This chapter examines two systems and builds simulation models and metamodels for them in much the same way as was done in Chapter 4 for the $M/M/s$ queuing system. In both cases, the metamodels produced are intuitively reasonable and also pass tests of validity. They are simpler and less costly to use than the simulation model and serve as auxiliary models to the simulation model, especially with regard to achieving an understanding of how the simulated variables work together.

EXAMPLE: A TIME-SHARED COMPUTER SYSTEM

Friedman (1989) simulated the time-shared CPU queuing system studied by Adiri and Avi-Itzhak (1969). This queuing system was first adopted into the realm of simulation research by Sargent (1976) and used extensively by Law (e.g., Law and Kelton 1991). The advantage of this system, in addition to the fact that solutions to several measures of system
performance are obtainable, is that the system is relatively complex and similar to a large class of queuing systems of real practical importance to simulation researchers in particular and management scientists in general.

The Simulation Model

The system consists of a single central processor and \( N \) terminals. Each terminal is in think state for an amount of time that is exponential with mean \( 1/\lambda \), at the end of which it sends to the CPU a job that has a service request for an amount of time which is exponential with mean \( 1/\mu \). Job requests to the CPU are served in round-robin fashion: each is allotted only a quantum of time before the next is served. The constant quantum is divided into two parts: \( \tau \), that due to the overhead of the round-robin scheduling and \( \theta \), that involved with actual processing.

SIMSCRIPT II.5 (CACI 1983) was used to build the simulation model of this system. Quantum overhead, \( \tau \), was set to .015 seconds for all runs. Similarly, the load on the system, given by \( \frac{N\lambda}{\mu} \), was maintained at .75 for all system variants simulated. This was done by setting the values of \( N \) and \( \mu \) and adjusting \( \lambda \) to ensure a load of .75.

Each independent replication was allowed to run until 1000 jobs had been processed, at which point the program was suspended and measurements