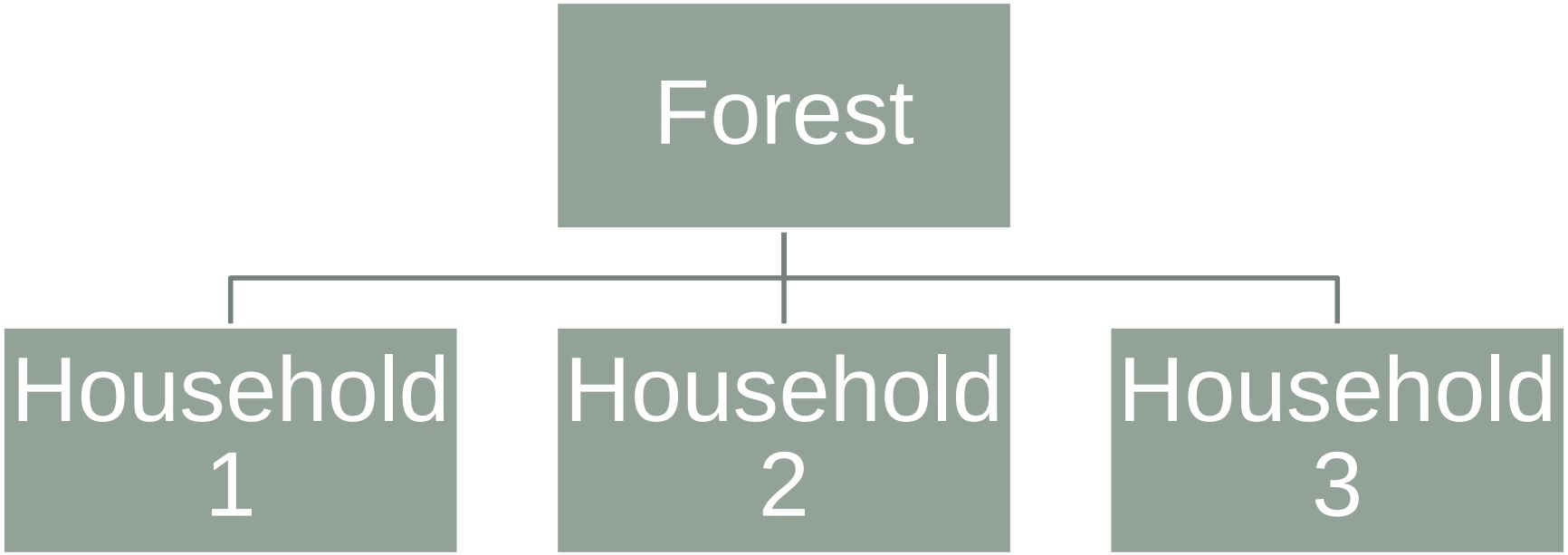
LINKING HOUSEHOLD TO OTHER DATA

Eric A. Coleman

CHALLENGES OF DATA STRUCTURE

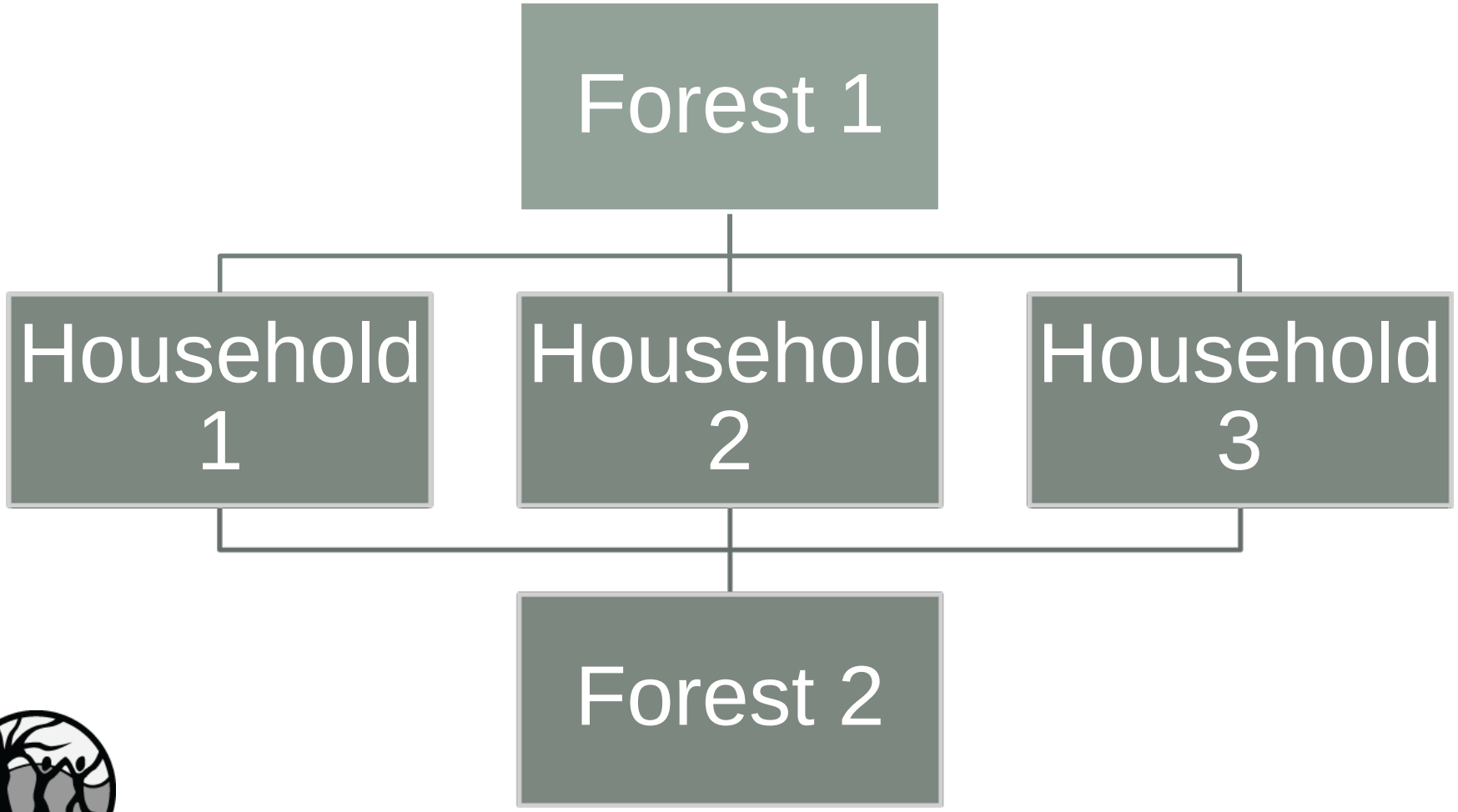
Households Nested within a Single Forest



Violation of traditional model assumptions

- When comparing across sites
  - Households in different forests uncorrelated with each other
  - Households in the same forest are correlated
- Intra-forest correlation
  - Rules likely similar across households
  - Behaviors likely to be similar in the same forest
    - At least more similar than behavior between different forests
  - Biophysical constraints probably similar

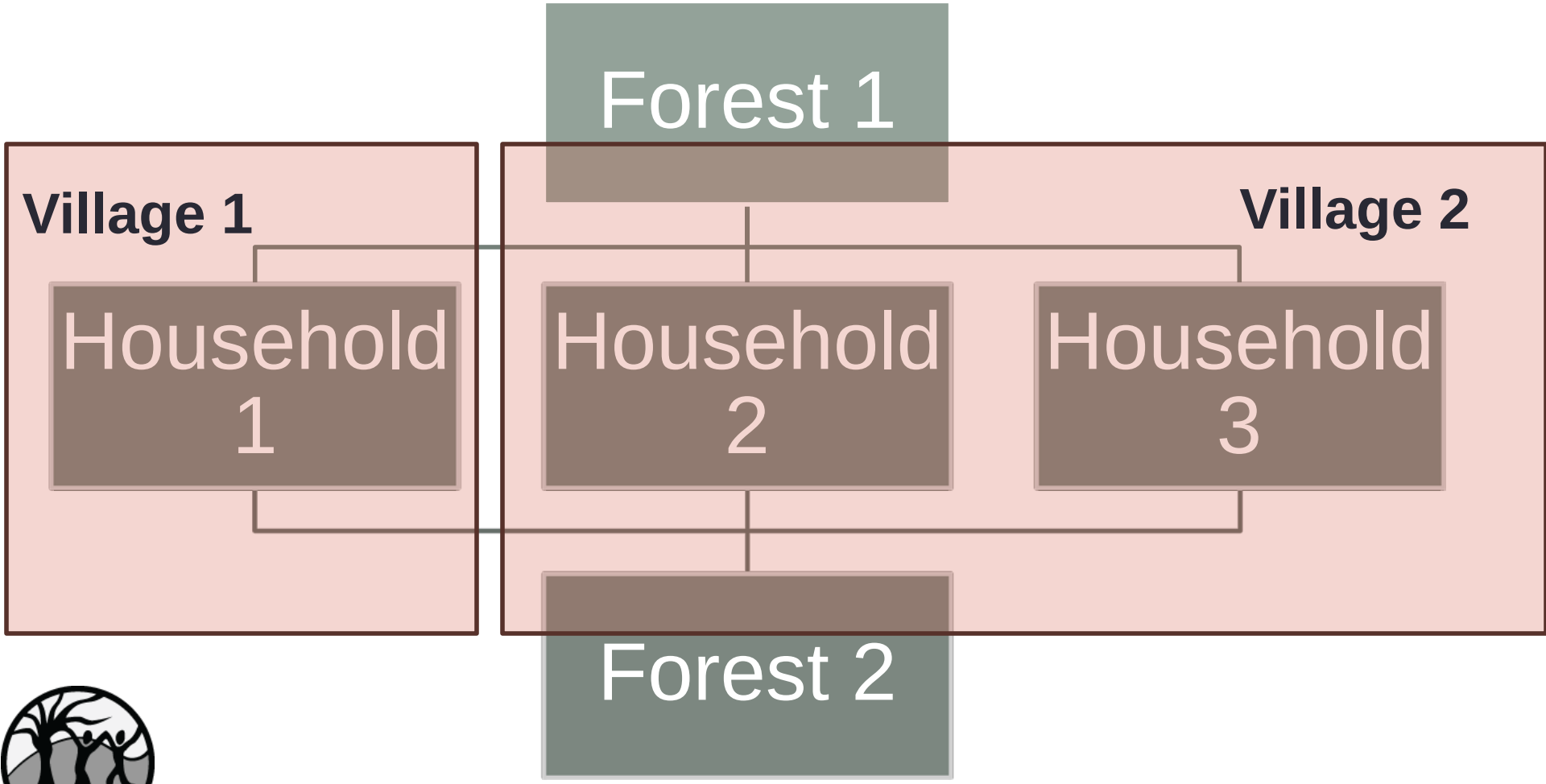
Households Cross-Nested to Multiple Forests



Violations of traditional model assumptions

- When comparing across forests
  - The behavior of a household within one forest is likely to be similar to their behavior in another forest
    - At least more similar than a completely different household within a completely different forest
  - The behavior of all households within a given forest is likely to similar
    - The problem discussed previously
- Cross-nesting
  - Forests nested within households
  - Households nested within forests

Add on village layer



What Outcome are you trying to explain?

- Household-Forest dyadic level
  - Household benefits derived from each forest
  - Use of each forest
  - Participation in each forest's governance
- Household level? (do not vary over forest)
  - Health, livelihoods
- Forest level? (do not vary over household)
  - Forest conditions like biomass, species diversity, etc.
  - Forest governance institutions

FOREST-LEVEL OUTCOMES

Overview

- We care about some aggregate measure of forest outcomes
  - We want to relate household-specific variables of interest to this outcome
- Example:
  - What is the relationship between household wealth and forest biomass? Do forests that have wealthier households nearby retain more biomass?
  - What is the relationship between household wealth inequality and forest biomass?
- Since the outcome is aggregate, we must find some way of aggregating household-level data to relate to each forest
  - Example: Average wealth or some measure of the variance in wealth

Formally…

- Average Wealth explains biomass
  - Let *i* denote a household
  - Let *j* denote a forest
  - Let n_j denote the number of households in forest *j*

$$Biomass_j = \beta_0 + \beta_W \frac{1}{n_j} \sum_{i=1}^{n_j} Wealth_{ij} + \varepsilon_j$$

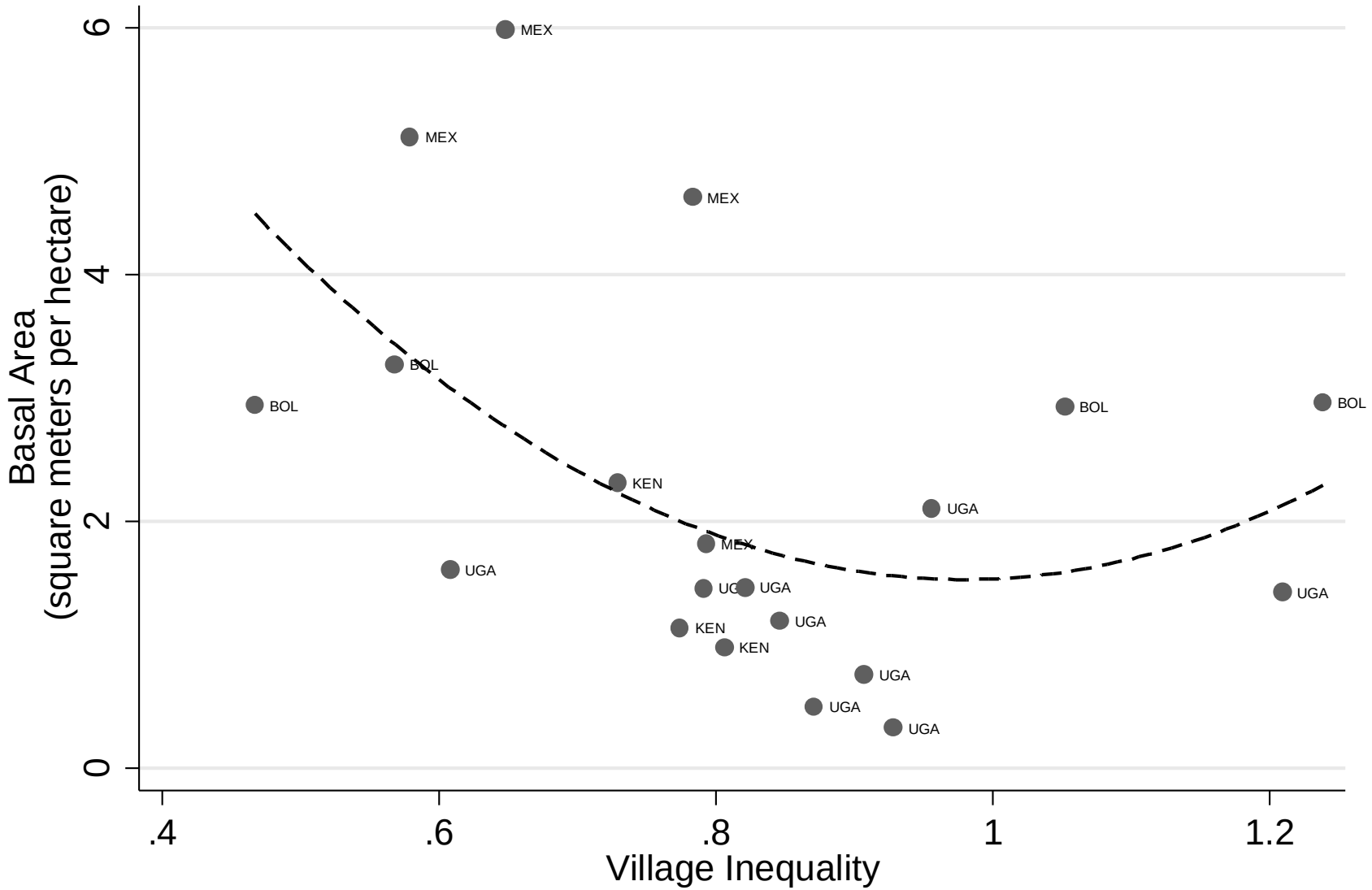
Average wealth across all households in forest *j*

Some things to note

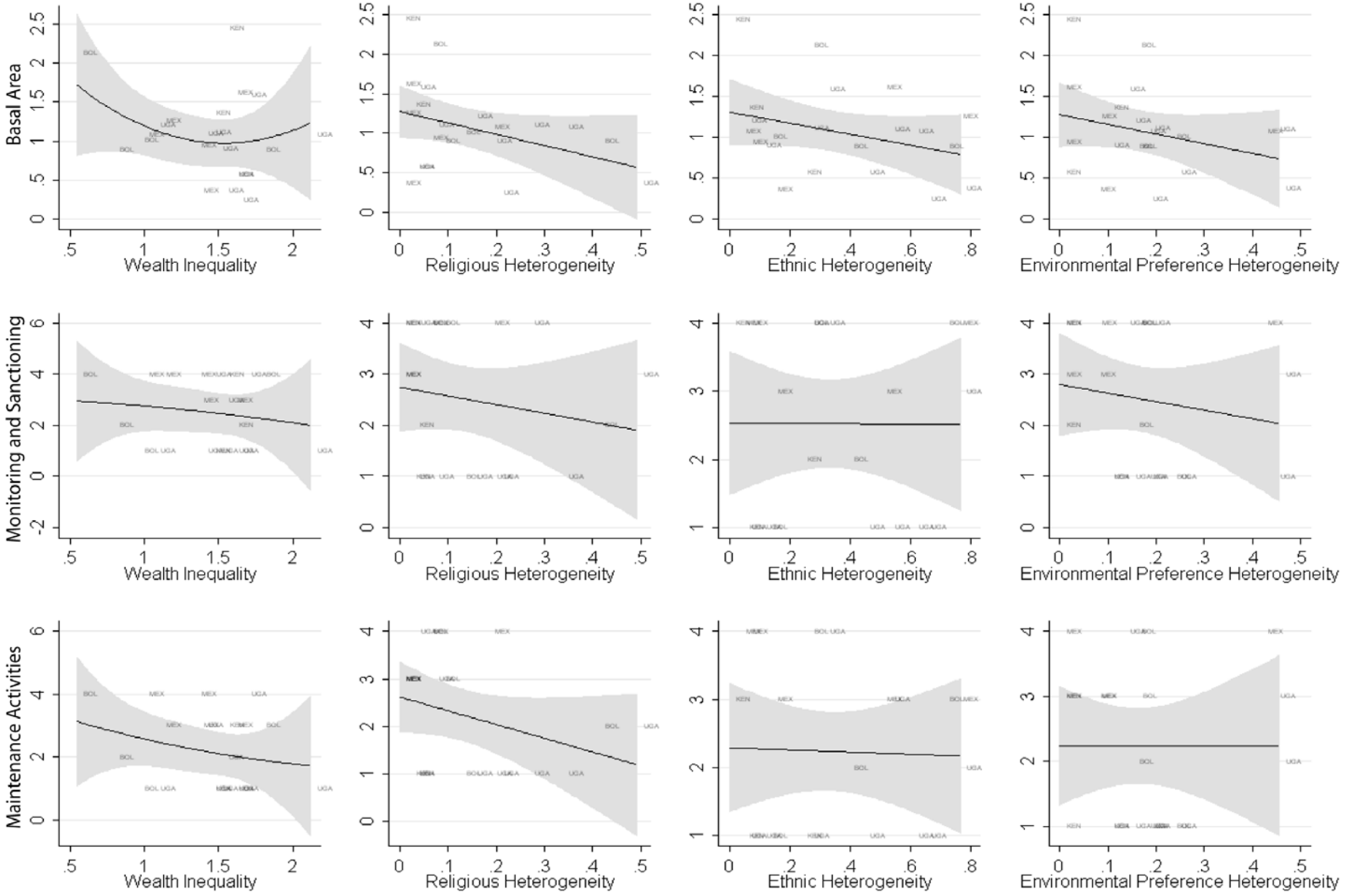
- You don’t just have to look at the mean level of wealth
  - If you want to look at inequality, you need some measure of how wealth is distributed among the households (like variance, GINI, etc.)
- The estimation strategy is straightforward
  - Fairly easy to calculate averages or variances of a variable across households in a forest
  - After this, just use OLS
    - May want to use WLS based upon the sampling intensity at each site
- Downsides
  - Expensive: many household surveys go into collecting a single data point in the analysis
  - Loose a lot of power—small sample size

An Application: Heterogeneity and Collective Action

- Theory
  - Heterogeneity and collective action (Mancur Olson)
  - Heterogeneous actors have different management preferences
- Measurement
  - How does one measure heterogeneity
    - Economic inequality? (assets)
    - Religious heterogeneity?
    - Ethnic heterogeneity?
    - Environmental preference heterogeneity?
  - How does one measure outcomes?
    - Forest Governance
      - Group monitoring and sanctioning
      - Group forest maintenance activities
    - Forest Conditions
      - Woody biomass



Basal Area = 12.37* - 22.159* x Village Inequality
+ 11.321** x Village Inequality^2


Basal Area
Monitoring and Sanctioning
Maintenance Activities
Wealth Inequality
Religious Heterogeneity
Ethnic Heterogeneity
Environmental Preference Heterogeneity

Lessons

- For the three types of outcomes and for all four measures of heterogeneity
  - Never observe a positive relationship between heterogeneity and forest outcomes
  - Either negative or non-significant
  - Biomass the most sensitive to heterogeneity
- Note, however…
  - These results are suggestive, but they rely on data aggregated up to only 23 forests
    - Although more than 1,200 surveyed households
  - We're still collecting household survey data and would need to expand this study to compare outcomes in more forests

HOUSEHOLD-FOREST DYADIC LEVEL OUTCOMES

Forest-Household Dyadic Data

- Dep variable: varies over each forest for each household
  - For example, the benefits a household gets from each forest
    - If you want to leverage the links between specific households and forest conditions/governance then ideally you need to link this in the measurement stage
  - We can deal with predictive variables at the household level, forest level, and at the dyadic forest-household level
- We still have to deal with the non-independence of observations if household are nested or cross-nested with forests
  - Model this non-independence explicitly
    - Hierarchal Linear Modeling, Mixed Modeling, Random Intercepts, Multilevel modeling
  - Note that the power to identify an effect of a variable at higher levels depends on the sample size at those higher levels

An Example…

- Biomass, Household Gender, and the household's property rights to each forest explains their benefits from each forest
  - Let *i* denote a household
  - Let *j* denote a forest

$$Benefits_{ij} = \beta_0 + \beta_1 Biomass_j + \beta_2 Gender_i + \beta_3 Property\ Rights_{ij} + \mu_j + \theta_i + \varepsilon_{ij}$$

Correlation within forests Correlation within households

Data

- Dependent Variable – *Benefits Index*
  - 45 point scale
    - The household rates the importance of the cash income, subsistence income, contribution to soil fertility, erosion control, and cultural/spiritual benefit they get in each nearby forest
    - You may have much more objective measures with your data
- Key Independent Variable – *Property Rights*
  - Household-forest level
  - Guttman scale, 0-6
- Additional control variables at Forest, household, and household-forest levels

A Note on Multi-Level Modelling

- Can be computationally burdensome
  - Some evidence that results can be sensitive to the search algorithm
    - Ordinary or Adaptive Quadrature with sufficient integration points
    - Should check convergence criteria, sensitivity to initial parameter estimates, identification
  - Cross-nested models especially burdensome
- Integrates nicely within a Bayesian framework

Bivariate Relationship

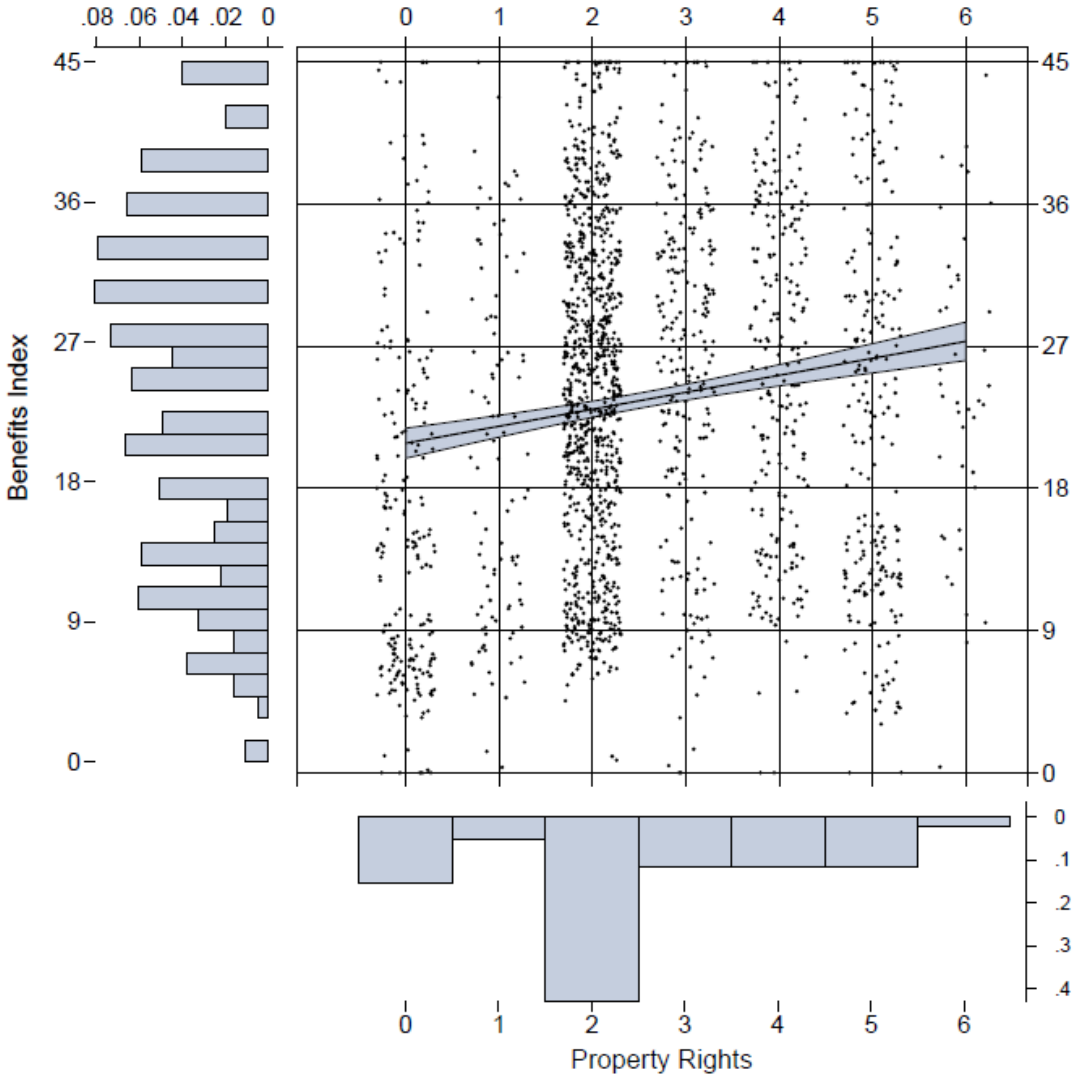


Table 3. Cross-Nested Property Rights Model for Benefit Index

	Model1 Coef (SE)	Model2 Coef (SE)	Model3 Coef (SE)
Fixed Part			
Property Rights	1.036*** (0.17)		
Landholder X Property Rights		1.254*** (0.20)	
Non-Landholder X Property Rights		0.469 (0.33)	
Ethnic Majority X Property Rights			1.433*** (0.20)
Ethnic Minority X Property Rights			0.262 (0.27)
Landholder	-0.595 (0.84)	-2.110* (1.10)	-0.618 (0.84)
Majority Ethnic Group	0.290 (0.59)	0.210 (0.58)	-2.356*** (0.91)
Control Variables	YES	YES	YES
Random Part			
χ^2	470.963***	460.021***	475.019***
σ_f	5.600*** (0.81)	5.477*** (0.80)	5.596*** (0.81)
σ_h	4.955*** (0.32)	0.566 (0.27)	0.505 (0.30)
$\sigma_{Property\ Rights}$		4.742*** (0.40)	4.770*** (0.39)
Model Statistics			
AIC	13842.769	13841.371	13832.399
χ^2	91.303***	95.993***	105.622***
N	1903	1903	1903

Thank You

- Krister Andersson for help with work on household heterogeneity and forest outcomes
- All the wonderful IFRI colleagues who painstakingly collect the data

Wealth Distribution by State