Poverty Dynamics and Synthetic Panel Methods

Peter Lanjouw (Vrije Universiteit Amsterdam)

Presentation for

Bangladesh Institute of Development Studies

Annual Conference 2024

December 10, 2024

DATA & EVIDENCE deep TO END EXTREME POVERTY

DEEP research on Poverty Dynamics

- Data and Evidence to End Extreme Poverty (DEEP)
  - Research Program administered by Oxford Policy Management and funded by UK FCDO.
- Program of research into poverty dynamics implemented out of University of Copenhagen and Vrije Universiteit of Amsterdam
  - Directors: Finn Tarp (U. Copenhagen) and Peter Lanjouw (VU Amsterdam)
  - Empirical analyses in Bangladesh, Ethiopia, India, Mozambique, Tanzania

DATA & EVIDENCE deep TO END EXTREME POVERTY

DEEP research on Poverty Dynamics: Approach

- Argument: Poverty assessments need to be reoriented
  - Away from “how to reduce poverty reduction within a context of generalized growth”
  - Towards “how to attenuate rising poverty in the face of generalized economic slowdown”
- This calls for a concerted focus on poverty dynamics
  - Who is vulnerable to falling back into poverty?
  - Who has been trapped in poverty during the “good years?”
- Ideally, we would draw on panel data for such analysis
  - Panel data rarely available, expensive, possibly non-representative

DATA & EVIDENCE deep TO END EXTREME POVERTY

Poverty Dynamics from Synthetic Panels

- Can widely available cross-sections indicate transitions in and out of poverty?
  - Dang, Lanjouw, Luoto and McKenzie (2014) introduce an approach yielding upper and lower bounds on mobility.
  - Dang and Lanjouw (2023) propose refinement to produce point estimates
    - At the cost of additional assumptions (potentially strong)
  - Rongen and Lanjouw (2024) explore methods to produce narrower bounds
  - Garcés-Urzainqui, Lanjouw and Rongen (2021) present overview of methodology

- Result: information on movements in and out of poverty, and chronic poverty
- Backbone of method: time-invariant household characteristics, vector x_i
  - Caste, religion, birth year, education of household head, region of birth, etc.

DATA & EVIDENCE deep TO END EXTREME POVERTY

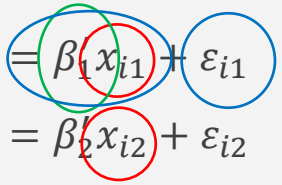
Mechanics of the method: Predicting round 1 income for a round 2 observation

1. Specify an income model using only time-invariant characteristics x_i :

$$y_i = \beta' x_i + \varepsilon_i$$

Round 1 income: $y_{i1} = \beta_1' x_{i1} + \varepsilon_{i1}$

Round 2 income: $y_{i2} = \beta_2' x_{i2} + \varepsilon_{i2}$



2. Estimate model using round 1 data (of other households) and take coefficients $\widehat{\beta_1'}$
3. By definition $x_{i1} = x_{i2}$ (time-invariant regressors)
4. Predict fitted round 1 income for round 2 households i :

$$\hat{y}_{i1} = \widehat{\beta_1'} x_{i2}$$

4. But: how to obtain estimate for round 1 error term ε_{i1} ?
   - This depends on the relationship between ε_{i1} and ε_{i2}

DATA & EVIDENCE deep TO END EXTREME POVERTY

Mechanics of the method:
Predicting round 1 income for a round 2 observation

Assumption 1:

Positive correlation ρ between errors ε_{i1} and ε_{i2} - on average

ρ is assumed to be between 0 and 1

Extreme cases used as bounds on mobility:

LOWER BOUND

- Perfect positive correlation between errors: $\rho = 1$
- Minimal variation in income
- Add scaled error of the round 2 observation of the household:

$$\hat{y}_{i1}^{LB} = \widehat{\beta_1'} x_{i2} + \gamma \hat{\varepsilon}_{i2}.$$

UPPER BOUND

- No correlation between errors: $\rho = 0$
- Larger variation in income
- Add randomly drawn error of a round 1 observation:

$$\hat{y}_{i1}^{UB} = \widehat{\beta_1'} x_{i2} + \widetilde{\hat{\varepsilon}_{j1}}$$

- Repeat R times and take average

DATA & EVIDENCE deep TO END EXTREME POVERTY

Mechanics of the method:
Predicting round 1 income for a round 2 observation

- ❖ Evaluate predicted upper and lower bound incomes against the poverty line
- ➢ Estimated poverty transition probabilities at household level
- ❖ Aggregate probabilities at country and subgroup level

Assumption 2:

Constant underlying population

- ➢ No large-scale migration
- ➢ Select stable age cohort, e.g. heads 25-55 years of age

DATA & EVIDENCE deep TO END EXTREME POVERTY

Empirical Application: Datasets

(Dang, Lanjouw, Luoto and McKenzie, JDE 2014)

- Choose two genuine panels from Vietnam and Indonesia:
- VLSS 1992/93 and 1997/98
  - Period over which poverty fell from 58% to 37%, more households exiting poverty than entering
  - Panel of approximately 4800 households
- Indonesian Family Life Survey 1997 and 2000 (IFLS2 and 3)
  - Static in terms of overall poverty levels, household moving into and out of poverty at similar rates
  - Panel of 7500 households

DATA & EVIDENCE deep TO END EXTREME POVERTY

Validation of method

- Randomly split each genuine panel into two sub-samples, A and B.
  - Use sub-sample A from round 1 and sub-sample B from round 2 as two repeated cross-sections.
  - Then carry out our method by using sub-sample A to impute round 1 values for sub-sample B, and compare to results we would get using genuine panel for sub-sample B.

DATA & EVIDENCE deep TO END EXTREME POVERTY

Choosing variables

- Consider a hierarchy of models which progressively employ more and more data that is sometimes, or potentially, collected retrospectively.
- Since we have the actual panel data to work with, we can force variables to be time-invariant by using round 1 variables.
- Start with a basic “traditional model”, and progressively add more regressors.

DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Models

1. (Basic Model): gender of head, age of head as of round 1, birthplace of head (rural/urban), whether the head ever attended primary school, education of head's parents, head's religion and ethnicity. ($R^2 = 0.20$)
2. Add locational dummies for where household was living in round 1.
3. Add community variables from round 1 (e.g. village has electricity, village has a paved road, community has a primary school)
4. Head's sector of work and education in round 1
5. Demographic variables from round 1 (household size, number of children)
6. Household's assets and housing quality as of round 1 – e.g. did household own TV, radio, what sort of roof and floor did it have? ($R^2 = 0.42$).

DATA & EVIDENCE deep TO END EXTREME POVERTY

Results: predicting *levels* of poverty

Data Source:	Round 1:	Lower Bound Basic	Lower Bound Full	Truth 95% CI		Upper Bound Basic	Upper Bound Full
IFLS	1997 Poverty Rate (P0):	0.149	0.159	0.145	0.188	0.164	0.150
VLSS	1992 Poverty Rate (P0):	0.611	0.592	0.597	0.682	0.558	0.578

Method gets levels fairly close

DATA & EVIDENCE deep TO END EXTREME POVERTY

Correlation between residuals

- Recall our claim was that the residuals would likely be positively autocorrelated, making our first method an upper bound, and that this correlation would shrink as we add more variables to the model.
- This is what we see:

Table 2: Correlation Between Round 1 and Round 2 Residuals

	1	2	3	4	5	6
Indonesia	0.474	0.466	0.464	0.452	0.408	0.348
Vietnam	0.653	0.575	0.563	0.539	0.523	0.420

Columns 1-6 build increasingly rich models of consumption.

DATA & EVIDENCE deep TO END EXTREME POVERTY

How well does the approach estimate overall movements into and out of poverty?

Table 3: Poverty Dynamics from "Pseudo" Panel and Actual Panel Data

Indonesia	Lower Bounds		Truth		Upper Bounds	
1997, 2000 Statuses	Basic	Full	95% CI		Basic	Full
Poor, Poor	0.115	0.105	0.047	0.070	0.024	0.037
Poor, Nonpoor	0.015	0.031	0.065	0.088	0.097	0.090
Nonpoor, Poor	0.021	0.030	0.065	0.088	0.111	0.099
Nonpoor, Nonpoor	0.848	0.832	0.759	0.801	0.766	0.774

Vietnam	Lower Bounds		Truth		Upper Bounds	
1992, 1998 Statuses	Basic	Full	95% CI		Basic	Full
Poor, Poor	0.360	0.322	0.275	0.360	0.227	0.288
Poor, Nonpoor	0.241	0.274	0.261	0.324	0.331	0.308
Nonpoor, Poor	0.000	0.039	0.034	0.060	0.138	0.077
Nonpoor, Nonpoor	0.398	0.366	0.300	0.386	0.305	0.327

For both countries, round 1 year is predicted, round 2 is "truth"

DATA & EVIDENCE deep TO END EXTREME POVERTY

Results seem encouraging

- Bounds not that wide:
  - Full model would lead us to estimate 3-9% of households in Indonesia and 27-31% of households in Vietnam exited poverty over 2 rounds.
  - Genuine panel would say 7-9% in Indonesia and 26-32% in Vietnam
- More detailed model for consumption with higher R^2 leads to narrower bounds
  - E.g. bounds of 0.021-0.111 using basic model vs (0.033-0.099) using full model for entry into poverty rate in Indonesia.

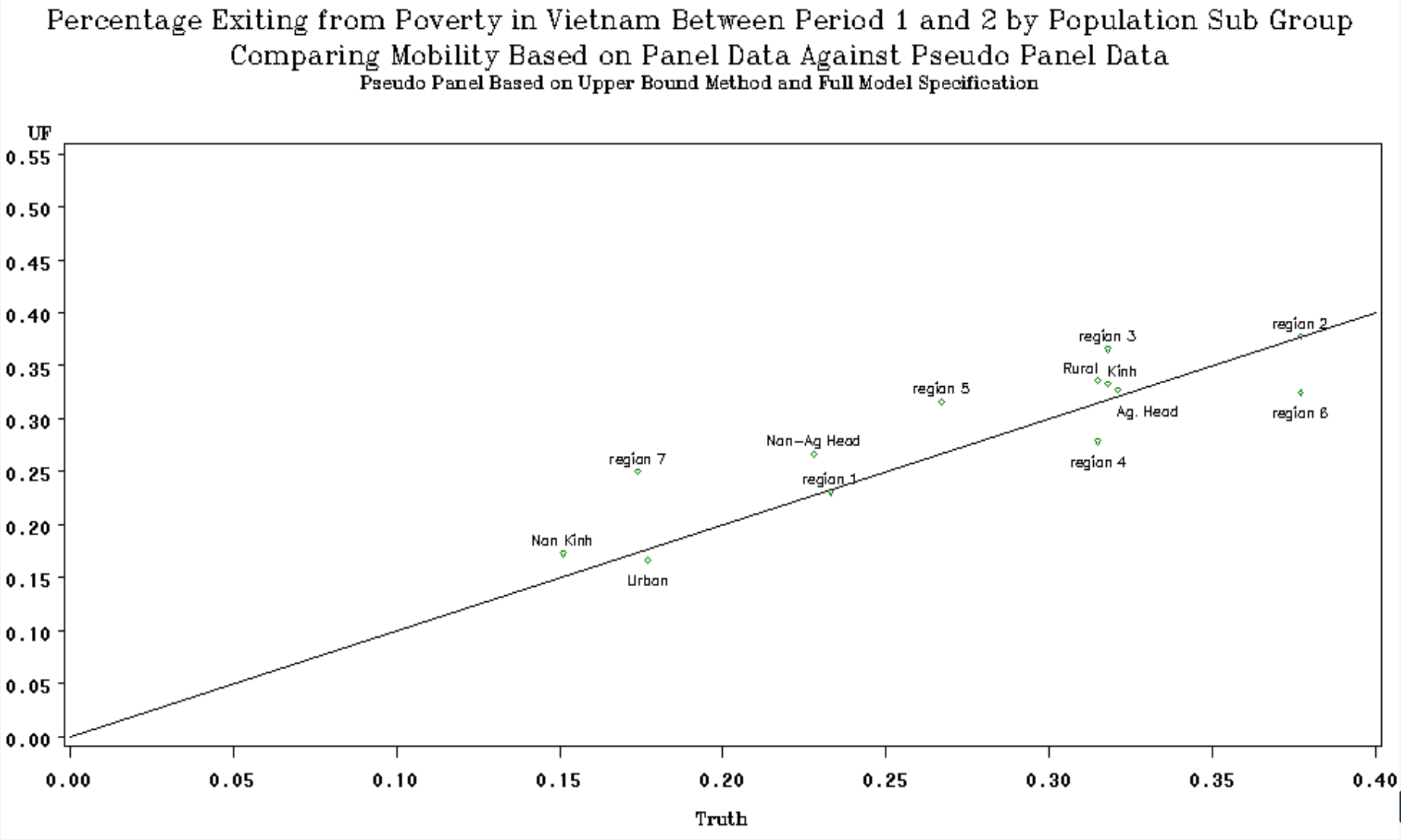
DATA & EVIDENCE deep TO END EXTREME POVERTY

What about describing profile of mobility?

- Would like to know which sub-groups of the population are more likely to be entering or exiting poverty
- E.g. are there regions where more people have entered or exited poverty? How does mobility vary by ethnic group or by education?
- We plot the full model upper bound estimates against the panel estimates to see whether ranking we get seems similar.

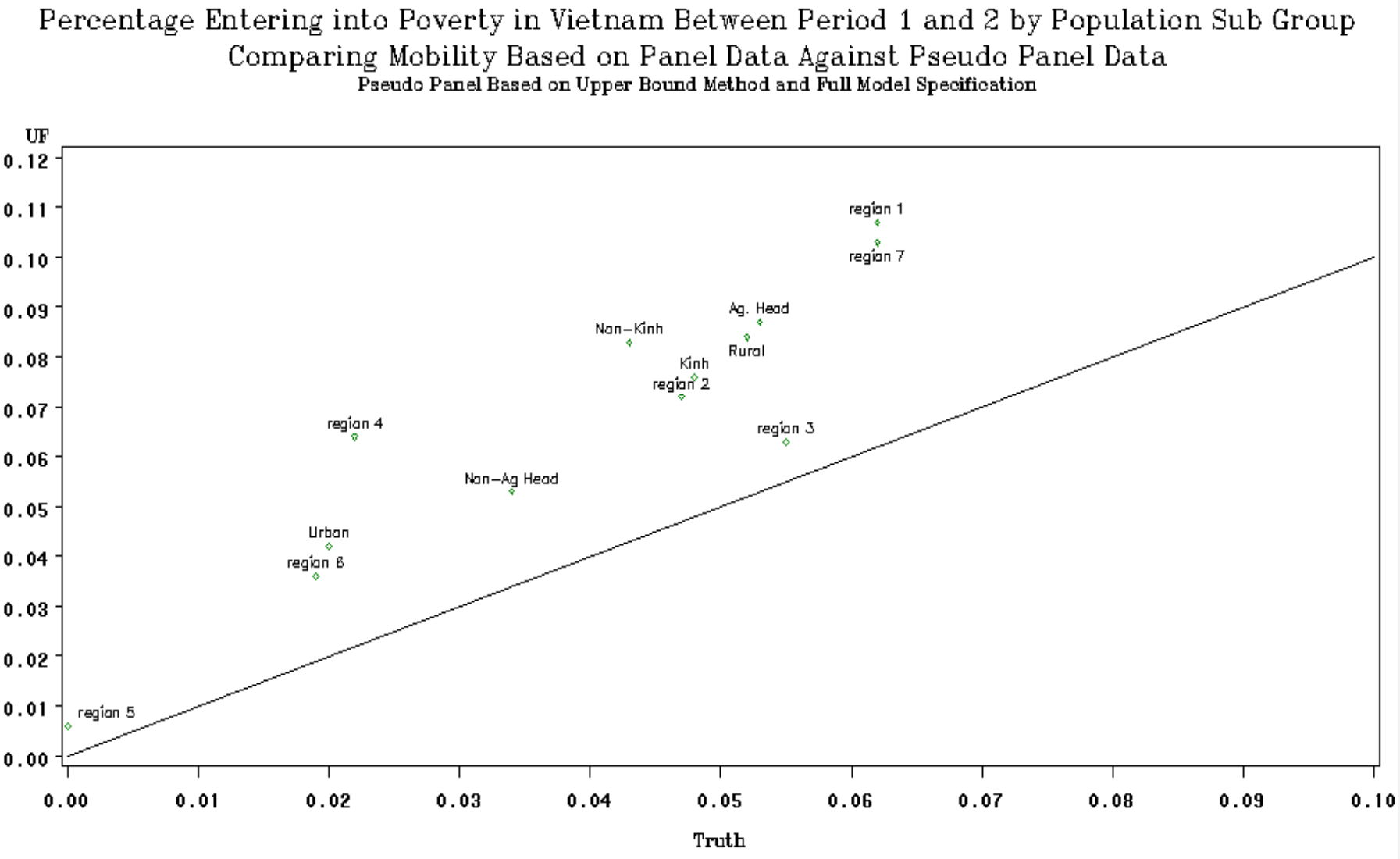
DATA & EVIDENCE deep TO END EXTREME POVERTY

Figure 1: Exiting Poverty



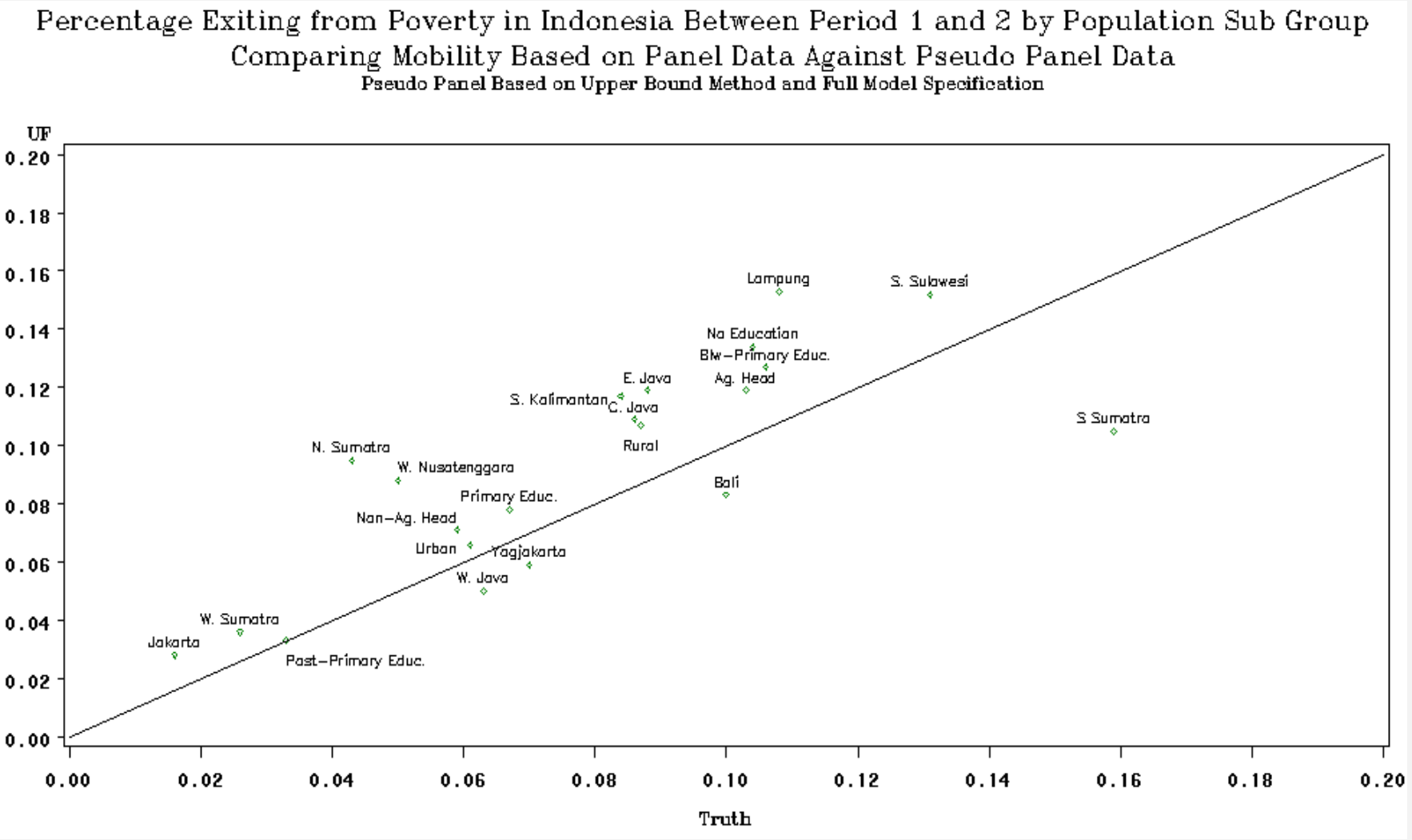
DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Figure 2: Entering Poverty



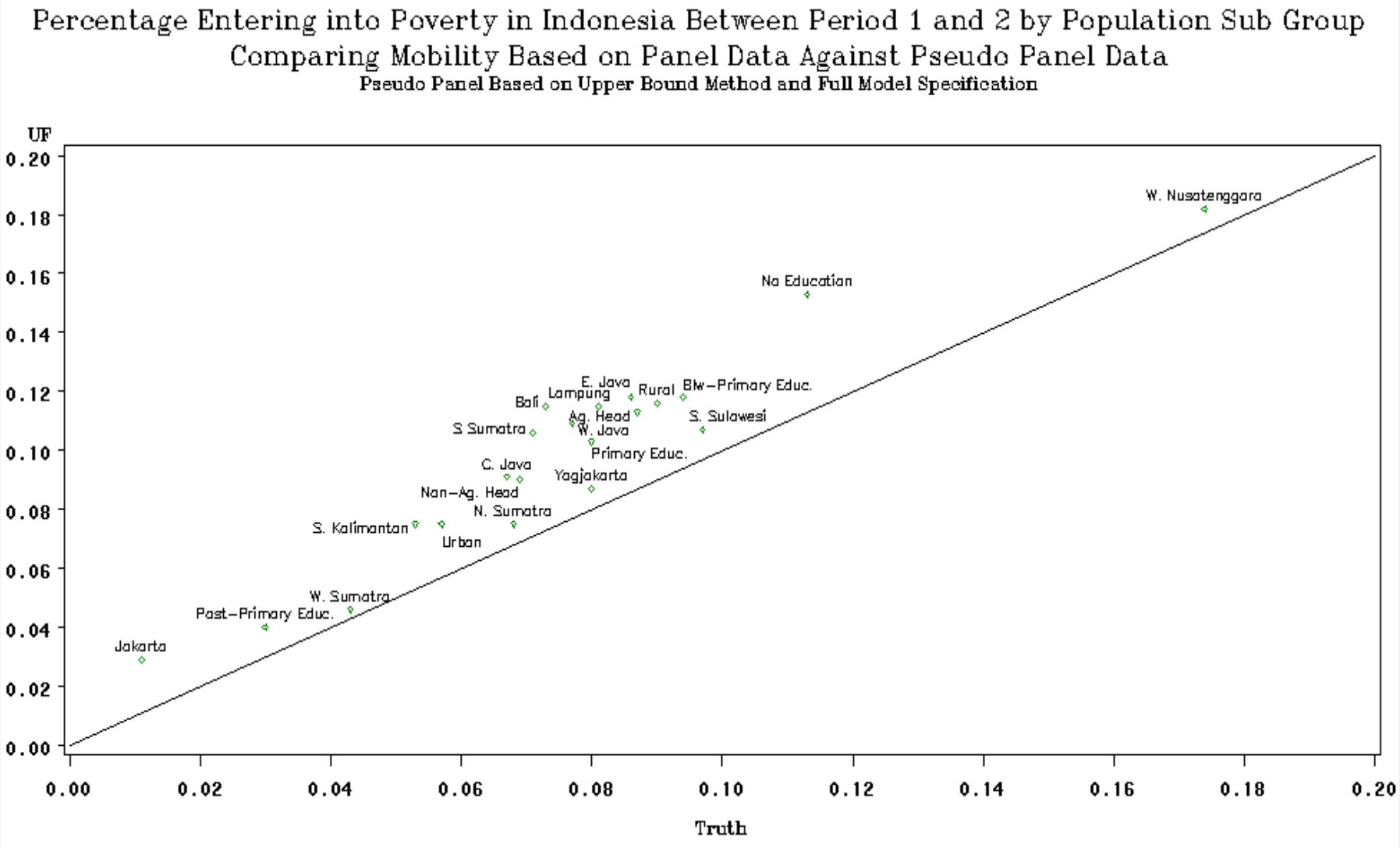
DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Figure 3: Exiting Poverty: Indonesia



DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Figure 4: Entering Poverty - Indonesia



DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Do we get the same rankings?

Correlation of Panel-Based Probability versus Pseudo-Panel Based Probability of Exiting Poverty Across Different Population Sub-groups				
	Vietnam (13 Population Sub-Groups)		Indonesia (21 Population Sub-Groups)	
	Pearson Correlation	Spearman Rank Correlation	Pearson Correlation	Spearman Rank Correlation
Lower bound method, basic model	0.399	0.480	0.331	0.388
Lower bound method, full model	0.650	0.770	0.189	0.170
Upper bound method, basic model	0.271	0.050	0.491	0.557
Upper bound method, full model	0.886	0.833	0.787	0.824

DATA & EVIDENCE deep TO END EXTREME POVERTY

Rankings

- Seems to give relatively similar rankings as genuine panel
- Particularly when upper bound full model is used
- When basic model is used, much less robust conclusions can be drawn

DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Imposing parametric assumptions

- Assume ε_{i1} and ε_{i2} have a bivariate normal distribution
  - With standard deviations $\sigma_{\varepsilon 1}$ and $\sigma_{\varepsilon 2}$
- ρ is the correlation coefficient between these two error terms (assumed positive)
- Bounds approach assumes ρ is either 1 or 0
- True value likely lies between these two values
- Applying a parametric approach one can “plug” in a value of ρ

DATA & EVIDENCE deep TO END EXTREME POVERTY

Estimating ρ

Dang et al (2014) show:

$$\rho = \frac{\rho_{y1y2}\sqrt{var(y_1)var(y_2)} - \beta_1' var(x)\beta_2}{\sigma_{\varepsilon 1}\sigma_{\varepsilon 2}}$$

Where ρ_{y1y2} is correlation of consumption between the two surveys.

DATA & EVIDENCE
deep
TO END EXTREME POVERTY

then

$$\mathrm{P}(y_{i1} < z_1 and\ y_{i2} > z_2 = \ \Phi_2(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{\varepsilon 1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{\varepsilon 2}}, -\rho)$$

Where $\Phi_2(.)$ is the bivariate normal cumulative distribution function (cdf)

DATA & EVIDENCE
deep
TO END EXTREME POVERTY

Imposing Parametric Assumptions

- Dang et al (2014) explore existing panel surveys to get a sense of plausible values for ρ.
  - Evidence from Indonesia, Vietnam, Chile, Nicaragua and Peru suggest ρ might range between 0.5-0.8
  - Evidence from EU-SILC data suggest ρ might range between 0.3-0.8.
- Dang and Lanjouw (2023) explore feasibility of proxying ρ_{y1y2} by calculating cohort-level correlation of consumption.
  - Show that this works well in validation studies for set of countries.
  - Herault and Jenkins (2019) find method to be problematic in UK and Australia data.
  - Difficult to assess *when* cohort-based approach will or will not work.

DATA & EVIDENCE deep TO END EXTREME POVERTY

Validation

- Draw on Vietnam VHLSS for 2006, 2008
- Apply method to cross-section component (approx 3600 hhs)
- Compare to “truth” in rotating panel sub-component (approx 4100 hhs)

DATA & EVIDENCE deep TO END EXTREME POVERTY

Vietnam Results

Poverty status 2006--> 2008	DLLM lower bound	Parametric lower Model 1	Truth	Parametric upper Model 1	DLLM upper bound	Parametric estimates ρ= 0.5	true ρ (0.6)	ρ= 0.8
Poor, Poor	12.1	12.5	7.6	4.8	6.1	7.1	7.7	9.2
			(0.4)					
Poor, Nonpoor	1.6	1.2	6.3	8.8	8.9	6.5	5.9	4.4
			(0.4)					
Nonpoor, Poor	0.2	0.2	4.3	7.9	8.1	5.6	5.0	3.4
			(0.3)					
Nonpoor, Nonpoor	86.1	86.1	81.9	78.5	76.8	80.8	81.4	82.9
			(0.6)					
N	3557	3557	4088	3557	3557	3557	3557	3557

DATA & EVIDENCE deep TO END EXTREME POVERTY

Conclusions/where to from here?

- Genuine panel data is rare, and even the best panels often smaller in scale & frequency than cross-sectional surveys.
- E.g. Indonesia IFLS is one of, if not the, best developing country panel out there
  - But not nationally representative
  - Sample size of around 7000 households
  - Low frequency
  - Vs SUSENAS
    - Annual, nationally representative (and representative at district level), around 200,000 households!

➢ Policymakers and academics do care about movements into and out of poverty- would be nice to be able to say something regularly and in most countries, even if what we can say is relatively basic.

➢ More experience needed:

- ➢ Rongen et al (2023), Rongen and Lanjouw (2024) apply method to Malaysia
- ➢ Dang and Lanjouw (2023) explore methodological refinements
- ➢ Etc.

DATA & EVIDENCE deep TO END EXTREME POVERTY

Conclusions/where to?

- We've provided a method of using repeated cross-sections to obtain bounds on movements into and out of poverty
  - Validated this against genuine panel data
  - Found the bounds can be narrow enough in practice to be useful
- However, method works best when full range of variables used, some of which are not typically asked retrospectively in surveys
  - But no reason why they can't be – and much cheaper to add a few of these questions than field a panel

  ⇒Seems worth experimenting with inclusion of some such questions in upcoming surveys.

DATA & EVIDENCE deep TO END EXTREME POVERTY