Advanced Microeconometrics 6SSPP393

Topic 2: Policy evaluation using panel data

Yonatan Berman and Elisa Cavatorta

Semester 2, 2023/24

Recap

- Unobserved heterogeneity causes pooled OLS to be biased
- Considering a simple model with one explanatory variable x_{it} (but also more generally), if the correct model is

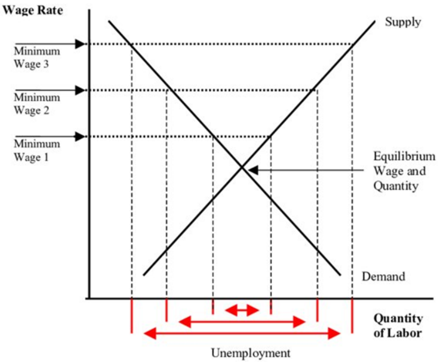
$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

  but if one estimates

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it} \ ,$$

  then $v_{it} = a_i + u_{it}$
- To use OLS we must assume $Cov\left(v_{It}, x_{it}\right) = Cov\left(a_i + u_{it}, x_{it}\right) = 0$, but this is not true if a_i and x_{it} are correlated.

Example 1: The effect of minimum wage on employment



- In February 1992 New Jersey (NJ) increased the state minimum wage from $4.25 to $5.05
- Card & Krueger (1994) wanted to analyse the impact of the minimum wage increase in New Jersey (NJ) and chose Pennsylvania (PA) as a comparison state

Example 1: The effect of minimum wage on employment

- Suppose C&K surveyed N fast food stores in NJ and PA *after* the minimum wage increase and run the regression model:

$$Y_{is} = \alpha_0 + \delta D_s + u_{is} \ ,$$

  where

  - Y_{is} is unemployment in store i in state s
  - $D_s \in \{1, 0\}$ and we think about it as a Treatment (and Control, or Comparison group) indicating exposure to the policy intervention; It is 1 in NJ after the policy change, and 0 otherwise

- In this model, all the unobserved heterogeneity is in u_i and it has mean zero, and we assume it is normally distributed $\left(u_i \sim N\left(0, \sigma^2\right)\right)$

Example 1: The effect of minimum wage on employment

- Assuming $Y_{is} = \alpha_0 + \delta D_s + u_{is}$ we make several modeling assumptions:
  - $E[u_{is}|D_{is}] = 0$: no omitted variable; this is like saying that the factors that we don't observe are on average the same in NJ and PA – *i.e.*, there is nothing in NJ and PA that can explain why NJ introduced the policy change, but not PA (treatment is randomly assigned within sample)
  - $E[Y_{is}|D_{is}] = \alpha_0$: This follows from the first assumption. The hidden implication is that at time $t = 0$, when both PA and NJ have $D_i = 0$, the unemployment levels in both states are similar on average
  - δ is constant: every unit is affected in the same way
- These are very strong assumptions – the first one in particular

Example 1: The effect of minimum wage on employment

- It is reasonable to think that the average unemployment, the outcome in absence of minimum wage laws (Y_{0i}), is determined by characteristics of each state s and by the month of the year t

- When the minimum wage law is introduced (Y_{1i}), the average unemployment shifts by some factor δ. We can write this assumption as:

$$E\left[Y_{0ist}|s,t\right] = \gamma_s + \lambda_t\,;\;\; E\left[Y_{1ist}|s,t\right] = \gamma_s + \lambda_t + \delta$$

- So the model for each fast food store i in state s at time t becomes

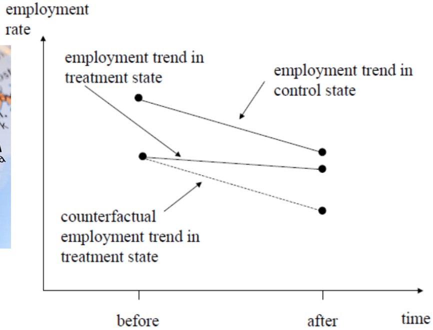
$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + u_{ist}$$

Example 1: The effect of minimum wage on employment

- Key modeling assumptions:
  - $E\left[u_{ist}|D_{st},\gamma_s,\lambda_t\right]=0 \rightarrow E\left[Y_{ist}|D_{st}=0,\gamma_s,\lambda_t\right]=\gamma_s+\lambda_t$: this is similar to saying that the unobserved factors that would cause problems when correlated with D_{st} are factors that need to vary across restaurants within state and within month. They may still exist (*e.g.*, some other policies that support employment) but this class contains less factors than the previous model
  - γ_s: There is state-level time-invariant heterogeneity; stores in each state are allowed to behave differently on average. Note that the state-level heterogeneity is constant over time
  - λ_t is constant across states: the period effect is constant across states; this means that the trend over time is the same across states. The assumption is imposing that the two states, *in absence of the policy*, would have had the same trend in unemployment over time. This is a key identifying assumption. It may well not be true
  - δ is constant – *i.e.*, every unit is affected in the same way

Example 1: The effect of minimum wage on employment

- The model requires a 2-period dataset: either a pooled cross-section of stores or a panel dataset
- C&K surveyed N fast food stores in NJ and PA before (February) and after (November) the policy change
- They assume that it is the levels of employment which evolve in the same way in PA and NJ



Example 1: The effect of minimum wage on employment

- For New Jersey:

$$
\begin{aligned}
E\left[Y_{ist}|s = NJ, t = Feb\right] &= \gamma_{NJ} + \lambda_{Feb} \\
E\left[Y_{ist}|s = NJ, t = Nov\right] &= \gamma_{NJ} + \lambda_{Nov} + \delta \\
E\left[Y_{ist}|s = NJ, t = Nov\right] - E\left[Y_{ist}|s = NJ, t = Feb\right] &= \lambda_{Nov} - \lambda_{Feb} + \delta
\end{aligned}
$$

- For Pennsylvania:

$$
\begin{aligned}
E\left[Y_{ist}|s = PA, t = Feb\right] &= \gamma_{PA} + \lambda_{Feb} \\
E\left[Y_{ist}|s = PA, t = Nov\right] &= \gamma_{PA} + \lambda_{Nov} \\
E\left[Y_{ist}|s = PA, t = Nov\right] - E\left[Y_{ist}|s = PA, t = Feb\right] &= \lambda_{Nov} - \lambda_{Feb}
\end{aligned}
$$

Example 1: The effect of minimum wage on employment

- The difference between NJ and PA between Nov and Feb is:

$$
\begin{aligned}
&(E\left[Y_{ist}|s = NJ, t = Nov\right] - E\left[Y_{ist}|s = NJ, t = Feb\right]) \\
&-(E\left[Y_{ist}|s = PA, t = Nov\right] - E\left[Y_{ist}|s = PA, t = Feb\right]) \\
&= \delta
\end{aligned}
$$

Example 1: The effect of minimum wage on employment

- This strategy and the parameter δ is called difference-in-differences (DiD)
- The model allows for state and time effects by taking the differences between means before and after the policy
- These parameters need not be estimated, but they are in the model by using a specific variation in the data for estimation
- The key identifying assumption is that of 'common trends': employment trends would have been the same in both states in the absence of the policy implementation

Exercise for next week

- Suppose the model is specified in terms of log of employment:

$$logY_{ist} = \gamma_s + \lambda_t + \delta D_{st} + u_{ist}$$

- What is now the assumption about the counterfactual trend? Can the assumption on 'common trend' using both logged outcome variable and non-logged outcome variable be true?

Using sample analogues

Variable	Stores by state: PA (i)	NJ (ii)	Difference, NJ − PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Figure 2: Findings from Card and Krueger (1994)

- The results of C&K suggest that minimum wages actually increased employment, but these results have been highly contested

Regression models for DiD

- There are two possible empirical approaches to estimate DiD with 2 period panel data:
  - Using an *interaction model* on the $N \times T$ observations
  - Using *first-differenced* data and estimate a model on $N \times T - 1$ observations
- The two approaches are equivalent for 2-period panel data but not for multiple time periods
- In a repeated cross-section, only the interaction model can be estimated. Sometimes, the quality of the data available does not allow to estimate the first-differenced estimator
- Also, using control variables is different in both specifications

Using an interaction model

- We should now adapt our theoretical approach into a regression equation
- We have two states and two periods and the model is:

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{ist} + u_{ist}$$

- This can be rewritten as:

$$Y_{ist} = \alpha + \gamma \cdot \mathbb{1}_{s=NJ} + \lambda \cdot \mathbb{1}_{t=Nov} + \delta \cdot \mathbb{1}_{s=NJ} \cdot \mathbb{1}_{t=Nov} + u_{ist}$$

Using an interaction model

- Taking conditional expectations for different states and periods for $Y_{ist} = \alpha + \gamma \cdot \mathbb{1}_{s=NJ} + \lambda \cdot \mathbb{1}_{t=Nov} + \delta \cdot \mathbb{1}_{s=NJ} \cdot \mathbb{1}_{t=Nov} + u_{ist}$ gives:

$$
\begin{aligned}
E\left[Y_{ist}|s = PA, t = Feb\right] &= \gamma_P A + \lambda_{Feb} = \alpha \\
E\left[Y_{ist}|s = PA, t = Nov\right] &= \gamma_P A + \lambda_{Nov} = \alpha + \lambda \\
E\left[Y_{ist}|s = NJ, t = Feb\right] &= \gamma_N J + \lambda_{Feb} = \alpha + \gamma \\
E\left[Y_{ist}|s = NJ, t = Nov\right] &= \gamma_N J + \lambda_{Nov} + \delta = \alpha + \gamma + \lambda + \delta
\end{aligned}
$$

Using First-differenced data

- If one has 2-period panel data (and the policy change occurs in the second period), the empirical model can be estimated on the data after first-differencing:

$$\begin{aligned} Y_{is2} - Y_{is1} &= \lambda_2 - \lambda_1 + \delta\left(D_{s2} - D_{s1}\right) + \left(u_{is2} - u_{is1}\right) \\ \Delta Y_i &= \lambda + \Delta D_i + \Delta u_i \end{aligned}$$

- Note that the intercept of the model is, in fact, the time trend and the state-effect has been differenced out

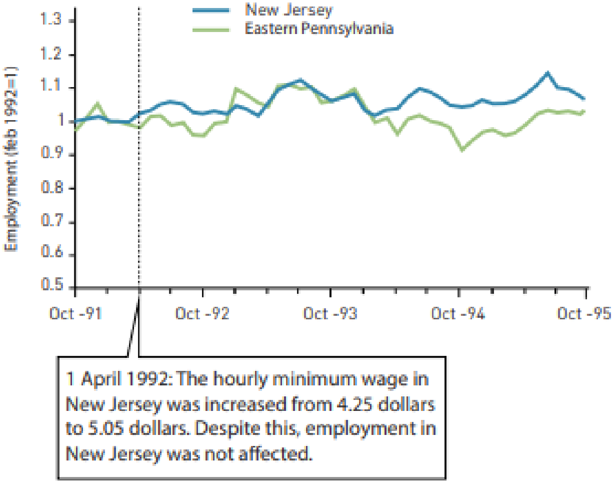
The issue with DiD: common or 'parallel' trends

- In the Card & Krueger example, the assumption that λ_t is the same across states ($=$ there is a common trend over time) is a key assumption
- If the trend in the two states is not the same, δ would not be distinguishable on its own (we say it cannot be identified)
- The DiD strategy hinges on the common trend assumption
- With 2-period panel data, the assumption is untestable (though some sensitivity analysis is possible)
- With multiple time period, the parallel trend assumption can be formally tested
- It must be said that failing to reject the null hypothesis of parallel trends does not equate to say the assumption holds for sure

When parallel trends assumption fails?

- *Different initial conditions*: changes in outcomes over time are a function of the differences in initial conditions between treated and comparison group (*e.g.*, general macroeconomic conditions in NJ and PA could have been different prior to the policy and react with lags, so stores in NJ (or PA) would have been on a very different trajectory and the trends would be different). Initial conditions may also affect the policy introduction: the key question is why they introduce the policy in the first place?
- *Heterogeneity of treatment impacts*: people with different observables may react differently to an intervention. Because of these differences in observables, it is reasonable to think different sets of people would have behaved differently in absence of the intervention: that is, people with different observables may have different trends in outcomes even without the intervention

When parallel trends assumption fails?



A formal test for parallel trends

- Suppose we have m pre-treatment periods
- We would like to test that at each pre-treatment period the difference between treatment can control is fixed
- A typical test involves lead effects to the treatment interacted with the treatment $(D = 1)$:
  - Let the treatment occur at $t = 0$, with the pre-treatment periods being $t = -1, -2, -3, \ldots, -m$

$$Y_{is(t)} = \gamma_s + \lambda_t + \sigma_{l=-m}^{-1} \alpha_{t=l} D_{is} + \delta_{t=0} D_{is} + u_{ist}$$

A formal test for parallel trends

- Using $Y_{is(t)} = \gamma_s + \lambda_t + \sigma_{l=-m}^{-1} \alpha_{t=l} D_{is} + \delta_{t=0} D_{is} + u_{ist}$ we get:

$$
\begin{aligned}
E\left[Y_{1(0)} - Y_{1(-1)} | D_s = 1\right] &= [\gamma_1 + \lambda_0 + \delta_0] - [\gamma_1 + \lambda_{-1} + \alpha_{-1}] = (\lambda_0 - \lambda_{-1} - \alpha_{-1}) + \delta_0 \\
E\left[Y_{0(0)} - Y_{0(-1)} | D_s = 0\right] &= [\gamma_0 + \lambda_0] - [\gamma_0 + \lambda_{-1}] = \lambda_0 - \lambda_{-1} \\
E\left[Y_{1(-1)} - Y_{1(-2)} | D_s = 1\right] &= [\gamma_1 + \lambda_{-1} + \alpha_{-1}] - [\gamma_1 + \lambda_{-2} + \alpha_{-2}] \\
&= (\lambda_{-1} - \lambda_{-2}) + (\alpha_{-1} - \alpha_{-2}) \\
E\left[Y_{0(-1)} - Y_{0(-2)} | D_s = 0\right] &= [\gamma_0 + \lambda_{-1}] - [\gamma_0 + \lambda_{-2}] = \lambda_{-1} - \lambda_{-2}
\end{aligned}
$$

A formal test for parallel trends

- $\alpha_{t=l}$ shows the average difference in trends between D in the pre-treatment periods leading up to the intervention relative to a period in discussion
- $\delta_{t=0}$ shows the average difference in outcomes for varying levels of D in the post-treatment period
- The assumption of parallel trend implies that

$$H_0 : \alpha_0 = \alpha_{-1} = \ldots = \alpha_{-m} = 0$$

- This is as F-test. If the H_0 is rejected, then the assumption of parallel trends is unlikely to hold. if you fail to reject, then there is evidence in favor of the parallel trend assumption, although we cannot say that it holds for sure (type II error)

Tests when pre-treatment data unavailable

- In some cases, there is only one observed point in time before the treatment in the data
- We can use *Placebo testing*
- Examples:
  - Use a different outcome variable, which you know is NOT affected by the treatment: Using the same control group and periods, one expects a DiD estimate equal to zero; if the DiD estimate is different from zero, there is a problem
  - Use the same outcome variable from a population that you know was NOT affected as a treatment group: One expects a DiD estimate equal to zero. If the DiD estimate is different from 0, the trends are unlikely to be parallel

When common trends are plausible?

- Group pre-intervention similarity in levels (and distributions): DiD will generally be more plausible if the treatment and control groups are similar in levels (averages and, ideally distributions) to begin with, not just in trends. Initial differences are an indicator of divergent trends. If there are differences, the key question is: why? And then, why should we think these differences would not impact trends?

- Absence of other shocks, background changes or 'special' periods: Even if no pre-treatment trend differences in outcome are detected, the existence of other policies, sectoral or social trends may induce changes in the post-treatment outcomes. The parallel trend assumption implies that these changes in outcomes are wrongly attributed to the effect of the treatment

What to do when the assumption fails?

- Allow for different trends in different groups within treatment
  - Here the concern is that units with different observables (*e.g.*, demographic characteristics) may have different trends, while instead we assume a common trend in the entire treatment and control group
  - A specification that relaxes this assumption includes interaction terms between the time fixed effects and observables characteristics measured pre-treatment: $\lambda_t \times X_{i(t=-m)}$. It is important that the characteristics are not outcomes themselves
- Allow for linear trend in different groups within treatment
  - The previous approach has a risk of over-fitting the data
  - An alternative is to still allow for different groups to have different pre-trends, but the trend is assumed to be linear – which makes it possible to estimate the potential bias due to the violation of parallel trends

Application: The impact of ‘16 and Pregnant’ TV show on teen childbearing (Kearney & Levine, AER 2015)

- K&L compare areas which had high pre-program viewership of MTV to areas with low pre-program viewership to show if the show led to differential changes in teen birth rates
- They have data on:
  - “designated market areas” j, which reflect the geography of the local television markets for which viewership data are available ($N = 205$)
  - The show first aired in June 2009 (treatment starts at Q3 of 2009) – the treatment is the rate of exposure to the show measured using US Nielsen ratings among 12–24 year-olds (viewership)
  - The outcome is birth rates, B_{jt}, occurring between January 2005 (Q1 of 2005) and December 2010 (Q4 of 2010). Birth rates are defined as ‘the number of births to girls between ages 15 and 19 in DMA j conceived in quarter t scaled by the population of women in each DMA’

Application: The impact of '16 and Pregnant' TV show on teen childbearing (Kearney & Levine, AER 2015)

- They estimate

$$ln\left(B_{jt}\right) = \beta_0 + \beta_1 Rate16P_j \times Post_t + \beta_2 U_{jy} + X'_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt}$$

  where t indexes quarters, j indexes media markets, and s indexes season of the year
- $ln\left(B_{jt}\right)$: teen birth rate in DMA j in quarter t
- $Post_t$: indicator variable for quarters after June 2009 when the show began
- U_{jy}: unemployment rate in DMA j in year y in which quarter t falls
- X_{jy}: percent non-Hispanic, black and Hispanic in j in year y
- θ_t: quarter fixed effects
- δ_{js}: DMA x season fixed effects to control for the seasonality that exists in teen birth rates

Baseline results

Table 1—Estimates of the Impact of *16 and Pregnant* Ratings on Teen Birth Rates

	OLS (1)	First stage (2)	IV (3)	Reduced form (4)
Dependent variable:	ln(birth rate)	*16 and Pregnant* ratings	ln(birth rate)	ln(birth rate)
16 and Pregnant ratings	−1.020*		−2.368**	
	(0.552)		(0.942)	
MTV ratings 2008–2009		1.511***		−3.581**
		(0.204)		(1.512)
Unemployment rate	−1.440***	−0.001	−1.487***	−1.485***
	(0.401)	(0.026)	(0.375)	(0.432)
F-statistic on omitted instrument		48.1		

Notes: The birth data used for this analysis represents quarterly birth rates by DMA for conceptions leading to live births between 2005 and 2010. The sample size in each model is 4,919 (205 DMAs, 24 quarters, and one observation was dropped because there were no teen births). Coefficients and standard errors (reported in parentheses) in birth rate regressions are multiplied by 100. Each model also includes the percentage of a DMA's female teen that is Hispanic and non-Hispanic black along with quarter and DMA × season fixed effects. Regressions are weighted by the relevant sample sizes for each outcome. Reported standard errors are clustered at the DMA level.

Figure 4: Table 1 from Kearney and Levine (2005) showing OLS and IV results

Testing parallel trends

- The authors conduct a test of parallel trends in the pre-treatment periods, using the following regression model, and cannot reject the F-test on α_i:

$$
\begin{aligned}
ln\left(B_{jt}\right) &= \beta_0 + \sum_{i=2005Q1}^{2008Q2} \alpha_i \left(Rate16P_j \times PreQ_i\right) + \sum_{i=2009Q3}^{2010Q4} \beta_i \left(Rate16P_j \times PostQ_i\right) \\
&+ \phi U_{jy} + X'_{jy}\gamma + \theta_t + \delta_{js} + \epsilon_{jt}
\end{aligned}
$$

- The first summation term counts from 18 quarters before the TV show was first aired to 5 quarters beforehand. The second summation term counts from the first quarter after the show started to 6 quarters afterwards

Testing parallel trends

- Since this controls for season fixed effect, they need to exclude four base period quarters, which they decide to be 2008Q3 to 2009Q2, essentially a year before the show began
- Hence, the reference reference time for β is the base year just before the show. The authors plotted the differences in birth rates associated with one point increase in MTV ratings, noting they are around 0%

Testing parallel trends

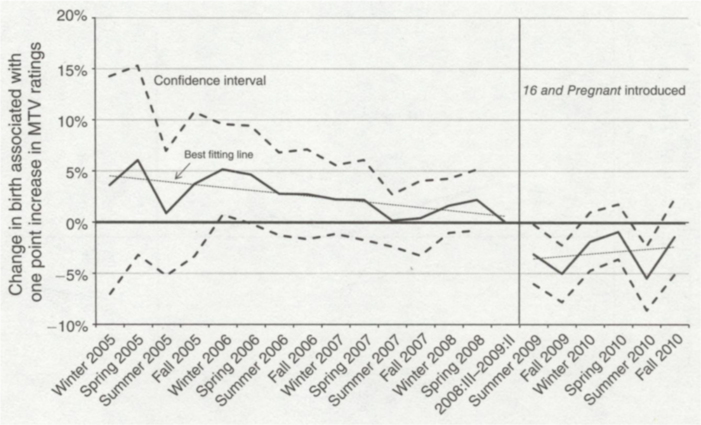
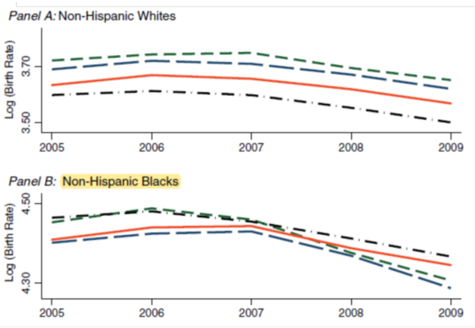


Figure 5: Figure 5 from Kearney and Levine (2005): reduced-form event study estimates of the impact of the show. It plots the coefficients of the interaction terms $(Rate16P_j \times PreQ_i)$ and $(Rate16P_j \times PostQ_i)$ and their confidence intervals.

Jaeger, Joyce and Kaestner (2019) re-analysis

- Jaeger, Joyce and Kaestner re-analyzed the K&L data, and argue that the pre-treatment viewership rates of MTV are correlated with factors like race and unemployment rates



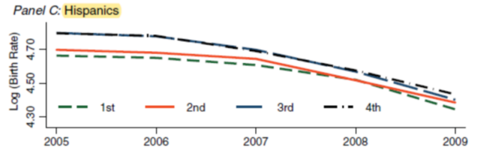


Figure 2. Log(Teen Birth Rate) by MTV ratings quartile and race.

Source: Authors' calculations using Bureau of the Census Census County Characteristics, National Vital Statistics System births, and Nielsen ratings data. The latter two data sources are confidential.

Notes: All panels shows weighted average annual Log(Birth Rate) for teenagers by quartile of MTV ratings where the weights are the female population aged 15-19 in the DMA. 2009 includes only the 2 quarters prior to the start of *16 and Pregnant*. Note that the y-axis scale is different in each panel.

Jaeger, Joyce and Kaestner (2019) re-analysis

- Using a specification that controls for differential time trends by an area's racial/ethnic composition or unemployment rate causes the result to disappear

	16 & Preg. Ratings or MTV ratings 2008—2009	Unemp. Rate	Percent Non–Hisp. Black	Percent Hispanic
Panel 1: Base model: Replication				
Reduced form	–3.581**	–1.485***	–2.305*	–2.812***
(KL)	(1.517)	(0.043)	(1.267)	(0.636)
Instrumental	–2.366**	–1.487***	–2.197*	–3.103***
variables (KL)	(0.941)	(0.375)	(1.197)	(0.553)
Panel 2: Base model plus covariates$_{2005}$ × Period Fixed Effects				
Reduced form	–0.404	–1.473***	–3.312***	–2.442***
	(1.504)	(0.360)	(1.012)	(0.553)
Instrumental	–0.238	–1.467***	–3.314***	–2.459***
variables	(0.800)	(0.318)	(0.914)	(0.489)
Panel 3: Base model plus covariates$_{2005}$ × Linear Trend				
Reduced form	–0.437	–1.414***	–3.295***	–2.470***
	–(1.309)	(0.349)	–(0.963)	–(0.536)
Instrumental	0.273	–1.410***	–3.302***	–2.484***
variables	(0.749)	(0.314)	(0.877)	(0.481)
Panel 4: Base model plus Race/Ethnicity$_{2005}$ × Linear Trend				
Reduced form	–1.237	–1.300***	–3.091***	–3.032***
	(1.494)	(0.429)	(1.150)	(0.631)
Instrumental	–0.776	–1.292***	–3.104***	–3.071***
variables	(0.847)	(0.384)	(1.046)	(0.561)

Figure 7: Table 1 from Jaeger, Joyce and Kaestner (2020): Impact when quarter

- Analyzing longer pre-treatment periods, JJK show that the pre-trends parallel trend test is rejected

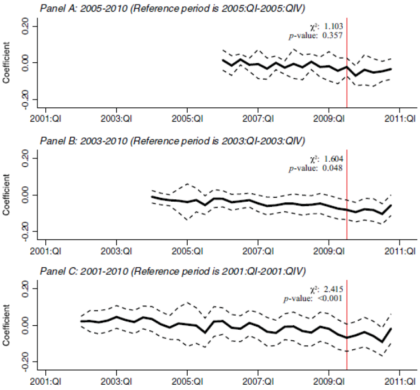


Figure 3. Reduced form event study: Extending the Pre-16 and Pregnant period.

Takeaway points

- Policy evaluation with after-policy data only are rarely convincing: groups are assumed to start from the same conditions and any other change occurring simultaneously with the introduction of the policy is (wrongly) attributed to the policy impact
- One can use data over time to allow for more flexible models
- The difference-in-difference estimator allows for time-fixed unobserved heterogeneity (different initial conditions) but relies on the assumption that the outcome evolves in the same way across treatment groups (common/parallel trends assumption)
- When $T = 2$ the parallel trend assumption is untestable
- When pre-intervention periods are available, a formal test is possible, but regardless of whether the test is passed or not, it is important to discuss whether it is reasonable to think the parallel trends assumption is justified

For next week

- Read JW Chapter 13 and 14
- Follow up on the papers mentioned today
- Exercise in slides (about correlation between error terms and the variance of individual fixed effects)
- For the seminar next week: Example 13.8 in JW Chapter 13; Exercise C12 in Chapter 13