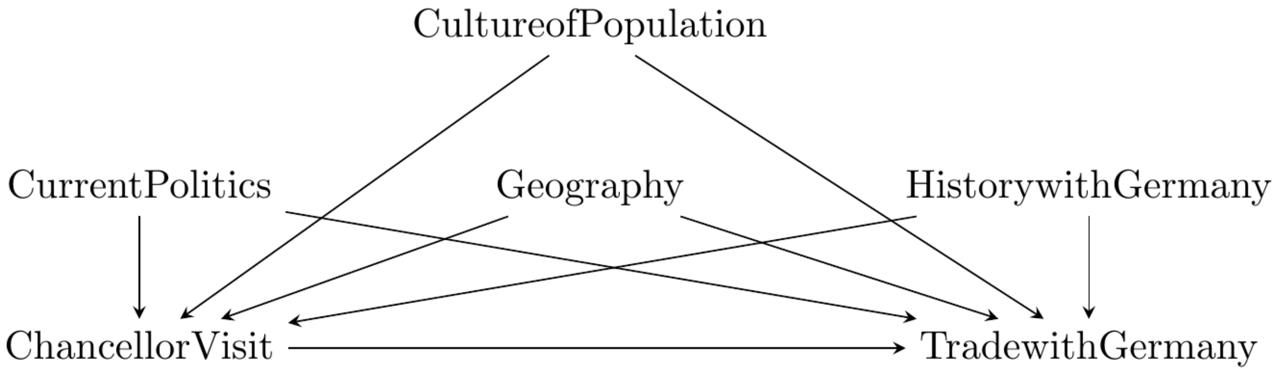
Motivation

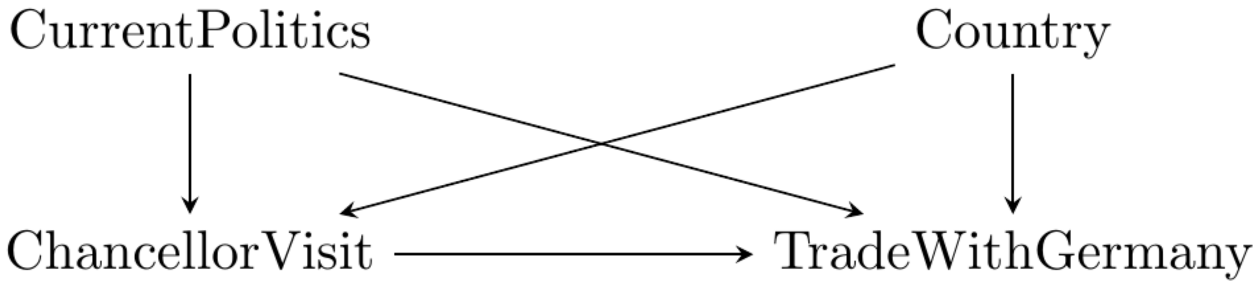
Panel data often looks like repeated observations of individuals (firms, countries) over time. More generally, panel data has a certain hierarchy (or levels) of observations. For example, if you are observing different individuals over a period of time, your first level of hierarchy is individual (i) and within that level you have further variation by years (t). These levels do not have to be individuals and time. It could be, for example, country by state, or state by city. A panel can also have more than two levels of hierarchy, e.g., country by state by city.

Panel data (or fixed effects) design allows one to control for observed and unobserved confounders "wholesale." It does so by controlling for the characteristic within which all other variables stay constant. In other words, instead of controlling for each confounder individually, which might not be feasible due to unobservables, we simply control for a single characteristic. All of the confounders, however, must remain constant within this characteristic.

Consider the following example (The Effect, chapter 16). Suppose we are interested in the causal effect of a visit by a German chancellor on a country's trade with Germany. We observe different countries over multiple years. One potential causal diagram might look like this.



The trade with Germany is affected by many things (culture, history of relations, geography, current politics), which also affect whether the German chancellor visits a country or not. Some of those confounders might be hard to measure or observe. However, many of them are fixed for a given country. We can therefore collapse all of those factors into a single variable: country. We would call that procedure adding *country fixed effects*. This would make our diagram look like this.



Geography, culture, and history got all absorbed by those *country* fixed effects. Current politics are clearly not fixed for a given country, so it remains a confounder we would need to control for to isolate the causal effect of a chancellor's visit. Remember that for this strategy to work we would need to have multiple observations over years for each country that provide some variation in the chancellor's visits. If the chancellor always visits some countries and never visits the rest, then controlling for country subsumes the treatment variable. We cannot identify the effect in this case. The treatment variable must vary within a country, i.e., the chancellor visits in one year but does not visit in another, for at least some countries.

Including fixed effects for a given characteristic (country, firm, individual) removes all the variation *between* those countries, firms or individuals. This leaves us with the variation *within*. If our fixed effect is individuals, and we are observing individuals over time, we are relying on the variation of treatment over time (within each individual) for identification.

Fixed Effects Regression

To be specific, suppose that we observe N units over T dates and we have no missing data. For each unit i at time t we observe an outcome Y_{it} and a binary treatment variable $X_{it} \in \{0,1\}$. A typical fixed effects regression model is

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it},$$

where α_i is a fixed but unknown intercept term for individual i (or i's fixed effect) and ϵ_{it} is the mean-zero error term. The fixed effect term captures all unobserved time-invariant confounders U_i, such that we can define it as $\alpha_i = h(U_i)$.

To identify the treatment effect, we assume that the error term is independent of the treatment conditional on the fixed effect:

$$\mathbb{E}[\epsilon_{it} \mid X_i, \alpha_i] = 0,\ \forall i, t,$$

where X_i is a $T \times 1$ vector of treatment variables for i. This assumption is equivalent to stating that $\mathbb{E}[\epsilon_{it} \mid X_i, U_i] = 0$.

The parameter β is the average contemporaneous effect of X on Y. Importantly, it is computed only for the units that have variation in the treatment variable over time (have within variation). It represents the average treatment effect for the units that have within variation:

$$\beta = \mathbb{E}[Y_{it}^1 - Y_{it}^0 \mid \mathbb{I}(0 < \sum_{t=1}^{T} X_{it} < T)].$$

A straightforward method to estimate the assumed regression model would be to simply treat each fixed effect as a coefficient on an indicator variable that encodes each unit. Then we estimate those coefficients using OLS. This method, however, will become problematic when there are many fixed effects, and especially if there are more than one type of fixed effects (e.g., two-way fixed effects). Notice that for our research question we do not even need the estimates of these fixed effects. We just need to control for them to block all the back-door paths from the treatment to outcome.

An alternative method is called the *within estimator*. It uses differencing to overcome the issue of dealing with too many fixed effects. Specifically, define the time averages of the outcome and treatment variables as

$$\bar{Y}_i \equiv \frac{1}{T}\sum_{i=1}^{T} Y_{it},$$
$$\bar{X}_i \equiv \frac{1}{T}\sum_{i=1}^{T} X_{it}.$$

Then

$$\bar{Y}_i = \alpha_i + \beta\bar{X}_i + \bar{\epsilon}_i.$$

Subtracting this expression from the regression model, we get

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + \epsilon_{it} - \bar{\epsilon}_i.$$

Therefore, we can estimate the treatment effect by simply running the regression of time-demeaned outcome on time-demeaned treatment.

The simple regression model we presented above can be extended along several dimensions. First, it is common to include time-varying observables that can potentially confound the causal relationship. This would lead to a following model, where Z is the time-varying observable confounder.

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Z_{it} + \epsilon_{it}.$$

Second, it is also common to include time fixed effects, in addition to unit fixed effects. This results in a so-called two-way fixed effects model:

$$Y_{it} = \alpha_i + \tau_t + \beta X_{it} + \epsilon_{it}.$$

However, the interpretation of β as a treatment effect is more complicated than researchers have previously assumed. Recent papers by Imai & Kim (2021) and Chaisemartin and D'Haultfœuille (2020) show the estimated effect in such models is a weighted sum of the average treatment effects in each group and period, with weights that may be negative. This might lead to an issue that the regression coefficient is negative while all the ATEs are positive.

Fixed Effects in Action

Marriage and Earnings

(Adapted from *The Mixtape, Chapter 8*).

A well-known fact from labor economics is that the earnings of married men are higher than the earnings of unmarried men. This relationship holds after controlling for observables. But is this relationship causal? What if there are unobserved confounders that affect both the marital status and earnings?

Cornell and Rupert (1997) attempt to answer this question by using panel data methods on the National Longitudinal Survey of Young Men. They estimate a version of the following regression model:

$$Y_{it} = \alpha_i + \delta M_{it} + \beta X_{it} + \epsilon_i t,$$

where the outcome is wage, the treatment variable is M (marital status), and X are other variables that change over time that might confound the treatment effect. The fixed effects α_i correspond to the characteristics that do not change over time, which might include both observables (race, gender) and unobservables (ability, family background).

Cornell and Rupert (1997) estimate both a feasible generalized least squares model and three fixed effects models with different sets of controls.

Dependent variable	FGLS	Within	Within	Within
Married (coef.)	0.083	0.056	0.051	0.033
S.E.	(0.022)	(0.026)	(0.026)	(0.028)
Education controls	Yes	No	No	No
Tenure	No	No	Yes	Yes
Quadratics in years married	No	No	No	Yes

Notice how the coefficient on *Married* drops with each specification and becomes insignificant in the last one.

Union Wages

(Adapted from *MH, Chapter 5*)

A classic question in labor economics is the effect of unions on wages. Workers who are union members typically earn more than non-union members. But is there a causal effect of union membership on wages? Or is simply because workers who join unions are also more skilled and would have earned more anyway?

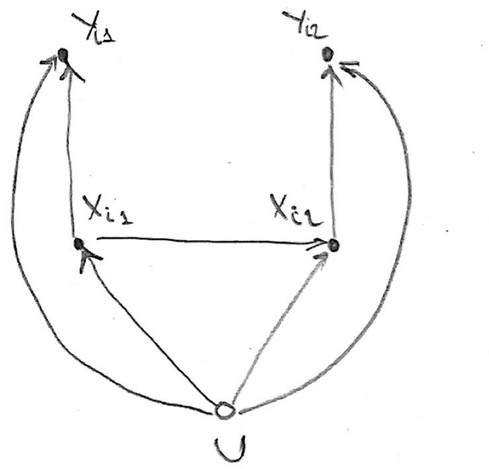
Freeman (1984) uses panel data methods and four data sets to estimate union wage effects. For each data set, the table shows the results from a fixed-effects regression and the corresponding pooled OLS estimates.

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

The cross section estimates are typically higher than the fixed effects estimates, which might suggest positive selection.

Causal Diagram of Fixed Effects

Let's review the assumptions of the regression we introduced previously using a DAG. Suppose that we are interested in the causal effect of X on Y and we have a panel dataset in which we have observations over individuals (indexed by i) and two time periods (index by $t = 1, 2$). We have a set of unobservable confounders U that can affect both the treatment and outcome. Consider the following causal diagram (adapted from from Imai & Kim (2019)).



Let's go over the assumptions behind the data-generating process that are shown in this DAG.

1. No unobserved time-varying confounder exists.
2. Past outcomes do not directly affect current outcome.
3. Past outcomes do not directly affect current treatment.
4. Past treatments do not directly affect current outcome.

The most critical assumption here is the first one. It means that the only unobserved confounders we have (summed up by U_i) are the ones that vary across individuals but *not* over time. On the graph, this assumptions corresponds to the absence of the time subscript on U. The confounder does affect, however, the treatment variable in each period, as well as the outcome variable.

The second assumption corresponds to the absence of arrows from previous outcomes Y to future outcomes. The third assumption corresponds to the absence of arrows from past outcomes to future treatments. Finally, the fourth assumption corresponds to no arrows from previous treatments to future outcomes.

In the fixed effects design, we do not control for U directly (it is unobserved) but rather for the variable (a characteristic) that subsumes U. How does controlling for that characteristic (and hence U) identify the causal effect of X on Y? Let's list all of the paths between X_1 and Y_1:

1. $X_1 \rightarrow Y_1$ (causal path)
2. $X_1 \leftarrow U \rightarrow Y_1$ (back-door path)
3. $X_1 \rightarrow X_2 \leftarrow U \rightarrow Y_1$ (back-door path)
4. $X_1 \rightarrow X_2 \rightarrow Y_2 \leftarrow U \rightarrow Y_1$ (back-door path)

Clearly, controlling for U closes all of the back-door paths between X_1 and Y_1. Also, the last two back-door paths have colliders. How about the effect of X_2 on Y_2?

1. $X_2 \rightarrow Y_2$ (causal path)
2. $X_2 \leftarrow U \rightarrow Y_2$ (back-door path)
3. $X_2 \rightarrow X_1 \leftarrow U \rightarrow Y_2$ (back-door path)
4. $X_2 \rightarrow X_1 \rightarrow Y_1 \leftarrow U \rightarrow Y_2$ (back-door path)

Again, controlling for U closes all of the back-door paths, plus the last two back-door paths are closed because they have colliders.

The key identifying assumption of fixed effects, no unobserved time-varying confounders, cannot be tested with the data. Other assumptions can be relaxed, to a certain degree. For example, Imai & Kim (2019)) show that one can relax assumptions 3 or 4 by including the lags of either outcome or treatment variables (respectively), or both.