SociaLab: A Census-based simulation tool for policy inquiry

COMPASS Seminar Series
5th April, 2019

Peter Davis and Roy Lay-Yee
Department of Statistics, COMPASS Research Centre
University of Auckland
- Features a full-scale, realistic, working simulation model of society based on demographic and social information and transitioning through time
- Contains a comprehensive description of the construction of the working model, together with details of a novel open-source micro-simulation method that will facilitate transfer, application and learning across sites
- Includes worked examples of key policy and substantive questions tested with the simulation model against real data
Outline

• Foundational - Peter
  – Inspiration, Objectives
  – Background, Framework
  – Counterfactuals

• Operational - Roy
  – Data, Statistical analysis, Simulation
  – Software

• Aspirational - Peter
  – Results
  – Strengths and limitations
  – Future
The Inspiration

• New Zealand
  – 1890-1920 a “social laboratory”
  – 1980-2010 a “transformational period”

• Canada
  – The Social Policy Simulation Database and Model
  – OpenM++ open source microsimulation platform
Three Objectives

• To construct a “whole-of-society” simulation model of New Zealand over the period 1981-2006 using microdata from the longitudinal 5-yearly Census

• To formulate and test policy counterfactuals about a period of far-reaching change

• To develop an Inquiry tool – SociaLab – that is both interrogable and visual
The Background

• Research programme in simulation at COMPASS
  – Marsden (2005) – residential segregation; partnership
  – MBIE – early life course (2009), knowledge laboratory (2013)
  – RSNZ, James Cook (2015) – “social laboratory”
  – MSD, Ernst and Young (2016) – vulnerable children investment

• Developments in data access at SNZ
  – Longitudinal Census (NZLC)
  – Remote access data facility (DataLab)
  – Integrated Data Infrastructure (IDI)
Conceptual Model

Simulation framework – at each time point

- Step 0: Previous values
- Step 1: Population dynamics: Exits death; emigration
- Step 2: Demographics age (+5), gender, ethnicity, region of birth
- Step 3: Living arrangements living alone, partnered, living with dependent children
- Step 4: Non-material Assets studying, training, highest education level, religion
- Step 5: Material Assets employment, personal income, household income, welfare receipt
- Step 6: Deprivation NZDep
- Step 7: Population dynamics: Entries immigration (years in NZ); birth

Direction of flow within ‘a year’ (i.e. at each time-point)
- Any prior factor can affect a subsequent factor in the sequence
- Values from the previous ‘year’ can affect current values
## Operational Detail

Simulation framework - showing variables simulated

<table>
<thead>
<tr>
<th>Step</th>
<th>Variables simulated</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Previous values</td>
<td>Age</td>
</tr>
<tr>
<td>1</td>
<td><em>Population dynamics – Exits</em>: death; emigration</td>
<td>All</td>
</tr>
<tr>
<td>2</td>
<td><em>Demographics</em>: age (+5); gender <em>(time-invariant)</em>; ethnicity <em>(time-invariant)</em>; region of birth <em>(time-invariant)</em></td>
<td>All</td>
</tr>
<tr>
<td>3</td>
<td><em>Living arrangements</em>: retain three separate variables conditional on “living alone”. Living alone – if yes, then partnered = no and living with dependent children = no. If not living alone, then partnership status (y/n); living with dependent children, i.e. age &lt;15 or &lt;18 if in full-time education or training (y/n)</td>
<td>0–14 (never living alone, nor partnered, nor living with dependent children) 15+</td>
</tr>
<tr>
<td>4</td>
<td><em>Non-material assets</em>: in full-time education or training; education (highest level) [personal factors]; religion [household factor]</td>
<td>0–14 (household factors only) 10–14 15+</td>
</tr>
<tr>
<td>5</td>
<td><em>Material assets</em>: employment; personal income (CPI-adjusted); welfare receipt [personal factors]; household income (CPI-adjusted) [household factor]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td><em>Standard of living</em>: deprivation; housing tenure [household factors]</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td><em>Population dynamics – Entries</em>: immigration (years in NZ: born in NZ/longer-term immigrant/recent immigrant&lt;5 years); birth (new-born in dwelling, aged 0–4)</td>
<td>All; Women 15–49</td>
</tr>
</tbody>
</table>
“Seven Ages”
(All the world’s a stage,
As You Like It, W. Shakespeare, First Folio, 1623)

• Early Childhood – health & thriving
• Childhood and Youth – education and readiness for life
• Young Adulthood – gaining & keeping employment
• Later Adulthood – settling into stable partnership
• Middle Adulthood – successfully raising families
• Older Life – retirement and successful ageing
• Later Life – the risks of dependency
## The Framework: Early-life trajectories

<table>
<thead>
<tr>
<th>Census</th>
<th>Age</th>
<th>Living arrangements</th>
<th>Education</th>
<th>Employment</th>
<th>Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>5</td>
<td><em>Family of origin</em></td>
<td><em>At school</em></td>
<td><em>NA</em></td>
<td><em>NA</em></td>
</tr>
<tr>
<td>1986</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>15</td>
<td><em>Live alone/with others</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>20</td>
<td><em>Partnering</em></td>
<td><em>Study-training</em></td>
<td><em>Employed or Unemployed or Home-maker</em></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>25</td>
<td><em>Having children</em></td>
<td></td>
<td></td>
<td><em>Own home or Rent</em></td>
</tr>
<tr>
<td>2006</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Framework: Mid-life trajectories

<table>
<thead>
<tr>
<th>Census</th>
<th>Age</th>
<th>Living arrangements</th>
<th>Education</th>
<th>Employment</th>
<th>Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>35</td>
<td>Live alone/with others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>40</td>
<td>Partnering</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>45</td>
<td>Having children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>55</td>
<td></td>
<td>Study-training</td>
<td>Employed or Unemployed or Home-maker</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td>Own home or Rent or Institution</td>
</tr>
</tbody>
</table>
## The Framework: Late-life trajectories

<table>
<thead>
<tr>
<th>Census</th>
<th>Age</th>
<th>Living arrangements</th>
<th>Education</th>
<th>Employment</th>
<th>Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>70</td>
<td><strong>Live alone/with others</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>75</td>
<td><strong>Partnering</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>80</td>
<td><strong>Having children</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td>Own home or Rent or Institution</td>
</tr>
</tbody>
</table>
“Capitals”

• Material
  o Employment
  o Income

• Non-material
  o Education (human)
  o Religion (cultural)
  o [Social]
  o [Functional/health]
The Counterfactuals

• “What If?” counterfactual scenarios
  – The liberalisation of immigration
  – Early childhood education, in-work family support
  – The “baby boomer” generation
  – The availability of life-course assets/capital
  – Future projections
Methods

• Data preparation
  – Harmonise Longitudinal Census data series
  – Missing data imputation (using MICE method)
  – Supplement with data on “exits” and “entries”

• Statistical analysis (regression)
  – Use inter-censal data to estimate transitions

• Simulation – reproduces Census parameters

• Interrogation software
  – base model vs. scenarios with adjusted settings
## Imputation

### Imputation models for ‘starting sample’: Adults (15+)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Type of model</th>
<th>Significant predictors ($p &lt; 0.05$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership</td>
<td>Logistic</td>
<td>Age, gender, NZ European/Other ethnicity, birth region, living alone, living with children, in study/training, education, employment, welfare receipt, personal income, household income, deprivation, housing tenure</td>
</tr>
<tr>
<td>Education</td>
<td>Ordinal</td>
<td>Age, Māori ethnicity, Pacific ethnicity, NZ European/Other ethnicity, birth region, years in NZ, living alone, partnership, living with children, in study/training, religion, welfare receipt, personal income, deprivation, housing tenure</td>
</tr>
<tr>
<td>Employment</td>
<td>Multinomial</td>
<td>Age, gender, Māori ethnicity, birth region, years in NZ, partnership, religion, welfare receipt, personal income</td>
</tr>
<tr>
<td>Welfare receipt</td>
<td>Logistic</td>
<td>Age, gender, Māori ethnicity, birth region, living alone, partnership, living with children, in study/training, education, employment, welfare receipt, personal income, household income, deprivation, housing tenure</td>
</tr>
<tr>
<td>Personal income</td>
<td>Linear</td>
<td>Age, gender, NZ European/Other ethnicity, living alone, partnership, living with children, in study/training, education, employment, welfare receipt, household income, housing tenure</td>
</tr>
<tr>
<td>Household income</td>
<td>Linear</td>
<td>Age, gender, birth region, living alone, partnership, living with children, in study/training, religion, education, employment, welfare receipt, personal income, deprivation, housing tenure</td>
</tr>
<tr>
<td>Deprivation</td>
<td>Ordinal</td>
<td>Māori ethnicity, Pacific ethnicity, NZ European/Other ethnicity, living alone, partnership, in study/training, education, welfare receipt, household income, housing tenure</td>
</tr>
<tr>
<td>Housing tenure</td>
<td>Logistic</td>
<td>Age, gender, Māori ethnicity, Pacific ethnicity, Asian ethnicity, NZ European/Other ethnicity, birth region, years in NZ, living alone, partnership, living with children, in study/training, education, employment, welfare receipt, personal income, household income, housing tenure</td>
</tr>
</tbody>
</table>
## Starting Sample (1981)

<table>
<thead>
<tr>
<th>Pair = 0, Year = 1</th>
<th>Time-invariant</th>
<th>Time-variant</th>
<th>Categorisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>(incremental)</td>
<td>raw</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>y</td>
<td>male/female</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>y</td>
<td>binaries: European-&amp;-other, Maori, Pacific, Asian</td>
<td></td>
</tr>
<tr>
<td>Number of years in NZ</td>
<td>(incremental)</td>
<td>categories: ‘born in NZ’, 5+ years, 0-4 years</td>
<td></td>
</tr>
<tr>
<td>Country of birth</td>
<td>y</td>
<td>region: NZ, Pacific, Asia, Europe, Americas, Middle East/Africa</td>
<td></td>
</tr>
<tr>
<td>New-born (in dwelling) (age 0-4)</td>
<td>y</td>
<td>yes/no</td>
<td></td>
</tr>
<tr>
<td>Living alone</td>
<td>y</td>
<td>yes/no</td>
<td></td>
</tr>
<tr>
<td>Partnership status</td>
<td>y</td>
<td>partnered-married (yes/no)</td>
<td></td>
</tr>
<tr>
<td>Living with dependent children</td>
<td>y</td>
<td>yes/no</td>
<td></td>
</tr>
<tr>
<td>Studying (in full-time education/training)</td>
<td>y</td>
<td>yes/no</td>
<td></td>
</tr>
<tr>
<td>Education (Highest level)</td>
<td>y</td>
<td>no qualification, school, post-school, tertiary</td>
<td></td>
</tr>
<tr>
<td>Religion</td>
<td>y</td>
<td>none, Christian, Other</td>
<td></td>
</tr>
<tr>
<td>Income (personal)</td>
<td>y</td>
<td>NZD - Consumers-Price-Index-adjusted to 2013 value</td>
<td></td>
</tr>
<tr>
<td>Income (household)</td>
<td>y</td>
<td>NZD - Consumers-Price-Index-adjusted to 2013 value</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>y</td>
<td>employed, unemployed, not in labour force</td>
<td></td>
</tr>
<tr>
<td>Welfare receipt</td>
<td>y</td>
<td>yes/no (income-tested benefits only)</td>
<td></td>
</tr>
<tr>
<td>Deprivation (area-based)</td>
<td>y</td>
<td>NZDep quintiles</td>
<td></td>
</tr>
<tr>
<td>Housing tenure</td>
<td>y</td>
<td>own / not own home</td>
<td></td>
</tr>
</tbody>
</table>
Predictive Equations with stochastic element

- Example: Probability of ‘being partnered’ at age 25-34 in 1986 (derived from logistic regression)

\[ P(\text{Outcome} = \text{Yes} | x_1, \ldots, x_n) = \frac{e^{\alpha + \beta_1 x_1 + \ldots + \beta_n x_n}}{1 + e^{\alpha + \beta_1 x_1 + \ldots + \beta_n x_n}} \]

where \( x_1, \ldots, x_n \) denote significant predictors (p<0.05), \( \beta_1, \ldots, \beta_n \) denote their coefficients, and \( \alpha \) is the intercept

- Main predictors: census-pair (1981-86), previous partnership status (in 1981), age, gender, ethnicity, income, religion, beneficiary, deprivation

- Stochastically assign ‘being partnered’ or ‘not’ – random number compared to probability (from predictive equation)
Simulation Framework – *across time points*

**BASE**

*(Starting Sample)*

**SIMULATED**


direction of flow across ‘years’
“Exits” and “Entries”

Simulation schedule

- Main module: simulation starting from 1981
- Exits: emigration and death – probabilities calculated from official statistics
- Entries: immigration and birth – random draws from census unit record data
- Calibrated to aggregate census numbers and composition
Simulation engine - inputs and outputs

Starting sample
Initial conditions at Year 1 ($Y_1$)

Statistical rules
How to transition

Simulation engine
$Y_1 \rightarrow Y_2 \rightarrow \ldots \rightarrow Y_i$
Stochastic yearly update of individual characteristics

Manipulation and testing

Results of Base simulation (status quo)

Results of Projection (future)

Results of Scenario (change)

Simario programmed in R
Shiny web interface
The Software

• Developed in R language (open source)
• **Simario** - to read data and implement simulation
  – source code at [https://github.com/kcha193/simarioV2](https://github.com/kcha193/simarioV2)
• **Shiny** web-based application - user-friendly tool for interrogation and visualisation
  – source code at [https://github.com/kcha193/SociaLabShiny](https://github.com/kcha193/SociaLabShiny)
  – application at [https://compassnz.shinyapps.io/SociaLabShiny](https://compassnz.shinyapps.io/SociaLabShiny)
• We endeavour to deposit as much as possible in the public domain
• We hope to share our software with both developers and end-users in research and policy communities
Summing up microsimulation

1. Conceptualisation
   1.1 Design simulation to mimic individual transitions through life course

2. Data preparation
   2.1 Build base file. 1% sample of >3 million Census 1981 = 30,174
   2.2 Harmonise 1981-2006 data series to generate usable inter-censal pairs
   2.3 1% sample of 11.4 million individual inter-censal pairs = 110,000
   2.4 Statistical analysis of inter-censal pair data to derive predictions of changes in individual states and behaviour through life course

3. Implementation
   3.1 Starting with base sample, apply predictive equations progressively from 1981 to 2006 – in effect reproducing Census over time
   3.2 Check these synthetic data against actual Census to make sure that we are reproducing it accurately “from the bottom up”

4. Application
   4.1 Design and test scenarios by varying relevant factors in data and probabilities
The Results (of counterfactuals, for overall population)

– The liberalisation of immigration
  • we alter proportion of immigrants to that prevalent before liberalisation (i.e. decrease) – result is: decreased income

– Early childhood education, in-work family support
  • we alter proportion of mothers in employment to that before these programs (i.e. decrease) – little impact

– The “baby boomer” generation
  • we alter employment level among “baby boomer” women to resemble that of women from previous generation (i.e. decrease) – results are: increased welfare dependency, decreased income (esp. for women)

– The availability of life-course assets/capital
  • we alter education level in earlier years to that in a recent year (i.e. increase) – results are: decreased welfare dependency, increased employment and income (esp. for Māori and women)
The Results
(of forward projections)

• Projection of current ‘base’ trends into the future
  - demographic ageing
  - increasing ethnic diversity
  - rising immigration
  - changing pattern of living arrangements departing from traditional norm
  - increasing secularisation
  - development of a more highly educated workforce
  - greater participation in paid employment
  - stabilising of levels of dependence on welfare benefits
  - rising average incomes
  - lower levels of deprivation
  - declining home ownership
Discussion

• Strengths
  – Dynamic microsimulation model
  – Whole of society, linked Census data
  – Inquiry tool for scenario testing – SociaLab

• Limitations
  – Census data “thin”, linkage rates incomplete
  – Counterfactual scenario tests not striking
  – Not fully delivered on analytical framework
The Future

- Include household structure
- Individual transitions through further censuses
- Use more administrative data (Integrated Data Infrastructure)
- Richer data (beyond Census)
- Deliver on analytical framework
- Open source tool and data for all (e.g. NGOs)
SociaLab: A Census-based simulation tool for policy inquiry

QUESTIONS, COMMENTS, DISCUSSION