

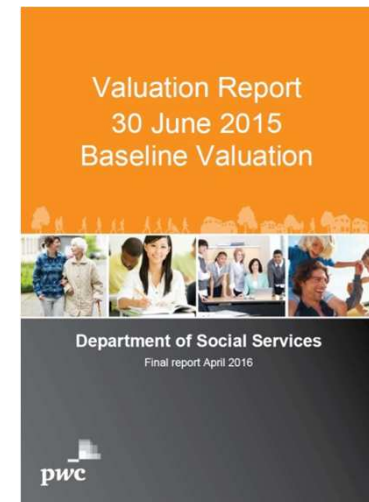
Using HILDA to understand long term changes in Australia's welfare system

*John Henstridge,
Emma Black, Donna Hill, Anna Munday*

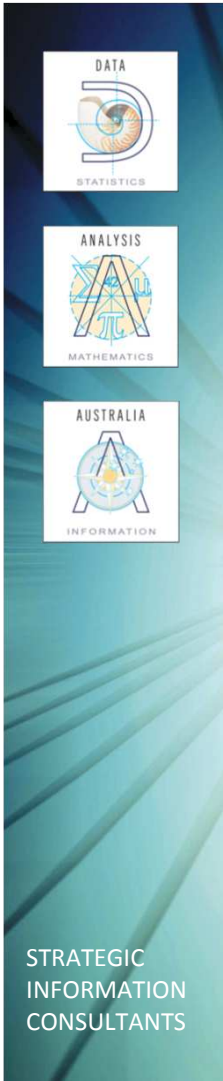
Data Analysis Australia

Background

- **Annually, over a third of Australia's budget goes to social security and welfare**
 - Almost \$160 billion for 2016-17
- **Actuarial valuation of social security and income support system**
 - Data Analysis Australia working in collaboration with PricewaterhouseCoopers

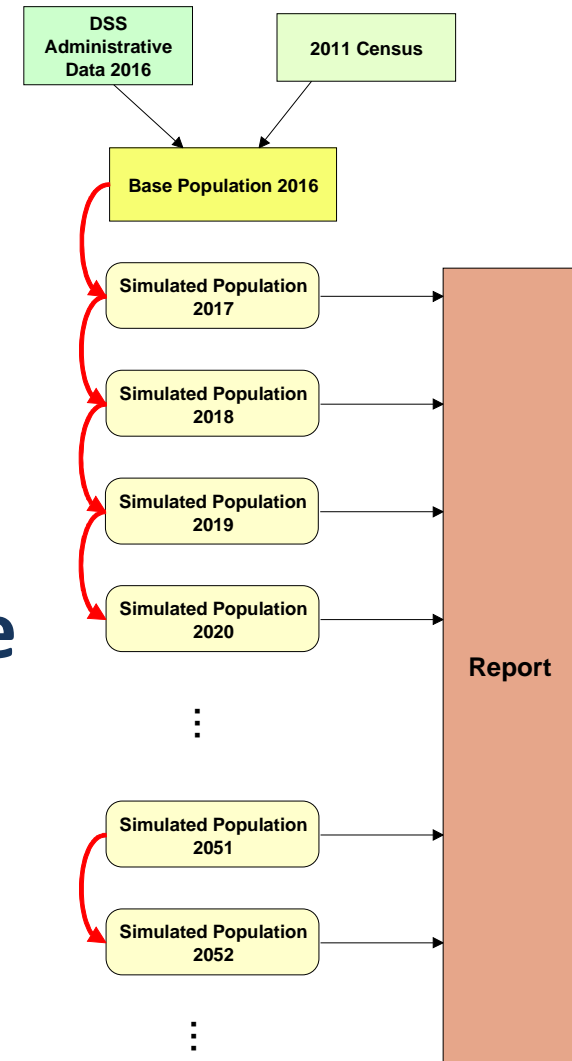


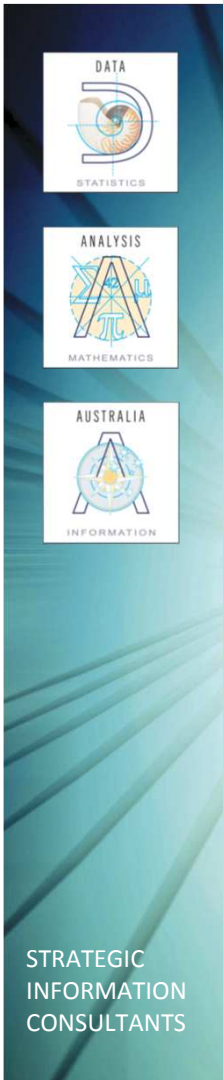
Source: <http://budget.gov.au/2016-17/content/glossies/overview/html/overview-13.htm#appb> Accessed 02/09/2016



Methodology

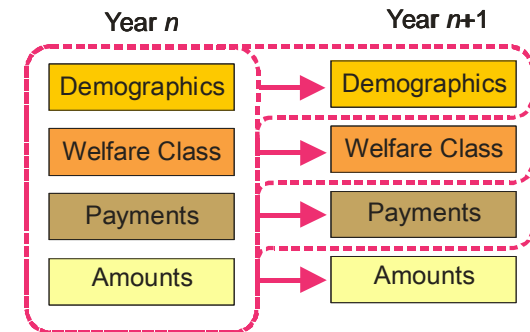
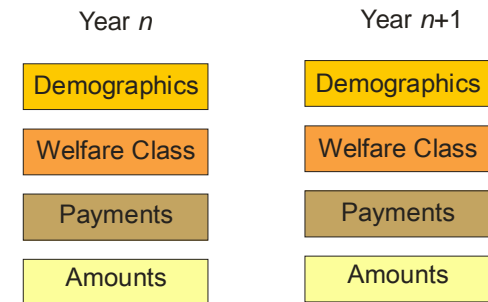
- **Start with current population**
 - All Australians
 - Including those not currently in the system
- **Simulate everyone's lifetime**
 - Including usage of social security
 - Transitions into, within and out of the system

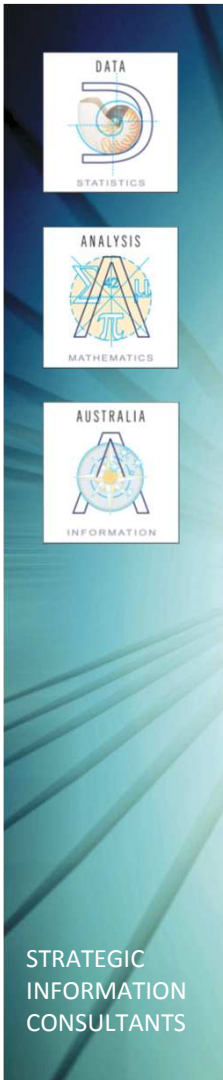




Simulation Structure

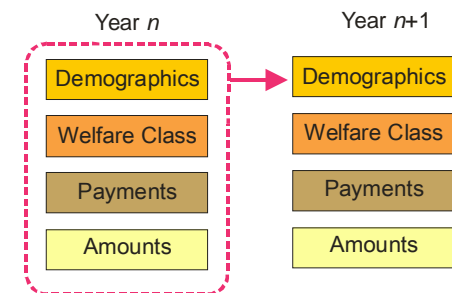
- **Four-stage process**
 - Changes in life circumstances (demographics)
 - Changes in “welfare class”
 - Which payments are received?
 - How much of each payment is received?
- **Focus on demographics**
 - Age, gender, relationships, family, education





Demographic Circumstances

- **Change over time**
 - Obviously age!
- **Social security often changes at the same time**
 - Have a baby - start to get Family Tax Benefit
 - Gain a qualification – stop getting studying payment
- **Therefore, first model demographic transitions**
 - Then model transitions in social security usage



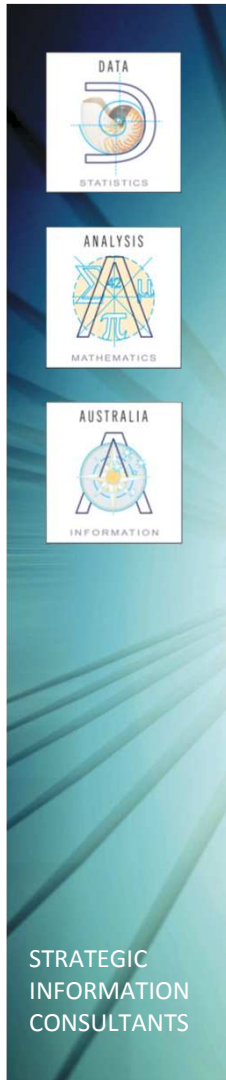


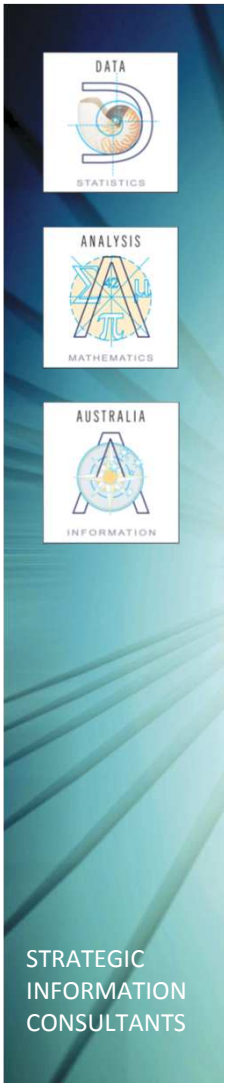
Demographic Modelling

- **Markov type models for transitions**
 - Generalised linear models to estimate transition probabilities
 - Needs longitudinal data
- **Two main data sources**
 - HILDA
 - Representative of all Australians, but a sample, 14 years' data
 - Department of Social Security administrative data
 - 100% but just those receiving benefits

Markov Model Approach

- **Next year's state predicted by only this year's state**
 - Predicted in terms of probabilities
 - State means the information you have
- **If the state contains enough information this will always work**
 - Made possible by *long term* longitudinal data

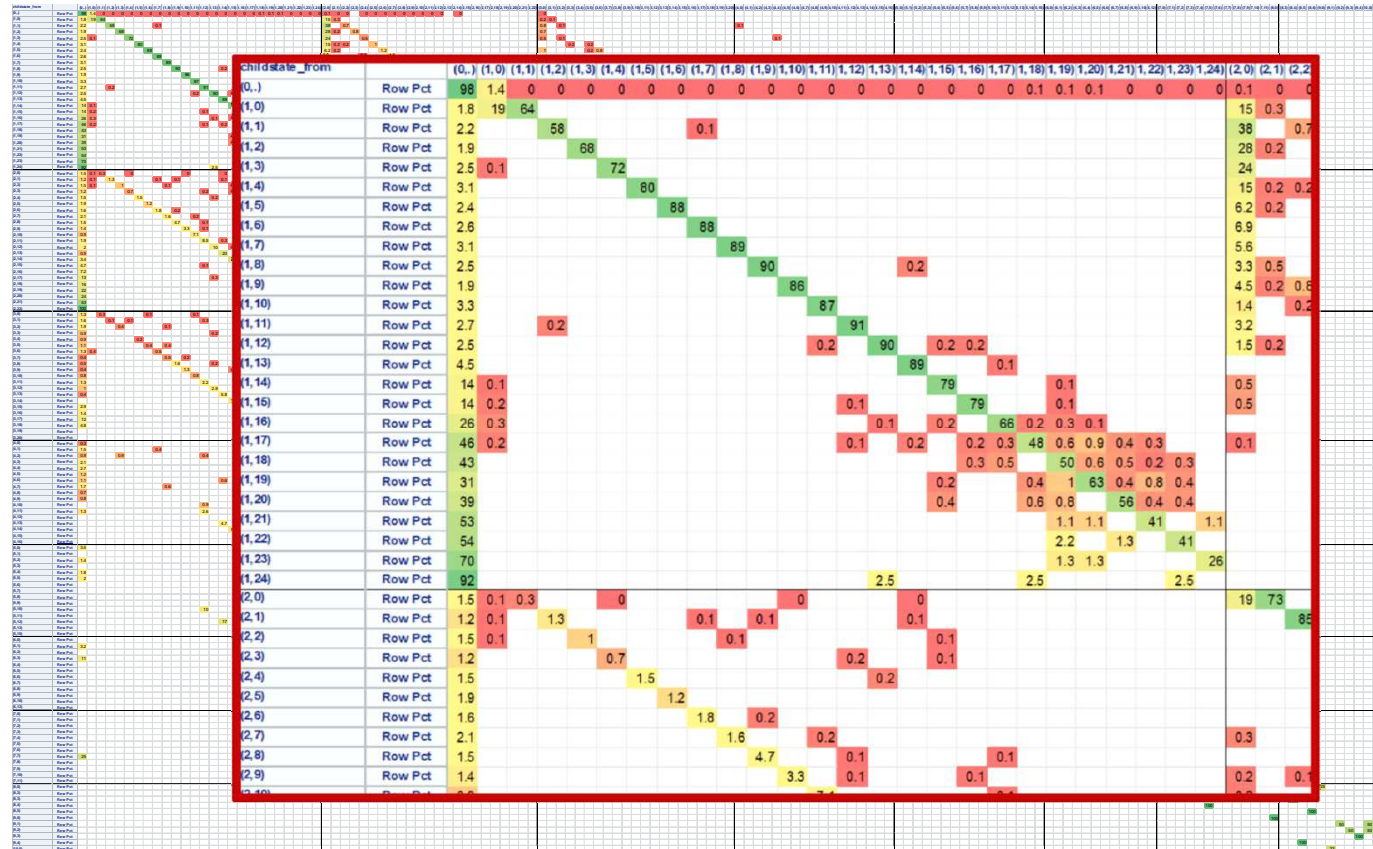


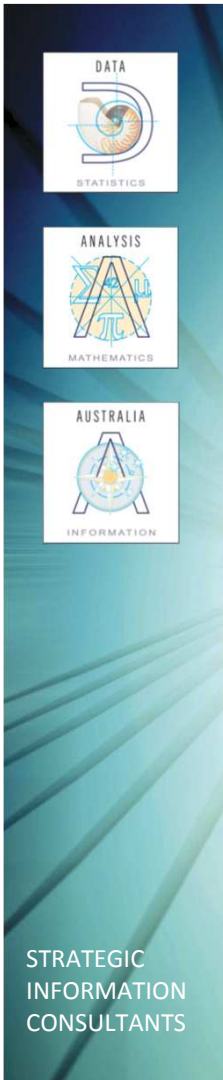


Example: Children, age of youngest

Year $n+1$

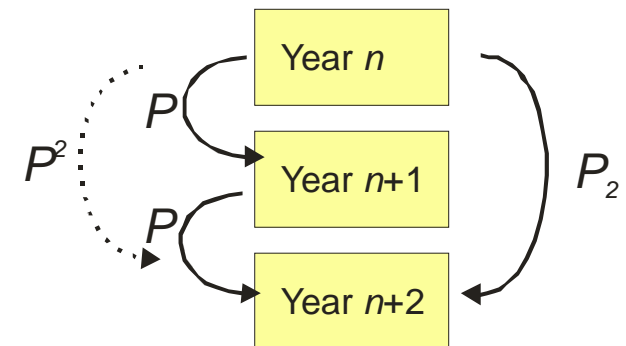
Year n





Testing the Markov structure

- Simple situation can use matrices
 - Empirical one step and two step models should be consistent
 - $P^2 = P_2$
 - Compare via eigenvalues and vectors



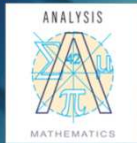
Markov Model

- **Matrix models not enough**
 - Far too many states
 - Gender \times age \times relationship status \times # children \times age of youngest \times
 - However simple gender \times age methods do provide “foundation” models
- **Multinomial models**
 - Need to parametrise – overall model has thousands of parameters
 - Largely fitted as generalised linear models



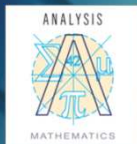
Why use HILDA?

- **DSS data only tells us about people already in the system**
 - Expect major biases
- **HILDA obvious source**
 - Representative of whole population
 - Enables inclusion of substantial history
- **But it has challenges**



Definitional Differences

- **Definitions vary between DSS and HILDA**
 - Data collected for very different reasons
- **DSS – an operational data set**
 - Records what is needed, when it is needed
 - Children not recorded until somebody claims a payment for them
 - Many people gain children at age one rather than at birth
 - But once recorded the children never leave
 - Educational achievement mainly when receiving study payments
- **HILDA – a research data set**
 - Annual cycle, updates most data regardless
 - Very good at recording children as soon as they are born
 - Also records when they leave home
 - Education constantly updated



Non-Response

- **HILDA**

- Only limited structural missingness
- Not “missing at random”
- Changes in life circumstance when respondents are most likely to be lost to HILDA (attrition)
 - Exactly what we are trying to measure
 - May underestimate rates of change

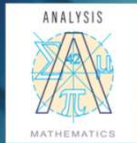
- **DSS**

- Only information required at the time is collected – structural missingness
 - E.g. Don’t need educational achievement after person off studying payment – graduation not recorded
 - Worse than that – data is how things were when last updated



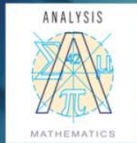
Sample Size

- **HILDA seems big**
 - ~10,000 households
 - ~25,000 individuals
 - 14 years
- **Still not big enough for the level of detail we want to include**
 - Complex models involving numerous terms
 - E.g. Age is a big driver of life circumstances, but relationship is not linear
 - Rapid changes over a few years
 - Need many age terms to capture the relationship
 - Needed to simplify models
 - Or resort to DSS data



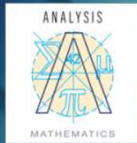
Differences Accumulate

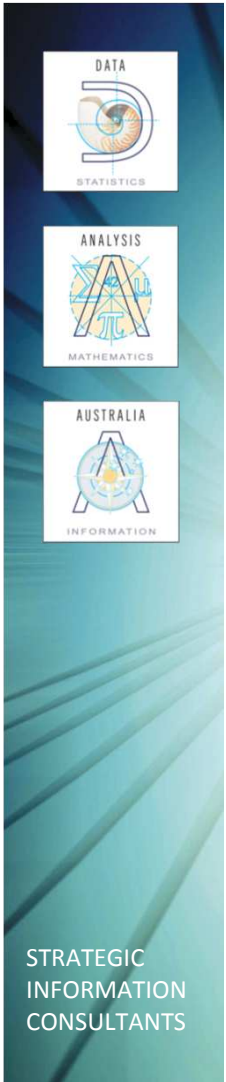
- **Simulating over many transitions, many years**
 - From current age until death
- **Small biases in transition probabilities significant**
 - A model being only slightly wrong has a big impact over a lifetime
 - Standard model selection methods may create some biases
 - Post hoc corrections necessary



Conclusions

- **HILDA unique source for some parts of the valuation model**
 - The only source of quality representative longitudinal data
- **Critical reference when considering DSS administrative data**
 - HILDA: research quality data
- ***HILDA not too big!***





Thank You

www.daa.com.au