

# **Practical tips for efficient generalised linear modelling using SAS under NT**

**Richard Bland**

## **Summary**

This is a preview note for a workshop. The workshop will cover:

- guidelines for memory usage and time taken with Genmod
- preparing data for more efficient running with Genmod
- tuning NT for greater performance
- practical demonstrations of performance improvements.

## 1. About Genmod

Genmod is a procedure which is included in the Stat module of SAS. It was introduced in version 6.09 and has been widely used within the insurance industry as one of the software products to offer a generalised linear modelling solution. As SAS becomes more of a standard within the industry, there has been a demand for more information on this product, particularly in relation to its performance and ways of enhancing that performance.

Linear modelling is a computationally difficult task, an iterative process requiring large scale matrix arithmetic. This is not made any easier by the way in which general insurers have used it for premium rating. For most insurance products, there are a large number of rating factors many of which may have relatively weak effects on the claim frequency and average claim size. To produce statistically significant models of these effects, large amounts of data must be supplied and this makes the task all the more difficult for both the hardware and software. If it were easy to create and solve linear models, there would be little concern about efficiency in the process, but where hardware is being pushed to the limit, efficiency becomes a matter of great importance for the modelling process.

## 2. Resource usage

One of the most common demands for information on Genmod relates to the memory usage and time taken by the process. Bluntly, most users wish to know how large and powerful a machine they will need for the task. One of the primary goals of the workshop is to show the dependencies of Genmod's resource usage on the dimensions of the model presented - the number of observations being used in the data, the number of factors being modelled and the number of levels which categorical factors have. These have been obtained by (a) testing a number of models of different dimensions, (b) confirming the formulae which have been deduced with the SAS developers in the USA.

## 3. Data preparation

One of the keys to efficient modelling is efficient preparation. Much time can be spent in data preparation, so speeding up this part of the modelling process can be a significant gain in itself. As importantly, however, is the performance effects which can be gained from presenting data efficiently to Genmod. The workshop will contain some discussion of (a) how the special features of models commonly used in premium rating can be exploited to

compress data used for modelling, and (b) how results from models can be efficiently combined with source data using efficient SAS macro techniques.

#### 4. Tuning NT

NT has become the standard operating system for PCs within the business environment, but many PCs found in the office are not appropriately set up for the demands of the heavy statistical user. Genmod will require absolute peak performance not only from the PC itself, but also from its operating system. The first step in this process is to look at ways of monitoring the performance of the PC, determining the resource usage of Genmod and finding the bottlenecks in the system. The second stage of performance enhancement is to look at ways of getting the most out of a high spec PC. Relatively simple alterations to the operating system (and ensuring that the right service packs are used) can significantly increase the size of models which can be attempted and improve the performance of very large models.

#### 5. Practical demonstrations

The workshop will include practical demonstrations to show how performance improvements can be implemented. Examples of SAS coding will be shown to illustrate the points made. This will include simple examples of Base SAS programs as well as more complex programs created with a code generation system. Participants are positively invited to bring examples of thorny SAS problems along to the workshop.