Maximum Likelihood Methods for the Social Sciences
POLS 510 · CSSS 510

Introduction to Multilevel Models

Christopher Adolph

Political Science *and* CSSS
University of Washington, Seattle

M C Escher *Belvedere* 1958

Multi-level data structures

Data are often structured into different "levels"

These levels are often hierarchical, but might also be non-nested

Response variable might be continuous, binary, count, etc.

Three examples:

1. Votes in a presidential election grouped by state
   - Two levels; nested
2. Test scores of students grouped by class, school, and district
   - Four levels; nested
3. Devolution of authority by country, policy area & policy tool
   - Two levels; non-nested

Models of multilevel data

$$y_{ij} = \alpha + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Suppose our data are for individuals i nested in groups j

For simplicity, assume y_{ij} is Normally distributed, so we could use LS

What problems might the model above pose?

Models of multilevel data

$$y_{ij} = \alpha + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Suppose our data are for individuals i nested in groups j

For simplicity, assume y_{ij} is Normally distributed, so we could use LS

What problems might the model above pose?

What about iid? Aren't members of the same group likely correlated?

What do we do?

Models of multilevel data

$$y_{ij} = \alpha + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Suppose our data are for individuals i nested in groups j

For simplicity, assume y_{ij} is Normally distributed, so we could use LS

What problems might the model above pose?

What about iid? Aren't members of the same group likely correlated?

What do we do?

Simplest solution: let each group have its own intercept

Models of variable intercepts

$$y_{ij} = \alpha_i + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Then there are a range of possibilities for group-specific intercepts:

Let α_i be a random variable with no systemic component.
This type of α_i known as a *random effect*:

$$\alpha_i \sim \mathrm{Normal}\left(0, \sigma^2_\alpha\right)$$

Models of variable intercepts

$$y_{ij} = \alpha_i + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Then there are a range of possibilities for group-specific intercepts:

Let α_i be a random variable with no systemic component.
This type of α_i known as a *random effect*:

$$\alpha_i \sim \mathrm{Normal}\left(0, \sigma_\alpha^2\right)$$

Let α_i be a systematic component with no stochastic component.
This type of α_i is known as a *fixed effect*:

$$\alpha_i = \alpha_i^*$$

Models of variable intercepts

$$y_{ij} = \alpha_i + \mathrm{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$$

Then there are a range of possibilities for group-specific intercepts:

Let α_i be a random variable with no systemic component.
This type of α_i known as a *random effect*:

$$\alpha_i \sim \mathrm{Normal}\left(0, \sigma_\alpha^2\right)$$

Let α_i be a systematic component with no stochastic component.
This type of α_i is known as a *fixed effect*:

$$\alpha_i = \alpha_i^*$$

Let α_i be a random variable with a unit-specific systematic component
This type of α_i known as a *mixed effect*:

$$\alpha_i \sim \mathrm{Normal}\left(\alpha_i^*, \sigma_\alpha^2\right)$$

Random effects

$$\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$$

Intuitive from a maximum likelihood modeling perspective

A unit specific error term

Assumes the units come from a common population,
with an unknown (estimated) variance, σ_α^2

In likelihood inference, estimation focuses on this variance, not on particular α_i's

Uncorrelated with $\mathbf{x}_i$ by design

Need MLE (or Bayes) to estimate

Random effects example

A (contrived) example may help clarify what random effects are.

Suppose that we have data following this true model:

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{it} + \alpha_i + \varepsilon_{it} \\
\alpha_i &\sim \mathcal{N}(0, \sigma^2_\alpha) \\
\varepsilon_{it} &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

with $i \in \{1, \ldots N\}$ and $t \in \{1, \ldots T\}$

It may help to pretend that these data have a real world meaning,
though remember throughout we have conjured them from thin air and `rnorm()`

Let's pretend these data reflect undergraduate student exam scores over a term for $N = 100$ students and $T = 5$ exams

Random effects example: Student aptitude & effort

Let's pretend these data reflect undergraduate student exam scores

over a term for $N = 100$ students and $T = 5$ exams:

$$
\begin{aligned}
\text{score}_{it} &= \beta_0 + \beta_1 \text{hours}_{it} + \alpha_i + \varepsilon_{it} \\
\alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2) \\
\varepsilon_{it} &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

with $i \in \{1, \ldots N\}$ and $t \in \{1, \ldots T\}$

The response is the exam score, score_{it}

and the covariate is the hours studied before the exam, hours_{it}

and each student has an unobservable aptitude α_i which is Normally distributed

Aptitude has the same (random) effect on each exam by a given student

Random effects example: Student aptitude & effort

Let's pretend these data reflect undergraduate student exam scores

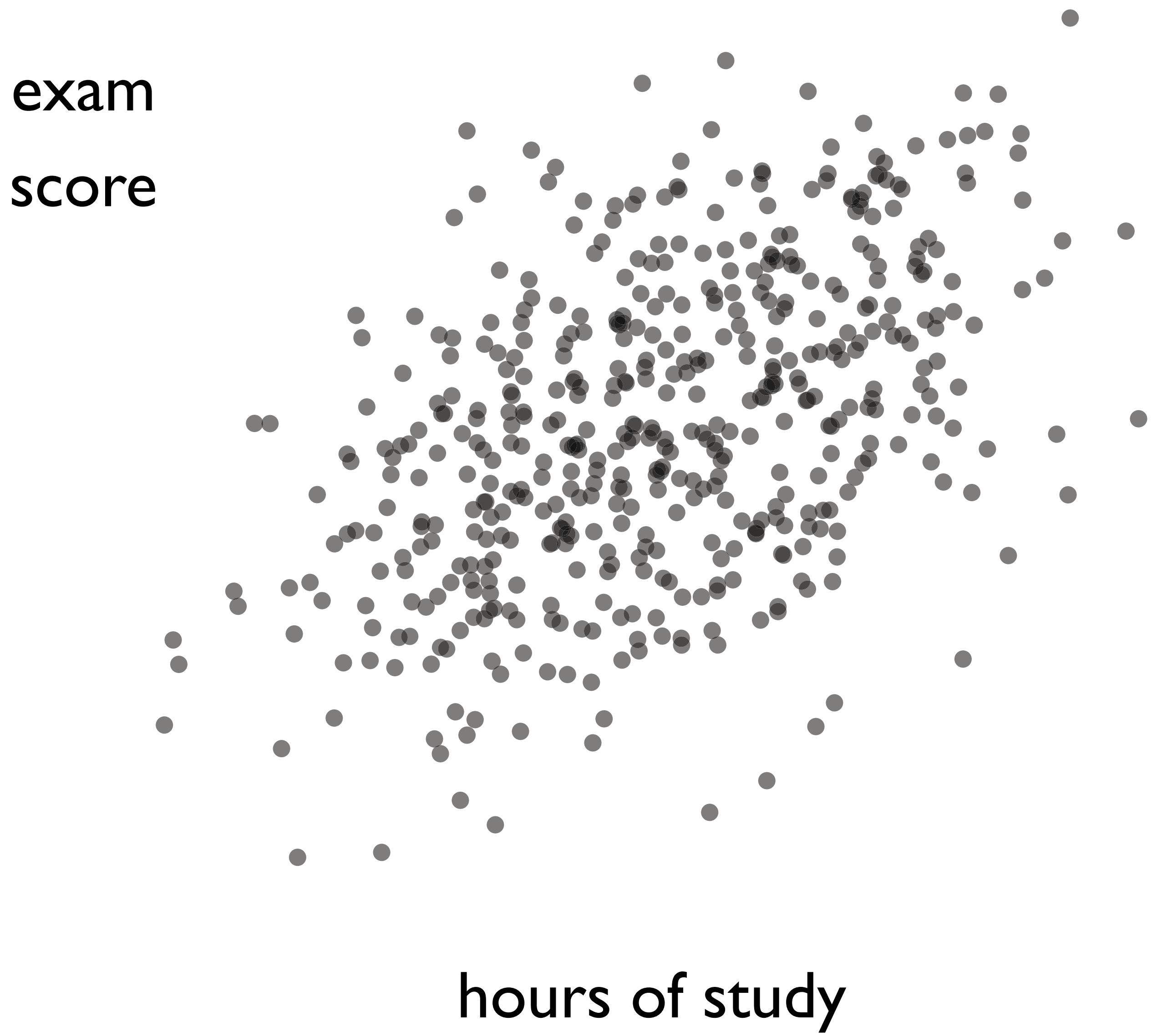
over a term for $N = 100$ students and $T = 5$ exams:

$$
\begin{aligned}
\text{score}_{it} &= 0 + 0.75 \times \text{hours}_{it} + \alpha_i + \varepsilon_{it} \\
\alpha_i &\sim \mathcal{N}(0, 0.7^2) \\
\varepsilon_{it} &\sim \mathcal{N}(0, 0.2^2)
\end{aligned}
$$

with $i \in \{1, \ldots 100\}$ and $t \in \{1, \ldots 5\}$

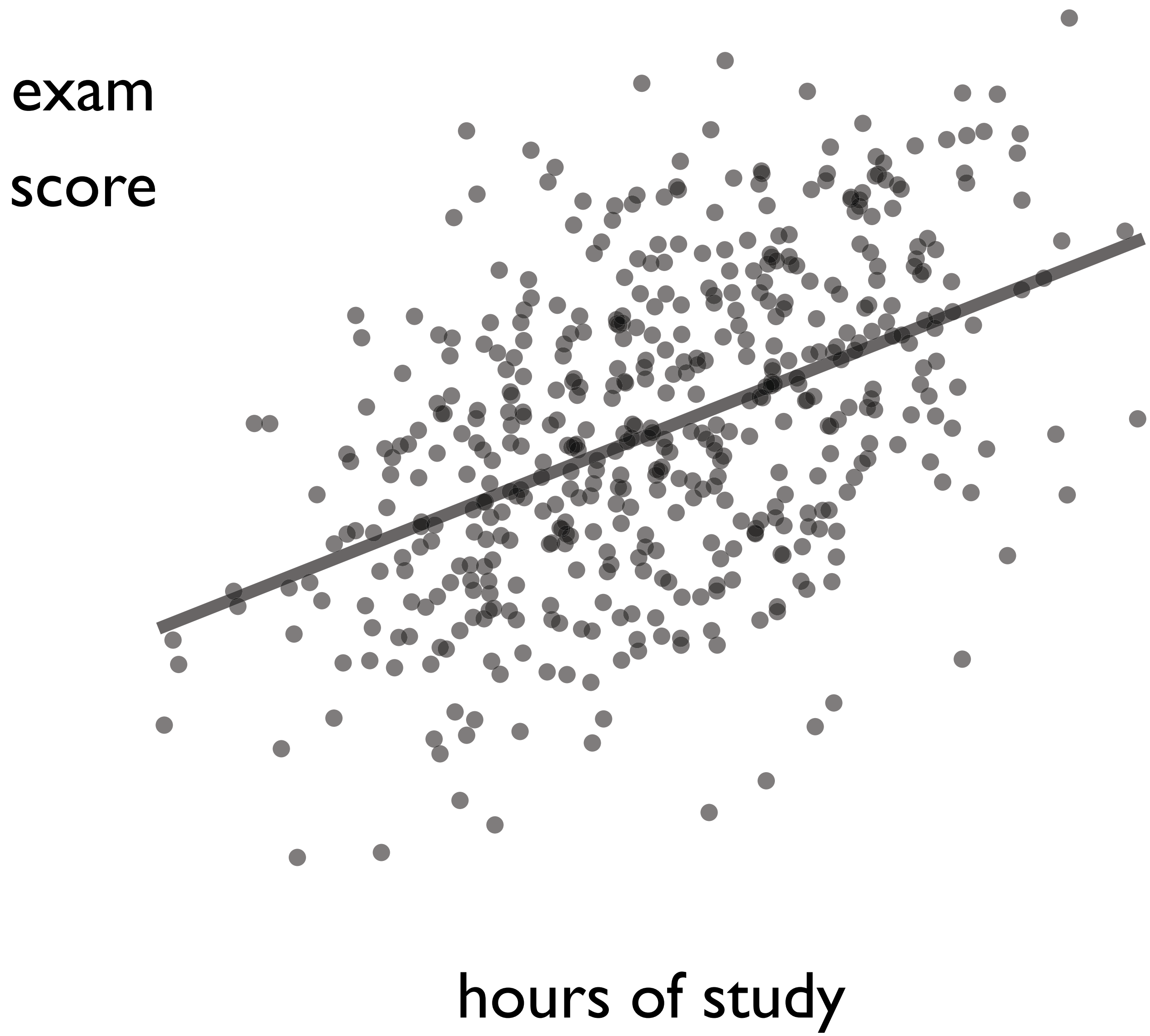
The above are the true values of the parameters I used to generate the data

Let's see what role the random effect α_i plays here



The 500 obervations

A relationship between effort & scores seems evident

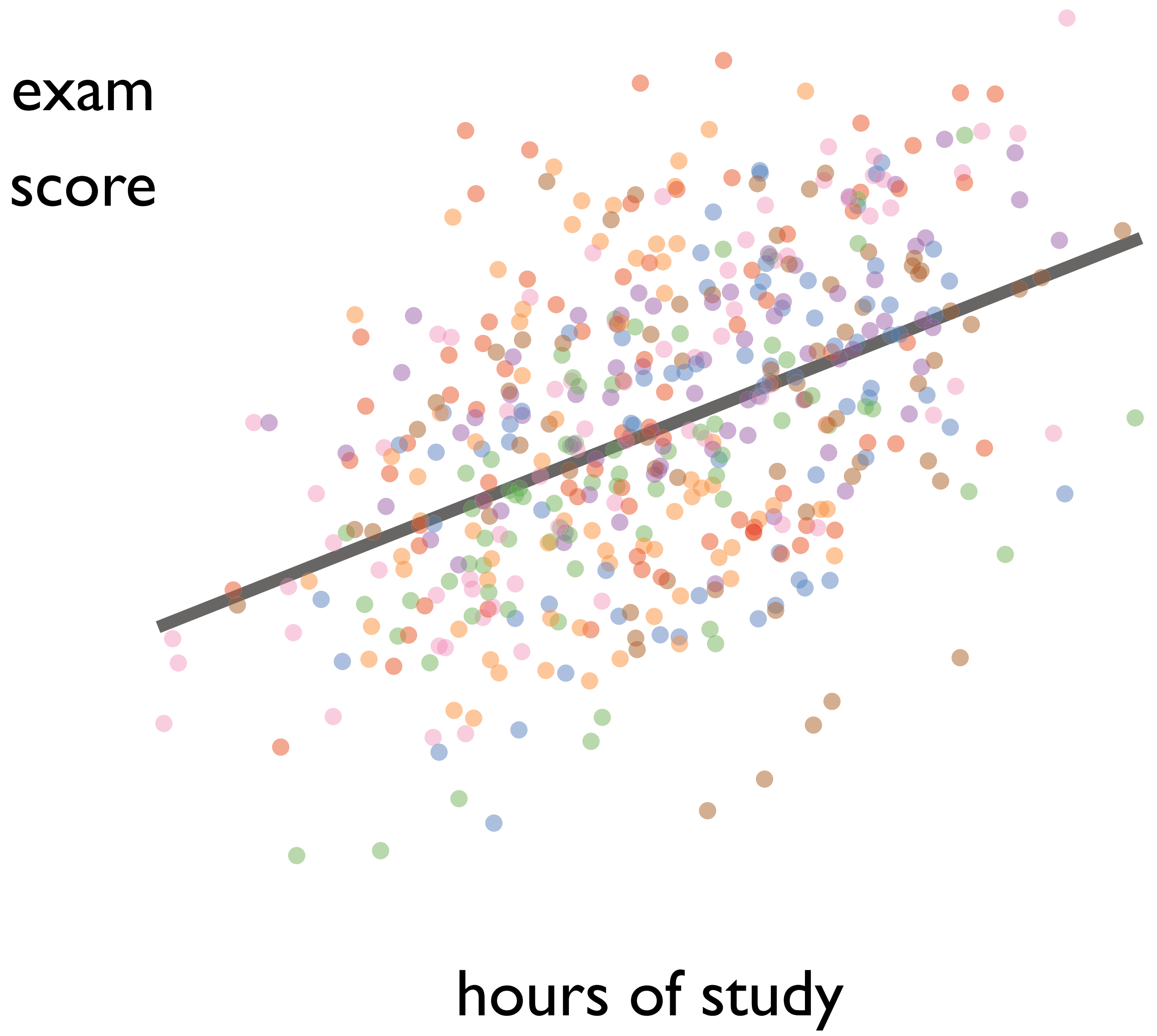


Let's summarize the relationship using the least squares $\hat{\beta}_1$

Approximately equal to the true $\beta_1 = 0.75$

Haven't discussed, used, or estimated the random effects yet

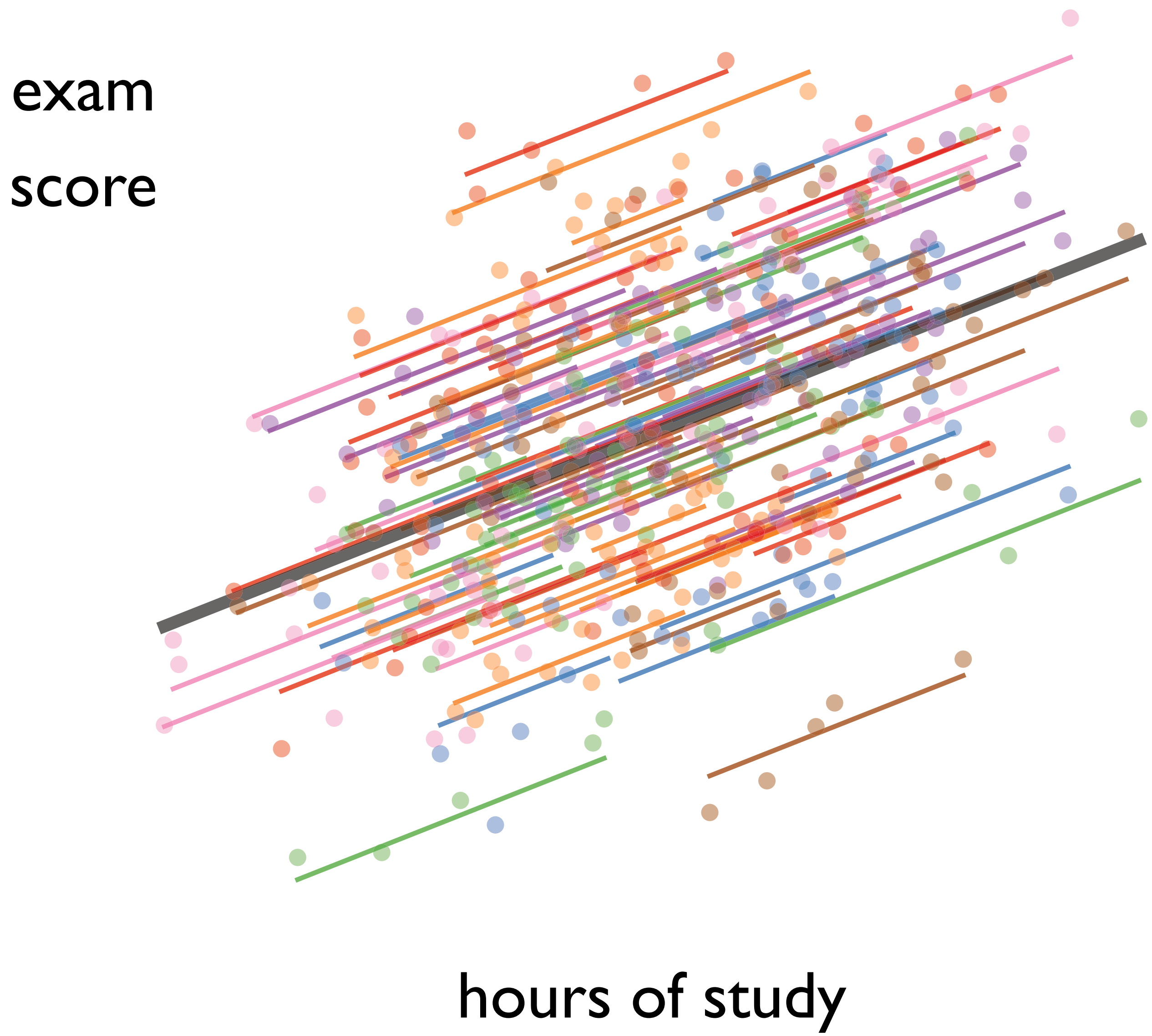
Do we need them?



Identified each of the 100 students using colored dots (we have 8 colors; they repeat)

Clear that each student's scores are tightly clustered

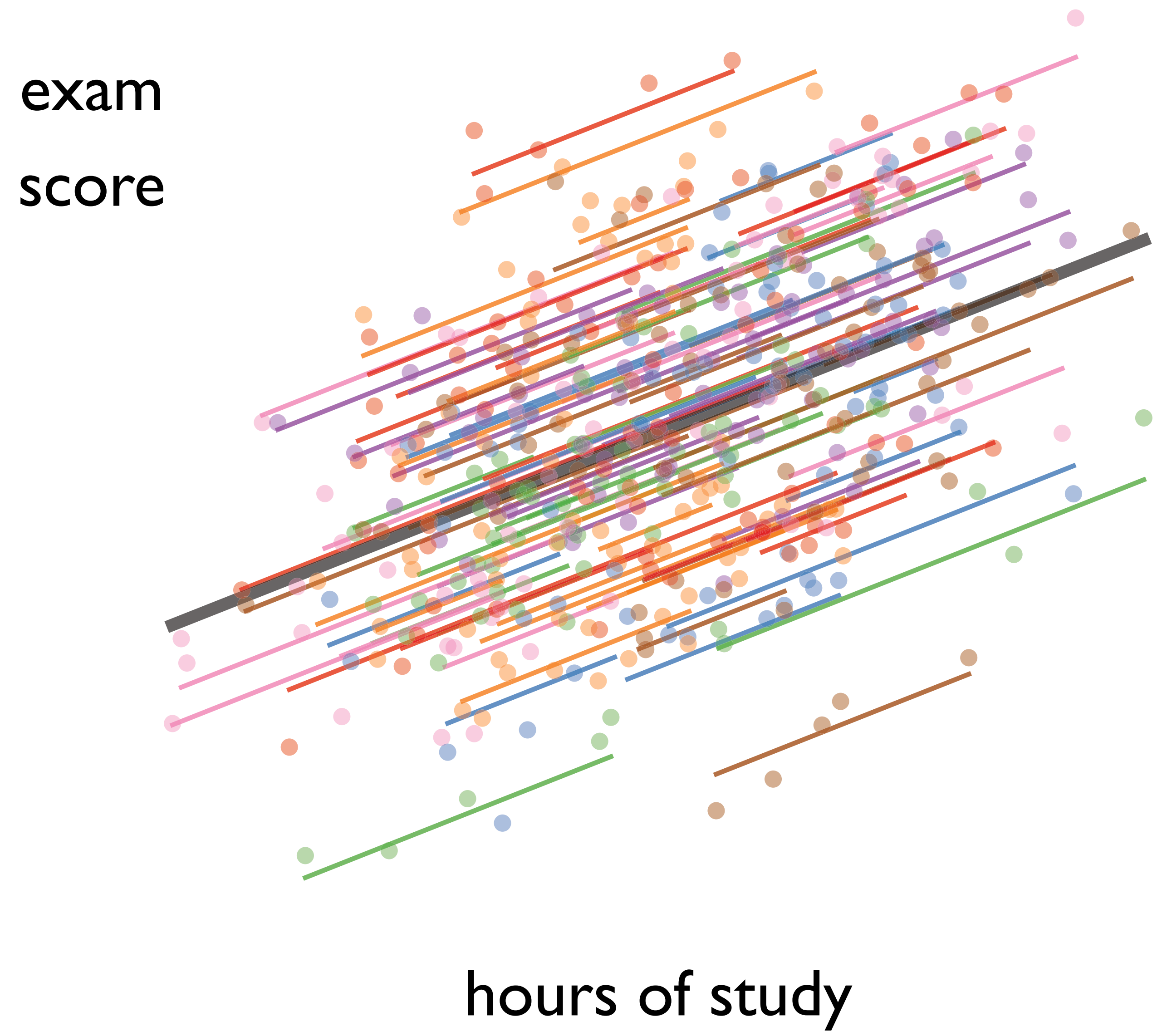
Note the student-level slopes



Each student follows the same regression line as the whole class, but with a unique intercept

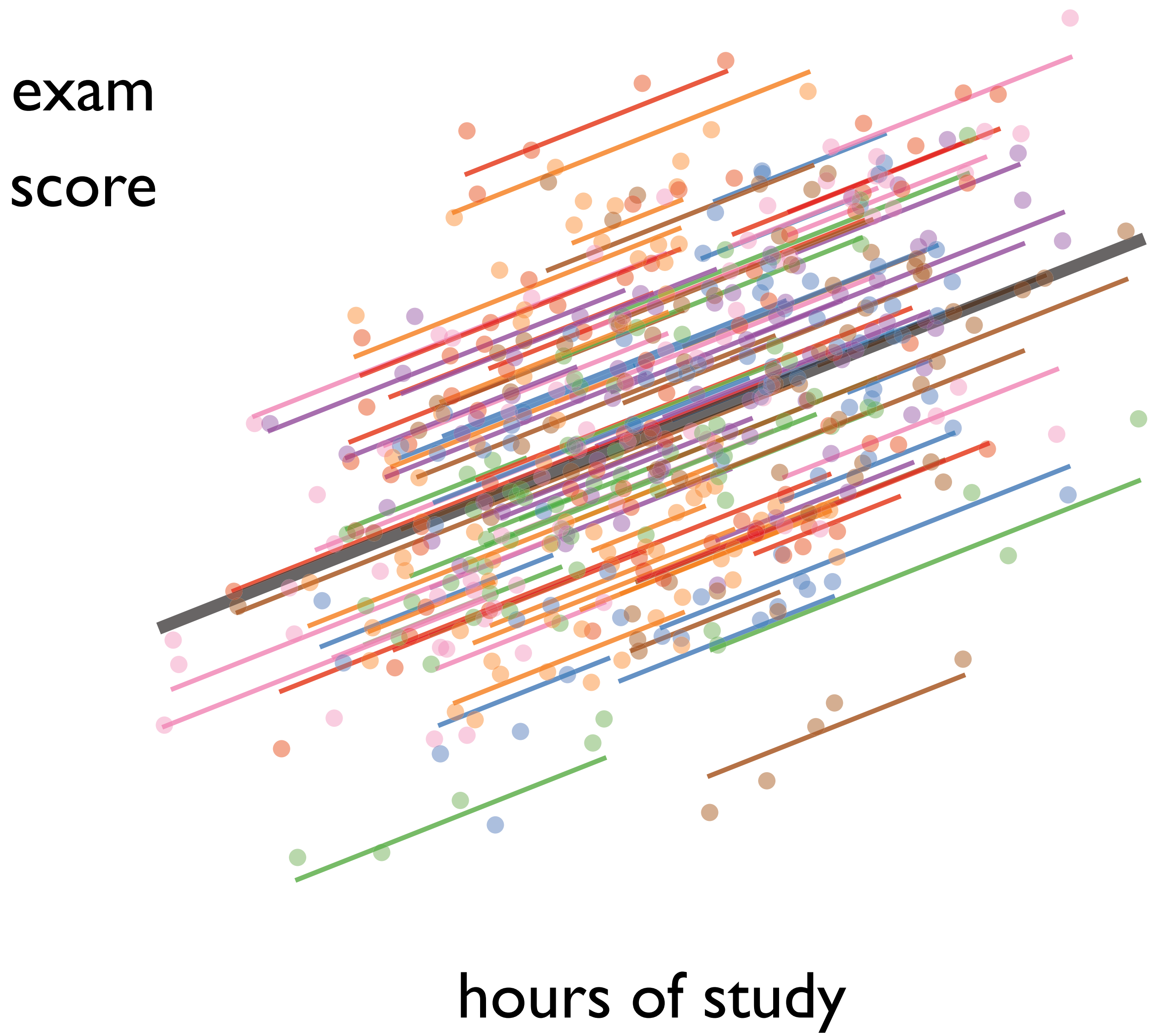
That intercept is the random effect α_i

It's also the average difference between student i's scores and the class-level regression line



The student random effect is the student-specific component of the error term

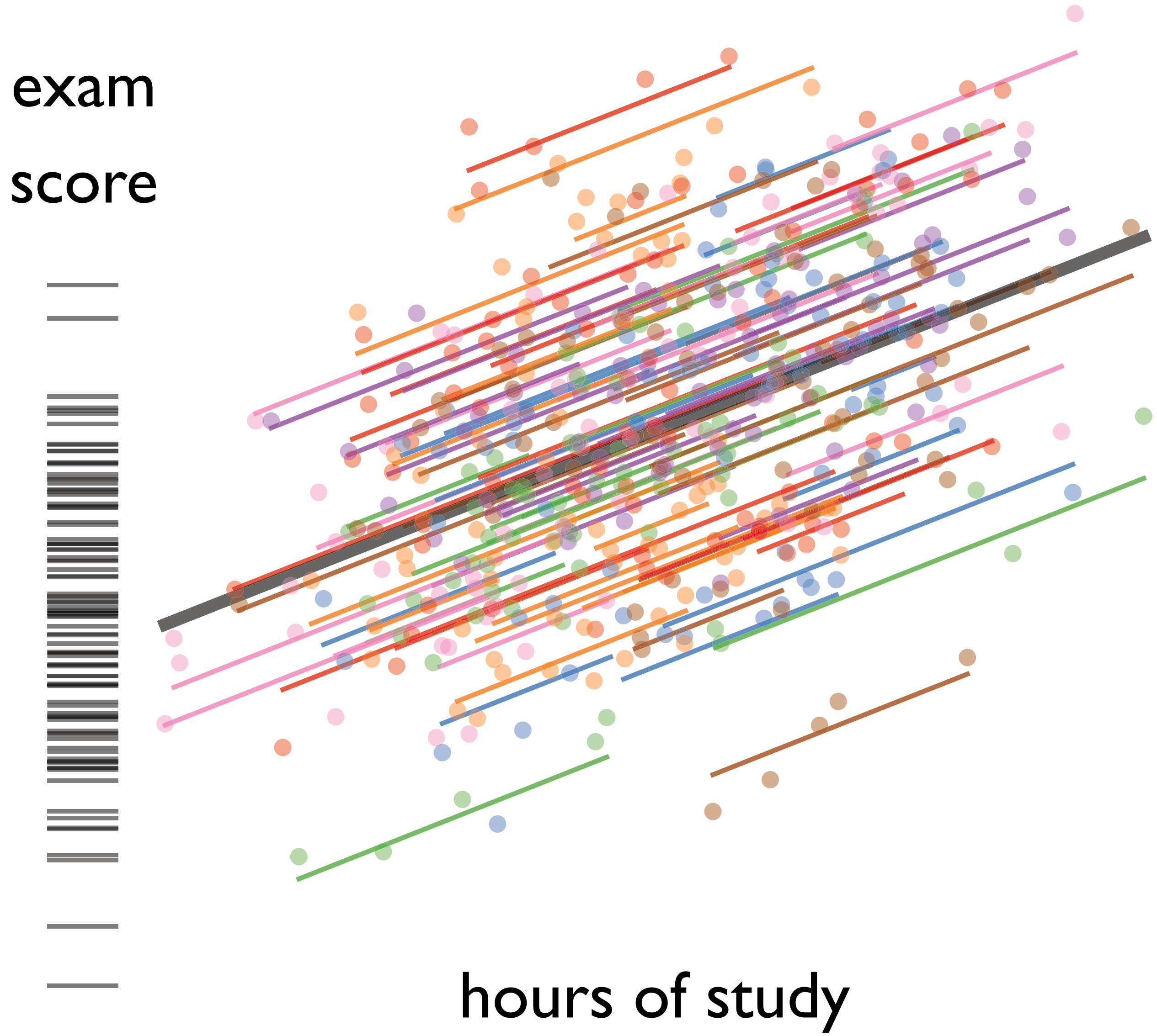
After we remove it, a student's scores across exams exhibit white noise variation around a student-specific version of the overall regression line



These random effects α_i reflect the portion of the error term that results from unmeasured student characteristics

I've labelled this random component "aptitude"

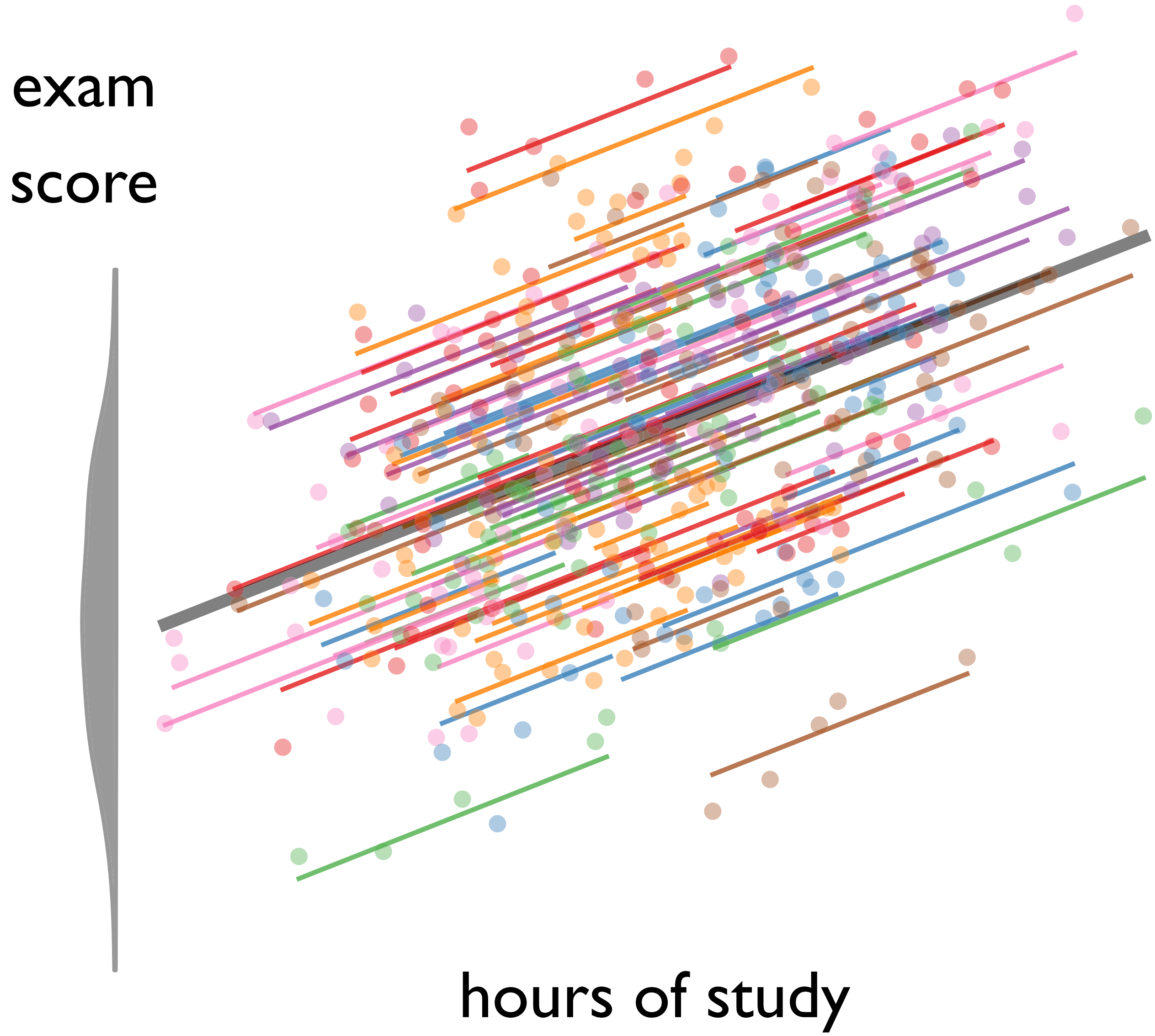
But that's is just a word for everything related to a student's ability



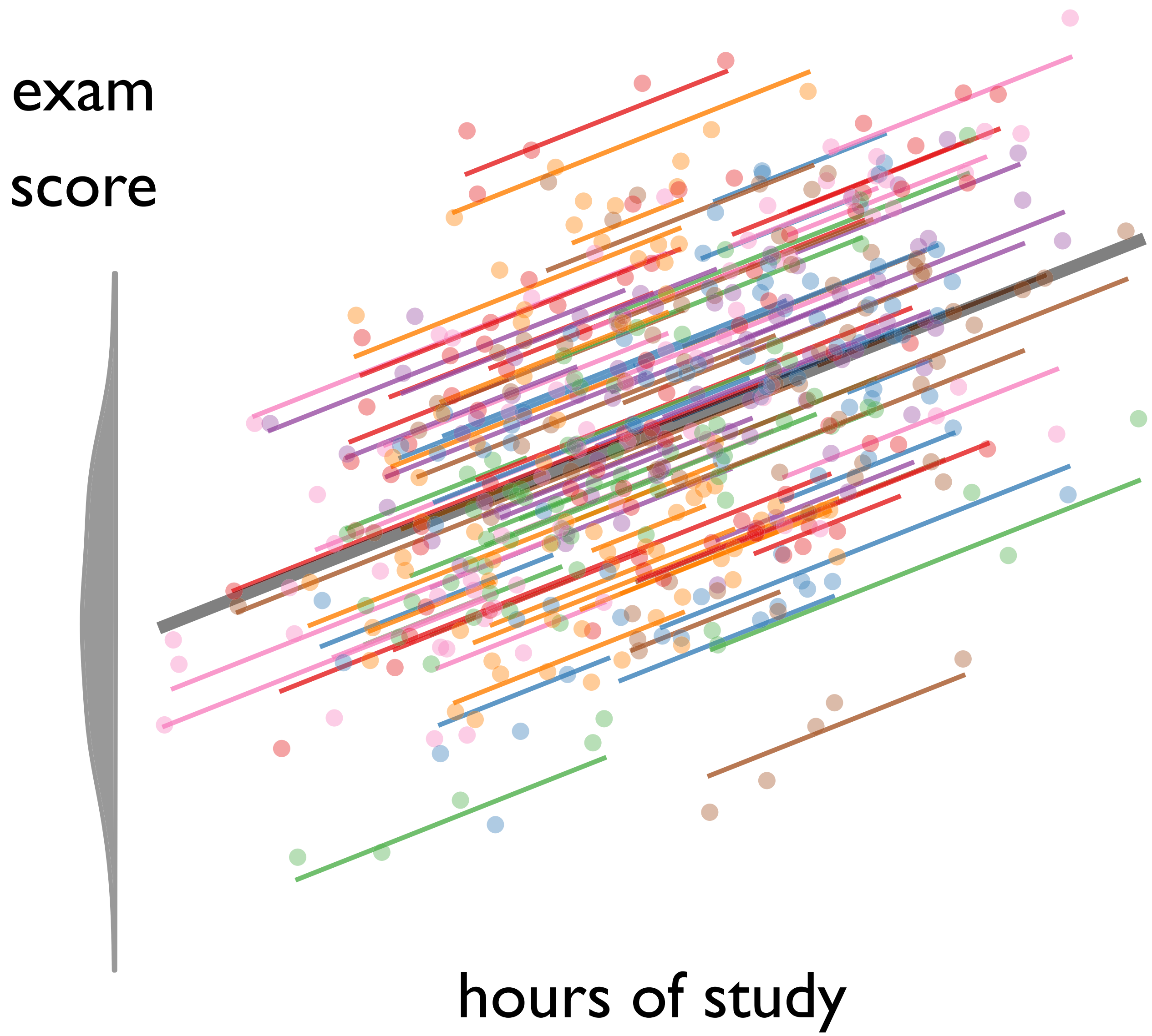
The distribution of the random effects is shown at the left

A plot of a marginal distribution on the side of a scatterplot is called a "rug"

NB: `tile` can make rugs using `rugTile()` traces



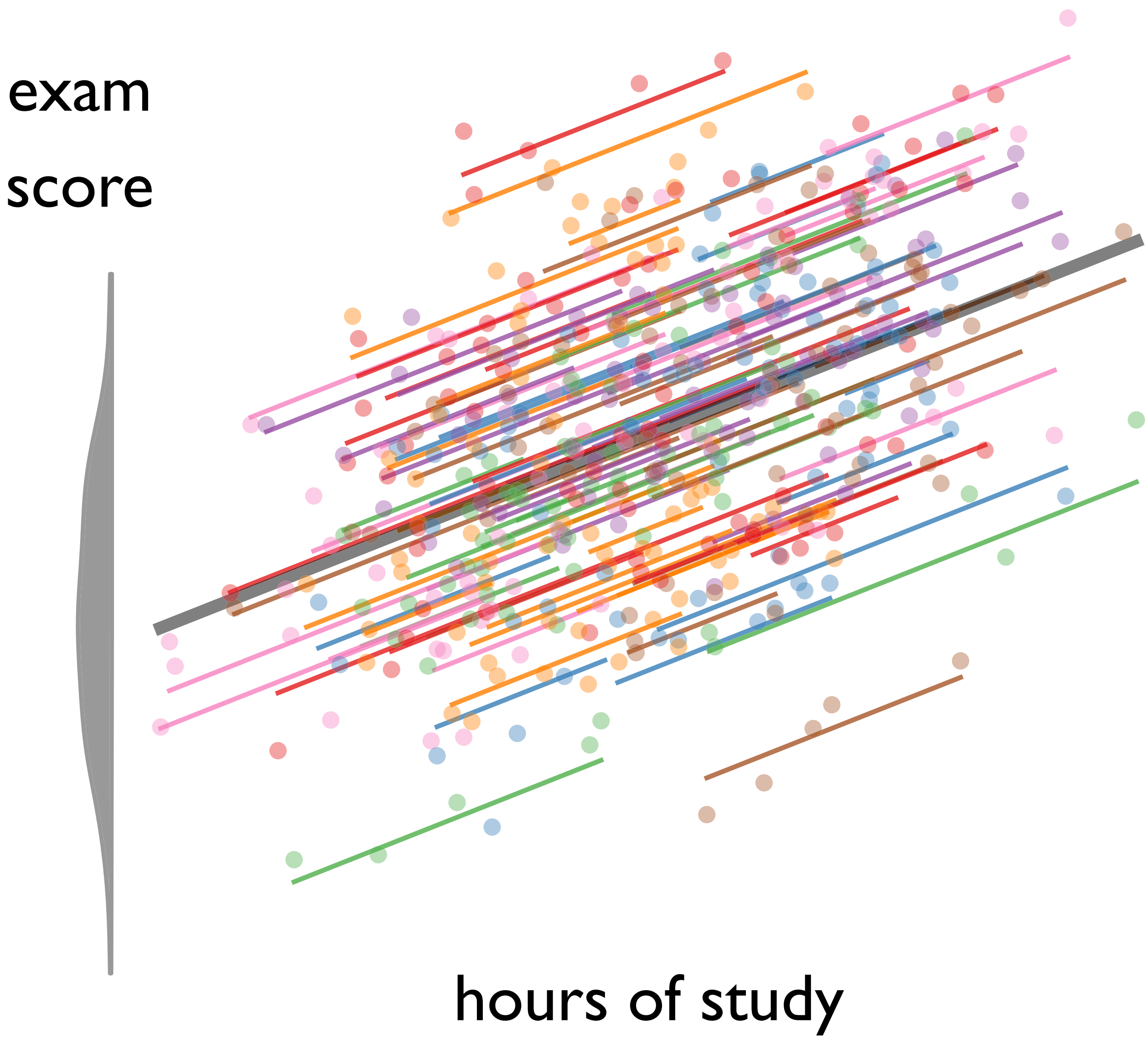
A density plot of the distribution of random effects suggests they are approximately Normal



Random effects are a decomposition of the error term into

1. a unit-specific part
2. an idiosyncratic part

Random effects are determined after we have the overall regression slope and cannot change that slope



The model is now hierarchical or multilevel

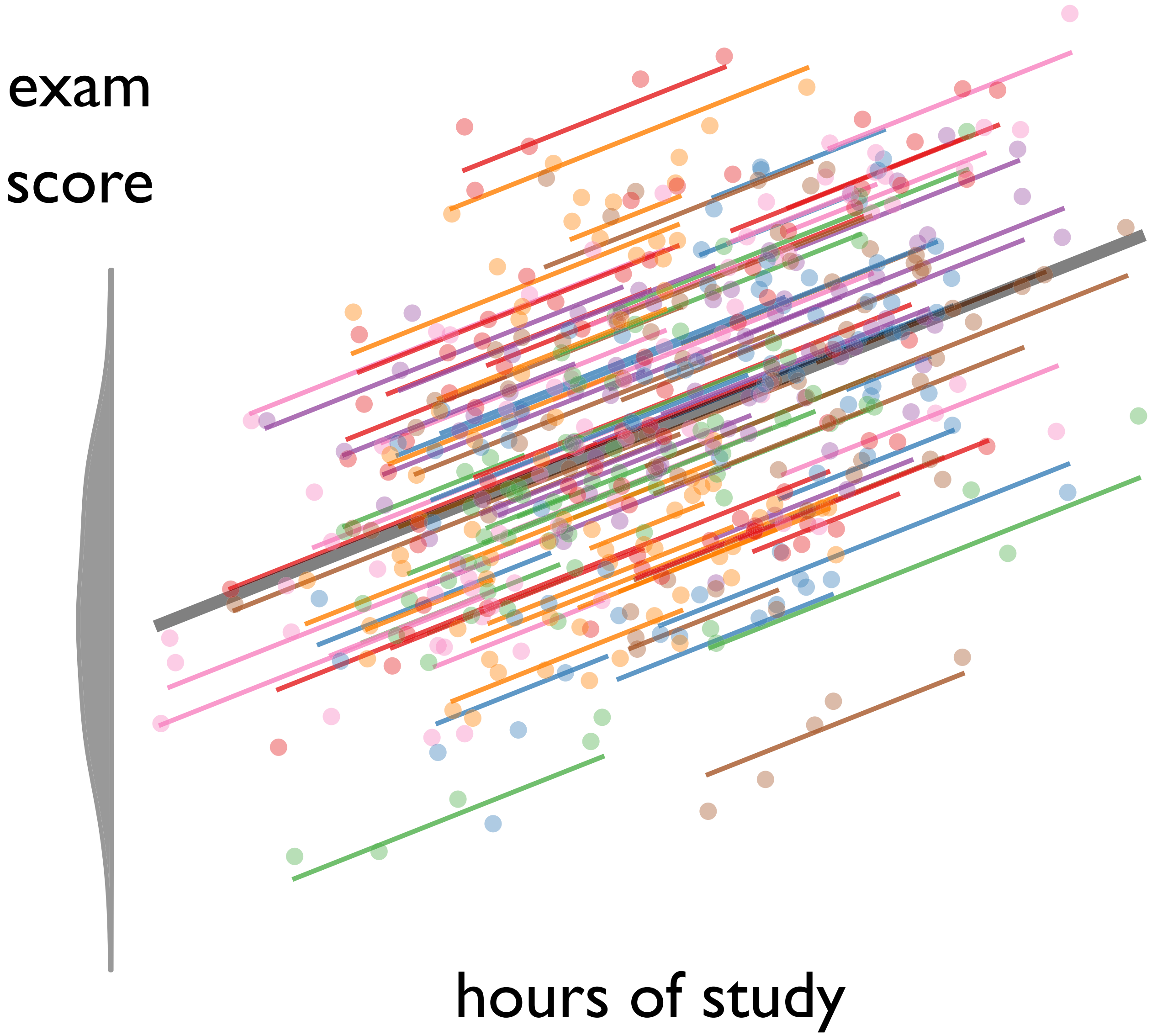
Level 1: Student level

sits above

Level 2: Student × exam level

There's random variation at both levels

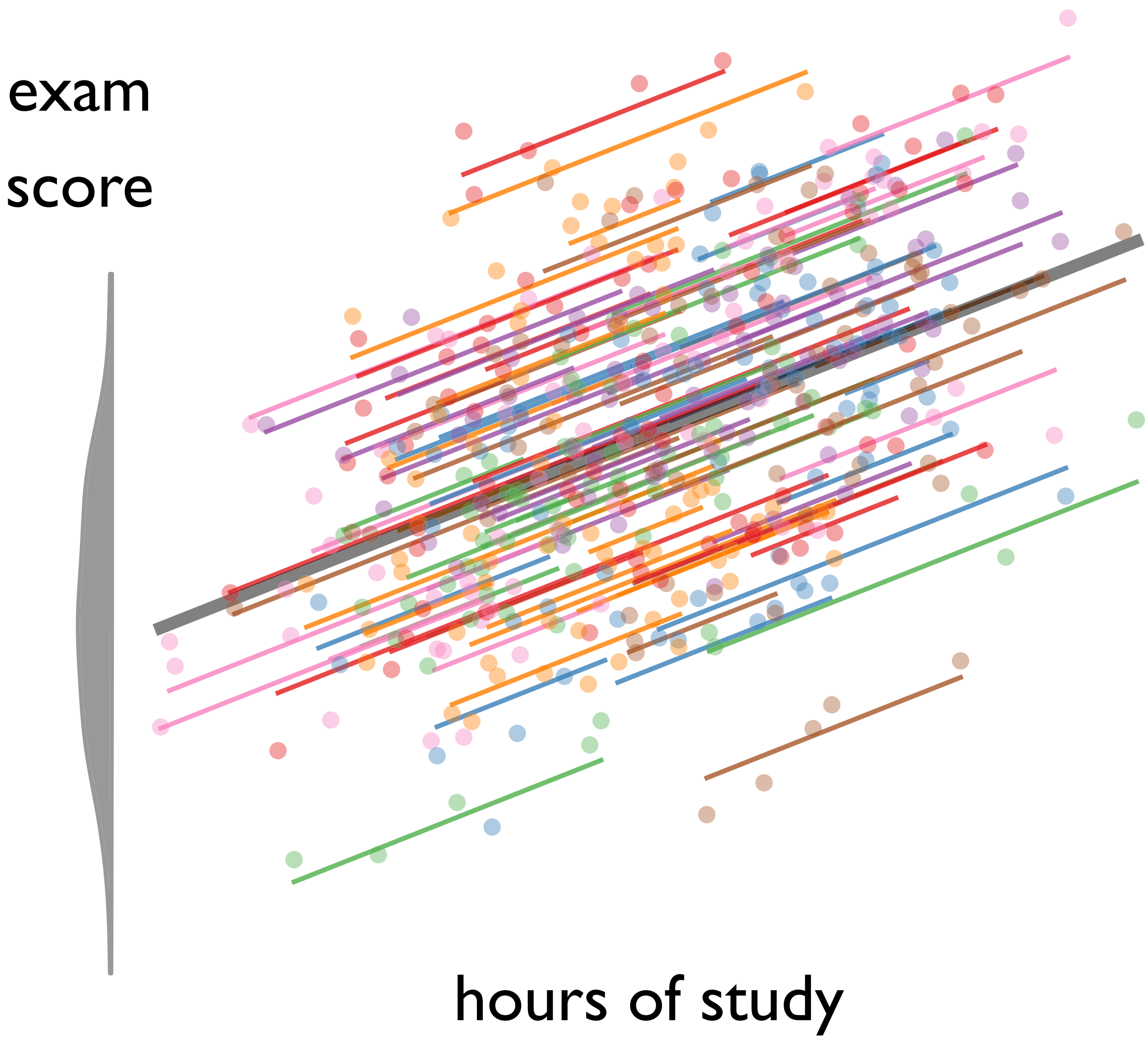
But mainly at the student level



Students randomly vary a lot: $\sigma_\alpha = 0.7$

Exams for a given student vary little: $\sigma_\varepsilon = 0.2$

Student level random effects comprise $100\% \times \sqrt{0.7^2/(0.7^2+0.2^2)} = 96\%$ of the total error variance



We haven't controlled for any omitted confounders

What if unmeasured ability were correlated with study effort?

Our $\hat{\beta}_1$ estimate would be biased

This bias persists even if we allow for random effects

Random effects example: Student aptitude & effort

Suppose that ability *is* correlated with effort

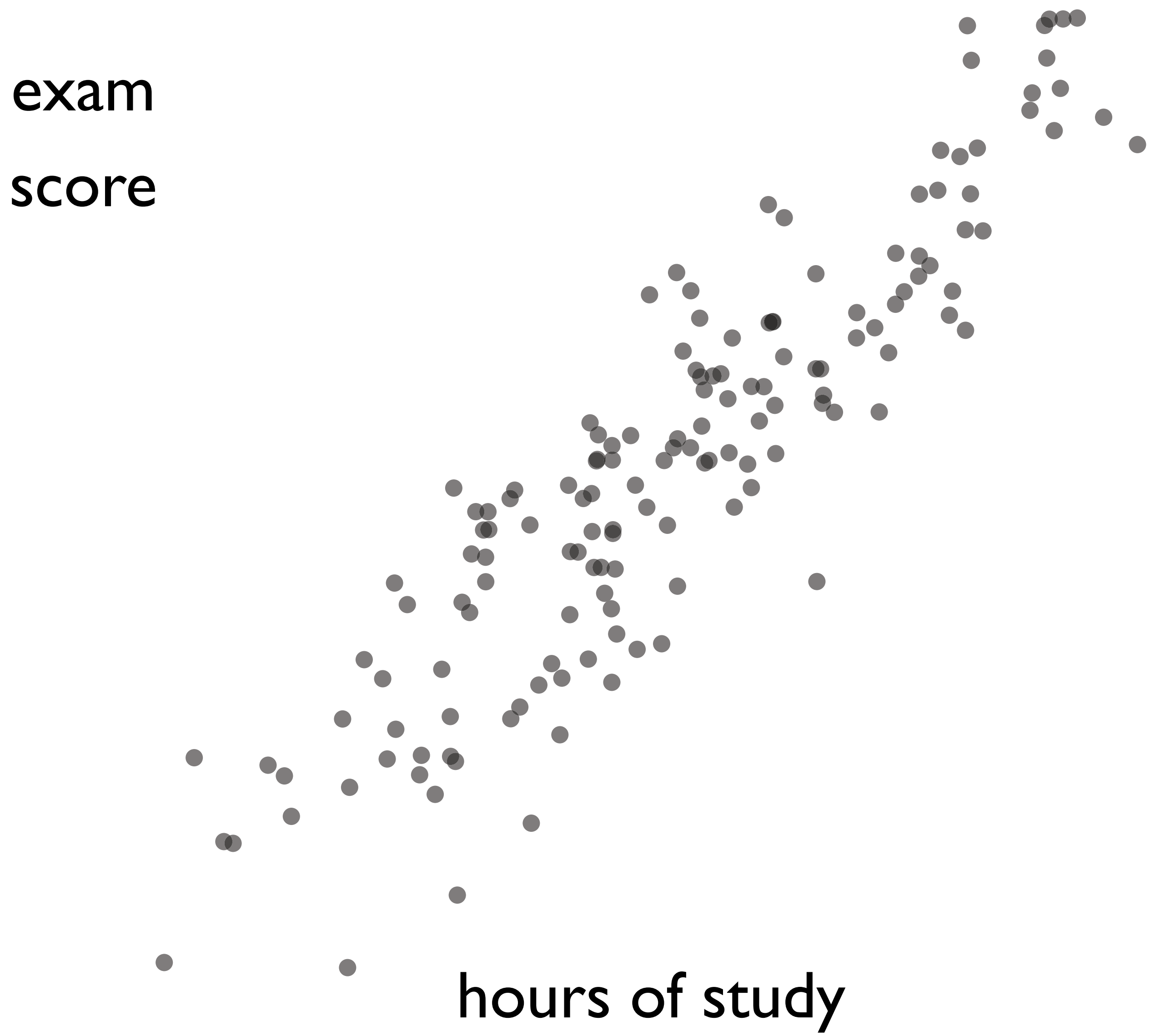
For example, perhaps high ability students rationally choose to study harder as their best available human capital investment opportunity

We have the same model, but now hours_{it} is a function of α_i:

$$\begin{aligned} \text{score}_{it} &= 0 + 0.75 \times \text{hours}_{it} + \alpha_i + \varepsilon_{it} \\ \text{hours}_{it} &= 0 + 0.5 \times \alpha_i + \text{uniform}(-0.7, 0.7) \\ \alpha_i &\sim \mathcal{N}(0, 0.7^2) \\ \varepsilon_{it} &\sim \mathcal{N}(0, 0.2^2) \end{aligned}$$

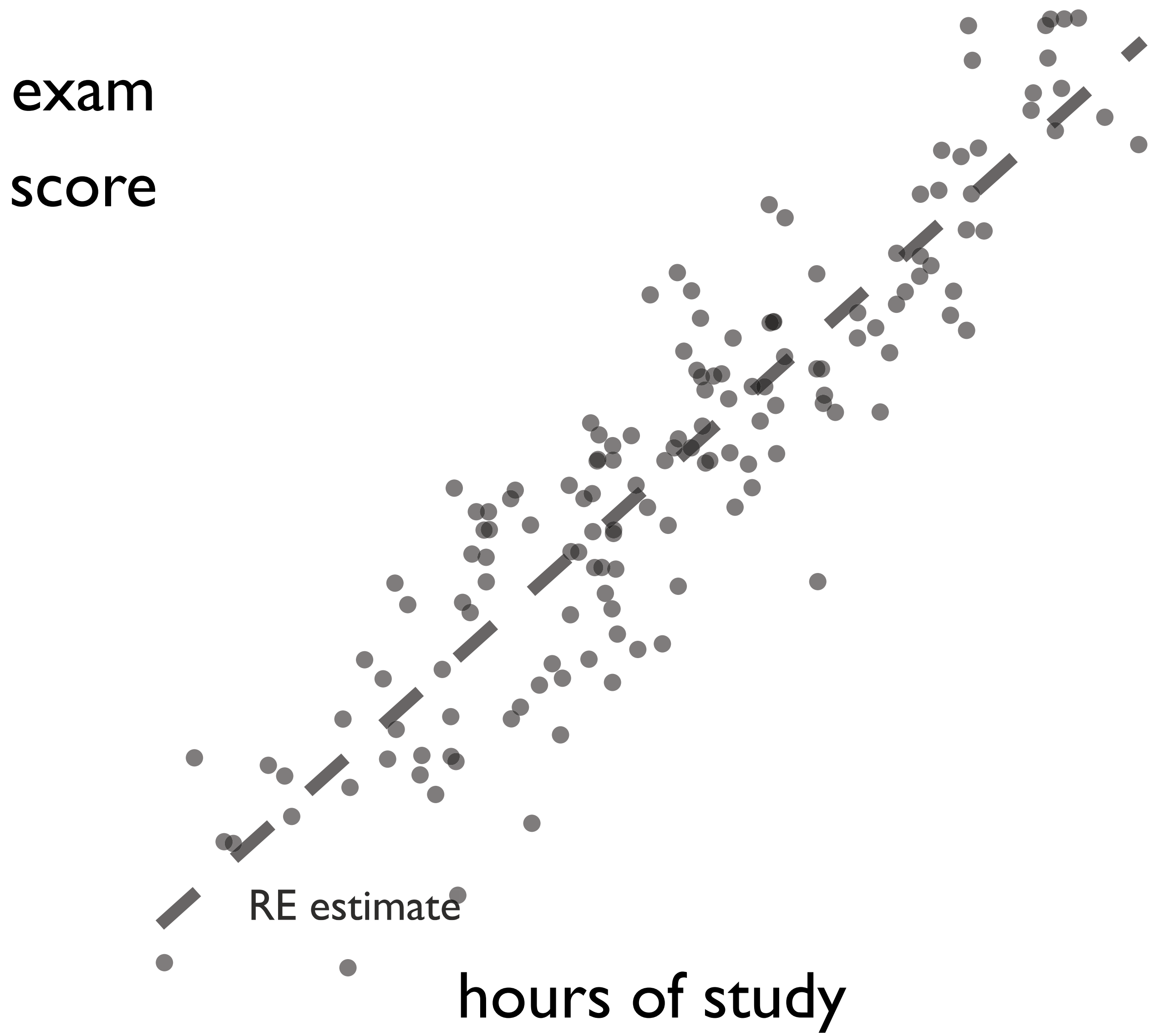
with $i \in \{1, \ldots 100\}$ and $t \in \{1, \ldots 5\}$

What happens when we estimate a treat α_i as a random effect and estimate $\hat{\beta}_1$?



I've shown only the first 30 students to make the graph easier to read

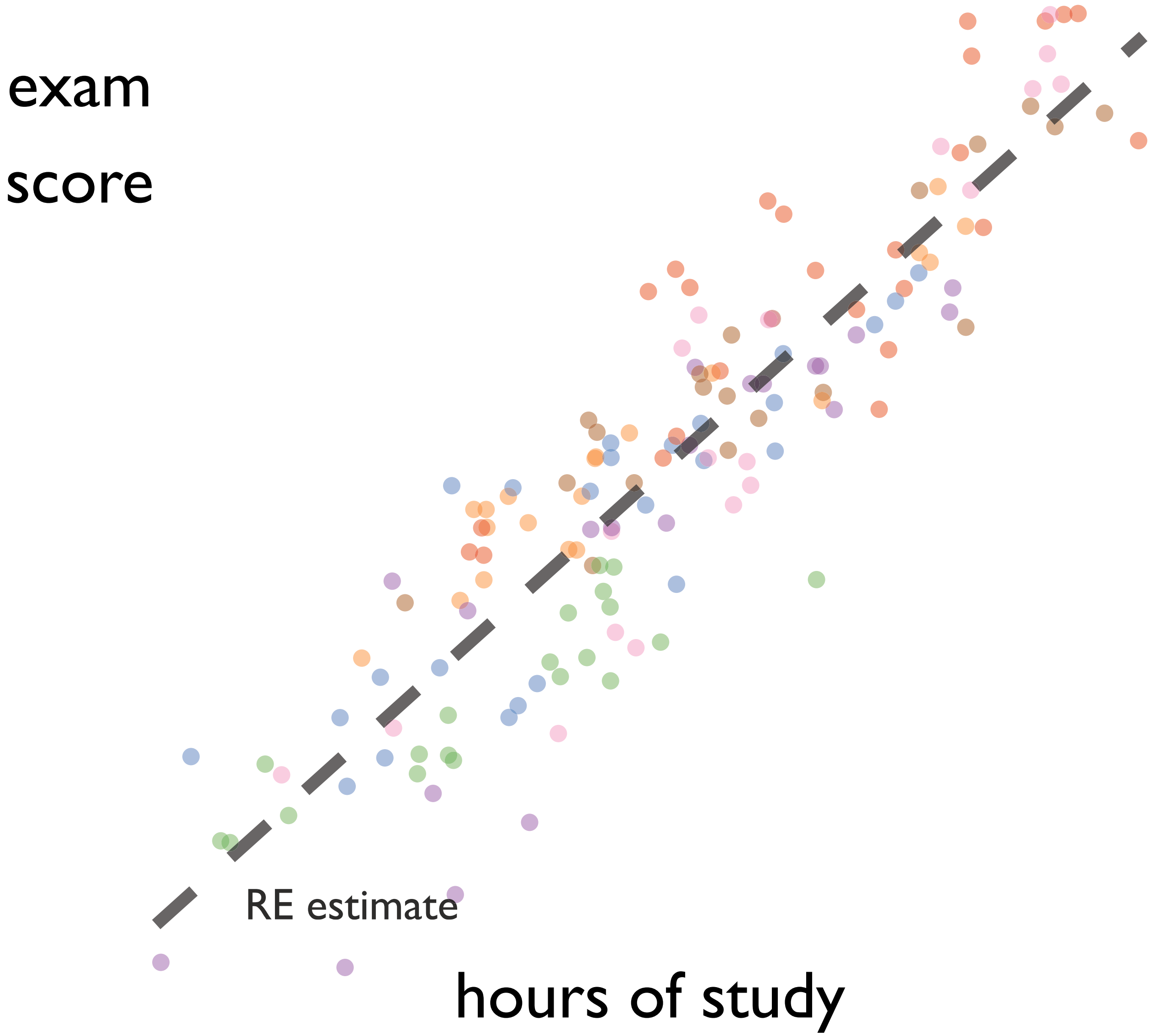
A *stronger* relationship between effort and grades seems evident



Least squares model finds $\hat{\beta}_1 \approx 1.6$

More than double the true value of 0.75!

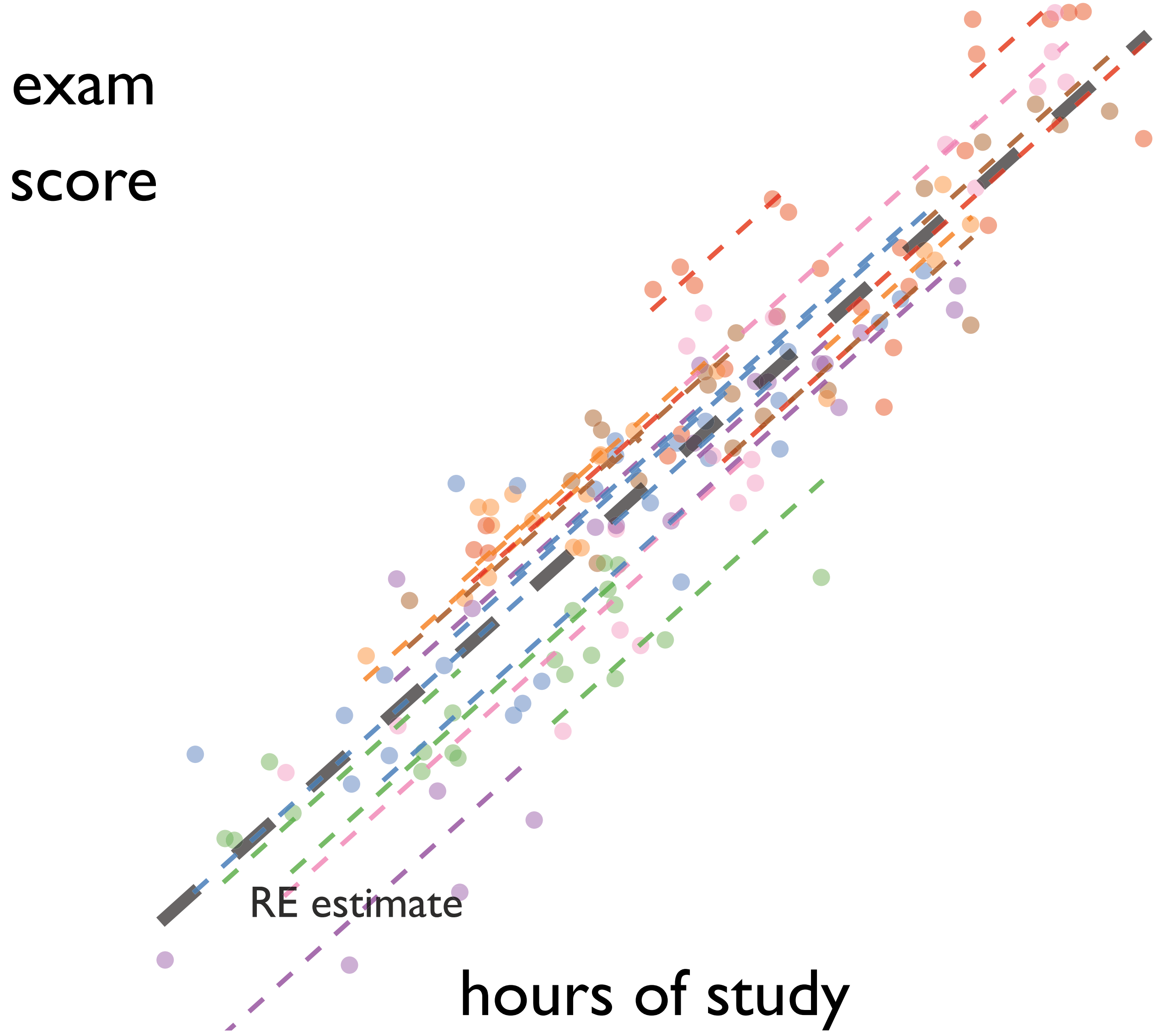
Where'd the bias come from?



With multilevel data, it helps to start at the lowest level

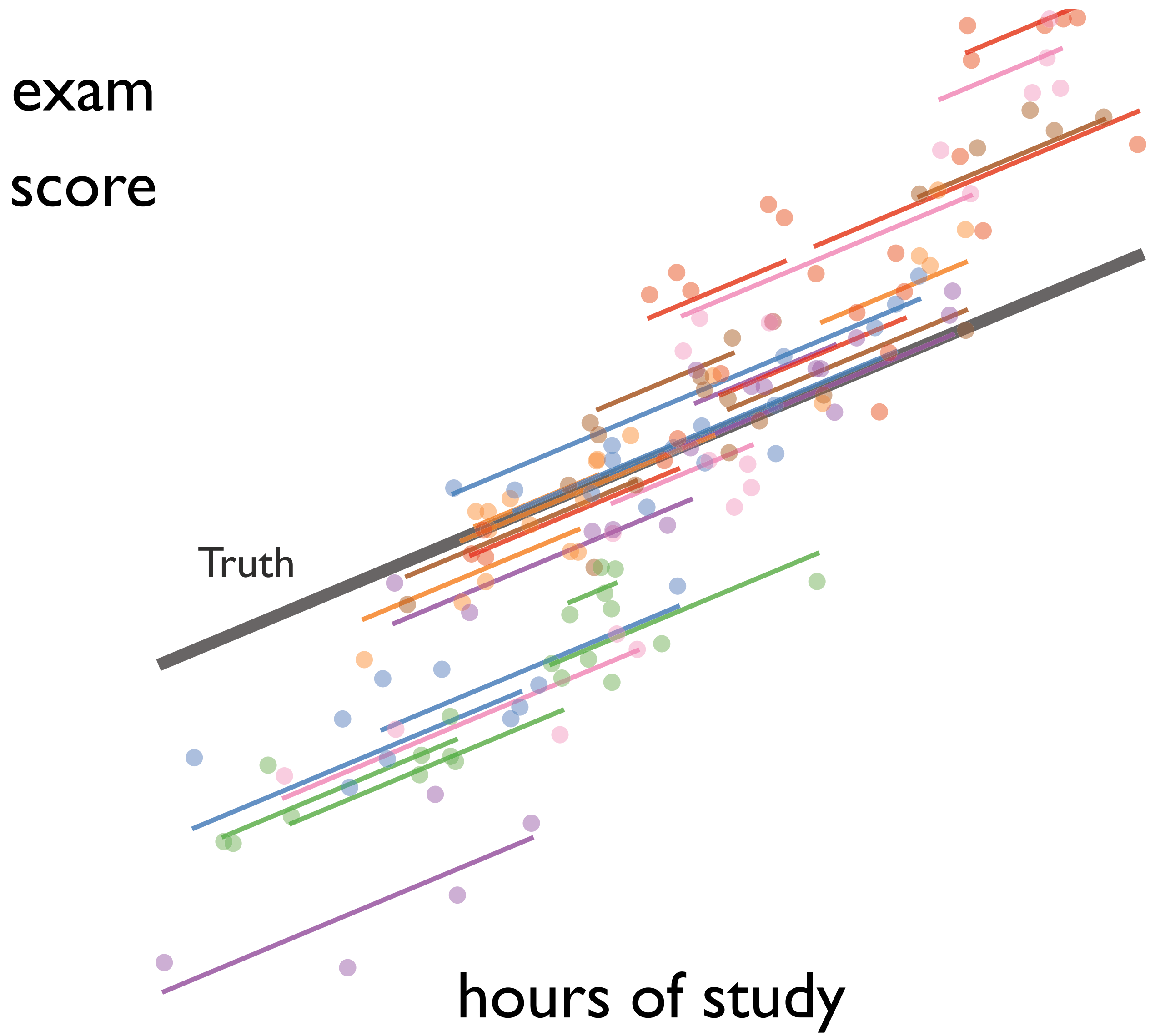
I've colored the points by student

A random effects model finds the student specific intercept *after* estimating the slope of the main regression line



The student specific relationships between effort and scores as estimated by a random effects model

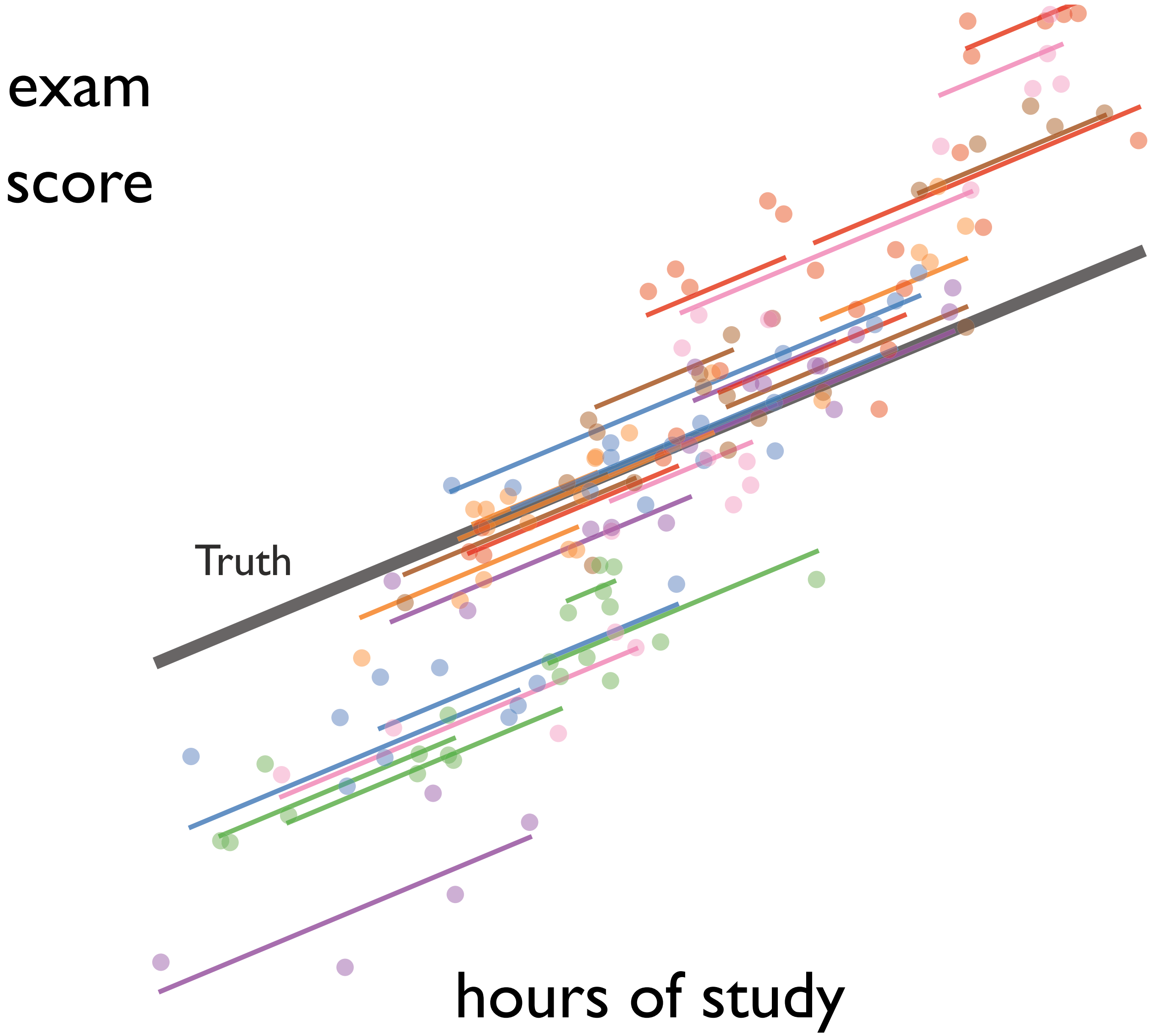
Are these estimates right?



Not even close

The *true* regression lines by student and overall

Random effects estimates of effort is *biased* because the student-specific effect is correlated with effort



Random effects are an inadequate model when the grouping indicator is correlated with our covariates

In this case we have *omitted variable bias*

We need a different model of α:
fixed effects

Fixed effects

$$\alpha_i = \alpha_i^*$$

Easiest to conceptualize in a linear regression framework

Easiest to estimate; just add dummies for each unit, and drop the intercept

Can be correlated with $\mathbf{x}_{it}$:
controls for any group-invariant omitted variable
(ie, controls for *all* group-specific variables, even unmeasurable ones!)

Indeed, that's usually the point.
Often included to capture unobserved variance potentially correlated with $\mathbf{x}$.

Irony: we're actually removing the across-group variation, not modeling it

Instead, we are:

- Assuming the same response in each group to changes in covariates
- Using only within-group variation in covariates to estimate parameters

Fixed effects

$$\alpha_i = \alpha_i^*$$

Fixed effects can't be added to models with perfectly group-invariant covariates
(Causes perfect collinearity of covariates)

Fixed effects specifications also incur an incidental parameters problem:
MLE is consistent as $I \to \infty$, but *not* as $J \to \infty$.

Fixed effects

$$\alpha_i = \alpha_i^*$$

Fixed effects can't be added to models with perfectly group-invariant covariates (Causes perfect collinearity of covariates)

Fixed effects specifications also incur an incidental parameters problem:
MLE is consistent as $I \rightarrow \infty$, but *not* as $J \rightarrow \infty$.

Monte Carlo experiments indicate small sample properties of fixed effects pretty good if $I > 5$ or so.

Fixed effects are common in studies where N is not a random sample, but a (small) universe (e.g., the industrialized countries).

Fixed effects

$$\alpha_i = \alpha_i^*$$

Fixed effects can't be added to models with perfectly group-invariant covariates (Causes perfect collinearity of covariates)

Fixed effects specifications also incur an incidental parameters problem:
MLE is consistent as $I \to \infty$, but *not* as $J \to \infty$.

Monte Carlo experiments indicate small sample properties of fixed effects pretty good if $I > 5$ or so.

Fixed effects are common in studies where N is not a random sample, but a (small) universe (e.g., the industrialized countries).

Sui generis: Fixed effects basically say "France is different because it's France," "America is different because it's America," etc.

Fixed effects example

Another example may help clarify what fixed effects are.

Suppose that we have data following this true model:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \beta_2 z_i + \varepsilon_{ij} \\
\varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

j indexes a set of M_i counties drawn from state i

There are $N = 15$ states total, and we drew $M_j = M = 15$ counties from each state

Fixed effects example

Suppose the data represent county level voting patterns for the US

(i.e., let's illustrate Gelman et al, *Red State, Blue State, Rich State, Poor State* w/ contrived data)

$$\begin{aligned} \text{RVS}_{ij} &= \beta_0 + \beta_1 \text{Income}_{ij} + \beta_2 \text{ConservativeCulture}_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

j indexes a set of M_i counties drawn from state i

Remember: the data I'm using are fake, and contrived to illustrate a concept simply

Gelman et al investigate this in detail with real data and get similar but more nuanced findings

We will review the real data later in this lecture

Fixed effects example: What's the matter with Kansas?

Suppose the data represent county level voting patterns for the US

(I.e., let's illustrate Gelman *et al*, *Red State, Blue State, Rich State, Poor State* using similar but contrived data)

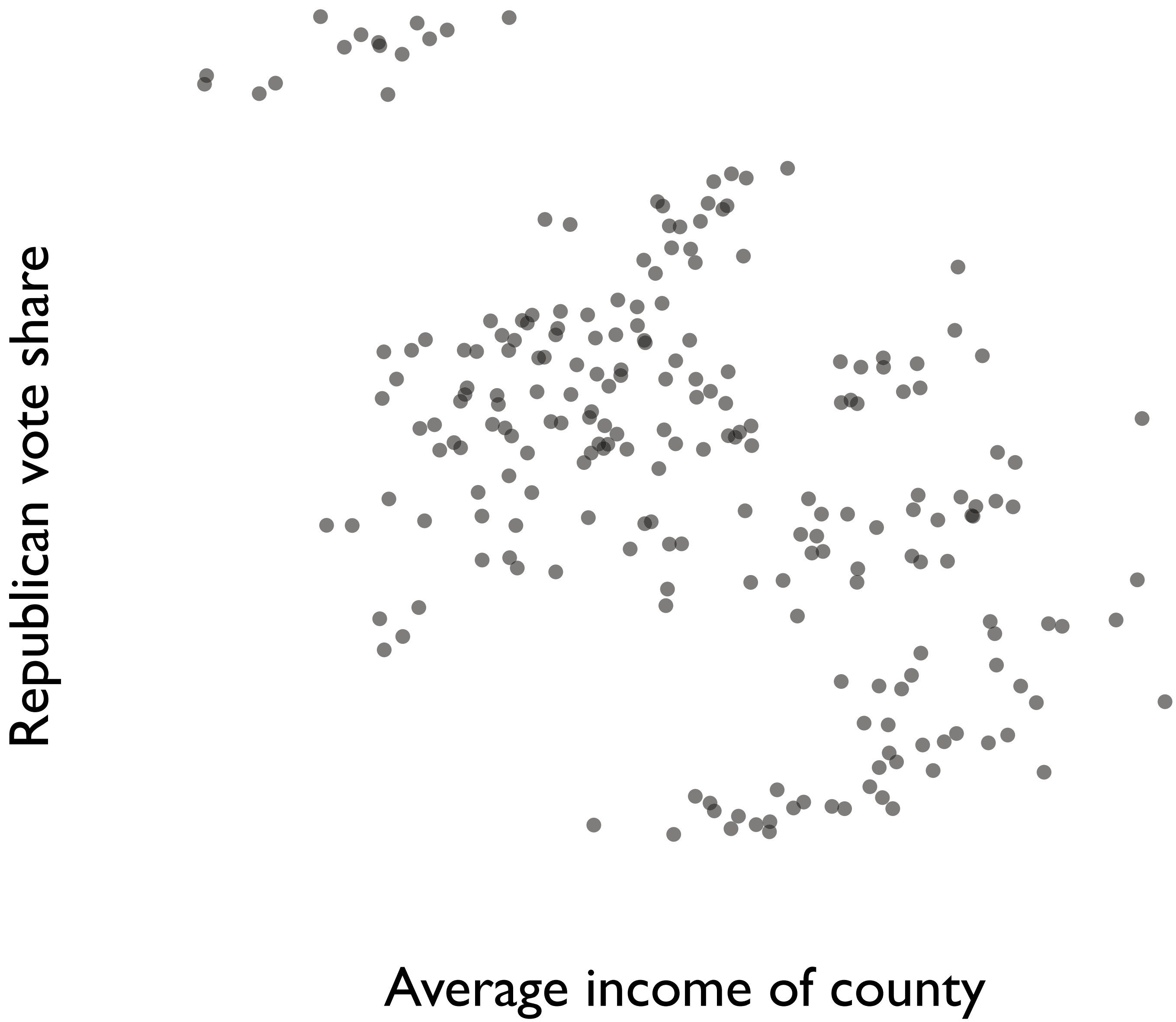
$$\begin{aligned} \text{RVS}_{ij} &= \beta_0 + \beta_1 \text{Income}_{ij} + \beta_2 \text{Conservatism}_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

A problem:
suppose we don't have (or don't trust) a measure of state-level Conservatism

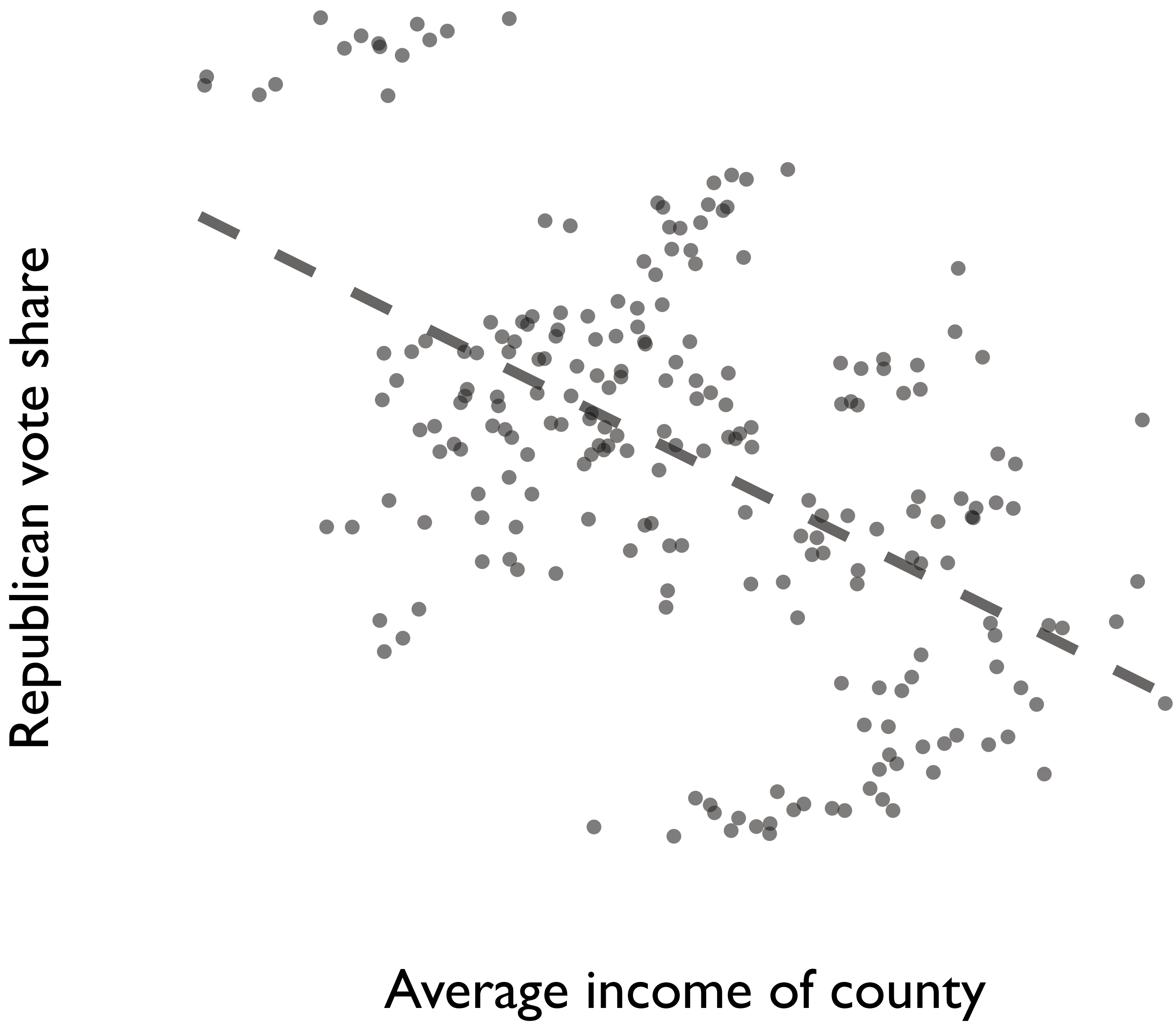
If we exclude it, or mismeasure it, we could get omitted variable bias in $\hat{\beta}_1$

This leads to potentially large misconceptions. . .



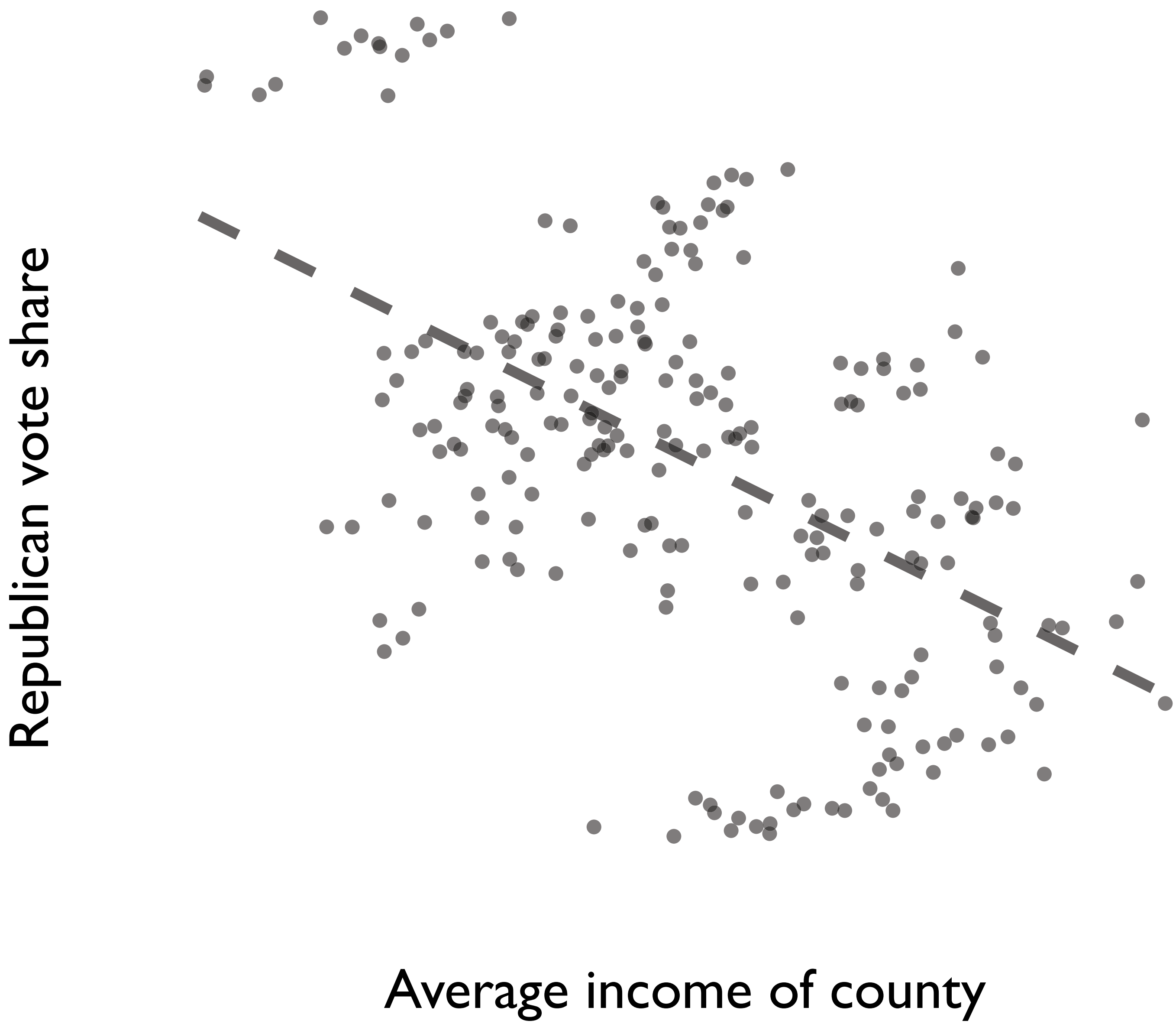
Suppose we observe 15 counties from each of 15 states (225 observations)

Our first cut is to estimate this simple linear regression: $y_{ij} = \beta_0 + \beta_1 \text{Income}_{ij} + \varepsilon_{ij}$



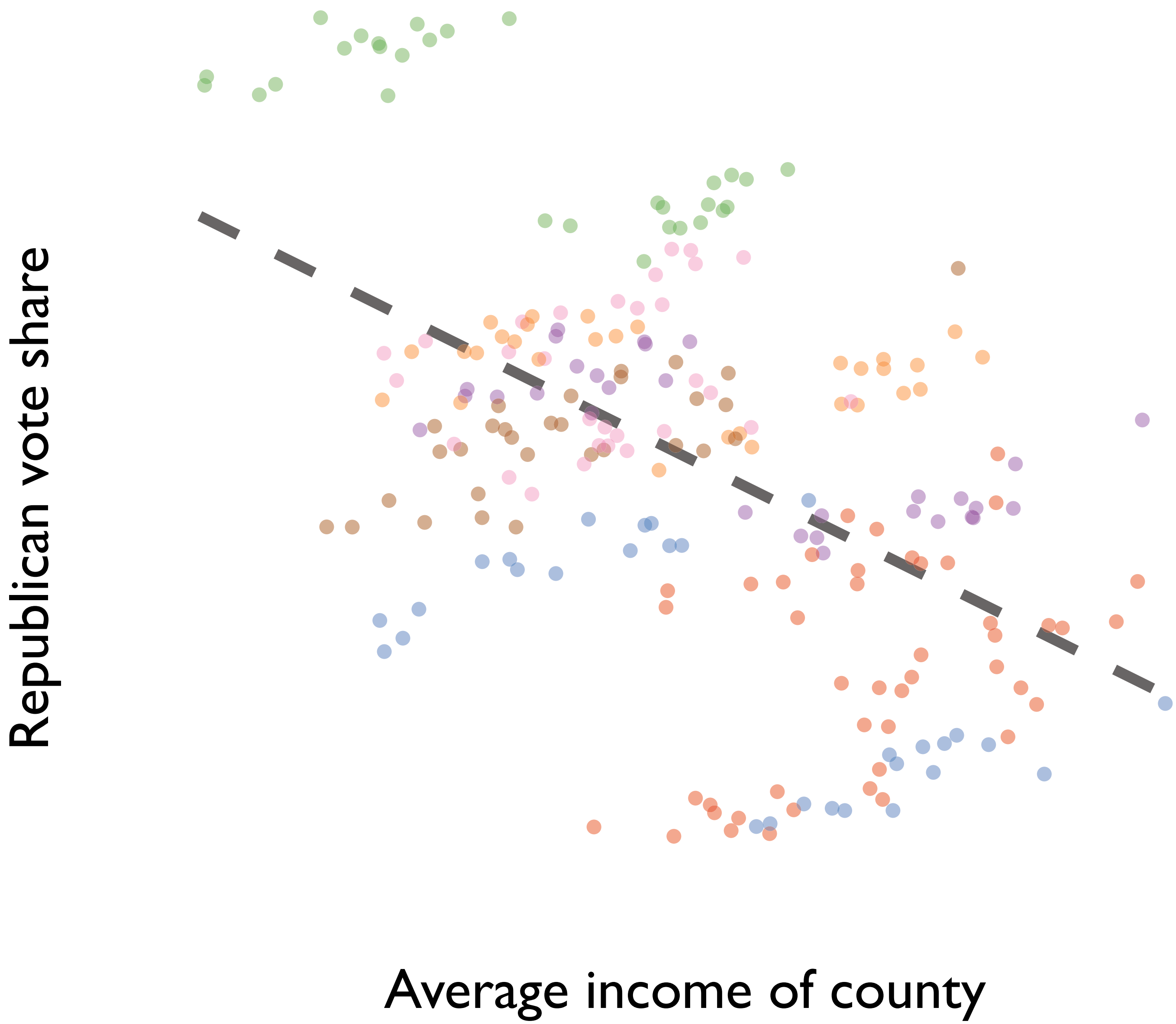
We find that $\hat{\beta}_1$ is negative:

Poor counties seem to vote more Republican than rich counties!



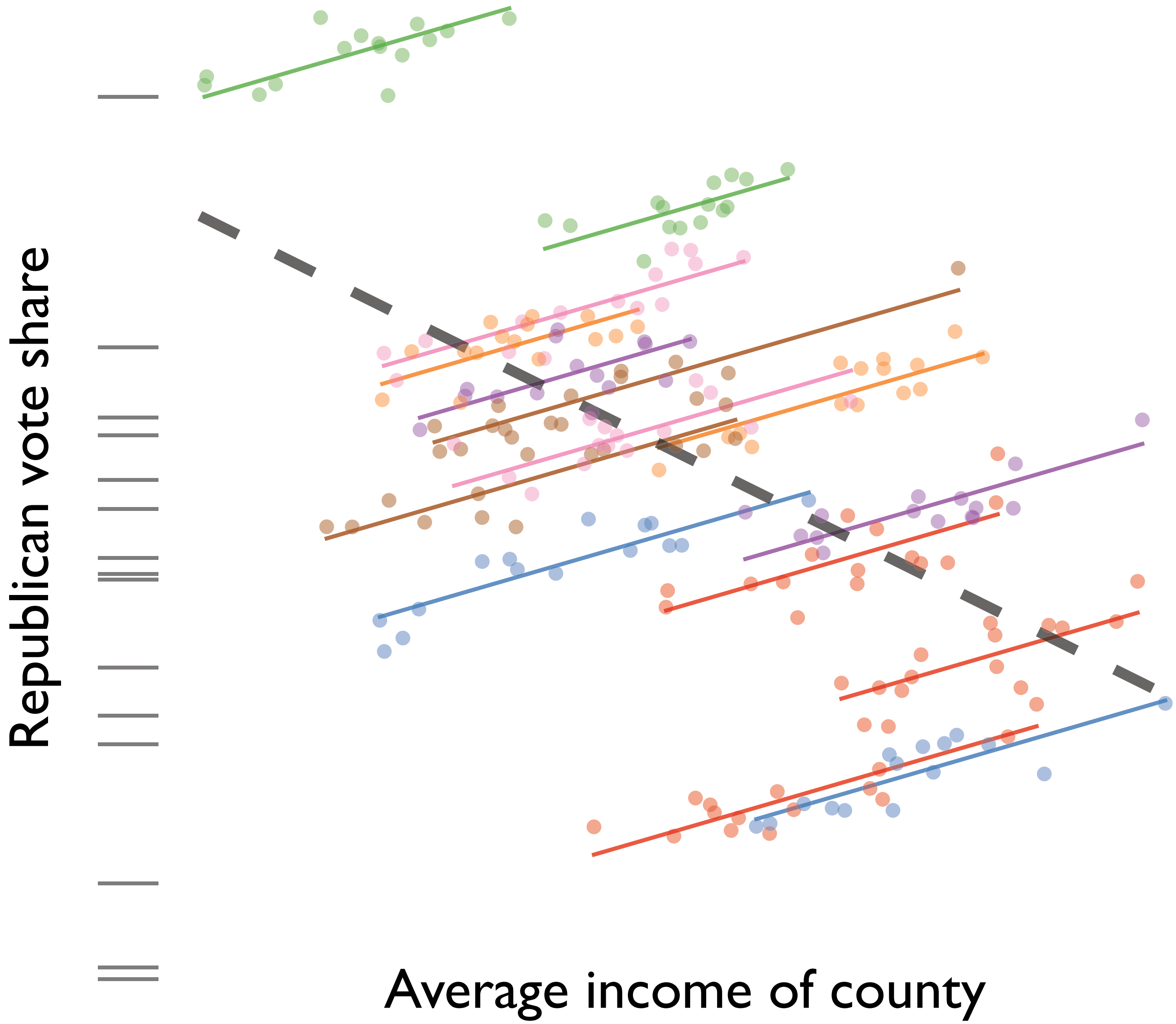
But Republican elected officials attempt to represent the affluent

What's the matter with (poor counties in) Kansas, as Thomas Frank asked?

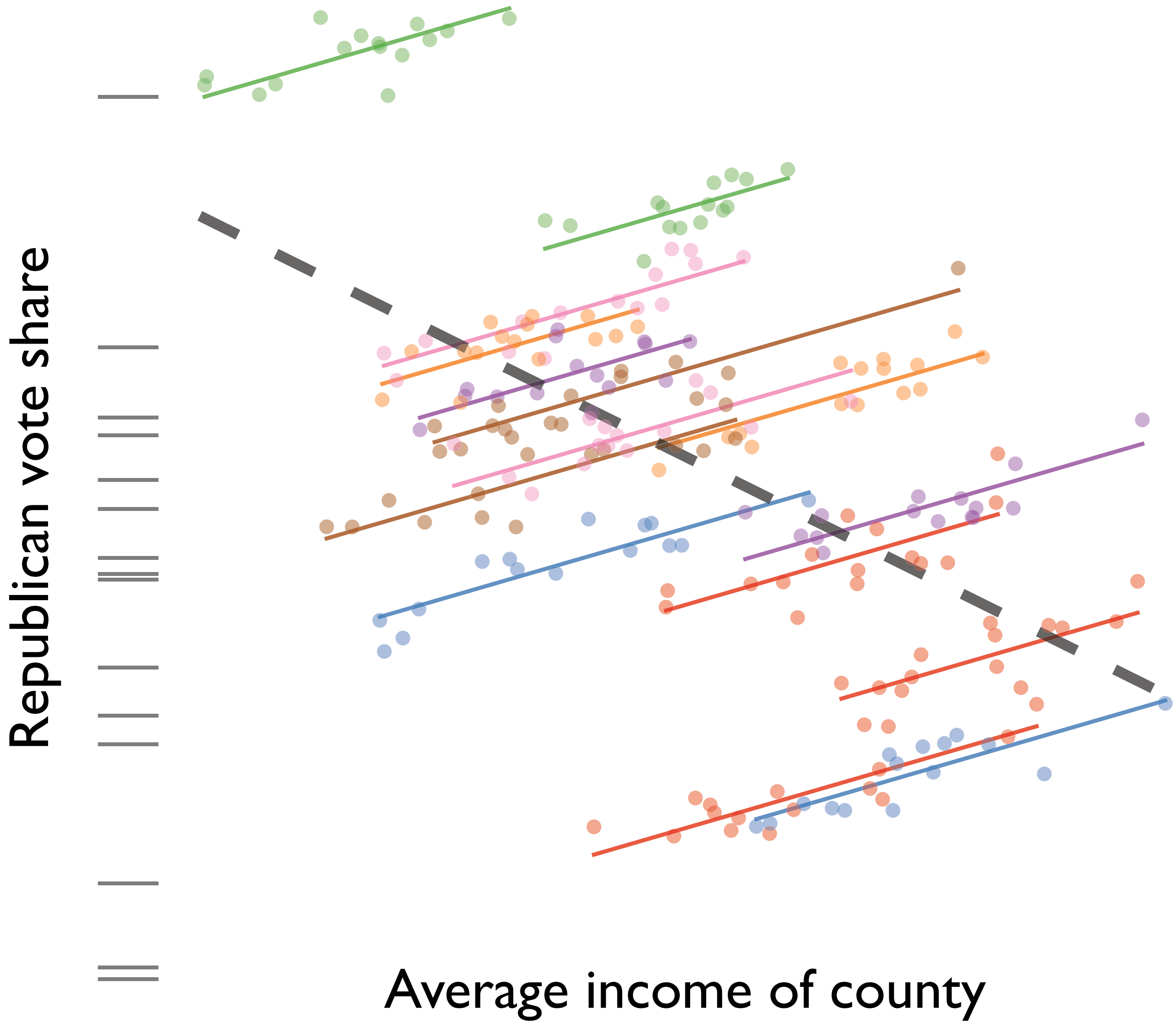


Let’s look at which observations come from which states

Clearly, counties from the same state are clustered

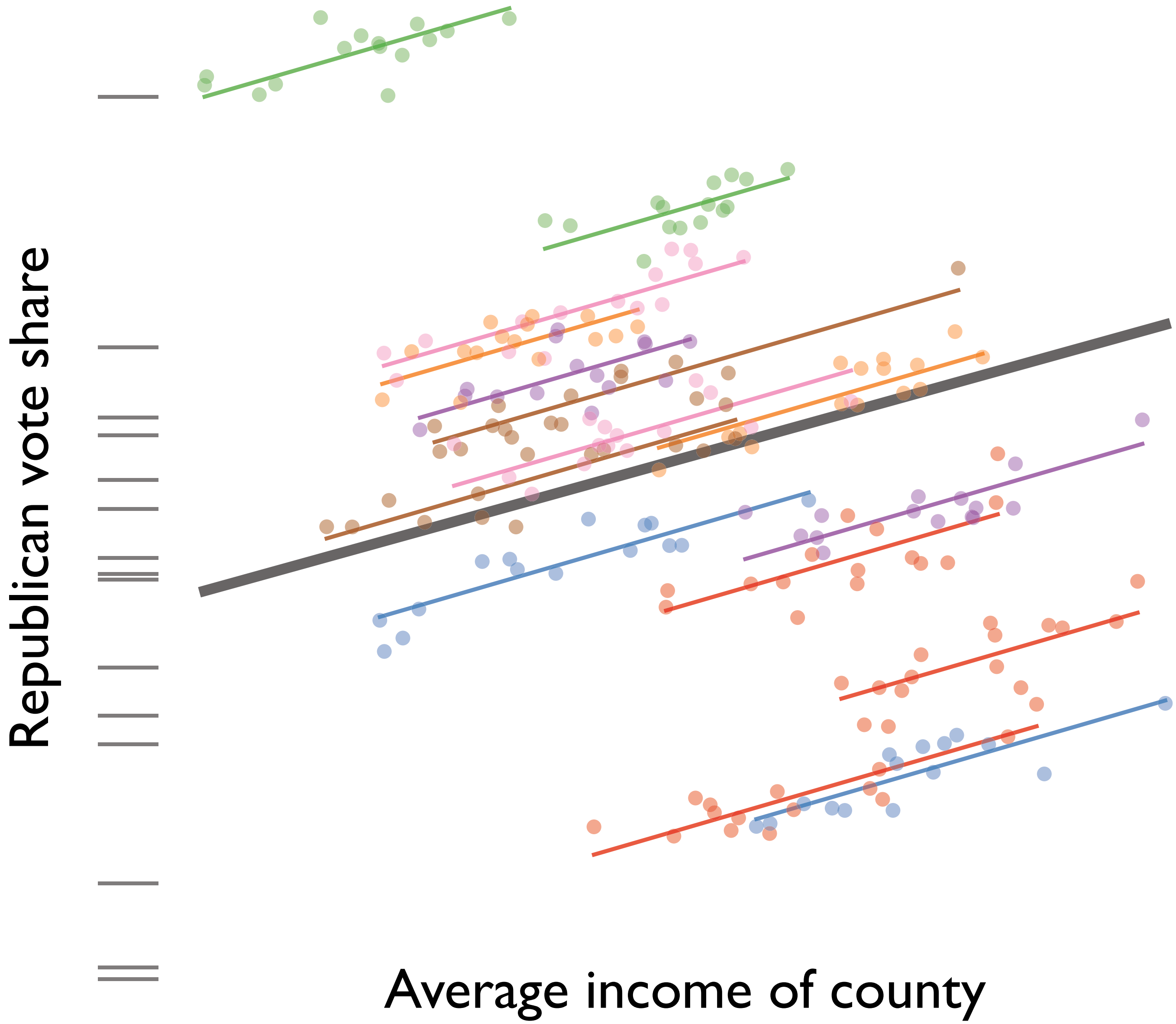


Within each state, there's a *positive* relationship between income & voting Republican



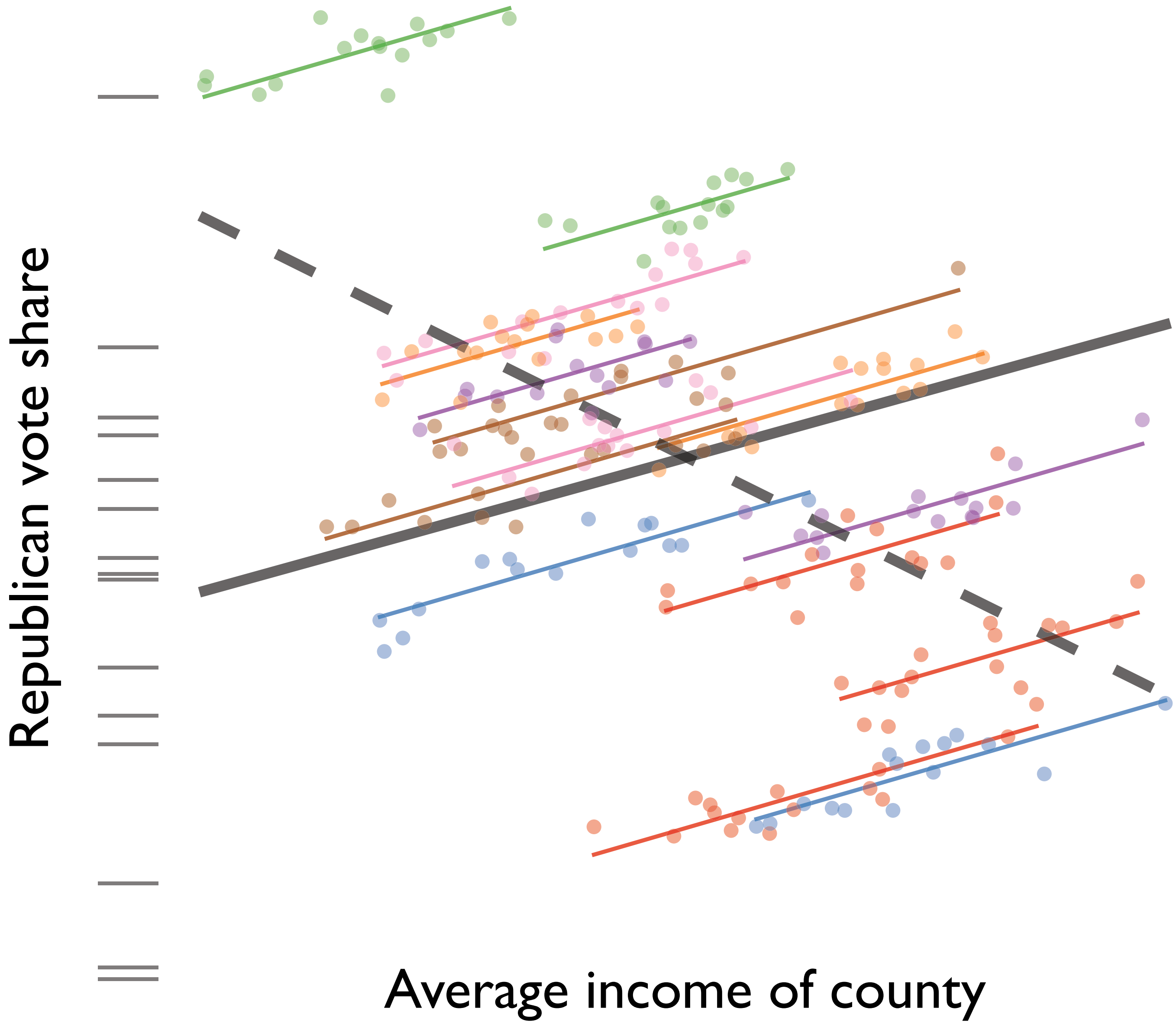
Suggests we need to control for variation at the state level,

either by collecting the state level variables causing the variation, or. . .



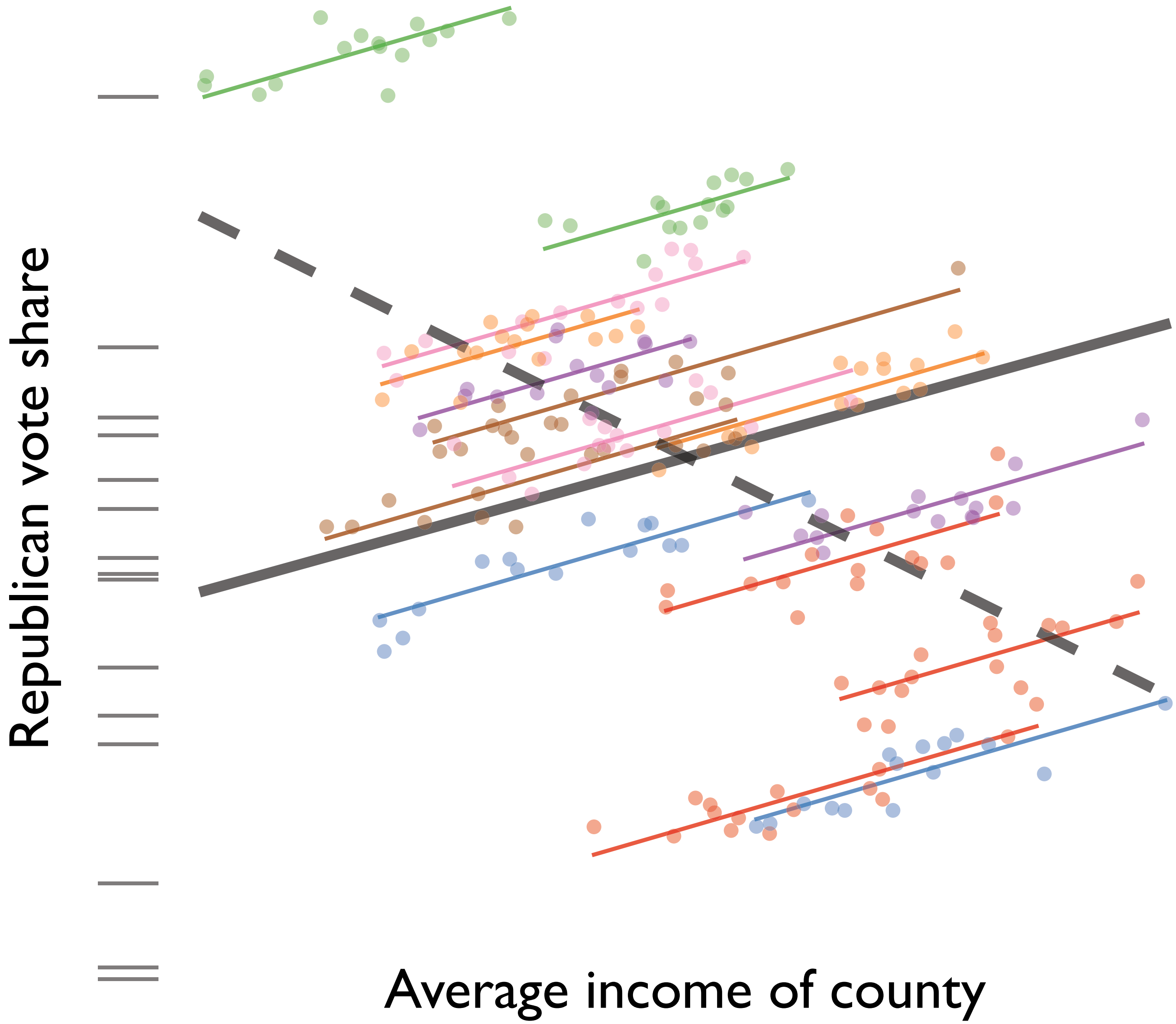
use brute force: add a dummy for each state to the matrix of covariates

to purge the omitted variable bias

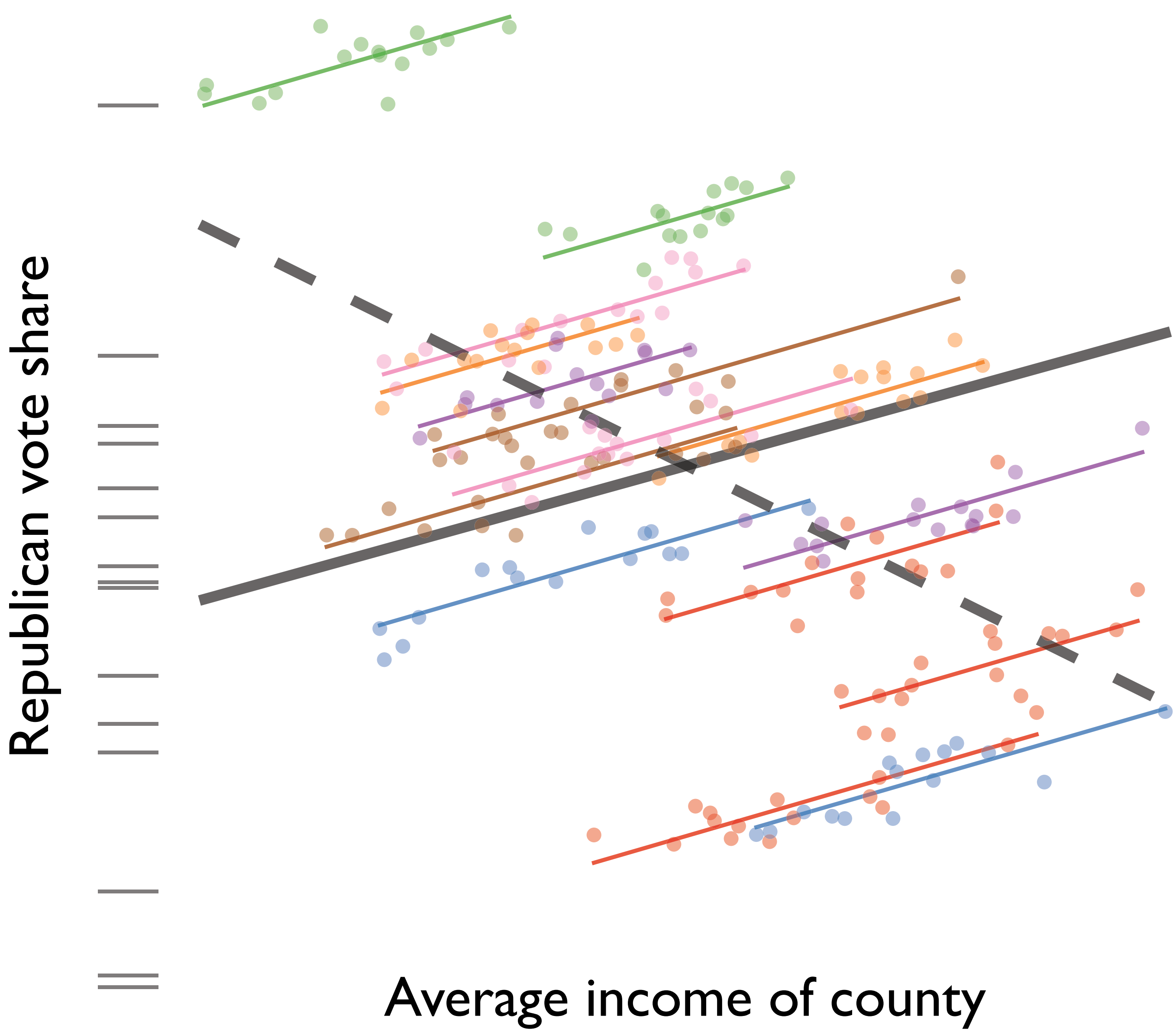


Controlling for state fixed effects, $\hat{\beta}_1$ flips signs!

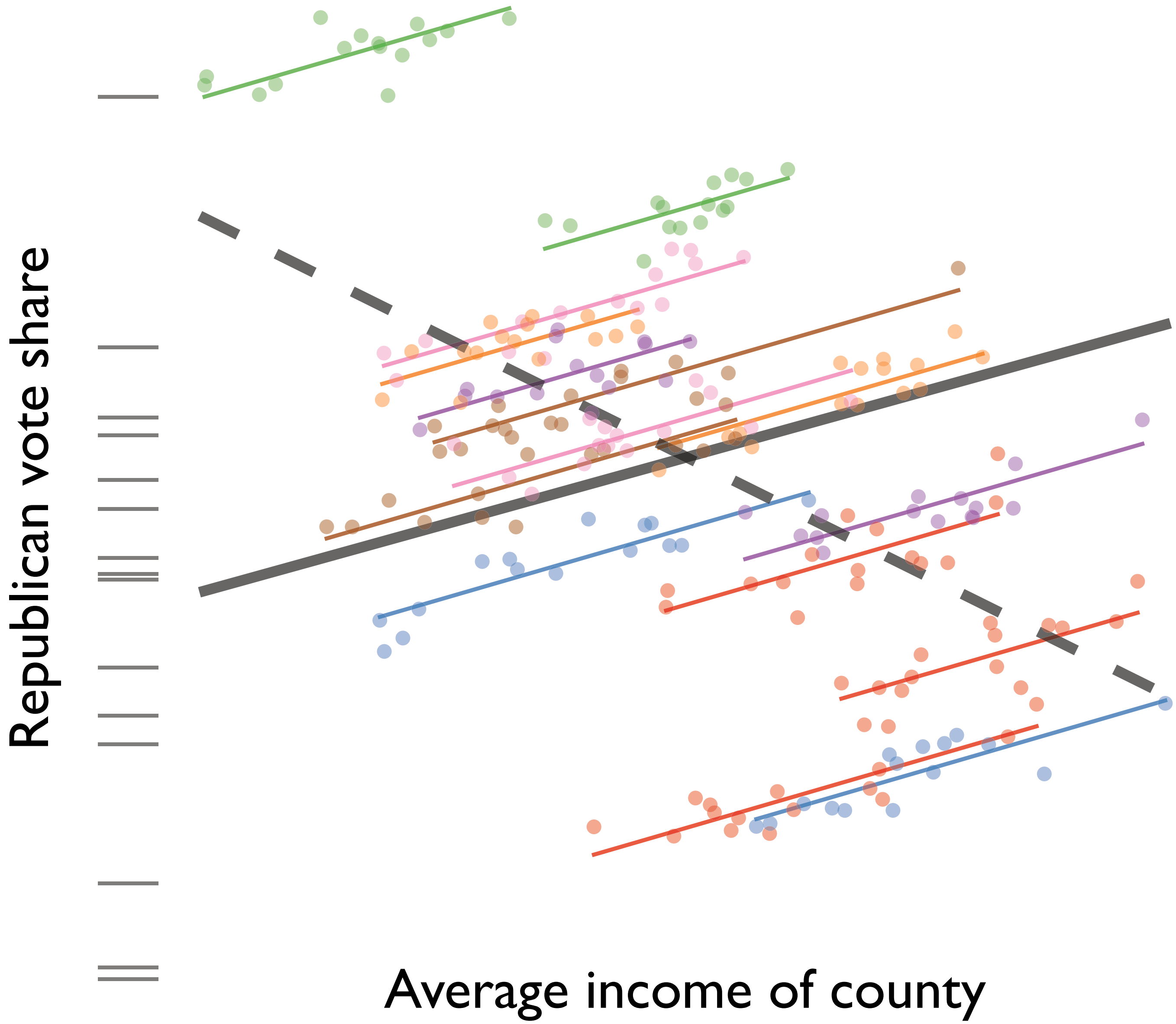
Including fixed effects for each state removes state-level omitted variable bias



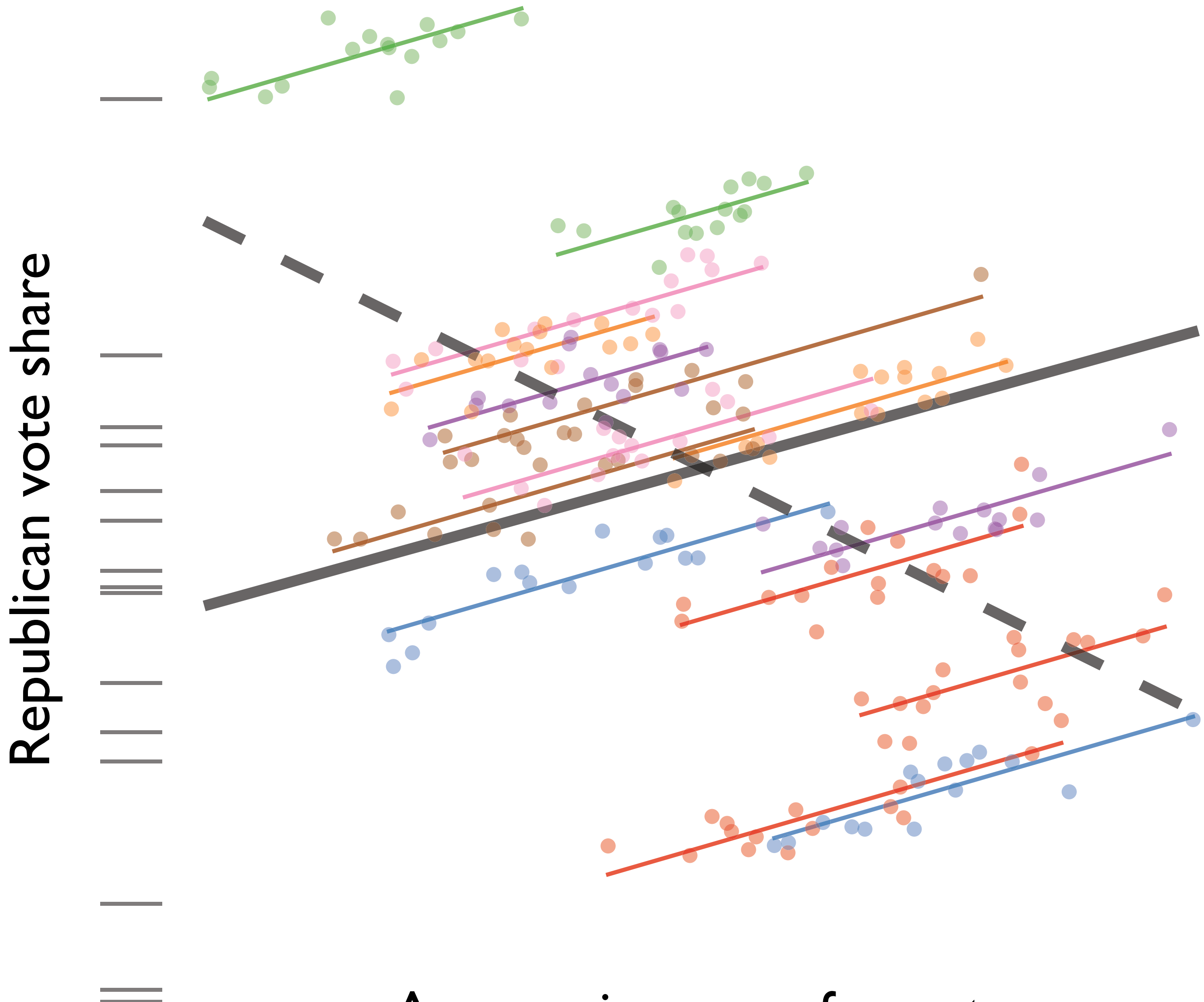
What's the matter with Kansas? On average, Kansans are more conservative than other Americans, but within Kansas, the same divide between rich and poor holds



What's the matter with Kansas? On average, Kansans are more conservative than other Americans, but within Kansas, the same divide between rich and poor holds *. . . or at least it did until 2016*

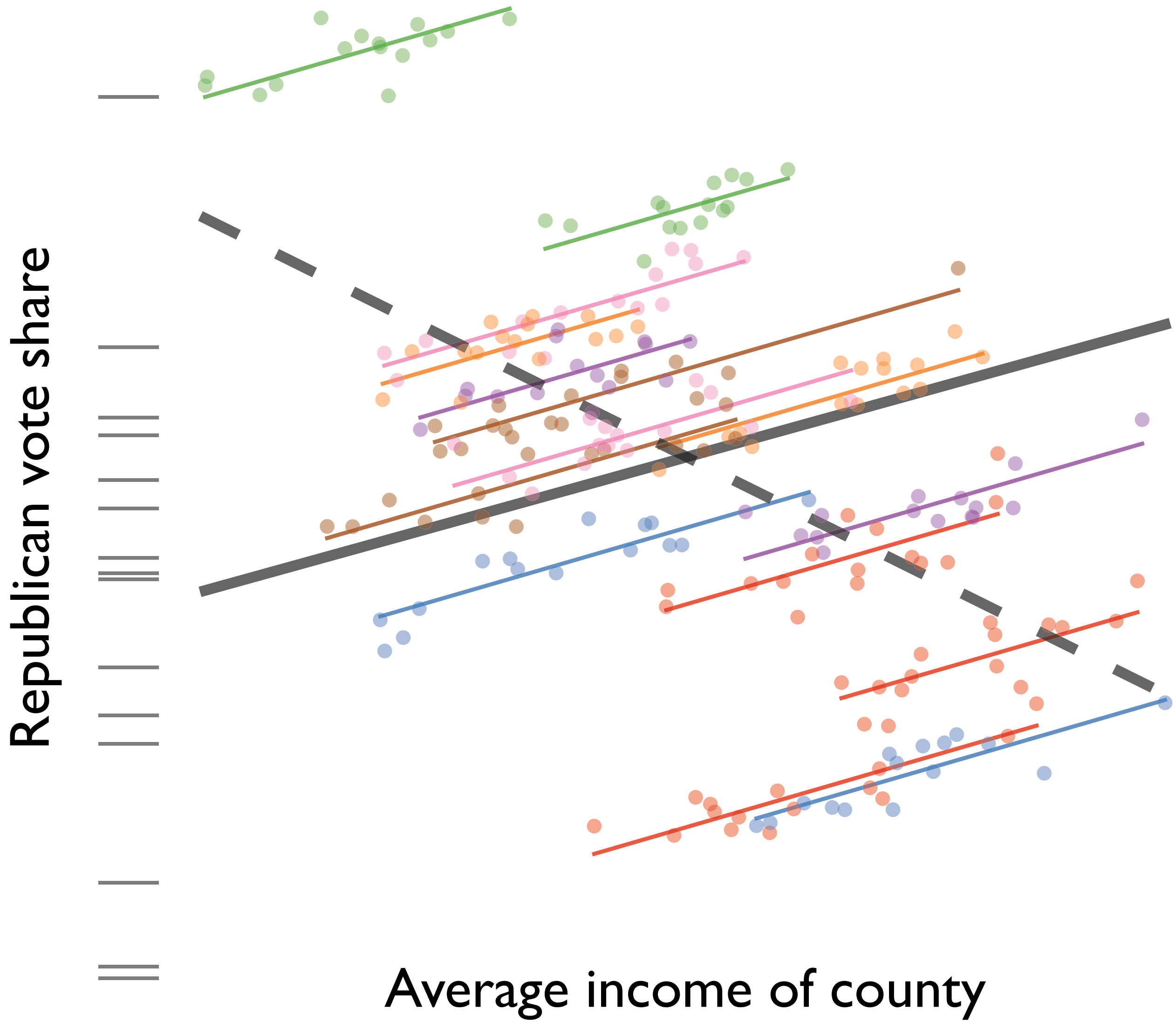


How are fixed effects different from random effects?



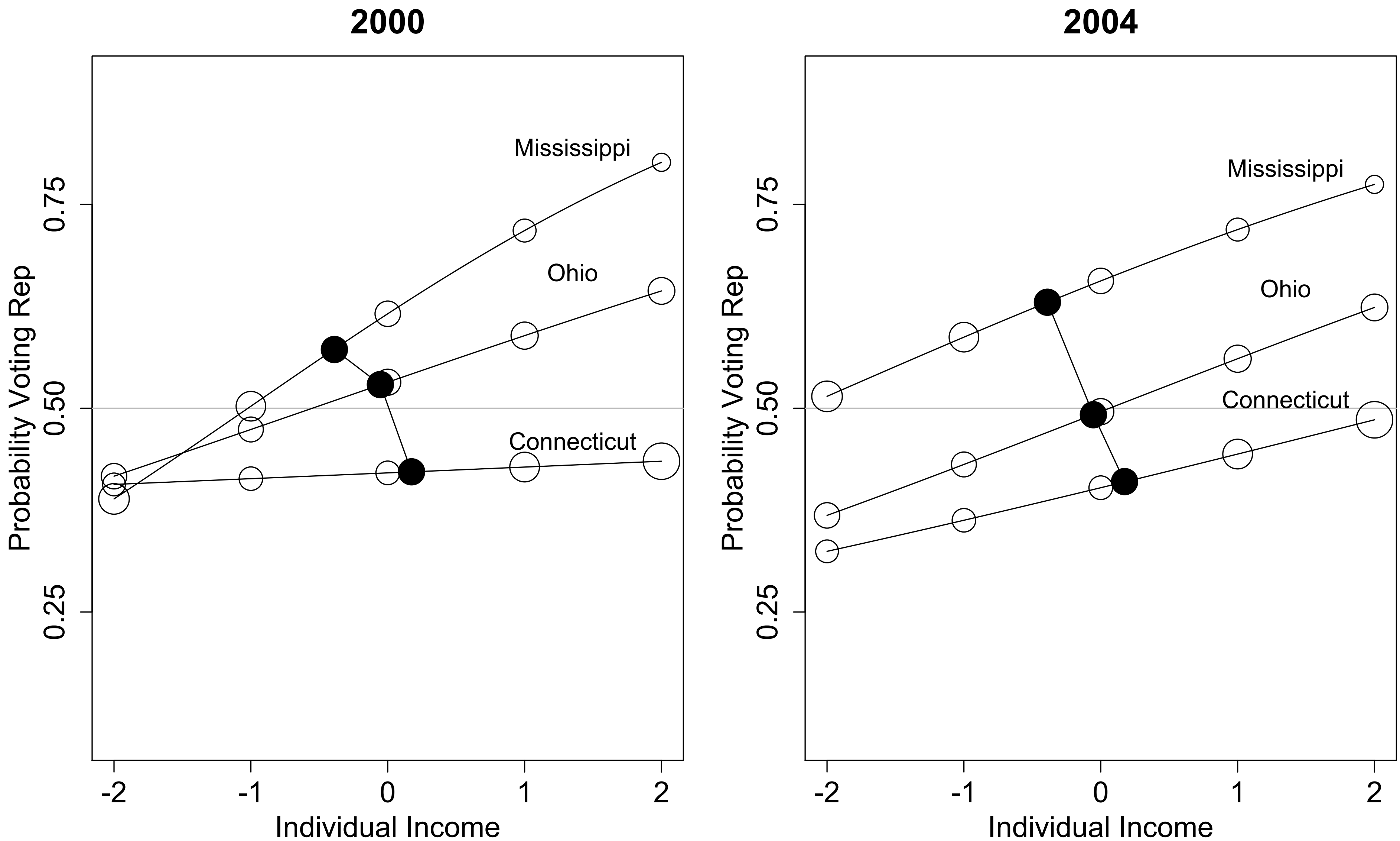
Fixed effects control for omitted variables *random effects don't*

Fixed effects don't follow any particular distribution *random effects do*



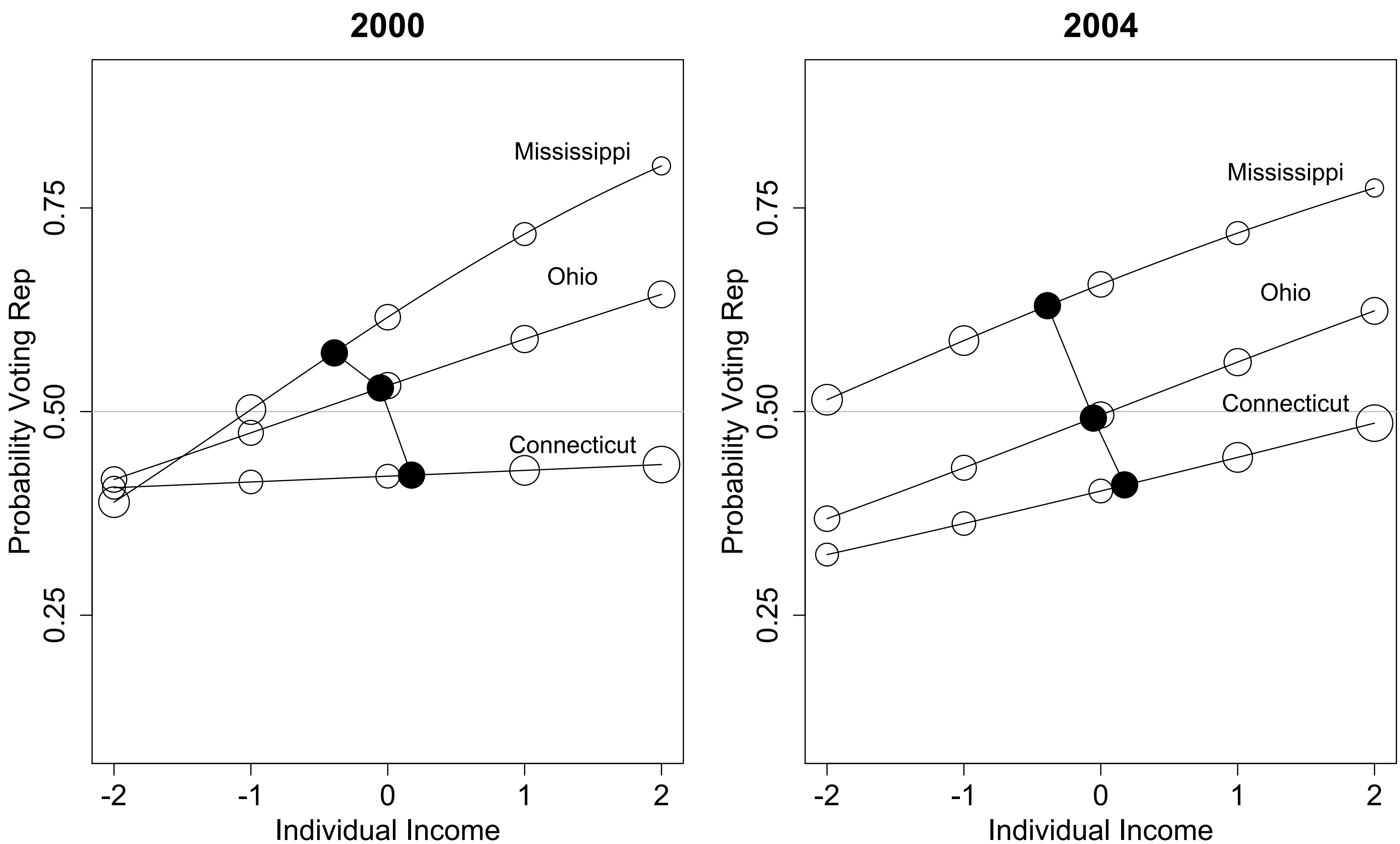
Aside 1 the above reversal is an example of the *ecological fallacy*, which says that aggregate data can mislead us about individual level relationships

Here, the pattern across states mislead us as to the pattern within states



Aside 2: Above are results on actual data from Gelman *et al*

This version of their model assume intercepts (but not slopes) vary by state



Aside 2 When Gelman *et al* allow slopes $\hat{\beta}_{1i}$ to vary across states, they find the rich-poor divide is actually *steeper* in poor states!

Red state, blue state

Running example from Gelman et al 2007, "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?", *QJPS* (also a book)

Gelman et al note a paradox:

1. Surveys show richer voters vote more Republican, ceteris paribus
2. Election data shows richer states vote more Democratic

Seek to disentangle relationship between income and party preference both between and within states

Use National Annenberg Election Study from 2000 and 2004

Red state, blue state

v_i measures whether the ith voter chose the Democrat $(v_i = 1)$ or Republican candidate $(v_i = 0)$

But ith voter of what group?

Start at top level of model, states indexed $1, \ldots j, \ldots J$

Within each state, there are voters $1, \ldots i, \ldots N_j$

Total number of observations: $\sum_{j=1}^{J} N_j$

Each state may have a different number of voters N_j

ML Estimation difficult if some N_j are small—will need Bayesian methods

Full pooling

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha + \beta \text{Income}_{ij}
\end{aligned}
$$

- Full pooling
- Assumes all voters regardless of state follow same model
- Intercept α and coefficient β the same regardless of state
- No special estimation issues; just use logit MLE
- Pooling assumption may be unreasonable; voters may vary by state

Full pooling

Assume we have several variables in memory in R:

`vote`	stacked vector of vote choices for all voters in all states
`income`	stacked vector of income for all voters in all states
`state`	stacked vector state indexes for all voters in all states

R code to estimate the full pooling model:

```
res <- glm(vote ~ income, family=binomial)
```

No pooling

$$
\begin{aligned}
v_{i1} &\sim \text{Bernoulli}(\pi_{i1}) \\
\pi_{i1} &= \text{logit}^{-1}(\mu_{i1}) \\
\mu_{i1} &= \alpha + \beta \text{Income}_{i1} \\
&\ldots \\
v_{iJ} &\sim \text{Bernoulli}(\pi_{iJ}) \\
\pi_{iJ} &= \text{logit}^{-1}(\mu_{iJ}) \\
\mu_{iJ} &= \alpha + \beta \text{Income}_{iJ}
\end{aligned}
$$

- No pooling
- Estimate J separate logits

No pooling

$$
\begin{aligned}
v_{i1} &\sim \text{Bernoulli}(\pi_{i1}) \\
\pi_{i1} &= \text{logit}^{-1}(\mu_{i1}) \\
\mu_{i1} &= \alpha + \beta \text{Income}_{i1} \\
&\dots \\
v_{iJ} &\sim \text{Bernoulli}(\pi_{iJ}) \\
\pi_{iJ} &= \text{logit}^{-1}(\mu_{iJ}) \\
\mu_{iJ} &= \alpha + \beta \text{Income}_{iJ}
\end{aligned}
$$

- Assumes each state is *sui generis*
- No common distribution governs voters in different states

No pooling

$$
\begin{aligned}
v_{i1} &\sim \text{Bernoulli}(\pi_{i1}) \\
\pi_{i1} &= \text{logit}^{-1}(\mu_{i1}) \\
\mu_{i1} &= \alpha + \beta \text{Income}_{i1} \\
&\ldots \\
v_{iJ} &\sim \text{Bernoulli}(\pi_{iJ}) \\
\pi_{iJ} &= \text{logit}^{-1}(\mu_{iJ}) \\
\mu_{iJ} &= \alpha + \beta \text{Income}_{iJ}
\end{aligned}
$$

- No special estimation issues; just use logit MLE repeatedly

May not be able to estimate all J equations if some N_j are small

No pooling

$$
\begin{aligned}
v_{i1} &\sim \text{Bernoulli}(\pi_{i1}) \\
\pi_{i1} &= \text{logit}^{-1}(\mu_{i1}) \\
\mu_{i1} &= \alpha + \beta \text{Income}_{i1} \\
&\dots \\
v_{iJ} &\sim \text{Bernoulli}(\pi_{iJ}) \\
\pi_{iJ} &= \text{logit}^{-1}(\mu_{iJ}) \\
\mu_{iJ} &= \alpha + \beta \text{Income}_{iJ}
\end{aligned}
$$

- No pooling assumption may be very inefficient if relationship between income and vote is somehow similar across states

No pooling

$$
\begin{aligned}
v_{i1} &\sim \text{Bernoulli}(\pi_{i1}) \\
\pi_{i1} &= \text{logit}^{-1}(\mu_{i1}) \\
\mu_{i1} &= \alpha + \beta \text{Income}_{i1} \\
& \ldots \\
v_{iJ} &\sim \text{Bernoulli}(\pi_{iJ}) \\
\pi_{iJ} &= \text{logit}^{-1}(\mu_{iJ}) \\
\mu_{iJ} &= \alpha + \beta \text{Income}_{iJ}
\end{aligned}
$$

R code to estimate the equation-by-equation "model":

```
res <- vector("list",J)
for (j in 1:J)
  res[j] <- glm(vote[state==j] ~ income[state==j], family=binomial)
```

Fixed effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij}
\end{aligned}
$$

- Blend of full pooling for β and no pooling for α
- Note that α now indexed by j: J parameters to estimate
- Assumes different states have different baseline tendencies to vote Democratic, but the relationship between voting and income is identical everywhere
- Fixed effects: Each state has a unique, *sui generis* intercept

Fixed effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij}
\end{aligned}
$$

- Purges omitted variable bias at the state level
- Can estimate using MLE (aside on consistency)

R code to estimate the fixed effects model:

```
stateFE <- makeFEdummies(state)    # in simcf package
res <- glm(vote ~ income + stateFE - 1, family=binomial)
```

Random effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij} \\
\alpha_j &\sim \text{Normal}(\mu_\alpha, \sigma^2_\alpha)
\end{aligned}
$$

- *Partial* pooling of intercepts, full pooling of coefficients
- Still assumes different states have different baseline tendencies to vote Democratic, but the relationship between voting and income is identical everywhere
- But α_j now includes a random effect.

Random effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij} \\
\alpha_j &\sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)
\end{aligned}
$$

- Could rewrite the systematic component to reflect this:

$$\mu_{ij} = \underbrace{\mu_\alpha + \varepsilon_{\alpha,j}}_{\text{RandomIntercept}} + \beta \text{Income}_{ij}$$

- Estimation by MLE if N_j's not too small; otherwise Bayesian MCMC

Random effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij} \\
\alpha_j &\sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)
\end{aligned}
$$

Advantages of partial pooling (also known as "random effects" or "mixed effects")

- Far less demanding than full fixed effects (estimating J α_j's)
- Still allows (and estimates) degree of heterogeneity
- Can decompose multiple levels of random effects to discern which levels driving the most variance in voting

Random effects regression

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta \text{Income}_{ij} \\
\alpha_j &\sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)
\end{aligned}
$$

R code to estimate the random intercept model:

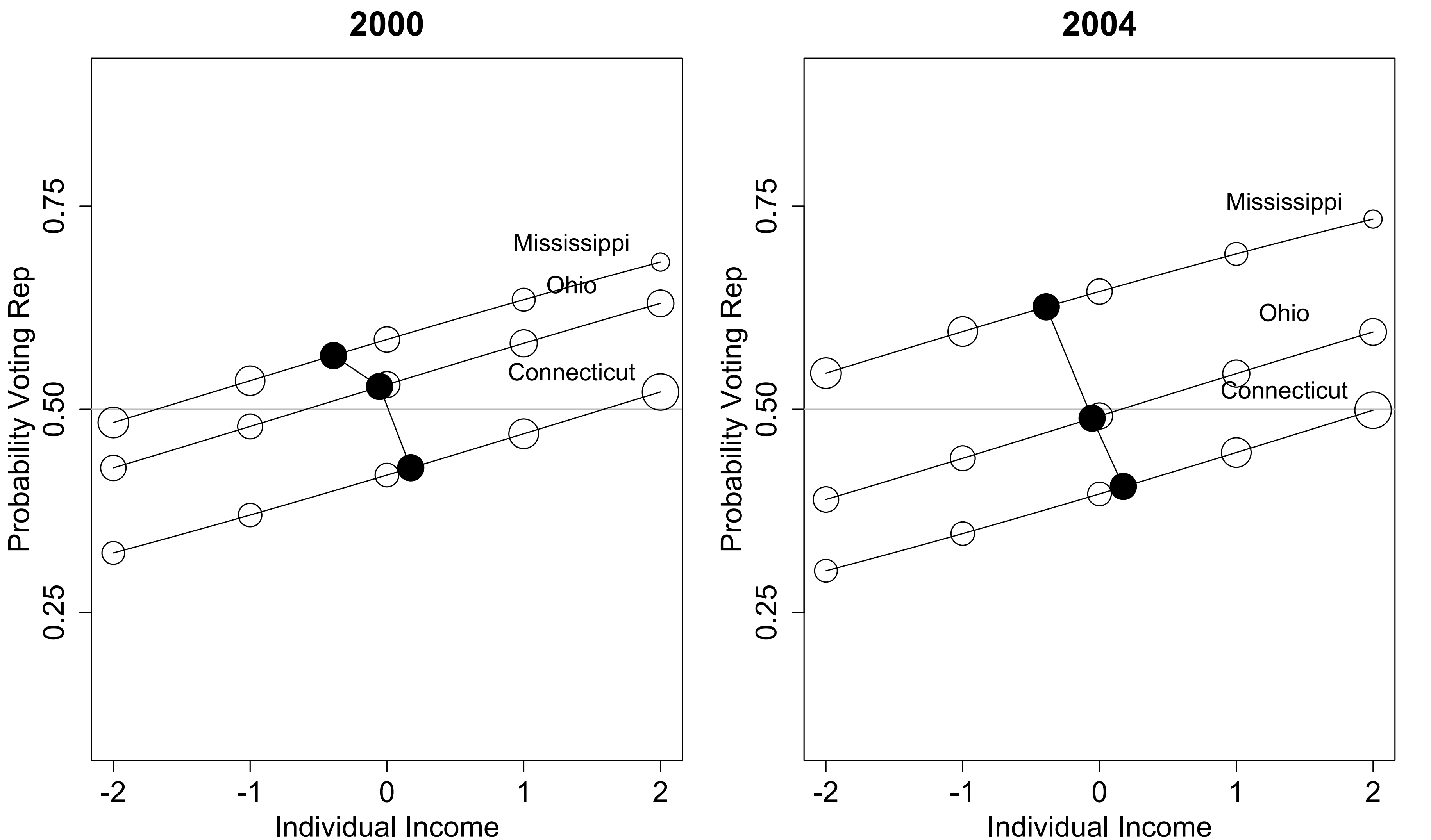
```
library(lme4)
res <- lmer(vote ~ income + (1|state), family=binomial)
```

Note the odd formula notation

In a formula, 1 represents a constant.

This says the model constant is conditioned on (randomly varies by) the grouping variable state

Random effects



Random interecept, random coefficient

$$\begin{aligned} v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\ \mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij} \end{aligned}$$

Random interecept, random coefficient

$$\begin{aligned} v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\ \pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\ \mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij} \\ \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} &\sim \text{Multivariate Normal}\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \sigma^2_\alpha & \rho\sigma_{\alpha,\beta} \\ \rho\sigma_{\alpha,\beta} & \sigma^2_\beta \end{bmatrix} \right) \end{aligned}$$

- We can make coefficients (in a linear model, slopes) random too
- Each group has a similar but randomly varying relation between votes & income

Random interecept, random coefficient

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij} \\
\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} &\sim \text{Multivariate Normal}\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \rho\sigma_{\alpha,\beta} \\ \rho\sigma_{\alpha,\beta} & \sigma_\beta^2 \end{bmatrix} \right)
\end{aligned}
$$

R code to estimate the random intercept, random coefficients model:

```
res <- lmer(vote ~ income + (1 + income|state), family=binomial)
```

Note the odd formula notation

In a formula, 1 represents a constant.

This says the model constant and coefficient for income are both conditioned on (randomly vary by) the grouping variable state

Fully conditional mixed effects

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij}
\end{aligned}
$$

Fully conditional mixed effects

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij} \\
\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} &\sim \text{Multivariate Normal}\left(\begin{bmatrix} \mu_{\alpha,j} \\ \mu_{\beta,j} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \rho\sigma_{\alpha,\beta} \\ \rho\sigma_{\alpha,\beta} & \sigma_\beta^2 \end{bmatrix} \right)
\end{aligned}
$$

Fully conditional mixed effects

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \alpha_j + \beta_j \text{Income}_{ij} \\
\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} &\sim \text{Multivariate Normal}\left(\begin{bmatrix} \mu_{\alpha,j} \\ \mu_{\beta,j} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \rho\sigma_{\alpha,\beta} \\ \rho\sigma_{\alpha,\beta} & \sigma_\beta^2 \end{bmatrix} \right) \\
\mu_{\alpha,j} &= \gamma_{\alpha,0} + \gamma_{\alpha,1}\text{AvgInc}_j \\
\mu_{\beta,j} &= \gamma_{\beta,0} + \gamma_{\beta,1}\text{AvgInc}_j
\end{aligned}
$$

Intercepts and/or coefficients could themselves covary with higher level covariates

Here, level two variables, like the average income of the state, help determine determine level one relationships

Fully conditional mixed effects

Note that we could rewrite α_j and β_j as a function of a systematic component and an error term:

$$\begin{aligned}
\alpha_j &= \gamma_{\alpha,0} + \gamma_{\alpha,1}\text{AvgInc}_j + \varepsilon_{\alpha,j} \\
\beta_j &= \gamma_{\beta,0} + \gamma_{\beta,1}\text{AvgInc}_j + \varepsilon_{\beta,j} \\
\begin{bmatrix} \varepsilon_{\alpha,j} \\ \varepsilon_{\beta,j} \end{bmatrix} &\sim \text{Multivariate Normal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_\alpha & \rho\sigma_{\alpha,\beta} \\ \rho\sigma_{\alpha,\beta} & \sigma^2_\beta \end{bmatrix} \right)
\end{aligned}$$

Fully conditional mixed effects

Substituting for α_j and β_j, we can rewrite this model as an interactive specification with random intercepts and coefficients:

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \left(\gamma_{\alpha,0} + \gamma_{\alpha,1}\text{AvgInc}_j + \varepsilon_{\alpha,j}\right) + \left(\gamma_{\beta,0} + \gamma_{\beta,1}\text{AvgInc}_j + \varepsilon_{\beta,j}\right)\text{Income}_{ij}
\end{aligned}
$$

Fully conditional mixed effects

Substituting for α_j and β_j, we can rewrite this model as an interactive specification with random intercepts and coefficients:

$$
\begin{aligned}
v_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\
\pi_{ij} &= \text{logit}^{-1}(\mu_{ij}) \\
\mu_{ij} &= \left(\gamma_{\alpha,0} + \gamma_{\alpha,1}\text{AvgInc}_j + \varepsilon_{\alpha,j}\right) + \left(\gamma_{\beta,0} + \gamma_{\beta,1}\text{AvgInc}_j + \varepsilon_{\beta,j}\right)\text{Income}_{ij} \\
&= (\gamma_{\alpha,0} + \varepsilon_{\alpha,j}) + (\gamma_{\beta,0} + \varepsilon_{\beta,j})\text{Income}_{ij} \\
&\quad + \gamma_{\alpha,1}\text{AvgInc}_j + \gamma_{\beta,1}\text{Income}_{ij} \times \text{AvgInc}_j
\end{aligned}
$$

Fully conditional mixed effects

R code to estimate the conditional intercept, conditional coefficients model:

```
# Make the second level variable
avg.income <- rep(NA,J)
for (j in 1:J)
  avg.income[j] <- mean(income[state==j])

# Create an interaction term at the first level
avg.income.expanded <- avg.income[state]

# Use interactions to mimic conditional coefficients
res <- lmer(vote ~ avg.income.expanded*income
                + (1 + income|state), family=binomial)
```

Fully conditional mixed effects