Measurement error – at the top, bottom, and middle

Stephen P. Jenkins (LSE)
Email: s.jenkins@lse.ac.uk

Presidential Lecture
ECINEQ-2021

LSE

What the lecture is about …

1. Brief overview of the many different types of measurement error …
2. … and then argue that different error problems have been emphasised at the top, bottom, and middle of the distribution
3. One example each about error at the top and at the bottom
4. Extended example about the middle: errors in employment earnings
5. Conclusions

Themes:

- Advantages of combining information from survey and administrative data – in various ways – for diagnosing and addressing error-related problems
- Importance of drawing out implications for inequality and poverty (and policies for statistical monitoring and redistribution per se)

Motivation (and acknowledgement)

I chose this topic because I think it's important and because, to paraphrase Molière's M. Jourdain, …

> « *Je parle d'*erreur de mesure *depuis plus de quarante ans sans le savoir et je vous suis très reconnaissant de m'avoir appris cela* » †

† : "I have been speaking in [prose] for more than forty years without knowing it and I am much obliged to you for having taught me that", *Le Bourgeois Gentilhomme*, 1670

LSE

Tony Atkinson and 'total error'

- On the importance of error(s) of various kinds and their implication, and an opportunity to salute an expert who worked on data and error a lot …
- … here's Tony Atkinson:

"… the Report stresses that any estimate—of level or of change—is surrounded by a margin of error. This is often lost from sight in public pronouncements, and it is important to convey to policy makers and other users that they are operating with numbers about which there is considerable uncertainty. Indeed, this could take us back to the position that nothing concrete can be said. However, the more positive response adopted here is the "total error" approach, which seeks to identify different potential sources of error and to attach an indication of their possible size."

Source: Atkinson (2017), *Monitoring Global Poverty*, pp. xv–xvi

LSE

Measurement error has many facets (1/2)

1. Construct validity: are we using the correct concept?
   - E.g. absolute versus relative inequality, absolute versus relative mobility, absolute versus relative poverty lines; family versus household as income-sharing unit; income versus expenditure as the 'welfare' measure; Canberra versus DINA 'income' measure
2. Comparability and coherence
   - E.g. different modes used across same collection instrument, or changes in mode over time
   - E.g. 'current' income versus 'annual' income measures
   - E.g. lack of data harmonisation more generally

Headings adapted from: Office for National Statistics (2013), *Guidelines for Measuring Statistical Output Quality*, version 4.1, which also includes discussion of other aspects of what counts as 'fit for purpose' for statistical estimates

Measurement error has many facets (2/2)

3. Accuracy and reliability

a) Sampling error
   - Standard errors and all that

b) Coverage error – does the data collection design cover all the population of interest?
   - E.g. individuals in private households versus all individuals; data on tax-payers versus all adults (i.e., also including non-taxpayers)

c) Non-response error: incomplete data
   - Unit and item non-response

d) Measurement error per se: inaccurate data
   - Data collected from respondents are not the 'true' values

e) Processing error: e.g. at data capture, coding, editing stages

f) Model specification error
   - E.g. models used for derivation of imputations, weights, or equivalence scales; parametric approaches to Inequality of Opportunity measurement; parametric models of whole income distribution (GB2 versus S-M?), etc.
   - E.g. error distribution modelling reported later

Measurement error and surveys: different emphases at **top**, bottom, and middle

Location in distribution	Type of error emphasised recently	Implications re survey	How addressed	Selected recent empirical applications
Top		Top income shares under-estimated	Start with admin data	• Atkinson, Piketty, Saez (etc.), now WID.world and DiNA
	Unit non-response in household surveys, income-related	Inequality under-estimated? (LCs cross)	Modify the survey weights (using model!)	• Korinek, Mistiaen, Ravallion
	Item non-response: income-related under-reporting (esp. of capital and business income) in household surveys	Inequality under-estimated	Use administrative data to correct top survey incomes using an imputation method (using a model!)	• Jenkins (Pareto II) • Burkhauser et al. (see example) • UK Department for Work and Pensions & Office for National Statistics (see example)

Three parts: (i) diagnosing the problem and (ii) addressing the problems (combining survey and administrative data); (iii) drawing out implications for statistical monitoring and/or redistributive policy

Top income under-coverage: UK example

- Diagnosis of problem via benchmarking survey against income tax data
- Survey under-coverage by income group: mean(survey) ÷ mean(tax) < 1

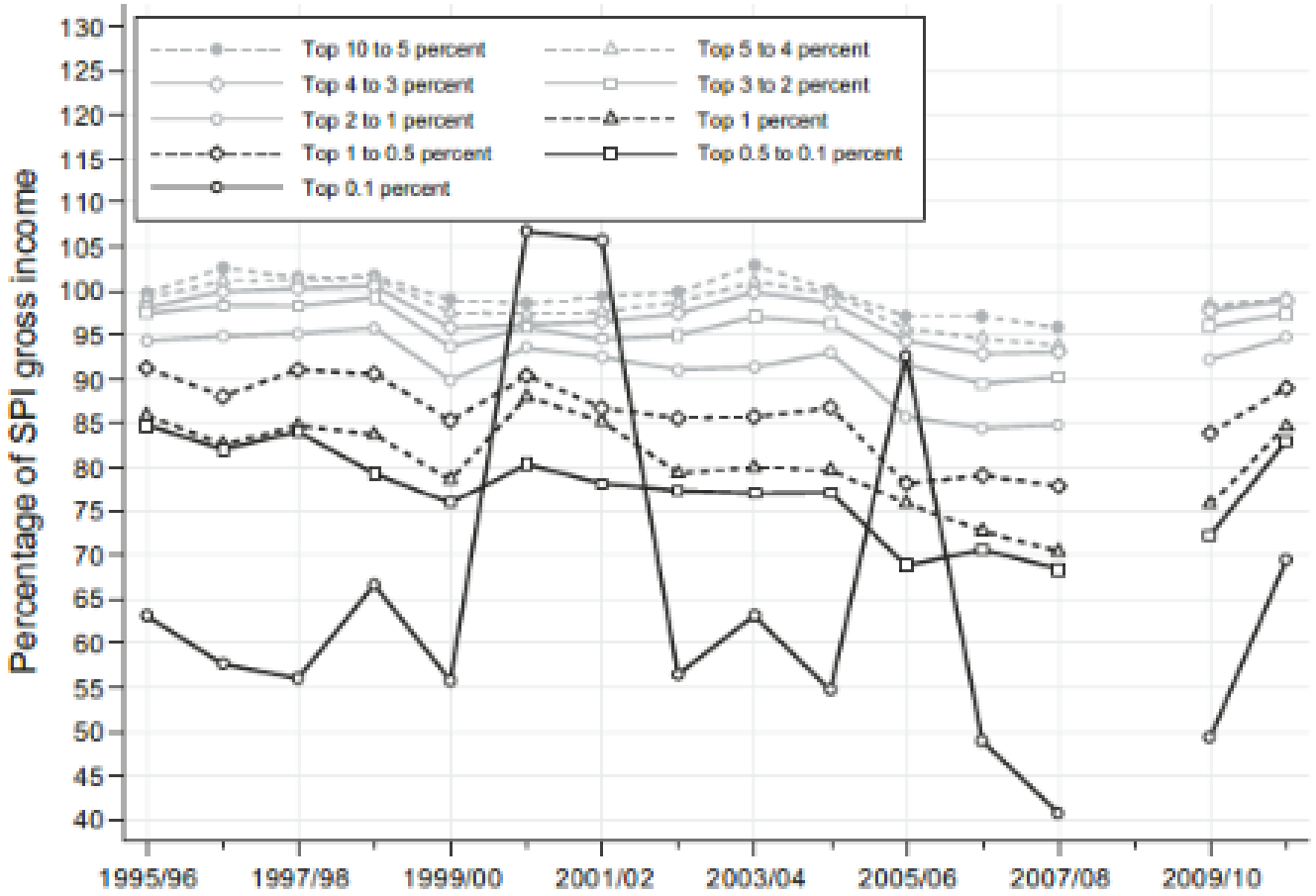


Fig. 3. Survey under-coverage of top incomes, by income group and year

Source: Burkhauser, Hérault, Jenkins & Wilkins, *OEP* 2018. SPI: **S**urvey of **P**ersonal **I**ncomes (income tax data)

Top income under-coverage: UK example

- Addressing the problem: replace very top survey incomes with cell-mean imputations from tax data, by top-quantile group: see chart
  - HBAI: unadjusted data. HBAI-SPI: DWP's HBAI adjustment.
  - HBAI2: our adjustment, going further down from top, and using more cell means

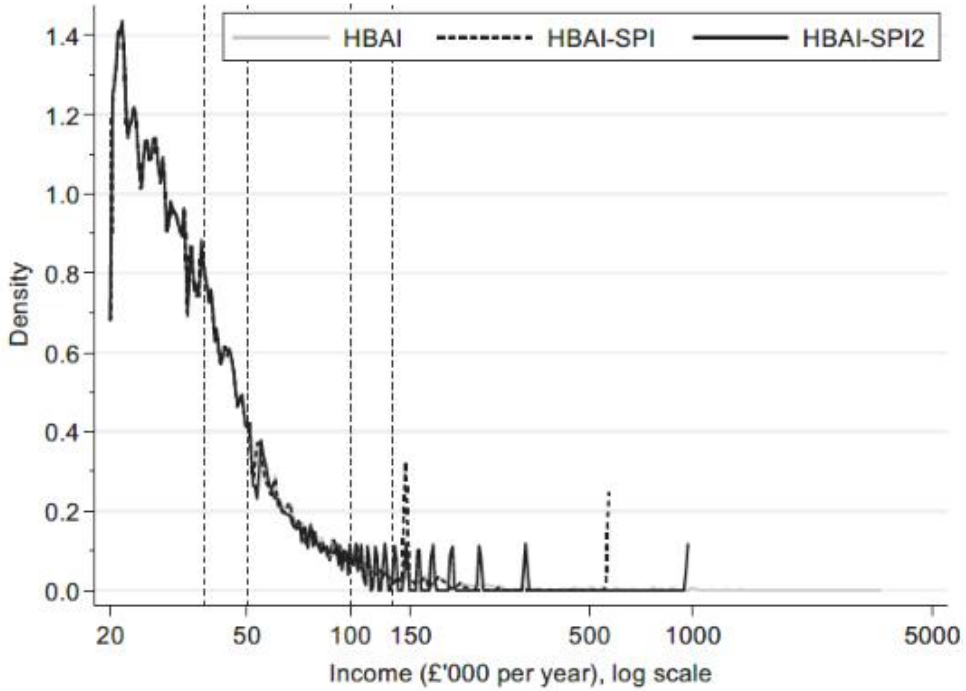


Fig. 4. Density estimates of the upper tails of the HBAI, HBAI-SPI, and HBAI-SPI2 income distributions, 2010/11

Notes: We calculate kernel density estimates for the distribution of log income using observations with income >£20,000 per year, using an Epanechnikov kernel and bandwidth of 0.008. The dashed vertical lines show the 90th, 95th, 99th, and 99.5th percentiles in the HBAI data. See Section 2 and Fig. 2 for explanations of the data series and acronyms.

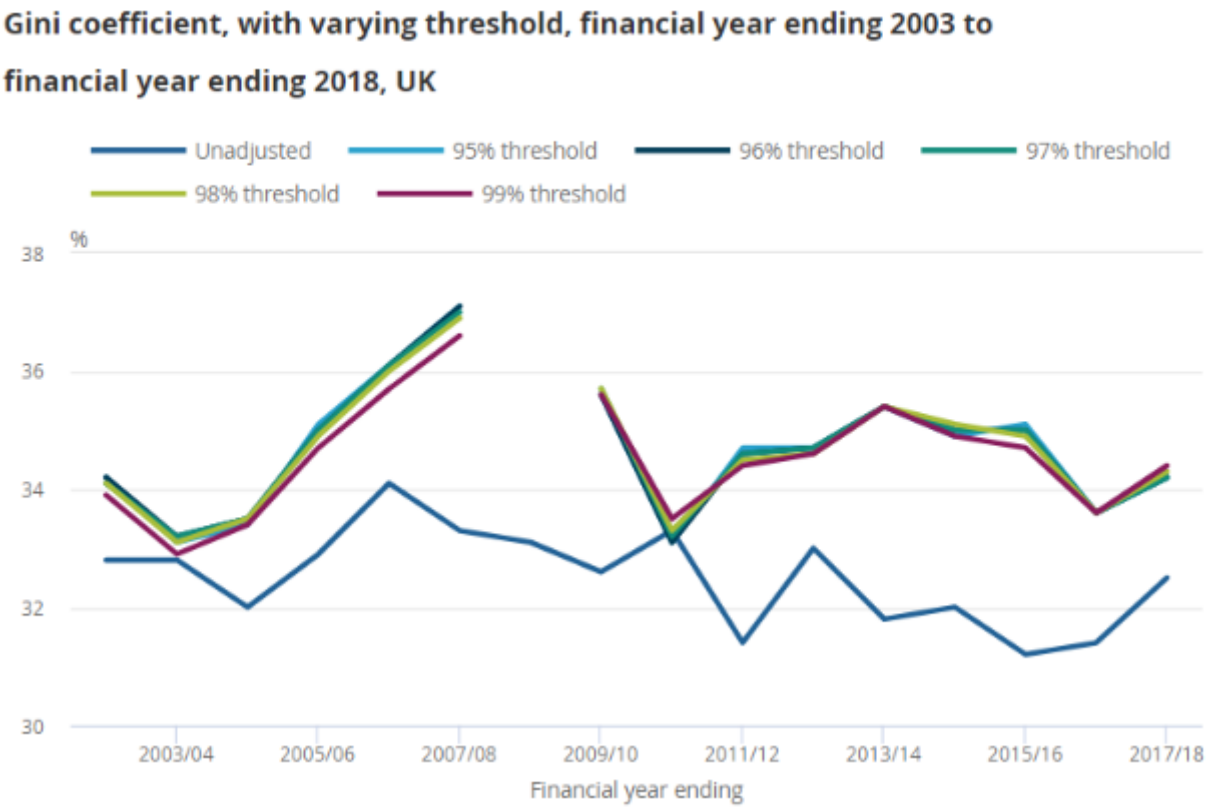
Source: Authors' calculations using FRS, HBAI, SPI, and WID data.

Source: Burkhauser, Hérault, Jenkins & Wilkins, *OEP* 2018

Top income under-coverage: UK example

- Implications for inequality: ONS's new official series is based on Burkhauser et al.'s methods
  - Main effect is to raise Gini level each year by 2–3 ppt, with small differences in trends
  - Effect on top-sensitive indices (including top 1% share) more dramatic

Source: Office for National Statistics (2019), Figure 3. See also latest ONS (2021) series here, with later years and slightly revised estimates



Measurement error and surveys: different emphases at top, **bottom**, and middle

Location in distribution	Type of error emphasised recently	Implications re survey	How addressed	Selected recent empirical applications
Bottom	Systematic under-reporting (of cash benefits) in household surveys	Poverty over-estimated	Use administrative data (on cash benefits) about aggregate gap, plus modelling	• USA: Meyer et al. (USA) • UK: Brewer et al.; Corlett (see example)

Three parts: (i) diagnosing the problem and (ii) addressing the problems (combining survey and administrative data); (iii) drawing out implications for statistical monitoring and/or redistributive policy

Bottom income under-coverage: UK example

- Diagnosis via comparisons of spending against income (for the same set of households): the ‘tick’ (‘check’) pattern …

Source: Brewer, Ethridge, and O’Dea, Economic Journal 2017, local kernel-weighted median regression lines

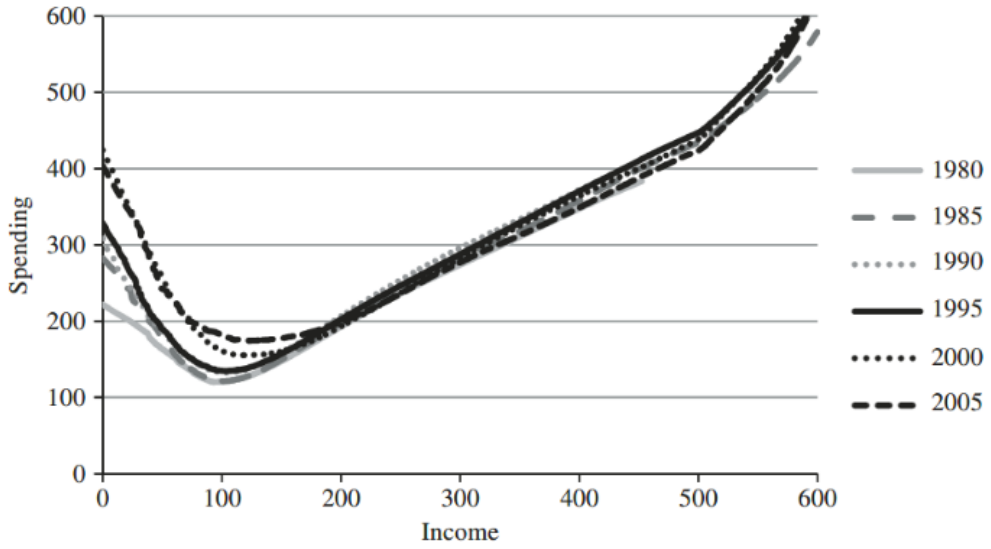


Fig. 5. *Median Expenditure by Period (Five Year Averages)*
Notes. Data are for the five years from the year shown in the legend. For other notes, see Figure 1.

- Most plausible explanation (Brewer et al.): under-reporting of cash benefits among very poorest households
  - Rather than: over-reporting of spending; dissaving due to inter-temporal smoothing
  - Backed by comparisons of benefit aggregates from survey and admin data

Bottom income under-coverage: UK example

- Impute missing income using information about aggregate shortfall from benefit administration data and modelling: mixture of (i) scaling up reported values; (ii) adding to recipient caseload; (iii) other imputations
- Big, policy-relevant, effects on poverty levels (less on trends)

Source: Corlett (2021) based on Corlett et al. (2018), *Living Standards Audit 2018*, Resolution Foundation

Figure 3 **Adding in the "missing" income would make a big difference**
Relative poverty, after housing costs, 2016—2017: GB

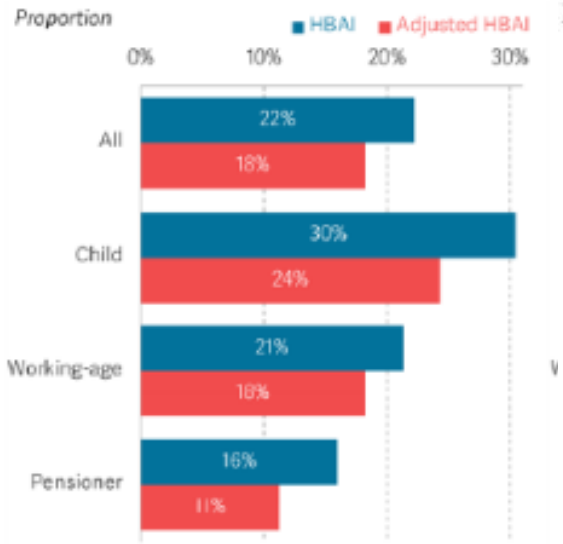
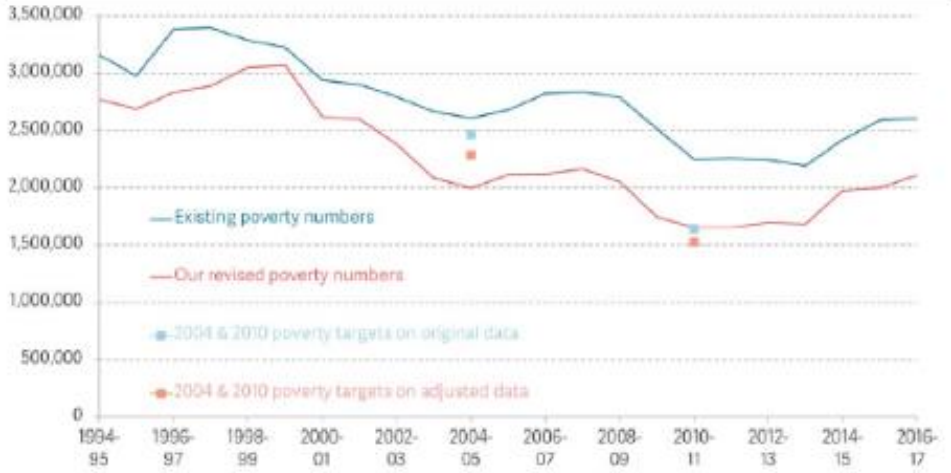


Figure 4 **With corrected poverty data, child poverty in the 2000s may have hit some of the Blair/Brown government's targets**
Number of children living in poverty before housing costs: GB



NB Adjustments also affect inequality levels: Gini lower by c. 2 ppt

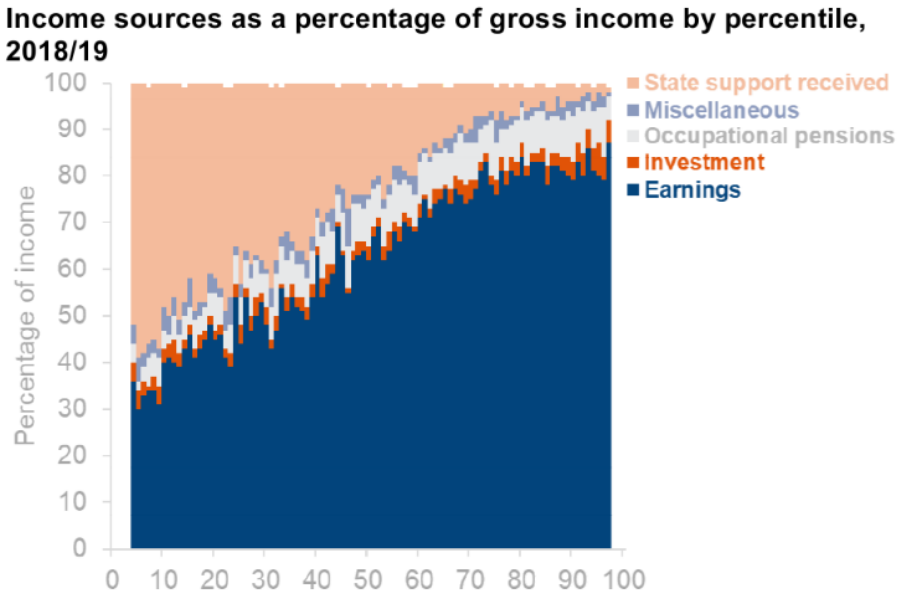
Measurement error: different emphases at top, bottom, and **middle**

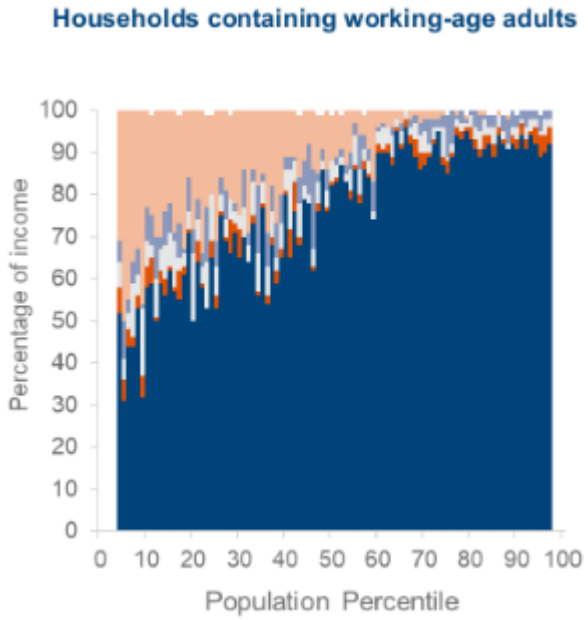
Location in distribution	Type of error emphasised recently	Implications re survey	How addressed	Selected recent empirical applications
Middle	Accuracy of survey data generally (and whether differential mis-reporting, i.e., 'mean reversion') Administrative data error?	Inequality little discussed	Use linked survey and administrative data (on employment earnings) for the same people	• Many labor economists' papers • Jenkins & Rios-Avila (see example)

Three parts: (i) diagnosing the problem and (ii) addressing the problems (combining survey and administrative data); (iii) drawing out implications for statistical monitoring and/or redistributive policy

Measurement errors in the middle

- The 'middle' is actually most of the distribution (in the UK)
  - Problems with survey at top refer to at most c. top 10%
  - Problems with survey at bottom (as revealed by 'tick' pattern shown earlier) refer to at most the bottom 3%–5%
- Labour earnings dominate households' income packages
  - ~ 60% of gross income at median (all households); ~ 80% of gross income at median at median for households with a working-age adult





Percentiles 1-3 and 98-100 are excluded because of large statistical uncertainty. Percentages may not always sum to 100% due to rounding.

Source: Department for Work and Pensions, *Households Below Average Income*, 2020

Classical error model and inequality

Suppose, as standard, a multiplicative error model:

$$\log(\text{survey income}) = \log(\text{true income}) + \text{error}$$

$$s_i = \xi_i + \varepsilon_i$$

$$\Rightarrow \text{var}(s) = \text{var}(\xi) + \text{var}(\varepsilon) + 2\text{cov}(\xi, \varepsilon)$$

Classical model: ξ_i and ε_i are uncorrelated

$$\Rightarrow \text{var}(s) = \text{var}(\xi) + \text{var}(\varepsilon)$$

Hence,

- Inequality of survey income over-estimates true inequality, according to variance of logs measure
- Result generalizes to all standard relative inequality indices: see Chesher and Schluter, *Rev Econ Stud* 2002

Mean-reverting errors and inequality

Suppose, instead, ξ_i and ε_i are correlated

$$\Rightarrow \text{var}(s) = \text{var}(\xi) + \text{var}(\varepsilon) + 2\text{cov}(\xi, \varepsilon)$$

Mean-reverting survey error: $\text{cov}(\xi, \varepsilon) < 0$

- E.g. over-report at bottom, under-report at the top

Hence (Gottschalk & Huynh, *REStat* 2010),

- If mean reversion sufficiently large, i.e., $\text{cov}(\xi, \varepsilon)/\text{var}(\varepsilon) < -0.5$, inequality of survey income **under**-estimates true inequality, according to variance of logs measure
- Trends: increases in inequality overstated if the variance of measurement error is increasing or if mean-reversion is declining
- [Do these results for non-classical errors generalize to other inequality measures?]

LSE

Key issues: errors, inequality, and data reliability (1/2)

1. Are survey errors mean-reverting or classical?
   - Answers depend on whether administrative data assumed to be error-free or not
   - 'First-generation' studies of labor earnings assume administrative data are error-free and found evidence of significant mean-reversion in survey earnings (in addition to substantial error variances)
     - US examples: Bound & Kreuger (1991), Bollinger (1998) Duncan & Hill (1985), Bound et al. (1994), Pischke (1995), Gottschalk & Huynh (2010), Kim & Tamborini (2014)
     - Non-US examples: AT (Angel et al., 2019); DK (Kristensen & Westergaard-Nielsen 2007);
     - UK examples: none
     - No similar studies of household income

Key issues: errors, inequality, and data reliability (2/2)

2. Information about the relative quality of survey and administrative measures is also important from a data collection point of view
   - Data substitution? If administrative data are much more reliable than the survey data, there are pay-offs to introducing methods that allow survey responses to be substituted by linked administrative data responses, not only for survey quality but also because respondent burden
     - Cf. Canadian Survey of Labour and Income Dynamics (SLID)

Errors, inequality, and data reliability:
Jenkins and Rios-Avila (2021*a*)

'Reconciling reports: modelling employment earnings and measurement error using survey and administrative data', IZA Discussion Paper 14405, May 2021

- First UK evidence about measurement errors in employment earnings in a field dominated by findings about the USA
- 'Second generation' study, i.e. one of few allowing for measurement errors in the administrative data as well as survey data
  - Kapteyn & Ypma (*JoLE* 2007), Abowd & Stinson (*REStat* 2013), Bingley & Martinello (*JoLE* 2017); Hyslop & Townsend (*JEBS* 2020); Bollinger et al. (WP 2018)
  - These studies find no mean reversion in survey measurement errors
- Econometric models with new features
  - Also examine UK-specific issue, i.e. 'reference period error' (current versus annual earnings measures)
  - Also allow error distributions to vary with observed characteristics – not discussed today

FRS-P14 linked dataset for 2011/12

Measures of employment earnings from UK Family Resources Survey (FRS) and HMRC's P14 administrative data for FRS respondents

HMRC: **H**er **M**ajesty's **R**evenue and **C**ustoms, the UK taxation authorities. P14: explained shortly

FRS-P14 linked dataset for 2011/12 on gross employment earnings: **FRS**

- FRS: large household survey (c. 20,000 households each year)
  - Source for DWP's annual *Households Below Average Income (HBAI)* reports
- Like other UK surveys, FRS uses a current measure of gross earnings with questions that refer to jobs in progress at the date of the interview
- For each job (up to 3), respondents are asked
  1. "What was the last amount received?", followed by
  2. "For what period does this amount refer"
     - Nine options including: 1 week (17%), fortnight, month (70%), 4 weeks (7%), year, etc., and 'other' (2%)
- Responses converted to "£ per week" pro rata by FRS data producers, which we convert to "£ per year": annualised earnings
- Survey measure of earnings for each respondent i, s_i, is the logarithm of total gross earnings (the sum across all jobs reported; annualised)

Reference period issues: current vs annual

- FRS earnings reference periods across respondents do not relate to a calendar-dated period that is common across respondents
  - Survey interviews can occur in any month during the financial year (the 12 months starting 5 April each year)
  - The reference period for the administrative data earnings measure is the financial year, i.e. annual (see below)
    - By contrast, in the US Annual Social and Economic Supplement to the CPS (CPS/ASEC), respondents provide information about earnings over the previous calendar year and this is also the reference period for the administrative data
- Hence, non-comparability between the annualised (survey data) and genuinely annual (admin data) earnings measures that analysis needs address head on, and we do this taking a model-based approach
  - Comparability error versus model error?

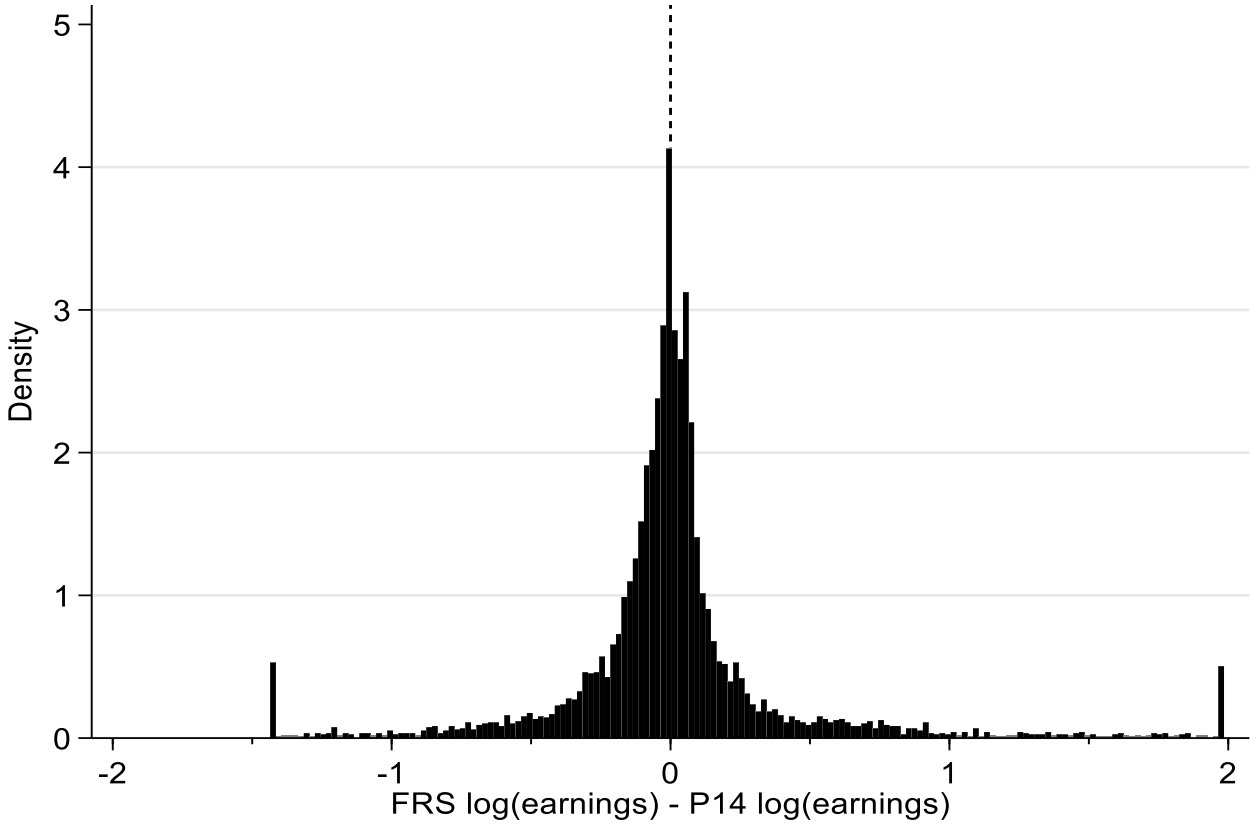
FRS-P14 linked dataset for 2011/12 on gross employment earnings: **P14**

- P14 data derived from records held by HMRC (UK tax authorities)
- Compiled from employers' year-end returns on P14 forms to HMRC about wages and salaries paid to employees and taxes and National Insurance contributions withheld
  - Cf. W-2 forms returned by US employers to the SSA
- Admin measure of earnings for each respondent i, r_i, is the logarithm of total gross earnings per year (the sum across all spells reported in 2011/12)
  - '*r*': think of '***r***egister' (as synonym for 'administrative')
  - Refer to "earnings" (rather than "log earnings") for brevity

FRS-P14 linked dataset for 2011/12 on gross employment earnings: **Linking**

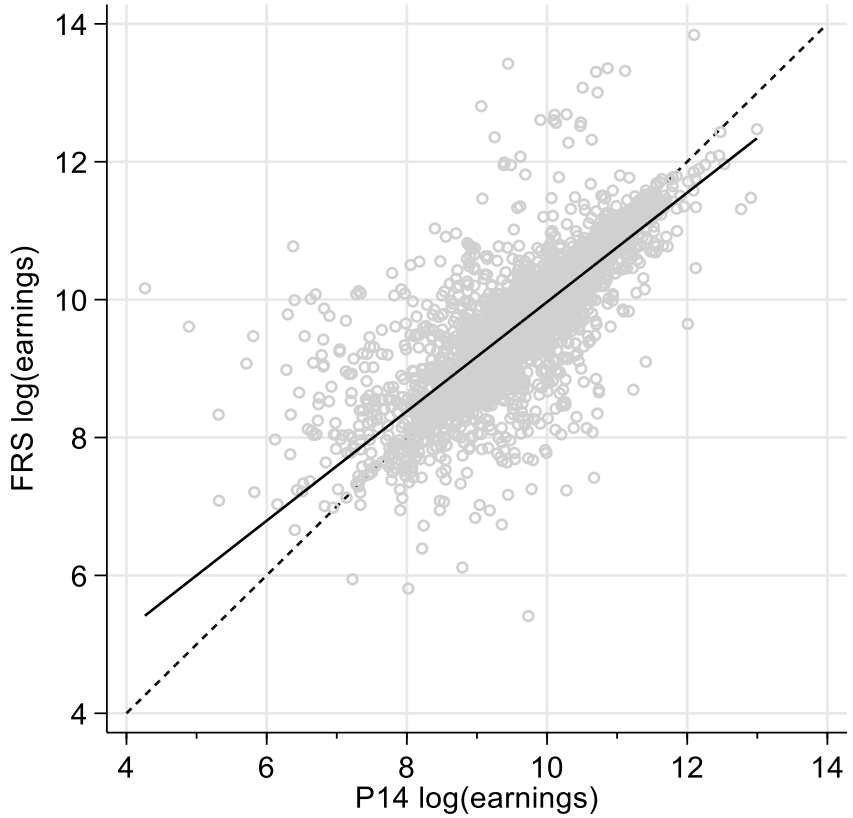
- The linked data we use are for:
  - FRS respondents in employment who gave their consent to record linkage and for whom DWP statisticians achieved a link, and …
  - Excluding some cases (exclusions standard in literature): 'self-employed', zero earnings in either source, obs with imputed/edited earnings
- Estimation sample:
  - $N = 5{,}971$ (2,595 men, 3,376 women)
  - Subsidiary analysis sample: $N = 3{,}564$ individuals aged 25–59, working full-time, not in any education; but similar findings (see Appendices)
- Consent and linkage = selective process?
  - Created inverse probability weights and used them to modify the FRS-survey weights; but findings robust to weighting, so no discussion here (see Appendices)
  - Results presented here are based on unweighted data

Distribution of differences between FRS and P14 log earnings ($s - r$)



Notes. Histogram with bin width = 0.02. Earnings differences are bottom-coded at $p1$ (–1.44) and top-coded at $p99$ (1.97) for purposes of presentation. **Summary statistics for $s - r$** (without bottom- or top-coding): mean, 0.016; $p5$, –0.579; $p10$, –0.315; $p50$, –0.005; $p90$, 0.331; $p95$, 0.714. Sample $N = 5,971$.

Relationship between FRS and P14 earnings



Notes. Solid line shows linear regression line with slope = 0.793 (SE 0.003). Sample N = 5,971.

If we treat the P14 data as 'truth', we see survey over-reporting at bottom and under-reporting at top = 'mean reversion' in survey measurement error

Econometric models

Finite mixture models, with latent classes defined by types of error present in survey and admin data

P14 earnings, r_i: 3 types of observation

Mixture of 3 types: P14 observations correctly linked with an FRS respondent (probability π_r) or mismatched, and correctly linked cases may be error-free (probability π_v) or contain measurement error:

- ($R1$) r_i equals i's true earnings, ξ_i
- ($R2$) r_i contains mean-reverting measurement error
- ($R3$) mismatch: r_i is the earnings of someone else in the full P14 dataset, ζ_i

$$r_i = \begin{cases} \xi_i \quad \text{with probability } \pi_r\pi_v & \text{(type } R1\text{)} \\ \xi_i + \rho_r(\xi_i - \mu_\xi) + v_i \quad \text{with probability } \pi_r(1-\pi_v) & \text{(type } R2\text{)} \\ \zeta_i \quad \text{with probability } (1-\pi_r) & \text{(type } R3\text{)} \end{cases} \qquad (1)$$

LSE

FRS earnings, s_i: 3 types of observation

Mixture: observations with error-free earnings; with measurement error; with error and reference period error

- ($S1$) s_i equals true earnings, ξ_i, with probability π_s
- ($S2$) s_i contains response error with a regression-to-the-mean component, with probability $(1-\pi_s)(1-\pi_\omega)$
- ($S3$) s_i as per $S2$ plus reference period error as well, with probability $(1-\pi_s)\pi_\omega$

where π_s: Pr(FRS earnings error-free), and

π_ω: Pr(FRS earnings include ref period error too)

$$s_i = \begin{cases} \xi_i \quad \text{with probability } \pi_s & \text{(type } S1\text{)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i \quad \text{with probability } (1-\pi_s)(1-\pi_\omega) & \text{(type } S2\text{)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i + \omega_i \quad \text{with probability } (1-\pi_s)\pi_\omega. & \text{(type } S3\text{)} \end{cases} \quad (2)$$

LSE

The general model has nine latent classes

Table 1. Groups (latent classes) in mixture factor model of FRS and P14 earnings

Group, j	Description	Types	Probability, $\pi_j = \ldots$
1	No error in P14 or in FRS earnings	$R1, S1$	$\pi_r \pi_\upsilon \pi_s$
2	No error in P14 earnings; error in FRS earnings	$R1, S2$	$\pi_r \pi_\upsilon (1-\pi_s)(1-\pi_\omega)$
3	No error in P14 earnings; error and reference period error in FRS earnings	$R1, S3$	$\pi_r \pi_\upsilon (1-\pi_s)\pi_\omega$
4	Error in P14 earnings; no error in FRS earnings	$R2, S1$	$\pi_r (1-\pi_\upsilon)\pi_s$
5	Error in P14 earnings; measurement error in FRS earnings	$R2, S2$	$\pi_r (1-\pi_\upsilon)(1-\pi_s)(1-\pi_\omega)$
6	Error in P14 earnings; measurement error and reference period error in FRS earnings	$R2, S3$	$\pi_r (1-\pi_\upsilon)(1-\pi_s)\pi_\omega$
7	Mismatched P14 earnings; no error in FRS earnings	$R3, S1$	$(1-\pi_r)\pi_s$
8	Mismatched P14 earnings; measurement error in FRS earnings	$R3, S2$	$(1-\pi_r)(1-\pi_s)(1-\pi_\omega)$
9	Mismatched P14 earnings; measurement error and reference period error in FRS earnings	$R3, S3$	$(1-\pi_r)(1-\pi_s)\pi_\omega$

Notes. π_s: probability FRS survey data are error-free. $1-\pi_r$: probability of data linkage mismatch. $1-\pi_\upsilon$: probability P14 administrative data contain measurement error.

Distributional assumptions

- True earnings, mismatch earnings, and errors are normally distributed:

$$\begin{pmatrix} \xi_i \\ \omega_i \end{pmatrix} = BVN\left(\begin{pmatrix} \mu_\xi \\ \mu_\omega \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \rho_{\xi\omega}\sigma_\xi\sigma_\omega \\ \rho_{\xi\omega}\sigma_\xi\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right),$$

$$\zeta_i \sim N(\mu_\zeta,\ \sigma_\zeta^2),\ \eta_i \sim N(\mu_\eta,\ \sigma_\eta^2),\ \text{and}\ \upsilon_i \sim N(\mu_\upsilon,\ \sigma_\upsilon^2)$$

where 'μ' denotes mean and 'σ' denotes SD

- Allows for a non-zero correlation between true earnings and reference period error $\rho_{\xi\omega}$; we expect < 0
- Normality: facilitates Maximum Likelihood estimation, and post-estimation derivations; commonly assumed
- Identification: see discussion in Jenkins & Rios-Avila (2021*a*)
- Methods for estimation and post-estimation, plus Stata programs: see Jenkins & Rios-Avila (2021*b*), IZA DP 14404

LSE

Parameter estimates

Constrained Extended model fits best

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo-likelihood	–8805.1	–8805.2	–9030.4	–9034.3
AIC	17644.1	17642.3	18086.7	18092.6
BIC	17757.9	17749.5	18173.8	18173.0
Reliability1 (r)	0.7405	0.7416	0.7552	0.7377
Reliability1 (s)	0.8101	0.8100	0.8395	0.8401
Reliability2 (r)	0.7398	0.7410	0.7144	0.6906
Reliability2 (s)	0.8188	0.8156	0.8072	0.8245

Notes. Sample $N = 5{,}971$ individuals within 4,874 households. Models based on a completely-labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$, see main text).

- The 2 Extended models fit much better than the 2 Full models ⇒ accounting for error in admin (P14) data is essential
  - I.e. to use less general models (3), (4) would lead to 'modelling error'
- The Constrained Extended model fits the best (AIC, BIC) and, in Extended model (1), $\hat{\rho}_{\xi\omega} = 0.03$ (SE 0.07), i.e. insignificantly different from zero
- Survey (FRS) data distinctly more reliable than the admin (P14) data
  - 'Reliability2': square of correlation between true and observed measure

LSE

Estimates (constrained Extended model)

		Estimate	(SE)		Estimate	(SE)
True earnings	μ_ξ	9.8077***	(0.0112)	σ_ξ	0.7243***	(0.0093)
Mismatch earnings	μ_ζ	8.0941***	(0.1687)	σ_ζ	1.2302***	(0.0862)
FRS meas error	μ_η	–0.0101***	(0.0029)	σ_η	0.0937***	(0.0066)
Ref period error	μ_ω	–0.2656***	(0.0478)	σ_ω	1.0081***	(0.1052)
P14 meas error	μ_υ	–0.0349	(0.0335)	σ_υ	0.3631***	(0.0257)
Mean-reversion	ρ_s	0.0073	(0.0040)	ρ_r	0.0908	(0.0609)
Probabilities	π_s	0.0517***	(0.0051)	π_1	0.0342***	(0.0024)
	π_ω	0.1093***	(0.0207)	π_2	0.5586***	(0.0338)
	π_r	0.9710***	(0.0056)	π_3	0.0685***	(0.0172)
	π_υ	0.6810***	(0.0464)	π_4	0.0160***	(0.0036)
				π_5	0.2617***	(0.0410)
				π_6	0.0321***	(0.0040)
				π_7	0.0015***	(0.0004)
				π_8	0.0245***	(0.0049)
				π_9	0.0030***	(0.0006)

Notes. Cluster-robust standard errors in parentheses (cluster is household).
Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
$1-\pi_s$: probability FRS survey data contain measurement error. π_ω: probability of survey reference period error.
$1-\pi_r$: probability of linkage mismatch. $1-\pi_\upsilon$: probability P14 administrative data contain measurement error.
Latent class probabilities: 3 largest are π_2 = Pr($R1$,$S2$), π_3 = Pr($R1$,$S3$), π_5 = Pr($R2$,$S2$)

Latent class probabilities: 3 largest are …

π_2 = Pr($R1$,$S2$) ≈ 56%

π_5 = Pr($R2$,$S2$) ≈ 26%

π_3 = Pr($R1$,$S3$) ≈ 7%

Factor distributions

- True earnings (ξ): mean and SD
  - Discussed shortly
- Mismatch earnings (ζ): mean and SD
  - Discussion in paper
- Pr(FRS meas error) ≈ 95%, $\mu_\eta \approx 0$, $\sigma_\eta \approx 0.09$
- Pr(P14 meas error) ≈ 32%, $\mu_\upsilon \approx 0$, $\sigma_\upsilon \approx 0.36$
- Pr(Ref period error) ≈ 11%, $\mu_\omega \approx -0.27$, $\sigma_\omega \approx 1.01$
  - Low prevalence of error, but … annualised (FRS) earnings under-estimate annual (P14) earnings on average; high error dispersion
- Mean-reversion absent: $\rho_s \approx 0$, $\rho_r \approx 0$

LSE

Inequality of observed earnings (FRS or P14) overestimates inequality of 'true' earnings

- SD of log earnings: 0.81 (FRS) vs. 0.73 (true) ⇒ *I*(FRS) 11% larger
- *p*90–*p*10 of log earnings: 2.01 vs. 1.86 ⇒ *I*(FRS) 8% larger
- *p*90/*p*10 of earnings levels: 7.7 vs. 6.5 ⇒ *I*(FRS) 18% larger
- Hence, also inequality of household income over-estimated
  - unless income component errors negatively correlated – are they?

Distributions of true and observed earnings

Log earnings	True (mixture model)	FRS data	P14 data
	(1)	(2)	(3)
Mean	9.81	9.77	9.75
*p*10	8.88	8.69	8.69
*p*50	9.81	9.83	9.84
*p*90	10.74	10.71	10.70
*p*50 – *p*10	0.93	1.14	1.14
*p*90 – *p*50	0.93	0.88	0.86
*p*90 – *p*10	1.86	2.01	2.00
Standard deviation	0.73	0.81	0.84

Take-aways from J & R-A paper

- Admin (P14) data are not error-free
  - Subject to measurement error and linkage mismatch
  - Distinctly less reliable than survey (FRS) data
    - Even though survey measurement error much more prevalent
- Reference period error
  - Bias (annualised current earnings measure under-estimates annual earnings); added noise; but prevalence relatively low
- Mean-reversion in measurement errors absent
  - As found in other second generation studies
  - Survey measure over-estimates 'true' inequality of employment earnings

- For further analysis, e.g., how error distributions differ across employees (models with covariates): see Jenkins and Rios-Avila (2021*a*)

Conclusions

Conclusions (1/3)

1. There are many different types of measurement error
   - Under-reporting at the top and at the bottom have received the most attention recently …
   - But other types of error are also consequential, …
   - E.g., inaccuracy in reporting, reference period errors, modelling error

Conclusions (2/3)

2. Measurement errors can have substantial impacts on estimates of inequality and poverty
   - Illustrations from top, bottom, and middle today: effects on levels greater than effects on trends?
   - Avoid generalizations: country and temporal contexts matter
     - E.g. UK situation doesn't apply everywhere else and its own situation changing; Nordic countries have very different data environment; etc.
     - E.g. many different types of admin data: income tax, cash benefits, social security contribution databases
   - Information gaps remain about …
     - Changes in the structure of measurement error over time (aside from aggregate under-reporting data)
     - Differences across countries
     - Errors in total household income versus errors in income components

Conclusions (3/3)

3. Administrative and survey data are complements, not substitutes
   - Avoid black/white descriptions: e.g. we should not assume that admin data are error-free and survey data error-ridden
   - Admin data have many strengths, including …
     - Long time series
     - Relatively good coverage of specific income ranges (very top and very bottom) and/or income sources
     - Increasing Real Time availability
   - But don't write the obituaries for survey data yet, especially since they too have particular advantages as well
     - Better measures of 'income' (conceptual validity); widespread availability; good coverage of most incomes
     - Also have extensive information about individual and household characteristics, essential for much analysis
   - Combine data … but cautiously
   - Linked data at the individual unit level are especially useful

4. Plenty of valuable research to be done on error-related issues, and ECINEQ is a very well-qualified collection of people to do the work!

LSE


"That's all Folks!"