



Department  
for Work &  
Pensions

# Pensim2

## Data fusion and simulation for pensions modelling

James Rees PhD

Thomas Dawson PhD

Model Development Division

Department for Work and Pensions

24<sup>th</sup> March 2015

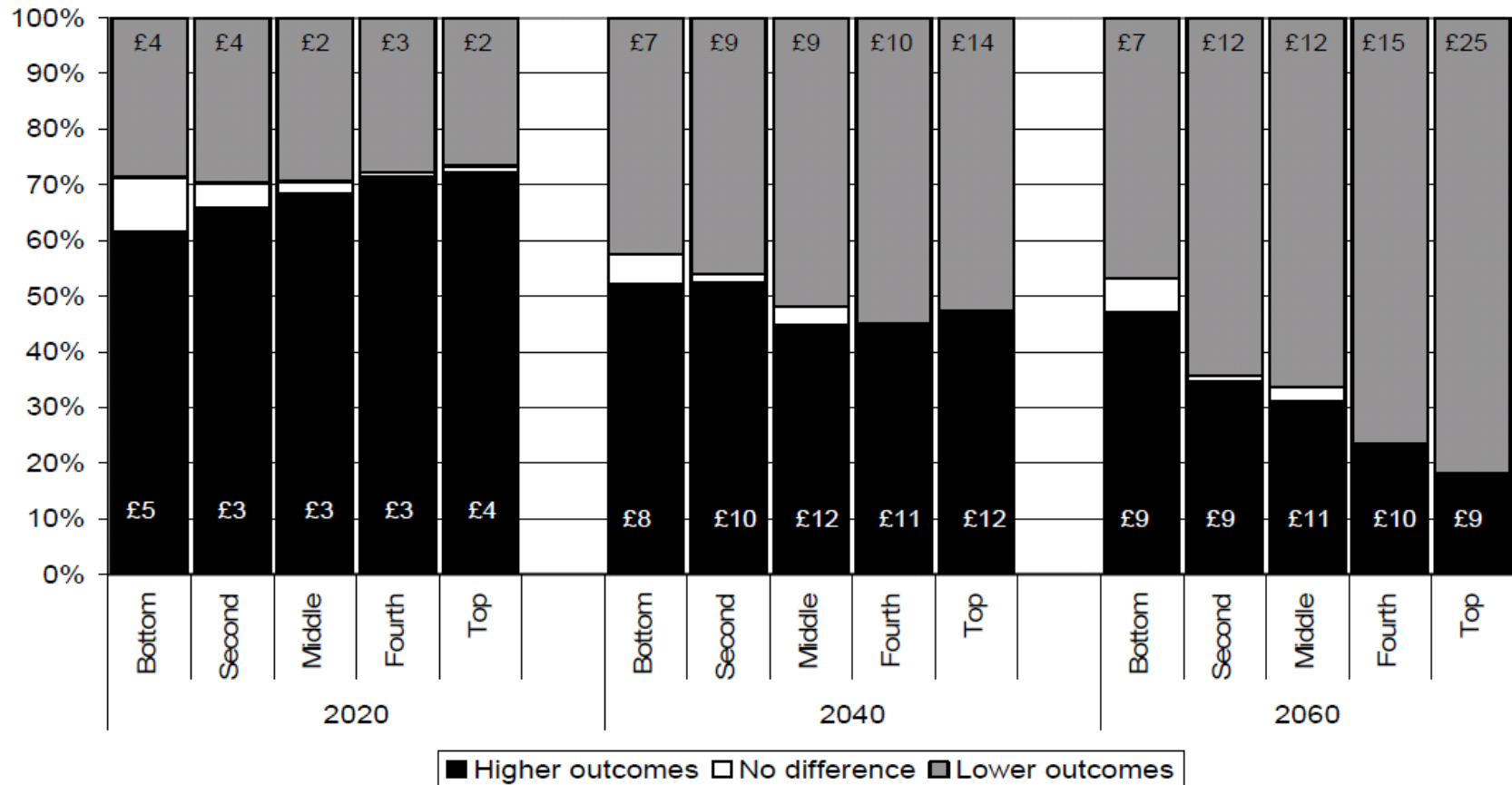
## Pensim2: Objectives and background

- Developed during work on the Turner Pensions Commission (2003-2005)
- Simulation of pensioner incomes and pension accruals to 2100
- Dynamic micro-simulation modelling
- Desktop tool, for pension policy analysts to carry out analysis of reforms under a variety of scenarios
- Flexible, easy to use, not a black box, developed in-house using Excel & SAS
- Used during the development of all major state & private pensions reforms since 2005

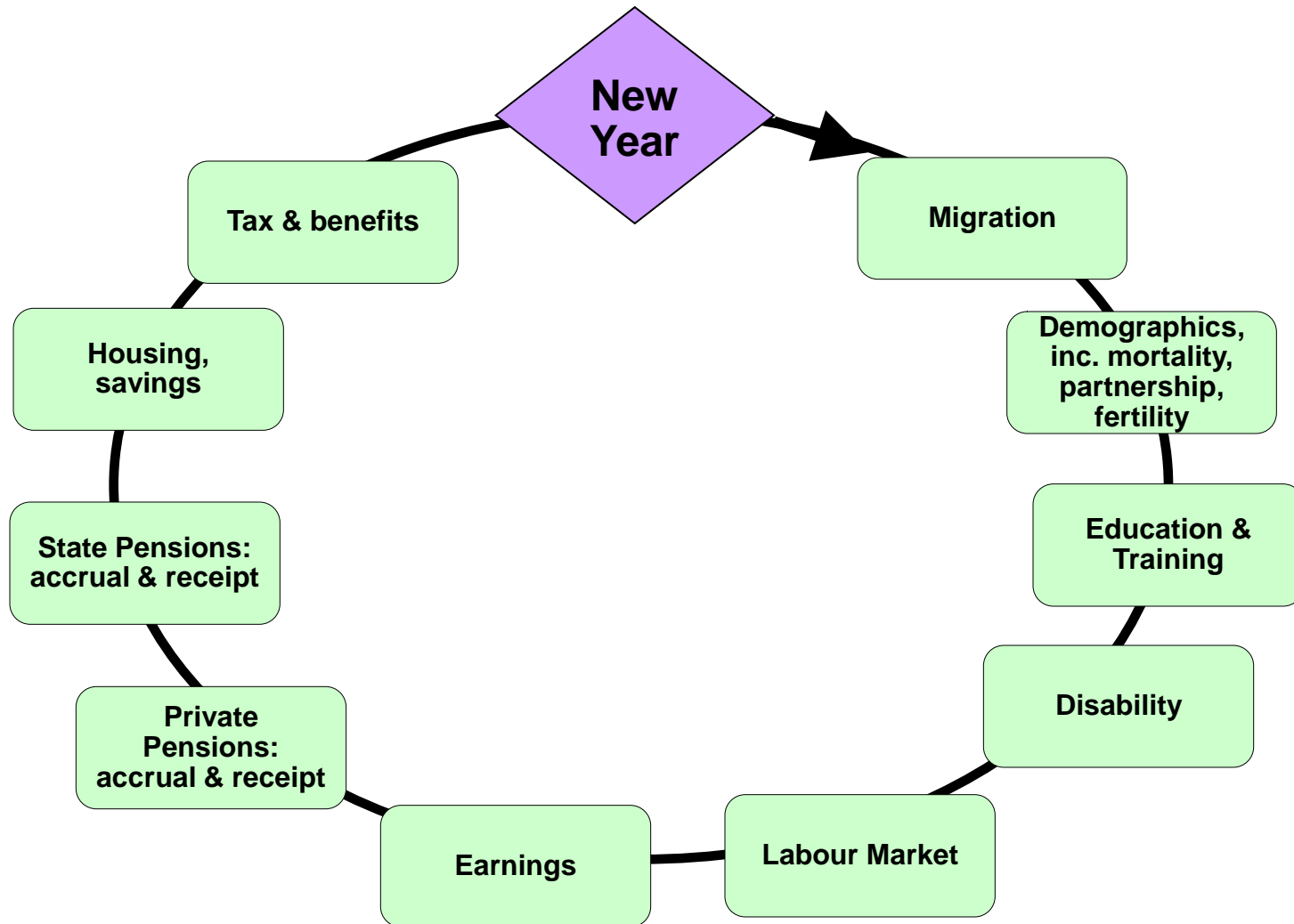
# Uses of Pensim2

Source: Pensions Act 2014 impact assessment

**Chart 3.4: Proportion of benefit units with changed notional outcomes due to single-tier reform compared to the baseline, by income quintile. Median weekly change (net incomes, 2013/14 earnings terms)**



# Processes simulated each year



## Key data sources

### Family Resources Survey (FRS)

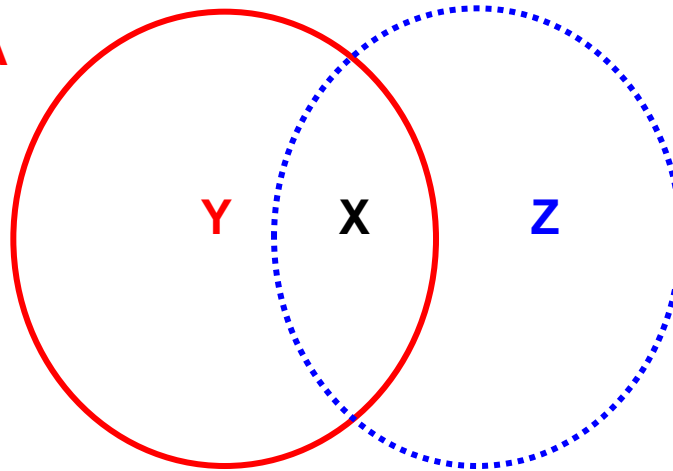
- Large sample (20-25K benefit units)
- Source of benefit unit structure
- Employment information
- Other contextual variables (e.g. disability, education level)
- Very limited information on working history

### Lifetime Labour Market Database (L2)

- 1% sample of National Insurance Recording data from HMRC
- Can be used to build up a longitudinal work history and NI contributions
- Also has information on private pension membership (contracted-out schemes only)
- No information on benefit unit structure

# Conditional Independence Assumption

Dataset A



Dataset B

(Sims et al., 1972)

(Sims, 1974)

Fusion of A and B is only valid if:

“Conditional on X, the Y and Z variables are independent”

i.e. If we were trying to predict Y from X and Z then we could omit the Z variables with no loss in predictive powers. (Likewise Z from X and Y)

Difficult to prove that the CIA holds – as the reason that we are fusing in the first place is that we don't have a dataset with X, Y and Z on it

## Bridging the gap?

### British Household Panel Survey (BHPS)

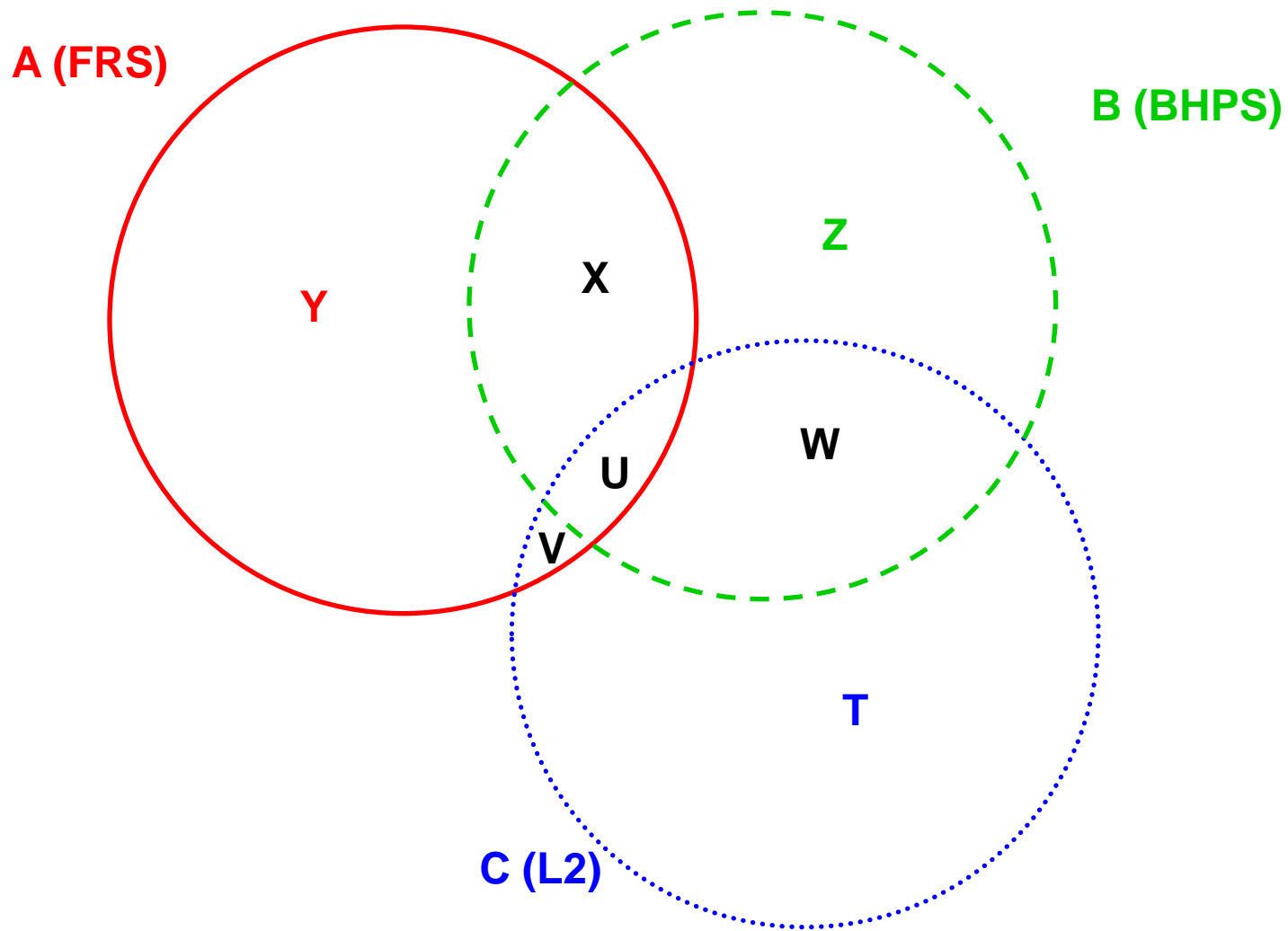
#### / Understanding Society

- Smaller sample (5K benefit units), but...
- Has information on work history back to 1991
- And also has some contextual information (benefit unit structure, etc)
- Therefore we can plausibly fuse this to the FRS and L2 without violating the CIA

### Wealth and Assets Survey

- Sample of around 30K benefit units
- Information on wealth from private pensions (occupational and personal)
- Limited information on work history

# Bridging the gap?



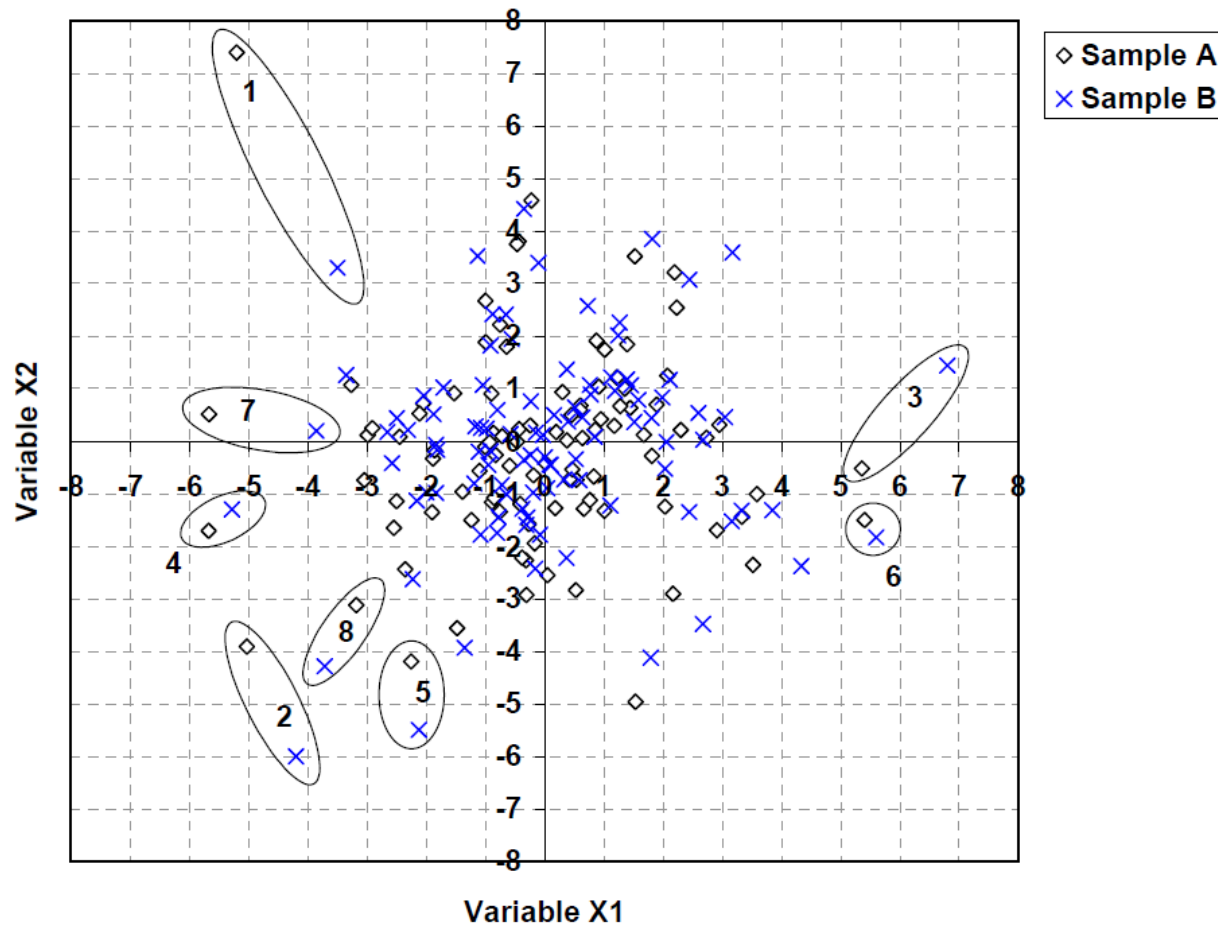


# Components of data fusion project

- Standard considerations, for example, see (Leulescuc, A. and Agafitei, M., 2013)
- Matching variables
- Distance function
- Sample subsets ('cells')
- Matching algorithm
  - Unconstrained matching (Rodgers, 1984)
  - Constrained matching approach (Rodgers, 1984; Rubin, 1986)
- Evaluation

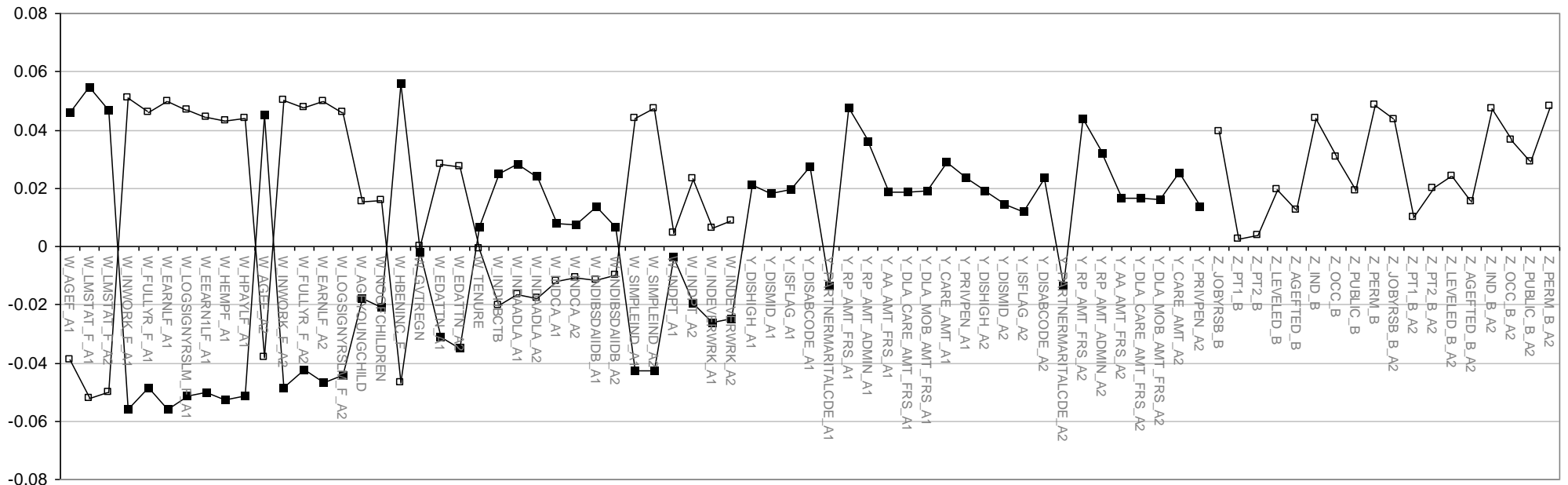
# Matching algorithm

The Order of Decreasing Distance algorithm (Redway, 2003).

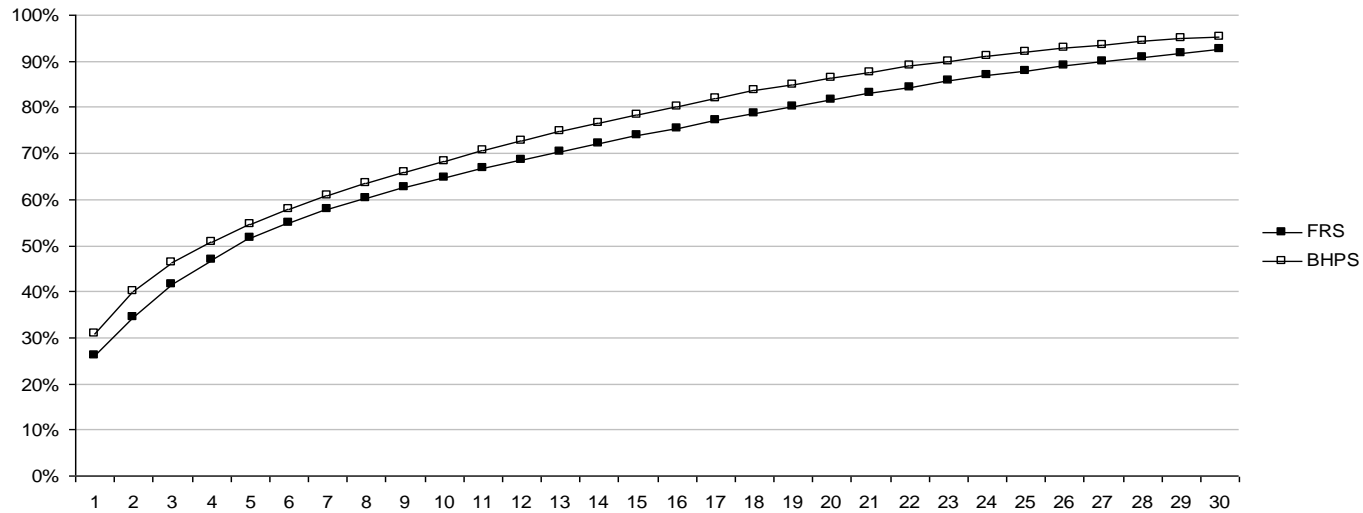
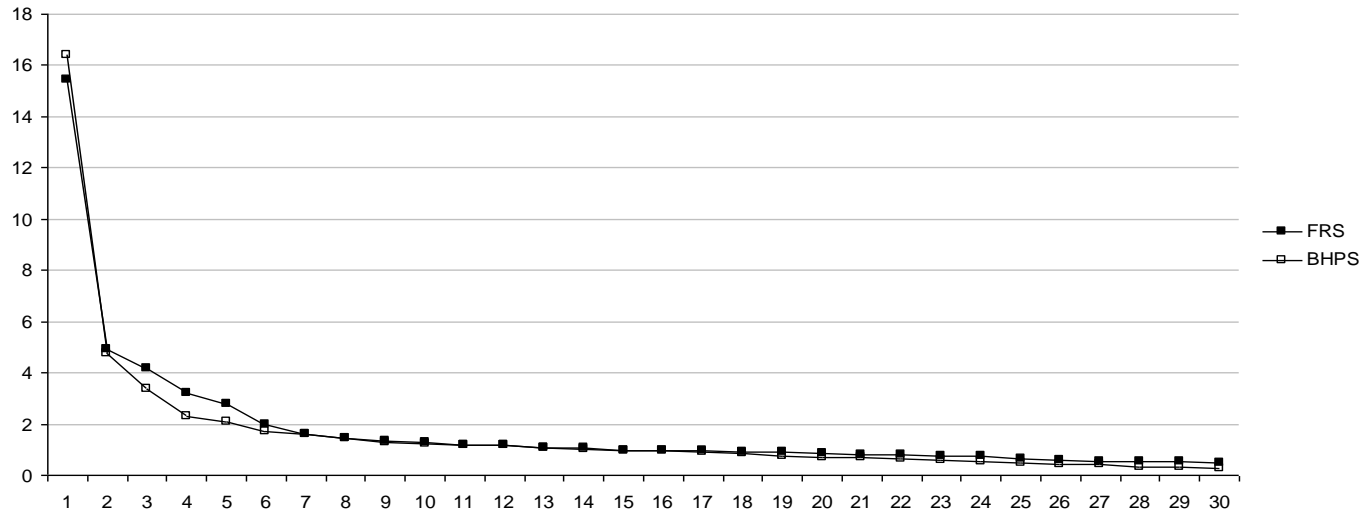


# Matching variables: principal components?

The first 'Principal component' of both FRS and BHPS data are almost identical

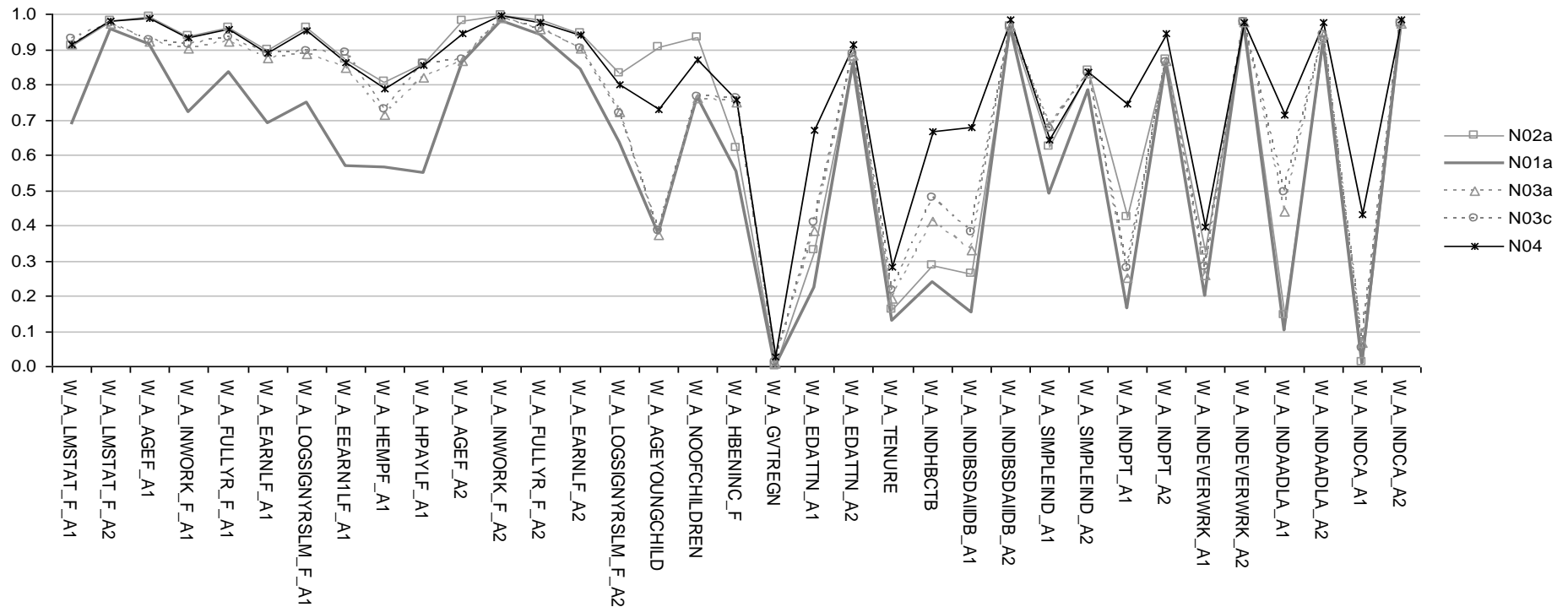


# Matching variables: principal components?



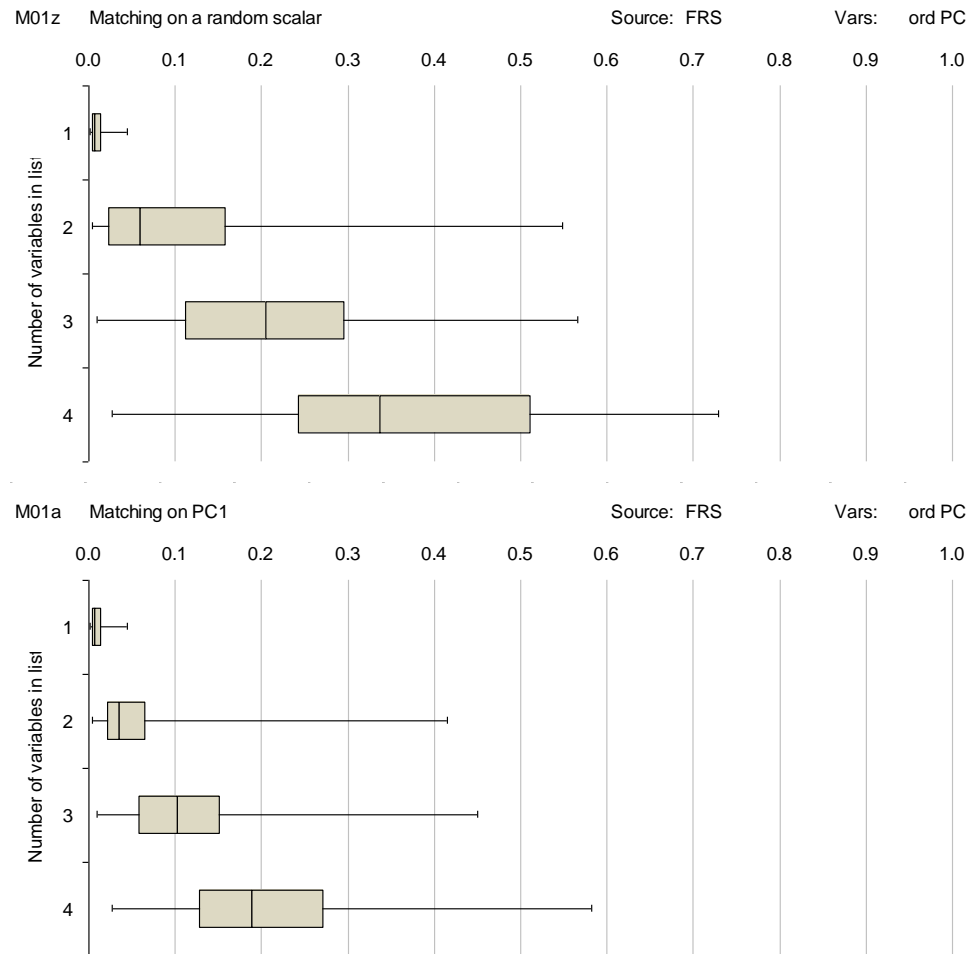
# Evaluating the fusion

## Correlations between variables common to both datasets



# Evaluating the fusion

Distance between multivariate distributions – Hellinger distance e.g. Leulescuc, A. and Agafiței, M. (2013)



## Pensim2 fusion

### Fusion 1: FRS-BHPS (benefit unit level fusion)

- Matching variables: age, partner age, earnings, partner earnings, years in job, partners years in job, number of children, age of youngest child...
- Subsets: couple status, age, sex, current employment status
- Distance: Euclidean
- Algorithm: ODD

### Fusion 2: FRS#BHPS – L2 (individual level fusion)

- Matching variables: age, earnings, historic NI contributions (class 1 or class 2)
- Subsets: age, sex
- Distance: Euclidean
- Algorithm: ODD

## Validation of the base data

### Comparison with alternative data sources

- Aggregate census data
- Wealth & Assets Survey: private pension wealth
- State Pension accrual
- National Statistics on pensioner incomes

### Forward simulation

- Discontinuities at the base data year
- Consistency with other modelling
- 'Response' to modelled reforms



# References

- Cohen, M. L. 'Statistical Matching and Microsimulation Models', in Citro, C. F. and Hanushek, E. A. (editors), (1991) 'Improving Information for Social Policy Decisions. The Uses of Microsimulation Modelling Vol. II', National Academy Press, Washington DC
- Department for Work and Pensions (2014) '[Pensions Act 2014 Annex A: Single Tier pension: Impact assessment](#)'
- Department for Work and Pensions (2012) '[Workplace Pension Reform: Baseline evaluation report](#)'
- D'Orazio, M. (2008) in 'Report of WP1. State of the art on statistical methodologies for integration of surveys and administrative data', *ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data*, available from: [http://cenex-isad.istat.it/archivio/Reports\\_of\\_the\\_project\\_workpackages/Report\\_of\\_WP1.pdf](http://cenex-isad.istat.it/archivio/Reports_of_the_project_workpackages/Report_of_WP1.pdf)
- Leulescuc, A. and Agafitei, M. (2013) 'Statistical matching: a model based approach for data integration', *Methodologies and Working Papers*, Eurostat, European Commission, available from: [http://epp.eurostat.ec.europa.eu/cache/ITY\\_OFFPUB/KS-RA-13-020/EN/KS-RA-13-020-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-13-020/EN/KS-RA-13-020-EN.PDF)
- Li, J. and O'Donoghue, C. (2012) 'A methodological survey of dynamic microsimulation models', *UNU-MERIT Working Paper No. 2012-002*
- Mackay, S. (2012) 'Evaluating approaches to Family Resources Survey data linking', Department for Work and Pensions, Working Paper 110, available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/222871/WP110.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/222871/WP110.pdf)
- Moriarty, C. and Scheuren, F. (2003) 'A Note on Rubin's Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations', *Journal of Business and Economic Statistics* (Vol. 21) No. 1, p. 65–73
- Moriarty, C. (2009) 'Regression-Based Statistical Matching: Past, Present and Future', Presentation at International Statistical Institute Conference 2009, available from: <http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0064.pdf>
- Redway, H. (2003) 'Data Fusion by Statistical Matching', *DWP internal paper*
- Rodgers, W. L. (1984) 'An Evaluation of Statistical Matching', *Journal of Business and Economic Statistics* (Vol. 2) No. 1, p. 91–102
- Rubin, D. B. (1986) 'Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations', *Journal of Business and Economic Statistics* (Vol. 4) No. 1, p. 87–94
- Scanu, M. (2008) in 'Report of WP2. Recommendations on the use of methodologies for the integration of surveys and administrative data', *ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data*, available from: [http://cenex-isad.istat.it/archivio/Reports\\_of\\_the\\_project\\_workpackages/Report\\_of\\_WP2.pdf](http://cenex-isad.istat.it/archivio/Reports_of_the_project_workpackages/Report_of_WP2.pdf)
- Shao, J. and Sitter, R. R. (1996) 'Bootstrap for Imputed Survey Data', *Journal of the American Statistical Association* (Vol. 91) No. 435, p. 1278–1288
- Sims, C. A., Peck, J. K., Budd, E. C., and Okner, B. A. (1972) Comments and responses, *Annals of Economic and Social Measurement* (1), p. 343–362
- Sims, C. A. (1974) 'Comment', *Annals of Economic and Social Measurement* (3), p. 395–397
- Soong, R. and de Montigny, M. (2001) 'The anatomy of data fusion', 2001 Worldwide Readership Research Symposium (Venice), available from: <http://www.zonalatina.com/WRRSfusion.pdf>
- Webber, D. and Tonkin, R. (2013), 'Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditures and material deprivation', *Methodologies and Working Papers*, Eurostat, European Commission, available from: [http://epp.eurostat.ec.europa.eu/cache/ITY\\_OFFPUB/KS-RA-13-020/EN/KS-RA-13-020-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-13-020/EN/KS-RA-13-020-EN.PDF)