CHAPTER 8

'Dead' recovery models

The first chapters in this book focussed exclusively on live encounter 'mark-recapture' models, where the probability of an individual being seen (encountered) on a particular sampling occasion was determined by 2 parameters: the probability the animal survived and the probability that an animal alive in state (r) at time (i) is alive and in state (s) at time ($i + 1$).

In this chapter, we move in a new direction altogether. Recall that 'classic' mark-recapture focuses on the problem of differentiating between (i) not seeing an animal because it is 'dead' (or permanently emigrated from the sample area) and (ii) simply 'missing' it, even though it is alive and in the sample area. In contrast, with 'dead recovery' analysis we are dealing with animals known to be dead (because they are recovered in the 'dead state', frequently in the process of harvest).

Echoing the seminal text by Brownie *et al*. (1985), it is sufficiently important to clearly distinguish between these two broad classes of sampling method (recovery and recapture) that we'll take a moment to elaborate on them. In the case of a recapture analysis, a single marked individual is potentially available for 'multiple encounters' – i.e., the individual may be 'seen' or 'recaptured' on more than one occasion. If you've worked through the preceding chapters of this book, this is entirely obvious to you. In contrast, in a recovery analysis, data are available on only a single, terminal 'encounter' (generally, the recovery event). Unlike recapture data, recovery data are treated as independent, mutually exclusive outcomes (i.e., a marked individual could be recovered in year 1, year 2, or not at all during the duration of the study). While this is a clear difference from a live encounter study, in fact, close examination shows a deep similarity between the two models. The distribution of 'dead recoveries' reflects the realization of a series of probabilistic events. Just as each live encounter in a live encounter history reflects the underlying survival and encounter processes, so too does the distribution of 'dead recoveries'.

8.1. 'Brownie' parameterization

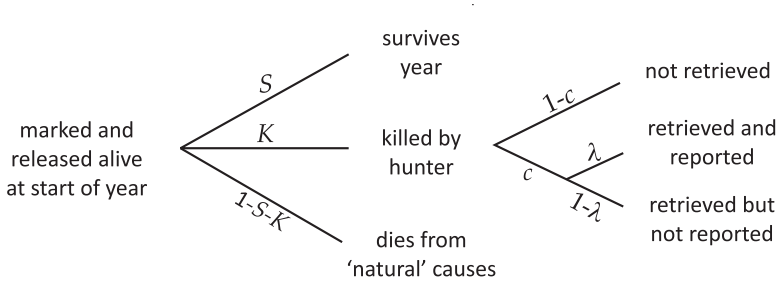
Consider the following example. An individual of a harvested ('exploited') species is marked and released alive. This newly marked individual can then experience one of 3 fates: (1) it can survive the year with some probability, (2) it can be 'harvested' (i.e., some 'action' leading to permanent removal) with some probability, or (3) it can 'die' from 'natural' causes (i.e., it might actually die from some reason other than harvest, or permanently emigrate the sampling area, at which point it appears dead. More on what constitutes the 'sampling area' for a dead recovery analysis in Chapter 9).

However, before this individual becomes 'dead recovery data', something else needs to happen – the 'harvest' needs to be 'reported'. This event reflects several underlying probabilistic events. Suppose you're a waterfowl hunter, and you shoot a bird from your blind (or 'hide' for much of the world). This

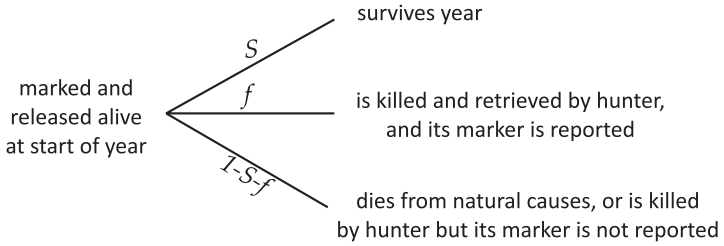
in itself does not constitute a recovery, since simply shooting the bird does not give us the information on who it is (i.e., its identification number). For this to happen, minimally (in most cases), the marked bird needs to be retrieved (i.e., physically handled, typically). Of course, there is some chance it won't be retrieved. If it is retrieved, however, then it might be 'reported' (i.e., the identification number submitted to some monitoring agency), or not. We'll let S equal the probability that the individual survives the year. We separate sources of mortality into 'hunter' and 'natural'. The probability that the individual dies from either source is simply $(1 - S)$. The probability that it dies due to hunting is K. Thus, the probability that it dies from natural causes is $(1 - S - K)$.

Now for the only real complication (that is simple enough in principle, but has several interesting implications we will discuss later in this chapter). Conditional on being shot (i.e., killed by hunting, with probability K), then one of 3 things can happen. The individual may not be retrieved (a fairly common occurrence with some types of harvest – individuals are in fact killed by harvest, but the dead animal is not physically retrieved). The probability of being retrieved is c. Thus, the probability of not being retrieved is $(1 - c)$. Conditional upon being retrieved, the hunter can either report the identification number (with probability λ), or not report the identification number (with probability $1 - \lambda$).

Let's put these probabilities together, using a 'fate diagram' (following Brownie *et al*. 1985).



Thus, recovery data supplies information directly (and directly is the key operative word here) about only those birds which are shot and reported. Thus, under this parameterization, not everything is estimable – only the product $Kc\lambda$ is estimable, but the component probabilities K, c and λ are not. Generally, the product $Kc\lambda$ (often written as $H\lambda$, where $H = Kc$ = harvest rate; the probability of being killed and retrieved by a hunter during the year) is referred to as the recovery rate, f.* Using these 'product' (summary) parameters, we can modify the preceding 'fate diagram' as follows:



Different assumptions about the parameters f and S give rise to the different models. In this sense, you can loosely (very loosely) think of f and S as the equivalents of p and φ for a live recapture analysis –

* We note that neither 'harvest rate' or 'recovery rate' are 'rates' in the strict sense of the word (which implies instantaneous rates of change). Strictly speaking, they should probably be referred to as 'harvest probability' and 'recovery probability', respectively. However, the use of the word 'rate' is traditional for these models.

clearly not in terms of what they represent, but in the fact that the 'encounter history' is defined by these 2 probabilities. Remember, the components of the recovery probability f (i.e., $Kc\lambda$) are not estimable without additional information (discussed later).

Let's see how these two primary parameters (f and S) combine to determine the expected numbers of bands recovered in a particular time period. The process is analogous to expressing the expected numbers of individuals with capture history '101101' as a function of the number released (R) and the underlying survival and recapture probabilities.

Suppose R_1 individuals are marked. How many recoveries are expected during the next year? Note, we're not asking specifically how many individuals are alive at the end of the 12 months following marking (although this can be derived, obviously), but rather, how many individuals will be (i) shot by hunters, (ii) retrieved, and (iii) reported? Look at the fate diagram on the preceding page. The probability that an individual is harvested, retrieved and reported (i.e., the individual is recovered) is simply f. Thus, the expected number of the R_1 released individuals we expect to be recovered in the first interval after marking is given simply as $R_1 f$.

If we assume for the moment that both survival and recovery probabilities are time-specific, then the expected number of recoveries are given as follows:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

Make sure you understand the connection between f, S, and the expected number of recoveries. S_i is the probability of surviving from time $(i-1)$ to time (i), whereas f (recovery rate) is the probability of being shot (i.e., not surviving), and then being retrieved and reported. So, f (recovery rate) combines the mortality event with two other events (retrieval and reporting). For example, for individuals marked in year 1, the number of expected dead recoveries in the second interval after marking is given as $R_1 S_1 f_2$. Why? Well, recall that the recovery parameter f is the probability of the mortality event. In order for the individual to be a dead recovery in the second interval, it has to survive the first interval (with probability S_1), and then be harvested, retrieved and reported (with probability f_2). Note that survival S does not appear on the diagonal.

Now, if you've already worked through the earlier chapters on mark-recapture, in looking at the table of expected number of recoveries (above), you probably recognize right away that there are reduced parameter models which can be fit. The expected recoveries shown in the preceding table reflect the expectations from a time-dependent model $\{S_t\ f_t\}$. Of course, you could fit model $\{S_t\ f_.\}$ – time dependence in survival only, or model $\{S_.\ f_.\}$ – constant survival and recovery probabilities, or a whole host of additional models. For the moment, let's quickly run through how you would fit the following 4 models: $\{S_t\ f_t\}$, $\{S_.\ f_t\}$, $\{S_t\ f_.\}$ and $\{S_.\ f_.\}$.

Historically, a subset of these models have been referred to by generic model names (for example, model $\{S_t f_t\}$ is referred to in Brownie *et al*. (1985) as Model 1). In the following, we note this historical connection – we suggest that in general you use a explicit model naming convention as we've used throughout the book (and as suggested in Lebreton *et al*. 1992). However, it is important to understand the historical naming conventions to allow you to easily read and interpret earlier papers and texts.

For individuals marked as adults, our models (and their corresponding legacy names) are:

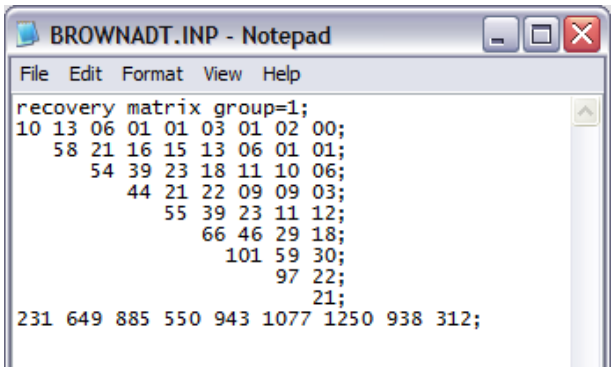
model	*legacy name*	*reference*
$\{S_t\ f_t\}$	Model 1	Brownie *et al*. (1985) pp. 15-20
$\{S_t\ f_.\}$	none	
$\{S_.\ f_t\}$	Model 2	Brownie *et al*. (1985) pp. 20-24
$\{S_.\ f_.\}$	Model 3	Brownie *et al*. (1985) pp. 24-30

You might be wondering about model $\{S_t\ f_.\}$? There is no corresponding model in Brownie *et al*. (1985) because this model (which assumes the recovery probability f is constant over time, while survival S varies) is seldom applicable to the waterfowl data sets for which the model set in Brownie *et al*. (1985) was developed.

8.1.1. Brownie models – PIM approach

To demonstrate how to fit these models using **MARK**, we'll use data set **`brownadt.inp`** (a subset of the **`brownie.inp`** data file distributed with **MARK**). **`brownadt.inp`** contains the recovery data for adult male mallards marked in the San Luis Valley in Colorado, from 1963 to 1971. The full data set (**`brownie.inp`**) contains data for both the adults and juveniles. For the moment, we'll look only at the adults.

Start **MARK**, and begin a new project by pulling down the '**`File`**' menu and selecting '**`New`**'. Select the file **`brownadt.inp`**. Before we go any further, let's have a look at the file. Again, the easiest way to do this is to click the '**`View file`**' button. Here's what **`brownadt.inp`** looks like:

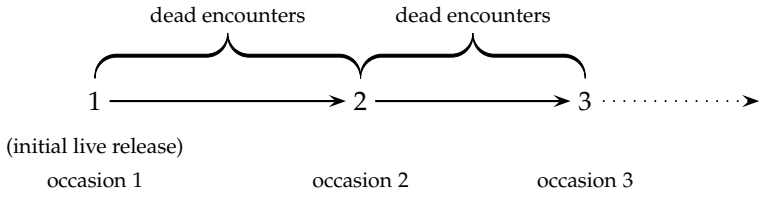


We see that the data are stored in 'classic' recovery matrix form. It is not necessary to format the data this way for a recovery analysis, but it is a traditional summary format. However, remember that using any sort of summary format, whether for a recovery analysis or for (say) mark-recapture analyses has the major disadvantage of not allowing individual covariates (since all individuals are lumped together in the summary). The other approach is to use the familiar encounter history format. **MARK** makes use of what we refer to as the 'LDLD' format to code dead recovery data (and joint live encounter-dead recovery data – this data type is covered in chapter 9). For more details on the 'LDLD' data format, see Chapter 2.

There are 9 'sampling occasions' in this data set, although we submit that occasions is not particularly useful as a reference term, since it is not accurate. In mark-recapture, the occasion is used to refer to the

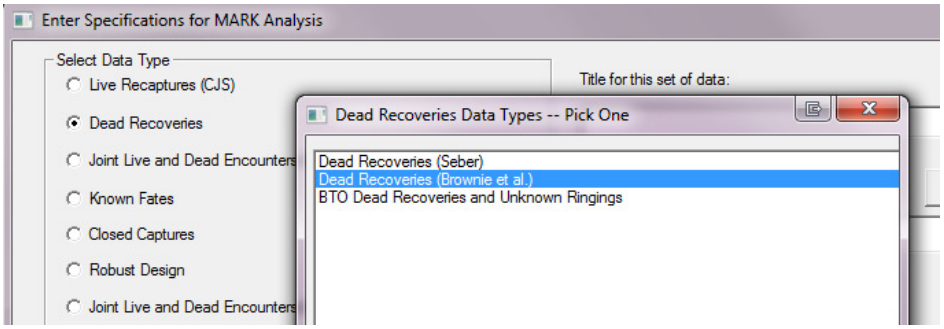
point in time (i.e, the sampling occasion) upon which a marked individual was encountered. Occasions were separated by intervals. In recovery analysis, the data refer to the total number of individuals recovered during the interval, and not at a particular occasion.

Consider the following diagram – individuals are initially marked and live released at (say) occasion 1, but are not encountered alive at any subsequent sampling occasion (2, 3,...). All subsequent encounters of individuals marked and live released at occasion 1 are encountered – once – as a dead recovery, in either period 1 (i.e., interval between occasion 1 and 2), period 2 (interval between occasion 2 and 3), and so on.



Thus, for dead recovery analysis, it is more appropriate to refer to the intervals themselves. In this example, we have 9 years ($l = 9$) of recovery data (as it turns out, ranging from 1963 to 1971). The bottom row indicates the number of newly marked individuals released at the start of each year (note that the year doesn't necessarily start with January 1 – it could be that 'year' refers to the 12-month interval between hunting seasons, for example). So, at the start of what we refer to as 1963, 231 newly marked adult mallards were released. Of these, 10 were recovered during the first 12 months following this release, 13 were recovered the next year, and so forth. Birds were marked and released each year of the study – in other words, there are $k = 9$ rows of recovery data in the data file (i.e., the recovery matrix is symmetric, $k = l$). This becomes important later on, so keep the fact that '$k = l$' in the back of your mind. Set the number of encounter occasions in **MARK** to 9.

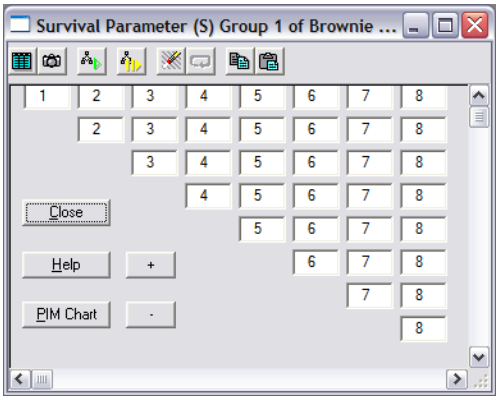
Now we need to select the data type. Remember, **MARK** 'can't tell' the sort of data (or analysis) you are interested in from the data – you have to 'tell it'. Now, if you look at the data type list in the **MARK** specification window, you'll see a radio-button corresponding to '**`Dead Recoveries`**'. If you select this radio-button, a small window will pop up as you to pick a dead recovery data type. Three are listed: '**`Dead Recoveries (Seber)`**', '**`Dead Recoveries (Brownie et al.)`**', and '**`BTO Dead Recoveries and Unknown Ringings`**'.

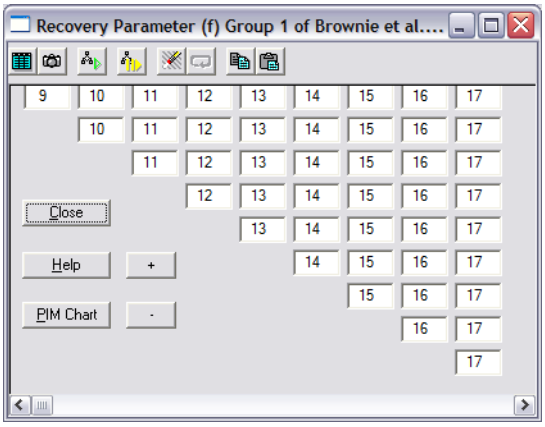


We're starting with the '**`Brownie`**' approach, even though it is not the first one presented in the **MARK** data type menu, simply because it is the 'classic' approach used in the vast majority of published recovery analysis. So, as shown, select the '**`Dead Recoveries (Brownie et al.)`**' data type from the list, and then click the '**`OK`**' button. You should now see the survival (S) PIM on the screen (just as with live encounter

– recapture – data, **MARK** defaults to opening up the 'survival' PIM). However, there are some subtle but important differences between the survival and recovery PIMs, at least when using the Brownie parameterization.

To explore this, let's also open up the recovery (f) PIM for comparison.





Woah – wait a second! These two PIMs don't have the same number of rows and columns – is this a mistake?! No! This is exactly the way it should be. Of course, now you need to consider *why* this is true. Look again at the table of expected recoveries, and the associated probability expressions – below (here, we are considering only 4 years ($l = k = 4$), but the principle is exactly the same):

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

The key is to look carefully at the probability expressions in each cell. Remember that in the case of live mark-recapture, the PIMs are (in effect) constructed from the subscripts of the parameters in the corresponding probability expressions. What about for dead recovery analysis? Look at the subscripting of the two primary parameters, S and f. If you look along the first row (the row with the greatest number of columns – i.e., years), we see that the subscripting for recovery probability f ranges from '1' to '4'. In contrast, we see that the subscripting for survival, S, ranges from '1' to '3' only. Thus, the PIM for S will necessarily be 'smaller' (i.e., reduced dimension) than the PIM for recoveries.

Make sure you understand why – the key is in the first year following the release of newly marked individuals. Consider the first cohort, where R_1 individuals are marked and released. As noted earlier, during that first year after marking and release, the expected number of individuals recovered is $R_1 f_1$ – there is no S term since S denotes survival. An individual cannot survive the interval and also be recovered during the interval (since a recovery implies mortality). The survival term S shows up only in years after the first year following marking (i.e., years 2, 3, 4,...). Why? Again, as noted earlier, this is because in order to be recovered in (say) year 2 after marking, the individual must have survived year 1 (thus, the expected number of recoveries in the second year after marking is $R_1 S_1 f_2$).

With a bit of thought, you might think that these 'asymmetric' PIMs might have implications for which parameters are individually identifiable. You would be correct – more on parameter identifiability in a moment. For now, let's proceed and run this model (we'll call it model '`S(t)f(t)`').

Here are the parameter estimates from fitting this model to the data:

Analysis of adult mallard data

Real Function Parameters of {S(t)f(t)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:S	0.5790621	0.1140631	0.3547412	0.7748865
2:S	0.6109507	0.0780945	0.4519835	0.7493747
3:S	0.6238588	0.0750657	0.4697793	0.7563827
4:S	0.8416860	0.1051312	0.5310696	0.9614770
5:S	0.6384904	0.0735474	0.4860772	0.7673382
6:S	0.5356839	0.0588983	0.4203960	0.6472803
7:S	0.5897947	0.0709194	0.4473406	0.7186258
8:S	0.5593780	0.1363263	0.3003884	0.7896348
9:f	0.0432900	0.0133899	0.0234478	0.0785724
10:f	0.0907043	0.0106739	0.0718427	0.1139101
11:f	0.0594177	0.0068767	0.0472870	0.0744172
12:f	0.0714050	0.0083896	0.0566096	0.0896994
13:f	0.0542005	0.0058384	0.0438320	0.0668502
14:f	0.0662084	0.0062061	0.0550341	0.0794607
15:f	0.0771753	0.0064394	0.0654624	0.0907803
16:f	0.0892152	0.0083208	0.0741953	0.1069247
17:f	0.0673077	0.0141848	0.0442910	0.1010212

We see a couple of things: as expected, there is one more recovery parameter estimated than survival parameters (8 survival, 9 recovery). And, we see that all of the parameters are separately estimable. More on parameter counting, and estimability in a moment.

Where do the estimates come from? While the formal development of the likelihoods for the data are presented in Brownie *et al*. (1985), we can develop an intuitive feel for the estimates by having another look at the expected number of recoveries, assuming time-variation in both survival and recovery:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

Let's focus in on the shaded elements in the column corresponding to recovery year 2. The first lighter shaded element is the number of recoveries of individuals marked in year (cohort) 1 that we *expect* to see in year 2 ($R_1 S_1 f_2$). Because these recoveries occur after the interval following marking, they are usually referred to as *indirect recoveries*.

The second element (slightly darker shading) is the number of recoveries of individuals marked in year (cohort) 2 that we *expect* to see in year 2 ($R_2 f_2$). Because these recoveries occur during the interval

immediately following marking, they are usually referred to as *direct recoveries*.* Note that in terms of the S and f parameters, these two elements differ only in terms of the parameter S_1 – the indirect recoveries are the product of number released and the recovery rate only, whereas the indirect recoveries requiring surviving to later sampling intervals.

But, the actual expected number is also a function of the number of newly marked individuals released (R_1, R_2). If the number of newly marked individuals released in each cohort was the same (i.e., $R_1 = R_2 = R$), the intuitive estimator for S_1 could be derived by taking the ratio of the two expectations:

$$\hat{S}_1 = \frac{R S_1 f_2}{R f_2} = S_1$$

Generally, though, the number of newly marked and released individuals, based on opportunistic samples from the population of interest, will vary from year to year. As a result, we need slightly more complex 'intuition' to understand where the estimates reported by **MARK** come from.

To start, it will help to look at the m_{ij} table (the basic idea was introduced in Chapter 5), for the ith release cohort, and the jth year of recoveries:

		year recovered				
year marked	*number marked*	1	2	3	$j = 4$	total
1	R_1	m_{11}	m_{12}	m_{13}	m_{14}	r_1
2	R_2		m_{22}	m_{23}	m_{24}	r_2
3	R_3			m_{33}	m_{34}	r_3
$i = 4$	R_4				m_{44}	r_4
total		m_1	m_2	m_3	m_4	$\sum_i r_i (= \sum_i m_i)$

where

m_{ij} = the number from the ith release that are recovered in year j (assuming tags are reported in the same year),

m_j = the total number of tags recovered in year j ,

r_i = the total number of tags recovered from the ith release.

Seber (1970) also introduced

z_i = the number of tags recovered after the ith release from the first $(i-1)$ releases that are not recovered in year i, but which are recovered later, $i = 2, 3, \ldots,$

T_i' = $m_i + z_i$, the number of tags recovered after the ith release from the first i releases $(i = 1, 2, \ldots)$.

The key random variables are r_i, m_i, and T_i'.

[While the meaning of the variables r_i and m_i are pretty straightforward, it is perhaps easier to grasp T_i' as the accumulated sum over t recovery samples as $T_i' = \sum_{j=i}^{t} m_{1j} + \sum_{j=i}^{t} m_{2j} + \cdots + \sum_{j=i}^{t} m_{ij}$. In other words, $T_1' = r_1$, and $T_i' = T_{i-1}' + r_i - m_{i-1}$. Thus, $T_2' = T_1' + r_2 - m_1 = r_1 + r_2 - m_1$, and so on.]

From Seber (1970), the probability of mortality in a given year i and having the tag reported (i.e., the recovery probability, f_j) is simply:

* Simply put: *direct recoveries* occur along the diagonal – the interval immediately following marking, while *indirect recoveries* occur off the diagonal – intervals later than the interval following marking.

$$f_i = \Pr[\text{ tag recovered in year } i \,|\, \text{tag recovered after } i\text{th release }] \times \Pr[\text{ tag recovered after } i\text{th release }].$$

Start with the second bit – the overall probability that the tag released in year i will be recovered. Following Seber (1970), we'll call this probability θ_i. The obvious estimate of this probability is simply the proportion of the total number released that are recovered by the end of the experiment: $\hat{\theta}_i = r_i/R_i$.

The other bit – the *conditional* probability that a tag will be recovered in year i, given that it is recovered at all after the ith release – is simply m_i/T_i'. Therefore, we can derive an estimator for f_i as

$$\hat{f}_i = \frac{m_i}{T_i'} \cdot \hat{\theta}_i.$$

What about survival, S_i? With a 'bit of algebra', Seber (1970) showed that

$$\hat{S}_i = \frac{T_i' - m_i}{T_i'} \cdot \frac{\hat{\theta}_i}{\hat{\theta}_{i+1}}.$$

Arguably, not particularly intuitive. But, we can (after a bit more algebra) re-write this expression as

$$\hat{S}_i = \frac{T_i' - m_i}{T_i'} \cdot \frac{\hat{\theta}_i}{\hat{\theta}_{i+1}} = \frac{\hat{\theta}_i - \hat{f}_i}{\hat{\theta}_{i+1}}.$$

Look again at the ratio we considered for $\hat{S}_1$:

$$\frac{R_1 S_1 f_2}{R_2 f_2} = \left(\frac{R_1 f_2}{R_2 f_2}\right) \cdot S_1.$$

In order to suffer mortality in interval 2, you need to survive interval 1. The denominator is the number released in cohort 2, times the recovery probability for interval 2 – in other words, the total number of recoveries from the second release cohort expected in interval 2. The numerator is the expected number of the first release cohort that survive the first interval, but die and are recovered in the second interval. To handle the possibility that $R_1 \neq R_2$, we consider a function of the ratio of the two (in fact, as expressed in the preceding expression).

But, instead of the relative numbers of individuals released, we're interested in the relative proportion of those individual that are recovered (die), and what fraction of those die in the second interval. Look again at

$$\hat{S}_i = \frac{\hat{\theta}_i - \hat{f}_i}{\hat{\theta}_{i+1}}.$$

You should now see the connection (at least at some intuitive level).

OK, enough algebra. Let's see if we can reconstitute the estimates for S_1 and f_1 that **MARK** reports. From a couple of pages back, **MARK** returns $\hat{S}_1 = 0.5790621$, and $\hat{f}_1 = 0.0432900$.

To work out these values 'by hand', we'll need to look at the *m*-array data for this data set:

Occ.	R(i)	Group 1 j= 1	2	3	4	5	6	7	8	9	Total
1	231	10	13	6	1	1	3	1	2	0	37
2	649		58	21	16	15	13	6	1	1	131
3	885			54	39	23	18	11	10	6	161
4	550				44	21	22	9	9	3	108
5	943					55	39	23	11	12	140
6	1077						66	46	29	18	159
7	1250							101	59	30	190
8	938								97	22	119
9	312									21	21

So, in year 1, there were $R_1 = 231$ total individuals newly marked and released, of which a total of $r_1 = 37$ were recovered by the end of the experiment, $m_1 = 10$ of them in the first year. And so on.

Let's start with $\hat{f}_1$. From above,

$$\hat{f}_i = \frac{m_i}{T_i'} \cdot \hat{\theta}_i.$$

From the preceding *m*-array table, $\hat{\theta}_1 = r_1/R_1 = 37/231 = 0.160173$. Since $T_1' = r_1 = 37$, then

$$\begin{aligned}\hat{f}_1 &= \frac{m_1}{T_1'} \cdot \hat{\theta}_1 \\ &= \frac{10}{37}\,(0.160173) \\ &= 0.0432900,\end{aligned}$$

which is what is reported by **MARK**.

For $\hat{S}_1$, recall that

$$\hat{S}_i = \frac{\hat{\theta}_i - \hat{f}_i}{\hat{\theta}_{i+1}}.$$

From the *m*-array, $\hat{\theta}_1 = 0.160173$ (above), and $\hat{\theta}_2 = r_2/R_2 = 131/649 = 0.201849$. So,

$$\begin{aligned}\hat{S}_1 &= \frac{\hat{\theta}_1 - \hat{f}_1}{\hat{\theta}_2} \\ &= \frac{0.1168832}{0.201849} \\ &= 0.5790624,\end{aligned}$$

which again is what is reported by **MARK**. And so on for the other estimates.

Whew! Let's let **MARK** do the 'heavy lifting' going forward (by working with the likelihood), and return to fitting the remaining models in our candidate model set. If you've worked through the preceding chapters of this book, it should be immediately obvious how to fit the other models in our candidate model set (again, the most efficient way is by manipulating the PIM chart).

Go ahead and run the remaining 3 models, and add the results to the browser:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{S(.)f(t)}	8655.1931	0.0000	0.76712	1.0000	10	46.2463
{S(t)f(t)}	8657.8302	2.6371	0.20522	0.2675	17	34.8258
{S(.)f(.)}	8662.2404	7.0473	0.02262	0.0295	2	69.3240
{S(t)f(.)}	8665.2441	10.0510	0.00504	0.0066	9	58.3032

We see clearly that model $\{S_{\cdot}\ f_t\}$ (Model 2, *sensu* Brownie *et al*. 1985) and model $\{S_t\ f_t\}$ (i.e., Model 1, *sensu* Brownie *et al*. 1985) are the 'best' two models out of the four in the model set (since they are clearly better supported by the data than are the other two models). Among these two models, model $\{S_{\cdot}\ f_t\}$ is almost 4 times better supported by the data than is the fully time-dependent model $\{S_t\ f_t\}$.

Using the classical 'model comparison' paradigm, the LRT between these two models confirms the 'qualitative result' from comparisons of the Akaike weights; the fit of model $\{S_{\cdot}\ f_t\}$ was not significantly different from that of model $\{S_t\ f_t\}$ ($\chi^2 = 11.42, P = 0.121$), so we accept model $\{S_{\cdot}\ f_t\}$ as our most parsimonious model, and conclude there is no 'significant' evidence of time-dependence in survival in these data.*

Now we come to the first challenge of the exercise – which we hinted at earlier in the initial discussion of the 'asymmetry' of the PIMs. How are the number of parameters determined? Which parameters are identifiable in each of the models?

8.1.2. Brownie models – the DM approach

In the preceding section, we constructed and fit some fairly standard models to our example data by manipulating the PIM structure underlying the model(s). Here, we demonstrate how you would build the same 4 models, using an approach based on the design matrix (DM). As introduced at length in Chapter 6, and applied in terms of age and cohort models in Chapter 7, in general you are encouraged to build models based on the DM, since it offers the most flexibility in terms of the types of models you might want to build.

In at least one respect, constructing Brownie models bears an apparent similarity with the age-based models introduced in Chapter 7. Recall that for individuals marked as young, subjected to subsequent live encounter, that there was one fewer year in the data for adults in the data than there was for juveniles (because, for individuals marked as young, there are no adults in the first year). Here, we are not dealing with a 'marked as young' data set, but because there is one more recovery survival probability than survival probability. This might suggest that you would approach constructing the DM in much the same way as we did for the age-based model(s) in Chapter 7.

There is an important difference, though. For the age-based models in Chapter 7, we were building the DM for 'young' and 'adults' for the *same parameter* (say, apparent survival, φ). So, we had a common intercept for the parameter, an 'age within sample' grouping variable (column) for 'young' and 'adult', followed by the time and (age × time) interactions. The time codings differed between the 'young' and 'adult' groups.

Here, though, for Brownie models, we have different numbers of years represented between two *different parameters*: survival and recovery. So, no common intercept. Each parameter has its own intercept, and respective time coding. So, in fact, building the DM-based Brownie models equivalent

* You can run a classical LRT – likelihood ratio test – in **MARK** simply by selecting the '**Tests | LR tests**' menu option. This option assumes you will correctly select only nested models, for which the LRT is valid (as discussed in Chapter 4).

to those we constructed in the previous sections using PIMs is comparatively straightforward. For the **brownadt.inp** data set,

Let's start with the most general model: $\{S_t\ f_t\}$. For these data, recall (from above) that we have 9 years of marking and recovery data. So, for model $\{S_t\ f_t\}$, that means 9 recovery parameters ($f_1 \rightarrow f_9$), and 8 survival parameters ($S_1 \rightarrow S_8$). Building the DM for model $\{S_t\ f_t\}$ should be straightforward – 9 recovery parameters, so 8 columns coding for recovery intervals (1 intercept, plus 8 for the time intervals = 9 total columns for recovery), and 8 survival parameters, so 7 columns coding for recovery intervals (1 intercept, plus 7 for the time intervals - 8 total columns for survival).

Here will speed things up by looking at the DM that **MARK** would construct for you, by using the '**Design | Full**' option:

Design Matrix Specification (B = Beta)

B1: S Int	B2: S t1	B3: S t2	B4: S t3	B5: S t4	B6: S t5	B7: S t6	B8: S t7	B9: f Int	Parm	B10: f t1	B11: f t2	B12: f t3	B13: f t4	B14: f t5	B15: f t6	B16: f t7	B17: f t8
1	1	0	0	0	0	0	0	0	1:S	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	2:S	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	3:S	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	4:S	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	5:S	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	6:S	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	7:S	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	8:S	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	9:f	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	10:f	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	11:f	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	12:f	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	1	13:f	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	1	14:f	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	15:f	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	16:f	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1	17:f	0	0	0	0	0	0	0	0

As expected, we see 8 total columns for survival (upper-left), and 9 total columns for recovery (lower–right). Also, note that we don't need to do anything special with respect to the time coding, to indicate that survival parameters are for years $1 \rightarrow 8$, while the recovery parameters are coded for years $1 \rightarrow 9$ – because it is the same 'first year' for both parameters. This was not the case for the age models introduced in Chapter 7, where for individuals marked as young, adults were present in the marked sample starting in the second year of the study (whereas the young individuals were marked in the first year). So, for age models, we had to pay attention to which age classes were present in which years. Not the case here for the Brownie analysis of these sample data.

If we run this model, and add the results to the browser (below) we see (perhaps not surprisingly) that we get the same fit to the data as the same model built using PIMs:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{S(.) f(t) PIM}	8655.1931	0.0000	0.38993	1.0000	10	46.2463	8635.1609
{S(.) f(t) PIM}	8655.1931	0.0000	0.38993	1.0000	10	46.2463	8635.1609
{S(t) f(t) PIM}	8657.8302	2.6371	0.10432	0.2675	17	34.8258	8623.7404
{S(t) f(t) - DM}	8657.8302	2.6371	0.10432	0.2675	17	34.8258	8623.7404
{S(.) f(.) PIM}	8662.2404	7.0473	0.01150	0.0295	2	69.3240	8658.2386

Building the remaining models in our candidate model set ($\{S_t\ f_.\}$, $\{S_.\ f_t\}$ and $\{S_.\ f_.\}$ should be easy enough – simply delete the appropriate time column(s) from this DM.

8.2. Counting parameters – Brownie parameterization

Let's start by having yet another look at the table of expected recoveries for the simpler 4 year study. As structured, this corresponds to model $\{S_t\ f_t\}$ – full time-dependence in both the survival and recovery parameters.

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

How many of these parameters are identifiable? The key to answering this question is to see whether or not there are any 'groups' of parameters that always occur together, and never apart. In the preceding table, we see that no such 'groups' exist – every parameter $(S_1 \rightarrow S_3)$ and $(f_1 \rightarrow f_4)$ occurs either alone or in unique combinations. As such, all 7 parameters are identifiable.

In general, for model $\{S_t\ f_t\}$, the number of identifiable parameters is $2k - 1$ (where k is the number of release cohorts). However, as we'll see in a minute, this isn't always the case.

What about model $\{S_\cdot\ f_t\}$? The probability statements for this model are:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S f_2$	$R_1 S S f_3$	$R_1 S S S f_4$
2	R_2		$R_2 f_2$	$R_2 S f_3$	$R_2 S S f_4$
3	R_3			$R_3 f_3$	$R_3 S f_4$
$k = 4$	R_4				$R_4 f_4$

In this case, all 5 parameters are identifiable – S and $(f_1 \rightarrow f_4)$.

Now, at this point you might be saying 'Gee...in both cases, all the parameters are identifiable...is this always the case?'. If only life were that simple! Consider the situation shown below:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
$k = 3$	R_3			$R_3 f_3$	$R_3 S_3 f_4$

The first notable feature is that $k \neq l$ (i.e., the number of rows in the recovery matrix, $k = 3$, is less than the number of columns – years of the study, $l = 4$). This sort of situation is not uncommon.

Marking individuals can be time consuming, and expensive, but collecting the recovery data is passive, inexpensive (generally), and continues as long as there is hunting, potentially well after marking has ended. In this case, recovery data were collected for $s = 1$ year ($s = l - k$) after the cessation of marking.

begin sidebar

Formatting the recovery matrix when $k \neq l$

When $k \neq l$ (typically when the number of years of marking is less than the number of years over which recovery data are collected – i.e., $k < l$), does this influence the structure if the data `.INP` file?

The answer, as you may recall from Chapter 2, is 'yes'. You need to add '0's for the 'missing elements' of the recovery matrix. For example, if $k = 3$, $l = 5$, the recovery matrix would look like:

$$\begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5; \\ & R_2 & R_3 & R_4 & R_5; \\ & & R_3 & R_4 & R_5; \\ & & & 0 & 0; \\ & & & & 0; \\ R_1 & R_2 & R_3 & 0 & 0; \end{array}$$

end sidebar

Now, if you read Brownie *et al*. (1985), you'd eventually come to a point where you're told

> *'In general, under Model 1 (i.e., $S_t\ f_t$), the parameters $f_1, f_2, \ldots, f_k$ and $S_1, S_2, \ldots, S_{k-1}$ are separately estimable, but if $s > 0$ (where $s = l - k$), only products such as $S_k f_{k+1}, S_k S_{k+1} f_{k+2}, \ldots, S_k S_{k+1}, \ldots, S_{k+s-1} f_{k+s}$ are also estimable, not the individual parameters S_{k+j-1}, and f_{k+j}, $j = 1, \ldots, s$.'*

OK, now to translate – look carefully at the table of probability expressions at the top of this page (for the time-dependent model $\{S_t\ f_t\}$, where $k < l$). We mentioned previously that the key to identifying inestimable parameters is to look for 'groups' of parameters that are never separated. Do we have any in this table? In fact, we do in this case. Notice that the parameters S_3 and f_4 always occur together as the product $S_3 f_4$ (i.e., whenever you find f_4 you always find S_3). So, they are not separately identifiable.

But you might say 'Well, S_2 and f_3 always occur together, as do S_1 and f_2, so are they identifiable?'. The answer in those cases is 'yes', because for those years (3 and 2, respectively), the last element of the column is simply the product of the number released and the recovery probability – no survival term. In contrast, in column 4, every element of the column has the products of the survival and recovery probabilities.

Why does this matter? It matters because it is these final elements of the columns 2 and 3 which allow you to estimate the various parameters. Also, with $k = 3$, columns 1 to 3 correspond to $l = 3$ (i.e., form a symmetrical recovery matrix), and thus all parameters are identifiable. In column 4, this is not the case, since all elements of column 4 contain at least one product in common ($S_3 f_4$).

Thus, in this example, S_1 and f_2 are separately identifiable, as are S_2 and f_3, but only the product $S_3 f_4$ is identifiable, so 5 parameters in total (4 individual, and 1 product). In general, estimates of the products are not of particular interest, since, for example, $S_3 f_4$ is the probability of surviving year 3 and being shot and reported in year 4.

However, non-identifiability can 'vanish' with a reduction in complexity of the model. You may recall this from the mark-recapture chapters, where non-identifiability did not occur in reductions from the

fully time-dependent model. The same is true here. If survival probability S is constant over time, for example, then

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S f_2$	$R_1 SS f_3$	$R_1 SSS f_4$
2	R_2		$R_2 f_2$	$R_2 S f_3$	$R_2 SS f_4$
$k = 3$	R_3			$R_3 f_3$	$R_3 S f_4$

In this case, because estimation of S is based on data from all years, there is no problem on non-identifiability – both S and all of the recovery parameters are estimable.

However, although everything is estimable, Brownie *et al*. (1985) notes that for years $> k$, estimates of recovery probability tend to be poor, because they are based on so few data. So, in this example, f_4 would likely be poorly estimated, since they are based entirely on recoveries from > 1 year after marking.

begin sidebar

Counting parameters in Brownie models: a different approach

If you're still confused about how to determine which parameters are estimable in Brownie models, here is another way of approaching the problem which might be more intuitive. Consider the following example recovery matrix, which is based on 4 release occasions where individuals are newly marked and released live:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

We'll introduce the approach by considering two 'problem' situations – (1) no recoveries in a given year, and (2) no mark-release effort in a given year.

We'll consider the problem of no recoveries in a given year first. For the preceding recovery matrix, the most direct way to get an estimate of S_1 is *algebraically*, by comparing the two cells in column 2 of the recovery matrix (above). You have information on f_2 from direct recoveries (along the diagonal), and information on the product of $S_1 f_2$ (based on the indirect recoveries from the first release cohort). This constitutes two equations in two unknowns, which is easily solved for S_1. If $f_2 = 0$ (as would be the case if there were no recoveries in year 2 of the study), then there is no information on S_1 in column 2.

However, looking at column 3, you can derive an estimate of S_1 from the combination of data from the top two cells in this column, and derive an estimate of S_2 from the combination of data from the bottom two cells in that column, assuming that there were recoveries in year 3. Normally you would not do these things, because S_1 and S_2 are also found in other cells in the model. The Brownie models use all information from all of the cells in the recovery matrix, to maximize precision. However, this 'algebraic' approach at least tells you whether the minimum data necessary for estimation of a particular parameter are available, given the absence of recoveries in one or more years of the study.

A somewhat more difficult problem arises when you also have years where you do not release any animals (as introduced earlier this chapter). In this case you are taking out an entire row of the recovery matrix.

For example, as shown in the following recovery matrix:

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		0	0	0
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$

In this example, we did not release any animals in year 2, and thus an entire row of the recovery matrix is set to 0. In this case, if you use the same algebraic approach described above, you will see that you lose your ability to estimate S_1 and S_2. You don't have direct recovery information on f_2 and therefore cannot use it to extract S_1 from the product $S_1 f_2$. In addition, you lose information on the product $S_2 f_3$, and therefore again cannot use it to algebraically 'solve' for S_2. The best you can do in this case is estimate the product $S_1 S_2$. In general, then, when you do not release animals in year t, you cannot get separate estimates of S_{t-1} and S_t.

end sidebar

Now that we've had a brief look at some of the considerations for counting parameters under the Brownie parameterization, let's return to the adult mallard example we have been working with. At this point, you should be able to figure out why model $\{S_t\ f_t\}$ (for example) has 17 identifiable parameters. Since $(k-l) = 9$, then we have $(k+l-1)$ identifiable parameters: 9 recovery probabilities, and 8 survival rates. For model $\{S_.\ f_t\}$ we have 10 identifiable parameters: 1 survival, and 9 recovery probabilities.

8.3. Brownie estimation: individuals marked as young only

In the preceding mallard example, we noted in passing that the data set consisted entirely of individuals marked as adults. What happens if you face the situation where you have only individuals marked as young? Can you still estimate survival and recovery probabilities? Are all parameters identifiable?

This general question is dealt with thoroughly in Brownie *et al.* (1985), pp. 112-115, and the associated paper by Anderson, Burnham & White (1985), reprinted in full as an Appendix in Brownie *et al.* (1985). These references should be consulted for a full treatment of the problem. Our motive here then is to 'test you' on your ability to determine which parameters are identifiable, and which are not. Paraphrasing Brownie *et al*. (1985), marking of young individuals only is often popular because it is often easier, and less expensive (young are typically easier to catch than adults or sub-adults).

However, in most cases (perhaps even in all cases), survival of young individuals is typically lower than the survival of older age classes, Also, first year (direct) recovery probabilities are typically higher than for older, adult individuals (this is to some degree a logically consistent statement, since some mortality, the complement of survival, is 'hidden' in recovery rate).

Given this, we first need to consider what an appropriate model would be for modeling recoveries from a sample of individuals marked as young. Brownie *et al*. (1985) describe a '`model H1`' as an appropriate model for these sorts of data (pp. 59-62). Its structure is shown at the top of the next page for a situation where $k = l = 4$.

expected recoveries – Model H_1

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f_1^*$	$R_1 S_1^* f_2$	$R_1 S_1^* S_2 f_3$	$R_1 S_1^* S_2 S_3 f_4$
2	R_2		$R_2 f_2^*$	$R_2 S_2^* f_3$	$R_2 S_2^* S_3 f_4$
3	R_3			$R_3 f_3^*$	$R_3 S_3^* f_4$
$k = 4$	R_4				$R_4 f_4^*$

In essence, `model H1` is model $\{S_{t/t}\ f_{t/t}\}$ – an age structured model with 2 age-classes with time-dependence for each class. If you worked through the preceding chapters on mark-recapture (Chapter 7 in particular), you should quickly recognize this structure, at least qualitatively. Along the diagonal, the recovery probabilities (denoted with an asterisk, *) reflect the recovery probabilities for young individuals, whereas the off-diagonal recovery probabilities (no asterisk) refer to recovery probabilities for adult age classes (remember that time, and thus age, increase going from left to right within a cohort – along a row). The survival rates marked with an asterisk (which form an internal diagonal within each column) represent survival during the first year of young individuals.

Now, what (if anything) can be estimated here? With a bit of thought, and looking carefully at the preceding table, you should see that the direct recovery probabilities f^* are estimable (recall that direct recovery probabilities are the recovery probabilities estimated for the first interval following marking). However, without extra information, S^* – the survival probabilities of young over the interval following marking – are not estimable, no matter what simplifying assumptions are made about how the probabilities vary over time.

Remember the trick is to look for parameter 'groups' that always occur together. Consider the following attempt to simplify the structure of this model in an attempt to 'make the parameters identifiable'. Assume that none of the 4 parameters (S, S^*, f and f^*) vary over time (i.e., model $S_{./.}\ f_{./.}$). The structure of this model would be (again assuming $k = l = 4$):

		year recovered			
year marked	*number marked*	1	2	3	$l = 4$
1	R_1	$R_1 f^*$	$R_1 S^* f$	$R_1 S^* S f$	$R_1 S^* S S f$
2	R_2		$R_2 f^*$	$R_2 S^* f$	$R_2 S^* S f$
3	R_3			$R_3 f^*$	$R_3 S^* f$
$k = 4$	R_4				$R_4 f^*$

Note that the parameters S^* and f always occur together as a product. In fact, this demonstrates why, even in this simple model, these two parameters cannot be separately estimated – only the product $S^* f$ can ever be estimated if no adults are marked.

Moral: don't mark only young individuals if you plan on using a dead recovery analysis alone to estimate parameters of interest – it is doomed to fail, and make your analysis quite difficult, if even possible. (Having said that, an approach combining data from dead recoveries and live encounters which can be applied to individuals marked as young only is described in the next chapter. Other

approaches, based on various assumptions and constraints, can be found in Catchpole *et al*. (1996, 1998), and references therein.)

8.4. Brownie analysis: individuals marked as young + adults

One of the unintended (yet important) messages of the preceding section was that recovery analysis of only individuals marked as young is ultimately futile. Of course, you should also understand that this statement is true only for recovery analysis – at least when contrasted to standard mark-recapture analysis, which has no such structural limits.

But, the question remains – how can you get age-specific estimates from a recovery analysis? The answer is, in fact, fairly straightforward – you mark both young and adults, and analyze their recovery data together. The reason we do this (as we'll see in a moment) is that the 'extra information' provided from the adults allows us to estimate some parameters we wouldn't be able to estimate using young alone.

The background for analyzing individuals marked both as young and adults using the Brownie parameterization is found in Brownie *et al*. (1985) – see pp. 56-115. As in Brownie *et al*. (1985), we'll start with a very general model – what is referred to as '`Model H1`' in the Brownie text (which we introduced in the preceding section). `Model H1` assumes (1) that annual survival, reporting and harvest probabilities are year-specific, (2) annual survival and harvest probabilities are age-dependent for the first year of life only, and (3) reporting probabilities are not dependent on the time of release.

As with the preceding discussion on individuals marked as young only, we'll let f_i^* be the recovery probability in year (i) for individuals marked and released as young in year (i). S_i^* will represent the survival rate for year (i) for individuals marked and released as young in year (i). f_i and S_i will represent the adult recovery and survival rates in year (i), respectively. Now, let's examine the structure of `Model H1`, again using a table of the probability expressions corresponding to the number of expected direct and indirect band recoveries.

		year recovered			
Year marked	*Number marked*	1	2	3	$l = 4$
		marked and released as adults			
1	R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
2	R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
3	R_3			$R_3 f_3$	$R_3 S_3 f_4$
$k = 4$	R_4				$R_4 f_4$
		marked and released as young			
1	M_1	$M_1 f_1^*$	$M_1 S_1^* f_2$	$M_1 S_1^* S_2 f_3$	$M_1 S_1^* S_2 S_3 f_4$
2	M_2		$M_2 f_2^*$	$M_2 S_2^* f_3$	$M_2 S_2^* S_3 f_4$
3	M_3			$M_3 f_3^*$	$M_3 S_3^* f_4$
$k = 4$	M_4				$M_4 f_4^*$

For marked adults, the assumptions of `Model H1` are the same as those of `Model 1` (i.e., model $S_t\ f_t$), so the expected recoveries from individuals marked as adults are the same under `Model H1` and `Model 1`.

For individuals marked as young, if M_1 are marked and released in the first year, then on the average we would expect $M_1 f_1^*$ recoveries in the first year after marking, and $M_1 S_1^*$ of the release cohort to survive to adulthood (i.e., to survive the year). At the start of the second year, M_2 new individual young are marked and released. In addition, the $M_1 S_1^*$ survivors from the first release cohort (now adults) are also released. The important thing to remember is that in the second year, these $M_1 S_1^*$ survivors will reflect the adult probabilities f_2 and S_2, giving on average $M_1 S_1^* f_2$ recoveries and $M_1 S_1^* S_2$ survivors. And so on for each successive cohort and recovery year.

From the table of expected recoveries for `Model H1` we see that the off-diagonal elements of the recovery matrix for individuals marked as young provide information about the adult probability parameters:

		year recovered			
Year marked	*Number marked*	1	2	3	$l = 4$
		marked and released as young			
1	M_1	$M_1 f_1^*$	$M_1 S_1^* f_2$	$M_1 S_1^* S_2 f_3$	$M_1 S_1^* S_2 S_3 f_4$
2	M_2		$M_2 f_2^*$	$M_2 S_2^* f_3$	$M_2 S_2^* S_3 f_4$
3	M_3			$M_3 f_3^*$	$M_3 S_3^* f_4$
$k = 4$	M_4				$M_4 f_4^*$

It is the presence of the 'adult' parameters in the off-diagonal cells that can be exploited to provide extra information needed to estimate parameters that might not be estimable otherwise.

Let's now consider `Model H1`. In fact, when you installed **MARK**, you'll find that this model (and 2 others) have already been 'done for you'. During the installation, a set of files named `BROWNIE.xxx` were extracted into the `\examples` sub-directory where **MARK** was installed. Open up `BROWNIE.DBF`. The results shown in the browser were derived by fitting the 3 models listed to the data in **`brownie.inp`**, which are in fact the mallard data from San Luis Valley, California we considered before – only now we're looking at both the recoveries for individuals marked as young and for individuals marked as adults.

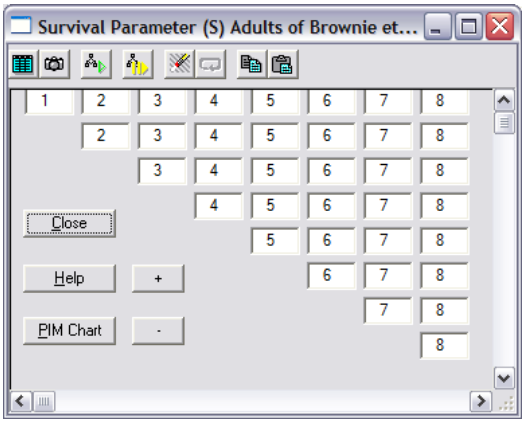
Before we continue, let's have a quick look at the `.INP` file format for these data – how do we put both the adult and young recovery matrices into the same input file? As shown below, it is very simple.

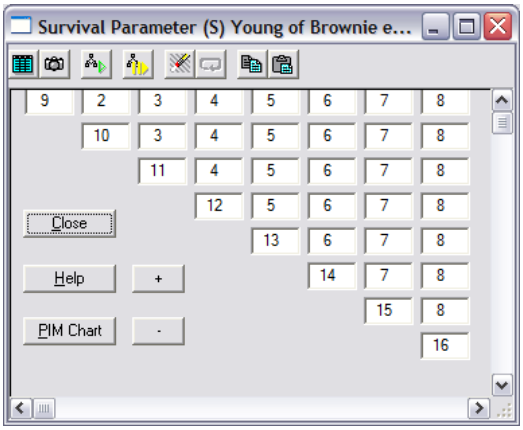
```
INPUT ---     /* San Luis Valley Mallards: Page 92, Brownie et al. 1985
INPUT ---    encounter occasions=9, groups=2 glabel(1)=Adults
INPUT ---    glabel(2)=Young  */
INPUT ---     recovery matrix group=1;
INPUT ---        10 13 06 01 01 03 01 02 00;
INPUT ---           58 21 16 15 13 06 01 01;
INPUT ---              54 39 23 18 11 10 06;
INPUT ---                 44 21 22 09 09 03;
INPUT ---                    55 39 23 11 12;
INPUT ---                       66 46 29 18;
INPUT ---                         101 59 30;
INPUT ---                             97 22;
INPUT ---                                21;
INPUT ---        231 649 885 550 943 1077 1250 938 312;

INPUT ---     recovery matrix group=2;
INPUT ---        83 35 18 16 06 08 05 03 01;
INPUT ---          103 21 13 11 08 06 06 00;
INPUT ---              82 36 26 24 15 18 04;
INPUT ---                153 39 22 21 16 08;
INPUT ---                   109 38 31 15 01;
INPUT ---                      113 64 29 22;
INPUT ---                         124 45 22;
INPUT ---                             95 25;
INPUT ---                                38;
INPUT ---        962 702 1132 1201 1199 1155 1131 906 353;
```

Near the top of the file, you'll see the two recovery matrices, for individuals marked as adults and young, respectively (the order is arbitrary, as long as you remember which one comes first). Note that the two recovery matrices are simply entered sequentially, each one preceded by a '`RECOVERY MATRIX GROUP=n`' statement. That's really all that's needed. The text that is `/* commented */` out is a holdover from the days when these data were run through **BROWNIE** (one of the original programs for running these sorts of data). **MARK** simply ignores the commented out text (as it should).

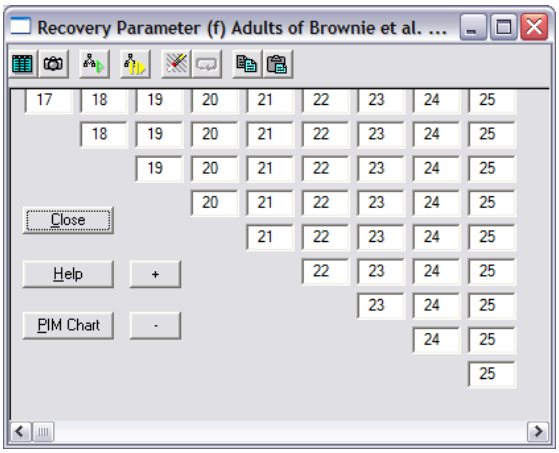
8.4.1. marked as young + adult − the PIM approach

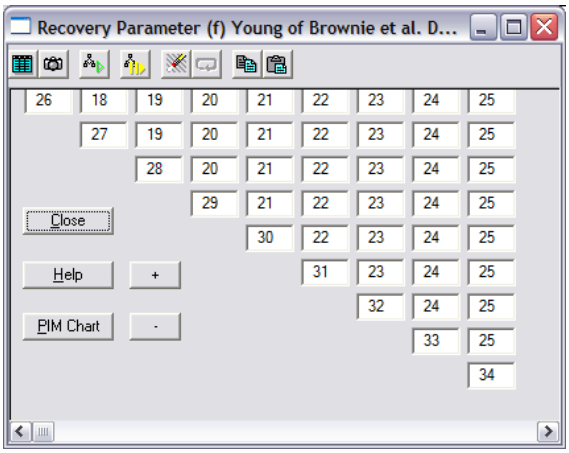
Now let's look at the models themselves. We'll start by constructing the models using PIMs. You might guess from inspection of the table of expected recoveries under `model H1` (shown on the preceding page) that this model is in fact model $\{S_{g*(t,\ t/t)}\ f_{g*(t,\ t/t)}\}$ – two age classes for both parameters for individuals marked as young, with time-dependence in each age class. This is model '`S(a*t)f(a*t)`' in the browser (although we prefer a more informative subscripting). The PIMs are shown below starting with survival, S, for marked as adults and marked as young respectively:





Now, the recovery PIMs, again for marked as adults and marked as young, respectively.





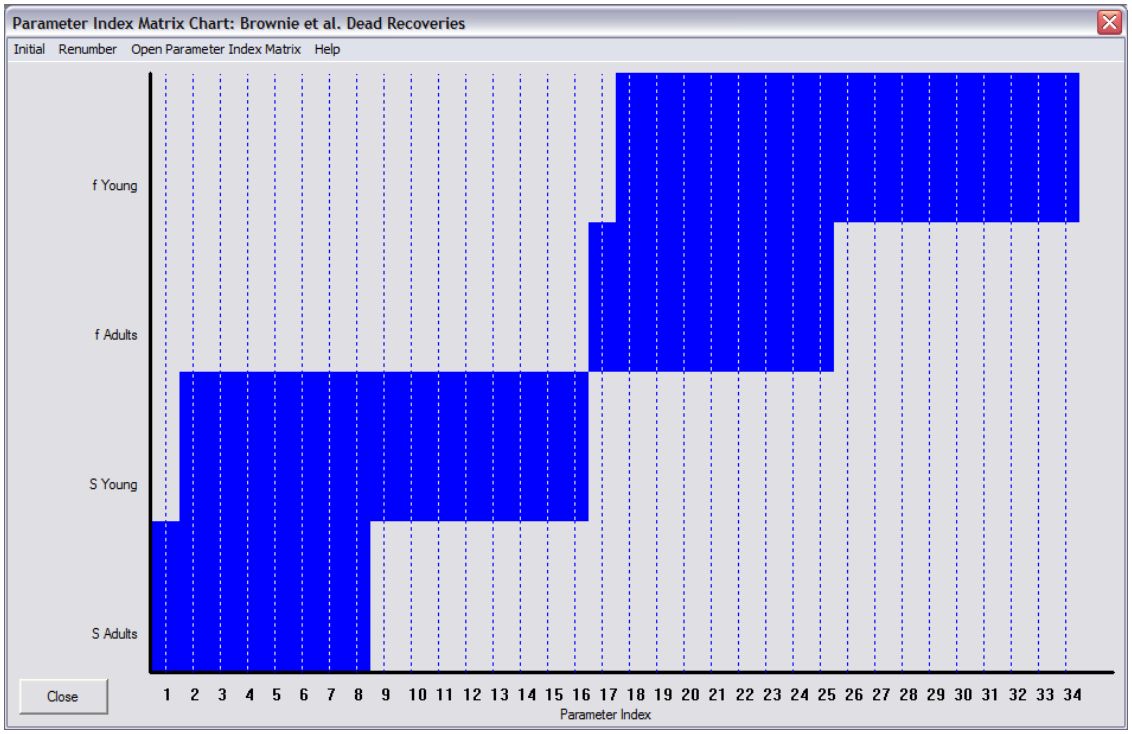
Note that there is no 'age structure' to the adult survival or recovery PIMs. This is because we do not *a priori* expect differences in the direct recovery or survival rate from the indirect probabilities for individuals marked as adults. In contrast, note the age-structure for survival and recovery PIMs for individuals marked as young. Again, the age-structure here is because we believe, *a priori*, that survival (and recovery) in the year following marking (i.e., the direct rates), will differ from the probabilities > 1 year after marking (when the surviving individuals are adults).

However, what is important to note here is that the parameter values appear to overlap. Consider the survival PIMs. For individuals marked and released as adults, it is a simple time-dependent PIM, with parameter indexing from 1 → 8. For individuals marked and released as young, there are 2 age-classes. The indexing for the first age-class (along the diagonal) goes from 9 → 16. However, off the diagonal, the indexing ranges from 2 → 8. In other words, off the diagonal, the indexing for the young individuals is the same as that for the adults. Why? Because off the diagonal, individuals marked as young are adults! Remember, time (= age) within cohort goes left to right. You have actually seen this before – it was discussed in some detail in Chapter 7.

Now, implicit in how the PIMs are indexed is the assumption (in this case) that adult survival S_i does **not** depend on whether or not the individual adult released (or entering) at occasion (i) was originally marked as an adult or not. As we will discuss in the next chapter, this may be a 'strong' (i.e., debatable) assumption in some cases.

What about the recovery PIMs? Again, much the same thing – simple time-dependence for adults (indexing ranging from 17 → 25), and age-structure with time-dependence in both age classes for young (26 → 34 along the diagonal for direct recovery probabilities, and 18 → 25 for the adult age-class).

To get a different (and perhaps more intuitive view) of the overlapping structure of the PIMs, simply take a look at the PIM chart (shown at the top of the next page), where you can see clearly the overlap between the 'marked as adult' and 'marked as young' PIMs:



Now, let's have a look at the results. Close the PIM chart and click on the '**View estimates**' button on the results browser toolbar:

Real Function Parameters of {Brownie - Model H1 - PIM}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:S	0.5790629	0.1140646	0.3547390	0.7748893
2:S	0.6390959	0.0759384	0.4815447	0.7714892
3:S	0.6259796	0.0735886	0.4747625	0.7560327
4:S	0.8689326	0.1050264	0.5210058	0.9758502
5:S	0.6531180	0.0727389	0.5008714	0.7793807
6:S	0.5550087	0.0583565	0.4397464	0.6646420
7:S	0.5761242	0.0668249	0.4428861	0.6991435
8:S	0.5654580	0.1343116	0.3083123	0.7916184
9:S	0.4737907	0.0597857	0.3600966	0.5902691
10:S	0.5089716	0.0702333	0.3740418	0.6426068
11:S	0.5533473	0.0670621	0.4212567	0.6783122
12:S	0.5944923	0.0720385	0.4493908	0.7247742
13:S	0.4801956	0.0612913	0.3634231	0.5991729
14:S	0.6550464	0.0726277	0.5028295	0.7809608
15:S	0.4669478	0.0682746	0.3384915	0.5999412
16:S	0.4099653	0.1183309	0.2103341	0.6444417
17:f	0.0432901	0.0133899	0.0234478	0.0785724
18:f	0.0855839	0.0091885	0.0692084	0.1053955
19:f	0.0590014	0.0061065	0.0481101	0.0721713
20:f	0.0673596	0.0071992	0.0545480	0.0829165
21:f	0.0520411	0.0050400	0.0430036	0.0628532
22:f	0.0632710	0.0054977	0.0533135	0.0749411
23:f	0.0789097	0.0060517	0.0678340	0.0916160
24:f	0.0888059	0.0080390	0.0742601	0.1058752
25:f	0.0673077	0.0141850	0.0442909	0.1010215
26:f	0.0862786	0.0090525	0.0701094	0.1057528
27:f	0.1467236	0.0133545	0.1224315	0.1748752
28:f	0.0724382	0.0077043	0.0587142	0.0890664
29:f	0.1273938	0.0096208	0.1096979	0.1474715
30:f	0.0909091	0.0083023	0.0758945	0.1085452
31:f	0.0978355	0.0087418	0.0819885	0.1163572
32:f	0.1096375	0.0092903	0.0927101	0.1292154
33:f	0.1048565	0.0101784	0.0865161	0.1265462
34:f	0.1076487	0.0164962	0.0793236	0.1445008

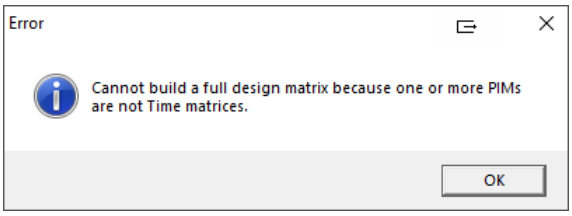
Note from the browser that 34 parameters were estimated for this model (`Model` H_1). If you look at the PIMs, you'll see that this is the total number of parameters in the structure of the model. And, the estimates (above) show that indeed, all 34 parameters are estimated.

Thus, under `Model` H_1, when $k = l$ (as it does in this example), all the parameters are estimable – including the young survival and recovery probabilities. Clearly, this is a significant improvement over the case using only individuals marked as young, where essentially nothing was estimable!

8.4.2. marked as young + adult − the DM approach

If you look at the PIM chart on the preceding page, the overlap of the PIMs for the adult age classes for both survival and recovery parameters might remind of 'similar' overlapping structures introduced in Chapter 7 (in particular, section 7.3.2). However, we need to remember that for this problem, we really do have 2 distinct groups: marked as young, and marked as adults. As such, the DM is quite straightforward.

As we did a preceding example where we analyzed recovery data from 'marked as adults' only, we might try to make use of the '**`Design | Full`**' option. However, if you do, **MARK** will respond with an 'error message':



The message is self-explanatory – **MARK** cannot automatically construct the 'full' design matrix unless the PIM is 'fully time-specific'. Here, the PIMs for the 'marked as young' individuals (shown a couple of pages back) have 'age structure' (i.e., are not fully time-specific): the indexing along the first diagonal differs from the corresponding indexing off the diagonal.

So, for this problem, you'll need to build the DM by hand. Simply select '**`Design | Reduced`**', accept the default 34 parameters, and proceed from there. Construction should be straightforward. Start with the linear model for survival, S. Let the variable '`AM`' represent the 'age of marking' (young, or adults):

$$\begin{aligned}\text{logit}(\hat{S}) &= \texttt{AM + TIME + AM.TIME} \\ &= \beta_1 + \beta_2(\texttt{AM}) + \beta_3(\texttt{t}_1) + \beta_4(\texttt{t}_2) + \beta_5(\texttt{t}_3) + \beta_6(\texttt{t}_4) + \beta_7(\texttt{t}_5) + \beta_8(\texttt{t}_6) + \beta_9(\texttt{t}_7) \\ &\quad + \beta_{10}(\texttt{AM.t}_1) + \beta_{11}(\texttt{AM.t}_2) + \beta_{12}(\texttt{AM.t}_3) + \beta_{13}(\texttt{AM.t}_4) + \beta_{14}(\texttt{AM.t}_5) + \beta_{15}(\texttt{AM.t}_6) + \beta_{16}(\texttt{AM.t}_7).\end{aligned}$$

Our linear model has 16 β parameters, matching the 16 parameters shown in the survival PIMs a couple of pages back.

Here is the completed DM corresponding to the survival parameter (S_i) for model H_1:

B1 S-int	B2 age marked	B3 t1	B4 t2	B5 t3	B6 t4	B7 t5	B8 t6	B9 t7	B10 a't1	B11 a't2	B12 a't3	B13 a't4	B14 a't5	B15 a't6	B16 a't7
1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0
1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0
1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0
1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0
1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0
1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Building the linear model and corresponding DM for the recovery parameter, f, follows the same process – except you need to remember that there is one more recovery parameter than survival

parameter. So, 18 total parameters for recovery, compared with the 16 total parameters for survival, for a total of (16 + 18) = 34 total parameters.

Once you've built the DM, run the model and add the results to the browser. What do we see?

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Brownie - Model H1 - PIM}	20718.5541	0.0000	0.99947	1.0000	34	68.4924
{Brownie - Model H1 - DM}	20733.6242	15.0701	0.00053	0.0005	34	83.5625

Hmmm – we see that the DM-based model is estimating only 33 parameters, instead of 34. So, we might decide to simply 'correct' the parameter count, by manually changing the number of parameters to 34.

In fact, for this problem, manually changing the number of parameters won't help. Look closely at the model deviance. It is different than the corresponding value for the PIM-based model. If the DM-based model is equivalent to the PIM-based model, they should be the same. In general, if the deviances are the same, but the number of parameters differ, then the reason is probably a parameter (or two) that are simply poorly estimated. On the other hand if as in this case the deviances are different, then there is a more fundamental issue.

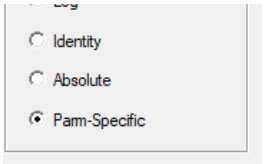
If you look at the estimates from the DM-based model, you'll get a clue as to what the problem is – look at the highlighted estimate for parameter 14 (below):

9:S	0.4737896	0.0597857	0.3600955	0.5902681
10:S	0.5089726	0.0702332	0.3740428	0.6426077
11:S	0.5533472	0.0670619	0.4212570	0.6783117
12:S	0.5944921	0.0720383	0.4493911	0.7247736
13:S	0.4801953	0.0612912	0.3634230	0.5991723
14:S	1.0000000	0.0000000	1.0000000	1.0000000
15:S	0.4669467	0.0682741	0.3384913	0.5999393
16:S	0.4099703	0.1183306	0.2103385	0.6444449

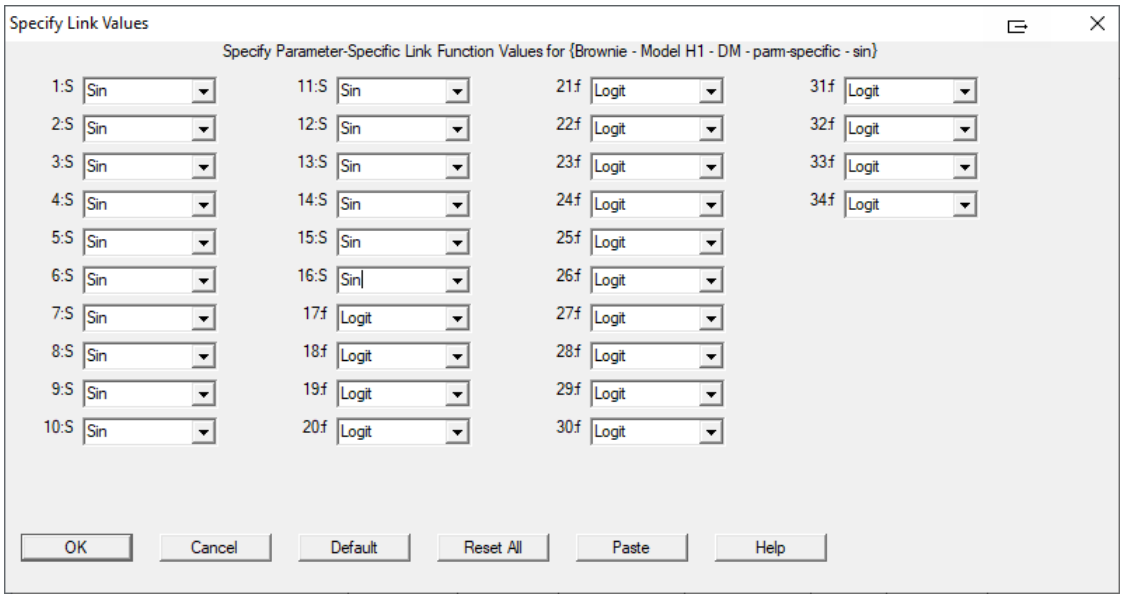
We see that parameter 14 is estimated 'at the boundary', 1.0. When we look at the estimates from the PIM-based model, parameter 14 was estimated as 0.6550464.

Why the difference? if you look at the structure of the model, you'll see that parameter 14 is an 'interior' parameter. Meaning, it doesn't occur at the beginning or end of the time series of data. As such, the reason the parameter is not well-estimated using the DM approach is not because of the structure of the DM itself (i.e., you won't solve the problem simply by changing the reference cell coding – see discussion of the dipper example in Chapter 6).

The alternative step to take is to change the link function from the default logit used with a DM. For example, you might try to enforce a sin link for (say) the survival parameters. How, since the sin link is 'greyed out' when you build a DM-based model. You can in fact 'over-ride' the default logit link, and select any link you want, by specifying a '**Parm-specific**' link function before running the model:



When you select this option, **MARK** will present you with a window allowing you to set the link function for any individual parameter. Here, we have specified the sin link for the survival parameters, and left the logit link applied to the recovery parameters:



If you fit this modified '**`Parm-specific`**' model to the data, you'll see that the problem seems to have been solved:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Brownie - Model H1 - PIM}	20718.5541	0.0000	0.49987	1.0000	34	68.4924
{Brownie - Model H1 - DM - parm-specific - sin}	20718.5541	0.0000	0.49987	1.0000	34	68.4924
{Brownie - Model H1 - DM}	20733.6242	15.0701	0.00027	0.0005	34	83.5625

We now have the same parameter counts, and (more importantly) the same model deviance as compared to the PIM-based model. And, the estimate for parameter 14 from this 'tweaked' model (using the sin link for the survival parameters) is 0.6550472, which is almost identical to the estimate from the PIM-based model.

OK, that seems to solve the problem. But, why did we have the problem in the first place? The answer lies in the fact that Brownie parameterized models can be somewhat 'twitchy' to fit. For example, note that the recovery parameter, f, is a probability that includes both 'death' (the mortality event) and 'reporting' (that the dead individual will be retrieved, and the mark reported/recorded). Because an animal cannot experience both recovery and survival in the same interval (think about it for a moment), then the parameters f and S are implicitly related as $f \leq (1 - S)$. OK, fine. However, there is nothing about the structure of Brownie parameterized models that numerically 'enforces' or imposes this relationship, which leads to the possibility that estimates of f and S that are logically impossible. While that does not seem to be the case here, it is indicative of the challenges you might run into with Brownie parameterized models.*

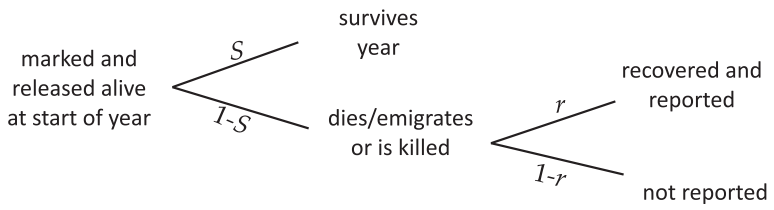
* In fact, for this particular problem, the underlying reason is 'weirdness' in the shape of the likelihood function for this parameter, that causes the logit link to 'fail'.

8.5. A different parameterization: Seber (*S* and *r*) models

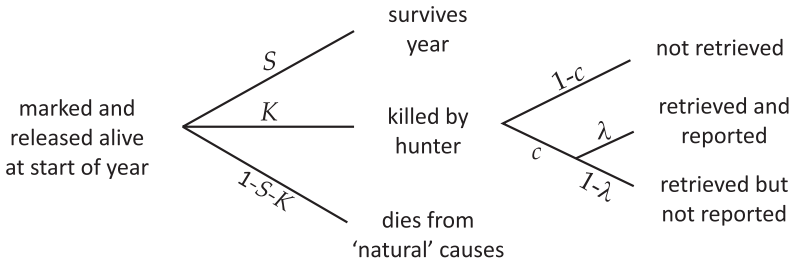
In this section we turn our attention to a rather different approach to the same questions, and the same data type – recoveries, but using a different parameterization, first described by Seber (1970) and later by Anderson *et al*. (1985) and Catchpole *et al*. (1995). In the Brownie parameterization, marked individuals are assumed to survive from one release to the next with survival probability S_i. Individuals may die during the interval, either due to hunting or due to 'natural' mortality. Individuals dying due to hunting (with probability K_i) may be retrieved and reported with some probability (c_i and λ_i, respectively). In contrast, individuals dying from natural causes are generally not retrieved and reported (in other words, the Brownie parameterizations assumes that dead encounter data reflect exploitation of the population).

Here, however, we introduce a new parameter r_i, for recovery probability, defined as the probability that dead marked individuals are reported during each period between releases, and (most generally) where the death is not necessarily related to harvest. Note that the recovery parameter r_i we're talking about here is **not** the same as the Brownie recovery probability f_i, which is the probability of being harvested, retrieved and reported during the period between releases.

The basic structure of the Seber parameterization is shown in the following diagram. Under this parameterization, a marked individual either (i) survives (with probability $S \rightarrow$ encounter history '10'), (ii) dies and is recovered and reported (with probability $(1 - S)r \rightarrow$ encounter history '11'), or (iii) dies and is not reported (either because it was not retrieved, or if retrieved, not reported), with probability in either case of $(1 - S)(1 - r) \rightarrow$ encounter history '10'.



Before we look at how to implement the Seber parameterization in **MARK**, let's take a moment to compare this parameterization with the Brownie parameterization we looked at earlier. First, clearly there must be some logical relationship between r and f. Recall that in the Brownie parameterization,



Consider the encounter history '11'. Under the Brownie parameterization, the expected probability of this event is $Kc\lambda$, which is traditionally referred to as f, the recovery rate ($f = Kc\lambda$). Under the Seber parameterization, the probability of the same '11' encounter history is $r(1 - S)$.

Since f and $(1-S)r$ are equivalent in 'probability space' (i.e., they both correspond to the same 'event' – e.g., $\text{P('11')}_{\text{Brownie}} = f$, $\text{P('11')}_{\text{Seber}} = (1-S)r$), we can write:

$$\because f_i = (1-S_i)r_i \qquad \therefore r_i = \frac{f_i}{(1-S_i)}.$$

As such, we can derive the expected cell probability expressions under the Seber parameterization simply by substituting $f_i = r_i(1-S_i)$ into the probability expressions for the Brownie parameterization:

Brownie

	year recovered			
number marked	1	2	3	$l = 4$
R_1	$R_1 f_1$	$R_1 S_1 f_2$	$R_1 S_1 S_2 f_3$	$R_1 S_1 S_2 S_3 f_4$
R_2		$R_2 f_2$	$R_2 S_2 f_3$	$R_2 S_2 S_3 f_4$
R_3			$R_3 f_3$	$R_3 S_3 f_4$
R_4				$R_4 f_4$

Seber

	year recovered			
number marked	1	2	3	$l = 4$
R_1	$R_1 r_1(1-S_1)$	$R_1 S_1 r_2(1-S_2)$	$R_1 S_1 S_2 r_3(1-S_3)$	$R_1 S_1 S_2 S_3 r_4(1-S_4)$
R_2		$R_2 r_2(1-S_2)$	$R_2 S_2 r_3(1-S_3)$	$R_2 S_2 S_3 r_4(1-S_4)$
R_3			$R_3 r_3(1-S_3)$	$R_3 S_3 r_4(1-S_4)$
R_4				$R_4 r_4(1-S_4)$

The preceding illustrates the algebraic and conceptual connection between the two parameterizations. In simplest terms, the parameter r_i is a reduced parameter – and is a function of two other parameters normally found in the Brownie parameterization. But, more pragmatically, what is the impact of the two parameterizations? Why use one over the other, or does it matter?

One of the motives for considering the Seber parameterization (using only S_i and r_i) is so that the encounter process can be separated from the survival process, entirely analogous to 'normal' mark-recapture. With the Brownie parameterization, the 2 processes are part of the f_i parameter (i.e., there is 'some survival' and 'some reporting/encounter' information included in recovery rate). As such, developing certain advanced models with **MARK** (by modifying the design matrix) is difficult, even illogical (on occasion) using the Brownie parameterization. For example, the Brownie parameterization does not lend itself to modelling of survival with covariates, as it is unclear how to model the survival portion of the f parameter with the same relationship as is used in the S parameter (which might be something you'd like to do).

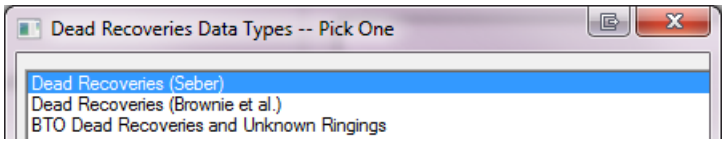
So, given the preceding, we should drop Brownie and use the Seber parameterization right? Well, perhaps not quite. First, under the Seber parameterization the last S_i and r_i are confounded in the time-dependent model, as only the product $(1-S_i)r_i$ (this is entirely analogous to the confounding of the final $\varphi_i p_{i+1}$ term in fully time-dependent model for live encounter CJS analysis). This has some implications for comparing and contrasting survival estimates for some constrained models – we'll deal with this in a moment.

Second, all of the parameters under the Seber parameterization are bounded [0, 1], which outwardly might seem like a benefit. However, parameter estimates at the boundary do not have proper estimates of the standard errors. The Brownie parameterization overcomes both these technical difficulties (for details, see the Brownie text).

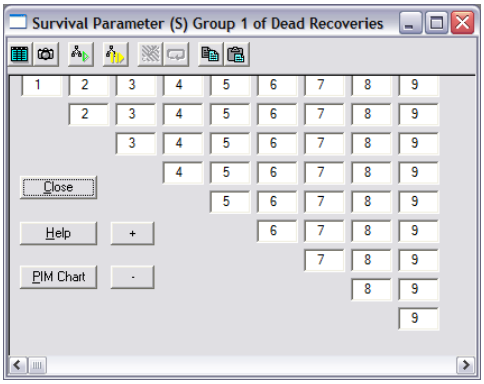
But finally, and perhaps more importantly (at least for exploited populations), the Seber parameterization does not allow you to separate 'harvest' mortality from 'natural' mortality, whereas the Brownie parameterization does. The Seber parameterization basically deals with 'mortality' as a whole, with no partitioning possible. In many cases, this can be an important limitation that you need to be aware of.

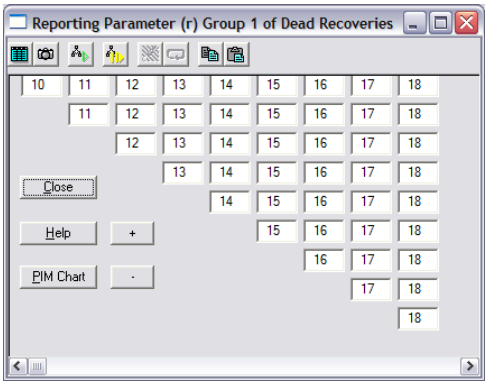
For the moment, though, we'll leave the comparison of these two models (and their respective pros and cons) for you to explore, and will concentrate on showing how to implement the reduced parameterization in **MARK**. In fact, if you've understood the way in which we applied the Brownie parameterization in **MARK**, you'll find this new approach very easy.

We'll demonstrate this using the **`brownadt.inp`** data set we analyzed earlier in the chapter. To specify the new parameterization, select '**`Dead recoveries (Seber)`**':



Look at the PIMs for the two parameters under the fully time-dependent model:





Note that unlike the Brownie parameterization, there are the same number (in absolute terms) of parameters (9) for each ($S_1 \rightarrow S_9$ and $r_1 \rightarrow r_9$).

Since the parameterization is analogous to 'normal' mark-recapture, then the question identifiability of parameters should pose no significant challenges. For example, for the fully time-dependent model $\{S_t r_t\}$ with 9 occasions, we expect 17 estimable parameters: $S_1 \rightarrow S_8$ and $r_1 \rightarrow r_8$, and the final product $r_9(1 - S_9)$. If you run this model in **MARK**, you'll see that 17 parameters are estimated.

8.5.1. Seber vs. Brownie estimates in constrained models: careful!

In the preceding section, we noted that under the Seber parameterization, the last S_i and r_i are confounded in the time-dependent model (analogous to the confounding of the final $\varphi_i p_{i+1}$ term in fully

time-dependent model for live encounter analysis). Recall that in live encounter models, the estimate of survival over the final interval can be obtained if the encounter probability on the last occasion is known. We saw that in models where survival varied over time, but encounter probability was held constant (i.e., $\{\varphi_t\ p_.\}$), all of the survival values were estimable, since a common, constant value for p was estimated for all occasions, including the terminal occasion.

However, while it would seem reasonable to use the same logic for recovery analysis using the Seber parameterization, care must be exercised – especially if you're comparing estimates from a model based on the Seber parameterization with those from a Brownie parameterization. Why? Simple – because the number of survival parameters estimable using the Brownie parameterization is always one less than the Seber parameterization! As such, comparing estimates from a model parameterized using the Seber parameterization can, for some models, be quite different than those from seemingly equivalent models parameterized using the Brownie parameterization.

This can be easily demonstrated by means of a numerical example. Consider the analysis of a simulated data set, 8 occasions, where $K = 0.2$, $c = 1.0$, and $\lambda = 0.4$. In other words, the probability of being harvested over a given interval is 0.2, probability of the harvested individual being retrieved is 1.0, and the probability that the harvested, retrieved individual is reported is 0.4. We'll assume all 3 parameters are constant over time. Under the Brownie parameterization, the recovery probability is $f = Kc\lambda = (0.2)(1.0)(0.4) = 0.08$.

Given these values, what is the recovery probability r under the Seber parameterization? Recall that

$$f_i = r_i(1 - S_i) \quad \text{and} \quad r_i = \frac{f_i}{(1 - S_i)}.$$

So, given f from the Brownie parameterization, then we can solve for r provided we have an estimate of S. Since $K = 0.2$, we know that survival probability is at least $(1 - 0.2) = 0.8$. However, this value is derived assuming the only source of mortality is harvest. What if there is some level of natural mortality, say $E = 0.1$? If we assume that harvest and natural mortality events are independent (i.e., temporally separated, or additive), then $S = (1 - K)(1 - E) = (0.8)(0.9) = 0.72$. So, given $f = 0.08$, and $S = 0.72$, then

$$\begin{aligned} r_i &= \frac{f_i}{(1 - S_i)} \\ &= \frac{0.08}{(1 - 0.72)} = 0.286. \end{aligned}$$

We'll assume no age structure, and 5,000 newly marked individuals on each occasion – the recovery data (in LD format) are contained in the file **seber-brownie.inp**.

We'll start our analysis by specifying the Brownie data type in the data type specification window. If we examine the default starting PIMs for the two parameters, we see that the survival PIM has 7 columns (corresponding to parameters $S_1 \rightarrow S_7$), while the recovery PIM has 8 columns (corresponding to parameters $f_1 \rightarrow f_8$). We run this model (i.e., model $\{S_t\ f_t\}$), and add the results to the browser.

Then, by modifying the PIMs, we construct a 'constrained' model, $\{S_t\ f_.\}$, where survival is allowed to vary over time, while the recovery probability is constant (remember – recovery probability under the Brownie parameterization includes information about mortality, since it is the product of kill probability K with the retrieval and reporting parameters c and λ, respectively. As such, a model where recovery probability is held constant, but where survival is allowed to vary over time has likely implications for how 'other sources of mortality' must vary). Model $\{S_t\ f_.\}$ then has only one recovery estimate, but the same 7 estimates for survival. So, constraining recovery f to be constant over time does not change the number of survival parameters S which are estimable.

The results for both models are shown below. We see that model $\{S_t\ f_t\}$ has 15 estimable parameters (8 recovery parameters + 7 survival parameters), whereas model $\{S_t\ f_.\}$ has only 8 estimable parameters (1 recovery parameter + 7 survival parameters).

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Brownie - S(t)f(.)}	61125.7146	0.0000	0.99063	1.0000	8	24.5108
{Brownie - S(t)f(t)}	61135.0370	9.3224	0.00937	0.0095	15	19.8253

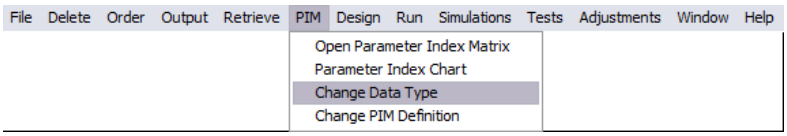
If we look at the parameter estimates from model $\{S_t\ f_.\}$

Real Function Parameters of {Brownie - S(t)f(.)}

rameter	Estimate	Standard Error	95% Confidence Lower	Upp
1:S	0.7065157	0.0252937	0.6546249	0.7535
2:S	0.7331216	0.0247148	0.6819903	0.7786
3:S	0.7238303	0.0240064	0.6744016	0.7683
4:S	0.6997713	0.0233873	0.6520410	0.7435
5:S	0.7051816	0.0242066	0.6556315	0.7503
6:S	0.6946485	0.0252099	0.6431340	0.7417
7:S	0.7389458	0.0289781	0.6783209	0.7916
8:f	0.0805426	0.0011235	0.0783678	0.0827

we see that the survival estimates are all fairly close to the true value of 0.72 (recall that in the true model under which the data were simulated, the true value for survival did not vary over time), and the estimated recovery probability is very close to the true value of 0.08. This is perhaps not surprising given the size of the data set, and that our fitted model is fairly close to the true model underlying these simulated data.

OK – fine. But now let's fit these same data using the Seber parameterization. We can do this easily in **MARK** by changing the data type from '**Brownie**' to '**Seber**'. We do this by selecting '**Change Data Type**' from the PIM menu:



and selecting the '**Dead recoveries**' data type from the list.

Now, if we examine the PIMs for the fully time-dependent model (i.e., model $\{S_t\ r_t\}$), we see that the PIMs for both parameters have 8 columns, corresponding to 8 parameters for survival, and 8 parameters for reporting rate, respectively. However, we also recall that the final two estimates of survival and reporting probabilities are confounded under the Seber parameterization, so we in fact have only 15 estimable parameters in this model. Fit this model to the data, and add the results to the browser:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Brownie - S(t)f(.)}	61125.7146	0.0000	0.98144	1.0000	8	24.5108
{Brownie - S(t)f(t)}	61135.0370	9.3224	0.00928	0.0095	15	19.8253
{Seber - S(t)r(t)}	61135.0370	9.3224	0.00928	0.0095	15	19.8253

Notice that models $\{S_t\ f_t\}$ (Brownie) and $\{S_t\ r_t\}$ (Seber) have **exactly** the same model deviances (19.8253), and number of estimated parameters (15). And, not surprisingly perhaps given this, you'll see that the estimates of survival from the Seber model are *identical* to those from the Brownie model, for the first seven estimates; the final estimate of survival from the Seber model is confounded with the final estimate of the reporting rate.

OK – so far, it seems as if the two parameterizations are equivalent. But now, let's try model $\{S_t\ r_.\}$ using the Seber parameterization. Recall that for this model, there are 9 estimable parameters (8 survival estimates + 1 reporting probability estimate). In contrast, for the 'equivalent' Brownie model $\{S_t\ f_.\}$ there are only 8 estimable parameters (7 survival estimates + 1 recovery probability estimate). So, unlike the case where we contrasted the fully time-specific models between the two parameterizations, here, the actual number of estimable parameters differs between the two models. This should suggest fairly strongly that these are not equivalent models.

As we can see after adding the results to the browser:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Brownie - S(t)f(.)}	61125.7146	0.0000	0.85011	1.0000	8	24.5108
{Seber - S(t)r(.)}	61129.4125	3.6979	0.13381	0.1574	9	26.2085
{Brownie - S(t)f(t)}	61135.0370	9.3224	0.00804	0.0095	15	19.8253
{Seber - S(t)r(t)}	61135.0370	9.3224	0.00804	0.0095	15	19.8253

that models {Seber - $S_t\ r_.$} and {Brownie - $S_t\ f_.$} are **not** equivalent; they have different deviances, and different numbers of estimable parameters.

If we compare our reconstituted parameter estimates from the Seber model (below) with those from the 'equivalent' Brownie model (preceding page),

Real Function Parameters of {Seber - S(t)r(.)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:S	0.7142337	0.0122936	0.6895424	0.7377083
2:S	0.7132212	0.0095683	0.6941071	0.7316028
3:S	0.7115103	0.0088103	0.6939398	0.7284669
4:S	0.7160142	0.0087768	0.6985020	0.7328982
...+....1....+....2....+....3....+....4....+....5....+....6....+....7.]..+....8....+				
5:S	0.7109462	0.0092731	0.6924377	0.7287771
6:S	0.7064264	0.0101995	0.6860455	0.7260126
7:S	0.7178115	0.0111038	0.6955484	0.7390574
8:S	0.7122758	0.0125209	0.6871256	0.7361805
9:r	0.2805204	0.0047519	0.2713017	0.2899277

we see that all of the survival estimates differ between the two models. (Note that the reporting probability estimate is fairly close to the true value of 0.286).

Now, in this particular example, you might suspect that the differences in the survival estimates between the two models (Brownie versus Seber) are 'not that big'. In fact, the relative 'closeness' of the estimates in this example owes more to the fact that the simulated data set is very large, and the underlying (generating) model is very simple (no time variation in any of the parameters).

To demonstrate this more graphically, let's reanalyze the recovery data for adult male mallards banded we considered earlier (contained in **`brownadt.inp`**). For these recovery data, fit models $\{S_t\ r_.\}$ (Seber) and $\{S_t\ f_.\}$ (Brownie), and add the results to the browser (top of the next page).

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Seber - S(t)r(.)}	8653.3953	0.0000	0.99733	1.0000	10	44.4485
{Brownie - S(t)f(.)}	8665.2441	11.8488	0.00267	0.0027	9	58.3032

We see that the model fits are not even remotely similar. This is reflected both in terms of the model deviances, but also (and more to the point we're trying to make here) in terms of the parameter estimates.

Here are the reconstituted estimates from the Seber and Brownie models, respectively:

Seber Estimates

Parameter	Estimate
1:S	0.7344295
2:S	0.5590948
3:S	0.6847692
4:S	0.6542361
5:S	0.6654351
6:S	0.6096429
7:S	0.5958613
8:S	0.5336050
9:S	0.6316335
10:r	0.1843101

Brownie Estimates

Parameter	Estimate
1:S	0.6251393
2:S	0.5725031
3:S	0.6367271
4:S	0.7291001
5:S	0.6546942
6:S	0.5783074
7:S	0.7027664
8:S	0.4746419
9:f	0.0703712

Several things to note. First, in the browser (above) we have both the $\{S_t\, r_.\}$ (Seber) and $\{S_t\, f_.\}$ (Brownie) models. You might be tempted to simply look at the AIC for each models, and conclude that there was 'overwhelming support' for the Seber parameterization. However, because the Brownie and Seber models are based on different likelihoods, you cannot legitimately compare the AIC values between the models (this is frequently an important consideration whenever you're comparing results from different data types fit to the same underlying data).

Second, and of particular interest, the estimates of survival (shown above) for the first 8 intervals which are estimable under both models are quite different – often dramatically so. For example, under the Seber model, the estimated survival probability for the first interval is 0.7344, whereas under the Brownie model, the estimate for the same interval is 0.625, a value which is almost 15% smaller!

So, the obvious question you might have is, 'which model yields the most robust estimates of survival?'. In considering this problem, there are a couple of things to keep in mind. First, for the Brownie model, the recovery probability f contains some information about mortality, and thus constraining either S or f to be constant while allowing the other parameter to vary with time makes implicit assumptions about the pattern of variation in other parameters. For example, for Brownie model $\{S_t\, f_.\}$, if K (kill rate) varies with time, then parameters c and λ must covary in such a way that the product $Kc\lambda$ (which equals the recovery probability f) does not vary. More likely, if c and λ are constant (as is often assumed), then constant S implies either that K is constant, or that natural mortality is compensatory (and not additive). Thus, it might be reasonable to wonder if model $\{S_t\, f_.\}$ is a reasonable model under the Brownie parameterization.

Second, it is important to remember that in the end, the Brownie and Seber models are fundamentally different models – the Seber model is more general (i.e., has more parameters) than the Brownie model, and thus will 'fit the data better' (i.e., have a smaller deviance).

And this is key – for most of our models, the Brownie and Seber parameterizations are effectively equivalent – and yield the same model fits and estimates for survival. For example, in the following browser (shown at the top of the next page) we show the model fits for our 4 standard models (models $\{S_t\, f_t\}$, $\{S_.\, f_t\}$, $\{S_t\, f_.\}$ and $\{S_.\, f_.\}$).

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{Seber - S(t)r(.)}	8653.3953	0.0000	0.48579	1.0000	10	44.4485
{Brownie - S(.)f(t)}	8655.1931	1.7978	0.19773	0.4070	10	46.2463
{Seber - S(.)r(t)}	8655.1931	1.7978	0.19773	0.4070	10	46.2463
{Brownie - S(t)f(t)}	8657.8302	4.4349	0.05290	0.1089	17	34.8258
{Seber - S(t)r(t)}	8657.8302	4.4349	0.05290	0.1089	17	34.8258
{Brownie - S(.)f(.)}	8662.2404	8.8451	0.00583	0.0120	2	69.3240
{Seber - S(.)r(.)}	8662.2404	8.8451	0.00583	0.0120	2	69.3240
{Brownie - S(t)f(.)}	8665.2441	11.8488	0.00130	0.0027	9	58.3032

Looking closely at the model results, we see that:

Brownie		*Seber*
$\{S_t\, f_t\}$	$\equiv$	$\{S_t\, r_t\}$
$\{S_.\, f_t\}$	$\equiv$	$\{S_.\, r_t\}$
$\{S_.\, f_.\}$	$\equiv$	$\{S_.\, r_.\}$
$\{S_t\, f_.\}$	$\not\equiv$	$\{S_t\, r_.\}$

In other words, only models $\{S_t\, f_.\}$ (Brownie) and $\{S_t\, r_.\}$ (Seber) are not equivalent – because these are the only two models which do not have the same number of estimable parameters.

And, thus, comparing survival estimates from these two models is analogous to comparing 'apples and oranges'. We leave the question of which set of estimates (Brownie or Seber) is least biased for you to explore – however, it is clear that you need to pay careful attention to the number of estimable parameters for a given model type if comparing estimates generated using either the Brownie or Seber parameterization.

8.6. Recovery analysis when the number marked is not known

If you look back in this chapter, you'll see that under the 'typical' application of recovery analysis, the number of recoveries expected over a given interval is equal to the number marked and released (R_i) times the survival and recovery probabilities.

However, under some marking schemes (for example, the marking or 'ringing' scheme that was used by the British Trust for Ornithology – the 'BTO'), the number marked and released is often unknown. What can you do in these cases?

To circumvent this problem, a ring recovery model is formulated where the recovery probability (using r_i from the reduced parameterization) is assumed constant by age class and year. Under this assumption, the survival probability can be estimated from the observed recoveries. How does this work?

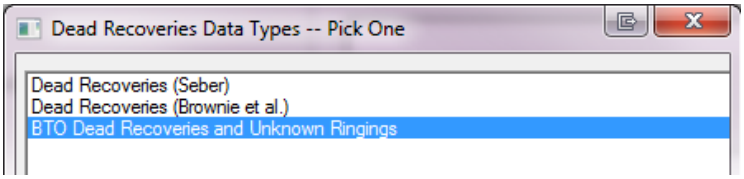
If we assume that r_i is a constant, then the cell probability for the j year of recoveries given k years of recoveries is

$$\frac{S_1 S_2 S_3 \ldots S_{j-1}(1 - S_j)}{1 - S_1 S_2 S_3 \ldots S_k},$$

where the denominator is 1 minus the probability of still being alive. Note in particular that recovery (r_i) does not appear in this expression.

Of the k survival probabilities, only $(k-1)$ are identifiable. Common approaches to achieve identifiability are to set $S_{k-1} = S_k$ or to set S_k equal to the mean of $S_1, S_2, \ldots, S_{k-1}$ using appropriate constraints in the design matrix. This model should only be used when you do not know the number of animals marked because you cannot evaluate the assumption of constant recovery probabilities with this model. If you know the number of individuals marked, use one of the 'normal' dead recovery models (Brownie or Seber) described earlier in this chapter.

To implement a BTO recovery analysis in **MARK**, simply select '**`BTO Dead Recoveries and Unknown Ringings`**':



What about the data file itself? Consider a 'typical' recovery matrix (in fact, the **`brownadt.inp`** file we've looked at previously). The last row of the `INP` file in this case reflect the number released in each year (cohort). What would you do to modify the format for the 'BTO' data type? Simple – delete the last line! [Why? Because for the 'BTO' data type, you don't actually know the number marked and released.]

Let's run this analysis using **MARK**, to get a more 'hands-on' sense of how the BTO data type analysis differs from the 'normal' dead recovery analyses we've already discussed. Start up **MARK**, and select **`brownadt.inp`**. View the file, by opening it in the default editor. Edit the file by deleting the last row of total number of marked and released individuals (which we're assuming we don't know). Save the edited file – calling it **`bto.inp`**. Re-select the file to analyze, this time picking **`bto.inp`**. Set the number of occasions to 9, and then make sure the '**`BTO Ring Recoveries`**' data type is selected.

To see quickly that we're working with something distinct from 'normal' recovery analysis, have a look at the PIM chart. The first thing you'll notice immediately is that there is only one parameter – S (survival). Why? Because recovery (r) is assumed to be constant, and is therefore not estimated. Or, in other words, since the recovery probability (i.e., r_i) does not factor in the expected cell probabilities, then you clearly don't need to estimate it (in fact, you can't!).

With only one parameter, then obviously all constraints are placed on survival only. Clearly, this is a significant limitation in your ability to analyze these data, since you cannot test any hypotheses concerning variation in recovery probability. Assuming a constant recovery rate is a necessary step to do anything with data collected in this way. Since the BTO has collected a lot of data over the past many decades, there has been a fair amount of work devoted to the theory of analyzing data of this type, where the number marked and released is unknown. However, despite those efforts, there are going to be limits to what you can do.

How do the estimates from this analysis, using the BTO data type, compare to those using 'normal' recovery analysis, where the number marked and released is known (recall that for these data, we actually do know the number marked and released)?

The estimates from model $\{S_t\}$ for the BTO data type are shown at the top of the next page. Notice that there are no standard errors reported for each of the estimates. This is because for the 'BTO' data type, the error variance around the estimates of S_i cannot itself be estimated under the constraint (assumption) of constant recovery rate.

Real Function Parameters of {S(t)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:S	0.7304926	0.0000000	0.7304926	0.7304926
2:S	0.5523693	0.0000000	0.5523693	0.5523693
3:S	0.6756720	0.0000000	0.6756720	0.6756720
4:S	0.6401287	0.0000000	0.6401287	0.6401287
5:S	0.6408258	0.0000000	0.6408258	0.6408258
6:S	0.5628517	0.0000000	0.5628517	0.5628517
7:S	0.5150371	0.0000000	0.5150371	0.5150371
8:S	0.3580048	0.0000000	0.3580048	0.3580048
9:S	0.2432110	0.0000000	0.2432110	0.2432110

How do these estimates of S_i compare to the values from the most parsimonious model fit to these data when number marked and released was known? Recall that we analyzed these data earlier in this chapter – referring back to that analysis, we see that the most parsimonious model was model $\{S_. \; f_t\}$. For this model, S was estimated as 0.638.

For the most parsimonious model with time dependence in S (model $S_t \; f_t$), the estimates of survival are

Real Function Parameters of {S(t)f(t)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:S	0.5790622	0.1140632	0.3547411	0.7748867
2:S	0.6109505	0.0780946	0.4519831	0.7493747
3:S	0.6238588	0.0750659	0.4697788	0.7563830
4:S	0.8416866	0.1051320	0.5310666	0.9614777
5:S	0.6384904	0.0735473	0.4860772	0.7673382
6:S	0.5356835	0.0588982	0.4203958	0.6472796
7:S	0.5897951	0.0709196	0.4473407	0.7186264
8:S	0.5593796	0.1363270	0.3003884	0.7896370
9:S	0.0302280	0.0000000	0.0302280	0.0302280

We see that the estimates are markedly different. Why? Because, as it turns out, the most parsimonious model(s) had time-dependence in the recovery parameter – clearly a 'violation' of the assumption of constant recovery probability required by the BTO data type analysis.

8.7. Recovery models and GOF

First the good news – GOF testing for recovery models is possible, and quite straightforward. Now the bad news (well, perhaps not 'bad' news, but something to note) – the type of GOF tests that are available to you depends on which parameterization you use (Brownie, or Seber):

- If you want to use the '**Brownie**' parameterization, you can test goodness of fit of the data to your general model using program **ESTIMATE** (Brownie *et al*. 1985). Program **ESTIMATE** can be called from within **MARK** (much as you can invoke program **RELEASE** from within **MARK**).
- Alternatively, if you're using the '**Seber**' parameterization, you can use either the bootstrap or median-$\hat{c}$ approaches (but not program **ESTIMATE**) for GOF testing.

But, suppose you've already fit your model set using the '**Brownie**' parameterization, but instead of using program **ESTIMATE** for the GOF, you'd like to estimate $\hat{c}$ using either the bootstrap or median-$\hat{c}$ approaches. Do you need to 'start over', and re-construct all your '**Brownie**' models using the equivalent '**Seber**' parameterization? The answer (thankfully) is 'no'. All we need to do is change the data type from '**Brownie**' → '**Seber**' for the general model, which we can do directly within **MARK** (see below).

In the following, we'll demonstrate the various steps needed for GOF testing for dead recovery models. We'll use the **brownadt.inp** data file we analyzed earlier in the chapter. Recall that our analysis for these data was based on the '**Brownie**' parameterization. Also, remember we want to derive the measure of fit (estimate of $\hat{c}$) for our most general model, which in this case is model $\{S_t\ f_t\}$.

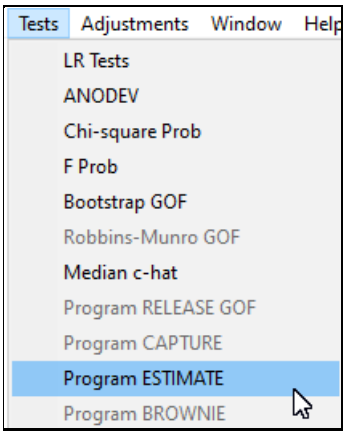
As noted above, for the '**Brownie**' parameterization, our only option for GOF testing is to run program **ESTIMATE** from within **MARK**. Program **ESTIMATE** provides basic GOF testing for several of the 'classic' models under the '**Brownie**' parameterization (think of **ESTIMATE** in some senses as the recovery equivalent of **RELEASE**).

Program **ESTIMATE** uses the 'classical' naming convention for models we noted earlier in this chapter:

model	*legacy name*	*reference*
$\{S_t\ f_t\}$	Model 1	Brownie *et al.* (1985) pp. 15-20
$\{S_t\ f_.\}$	none	
$\{S_.\ f_t\}$	Model 2	Brownie *et al.* (1985) pp. 20-24
$\{S_.\ f_.\}$	Model 3	Brownie *et al.* (1985) pp. 24-30

Under this model naming convention, model $\{S_t\ f_t\}$ is '`Model 1`'. To run **ESTIMATE**, you don't need to make any particular model in the browser 'active', since **ESTIMATE** simply fits a series of 'built-in' models, regardless of the models you have in your browser.* One of these models is '`Model 1`'.

To run **ESTIMATE**, simply pull down the '**Test**' menu, and select '**Program Estimate**':



After a few seconds, you'll be dumped into the Notepad, which will present the results of the **ESTIMATE** analysis. You'll want to find the part of the output pertaining to '`Model 1`'. After a bit of scrolling, you'll find the following results for '`Model 1`' (shown at the top of the next page). At the bottom of the output, you'll see that the observed χ^2 statistics for '`Model 1`' is 31.57, with 25 df. The P-value of observing a χ^2-value larger than 31.57 is 17.1%. Using (χ^2/df) as an estimate of $\hat{c}$, then $\hat{c}$ = (31.57/25) = 1.263.

* This means that unless your general model is one of the 'built-in' models that **ESTIMATE** is fitting to your data, you're out of luck.

```
YEAR NUMBER                                    RECOVERY MATRIX
     BANDED
____ ______                                    _______________

   1   231    10.   13.    6.    1.    1.    3.    0.    0.    3.
   2   649          58.   21.   16.   15.   13.    6.    0.    2.
   3   885                54.   39.   23.   18.   11.   10.    6.
   4   550                      44.   21.   22.    9.    9.    3.
   5   943                            55.   39.   23.   11.   12.
   6  1077                                  66.   46.   29.   18.
   7  1250                                        101.   59.   30.
   8   938                                              97.   22.
   9   312                                                    21.

MATRIX OF EXPECTED VALUES -- ASSUMING TIME-SPECIFIC SURVIVAL AND RECOVERY RATES   (MODEL 1)

           10.0  12.1   4.9   3.6   2.3   1.8   0.0   0.0   2.2
                 58.9  23.6  17.7  11.3   8.8   5.5   0.0   5.3
                       52.6  39.4  25.2  19.6  12.3   8.4   3.5
                             39.3  25.1  19.6  12.2   8.3   3.5
                                   51.1  39.9  24.9  17.0   7.2
                                         71.3  44.5  30.4  12.8
                                               96.5  65.8  27.8
                                                     83.7  35.3
                                                           21.0

MATRIX OF CHI-SQUARE VALUES -- ASSUMING TIME-SPECIFIC SURVIVAL AND RECOVERY RATES   (MODEL 1)

           0.00  0.06  0.27  1.92  0.76  0.78  0.00  0.00  0.27
                 0.01  0.28  0.16  1.22  2.00  0.05  0.00  2.08
                       0.04  0.00  0.19  0.14  0.13  0.32  1.73
                             0.57  0.67  0.30  0.85  0.05  0.08
                                   0.30  0.02  0.14  2.10  3.27
                                         0.39  0.05  0.06  2.10
                                               0.21  0.70  0.18
                                                     2.12  5.02
                                                           0.00

(FREQUENCIES WERE COMBINED WHERE EXPECTED VALUES WERE SMALL)


TEST OF THE NULL HYPOTHESIS THAT THE DATA FIT MODEL 1 -- ASSUMING TIME-SPECIFIC SURVIVAL AND RECOVERY RATES

   CHI-SQUARED VALUE (SAMPLE) =   31.57
   THEORETICAL CHI-SQUARE VALUE AT THE 5% LEVEL =
                                                  37.70
   DEGREES OF FREEDOM =     25

PROBABILITY OF A CHI-SQUARE VALUE LARGER THAN   31.57  =   0.17076623
```

But, what if we wanted to use either the bootstrap or median-$\hat{c}$ approaches, rather than program **ESTIMATE**? Recall that both the bootstrap or median-$\hat{c}$ GOF tests are available only for the '**Seber**' parameterization. If your models are already constructed using the '**Seber**' parameterization, then you simply make the general model active in the browser (by right-clicking and retrieving it), and then proceeding as per normal.

However, if your models are constructed using the '**Brownie**' parametrization, as in the present example, then you first need to change the data type for the general model from '**Brownie**' → '**Seber**'. As demonstrated earlier in this chapter, this is easy to do – simply make the general model active in the browser (by right-clicking and retrieving it), and then select '**PIM | Change data type**'. **MARK** will present you with a selection of data types which are consistent with the data contained in the PIM.

In this case, there are only two such data types: the '**Dead recoveries (Seber)**' (i.e., the S and r Seber parameterization, and the '**Dead recoveries (Brownie et al.)**' (our current data type). We want to switch to the Seber 'S and r' data type, so pick the '**Dead recoveries (Seber)**' option from the list. You won't see anything happen, but you'll now be able to run a model under the 'S and r' parameterization. The model we want to run is model $\{S_t\ r_t\}$, which is equivalent to model $\{S_t\ f_t\}$. If you want look at the PIM chart, you'll see that the general model is now parameterized in terms of S and r – i.e., the '**Seber**' parameterization.

Once you've changed the data type, go ahead and run the model, and call it model '`S(t)r(t)`'. Add the results to the browser. You should observe that the AIC, deviance and the number of parameters are identical to that reported for model $\{S_t\ f_t\}$. Now, all you need to do is run a bootstrap or median-$\hat{c}$ GOF test on this new model $\{S_t\ r_t\}$. The mechanics for both tests were covered in detail in Chapter 5.

Based on 1,000 bootstraps, we found that approximately 21% of the bootstrapped deviances were greater than the observed deviance for model $\{S_t\ r_t\}$, indicating adequate fit. Recall from our program **ESTIMATE** GOF analysis that the observed χ^2 statistics for `Model 1` was 31.57, with 25 df. The P-value of observing a χ^2-value larger than 31.57 is 17.1%, which is comparable to the 21% value observed from the bootstrap analysis. Further, both our bootstrapped and median-$\hat{c}$ estimates for $\hat{c}$ (1.153, and 1.110, respectively) are consistent with the estimate of $\hat{c}$ from the **ESTIMATE** analysis (31.57/25 = 1.263).

Taken together, this would suggest some level of equivalence between the approach based on program **ESTIMATE**, applied to the general model under the '**Brownie**' parameterization, and the bootstrap and median$\hat{c}$ approaches, under the '**Seber**' parameterization. Such a conclusion should be approached cautiously. One thing the **ESTIMATE** output does give you is the *relative* contribution of each element of the recovery matrix to the overall model χ^2, analogous to partitioning the data into the contingency tables that we used with program **RELEASE** for live encounter data. Careful examination of these tables can sometimes help you diagnose lack of fit.

8.8. Summary

Recovery models are more common than you think, and not simply restricted to 'harvested' species. It is worth spending some time getting comfortable with the theory, and the different implementation of recovery analysis in **MARK**. In the next chapter, we'll actually combine 'dead recovery' models with 'live encounter' models. As you'll see, this 'joint' estimation can in some cases make it possible to tease apart sources of apparent mortality in novel and potentially useful ways.

8.9. References

Anderson, D. R., Burnham, K. P., and White, G. C. (1985) Problems in estimating age-specific survival rates from recovery data of birds ringed as young. *Journal of Animal Ecology*, **54**, 89-98.

Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985) *Statistical Inference from Band-Recovery Data – A Handbook*, 2nd Edition. Resource Publication No. 156. Washington, D.C. – Fish & Wildlife Service, U.S. Department of the Interior.

Catchpole, E. A., Freeman, S. N., and Morgan, B. J. T. (1995) Modeling age variation in survival and reporting rates for recovery models. *Journal of Applied Statistics*, **22**, 597-609.

Catchpole, E. A., Freeman, S. B., and Morgan, B. J. T. (1996) Steps to parameter redundancy in age-dependent recovery models. *Journal of the Royal Statistical Society*, **58**, 763-774.

Catchpole, E. A., Morgan, B. J. T., and Freeman, S. B. (1998) Estimation in parameter redundant models. *Biometrika*, **85**, 462-468.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67-118.

Seber, G. A. F. (1970) Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika*, **57**, 313-318.