



## Two-level Poisson model with offset variable

### Contents

3.1	TWO-LEVEL MODELS FOR COUNT OUTCOMES FROM ASPART DATA .....	1
3.1.1	<i>The data</i> .....	1
3.1.2	<i>A 2-level Poisson log model with an offset variable</i> .....	2

### 3.1 Two-level models for count outcomes from ASPART data

#### 3.1.1 The data

The data for this example are taken from a paper by McKnight and Van Den Eeden (1993), who reported on the number of headaches in a two treatment, multiple period crossover trial. Specifically, the number of headaches per week was repeatedly measured for 27 patients. Following a seven day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods according to the double-blind crossover treatment design. Each treatment period was separated by a washout day. The sample size is 122. Data for the first 10 observations of all the variables used in this section are shown below in the form of a SuperMix spreadsheet window for **aspart.ss3**.

	(A)_ID	(B)_Headac	(C)_Period1	(D)_Period2	(E)_Period3	(F)_Period4	(G)_DrugAsp	(H)_NPerio	(I)_NTDdays
1	2.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00	7.00
2	2.00	5.00	1.00	0.00	0.00	0.00	1.00	3.00	7.00
3	2.00	2.00	0.00	1.00	0.00	0.00	0.00	3.00	7.00
4	5.00	3.00	0.00	0.00	0.00	0.00	0.00	5.00	7.00
5	5.00	0.00	1.00	0.00	0.00	0.00	1.00	5.00	7.00
6	5.00	2.00	0.00	1.00	0.00	0.00	0.00	5.00	7.00
7	5.00	0.00	0.00	0.00	1.00	0.00	1.00	5.00	7.00
8	5.00	0.00	0.00	0.00	0.00	1.00	0.00	5.00	7.00
9	13.00	7.00	0.00	0.00	0.00	0.00	0.00	5.00	7.00
10	13.00	7.00	1.00	0.00	0.00	0.00	1.00	5.00	7.00

The variables of interest are:

- ID is the patient ID (27 patients in total).

- Headache is the number of headaches during the week (from 0 to 7).
- Period1 is a period 1 treatment indicator (1 for the first treatment period and 0 otherwise).
- Period2 is a period 2 treatment indicator (1 for the second treatment period and 0 otherwise).
- Period3 is a period 3 treatment indicator (1 for the third treatment period and 0 otherwise).
- Period4 is a period 4 treatment indicator (1 for the fourth treatment period and 0 otherwise).
- DrugAsp indicates the type of drug being used for the treatment, (0 = placebo and 1 = aspartame). 75 observations used placebo and 47 used aspartame.
- Nperiods is the number of periods the individual was observed (from 2 to 5).
- NTDays is the number of treatment days in the period (from 1 to 7).

### 3.1.2 A 2-level Poisson log model with an offset variable

#### 3.1.2.1 The model

In a standard Poisson model, it is assumed that the counts were all observed for the same number of days. However, this was not the case since the number of treatment days in the period did vary to some degree. Most of the counts were based on the full seven days in the week; however, some observations were made only for 1 day in the given week. To take this into account, we need to specify a so-called OFFSET variable. The offset variable indicates the amount of time that each count is based on. If OFFSET = no is specified, as was the case in the previous example, SuperMix assumes that all counts are based on the same amount of time.

The offset variable is introduced into the Poisson model in the following way:

$$\log\left(\hat{\lambda}_{ij}\right) = \log(\text{offset variable}) + \left[\mathbf{x}_{ij}' \mathbf{b}_i\right]$$

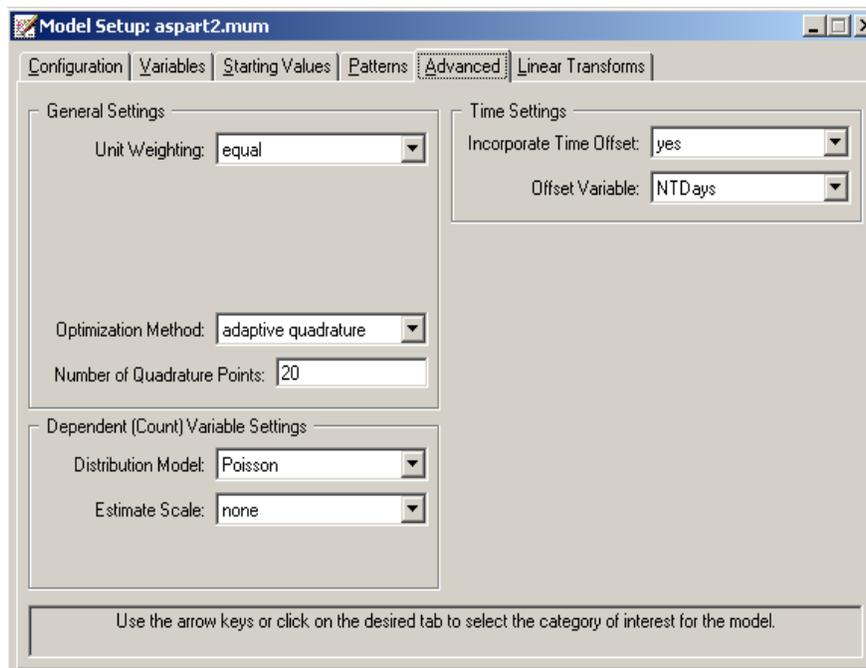
where  $\mathbf{x}_{ij}$  represent the values of the covariates corresponding to level-1 unit  $j$  nested within level-2 unit  $i$  and  $\mathbf{b}_i$  denotes the coefficient vector containing both fixed and random effects.

In the current situation, the variable NTDays is the appropriate choice as the OFFSET variable. The model to be fitted to the data now changes to:

$$\begin{aligned} \log(\text{Headache}_{ij}) = & \log(\text{NTDays}) + (\beta_0 + \beta_1 \times \text{Period1}_{ij} + \beta_2 \times \text{Period2}_{ij} \\ & + \beta_3 \times \text{Period3}_{ij} + \beta_4 \times \text{Period4}_{ij} + \beta_5 \times \text{DrugAsp}_{ij} + v_{i0}). \end{aligned}$$

### 3.1.2.2 Setting up the analysis

To create the model specifications for this model, start by opening **aspart.ss3** in a SuperMix spreadsheet window and using the **Open Existing Model Setup** option on the **File** menu to open the **Model Setup** window for **aspart1.mum**. On the **Configuration** screen, extend the title in the **Title 1** text box by adding the string "with Offset Variable." Next, click on the **Advanced** tab of the **Model Setup** window. Select **yes** from the **Incorporate Time Offset** drop-down list to activate the **Offset Variable** drop-down list box. Select the variable **NTDays** from the drop-down list of **Offset Variable** to produce the following **Advanced** tab.

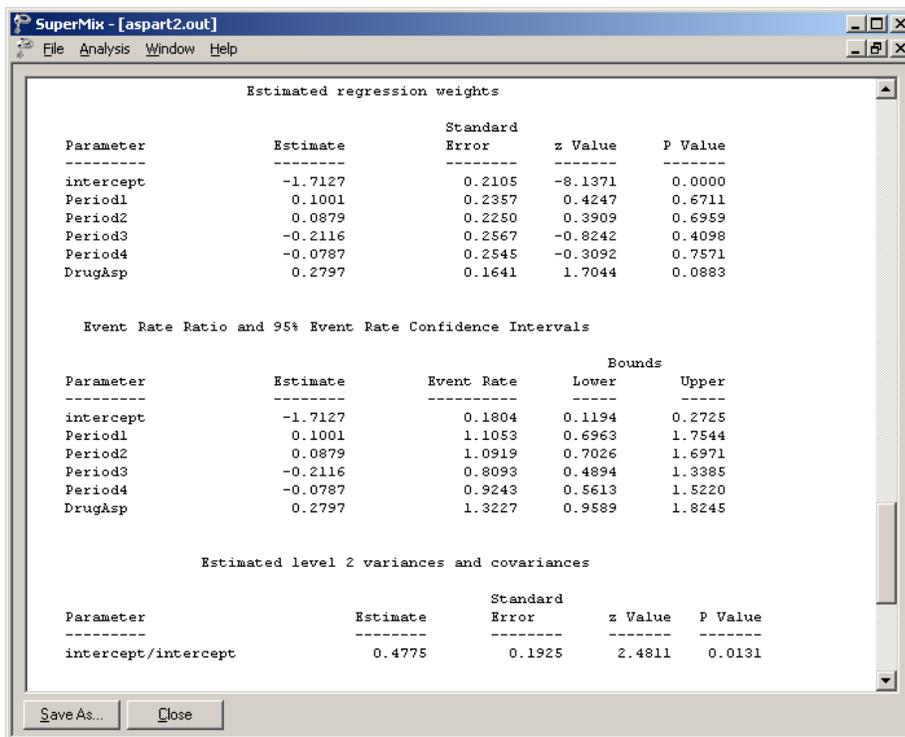
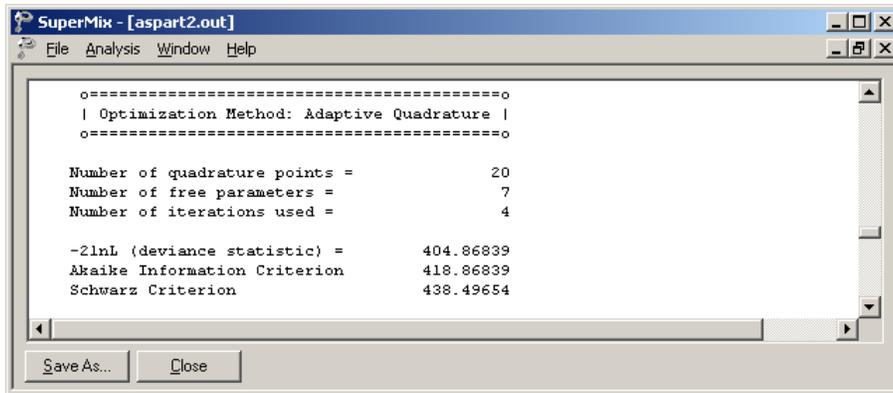


Save the changes to the file **aspart2.mum** by using the **Save As** option on the **File** menu and select the **Run** option on the **Analysis** menu to produce the output file **aspart2.out**.

### 3.1.2.3 Discussion of results

#### Fixed and random effect results

Portions of this output file are shown below. Results for this model differ from those obtained for the model without offset variable discussed in the previous section. While the overall trend in predictor coefficient estimates is similar, the intercept is now estimated as  $-1.7127$ , compared to  $0.2572$  previously. The variance in intercept over patients for this model is estimated as  $0.4775$  compared to  $0.4290$  previously.



### 3.1.2.4 Interpreting the results

#### Estimated outcomes for groups: unit-specific results

The expected number of headaches can be obtained in the following fashion. First, we substitute the estimated coefficients in the model formulation

$$\begin{aligned}\log\left(\widehat{\text{Headache}}_{ij}\right) &= \log\left(\text{NTDays}_{ij}\right) + \left(\widehat{\beta}_0 + \widehat{\beta}_1 \times \text{Period1}_{ij} + \widehat{\beta}_2 \times \text{Period2}_{ij}\right. \\ &\quad \left. + \widehat{\beta}_3 \times \text{Period3}_{ij} + \widehat{\beta}_4 \times \text{Period4}_{ij} + \widehat{\beta}_5 \times \text{DrugAsp}_{ij}\right) \\ &= \log\left(\text{NTDays}_{ij}\right) + (-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij} \\ &\quad - 0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}),\end{aligned}$$

or, after taking exponents on both sides, as

$$\begin{aligned}\widehat{\text{Headache}}_{ij} &= \text{NTDays}_{ij} \times \exp(-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij} \\ &\quad - 0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}).\end{aligned}$$

As most observations had a value of  $\text{NTDays} = 7$ , we start by considering typical patients with a full set of treatment days. We also assume that  $\text{DrugAsp} = 1$ , in other words, that aspartame rather than a placebo was administered.

During the first treatment period, we find that for such a patient

$$\begin{aligned}\widehat{\text{Headache}}_{ij} &= 7 \exp(-1.7127 + 0.1001 + 0.2797) \\ &= 7 \exp(-1.3329) \\ &= 1.8460.\end{aligned}$$

The expected numbers of headaches for a typical patient (again with  $\text{NTDays} = 7$  and  $\text{DrugAsp} = 1$ ) for the second, third, and fourth treatment periods are calculated as

$$\begin{aligned}\widehat{\text{Headache}}_{ij} &= 7 \exp(-1.7127 + 0.0879 + 0.2797) \\ &= 1.8236,\end{aligned}$$

$$\begin{aligned}\widehat{\text{Headache}}_{ij} &= 7 \exp(-1.7127 - 0.2116 + 0.2797) \\ &= 1.3516,\end{aligned}$$

and

$$\begin{aligned}\widehat{\text{Headache}}_{ij} &= 7 \exp(-1.7127 - 0.0787 + 0.2797) \\ &= 1.5437\end{aligned}$$

respectively.

For a typical patient with only 5 treatment days, the expected numbers of headaches in each of the four treatment periods are 1.3186, 1.3026, 0.9654, and 1.1027 respectively.

When the expected numbers of headaches for a typical patient receiving aspartame under the Poisson model without offset variable (see previous section) and the Poisson model with offset variable are compared, we clearly see the impact of the inclusion of the offset variable on the estimated coefficients. These results are shown in Table 5.4.

**Table 5.4: Comparison of results for Poisson models**

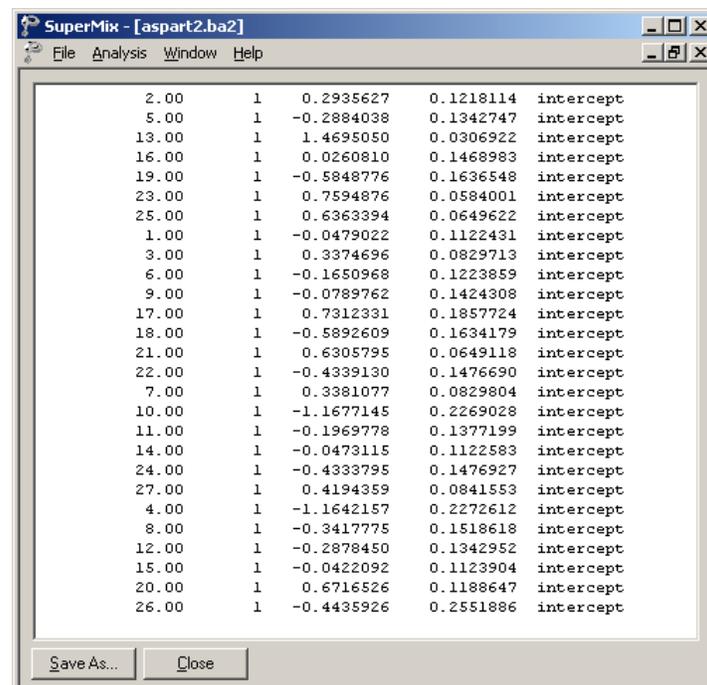
<b>Period</b>	<b>Without offset variable</b>	<b>With offset variable (NTDays = 7)</b>	<b>With offset variable (NTDays = 5)</b>
1	1.7385	1.846	1.3186
2	1.6600	1.8236	1.3026
3	1.2784	1.3516	0.9654
4	1.3677	1.5437	1.1027

## Level 2 Bayes results

As requested during the model specification stage, the empirical Bayes estimates of the random effects are written to the file **aspart2.ba2**. The first few lines of this file are shown below.

The file **aspart.ba2** contains five pieces of information per individual:

- the individual's ID,
- the number of repeated observations for that individual,
- the empirical Bayes estimate for that individual (which is the mean of the posterior distribution),
- the associated posterior standard deviation, and
- the name of the relevant random coefficient.



ID	Number of observations	Estimate	Standard deviation	Coefficient name
2.00	1	0.2935627	0.1218114	intercept
5.00	1	-0.2884038	0.1342747	intercept
13.00	1	1.4695050	0.0306922	intercept
16.00	1	0.0260810	0.1468983	intercept
19.00	1	-0.5848776	0.1636548	intercept
23.00	1	0.7594876	0.0584001	intercept
25.00	1	0.6363394	0.0649622	intercept
1.00	1	-0.0479022	0.1122431	intercept
3.00	1	0.3374696	0.0829713	intercept
6.00	1	-0.1650968	0.1223859	intercept
9.00	1	-0.0789762	0.1424308	intercept
17.00	1	0.7312331	0.1857724	intercept
18.00	1	-0.5892609	0.1634179	intercept
21.00	1	0.6305795	0.0649118	intercept
22.00	1	-0.4339130	0.1476690	intercept
7.00	1	0.3381077	0.0829804	intercept
10.00	1	-1.1677145	0.2269028	intercept
11.00	1	-0.1969778	0.1377199	intercept
14.00	1	-0.0473115	0.1122583	intercept
24.00	1	-0.4333795	0.1476927	intercept
27.00	1	0.4194359	0.0841553	intercept
4.00	1	-1.1642157	0.2272612	intercept
8.00	1	-0.3417775	0.1518618	intercept
12.00	1	-0.2878450	0.1342952	intercept
15.00	1	-0.0422092	0.1123904	intercept
20.00	1	0.6716526	0.1188647	intercept
26.00	1	-0.4435926	0.2551886	intercept

Since they are estimates of  $b_{i0}$  for each individual, the empirical Bayes estimates are expressed on the standard normal scale. Inspection of these estimates indicates that subject 13 has a very high score. This person's estimate of 1.043 (with standard deviation .016) suggests a very high level of headaches. This agrees well with the raw data, which reveals that this person recorded 7 headaches on four occasions and 6 on the only other occasion.

## Graphical displays

Figure 5.5 is a comparison (represented by a dotted line) of the predicted average number of headaches reported by each patient when taking a placebo (left axis) as opposed to the predicted average number when the treatment is aspartame (right axis). From the graphical display, it appears as if all of the lines (each representing a patient) have a positive slope. The slopes become steeper as the number of headaches increases. This suggests an increase, albeit small, in the expected average number of headaches when aspartame is used. Note that patient 13, who reported a consistently high number of headaches at all occasions, was excluded from this graph.

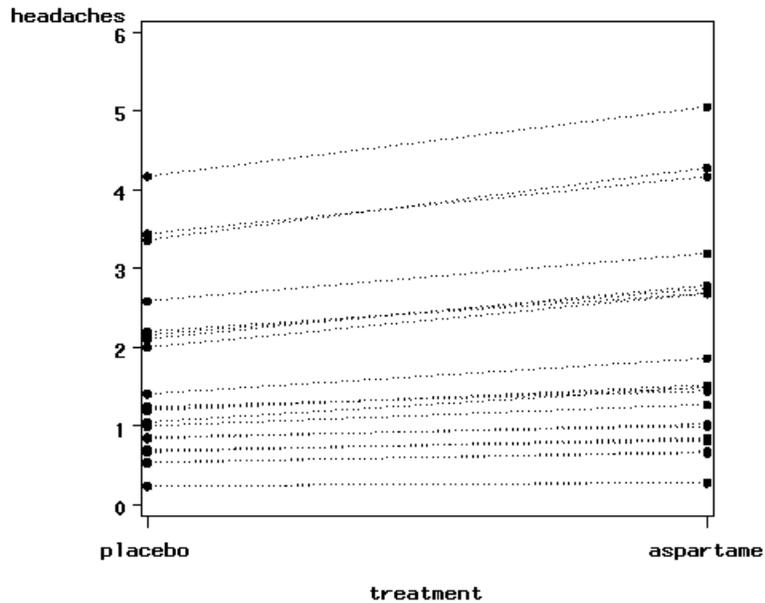
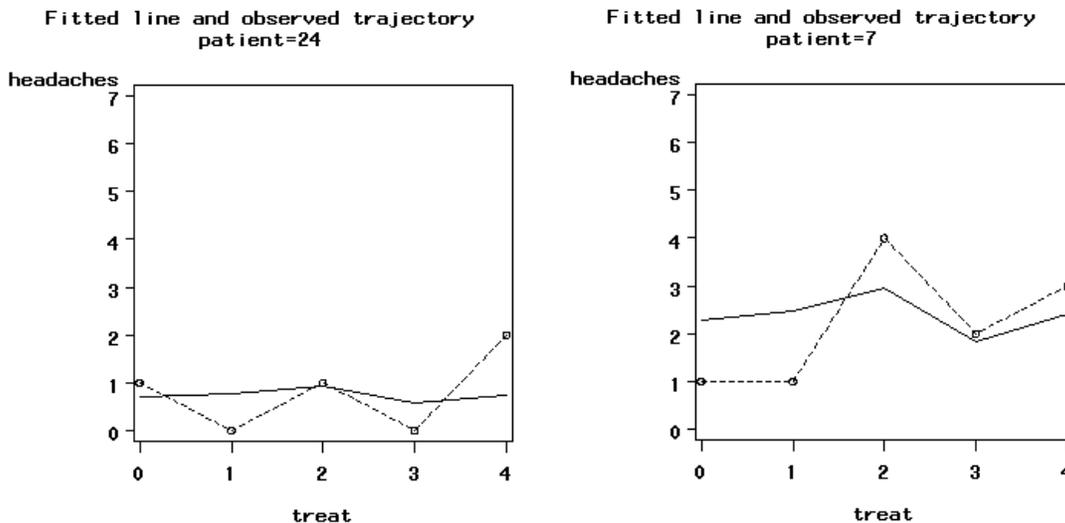
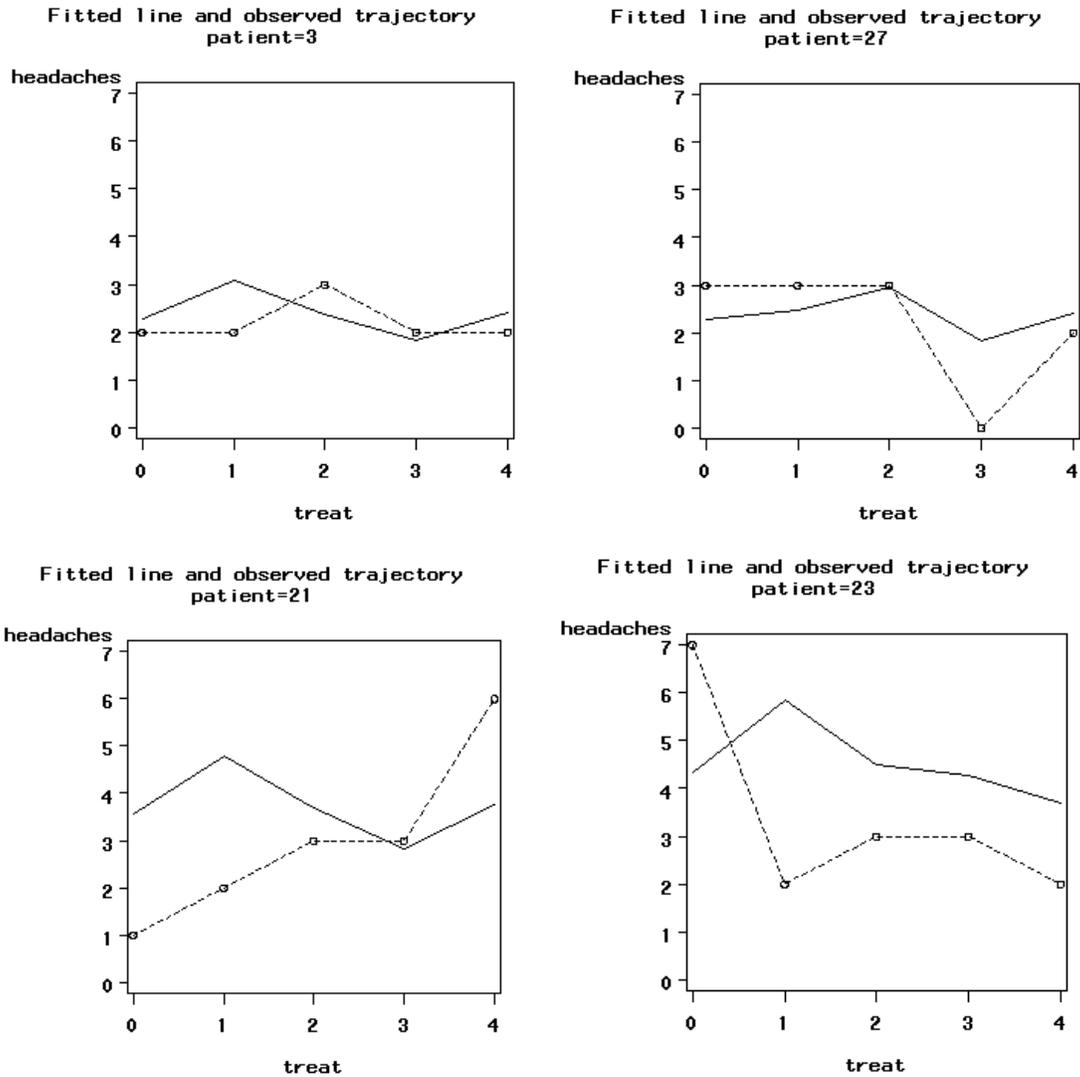


Figure 5.5: Predicted average number of headaches for placebo and aspartame





**Figure 5.6: Fitted and observed trajectories**

Figure 5.6 is a graphical display of the fitted trajectory (solid line) and observed trajectory (dotted line) for a sample of 6 patients. These displays are ordered from a patient who reported a relatively small number of headaches at the different treatment occasions to one who reported a relatively high number of headaches at the treatment occasions. A study of the fitted and observed trajectories reveals that, in general, the model fit is best when the number of headaches is smaller and becomes less accurate as the number of headaches increases. For patient 13, who is not represented in the graphical display, the number of predicted headaches is almost twice the number observed.

The fitted lines were obtained as

$$\widehat{\text{Headache}}_{ij} = \text{NTDays}_{ij} \times \exp(-1.7127 + 0.1001 \times \text{Period1}_{ij} + 0.0879 \times \text{Period2}_{ij} - 0.2116 \times \text{Period3}_{ij} - 0.0787 \times \text{Period4}_{ij} + 0.2797 \times \text{DrugAsp}_{ij}) + \widehat{b}_{i0}.$$

where  $\widehat{b}_{i0}$  is obtained from the **aspart2.ba2** file, shown previously in this section.