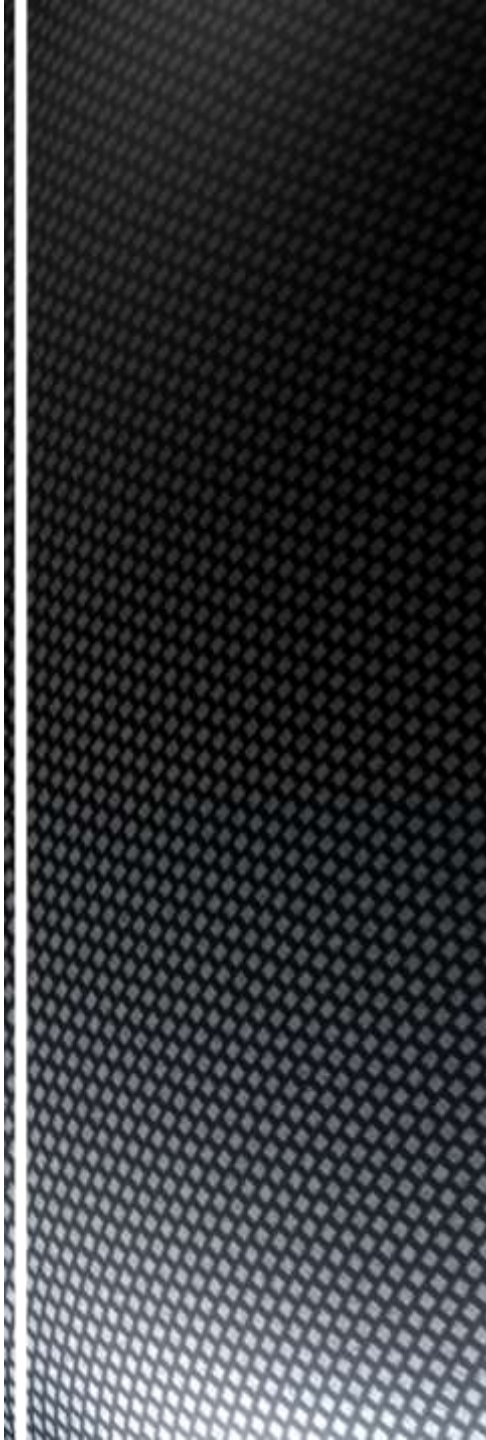
Simulating Histories for Dynamic Microsimulation Models

Jinjing Li

Maastricht Graduate School of Governance

Cathal O'Donoghue

Teagasc RERC, Ireland



"We've done a computer simulation of your projected performance in five years. You're fired." by J.P. Rini

Microsimulation Has High Data Requirement

The Dilemma of the Base Dataset

- What Dataset do we Need for Dynamic Microsimulation
  - Age, Education, Job, Earnings
  - Marriage, Children, Health, Spatial Information
  - Other relevant Information (everything)
  - Life time information
- What dataset Do we Have
  - Cross sectional Survey/Census
  - Panel Survey
  - Administrative data
- Difficult to choice (Cassells et al, 2006; Zaidi and Scott, 2001; Zaidi 2001,2004)

What are the Alternatives

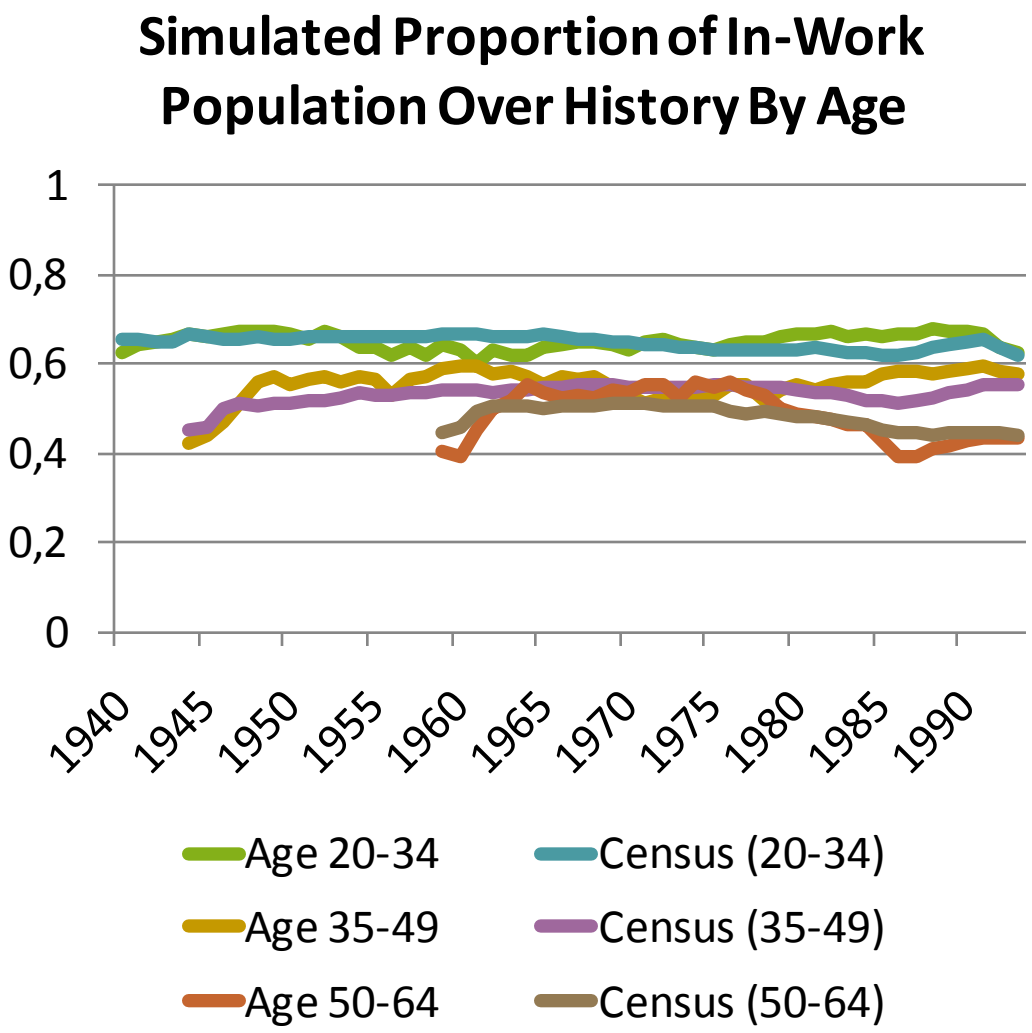
- Statistical Matching
  - DYNASIM (CPS with Administrative data)
  - DYNASIM3 (SIPP with PSID)
  - PENSIM2 (Mix of FRS, BHPS, LLMDB)
- Synesthetic Simulation
  - DYNANCAN
  - CORSIM

LIAM Model and its Data Requirement

- Our model LIAM
  - A Dynamic MSM for life cycle pension analysis and reform evaluation based on LII (Living in Ireland Survey, similar to ECHP/EU-SILC)
  - Require demographic info(age, education attainment, marriage etc), employment info, and pension info
  - Requires history since the potential pensioner's birth year
- Neither method discussed earlier meets the requirement of because
  - No admin record goes back to 40s in Ireland
  - Need consistent history for pension projection

- Has to be consistent with current individual profile
- Should be close to what actually happened in the history for all key variables
- Has to be primarily based on LII and other public statistics (mainly macro)

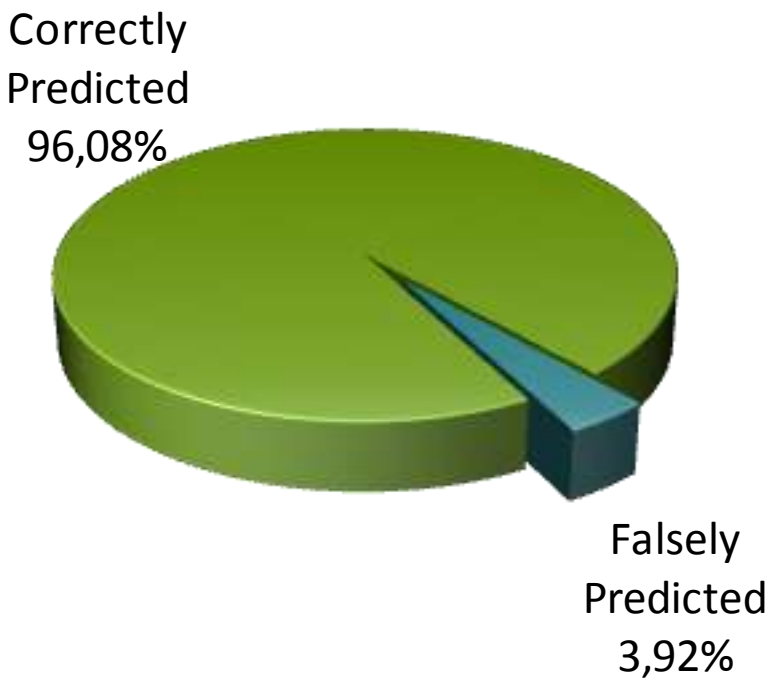
Historical re-creation seems to be the only workaround



A Preview on the Results

- It replicates the History

Irish State Contributory Pension has requirements on length, density, and timing.



A Preview on the Results
- It replicates the State Pension Eligibility

How the History was Modelled

- Exploit the Retrospective Variables
  - the year when they started their current job and the job duration
  - the year and age when they first started to work
  - the number of years spent in full-time education, employment, unemployment, illness or disability
  - home caring or retirement since the age of 10
  - the year when they became unemployed, if currently unemployed

How the History was Modelled

- Exploit the Retrospective Variables Further
  - Maternity leave
  - Recent unemployment prior to the base year
  - Some pension membership information

- Deterministic Simulation
  - Directly calculated from retrospective variables
- Semi-Stochastic Simulation
  - Indirectly calculated from retrospective variables in combination with assumptions and external Data
- Stochastic Simulation
  - Microsimulation in a reversed timeline

A Mixtures of Methodologies in Back Simulation: Three Steps

- Discrete Variable Simulation
  - Model Requirement: Able to capture stable individual effect
  - A logit variation of Mundlak (1978)

$$\log(p_i) = \alpha + \beta X_i + \varepsilon_i$$

$$\varepsilon_{it} = u_i + v_{it}$$

  - Extract individual effects (u_i) from the panel residuals

$$\log(p_i) = \alpha + \beta X_i + d(\overline{\varepsilon_{it}}) + \varepsilon_i$$

- Continuous Variable Simulation
  - Random-effects Panel estimates

Back Simulation Estimation Method

- More difficult than forward simulation due to the extra constraints
- To simulation pension eligibility correctly, you need to get **length**, **density**, and **timing** correct in career profile
- Dual alignments (Cross-sectional and Longitudinal) Design

Alignment Implement of Back Simulation

How did we Align the Data

- Cross-sectional (and Partial longitudinal) Alignment on Yearly Basis
  - Continuous variables (ratio)
  - Discrete variables (rank the probability on **condition of the longitudinal consistency**)
- Further Adjustment is still required for complex longitudinal alignment
  - e.g. Pension eligibilities
  - Longitudinal statistics is required

Longitudinal Alignment

An example: State Contributory Pension

- Introduced in 1953 for employed, 1988 for self employed
- Need to pay the first contribution before certain age (depends on cohort)
- Requires a contribution density of at least 10/52

Year	1950	1951	1952	1971	1972	1973	Total Years of Work	Rate	Simulated Eligible	Actual Eligibility
Individual A	●	○	●	○	○	●	3	33.33%	Yes	No
Individual B	●	○	●	●	○	●	4	66.67%	Yes	Yes
Individual C	●	●	●	○	○	○	3	00.00%	No	Yes
Yearly Aggregate	3	1	3	1	0	2			(1/3 correct)	

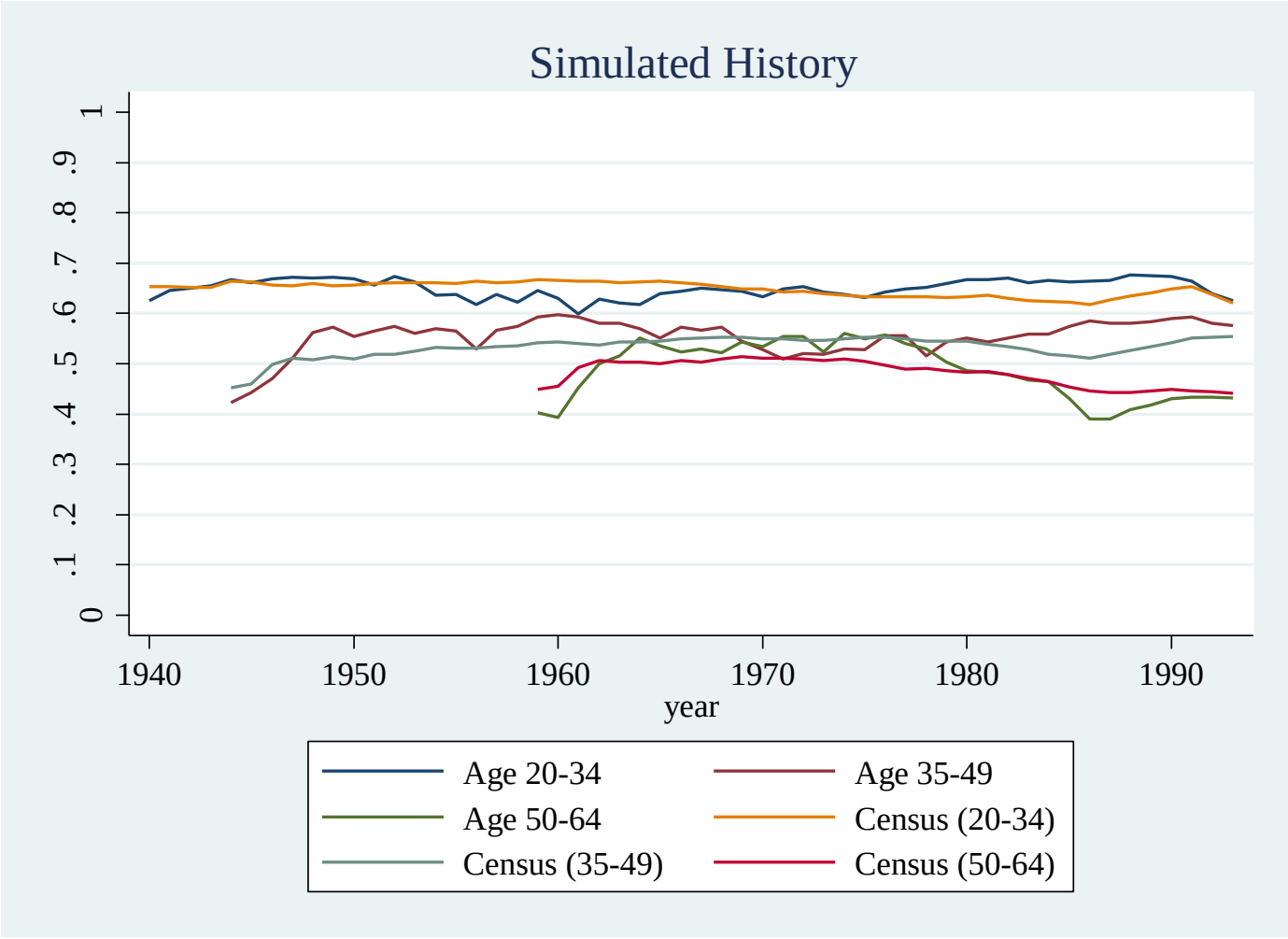
Longitudinal Alignment

An example: State Contributory Pension

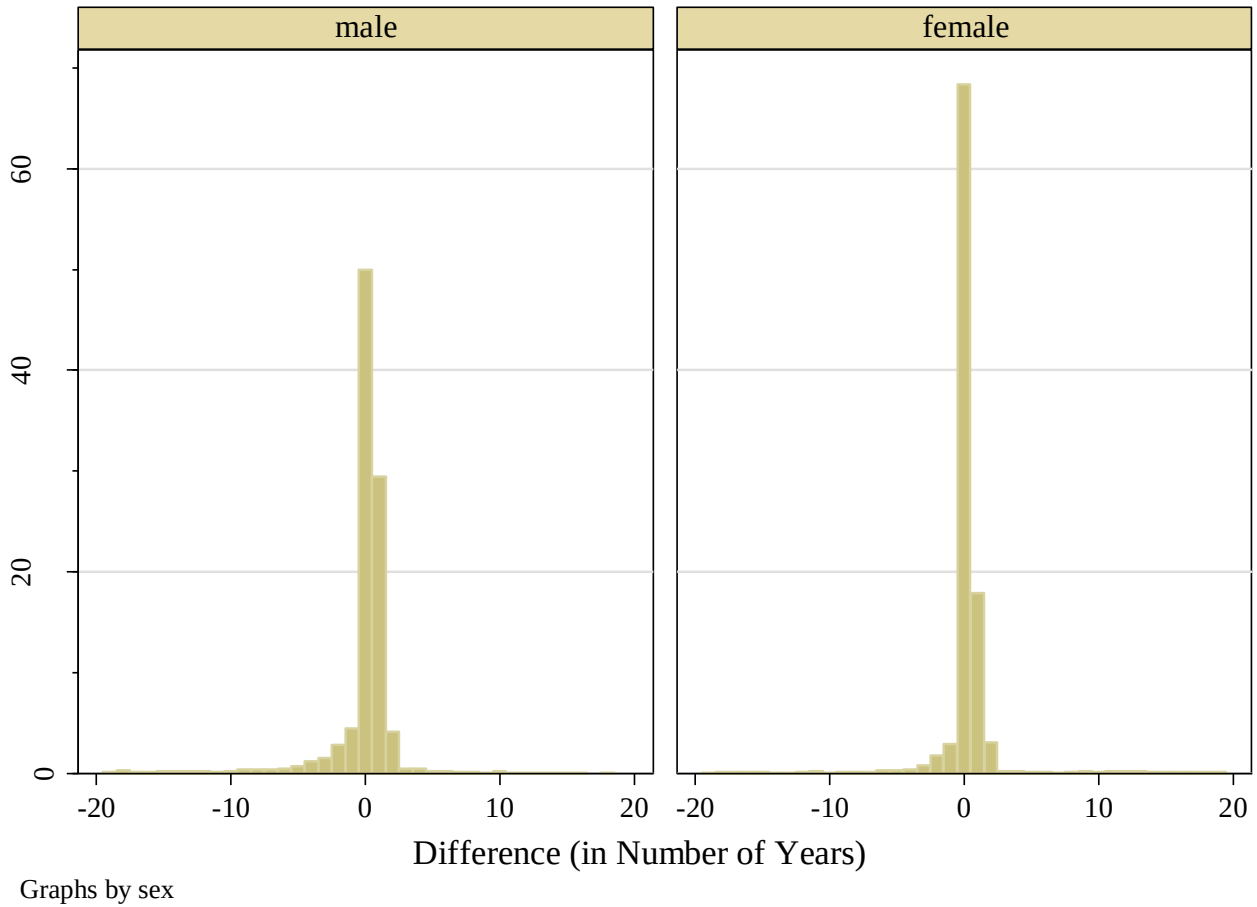
- Introduced in 1953 for employed, 1988 for self employed
- Need to pay the first contribution before certain age (depends on cohort)
- Requires a contribution density of at least 10/52

Year	1950	1951	1952	1971	1972	1973	Total Years of Work	Rate	Simulated Eligible	Actual Eligibility
Individual A	●	●	●	○	○	○	3	00.00%	No	No
Individual B	●	○	●	●	○	●	4	66.67%	Yes	Yes
Individual C	●	○	●	○	○	●	3	33.33%	Yes	Yes
Yearly Aggregate	3	1	3	1	0	2			(3/3 correct)	

Simulated History Result



Simulated History Result

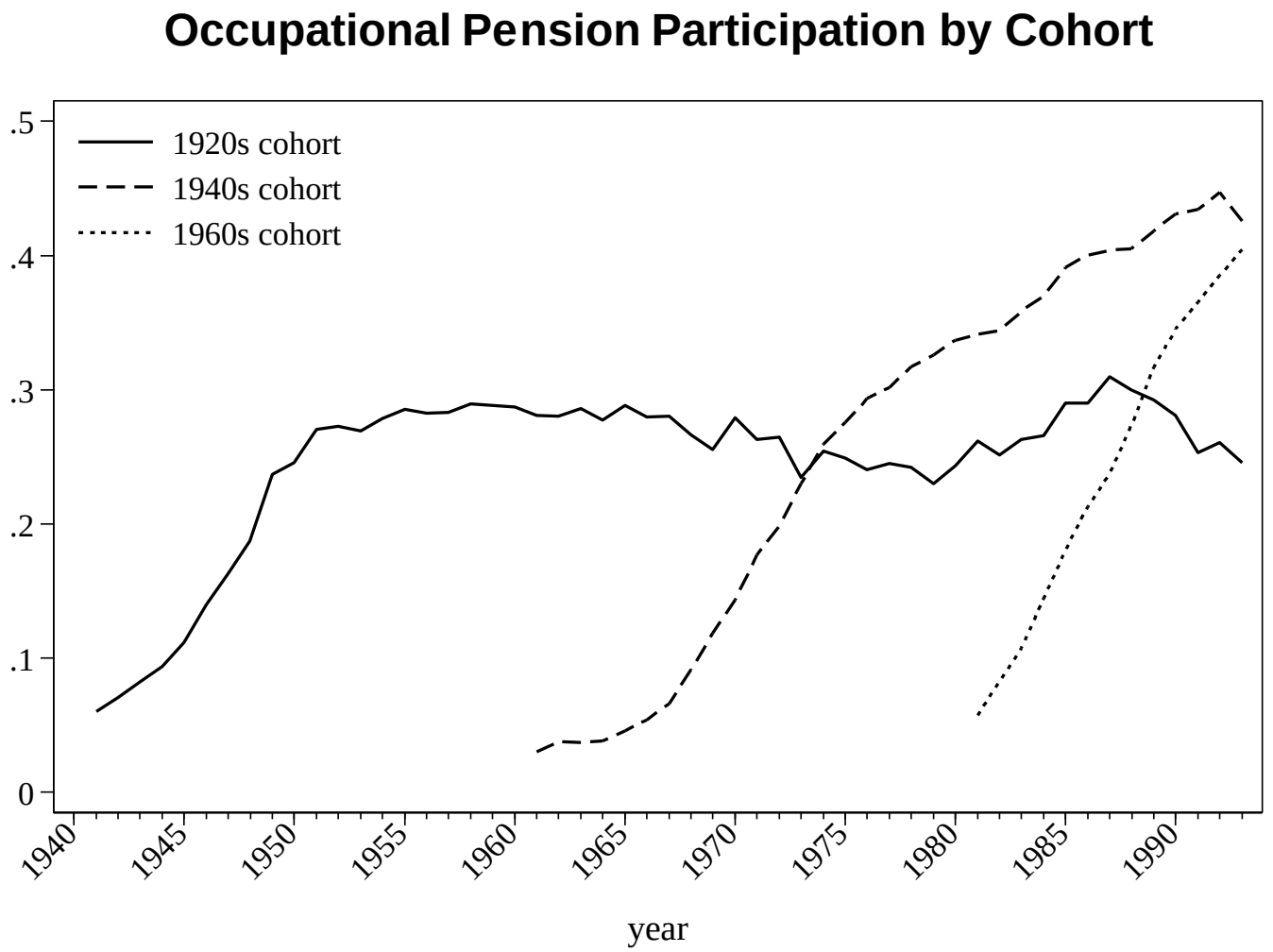


Simulated History Result

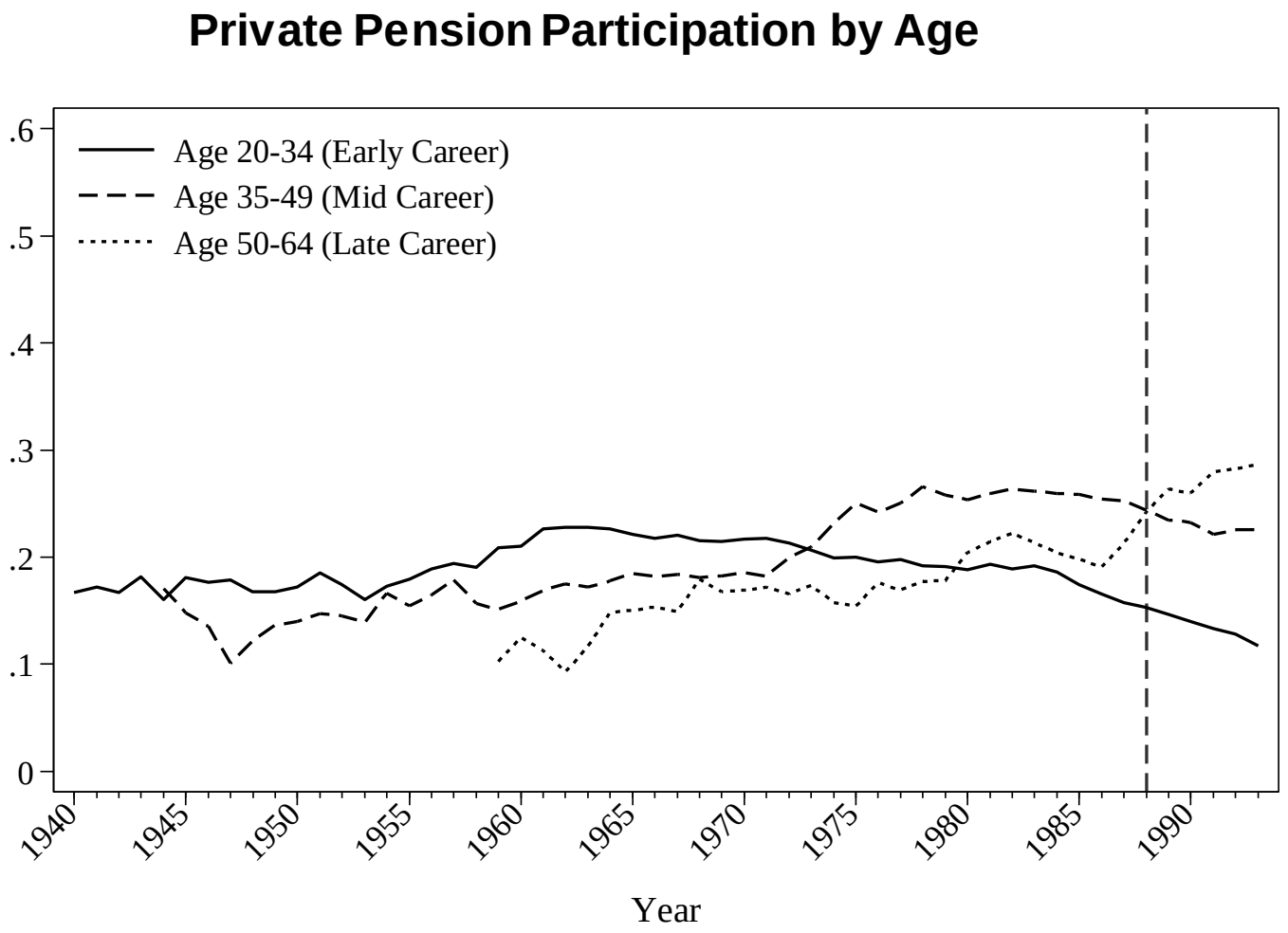
Percentage of Correctly Simulated Eligibility

Pension Type	Correctly Simulated	Number of Observations*
Contributory State Pension	96.08%	9343
Occupational Pension	98.25%	10030
Private Pension	97.36%	1706(Year 2000 onwards)

Some Other Findings



Some Other Findings



1988 is the year when the last major state pension reform took place during the simulated time period

- A generic back simulation framework which could be adaptive to many datasets
  - Eg. ECHP, BHPS, GSOEP, FFS
- Possibilities of life cycle based modelling simulation without the scared long panel

The Implication of Simulated Historical Panels

- An example: Option Value Based Retirement Choice Models
  - Option value (Stock & Wise, 1990) could be modelled on utility maximizing process
  - It requires data (past & future)
  - The dynamics of pension in the past and project future could be explicitly modelled
  - Could incorporate Household Based Behaviour

The Implication of Simulated Historical Panels

Thank you