

# Dealing with incomplete household panel data in microsimulation models

*Joachim R. Frick, Markus M. Grabka, Olaf Groh-Samberg  
(SOEP, DIW Berlin)*

2nd General Conference of the International Microsimulation Association (IMA)  
"Microsimulation: Bridging Data and Policy"  
Ottawa, Canada, June 8-10, 2009

# Motivation

- **Household panel surveys**
  - increasingly relevant source for static and dynamic micro-simulation
    - Interaction between HH members: e.g. labor supply decisions of couples
    - Aggregation at HH-level, e.g. household income aggregates
  - collecting complete info from all adult respondents in HH panels is crucial
- **But: Partial Unit-Non-Response (PUNR)**
  - Incomplete coverage of multi-respondent households (*no problem for single HH*)  
**= underestimation of household income aggregates !**
- **Open questions**
  - How does PUNR affect cross-sectional and longitudinal analyses ?
  - How can we deal with PUNR, esp. in panel surveys ?
  - Does the choice of the treatment affect micro-simulation models ?

# Dealing with PUNR

- Version 1: ignoring* PUNR for income measure, only  
 $Y(PUNR)=0$  &  $Needs(PUNR)>0$       *assumption:*  $Y(PUNR)=0$
- Version 2:* deleting non-responding individuals from PUNR-households by **adjusting equivalent scale**  
 $Y(PUNR)=0$  &  $Needs(PUNR)=0$       *assumption:*  $Y(PUNR) \sim Y(noPUNR)$  within PUNR-HHs
- Version 3: deleting* PUNR-households altogether **and re-scaling** of weights for households at risk of PUNR  
 $PUNR-HH=.$       *assumption:*  $Y(PUNR-HH) \sim Y(noPUNR-HH | \#adults>1)$
- Version 4: imputation* of annual incomes of PUNR  
 $Y(PUNR)>0$  &  $Needs(PUNR)>0$       *assumption:*  $Y(PUNR)=f(X)+e$

# Outline

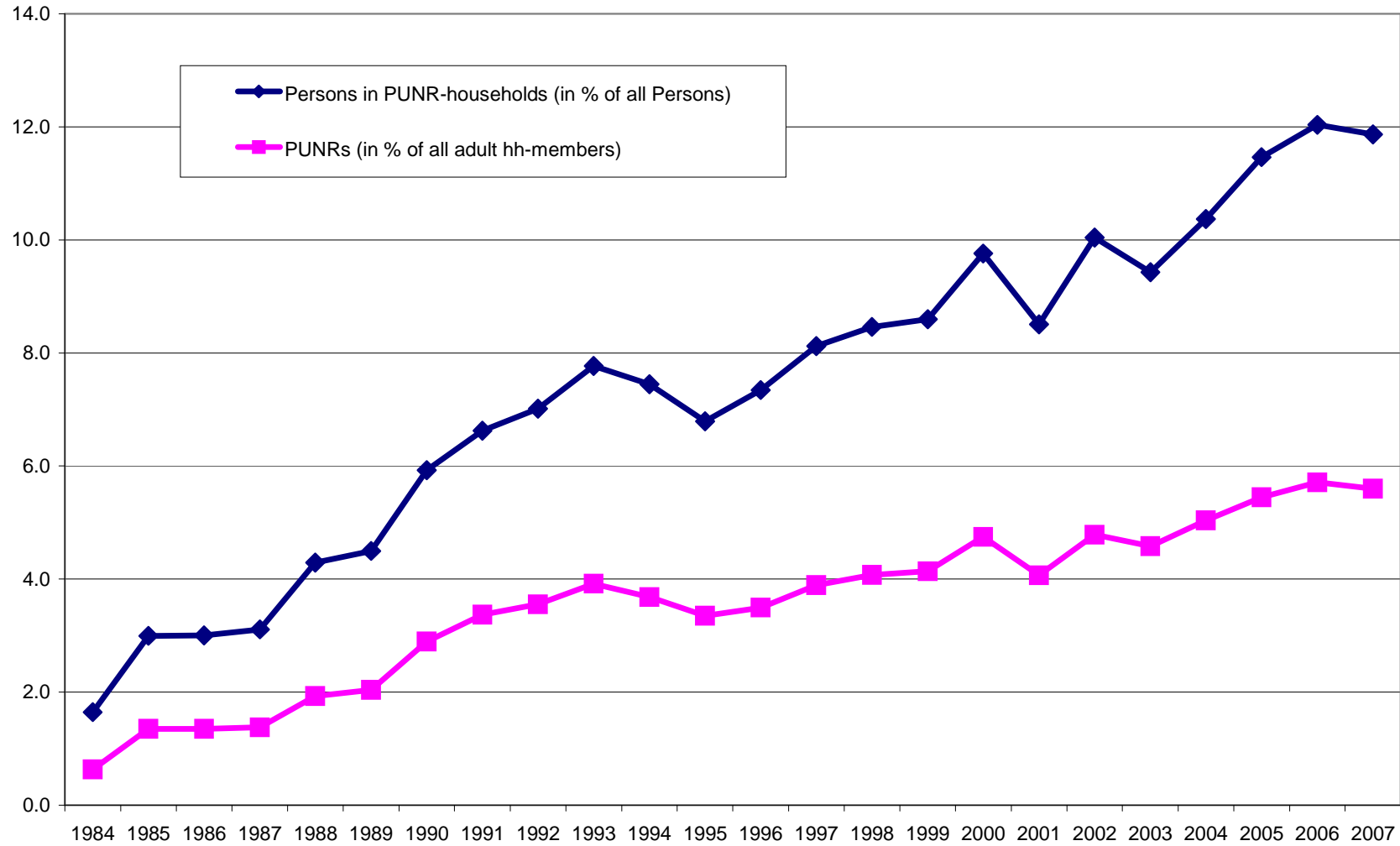
1. Incidence and Selectivity of PUNR in the German SOEP
2. The Imputation Process
3. The Impact of PUNR-Treatment on inequality & poverty
4. PUNR-treatment in micro-simulation (child benefit reform)
5. Conclusion

# 1. Incidence and Selectivity of PUNR in the German SOEP

# The German SOEP

- started in 1984 incl. oversampling of migrants
  - additional samples started in 1990 (East), 1994/95 (new immigrants), 1998/2000/2006 (refreshment and innovation samples), 2002 (high income sample)
- annual interviews
  - household
  - all individual household members (17+)
  - mixed mode: PAPI, CAPI, self-administered
- more information: <http://www.diw.de/soep>

# Incidence of PUNR (%)



## Selectivity of PUNR – Results from a pooled probit regression

|  |  | Coeff          |
|--|--|----------------|
| sex  | female   | -0.443 ***     |
| age groups (ref. 56-65 yrs of age)                       | 17-24  | -0.003         |
|  | 25-40  | -0.053 *       |
|  | 41-55  | -0.210 ***     |
|  | 66 and over                                      | -0.304 ***     |
|  | relation to household head (ref: household head) | partner        |
|  | child  | 0.984 ***      |
|  | other  | 1.700 ***      |
| migration background (ref: autochthonous)                | native born foreigners                           | 0.204 ***      |
|  | foreign born Germans                             | -0.264 ***     |
|  | foreign born foreigners                          | -0.347 ***     |
| East Germany (ref: West Germany)                         | East Germany                                     | -0.245 ***     |
| change in household composition (ref: no change)         | change   | 0.076 ***      |
|  | first wave                                       | 0.528 ***      |
| no. of adults in household (ref: 2)                      | 3 adults   | 0.052 *        |
|  | 4+ adults  | 0.125 ***      |
| no. of children in household (ref: none)                 | 1 child  | -0.078 ***     |
|  | 2 children                                       | -0.163 ***     |
|  | 3+ children                                      | -0.170 ***     |
|  | home owner (ref: tenant)                         | home owner     |
| highest education of hh head/partner (ref: intermediate) | lower secondary                                  | 0.039          |
|  | higher secondary                                 | -0.102 **      |
|  | tertiary   | -0.124 **      |
| item non-response monthly household income (ref: no)     | yes  | 0.396          |
| Control dummies for survey year                          | yes  |                |
| Constant   |  | -1.779         |
| Pseudo-R2=   |  | 0.1568         |
| Observations / Individuals                               |  | 325414 / 45038 |

Robust standard errors. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Source: SOEP 1985-2007. Authors' calculations.

→ Indication for underreporting of economically active HH members

## 2. Imputation of PUNR

# Imputation strategy

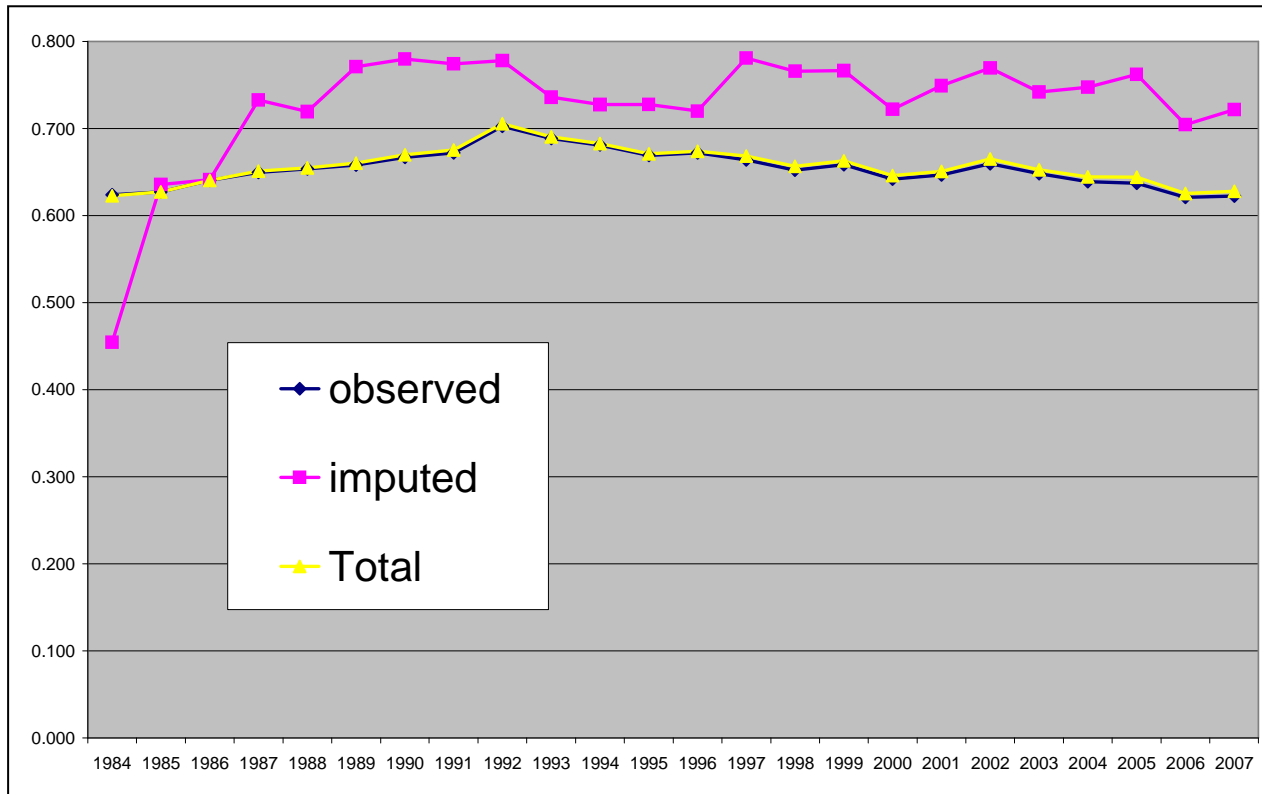
## Imputation of 6 individual gross income components

(comparable to detailed information as collected in standard questionnaire):

1. labour income (all),
2. pensions (all),
3. unemployment benefits (all),
4. student grants,
5. maternity leave transfers,
6. private transfers

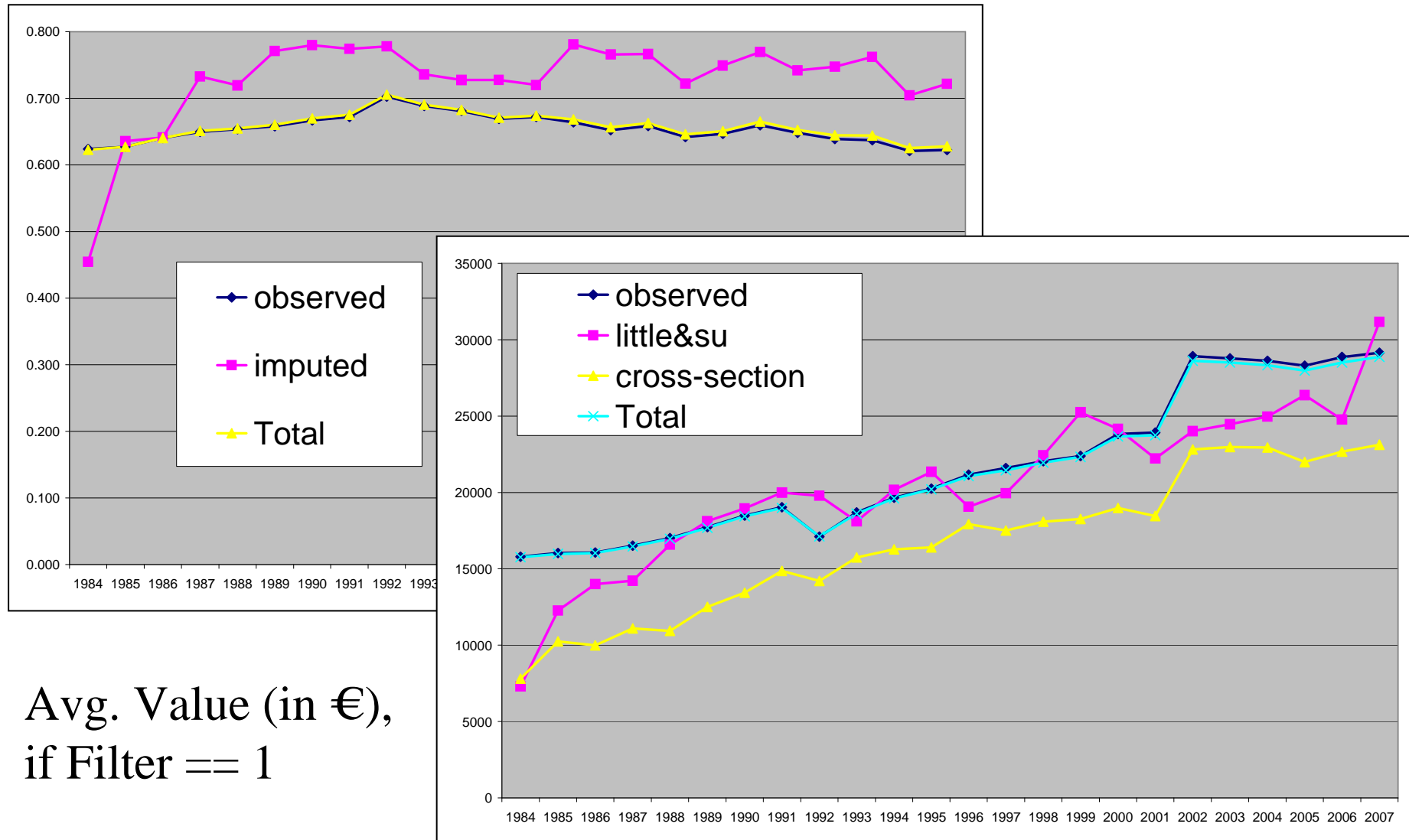
1. Step: Filter imputation ( $Y > 0$ ) using available longitudinal information
2. Step: Imputation of metric value | Filter = 1
  - a) longitudinal imputation (Little & Su 1989)
  - b) otherwise cross-sectional, regression based imputation
3. Step: Integration of (imputed) gross income components into the standard tax and SSC simulation module  
→ PUNR-adjusted Post-Government Household Income

# Imputation Results: Gross Annual Labour Income



Population  
 Share (in %)

# Imputation Results: Gross Annual Labour Income



Avg. Value (in €),  
if Filter == 1

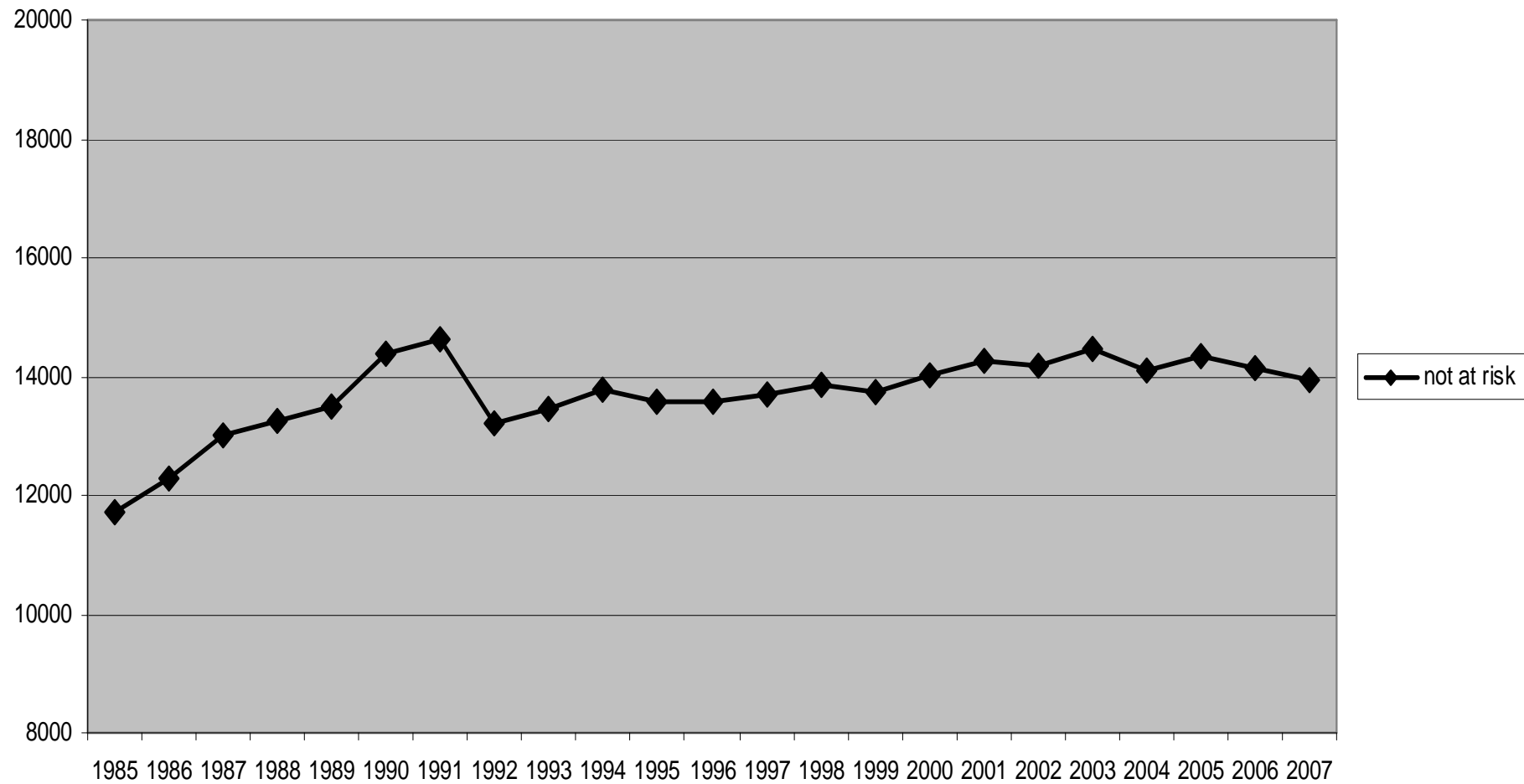
## 3. The impact of PUNR-treatment

- Relevant Income Variable: Equivalent post-government income
  - after simulation of direct taxes and SSC
  - including imputed rent
  - applying modified OECD scale (1; 0.5; 0.3)
  - in prices of year 2000
  - Relative income poverty (PL = 60 % of national median income)
- Comparison of four PUNR-treatments
  - 1: ignoring PUNR in income measure
  - 2: adjusting equivalent scale
  - 3: deleting PUNR-households and re-scaling weights for households at risk of PUNR
  - 4: imputation of annual incomes of PUNR

# (a) Income Levels

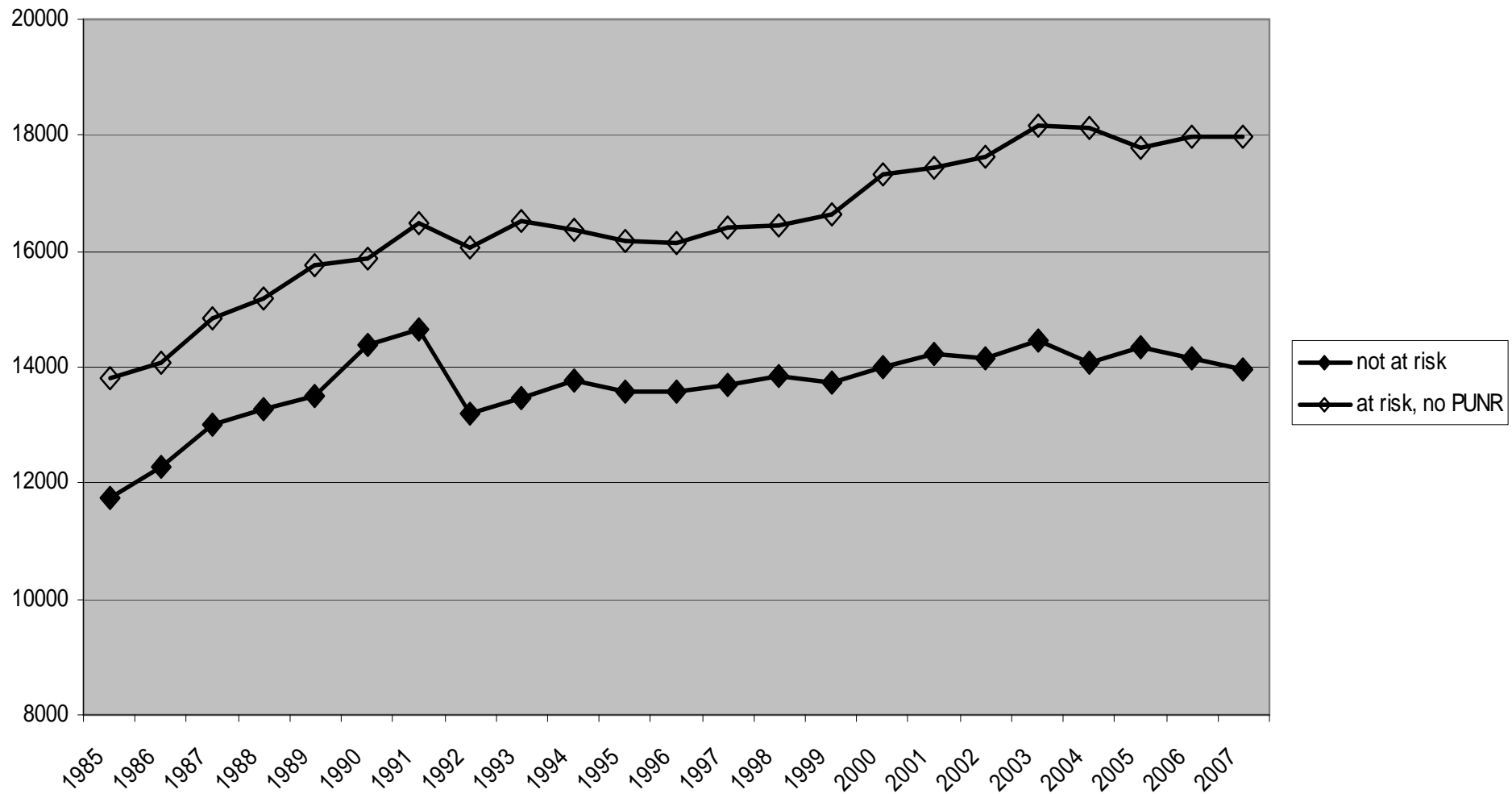
# Equivalized Post-Gov't Income by PUNR-Treatment

- Median in € -



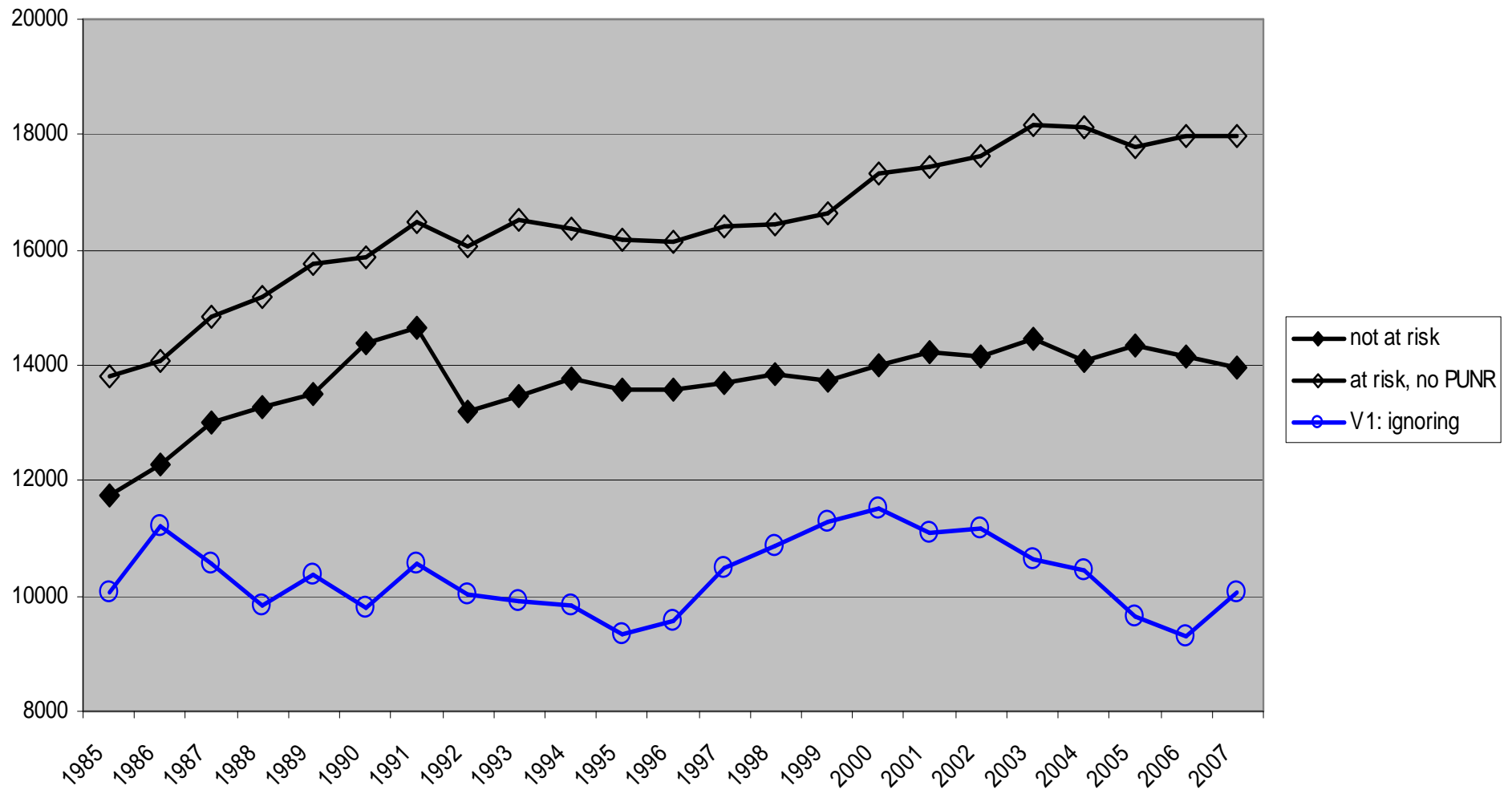
# Equivalized Post-Gov't Income by PUNR-Treatment

- Median in € -



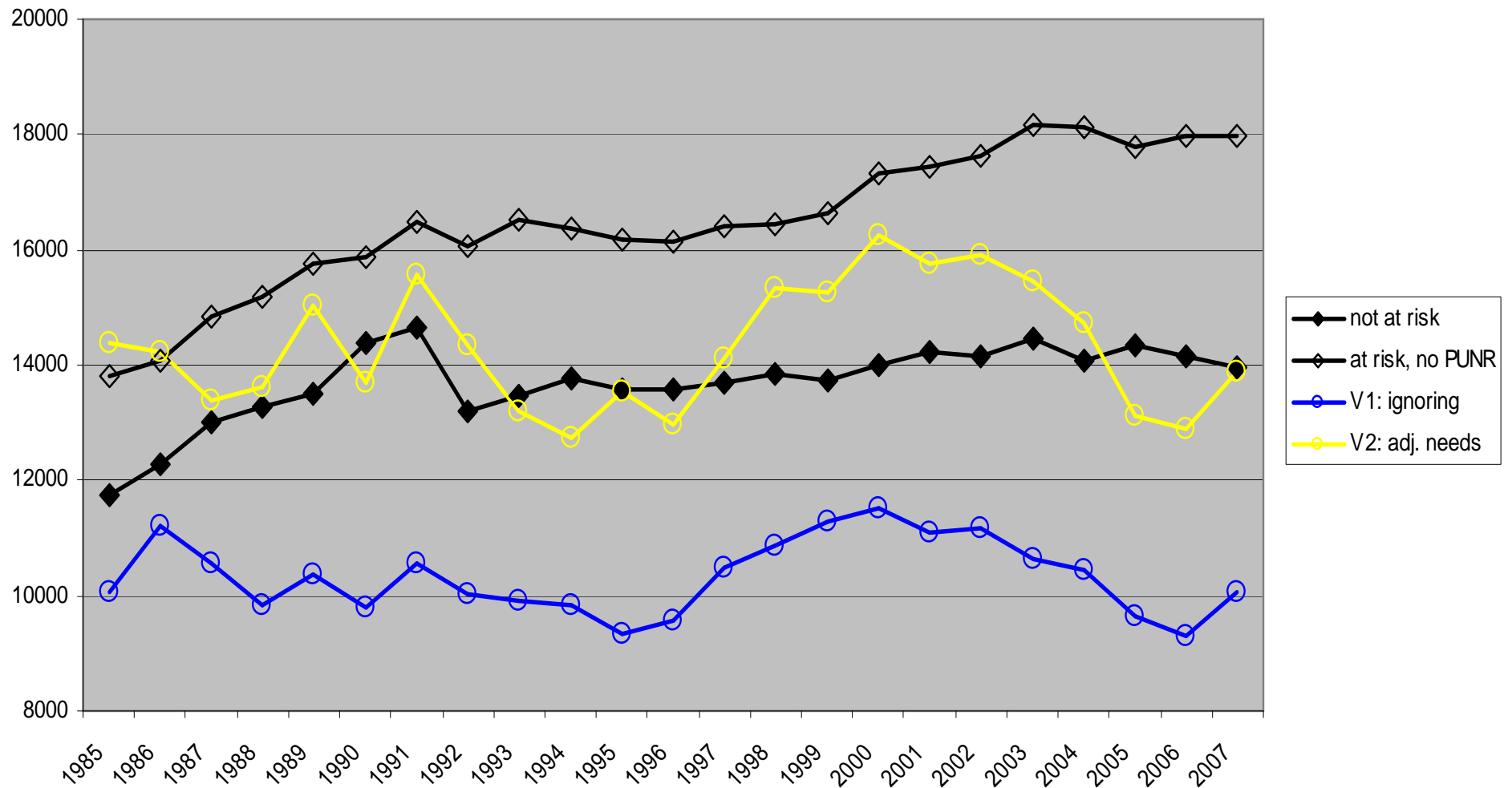
# Equivalized Post-Gov't Income by PUNR-Treatment

- Median in € -



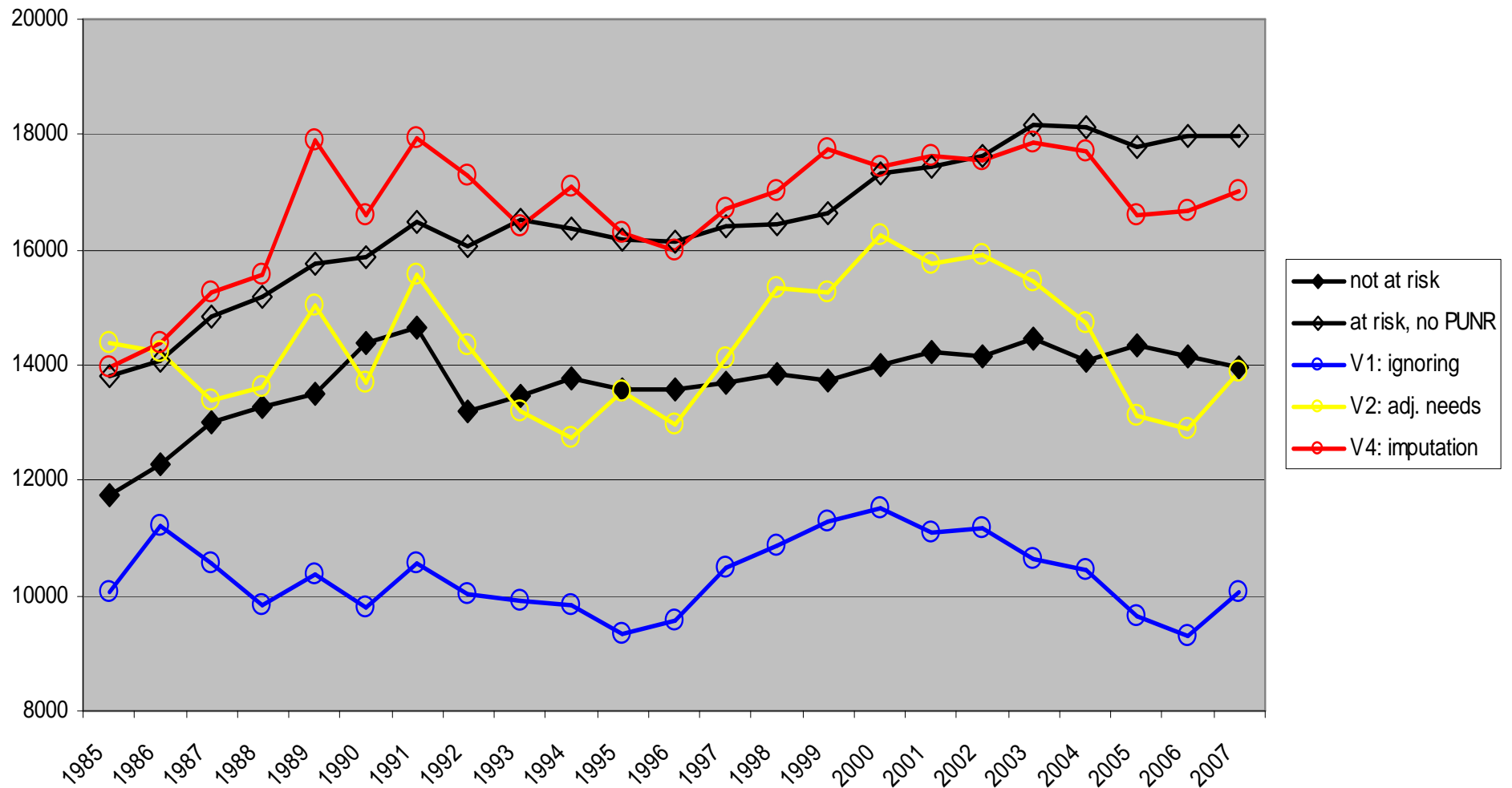
# Equivalized Post-Gov't Income by PUNR-Treatment

- Median in € -



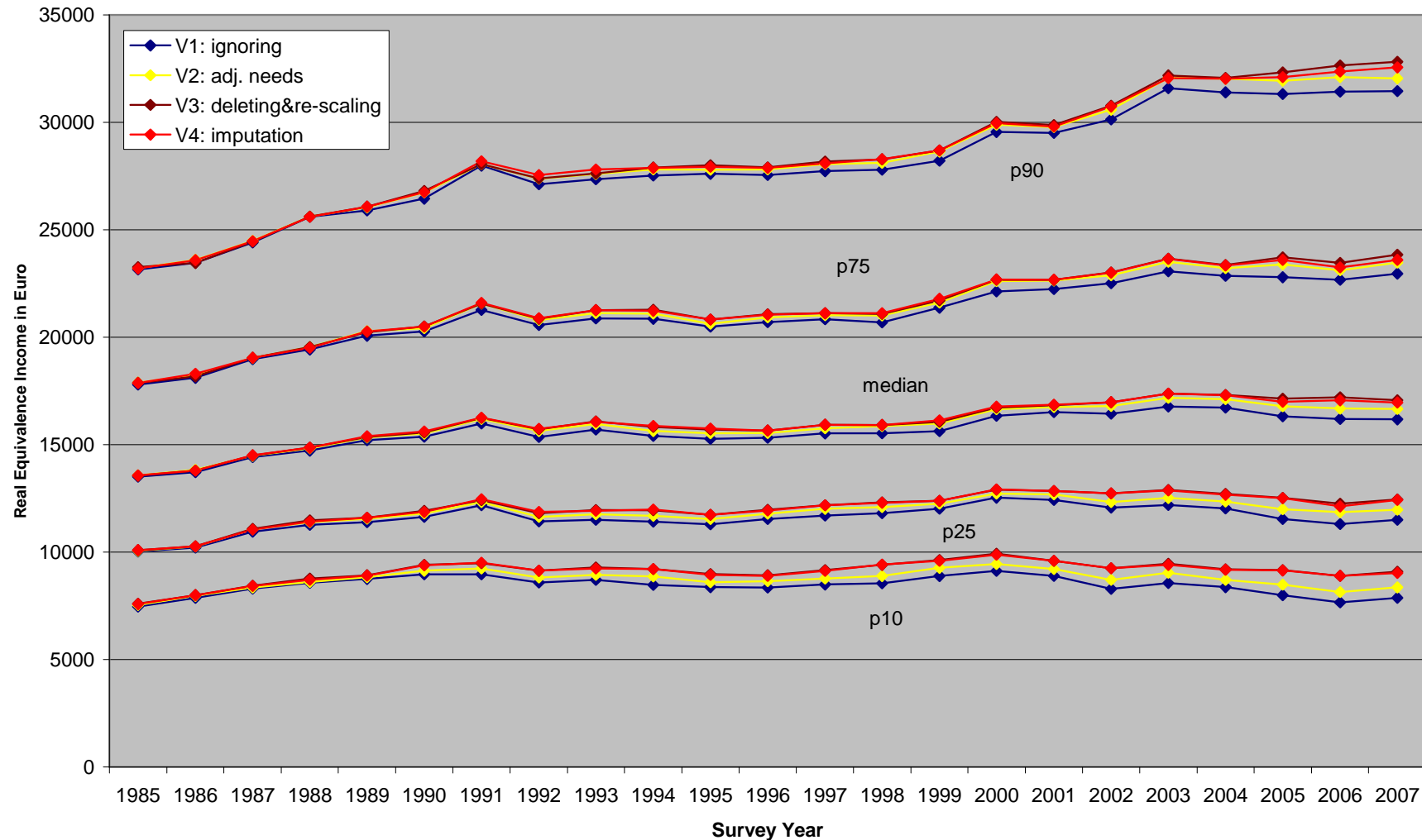
# Equivalized Post-Gov't Income by PUNR-Treatment

- Median in € -



# Equivalized Post-Govt Income by PUNR-Treatment

- Median in € - **Results for total population**



Source: SOEP 1984-2007 (until 1991 West Germany, only). In prices of 2000.

## (b) Inequality & Poverty

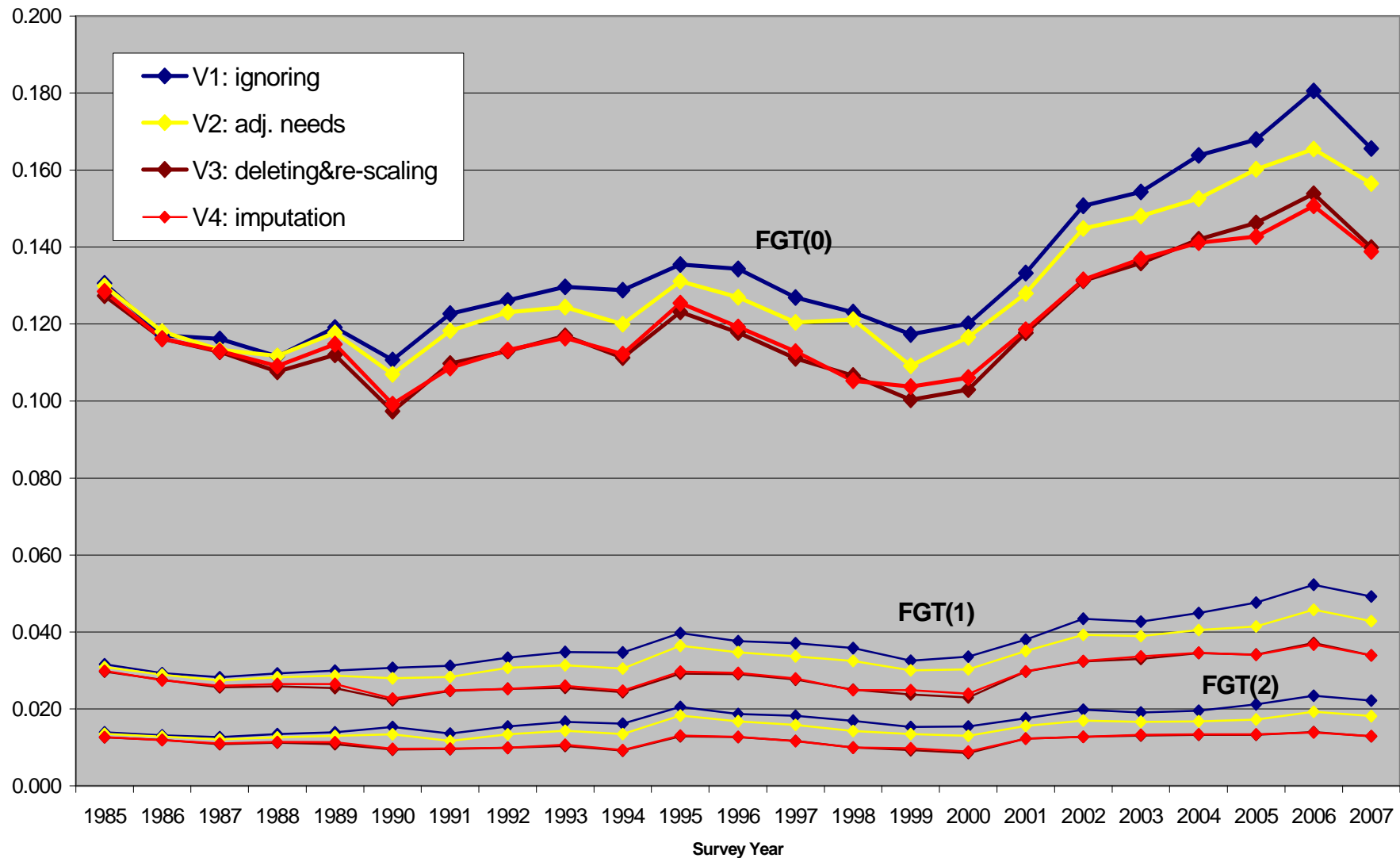
# Hypothetical Effects of PUNR on Relative Poverty and Inequality

- any accounting for PUNR yields *higher* average incomes
    - increase in relative poverty threshold
    - c.p. increasing poverty (risk) among non-PUNR-HH
  - versions 2 (adjust equiv. scale) and 4 (imputation) yield higher equivalent income among PUNR-HH
    - c.p. decreasing poverty risk rates for PUNR-HH
- overall effect on poverty at a *given point in time* unclear
- however, due to increase in the incidence of PUNR over time, *poverty trends* will be affected as well
  - effects on the *socio-demographic structure* of poor HH:
    - increase in poverty for all HH not at risk of PUNR (singles, lone parent families) due to increase in Pov-line



# Poverty by PUNR-Treatment

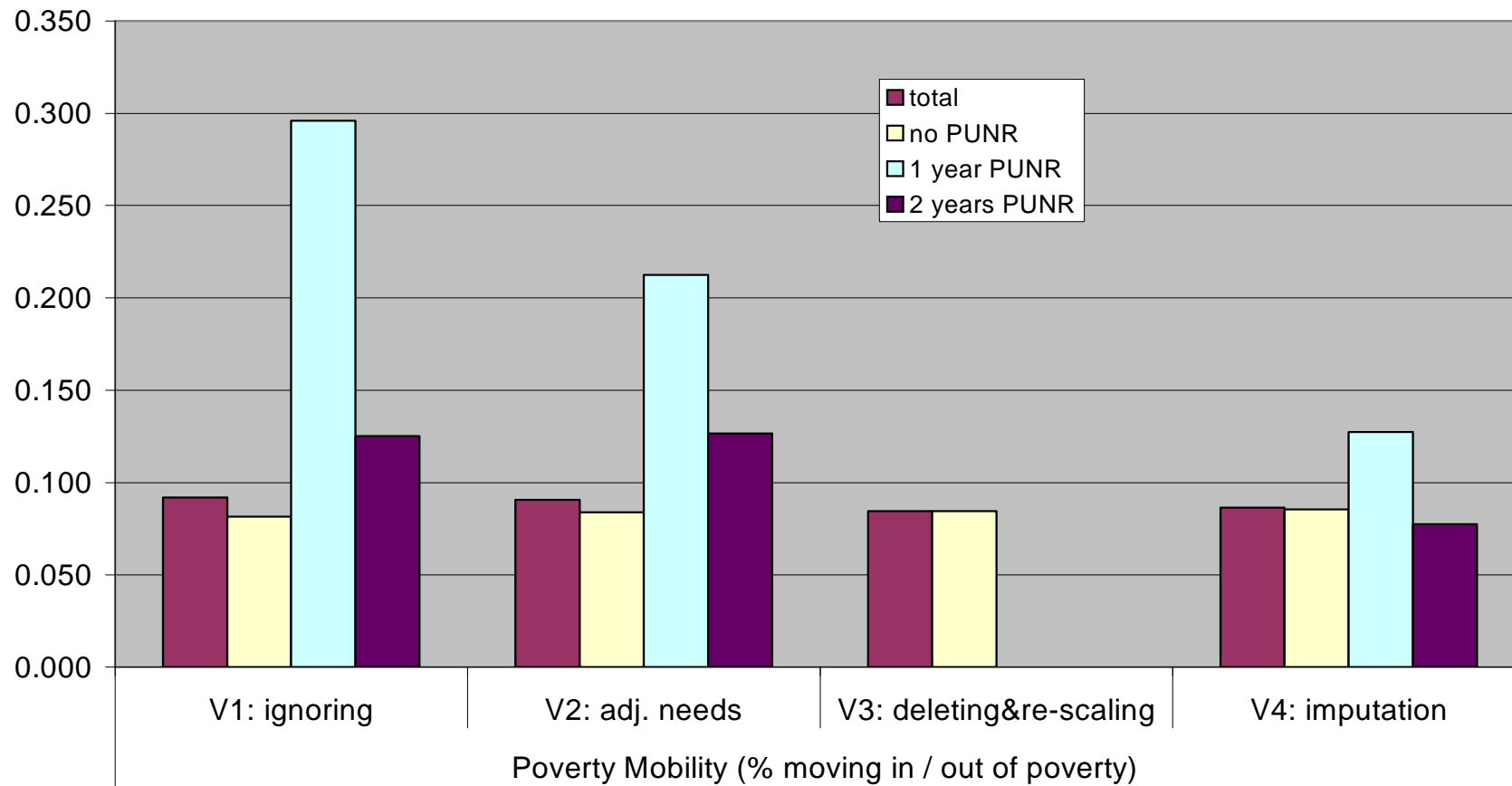
## Poverty Measures



## (c) Poverty Mobility

# Poverty Dynamics by PUNR-treatment

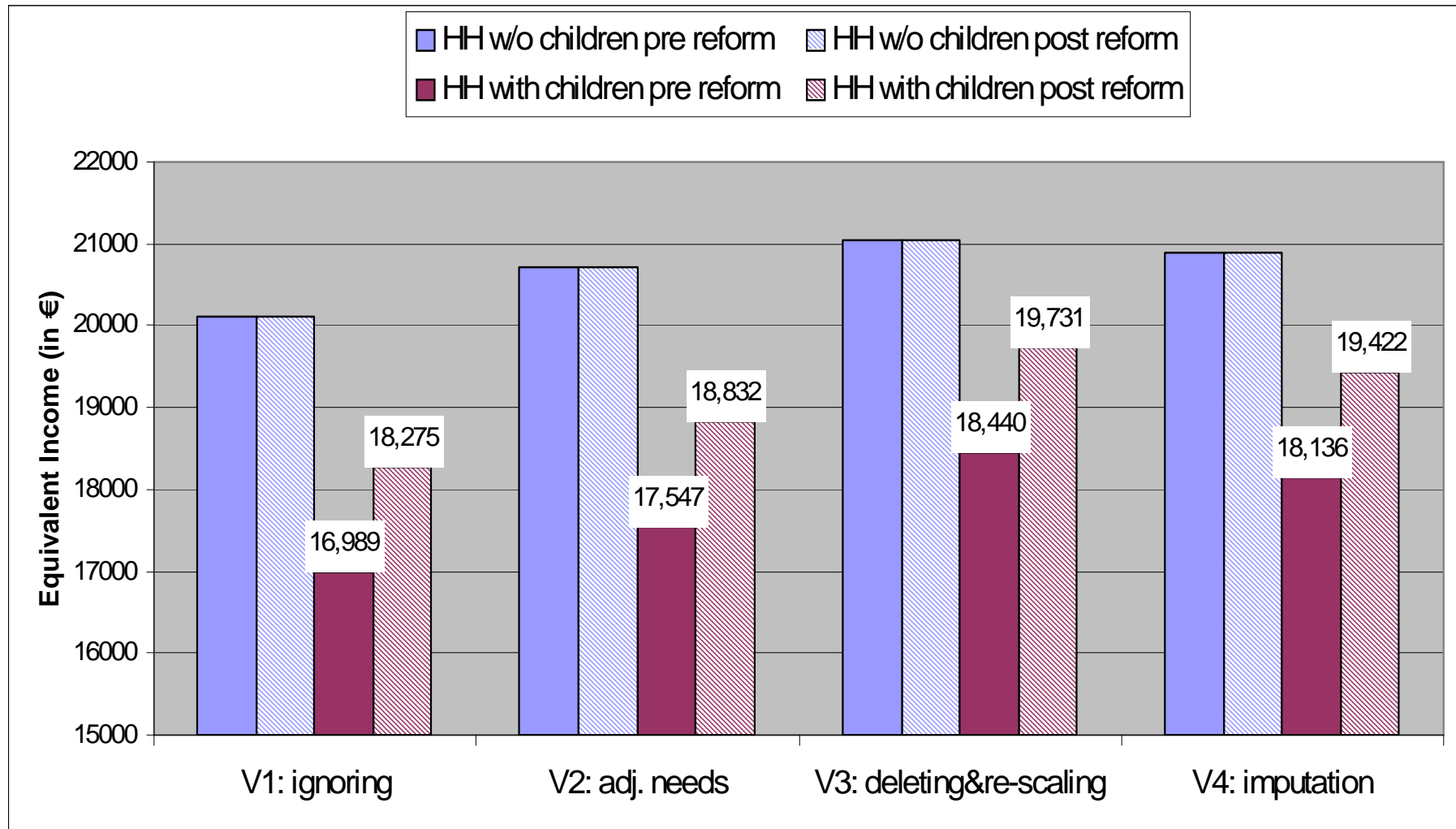
Year-to-Year Poverty mobility by PUNR-treatment (avg. 1985-2007)



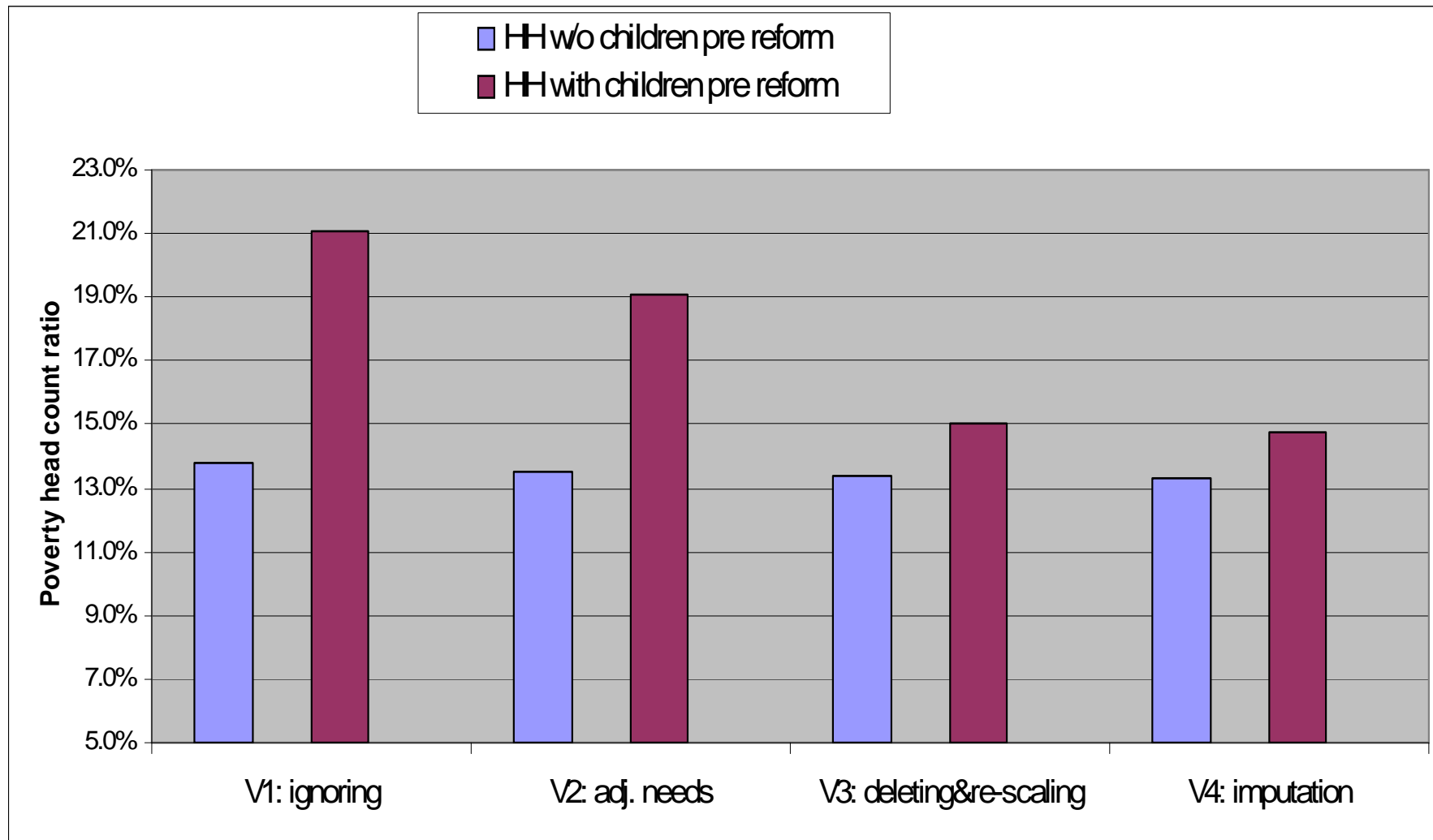
## 4. Micro-Simulation

- Reform: Increase child benefits by € 150 / month per child ( $\leq 16$  years of age)
- Comparing pre- and post simulated child benefit reform
  - focus: poverty reduction effect
- Concept of relative income poverty
  - Standard approach in EU: PovLine = 60 % of national median income
  - Dynamically adjusted

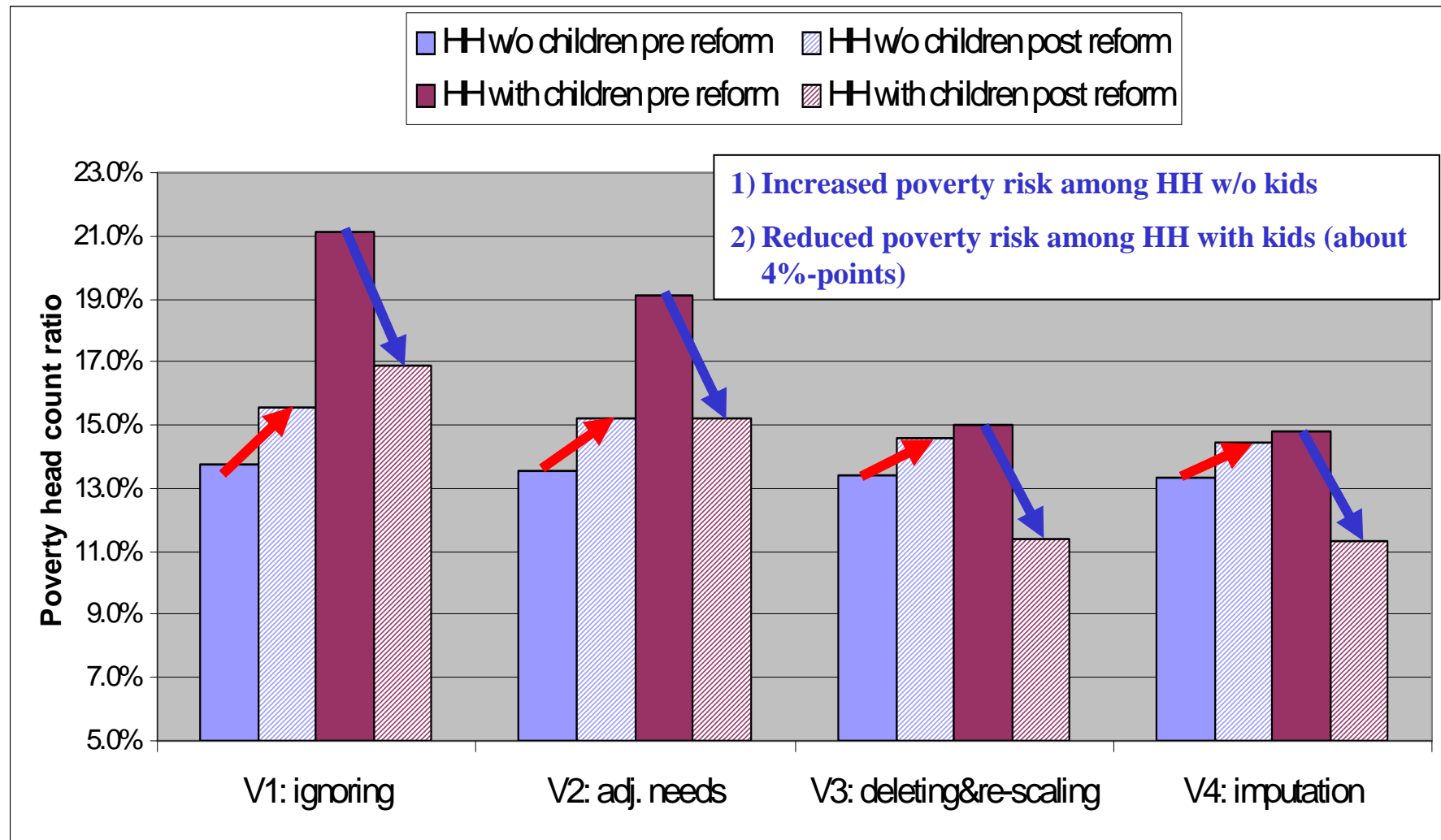
## Income effects of simulated child benefit reform, by household type and PUNR treatment (2007)



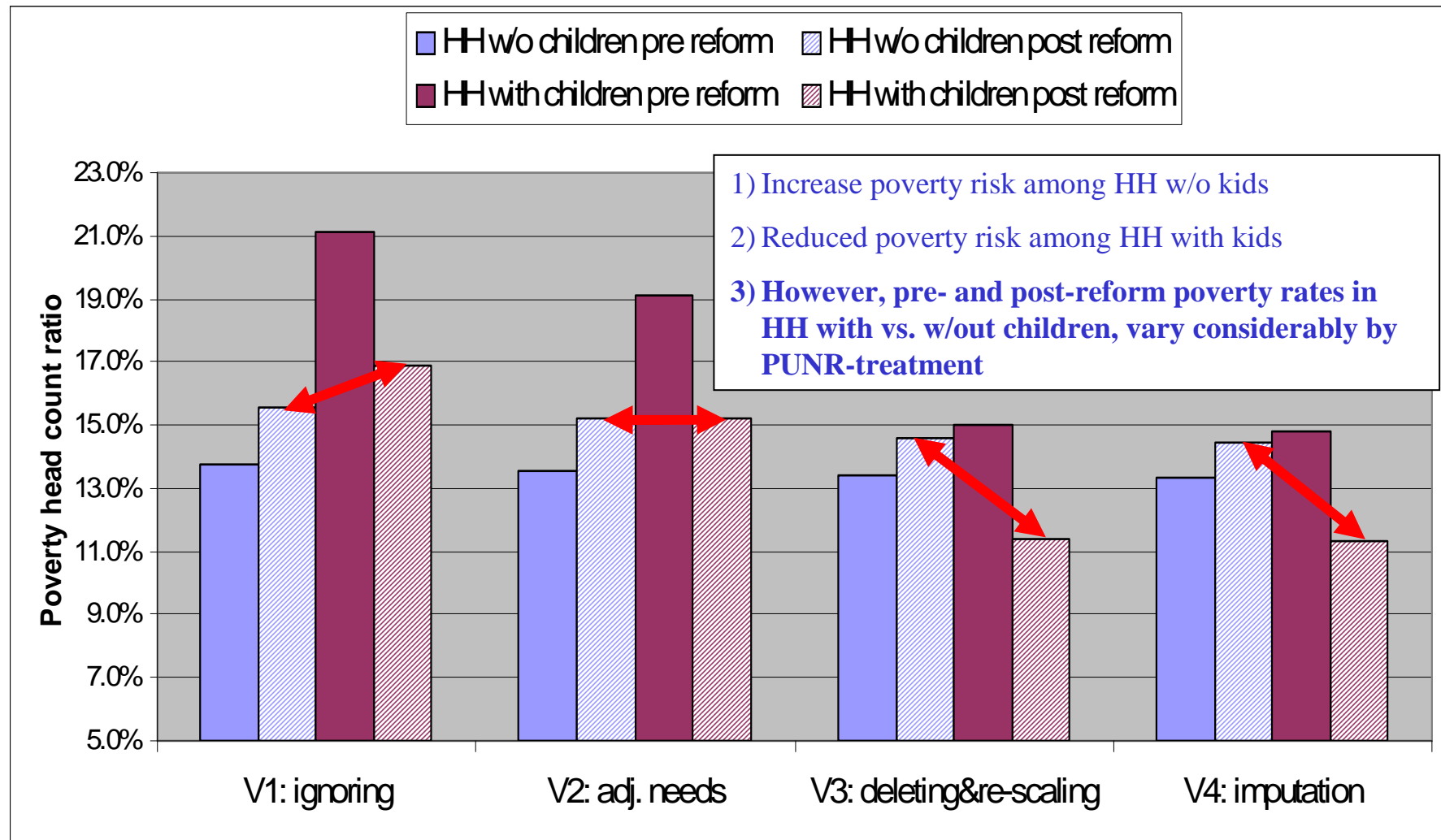
## Poverty effects of simulated child benefit reform, by household type and PUNR treatment (2007)



## Poverty effects of simulated child benefit reform, by household type and PUNR treatment (2007)



## Poverty effects of simulated child benefit reform, by household type and PUNR treatment (2007)



# 5. Conclusion

- Facts
  - PUNR is of increasing relevance in household panels
  - PUNR is selective
  - *Consequence*: Bias in aggregated household income measures
- PUNR-treatments
  - Ignoring is “far from perfect”
  - Equivalence scale adjustments do not sufficiently correct for this bias
  - Deleting PUNR and re-scaling weights as well as (complex) imputation of missing income components yield similar results
  - However, imputation allows to maintain the entire survey population  
→ most relevant for longitudinal (mobility) research
- Looking ahead
  - How does the choice of PUNR-treatment affect cross-national comparability
    - most important for comparative microsimulation databases, e.g. Euromod [ECHP, EU-SILC]
  - What does this mean for designing incentives in HH panel surveys ?
  - Multiple imputation

# Comments welcome

Joachim R. Frick

[jfrick@diw.de](mailto:jfrick@diw.de)

SOEP-hotline

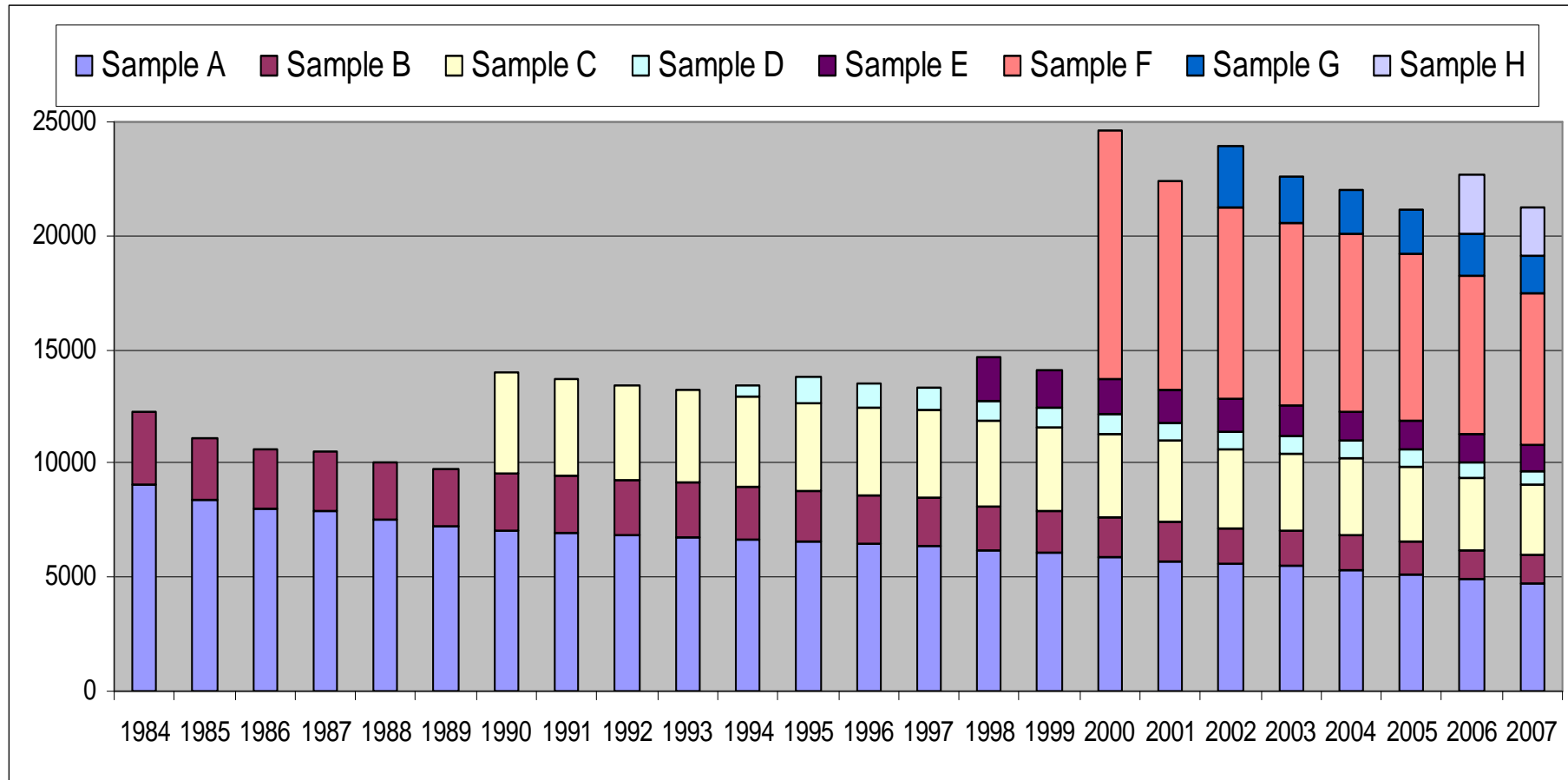
[soepmail@diw.de](mailto:soepmail@diw.de)

SOEP-homepage

[www.diw.de/gsoep](http://www.diw.de/gsoep)

# SOEP – Development of Sample Size

Number of successfully interviewed persons by sample



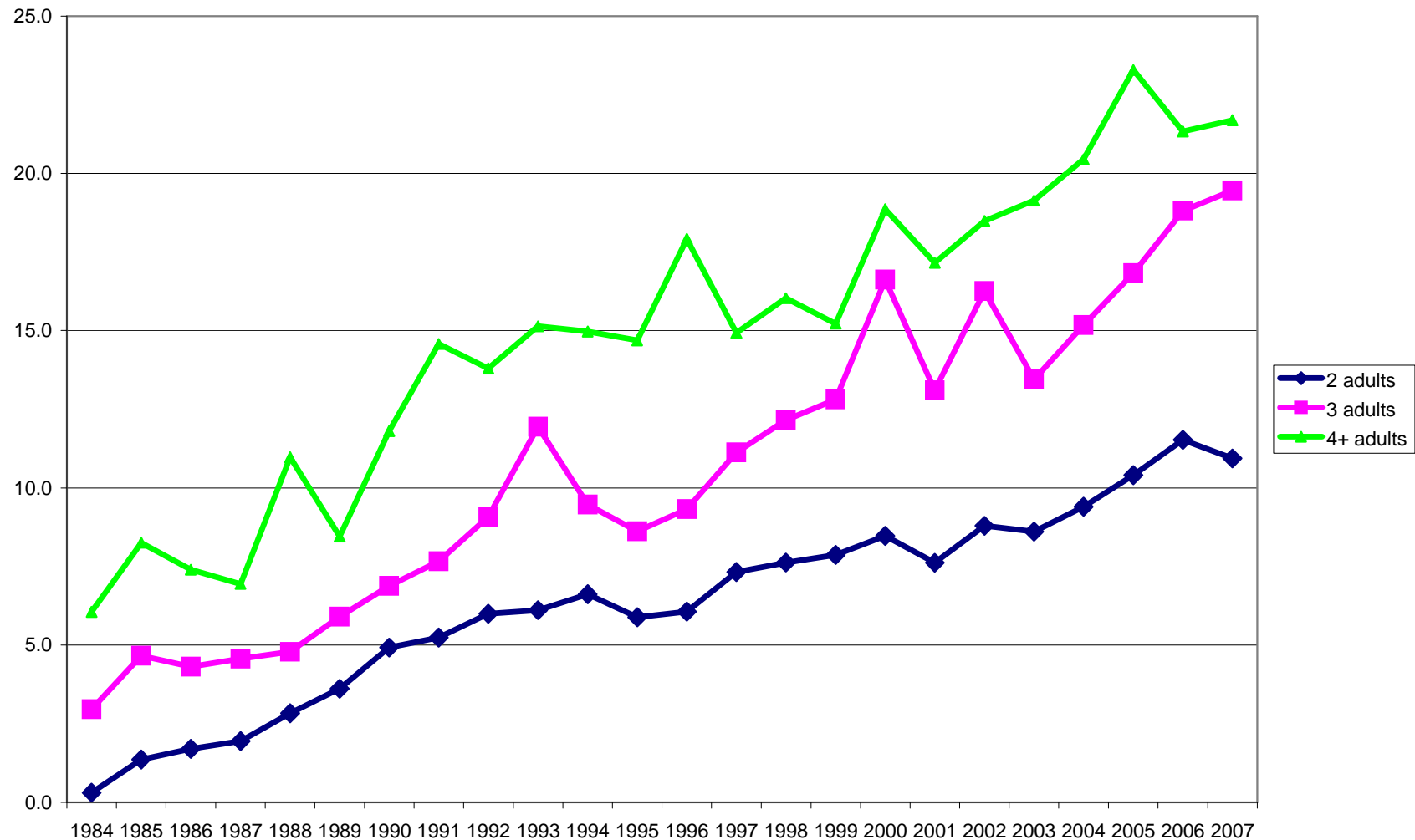
# PUNR in the SOEP

|                    | Persons       | Person-<br>Years,<br>total | <i>Thereof:</i><br>Years<br>PUNR | Average<br>Years in<br>SOEP |
|--------------------|---------------|----------------------------|----------------------------------|-----------------------------|
| never PUNR         | <b>43,068</b> | <b>343,622</b>             | <b>0</b>                         | <b>8.0</b>                  |
| PUNR &<br>Non-PUNR | <b>4,348</b>  | <b>35,817</b>              | <b>8,737</b>                     | <b>8.2</b>                  |
| always PUNR        | <b>2,455</b>  | <b>6,454</b>               | <b>6,454</b>                     | <b>2.6</b>                  |
| Total              | <b>49,871</b> | <b>385,893</b>             | <b>15,191</b>                    | <b>7.7</b>                  |

Population: adult members (17+) in interviewed households 1984-2007

*Source: SOEP 1984-2007*

## Incidence of PUNR by #adults in HH (%)



# 1) Filter Imputation ( $Y_k > 0$ )

- successive conditional regressions
  1. pension
  2. unemployment
  - 3.-5. maternity benefits, student grants, private transfers
  6. labour income
- binary regression model (probit)
- Dep. Variable: Filter = 1, if  $P(Y_k > 0) > \text{Cut-off}$ 
  - random cut-off is normally distributed with mean 0.5, stdev. 0.2
- RHS-Variables:
  - *Longitudinal information* on receipt of  $Y_k$ , if at all available
  - *demographic variables*: age-groups, type of hh, no. of children, person in need of care, relation to head of hh, region, community size, SOEP-sample, change in hh-composition
  - *social structure*: home owner, (highest) education of head of hh or spouse/partner, education of children, migration background
  - *income variables*: screener income (net household incomes at month of interview), public transfers (housing subsidies, social assistance, etc.), aggregated observed individual incomes as a share of household screener income
  - *filter(s)* considering receipt of other income components (successive inclusion)

## 2a) Longitudinal Imputation | Filter=1

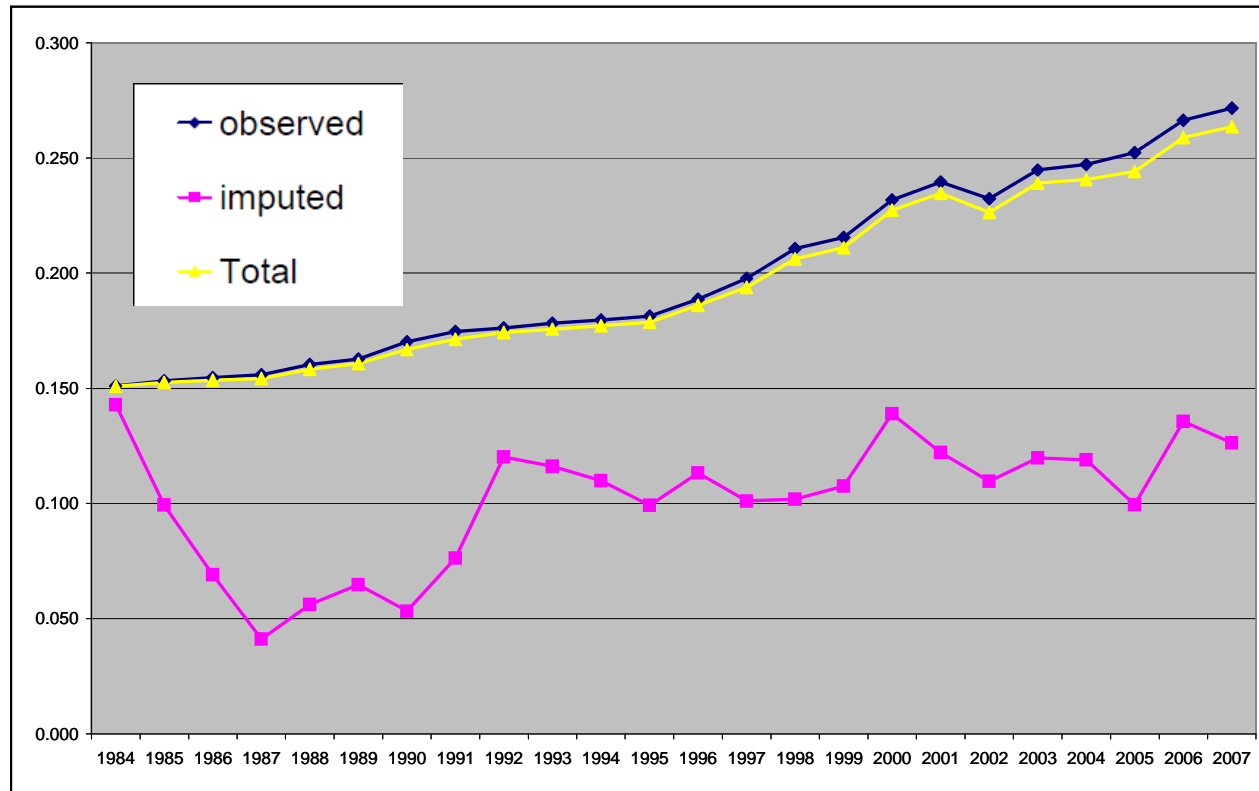
- “Row-and-column” imputation (Little & Su, 1989):
  - (1) row effect (longitudinal: within-person avg. Y position)  $r_i = \frac{1}{k} \sum_{j=1}^k \frac{Y_{ij}}{Y_j}$
  - (2) column effect (x-sectional: avg. yearly income)  $c_t = Y_t / \frac{1}{k} \sum_{i=1}^k Y_i$ 
    - expected income position of individual  $i$  at time  $t$ :  $E_{it} = r_i * c_t$
  - (3) stochastic term (=deviation from observed value of nearest neighbour  $n$ )
    - imputed income  $Z_{it} = E_{it} * [Y_{nt} / E_{nt}]$
- PUNRs with valid interview(s) within 7 year-window
  - three years before and after non-response
  - *weighted by distance* to the year of missing information
  - applies to about roughly 1/3 of all PUNRs
  - L&S is also the standard imputation procedure for imputation of INR

## 2b) Cross-sectional Imputation | Filter=1

Imputation of annual income  $Y_k$  |  $Y_k > 0$

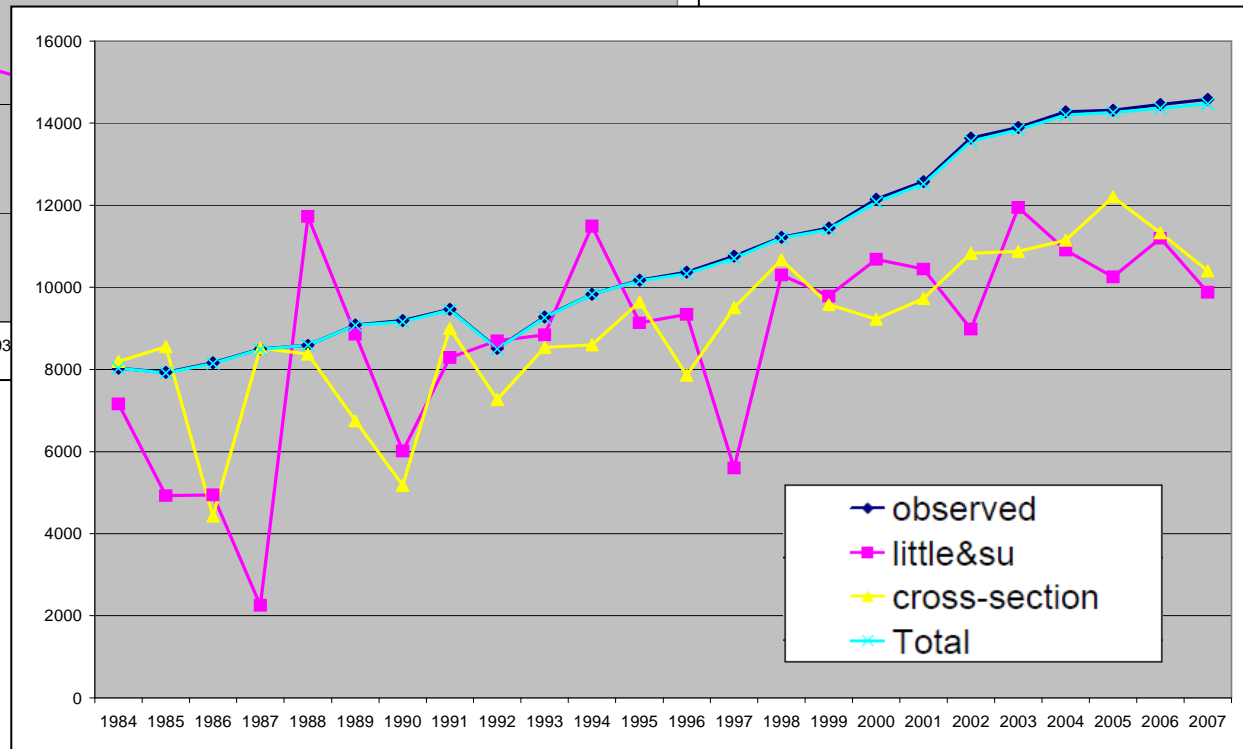
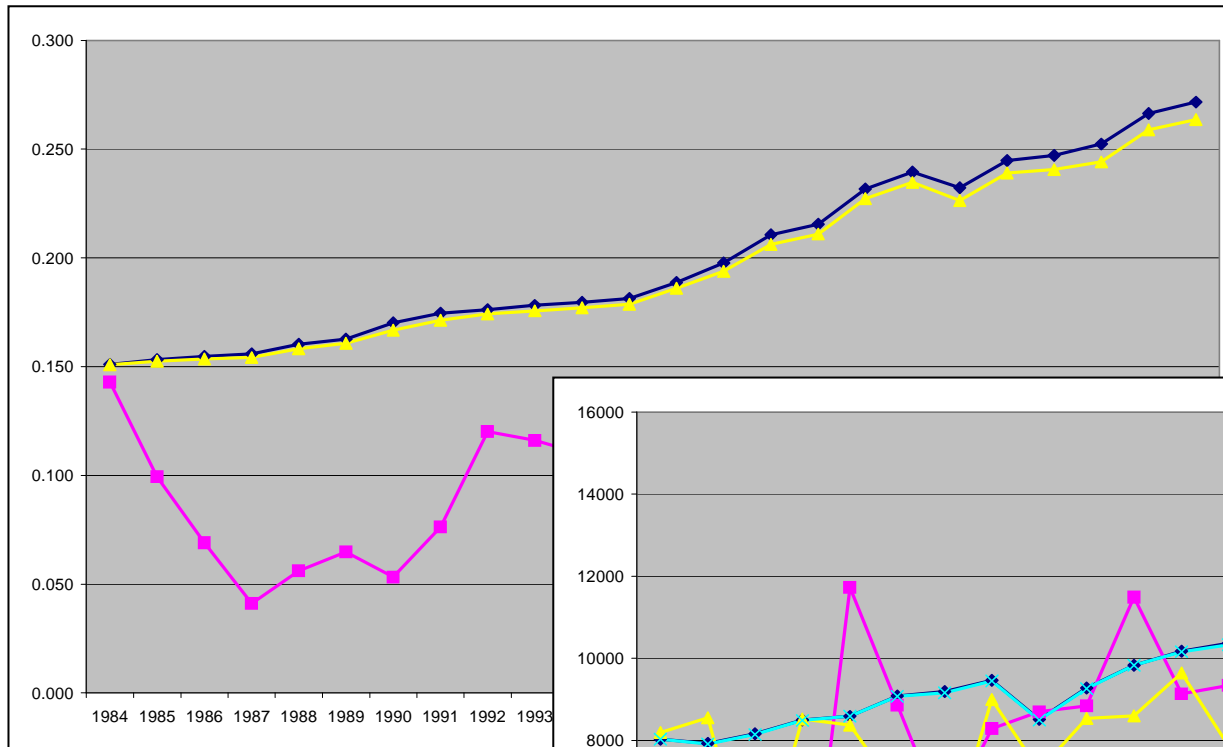
- OLS-regression, conditional on full set of filters
- Dep. Variable:  $\log(Y_k)$
- RHS-variables (see Filter imputation)
- stochastic term: randomly drawn residuals from observed population
- separate estimations for men and women (esp. labour)

# Imputation Results: Gross Annual Pensions



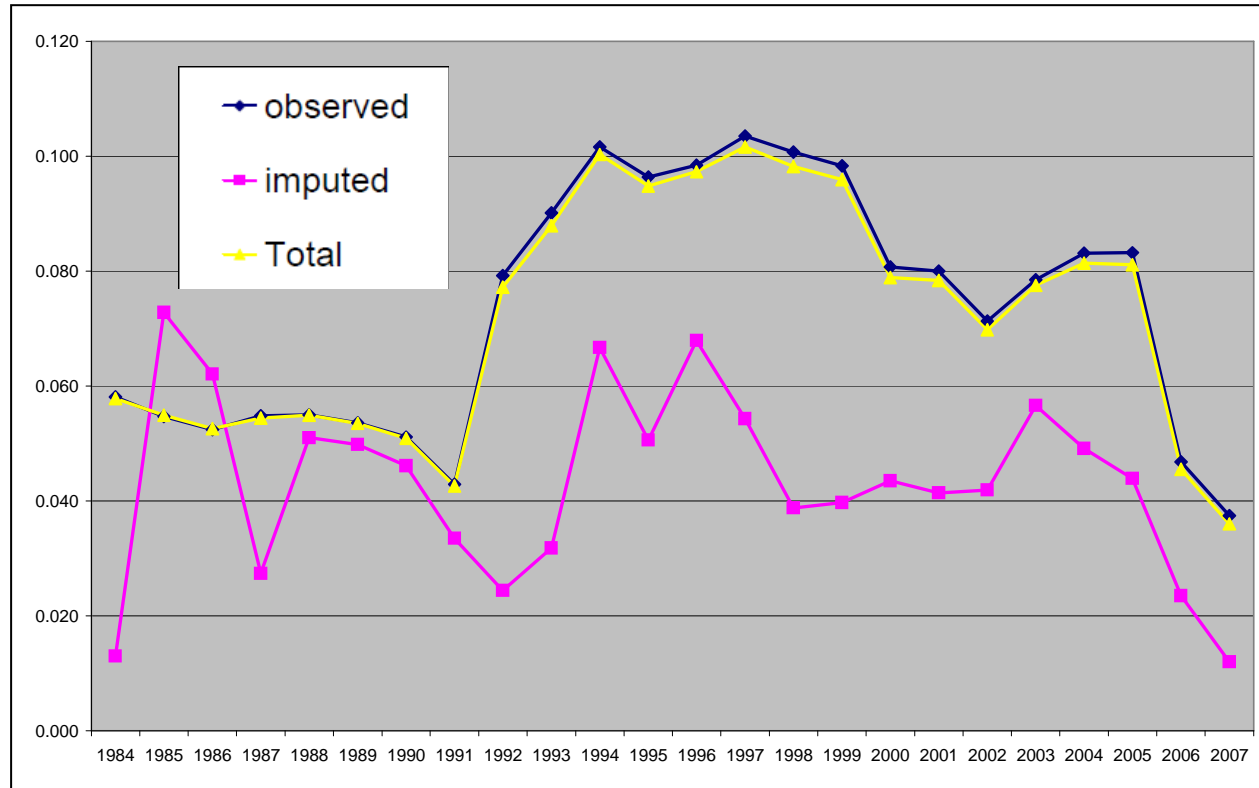
Population  
Share (in %)

# Imputation Results: Gross Annual Pensions



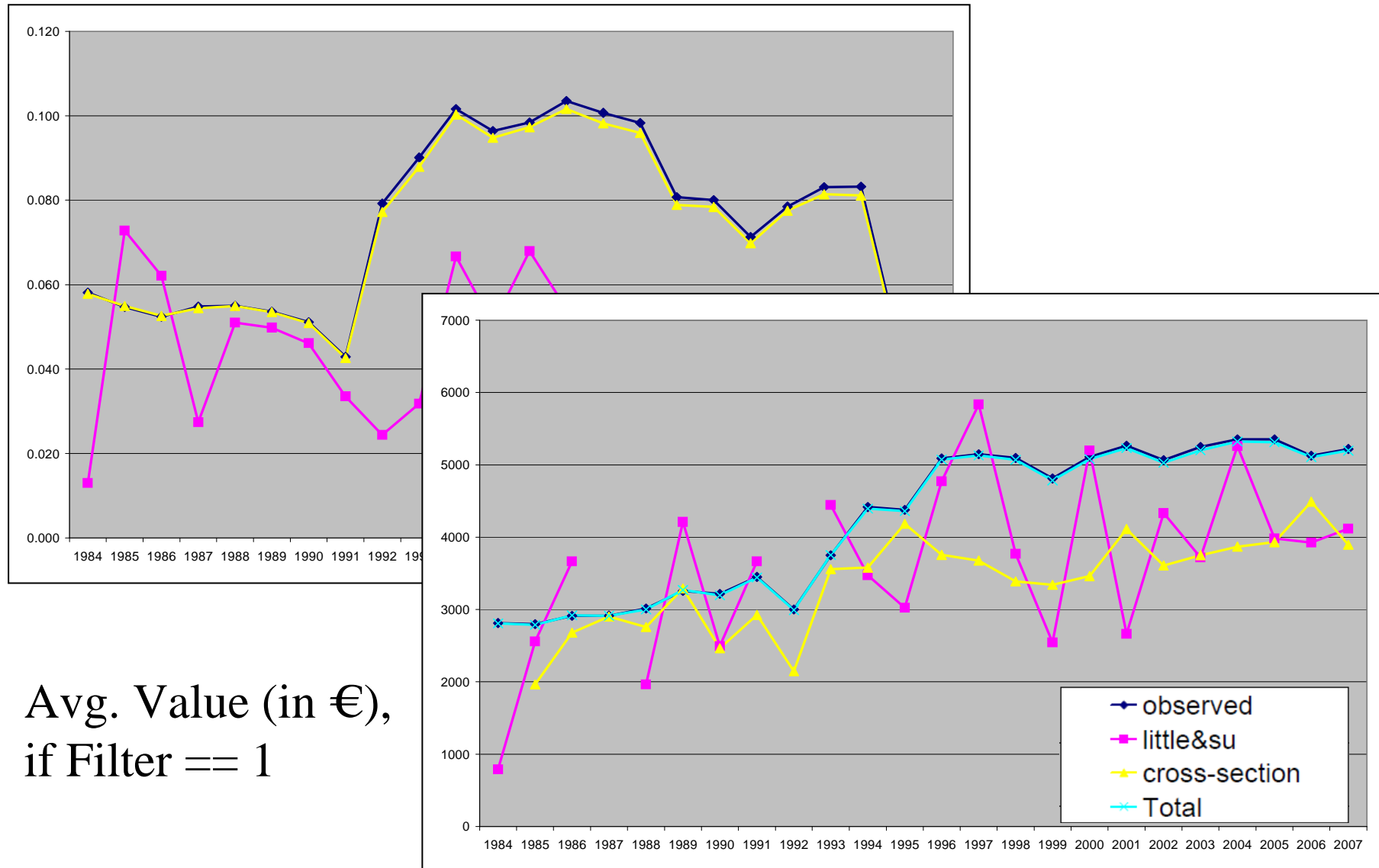
Avg. Value (in €),  
if Filter == 1

# Imputation Results: Unemployment Benefits



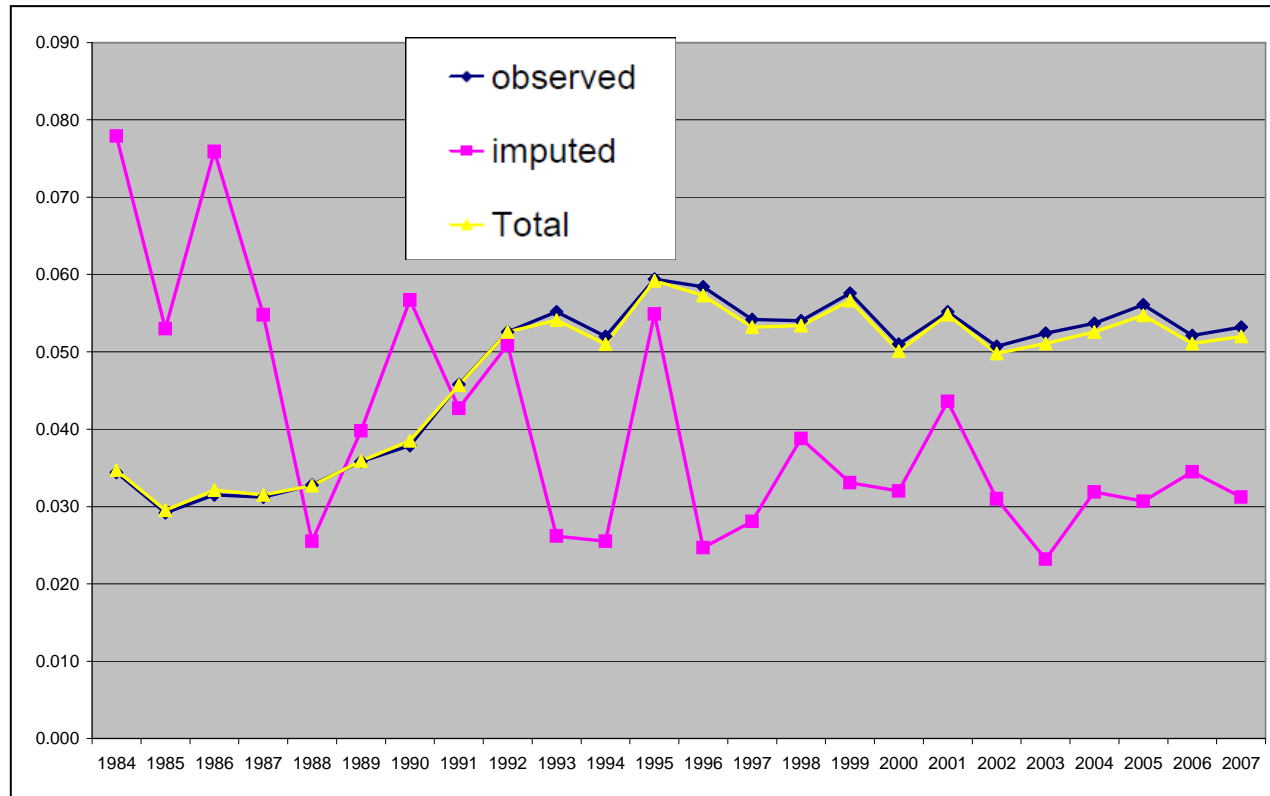
Population  
Share (in %)

# Imputation Results: Unemployment Benefits



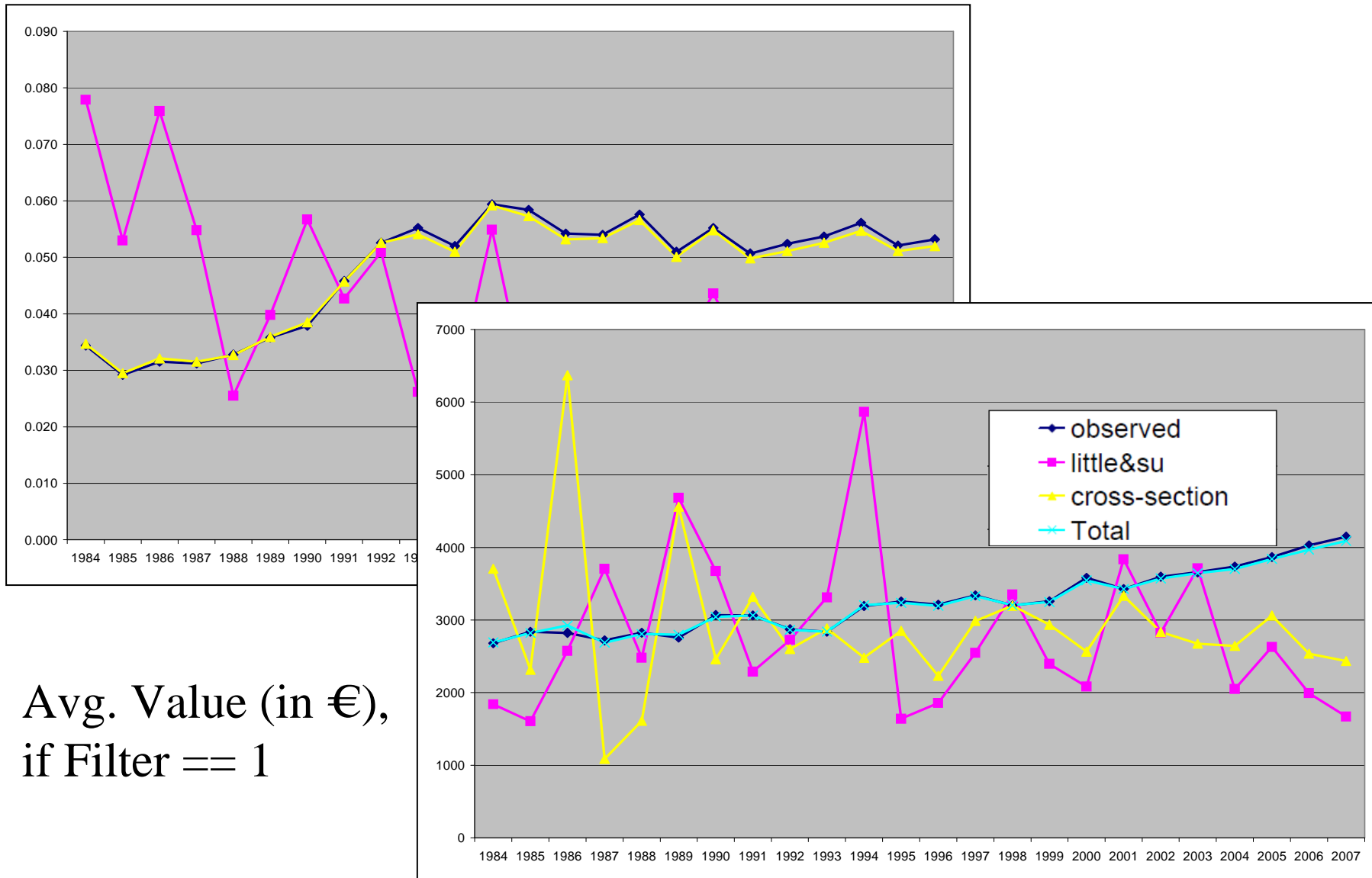
Avg. Value (in €),  
if Filter == 1

# Imputation Results: Other Transfer Income



Population  
Share (in %)

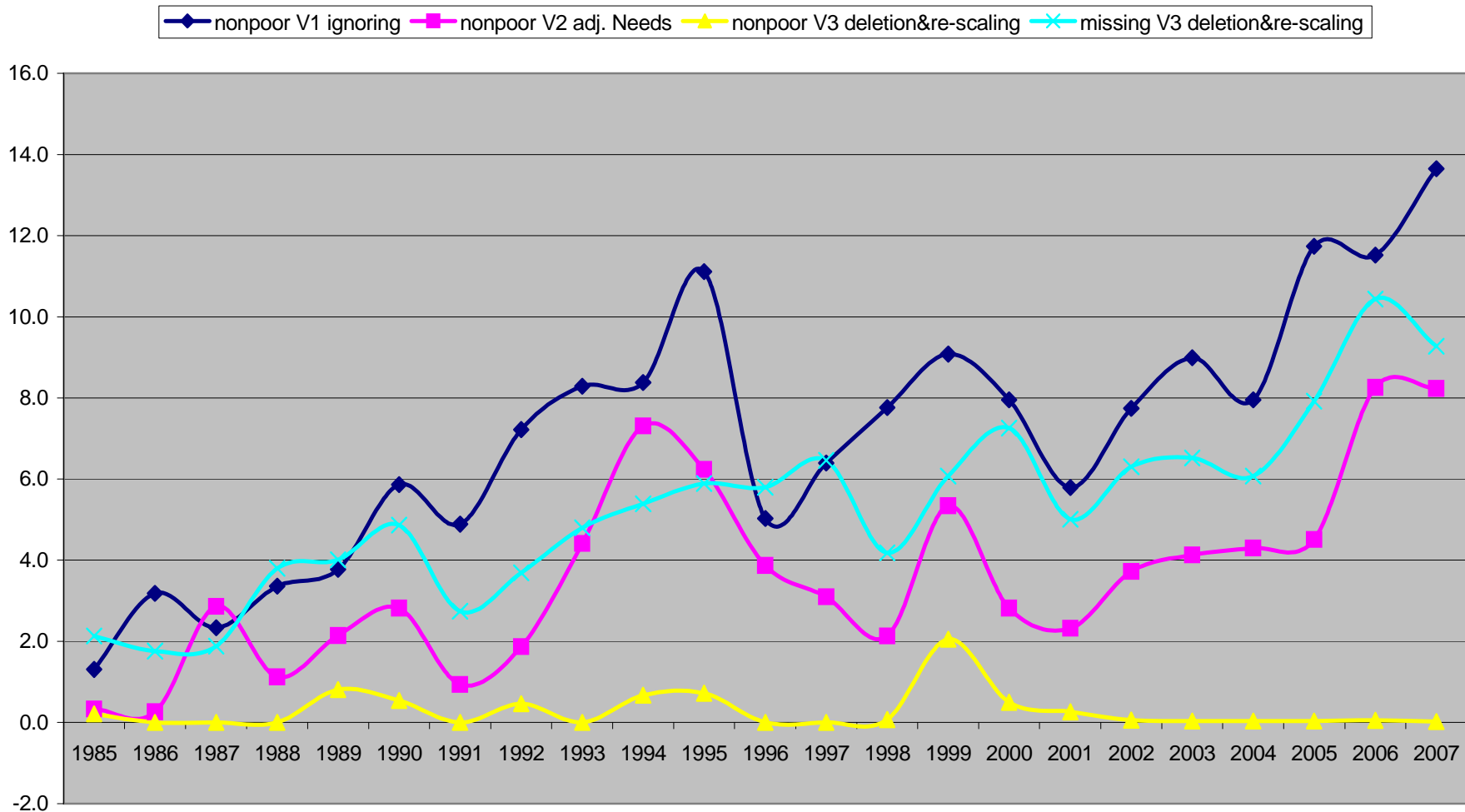
# Imputation Results: Other Transfer Income



Avg. Value (in €),  
if Filter == 1

# Consistency of poverty status by PUNR treatment

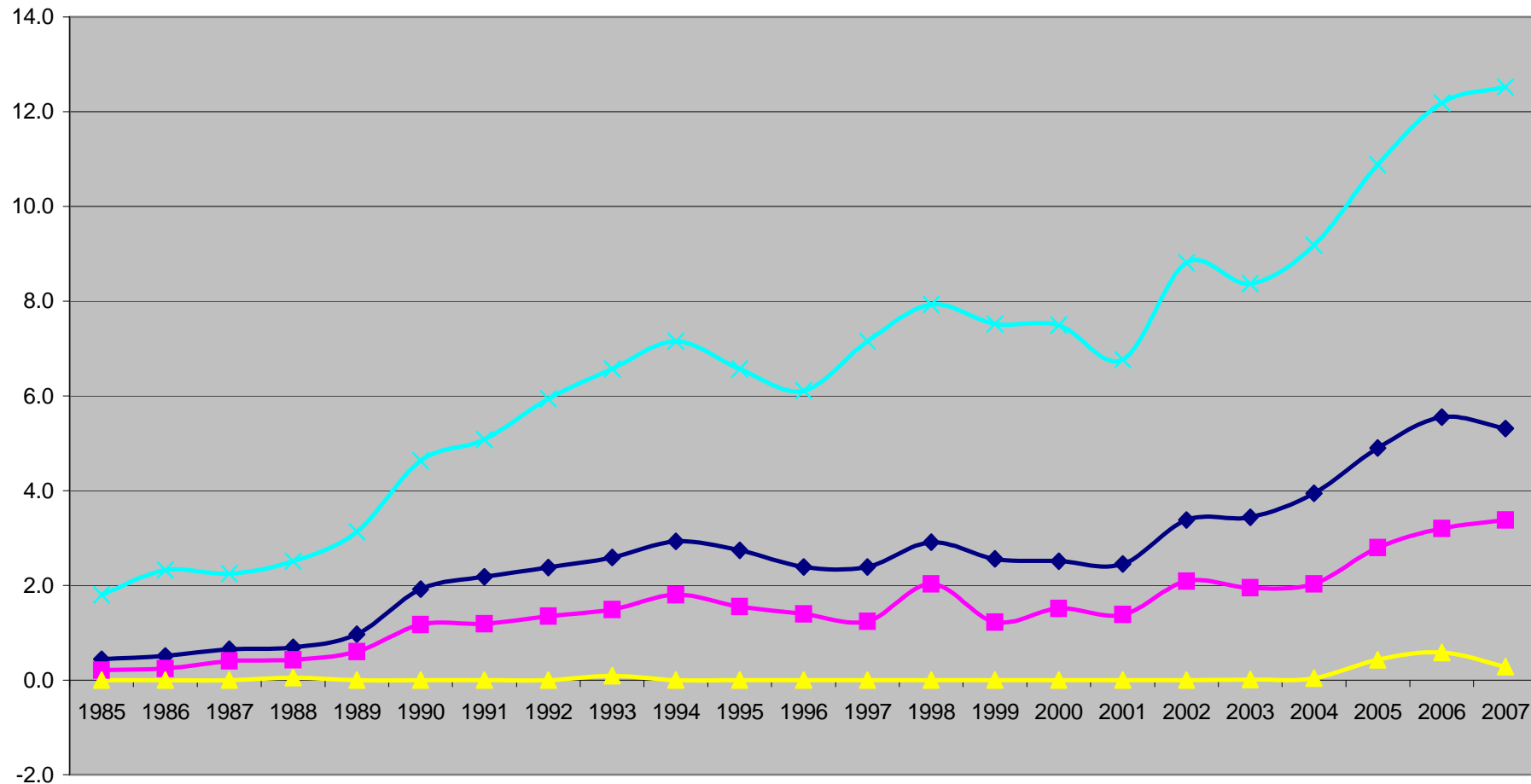
Poor according to V4 (Imputation) -  
 Non-Poor according to Versions 1-3, Persons in % (weighted)



# Consistency of poverty status by PUNR treatment

Non-Poor according to V4 (Imputation) -  
 Poor (or Missing) according to Versions 1-3, Persons in % (weighted)

poor V1 ignoring    poor V2 adj. Needs    poor V3 deletion&re-scaling    missing V3 deletion&re-scaling

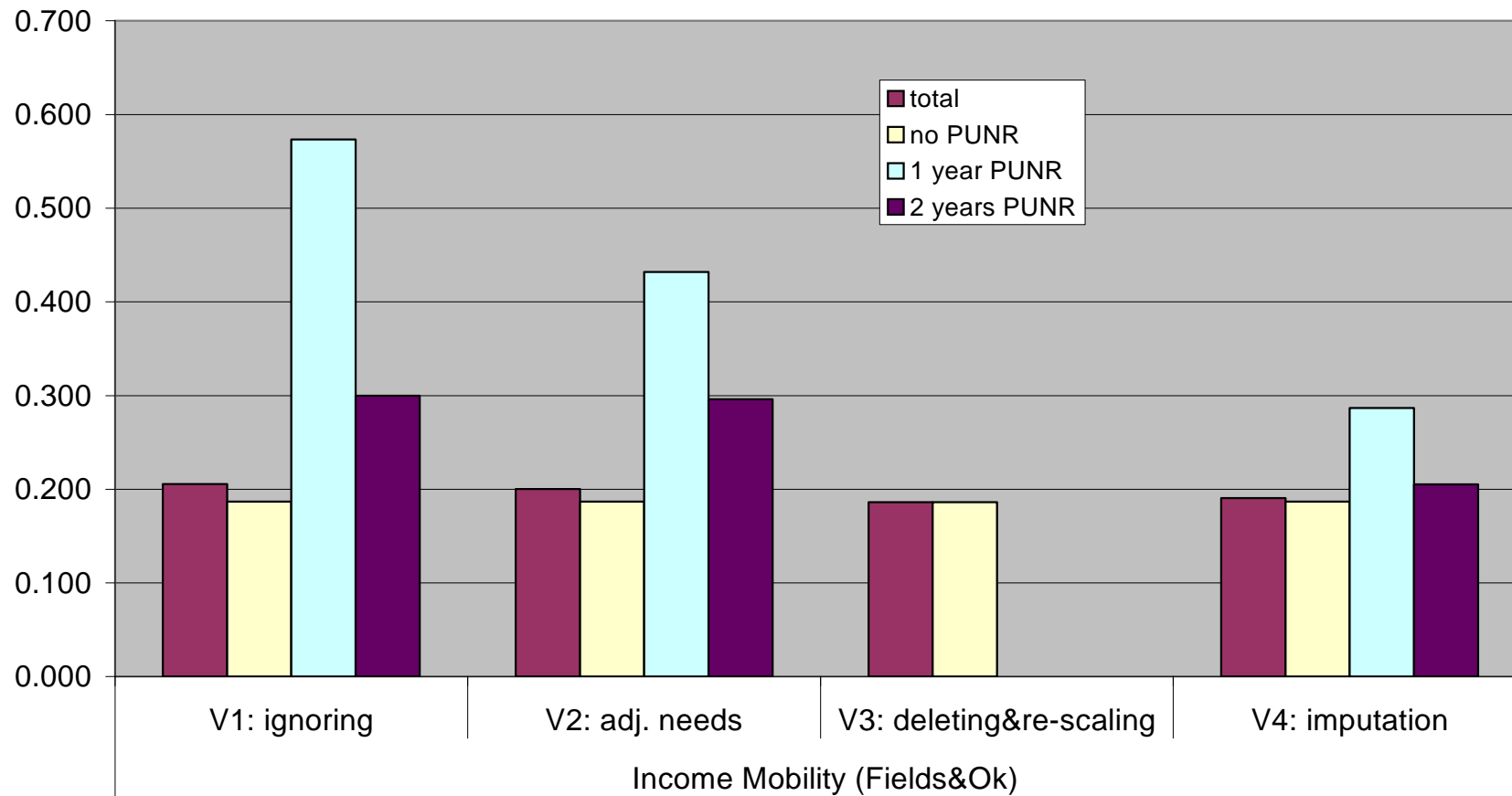


## (c) Income & Poverty Mobility

- Poverty Dynamics
- Wave-to-Wave Mobility (Fields & Ok 1999, Shorrocks 1979)
- 5-Wave Income Mobility (Shorrocks 1979)

# Income Mobility by PUNR-treatment: Fields & Ok (1999)

Year-to-Year Income Mobility by PUNR-Treatment (avg. 1985-2007)



# Income Mobility by PUNR-treatment: Shorrocks (1978 ), based on Gini

