Economic Simulations in Ada

Graham Stark
Howard Reed
Jocelyn Paine
Stuart Rossiter

University of Southampton, UK
Virtual Worlds Research, UK
Landman Economics

Outline

- Background – previous work/microsimulation
- Design of the model – data/forecasting/modelling
- Problems – data problems/modelling inadequacies
- Implementation – programming the model/databases/testing strategy/web interface
- Possible future developments
- Let's have a play

The stuff we do

When taxes or state benefits are made more or less generous:

- What kinds of people gain or lose?
- How much does it cost (or save)
- What are the economic effects? Are markets more or less distorted after the change?

It's not physics!

- Few universally agreed theories;
- Politics everywhere;
- Often inaccurate/inconsistent data

But it's important, can affect all our lives, and we do our best

Microsimulation

Most countries have some large scale economic surveys

- Belgium has EU/SILK
- UK has Family Resources Survey, Understanding Society, and others
- These survey a representative group of households, and ask them about their incomes, savings, family size, housing costs ...

We take those surveys and write programs that apply the rules of the income tax system, pension system and so on to each of the households. Summing over the households should then give you the effects we're interested in. But...

- Data can be unrepresentative/poorly recorded
- Rules might not apply – dodging taxes, not claiming benefits, administrative mistakes..
- Behaviour might change (work harder, stop buying taxed goods..)

Again, it's not physics: just be honest, fix the problems you can

Us and Ada (1)

We've been doing this since the 1980s...

- Started at the Institute for Fiscal Studies (IFS)
- Most people's background in economics (also statistics, mathematics), not programmers
- Early models were in Fortran 66
- After a particularly nasty experience crashing a model on live television (1987 BBC Budget program) I resolved to find a better way.

Us and Ada (2)

- Moved to Pascal in 1988
- Other aspects of professional development (if such we are) came later – unit testing version control;
- And we've never quite managed proper design documents – always unanticipated difficulties with the data, no comprehensive theory to guide us
- But Pascal was a huge leap forward
- Over the years, moved to Modula-2 and then back to Pascal (Delphi)

Us and Ada (3)

- Started own company in 2003.
- Agreed not to use any IFS code;
- Looked around for a new language
- Tried Java, Python, Fortran 9x, often as for-hire programmer in early days
- Soon settled on Ada and have stuck with it!

Us and Ada (4)

- The languages usually used in economics, finance and the like are good for complicated calculations on simple data – vectors, matrices – Fortran, Gauss, R, etc.
- Our models are simple calculations on complicated data
  - Households have multiple characteristics: region, tenure type, housing costs...
  - And contain families
  - Which contain people
  - Who have ages, employment status, health stage, multiple sources of income and wealth
  - Fiscal systems parameters are similar: tax allowances, rates, bands, benefit levels ages, employment statuses, etc. to which they apply
- Ada is perfect for modelling these
  - Capture the structures and the actual algorithms are usually very easy

Us and Ada (5)

- Models we've built this in Ada:
- Code available at github.com/grahamstark/
- Simulations of the Legal Aid system for Scottish, English and Irish legal aid boards;
- Affordability calculator for charities commission
- Latest: Social Care Simulations for Welsh Governments and (in development) England

Us and Ada (6)

- Ada was intended as a general-purpose language, replacing other languages all across Government, not just as a language for engineering applications
- Has facilities for reading Cobol files, modelling Decimal types, etc. - probably not used in missiles
- As well as simulations, I think it could make good sense as a language for (e.g.) finance, accounting etc.
-

Good things about Ada

- Clarity
- Records, enumerated types and classes are perfect for the complex data we use
- Some fantastic libraries (AWS, XML, Simple Components..)
- Multitasking (always nice for simulations)
- Great compiler
- Lively, very supportive community

Not so Good ..

- Later languages (Python, Modula) maybe did the whole package thing better
- String handling!!!
- Relatively few native packages (though we've found importing non-native packages very easy)
- After using Ada for 10+years, there are still bits I don't understand/can't predict the rules for. I may be alone, though..
- Heresy: I think some typing rules are too strict, especially in assignments. I prefer the rules in Pascal (and Java).

A Case Study: Our Social Care Simulations

- Last year we built an enormous (by our standard) simulation of the Welsh Social Care system
- Presently we're building a similar (but more advanced) model for England
-

Background (2)

- Main interest is in Paying for social care
- residential care typically costs approx £600-£800 per week;
- System has a simple means test based on needs, income and assets;

  Non-residential care typically £20 per hour
- Dilnot Report (for England) – proposed £35,000 lifetime care costs – being implemented at £72,000 – and a £100,000 capital limit (from £24,000);
- Costing of Dilnot used simulation model from PSSRU – approximately £2bn pa
- Separate system in Wales – care home based on simple income and capital means test; domestic care allocated by need, with some 'passporting' and a £55pw maximum payment for LA supplied services
- New legislation in Wales from 2016; see http://wales.gov.uk/docs/dhss/report/130319payingforcarereporten.pdf

Requirements

- A model to capture the likely costs and distributional effects of reforms to social care finance for up to 20 years ahead, as the Welsh Population ages;
- Concentrating on the elderly;
- General taxation and most of the state benefit system out-of-scope;
- Had to be directly usable by Welsh Government Staff.

Earlier Work

- PSSRU have a microsimulation model of England social care (Forder et al 2010)
- Used for Dilnot Report (a bit);
- Implemented in SAS; closed source;
- Data sources: BHPS (British Household Panel Survey) and ELSA (English Longitudinal Study of Ageing)
- Forecasting strategy: mainly uses data reweighting – if (say) 500,000 elderly now, and 1,000,000 forecast in 10 years, double the weight each elderly person has in the totals when simulating for 2024.
- ELSA is a big advantage since it includes the care home population – no such dataset for Wales – BHPS follows household members only.

Data - BHPS

- We use data from Understanding Society – formerly the British Household Panel Survey
- Follows a sample of British households from 1991 onwards;
- Understanding Society - larger sample from 2010 onwards;
- Contains, incomes, some simple wealth measures; lots of health and wellbeing measures; demographic transition data (deaths, changes in health, leaving households, etc.)
- Is 'oversampled' in Wales and Scotland, but sample still quite small (~2,300 households in the years 2000 - 2010)

Design – Forecasting Strategy

- We want to make a projected dataset for the next 20 years or so using use the BHPS/Understanding Society data;
- BHPS has information on demographic transitions (death, sickness, moving to care homes..). We can model the likelihood of these things happening given a person's current state (age, health, income, etc..)
- Use the resultant probabilities to age the population – add one year to everyone's age, calculate the new state for everybody given the new ages, re-arrange households to account for deaths, moving out;
- Advantages – our households have a history – we can model cumulative payments, 'transitional protection', wealth depletion, etc;
- Disadvantages (over reweighting) hard to get data to track known (or believed) demographic trends.

Forecasting - Transition Probabilities - Example

- Likelyhood of death for over 60 males - Probit regression on Wales–only BHPS data

```
Probit regression                                 Number of obs   =       1949
                                                  LR chi2(7)      =      63.54
                                                  Prob > chi2     =     0.0000
Log likelihood = -225.71496                       Pseudo R2       =     0.1234

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -2.492184   3.067242    -0.81   0.416    -8.503869      3.5195
       age_2 |    3.20898   3.976602     0.81   0.420    -4.585016    11.00298
       age_3 |  -1.344959   1.710742    -0.79   0.432    -4.697952    2.008034
      hl_exc |  -.2429036    .214984    -1.13   0.259    -.6642646    .1784574
     hl_good |    -.50586   .1661655    -3.04   0.002    -.8315383   -.1801816
     hl_poor |    .152715   .1737817     0.88   0.380    -.1878908    .4933209
    hl_vpoor |   .7425185   .2235365     3.32   0.001      .304395    1.180642
       _cons |   61.27476   78.48694     0.78   0.435    -92.55682    215.1063
------------------------------------------------------------------------------
```

Forecasting

In total we have similar estimates for:

- Hours Of Care Received
- Health(Improving/Worsening)
- Receipt of Informal Care
- Activities of Daily Life (ADL)- Base Levels and worsening/improving – ADL is a simple set of capabilities used in assessing social care needs;
- Death;
- Household Splitting
- Working/Retiring/Returning to Work
- Moving From renting to buying accommodation or back again (not used currently)
- Predicted Wealth (using Wealth and Assets Survey)
- Moving To Care Home (using ELSA)
- Receipt of various health-related contingent benefits
- Demand for Private Care

The Care Home Population

- Since the BHPS covers only people in households, and ELSA covers England only, we have no base data on the Welsh Residential Care home population;
- However, we do have probabilities of BHPS members moving to care homes (actually estimated from English ELSA data);
- We generate an initial stock of care-home residents from the BHPS by duplicating the household members with the highest modelled probability of moving to a care home until some initial population targets are met;
- Thereafter the care-home population is depleted by death probabilities for care-home residents taken from (Forder et, al 2008) and replenished by householders moved using the Probit transition probabilities.

Forecasting – differential uprating

- In addition to this, incomes, prices, rents, interest rates are likely to change;
- We use OBR annual forecasts released at Budget time;
- At least this bit is someone else's mistake..

Monte-Carlo simulation

- We can get round some of the problems with our ageing routines by using Monte-Carlo simulation – run the model multiple times but introducing different random standard normal errors into each probit for each iteration (and a similar procedure for the OLS regressions). So you get slightly different deaths, retirements, etc. on each iteration
- This works well, except:
  - Doesn't play nicely with the dynamic reweighting routines which want to force the same population targets on any sample;
  - Makes an already slow model much slower.

Forecasting – Dynamic Reweighting

- Unfortunately, this ageing strategy doesn't always work well:
  - Welsh BHPS sample is small;
  - BHPS/ELSA transitions data may well be unreliable – people leave the sample in the years before death or moving into care so our models of these things under estimate the true transition probabilities;
- We struggled for a long time to get the ageing right;
- Also, we do very little modelling of the young end of the population – marriage and cohabiting, starting work, childbirth, etc.. We need at least something there;
- Solution – supplement ageing with reweighting.
- Use external forecasts of populations (Daffodil)
- If our ageing process produces 100 households of some type and we need 200 to match some population forecast, just double the weight given to each.

Dynamic Reweighting (2)

- Actually, a bit more to it than that;
- Demographers produce forecasts for many things: populations by age; numbers needing care; numbers of households etc.;
- They may be wrong, but we should be consistent with them unless we have strong priors;
- We use a technique (Creedy 2008, Deville et. al. 1996) that allows us to produce weights that make our sample hit any number of targets exactly (we use number of households; populations by age and gender; numbers needing care, care home population);
- There are an infinite number of weights that will hit the targets; the algorithm picks the set closest (in some sense) to uniform weights;
- The routines for this are built in to the model so you can explore the impacts of different population forecasts;
- We get out target forecasts from Daffodil: http://www.daffodilcymru.org.uk/.

Design – Modelling Strategy

- Once we have our aged, uprated and reweighted sample, we can apply the rules of the social care means-tests and the benefit system to it;
- Some parts are straightforward and mechanistic (tax credits, the capital and income means-tests for residential social care);
- Currently big chunks of the tax and benefit system are simply left alone – recorded receipts and payments of income tax, NI, Housing Benefit are simply uprated and carried forward.
- Other parts – e.g. domestic help from the Local Authority are dealt with by using the generated household characteristics to rank people by need – funding then goes to the most needy until some user-selected needs threshold or budget limit is hit.

BHPS – Problems

- We've already discussed the small BHPS Wales sample, the unreliable demographic transitions data and the lack of a care-homes sample;
- Also, compared to other UK surveys such as the Family Resources Survey, information on income and wealth seems very poor. Often, recorded receipts of benefits were for amounts that weren't possible in that year; grossed up household payments of taxes and receipts of benefits look very different from official totals (this is a problem with all survey data but seems particularly acute here);
- We got round BHPS wealth data problems by imputing from Wealth and Assets Survey (WAS);
- In retrospect, a better strategy would have been to estimate all the transitions from BHPS or ELSA but apply them to FRS/WAS data.

Modelling – Problems

- modelling of benefits is relatively crude compared to that found in a conventional tax-benefit model;
- Retaining recorded receipts for direct taxes and some benefits means we get very out of step as the household ages;
- Wealth evolution is very poorly modelled;
- Costs of care are crudely handled, with single flat-rate costs;
- And more...

Implementation

- The computer model that implements all this is big:
  - 60,000 lines in 150 files in the model implementation itself;
  - Plus a similar amount in general purpose library code;
  - Plus database scripts; scripts to start and stop the web version; templates and static html pages for the web interface;
- Plus lots of 3rd party code (web server; xml handling; database interfaces, etc..)
- Model is open-source, available on GitHub:

  https://github.com/grahamstark

Implementation (2)

- The model is written in the Ada language;
  - Pascal-like language sponsored by the US Department of Defence;
  - Mainly used in missiles, H-bombs etc.; comparatively rare in application programming, outside of US Government and some banks;
  - But we like it:
    - Many years experience writing simulations in Pascal/Delphi (IFS tax benefit model);
    - Very expressive language; easy to read; hard to make common mistakes;
    - Lots of nice libraries (numerics, web, database ..)
    - Very high-quality Open Source compiler (libre.gnat.org)
    - Very good, active community (comp.lang.ada)
- Learn more:http://university.adacore.com/
- Regressions were done in Stata;
- Postgres database;
- Some scripts in Ruby, Bash and Python.

Database

- The original design didn't use a database to hold data, results, or anything else – everything was in simple flat files;
- Adding Monte Carlo simulations made relying on flat files for everything too complicated;
- We use a Postgres database;
- Some custom PSQL scripts produce summary statistics for multi-iteration runs (10th, 90th percentiles, etc.)
- Database implementation is poor and very slow – much faster version is under development.

Test-first

- Most of the model was developed test-first;
  - Write a Unit Test for each function we wanted the model to have – specify behaviour in standard and edge-cases;
  - Then write enough actual code to make all the tests pass;
  - Then you're done;
  - As the model passed through many iterations, the test suite became less comprehensive and is now sometimes wrong – needs major clean up.

Web Interface

- A requirement was that the model be fully usable by Welsh Government Officials without our intervention;
- So we mounted it on a web server and built a web interface for it;
- We use Ada Web Server for this;
- The result is somewhat unpolished but really quite powerful:
  - All the model parameters are fully controllable; changes can be made in bulk for multiple years (e.g. set the pension age to 70 from 2015 onwards);
  - You can drill down from aggregate results to individual BHPS households, which can be tracked through time;
  - The model has its own batch queuing system, using Ada's tasking facilities;
- There's now almost as much interface code as actual modelling code.

Future Developments

- There is currently a limited amount of development work going on;
  - I'm rewriting (unfunded!) the database code to address the chronic speed problems the database introduced. That's close to completion;
  - Some paid maintenance and deployment work is happening – up-to-date forecasts from OBR and Daffodil, and updated parameters will be added in the next couple of weeks;
  - The model has recently been deployed on the Cloud (Microsoft Azure);
  - Much of the documentation has recently been updated – thanks Stuart! and a paper for the Journal of Microsimulation is in the works (just...)

Future Developments - Data

- Having come this far I feel this model has a lot of potential
- I'd like to use better datasets – the Family Resources Survey, ELSA, the Wealth and Assets and the Living Costs and Food Survey. (Possibly alter the model so it can switch between datasets without reprogramming);
- Transition estimation would probably still be based on BHPS/Understanding Society/ELSA, though;
- Extend the ageing routines and estimates so they can model more of the young end of the population births, marriages, etc. Perhaps also immigration.. I'm unclear how to do any of this, though;
- Since much of social care and some of social security is devolved locally, I'd like to adapt the model to do simulations for a single Region or Local authority. We already have a prototype of this: http://projectsvr.virtual-worlds-research.com/la/

Future Develpments - modelling

- Add a complete set of conventional tax benefit model routines – income tax, NI, VAT, all tax credits, housing benefits, etc..
- Model pensions;
- Model Piketty-style wealth and asset accumulation;
- And much else, including improving and extending the existing routines.

Future Develoments - usability

- I'd love to make this model, or something very like it, available to the general public;
- I believe funding could be available for this;
- Here's one we did earlier: http://www.bized.co.uk/virtual/economy/index.htm

The Model In action

- Plan A – in the cloud:
  http://wgcaremodel.cloudapp.net/wsc2/
- Plan B – some screenshots:

  http://virtual-worlds.biz/demonstrations/wsc/

References:

CREEDY, John. Survey Reweighting for Tax Microsimulation Modelling. September 2003. Available : www.treasury.govt.nz/publications/research-policy/wp/2003/03-17/twp03-17.pdf

FORDER, Julian and FERNANDEZ, Jon-Luis. Length of Stay in Care Homes. January 2011. Available :

http://www.pssru.ac.uk/publication-details.php?id=3211

DEVILLE, Jean-Claude and SARNDAL, Carl-Erik. Calibration Estimators in Survey Sampling. 1992.

http://www.jstor.org/stable/2290268