APPENDIX-B
AN ALGORITHM FOR LEAST SQUARES ESTIMATION OF NON-LINEAR PARAMETERS WITH MODEL MIGRATION SCHEDULES

This appendix briefly illustrates the mathematical programming procedure used to estimate the parameters of the model migration schedules. The functional minimization procedure (non-linear least square principle) may be defined as the search for the "best" parameter values for the function:

\[ M(x) = A_1 \exp(-\alpha_1 x) + A_2 \exp(-\alpha_2 (x-\mu_2)) - \exp(-\lambda_2 (x-\mu_2)) ] + A_3 \exp(-\alpha_3 (x-\mu_3)) - \exp(-\lambda_3 (x-\mu_3)) + C \quad ;x=0, 1, 2, \ldots, n \quad (B.1) \]

in the sense that a pre-defined objective function is minimized when the parameters take on these values.

Since, no single method has emerged which is best for the solution of all non-linear programming problems (Bird, 1974, PP.84), an iterative method with the adoption of specific convergence criteria are usually used, such as Gauss-Newton procedure, Steepest descent method, Quasi-Newton algorithms etc.

But, if the columns of the matrix \( V^{(r)} = (V^{(r)}_{ij}) \), are multicollinear, than the problem would arise. In such
situations the Gauss-Newton procedure becomes deadly slow, requires excessive step halving and becomes practically unusable.

Therefore, such situations are handled very well by a procedure called Marquardt's Algorithm for least squares estimation of parameter vector $\theta^*$ that minimizes the following objective function:

$$S^2(\theta) = \sum_{i=1}^{n} (M(x_i) - f(x_i, \theta))^2 \quad (B.2)$$

(Note, that the model used here is $M(x_i) = f(x_i, \theta) + \epsilon_i$, where $i=1, \ldots, n$, and assumed that $E(\epsilon) = 0$ and $E(\epsilon \epsilon') = \sigma^2 I$).

Assuming adequate differentiability of $f$, and setting partial derivatives of (B.2) with respect to each $\theta_j$ equal to zero, we get,

$$\sum_{i=1}^{n} (M(x_i) - f(x_i, \theta)) v_{ij} = 0 \quad (B.3)$$

for all $j = 1, 2, \ldots, k$, where $v_{ij} = \partial f(x_i, \theta)/\partial \theta_j$. Here (B.3) is a system of $k$ equations but, which can not be solved by the usual simple procedure called "solving system of equations". Therefore, an iterative procedure needs to be applied in case of such problems. One such procedure, which is most popular is the Gauss-Newton procedure.
B.1 GAUSS-NEWTON PROCEDURE:

Assuming adequate differentiability of \( f \), we can approximate it using linear terms of a Taylor expansion around a value \( \theta^{(0)} \) of \( \theta \) as follows:

\[
f(x_i, \theta) = f(x_i, \theta^{(0)}) + \sum_{j=1}^{k} v_{ij}^{(0)} (\theta - \theta^{(0)})
\]  

(B.4)

for \( i = 1, 2, \ldots, n \) where \( v_{ij}^{(0)} \) is the partial derivative evaluated at \( \theta^{(0)} \). Hence,

\[
M(x_i) - f(x_i, \theta) = e_i^{(0)} - \sum_{j=1}^{k} v_{ij}^{(0)} \tau_j^{(0)}
\]  

(B.5)

Where \( e_i^{(0)} = (e_i^{(0)}, \ldots, e_n^{(0)}) \), with \( e_i^{(0)} = M(x_i) - f(x_i, \theta^{(0)}) \)

and \( \tau^{(0)} = (\tau_1^{(0)}, \ldots, \tau_k^{(0)}) = \theta - \theta^{(0)} \). Therefore, if the approximation (B.4) is close, minimizing \( S^2(\theta) \) is approximately equivalent to minimizing

\[
\sum_{i=1}^{n} (e_i^{(0)} - \sum_{j=1}^{k} v_{ij}^{(0)} \tau_j^{(0)})^2 = (e^{(0)} - V^{(0)} \tau^{(0)})' (e^{(0)} - V^{(0)} \tau^{(0)})
\]  

(B.6)

with respect to \( \tau^{(0)} \). Where \( V^{(0)} \) is the matrix \( (v_{ij}^{(0)}) \) of \( v_{ij}^{(0)} \)'s.

Since \( V^{(0)} \tau^{(0)} \) is linear in \( \tau^{(0)} \), (B.6) is obviously minimized by the linear least squares estimation \( t^{(0)} \) of \( \tau^{(0)} \), i.e., by

\[
t^{(0)} = [(V^{(0)})' V^{(0)}]^{-1} (V^{(0)})' e^{(0)}
\]

Therefore, if the approximation (B.4) is reasonably good, one
should expect that

\[ S^2(\theta^{(0)} + t^{(0)}) \leq S^2(\theta^{(0)}) \]  \quad (B.7)

Unfortunately, the expression (B.4) is not always good approximation and (B.7) does not always holds true. But, whether it holds true or not, we can mechanically continue with the procedure. At the next iteration of the procedure, we would start with \( \theta^{(1)} = \theta^{(0)} + t^{(0)} \) and do exactly what we did earlier, only with \( \theta^{(1)} \) instead of \( \theta^{(0)} \). These iteration are continued until a suitable criterion is met. Practically, iterations are stopped when either

1. \[ \frac{S^2(\theta^{(r)}) - S^2(\theta^{(r-1)})}{S^2(\theta^{(r)}) + \delta^*} < \delta \]

   where \( \delta^* \) and \( \delta \) are some very small numbers (e.g, \( 10^{-6} \)) or so), or

2. a specified number of iterations, or till to get converged values of \( \theta \).

B.2 MARQUARDT PROCEDURE:

Since at each step, the Gauss-Newton procedure uses a linear least squares estimate \( t^{(0)} = [(V^{(r)})^T V^{(r)}]^{-1} (V^{(r)})^T e^{(r)} \), therefore, if the matrix \( V^{(r)} \) is ill conditioned (i.e if the columns of \( V^{(r)} \) are multicollinear ) the problems would arise. In ill conditioned cases the Gauss-Newton procedure becomes practically unusable. To over come this difficulty Marquardt
(1963) has suggested a procedure in which at each iteration, 
\( \tau^{(r)} = \theta^{(r+1)} - \theta^{(r)} \) is estimated by
\[
t^{(r)} = (V^{(r)\top} V^{(r)} + CD)^{-1} (V^{(r)\top} e^{(r)})
\] (B.8)
Where D is the diagonal matrix consisting of the diagonal entries in \((V^{(r)}) (V^{(r)\top})\), and C is an arbitrary number (This procedure was actually suggested by Levenberg, 1947, but in place of cD it was cI in (B.8), where I is a unit matrix). In actual use, c is usually set equal to a very small number like \(10^{-3}\) or so. c is scaled up by some factor (say 10) if 
\[ S^2(\theta^{(r+1)}) > S^2(\theta^{(r)}) \] otherwise, c is scaled down by (say 0.1) for the next iteration. The iterations are continued until the convergence is met.