Benefit Reform – a Dynamic Microsimulation Approach Using Administrative Data

Simon C Gault
Model Development Unit, Strategy Directorate, Department for Work and Pensions, Caxton House, Tothill Street, London SW1H 9NA,United Kingdom
Email: Simon.gault@dwp.gsi.gov.uk

Thomas Jackson
Model Development Unit, Strategy Directorate, Department for Work and Pensions, Caxton House, Tothill Street, London SW1H 9NA,United Kingdom
Email: thomas.jackson@dwp.gsi.gov.uk

Abstract :

INFORM is a dynamic microsimulation model of future benefit receipt based entirely on administrative data. We take a recent sample of Department for Work and Pensions (DWP) benefits data which represents the current population in receipt of one or more DWP benefits and then simulate future changes in entitlement including movement between or within benefits, spells without benefit and completely new entitlements. The model was initially developed for forecasting of benefit caseloads and combinations of receipt and must incorporate significant benefit reforms planned over the coming years.

This paper outlines the development and methodology behind INFORM and looks in particular at how past administrative benefits data has been used to simulate significant changes to the same benefits in the future as well as the introduction of new benefits. While this has been a necessity in order to produce a satisfactory forecasting model covering known reforms, it has also established the potential of the model for contributing to analysis of policy options in the future. The paper examines the problems of simulating major reforms given the limited scope of administrative data and presents particular examples of the solutions found, highlighting the importance of benefit history and the interactions between receipt of particular benefits which is one of the key functions of the model.

1 Overview

This paper describes INFORM[1], a dynamic microsimulation model used for benefit caseload forecasting that simulates basic demographic events and transitions for current and recent recipients of nine working age benefits administered by the UK Department for Work and Pensions (DWP). Section 2 provides a summary description of the model, the microdata on which it relies and the modelling approach used. It also looks at the strengths and limitations of the model and future potential for expansion and improvement.

Sections 3 and 4 then examines in more detail how the model developers and users have attempted to overcome the limitations of the model when faced with significant change – the examples used being a major policy change and the impact of the recent economic downturn. Section 5 concludes. Appendices provide some additional technical detail on the generic modelling architecture used (Genesis), as well as specific techniques used for INFORM.

---

[1] **IN**tegrated **FOR**ecasting **M**odel

2 Description of INFORM

2.1 Functionality and purpose

The primary purpose of INFORM is to produce medium term (four to six year) forecasts of caseloads of the included benefits and of benefit combinations. It does not attempt to model financial amounts for fixed rate or income related benefits, relying instead on external “average amount” models although where possible it does flag receipt of (and transition between) different components, additions etc.

The simulations are carried out on a discrete monthly basis for each individual in the base microdata file which is a sample of those currently in receipt of one or more of the benefits in addition to those who have been in receipt within the past five years. For each month INFORM models onflow and exit for each benefit as well as internal transitions within a benefit. In any given month a certain number of cases will exit the model (through reaching state pension age, dying, or being “benefit free” for five years) and a separate group will enter the model as individuals new to benefit receipt (at least in the past five years). Table 1 below lists the benefits included in INFORM.

TABLE 1 - WORKING AGE BENEFITS MODELLED IN INFORM
Jobseeker’s allowance (JSA) is for people who are unemployed or working less than 16 hours a week and who are available, and actively looking, for work.
Incapacity Benefit (IB) is for people unable to work because of illness or disability, usually requiring sufficient national insurance contributions to qualify.
Employment and support allowance (ESA) replaced IB for new cases from October 2008. ESA also replaced income support paid on the grounds of incapacity.
Disability living allowance (DLA) is a benefit for adults and children with disabilities. It is for people who need help looking after themselves and those who find it difficult to get around. DLA is not means tested and paid on top of other entitlements in most cases. Different components and rates are paid for help with personal care and for mobility.
Income support (IS) is a means-tested or income-related benefit intended to provide for basic living expenses. It can be paid on its own or top up other benefits. IS is for people who are not required to sign on for work through ill-health or disability (until October 2008 when IS was replaced by means tested ESA) or people who are carers or lone parents.
Carer’s allowance (CA) is a benefit for people who regularly spend at least 35 hours a week caring for a severely disabled person.
Pension Credit (PC) is a means-tested benefit for people aged 60 or over (although INFORM is a working age model PC is included for male claimants aged 60-64)
Bereavement allowance, **widowed parent’s allowance** and **widow’s benefit** (pre 2001 entitlement) are all bereavement related benefits for those of working age.

2.2 Model overview

Figure 1 shows the basic process flow over a monthly simulation within INFORM. Mortality is the initial process and then surviving cases are aged with some exited if they are no longer classed as ‘Working age’. Remaining cases then go through a sequence of benefit specific processes, determining the status of each individual for each benefit for that month (see 2.5 below). A check is then made for individuals who have not received any benefit for five years (such cases exit the model). The five year limit was based on the limited historic scope of the data. Finally a process is run to append a cohort of completely new benefit recipients to the microdata, using pre-determined counts (see 2.6 below). The process then repeats.

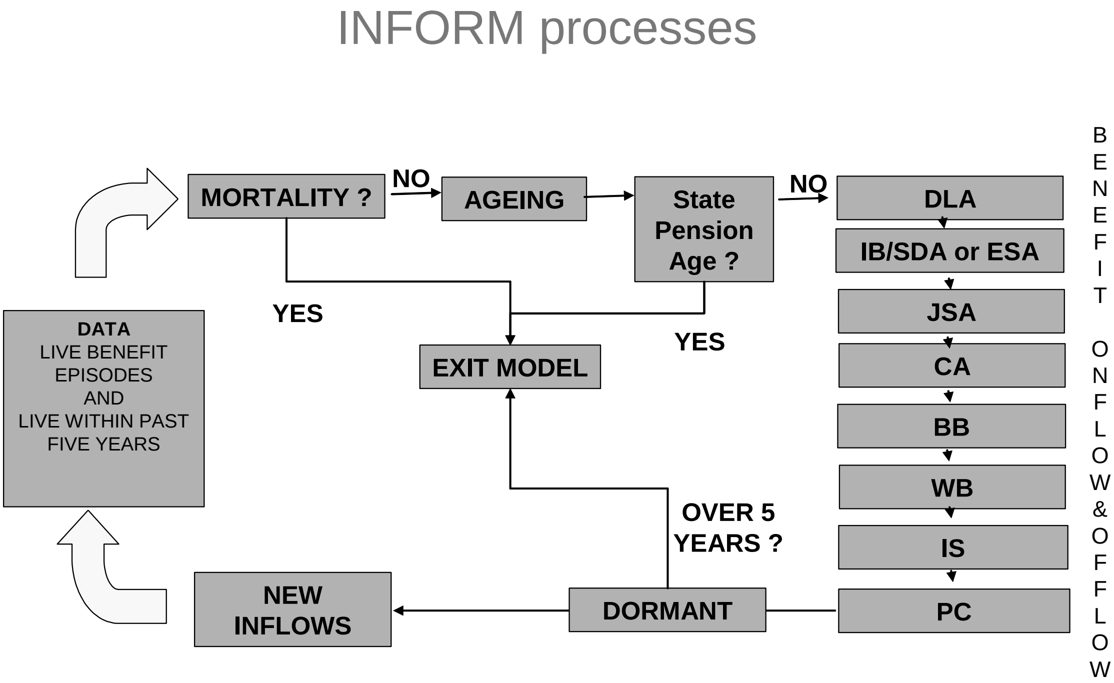


Figure 1

2.3 Model data

The model basedata are derived from DWP administrative data covering the working age population in receipt of benefits back to 1999, with data being available with around an eight month lag. These data are based on a series of scans of administrative systems (variable by benefit but generally four to six weekly) which are linked by a unique identifier for each individual providing a longitudinal database of benefit recipients. The data are made available on a quarterly basis but for the purposes of INFORM a 5% sample is taken and manipulated to show monthly transitions. This monthly dataset is then used as the source for the analysis underpinning the various transition equations used in the model as well as for the basedata used as the model starting point. For speed purposes the actual model runs currently use only a 1% sample of around 110,000 individuals (of which around 55,000 are initially in receipt of one or more benefits).

2.4 Architecture

INFORM is written using the Genesis system, developed within DWP, for creating Discrete Dynamic Microsimulation Models (DDMS). This system allows the developer and user a standardised Excel based interface to all model processes and actions although Genesis itself is written in SAS.

Genesis has been used as the basis for a variety of DDMS models within DWP and it allows rapid development of working versions of models which can then be improved as requirements and resources allow. It encourages a modular approach to development where the model is broken down into major processes and then individual actions within those processes. Initial, simple versions of actions (for example a simple probability matrix for onflow to a benefit) can then easily be replaced at a later point by a more complex approach (for example a multivariate model). Most discrete events are simulated stochastically using *Monte Carlo* techniques and so in the default case repeat runs using the same basedata, equations and assumptions will produce different results. However optional functionality is available to ensure that the stream of random numbers used on two separate runs of the model is identical. The Genesis architecture also has built in functionality to allow alignment to target populations or rates. Appendix I provides additional background on Genesis.

2.5 Modelling benefit receipt

After extracting the 5% sample from the published administrative datasets, adjusted datasets are derived that standardise the basic approach to modelling benefit receipt so every individual for every benefit and every period (month) must be in one of six possible states as shown in Figure 2.

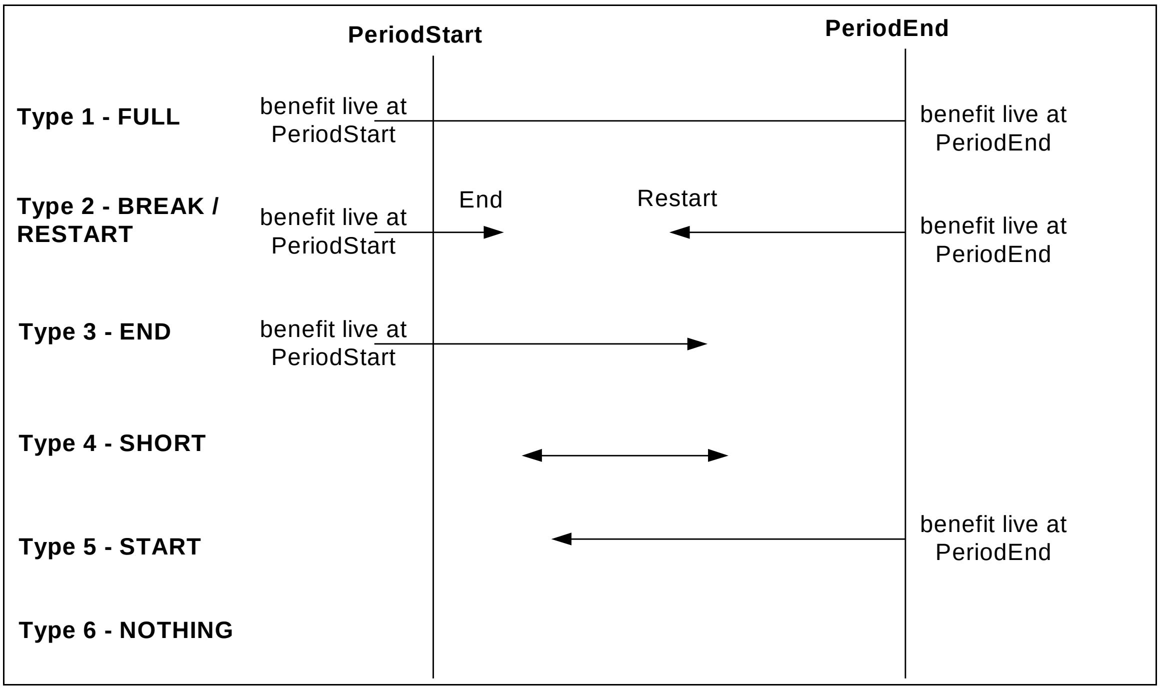


Figure 2

The same approach is then followed for the simulation period, except that for simplicity states 2 and 4 are only simulated for JSA. This allows the same basic structure for each benefit module as a series of event decisions as shown below. For most benefits there is a sequence of two binary choices as in Figure 3, which establish the transition type for the given individual, benefit and month.

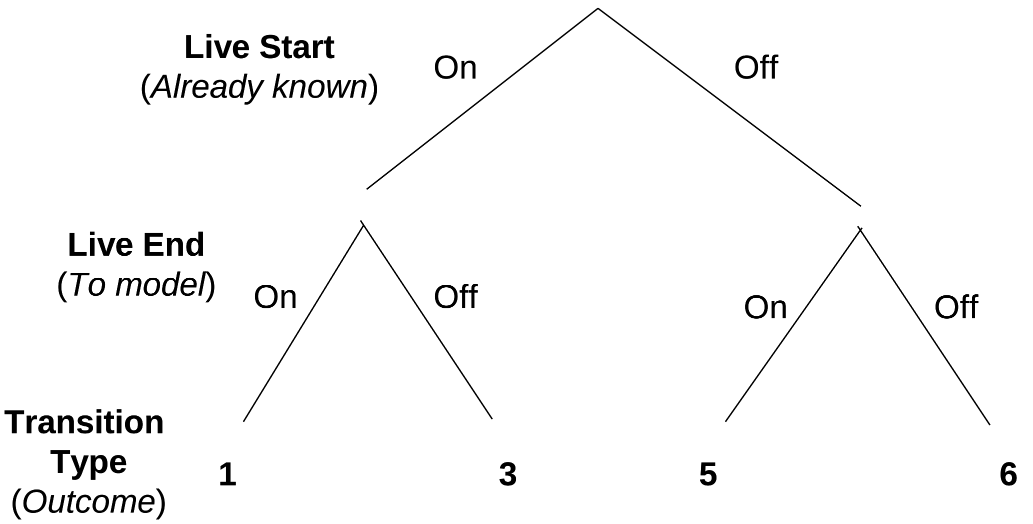


Figure 3

For JSA additional binary events are modelled to allow an individual to onflow and exit in the same month (states 2 and 4) as shown in Figure 4.

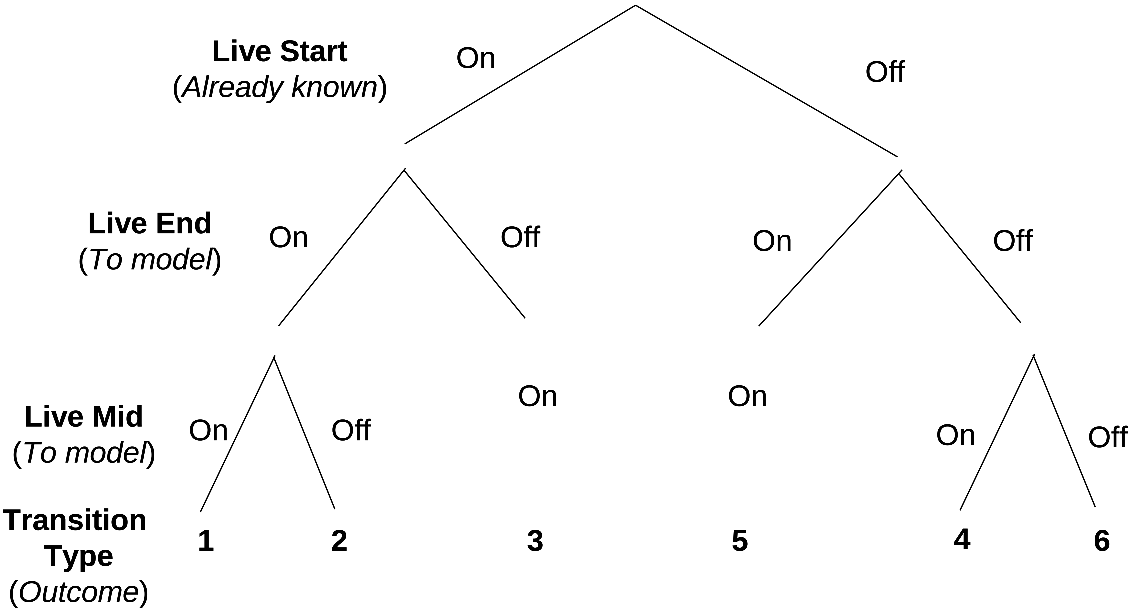


Figure 4

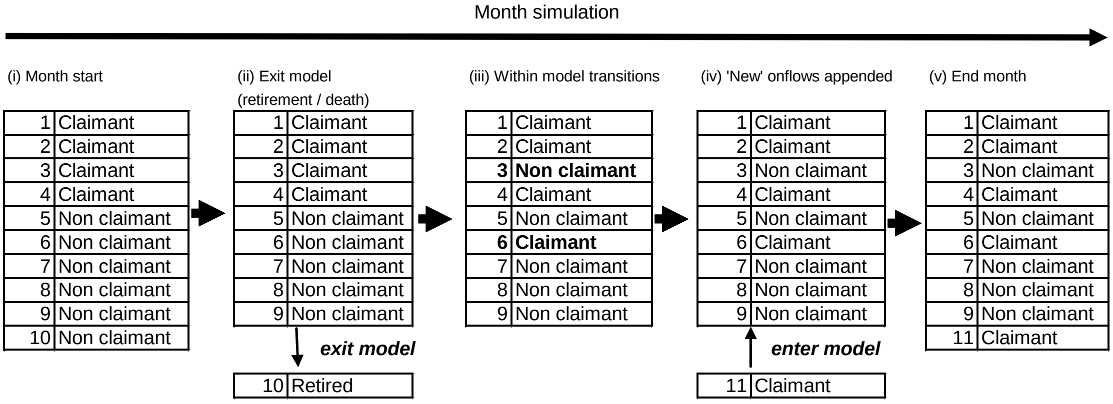
Further events such as transitions between benefit rates and entitlement to premia are also modelled depending on the benefit and individual.

The events are determined with reference to basic demographic and benefit related characteristics of the individual in the current and previous months – one of the key gains relative to previous single benefit models being the ability to model benefit interaction effects and to see individuals as multiple benefit recipients. Appendix III provides an outline of the key events modelled and the sort of variables used.

2.6 Changes to model population

The model population at the end of each simulated month consists of those working age individuals currently in receipt of one or more benefits as well as those who have been in receipt at some point in the past five years. In each month a group of individuals will leave the model (or more accurately cease to be simulated further) – this will be due to death, reaching state retirement age or not having received a benefit for five years.

Because the model dataset represents only those currently or recently on benefit it is also necessary to simulate the addition of a new set of individuals who represent those completely new[2] to benefit receipt in that month. These cases are actually appended to the model dataset each month by ‘cloning’ from a separate dataset of representative cases. It is a current requirement that the numbers (and splits by characteristic such as benefit combination, gender, age etc) required each month are pre-determined before a particular run of the model. The dataset of potential ‘clones’ has count variables showing how many times each is to be re-used in each future month.



(i) Month start

1	Claimant
2	Claimant
3	Claimant
4	Claimant
5	Non claimant
6	Non claimant
7	Non claimant
8	Non claimant
9	Non claimant
10	Non claimant

(ii) Exit model (retirement / death)

1	Claimant
2	Claimant
3	Claimant
4	Claimant
5	Non claimant
6	Non claimant
7	Non claimant
8	Non claimant
9	Non claimant
10	Retired

(iii) Within model transitions

1	Claimant
2	Claimant
3	**Non claimant**
4	Claimant
5	Non claimant
6	**Claimant**
7	Non claimant
8	Non claimant
9	Non claimant

(iv) 'New' onflows appended

1	Claimant
2	Claimant
3	Non claimant
4	Claimant
5	Non claimant
6	Claimant
7	Non claimant
8	Non claimant
9	Non claimant
11	Claimant

(v) End month

1	Claimant
2	Claimant
3	Non claimant
4	Claimant
5	Non claimant
6	Claimant
7	Non claimant
8	Non claimant
9	Non claimant
11	Claimant

Figure 5

Figure 5 illustrates the process using a simplified model with at stage (i) an initial population of ten cases and a single benefit which four of the ten are claiming. Early in the monthly simulation (stage (ii)) some cases will exit the model – illustrated here by case 10 ceasing to be working age. At stage (iii) various transitions are modelled and the example shows case 3 off-flowing from the benefit and case 6 onflowing. At stage (iv) the ‘new’ onflows are appended – illustrated here by case 11 being added to the model population. Finally at stage (v) the month is complete. This illustrates an important but artificial distinction each month between those who onflow to a benefit from within the existing model population (such as case 6 via transition equations) and those who onflow as a completely new case (such as case 11 as a ‘new’ onflow appended to model data) and whose numbers are pre-determined by an external assumption.

It is hoped that a future enhancement to the model will allow the process of determining the number of ‘new’ onflows required to be part of the simulation process.

2.7 Use of Alignment

Onflows alignment

---

[2] In this context ‘completely new’ means no receipt within the past five years

The benefit forecasting analysts who commissioned the model stipulated the need to have (optional) control over the total number of onflows to a given benefit in a given month. This allows them to ensure that a given model run is consistent with specific onflow assumptions they will have negotiated with a variety of stakeholders. Given that onflows to a given benefit are a combination of within model transitions and appending of new cases this requirement is met through a combination of the forecasts of new onflows (see above) and use of the alignment functionality built into INFORM via the Genesis architecture. Given a requirement of a total number of onflows to a benefit in a month, subtracting the pre-determined new onflows will give the number of 'within model' onflows expected. Combining this with a calculation of the model population at risk of onflow then provides a required onflow rate which can be aligned to.
This approach is used for all the benefits except Job Seekers Allowance (JSA) where user requirements mean that alignment must be to a forecast caseload and not onflows. Thus for JSA a within model calculation is made of the onflow rate required to hit a caseload target after allowing for new onflows and exits. Of course use of alignment in such a way, while required for some forecast runs, also risks imposing unrealistic restrictions on the model while also limiting the capability to properly reflect interaction effects between benefits – for example on onflows. For this reason more effort is now being made to realistically model without alignment, or with alignment restricted to certain benefits. This is discussed more in section 4.1 .

Other alignment
Alignment is also used in areas of the model other than benefit onflows, for example in modelling planned policy changes and again this is explored in more detail later in the paper, particularly in section 3.5.

2.8 Model strengths

INFORM is primarily used for caseload and expenditure forecasts submitted to HM Treasury for budgetary purposes and this makes recent administrative data an appropriate data source. Even the 1% sample used for the simulation still allows reliable breakdowns for relatively small groups and the availability back to 1999 allows the build up of a reasonable longitudinal database from which to derive transition equations.

Prior to INFORM the forecasting models used have been almost exclusively for single benefits - largely cell based models developed in Excel but gradually being replaced by microsimulation models using the Genesis architecture. The key drivers behind the development of INFORM also illustrate its strengths:

- To allow impacts of policy changes affecting multiple benefits to be forecast within a single model as opposed to using several models plus multiple off model adjustments;
- resource savings from integrating several forecasting models into one;
- increased flexibility of outputs;
- availability of reliable data providing the longitudinal multiple benefit microdata source required ;

- increased focus on individuals as recipients of multiple Government services / benefits rather than a ‘silo’ approach.

These drivers still apply and the development process has yielded further benefits. In particular the datasets derived for the models microdata and transition analysis have become a general resource for analysing trends in benefit receipt while the analysis used to derive various equations and assumptions has added significantly to the general knowledge base as well as being central to the model.

2.9 Model limitations

While benefiting from the strengths of administrative data INFORM also suffers from the inherent weaknesses, particularly the lack of wider contextual and socio-demographic detail covering individuals present and past circumstances. There is no information on employment history, educational attainment, income or wealth and only indirect coverage of disability and sickness (via specific benefits) and family status. While this clearly limits the explanatory power of the model equations used for benefit transitions it has still been possible to incorporate useful benefit interaction effects as well as the impact of an individuals benefit history. While there are no plans to dramatically increase the detail available for individuals (for example by merging with survey data) there is scope for merging on additional administrative data (see 2.10).

INFORM has no macroeconomic linkages built in and so is reliant on externally imposed assumptions for new onflows as well as alignment to incorporate significant anticipated changes – the current economic downturn being a key example explored more fully in section 4. As with any model of this type there is an assumption that models of transitions such as benefit onflow and outflow estimated on a given sample of historic data remain valid into the future and again major shocks such as recession will call this into question. Similarly while the scope of the data limits the extent to which planned policy changes can be modelled at a detailed level it is possible to incorporate estimated impacts from other models – for example as described for modelling the new ESA benefit in section 3.

2.10 Future development

The benefits microdata used for INFORM is part of a wider database which also includes a variety of other administrative sources, linked by a unique identifier, from both DWP and other government departments. This gives the potential to add additional data about individuals covering areas such as national insurance history, employment and involvement with welfare programmes. As well as increasing the richness of data for modelling benefit receipt for forecasting purposes this may also extend the usefulness of INFORM into other policy areas. Projects are already underway to project future performance against Public Service Agreement employment targets relating to the amount of time individuals spend off working age benefits after exiting and also to integrate tax credit datasets which will add detail about in work benefits and lone parent status.

Possible future projects include extending coverage to pensioner benefits, adding data on employment and destinations when exiting benefit and incorporating additional data on employment history and national insurance contributions. In terms of functionality the model would also benefit from incorporating the calculations of ‘new’ onflows required each month (see 2.6) into the simulation process rather than being pre-determined.

Significant improvements to the timeliness and frequency of administrative data scans are also planned which should make the analysis and the model much more responsive to recent developments.

3 Modelling policy change

3.1 Introduction

This section describes how a major policy change – the introduction of a new ill health / disability benefit – was incorporated into INFORM, showing the different techniques used to simulate a completely new benefit in a model based entirely on administrative data

3.2 The new benefit

Employment and support Allowance (ESA) is the new benefit paid to people whose ability to work is limited by ill health or disability and from October 2008 it replaced both incapacity benefit (IB) and income support (IS) claimed on the grounds of incapacity for new customers. ESA retains features of both those benefits, having two elements: contributory[3] ESA (which is similar to IB) and income-related ESA (which is similar to IS paid on the grounds of incapacity). However ESA also has some key differences. Firstly it is a single benefit and so should be simpler to administer. Secondly ESA is not paid simply because you are found to be incapable of work. The previous strict divide between those capable and not capable of work was seen as not reflecting reality and as a potential disincentive for those wanting to move from benefits to work. Consequently ESA is paid when you have a 'limited capability for work' and claimants are divided into two groups : the 'support group' and the 'work-related activity group'. The group then determines the amount of benefit and the responsibilities a claimant will need to meet. New ESA claimants enter a 13-week 'assessment phase, during which a lower rate of benefit is paid and a 'work capability assessment' (WCA) is undergone which determines both entitlement to ESA as well as whether the 'support' or 'work-related' group is appropriate.

3.3 Modelling issues

The central issue for INFORM development was how to reliably incorporate and forecast a benefit which did not yet exist in a model based purely on historical administrative data, the key issues being how best to simulate:

- onflows to ESA;
- exits from ESA;
- distinguishing contributory and income based cases;
- allocating to the 'support' and 'work-related' groups;
- interactions of ESA receipt with other benefits.

The following sections below examine some of these issues and the approaches used in more detail.

---

[3] based on previous national insurance contributions and not means tested

3.4 Onflows

While ESA did have distinct differences from the benefits it replaced it was still similar enough that it was considered reasonable to use as a starting point assumptions and data relating to historic receipt of IB and IS. For example completely new onflows to ESA were simulated by cloning cases who had previously flowed onto IB and IS (incapacity related). Similarly logit equations used for within model onflow and exit to/from IB and IS were used as the basis for the equivalent actions for ESA, although the exit models had to be re-estimated, omitting variables unique to the previous regime. In addition entitlement to IB was to be maintained for existing cases as well as for cases with recent live claims (known as ‘linked’ cases).

3.5 Exits

The WCA used to allocate cases to ‘support’ or ‘work-related activity’ is more complex than the previous personal capability assessment used for IB and was expected to lead to overall higher exit rates than IB, biased towards the less disabled. In addition a pre-existing policy to aid Incapacity Benefit and (once introduced) Employment and Support Allowance recipients in returning to work was being expanded from initial pilot areas to all recipients. This ‘Pathways to Work’ policy includes mandatory work-focused interviews, a range of programmes to help the customer in preparing for work, and a return to work credit and the assumed impact (based on the estimated effect from pilot areas) was for an increase in exit rates during the first six months, with slightly reduced exit rates at longer durations. These assumptions were available only as changes to the exit rates banded by duration, and to create absolute alignment rates would have been complicated by the fact that the INFORM exits regressions have different coverage – for example modelling exit due to mortality and pension age separately. The INFORM equations also took account of differential exit by severity of disability.

This example is one of a number of occasions where a policy change was introduced, in which the estimates of the impact were available only as changes to the transition rates (i.e. the proportions of the pool of potential transitions that actually undergo the transition) rather than as the absolute transition rates themselves. These situations generally occur where the impact assessment of a policy change is made using a cell-based model and overall alignment rates are either unavailable or unsuitable. Rather than re-estimating all policy impacts in a format directly applicable to INFORM, a methodology was developed for aligning the change in the transition rate, rather than the actual transition rate, as a simple and practicable option. This ‘delta-alignment’ approach is described in more detail in Appendix II. As well as being able to incorporate the external estimates of changes in exit rates it was also able to retain the benefits of unaligned exit equations being able to pick up changes driven by trends in explanatory variables. Figure 6 and Figure 7 show the impact of the alignment change on the forecast caseloads and flows.

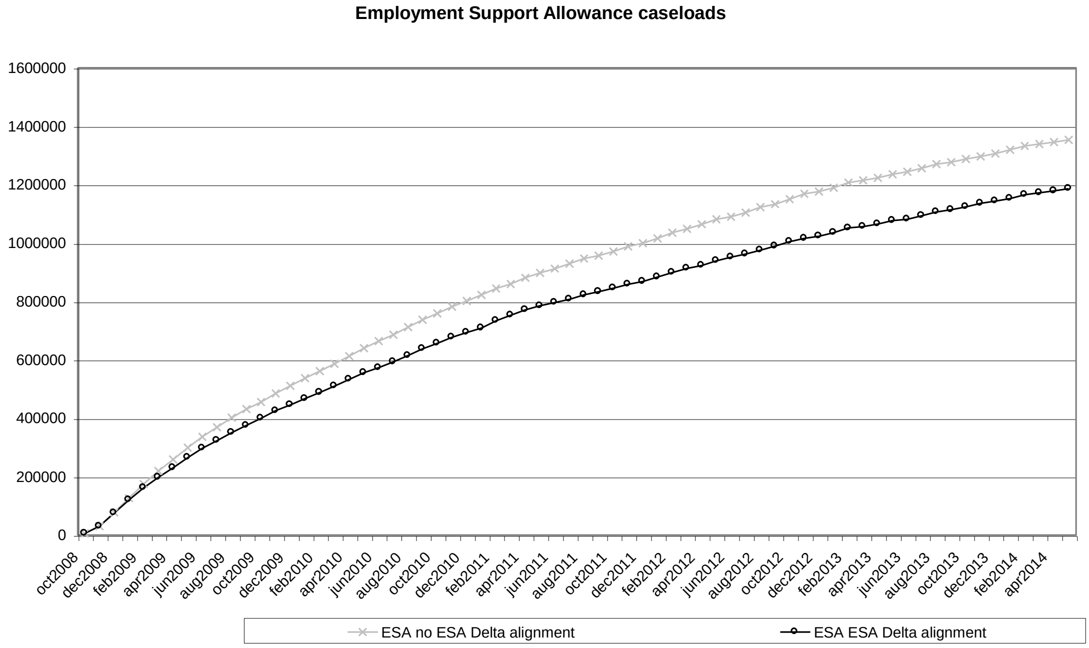


Figure 6

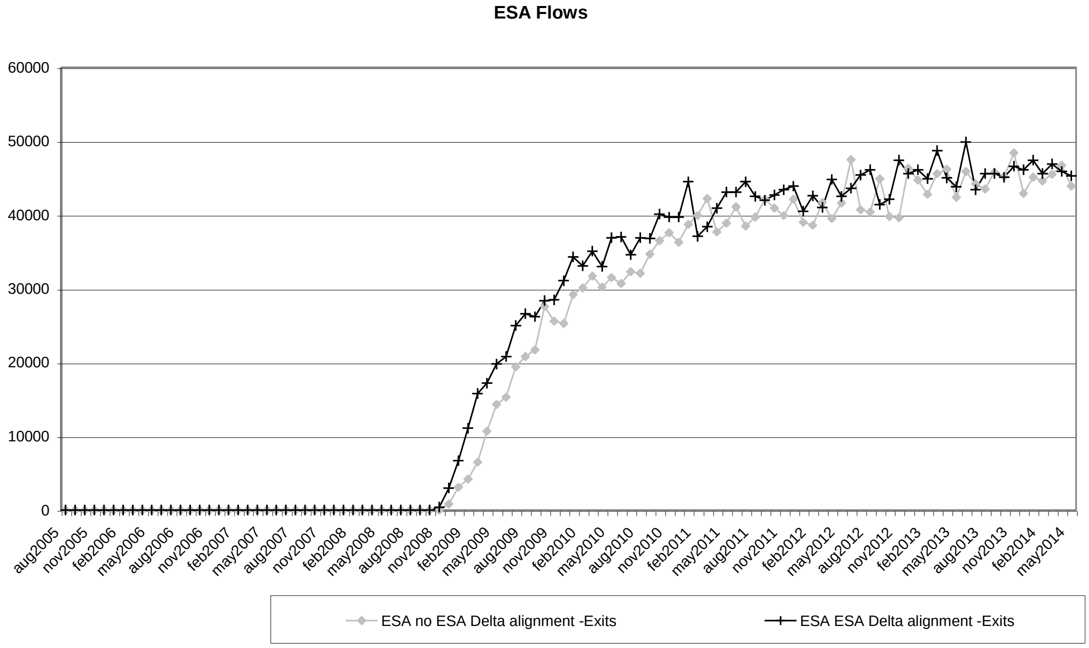


Figure 7

3.6 ‘Ghost of IS’

Under the pre ESA regime IB cases with insufficient contributory records would be entitled to income related IS, leading to a large number of cases entitled to both IB and IS. Those entitled to IS who then became entitled to IB would simply move between statistical groups within the IS caseload. Similarly a case who lost IB but retained entitlement to IS for some other reason would change group. However under the new system it is not possible to receive both ESA and IS, which introduced a more significant modelling challenge – essentially how to continue to utilise the characteristics and transition equations (for example interactions with other benefits) modelled under the IS based system in the new ESA regime.

The various transition equations used within INFORM rely on the implicit assumption that the relations between the dependent and independent variables remain valid throughout the forecast period. Since IS receipt was used as a variable in several of those equations it was necessary to ensure that ESA cases with the same characteristics (those who would have been on IS under the previous system) behaved and impacted in the model in a similar way. The solution was to introduce a new ‘ghost of IS’ flag which would be set to represent cases on ESA who would previously have been on IB and IS (those with insufficient contributions or alternative income sources). This meant that a case on income based ESA would be treated exactly the same as a IB/IS case when appearing in a transition equation – to the point of assuming that the coefficient that would have applied to IB/IS receipt now applied to ‘ghost of IS’. For onflows from outside the model the new flag was simply set for those cloned cases who had previously onflowed to IB and IS at the same time. Cases onflowing to ESA from within the model population would have the same probability of having the ‘ghost of IS’ flag set as the equivalent historic IB cases had of flowing onto IS.

Similarly cases with the ‘ghost of IS’ that leave ESA are given a probability of then onflowing to IS, to represent cases who would have been observed historically as simply moving between groups in the overall IS caseload.

The flow diagrams below illustrate the process of modelling onflow to IB and IS before the reform and to ESA after (note both apply to onflows from within the model population).

Pre-ESA: Not live on IB or IS at start

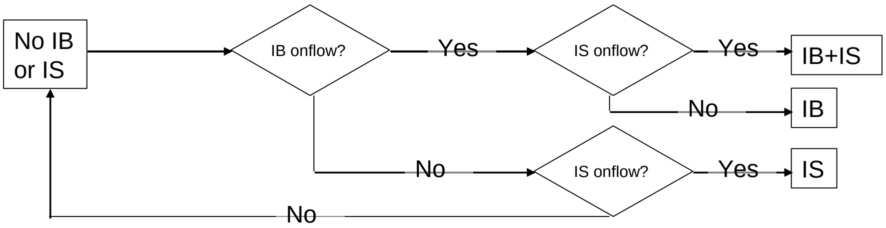


Figure 8

Figure 8 shows how under the old regime there is an initial decision as to onflow to IB, followed by a decision as to onflow to IS. Both decisions are modelled using logits derived from historic data with separate equations used for the IS onflow decision dependent on receipt of IB.

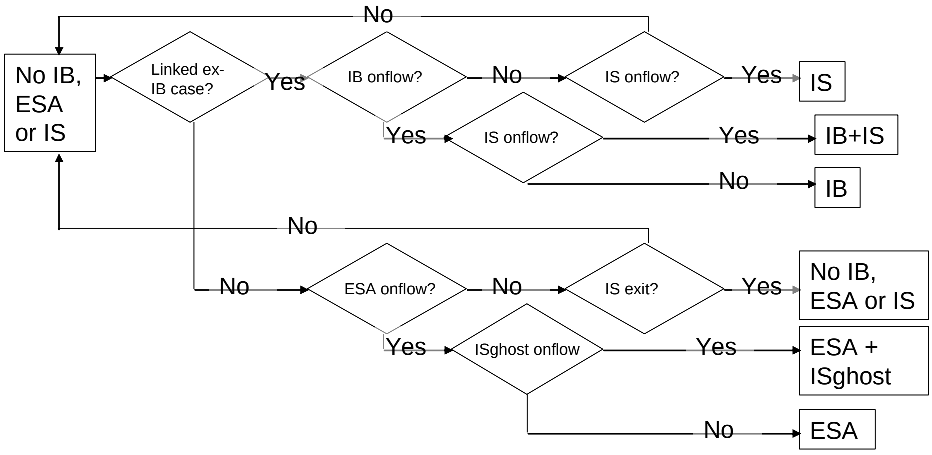

Post-ESA: Not live on IB, ESA or IS at start

Figure 9

After the introduction of ESA (Figure 9) the onflow decision is split into those with recent IB spells (who remain eligible to flow onto IB) and others who can now only flow onto ESA – however the same equations are used for both, with alignment used if required to control the number of onflows. Those who flow onto IB can then still flow onto IS (in the same or a later period). Those who flow onto ESA have the opportunity to acquire the ‘ghost of IS’ flag which can then be used as the equivalent of IS receipt in equations elsewhere within INFORM.

3.7 Conclusions

When a model is exclusively based on administrative data there is a concern that the lack of detail beyond that recorded in the administrative system will severely restrict the scope of modelling significant policy changes, for example limited information on potential new recipients. While the similarities between ESA and IB/IS were helpful in allowing key assumptions to be made about similarities in recipients, there were still some significant differences to be overcome and the success in using a variety of techniques to implement a new benefit within INFORM was a indicator of potential usefulness in the future beyond pure budgetary forecasts.

4 Modelling the Downturn

As previously described, much of the modelling in INFORM focuses on the interactions between different benefits and the importance of individual characteristics such as recent benefit history in driving paths through the benefit system. While factors such as seasonality in benefit transitions are allowed for there is an implicit assumption that relationships derived from recent administrative data will remain valid through the forecast period. For example if a forecaster expects a significant change in onflows to a benefit (for example due to a policy change or other exogenous disturbance) then this is achieved through a combination of alignment of within model onflows and adjustments to onflows of completely new cases, with the equations within the model acting to order characteristics of onflows rather than control numbers.

4.1 Use of alignment

In the earlier stages of development considerable reliance was placed on the role of alignment. The ability to align onflows was a key user requirement but risked deflecting attention from the need for the model to be as realistic as possible even if alignment was not used. It also detracted from one of the key benefits originally envisaged from developing an integrated benefit model – i.e. the power to simulate interactions between benefits. In a scenario where all benefit onflows are aligned to agreed forecasts then this power is constrained.

During the period of relative economic stability when INFORM was first developed this was not such a significant issue and it was important to use alignment to ensure consistency of assumptions with existing models for validation purposes. However the onset of the global economic down turn and the significant impact on forecasts for out of work benefits prompted an increased emphasis on the potential impact on other benefits. This required a combination of approaches, some of which were already part of the model architecture and others requiring more innovative modifications.

4.2 Limitations of data

The period of administrative data available (1999 – 2008) for model estimation was of limited use in predicting the impact of a major recession but if other sources could be used to predict, for example, expected changes in JSA outflow rates these could then be applied to the transition rates being generated by the equations within INFORM. The model would allow these adjustments to be applied differentially to different groups if required and over different time profiles while maintaining the power of the original equations in taking account of different characteristics – although this assumes that the relative rankings in exit probabilities implied by the equations was still valid.

As described in section 2.7 a requirement of model users is that INFORM be capable of aligning to specific JSA caseload scenarios – so for JSA the constraint was a target caseload rather than target onflows as for other benefits. The intention was to model various JSA scenarios while letting the other benefits run unaligned and then consider the different impacts on other benefit caseloads. By default the model could be made to hit a target caseload in any given month by calculating exits based on the appropriate equation and then aligning onflow rates as required. However under a scenario of rapidly increasing JSA caseload this was clearly unrealistic and would have led to excessive amounts of churn– particularly as the onflows required would be generated from within the model population. Figure 10 illustrates this by showing the unrealistic episode counts generated within the simulated JSA caseload [4] (e.g. the line showing '5+' is those cases within the caseload who are on their fifth or more JSA episode).

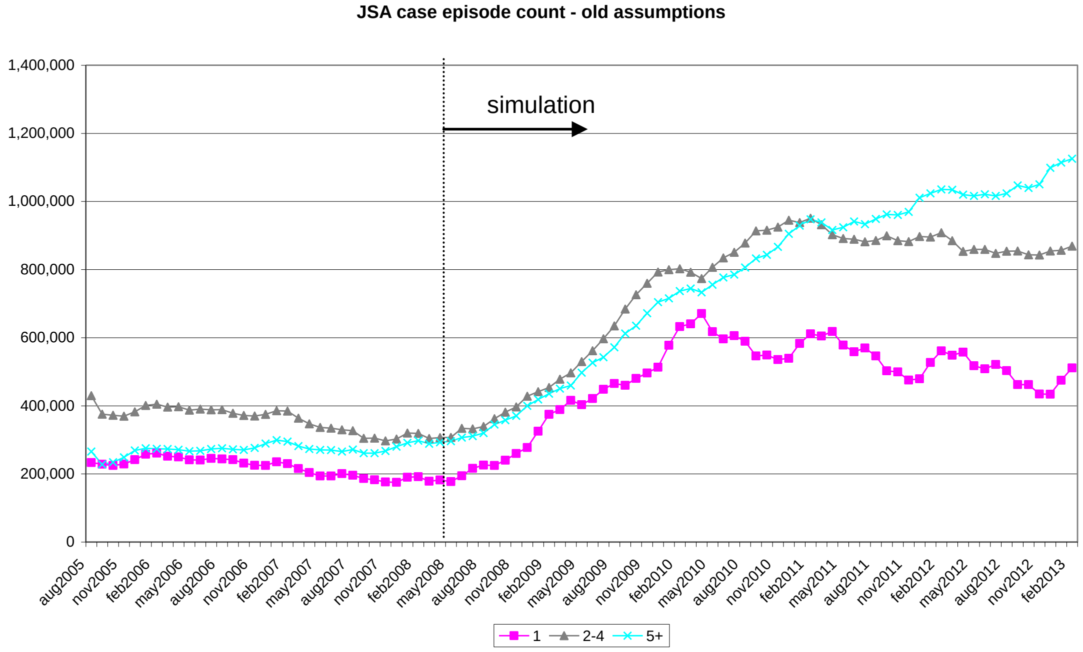


Figure 10

It was therefore necessary to come up with adjustments that would simulate more realistic changes to both numbers and characteristics of JSA flows.

4.3 Onflows

Within INFORM benefit onflows in any given forecast month are a combination of new recipients appended to the model and within model cases moving back onto benefit. It was felt that the recession would be likely to lead to an increase in the proportion of onflows who had no or little previous engagement with the benefit system – particularly for JSA. This view was supported by emerging statistics on previous claims for new cases and so a significant change was made to the forecast profile of the proportion of 'new onflows' within overall JSA onflows.

This would reduce to some extent the reliance on churn of existing cases to align to caseloads as well as provide a more realistic profile in terms of simulated paths through the benefit system. Figure 11 shows the impact of

---

[4] Note that the caseloads implied in the charts do not represent an official forecast and are simply one possible scenario

this change – although the absolute number of cases experiencing their first JSA episode increases there is still excessive churn as the simulation progresses.

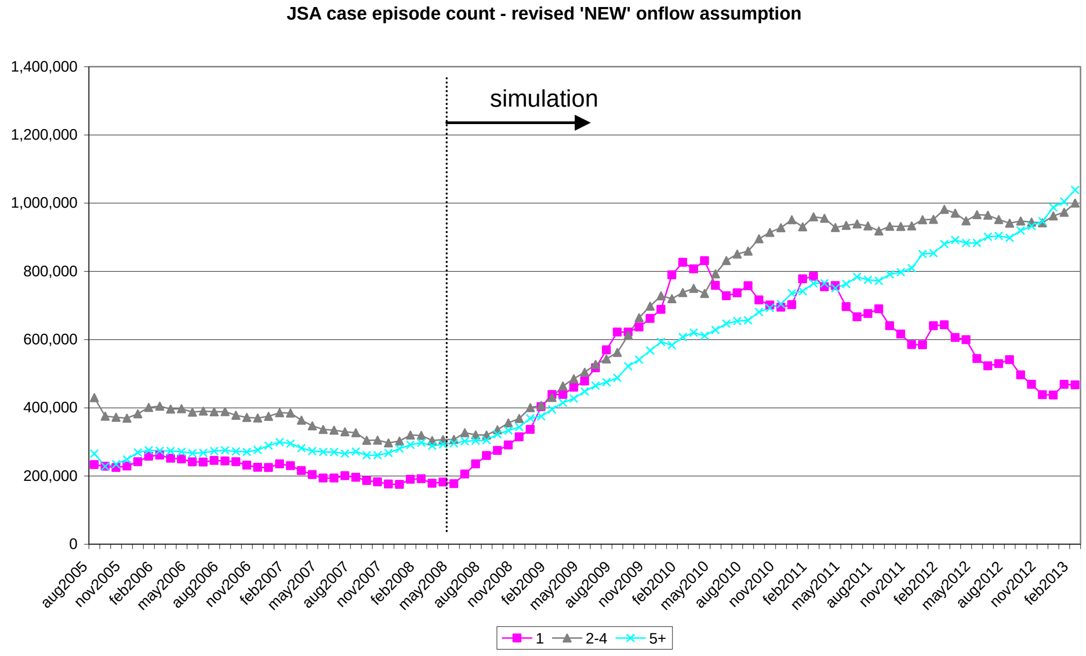


Figure 11

4.4 Exit rate changes

One key requirement was to allow differential adjustments to benefit outflow probabilities for different groups while retaining as much as possible of previously estimated econometric models. While a key lever in attempting to model the downturn was the ability to increase onflows to benefit it was not realistic to assume that outflow rates would be unaffected. In fact to do so and also meet caseload alignment targets would lead to unrealistic 'churn' of JSA cases in the simulation.

Separate, quarterly, stock-flow based spreadsheet models were used to provide estimates of the expected impact to onflows and outflows with flows within a quarter derived from the difference between stock in one quarter and that in the previous quarter at a duration three months shorter. However this could not provide outflow estimates for the first quarter from onflow.

Due to time constraints, the simple approach was taken of deriving alignment totals by applying estimated changes in the exit rates to average exit rates observed over the last two years. Since the stock-flow spreadsheet model could not provide an estimated rate for the first quarter, the assumption was made that the change in the first quarter was equal to the change estimated for the second quarter, although to avoid a lag between the start of the downturn and the change in the exit rates these changes were applied to the first quarter three months earlier than the date they were calculated from for the second quarter. Sixty six alignment groups were used, split by banded age and gender, which was driven by the breakdown in the estimated changes from the spreadsheet model

Figure 12 shows the episode counts after the exits change was applied resulting in a more reasonable caseload with the overall increase in caseloads driven much more by increases in those experiencing their first JSA episode.

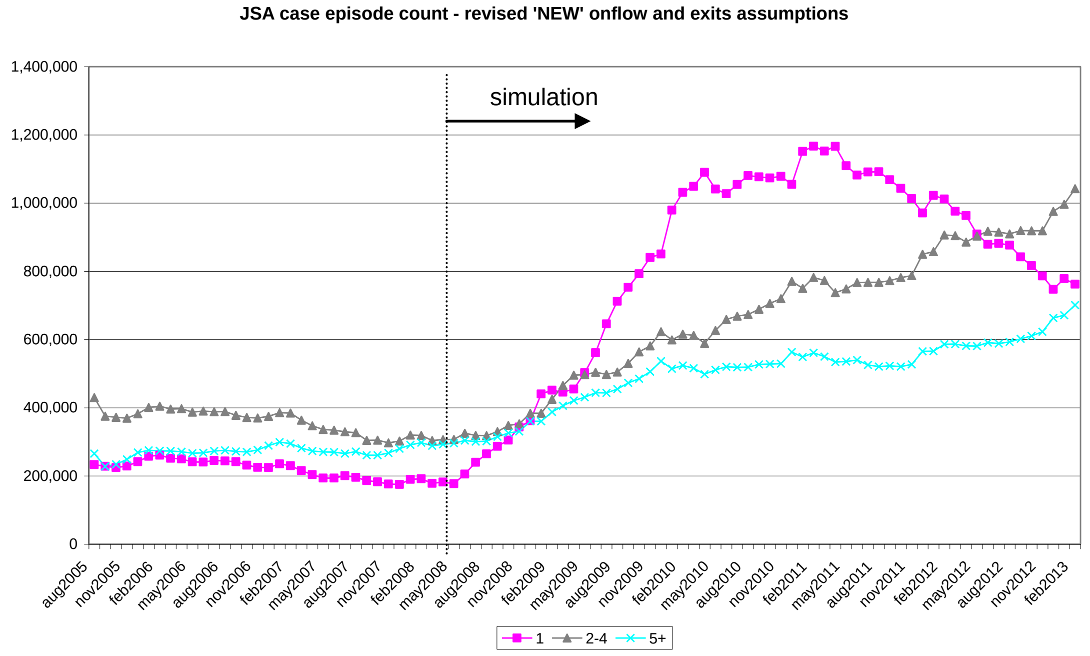


Figure 12

4.5 Benefit interactions

A key objective of this work was to get a better feel for the expected impact on other benefits of projected increases in unemployment and thus JSA – in particular assumptions concerning the proportion of cases who exit JSA and flow onto sickness benefits (ESA) possibly also triggering increases in disability benefits such as DLA. Historically a fairly stable proportion of cases who exit from JSA have then started sickness benefit episodes and so there is considerable interest in whether this will be sustained during the downturn and also the switch from IB to ESA.

Figure 13 illustrates how JSA scenarios changed as the severity of the down turn became clearer (although again neither scenario is intended to reflect an official view).

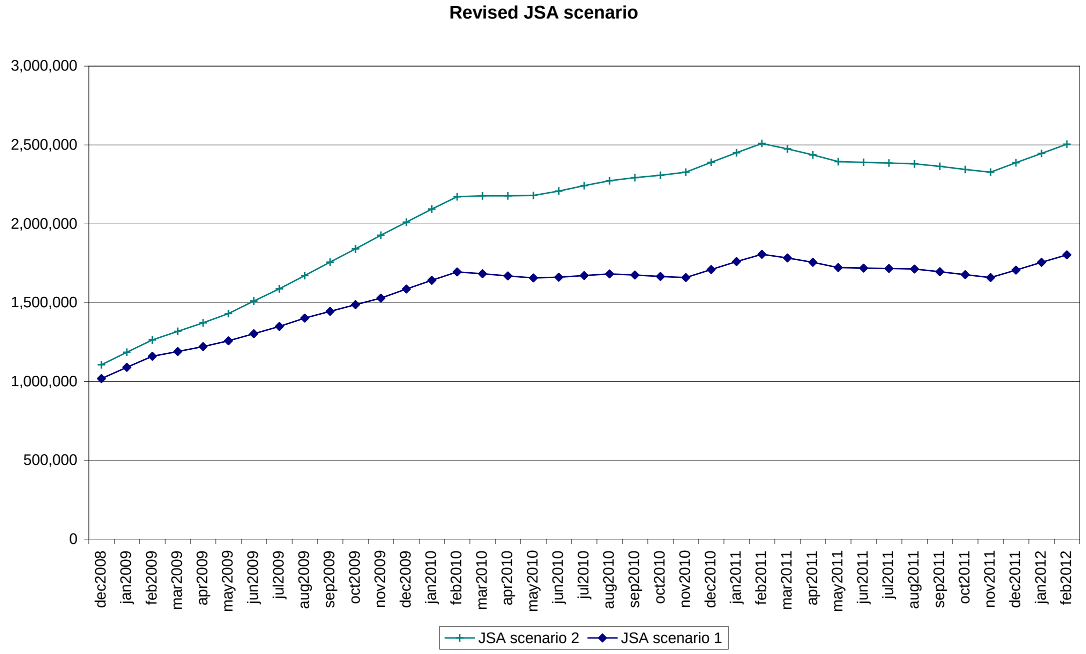


Figure 13

Figure 14 then shows the impact on ESA onflows of both the JSA scenario change and then the modifications as described in 4.3 and 4.4 above. The *'JSA scenario1'* and '*JSA scenario2'* series show the growth in ESA onflows under the higher JSA scenario, reflecting relationships between the two embedded in the model equations**.** The first modification (described in 4.3) and shown by the series *'Revised 'new' JSA onflows'* (without adjustments to exit rates) leads to an increased proportion of cases with no benefit history and short duration who will be relatively likely to exit JSA quite quickly. While this will reduce churn in the JSA caseload to some extent it also appears to drive higher ESA onflows – most likely through the equations for ESA onflow being sensitive to previous JSA and recent JSA exit. The further modification to reduce JSA exit rates (series *'Revised 'new' JSA onflows and exits'* corrects for the excessive churn (see 4.4) and also reduces the impact on ESA onflows. While this serves to illustrate the sort of interactions of interest and modelled within INFORM, more detailed analysis is required to investigate the precise drivers of the different observed effects, as discussed below.

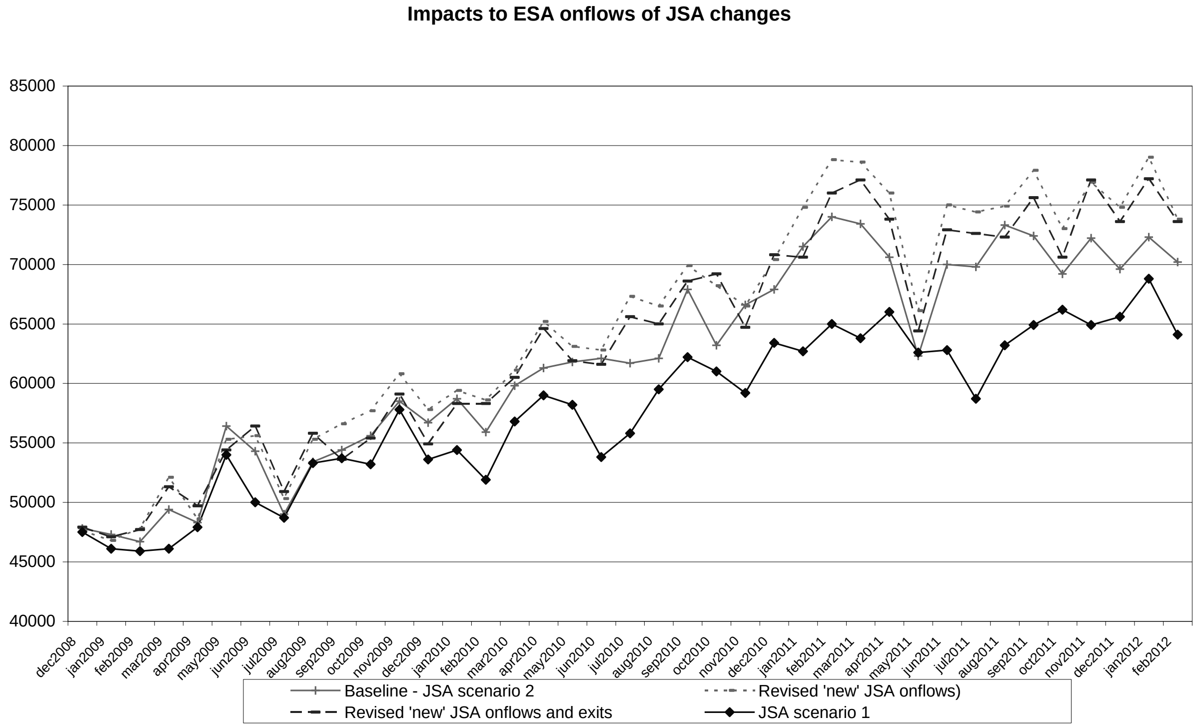


Figure 14

4.6 Further work

There is clearly more work that can be done in this area, particularly as more administrative data becomes available on the characteristics of JSA onflows and exits as the downturn progresses, relative to the previous more stable economic climate. It will be important to investigate the extent to which previous relationships, as embodied in the equations used in INFORM, remain valid in a period of rapid economic change. Of particular interest will be expectations for recovery – for example, given the limitations of the administrative data, the extent to which we can differentiate those individuals with higher human capital who are likely to return to employment and stay there from those who regularly cycle in and out of employment. More analysis is also needed of benefit interaction effects such as those with ESA explored above and whether they are sustained.

5 Conclusions

While it is still relatively early days in the usage and development of INFORM it is still on track to deliver the efficiency and other improvements anticipated– in particular the ability to model benefit interaction effects and to take account of far more relationships between individual characteristics and benefit histories than the single benefit models it is replacing.

The importance of the Genesis architecture in developing INFORM cannot be overstated. As described above INFORM does have limitations, particularly in the limited scope of the administrative data used and if developing from scratch it would probably not have been considered cost effective. However the standardised development system and interface, rapid prototyping and alignment features Genesis provides make special purpose models such as INFORM more feasible. Moreover as has hopefully been shown in this paper through a variety of techniques it has proved possible to overcome some of the inherent limitations the model in areas such as modelling major policy change as well as incorporating the impact of a massive change in the economic climate relative to the period on which the model and its equations were estimated.

The intention is now to build on this initial success and develop both the functionality and scope of INFORM beyond its initial benefit forecasting role.

APPENDICES

I. Genesis

Overview

Genesis is a generic architecture for implementing dynamic simulation models, developed by the Information and Analysis Directorate, Department for Work and Pensions (IAD, DWP).

Genesis is a flexible architecture, developed to allow dynamic micro simulation modelling to be carried out for any application, without requiring modification to the model code.

Through parameter spreadsheets, users specify a series of operations to be performed on data (processes and actions). The Genesis model reads these spreadsheets and generates code to perform the required operations on the base data. The location of the base data is specified in the Global spreadsheet within the Global workbook.

Because the base data, the operations, and the order in which the operations should be carried out, can all be specified in either spreadsheets or the SAS controlling program, no re-coding is necessary to implement a range of models.

Architecture

The model is made up of two distinct areas as shown in Figure 15, the SAS code that makes up the generic Model Engine and the parameter spreadsheets held in Excel.

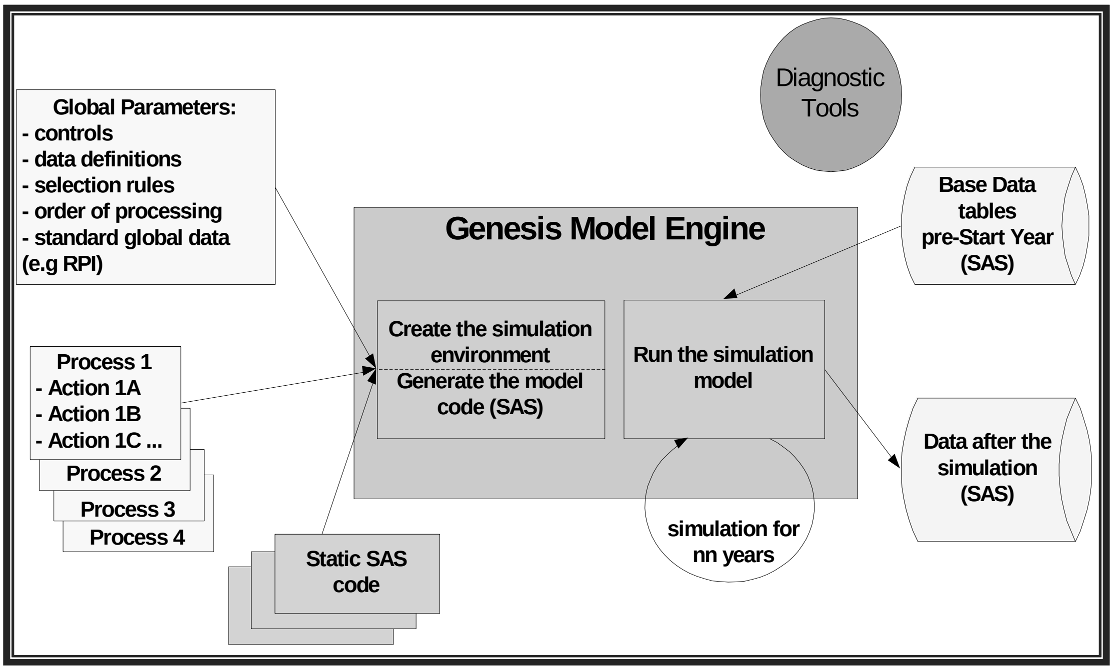


Figure 15

The Genesis system can be briefly described in the following way:-

- A simulation can be run for a defined number of periods
- For each period a number of life processes occur, for example fertility, changing labour market status, mortality etc
- The order that the life processes are run is flexible and is controlled by the analyst
- Each process is composed of a number of related actions.

The Model Engine is written in SAS and operates against SAS datasets held on a remote UNIX server. This server provides greater processing power to cope with the large datasets that are created during a Genesis run.

It takes the processing rules, general parameter tables and static code that are indicated in the parameter spreadsheets and then converts this information into a specific model. The Model Engine then produces the actual model code for this simulation and runs it against the main data tables.

The actual SAS code produced for a particular simulation run is saved and can be checked if necessary.

The parameter spreadsheets hold all of the information that is required by the Model Engine to run a Genesis simulation. The spreadsheets indicate the type of processes that are to be run, what order they are to be run, what data the processes should be applied to and where this data is located.

The spreadsheets hold the information about the specific economic algorithms that are to be used to simulate values for variables throughout a simulation run. It includes details about which explanatory variables, coefficients and selection criteria are to be used to simulate a value for that dependent variable.

Static code

It is possible to include static SAS code within the simulation model. This may be preferable if complex functionality is required that cannot be satisfied easily using the generic architecture.

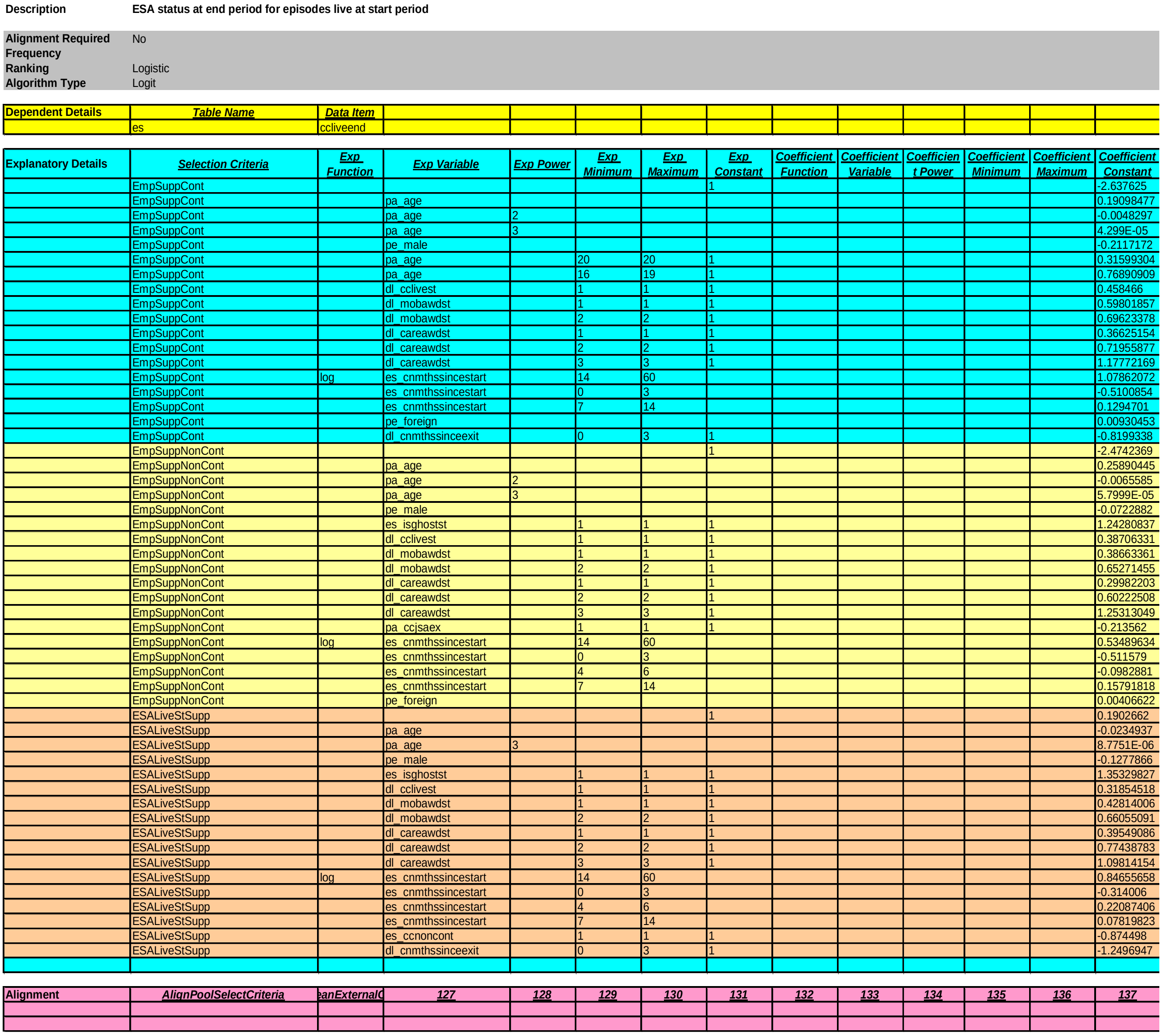
Static code may be written for a variety of purposes and may be specific to a particular iteration of Genesis.

When static code is run the only processing that takes place is to run that piece of SAS code. No other processing by Genesis takes place and all global and local selection criteria are ignored. It is the responsibility of the user writing the static code to ensure that any loops required are added to their static code. Similarly any filters required must be coded into the programme, as no other selection criteria will be applied.

II. Alignment and ‘Delta Alignment’

Alignment is part of the simulation process. It ensures that where a value has been set for a particular dependent variable, the value falls within an acceptable probability of it occurring in real-life. For example, alignment for a fertility process could be used to ensure that the number of babies born within a particular year in the simulation is consistent with the number of babies born within the general population.

Figure 16 shows an example Genesis action – in this case that used to determine exits from Employment Support Allowance. Three different logit models are shown, the appropriate one chosen for each case dependent on characteristics in the ‘SelectionCriteria’column. The alignment section at the bottom shows where alignment targets dependent on characteristics could be set if required. This example generates a probability of survival.



Description	**ESA status at end period for episodes live at start period**

Alignment Required	No
Frequency	
Ranking	Logistic
Algorithm Type	Logit

Dependent Details	***Table Name***	***Data Item***											
	es	ccliveend											

Explanatory Details	***Selection Criteria***	***Exp Function***	***Exp Variable***	***Exp Power***	***Exp Minimum***	***Exp Maximum***	***Exp Constant***	***Coefficient Function***	***Coefficient Variable***	***Coefficien t Power***	***Coefficient Minimum***	***Coefficient Maximum***	***Coefficient Constant***
	EmpSuppCont						1						-2.637625
	EmpSuppCont		pa_age										0.19098477
	EmpSuppCont		pa_age	2									-0.0048297
	EmpSuppCont		pa_age	3									4.299E-05
	EmpSuppCont		pe_male										-0.2117172
	EmpSuppCont		pa_age		20	20	1						0.31599304
	EmpSuppCont		pa_age		16	19	1						0.76890909
	EmpSuppCont		dl_cclivest		1	1	1						0.458466
	EmpSuppCont		dl_mobawdst		1	1	1						0.59801857
	EmpSuppCont		dl_mobawdst		2	2	1						0.69623378
	EmpSuppCont		dl_careawdst		1	1	1						0.36625154
	EmpSuppCont		dl_careawdst		2	2	1						0.71955877
	EmpSuppCont		dl_careawdst		3	3	1						1.17772169
	EmpSuppCont	log	es_cnmthssincestart		14	60							1.07862072
	EmpSuppCont		es_cnmthssincestart		0	3							-0.5100854
	EmpSuppCont		es_cnmthssincestart		7	14							0.1294701
	EmpSuppCont		pe_foreign										0.00930453
	EmpSuppCont		dl_cnmthssinceexit		0	3	1						-0.8199338
	EmpSuppNonCont						1						-2.4742369
	EmpSuppNonCont		pa_age										0.25890445
	EmpSuppNonCont		pa_age	2									-0.0065585
	EmpSuppNonCont		pa_age	3									5.7999E-05
	EmpSuppNonCont		pe_male										-0.0722882
	EmpSuppNonCont		es_isghostst		1	1	1						1.24280837
	EmpSuppNonCont		dl_cclivest		1	1	1						0.38706331
	EmpSuppNonCont		dl_mobawdst		1	1	1						0.38663361
	EmpSuppNonCont		dl_mobawdst		2	2	1						0.65271455
	EmpSuppNonCont		dl_careawdst		1	1	1						0.29982203
	EmpSuppNonCont		dl_careawdst		2	2	1						0.60222508
	EmpSuppNonCont		dl_careawdst		3	3	1						1.25313049
	EmpSuppNonCont		pa_ccjsaex		1	1	1						-0.213562
	EmpSuppNonCont	log	es_cnmthssincestart		14	60							0.53489634
	EmpSuppNonCont		es_cnmthssincestart		0	3							-0.511579
	EmpSuppNonCont		es_cnmthssincestart		4	6							-0.0982881
	EmpSuppNonCont		es_cnmthssincestart		7	14							0.15791818
	EmpSuppNonCont		pe_foreign										0.00406622
	ESALiveStSupp						1						0.1902662
	ESALiveStSupp		pa_age										-0.0234937
	ESALiveStSupp		pa_age	3									8.7751E-06
	ESALiveStSupp		pe_male										-0.1277866
	ESALiveStSupp		es_isghostst		1	1	1						1.35329827
	ESALiveStSupp		dl_cclivest		1	1	1						0.31854518
	ESALiveStSupp		dl_mobawdst		1	1	1						0.42814006
	ESALiveStSupp		dl_mobawdst		2	2	1						0.66055091
	ESALiveStSupp		dl_careawdst		1	1	1						0.39549086
	ESALiveStSupp		dl_careawdst		2	2	1						0.77438783
	ESALiveStSupp		dl_careawdst		3	3	1						1.09814154
	ESALiveStSupp	log	es_cnmthssincestart		14	60							0.84655658
	ESALiveStSupp		es_cnmthssincestart		0	3							-0.314006
	ESALiveStSupp		es_cnmthssincestart		4	6							0.22087406
	ESALiveStSupp		es_cnmthssincestart		7	14							0.07819823
	ESALiveStSupp		es_ccnoncont		1	1	1						-0.874498
	ESALiveStSupp		dl_cnmthssinceexit		0	3	1						-1.2496947

Alignment	***AlignPoolSelectCriteria***	***eanExternalC***	***127***	***128***	***129***	***130***	***131***	***132***	***133***	***134***	***135***	***136***	***137***

Figure 16

Delta-Alignment

When attempting to forecast changes, for example in benefit onflow of outflow rates, it is often the case that estimates of the impact are available only as changes in the rates (i.e. the proportion of the pool of potential transitions that actually undergo the transition) rather than as the absolute transition rates themselves. Often the impact assessments will have been made using a cell-based model where overall alignment rates are either unavailable or unsuitable. Rather than re-estimating all policy impacts in a format directly applicable to INFORM, we developed the methodology for aligning the change in the transition rate as a simple and practicable option.

INFORM makes extensive use of unaligned regressions, and we would not wish without good reason to lose the benefit this gives us in picking up changes driven by trends in the explanatory variables. One of the key areas in which INFORM is to be used is in assessing the impacts of changes in one working age benefit across all of the others. In order to do this, it is necessary to run the model unaligned so far as possible. Purely aligning the changes rather than absolute rates allows us to do this while retaining the estimated impact of policy changes.

The methodology for performing the alignment is very simple. A first pass of the regression generates the unaligned transition rates, broken down by any required subgroups. These rates are then transformed, within the limitation that the transformed rates can neither be greater than one or less than zero, and these transformed rates are used as the alignment targets for a second pass of the regression. The same random numbers are used in both passes of the regression.

Genesis uses alignment by sorting, but there is no reason that this methodology could not be applied to any other alignment algorithm as the only difference from traditional alignment is to calculate the alignment target at runtime.

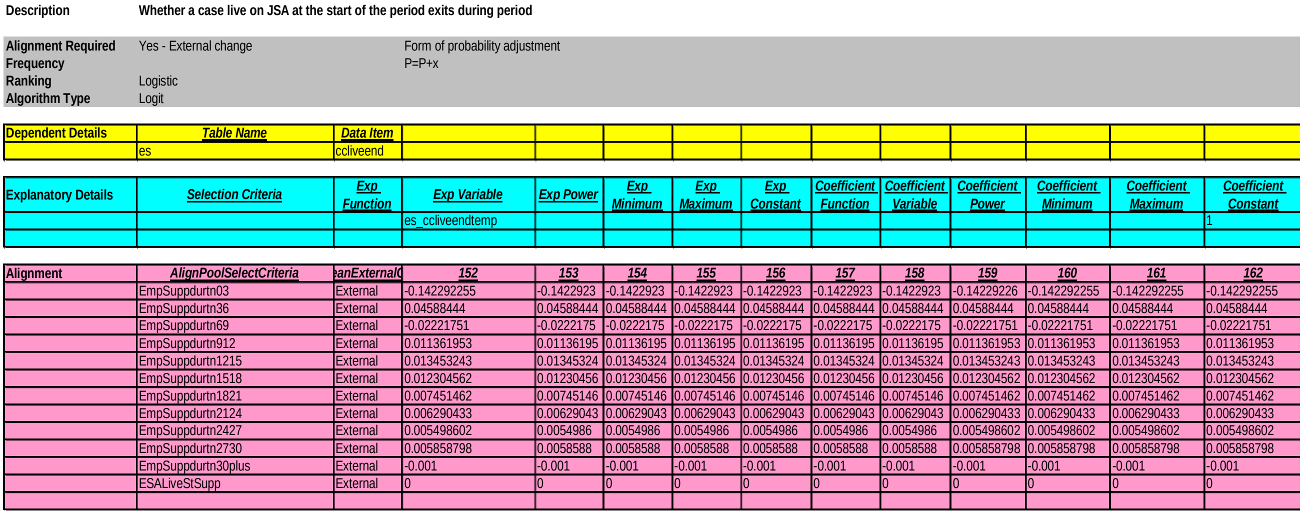
Implementation within Genesis

The Genesis engine generates the model code from a user interface of Excel spreadsheets. The engine does not support Delta-alignment, so the changes required for this are made via static code. The technique is ultimately intended to be accessible to the more advanced Genesis users rather than just model developers, so a spreadsheet tool was created to automate the process of writing the static code that so far as possible resembles the Genesis spreadsheets.

To simplify this tool, the limitation was imposed that it only accepts a single parameter to the regression, and so it requires the full regression to be run before the static code process with the modification that rather than generating transitions it simply outputs the logits. These logits are then used as the single regression parameter in the static code generator.

The tool allows time-varying alignment split by any required subgroups, with the changes in the form either of adding to / subtracting from or multiplying by the unaligned rate.

Figure 17 shows an example sheet, in the same style as a Genesis action, used to generate code to create the revised alignment targets. The changes can be varied by characteristics and period of simulation. In this example the change is additive. For example in period 152 the ‘raw’ survival rate for those cases in the ‘*Empsuppdurtn03*’ group (short duration) is reduced by an absolute reduction of around 0.142.



Description Whether a case live on JSA at the start of the period exits during period

Alignment Required	Yes - External change	Form of probability adjustment
Frequency		P=P+x
Ranking	Logistic	
Algorithm Type	Logit	

Dependent Details	***Table Name***	***Data Item***											
	es	ccliveend											

Explanatory Details	***Selection Criteria***	***Exp Function***	***Exp Variable***	***Exp Power***	***Exp Minimum***	***Exp Maximum***	***Exp Constant***	***Coefficient Function***	***Coefficient Variable***	***Coefficient Power***	***Coefficient Minimum***	***Coefficient Maximum***	***Coefficient Constant***
			es_ccliveendtemp										1

Alignment	***AlignPoolSelectCriteria***	***eanExternalC***	***152***	***153***	***154***	***155***	***156***	***157***	***158***	***159***	***160***	***161***	***162***
	EmpSuppdurtn03	External	-0.142292255	-0.1422923	-0.1422923	-0.1422923	-0.1422923	-0.1422923	-0.1422923	-0.14229226	-0.142292255	-0.142292255	-0.142292255
	EmpSuppdurtn36	External	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444	0.04588444
	EmpSuppdurtn69	External	-0.02221751	-0.0222175	-0.0222175	-0.0222175	-0.0222175	-0.0222175	-0.0222175	-0.02221751	-0.02221751	-0.02221751	-0.02221751
	EmpSuppdurtn912	External	0.011361953	0.01136195	0.01136195	0.01136195	0.01136195	0.01136195	0.01136195	0.011361953	0.011361953	0.011361953	0.011361953
	EmpSuppdurtn1215	External	0.013453243	0.01345324	0.01345324	0.01345324	0.01345324	0.01345324	0.01345324	0.013453243	0.013453243	0.013453243	0.013453243
	EmpSuppdurtn1518	External	0.012304562	0.01230456	0.01230456	0.01230456	0.01230456	0.01230456	0.01230456	0.012304562	0.012304562	0.012304562	0.012304562
	EmpSuppdurtn1821	External	0.007451462	0.00745146	0.00745146	0.00745146	0.00745146	0.00745146	0.00745146	0.007451462	0.007451462	0.007451462	0.007451462
	EmpSuppdurtn2124	External	0.006290433	0.00629043	0.00629043	0.00629043	0.00629043	0.00629043	0.00629043	0.006290433	0.006290433	0.006290433	0.006290433
	EmpSuppdurtn2427	External	0.005498602	0.0054986	0.0054986	0.0054986	0.0054986	0.0054986	0.0054986	0.005498602	0.005498602	0.005498602	0.005498602
	EmpSuppdurtn2730	External	0.005858798	0.0058588	0.0058588	0.0058588	0.0058588	0.0058588	0.0058588	0.005858798	0.005858798	0.005858798	0.005858798
	EmpSuppdurtn30plus	External	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	ESALiveStSupp	External	0	0	0	0	0	0	0	0	0	0	0

Figure 17

III. Determinants of major events simulated by INFORM

One of the key justifications for developing INFORM was the additional capability it would give for modelling the interactions between working age benefits in a single model. The table below provides a summary of some of the major transitions and flows modelled in INFORM and the variables used, showing the nature of relationships with other benefits and benefit history.

Transition	Variables used to determine
Onflow to Job Seekers Allowance	Age, gender, previous JSA (how many times), current DLA, current IS, current CA, seasonality, transfer from IB
Outflow from Job Seekers Allowance	Age, gender, current DLA, previous IS, duration, episode count, contributory status, seasonality
Onflow to Incapacity Benefit / Employment Support Allowance	Age, gender, transfer from IS/JSA, duration on benefit, previous JSA, previous DLA, recent exit from DLA, DLA award type
Outflow from Incapacity Benefit / Employment Support Allowance	Age, gender, current IS/'IS ghost', DLA award type, duration
Onflow to Disability Living Allowance	Age, gender, current IS/'IS ghost', current JSA, current IB/ESA, previous IB/DLA, lone parent, recent IB exit
Outflow from Disability Living Allowance	Age, gender, DLA award type, duration, current IB/ESA, current IS / 'IS ghost'
Onflow to Income Support	Age, gender, current IB, seasonality, previous IS, number of previous episodes
Outflow from Income Support	Lone parent (and age youngest child), age, gender, seasonality, current DLA, duration, seasonality
Mortality	Age, gender, DLA receipt and award type