Model Specifications

EC 320: Introduction to Econometrics

Philip Economides
Winter 2022

Prologue

Housekeeping

Data Projects are in!

- I will have those graded by the weekend
- Points based on the quality of submissions and presentation of results

Problem Set 5

- Due Monday, March 7th

Model Specification

Concerns the following questions of a given model:

What are the consequences of including in the regression model a variable that should not be there?

Multicollinearity

What are the consequences of excluding a variable that should be featured?

Omitted Variable Bias

How do we test restrictions to model specifications?

t and F tests

Model Specification

What happens if you have difficulty finding data on a variable and use a proxy instead?

Today's lesson

We will then go a useful approach to know for policy analysis: the **difference-in-difference** model

Proxies in Model Specifications

Proxies

Suppose you are considering the following model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

While X_1 is observed, lets consider a case where X_2 is an unobserved variable. Cases of unobservable data could include:

- Vaguely defined with no explicit measure (e.g. quality of education)
- Intangible and cannot be quantified (e.g. utility, ability)
- Requires so much time/energy that measurement is infeasible
- Confidentiality, privacy concerns may limit observed data availability

Proxies

Rather than exclude the unobservable, you may wish to use an adequate **proxy** variable for X_2.

A **proxy** variable is a substitute variable that may reasonably be supposed to maintain similar properties to our missing variable.

Example: For quality of education, we could use the student-staff ratio to have a measure of how many resources the educational institution makes available to students. Where quality is high, student-staff ratios are low.

Proxies

Returning to the model, our true data generating process for Y:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

In the case where we have no data on X_2, suppose we have an **ideal proxy** for it such that there exists an *exact linear relationship* X_2 *and* Z:

$$X_2 = \lambda + \mu Z,$$

where λ and μ are fixed, unknown constants. We cannot estimate them by running a regression of the above relationship, since we have no data on X_2. **Let's sub in our expression and see what using Z achieves**.

Inference using Proxies

Using $X_2 = \lambda + \mu Z$,

$$\begin{aligned} Y =& \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \\ =& \beta_0 + \beta_1 X_1 + \beta_2 (\lambda + \mu Z) + u \\ =& (\beta_0 + \beta_2 \lambda) + \beta_1 X_1 + \beta_2 \mu Z + u \\ =& \quad \alpha \quad\quad + \beta_1 X_1 + \quad \gamma_2 Z + u \end{aligned}$$

1. β_1, its standard error and t-stat will be same as if X_2 used
2. R^2 will be the same as if X_2 was used instead of Z
3. The coefficient of Z, γ_2, will be estimate of $\beta_2 \mu$, and not possible to estimate β_2 directly
4. t-stat for γ_2 same as β_2, so able to assess significance of X_2
5. Not possible to estimate β_0 since we now only see α

Risks of using Proxies

Validity of all the subsequent takeaways rely on the condition that Z is an ideal proxy for X_2

- In practice, unusual to find a proxy that is exactly linearly related to our missing variable
- If the relationship is close to linear, the results will hold **approximately**
- The biggest problem faced is that there is never any manner by which to test this condition of whether the approximated difference is sufficiently small
- Critical thinking in explaining your logical link between X_2 & Z and accepting that you are relying on a strong assumptions are often required to convince an audience that a proxy is indeed valid

Difference-in-Difference (DiD)

Non-Random Treatment

Previously we'd talked about binary categorical variables, known as dummies. An important application of **dummy variables** is to study the impact of a treatment.

The estimation of treatment effects is important in a wide range of fields.
Examples:

- impact of cash transfers on child health
- effect of class size on student achievement

In clinical trials of health interventions, a common question is whether use of a medicine will improve health outcomes.

Non-Random Treatment

Ideally treatment effects are evaluated using **randomized controlled trials**.

If, for convenience, there is only one treatment level, for those treated, then this can be captured by a binary dummy variable, and a simple statistical model can be used:

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

for $i = 1, \ldots, n$, where $Di = 1$ if treated and $Di = 0$ if not treated, and where y_i is the outcome for individual i.

Random assignment: D_i and u_i are **independent (A3.)**, and an OLS regression will yield an **unbiased** estimate $\hat{\beta}_1$ of β_1. Our effect is the diff between mean outcomes for treated and untreated groups, $\hat{\beta}_1 = \bar{y}_1 - \bar{y}_0$. Also known as the **differences estimator**.

Non-Random Treatment

Often we are analyzing an intervention (or *treatment*) that takes place **over time**, and we have data **both before and after** the treatment took place.

If the policy intervention affected everyone and we have data both in pre- and post-policy periods, might consider the **differences estimator**

> **Problem:** Implicitly assumes that there are not other reasons for the change in mean outcomes before and after the treatment

In this case the dummy variable D_i in effect stands for the impact of **all factors**, including the treatment, that are different between the pre- and post- periods.

Non-Random Treatment

Since **other factors** may be partially responsible for the change in means, this estimator may be a biased estimate of the treatment effect.

This point emphasizes importance of control group that does not receive the treatment. Occasionally policy changes have randomization built in **(randomized field experiments)**.

> **Example:** Oregon Health Plan (OHP) experiment of 2008. Oregon that year decided to expand coverage of its version of Medicaid to groups not previously eligible, but for reasons of expense this expanded coverage was rationed by a lottery: somewhat under half of those who registered were randomly selected and invited to apply for expanded coverage. It is then possible to estimate the treatment effect on various outcomes by looking at differences in the sample means of the two groups.

DiD: Concept

In **most cases**, this randomization aspect to policy change does not occur. Due to the paucity of randomized field experiments, applied econometricians have shown great interest in **natural experiments**.

> **Setting:** There is a discrete change in policy that affects only part of the population of interest, and that there are both pre- and post-treatment observations for both those who received treatment and those who did not.

Although the policy change occurs in a non-experimental setting, in some cases the change may be introduced in a way that makes it appear **as if** the treatment was **randomly assigned**.

The **difference-in-difference (DiD)** method recognizes that in absence of random assignment, treatment and control groups are likely to differ for many reasons.

DiD: Structure

Suppose we have data on specific individuals $i = 1, \ldots, n$ in each of two time periods $t = 1, 2$ that correspond to before and after the treatment was applied to some of the individuals. A simple statistical formulation of the model is then:

$$Y_{it} = \beta_0 + \beta_1 T_t + \beta_2 D_i + \beta_3 (T_t \times D_i) + u_{it}$$

for $i = 1, \ldots, n$, and $t = 1, 2$. (Note that we have $2n$ data points).

Here $T_t = 1$ if $t = 2$ and $T_t = 0$ if $t = 1$, while $D_i = 1$ if individual i receives treatment and $D_i = 0$ if individual i does not receive treatment.

DiD: Structure

In the **pre-treatment period**, $t = 1 \implies T_t = 0$, we have

$$E(y_{i1}|D_i = 0) = \beta_0 \text{ and } E(y_{i1}|D_i = 1) = \beta_0 + \beta_2$$

In the above specification, we are allowing for the possibility that in the quasi-experiment there was imperfect control, in the sense that those treated had different pre-treatment means than those not treated.

In the **post-treatment period**, $t = 2 \implies T_t = 1$, we have

$$E(y_{i2}|D_i = 0) = \beta_0 + \beta_1 \text{ and } E(y_{i2}|D_i = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

The **treatment effect** corresponds to β_3: This is because

$$\begin{aligned} &\text{Diff Post-Period: } E(y_{i2}|D_i = 1) - E(y_{i2}|D_i = 0) = \beta_2 + \beta_3 \text{ and} \\ &\text{Diff Pre-Period: } E(y_{i1}|D_i = 1) - E(y_{i1}|D_i = 0) = \beta_2 \end{aligned}$$

DiD: Example

"Monetary Intervention in the Great Depression" by **Gary Richardson and William Troost (JPE, 2009)**.

Consider the 1930s bank failures. Central Bank can prevent bank runs and bank failures by acting as lender of last resort to solvent banks.

The twelve different regional Federal Reserve Districts reacted in the 1930s to potential bank failures in different ways.

- Atlanta FRB (sixth district) and St Louis FRB (eighth district) followed radically different policies wrt bank runs & border between two regions runs East-West through the center of Mississippi
- Makes for a natural experiment, comparing outcomes for bank failures in Mississippi banks in the two districts

DiD: Example

The **Sixth District policy** favored using the FRB lender of last resort role to lend to troubled banks, while **Eighth District policy** took the view that credit should be restricted in recessions.

R&T analyze this using a **DiD approach**.

Bank Losses by Policy

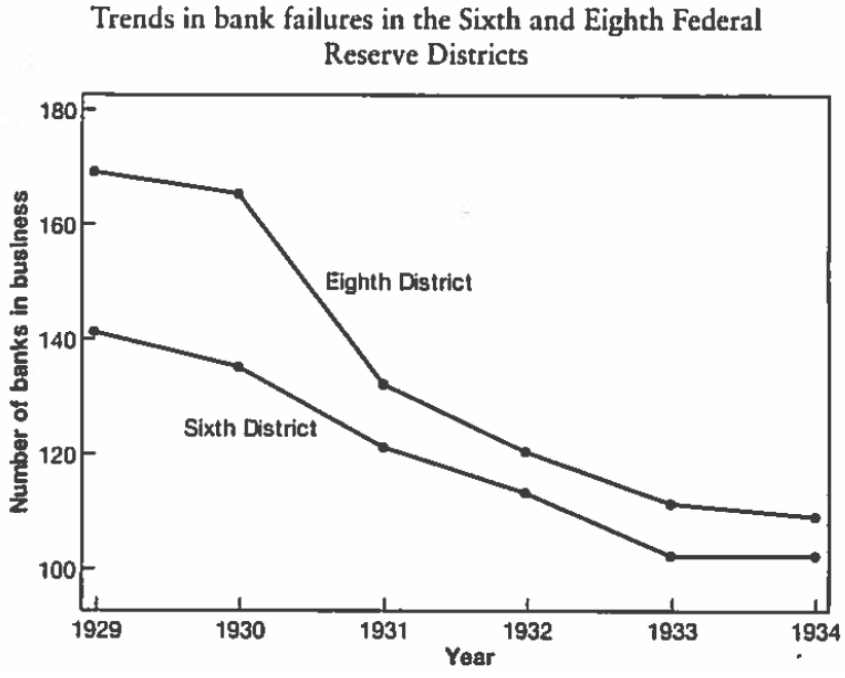
Variable	8th District	6th District	Diff., 6th-8th
No. of Banks open 1930	165	135	-30
No. of Banks open 1931	132	121	-11
Changes in banks open	-33	-14	19

Estimated to have saved 19 banks, 14% of Sixth District in 1930.

DiD: Required Assumption

A central question in assessing quasi-experiments is the **validity** of the control group, so that the treatment is really *as if* it was randomly assigned.

A key issue is the implicit **common trends** assumption: in the absence of the *"treatment"*, would the number of open banks in the two districts have been expected to evolve in the same way?



DiD: Interpretation

DiD Model Results

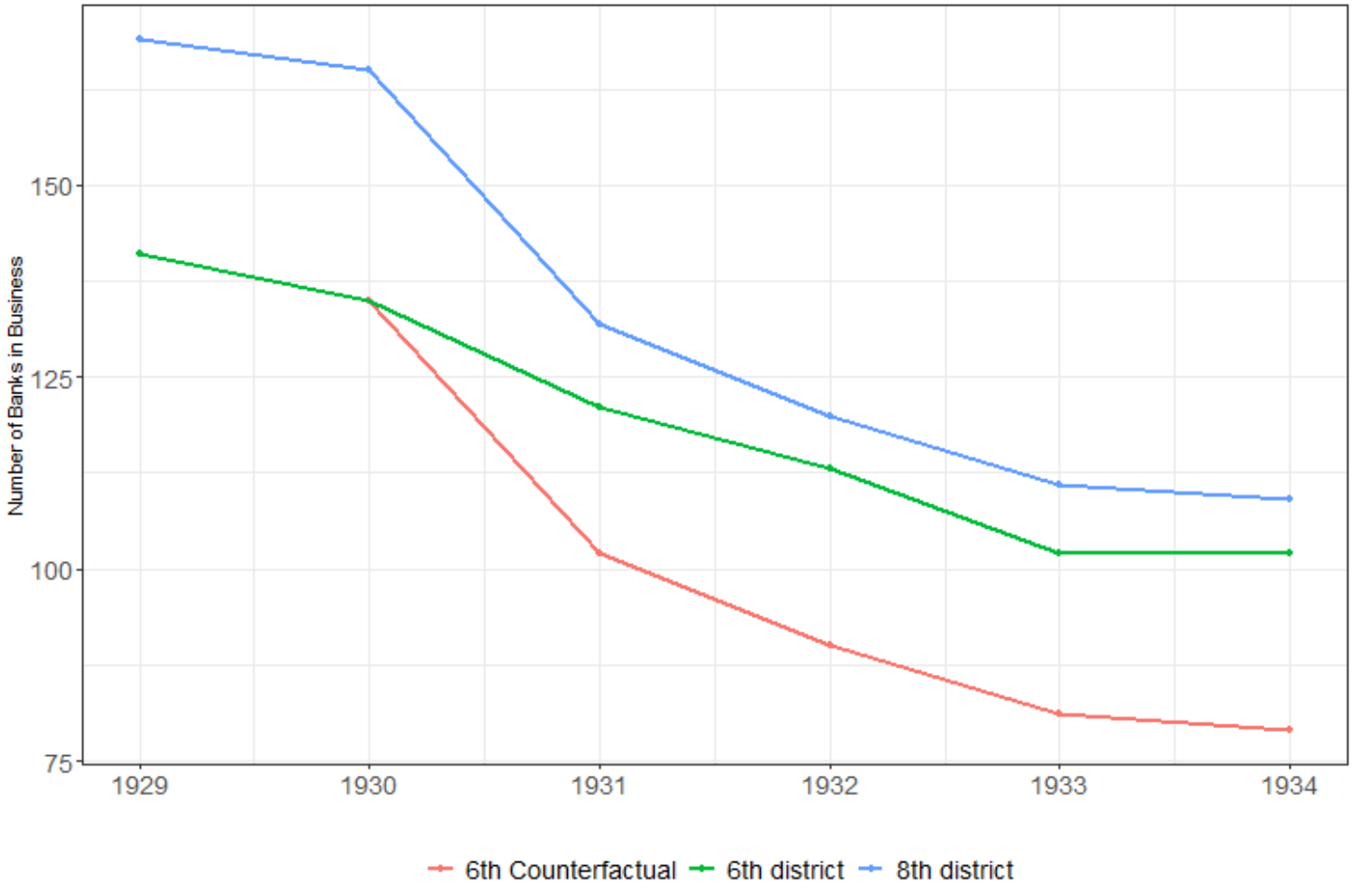
One can use this data to get an estimate of the treatment effect $\hat{\beta}_3$ and a standard error $SE(\hat{\beta}_3)$ for the effect based on these twelve data points:

$$Y_{dt} = 167 - 49T_t - 29D_d + 20.5(T_t \times D_d) + u_{dt}$$

Here T_t is coded 1 for 1931 or later, and $D_d = 1$ for the 6th District. With a t-score of 1.9 and using a two-tailed test, significant at 10% level.

This is about the most evidence one could hope for with twelve data points.

DiD: Interpretation



Let's do the same **DiD** with other outcome variables!

DiD: Interpretation

	1929	1933	Difference (1933-1929)
Panel A. Number of Wholesale Firms			
Sixth Federal Reserve District (Atlanta)	783	641	-142
Eighth Federal Reserve District (St. Louis)	930	607	-33
Difference (Sixth-Eight)	-147	34	**181**
Panel B. Net Wholesale Sales ($ million)			
Sixth Federal Reserve District (Atlanta)	141	60	-81
Eighth Federal Reserve District (St. Louis)	245	83	-162
Difference (Sixth-Eight)	-104	-23	**81**

Notes: This table presents a DiD analysis of Federal Reserve liquidity effects on the number of wholesale firms and the dollar value of their sales, paralleling the DiD analysis of lquidity effects on bank activity.

This **policy change** seems to have prevented a harsher impact on existing firms, when assessing sales performance and numbers of firms.

DiD: MLDA Laws

MLDA Laws

This DiD estimates MLDA-induced deaths among 18-20 year olds, from 1970-1983.

1933: End of the federal alcohol Prohibition. Most states regulated the **Minimum Legal Drinking Age** to 21. However, some states including Kansas, New York and North Carolina maintained a drinking age of 18.

1971: Voting ages were reduced to 18, and many states coincided this policy change of MLDA to 18. Arkansas, California, and Pennsylvania are among states that kept MLDA at 21.

1984: National Minimum Drinking Age Act punished states set to 18 by withholding federal highway construction aid.

What happened between 1970 and 1984 when drinking age was reduced?

MLDA: DiD model

Dependent variable: death rates of 18-20 year olds in state s and year t. For simplicity, sample includes only Alabama (treated) and Arkansas (control).

$$Y_{st} = \beta_0 + \beta_1 \mathrm{D}_s + \beta_2 \mathrm{T}_t + \beta_3 (D_s \times T_t) + u_{st},$$

where D_s indicates whether a given state is Alabama, and T_t is a post-period dummy for whether observations are from 1975 onward when Alabama introduced an MLDA of 19.

Key Assumption: Pre-policy (prior to 1975), Alabama and Arkansas shared parallel trends in their death rates of 18-20 year olds.

MLDA: DiD model

What about the other states? Would we not want to use more observations? Ideally, yes, but this will introduce further econometric challenges.

- States did not all lower to 18 at the same time, meaning there is no common post-treatment period.
- Unobservable differences exist across states, constant across time.
- Absense of a common treatment variable in a multi-state setting, since states chose ages of 18,19 and 20.

To solve for these issues, use **year dummies**, **state fixed effects**, and replace $T_t \times D_d$ with a **common treatment variable**, $LEGAL_{st}$, which measures the proportion of 18-20 year-olds allowed to drink in state s and year t.

MLDA: DiD model

The **multistate regression** DiD model looks like:

$$Y_{st} = \alpha_0 + \alpha_1 LEGAL_{st} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \alpha_k \text{State}_{ks} + \sum_{j=1971}^{1983} \alpha_j \text{Year}_{jt} + \nu_{st},$$

where now we are using state and year dummies to capture fixed effects, which uses Alabama as our reference state and 1970 as our reference year.

State-Year Panel Data: Including District of Col. we have 51 "states" and 14 years, giving us 714 observations.

Regression results suggest that there was a significant increase in deaths, largely attributed to motor vehicle accidents.

MLDA: Results

Dependent Variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	0.59	0.47	1.49	1.26
	(0.59)	(0.79)	(0.88)	(0.89)
All internal causes	1.33	0.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State Trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Notes: This table reports regression DiD estimate of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18-20 year-olds. The table shows coefficients on the proportion of legal drinkers by state and year from models controlling for state and year fixed effects. The model used to construct (2) and (4) include state-specific linear time trends.

MLDA: Assumptions

The inclusion of many states and many years of observations allows us to relax the **common trends** assumption.

Including **state trends** violates this assumption, by allowing for a degree of non-parallel evolution in outcomes between state in the absense of a treatment effect.

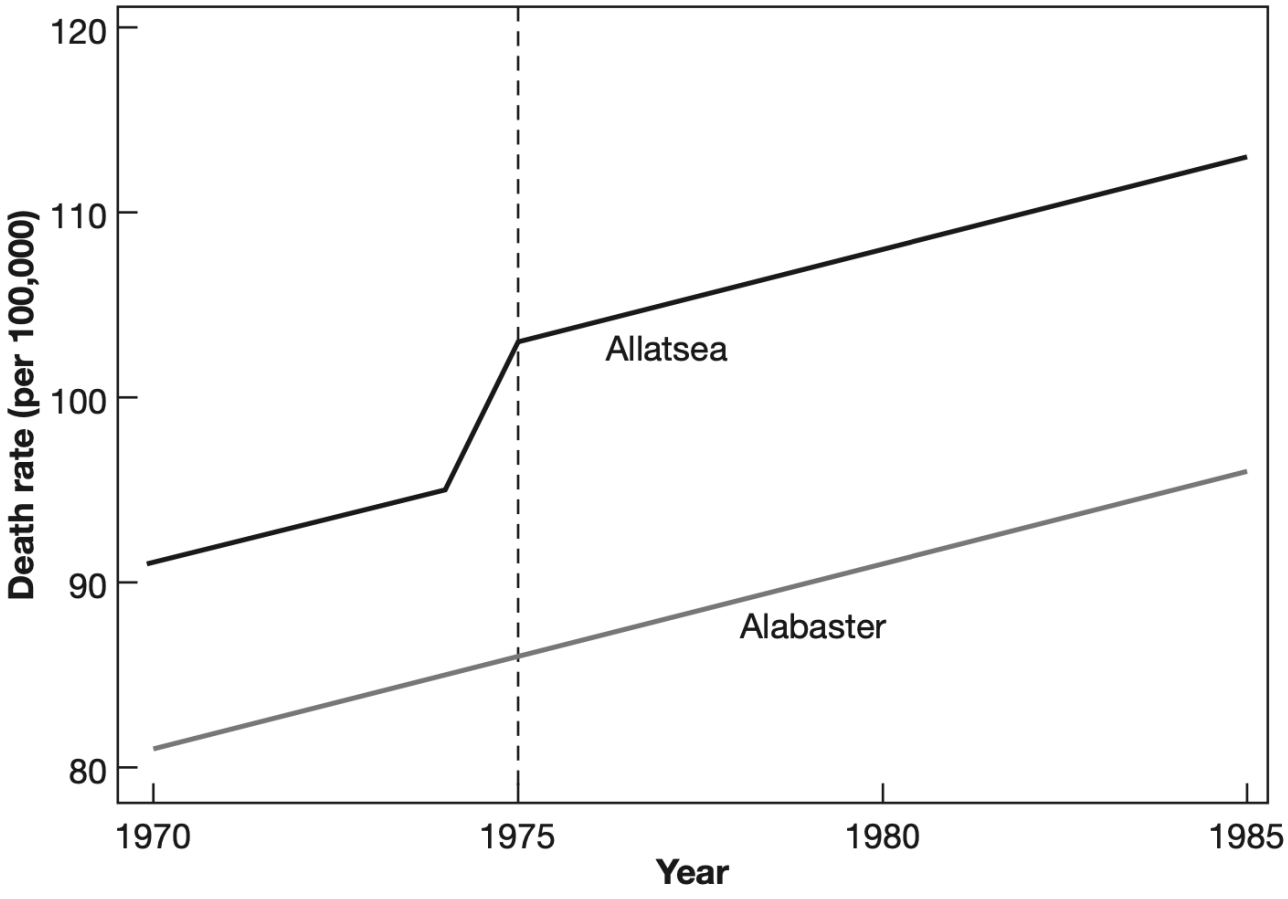
With state trends factored in:

$$Y_{st} = \alpha_0 + \alpha_1 LEGAL_{st} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \alpha_k \text{State}_{ks} + \sum_{j=1971}^{1983} \alpha_j \text{Year}_{jt} + \sum_{k=\text{Alaska}}^{\text{Wyoming}} \theta_k (\text{State}_{ks} \times t) + \varepsilon_{st},$$

What does this allow us to address?

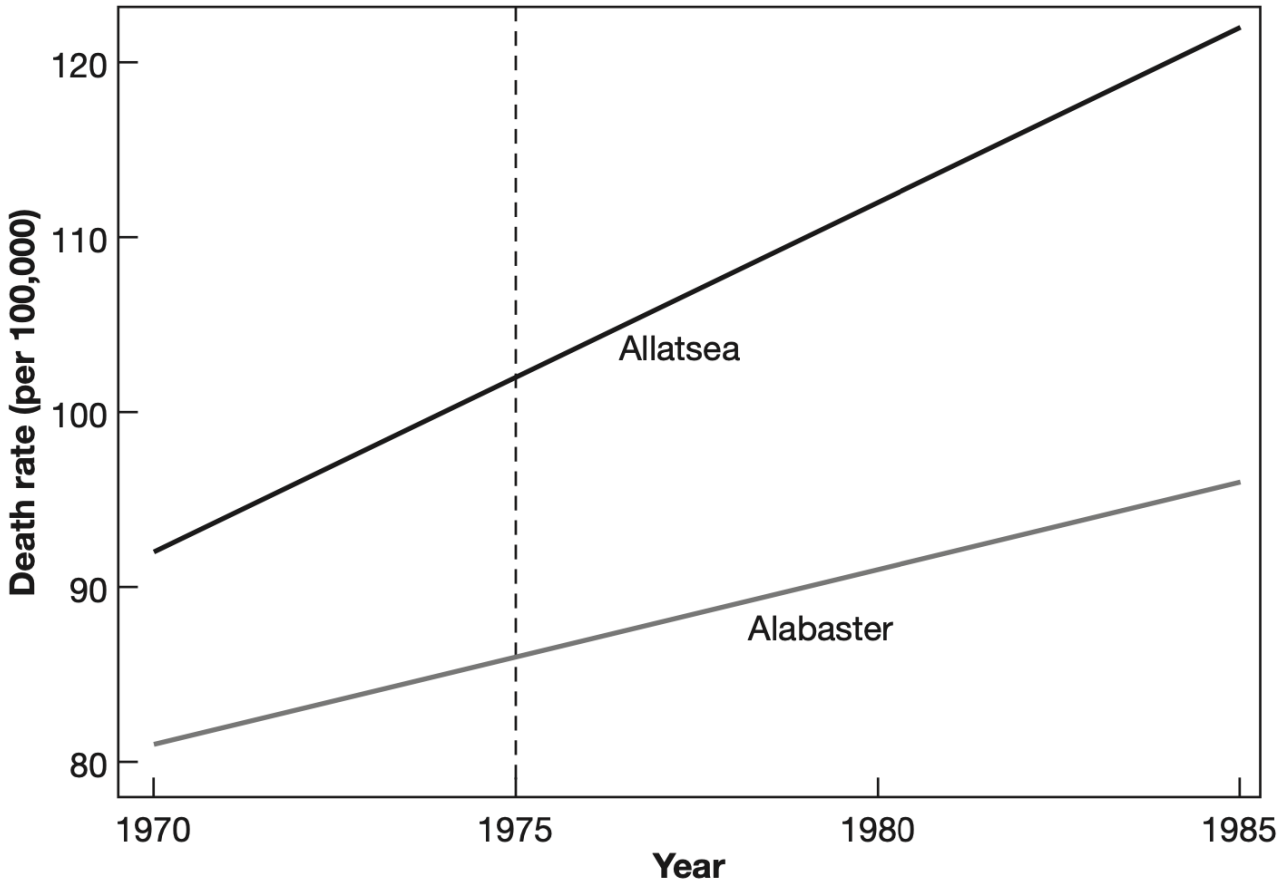
MLDA: Assumptions

FIGURE 5.4
An MLDA effect in states with parallel trends



MLDA: Assumptions

FIGURE 5.5
A spurious MLDA effect in states where trends are not parallel



MLDA: Assumptions

FIGURE 5.6
A real MLDA effect, visible even though trends are not parallel