An Overview of Distributional Modeling, Focusing on Dynamic Microsimulation

Melissa Favreault

The Urban Institute

Washington, DC

February, 2011

Overview of presentation

- Definitions
- Overview of dynamic microsimulation
  - Urban Institute models as examples, not focus
- Simulation analyses using these models
  - What models make the most sense for which applications?
- Strengths and weaknesses / challenges
- Sample output from analyses
- Lessons learned & discussion

What is microsimulation?

- "Micro": Based on a large file with data on individuals
  - Usually a nationally representative sample
- "Simulation": Computer programs that mimic the rules of
  - government programs
  - individual and family behaviors

Why use microsimulation?

- Concern with the distributional consequences of changes to policies / programs
- Representative worker analyses, while helpful for illustration, do not capture diversity, complexity in individual life paths
  - Efforts to improve representative workers based on patterns in longitudinal data
- Can it inform long-run cost estimates also?

Types of microsimulation

- **"Static"**: focused on the present or short-run changes
  - Underlying sample changed only modestly over time (for example through reweighting)
  - Many tax and transfer models
- **"Dynamic"**: focus on the long-run
  - Units are aged year by year
  - Most Social Security models are of this sort
    - Some shorter-run Social Security analyses use static (Herd)

UI Modeling Projects

- Social Security focus:
  - DYNASIM (on-going)
  - MINT (on-going, collaboration with SSA)
  - POLISIM (2003-2007, collaboration with SSA, my knowledge is dated)
- Others (Tax/Transfer):
  - TRIM
  - Tax Policy Center

Other Dynamic Modeling Projects

- Social Security focus:
  - CBOLT: Extensive use for baselines and policy alternatives
  - Policy Simulation Group (Holmer's models): Used by GAO
  - DEMSIM (Gokhale's model)

Structure of a Dynamic Microanalytic Simulation Model

- Starting sample
- Aging algorithms
- Alignment of outcomes from aging process
- Benefit calculators

Compare Models

Model	DYNASIM	MINT6	POLISIM (2007)
Starting data	1990-1993 SIPP matched to PSID and CPS/SER	2001-2004 SIPP matched to administrative data	1980 PUMS, matched to CPS or SIPP/SER
Observations	110,000 (base) 330,000 (horizon)	75,000 core, 325,000 over extended	420,000 (typical)
Sample	All individuals	1926-1975 birth cohorts (plus 1976-2070)	All individuals
Projection period	1993 to 2085	1993 to 2099	1980 to 2090
Alignment	Birth, death, DI, employment, wage growth, immig.	DI, death, wage growth, immigration	Most outcomes

DYNASIM Starting Sample

- Based on a sample of all individuals in the 1990-1993 SIPP panels
- Start with approximately 110,000 persons (can do more or fewer)
- End up with over 300,000
- Individuals in the sample are matched to synthetic earnings histories from PSID and CPS/SER match using complex algorithm

DYNASIM Aging Modules

- Model simulates income and demographic events from 1993 to 2085
- Core:
  - Birth / school / leave home
  - Marriage / divorce
  - Health / work limits / limits in IADLs/ADLs / DI take-up
  - Death
  - Earnings
  - Immigration / emigration
  - Pension coverage (DB and DC)
  - Financial assets and home equity
  - Retirement and benefit take-up
- Postprocessor (benefit calculators):
  - Social Security and SSI
  - Pension amounts
  - Co-resident (non-spouse) family income
  - Total incomes, poverty, federal income taxes

Sample Aging Module

- Regression equations / “splice” or statistical match
  - Logit and probit for discrete outcomes: entry and exit where realistic
  - Linear regression (often with complex errors) for continuous outcomes
  - **Equations only include outcomes the model predicts**
    - **Complexity increases with each outcome introduced**
- Alignment
  - OASDI Trustees’ Assumptions
- Rule-based processes

Birth Module

- Logistic regression equations from NLSY
  - Marital status- and parity-specific equations
  - Covariates include factors like education/school status, time since marriage/last birth, whether mom was a teen mother, sex composition of kids if parity three +
- Alignment
  - OASDI Trustees' intermediate assumptions
- Probability of having twins or triplets
  - Varies by age of mother
- Assign the sex of the baby/babies
  - Varies by race of mother

Alignment / Calibration

- Technical adjustment to meet aggregate targets
  (e.g., from Trustees, Vital Statistics)
  - Variety of techniques
- Minimize alignment in the short-term/historical
  - Necessary because small deviations in rates can compound when processes feed into one another
- Longer-term, it's a mechanism for "steering"
  - Means for integrating judgment
  - Our conversation last month about assumptions

Microsimulation as Sampling

- Errors in starting sample (even a census)
- Errors in parameters (confidence interval)
- Pure Monte Carlo variability

- Supports alignment—can think of it as estimation subject to a constraint

Benefit Calculators

- OASDI
  - Mimic the law in extensive detail
- SSI with stylized state supplements
- Federal and state income taxes
  - Requires extensive assumptions about the tax code going forward

Strengths

- Internal consistency
  - Social Security benefits are a direct function of earnings and disability and marital histories
  - Benefits never aligned
    - If benefits do not line up with historical, need to figure out why
- Straightforward to add up and disaggregate
- Interactions often revealed
  - For example, between programs and between taxes and benefits (e.g., OASDI benefits net of income taxes)
  - RET

Strengths

- Makes simulation of alternatives straightforward
  - Change the law
  - Change behavior (work, benefit claiming) where changes to the law are large enough to be likely to elicit a response
    - Needs to be distributed among those responding
- Often requires policymakers to get extremely specific about intentions, beneficiary subgroups
  - Helps minimize hand-waving over DI, for example

Simulating Alternatives

- Alternative baselines
  - Demographics
  - Economics
  - “Scenario planning”
- Social Security alternatives
  - Formula changes
  - Earnings sharing
  - Personal accounts (including annuitization)
  - Rough cost-neutrality possible

Distributional Output

- Winners and losers by detailed subgroup
  - Average/median size of gains/losses
  - Compared to what? (scheduled/payable)
- Poverty/near-poverty
- Various groups' shares of benefits / taxes
  - Equity across family types

Technical Challenges

- Extensive input data demands
  - Need high-quality longitudinal data
  - PSID, NLS/NLSY, HRS
  - SIPP matched to administrative data incredibly valuable
    - Access and disclosure review challenges
- Extensive output data demands
  - Need tools to organize massive files:
    - Validation tools to display trends over time, across groups
    - Equation-specific tools
  - Need to summarize data in a way that informs busy policymakers

Technical Challenges

- Small groups may be difficult to represent well
- Small offsetting deviations can have large effects on costs / solvency
- Computationally burdensome
  - Problem has diminished greatly in the recent past
  - Still an issue that should be (re)considered when designing and updating
    - UI recently moved most of our SAS postprocessor into FORTRAN
- Need trained analysts, resources

Substantive Challenges

- Interdependencies between processes
  - Model processes jointly
  - Errors correlated
  - Different models confront this in different ways
- Modeling complex outcomes like earnings very difficult, social science still evolving
  - Both cross-sectional and longitudinal consistency
  - Outliers in a regression context
  - Some rely on hybrid splicing / equation approach

Lessons Learned

- Where practical, use most recent database as an input file
  - Detailed files with life histories minimizes imputation
- Use real (ideally administrative) matched earnings histories where possible
- Equations / matching to age a balancing act
  - Separate equations versus extensive interaction terms
    - Use statistical analyses to guide these choices
  - Age-centered regression (Sabelhaus and Walker 2009)
  - Updating regularly can be very valuable

Lessons Learned

- Even in current world of relatively cheap computer power, choices about computing environment are not trivial
  - They affect
    - Sample size
  - They also affect developers' capacity to:
    - Correct errors
    - Change baselines
    - Perform detailed sensitivity analyses

Recommendations

- These models are expensive and complex to build, so it is a good investment to learn from others
  - Can SSA leverage investments across groups (OCACT/ORES)?
  - Stay on top of the literature, communicate with developers from other groups
- Importance of transparency
  - Analysts from Congressional Research Service used DYNASIM, very valuable to have more sets of eyes on the output

TPAM 2007 Recommendations

- Imperative to accelerate development of new models to augment cell-based projections
- Not of the view that microsimulation should replace segmented model, but rather that it should augment it by:
  - Increasing understanding of interrelationships
  - Illustrating impacts for individuals
  - Adding transparency to implicit assumptions
    - Facilitates debate / discussion in the research community

Discusssion

- Look forward to hearing about current state of OACT efforts in these areas
  - Challenges
  - Successes
  - Future plans

Questions for OACT

- Could you please update us on the current status of OACT's microsimulation efforts (including both large-scale dynamic modeling efforts and the development of micro-level databases that facilitate, for example, development of benefit projection estimates)?

Questions for OACT

- Could you also describe any future development plans?
  - We are particularly interested in understanding the relationship between the main (segmented) model and the dynamic microsimulation model under development.
  - What are the implications for the development of benefit projections as the dynamic model progresses?

Questions for OACT

- One of the biggest challenges that microsimulation developers face is the estimation of lifetime earnings histories. Can you tell us about OACT's current approach and how it compares with the literature and other similar models (e.g., CBOLT, DYNASIM, MINT)?
- Are there ways that members of the Technical Panel could offer developers guidance in meeting any microsimulation development challenges?

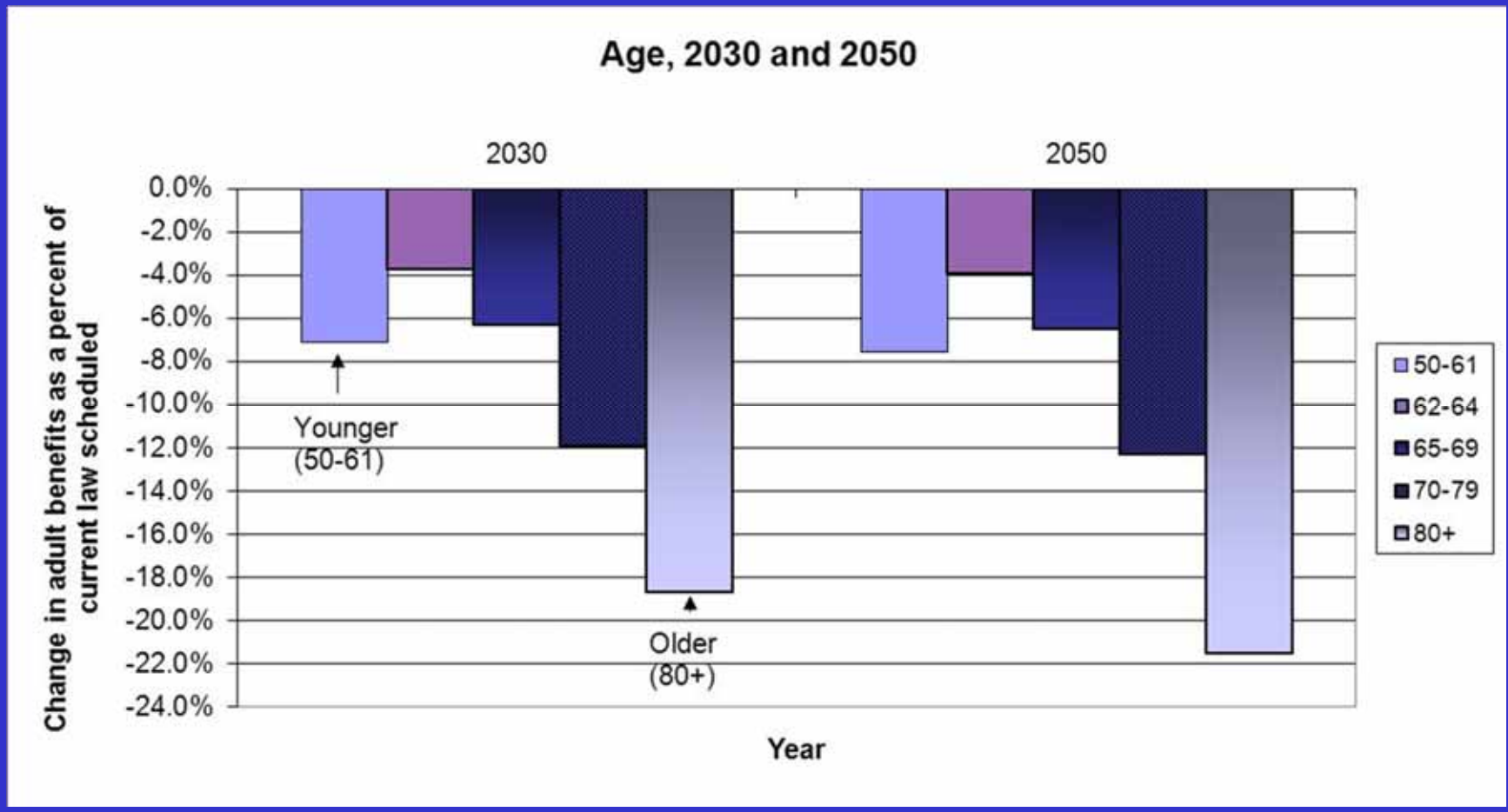
Questions for OACT

- How does OCACT envision the role of alignment in its dynamic microsimulation models?
- Is there potential to take advantage of synergies across dynamic microsimulation modeling projects in different parts of the agency (most notably MINT)?

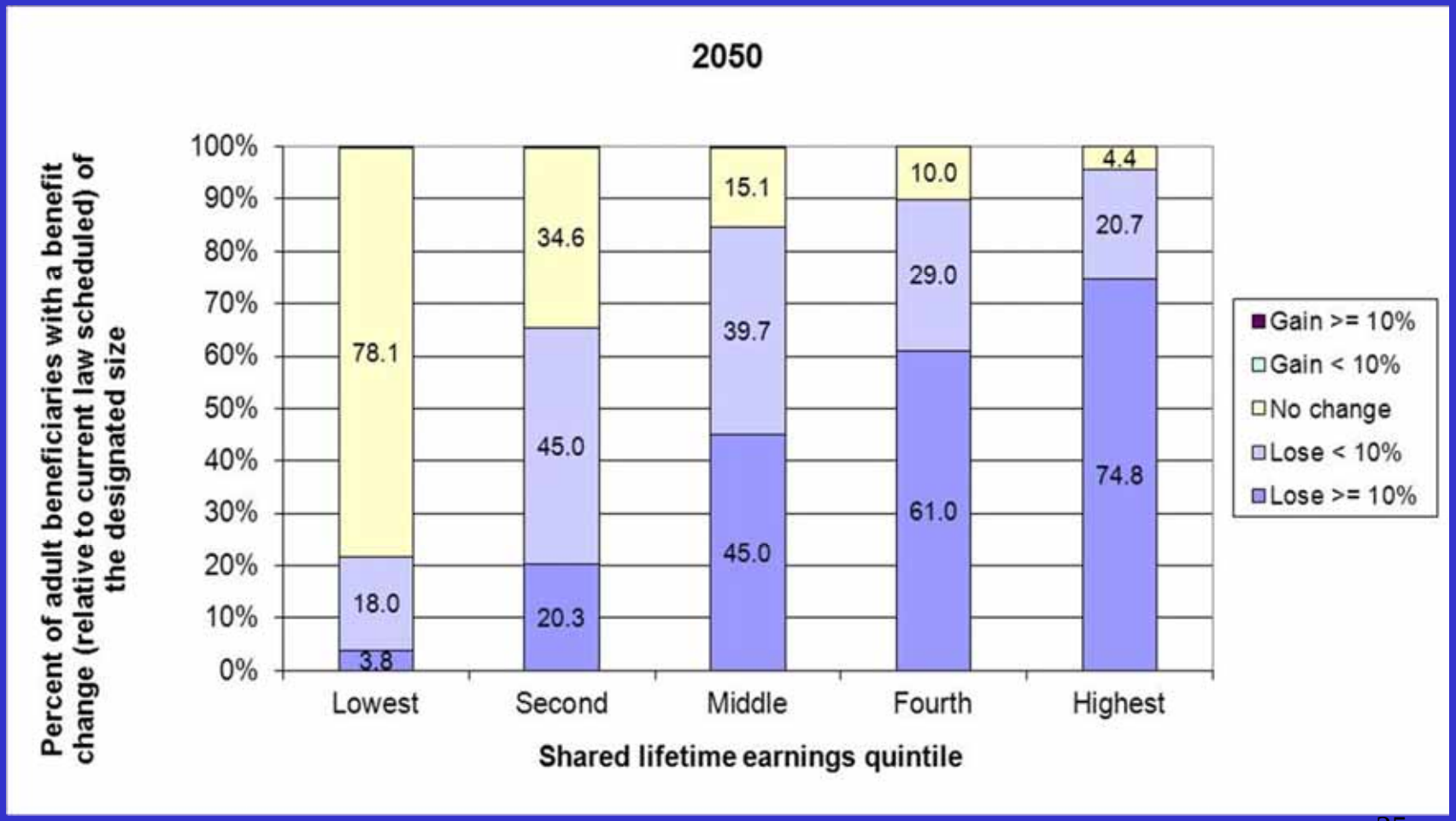
Disclaimer

Views expressed are my own and not those of the Urban Institute, its board, or sponsors.

Distributional Output



Distributional Output



Subgroups Examined

- Men vs. women
- Marital status (history) groups
- Race / ethnicity / nativity
- Education
- Lifetime earnings or income (quintiles, deciles, etc.) and work experience
- Disabled
- Parents
- Interactions