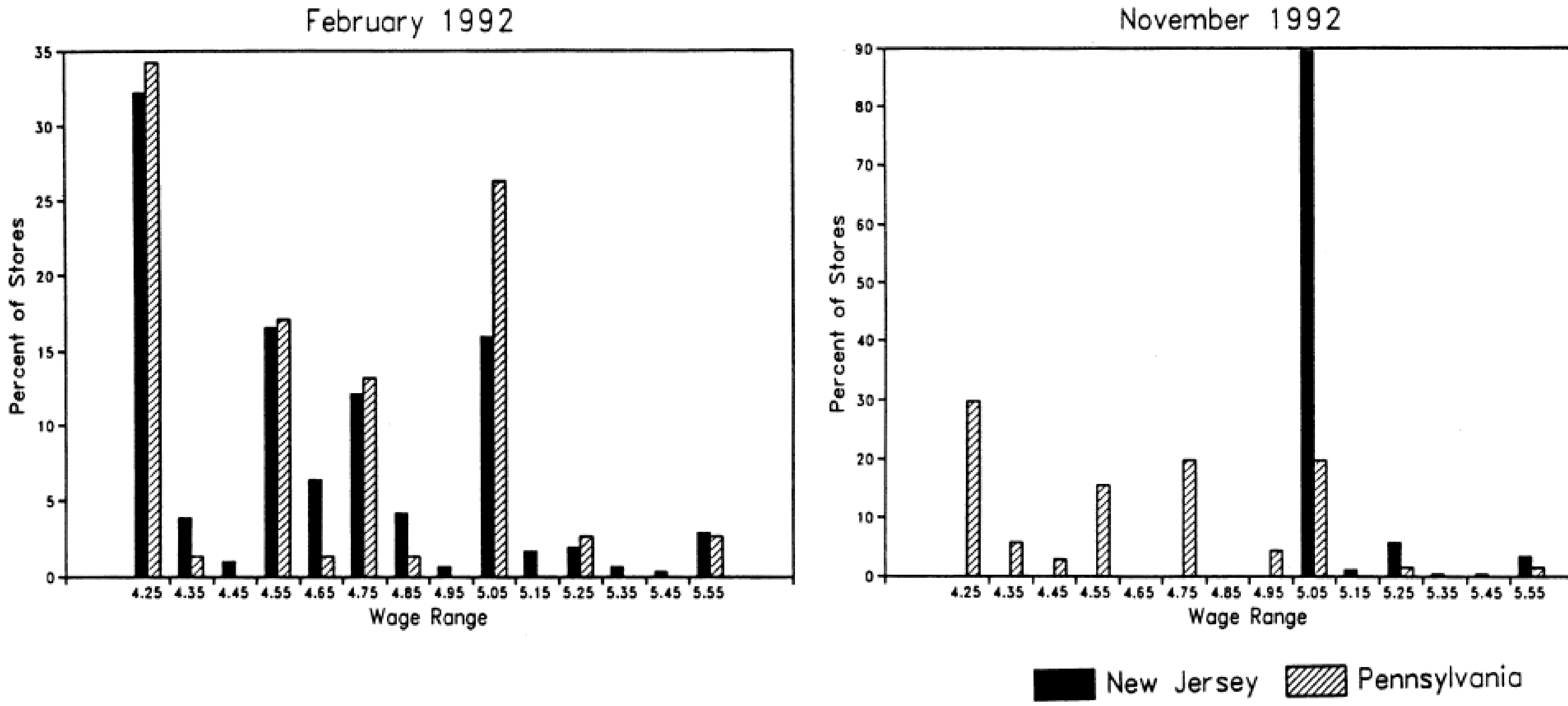
Causal inference

Part II: Difference In Difference and Instrumental Variables

Difference in difference

Card & Krueger (1995,AER)

- Rise in minimum wage from 4,2$ to 5,05$ in April 1992 in the State of New Jersey.
- Research question: impact on unskilled labour demand?
- Rise decided in 1990, but economic recession in 1992 led to an unsuccessful attempt to abort the measure.
  - => it makes sense to think that the shock was exogenous (unanticipated).
- Compare employment before and after the measure.
- Compare employment trend in New Jersey and Pennsylvania


February 1992
Percent of Stores
35
30
25
20
15
10
5
0
4.25
4.35
4.45
4.55
4.65
4.75
4.85
4.95
5.05
5.15
5.25
5.35
5.45
5.55
Wage Range
November 1992
90
80
70
60
50
40
30
20
10
0
New Jersey
Pennsylvania

Results

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)

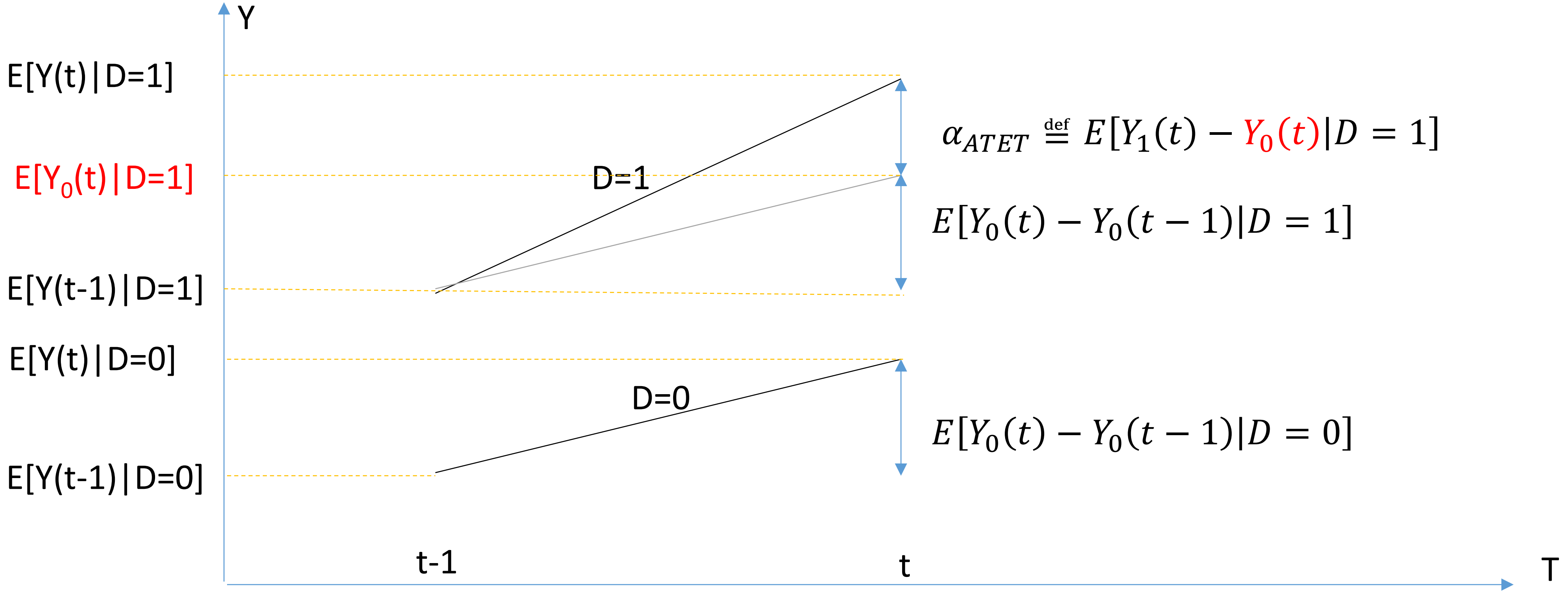
Selection on Unobservables

- Maybe potential outcomes (employment with and without minimum wage increase) are affected by unobserved characteristics (such as skills, labour market structure, business cycle).
- Therefore, use an identification strategy based on unobserved characteristics.

Notation

- Two groups:
  - D=1 Treated units
  - D=0 Control units
- Two periods:
  - t-1 Pre-treatment period
  - t Post-treatment period
- Potential outcome $Y_d(t)$
  - $Y_{1i}(t)$ outcome unit i attains in period t when treated between t and t-1
  - $Y_{0i}(t)$ outcome unit i attains when control between t and t-1

Parallel trend assumption



Assumption: $E[Y_0(t) - Y_0(t-1)|D = 1] = E[Y_0(t) - Y_0(t-1)|D = 0]$

Treatment only affects period t => $E[Y_0(t-1)|D = 1] = E[Y(t-1)|D = 1]$

$$\Rightarrow \alpha_{ATET} \stackrel{\text{def}}{=} E[Y_1(t) - Y_0(t)|D = 1] = E[Y_1(t)|D = 1] - E[Y_0(t)|D = 1]$$
$$= \{E[Y(t)|D = 1] - E[Y(t)|D = 0]\} - \{E[Y(t-1)|D = 1] - E[Y(t-1)|D = 0]\}$$

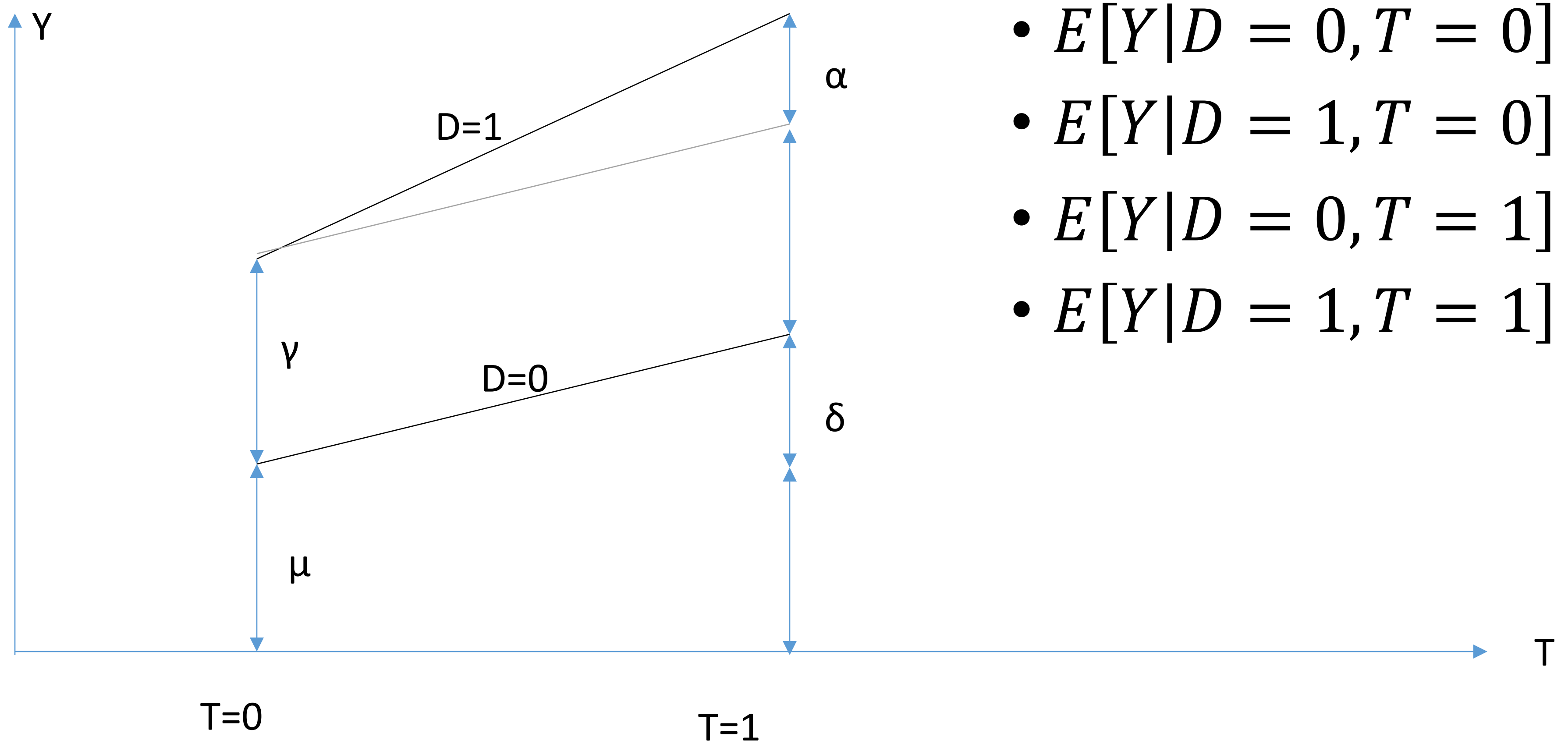
Parallel trend assumption

DID estimator

- $\hat{\alpha}_{ATET}$
$$= \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(t) - \frac{1}{N_0}\sum_{D_i=0} Y_i(t)\right\} - \left\{\frac{1}{N_1}\sum_{D_i=1} Y_i(t-1) - \frac{1}{N_0}\sum_{D_i=0} Y_i(t-1)\right\}$$
$$= \frac{1}{N_1}\sum_{D_i=1}[Y_i(t) - Y_i(t-1)] - \frac{1}{N_0}\sum_{D_i=0}[Y_i(t) - Y_i(t-1)]$$
- The same result is obtained using OLS with dummy T=0 at t-1 and T=1 at t:

$$Y = \mu + \gamma D + \delta T + \alpha_{ATET}(DT) + \epsilon$$

Graphic representation of OLS with dummies



- $Y = \mu + \gamma D + \delta T + \alpha(DT) + \epsilon$
- $E[Y|D = 0, T = 0] = \mu$
- $E[Y|D = 1, T = 0] = \mu + \gamma$
- $E[Y|D = 0, T = 1] = \mu + \delta$
- $E[Y|D = 1, T = 1] = \mu + \gamma + \delta + \alpha$

Add explainatory variables

- $Y = \mu + \gamma D + \delta T + \alpha(DT) + X\beta + \epsilon$
  - If many confounders, X is a matrix with k columns and beta a vector with k rows
- Problem: time-invariant X are impossible (its effect is captured by gamma)
- However, if X is time-variant, X may be affected by treatment =>causal relationship between explainatory variables
- Solution if many periods: work with first difference

Multiple groups and time periods

- Imagine that you have panel data for 5 years and 6 states and a comparable minimum wage increase was introduced at different times in different states. Panel with 3 dimensions: treatment, country and time. Regress:

$$Y = \mu + \sum_{states} \gamma_i D_i + \sum_{periods} \delta_t D_{time} + \alpha D_{treated} + X\beta + \epsilon$$

- The i-th state at the t-th time writes:

$$Y_{it} = \mu + \gamma_i + \delta_t + \alpha D_{treated_{it}} + X_{it}\beta + \epsilon_{it}$$

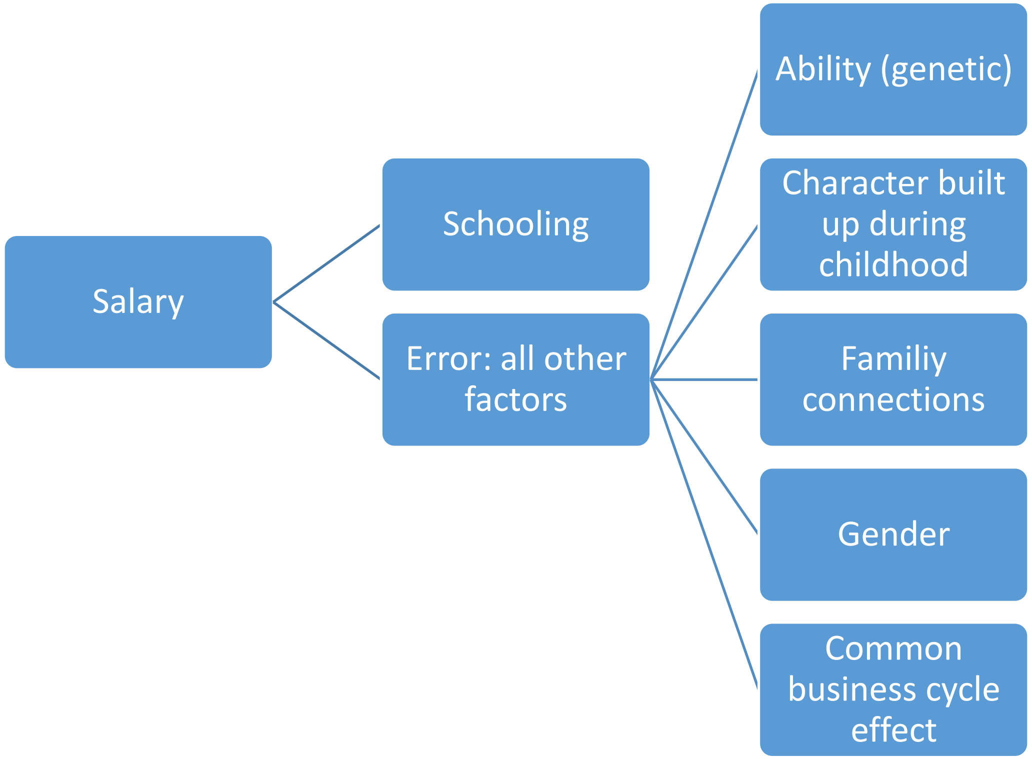
- One parameter for each time period and state
- Adjust standard errors for temporal dependence
- Assumes the same effect in every state $\alpha_i = \alpha$

Regression with fixed time and individual effects

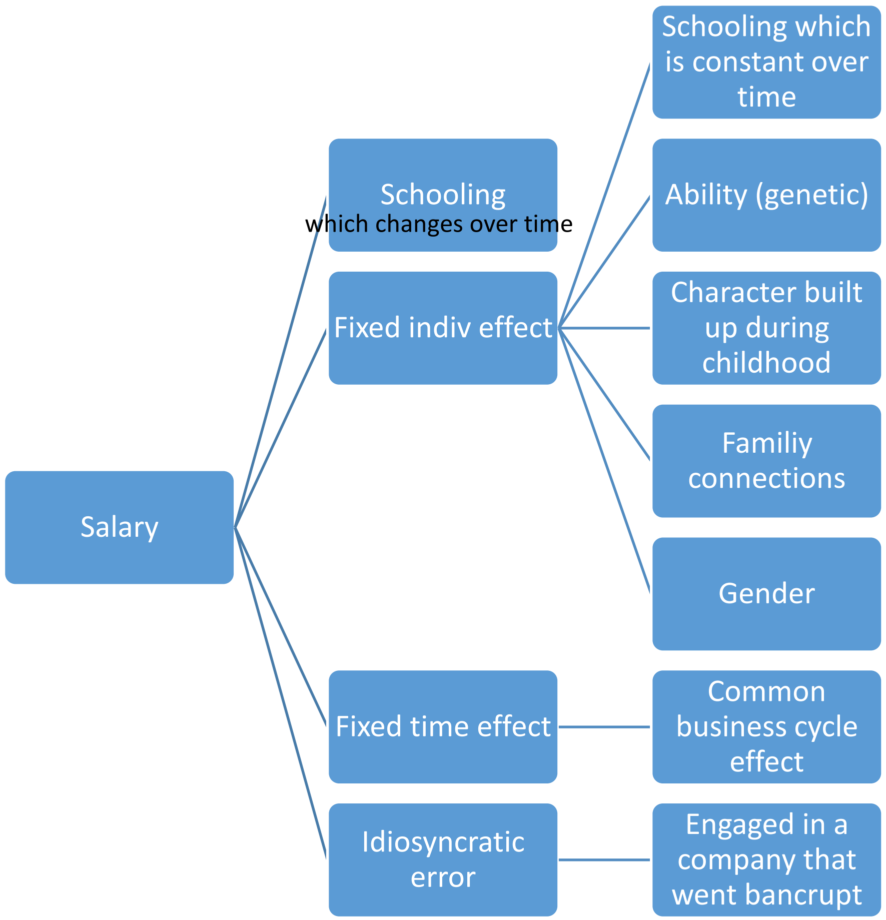
- Until now we had a panel with 3 dimensions, now we look at only 2 dimensions. Ex 10 companies followed over 5 years.
- Recall: regression with fixed individual effects:
  - $Y_{it} = \mu + \gamma_i + X_{it}\beta + \epsilon_{it}$
  - Avoids omitted variable bias from any time-invariant company characteristics (country effect under unchanged policy, sector effects that do not interact with X's…)
- Regression with fixed individual and time effect (=2 way error component model).
  - $Y_{it} = \mu + \gamma_i + \delta_t + X_{it}\beta + \epsilon_{it}$
  - Avoids omitted variable bias from any time invariant characteristics (ex. country) and any time effects (ex business cycle) that are common to all companies
  - The fixed effects subtract parallel time trends like in DID=> $X_{it}\beta$ only driven by differences between companies that change over time after common trends are subtracted
  - X must be company specific and change over time (to avoid perfect colinearity)
  - N-1 + T-1 degrees of freedom lost =>efficiency loss (higher standard errors)

What factors could cause endogeneity?

- Regress $Salary = \alpha + \beta Schooling + \epsilon$



2 way fixed effects



- You only measure for those people who take schooling while they are working
- => Fixed effect leaves out potential bias but also a lot of interesting information, certainly when information with strong auto-correlation.

The effect of trade union membership on wage (Freeman 1984)

Table 5.1.1: Estimated effects of union status on log wages

Survey	Cross section estimate	Fixed effects estimate
May CPS, 1974-75	0.19	0.09
National Longitudinal Survey of Young Men, 1970-78	0.28	0.19
Michigan PSID, 1970-79	0.23	0.14
QES, 1973-77	0.14	0.16

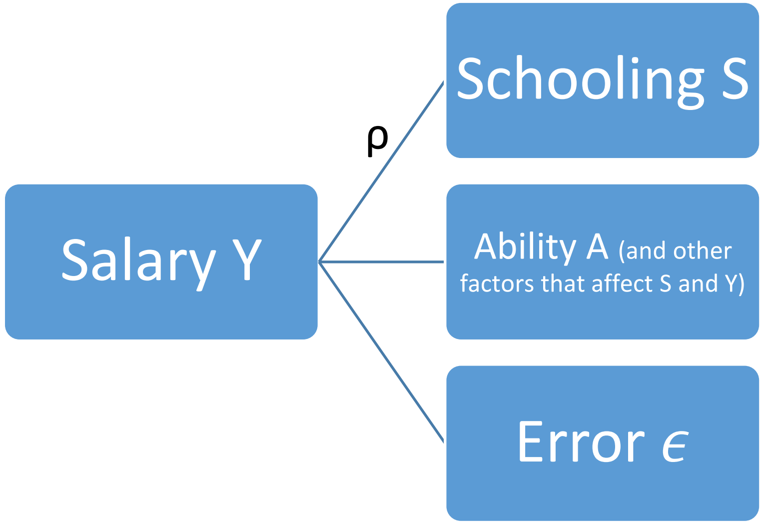
Notes: Adapted from Freeman (1984). The table reports cross-section and panel estimates of the union relative wage effect. The estimates were calculated using the surveys listed at left. The cross-section estimates include controls for demographic and human capital variables.

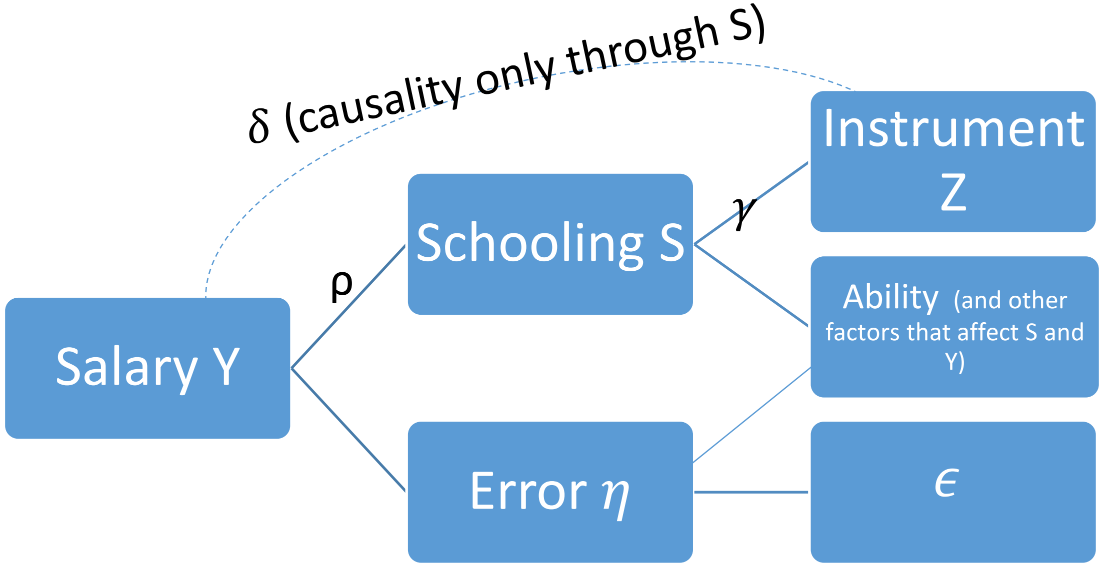
Disadvantages of 2 way fixed effects

- Trade union membership data are highly persistent (a worker who is a trade union member this year is likely to be member next year) => big attenuation bias from measurement errors
- Fixed effect ‘erases’ out a lot if interesting information: only the effect for workers that become member or disaffiliate is measured. Difference between members that are allways affiliated (the most combattive ones?) and members that are never affiliated (the closest to the management?) have no effect on the estimate.
- Fixed effect assumes that effects are fixed: no interaction (ex. effect of downturn is the same for everybody, high-skilled and low-skilled alike)

Instrumental variables

Wald estimator



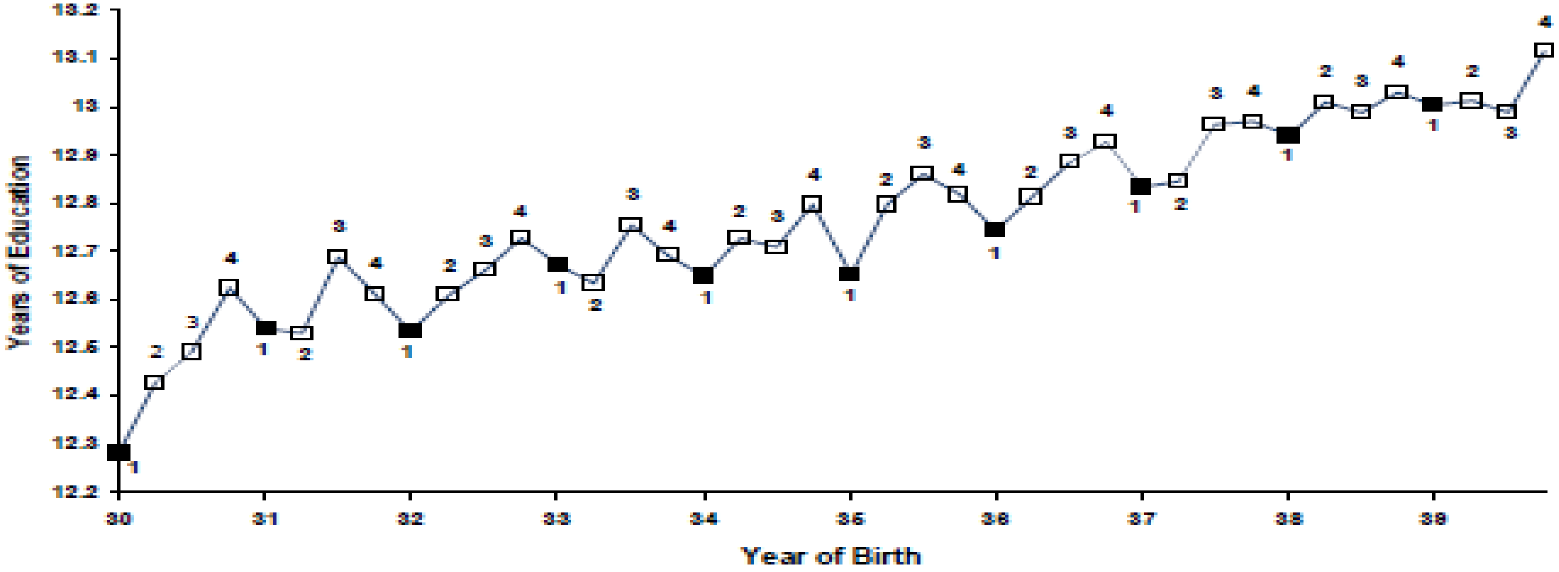


- Want to estimate $Y = \alpha + \rho S + \beta A + \epsilon$ but ability is unobservable
- Imagine a binary instrument correlated with schooling but independent from Ability (and any other factors that affect S and Y)
  - $cov\,(Z, \eta) = 0$
  - $\Leftrightarrow$ Z mimicks a random assignement i.e. potential outcome $Y_{0i}, Y_{1i} \perp Z$
- No covariates: the only effect of the instrument is through the causal variable of interest (will be relaxed)
- Instrument has no direct effect on salary, instrument affects salary only via one causal path, which goes over schooling $\delta = \gamma \times \rho$
- Wald estimator $\rho = \frac{\delta}{\gamma} = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(S|Z=1) - E(S|Z=0)}$

Angrist and Kreuger (1991)

- Use date of birth as an instrument for schooling.
- Most states require children to enter school in the calendar year in which they turn 6.
  - Children born in Oct, Nov, Dec enter school shortly before 6,
  - whereas children born in Jan, Febr, March enter school around 6,5.
- By contrast, legal age of school dropout is 16.

A. Average Education by Quarter of Birth (first stage)



B. Average Weekly Wage by Quarter of Birth (reduced form)

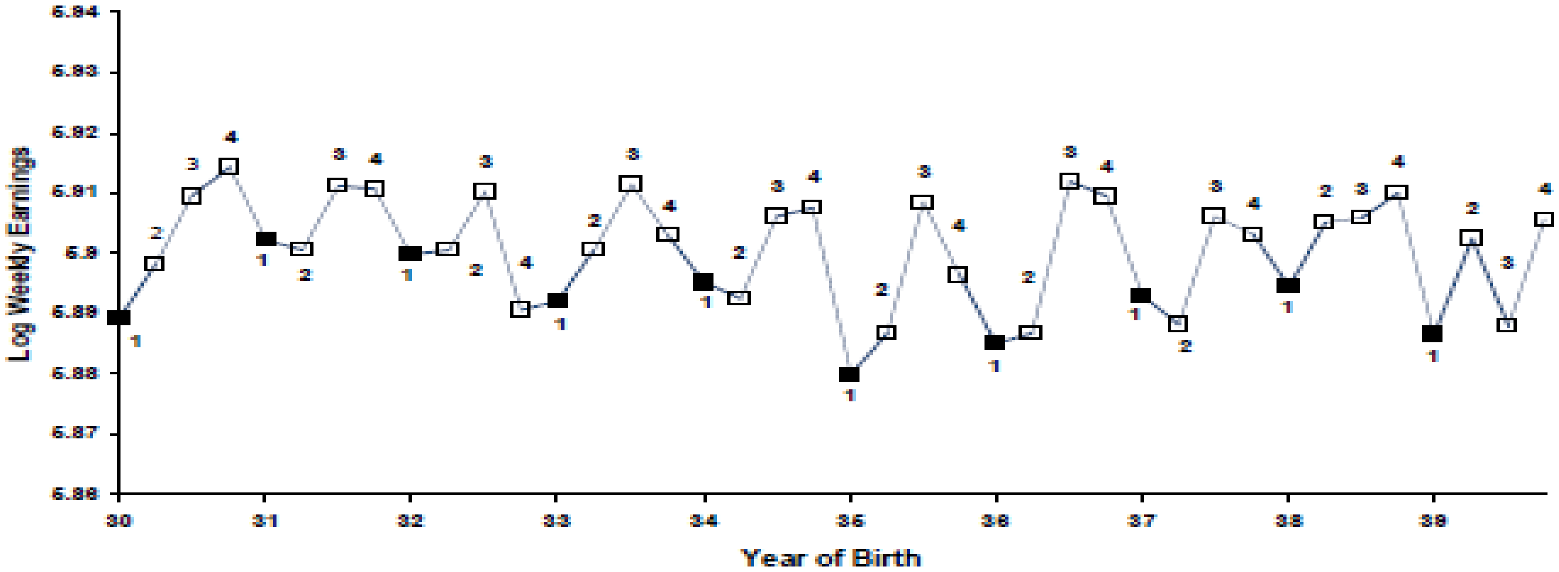


Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

Notes: Adapted from a re-analysis of Angrist and Krueger (1991) by Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930-39 birth cohorts in the 1980 Census 5 percent file. The sample size is 329,509.

Effect of Vietnam service on earnings (Angrist 1990)

- What are possible confounders affecting both the propobability of going to vietnam and earnings?
  - Social status
  - Race …
- In every cohort of 19 years old, each birthday was assigned a random sequence number. Birthdays with a number below a treshold were draft-eligible, above a treshold were non draft-eligible.
- Non-eligible persons could go to Vietnam and many eligible persons did not go to Vietnam. But eligibility is correlated with Vietnam service.

Rich draft-eligible men may have a lower probability to serve than poor draft-eligible men…

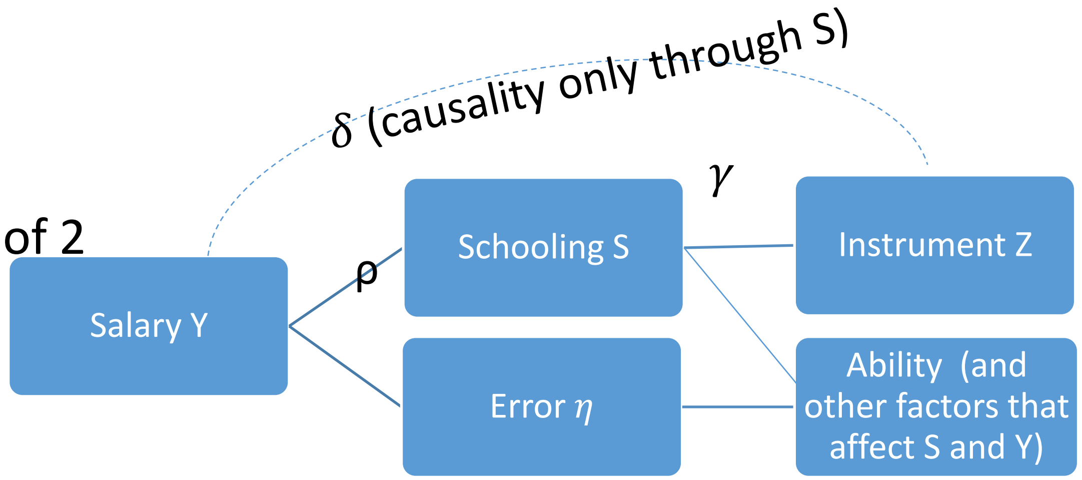
- Does this make the instrument invalid?
- Social status affects both salary and the probability to go to Vietnam, that’s why we need an instrument in the first place.
- But status does not affect the probability to be draft-eligible(both rich and poor have the same probability to be draft-eligible).That’s why the instrument is valid.
  - i.e. there are no common drivers that affect both draft-eligibility and salary.
- Even stronger, the instrument is randomized,
  - i.e. the potential outcome (salary if one would/wouldn’t have been eligible) is independent of treatement (eligibility).
- There may be heterogenous effects. Ex effect of going to Vietnam on salary may be lower (or higher) for poor people.
  - Therefore, the Treatement Effect of the Treated (ATET) will be lower (higher) than the Average Treatement Effect (ATE).
  - The instrument estimates the ATET.
  - The estimated (and real) effect will therefore depend on discrimation mechanisms affecting the relationship between draft-eligibility and going to Vietnam.
- Remark: If draft-eligible young people study longer to avoid going to Vietnam, the instrument is biased (earnings modifying draft avoidance)
  - There is a second causal way (through study) in which instrument affects salary

Table 4.1.3: Wald estimates of the effects of military service on the earnings of white men born in 1950

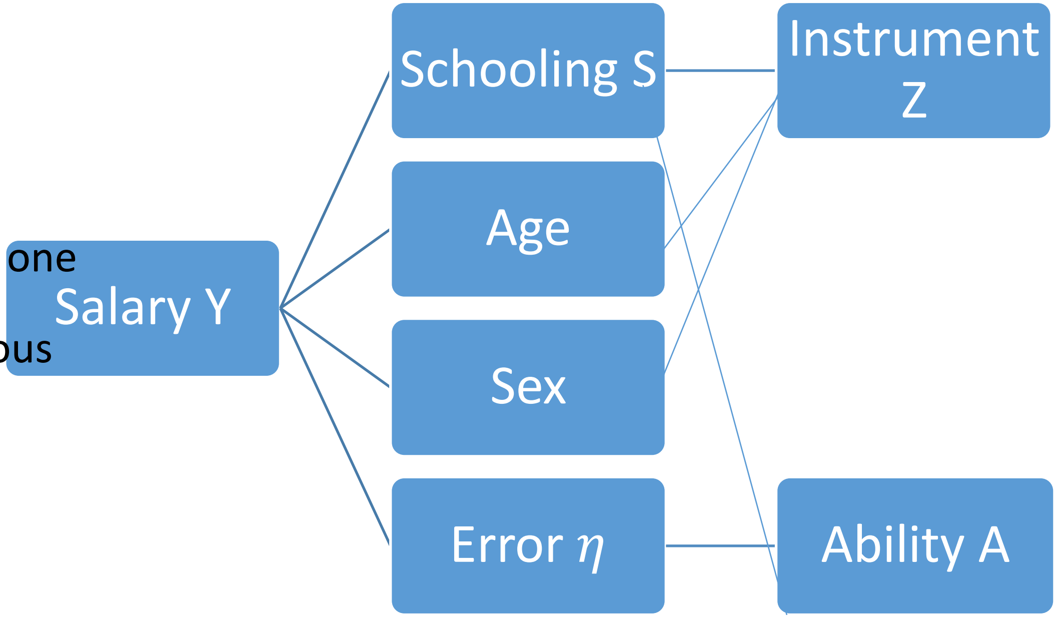
Earnings year	Earnings		Veteran Status		Wald Estimate of Veteran Effect
	Mean	Eligibility Effect	Mean	Eligibility Effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8 (210.5)	0.267	0.159 (0.040)	-2,741 (1,324)
1971	3,338	-325.9 (46.6)			-2050 (293)
1969	2,299	-2.0 (34.5)			

Notes: Adapted from Angrist (1990), Tables 2 and 3. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.

2 Stages Least Squares



- If the instrument is a continuous variable, use a system of 2 equations
  - 'first stage' equation $S = \alpha + Z\gamma + \upsilon$
  - 'second stage' equation $Y = \alpha^* + Z\delta + \epsilon$
  - $\rho = \frac{\delta}{\gamma} = \frac{cov(Y,Z)/varZ}{cov(S,Z)/varZ} = \frac{cov(Y,Z)}{cov(S,Z)}$
- Independence of unobserved confounders (ability) conditional on covariates is sufficient => add observed confounders X
  - 'first stage' equation $S = \alpha + Z\gamma + X\beta + \upsilon$
  - 'second stage' equation $Y = \alpha^* + Z\delta + X\beta^* + \epsilon$
  - $\rho = \frac{\delta}{\gamma} = \frac{cov(Y,\tilde{Z})}{cov(S,\tilde{Z})}$ $(with\ Z = X\beta^{**} + \tilde{Z})$
  - Possibility to use several endogenous variables and more than one instrument per endogenous var
  - Matrix formula $\hat{\beta} = (Z'X)^{-1}Z'Y$ (X= endogenous and exogenous variables, Z= instruments and exog var)



- Command stata: ivregress y x1 x2 (X1=Z)
- A good instrument:
  - Has a clear effect on the X it is instrumenting (avoid 'weak instruments')
  - Instrument must be as good as randomly assigned $E[Z\eta|X] = 0$
  - The only causal link runs over schooling: no direct effect on Y

Table 4.1.1: 2SLS estimates of the economic returns to schooling

	OLS		2SLS					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Years of education	0.075 (0.0004)	0.072 (0.0004)	0.103 (0.024)	0.112 (0.021)	0.106 (0.026)	0.108 (0.019)	0.089 (0.016)	0.061 (0.031)
Covariates:								
Age (in quarters)								✓
Age (in quarters) squared								✓
9 year of birth dummies		✓			✓	✓	✓	✓
50 state of birth dummies		✓			✓	✓	✓	✓
Instruments:			dummy for QOB=1	dummy for QOB=1 or QOB=2	dummy for QOB=1	full set of QOB dummies	full set of QOB dummies int. with year of birth dummies	full set of QOB dummies int. with year of birth dummies

Notes: The table reports OLS and 2SLS estimates of the returns to schooling using the the Angrist and Krueger (1991) 1980 Census sample. This sample includes native-born men, born 1930-1939, with positive earnings and non-allocated values for key variables. The sample size is 329,509. Robust standard errors are reported in parentheses.

Weak instruments

- 2SLS is based on asymptotic theory
- 2SLS estimator is consistent but biased => you need a big sample.
- If there is only one instrument, the median of the estimator is unbiased. The more (overidentifying) instruments, the greater the bias.
- If the instrument is weak (correlation between Z and X is very low) the bias is much more important.
  - => useful to report first stage regression
- Study of Angrist and Kreuger has been critisized for being a weak instument despite a sample size of 329 000 (Imbens, Rosenbaum 2005).
- Also, the measured effect of schooling is not the effect for the whole population, only for the low skilled. (For high skilled people, period of birth and schooling are uncorrelated).