Lecture 7B: Difference in Differences

Applied Micro-Econometrics,Fall 2020

Zhaopeng Qu

Nanjing University

12/3/2020

Section 1

Difference in Differences

Subsection 1

Introduction

Difference in Differences: Introduction

- DD(or DID) is a special case for "twoway fixed effects" under certain assumption, which is one of most popular research designs in applied microeconomics.
- It was introduced into economics via Orley Ashenfelter in the late 1970s and then popularized through his student David Card (with Alan Krueger) in the 1990s.

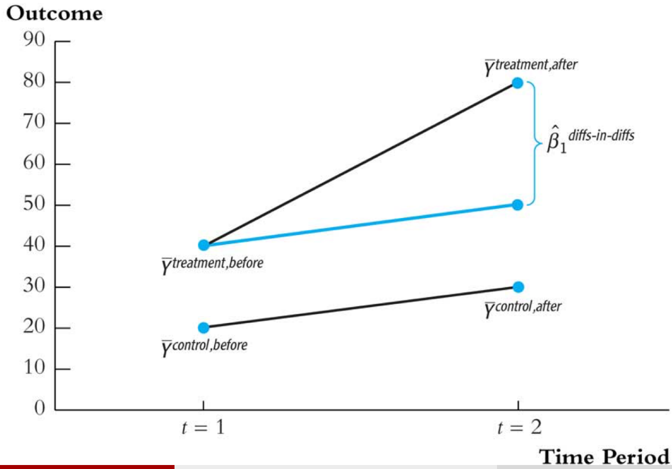
RCT and Difference in Differences

- A typical RCT design requires a causal studies to do as follow
  1. Randomly assignment of treatment to divide the population into a "treatment" group and a "control" group.
  2. Collecting the data at the time of post-treatment then comparing them.
- It works because *treatment* and *control* are randomized.
- What if we have the treatment group and the control group, but they are not fully randomized?
- If we have observations across two times at least with one before treatment and the other after treatment, then an easy way to make causal inference is **Difference in Differences(DID)** method.

DID estimator

- The DID estimator is

$$\hat{\beta}_{DID} = (\bar{Y}_{treat,post} - \bar{Y}_{treat,pre}) - (\bar{Y}_{control,post} - \bar{Y}_{control,pre})$$



Subsection 2

Card and Krueger(1994): Minimum Wage on Employment

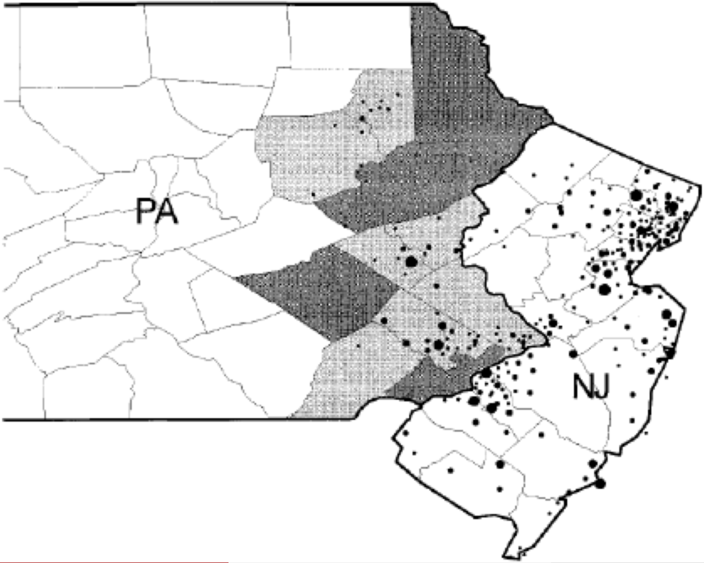
Introduction

- Theoretically,in competitive labor market, increasing binding minimum wage decreases employment.But what about the reality?
- Ideal experiment: randomly assign labor markets to a control group (minimum wage kept constant) and treatment group (minimum wage increased), compare outcomes.
- Policy changes affecting some areas and not others create natural experiments.
  - Unlike ideal experiment, control and treatment groups here are not randomly assigned.

Card and Krueger(1994): Backgroud

- Policy Change: in April 1992
  - Minimum wage in New Jersey from \$4.25 to \$5.05
  - Minimum wage in Pennsylvania constant at \$4.25
- Research Design:
  - Collecting the data on employment at 400 fast food restaurants in NJ(treatment group) in Feb.1992 (before treatment)and again November 1992(after treatment).
  - Also collecting the data from the same type of restaurants in eastern Pennsylvania(PA) as control group where the minimum wage stayed at \$4.25 throughout this period.

Card & Krueger(1994): Geographic Background



Card & Krueger(1994): Model Graph

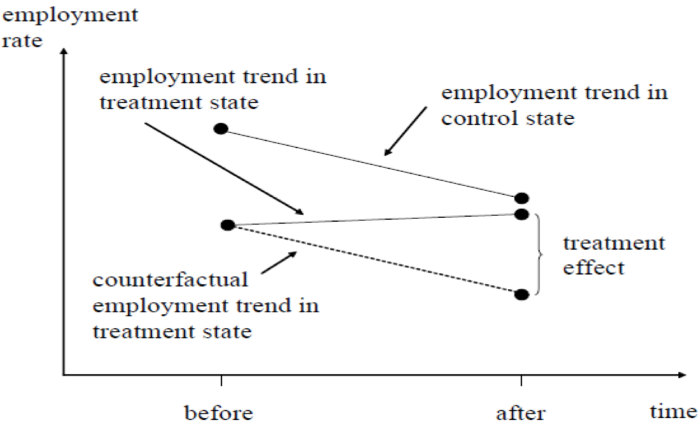


Figure 5.2.1: Causal effects in the differences-in-differences model

Card & Krueger(1994):Result

Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

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

Notes: Adapted from Card and Krueger (1994), Table 3. The

Regression DD - Card and Krueger

- DID model:

$$Y_{ts} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + u_{its}$$

  - NJ is a dummy equal to 1 if the observation is from NJ,
  - otherwise equal to 0(from Penny)
  - d is a dummy equal to 1 if the observation is from November (the post period),
  - otherwise equal to 0(Feb. the pre period)
- Which estimate coefficient does present DID estimator?

Regression DD - Card and Krueger

- A 2×2 matrix table

		treat or control	
		NJ=0(control)	NJ=1(treat)
pre or post	d=0(pre)	α	$\alpha+\gamma$
	d=1(post)	$\alpha+\lambda$	$\alpha+\gamma+\lambda+\delta$

- Then DID estimator

$$
\begin{aligned}
\hat{\beta}_{DID} &= (\bar{Y}_{treat,post} - \bar{Y}_{treat,pre}) - \\
&\quad (\bar{Y}_{control,post} - \bar{Y}_{control,pre}) \\
&= (NJ_{post} - NJ_{pre}) - (PA_{post} - PA_{pre}) \\
&= [(\alpha+\gamma+\lambda+\delta) - (\alpha+\gamma)] - [(\alpha+\lambda) - \alpha] \\
&= \delta
\end{aligned}
$$

Subsection 3

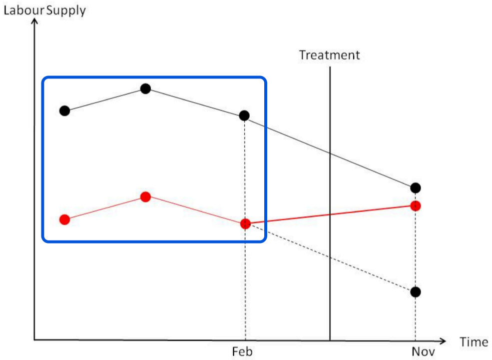
Key Assumption For DID

Paralled Trend

- A key identifying assumption for DID is: **Common trends** or **Parallel trends**
    - Treatment would be the same "trend" in both groups in the absence of treatment.
- This doesn't mean that they have to have the same mean of the outcome.
- There may be some unobservable factors affected on outcomes of both group. But as long as the effects have the same trends on both groups, then DID will eliminate the factors.
- It is difficult to verify because technically one of the parallel trends can be an unobserved counterfactual.

Assessing Graphically

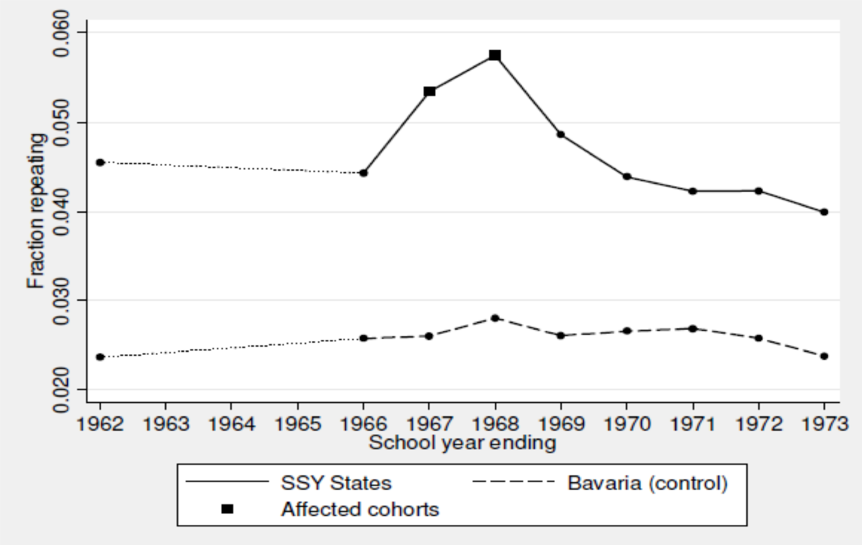
- **Common Trend**: It is difficult to verify but one often uses pre-treatment data to show that the trends are the same.
  - If you only have two-period data, you can do **nothing**.
  - If you luckly have multiple-period data, then you can show something graphically.



An Encouraging Example: Pischeke(2007)

- Topic: the length of school year on student performance
- Background:
  - Until the 1960s, children in all German states except Bavaria started school in the Spring. In 1966-1967 school year, the Spring moved to Fall.
  - It make two shorter school years for affected cohort, 24 weeks long instead of 37.
- Research Design:
  - Dependent Variable: Retreating rate
  - Independent Variable: spending time on school
  - Treatment group: Students in the German **States except Bavaria**.
  - Control group: Students in **Bavaria**.

An Encouraging Example: Pischeke(2007)



An Encouraging Example: Pischeke(2007)

- This graph provides strong visual evidence of treatment and control states with a common underlying trend.
- A treatment effect that induces a sharp but transitory deviation from this trend.
- It seems to be clear that a short school years have increased repetition rates for affected cohorts.

Section 2

Extensions of DID

Subsection 1

Extensions in Multiple Periods and Groups

A Simple DID Regression

- The simple DID regression

$$Y_{ist} = \alpha + \beta(Treat \times Post)_{st} + \gamma Treat_s + \delta Post_t + u_{ist}$$

  - $Treat_s$ is a dummy variable indicate whether or not is **treated**.
  - $Post_t$ is a dummy variable indicate whether or not is **post-treatment** period.
  - γ captures the outcome gap between treatment and control group that *are constant over time*.
  - δ captures the outcome gap across post and pre period that *are common to both two groups*.
  - β is the coefficient of interest which is the *difference-in-differences* estimator
- Note: *Outcomes are often measured at the individual level i,while treatment takes place at the group level s*.

A Simple DID Regression with Covariates

- Add more covariates as **control variables** which may reduce the residual variance (lead to smaller standard errors)

  $$Y_{ist} = \alpha + \beta(Treat \times Post)_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

  - X_{ist} is a vector of control variables. Γ is the corresponding estimate coefficient vector.

- X_{ist} can include individual level characteristics and time-varying measured at the group level.
- Those time-invariant Xs may not helpful because they are part of fixed effect which will be differential.
- Time-varying Xs may be problematic if they are the outcomes of the treatment which are **bad controls**.
- So *Pre-treatment covariates* which could include Xs on both group and individual level are more favorable.

DID for different treatment intensity

- Study treatments with different treatment intensity. (e.g., varying increases in the minimum wage for different states)

A Simple DID Regression with More Periods

- We can slightly change the notations and generalize it into

$$Y_{ist} = \alpha + \beta D_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

  - Where D_{st} means $(Treat \times Post)_{st}$

- Using Fixed Effect Models further to transform into

$$Y_{ist} = \beta D_{st} + \alpha_s + \delta_t + \Gamma X'_{ist} + u_{ist}$$

  - α_s is a set of groups fixed effects, which captures $Treat_s$.
  - δ_t is a set of time fixed effects, which captures $Post_t$.

- Note:
  - Samples enter the treatment and control groups **at the same time**.
  - The frame work can also apply to **Repeated(Pooled) Cross-Section Data**.

Subsection 2

Loose or Test Common Trend Assumption

Add group-speicific time trends

- This setting can eliminate the effect of group-specific time trend in outcome on our DID estimates

$$Y_{ist} = \beta D_{st} + \alpha_s + \delta_t + \tau_{st} + \Gamma X'_{ist} + u_{ist}$$

- τ_{st} is group-specific dummies multiplying the time trend variable t, which can be quadratic to capture some nonlinear trend.
- The **group specific time trend** in outcome means that treatment and control groups can follow different trends.
- It make DID estimate more robust and convincing when the pretreatment data establish a clear trend that can be extrapolated into the posttreatment period.

Add group-speicific time trends

- Besley and Burgess (2004),"Can Labor Regulation Hinder Economic Performance? Evidence from India", *The Quarterly Journal of Economics*.
  - Topic: labor regulation on businesses in Indian states
  - Method: Difference-in-Differences
  - Data: States in India
  - Dependent Variable: log manufacturing output per capita on states levels
  - Independent Variable: Labor regulation(lagged) coded $1 = pro - worker$;$0 = neutral$;$-1 = pro - employer$ and then accumulated over the period to generate the labor regulation measure.(Convincing?)

TABLE 5.2.3
Estimated effects of labor regulation on the performance of firms in Indian states

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	−.186	−.185	−.104	.0002
	(.064)	(.051)	(.039)	(.020)
Log development expenditure per capita		.240	.184	.241
		(.128)	(.119)	(.106)
Log installed electricity capacity per capita		.089	.082	.023
		(.061)	(.054)	(.033)
Log state population		.720	0.310	−1.419
		(.96)	(1.192)	(2.326)
Congress majority			−.0009	.020
			(.01)	(.010)
Hard left majority			−.050	−.007
			(.017)	(.009)
Janata majority			.008	−.020
			(.026)	(.033)
Regional majority			.006	.026
			(.009)	(.023)
State-specific trends	No	No	No	Yes
Adjusted R^2	.93	.93	.94	.95

Within control group – DDD(Triple D)

- More convincing analysis sometime comes from higher-order contrasts: **DDD** or **Triple D** design.
  - Build the third dimension of contrast to eliminate the potential bias.
- e.g: Minimum Wage
  - Treatment group: Low-wage-workers in NJ.
  - Control group 1: High-wage-workers in NJ.
  - Assumption 1: the low wage group would have the same trends as high wage group if there were not the new law.
  - Control group 2: Low-wage workers in PA.
  - Assumption 2: the low wage group in NJ would have the same trends as those in PA if there were not the new law.
- It can loose the simple *common trend* assumption in simple DID.

Within control group – DDD(Triple D)

- Jonathan Gruber (1994), "The Incidence of Mandated Maternity Benefits", *American Economic Review*
  - Topic: how the *mandated maternity* benefits affects female's wage and employment.
  - Several state government passed the law that mandated childbirth be covered comprehensively in health insurance plans.
  - Dependent Variable: log hourly wage
  - Independent Variable: mandated maternity benefits law
- Econometric Method: **Triple D**
  1. DID estimates for treatment group (women of childbearing age) in treatment state v.s. control state before and after law change.
  2. DID estimates for control group (women not in childbearing age) in treatment state v.s. control state before and after law change.
  3. DDD DDD estimate of the effect of mandated maternity benefits on wage is $(1)-(2)$

Within control group – DDD(Triple D)

- DDD in Regression

$$Y_{isct} = \beta D_{sct} + \alpha_s + \gamma_c + \delta_t + \lambda_{1st} + \lambda_{2sc} + \lambda_{3ct} + \Gamma X'_{icst} + u_{isct}$$

- α_s:a set of dummies indicating whether or not treatment state
- δ_t: a set of dummies indicating whether or not law change
- γ_c: a set of dummies indicating whether or not women of childbearing age

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

The Event Study Design: Including Leads and Lags

- If you have a multiple years panel data, then including leads into the DD model is an easy way to analyze pre-treatment trends.
- Lags can be also included to analyze whether the treatment effect changes over time after assignment.
- The estimated regression would be

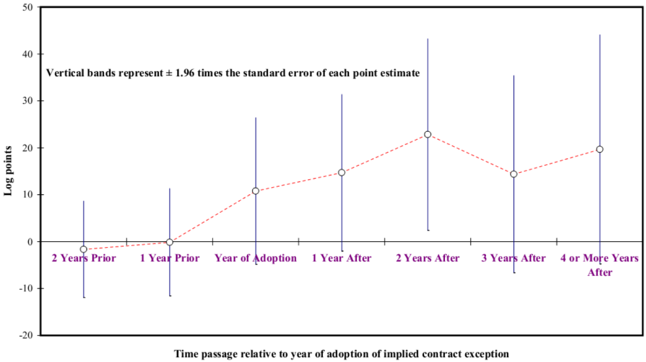
$$Y_{its} = \alpha_s + \delta_t + \sum_{\tau=-q}^{-1} \theta_\tau D_{st} + \sum_{\tau=0}^{p} \delta_\tau D_{st} + X_{ist} + u_{its}$$

- Treatment occurs in year 0
- Includes q leads or anticipatory effects
- Includes p leads or post treatment effects

Study including leads and lags – Autor (2003)

- Autor (2003) includes both leads and lags in a DD model analyzing the effect of increased employment protection on the firm's use of temporary help workers.
- In the US employers can usually hire and fire workers at will.
- U.S labor law allows 'employment at will' but in some state courts have allowed a number of exceptions to the doctrine, leading to lawsuits for 'unjust dismissal'.
- The employment of temporary workers in a state to dummy variables indicating state court rulings that allow exceptions to the employment-at-will doctrine.
- The standard thing to do is normalize the adoption year to 0
- Autor(2003) then analyzes the effect of these exemptions on the use of temporary help workers.

Study including leads and lags – Autor (2003)



- The leads are very close to 0: Common trends assumption may hold.
- The lags show that the effect increases during the first years of the treatment and then remains relatively constant.

Subsection 3

Other Issues

Standard errors in DD strategies

- Many paper using DD strategies use data from many years: not just 1 pre and 1 post period.
- The variables of interest in many of these setups only vary at a group level (say a state level) and outcome variables are often serially correlated
- In the Card and Krueger study, it is very likely that employment in each state is not only correlated within the state but also serially correlated.
- As Bertrand, Duflo and Mullainathan (2004) point out, conventional standard errors often severely *understate* the standard deviation of the estimators – standard errors are biased downward.

Standard errors in Practice

- Simple solution:
  - Clustering standard errors at the group level,but the number of groups does matter.
  - It may also cluster at both the group level and time level.
- Other solutions: Bootstrapping

Other Threats to validity

- Non-parallel trends
- Other simultaneous shock
- Functional form dependence
- Multiple treatment times

Non-parallel trends

- Often policymakers will select the treatment and controls based on pre-existing differences in outcomes – practically guaranteeing the parallel trends assumption will be violated.
- "Ashenfelter dip"
  - Participants in job trainings program often experience a "dip" in earnings just prior to entering the program.
  - Since wages have a natural tendency to mean reversion,comparing wages of participants and non-participants using DD leads to an upward biased estimate of the program effect.

DD with multiple treatment times

- What happens if we have treated units who get treated at different times?
- The simple DID model

$$Y_{ist} = \alpha + \beta D_{st} + \gamma Treat_s + \delta Post_t + \Gamma X'_{ist} + u_{ist}$$

- But now DT_{it} can turn from 0 to 1 at different times for different units.
- **Caution**: this specification gets you a weighted average of several comparisons.This may not be exactly what you want!

Function Forms

- So far our specifications of DID regression equation is linear, but what if it is wrong?
- Several nonparametric or semi-parametric methods can be used
  - Matching DID: Propensity Score Matching and Kernel Density Matching DID
  - Semiparametric DID

Checks for DD Design

- Very common for readers and others to request a variety of "robustness checks" from a DID design.
- Think of these as along the same lines as the leads and lags
    - Falsification test using data for prior periods
    - Falsification test using data for alternative control group(kind of triple DDD)
    - Falsification test using alternative "placebo" outcome that should not be affected by the treatment

Wrap up

- Difference-in-differences is a powerful horse in our toolbox to make causal inference.
- The key assumption is common trend which is not easy to testify using data.
- Noting that using the right way to inference the standard error.