences with Covariates

For the same unit, subtract the pre-period equation from the post-period equation:

$$\Delta Y_i = \tau + \delta T_i + \sum_{k=1}^K \theta_k X_{ik} + \Delta u_i.$$

- α , γT_i , and $\sum_{k=1}^K \beta_k X_{ik}$ cancel.
- τ , δ , and θ_k retain exactly the same meanings.
- Card–Krueger Table 4 uses employment changes as the outcome, with chain and ownership indicators entered directly as controls.

Here $\Delta Y_i = Y_{i,post} - Y_{i,pre}$ and $\Delta u_i = u_{i,post} - u_{i,pre}$. Before the policy, both $Post_t$ and $T_i \times Post_t$ are zero. After the policy, $Post_t = 1$ and $T_i \times Post_t = T_i$. Since the characteristics are held fixed, their baseline contributions cancel while their post-period interactions leave $\theta_k X_{ik}$.

Thus the covariate coefficients in the change regression are the θ_k **coefficients** from the levels model. Chain and ownership indicators allow different types of restaurants to experience different employment changes.

Card and Krueger estimate the change regression directly. The preceding levels equation shows how that specification can be represented with period interactions. Adding the fixed characteristics to a levels regression only as main effects would allow baseline differences but would not allow these covariate-specific changes.

For a balanced two-period sample with these same fixed covariates and all the displayed interactions, OLS in levels and OLS in changes give the same treatment-effect point estimate. Their standard errors must account for the different representations of repeated observations and may use different finite-sample corrections.

Card and Krueger: Table 4

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap ^a	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership ^b	no	yes	no	yes	yes
4. Controls for region ^c	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls ^d	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are -0.237 and 8.825 , respectively. All models include an unrestricted constant (not reported).

^aProportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

^bThree dummy variables for chain type and whether or not the store is company-owned are included.

^cDummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

^dProbability value of joint F test for exclusion of all control variables.

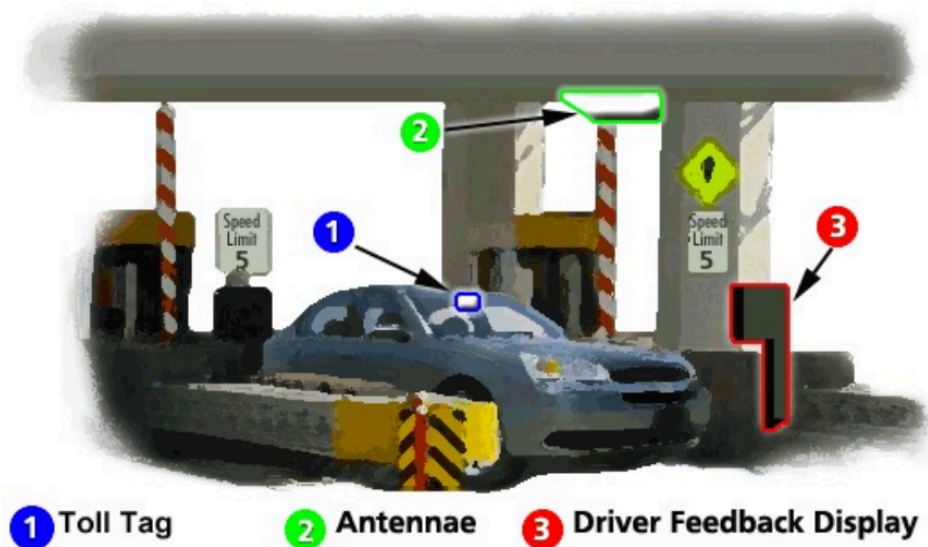
For discussion: Compare columns (i) and (ii). Why might chain and ownership matter? Separate the change in sample from the effect of adding controls.

The dependent variable is the change in employment. Column (i) regresses that change on the New Jersey indicator and an intercept. Column (ii) adds indicators for chain and company ownership. These allow average employment changes to differ across those types of restaurants.

The sample is smaller than the balanced sample used in Table 3. The paper restricts the subsequent regression analysis to restaurants with the employment and wage information needed for those comparisons. Consequently, a difference between Table 3 and Table 4, column (i), can arise from changing the sample. The comparison between columns (i) and (ii) isolates the addition of the chain and ownership controls on the same sample.

Both columns give positive New Jersey estimates. Adding controls changes the magnitude only modestly. This is useful evidence about those measured differences, while the causal interpretation still depends on the comparison group's counterfactual employment trend.

Currie and Walker: The Policy



As your vehicle enters the toll lane, the toll tag (1) that is mounted on your vehicle's windshield is read by the antennae (2). As your vehicle passes through, your *E-ZPass*® account is charged the proper amount. Feedback is provided to you on an electronic display (3). If your vehicle does not have a toll tag, the system classifies you as a violator and cameras take photos of your license plate for processing.

E-ZPass allows electronic toll collection, reducing stops and queues at toll plazas. The study asks whether the resulting changes improve infant health nearby.

The mechanism connects transport policy to local exposure: less queuing and idling may reduce vehicle emissions near a toll plaza. Pregnancy provides a period during which changes in exposure could affect birth outcomes. This gives us a second application of the same comparison of changes, with a very different policy and outcome.

Currie and Walker: Data and Identification

For discussion: Define the treatment and control groups, the timing of treatment, and the birth outcomes. Why compare mothers near the same road network? State parallel trends for this design.

The researchers link geocoded birth records in New Jersey and Pennsylvania to toll-plaza locations and E-ZPass adoption dates. Their central comparison uses mothers within 2 km of a toll plaza and mothers between 2 and 10 km away, with the latter also living within 3 km of a major highway.

The control group helps account for changes affecting mothers living near major roads. The assumption is that birth outcomes near toll plazas would have changed like those in the comparison areas without E-ZPass. The published analysis includes multiple locations and adoption dates. We use the near-versus-far, before-versus-after contrast to understand its logic here. The fuller treatment of timing and multiple periods follows in the next sessions.

Where Are the Comparisons Made?

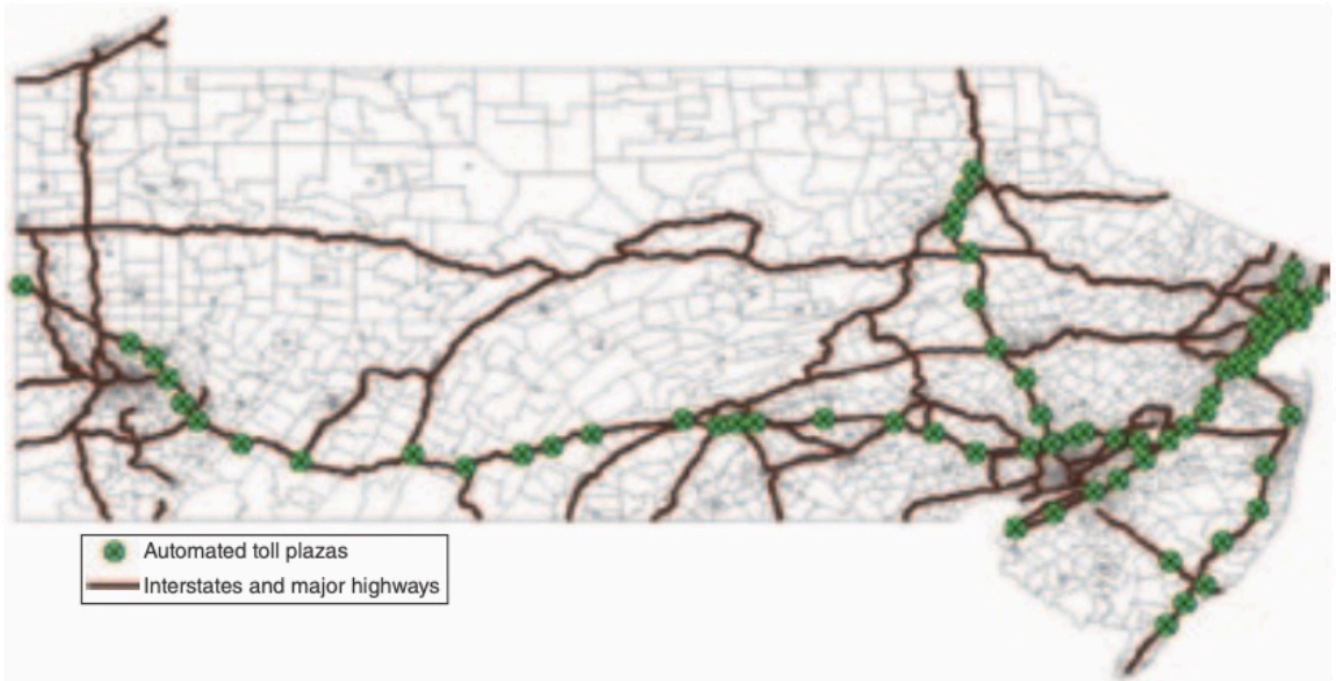


FIGURE 1. LOCATIONS OF TOLL PLAZAS AND MAJOR ROADWAYS IN NEW JERSEY AND PENNSYLVANIA

Currie and Walker (2011), Figure 1.

The map places the treatment and comparison areas in their regional setting. Proximity makes some shared economic and environmental changes plausible, but it also raises questions about spillovers. For example, changes in traffic patterns could affect nearby comparison areas as well as the area immediately around a toll plaza.

Currie and Walker: Table 3

TABLE 3—REGRESSIONS OF BIRTH OUTCOMES ON E-ZPASS ADOPTION
(Difference-in-Difference Specification)

	Prematurity (1)	Prematurity (2)	LBW (3)	LBW (4)
<i>Panel 1</i>				
<2km toll × after E-ZPass	−0.0085 [0.0039]**	−0.0086 [0.0034]**	−0.0094 [0.0032]**	−0.0093 [0.0028]**
R^2	0.0044	0.0034	0.0032	0.0028
<i>Panel 2</i>				
<1.5km toll × after E-ZPass	−0.0088 [0.0051]*	−0.0098 [0.0048]**	−0.0077 [0.0035]**	−0.0084 [0.0032]**
R^2	0.0042	0.0048	0.0035	0.0032
Maternal characteristics	No	Yes	No	Yes
Observations	405,802	405,802	409,673	409,673

Notes: Each coefficient is from a different regression. All regressions also included controls for being within 2 km (or 1.5 km) of a toll plaza, year of birth, month of birth, toll plaza indicators, an indicator for post E-ZPass, and distance to highway. Maternal characteristics include: mother black, mother Hispanic, mother education (<hs, hs, some college, college +), mother age (19–24, 25–34, 35+), smoking, multiple birth, gender, and birth order, and indicators for missing values. Standard errors are in brackets.

** Indicates that the estimate is statistically significant at the 95 percent level of confidence.

* Indicates significance at the 90 percent level of confidence.

For discussion: Interpret the prematurity and low-birth-weight estimates in Panel 1. Distinguish percentage points from percentages. How do the estimates change when maternal characteristics are included?

Prematurity and low birth weight are binary outcomes. A coefficient of -0.009 in a linear probability model represents a decline of 0.9 percentage points. A percentage reduction requires dividing that change by a relevant baseline probability. Always state the baseline used.

Panel 1 uses the 2 km treatment radius. The columns compare specifications with and without maternal characteristics. Panel 2 changes the treatment radius to 1.5 km. This comparison helps assess how sensitive the estimates are to the definition of local exposure.

For the class discussion, distinguish three questions: the direction and size of the estimate, its uncertainty, and the assumptions needed to interpret it causally. A precise coefficient addresses the second question but cannot settle the third.

Currie and Walker: Table 2

TABLE 2—TESTING THE VALIDITY OF THE RESEARCH DESIGN: REGRESSIONS OF MATERNAL CHARACTERISTICS ON E-ZPASS ADOPTION
(Difference-in-Difference Specification)

	Black (1)	Hispanic (2)	Mother yrs. ed (3)	Dropout (4)	Teen mother (5)	Mother smoked (6)	Housing sale price (7)
<i>Panel 1</i>							
<2 km toll × after E-ZPass	−0.011 [0.011]	−0.01 [0.010]	0.037 [0.040]	−0.007 [0.005]	−0.001 [0.005]	0.005* [0.003]	0.149 [0.103]
Observations	397,201	406,641	406,198	397,201	412,884	402,590	252,343
<i>Panel 2</i>							
<1.5 km toll × after E-ZPass	−0.014 [0.055]	−0.01 [0.011]	0.013 [0.010]	−0.003 [0.006]	0.001 [0.003]	0.007** [0.003]	0.031 [0.106]
Observations	397,201	406,641	406,198	397,201	412,884	402,590	252,343

Notes: Each coefficient is from a separate regression. Each coefficient in columns 1–6 is from a regression that also included controls for being within 2 km (or 1.5 km) of a toll plaza, year of birth, month of birth, indicators for each toll plaza, an indicator for post E-ZPass at nearest toll plaza, and distance to highway. Housing sale price regressions in column 7 include year and month of sale, indicators for nearest toll plaza, an indicator for condo units, distance to highway, municipality fixed effects, square footage (in categories including dummies for missing), and age (in categories, including dummies for missing). Standard errors are in brackets.

** Indicates that the estimate is statistically significant at the 95 percent level of confidence.

* Indicates significance at the 90 percent level of confidence.

For discussion: Could changes in who lives near toll plazas explain the health results? What does this table tell us, and what uncertainty remains?

The authors apply the comparison of changes to maternal characteristics. If E-ZPass led different families to move into areas near toll plazas, changes in infant health could partly reflect the changing composition of births observed there.

Small and imprecise changes in measured characteristics are reassuring to the extent that they rule out substantively important changes. They still leave room for changes in unmeasured characteristics. Consider the estimates and their uncertainty rather than treating a set of insignificant coefficients as proof of comparability.

Housing-market responses offer another way to investigate sorting. The timing matters: a policy could have little immediate effect on residential composition but affect it over a longer period.

Currie and Walker: Table 7

TABLE 7—DIFFERENCE-IN-DIFFERENCES ESTIMATES OF EFFECTS OF E-ZPASS ON POLLUTION

	NO ₂ All control monitors (1)	SO ₂ All control monitors (2)	NO ₂ Random control 1 (3)	NO ₂ Random control 2 (4)	NO ₂ Random control 3 (5)	NO ₂ Random control 4 (6)	NO ₂ Random control 5 (7)
<2km toll × after E-ZPass	−0.108 [0.019]**	0.053 [0.034]	−0.208 [0.028]**	−0.090 [0.024]**	−0.065 [0.017]**	−0.181 [0.023]**	0.018 [0.038]
Observations	84,159	128,513	6,361	6,449	6,453	6,448	6,421

Notes: Each coefficient is from a separate regression. Columns 1 and 2 use all monitors over 2 km from a toll plaza as controls. Columns 3–7 each use a randomly selected control monitor. Regressions also included controls for being within 2 km of a toll plaza, year of birth, month of birth, indicators for each toll plaza, an indicator for post E-ZPass at nearest toll plaza, and distance to highway. Dependent variable is the log daily mean pollution level for the indicated pollutant. Standard errors are in brackets.

** Indicates that the estimate is statistically significant at the 95 percent level of confidence.

* Indicates significance at the 90 percent level of confidence.

For discussion: Compare the two pollutants. How does this evidence support the proposed mechanism? What does having only one monitor near a toll plaza imply?

The dependent variables are log daily mean pollutant levels. Column (1)'s coefficient of -0.108 corresponds approximately to a 10.8% reduction in nitrogen dioxide. The exact proportional change is $\exp(-0.108) - 1$, or about a 10.2% reduction. The sulfur dioxide comparison is useful because traffic is a much more important source of the first pollutant.

There is only one monitor within 2 km of a toll plaza. Many observations over days provide information about that location, but they do not create many independently treated locations. This limits the strength and generality of the pollution evidence.

The pollution analysis supports a mechanism linking E-ZPass to health. Estimating the causal effect of pollution itself would additionally require an appropriate first stage, an exclusion restriction ruling out other channels from E-ZPass to health, and a compatible population and exposure measure.

What Can Threaten the Comparison?

- **Different underlying trends:** a local shock changes the treated group's outcome independently of the policy.
- **Anticipation:** behaviour changes before the designated treatment date.
- **Spillovers:** the policy affects the control group.
- **Composition:** the people or establishments observed change over time.

In Card–Krueger, ask whether regional economic developments could have affected employment differently across states and whether restaurants adjusted before April. In Currie–Walker, ask whether traffic was displaced and whether the composition of mothers changed.

Treating closures correctly illustrates why composition requires care. Zero employment at a closed restaurant is a meaningful outcome. Dropping closed restaurants would change the question to employment among survivors, whose survival may itself have been affected by treatment.

Controls and Evidence on Trends

Use institutional knowledge to motivate the comparison. Examine earlier trends when the data allow it.

Choose controls with a clear role in the design. Variables affected by the policy may remove part of the effect we want to estimate or introduce selection.

Adding controls can account for measured differences in how outcomes would otherwise change. It requires a corresponding argument that the remaining untreated trends are comparable conditional on those controls. Similar baseline levels alone cannot supply that argument.

With only one pre-policy period, we cannot compare pre-policy slopes. With several, plots can reveal differences before treatment and help identify anticipation. A failure to reject a pre-trend test may also reflect limited precision. Event studies will give us a systematic way to examine these patterns in the next session.

Uncertainty and the Level of Treatment

Many individuals can share the same policy and the same shocks.

- Restaurants share state-level economic conditions.
- Repeated observations on a restaurant can have correlated disturbances.
- Daily observations at a monitor can share persistent local conditions.

The number of observations is different from the number of independent policy comparisons.

Heteroskedasticity-consistent standard errors allow the disturbance variance to differ across observations but ordinarily retain independence across observations. Clustered standard errors also allow disturbances to be correlated within a specified group.

In the exercise, clustering by restaurant accounts for the two observations from each restaurant. It does not account for shocks shared by different restaurants in the same state. With only two states, usual methods that rely on many independent state clusters cannot be justified simply by requesting state-clustered standard errors. The reported uncertainty should therefore be read alongside the institutional argument and the limited number of policy comparisons.

From Four Means to Python

For data with one row per unit and period:

```
import statsmodels.formula.api as smf

model = smf.ols("outcome ~ treated * post", data=long).fit(
    cov_type="cluster", cov_kws={"groups": long["unit_id"]}
)
```

`treated * post` includes both indicators and their interaction. The interaction coefficient is the unadjusted DiD.

Here `treated` identifies the treatment group in both periods and `post` identifies the later period in both groups. The variable `unit_id` links repeated observations on the same unit. Replace these generic names with the actual columns in your data.

The post-class notebook supplies the import and reshaping code for Card–Krueger. You calculate the means, plot the counterfactual, and interpret the coefficients. Using the same sample at each step is essential for verifying the equivalence.

Changes as the Dependent Variable

With the same units observed twice, subtract each unit’s pre-period equation from its post-period equation:

$$\Delta Y_i = \tau + \delta T_i + \Delta u_i.$$

The intercept is the control group’s average change. The coefficient on T_i is the difference between the groups’ changes.

The constant α and the time-invariant group component γT_i cancel. The post indicator changes from zero to one. Treatment exposure also changes from zero to one for the treatment group, while remaining zero for controls.

This gives a second regression implementation of the same 2×2 comparison. Use the same balanced sample to verify equality of the point estimates. Standard-error corrections can differ slightly between implementations, so exact equality of coefficients does not imply numerical identity of every reported statistic.

What a Credible DiD Design Needs

1. Clear policy timing and a well-defined outcome.
2. A plausible comparison group and an explicit parallel-trends argument.
3. Evidence on other changes, anticipation, spillovers, and composition.
4. Uncertainty appropriate to the available independent comparisons.

The control group supplies the **change** used to construct the treated group’s counterfactual.

References

Card, D., and A. B. Krueger (1994). [Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania](#). *American Economic Review*, 84(4), 772–793.

Currie, J., and R. Walker (2011). [Traffic Congestion and Infant Health: Evidence from E-ZPass](#). *American Economic Journal: Applied Economics*, 3(1), 65–90.

Angrist, J. D., and J.-S. Pischke (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Chapter 5.

Angrist, J. D., and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Chapter 5. Advanced reading.

Cunningham, S. *Causal Inference: The Remix*. Chapter 9, linked above.

Huntington-Klein, N. *The Effect: An Introduction to Research Design and Causality*. Chapter 18, linked above.