Inferring the distribution of financial household wealth from capital income and, scarce, wealth data.

Raun van Ooijen [a,b] Rob Alessie [a,b]

[a] *University of Groningen, the Netherlands*
[b] *Netspar, the Netherlands*

Netspar Pension Day, 14 October 2016
Preliminary

university of groningen

Motivation

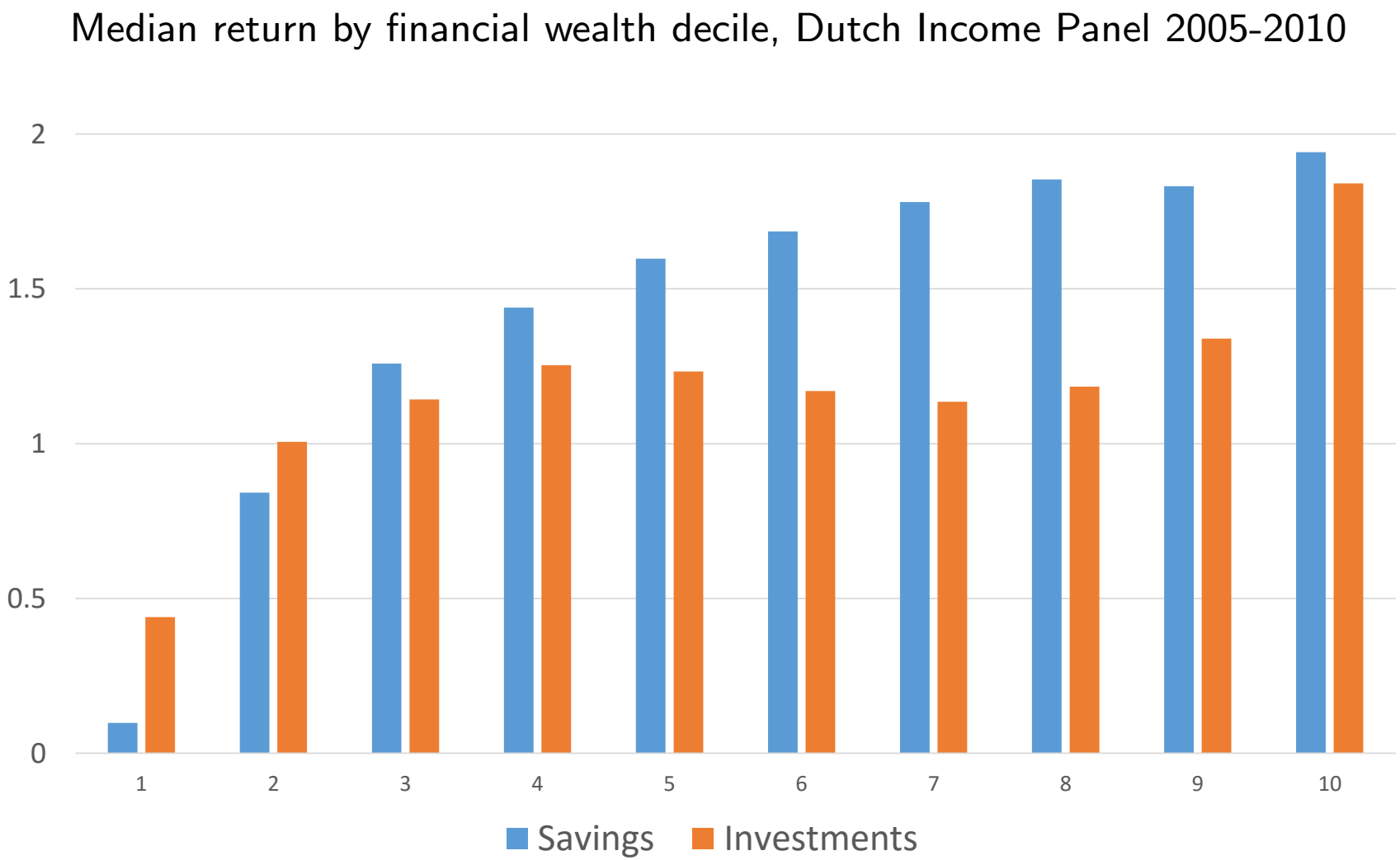
- There is a growing interest in household wealth data (Piketty) to examine trends in inequality and saving behavior.
- Requires reliable (longitudinal) micro data on household wealth.
- Survey datasets like GSOEP and SCF do not cover well the top of the wealth distribution.
- Surge in the use of administrative tax records on wealth which are only available in countries with a comprehensive wealth tax (not in the US)
- Saez and Zucman (2016) reintroduced the capitalization method (see e.g. Greenwood 1980, Wolff, 1983), to infer the value of assets from
  1. Administrative micro data on different sources of capital income from U.S. income tax data (interest income, dividends etc.) ,.
  2. Asset class specific *capitalization factors*

Capitalization method

$$\overline{\text{Return}} = \frac{\sum_i^N \text{Interest}_i}{\sum_i^N \text{Savings deposits}_i}$$

- Capitalization factor for saving accounts: $1/\overline{\text{Return}}$
- Capitalization factor obtained from the household Flow of Funds statements of the National Accounts
- Predicted saving account balance for household i: $\frac{\text{Interest}_i}{\overline{\text{Return}}}$
- Assumption: Within a given asset class, all households have the same return.
- Overestimates inequality if the return on a type of asset increases in wealth.

Heterogenous median returns



This paper

- We use data from the Dutch Income Panel 2001-2010 (IPO)
  - Representative sample of approximately 100,000 reference persons (age 18+) and members of the person's household.
  - Administrative data on
    1. Capital income (e.g.interest income, bond yields and dividends provided by Financial Institutions) and other sources of income
    2. Dummy variable indicating whether the households pays wealth tex ('box 3')
    3. Self-reported amounts in (1) saving accounts, 2) investments (stock, bonds and mutual funds) for households who are not exempt from wealth ('box 3') taxes.
    4. Some demographic characteristics
    5. From 2005 onwards, Assets and liabilities (source: Financial Institutions)

- Goal paper: predict for 2001-2004 the amounts invested in 1) saving accounts and 2) investments

- We propose an alternative for the capitalization method which only uses admin data on capital income (besides the capitalization fsctors)

- Our method uses all relevant information available in IPO (see point 1, 3 and 4 above)

- Moreover, our method exploits the panel feature of the IPO data: we use info

Our method has several advantages:

- Unobserved heterogeneity is taken into account
- Exploits the large persistence in wealth by "working backward"
- Captures heterogeneity by using a rich set of controls
- Does not rely on uniform implied rates of return

Wealth tax in the Netherlands

- $\geq$ 2001 Wealth tax of 1.2% on individual held assets except owner occupied housing. Capital income not taxed
- Wealth exemption of euro 24,437 (euro 48,848 for couples).

Dutch income panel 2001 – 2010

- See above
- Financial assets:
  - **savings** (checking and savings accounts and deposits);
  - **risky assets** (stocks, bonds and mutual funds).
  - Non-financial assets: business equity and stocks from a substantial holding, real estate, owner occupied housing.
  - Mortgage debt and other debt.

Non-exempt households: Filed Tax-form (box 3) *2001 - 2010*	All households: Records from financial institutions *2001 - 2010*	*2005 - 2010*
Savings (reported)	Interest	Savings (actual)
Risky assets (reported)	Dividend & Bond yield	Investments (actual)

Actual versus reported amounts (1)

- Let

$bankteg_{it}$= reported saving account balance
$balsav_{it}$= actual saving account balance
$y^*_{it} = balsav_{it} - bankteg_{it}$

$risky\_assets_{it}$= reported amount invested in risky assets
$balshabon_{it}$= actual amount invested in risky assets
$z^*_{it} = balshabon_{it} - risky\_assets_{it}$

Actual versus reported amounts (2)

Distribution of y_{it}^* (saving accounts) and z_{it}^* (risky assets) in 2005

Percentiles	Distribution of y_{it}^* $bankteg_t = 0$	$bankteg_t > 0$	Percentiles	Distribution of z_{it}^* $risky\_assets_t = 0$	$risky\_assets_t > 0$
10%	0	0	10%	0	0
25%	1114	0	25%	0	0
50%	5832	3863	50%	0	0
75%	16607	17633	75%	0	6330.5
90%	29745.67	40953	90%	3783	29999.32
Mean	11233.93	11882.62	Mean	2401.10	9907.77
Obs	63471	25147	Obs	75194	13424

Actual versus reported amounts (3)

Differences between actual and reported wealth

Year	$bankteg_t = 0$			$bankteg_t > 0$		
	Obs $y_{it}^* < 0$	Obs $y_{it}^* = 0$	Obs $y_{it}^* > 0$	Obs $y_{it}^* < 0$	Obs $y_{it}^* = 0$	Obs $y_{it}^* > 0$
2005	0	10355	53116	223	7528	17396
2006	0	10464	53034	136	7945	18231
2007	0	9591	53406	41	8296	19370
2008	0	9245	54554	39	8589	19511
2009	0	5483	60280	2221	6095	19784
2010	0	5099	62805	10182	7478	12313

(a) Checking and savings accounts

Year	$risky\_assets_t = 0$			$risky\_assets_t > 0$		
	Obs $z_{it}^* < 0$	Obs $z_{it}^* = 0$	Obs $z_{it}^* > 0$	Obs $z_{it}^* < 0$	Obs $z_{it}^* = 0$	Obs $z_{it}^* > 0$
2005	0	63032	12162	639	7849	4936
2006	0	64741	11543	576	7748	5202
2007	0	66478	10809	510	7901	5006
2008	0	68669	10506	411	7527	4825
2009	0	69968	11361	1235	5307	5992
2010	0	74623	10771	5027	4955	2501

(b) Risky Assets

Two part model: part 1

Probit model to predict any discrepancy between the reported amount and the actual amount.

$$P(y_{it} = 1|\boldsymbol{x}_{it}) = \Phi(\boldsymbol{f}'_{it}\boldsymbol{\beta} + \boldsymbol{s}'_{i,t+1}\boldsymbol{\lambda} + y_{it+1}\gamma),$$

$$y_{it} = \begin{cases} 1 & \text{if actual amount}_{it} = \text{self-reported amount}_{it} \\ 0 & \text{if actual amount}_{it} \neq \text{self-reported amount}_{it} \end{cases}$$

Variables:

- $\boldsymbol{f}'_{it}$ = Income derived from owning an asset class;
  Self-reported amount of an asset class,
- $\boldsymbol{s}_{i,t+1}$ = Lead value of the actual amount of an asset class.

Imputation of the value of a financial asset

Predict the value of savings in case of a discrepancy ($Y_{it} = 0$).

$$\ln s_{it}^{*} = f_{it}'\beta + \boldsymbol{s}_{i,t+1}^{\prime *}\boldsymbol{\lambda} + \boldsymbol{d}_{it}{}'\boldsymbol{\delta} + \alpha_i + v_{it}$$

$$\mathrm{E}(\alpha_i|\boldsymbol{z}_i) \neq 0, \qquad v_{it} = \rho v_{i,t+1} + e_{it}; \qquad e_{it} \sim \mathcal{N}(0, \sigma_{e,it}^2).$$

Variables:

- $\ln s_{it}^{*}$ = Log amount of an asset class;
- $\boldsymbol{f}_{it}'$ = Income derived from owning an asset class;
  Self-reported amount of an asset class;
- $\boldsymbol{s}_{i,t+1}$ = Lead value of the actual amount of an asset class;
- $\boldsymbol{d}_{it}{}'$ = Socio-demographic variables.

Imputation of the value of a financial asset

Backward imputation for the years 2001 – 2004:

$$\hat{s}_{i,t} = \exp\left(\boldsymbol{d}'_{i,t}\hat{\boldsymbol{\delta}} + f'_{i,t}\hat{\boldsymbol{\beta}} + \boldsymbol{s}'_{i,t+1}\hat{\boldsymbol{\lambda}} + \hat{a}_i + \hat{v}_{it}\right).$$

Account for **heteroscedasticity** and **serial correlation**.

Predicted and empirical distribution of savings in 2005

	Non-Exempt (Box 3)			Exempt		
	s_t	$\hat{s}_t$	Capital-ization	s_t	$\hat{s}_t$	Capital-ization
5%	8.3	8.0	2.9	0	0	0
10%	17.4	15.9	8.5	0	0	0
25%	38.0	37.7	26.6	1.4	1.8	0
50%	71.2	70.9	57.9	6.7	7.6	4.0
75%	123.5	123.7	114.1	18.6	18.5	14.5
90%	209.2	212.6	221.3	34.0	33.6	30.7
95%	301.2	308.8	341.6	43.7	46.2	43.2
Mean	106.8	107.7	104.8	12.9	13.6	11.3
Std.Dev.	167.2	176.5	195.2	19.0	20.5	23.7
Skewness	14.5	18.1	12.3	8.8	16.1	15.5
Kurtosis	419.4	684.8	325.0	239.1	1149.5	629.3
Spearman's ρ	-	0.92	0.68	-	0.78	0.82

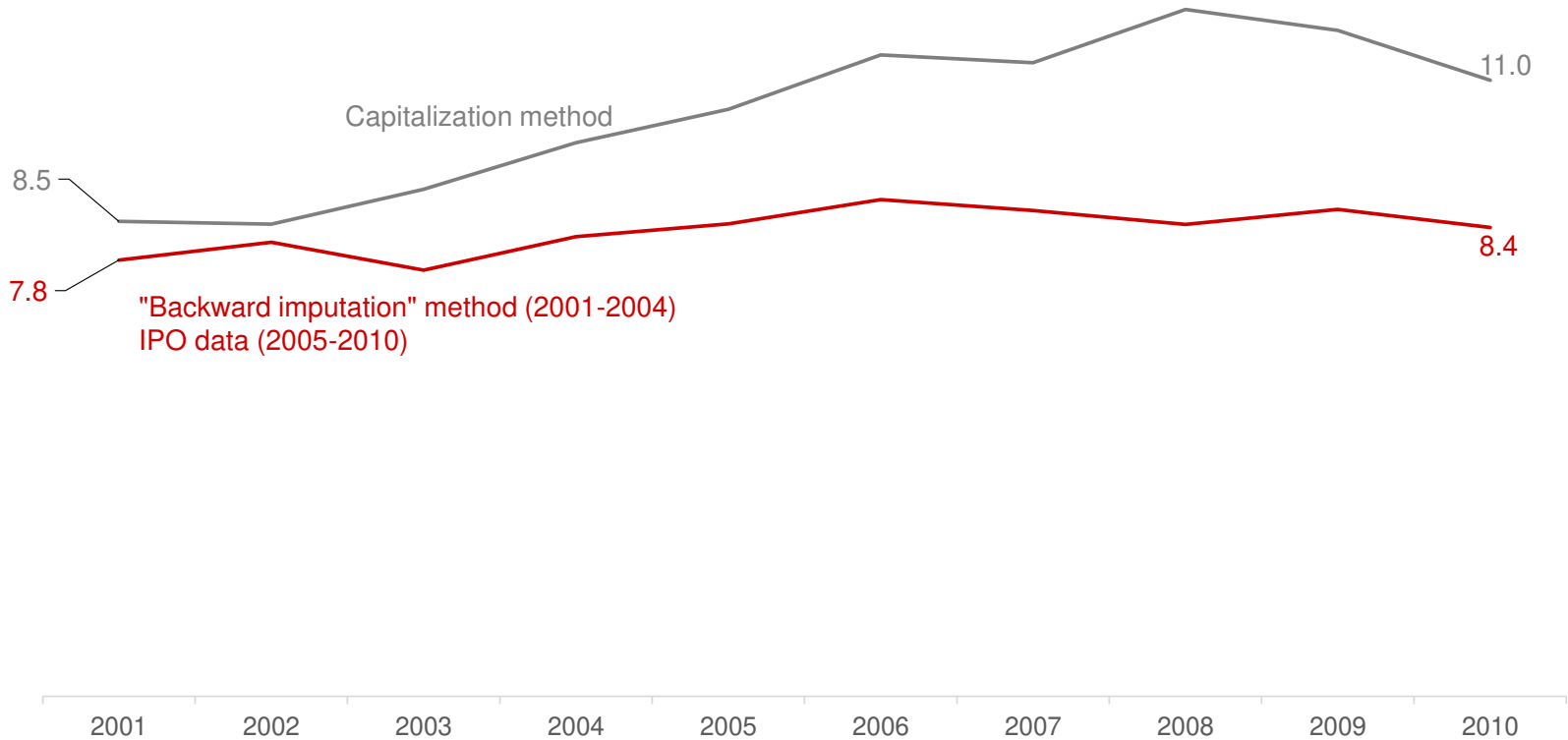
Pseudo R^2 of predicted discrepancy $= 0.339$

Predicted and empirical distribution of investments in 2005

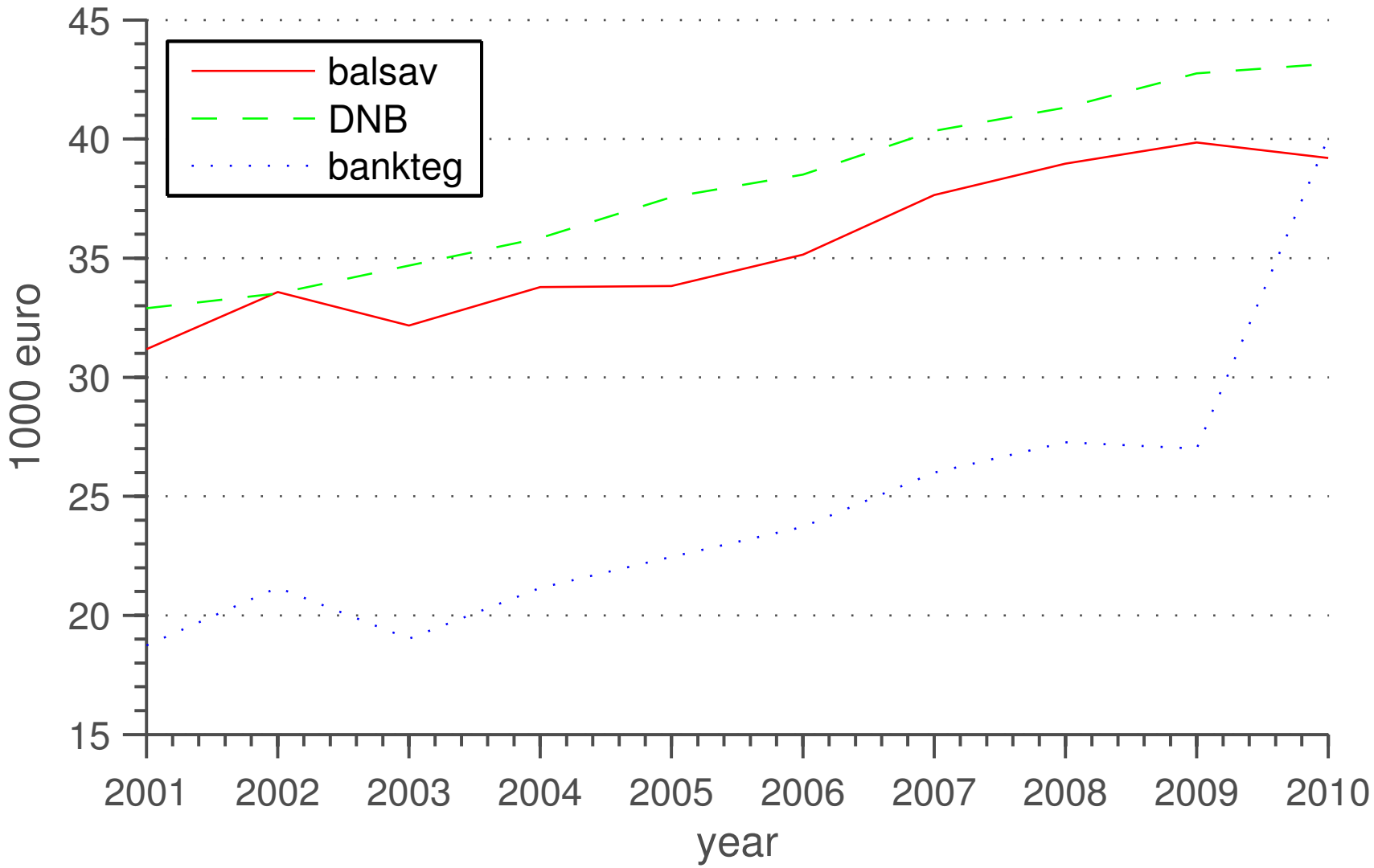
	Non-Exempt (Box 3)			Exempt		
	s_t	$\hat{s}_t$	Capital-ization	s_t	$\hat{s}_t$	Capital-ization
5%	0	0	0	0	0	0
10%	0	0	0	0	0	0
25%	0	0	0	0	0	0
50%	10.7	10.5	2.3	0	0	0
75%	60.7	60.1	32.2	0	0	0
90%	163.9	163.8	124.2	6.1	5.0	2.7
95%	290.0	310.5	278.4	15.1	12.7	9.0
Mean	75.3	79.7	66.3	2.9	2.7	2311
Std. Dev.	267.1	324.7	299.7	24.2	23.3	54.0
Skewness	12.3	14.9	14.3	65.6	99.2	131.1
Kurtosis	226.2	315.0	300.7	6575.1	15033.7	19971.1
Spearman's ρ	-	0.96	0.68	-	0.81	0.82

Pseudo R^2 of predicted discrepancy $= 0.504$

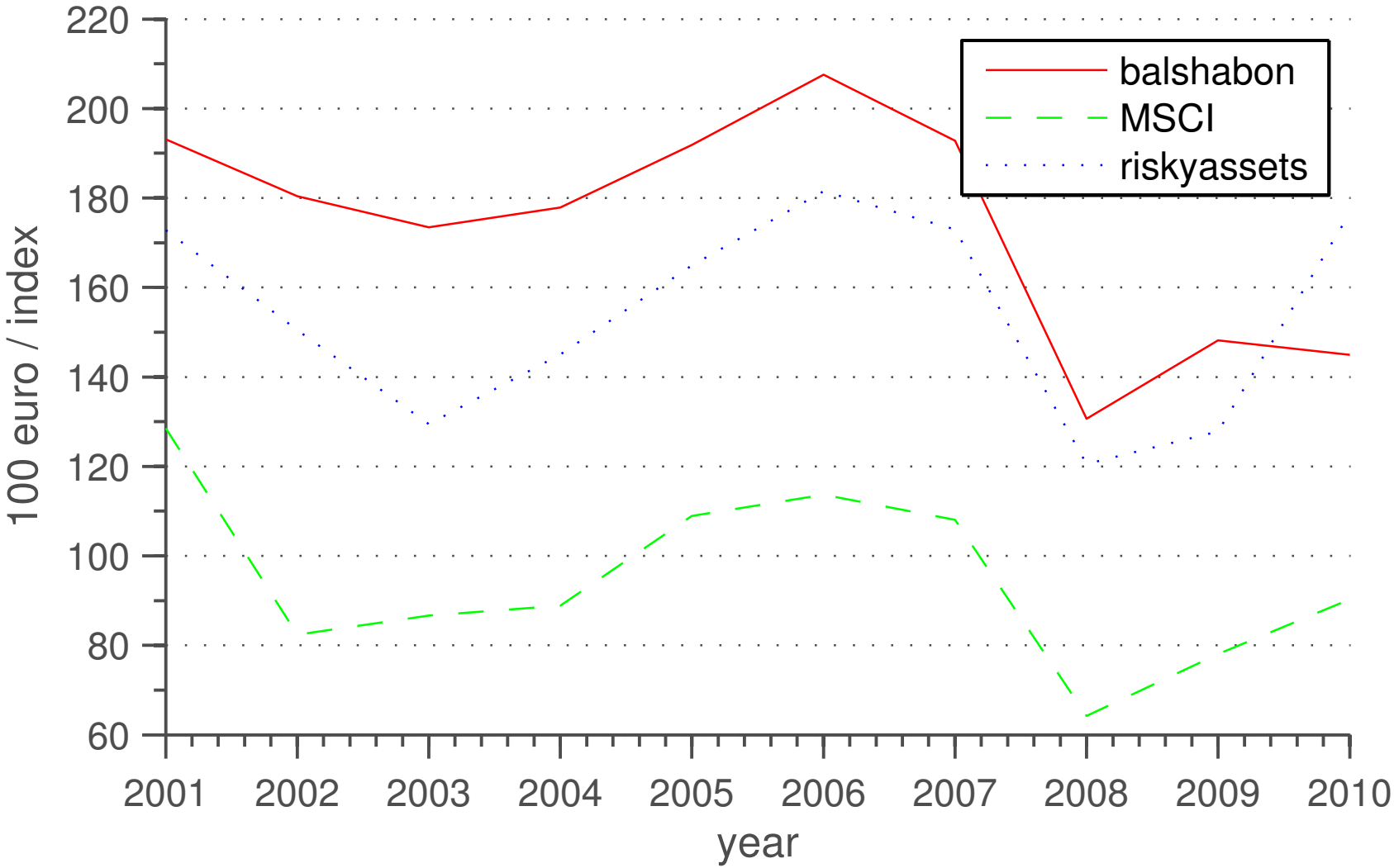
Trend in wealth inequality: p90/p50 ratios



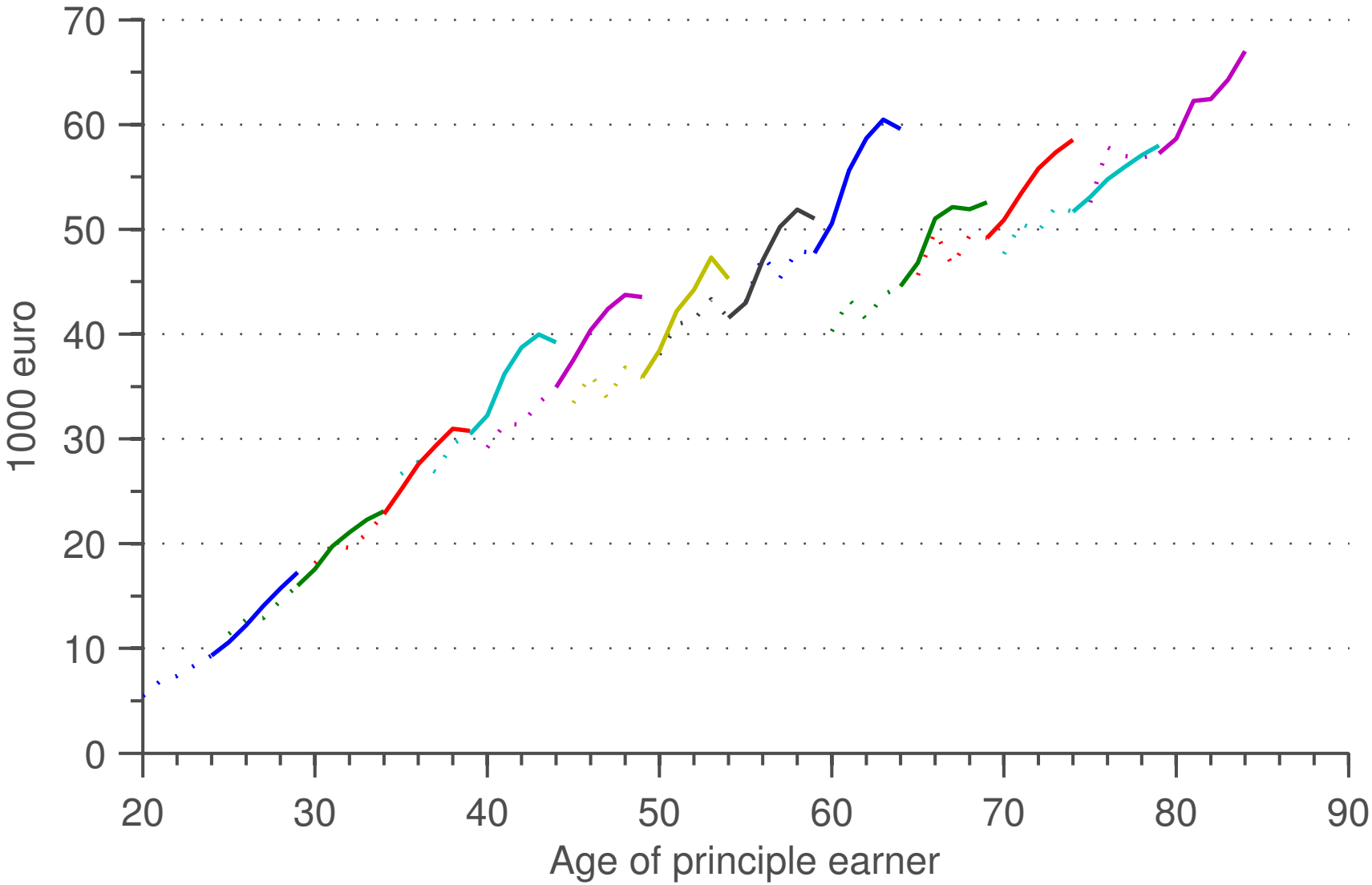
Verification of trends with external sources – checking and saving accounts



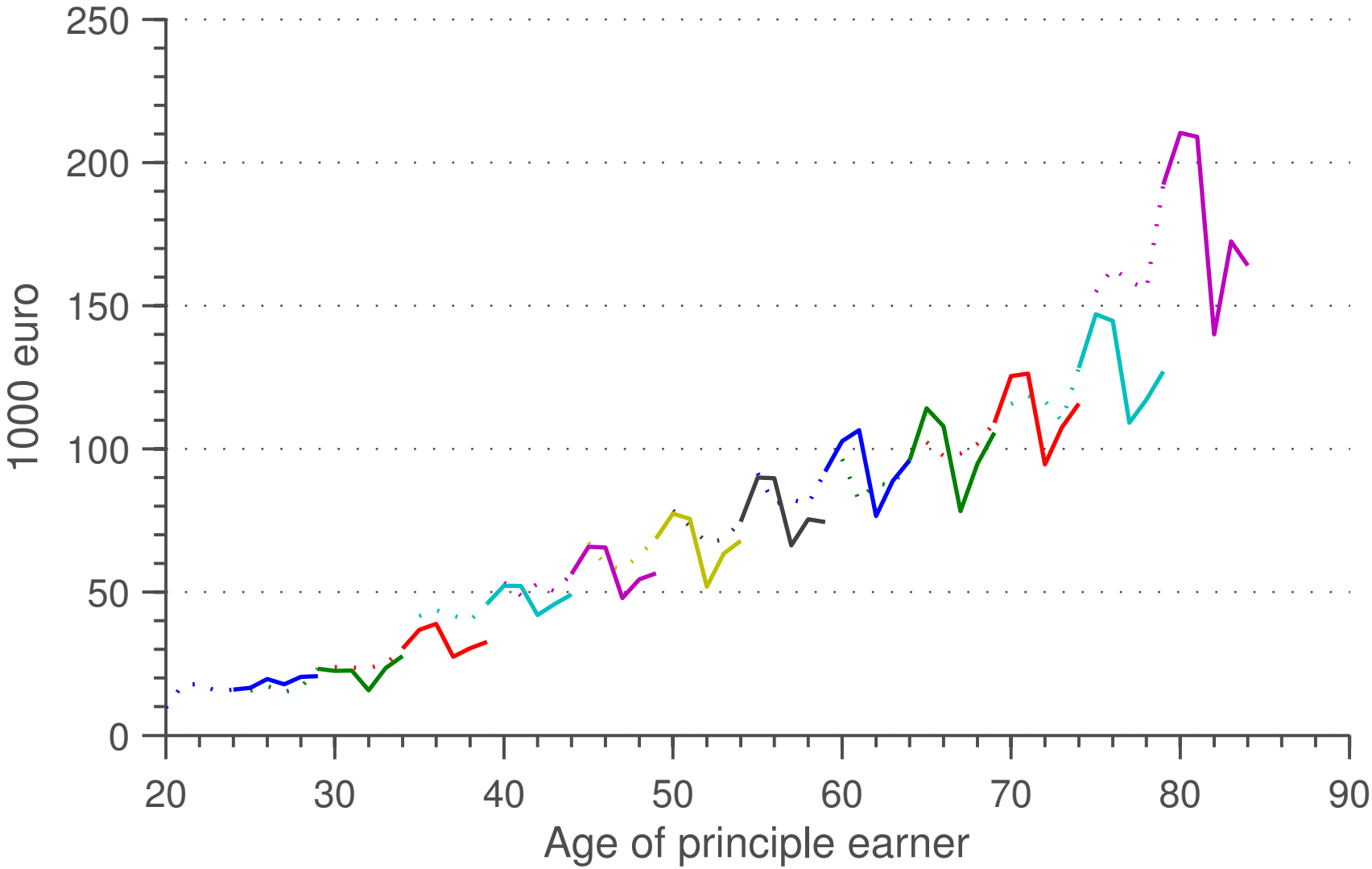
Verification of trends with external sources – risky assets



life cycle patterns



Verification of trends with external sources – and life cycle patterns



Conclusion

- Impute previous panel waves of financial wealth, using tax records on capital income AND reported wealth ('box 3').
- In-sample verification, and comparison with external sources, indicates that the method works well.
- The level of and trend in (our estimate of) inequality seems more moderate than an estimate based on the capitalisation method.
- An adapted version of our model can be used for imputing wealth in surveys with infrequent wealth modules (SOEP, early PSID, HILDA...)
- Things to do
  - Adapt the econometric model:
    - remove the rhs variable s_{it+1} (Lead value of the actual amount of an asset class)
    - The in-sample verification suggests estimation of the highest percentiles of the distribution of risky assets
  - Backward imputation of the asset items (business equity, stocks of substantial holding