GOV 2001/ 1002/ E-200 Section 7 Zero-Inflated models and Intro to Multilevel Modeling[1]

Anton Strezhnev

Harvard University

March 23, 2016

[1]These section notes are heavily indebted to past Gov 2001 TFs for slides and R code.

Logistics

Reading Assignment- Becker and Kennedy (1992), Harris and Zhao (2007) (sections 1 and 2), and Esarey and Pierce (2012)

Re-replication- Due by 6pm Wednesday, March 30th on Canvas.

Re-replication

Re-replication- Due **March 30** at 6pm.

- You will receive all of the replication files from another team.
- Re-replication teams will be posted on Canvas. We'll download everyone's files and put them on a Dropbox to access. Download the paper that you have been assigned to.
- Go through the replication code and try to improve it in any way you can.
- Provide a short write-up of thoughts on the replication and ideas for their final paper.

Re-replication

Re-replication- Due **March 30** at 6pm.

- You will receive all of the replication files from another team.
- Re-replication teams will be posted on Canvas. We'll download everyone's files and put them on a Dropbox to access. Download the paper that you have been assigned to.
- Go through the replication code and try to improve it in any way you can.
- Provide a short write-up of thoughts on the replication and ideas for their final paper.
- Aim to be helpful, not just critical!

OVERVIEW

- In this section you will...

OVERVIEW

- In this section you will...
  - learn how to extend count models to account for DGPs with lots of zeroes

Overview

- In this section you will...
  - learn how to extend count models to account for DGPs with lots of zeroes
  - learn how to incorporate group-level effects in GLMs

Overview

- In this section you will...
  - learn how to extend count models to account for DGPs with lots of zeroes
  - learn how to incorporate group-level effects in GLMs

OUTLINE

Zero-inflated models

"Fixed" and "Random" Effects Models

What is Zero-Inflation?

Consider our DGP for a count model.

WHAT IS ZERO-INFLATION?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero?

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:
  1. Some data are spoiled or lost

WHAT IS ZERO-INFLATION?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:
  1. Some data are spoiled or lost
  2. Survey respondents put "zero" to an answer on a survey just to get it done.

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:
  1. Some data are spoiled or lost
  2. Survey respondents put "zero" to an answer on a survey just to get it done.
  3. Units never get a chance to accumulate events

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:
  1. Some data are spoiled or lost
  2. Survey respondents put "zero" to an answer on a survey just to get it done.
  3. Units never get a chance to accumulate events e.g. counts of conflict deaths during peacetime

What is Zero-Inflation?

Consider our DGP for a count model.

- What if we knew that something in our data did not fit the data-generating process?
- For example, what if we thought that some of our data were systematically zero rather than randomly zero? This could be when:
  1. Some data are spoiled or lost
  2. Survey respondents put "zero" to an answer on a survey just to get it done.
  3. Units never get a chance to accumulate events e.g. counts of conflict deaths during peacetime

If we have an "excess" of zeroes, our modeling assumption will be wrong.

A Working Example: Fishing

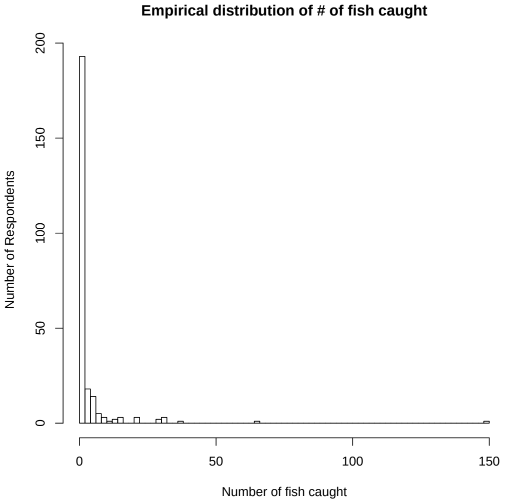
You're trying to figure out how many fish people caught in a lake from a survey. People were asked:

- How many children were in the group
- How many people were in the group
- How many fish they caught

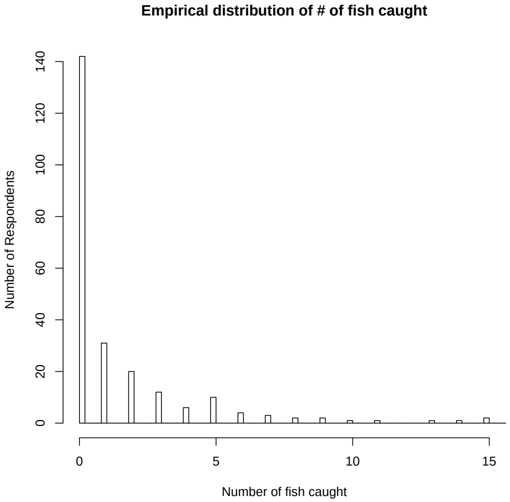
A Working Example: Fishing

The problem is, some people didn't even fish! These people have systematically zero fish.

A Working Example: Fishing



A Working Example: Fishing



Zero-Inflated Poisson Model

Solution: Model a multi-stage data generating process

Zero-Inflated Poisson Model

Solution: Model a multi-stage data generating process

We're going to assume that whether or not the person fished Q_i is the outcome of a Bernoulli trial.

Zero-Inflated Poisson Model

Solution: Model a multi-stage data generating process

We're going to assume that whether or not the person fished Q_i is the outcome of a Bernoulli trial.

$$Q_i \sim \text{Bernoulli}(\psi_i)$$

ψ_i is the probability that you do not fish.

Zero-Inflated Poisson Model

Solution: Model a multi-stage data generating process

We're going to assume that whether or not the person fished Q_i is the outcome of a Bernoulli trial.

$$Q_i \sim \text{Bernoulli}(\psi_i)$$

ψ_i is the probability that you do not fish.

This is a **mixture model** because our data is a mix of these two types of groups each with their own data generation process.

ZERO-INFLATED POISSON MODEL

Given that you fished, the Poisson model is what we have done before:

ZERO-INFLATED POISSON MODEL

Given that you fished, the Poisson model is what we have done before:

1. $Y_i | Q_i = 1 \sim \text{Poisson}(\lambda_i)$

Zero-Inflated Poisson Model

Given that you fished, the Poisson model is what we have done before:

1. $Y_i | Q_i = 1 \sim \text{Poisson}(\lambda_i)$
2. $\lambda_i = e^{X_i \beta}$

Zero-Inflated Poisson Model

Given that you fished, the Poisson model is what we have done before:

1. $Y_i | Q_i = 1 \sim \text{Poisson}(\lambda_i)$
2. $\lambda_i = e^{X_i \beta}$
3. $Y_i | Q_i = 1$ and $Y_j | Q_j = 1$ are independent for all $i \neq j$.

Zero-Inflated Poisson Model

Given that you fished, the Poisson model is what we have done before:

1. $Y_i|Q_i = 1 \sim \text{Poisson}(\lambda_i)$
2. $\lambda_i = e^{X_i\beta}$
3. $Y_i|Q_i = 1$ and $Y_j|Q_j = 1$ are independent for all $i \neq j$.

Zero-Inflated Poisson Model

Given that you *did not* fish, what is the model?

Zero-Inflated Poisson Model

Given that you *did not* fish, what is the model?

It's just a constant 0!

$$Y_i|(Q_i = 0) = 0$$

Zero-Inflated Poisson Model

Given that you *did not* fish, what is the model?

It's just a constant 0!

$$Y_i|(Q_i = 0) = 0$$

and the probability that Y is 0 given $Z_i = 0$:

$$P(Y_i = 0|Q_i = 0) = 1$$

Zero-Inflated Poisson Model

So we can write Y_i as a mixture of two DGPs

$$Y_i \sim \begin{cases} 0 & \text{with probability} \quad \psi_i \\ \text{Poisson}(\lambda_i) & \text{with probability} \quad 1 - \psi_i \end{cases}$$

Zero-Inflated Poisson Model

So we can write Y_i as a mixture of two DGPs

$$Y_i \sim \begin{cases} 0 & \text{with probability} \quad \psi_i \\ \text{Poisson}(\lambda_i) & \text{with probability} \quad 1 - \psi_i \end{cases}$$

And we can put covariates Z_i on ψ_i using a logit model:

$$\psi_i = \frac{e^{z_i \gamma}}{1 + e^{z_i \gamma}}$$

Zero-Inflated Poisson: Deriving the Likelihood

The likelihood function is proportional to the probability of Y_i:

Zero-Inflated Poisson: Deriving the Likelihood

The likelihood function is proportional to the probability of Y_i:

$$Pr(Y_i|X_i, Z_i, \gamma, \beta) = \left[\psi_i + (1-\psi_i)e^{-\lambda_i}\right] \mathcal{I}(Y_i = 0) \times \\ \left[(1-\psi_i)\frac{\lambda_i^{Y_i} e^{-\lambda_i}}{Y_i!}\right] \mathcal{I}(Y_i > 0)$$

Zero-Inflated Poisson: Deriving the Likelihood

The likelihood function is proportional to the probability of Y_i:

$$Pr(Y_i|X_i, Z_i, \gamma, \beta) = \left[\psi_i + (1-\psi_i)e^{-\lambda_i}\right] \mathcal{I}(Y_i = 0) \times \left[(1-\psi_i)\frac{\lambda_i^{Y_i} e^{-\lambda_i}}{Y_i!}\right] \mathcal{I}(Y_i > 0)$$

$$Pr(Y_i|X_i, Z_i, \gamma, \beta) = \left[\frac{e^{Z_i\gamma}}{1+e^{Z_i\gamma}} + \frac{1}{1+e^{Z_i\gamma}} e^{-e^{X_i\beta}}\right] \mathcal{I}(Y_i = 0) \times \left[\frac{1}{1+e^{Z_i\gamma}} \frac{e^{Y_i X_i \beta} e^{-e^{X_i\beta}}}{Y_i!}\right] \mathcal{I}(Y_i > 0)$$

Zero-Inflated Poisson: Deriving the Likelihood

Multiplying over all observations and taking the log, we get:

$$\ell(Y|X,Z,\gamma,\beta) = \sum_{Y_i=0} \log\left[\frac{e^{Z_i\gamma} + e^{-e^{X_i\beta}}}{1+e^{Z_i\gamma}}\right] + \sum_{Y_i>0} \log\left[\frac{e^{Y_iX_i\beta}e^{-e^{X_i\beta}}}{Y_i!(1+e^{Z_i\gamma})}\right]$$

Zero-Inflated Poisson: Deriving the Likelihood

Multiplying over all observations and taking the log, we get:

$$\ell(Y|X,Z,\gamma,\beta) = \sum_{Y_i=0} \log\left[\frac{e^{Z_i\gamma} + e^{-e^{X_i\beta}}}{1+e^{Z_i\gamma}}\right] + \sum_{Y_i>0} \log\left[\frac{e^{Y_iX_i\beta}e^{-e^{X_i\beta}}}{Y_i!(1+e^{Z_i\gamma})}\right]$$

$$\ell(Y|X,Z,\gamma,\beta) = \sum_{Y_i=0}\left[\log(e^{Z_i\gamma} + e^{-e^{X_i\beta}}) - \log(1+e^{Z_i\gamma})\right] + \sum_{Y_i>0}\left[Y_iX_i\beta - e^{X_i\beta} - \log(Y_i!) - \log(1+e^{Z_i\gamma})\right]$$

Zero-Inflated Poisson: Deriving the Likelihood

Combining terms and dropping additive constants:

Zero-Inflated Poisson: Deriving the Likelihood

Combining terms and dropping additive constants:

$$\ell(Y|X,Z,\gamma,\beta) = \sum_{Y_i=0}\left[\log(e^{Z_i\gamma} + e^{-e^{X_i\beta}})\right] +$$

$$\sum_{Y_i>0}\left[Y_i X_i\beta - e^{X_i\beta}\right] - \sum_{i=1}^{n}\log(1+e^{Z_i\gamma})$$

Zero-Inflated Poisson: Deriving the Likelihood

Combining terms and dropping additive constants:

$$\ell(Y|X,Z,\gamma,\beta) = \sum_{Y_i=0}\left[\log(e^{Z_i\gamma} + e^{-e^{X_i\beta}})\right] +$$

$$\sum_{Y_i>0}\left[Y_iX_i\beta - e^{X_i\beta}\right] - \sum_{i=1}^{n}\log(1+e^{Z_i\gamma})$$

How many parameters are we estimating?

LET'S PROGRAM THIS IN R

Load and get the data ready:

```
fish <- read.table("http://www.ats.ucla.edu/stat/data/fish.csv",
  sep=",", header=T)
X <- fish[c("child", "persons")]
Z <- fish[c("persons")]
X <- as.matrix(cbind(1,X))
Z <- as.matrix(cbind(1,Z))
Y <- fish$count
```

LET'S PROGRAM THIS IN R

Write out the Log-likelihood function

```
ll.zipoisson <- function(par, X, Z, Y){
  beta <- par[1:ncol(X)]
  gamma <- par[(ncol(X)+1):length(par)]

  ## Which indices are Y = 0?
  yzero <- Y==0
  ynonzero <- Y > 0

  ## First part of likelihood
  lik <- -sum(log(1 + exp(Z%*%gamma))) + # Sum over all N
    sum(log(exp(Z[yzero,]%*%gamma) + exp(-exp(X[yzero,]%*%beta)))) + #Y = 0
    sum(Y[ynonzero]*X[ynonzero,]%*%beta - exp(X[ynonzero,]%*%beta)) #Y > 0

  return(lik)

}
```

LET'S PROGRAM THIS IN R

Optimize to get the results

```
par <- rep(1,(ncol(X)+ncol(Z)))
out <- optim(par, ll.zipoisson, Z=Z, X=X,Y=Y, method="BFGS",
             control=list(fnscale=-1), hessian=TRUE)

out$par
[1] -0.3551572 -1.3403588  0.8614703  0.5308458 -0.2966534
```

Plotting to See the Relationship

These numbers don't mean a lot to us, so we can plot the predicted probabilities of a group having not fished (i.e. predict ψ.

PLOTTING TO SEE THE RELATIONSHIP

These numbers don't mean a lot to us, so we can plot the predicted probabilities of a group having not fished (i.e. predict ψ.

First, we have to simulate our gammas:

```
varcv.par <- solve(-out$hessian)
library(mvtnorm)
sim.pars <- rmvnorm(10000, out$par, varcv.par)
# Subset to only the parameters we need (gammas)
# Better to simulate all though
sim.z <- sim.pars[,(ncol(X)+1):length(par)]
```

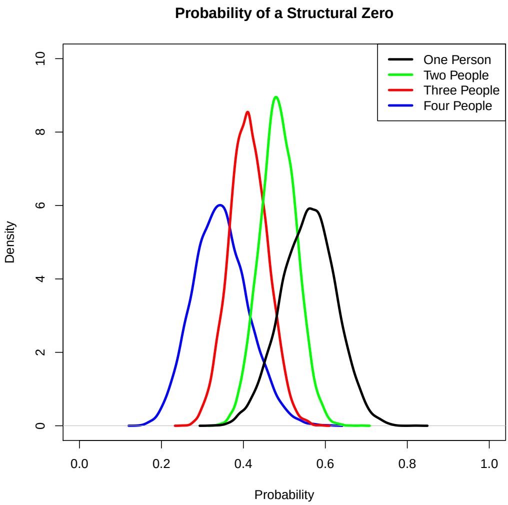
Plotting to See the Relationship

We then generate predicted probabilities of not fishing for different sized groups.

```
# Calculate the predicted probabilities for four different group sizes
person.vec <- seq(1,4)
# Each row is a different vector of covariates (our setx)
Zcovariates <- cbind(1, person.vec)
# Calculate predicted probability (inverse logit) for each set of covariates
# Remember that for each group size, this is a vector of 10000 predicted
     probabilities
exp.holder <- matrix(NA, ncol=4, nrow=10000)
for(i in 1:length(person.vec)){
  exp.holder[,i] <- exp(Zcovariates[i,]%*%t(sim.z))/(1+exp(Zcovariates[i,]%*%t(sim.z)
       ))
}
```

Plotting to See the Relationship

Using these numbers, we can plot the densities of probabilities, to get a sense of the probability and the uncertainty around those estimates



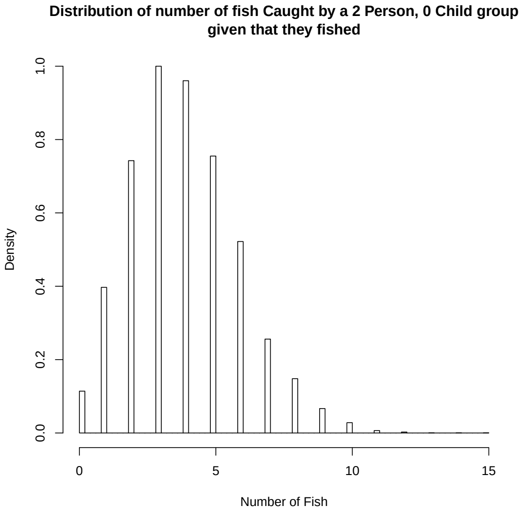
HOW MANY FISH DID SOMEONE CATCH GIVEN THAT THEY FISHED?

We can also simulate to get a distribution or expectation of $Y|Z = 1$ - let's do this for the "median" group of 2 with no children

```
# Let's plot the count distribution conditional on $Z_i = 1$
# Subset our simulation to simulated Xs
sim.x <- sim.pars[,1:ncol(X)]

# Calculate distribution of fish caught | fishing
# Set covariates to median # of children and median # of people
Xcovariates <- apply(X, 2, median)
# Remember that for each group size, this is a vector of 10000 predicted
     probabilities
set.seed(02138)
fish <- rep(NA, nrow=10000)
fish <- rpois(10000,exp(Xcovariates%*%t(sim.x)))
```

How many fish did someone catch given that they fished?



OUTLINE

Fixed and Random Effects

- Most unclear terminology in all of statistics.

Fixed and Random Effects

- Most unclear terminology in all of statistics. At least 5 definitions across different authors (See Andy Gelman's comments here)

Fixed and Random Effects

- Most unclear terminology in all of statistics. At least 5 definitions across different authors (See Andy Gelman's comments here)
- We're going to focus on fixed vs. random effects in *hierarchical* or *multi-level* models.

Fixed and Random Effects

- Most unclear terminology in all of statistics. At least 5 definitions across different authors (See Andy Gelman's comments here)
- We're going to focus on fixed vs. random effects in *hierarchical* or *multi-level* models.
- For these purposes, "fixed" effects are treated as constants while "random" effects are modeled as random variables.
- **Key point:** Ignore the labels, look at what the researcher actually wants to do.

MULTI-LEVEL MODELS

- We often have observations of units that are nested within some grouping structure.

MULTI-LEVEL MODELS

- We often have observations of units that are nested within some grouping structure.
  - Students, nested in classrooms, nested in schools.

MULTI-LEVEL MODELS

- We often have observations of units that are nested within some grouping structure.
  - Students, nested in classrooms, nested in schools.
  - Voters, nested in districts, nested in states.

Multi-level models

- We often have observations of units that are nested within some grouping structure.
  - Students, nested in classrooms, nested in schools.
  - Voters, nested in districts, nested in states.
  - Patients, nested in hospitals

Multi-level models

- We often have observations of units that are nested within some grouping structure.
  - Students, nested in classrooms, nested in schools.
  - Voters, nested in districts, nested in states.
  - Patients, nested in hospitals
- Our data are now indexed by i (unit) and j (cluster).

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}.

A SIMPLE TWO-LEVEL MODEL

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

- β_0 and β_1 are our usual fixed intercept/slope coefficients, and ϵ_{ij} is our individual-level error term.

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

- β_0 and β_1 are our usual fixed intercept/slope coefficients, and ϵ_{ij} is our individual-level error term.
- But what's η_j, our group-level term?

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

- β_0 and β_1 are our usual fixed intercept/slope coefficients, and ϵ_{ij} is our individual-level error term.
- But what's η_j, our group-level term? Depends on how we model it!

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

- β_0 and β_1 are our usual fixed intercept/slope coefficients, and ϵ_{ij} is our individual-level error term.
- But what's η_j, our group-level term? Depends on how we model it!
  - "Fixed" effect: η_j is an unmodeled constant (like β_0)

A simple two-level model

- We observe outcome Y_{ij} and covariate X_{ij}. The simplest two-level linear model would look like:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \eta_j + \epsilon_{ij}$$

- β_0 and β_1 are our usual fixed intercept/slope coefficients, and ϵ_{ij} is our individual-level error term.
- But what's η_j, our group-level term? Depends on how we model it!
  - "Fixed" effect: η_j is an unmodeled constant (like β_0)
  - "Random" effect: η_j is a random variable (like ϵ_{ij})

Example: 2012 U.S. Election Results and Income

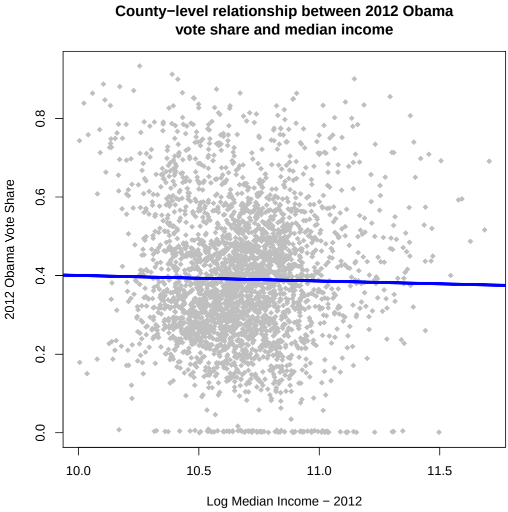
EXAMPLE: 2012 U.S. ELECTION RESULTS AND INCOME

- Suppose we model Y_{ij}, the county-level vote share for Barack Obama in the 2012 U.S. election using X_{ij}, the county median income in 2012 (as estimated by the Census Bureau)

Example: 2012 U.S. Election Results and Income

- Suppose we model Y_{ij}, the county-level vote share for Barack Obama in the 2012 U.S. election using X_{ij}, the county median income in 2012 (as estimated by the Census Bureau)
- Here i indexes counties and j indexes states.

MODEL 1: NO GROUPING



MODEL 2: "FIXED" EFFECTS

- Simple model assumes no difference in "cluster" effects.

MODEL 2: “FIXED” EFFECTS

- Simple model assumes no difference in “cluster” effects. But we know counties within states likely have correlated outcomes.

Model 2: "Fixed" Effects

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.

Model 2: "Fixed" Effects

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.

Model 2: "Fixed" Effects

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.
- Easiest way is to estimate a "fixed effect" parameter for each state.

MODEL 2: "FIXED" EFFECTS

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.
- Easiest way is to estimate a "fixed effect" parameter for each state. Changes model to have a unique intercept for each grouping rather than one single intercept.

MODEL 2: "FIXED" EFFECTS

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.
- Easiest way is to estimate a "fixed effect" parameter for each state. Changes model to have a unique intercept for each grouping rather than one single intercept.
  - Advantage: For linear models, can still use OLS

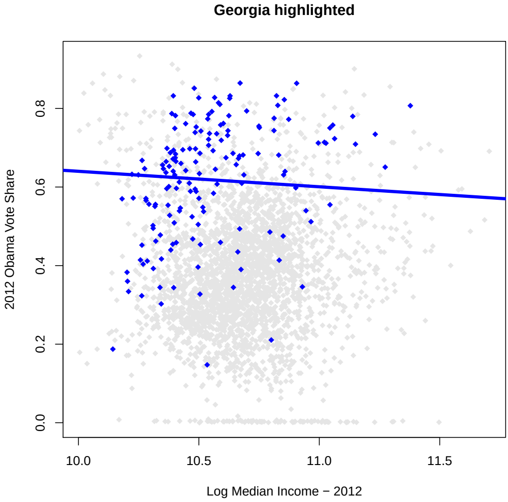
MODEL 2: "FIXED" EFFECTS

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.
- Easiest way is to estimate a "fixed effect" parameter for each state. Changes model to have a unique intercept for each grouping rather than one single intercept.
  - Advantage: For linear models, can still use OLS
  - Disadvantage: Can't include cluster-level predictors

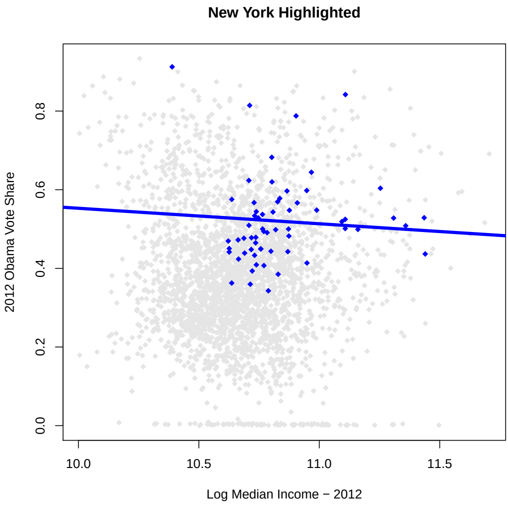
Model 2: "Fixed" Effects

- Simple model assumes no difference in "cluster" effects. But we know counties within states likely have correlated outcomes. Some of this correlation is due to a common "state-level" influence.
- We can improve the model by modeling this state-level factor.
- Easiest way is to estimate a "fixed effect" parameter for each state. Changes model to have a unique intercept for each grouping rather than one single intercept.
    - Advantage: For linear models, can still use OLS
    - Disadvantage: Can't include cluster-level predictors (perfect co-linearity).

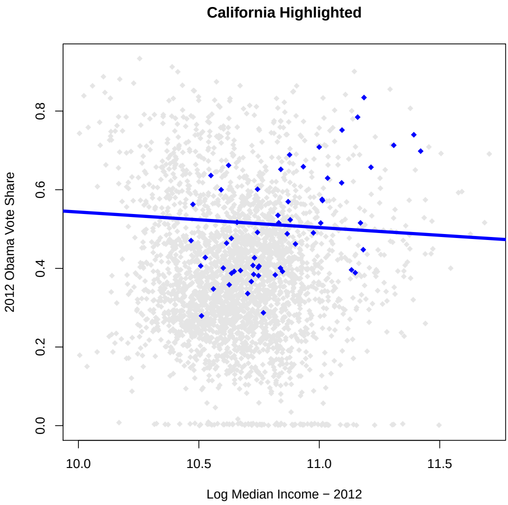
Model 2: "Fixed" Effects



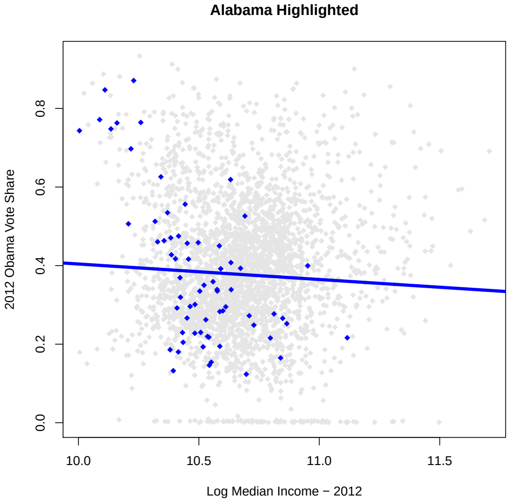
Model 2: "Fixed" Effects



Model 2: "Fixed" Effects



MODEL 2: “FIXED” EFFECTS



MODEL 3: “RANDOM” EFFECTS

- Alternative is to assume group-level effect η_j comes from some underlying distribution (like the error term).

Model 3: "Random" Effects

- Alternative is to assume group-level effect η_j comes from some underlying distribution (like the error term).
- For linear random effect, typical assumption is that $\eta_j \sim \mathcal{N}(0, \sigma_\eta^2)$.

MODEL 3: "RANDOM" EFFECTS

- Alternative is to assume group-level effect η_j comes from some underlying distribution (like the error term).
- For linear random effect, typical assumption is that $\eta_j \sim \mathcal{N}(0, \sigma_\eta^2)$. η_j are i.i.d. across groupings and is independent of the individual error term ϵ_{ij}.
- **Intuition**: We're decomposing the overall error into "between-group" variance (σ_η^2) and a "within-group" variance (σ^2)

ESTIMATING RANDOM EFFECTS

- Maximum likelihood methods still work...

Estimating random effects

- Maximum likelihood methods still work...fairly well...

ESTIMATING RANDOM EFFECTS

- Maximum likelihood methods still work...fairly well...for simple models.

ESTIMATING RANDOM EFFECTS

- Maximum likelihood methods still work...fairly well...for simple models. The `lme4` package in `R` has routines: `lmer` and `glmer` to fit mixed effects models (models with both "fixed" and "random" effects) by ML.

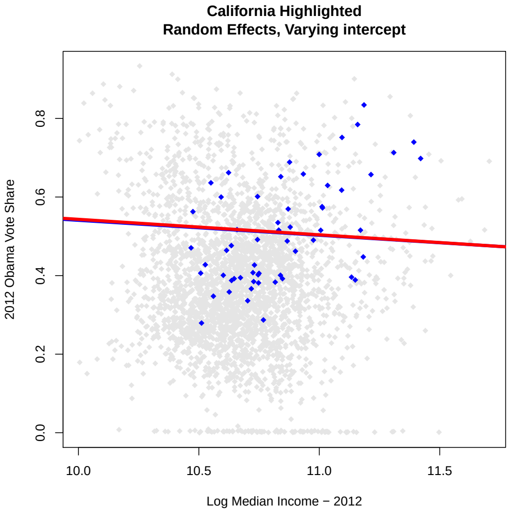
ESTIMATING RANDOM EFFECTS

- Maximum likelihood methods still work...fairly well...for simple models. The `lme4` package in `R` has routines: `lmer` and `glmer` to fit mixed effects models (models with both “fixed” and “random” effects) by ML.
- As you get smaller sample sizes, fewer clusters or more complicated variance structures, turn to Bayesian approaches.

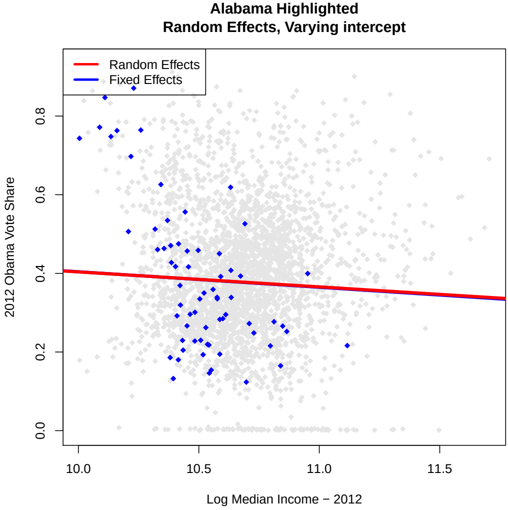
ESTIMATING RANDOM EFFECTS

- Maximum likelihood methods still work...fairly well...for simple models. The `lme4` package in `R` has routines: `lmer` and `glmer` to fit mixed effects models (models with both "fixed" and "random" effects) by ML.
- As you get smaller sample sizes, fewer clusters or more complicated variance structures, turn to Bayesian approaches. Increasingly popular with easy-to-use software - see Gelman and Hill (2007)

Model 3: Random intercept model



Model 3: Random intercept model



VARYING SLOPE AND INTERCEPT – "FIXED EFFECTS"

- May want to allow slope and intercept to vary across groups.

VARYING SLOPE AND INTERCEPT – "FIXED EFFECTS"

- May want to allow slope and intercept to vary across groups.

$$Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + \eta_j + \epsilon_{ij}$$

Varying Slope and Intercept – "fixed effects"

- May want to allow slope and intercept to vary across groups.

$$Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + \eta_j + \epsilon_{ij}$$

- Under a "fixed effects" assumption, how do we estimate this?

Varying Slope and Intercept – "fixed effects"

- May want to allow slope and intercept to vary across groups.

$$Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + \eta_j + \epsilon_{ij}$$

- Under a "fixed effects" assumption, how do we estimate this? Estimate a unique slope/intercept parameter for each j –

Varying Slope and Intercept – "fixed effects"

- May want to allow slope and intercept to vary across groups.

$$Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + \eta_j + \epsilon_{ij}$$

- Under a "fixed effects" assumption, how do we estimate this? Estimate a unique slope/intercept parameter for each j – for linear model, we can just use OLS with interaction between X and the group dummies.

Varying Slope and Intercept – "fixed effects"

- May want to allow slope and intercept to vary across groups.

$$Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + \eta_j + \epsilon_{ij}$$

- Under a "fixed effects" assumption, how do we estimate this? Estimate a unique slope/intercept parameter for each j – for linear model, we can just use OLS with interaction between X and the group dummies.
- Point estimates identical to also subsetting the data by group and estimating j separate regressions.

Varying Slope and Intercept – “random effects”

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.

Varying Slope and Intercept – “random effects”

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to “partially” pool?

Varying Slope and Intercept – "random effects"

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to "partially" pool? Can instead put a distribution on β_{1j}:

$$\beta_{1j} \sim \mathcal{N}(\beta_1, \sigma_\beta^2)$$

VARYING SLOPE AND INTERCEPT – "RANDOM EFFECTS"

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to "partially" pool? Can instead put a distribution on β_{1j}:

$$\beta_{1j} \sim \mathcal{N}(\beta_1, \sigma_\beta^2)$$

- **Intuition:** We let β_{1j} vary across groups j, but within a certain range (not too far from the common group mean β_1).

Varying Slope and Intercept – "random effects"

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to "partially" pool? Can instead put a distribution on β_{1j}:

$$\beta_{1j} \sim \mathcal{N}(\beta_1, \sigma_\beta^2)$$

- **Intuition:** We let β_{1j} vary across groups j, but within a certain range (not too far from the common group mean β_1).
- Can either estimate σ_β^2 or fix it to begin with.

VARYING SLOPE AND INTERCEPT – "RANDOM EFFECTS"

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to "partially" pool? Can instead put a distribution on β_{1j}:

$$\beta_{1j} \sim \mathcal{N}(\beta_1, \sigma^2_\beta)$$

- **Intuition:** We let β_{1j} vary across groups j, but within a certain range (not too far from the common group mean β_1).
- Can either estimate σ^2_β or fix it to begin with. What happens if we set $\sigma^2_\beta = 0$?

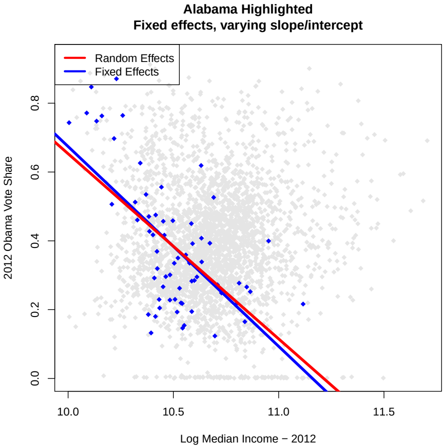
VARYING SLOPE AND INTERCEPT – "RANDOM EFFECTS"

- Under fully interacted FE model, data from one group tells us *nothing* about slope/intercept in another group.Sacrificing efficiency for less bias.
- What if we want to "partially" pool? Can instead put a distribution on β_{1j}:

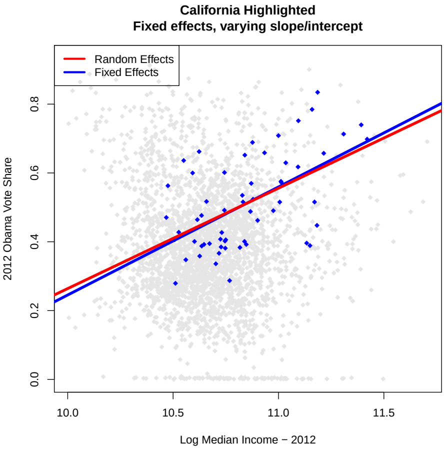
$$\beta_{1j} \sim \mathcal{N}(\beta_1, \sigma_\beta^2)$$

- **Intuition:** We let β_{1j} vary across groups j, but within a certain range (not too far from the common group mean β_1).
- Can either estimate σ_β^2 or fix it to begin with. What happens if we set $\sigma_\beta^2 = 0$? We get the fully pooled slope model again!

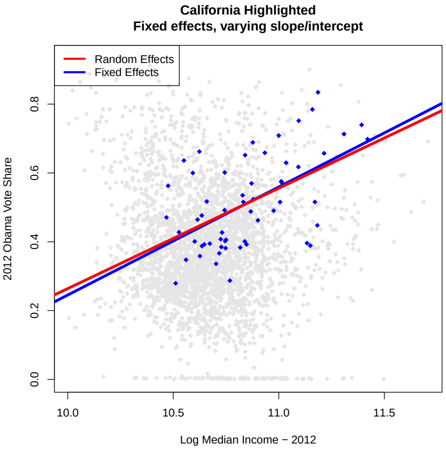
COMPARING FE AND RE



COMPARING FE AND RE



Comparing FE and RE



Not very different (due to lots of data), but can see RE slope slightly attenuated towards 0.

How to think about hierarchical models

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)
  - No pooling = higher variance, but less bias + capture interesting variation

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)
  - No pooling = higher variance, but less bias + capture interesting variation
- Pooling is a continuum:

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)
  - No pooling = higher variance, but less bias + capture interesting variation
- Pooling is a continuum:
  - All pooling: Pooled slope and intercept: $Y_{ij} = \beta_0 + \beta_1 X_{ij}$

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
    - Pooling = lower variance, but more bias (stronger assumptions)
    - No pooling = higher variance, but less bias + capture interesting variation
- Pooling is a continuum:
    - All pooling: Pooled slope and intercept: $Y_{ij} = \beta_0 + \beta_1 X_{ij}$
    - Some pooling: Pooled slope, group-varying intercept: $Y_{ij} = \beta_{0j} + \beta_1 X_{ij}$

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)
  - No pooling = higher variance, but less bias + capture interesting variation
- Pooling is a continuum:
  - All pooling: Pooled slope and intercept: $Y_{ij} = \beta_0 + \beta_1 X_{ij}$
  - Some pooling: Pooled slope, group-varying intercept: $Y_{ij} = \beta_{0j} + \beta_1 X_{ij}$
  - No pooling: Group-varying slope and intercept: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij}$

How to think about hierarchical models

- When fitting a model on data comprised of units coming from groupings, you are always considering how much *pooling* across groups to do.
  - Pooling = lower variance, but more bias (stronger assumptions)
  - No pooling = higher variance, but less bias + capture interesting variation
- Pooling is a continuum:
  - All pooling: Pooled slope and intercept: $Y_{ij} = \beta_0 + \beta_1 X_{ij}$
  - Some pooling: Pooled slope, group-varying intercept: $Y_{ij} = \beta_{0j} + \beta_1 X_{ij}$
  - No pooling: Group-varying slope and intercept: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij}$
- Multilevel modeling provides a compromise between estimating fully separate models and assuming a constant coefficient across all groups.

QUESTIONS

Questions?