Decomposition of inequality measures calculated from NSS micro data to identify age effects

SSER's Research Methodology Workshop on
Using Large-scale Survey Data
Takashi Kurosaki (Dec 22, 2016) kurosaki@ier.hit-u.ac.jp

This Lecture

(1) Concept of inequality measures: Axiomatic approach; popular measures; correspondence with microeconomic theory

(2) Inequality measurement using NSS-type household data: Cluster, strata, weight; inequality profile analysis

(3) Within-between decomposition of inequality: Concept, results from India

(4) Age effects in within-cohort inequality: Definition; why important

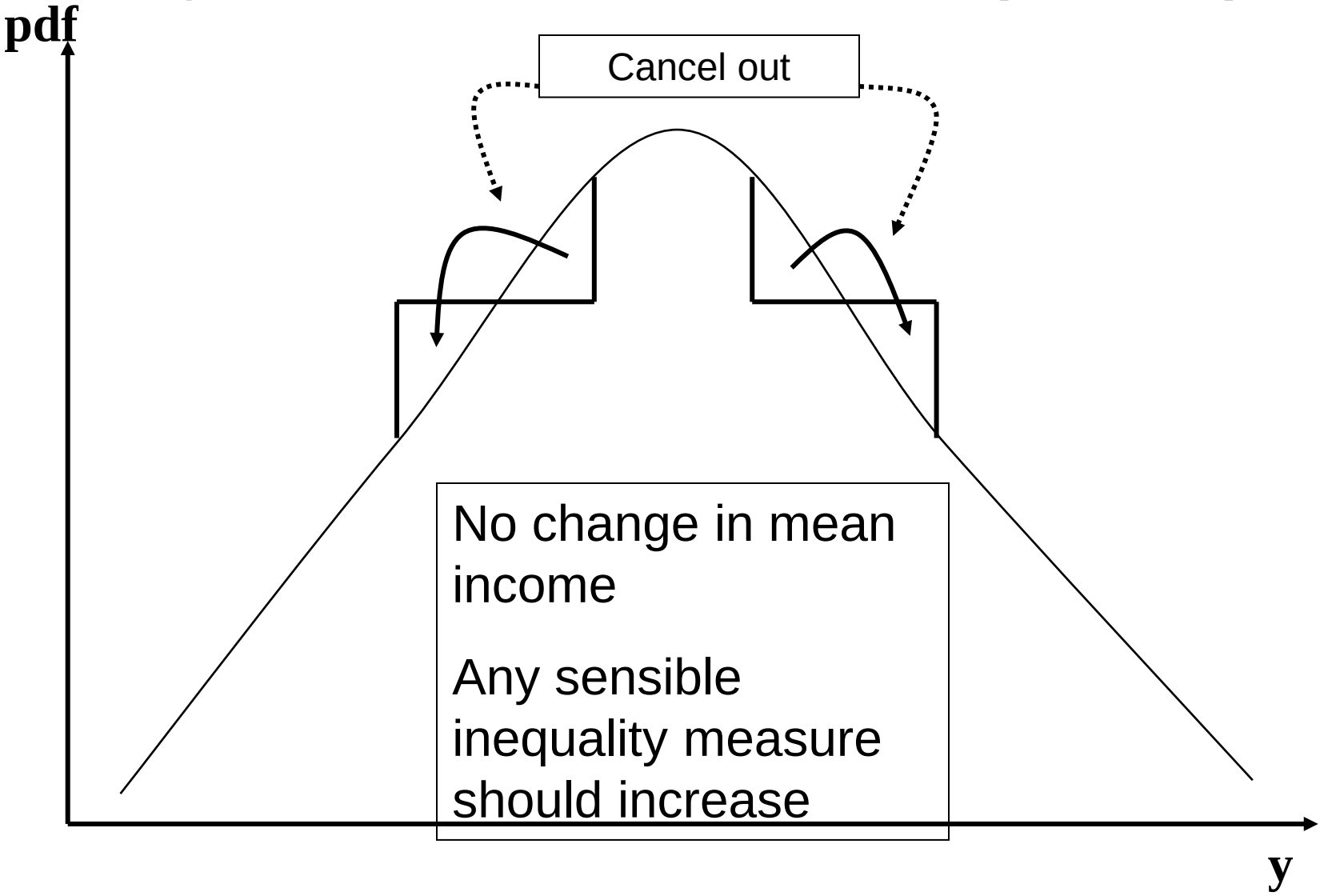
(5) Estimating age effects using Indian NSS data: Empirical model; identification of age effects; empirical results

(1) Concepts of inequality measures

Axioms to be satisfied by standard measures of inequality:

- **Anonymity or symmetry**
- **Mean independence or income homogeneity**
- **Population independence**
- **Pigou-Dalton transfer principle**
- **Subgroup decomposable**
- **etc.**

Pigou-Dalton transfer principle



Frequently-used measures of inequality satisfying the axioms
(except for subgroup decomposability)

- Gini coefficient
- Coefficient of variation
- Theil index
- Generalized entropy (GE) index
- Atkinson measure
- Log variance (a non-normalized Atkinson measure when the inequality aversion parameter is 1)

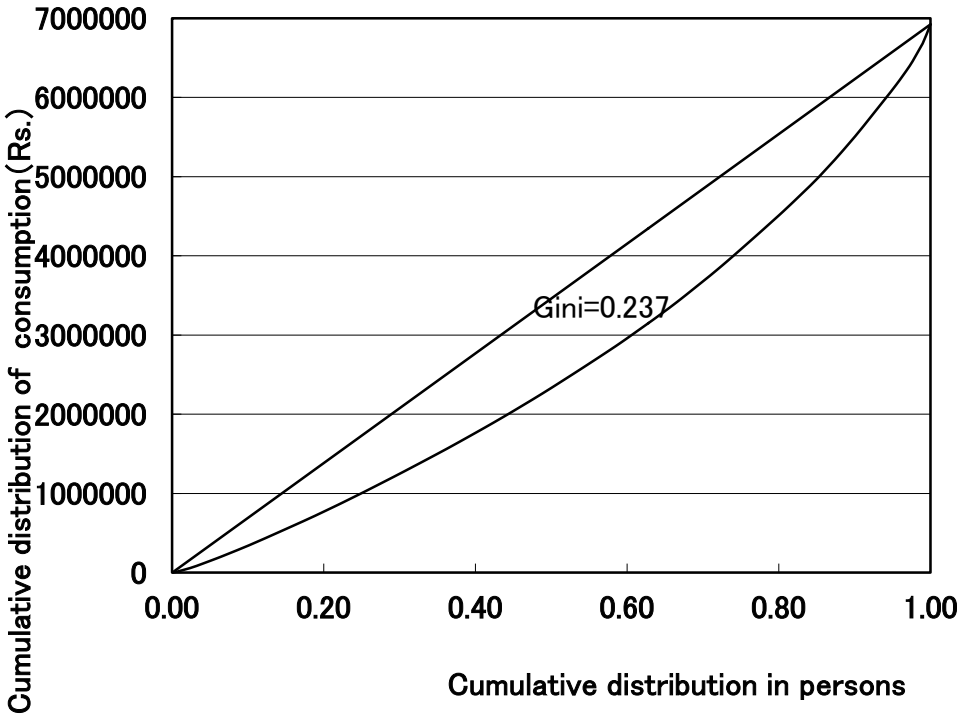
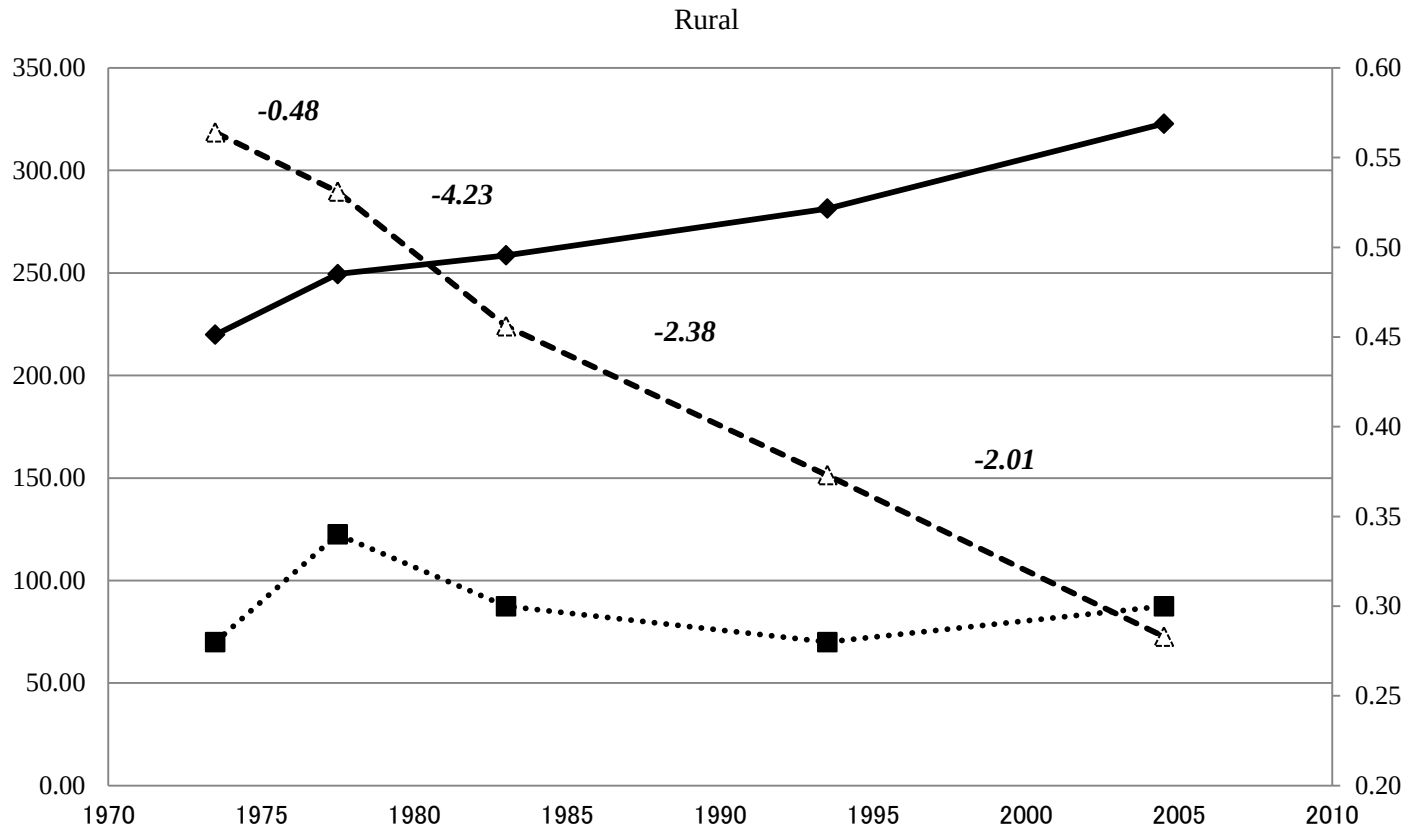
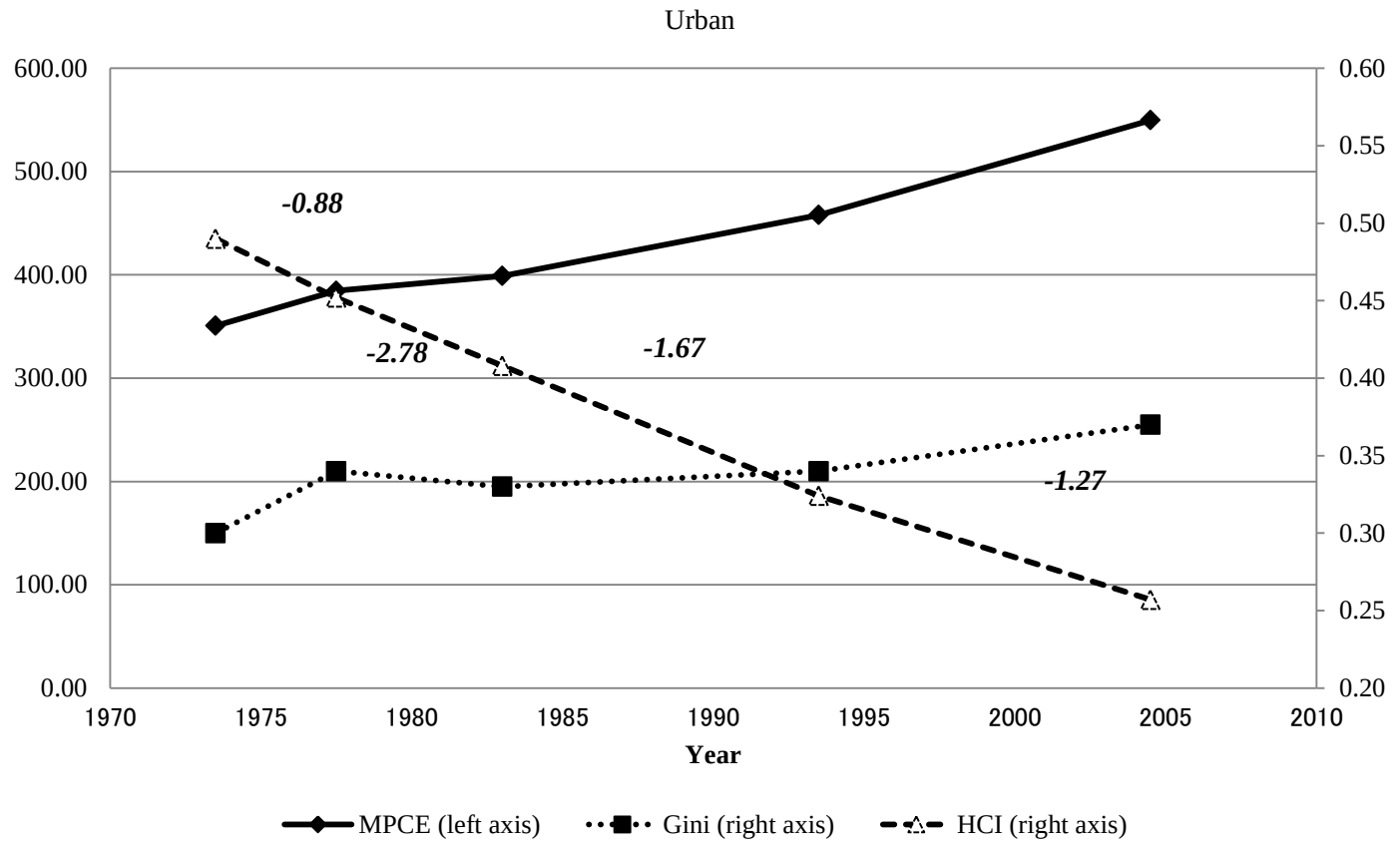


Figure 1: Trends in Growth, Inequality, and Poverty



- Original data: *MPCE* (monthly per-capita consumption expenditure) in the NSS thick and comparable rounds. Average real MPCE (solid line), Gini index (dotted line), Poverty headcount index (broken line)
- Negative numbers in italic: *Poverty elasticity to growth*
- *Inequality increased in more recent years*

Figure 1: Trends in Growth, Inequality, and Poverty



- Similar changes in urban areas
- Inequality increased in recent years (not only in India, but global phenomenon)

Advantages of Atkinson inequality measure (and thus the log variance)

- Subgroup decomposable
- Microeconomic foundation with a social welfare function in the form of CES
- 3 inequality measures of (1) GE(0), (2) Atkinson measure with the inequality aversion parameter at 1, and (3) the log variance have exactly the same information (each is a monotonic transformation of another)
- (Especially the log variance:) Easily applicable to micro data of NSS type household surveys

Atkinson inequality measure

(ε is the inequality aversion parameter)

(1) Definition:

$$A_\varepsilon(y_1,\ldots,y_N)=\begin{cases}1-\frac{1}{\mu}\left(\frac{1}{N}\sum_{i=1}^{N}y_i^{1-\varepsilon}\right)^{1/(1-\varepsilon)} & \text{for } 0\le\epsilon\neq 1\\ 1-\frac{1}{\mu}\left(\prod_{i=1}^{N}y_i\right)^{1/N} & \text{for } \varepsilon=1,\end{cases}$$

(2) Associated social welfare function:

$$\begin{aligned}W_\epsilon(y_1,\ldots,y_N) &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1-\epsilon}y_i^{1-\epsilon}, \qquad 0\le\epsilon\neq 1\\ &= \frac{1}{N}\sum_{i=1}^{N}\ln y_i, \qquad \epsilon=1\end{aligned}$$

(3) Subgroup decomposability:

$$A_\varepsilon(y_{gi}:g=1,\ldots,G,i=1,\ldots,N_g)=\sum_{g=1}^{G}w_gA_\varepsilon(y_{g1},\ldots,y_{g,N_g})+A_\varepsilon(\mu_1,\ldots,\mu_G)$$

(2) Inequality measurement using NSS-type household data <Practical tips>

(1) Use the natural log of expenditure as the measure of individual expenditure

(2) Use sample weights (multiplier) and cluster/stratefication structure to analyze the distribution of $\ln(y_i)$

(3) The mean of $\ln(y_i)$ shows the average welfare level

(4) The variance of $\ln(y_i)$ shows the inequality level

(5) (3) and (4) can be done for sub-groups

Clustering, stratification, and weighting in NSS-type household data

- **Clustering & stratification:**
  - ✓ **The groups about which representative information is required are fixed as strata (e.g., urban vs. rural; NSS regions etc.)**
  - ✓ **Sample households are randomly drawn from randomly chosen clusters (primary sampling unit: PSU, usually a village in rural areas and a ward in urban areas). The sample PSUs may be randomly chosen from the next level which is also randomly chosen (3-stage sampling)**
- **Weighting (multipliers):**
  - ✓ **The sampling probability is set higher for categories whose population number is expected to small**
  - ✓ **For example, the sampling probability is higher for wealthy villages and households than no-wealthy villages and households.**

(3) Within-between decomposition of inequality

Subgroup decomposability:

$$A_\varepsilon(y_{gi} : g = 1, \ldots, G, i = 1, \ldots, N_g) = \sum_{g=1}^{G} w_g A_\varepsilon(y_{g1}, \ldots, y_{g,N_g}) + A_\varepsilon(\mu_1, \ldots, \mu_G)$$

within-inequality **between-ineq**

This is for Atkinson measure, but can be applicable to all subgroup-decomposable measures such as GE and Theil

Between and within inequality

Calculated from NSS 2004-05 microdata, with proper weights & clusters

	Mean	Std. Err.	[95% Conf.	Interval]
MPCE				
ST	463.1923	5.721867	451.9765	474.408
SC	531.9577	4.621053	522.8997	541.0157
OBC	625.8425	4.611684	616.8029	634.8821
others	919.5001	11.33248	897.2867	941.7135

Variable	Observed	Bias	Std. Err.	[95% Conf.	Interval]
Varlogs of MPCE					
ST	.2282296	.0012636	.0056322	.2170541	.2394052
SC	.2108154	.0000549	.0049853	.2009234	.2207074
OBC	.2420121	-.000184	.0032216	.2356197	.2484046
others	.3330572	.0009146	.0048056	.3235219	.3425925

Relative importance of between-group inequality in India

- The within component shows the weighted sum of within-group inequality for all groups
- The between component shows the level of inequality in the population if everyone within the group had the same (group-average) consumption
- Results in Table 5.5 of Kurosaki (2011): Caste-based between-inequality is the largest; Regardless of the choice of decomposition axes, within-inequality is much larger than between-inequality.

	Within-group inequality		Between-group inequality	
	GE(0)	(%)	GE(0)	(%)
Level of decomposition	(2)		(3)	
Urban versus rural	0.1709	(97.1)	0.0051	(2.9)
States and UT	0.1610	(91.5)	0.0149	(8.5)
Religion	0.1728	(98.2)	0.0032	(1.8)
SC, ST, OBC, and others	0.1595	(90.6)	0.0165	(9.4)

(4) Age Effects in Within-Cohort Inequality

Example from U.S. (Storesletten et al. 2004)

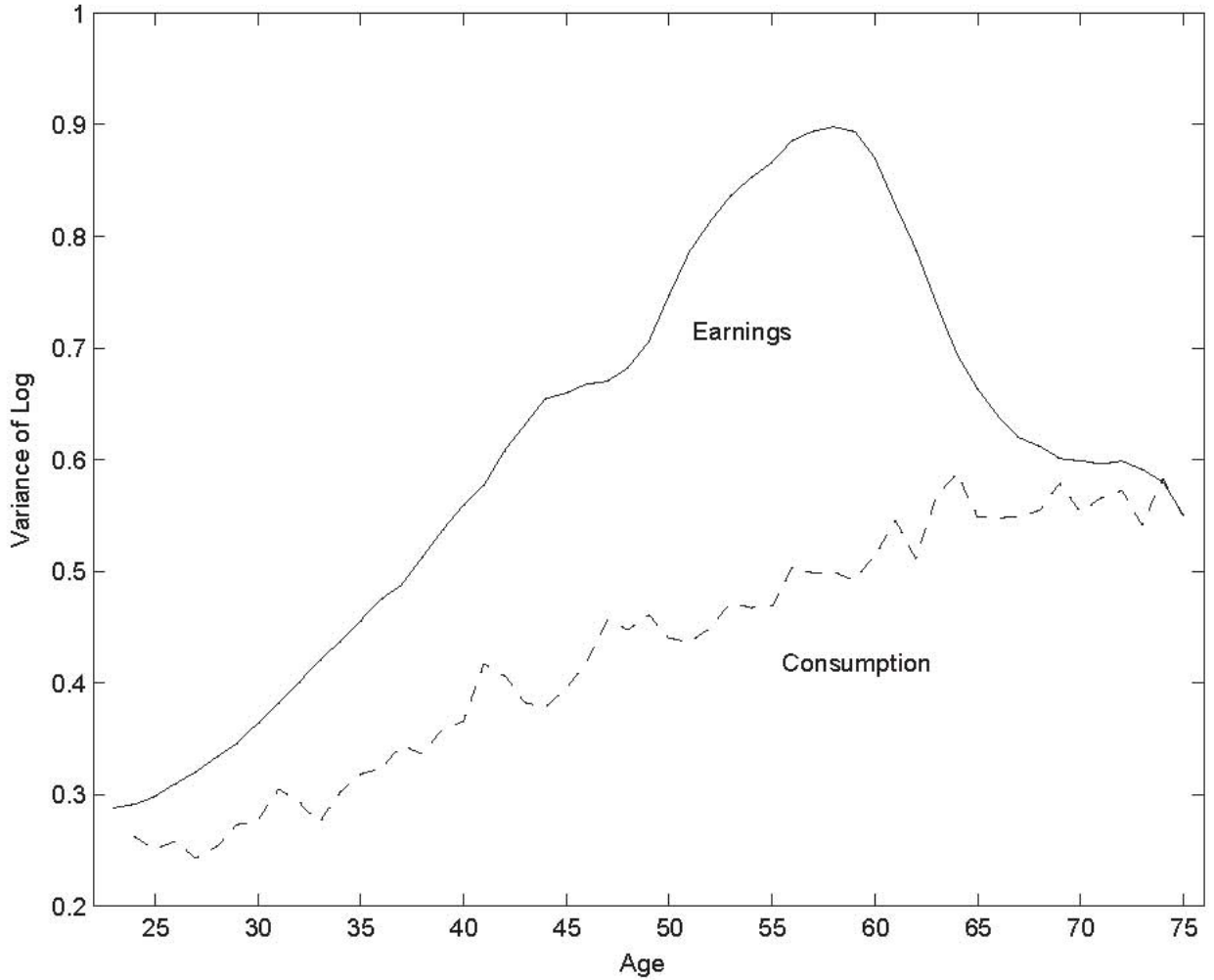


Fig. 1. The graphs represent the cross-sectional variance of the logarithm of earnings and consumption.

Example from Deaton and Paxson (1994)

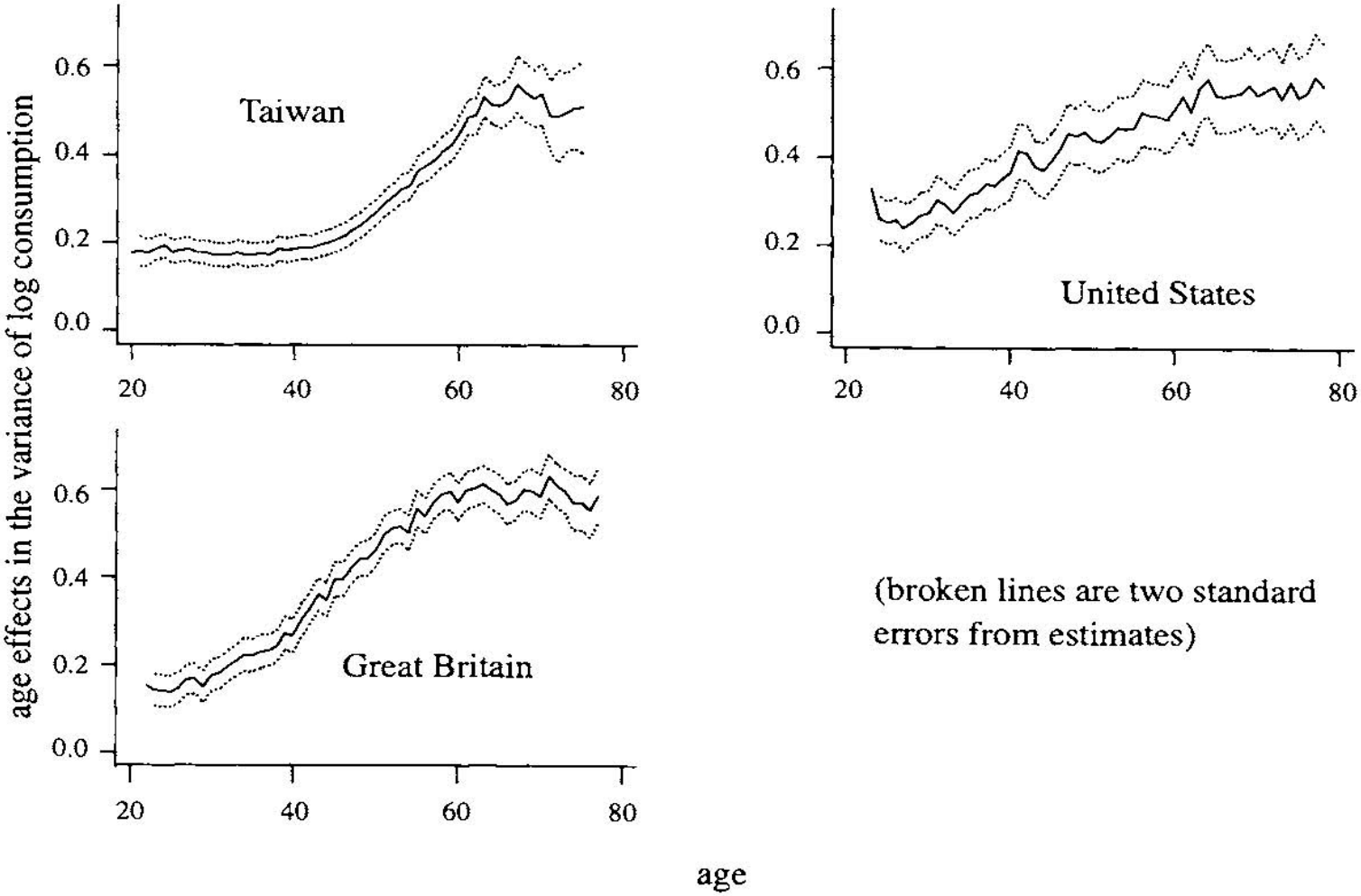


FIG. 4.—Age effects (and confidence bands) for the variance of log consumption

Increasing age effects found in developed and middle income developing countries:

- **Consumption inequality increases with age**
- **Income inequality increases with age (up to retirement age)**
- **Income inequality is larger than consumption inequality**
- **Slope of age effect is steeper on income inequality**
- **Broadly consistent with consumption smoothing (short-term and life-cycle) under permanent income hypothesis**

The figures were drawn using repeated cross-section data similar to NSS, not using panel data of households

- Each cross-section is nationally representative
- From a cross-section data in 2000 for example, using subgroup-decomposable inequality measures, we can estimate inequality level of a cohort defined by the age (which is the same as birth year or "cohort"), whose sum is equal to the within-inequality in the whole economy
- Repeat the same using a cross-section data in 2005. The within-inequality of the age 45 group from this dataset should be comparable to the within-inequality of the age 40 group in the 2000 survey. Because both were born in year 1960.
- By comparing the two inequality levels, we can infer the change that occurred to the born-1960 group between the age 40 to 45!

Research Questions

- **Is the dynamic pattern in age effects the same regardless of socio-economic background?**
- **What kind of dynamic household models can explain the dynamics of within-cohort inequality? (modeling question)**
- **How can we identify the model from data usually available in LDCs (nationally-representative repeated cross-section data; limited panel data; different types of measurement errors)? (diagnostic question)**

=> From addressing these questions, we can have detailed diagnostics of poverty and vulnerability with data typically available in LDCs, yielding rich policy implications.

(5) Estimating age effects using Indian NSS data <Empirical specification>

Equation (1) = Cohort-level model by Deaton and Paxson (1994)

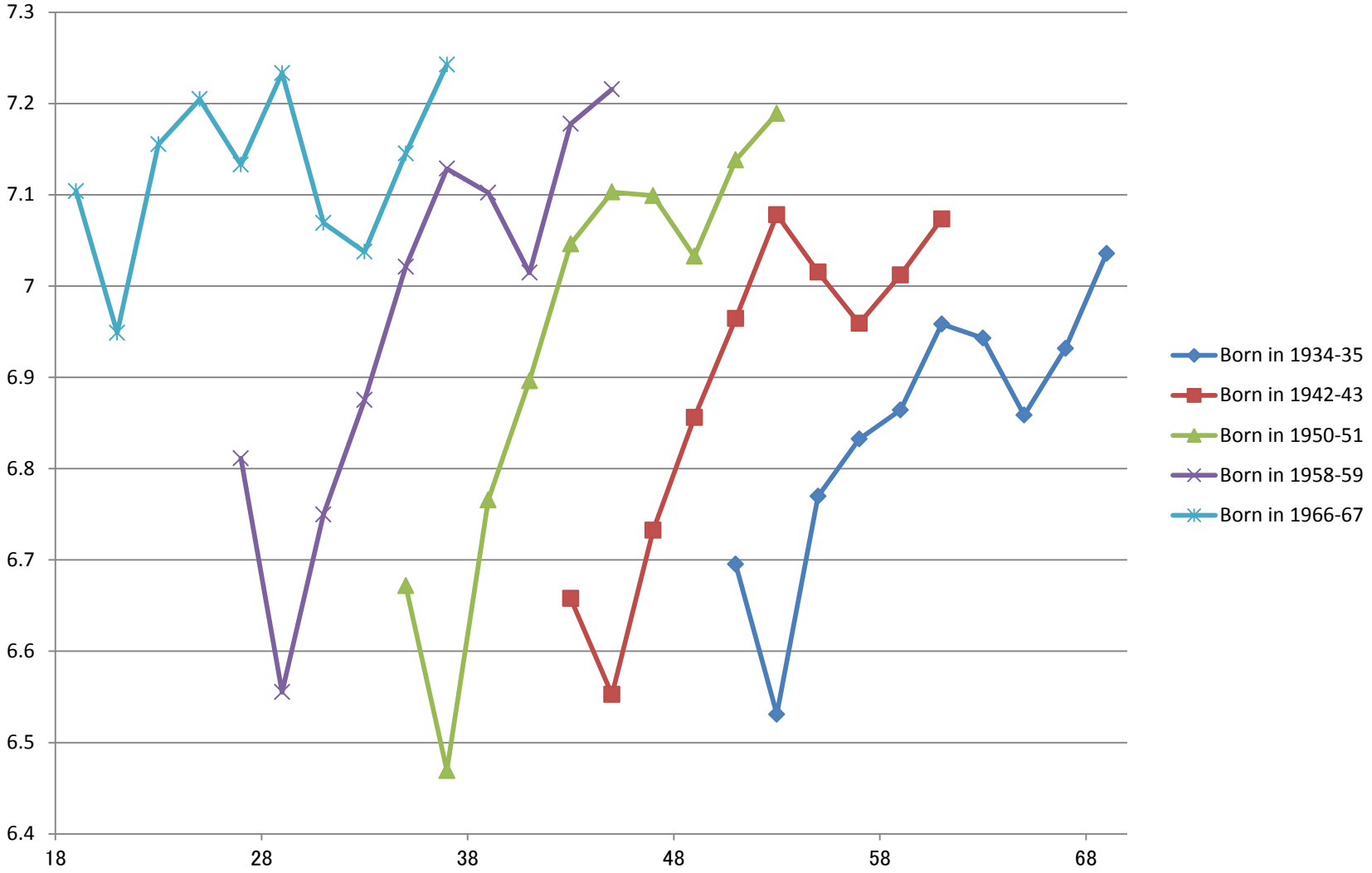
- **Dependent variable: $Ineq_{gt}$ (Variance of $\ln c_{igt}$ for *i* belonging to cohort *g*)**

  (*i*=individual hh, *g*=cohort of the comparable group, *t*=survey round)

- **Explanatory variable: Fixed effects of household head's age, Fixed effects of household head's cohort**
- **"Age-effects in log c variance" = fitted values of age fixed effects**
- **Advantage: Can be applicable to other measures of inequality**

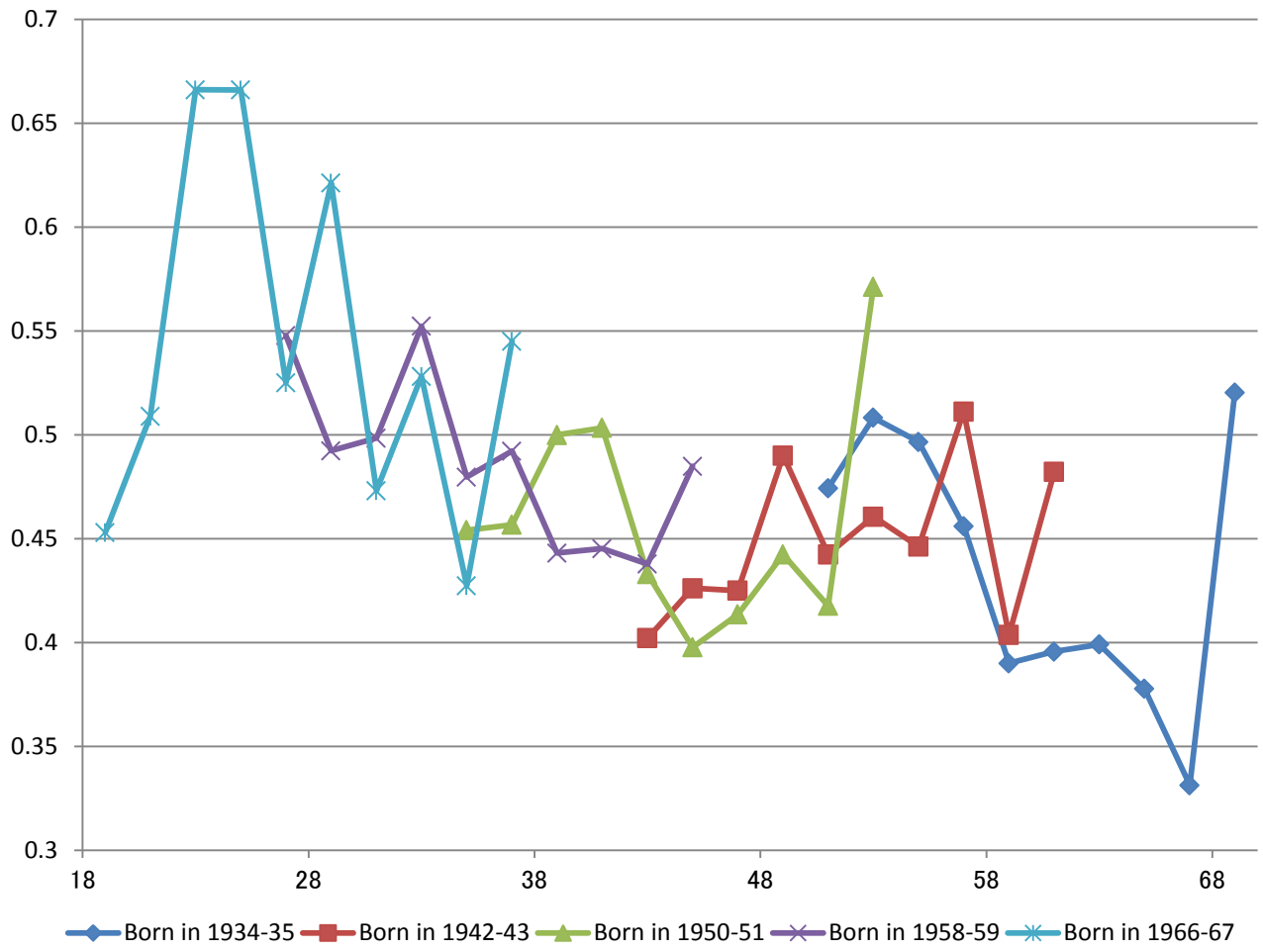
 # In effect, this is a decomposition of observed inequality into factors attributable to the age and cohort of household heads

Growth within Cohort (Thai SES data, Fig.0-2)



How does equation (1) work?

- **Data: Inequality for cohort *g* in year *t* (Fig.0-1. for Thailand)**
- **Plot the time series for each g against age (not survey year)**
- **For each cohort, the fixed effect adjusts the curve in a parallel way**
- **Thus the age fixed effects smoothly connects the shape of different cohorts.**



Extension to household-level regression model (Equation (2))

- Dependent variable: $(\ln c_{igt} - \ln c_{gt})^2$
  (*i*=individual hh, *g*=cohort of the comparable group, *t*=survey round)
- Explanatory variable: F.E. of hh head's age, F.E. of hh head's cohort, and hh-level controls X_{it} (demographic controls, region fixed effects, etc.)
- "Age-effects in logc variance" = fitted values of age fixed effects
- Advantage: Nests equation (1) when there is no X_{it} ; Flexible controls for the household-level characteristics; Gain in statistical efficiency; Flexible for (potentially complicated) survey designs.

Data from India

NSS conducted by the National Sample Survey Organization, Govt of India

- Representative at national, state, and NSS region level
- 5 rounds available: 1983 (38th NSS), 1987/88 (43rd NSS), 1993/94 (50th NSS), 1999/2000 (55th NSS), 2004/05 (61st NSS)
- Each round contains more than 100,000 hhs
- Real per-capita consumption (c_{igt}) = as in Thai.
- Advantage: Huge sample size; Very detailed information on consumption items (but no income info at all)
- Due to non-comparability, 4 rounds are used, excluding 55th (1999/2000)

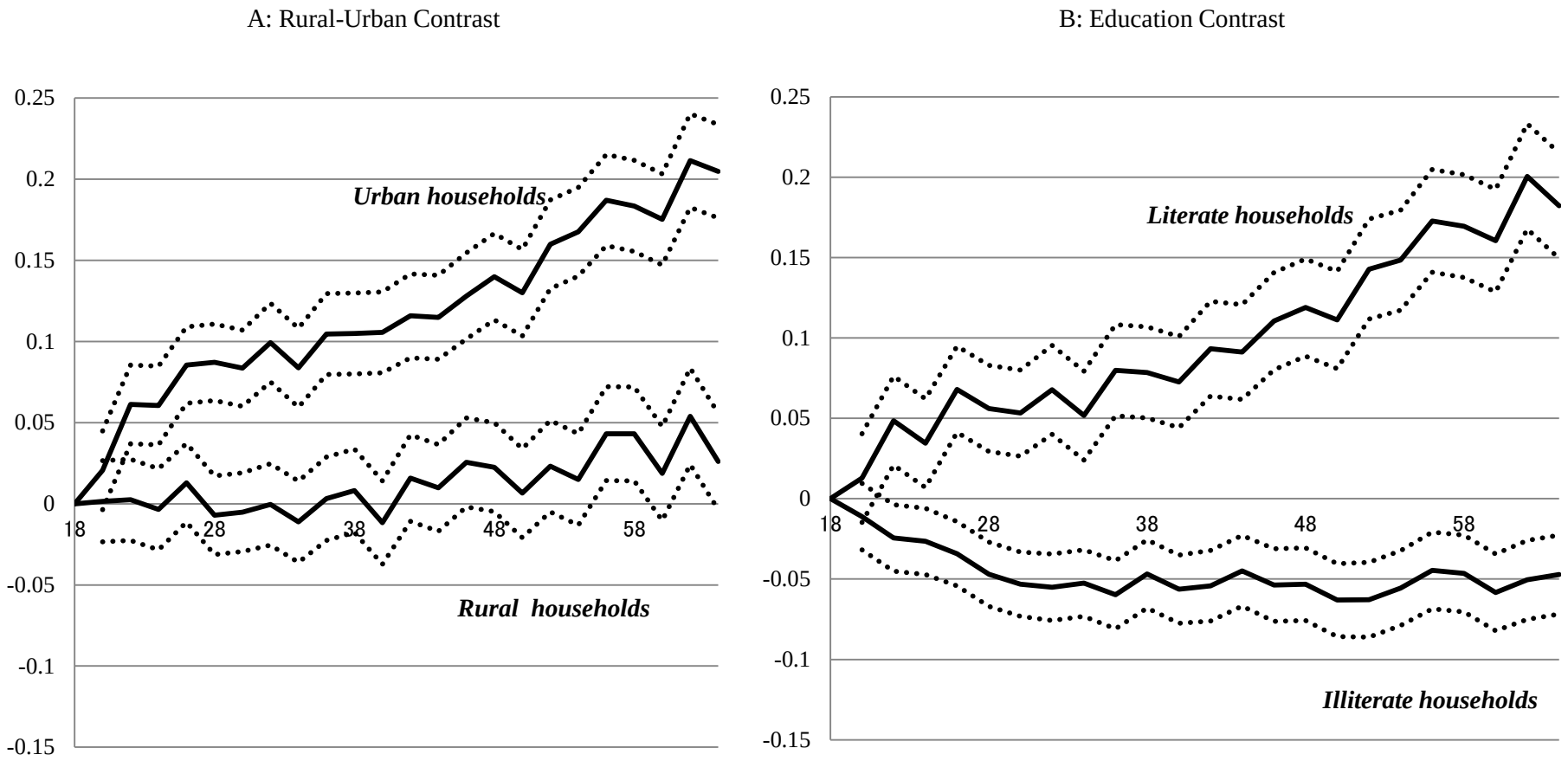


Figure 3: Age Effects in Within-Cohort Consumption Inequality, India

- **As in developed countries, urban and literate households show increasing age effects in within-cohort inequality.**
- **Among rural households, age effects are almost flat. *Among illiterate households, age effects are decreasing!***

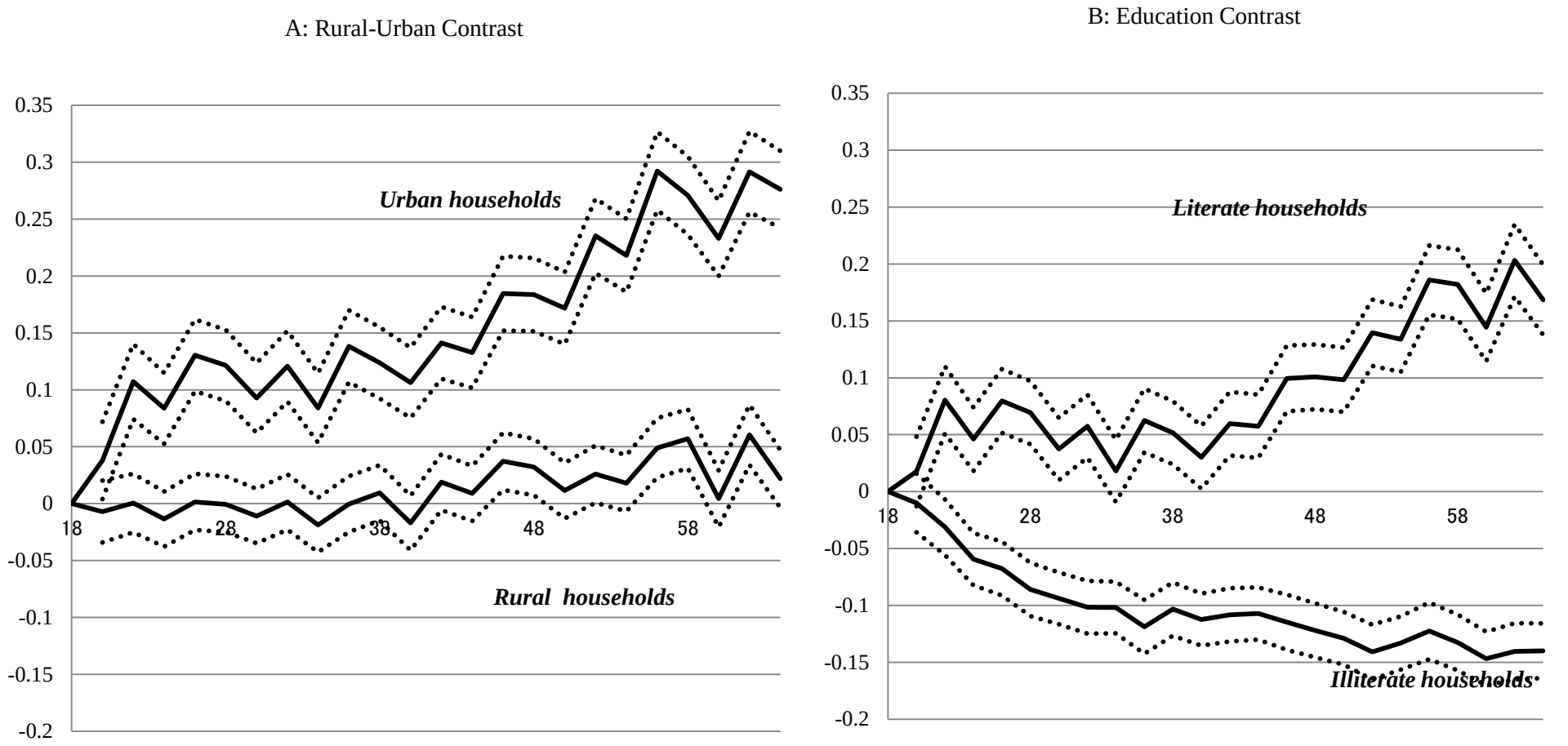


Figure 4: Age Effects (based on micro-data)

- **More statistically-efficient model based on micro-data shows similar results. Therefore, *the decreasing age effects in within-cohort inequality among illiterate households are robust.* (Similar contrasts found in Thailand and Pakistan: see Kurosaki et al. 2010)**

Why these findings are important?

- "Inequality is decreasing with age" is good? Not necessarily so. It can indicate the inefficiency of intertemporal resource allocation among less educated households, contributing to more poverty!
- From the same NSS microdata, the average MPCE among rural/illiterate households did not show increasing age effects. Thus, rural and illiterate households are left out from dynamic change in Indian economy, trapped in persistent poverty with substantial variability of consumption due to idiosyncratic transient shocks.

From a viewpoint of microeconomic theories, a flat age effects pattern is easy to be explained. It is a challenge to explain the decreasing age effects.

Concluding remarks about inequality in India using NSS microdata

- Economic inequality is increasing in India resulting in persistent poverty.
- The spatial and social disparity is huge but within-group inequality is by far the larger.
- Illiterate households are more likely to constrained in intertemporal resource allocation, resulting in an abnormal pattern of decreasing within-cohort inequality across age.
- These analysis are based on MPCE microdata in NSS. To verify the interpretation above, we need more detailed and panel information on income and assets.

=> It is of great importance to link detailed village-level case studies and NSS microdata through intermediate kind of datasets (i.e., panel data of household income/consumption/asset, with reasonably large *N* and *T*).

Summary of this lecture

(1) Concept of inequality measures: Atkinson measure, esp. log variance is recommended

(2) Inequality measurement using NSS-type household data: Be sure to use ln(MPCE) including imputed values; adjust for cluster, strata, and weight

(3) Within-between decomposition of inequality: Traditional decomposition; within inequality important in India than usually thought

(4) Age effects in within-cohort inequality: Can be estimated from repeated cross-section data and informative about the source of inequality

(5) Estimating age effects using Indian NSS data: Developed countries pattern applicable only to urban and educated households; rural and non-educated households may be trapped in low-level stagnation in there welfare

References

- Atkinson, Anthony B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory* 2: 244-263.
- ----- (2014), *Inequality: What Can Be Done?* Harvard, Massachusetts: Harvard University Press.
- Deaton, Angus and Christina Paxson (1994), "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102(3): 437-467.
- Kurosaki, Takashi, Kyosuke Kurita, and Ethan Ligon (2010), "Intertemporal Choice and Inequality in Low-Income Countries: Evidence from Thailand, Pakistan, and India," Paper presented at the Northeast Universities Development Consortium (NEUDC) Conference, MIT, November 6-7, 2010.
- Kurosaki, Takashi (2011), "Economic Inequality in South Asia," *Routledge Handbook of South Asian Economics*, edited by Raghbendra Jha, Oxon, UK: Routledge, pp.61-75.
- Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron (2004), "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 51(3): 609-633.