Statistics Canada
www.statcan.gc.ca

Microsimulation and public policy. Issues and prospects

Auckland, Feb 2012

Statistics Canada Statistique Canada

martin@spielauer.ca

Canada

Outline

- What is Microsimulation?
- Microsimulation at Statistics Canada
- Examples
- Conclusions

Why do we simulate?

Perhaps consciousness arises when the brain's simulation of the world becomes so complete that it must include a model of itself. - *R. Dawkins, in "The selfish gene"*

- Simulation = creation and the use of models pursuing a purpose:
  - Exploration
  - Prediction
  - Problem-solving
  - Training
  - Explanation and theory building
  - Raising consciousness
- Many policy questions can only be answered using simulation

What is Microsimulation?

- Social science microsimulation: Computer-simulation of a society or economy in which the population is represented by a large sample of its individual members and their behaviors

When does MS make sense?

If individuals are different and differences are given importance

- No single representative agent
- Study and projections of distributions

If behaviors are more stable or better understood on micro level

- Non-linear tax and benefit rules
- Composition effects

If individual histories matter

- Non-Markov processes: agents have a memory
- Pension projection: contribution history

Where is it used?

- Tax-Benefit analysis
- Pension
- Education
- Population projections
- Health
- Spread of diseases
- Care
- Personnel planning

Why MS at Statistics Canada?

Increase (policy) relevance of data collected by STC

- Data from a single source often can't answer all questions related to a change in policy
- MS allows (and requires) the integration of various data into a coherent platform

Makes data more accessible

- Confidential micro-data might be releasable when integrated into a synthetic database

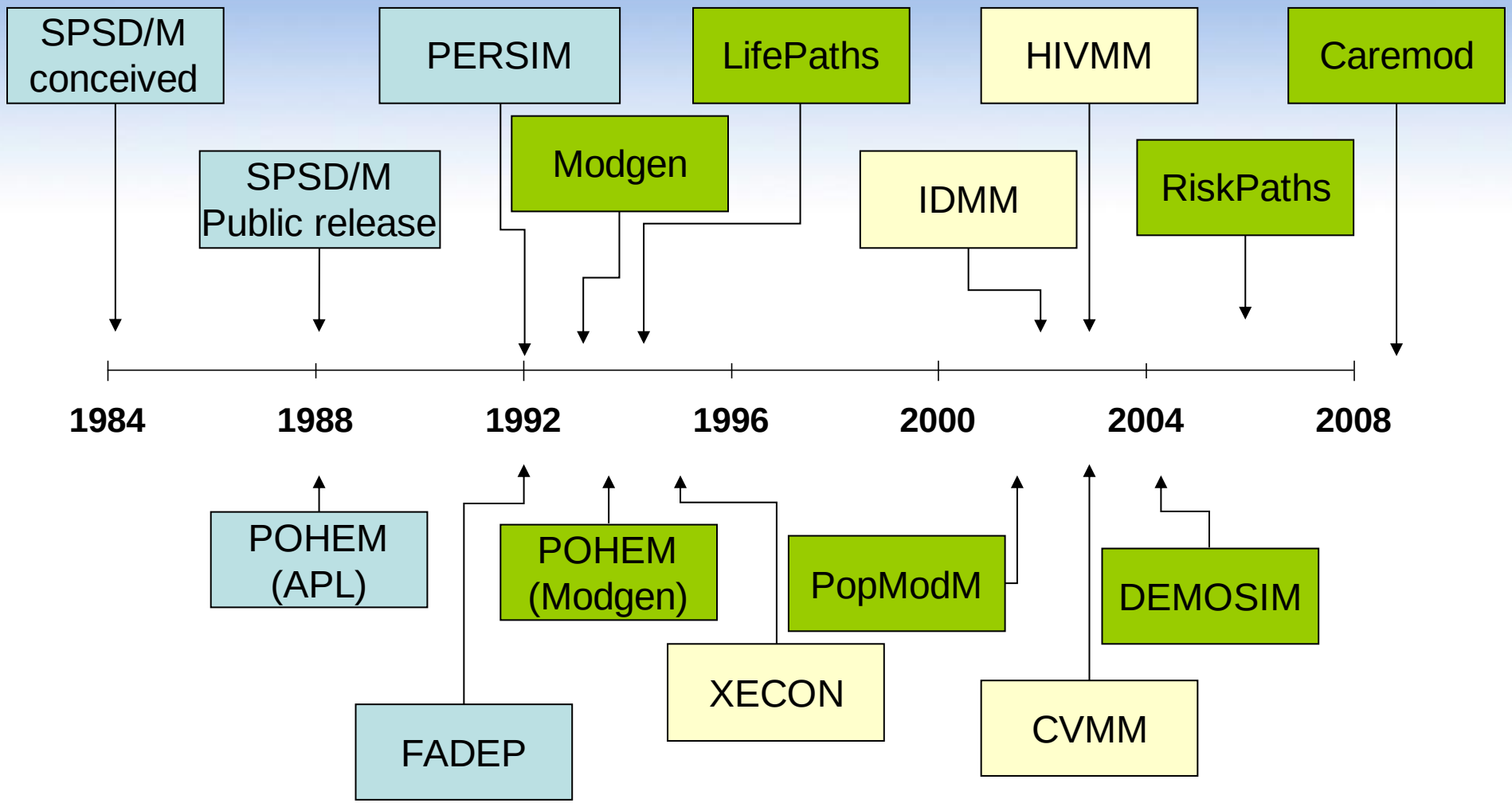
Data quality

- Detection of data gaps and inconsistencies
- Internal data analysis and modeling expertise together with close client relations feed back

Microsimulation at Statistics Canada

- Long history going back to 1984
- Large variety of modes (15+; e.g.:)
  - Static tax benefit model: SPSD/M
  - Large “multi-purpose” socio-economic model: LifePaths
  - Population projection model: Demosim
  - A growing family of health models: Pohem
- Modgen modeling technology and programming language
  - “Common language” at STC
  - Worldwide use
- Provision of modeling support, collaborations & consultancies
  - Academia – Governments – International Organizations

Time Line


SPSD/M conceived
SPSD/M Public release
PERSIM
Modgen
LifePaths
IDMM
HIVMM
RiskPaths
Caremod
1984
1988
1992
1996
2000
2004
2008
POHEM (APL)
FADEP
POHEM (Modgen)
XECON
PopModM
CVMM
DEMOSIM

Who at STC develops MS models?

Modeling Division

- Socio-economic models: LifePaths, SPSD/M
- Modeling technologies: Modgen
- Internal and external modeling support, model prototypes

Health Analysis Division

- Fast growing family of health and disease models
- Various academic and other collaborations

Demography Division

- Demographic projection models

Human Resources Development Division

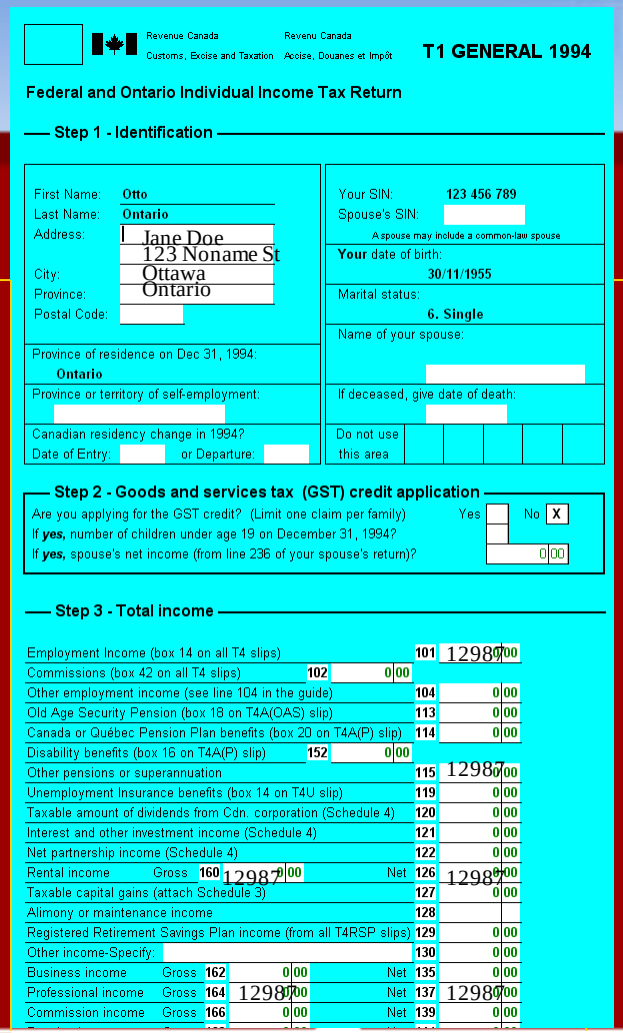
- Personnel models

STATIC MICROSIMULATION
Example: SPSD/M

- Model for analysis of tax-transfer changes in a given year
- Federal & provincial transfers and taxes
  - Static accounting model - no behavioural change
  - Includes tax and transfer rules from 1991 to 2015
- Tax and transfer model is controlled by over 1600 parameters
- Includes (11) commodity taxes & consumption
- Starting database: synthetic database (~200,000 individuals) created using 4 datasets

SPSD/M

	Age	Income	Taxes
	237	37	32,000	10,200
	237	33	12,350	360
	237	11	0	0
	237	7	0	0
	422	59	55,876	16,800
	422	55	421	0
	142	79	8,700	0
	376	19	14,000	0
	30 M		438B	



Revenue Canada, Customs, Excise and Taxation — Revenu Canada, Accise, Douanes et Impôt

T1 GENERAL 1994

Federal and Ontario Individual Income Tax Return

Step 1 - Identification

First Name: Otto
Last Name: Ontario
Address: Jane Doe / 123 Noname St
City: Ottawa
Province: Ontario
Postal Code:

Province of residence on Dec 31, 1994: Ontario
Province or territory of self-employment:
Canadian residency change in 1994? Date of Entry: or Departure:

Your SIN: 123 456 789
Spouse's SIN:
A spouse may include a common-law spouse
Your date of birth: 30/11/1955
Marital status: 6. Single
Name of your spouse:
If deceased, give date of death:
Do not use this area

Step 2 - Goods and services tax (GST) credit application

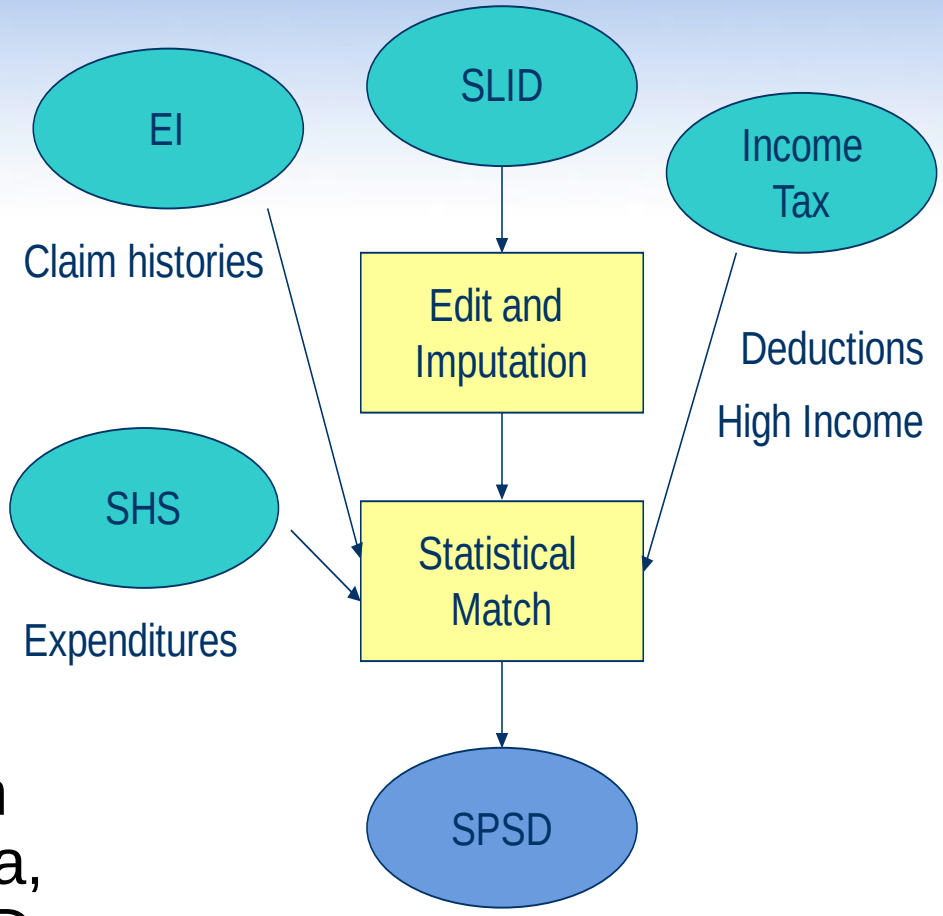
Are you applying for the GST credit? (Limit one claim per family) Yes [ ] No [X]
If *yes*, number of children under age 19 on December 31, 1994?
If *yes*, spouse's net income (from line 236 of your spouse's return)? 0 00

Step 3 - Total income

Employment Income (box 14 on all T4 slips)			101	12987	00
Commissions (box 42 on all T4 slips)	102	0 00			
Other employment income (see line 104 in the guide)			104	0	00
Old Age Security Pension (box 18 on T4A(OAS) slip)			113	0	00
Canada or Québec Pension Plan benefits (box 20 on T4A(P) slip)			114	0	00
Disability benefits (box 16 on T4A(P) slip)	152	0 00			
Other pensions or superannuation			115	12987	00
Unemployment Insurance benefits (box 14 on T4U slip)			119	0	00
Taxable amount of dividends from Cdn. corporation (Schedule 4)			120	0	00
Interest and other investment income (Schedule 4)			121	0	00
Net partnership income (Schedule 4)			122	0	00
Rental income Gross	160	12987 00	Net 126	12987	00
Taxable capital gains (attach Schedule 3)			127	0	00
Alimony or maintenance income			128		
Registered Retirement Savings Plan income (from all T4RSP slips)			129	0	00
Other income-Specify:			130	0	00
Business income Gross	162	0 00	Net 135	0	00
Professional income Gross	164	12987 00	Net 137	12987	00
Commission income Gross	166	0 00	Net 139	0	00

SPSD/M: Data integration

- Accessibility: Releasable synthetic data from 4 sources
- Quality: Added insight in differences of data sources
- Testable quality improvement: Experiments running models on main single original data
- Relevance: many users!



- STC provides tool and training
- Clients run own scenarios
- Users across political spectrum including political parties, media, research institutions and Gov. Dep.

Technology

Dynamic MS introduced 1957

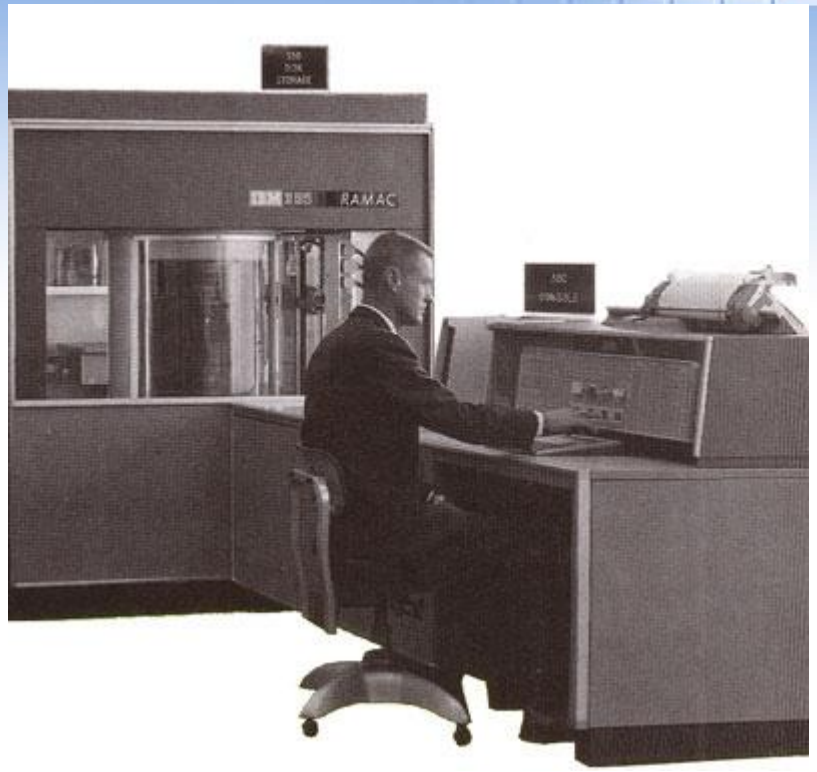
Guy Orcutt: “A new type of socio-economic system”

Impressive progress:

- Hardware
- Statistical software
- Data: Longitudinal; administrative

Programming

- Reinventing the wheel?

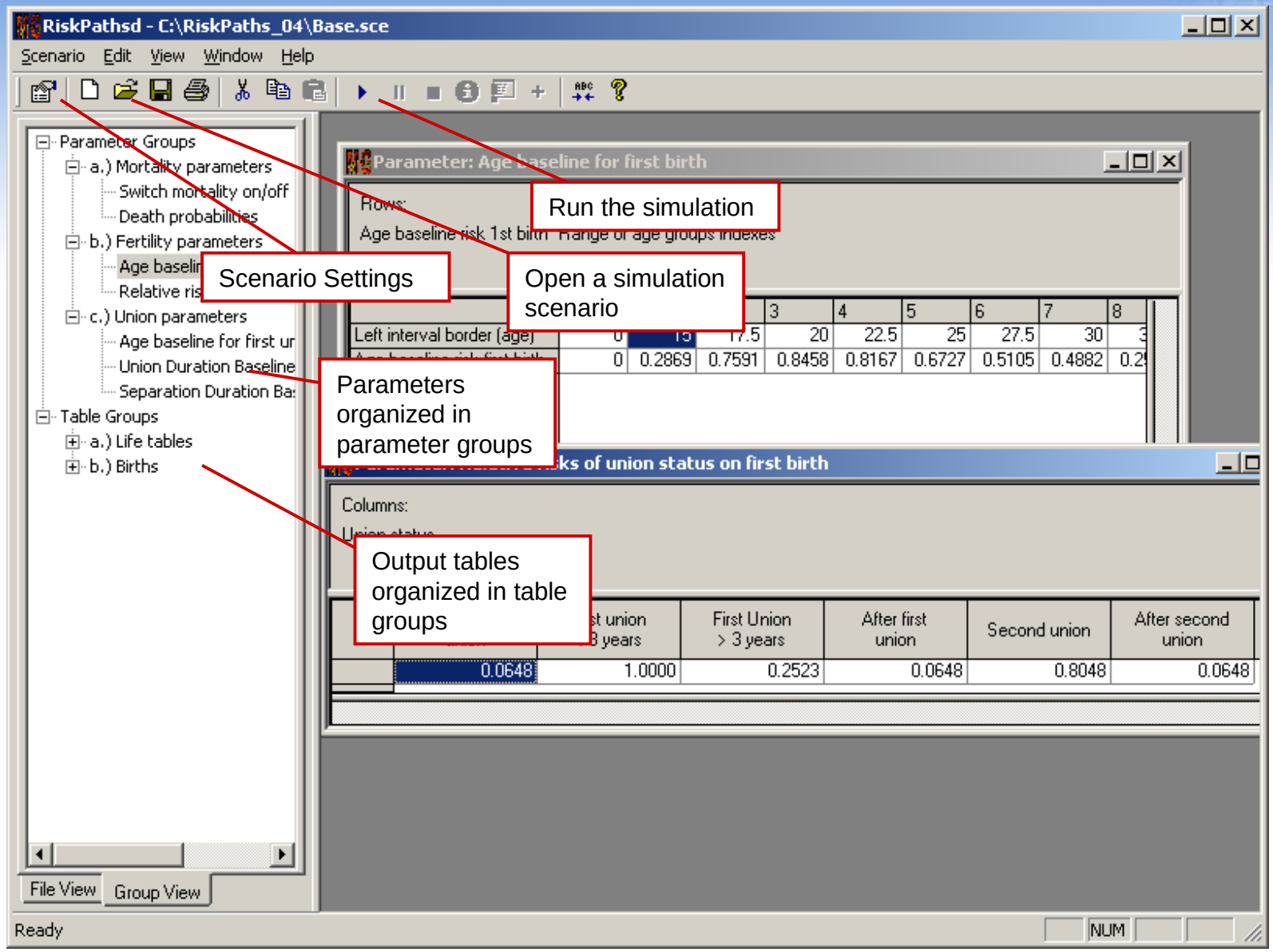


1956 IBM 305 RAMAC
First computer with a disk drive - fifty 24" discs – total 4.4 MB total; $35,000/year

Modgen technology

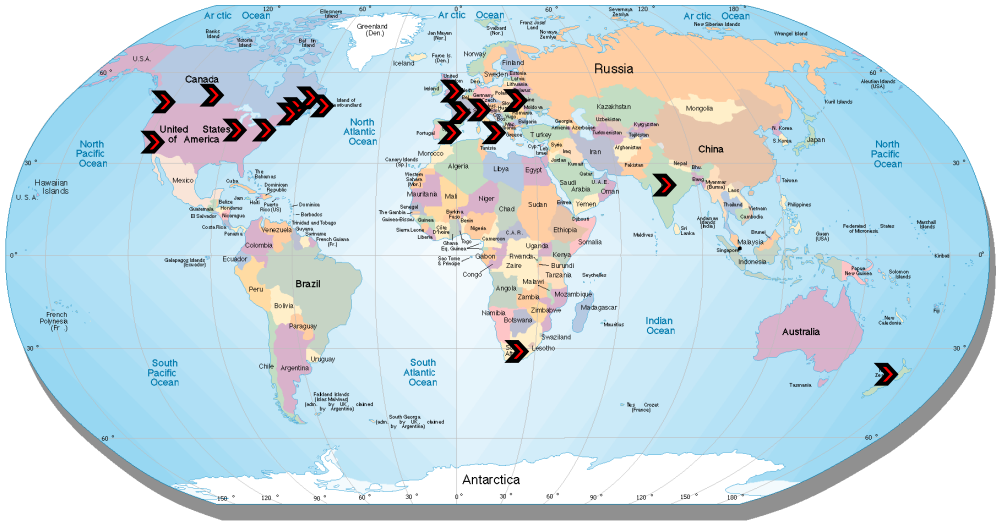
- Modgen is a generic MS programming language
  - Continuous time and discrete time
  - Interacting and non-interacting populations
- It is free (builds on MS C++); support contracts available
- Efficient applications: fast; large populations; multithreading
- Efficiency coding: No advanced programming skills required
- User-friendly: Creates a stand-alone model with a complete visual interface and detailed model documentation.
- Powerful tabulation facilities, Standard errors and coefficients of variation for tables; Visualization of individual life courses

Modgen: User interface



Modgen: Conclusions

- Common language for MS at STC: more flexible team
- Analysts as opposed to programmers can develop and maintain models
- Efficiency in development of models
  - Avoid re-inventing the wheel
  - Shared high standards in model speed. UI, documentation
  - Less error prone
- Fast creation of prototypes one of the methods used to develop new markets
- Vehicle for teaching MS at universities



Microsimulation and Demography

- Demographic research showcases trends in social sciences and the transformative potential of microsimulation
  - From structure to processes
  - From macro to micro towards multi-level integration
- Traditional Macro projections limited to few variables. Miss need for more detailed projections; poor micro foundation
- Micro-analysis falls short of being able to assess overall impact of micro-level behaviours and changes.

- Starting from Macro level: MS can replicate macro models which can be stepwise refined accounting for heterogeneity
- Starting from Micro level: MS is the tool to add synthesis to analysis: models of single processes brought together to simulate macro dynamics

Microsimulation and Demography

- Population projections showcase one of the fundamental trade-offs in microsimulation: detail versus prediction power
  - Detail leads to specification randomness: too many variables produce noise
  - Simplicity leads to misspecification errors: too few variables, i.e., too simple models
  - Microsimulation allows scaling of models by theoretical considerations

Demosim: Context

General:

- Rapid increase in the ethno-cultural diversity
- Demand for detailed projections going beyond age and sex

Institutional:

- Funded by external federal departments (PCH, HRSDC, CIC and INAC) responsible of programs related to Employment Equity, Labour market integration, and fight against racism and discrimination

Results

- Released (2011) “Projections of the diversity of the Canadian population, 2006-2031”
- Follow-up studies, e.g. for Aboriginal Population

Demosim – Base Population & Variables

- Starting population is a 20% sample of the census (7 Mill)

- Age
- Sex
- Place of residence (33 metropolitan areas, provinces, on-off Indian reserves)
- Place of birth
- Generation status
- Visible minority group
- Mother tongue
- Aboriginal identity & Registered Indian Status
- Religious denomination
- Highest level of schooling
- Labour force participation

Demosim - “Typical modeling strategy”

- Many models follow a 2 step approach
  - Base risks / probabilities: equivalent macro models
  - Relative risks / probabilities: detailed accounting for individual characteristics

  - Fertility
  - Mortality
  - Emigration
  - Labour force participation
  - Education (indirect)
- Allows to combine robust with detailed data sources
- Supports scenario building

Demosim: Conclusions

- Demographic microsimulation projection meets demand for more detailed projections while leaving models transparent
- The project demonstrated the strength of microsimulation in specialized projection models
- Demonstration that MS models can be developed fast with modest input of resources
- High confidence in results as all aggregate outcomes closely resemble macro projections (and differences are explainable)
- Transparent assumptions; scenarios created by clients
- Very high media coverage

Policy maker's perspective

Attractive features

- Virtual world: Test and fine-tune new policies
- Detail: Policy rules at any level of detail; distributions
- Longitudinal perspective: Sustainability issues

Transformative potential:

- Shift in focus on sustainability issues and time dimension
- Better understanding of trade-offs and distributions
- Accounting: Integration to social accounting framework

Many challenges:

- Management of complexity and randomness
- Theoretical foundation

Life Course Perspective

MS fits naturally into dominant research paradigms

- Individual agency: Modeling of decisions and events in context and on level they take place
- Interaction between careers: Life-course interactions between parallel careers within a changing socio-economic context
- Interaction between individuals: Family and social networks, Intergenerational and other transfers and transmission

LifePaths

- LifePaths simulates a large sample of detailed individual life courses which together represent the Canadian population in its socioeconomic diversity
- Dynamic – large - multi-purpose
- Family context: simulation of spouses, children, grandchildren

- Continuous time model
- Case based model
- Synthetic starting population
- Historic depth: first actors born 1871

LifePaths – Behaviours & Systems

- Family demographics
- Education
- Health
- Employment
- Income and earning
- Home ownership
- Social insurance
- Benefits
- Taxes
- Public pension plans
- Senior benefits (OAS, SPA, GIS)
- Private pension plans

LifePaths – Data

- The content of LifePaths is largely embodied in its behavioral equations estimated from micro data
- Challenge: both longitudinal and cross-sectional consistency
- LifePaths makes use of a multitude of data sources.
- LifePaths can be viewed as a data integration exercise

LifePaths application example: Replacement adequacy of retirement income

Motivation: concerns about the “replacement adequacy” of the retirement income system for future retirees

- declining RPP coverage for current workers
- perceived inadequate take-up of RRSP saving

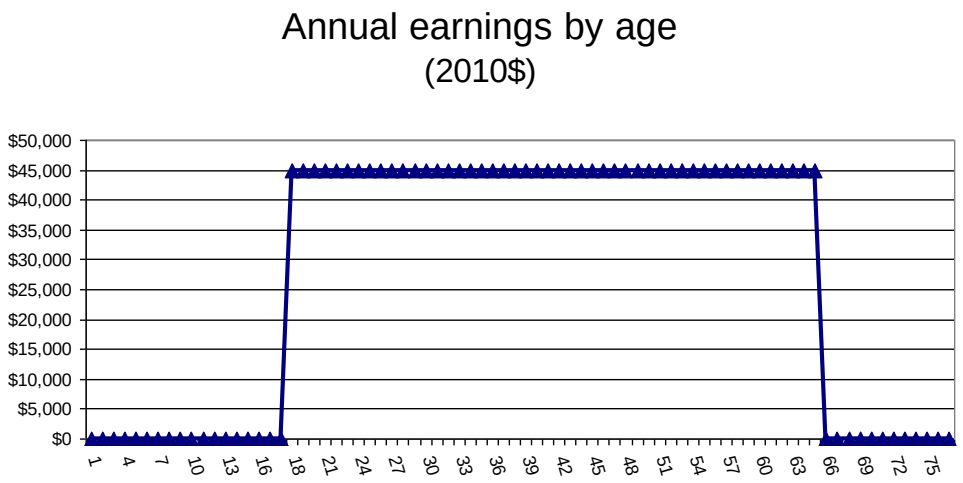
Why LifePaths:

- Provides detailed distributional results:
- Complete lifetimes observable: detailed earning histories; flexibility in measuring pre-retirement earnings, e.g. best 10 years
- Allows exploration of “what-if” questions

Example: Replacement adequacy of retirement income

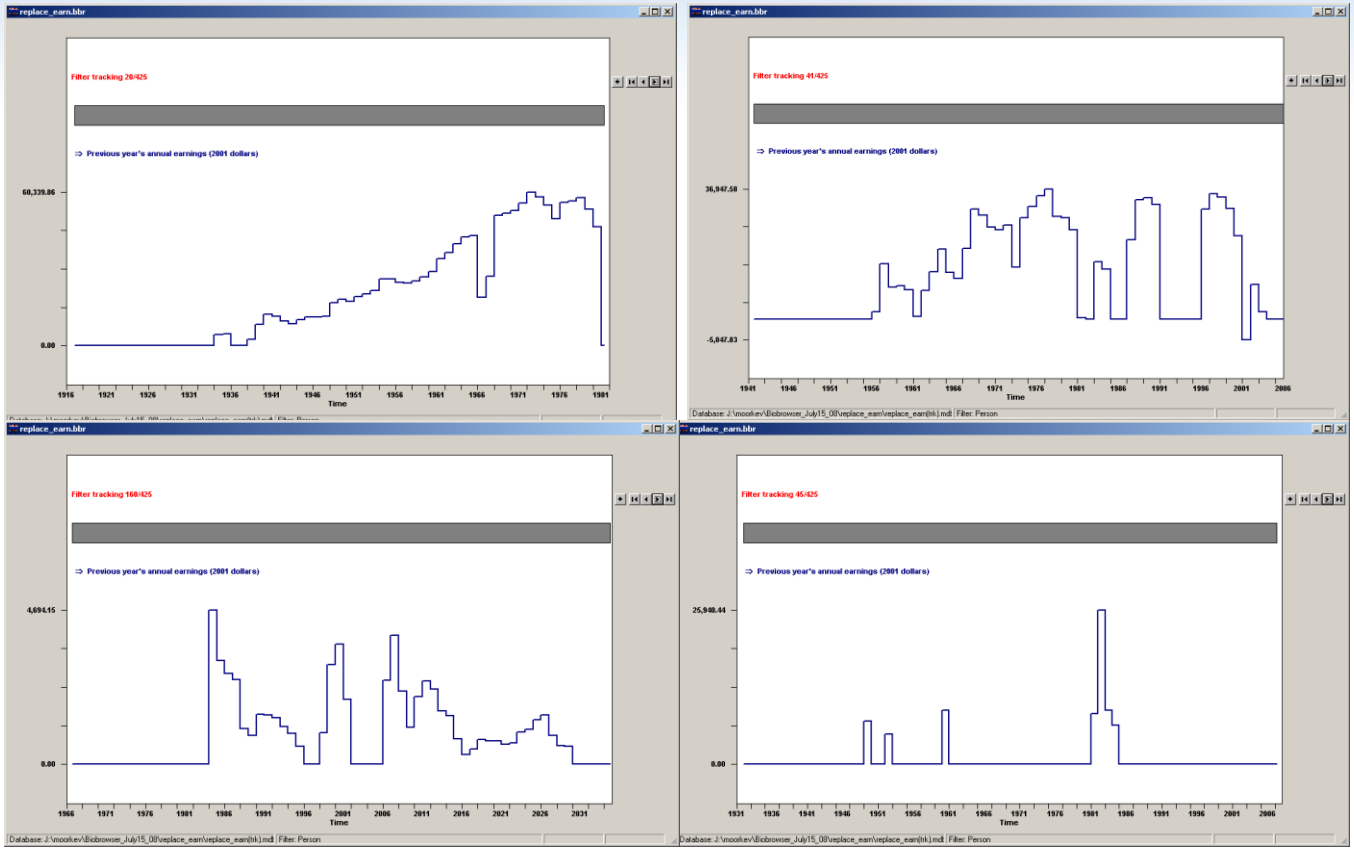
Traditional “typical case” analysis

- Starts work at age 18
- Earns the average wage every year
- Stops working at age 65
- C/QPP provides a pension that replaces 25% of his/her average earnings



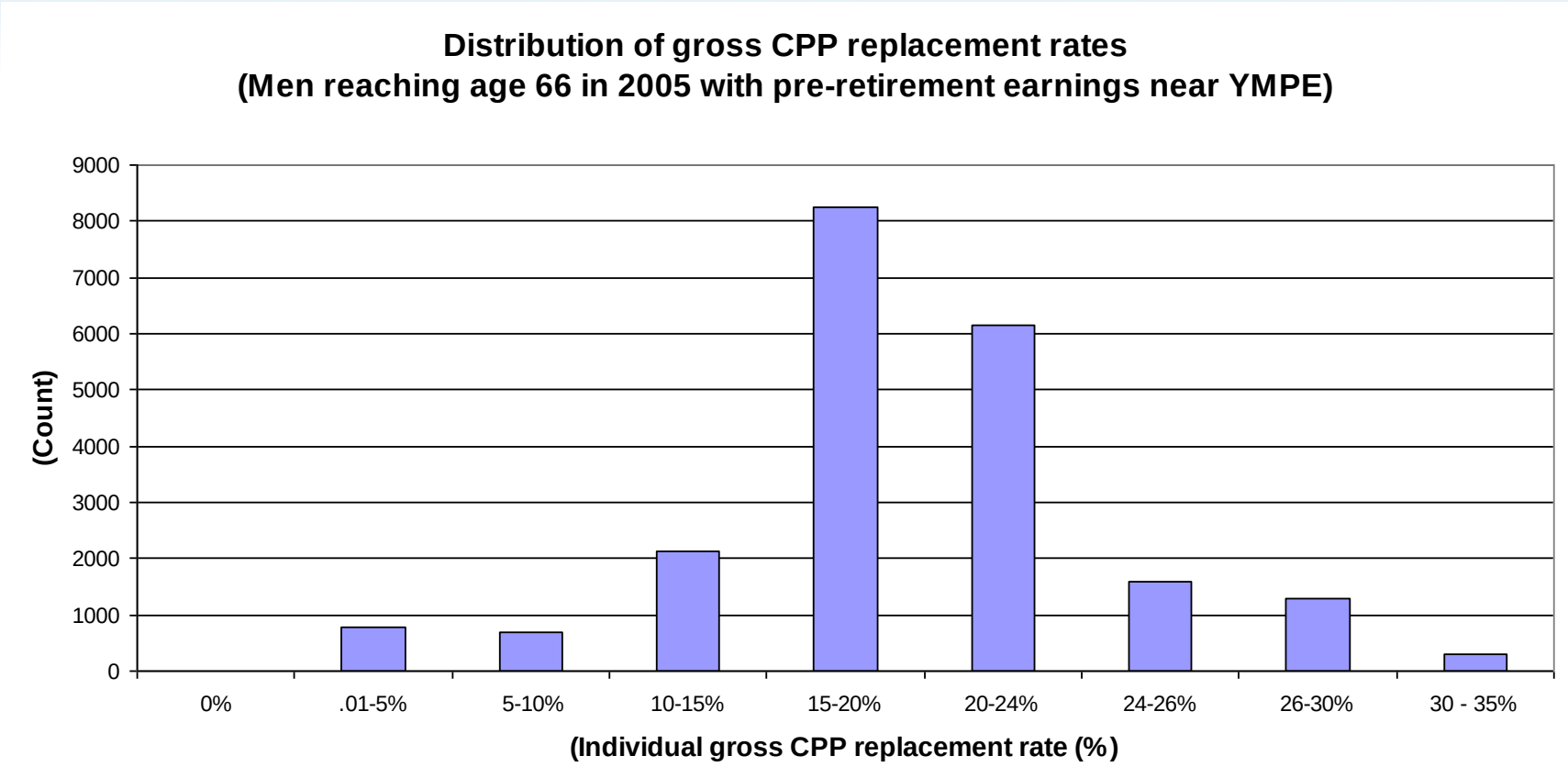
Example: Replacement adequacy of retirement income

LifePaths analysis: realistic careers



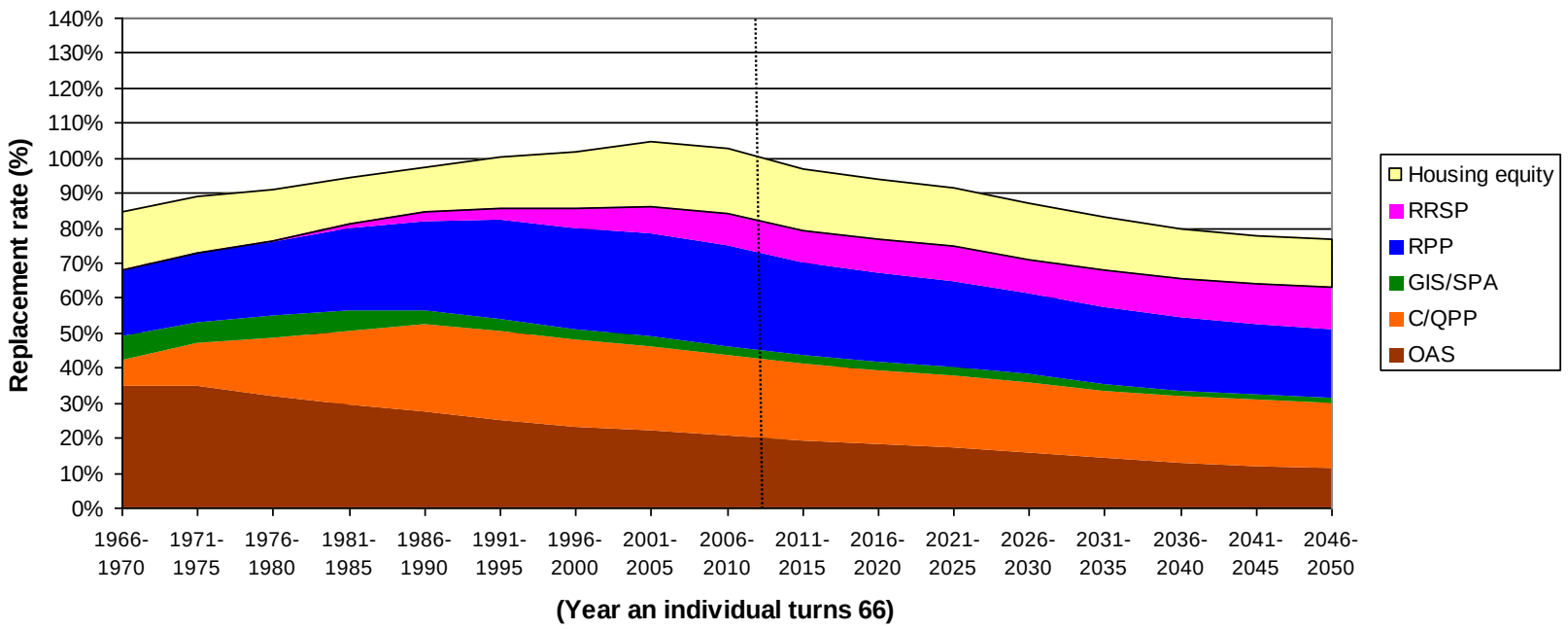
Example: Replacement adequacy of retirement income

Distributional analysis



Example: Replacement adequacy of retirement income

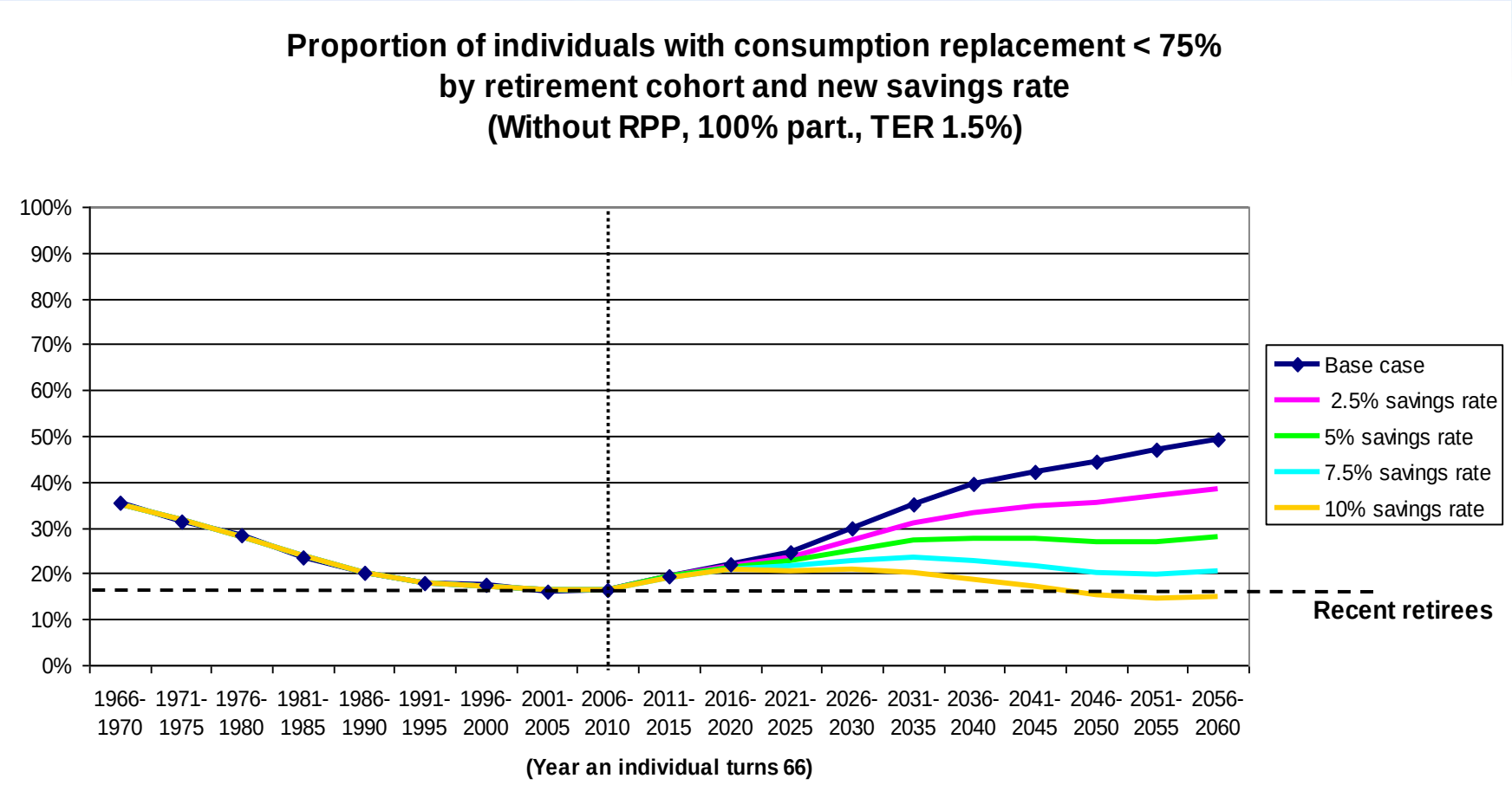
Cohort analysis



Source: Kevin D. Moore, William Robson, Alexandre Laurin (2010) Canada's Looming Retirement Challenge; C.D. Howe Institute Commentary Pension Papers

Example: Replacement adequacy of retirement income

What-if behavioral analysis



LifePaths: Benefits & Lessons

- Complex “general purpose” models come at a price, they require strong leadership, a professional “permanent” team and robust funding
- Policy needs ebb and flow; balancing updates and core development with client’s applications is a difficult task.
- Modularity and a technology that allows modellers program their models crucial
- Many spin-offs (which have to be valued)
  - Expertise; “training lab”
  - Readiness: base for new applications
  - Contribution to quality and relevance of data
- Management of complexity (and ambition) difficult; mixed international experience and perception

Conclusions

- There are many reasons for expecting a boom in MS
  - Many policy questions require micro-simulation
  - MS is a logical next step complementing data analysis
  - MS will probably replace many cell-based models: population projection, actuarial models, disease models
  - Technologies like Modgen make MS accessible
- Microsimulation is still under-used and has to find its way into the standard toolbox of social scientists and into classrooms
- Students pick up the ideas and develop skills fast
- Microsimulation has high data demands. Statistical offices are a logical collaborator in microsimulation projects; In return MS can enhance the relevance, quality, and accessibility of data