

Evaluation of Alternative Income Imputation Methods for the HILDA Survey

Rosslyn Starick and Nicole Watson
19 July 2007



Missing income (% of respondents with non-zero income)

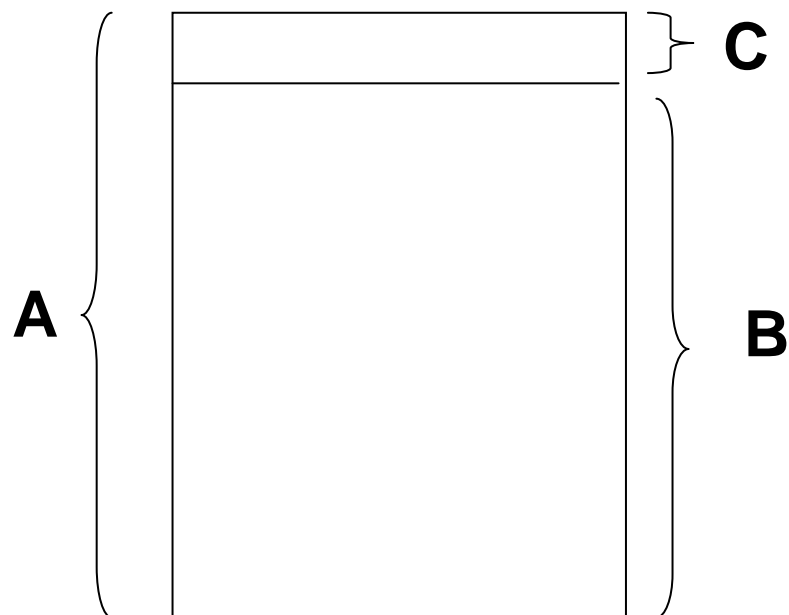
FY income component	W1	W2	W3	W4	W5
Wages and salaries	7.8	6.8	5.4	3.7	4.8
Gov't benefits / pensions	2.1	2.1	1.3	2.0	2.4
Business income	29.2	28.6	27.4	19.4	21.9
Investment income	19.4	17.9	14.2	11.5	12.1
Other income	5.9	8.9	6.3	6.4	10.2
Total respondent income	14.8	14.2	11.5	9.2	11.2
Total household income	29.4	28.0	24.0	21.9	23.7

Imputation strategy in HILDA

- Impute at person-level:
 - Item non-response
 - Wave non-response in responding households
 - Unit non-response in responding households
 - Impute at derived variable level
- Sum to get household-level income
- Revised each release

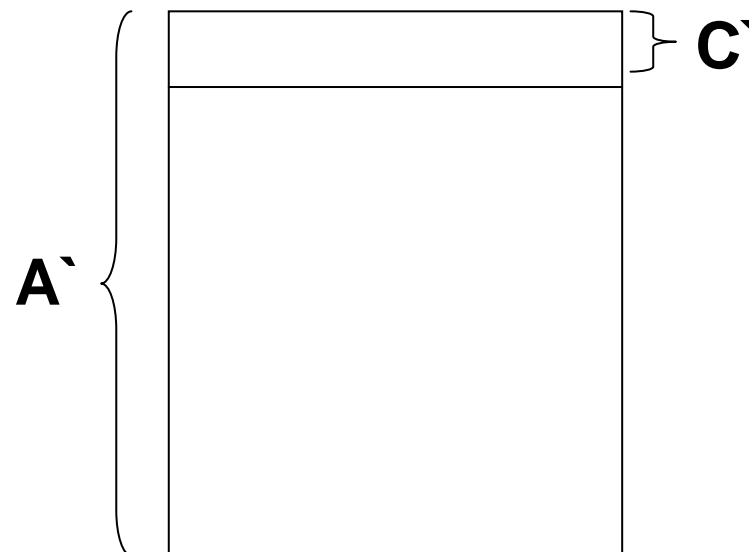
Simulation datasets

x30



HILDA data

A: entire HILDA dataset
B: 'complete' cases
C: true missing values



Evaluation data

A': evaluation dataset
C': imputed values

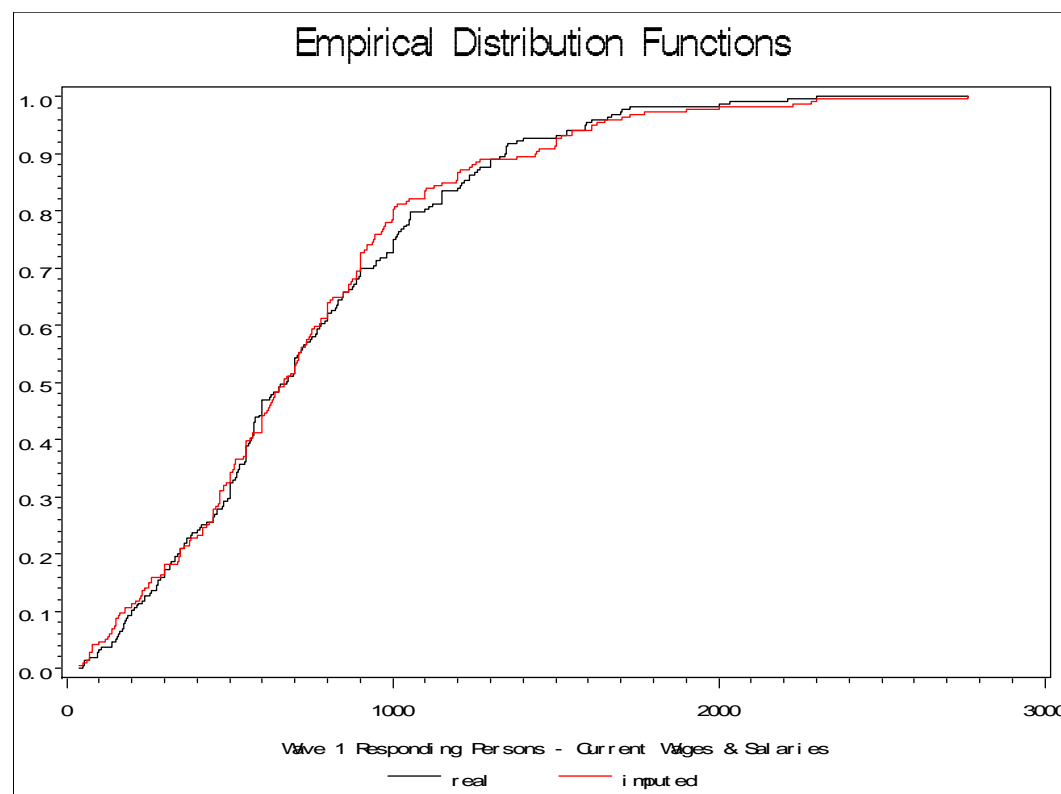
Evaluation criteria: Predictive accuracy

Smallest difference between true and imputed values:

1. Correlation*
2. Regression coefficient*
3. Correlation across waves
4. Euclidean distance between multiple components
5. Correlation across multiple components

Evaluation criteria: Distributional accuracy

1. Smallest max distance btw empirical distributions*
2. Closest reproduction of income mobility between waves

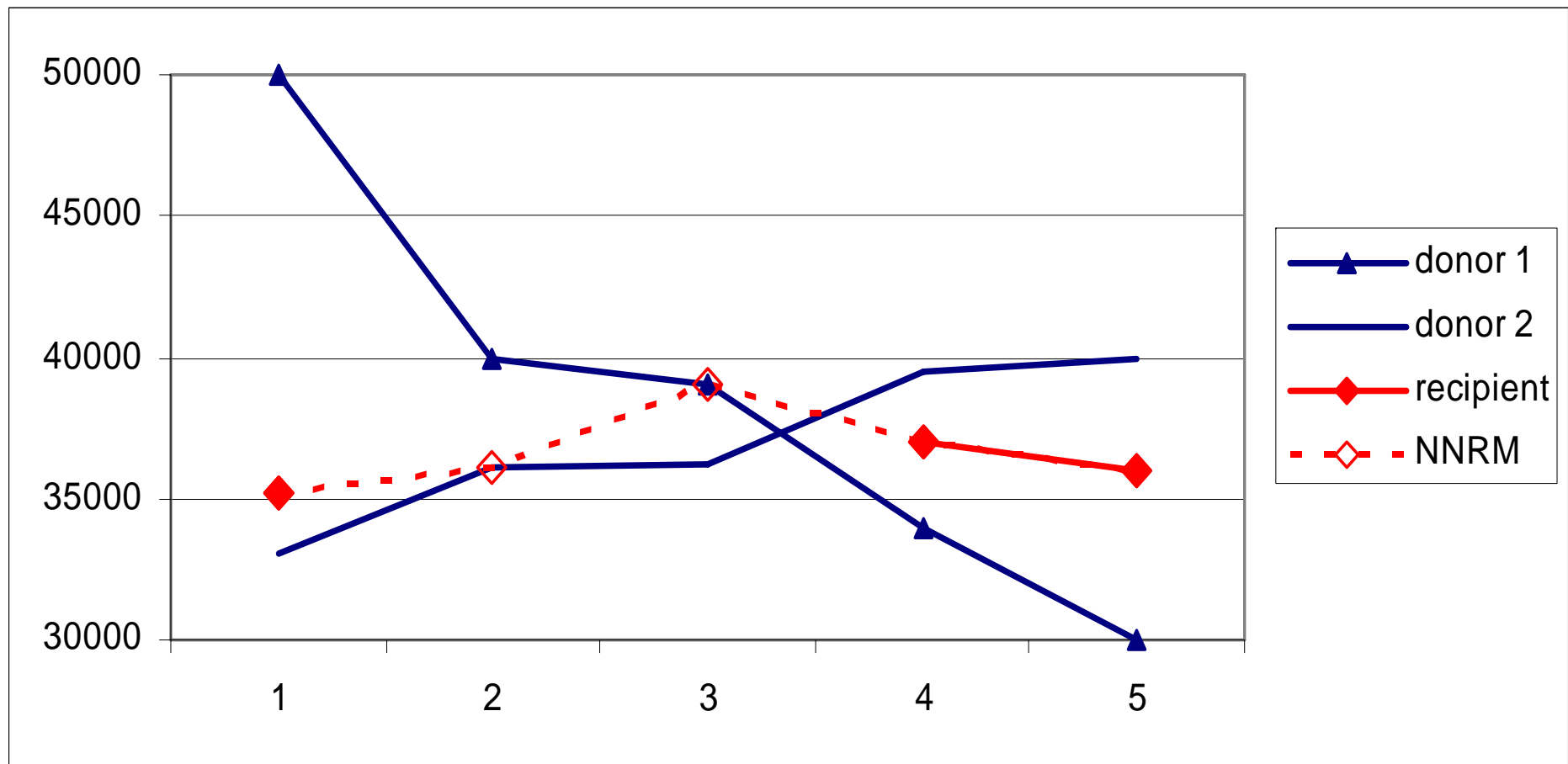


Evaluation criteria: Estimation accuracy

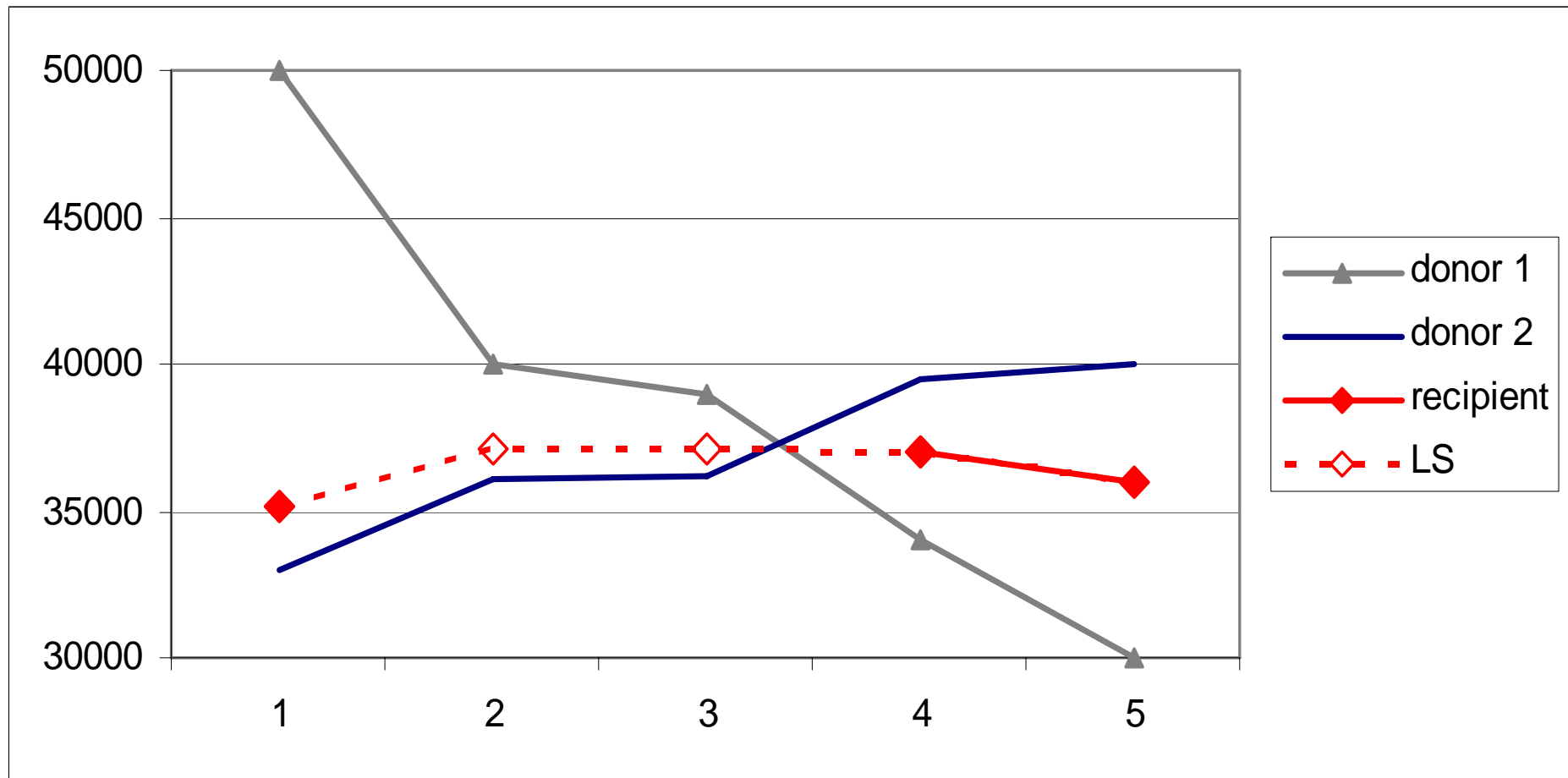
Smallest difference between true and imputed estimates of:

1. Means*
2. Standard deviations*
3. Skewness*
4. Kurtosis*

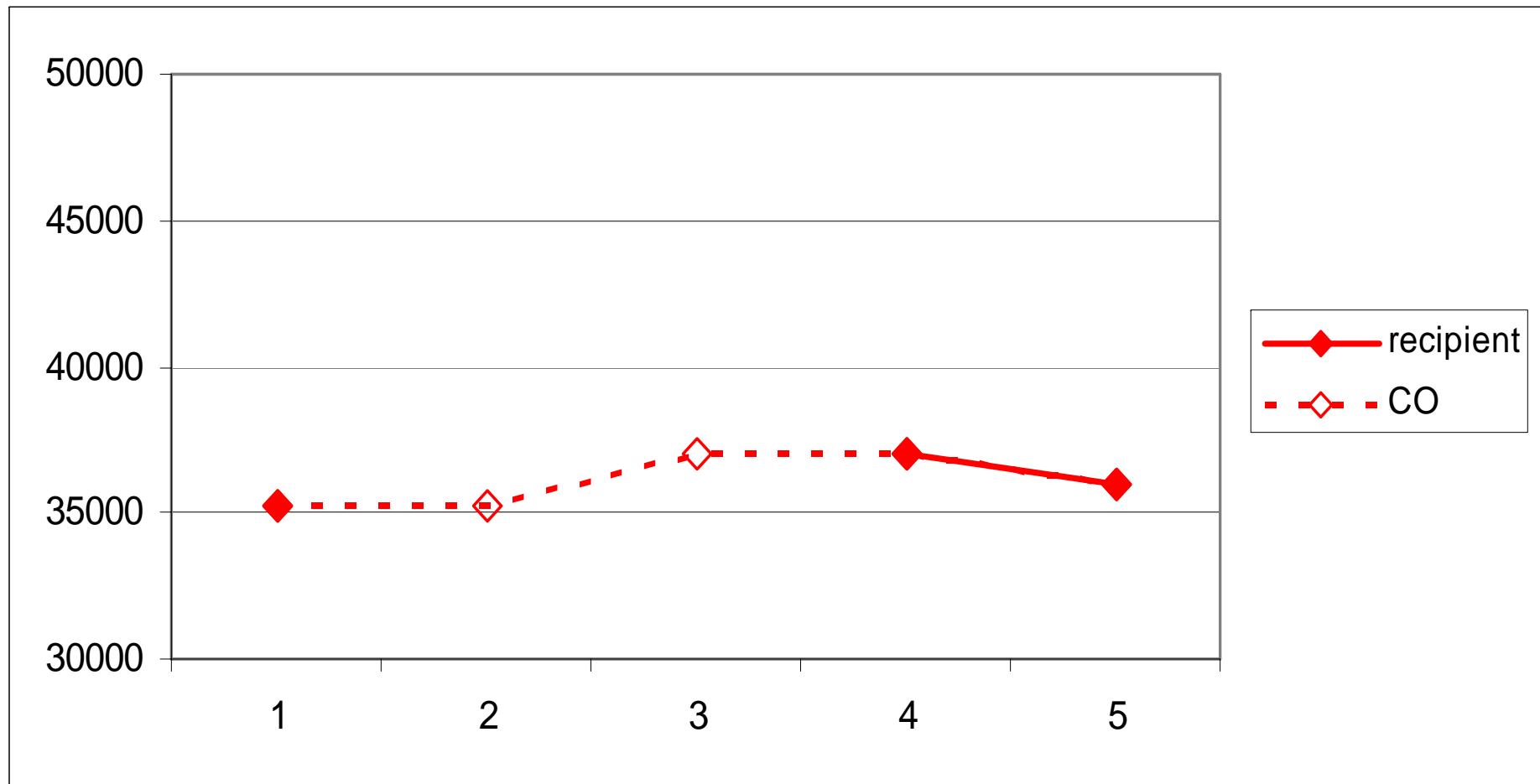
Methods: NNRM and Hotdeck



Methods: Little and Su



Methods: Carryover



Methods evaluated

	Fallback->	NNRM	Hotdeck
Longitudinal nearest neighbour regression method		✓	
Longitudinal hotdeck method			✓
Basic Little and Su		✓	✓
Little and Su with imputation classes		✓	✓
Little and Su via key variable		✓	✓
Little and Su via distance function		✓	✓
Last value carried forward		✓	✓
Random carryover method		✓	✓
Population carryover method		✓	✓

Results (1)

- No one method stands out
- Carryover methods
 - Good for level estimates but generally very poor at preserving distribution of change
 - Population Carryover better theoretically, but rarely statistically better than Random Carryover
 - Both better than the Last Value Carried Forward
 - For non-respondents, carryover methods were often better than other methods (zero/non-zero problem?)

Results (2)

- Little and Su methods
 - Generally better at preserving change between waves
 - Provided reasonably level estimates
 - Imputation classes helped where good association with component
 - Otherwise basic method was better
 - No real gains from extra complexity of multivariate options

Results (3)

- Longitudinal NNRM and Hotdeck methods
 - Good for level estimates for wages and salaries and for benefits
 - Poor for change estimates
- Fallback method
 - NNRM and Hotdeck often indistinguishable
 - When used more often
 - NNRM slightly better on distributional accuracy
 - Hotdeck slightly better on predictive accuracy

Recommended methods

Variable	Respondents	Non-respondents
Wages and salaries	LS with imp class	LS with imp class
Gov't benefits / pen's	LS with imp class	LS with imp class
Business income	Basic LS	Pop. Carryover
Interest income	Basic LS	Pop. Carryover
Dividends / royalties	Basic LS	Pop. Carryover
Rental income	Basic LS	Pop. Carryover
Private transfers	Basic LS	Pop. Carryover
Total FY income	Total of components	Total of components

Further work

- Alternative imputation classes for Little and Su
- Investigate other imputation methods
- Vary response mechanism