Imputation of Pension Accruals and Investment Income in Survey Data

Andrew Aitken & Martin Weale

14th December 2017

First WID.World Conference, Paris

ES
CO
E

ECONOMIC
STATISTICS
CENTRE
OF
EXCELLENCE

Objective

- Construct a democratic measure of income growth

- Give equal weight to the income growth of each household, deflated using a democratic price index

- Use a method of stochastic imputation which largely replicates the distributional properties of the source data
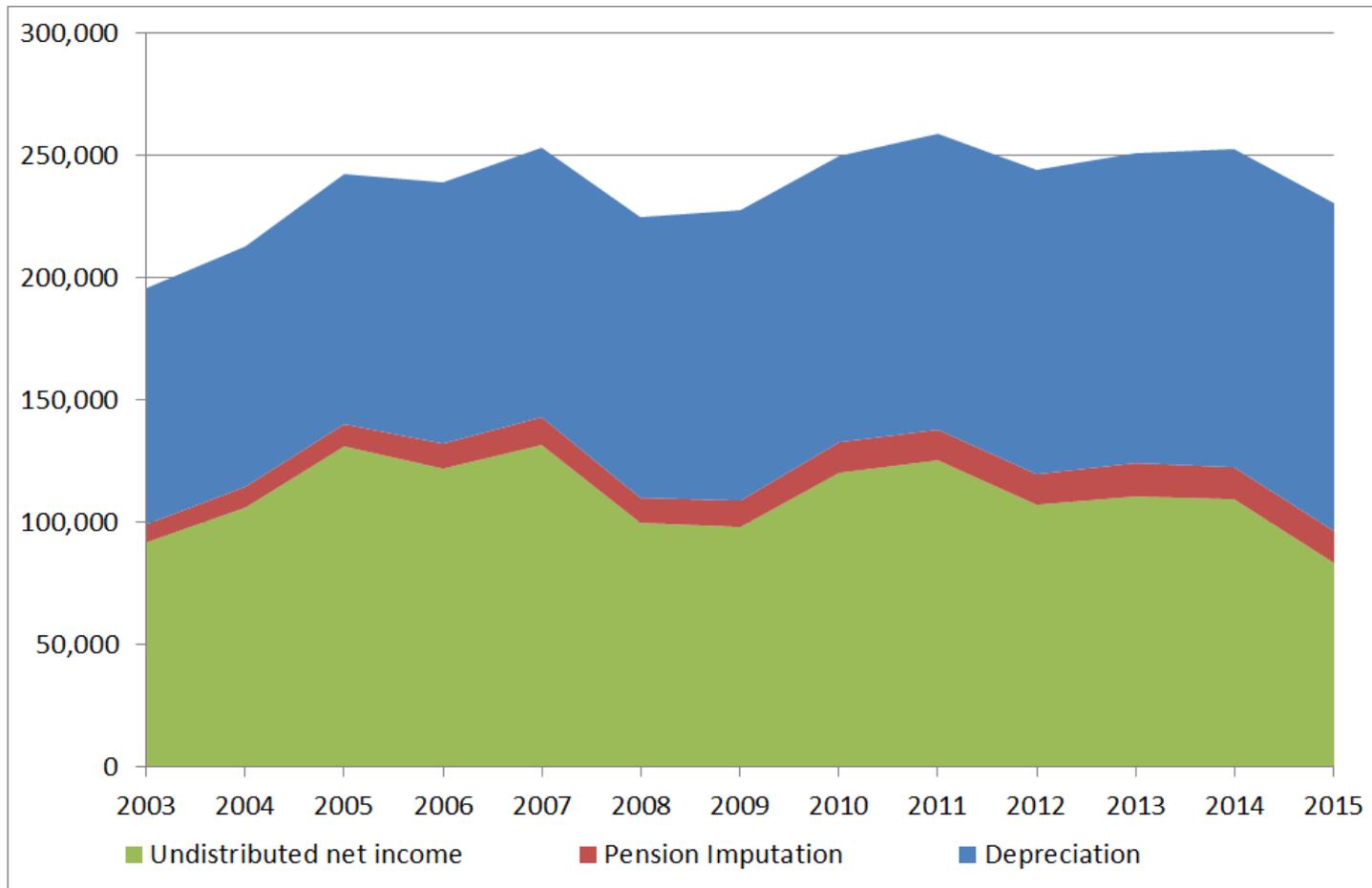
Two imputation issues

- Apparent under-reporting in the Living Costs and Food Survey relative to macro sources

- The need to allocate undistributed income of corporations and pension funds to households

- Both require modelling- the first stochastically and the second on the basis of plausible covariates.

The scale of misreporting

Component	National Accounts Total	Microsource Total	Coverage Rate (%)
Macro resources (received):			
Operating surplus	130,150	68,060	52
Mixed income	110,469	63,274	57
Wages and salaries	711,054	663,206	93
Net property income received	149,811	34,396	23
Social benefits other than STiK	332,504	231,013	69
Social transfers in kind	273,509	179,603	66
A Total	**1,707,497**	**1,239,552**	**73**
Macro uses (paid):			
Current taxes on income and wealth	195,524	142,923	73
Employers actual social contributions	136,091	59,606	44
Households social contributions	67,528	62,945	93
B Total	**399,143**	**265,474**	**67**
Household Disposable Income (A-B)	**1,308,354**	**974,078**	**74**
Memo: Gross Prop. Inc. excl. Rent	75,903	21,651	29

Source: Office for National Statistics and own calculations

Undistributed income of corporations (£mn)

Pension accruals and components of household investment income (£mn)
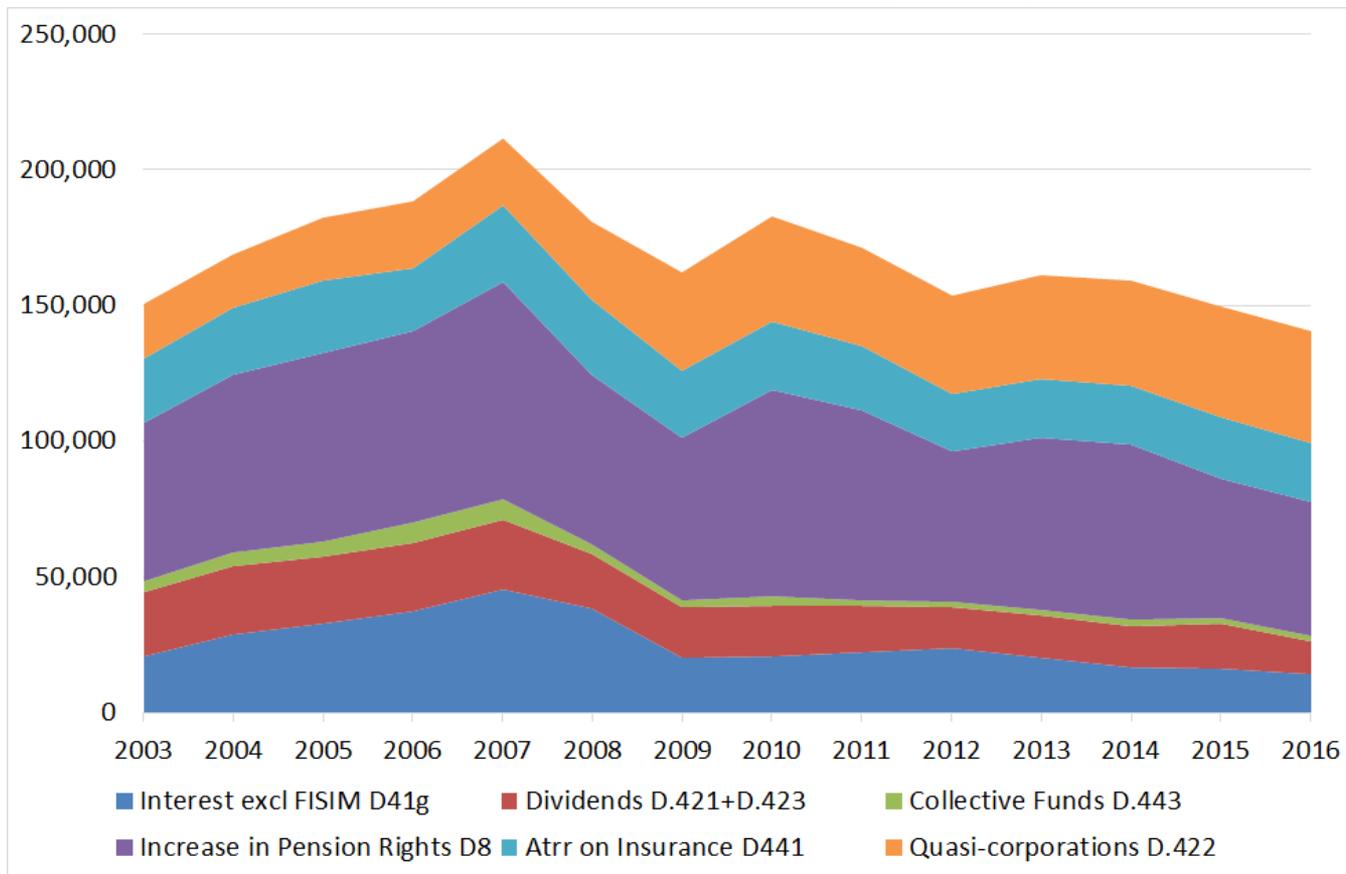
Note that income from "quasi-corporations" is income from partnerships perhaps better seen as mixed income than investment income

Imputation issues and approaches

- Scaling widely used (e.g. in ONS work on consumption)
- Scaling preserves zeroes
- Scaling will not work for sources of income omitted from LCFS- e.g. undistributed accruals to pension funds.
- We found a higher proportion of zeros in LCFS than in other sources (e.g. SPI and HBAI)
- Need to model both the probability of a non-zero receipt and the magnitude of the receipt conditional on being non-zero
- In contrast to scaling, this has to be stochastic - there is not going to be any covariate which exactly identifies non-zero recipients in HBAI or SPI

Heckman modelling

- Could use Heckman's approach to model jointly the probability of receiving interest/dividends and the amount conditional on receipt

- No obvious exclusion restriction: the model has to be identified by making the assumption of joint normality

- The distribution in fact departs substantially from normality

- This may not matter for the coefficients but it does for the stochastic imputation

Categorical imputation using Ordered Probit Models (i)

- We adopt a more flexible approach structured round an ordered probit model

- We convert the data in our source datasets (SPI *for investment income*/WAS *for pensions*) into a large number of categories (89 for investment income and 32 for pensions) and fit ordered probit models to these

- Covariates have to be variables available both in the source surveys and in LCFS

- Simulating these models provides stochastic categorical estimates which can be imputed into LCFS

Categorical imputation using Ordered Probit Models (ii)

- Compute a fitted value for each latent variable, and add on random terms from the multivariate normal distribution

- Each latent variable is allocated to the relevant category underpinning the probit model

  – Where it lies between 2 cut points, the distance between 2 categories is interpolated on the basis of the latent variable

The Upper Tail

- Reconciliation with the macro data requires appropriate handling of the upper tail

- Use a Pareto type-1 distribution for observations $x_i > x_m$ of the form:

- $1 - F(x) = (x_m/x)^\alpha$ with $\alpha > 0$

- where the expected value conditional on $x > x_m$ is $x_m \alpha / (\alpha - 1)$ if $\alpha > 1$ but infinite otherwise

- The expected value is used for imputed observations in the top category

Individuals and households

- SPI is based on tax records and provides data on individuals but not households

- This is because income tax is levied on individuals

- WAS and LCFS provide both individual and household data

- Investment income is imputed on an individual basis while pension rights are imputed on a household basis

ECONOMIC
STATISTICS
CENTRE
OF
EXCELLENCE

A COLLABORATION WITH

Office for
National Statistics

Taxation

- Revisions to tax paid need to be consistent with revisions to taxable household income

- We calculate each individual's tax bill on the basis of their income as recorded in LCFS and then recalculate it in the light of the imputations we make

- We add the difference on to the LCFS figure for tax paid

Covariances

- Need to take into account correlation between random components of imputed variables

- Use best source of data for pension wealth (WAS) and investment income (SPI), therefore not able to jointly estimate our models to estimate correlations simultaneously with parameters

- Estimate a correlation matrix using WAS (which does allow joint estimation but is not the ideal source) for the random components

Pension income

- Use ordered probit with waves 3 and 4 of WAS to allocate pension and insurance income to categories

- Include age, age^2, No. adults, No. children, tenure type, marital status, labour or pension income

- Estimate separately for under 65 (with & without labour income) and over 65 (with & without pension income)

- Waves 1 and 2 do not provide satisfactory income measures for use as covariates

Pension income

- Compare the performance of the Heckman and Ordered Probit approaches with wave 4 of WAS

- Assess the ability of the models to match the distribution of pension rights in the data.

- Examine both the full ordered probit model and the model relying on dummy variables only

E S
C O
E

ECONOMIC
STATISTICS
CENTRE
OF
EXCELLENCE

A COLLABORATION WITH

Office for
National Statistics

The distribution of pension rights simulated for 2013 using Heckman and ordered probit models applied to WAS data
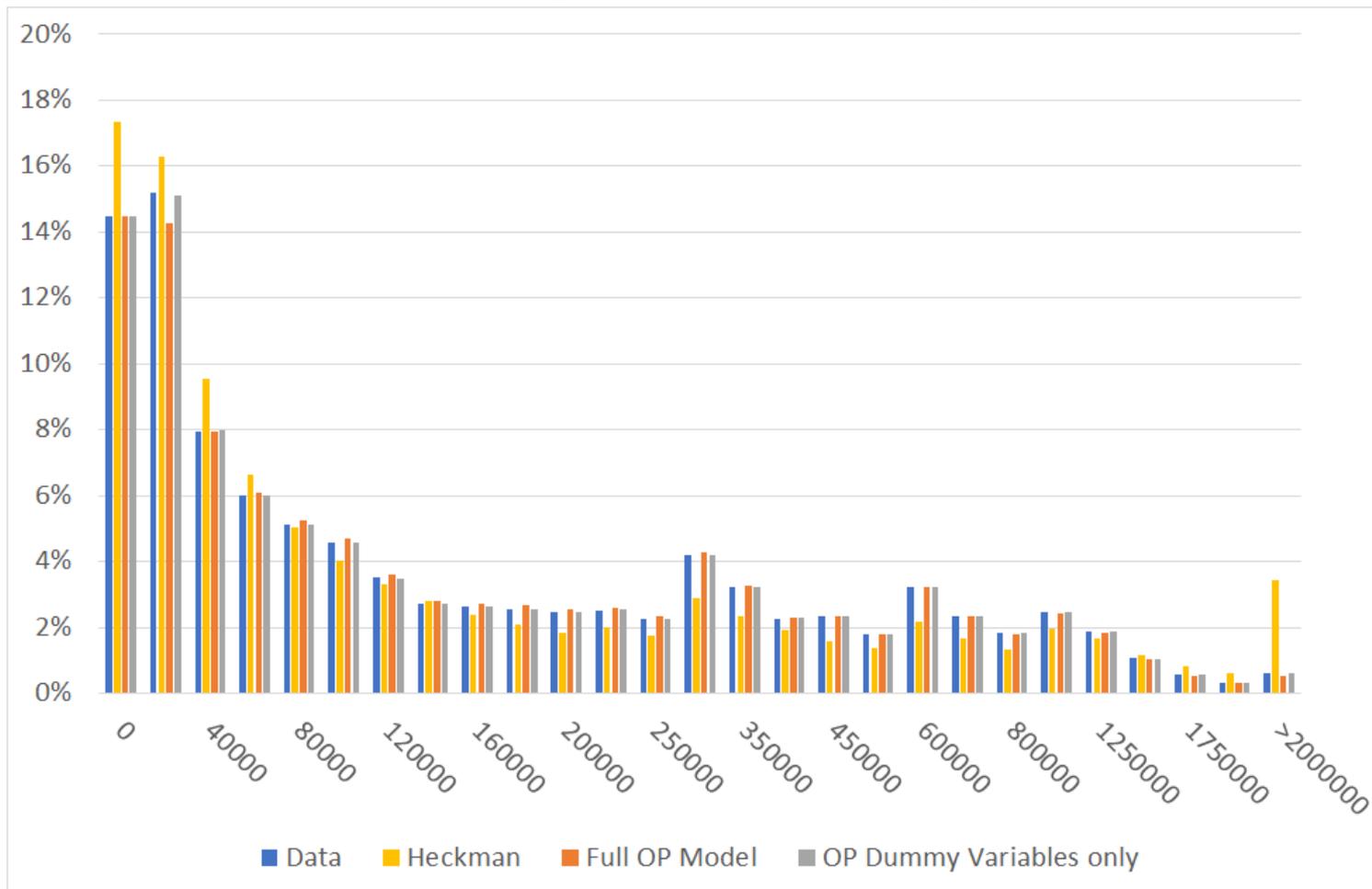
Investment income

- Use ordered probit with SPI to allocate investment income to categories
    - Include age bands, log labour income, regional dummies
    - Estimate separately for men and women and by year
    - Currently working on imputing dividends and interest income separately

The distribution of investment income in the 2013 SPI and the distribution fitted by the ordered probit model
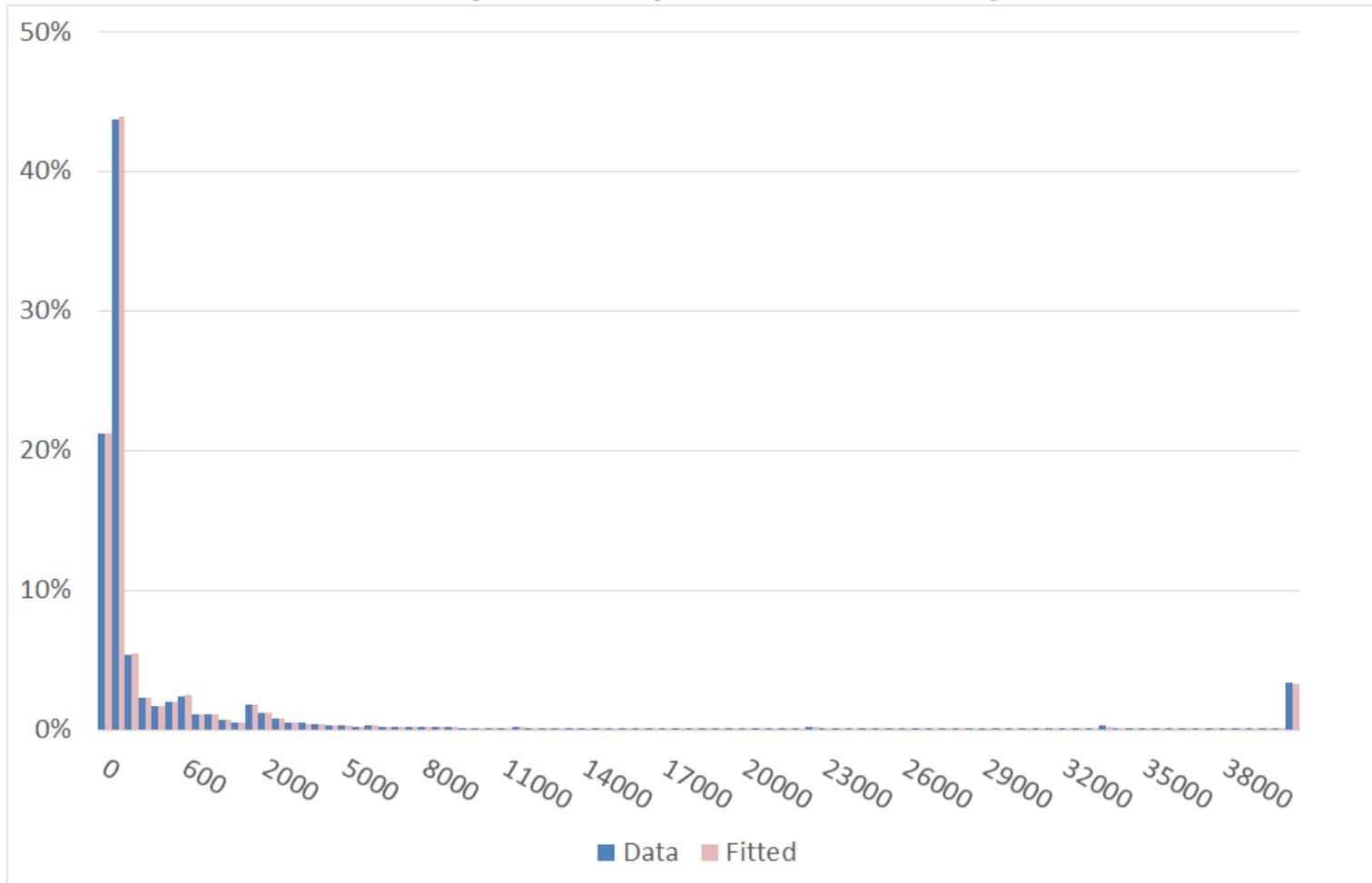
Covariances implementation (i)

- Assuming few households have more than two adult members, three correlations are needed

- ρ_{12} - the correlation between the latent variables driving investment income for each of the two adults

- ρ_{13} - the correlation between the latent variables driving investment income of the first adult and that driving pension rights

- ρ_{23} - the correlation between the latent variables driving investment income of the second adult and that driving pension rights

Covariances implementation (ii)

- Base covariances on coarse multivariate OP models fitted to WAS. Use financial asset holdings of first and second household members as proxies for investment income, together with household holding of pension rights.

- The model cannot be estimated for all types of household

- We use the estimated correlations we can find and take the arithmetic average

Covariances implementation (iii)

	Wave 3			Wave 4		Mean
	<65 Empl Inc	<65 No Empl Inc	>64 Pens Inc	<65 Empl Inc	< 65 No Empl Inc	
ρ_{12}	0.78	0.88	0.80	0.78	0.88	0.82
ρ_{13}	0.24	0.42	0.10	0.23	0.43	0.28
ρ_{23}	0.25	0.47	0.08	0.22	0.44	0.29

There is a strong correlation between the investment income of the two household members with possibly material implications for household income inequality. Correlations between investment income and pension rights are much weaker.

Simulations

- Examine the effect of including imputed pension and investment income on measures of inequality such as Gini & geometric mean of income

- Present results from 5 simulations

  - Preliminary due to top-coding of labour income in LCFS data

Estimates of the Gini coefficient with different definitions of income: 2006-2013
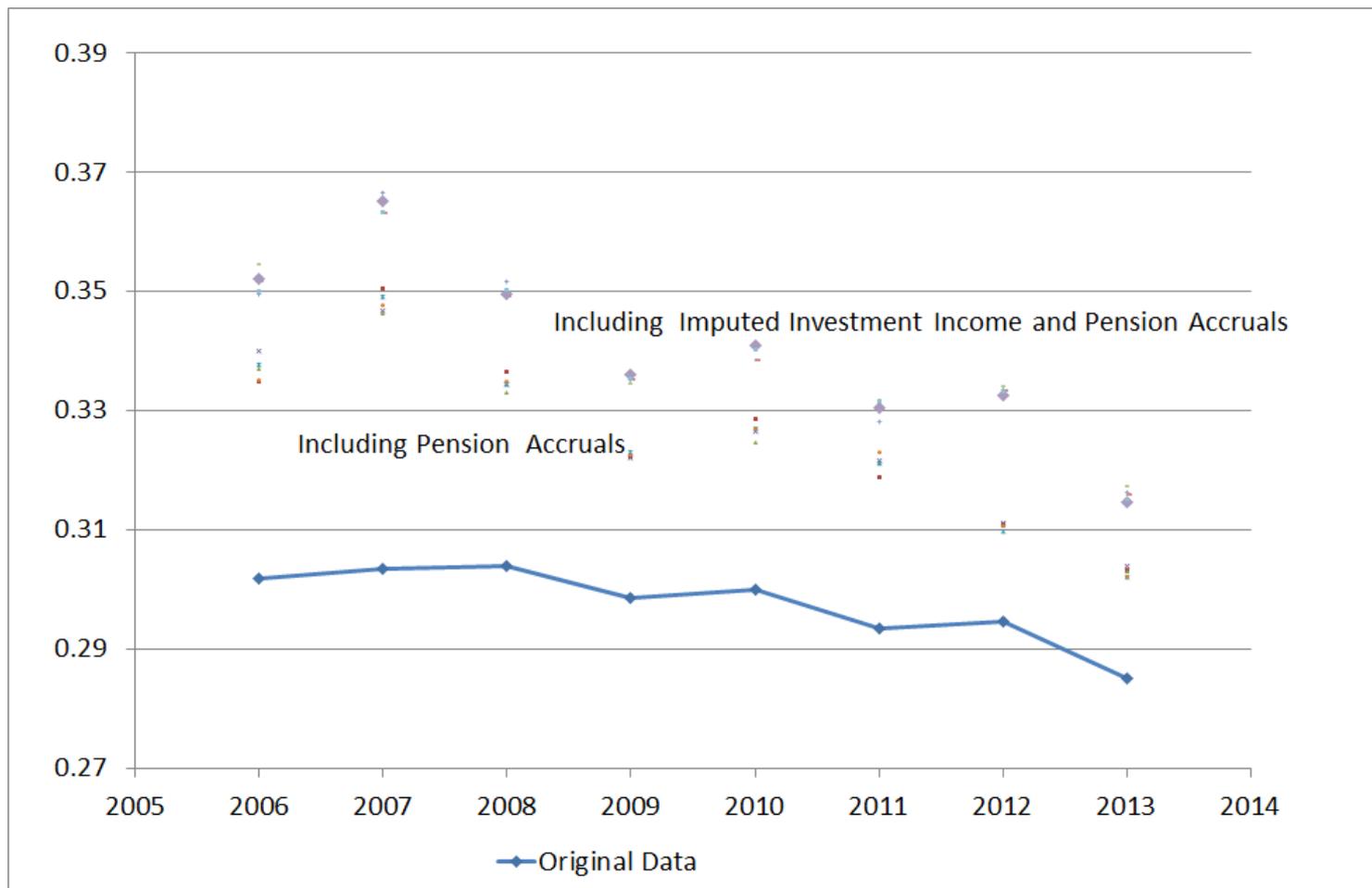
The Geometric mean of equivalised household income (£p.a.) with different definitions of income
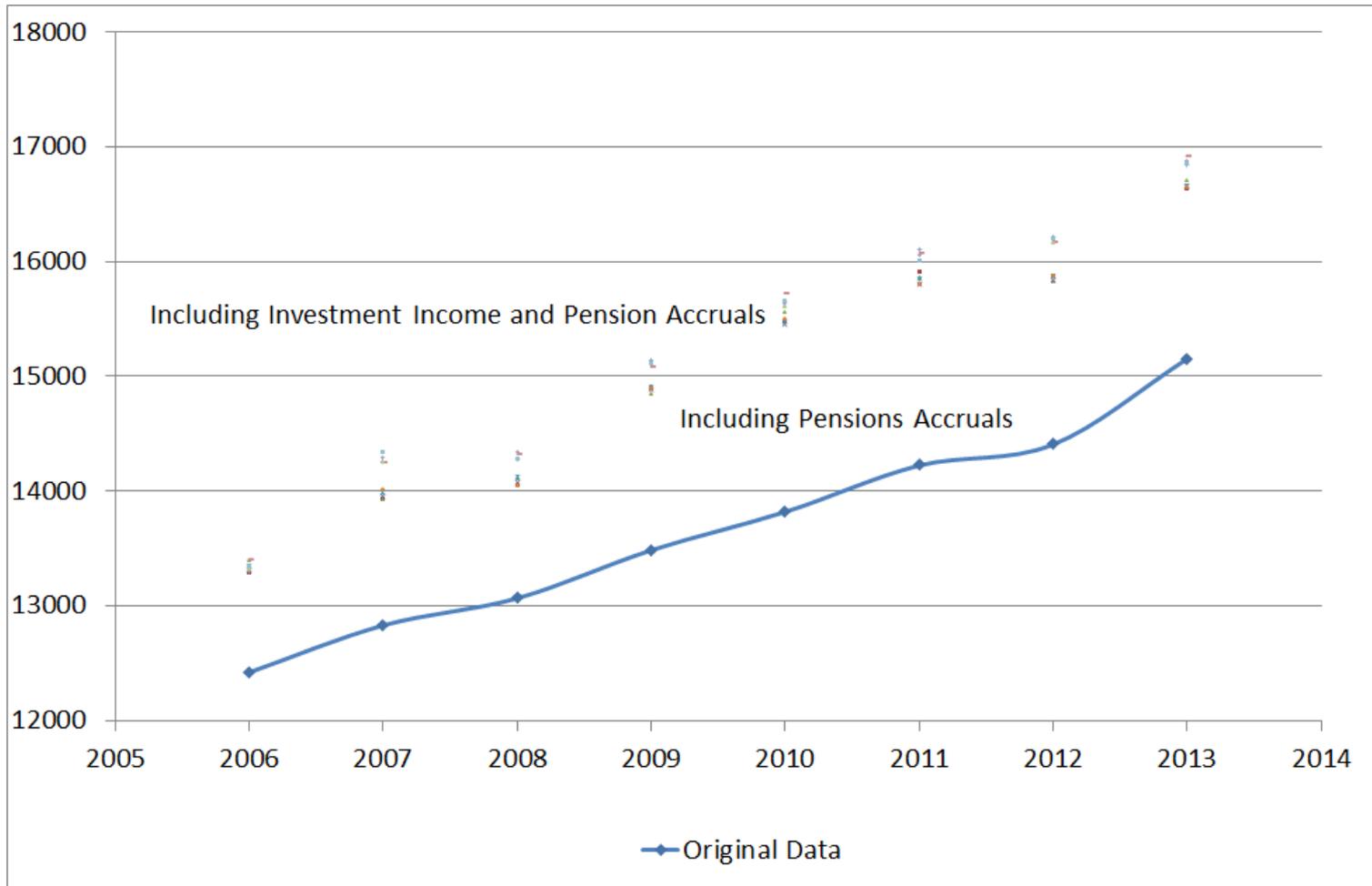
Future work

- Currently using top-coded version of LCFS, waiting for access to full version of data

- Investigate further the difference in investment income reported in the SPI and in the national accounts

- Imputing dividends and interest receipts separately

ESCO
ECONOMIC
STATISTICS
CENTRE
OF
EXCELLENCE

A COLLABORATION WITH
Office for
National Statistics

Merci

www.escoe.ac.uk