



1 SUMMARY

To calculate the best, **weighted least-squares fit, to given data by a cubic spline** with pre-assigned knots, i.e. if (x_i, y_i) , $i=1,2,\dots,M$ are the data points with weights w_i , $i=1,2,\dots,M$ and y_i , $i=1,2,\dots,N$ are the given knots, this subroutine finds the cubic spline $s(x)$, having knots ξ_i , which minimizes

$$\sum_{k=1}^M [w_k (y_k - s(x_k))]^2 \quad N \geq 2, M \geq 1$$

Note that the spline is defined only in the interval between the knots, i.e. $\xi_1 < x < \xi_N$, and this interval must include all data positions.

ATTRIBUTES — **Version:** 1.0.0. **Types:** VB05B, VB05BD **Calls:** FD05. **Original date:** 1966. **Origin:** M.J.D. Powell.

2 HOW TO USE THE PACKAGE

2.1 The argument list and calling sequence

The single precision version

```
CALL VB05B(M,N,XD,YD,WD,RD,XN,GN,IPRINT,W)
```

The double precision version

```
CALL VB05BD(M,N,XD,YD,WD,RD,XN,GN,IPRINT,W)
```

Note that XD, YD, WD, RD, XN, FN, GN, W are all one-dimensional arrays which must be declared in the user's calling program. Only the variables FN, GN, WN, RD are changed by this subroutine, all others retain their input values. The input parameters, set by the user's program before calling VB05B are as follows.

M is an INTEGER specifying the number of data points (x_k, y_k) .

N is an integer specifying the number of knots k .

XD is a REAL (DOUBLE PRECISION in the D version) array, of length at least M, containing in its first M elements the abscissas of the data points (x_k, y_k) , i.e. $XD(k) = x_k$, $k=1,2,\dots,M$.

YD is a REAL (DOUBLE PRECISION in the D version) array of length at least M, containing in its first M elements the data values y_k , i.e. $YD(k) = y_k$, $k=1,2,\dots,M$.

WD is a REAL (DOUBLE PRECISION in the D version) array, of length at least M, containing the values of the weights w_k associated with the data, i.e. $WD(k) = w_k$, $k=1,2,\dots,M$.

XN is a REAL (DOUBLE PRECISION in the D version) array of length at least N, containing the knot positions ξ_k , i.e. $XN(k) = \xi_k$, $k=1,2,\dots,N$

IPRINT is an INTEGER parameter to request printing by subroutine VB05B/BD as follows:

if IPRINT > 0 details of the fit will be printed

if IPRINT ≥ 0 no printing other than possible diagnostics, see section 3.

W is a one-dimensional REAL (DOUBLE PRECISION in the D version) array of length at least 2N+6M+13. The input values in this array will not be used by VB05B/BD but they will be destroyed. W is a work area for the subroutine.

The output parameters, set by VB05B to define the fit, are as follows.

RD is a REAL (DOUBLE PRECISION in the D version) array of length M, in which VB05B/BD stores the residuals, i.e.

$$RD(k) = YD(k) - s(XD(k)) = y_k - s(x_k), k=1,2,\dots,M.$$

FN is a REAL (DOUBLE PRECISION in the D version) array of length N, set by VB05B/BD to the values of the spline $s(x)$ at the knots, i.e. $FN(k) = s(\xi_k)$, $k=1,2,\dots,N$.

GN is a REAL (DOUBLE PRECISION in the D version) array of length N, set to the values of the first derivative of the spline at the knots, i.e. $GN(k) = s'(\xi_k)$, $k=1,2,\dots,N$.

Note : the spline $s(x)$ is defined completely by N, XN, FN, GN, and can be evaluated for any value of x by Harwell Library subroutine TG01B/BD.

3. DIAGNOSTIC PRINTING

If the data are too sparse, for example the number of data points, M is small compared with the number of knots, N (for instance $M < N+2$), then the minimising spline is not uniquely determined by the least squares condition. The ambiguity is resolved by effectively removing certain knots, i.e. foregoing the requirement that $s(x)$ should have a continuous third derivative at these knots. Another irregular case is when $M < 4$ in which case $s(x)$ degenerates to a single polynomial of degree less than M. In both these cases a suitable diagnostic is printed stating which knots have been forced to be inactive. These knots are also planted in the leading elements of the array W so that the user's program can recognise them if necessary.

4 RESTRICTION

The sizes M and N and the orderings of all points are subject to the following restriction

$$M \geq 1$$

$$N \geq 2$$

$$\xi_1 < \xi_2 < \dots < \xi_{N-1} < \xi_N$$

$$\xi_1 < x_1 < x_2 < \dots < x_{M-1} < x_M < \xi_N.$$

5 METHOD

The required spline fit is regarded as a linear combination of Fundamental Splines (which are non-zero only in the interval between 5 consecutive knots). Consequently the least squares equations have at most 4 non-zero elements in each row. This accounts for the small amount of work space required, and allows an efficient algorithm to be used for the solution of the least squares equations. Householder orthogonal transformations are applied to the least squares matrix of coefficients, reducing it to an upper triangular form. A back-substitution yields the required multipliers of the fundamental splines. The elements FN, GN are then obtained by straightforward calculations.