Appendix D
Dynamic Programming

Fundamentally, dynamic programming addresses how to make optimal decisions over time. While it can be applied to both deterministic and stochastic problems, our focus here is on stochastic problems because making decisions under uncertainty is central to revenue management. Our treatment largely follows that in Bertsekas [57] with some slight variations in notation. We summarize only the key results for the discrete-state, discrete-time, finite-horizon problem, again because it is the most frequently encountered one in RM. The reader is referred to Bertsekas [57] for an extensive treatment of other cases of dynamic programming and a discussion of further theoretical and computational issues.

Elements of a Dynamic Program

Dynamic programming involves the optimal control of a *system* over time. The system is dynamic and its *state* evolves over time as a function of both *control decisions* and *random disturbances* according to a *system equation.* The system generates *rewards* that are a function of both the state and the control decisions. The objective is to find a *control policy* that maximizes the total expected rewards from the system.

There are T time-periods. Time is indexed by t and the time indices run forward, so $t = 1$ is the first period and $t = T$ is the last period. The key elements of a dynamic program and related technical assumptions are

$\mathbf{x}(t)$ The *system state.* Assumed to be discrete and belonging to a finite-state space S_t.

$\mathbf{u}(t)$ The *control decision.* Assumed discrete and constrained to a finite set, $U_t(\mathbf{x}(t))$, that may depend on time t and the current state $\mathbf{x}(t)$.

$\mathbf{w}(t)$ The *random disturbance.* Assumed to be a discrete random variable (or vector) with known distribution, belonging to a countable state space W_t. The disturbances $\mathbf{w}(t), t = 1, \ldots, T$ are independent.

$\mathbf{f}_t(\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t))$ A *system function,* which determines the next state as a function of the current state $\mathbf{x}(t)$, the decisions $\mathbf{u}(t)$ and disturbance $\mathbf{w}(t)$, according to the *system equation:*

$$\mathbf{x}(t + 1) = \mathbf{f}_t(\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t)).$$