WARMUP OBSERVATIONS
Where do the response dynamics come from?

Suppose:

\[
\begin{align*}
\Delta y_t &= \beta \Delta s_t + \rho \Delta y_{t-1} + u_t^y \\
\Delta s_t &= \theta \Delta s_{t-1} + u_t^s
\end{align*}
\]

\[u_t \sim D(0, I)\]

It's easy to see that:

\[
R_{ys}(h) = \beta \theta^h + \beta \theta^{h-1} + \ldots + \beta \theta \rho^{h-1} + \beta \rho^h
\]

\[\text{due to policy persistence} \]

\[R_{ss}(h) = \theta^h\]

Setting \(\theta = 0\):

\[
R_{ys}^*(h) = \beta \rho^h \quad \text{internal propagation}
\]
Actual response versus response conditional on future treatments

Simulated data from counterfactual.do

$R_{ys}$: dashed blue line; $R_{ss}$: purple. Lighter colors are $R^{*}_{ys}$ and $R^{*}_{ss}$.
Suppose that instead of $R_{ss}(h) = \theta^h$, you feed counterfactual $R_{ss}^c(h)$
Can show that:

$$
\begin{align*}
R_{ys}^c(0) &= R_{ys}^c(0)R_{ss}^c(0) \\
R_{ys}^c(1) &= R_{ys}^c(0)R_{ss}^c(1) + R_{ys}^c(1)R_{ss}^c(0) \\
R_{ys}^c(2) &= R_{ys}^c(0)R_{ss}^c(2) + R_{ys}^c(1)R_{ss}^c(1) + R_{ys}^c(2)R_{ss}^c(0) \\
R_{ys}^c(3) &= R_{ys}^c(0)R_{ss}^c(3) + R_{ys}^c(1)R_{ss}^c(2) + R_{ys}^c(2)R_{ss}^c(1) + R_{ys}^c(3)R_{ss}^c(0)
\end{align*}
$$

Example: counterfactual.do
Illustration: recovering the original response

Using original treatment path

\( \mathcal{R}_{ys} \): solid blue line; \( \mathcal{R}^*_y \mathcal{R}_{ss} \): dashed line
Counterfactuals in practice

Estimate usual LP but control for future treatments:

\[ y_{t+h} - y_{t-1} = a_h + b_h \Delta s_t + \sum_{j=1}^{h} c_{jh} \Delta s_{t+j} + d_h \Delta y_{t-1} + e_h \Delta s_{t-1} + v_{t+h} \]

Then \( b_h \) is an estimate of \( R^*_y(s)(h) \)
\( R^c_{ss}(h) \) is supplied by user given particular counterfactual
Counterfactual response versus original response
Simulated data from counterfactual.do

\[ \mathcal{R}_{ys}: \text{solid blue line}; \quad \mathcal{R}_{ys}^c = \mathcal{R}_{ss} + 0.25: \text{dashed blue} \]
Are these operations valid?

Mechanically: yes; causally: ?

- **Conditioning on future treatments**: are they randomly assigned?
- **Counterfactual treatment path**: how different from \( R_{ss}(h) \)?

Note:

\[
\left( R_{ss}^* - \hat{R}_{ss} \right)' \Sigma_{ss}^{-1} \left( R_{ss}^* - \hat{R}_{ss} \right) \rightarrow \chi^2_H
\]

IMPULSE RESPONSE HETEROGENEITY:
KITAGAWA–OAXACA–BLINDER DECOMPOSITIONS
Cloyne, Jordà, and Taylor (2020). Decomposing the fiscal multiplier
Potential outcomes: A static setup first  
Borrowing from applied micro

Think of observed $y$ as coming from a latent mixture:

$$y = (1 - s) y_0 + s y_1 = y_0 + s (y_1 - y_0); \quad s = 0, 1$$

Assumption:

$$y_i \sim f(\mu_j; \sigma_j); \quad j = 0, 1 \quad \text{unobservable random variables}$$

we would like: $E(y_1 - y_0)$ average treatment effect
Assume linear model for latent variables: $y_j; j = 0, 1$

let $y_j = \mu_j + v_j$, $E(v_j) = 0, j = 0, 1$. $v_j$ captures heterogeneity

let $v_j = (x - \mu_x) \gamma_j + \epsilon_j$ with $E(\epsilon_j) = 0$ and $E(\epsilon_j|x) = 0$

then:

$$E_x[E(y_1|s = 1; x) - E(y_0|s = 0; x)] = [\mu_1 + E_x[E(x - \mu_x|s = 1)] \gamma_1]$$

$$- [\mu_0 + E_x[E(x - \mu_x|s = 0)] \gamma_0]$$

add/subtract counterfactual: $E_x[E(x - \mu_x|s = 1)] \gamma_0$

$$ATE = (\mu_1 - \mu_0) + E_x[E(x - \mu_x|s = 1)] (\gamma_1 - \gamma_0) +$$

$$E_x [E(x - \mu_x|s = 1) - E(x - \mu_x|s = 0)] \gamma_0$$
Kitagawa-Oaxaca-Blinder decomposition components

recall:

\[
ATE = (\mu_1 - \mu_0) + \underbrace{E_x[E(x - \mu_x|s = 1)]}_\text{direct} (\gamma_1 - \gamma_0) + \underbrace{E_x [E(x - \mu_x|s = 1) - E(x - \mu_x|s = 0)]}_{\text{indirect}} \gamma_0
\]

**direct**: ATE under random assignment

**indirect**: treatment spillovers on covariates

**composition**: failure of random assignment? small sample bias
interesting null hypotheses
linear case, still working through applied micro motivation

\[ y_i = \mu_0 + (x_i - \bar{x}_0) \gamma_0 + s_i [\beta + (x_i - \bar{x}_1) \theta] + \omega_i \]

note: \( \beta = \mu_1 - \mu_0; \ \theta = \gamma_1 - \gamma_0; \) and \( \omega_i = \epsilon_{0,i} + s_i (\epsilon_{1,i} - \epsilon_{0,i}) \)

hence:

- \( H_0 : \beta = 0 \) null of no direct treatment effect
- \( H_0 : \theta = 0 \) null of no indirect effect
- \( H_0 : E(x|s = 1) - E(x|s = 0) = 0 \) null of no composition effect
- \( H_0 : \gamma_0 = 0 \) null of random assignment
  (hence no composition effect possible)
What does this mean for local projections?

let \( y_t = (y_t, y_{t+1}, \ldots, y_{t+H}) \) and \( y \) denote the associated r.v.

assume conditional mean independence

let \( E(y_s) = \mu_s \) for \( s \in \{0, 1\} \), wlog \( y_s = \mu_s + v_s \)
under linearity \( v_s = (x - \mu_x) B_s + \epsilon_s \), then:

\[
E(y_s | x) = \mu_s; \quad E(v_s) = 0; \quad E(\epsilon_s | x) = 0; \quad s \in \{0, 1\}
\]

note: Angrist et al. (2017) assume stronger conditional ignorability

hence:

\[
y_{t+h} = \mu_0^h + (x_t - \bar{x}) \gamma_0^h + S_t \beta^h + S_t (x_t - \bar{x}) \theta^h + \omega_{t+h};
\]

usual local projection \quad Kitagawa term

\[
h = 0, 1, \ldots, H; t = h, \ldots, T.
\]
Kitagawa decomposition components

recall:

\[ y_{t+h} = \mu_0^h + (x_t - \bar{x})\gamma_0^h + s_t \beta^h + s_t (x_t - \bar{x})\theta^h + \omega_{t+h}; \]

usual local projection

Kitagawa terms

\[ h = 0, 1, \ldots, H; \ t = h, \ldots, T. \]

direct effect:

\[ \hat{\mu}_1^h - \hat{\mu}_0^h = \hat{\beta}^h \]

indirect effect:

\[ (\bar{x}_1 - \bar{x})(\hat{\gamma}_1^h - \hat{\gamma}_0^h) = (\bar{x}_1 - \bar{x})\hat{\theta}^h \]

composition effect:

\[ (\bar{x}_1 - \bar{x}_0)\hat{\gamma}_0^h \]

ergodicity: needed to ensure \( \bar{x} \rightarrow \mu_x \)
Implications: state-dependence

note: suppose $x = x^*$ then total response is:

$$E(y_1|x^*, s = \delta) - E(y_0|x^*, s = 0)$$

$$= \delta \mu_1 + \delta [x^* - E(x)] \gamma_1 - \{ \mu_0 + [x^* - E(x)] \gamma_0 \}$$

$$= \delta \beta + \delta [x^* - E(x)] \theta,$$

remarks:

- dependence on $x^*$ is only partial equilibrium
- need identification (instruments) for $x$
- usual single variable stratification omits other terms in $x \rightarrow$ bias
Example from a previous experiment: two episodes

how effective was monetary policy in ...

1. November 1987 (post-stock market crash)
   - stocks 23% lower by end of October
   - Fed lowered funds rate 50bps

2. February 1996 (middle of a long expansion)
   - middle of stable funds rate

idea: two different scenarios, but similar policy paths → differences not due to different policy
funds rate path nearly identical and to baseline
baseline is average over the sample

Federal funds rate

(a) November 1987
(b) February 1996
policy unable to boost activity post-1987 crash
policy as usual February 1996

Industrial production

(a) November 1987

(b) February 1996
Example: GDP response to fiscal policy varies with monetary stance
Cloyne, Jordà, and Taylor 2023
Variation in the multiplier by horizon and stance

![Graph showing variation in the multiplier by horizon and stance. The x-axis represents standard deviations, and the y-axis represents the multiplier. The graph includes lines for different horizons (h=0, h=2, h=3) and shows how the multiplier varies with horizon and stance (looser vs. tighter).]
Time varying estimates of the multiplier
Response of real GDP to 1pp fiscal consolidation

90% error bands
Response of real GDP to 1pp fiscal consolidation

90% error bands
Response of real GDP to 1pp fiscal consolidation

90% error bands
PANEL DATA APPLICATIONS
DIFFERENCES-IN-DIFFERENCES WITH LPs
DUBE, GIRARDI, JORDÀ AND TAYLOR
TWFE implementation of DiD (static or distributed lags) can be severely biased.
  - Estimate is an average with possibly negative weights. Bad!
LP-DiD = local projections + clean controls (Cengiz et al 2019)
  - No negative weights. Good!
  - Simple reweighting to recover ATT
Background

Difference-in-Differences (DiD)

2x2 Setting

Staggered Setting

(Visual examples from Goodman-Bacon, 2021)
The conventional (until recently) DiD estimator: TWFE

- let $P_t = 1$ for post, 0 for pre; $A_i = 1$ for treated, 0 for control.

- Static TWFE

$$y_{it} = \alpha_i + \delta_t + \beta^{TWFE} D_{it} + \epsilon_{it}; \quad D_{it} = P_t \times A_i$$

- Event-study (distributed lags) TWFE

$$y_{it} = \alpha_i + \delta_t + \sum_{m=-Q}^{M} \beta_{m}^{TWFE} D_{it-m} + \epsilon_{it}$$

- OK in the 2x2 setting, or when treatment occurs at the same time.

- Biased even under parallel trends with staggered treatment, if treatment effects are dynamic and heterogeneous.
The problems with TWFE in the staggered setting

- TWFE as weighted-average of 2x2 comparisons (Goodman-Bacon 2021)
  1. Newly treated vs Never treated;
  2. Newly treated vs Not-yet treated;
  3. Newly treated vs Earlier treated.
The problems with TWFE in the staggered setting

- TWFE as a weighted-average of cell-specific ATTs (de Chaisemartin & D’Haultfoeuille 2020)

\[
E \left[ \hat{\beta}^{TWFE} \right] = E \left[ \sum_{(g,t):D_{gt}=1} \frac{N_{g,t}}{N_1} w_{g,t} \Delta_{g,t} \right]
\]

→ Weights can be negative!
LP-DiD Estimator
No Covariates, Outcome Lags

\[ y_{i,t+k} - y_{i,t-1} = \beta^k_{LP-DiD} \Delta D_{it} \quad \} \text{treatment indicator} \]
\[ + \delta^k_t \quad \} \text{time effects} \]
\[ + e^k_{it} ; \quad \text{for } k = 0, \ldots, K. \]

restricting the sample to observations that are either:

\[ \begin{cases} 
\text{treatment} & \Delta D_{it} = 1, \\
\text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k. 
\end{cases} \]

Key advantage of LP over distributed lags TWFE formulation of DiD:
differencing is in outcomes, not treatments.
LP-DiD Estimator

\[ y_{i,t+k} - y_{i,t-1} = \beta_k^{LP-DiD} \Delta D_{it} \]
\[ + \sum_{p=1}^{P} \gamma_{0,p}^k \Delta y_{i,t-p} \]  \{ treatment indicator \}
\[ + \sum_{m=1}^{M} \sum_{p=0}^{P} \gamma_{m,p}^k \Delta x_{m,i,t-p} \]  \{ outcome lags \}
\[ + \delta_{t}^k \]  \{ covariates \}
\[ + e_{it}^k \]  \{ time effects \}
\[ ; \]
for \( k = 0, \ldots, K \).

restricting the sample to observations that are either:

\[
\begin{cases}
\text{treatment} & \Delta D_{it} = 1, \\
\text{clean control} & \Delta D_{i,t+h} = 0 \text{ for } h = -H, \ldots, k.
\end{cases}
\]
An equivalent specification to implement LP-DiD

- Instead can use dummies to rule out unclean controls

\[ y_{i,t+k} - y_{i,t-1} = \beta^k L\!P\!-
\!D\!i\!D \Delta D_{it} + \theta^k UC_{i,t} \]
\[ + \sum_{p=1}^P \gamma^k_{0,p} (1 + \rho^k_{0,p} UC_{i,t}) \Delta y_{i,t-p} \]
\[ + \sum_{m=1}^M \sum_{p=0}^P \gamma^k_{m,p} (1 + \rho^k_{m,p} UC_{i,t}) \Delta x_{m,i,t-p} \]
\[ + \delta^k_t (1 + \phi^k_t UC_{i,t}) \]
\[ + e_{it}^k; \]  

\( UC_{it} = 1 \) if previously treated.

- With absorbing treatment, \( UC_{it} = \sum_{j=-H(j\neq0)}^k \Delta D_{i,t+j} \)
Simulation Evidence

- **N=500 units; T=50 time periods.**
- **DGP:**
  \[ Y_{0it} = \rho Y_{0,i,t-1} + \lambda_i + \gamma_t + \epsilon_{it}; \quad -1 < \rho < 1; \quad \lambda_i, \gamma_t, \epsilon_{it} \sim N(0, 25) \]
- Binary staggered treatment.
- **TE grows in time for 20 periods, and is stronger for early adopters.**
  1. **Exogenous treatment**
     - Units randomly assigned to 10 groups of size \(N/10\)
     - One group never treated; others treated at \(\tau = 11, 13, 15 \ldots, 27\).
  2. **Endogenous treatment**
     - Probability of treatment depends on past outcome dynamics.
     - Negative shocks increase probability of treatment.
     - Parallel trends holds only conditional on outcome lag.
Average estimates and 95% and 5% percentiles from 200 replications.
Simulation Evidence
endogenous treatment scenario

Average estimates and 95% and 5% percentiles from 200 replications.

- Leblebicioglu & Weinberger (2020) use static & event-study TWFE to estimate effects on the labor share.
- Negative effect of *inter-state* banking deregulation ($\approx -1\text{p.p.}$).
- No effect of *intra-state* branching deregulation.
TWFE estimates

- Negative effect from inter-state
- No effect from intra-state
Forbidden comparisons in the TWFE specification

- TWFE uses ‘forbidden’ comparisons: earlier liberalizers are controls for later liberalizers.


- Contribution of unclean comparisons to TWFE estimates:
  - 36% for inter-state banking deregulation;
  - 70% for intra-state branching deregulation.
Effect of banking deregulation on the labor share
LP-DiD estimates

(a) Inter-state banking deregulation

(b) Intra-state branching deregulation

- LP-DiD avoids unclean comparisons & allows controlling for y lags.
- Negative effect of inter-state branching deregulation is confirmed.
- But also intra-state branching deregulation has negative effect.