Cohort analysis

3.1 Cohort analysis

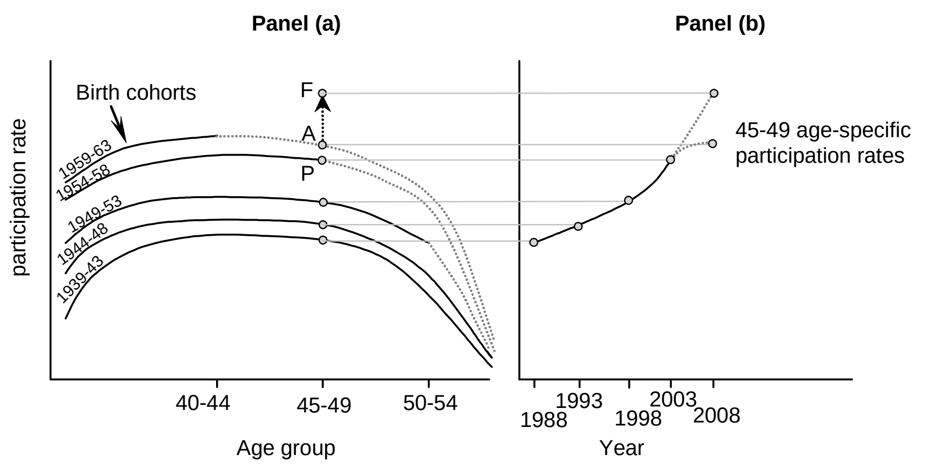
The labour market behaviour of people born in different years (so-called ‘cohorts’) can be quite different. This has implications for forecasting future patterns in labour supply. These generational variations may reflect their:

- different social attitudes (for example, attitudes to the role of women in the workforce after marriage or childbirth);
- varying aptitudes (due to different levels of education and different lifetime exposures to technology and opportunities for learning by doing); and
- the enduring effects of historical events (such as higher disability rates among combatants in the world wars or the potentially ‘scarring’ effects of mass unemployment).

Cohort effects may sometimes enable more accurate projections of participation rates. Using hypothetical data, figure 3.1 contrasts the picture of participation rates that can be given by cohort (panel a) versus age-specific information (panel b). Age-specific participation rates reflect the participation rates of different cohorts. In the example below, the observation for 2003 for those aged 45-49 years relates to the 1954-58 birth cohort, while that for 1998 for the same age group relates to the 1949-53 birth cohort.

Based on the past values of the age-specific participation rate shown in panel (b), it would appear likely that the participation rate in 2008 for 45-49 year olds would be around that shown at F. However, this ignores the information given by data in a cohort form. As shown in panel (a), the participation rate for 45-49 year olds in 2008 reflects the labour market behaviour of the 1959-63 birth cohort. In the hypothetical data, this cohort’s participation rate is higher for each age than the previous cohort, but by less of an increment than the 1954-58 cohort over the 1949-53 cohort. A reasonable forecast of the participation rates for the 1959-63 cohort when aged 45-49 years is A, considerably less than the ‘naïve’ forecast based on extrapolating age-specific participation rates over time. Of course, in other circumstances, using a cohort approach may result in a higher projected estimate than one based on past trends of age-specific participation rates, but the point remains that cohort data may be helpful in providing more credible projections.

Figure 3.1 **How understanding cohort effects can provide better projections**



3.2 Cohort data

While true longitudinal data on participation rates are not available, it is possible to construct a synthetic panel of data.[1] However, several data deficiencies must be addressed. Data on participation rates by age are incomplete. For some periods, five year age groups are available, while for other periods, 10 year age groups are available (for example, for data from 1965 to 1977, participation rates are available for 25-34 year olds rather than separately for 25-29 and 30-34 year groups, as are available for other years). Moreover, data for some periods are missing altogether. Thus, yearly data are available from 1965, but for earlier years back to 1901, only infrequent census data are available. These data inadequacies represent an obstacle to analysis of the participation rates of people of given birth years over their lifetimes.

A method for resolving these data inadequacies is to use cubic spline smoothing and other interpolation techniques to fill in the missing gaps (box 3.1). This provides a dataset for examining cohort issues over long time frames. It is the basis for

[1] The panel data set is synthetic because it does not compare the same people over time. For example, the change in participation of women born in 1949-53 between ages 40-44 and 45-49 years is calculated by comparing women aged 40-44 years in 1998 with women aged 45-49 years, five years later. Some women present in the 40-44 group in 1998 will have died or emigrated by 2003, while some women present in the 45-49 group in 2003 will have come from overseas during the last five years.

graphical analysis, such as figure 3.7 in chapter 3. More reliable data over shorter periods have been used for the actual projections in chapter 3.

Box 3.1 **Interpolation methods for deriving cohort data**

For the various gaps in the availability of participation rates for quinquennial age groups between 1965 and 1977, a regression approach was used. A stable relationship was found between the participation rates of five year age groups and (among other variables) the ten year age groups in which they are subsumed. For example, for the period for which complete quinquennial group data were available, the following simple relationship was found for the participation rate for females aged 25 to 29 years (with t statistics in parentheses):

$$PR_{\text{females 25-29}} = \underset{(4.9)}{0.101} + \underset{(26.7)}{1.125}\ PR_{\text{females 25-34}} - \underset{(3.5)}{0.211}\ PR_{\text{females 20-24}}$$

This explained 99.7 per cent of the variation in participation rates for this group. This was then used to impute the female participation rate for 25-29 year olds for the years in which data were missing. Other missing quinquennial data were imputed in a similar way.

A standard cubic spline was used to interpolate data for the various missing years between 1901 to 1964. This will obviously fail to take account of business cycle impacts in missing years, and this must be noted as a limitation to the kind of analysis that can be undertaken using (at least parts of) the constructed dataset. Given the obviously approximate nature of the interpolated data, birth year cohorts were chosen in a way that favoured interpolated years that are close to, rather than far away from, years where observed data are available.

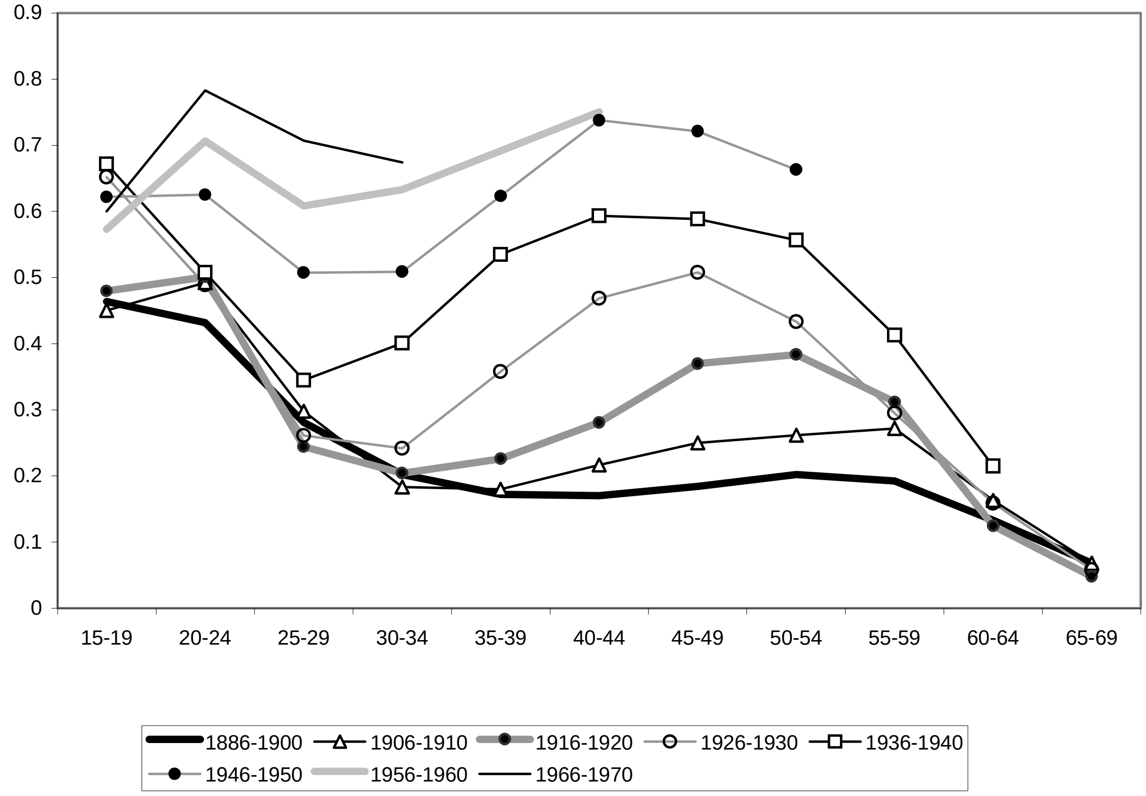
Snapshots of lifetime participation rates for people being born in each decade from the late 19th century to the end of the 20th century (figures 3.2 and 3.3) reveal the significant shifts that have occurred, particularly for females. (figure 3.7 in chapter 3 presents more fine-detailed 3D plots over an even longer period.)

Statistical evidence about the importance of cohort effects

The importance of cohort effects revealed by the qualitative results shown in section 3.3 of chapter 3 are confirmed by econometric analysis. On the basis of recent Australian data based on single years of age (rather than age groups in the data above), Ravindiran et al. (2002) used panel data analysis to substantiate that female cohort effects were strong and positive, while male cohort effects were weak and negative.[2] Participation with age followed the usual inverted-u shape — as above. A third general contributor to changes in participation rates — ‘year’ effects brought about by the business cycle — were relatively unimportant, particularly for men. Panel data analysis[3] undertaken by the Commission of the (unbalanced) data set shown in figure 3.7 in chapter 3 revealed broadly similar qualitative results to Ravindiran et al.

Figure 3.2 **Lifetime labour participation rates for females**[a]

Australia, females born in decade waves from 1886-1900 to 1966-1970



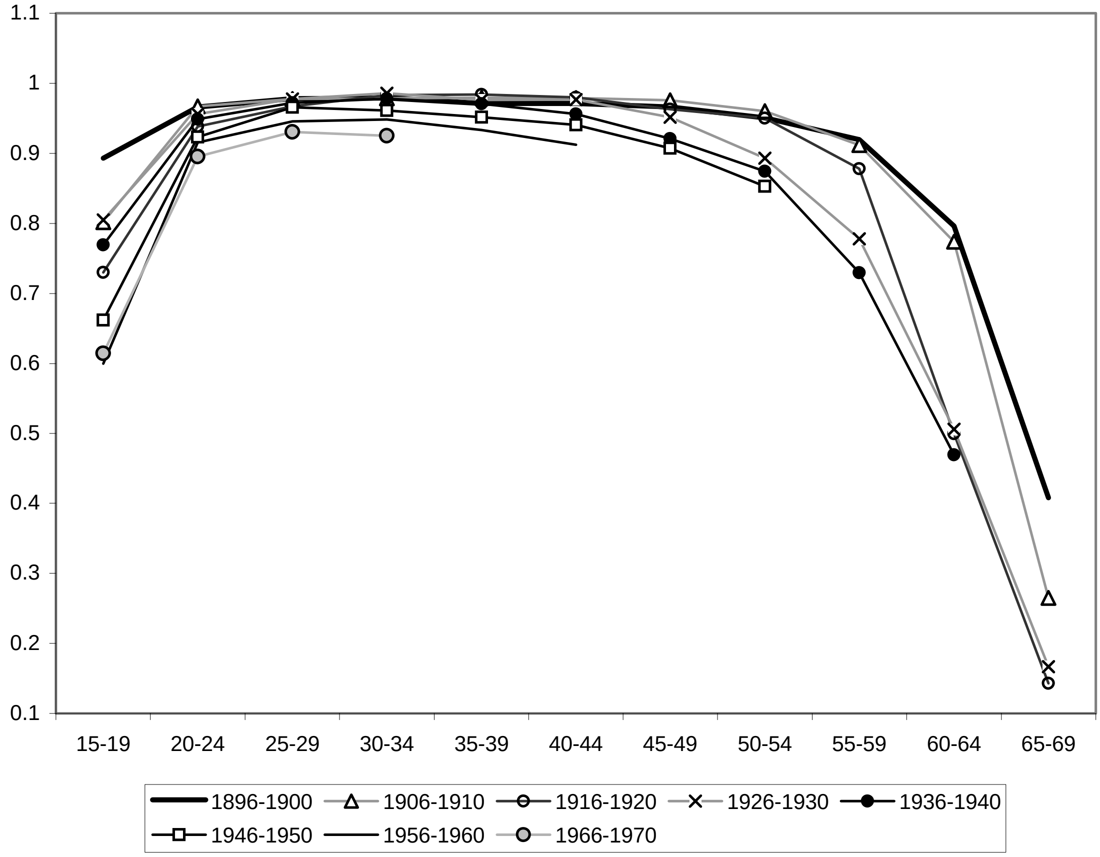
a Data on participation rates for those aged 10 to 14 years are excluded because they confuse the picture provided by other data in the graph. For this age group, participation rates have always been small, and following legislative changes relating to child labour, effectively set to zero in the mid 20th century.

Data sources: Original data sources are ABS (*The Labour Force, Australia, Historical Summary, 1966 to 1984* Cat. no. 6204.0; and *The Labour Force, Australia*, Cat. no. 6203.0) and Withers et al (1985). The Commission used various interpolation methods to construct some data.

2 Their data set relates to people born between 1937 and 1957 and charts their labour market involvement to 2001.

3 A fixed effects model was estimated (and appeared justified by the relevant test statistics compared with a random effects model).

Figure 3.3 **Lifetime labour participation rates for males**[a]
Australia, males born in decade waves from 1886-1900 to 1966-1970



[a]See **a** in figure above.

Data sources: See data sources in figure above.

3.3 Exits and entries in a cohort model

Cohort methods for forecasting participation rates rely on measures of exit from and entry to the labour market as cohorts age.

The most straightforward method for incorporating cohort effects into labour supply projections is to take account of the shape of the lifetime labour participation profile for each successive cohort and to extrapolate on that basis.[4] An example is the extrapolation shown in figure 6.1 for the 1959-63 cohort for age 45-49 years.

---

4 Another approach — using neural networks — was also trialed, but ultimately not used. There is a highly complex non-linear relationship between participation rates and age and cohort effects. Neural network models have the advantage that they do not impose excessive structure on the data and may capture these complex effects. However, while a range of neural network models

The shape of the lifetime participation profile is determined by the pattern of entry and exit from the labour force as a cohort ages. Given that reliable labour force data is usually only available for people in five year age ranges (people aged 15-19 years, 20-24 years and so on), entry rates must be calculated over five year periods. Accordingly, where participation rates are rising for a given cohort, the entry rate at time t is defined as the net addition to the labour force relative to the initial number of people who were *not* in the labour force five years previously:

$$Entry^{t}_{x,x+4} = \frac{increase\ in\ labour\ force\ from\ t-5\ to\ t\ of\ people\ aged\ x+1\ to\ x+9\ years\ at\ t}{number\ of\ people\ aged\ x\ to\ x+4\ years\ not\ in\ the\ labour\ force\ at\ t-5}$$

It should not be assumed that all these entries actually occur at time t — clearly, entries occur smoothly over the period from t-5 to t. The entry rate measures the completed number of entries *by* time t of the cohort aged x to x+4 years five years previously.

Estimating entry (and exit) rates is complicated by net migration and deaths, which mean that the civilian population aged between x+5 and x+9 at time t is different (generally by a small margin) from that aged between x and x+4 at time t-5. The impact of sample attrition and addition can largely be removed by assuming that the observed participation *rate* for people aged x+5 to x+9 at time t would be the same as that which would be observed had no sample attrition or addition occurred. In that case, the entry rate can be estimated as:

$$Entry^{t}_{x,x+4} = \frac{PR^{t}_{x+5,x+9} \times CPOP^{t-5}_{x,x+4} - LF^{t-5}_{x,x+4}}{NLF^{t-5}_{x,x+4}}$$

where PR, CPOP and NLF denote participation rate, civilian population and ‘not in the labour force’, respectively. Dividing through by CPOP and noting that NLF/CPOP = 1-PR, then the entry rate can be re-expressed as:

$$Entry^{t}_{x,x+4} = \frac{PR^{t}_{x+5,x+9} - PR^{t-5}_{x,x+4}}{1 - PR^{t-5}_{x,x+4}}$$

However, it is not feasible for some people in the civilian population to enter the labour force, so for the purpose of defining entry rates, NLF is defined as the difference between the maximum number of people in the civilian population who could be in the labour force (CPOP*) and the observed labour force. In that case,

produced credible short run forecasts, their long run forecasts were sometimes implausible. For example, the models forecast participation rates for 2050 approaching 100 per cent for females aged 30-34 years (at the one extreme) and 5 per cent for males aged 65-69 years (at the other).

NLF/CPOP = CPOP*/CPOP – LF/CPOP, so that the entry rate can be represented as:

$$Entry^{t}_{x,x+4} = \frac{PR^{t}_{x+5,x+9} - PR^{t-5}_{x,x+4}}{PR^{*} - PR^{t-5}_{x,x+4}}$$

where PR* is the maximum potential participation rate (with Burniaux et al. 2003 using PR*=0.99 and 0.95 for men and women respectively) and x are ages for quinquennial age groups. For example, for the 1954-58 cohort, the entry rate from 1998 to 2003 is the participation rate for 45-49 year olds in 2003, less the rate for 40-44 year olds in 1998, over the potential number of people who could enter the labour force in this group.

Where participation rates are falling for a given cohort, the *exit rate* is defined as the net reduction in the labour force relative to the number of people who were initially in the labour force in that cohort:

$$Exit^{t}_{x,x+4} = \frac{PR^{t-5}_{x,x+4} - PR^{t}_{x+5,x+9}}{PR^{t-5}_{x,x+4}}$$

3.4 Cohort exit and entry rates

Fixed exit and entry rates

The analysis by Burniaux et al. (2003) employs fixed entry and exit ratios — based on the last observed values of these ratios. If entry and exit rates remain fixed at their current values, then only data from 1999 and 2004 are required to estimate future participation rates for each of the age groups for five year intervals into the future. This has the advantage of computational simplicity and would be appropriate if exit and entry rates were non-trending.

For example, the participation rate for the 1960-64 female cohort when aged 45-49 years (that is, in 2009, five years after 2004) can be estimated as from the appropriate exit rate as:

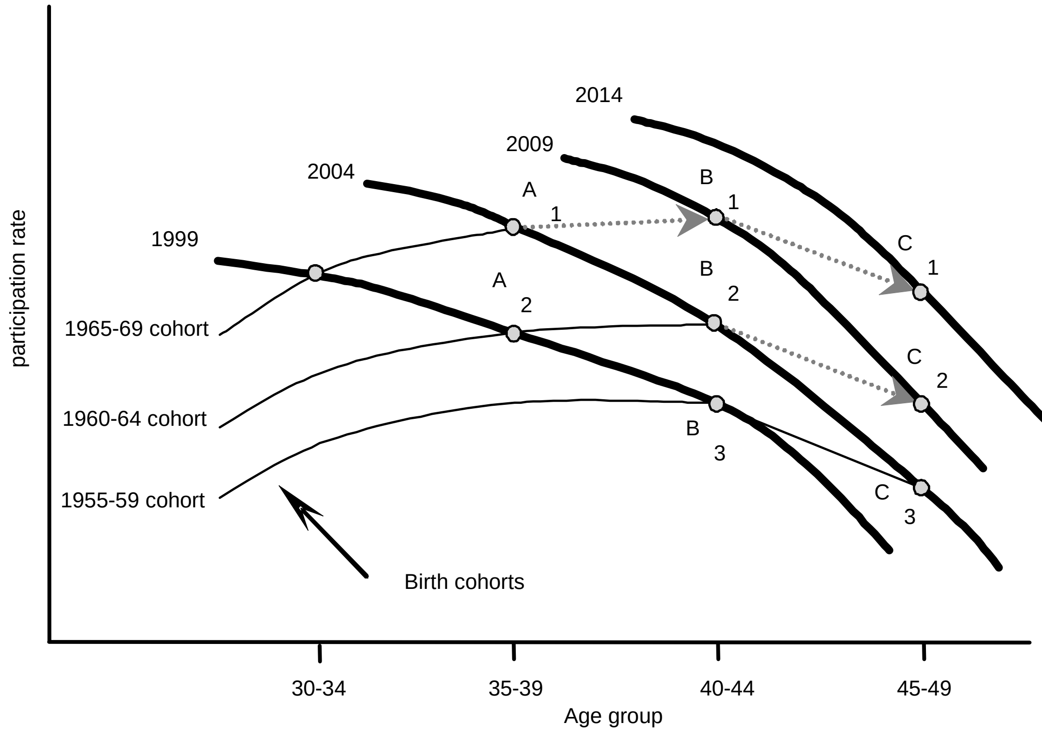
$$PR^{2009}_{1960-64\ cohort} = \left\{1 - \frac{PR^{1999}_{40-44} - PR^{2004}_{45-49}}{PR^{1999}_{40-44}}\right\}_{1955-59\ cohort} \times PR^{2004}_{1960-64\ cohort}$$

which is equivalent to $C_3/B_3 \times B_2$ in figure 3.4.

Similarly, the participation rate for the younger female cohort born in 1965-1969 when aged 40-44 years (again in 2009) can be estimated from the appropriate *entry rate* (noting from figure 6.4 that A1 to B1 rises so that entry has occurred) as:

$$PR^{2009}_{1965-69\ cohort} = PR^{2004}_{1965-69\ cohort} + \left\{\frac{PR^{2004}_{40-44} - PR^{1999}_{35-39}}{PR^{*} - PR^{1999}_{35-39}}\right\}_{1960-64\ cohort} \times (PR^{*} - PR^{2004}_{1965-69\ cohort})$$

Figure 3.4 **The dynamic cohort method for projecting participation rates**



Data source: Burniaux et al. (2003).

This leaves gaps between each of the years 2004, 2009, 2014, 2019 and so on. These gaps can be completed using interpolation.[5]

Time varying entry and exit rates

As Burniaux et al. (2003) note, it would be desirable to also produce forecasts of participation rates that allow entry and exit rates to evolve over time. The future

5 If fixed entry/exit rates are used, linear interpolation is probably most appropriate for several reasons. First, it is easier to implement than spline methods and is justified by the small differences between rates observed at five year intervals. Second, the participation rates converge to a set value for each of the age groups, and linear interpolation gives exact results when this occurs, whereas spline methods do not.

evolution of exit and entry rates will determine what happens to the labour force involvement of present cohorts. This section sets out the method used for estimating such dynamic exit and entry rates for Australia. Such time varying exit and entry rates were used in the projections in chapter 3.

The pattern of exit and entry rates in the past time is sometimes quite erratic, making it hard to forecast their future paths. Smoothing helps to reduce noise (but at the risk of producing spurious trends). Exit rates using smoothed data are shown in figure 3.5. Where exit rates are negative, it means that entry is occurring, and that it is more informative to graph the entry rate. For example, the comparable entry rate for people aged 15-19 years is shown in figure 3.6.

The data reveal that some rates appear to have stabilised — as in exit rates for females aged 25-29 years and entry rates for females aged 15-19 years. Others appear to still be trending. For example, exit rates for older workers of both genders (aged 55-64 years) have generally been dropping, and this may continue.

Clarifying bounds for exit and entry rates in dynamic models

There are bounds on potential exit rates, which are important when projecting dynamic exit rates. When net exits occur, then the exit rates (*Exit*) is bounded: $0 \le Exit \le 1$. When all members of a cohort exit the labour force over a period, then X=1, the highest possible bound. When there is no net attrition, *Exit* = 0. If there is negative net exits (entry) then, in theory, in the extreme $Exit \to -\infty$.

Given knowledge of the future path of *Exit*, forecasts of participation rates can be made by noting that:

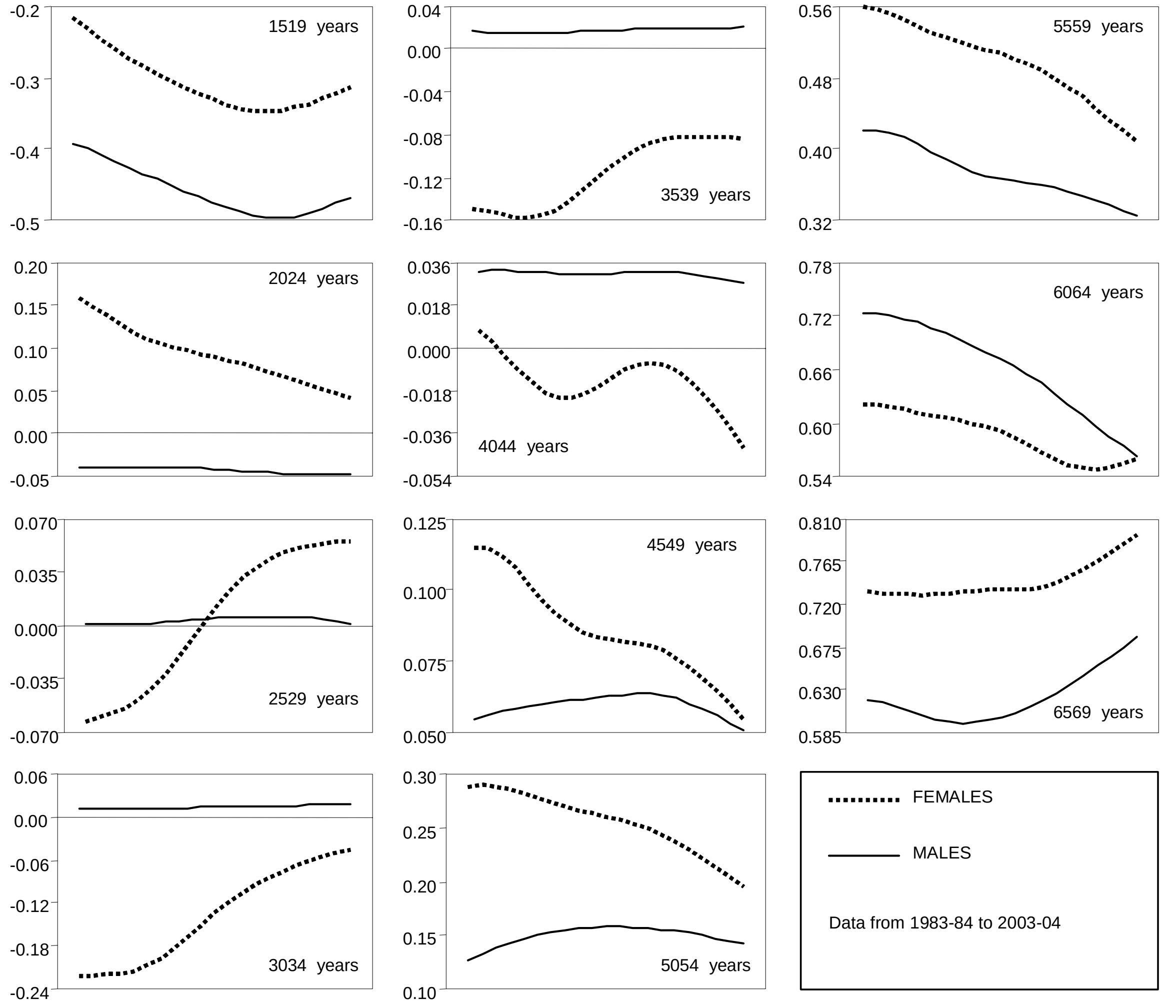
$$PR^{t}_{x+5,x+9} = (1 - Exit_{a,a+4,t}) \times PR^{t-5}_{x,x+4}$$

While in theory negative net exit rates cover cases where net entry occurs, it is useful to also set out a direct measure of entry rates (*Entry*). These are defined relative to the stock of people that could be, but are not, in the labour force.

Where net entry occurs, then $0 \le Entry \le 1$. *Entry* is at its maximum of one when $PR^{t}_{x+5,x+9} = PR*$. When there is no net entry, *Entry* = 0. If there are negative net entry (positive net exits), then in the extreme it is possible that $Entry \to -\infty$. Forecasts of participation rates based on these entry rates are:

$$PR_{a+5,a+9,t} = Entry_t \times (\max - PR_{a,a+4,t-5}) + PR_{a,a+4,t-5}$$

Figure 3.5 **Exit rates based on smoothed participation rates**[a]
1983-84 to 2003-04

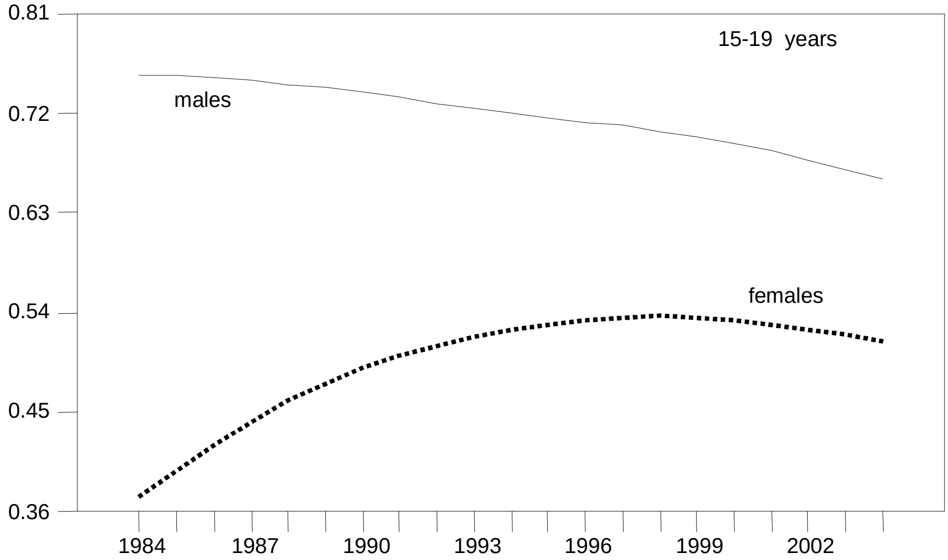


[a]Participation rates were smoothed using a Hodrick Prescott filter. Note that an entry or exit rate for age x mean the rate at which a cohort aged x in year t-5 enter or exit the labour force t years later.

Data source: ABS (*Labour Force, Australia, Detailed*, Cat. no. 6291.0.55.001).

Figure 3.6 **Entry rate based on smoothed participation rates**[a]

15-19 year olds, 1983-84 to 2003-04



a Participation rates were smoothed using a Hodrick Prescott filter. Note that an entry rate for age x mean the rate at which a cohort aged x in year t-5 enters the labour force t years later.

Data source: ABS (*Labour Force, Australia, Detailed*, Cat. no. 6291.0.55.001).

Because both rates can encompass negative values, it would be possible to forecast *either* entry rates *or* exit rates and use just the single measure to project participation rates.

However, there is a major practical obstacle to forecasting just the exit rate or the entry rate alone and then deriving participation rates. This arises because there is no single observed value of exit rates at the sensible upper bound of entry rates and vice versa. For example, a rise in the participation rate for a cohort from 0.93 to 0.945 over a five year period gives an entry rate of 0.747 and an exit rate of -0.0159 (for a PR* of 0.95). However, a rise in the participation rate from 0.945 to 0.96, which violates the condition for PR*, gives the same value for the exit rate, whereas the value for the entry rate clearly discloses that the upper bound has been exceeded (table 3.1). It is even possible that a seemingly sensible value for the exit rate might result in a participation rate that exceeded one, let alone PR*. Similarly, the same indeterminacy arises when predicting participation rates from negative values for entry rates — a seemingly innocuous value for an entry rate may generate even negative participation rates.

Table 3.1 **No single measure of exit or entry gives determinate results**[a]

PR_t	PR_{t-5}	*Entry rate*	*Exit rate*
0.945	0.930	0.747	-0.0159
0.960	0.945	3.000	-0.0159
0.059	0.180	-0.157	0.6700
-0.080	0.059	-0.157	2.3486

a These calculations were made with PR*=0.95.

Source: Commission estimates.

Given these practical issues, the best approach is to model exit rates when long-run exit rates are likely to be positive and to model entry rates when long run entry rates are likely to be positive.[6] This ensures that forecast participation rates are appropriately bounded.

The fact that appropriately defined exit and entry rates must be bounded suggests that linear and log trend models are not appropriate for projection. In order that the long run does not violate bounds, various s-shaped growth curves were considered appropriate. Ultimately, the analysis used Richards' curves (a generalised logistic), since these curves allow for more flexible inflection points than other common s-shaped curves, such as Gompertz or logistic curves (technical paper 2).

3.5 Modelling and projection strategy

For each age, gender and jurisdiction, the following modelling and projection strategy was adopted.

- Participation rates were smoothed using the Hodrick-Prescott filter to reduce noise associated with the business cycle.
- Exit and entry rates were derived for each quinquennial age group for fiscal years from 1983-84 to 2003-04 (noting that estimating these requires data on smoothed participation rates from 1978-79). These were graphed. Where it appeared likely that an entry (exit) rate was likely to be positive in the long run, then the series was modelled as an entry (exit) rate.
- The smoothed series' were differenced and examined for any significant turning points to identify structural breaks in the series. Since an s-shaped curve must

6 It may still be the case that the historical data on which these forecasts are based will have some negative values for exit or entry rates, but so long as the long run rate is positive, this does not present problems for forecasting analysis. Indeed, a modelling approach that omits data points that are negative would generate potentially significant biases.

either show decline or fall throughout its course, any sign changes in the smoothed series reveal breaks. (However, small changes relative to the mean were ignored as statistical noise.) The starting year for estimation occurred after any structural break.

- The smoothed series' were differenced twice to examine acceleration and de-acceleration of participation rates over time. This identified the inflection points in the s-shaped curves (in some case these were forecast with extrapolation methods — but in all cases, plausible inflection points could be identified). These inflection points were used in the estimation of the s-shaped curves.
- Reasonable priors about appropriate maximum or minimum participation rates were formulated, since the non-linear least squares estimation of s-shaped curves can sometimes result in long run rates that are not credible. These limits were used to constrain the estimation results — but in many cases they were not binding. Accordingly, the results largely represent a statistical approach to projection rather than subjective views about how exit and entry rates might evolve.
- Exit and entry rates were modelled as Richards curves. The rates were estimated using non-linear least squares subject to maximum (or where participation rates were falling, minimum) limits on the long run participation rates (technical paper 2).
- The projected entry and exit rates were then used to estimate participation rates by using the formulae above.