OPERATIONS RESEARCH CENTER
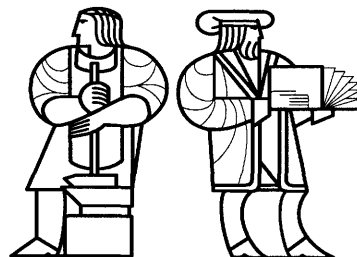
Working Paper

Extensions to the Dynamic Requirements Planning Model

by
John Ruark

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

*Extensions to the Dynamic Requirements Planning
Model*

by
John Ruark

OR 326-98   April 1998

Extensions to the Dynamic Requirements Planning Model

John D. Ruark

Massachusetts Institute of Technology
Cambridge, MA 02139
ruark@post.harvard.edu

A recent development in requirements planning for multi-stage production-inventory systems is the Dynamic Requirements Planning (DRP) model of Graves, Kletter, and Hetzel (1994). The DRP model captures the impact of a flexible, dynamic forecast model of demand on the variance of planned production levels and forecasted inventory levels given a specified policy. This paper examines three potential extensions to the original DRP model to account for periodic production, lot-sizing, and capacitated production.

Introduction

A recent development in requirements planning for multi-stage production-inventory systems is the Dynamic Requirements Planning (DRP) model of Graves, Kletter, and Hetzel (1994). This model captures various characteristics of the planning process, particularly the impact of a flexible, dynamic forecast model of demand on the variance of planned production levels and forecasted inventory levels given a specified production policy. The steady-state performance measures analyzed are simple to express and easy to compute. It is important to obtain measures of variance of production and inventory in order to set parameters such as safety stock levels, yet there is little literature on these measures for multi-stage production-inventory systems.

Researchers at General Motors, along with the authors, have identified several limitations in the original DRP model, and have expressed interest in extending the DRP model framework in order to overcome these limitations. The desired end-product is a model which retains the flexibility of the demand and production specification of the DRP model while incorporating consideration of periodic or batched production, lot-sizing, production capacity constraints, and inventory shortages.

We have begun to develop these extensions and to identify avenues of exploration which through further research could satisfy the desired end-product requirements of flexibility and ease of computation. This paper is the result of my work at the General Motors Technical Center during the summer of 1994 and at the Operations Research Center at MIT this year.

This paper will review the DRP model and will contextualize it and the desired extensions within the literature. It will then present our current results on extending the DRP framework in the following directions:

Periodic Production. Due to its steady-state nature, the DRP model treats every period in the same manner; in particular, if there can be production in any period, there can be production every period. We have adapted the model to permit production cyclically, only every P periods, for a given P.

Lot-sizing. The original DRP model assumes continuously-valued demand and production levels. It is possible to incorporate lot-sizing effects if the analysis may be done with some integrality constraints. How to add such constraints and then analyze them is an important issue. Another way to model lot-sizing is by treating non-lot-sized production plans as inputs into a traditional lot-size inventory model and deriving performance characteristics. Some other possibilities are mentioned, but not explored.

Capacitated Production. As a design and analysis tool, the DRP model does not include any capacity constraints; one of its uses is to determine what capacity levels should be, based on production variance levels in an uncapacitated environment. However, some interesting results can be achieved when capacity is incorporated into the model.

While useful themselves, these results generate more questions than they answer. We conclude with an appendix of some models and directions we partially explored but which remain for further analysis.

1

■ 1. Overview of Dynamic Requirements Planning

This section provides a brief overview of the major assumptions and results that we will need from the Dynamic Requirements Planning (DRP) model for single stage analysis. Where possible, I have copied the original equations as (DRP #) from Graves, Kletter, and Hetzel (1994).

For our purposes, the main results of DRP are the equations for variances for certain performance measures of variance at a processing stage which plans production based upon forecasted demand according to a linear control rule we will specify shortly. In each period the forecast process generates a new forecast for demand for the current period and the H subsequent periods; H is called the horizon of the forecast process. These forecasts are determined by a revision to the previous period's forecast by a revision process. Thus:

- $f_t(t + i)$ is the forecast in period t of demand for period t+i, for $i = 0, ..., H$,

- $\Delta f_t(t + i) = f_t(t + i) - f_{t-1}(t + i)$ is the forecast revision in period t for demand in period t+i, for $i = 0, ..., H$,

- $f_t = [f_t(t), ..., f_t(t + H)]^T$ is the (H+1) vector of forecast demand over the horizon at period t, and $\Delta f_t = [\Delta f_t(t), ..., \Delta f_t(t + H)]^T$ is the (H+1) vector of forecast revisions over the horizon at period t.

It is assumed that beyond the horizon the forecasted demand is μ, which is the average demand per period; thus, $f_t(t + i) = \mu$ for $i > H$ for all t. Forecast revisions are assumed to be i.i.d with $E[\Delta f_t] = 0$ and $Var[\Delta f_t] = \Sigma$; Σ is the covariance matrix of the forecast revision process. Justification for an expected revision of zero is provided in Graves, Kletter, and Hetzel (1994), as well as Heath and Jackson (1994) and Hausman (1969).

From this revision process, the planned production outputs are determined. Mirroring the forecast process, new planned production outputs for each period over the horizon are generated.

- $F_t(t + i)$ is the plan in period t for production in period t+i, for $i = 0, ..., H$,

- $\Delta F_t(t + i) = F_t(t + i) - F_{t-1}(t + i)$ is the plan revision in period t for production in period t+i, for $i = 0, ..., H$,

- $F_t = [F_t(t), ..., F_t(t + H)]^T$ is the (H+1) vector of planned production over the horizon at period t, and $\Delta F_t = [\Delta F_t(t), ..., \Delta F_t(t + H)]^T$ is the (H+1) vector of planned production revisions over the horizon at period t.

These planned outputs are determined by a planned production revision, which itself is determined from the demand forecast revisions, as follows:

$$\Delta F_t(t + i) = \sum_{j = 0}^{H} w_{ij} \Delta f_t(t + j) \qquad \text{(DRP 6)}$$

or, in matrix notation,

$$\Delta F_t = W \Delta f_t \qquad \text{(DRP 7)}$$

where we call $W = \{w_{ij}\}$ the weight matrix. It is assumed that $0 \le w_{ij} \le 1$ and that $\sum_{i=0}^{H} w_{ij} = 1$ (the columns of W each sum to unity), which implies that the forecast of inventory for the end of the horizon is constant (equation DRP 5).

Performance Measures

Covariance of planned production revisions. The expectation and covariance of planned production revisions are shown to be

$$E[\Delta F_t] = 0$$
$$Var[\Delta F_t] = W \Sigma W^T \qquad \text{(DRP 14)}$$

These values can serve as the first two central moments for a demand process which feeds the stage upstream of the current stage. Since the characterization of these parameters is identical to that of the demand process (expectation zero and some constant covariance matrix), the "output" of the model, the planned production revisions, is stochastically similar to the "input" of the model, the forecast demand revisions. This makes the multistage analysis of a network simpler.

Variance of production. The expectation and variance of actual production (actual production output in period t is given by $F_t(t)$) are given by

$$E[F_t(t)] = \mu$$
$$Var[F_t(t)] = tr(W \Sigma W^T) \qquad \text{(DRP 12)}$$

These values characterize the actual performance of the stage.

Variance of inventory. The expectation and variance of actual final goods inventory $I_t(t)$ are shown to be

$$E[I_t(t)] = SS$$
$$Var[I_t(t)] = \sum_{k=0}^{H} \sum_{i=0}^{k} \sum_{j=0}^{k} a_{ij} \qquad \text{(DRP 20)}$$

where $A = \{a_{ij}\} = [W - I] \Sigma [W - I]^T$, and SS is the specified base-stock level. In general, SS will be chosen based upon the variance of the inventory.

3

■ 2. General Motors Research and DRP

The motivation for this research came from the generous researchers at the General Motors Research Laboratories in Warren, Michigan. Having already established a working relationship with Professor Graves and David Kletter (co-author of the DRP paper), they wished to examine the potential for DRP's application to their research efforts. Early analysis led them to the conclusion that the while DRP provides all the benefits discussed in this paper and the original paper, the model ignores some factors which are deemed of high importance for General Motors. My task for the summer was to determine feasible avenues of exploration for overcoming these identified limitations of the DRP model.

The initial thoughts about using DRP were based on its applicability to a stamping plant environment. A stamping plant is a mid-level processing and assembly center, where rolls of steel are cut into blank pieces (the blank and shear process) which are then stamped on press lines to make panels, roofs, hoods, and other numerous parts. Various subassemblies such as doors are made at stations called metal assemblies. For a more concrete description of a stamping plant and an analysis of lot-sizing and scheduling at such a plant, see Kletter (1994).

Our viewpoint of the stamping plant thus drove most of our thoughts about the DRP model. It gave us two contexts in which to place various alternatives for how DRP might fit with General Motors' needs. First was the relationship between the stamping plant and other plants in the supply chain. The demand process can be viewed as the assembly schedule handed down to the stamping plant from the assembly plant; the upstream stage in the process is an aggregation of the steel companies from which the plant received its steel rolls. Second was the internal operation of the stamping plant, where the stages are the blank and shear, the stamping press lines, and metal assembly. Put together, these two views form a three stage system (blank and shear, the press lines, and metal assembly), with an external demand from the assembly plant feeding the metal assembly stage, and an external raw materials provider (steel companies) for the blank and shear.

A central problem for this stamping plant was the instability of its forecasts to its suppliers. Even though the assembly plant's requirements were relatively stable, as the requirements moved upstream production became more variable and unpredictable. This was due to batching, because parts shared machine resources for production, and lot-sizing effects, because of lot-size runs at press and of the sheer size of a roll of steel, which generated hundreds of blanks of a particular size. (There was not much commonality of the blanks as sources for stamping; two different types of stamped parts normally used different types of blanks.) The standard concerns associated with unstable forecasts of requirements were supplemented with a problem particular to this plant, which was that they had to freeze their orders four weeks in advance due to contractual obligations with their steel suppliers. The standard GM forecast horizon was twelve weeks, so this fro-

zen period represented a significant fraction of their horizon.[1]

Seemingly, DRP would be able to serve as a planning or performance analysis tool for this system. However, GM recognized four limitations of the DRP framework which would hinder its applicability to the stamping plant model. First, the processes in the DRP analysis are steady-state and do not differentiate between particular periods. It is not possible to directly codify a production schedule which allows production in some periods but not in others. For example, there is the Fixed Time, Variable Quantity (FTVQ) schedule, which is of great interest to General Motors researchers. In a FTVQ policy, production of all parts is broken down into cycles, and each part or product is processed at the same point in the cycle every time; enough is produced to cover orders or lead time demand. If the cycle is one week, then a part might be setup to run at 1:00pm every Tuesday, and may run to 4:00pm or 5:00pm depending on the number of parts needed. In this paper, this is called a batching phenomenon, because on a single machine different parts are run in batch mode, although not in fixed lot sizes.

Second, the processes in the DRP analysis are continuously valued, and the relationships between production revisions and forecast revisions is linear. While this greatly eases the analysis and computational effort, it does not allow ready analysis of production policies which incorporate lot-sizing effects. The analog to the FTVQ is Variable Time, Fixed Quantity (VTFQ), which is a lot-sizing policy. Here, a part is produced whenever it is needed (and can be scheduled in relation to other parts and their demands), but once a production run is started, a fixed amount is produced. An alternative is once a production run is started, any number of lots are run, where the lots are of a fixed size. Overcoming these first two limitations by similar analytical means would make the comparison of FTVQ and VTFQ policies comparable within the DRP framework; such comparisons are of great interest.

Thirdly, the DRP model does not directly incorporate analysis of production capacity constraints. A defense of this limitation is that DRP may be used as a capacity planning tool, to determine variance of production to then set capacities. However, this does not discount the value of modeling capacity within the DRP framework. For instance, examining networks of stages with capacity would be more accurately analyzed if the model itself included capacity constraints than if it did not. Also, most of the stages in the stamping plant are heavily loaded, and examining the impact of capacity constraints is important because often the constraints are imposed rather than decided. One use of the DRP tool would be to use it as a submodel in a larger system which examined when to run multiple product families on a single capacitated machine. There could be iterations between the larger system, which might impose production capacities on individual families (by specifying how much processing time they will have), and the DRP submodel, which returns

---

1. It would be interesting to examine their forecast data to determine what percentage of the forecast variability is accounted for in the final four weeks before demand is realized; such an analysis on different data is summarized in Heath and Jackson (1994).

5

performance results given the capacities. However, to even begin to think about doing that, DRP would have to include capacity constraints.

Finally, the buffers between stages in a DRP system never cause blockage or starvation of processors. At GM's particular stamping plant, some metal assemblies were occasionally starved, and there was interest in trying to capture that within the DRP framework. The same defense as used for the previous can be used for this limitation, namely that DRP is meant to serve as a planning tool for safety stock determination, and that for those times when inventory would stock out, some extraordinary means are undertaken (such as expediting or out-sourcing) to make sure a processor is never starved; see Graves (1988) for a discussion of such assumptions. Again, while this is true, there was an expressed interest in trying to capture inventory shortages.

To recap, the four extensions we would like to add to the DRP framework are:
- Production batching
- Production lot-sizing
- Production capacity constraint
- Inventory starvation

State of the Research

Before examining some of the features of the DRP in the context of the literature in the next section, it is best to quickly summarize the results to date.

Production batching has been sufficiently incorporated into the DRP framework. We have extending the way the weight matrices are defined to permit production in fixed cycles, say once every P periods. These results are presented in the fourth section.

The production lot-sizing problem has been more difficult to solve, and we have had to this point to use different tools to approach the problem. There are some promising directions, but no finished, verifiably correct results. One direction is described in the fifth section, while other, unexplored options are mentioned in the appendix.

Production capacity has proved even more difficult. We have some results for the case where demand is independent and production does not introduce any dependencies. These are obtained by approximating the production process as a diffusion and analyzing the resulting process. We have run into some difficulty, though; these results are summarized in section six.

We have not explored the inventory starvation/shortages problem at all. The paper concludes with an appendix which describes, for the sake of completeness, some (apparently) failed attempts and some unexplored possibilities.

6

▓ 3. Literature Review

At the heart of production analysis lie three components. These are the specification of the demand process, the production policy or the set of feasible production policies, and some cost/ profit/value structure for these policies. Factors such as the echelon structure, lead times, procurement and stocking policies, setup times and configuration, and push or pull production can in a holistic sense be captured within the set of feasible production policies.

The dynamic requirements planning model (Graves, Kletter, and Hetzel, 1994) is an analysis tool for which the demand process and production policies are specified. (The third section of their paper examines a particular cost structure applied to the basic DRP model.) In particular, the demand process is characterized as stochastic but stationary, with forecast revisions on a rolling forecast horizon. The forecast revisions specify the change between the forecast made in the previous period with the forecast made in the current period in forecasted demand for some point in the future. The revisions are assumed to be zero-mean and may be correlated. These correlations permit either independent or correlated demand process, and since the two are characterized by the same demand process and undergo the same production policies, the DRP model facilitates the direct comparison of independent demand and correlated demand. Similarly, the DRP model enables the analysis of different policies under correlated demand. In particular, if there is a policy for which there exist results for independent demand and which can be modeled within the DRP framework, then the performance of the policy under correlated demand can be analyzed, using independent demand processes as a benchmark between the DRP model and the original results.

The specification of the production policy, as I have defined it above, is locally precise and globally abstract in the DRP model. Namely, it focuses on a single production stage without considering the system architecture surrounding this stage. This is accomplished by retaining a simple consistency between the demand process seen by the single stage, and the raw materials requirements process generated by the single stage as a result of its production policies. In particular, the requirements process generated by the stage has the same mathematical form as its demand process, and so can function without alteration as the demand process for another stage. Furthermore, stages may be designed and interpreted as representing any part of the production system, from transportation to stocking and shipping. Within each stage, the production policy is well defined (by the weight matrix, as explained in the previous section). Thus, the DRP model permits wide latitude in system and echelon structure modeling, but requires fitting particular production policies within the DRP single stage analysis framework.

This, in general, is a desirable property of multi-stage production models: that the stages (or echelons) decouple from each other and thus may be analyzed in isolation. In the DRP model, this is accomplished by assuming that inventory levels may be negative; that is, there are no constraints on the levels of inventories between stages. Since between each stage there is an inven-

tory, not having constraints on these inventories entirely decouples the system. This does not consider the case of assembly and distribution; it requires more details to determine how demand processes are generated at distribution stages and how requirements needs are defined at assembly stages. Furthermore, it would require extra work to compute steady-state performance measures for a production network with a cycle, since the standard DRP results assume all demand processes for a given stage are known prior to analyzing that stage; in a cycle, there's nowhere to start, and some iterative convergence method might be required.

The DRP model is thus characterized by three fundamental features. The first is the decomposition of a multi-stage production architecture into its single-stage components. This is accomplished by the second feature, the definition of a demand process, and the third feature, the specification of a single-stage production policy based on the demand process, so that this decomposability is maintained across stages. It is interesting to examine how these three features are utilized in other models in the literature.

The Demand Process

The papers by Heath and Jackson (1994) and Jackson, Heath, and Muckstadt (1993), describe what is essentially the same demand process as in the DRP model. In these papers, they call it the Martingale Model of Forecast Evolution (MMFE), because forecasted demand for any given demand period is a martingale process (because the revisions have zero expectation)[2]. In Heath and Jackson (1994), a large portion of the paper is devoted to explaining and justifying this type of demand process. They emphasize the importance of examining correlations in demand when investigating production and distribution flexibility potential and in analyzing production and inventory plans. They propose two types of MMFE's, the additive and the multiplicative. The *additive MMFE* is exactly like the DRP demand process, except Heath and Jackson do not specifically define a finite forecast horizon *a priori*. The *multiplicative MMFE* is similar; instead of adding the new forecast revision to the previous forecast, the previous forecast is multiplied by the forecast revision to obtain the new forecast. This enables a stationary revision process to generate changes in forecasts which are proportional to the size of the previous forecast (if the forecasts were 50 and 100 for two separate products, and both had forecast revisions of 1.1, then the new forecasts would be 55 and 110, so that one increases by 5, the other by 10, even though the revision factor was the same). They claim this model permits the analysis of cyclical demand which is not necessarily stationary. The problem with this is that the revision process is still stationary; while there may be a cyclical demand, it may not follow the cycles of any real demand

---

2. I.e., if the forecast made at time t-i for demand in period t is $f_{t-i}(t)$ then because $E\left[\Delta f_{t-i}(t)\right] = 0$ and

$f_{t-i}(t) = f_{t-i-1}(t) + \Delta f_{t-i}(t)$, then $E\left[f_{t-i}(t)|f_{t-i-1}(t) = x\right] = x$, and so the forecasted demand for period t forms a martingale process for i decreasing (i.e., for periods closer to the demand realization).

because the revisions cannot be forced to increase demand at the appropriate times and then decrease it at the end of a real cycle.

Heath and Jackson examine a well defined production architecture with production policies generated from a linear program which uses the forecasted demand as its inputs. They implement a simulation study to calculate expected costs over long simulation times in order to compare two different forecasting methods. In these two respects their analysis differs from the DRP analysis, which leaves the system architecture open-ended and defines the single-stage policy differently, and which uses a direct analysis to compute performance levels. Optimization of system-wide parameters is ruled out of the question because of the stochastic nature of the demand and the presence of a linear program that's used every period to determine production levels. The MMFE models lend themselves to easy forecast generation for simulation.

Jackson, Heath, and Muckstadt (1993) use the multiplicative MMFE model to feed a cost-driven simulation which examines customer service, inventory levels, and the need for expediting. Comparisons are made between demand forecasts of different variability to show that as forecasts are more variable, costs rise and fill rates decrease. However, with all reported fill rates above 99.3%, the capacity and expediting constraints on the system may not have been tight enough to bring out the worst in highly variable forecasts.

As Heath and Jackson note, their model of forecast evolution, and thus the DRP's model, is a generalization of the model presented by Hausman (1969). In that paper, Hausman develops a dynamic program to optimize costs for a quasi-markov system in which the state variable is replaced by a forecast of the state variable, and the forecasts are revised each period by a multiplicative factor which is lognormally distributed. (A system is quasi-markov is the state variable in the next period depends only on the state variable in the current period, and on some other control factor in the current period; thus inventory levels could be a state variable, where stochasticity is imposed by a demand process on that inventory, and the control variable would be how much to product each period to fill the inventory.) Hausman considers mainly the cases where demand is actually realized only in the final period, and the state variables each period are the forecasts for that final period's demand. The MMFE multiplicative model is thus a generalization in that it permits demand in every period and defines the forecast over an arbitrary horizon. The model of Hausman and Peterson (1972) is a specific application of Hausman's quasi-markov forecast model to a style-goods problem; in that paper, they show how to structure costs for the dynamic program, and develop some heuristics for approximating an optimal solution.

The Production System Architecture

Let us now turn our attention to the system architecture, the connection of various production stages, sources of supply, and stocking locations. In the DRP model, the architecture is characterized by complete decomposability of stages, where each stage is modeled mathematically identi-

cally. An ancestor of this architecture is the tactical planning model (TPM) of Graves (1986). In this paper, a job shop is modeled as a network of workstations (stages) which receive work, measured in a uniformized workload unit, from other workstations or from the outside, process this workload based on a simple linear control rule, and then send the completed work onto other workstations or to the outside. The control rule simply states, given a certain level W of work to do, complete p*W where p is between 0 and 1, and constant for that workstation. Since the work has been uniformized into standard units, each workstation operates identically except for the linear control rule and the routing of work into and out of the workstation. The TPM assumes a discrete time model, an infinite history for steady-state analysis, continuous workload levels, no production capacity, and Markovian workflows. These same assumptions are made in the DRP model. Furthermore, the TPM allows for stochastic noise to impact workload levels; this noise can be correlated across workstations, and the analysis takes these correlations into account. The TPM is similar to queueing network models in its routing probabilities, but the correlation of noise across workstations gives it some interesting potential.

If we assume there are no cycles in the TPM, then it decomposes exactly as the DRP model does; that is, for any given stage, if the input processes are known with certainty (in steady-state), then the output steady-state process may be determined, and that may feed all further downstream stages. Thus, conceptually the analysis moves from upstream to downstream. However, one of the strong points of the TPM is its arbitrary routing schemes, and cycling may be an important part of routing (for rework or repair, reprocessing, or simply reuse of a workstation for a different purpose). The TPM easily handles this analysis by doing the computations for steady-state distributions for the entire system at the same time. Thus, even though the stages decompose if there are no cycles, the analysis of the entire system is, in general, simple enough to do without decomposition. We expect that this is something the DRP model cannot do so easily, because of the complexity of its demand process, although the approach might be similar.

In an extension to the TPM (Graves, 1988), a single workstation with a production capacity constraint is analyzed by approximating the station's queue by a diffusion process which is fed by the arrival of work to the station and processed by the linear control rule in the TPM. We will use this result as a first-cut approximation to the capacity constraints in the DRP model. It works reasonably well because of the similar assumptions made in the TPM and the DRP. The extension assumes the interarrival process is independent, which is a problem; under correlated demand, the DRP model with capacity, as based on the TPM extension, does not perform well because of that independence assumption.

The literature review in Graves (1986) discusses briefly two classical queueing models, the Jackson network and the Queueing Network Analyzer (QNA), which treat networks of processors as the TPM does. The behavior of each processor decomposes from the other processor; the processors are related only by the average arrival rates (and in the case of the QNA, coefficients of

variation) between them. The Jackson network models precisely M/M/m processors, while the QNA models approximately G/G/m processors (Whitt, 1983). These two systems are based on continuous-time operation, which is a slight difference.

The Production Policy

The method of using weight matrices to generate planned production outputs which can then be used as demand forecasts for upstream stages is an effective way to retain the decomposability of stages across a network. The models discussed in the previous section have the same feature. For general networks, it is interesting to examine whether some production policies within the realm of periodic-review extend easily to multistage analysis or not.

The safety stock model (Graves, 1988) uses a linear control rule with a smoothing constant just like the TPM. The focus of this model is to determine the requisite levels of safety stock for specified service levels. A single stage follows a base-stock policy, producing to fill depleted demand according to the given linear control rule. The stage induces demand for its raw materials, and this demand may feed other stages in a network. Because the model produces in each period, and because the production requirements process, may be characterized similarly like the demand process, the stages decompose. This in general is true of a network of processors which are not restricted to ordering their raw materials in batches or fixed lots, and which produce every period.

For example, the presence of a simple $<Q,r>$-ordering stage in a network might pose problems for the stages upstream of the $<Q,r>$ stage. This is because the stage orders in multiples of Q; if the external demand process is a continuous random variable, then it becomes necessary to accommodate both types of demand. It may be necessary to approximate a $<Q,r>$ ordering policy to two moments by a normal distribution, or something similar.

■ 4. Fixed Time, Variable Quantity Policy Analysis

We wish to examine production policies of the form "produce any amount once every P periods." Thus, production periods form cycles of producing one period and not producing for (P-1) periods. We say that periods in which there will be production are production periods, and those that are not are non-production periods. The production decision becomes how much to produce in the future production periods.

The standard DRP model cannot exactly capture this policy, because in steady-state every period is treated the same, because the weight matrix remains constant from period to period. So the DRP model cannot capture the following effect: If the current period and the next two periods are non-production, production, and non-production periods respectively, then it is necessary to not assign production revisions to the current period or to two periods from now, while production revisions may be assigned to the next period. This would imply the first and third rows of the

11

weight matrix should be zero, and the second row might be non-zero. If the clock advances, and the current period becomes what was the next period, then the system changes. Production is now possible in the current period but not the next period, implying the first row of the weight matrix may be non-zero while the second row must be zero. Thus, there is an inconsistency in the requirements of the weight matrix when we try to model this production policy. It seems there should be different weight matrices from which to choose.

There are two methods of attacking this problem. The first is to approximate this policy with one weight matrix and use the standard DRP model. The second is to extended the DRP model to explicitly account for this production policy. There are some advantages and disadvantages to both methods, and these will be discussed at the end of this section. In order to analyze this policy, it is necessary to modify the production horizon. How this is accomplished is discussed in the following section.

Changing the Planned Production Horizon

This section describes a simple modification to the DRP model which allows the analysis of the case where the planned production horizon is not the same as the demand forecast horizon. The motivation for this change comes from examining the impact of leadtimes on a system. With DRP, modeling a leadtime of L requires that the first L rows of the weight matrix be zero. If the demand forecast horizon is H, this implies that H+1 periods of forecast revisions can be allocated over only (H+1-L) production periods (as specified by the last H+1-L rows of the weight matrix). It would be preferable to have H+1 periods of planned production over which changes can be allocated. (In fact, a strict pull system with nonzero leadtime requires this.) This can be handled by making the planned production horizon H+L periods, so that the first L periods are fixed (by the leadtime) while the remaining H+1 periods are revisable.

The basic DRP model imposes only three constraints on the production policy for a single stage. The first is that planned production has the same horizon as the forecasted demand. Second, each element of the weight matrix is assumed (expected) to be between 0 and 1. Third, the columns of the weight matrix must sum to 1. We are able to maintain the latter two constraints, which preserve end-of-horizon expected inventory at the base stock level, while setting the planned production horizon to any value.

In particular, let the planned production horizon be P. There are now two horizons, H and P. We define a new system horizon S=max(H,P). Then extend the demand forecast revision vector and the planned production revision vector from (H+1) and (P+1) dimensions, respectively, into (S+1) dimensions by zero filling the last elements as needed. The only change we need to make to the DRP model is to say that the weight matrix is now a (S+1) by (S+1) matrix rather than an (H+1) by (H+1) matrix. Thus, $\Delta F_t = W \Delta f_t$ still applies. In fact, all of the performance results

from the original DRP analysis apply, where now the covariance matrix of production is (S+1) by (S+1) rather than (H+1) by (H+1). This implies that the stage upstream and adjacent to this one sees a forecast demand with horizon S rather than H. Thus, each stage in a multistage network may experience different horizons; if this is not desired, then the production horizon for each stage could be set equal to the maximum of all natural production horizons, and the weight matrix can be zero filled for any stage for which its production horizon is less than the maximum.

The weight matrix must be square instead of being (P+1) by (H+1) because the effective demand horizon and production horizon must be the same. What this means is that none of the DRP equations change *per se*; instead, this discussion is one of implementation. Namely, given a forecast horizon and a potentially different production horizon, how do we set up the DRP system horizon and the weight matrix at a single stage?

We'll demonstrate this method for the case where we want to have an arbitrary policy with a fixed, known positive lead time of L and a forecast horizon of H. We set the production horizon equal to (H+L) and the weight matrix to the following (H+L+1) by (H+L+1) matrix:

$$W = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & & & \ldots & \ldots & & \ldots \\ 0 & 0 & \ldots & 0 \cdot & 0 & \ldots & 0 \\ w_{L,0} & w_{L,1} & \ldots & w_{L,H} & 0 & \ldots & 0 \\ w_{L+1,0} & w_{L+1,1} & \ldots & w_{L+1,H} & 0 & \ldots & 0 \\ \ldots & & & \ldots & \ldots & & \ldots \\ w_{L+H-1,0} & w_{L+H-1,1} & \ldots & w_{L+H-1,H} & 0 & \ldots & 0 \\ w_{L+H,0} & w_{L+H,1} & \ldots & w_{L+H,H} & 1 & 1 & 1 \end{bmatrix} \tag{1}$$

where rows 0 through L-1 are zero, and columns H+1 through P+1 are zero in all elements except the last, which is one. The last columns are defined that way to satisfy the constraint that the sums of the columns equals 1; they will never be a factor in the calculation of the planned production revision because $\Delta f_t(t+i) = 0$ for $i > H$, and so those one's will *always* be multiplied by zero when computing production revisions.

If we want this to be a strict pull system, then set $w_{ij} = 1$ if $i = L + j$, $w_{ij} = 0$ otherwise. It is possible to show then that this weight matrix will correctly model a pull system with positive lead time; had we restricted ourselves to a production horizon of H periods, we could not model such a system correctly.

Extending DRP to account for FTVQ

As usual, we have the demand forecast horizon is H_d and production horizon H_p; these together define the stage's system horizon $H_s = \max(H_d, H_p)$. Suppose production occurs every P periods as defined above. Furthermore, assume without loss of generality that the system horizon, H_s, satisfies the condition that $H_s + 1$ is an integral multiple of P (i.e, the number of rows in the weight matrix is an integral multiple of P). This may be assumed because we can extend any production horizon out to the next value of H_s which satisfies the condition.

In each period of operation, the number of periods until the next production period is known. In particular, if the current period is a production period, the number of periods is zero; if production occurs in two weeks, the number of periods until the next production period might be two. The next production period as a relative offset from the current period defines all future production periods as offset from the current period. So, define $\pi(t)$ as the number of periods until the next production period at time t. Then all future production periods occur $(\pi(t) + Pi)$ periods from time t for i=0,1,2,.... As mentioned above, this information implies what the structure of a weight matrix would look like if we want to restrict production revisions to only production periods. In particular, a row j must be zero if $j \neq \pi(t) + Pi$ for all integers i. The other rows, for which $j = \pi(t) + Pi$, may be non-zero. The weight matrix must still satisfy the conditions that the elements are between zero and one and that the columns sum to one. The implication is that for each period there will be a different weight matrix.

Note, however, that the range of $\pi(t)$ is 0,...,P-1; because there will always be production within no more than P-1 periods from now (if there is a production period P periods from time t, then time t is also a production period and $\pi(t) = 0$). Furthermore, $\pi(t)$ decreases by one every period until it reaches zero, when it then becomes P-1. Thus, $\pi(t+1) \equiv (\pi(t) - 1) \mod P$. So, $\pi(t)$ scrolls through its P values cyclically.

In the standard DRP model, the weight matrix is assumed constant for every period. This facilitates a steady-state analysis. We make a similar assumption here, which is that for any given value of $\pi(t) = j$, there is one weight matrix, W_j. There are therefore P different weight matrices, W_i for i=0,...,P-1, and for any given time t the weight matrix used is given by $W_{\pi(t)}$. Thus, we will cycle through the weight matrices in the order W_0, W_{P-1}, ..., W_1, W_0, ...

Example. A sample weight matrix set, for P=3 and $H_s = 5$ is shown in Figure 1.

Now, suppose $\pi(t) = 0$, so that period t is a production period. The production revision equations for periods t through t+3 are:

$$\Delta F_t = W_0 \Delta f_t$$
$$\Delta F_{t+1} = W_2 \Delta f_{t+1}$$
$$\Delta F_{t+2} = W_1 \Delta f_{t+2} \tag{4}$$
$$\Delta F_{t+3} = W_0 \Delta f_{t+3}$$

Another issue is what planned production is for those periods which are beyond the production horizon. In the original DRP model, $F_t(t+i) = \mu$ for $i > H_s$; that is, initially plan on producing the mean demand in each period. In this extension, we must have for $i > H_s$

$$F_t(t+i) = \mu P \qquad \text{if} \qquad \pi(t+i) = 0$$
$$F_t(t+i) = 0 \qquad \text{if} \qquad \pi(t+i) \neq 0 \tag{5}$$

Thus, in production periods production must be sufficient to account for the entire expected demand over a single production cycle.

Variance of Production

We wish to compute both the variance of production for any given period and also the covariance of production revisions, which are passed upstream to the stage's source. The covariance of revisions is a tricky situation, because all periods are no longer identical. This is related to the issue of synchronizing stage production periods; this is discussed at the conclusion to this section.

We can, however, easily compute the variance of production for any period. Obviously, the variance of production for non-production periods is zero, since no production occurs. To com-

$$W_0 = \begin{bmatrix} w_{00} & w_{01} & w_{02} & w_{03} & w_{04} & w_{05} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ w_{30} & w_{31} & w_{32} & w_{33} & w_{34} & w_{35} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ w_{10} & w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ w_{40} & w_{41} & w_{42} & w_{43} & w_{44} & w_{45} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{and} \tag{2}$$

$$W_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ w_{20} & w_{21} & w_{22} & w_{23} & w_{24} & w_{25} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ w_{50} & w_{51} & w_{52} & w_{53} & w_{54} & w_{55} \end{bmatrix}. \tag{3}$$

Figure 1. Sample weight matrix set for P=3 and H_s=5.

pute the variance of production in production periods, we may follow the DRP analysis, taking into account our special weight matrix structure.

We want to find $\mathrm{Var}(F_t(t)|\pi(t) = 0)$, which is the variance of production at period t, conditioned on t being a production period, in steady-state. Working as in with DRP, we have:[3]

$$F_t = BF_{t-1} + \Delta F_t \qquad (6)$$

By back substitution, we have

$$F_t = B^{H_s+1}F_{t-H_s-1} + \bar{\mu} + \sum_{i=0}^{P-1}\sum_{j=0}^{\overline{\frac{H_s+1-P}{P}}} B^{Pj+i}W_{P-1-i}\Delta f_{t-Pj-i} \qquad (7)$$

where $\bar{\mu} = \{\mu_i\}$, $i = 0, ..., H_s$ and $\mu_i = \mu P$ if $i \equiv 0 \bmod P$, 0 otherwise. Thus, as in the DRP analysis,

$$\mathrm{Var}(F_t|\pi(t) = 0) = \sum_{i=0}^{P-1}\sum_{j=0}^{\overline{\frac{H_s+1-P}{P}}} (B^{Pj+i}) (W_{P-1-i}) \Sigma (W_{P-1-i})^T (B^{Pj+i})^T. \qquad (8)$$

Where Σ is the covariance matrix of Δf_t. After some manipulation we find

$$\boxed{\mathrm{Var}(F_t(t)|\pi(t) = 0) = \mathrm{tr}(W\Sigma W^T) \qquad \text{where} \qquad W = \sum_{i=0}^{P-1} W_i} \qquad (9)$$

Variance of Final Goods Inventory

To determine the variance for inventory, it is necessary to examine any arbitrary period, not just production-periods. Starting with the standard DRP inventory transition equation, accounting for the new weight matrices, we have

$$I_t = T[W_{\pi(t)} - I]\Delta f_t + BI_{t-1} + ssU_{H_s+1}. \qquad (10)$$

Moving back one period in time,

$$I_{t-1} = T[W_{(\pi(t)+1) \bmod P} - I]\Delta f_{t-1} + BI_{t-2} + ssU_{H_s+1}. $$

After back substitution, as in DRP, we get

$$I_t = \sum_{i=0}^{H_s} B^i T[W_{(\pi(t)+i) \bmod P} - I]\Delta f_{t-i} + \underline{ss} \qquad (11)$$

---

3. Note the absence of μU_{H+1}, because period t+H is not a production period if $\pi(t) = 0$.

This is similar to the DRP result, except for the cycling weight matrices.

Now, to compute steady-state results, we need to following observation: Every P periods, the same weight matrix is used, and since Σ is constant, those periods which use the same weight matrix are stochastically identical. Thus, we can partition the set of periods into P sets, where period t is in set i if at period t the next production period is t+i; i.e., if $\pi(t) = i$. So, to compute variance of FGI over all periods (in steady-state), we can compute variance of FGI in steady-state for each of these P sets and then combine them into the variance over all periods. I.e., let

$$V_k = Var(I_t(t)|\pi(t) = k). \tag{12}$$

This defines the variance of FGI inventory in those periods with $\pi(t) = k$. Now, let

$$V = Var(I_t(t)) = \sum_{i=0}^{P-1} Var(I_t(t)|\pi(t) = k)Prob(\pi(t) = k) \tag{13}$$

where the second equation is just the total probability rule. Now, since the $\pi(t)$ cycle predictably from P-1 to 0, we have $Prob(\pi(t) = k) = 1/P$, i.e., $\pi(t)$ is uniformly distributed in steady state. So, from (13),

$$Var(I_t(t)) = \frac{1}{P}\sum_{i=0}^{P-1} V_k. \tag{14}$$

Now it remains to calculate the V_k. We may use (11) for the case where $\pi(t) = k$. Thus,

$$V_k = Var(I_t(t)|\pi(t) = k) = \sum_{i=0}^{H_s} B^i T [W_{(k+i) \bmod P} - I] \Sigma [W_{(k+i) \bmod P} - I]^T T^T (B^i)^T$$

and so from (14),

$$V = \frac{1}{P} \sum_{k=0}^{P-1} \sum_{i=0}^{H_s} B^i T [W_{(k+i) \bmod P} - I] \Sigma [W_{(k+i) \bmod P} - I]^T T^T (B^i)^T$$

If we reorder the summation and pull out the B and T matrices,

$$V = \frac{1}{P} \sum_{i=0}^{H_s} B^i T \left(\sum_{k=0}^{P-1} [W_{(k+i) \bmod P} - I] \Sigma [W_{(k+i) \bmod P} - I]^T \right) T^T (B^i)^T$$

Note that for any given i, the inner sum will cycle through all P weight matrices. Thus following a similar analysis taken with Equation (DRP 19) in the DRP paper, let

$$A^{(m)} = \{a_{ij}^{(m)}\} = [W_m - I] \Sigma [W_m - I]^T \text{ for } m = 0, ..., P-1 \tag{15}$$

17

and so

$$V = \frac{1}{P} \sum_{i=0}^{H_s} B^i T \left(\sum_{m=0}^{P-1} A^{(m)} \right) T^T (B^i)^T = \frac{1}{P} \sum_{i=0}^{H_s} \sum_{j=0}^{i} \sum_{k=0}^{i} \sum_{m=0}^{P-1} a_{ij}^{(m)} \qquad (16)$$

This result may be used to calculate V quickly. Another turn in the development, namely going from equation (13) and (14) using the DRP single-weight matrix result would give us

$$V = \frac{1}{P} \sum_{m=0}^{P-1} \left(\sum_{i=0}^{H_s} \sum_{j=0}^{i} \sum_{k=0}^{i} a_{ij}^{(m)} \right), \qquad (17)$$

which, while just a reordering of summations from (16), shows that the steady-state variance in inventory using cycling weight matrices is just the average of the P variances of inventory if we were to use only one matrix exclusively each time.

Advantages and Disadvantages

The previous section showed the development of the second method of handling the FTVQ policy; as mentioned before, another method to deal with FTVQ is to approximate it by a single weight matrix and use the original DRP model. It is worthwhile to compare the two.

In terms of variance of production and inventory for a given period, the extension provides more precise results because of the multiple weight matrices. The extension also distinguishes between production and non-production periods. For example, the original DRP model will provide a variance of production for all periods; it is then necessary to juggle this result into something meaningful in relation to the production periods.

The main disadvantage with the extension in regard to the original DRP model is with the interface between the stage using FTVQ and its upstream supplier stage. The planned production output from, and thus the required demand for, a stage using FTVQ has very special characteristics, as we have examined above; namely, production is planned only for production periods. How this special information is transmitted upstream is an area for further research. There might be several ways of dealing with this. One would be to synchronize the stages, so that upstream stages only produce in the downstream's production stages; it would then be possible to redefine the time period for the upstream stage to have a length of P time units for the downstream stage. Another way of dealing with this problem is to keep the upstream stage ignorant of the actual policy used at the FTVQ stage, and pass the upstream stage a demand forecast which does not include information about when the production periods are. Then, an extra buffer could be placed between the two stages which would hold inventory designed to account for the timing differences between the two stages. This kind of buffer is explored in more detail in the lot-sizing analysis.

This extension provides an easy way to generate variance of production and inventory for the case where production occurs only in specified, regular interval periods. Such analysis does not lend itself to a multistage analysis. This is a common drawback in the literature, due to timing synchronization problems between stages. However, we still think this is a useful extension.

■ 5. Lot-Sizing and DRP - VTFQ Analysis

This section describes the method by which a lot-size model for a single stage can be viewed as a periodic-review $<Q,r>$ inventory system (like the $<nQ,r>$ system of Hadley and Whitin, 1963). This is one means of approaching the problem of incorporating lot-sizes into the DRP framework. Some others, which are based on modifying the planned production transition equations, are discussed in the appendix.

Modeling the production output of a single stage is accomplished by keeping track of the difference between the cumulative actual production and the cumulative planned production to a given time. In particular, let

$$n(t) = \text{number of batches produced in period } t$$

$$\pi(t) = n(t)L = \text{actual production in period } t$$

$$\Pi(t) = \sum_{i=0}^{t} \pi(i) = \text{cumulative actual production through period } t$$

$$\Psi(t) = \sum_{i=0}^{t} F_i(i) = \text{cumulative planned production through period } t$$

$$J(t) = \Pi(t) - \Psi(t)$$

(18)

where L is the lot size.

The decision to be made is how many lots to produce each period, namely, to choose n(t), given the planned production vector for the current horizon, F(t), and given all data for the past, from period 0 to period t. We make the following assumption:

Lot-size Policy Assumption: In each period, n(t) is chosen so as to keep the cumulative actual production through period t greater than the cumulative planned production through period t, but not greater than a single lot from the cumulative planned production. That is, n(t) is chosen so that $J(t) = \Pi(t) - \Psi(t) \in (0, L]$. The transition equation for J(t) is $J(t) = J(t-1) - F_t(t) + n(t)L$.

This assumption defines an achievable policy. If J(t) would be less than zero, produce enough lots to bring it nonpositive but less than L; if J(t) would be greater than L, "produce" negative lots to bring it below L but greater than zero. We can "produce" negative lots in the same way that the planned production can be negative in the original DRP model.

The claim is that this production policy (produce lots to keep J(t) in a specified range) can be

modeled as a periodic-review <Q,r> inventory system which can receive negative demands and therefore produce negative lots. The periods in the <Q,r> model are the same periods as in the original model. The demand in period t to the <Q,r> model is given by the planned production for period t, $F_t(t)$. The inventory level of the <Q,r> model is given by J(t). Finally, Q=L, the lot size, and r=0, namely, we reorder (produce) when J(t) is less than zero. Finally, because demand can be negative, we want to allow the production of negative numbers of lots. This, in some sense, adds a reorder point (for negative production) at inventory position L, but this doesn't affect the analysis of the <Q,r> system, since the spirit of the <Q,r> model is that the inventory position is maintained within the range (r,r+Q].

The problem with using existing <Q,r> literature is that the DRP planned production process is very specifically defined and is not necessarily independent over time. Thus, the analysis of the <Q,r> model must be specific to the DRP planned production process. In some instances (such as independent demand under a pull production policy), planned production is independent, but it is not independent if the demand process is dependent or if there is smoothing in the weight matrix. If we make the assumption that DRP demand forecast revisions are multivariate normal random vectors, then the planned production revisions will be multivariate normal random vectors. Thus, the demand process to the <Q,r> model is multivariate normal.

Results

The result we hope to get is that J(t) is uniformly distributed over (0,L] in steady-state. If this is true, then we can determine the probability of producing in a given period in steady state, since we know the variance of planned production. Namely, if at time t the lot position is J(t),

$$
\begin{aligned}
n(t) > 0 \qquad &\text{if} \qquad F_t(t) \geq J(t-1) \\
n(t) < 0 \qquad &\text{if} \qquad F_t(t) < J(t-1) - L
\end{aligned}
\tag{19}
$$

The probabilities for these events on n(t) are:

$$
\begin{aligned}
\text{Prob}\,[n(t) > 0] \;&=\; \frac{1}{L}\int_0^L \left(1-\Phi\!\left(\frac{y-\mu}{\sqrt{\mathrm{tr}(W\Sigma W^T)}}\right)\right)dy \\
\text{Prob}\,[n(t) < 0] \;&=\; \frac{1}{L}\int_0^L \Phi\!\left(\frac{y-L-\mu}{\sqrt{\mathrm{tr}(W\Sigma W^T)}}\right)dy
\end{aligned}
\tag{20}
$$

where $\Phi(z)$ is the standard normal distribution function. The integrand in the first equation represents the probability that planned production is greater than J(t), Prob $[F_t(t)] > J(t)$, and in the second equation represents the probability that planned production is less than J(t)-L, Prob $[F_t(t)] \leq J(t) - L$. The probability of production in steady-state is the sum of the two equations in (20).

We have not shown that the inventory position is uniform, but in some (statistically speaking, not enough) simulations, a histogram of inventory position over time appeared no less uniform with correlated demand than with uncorrelated demand. It clearly requires that the demand process is continuously valued (or at least dense over its range) and has nonzero variance (otherwise the position would be entirely predictable anyway). Intuitively, we can argue that the position should be uniform. In the <Q,r> context, the inventory position at any time is the starting inventory minus cumulative demand to that time, modulo Q and shifted up by R. It would be necessary then to show that the cumulative demand module Q is uniformly distributed on (0,Q]. This works for independent processes because of the Markovian nature of the inventory position. Why would it fail for dependent processes? Perhaps if some correlation in the process caused the position to be biased, skewed, bimodal, or any way nonuniform. A peak in a non-uniform distribution might be caused by the sequence of events of the position being at a peak and then a long, correlated series of demand which keeps the inventory position there (say, demand is around Q for a while), thus throwing the distribution out of balance. But we remember that the forecast revisions have expected value zero and are independent of the forecast themselves, and this hopefully implies that if there's a possibility the inventory position could get stuck at a particular value, then the possibility of such an event occurring is independent of the actual inventory position, and so it could happen at any inventory position with equal probability, and thus all imbalance would probabilistically be balanced, and the distribution would be uniform.

■ 6. Capacity Analysis Under Independence

This section presents an analysis of production capacity for the special case where demand to the stage is independent and the production policy yields independent planned production outputs. The analysis is based on an extension to the single-stage Tactical Planning Model (Graves, 1986, 1988).

To a large degree, the DRP is intended as a design tool to answer production planning questions. One such question is, given a demand forecast process and a production policy via a weight matrix, how much capacity does the stage need in order to produce under capacity with some probability? This question can be answered by computing the variance of production, and on the assumption that demand forecast revisions are normal, the variance of production can be used to look up a required capacity level. Thus, since the DRP is intended to answer questions about desired levels of capacity, there seems to be little need to include capacity itself in the analysis. However, it may be interesting to examine the impact of a specified capacity level in one stage on other stages within the system. Thus, it is desirable to include the potential for production capacities in a DRP single-stage analysis which preserves its extension to multistage analysis.

Single-stage capacity analysis

We have some results for the single-stage with capacity, but its potential for multistage analysis is limited. This is because even if there are such limiting restrictions on the stage that the planned production outputs $F_t(t)$ is independent over t, if a capacity is imposed on this planned production and if planned production output is ever greater than the capacity (which is when the problem becomes interesting), the actual production output will invariably not be independent. Without further approximations or limitations, this means that none of the stages upstream of the capacitated stream could be capacitated, since they all see a dependent demand process. Another reason is that it is unclear at this time exactly how to specify a covariance matrix for the actual capacitated production revisions. The response to this is just as in the lot-sizing model; it is possible to use the desired planned production outputs from the current stage as the demand process for the upstream stage, even though the actual production for the current stage is different. This would then result in some extra buffering between the upstream stage and the current stage.

To proceed with the analysis, we can essentially view a single stage of the DRP model as a black-box which gives us various outputs. These include the covariance matrix of planned production outputs, the planned production output vectors themselves, planned production starts, etc. We take one of these many outputs, namely the planned production output for the current period, $F_t(t)$. This represents the actual production output in the original, uncapacitated DRP model. The sequence $F_t(t)$ is a random, stationary sequence with mean μ and variance $\text{tr}(W\Sigma W^T)$. We have assumed that this sequence is independent over t; this is clearly a most restrictive assumption.

This sequence is then used as the arrival process to a single stage in the capacitated tactical planning model. Here's how the tactical planning model works: Arrivals to a server represent work to be completed at that server; these arrivals queue in a buffer prior to service. Each period, the server completes an amount of work proportional to the size of the queue, or the capacity, whichever is less. The transition equations, with the interface to the DRP model, are:

$$A_t = F_t(t)$$
$$Q_t = Q_{t-1} + A_t - P_{t-1}$$
$$P_t = \min(\alpha Q_t, \text{CAP})$$

(21)

where A_t is the number of arrivals in period t (assumed to occur at the start of a period), Q_t is the size of the queue at the start of time t, P_t is the actual production in time t (which occurs at the end of a period), CAP is the production capacity, and α, called the smoothing parameter, is the fraction of the queue's work that is completed in time t. We assume $\alpha = 1$, which means that each period the server tries to produce everything in the queue; it cannot when the queue size is larger than the capacity. This is consistent with how the DRP stage might behave in a capacitated

22

production environment. Namely, in each period the stage tries to produce the planned production output for that period. If it cannot, it queues the excess planned outputs into the next period.

The analysis for the DRP model in this instance reduces to analyzing a single-stage tactical planning model with capacity. This is done in Graves (1988) by approximating the discrete time model with a continuous-time model which yields, upon taking appropriate limits, an approximation of the queue size, Q_t, by a diffusion process. The density of the queue size is determined, and from that the density for production can be determined. The results are reproduced here for the specific case where $A_t = F_t(t)$ and $\alpha = 1$.

Approximation by continuous-time process. The continuous time transition equations are:

$$Q(t+\tau) = Q(t) + F(t, t+\tau) - P(t, t+\tau)$$
$$P(t, t+\tau) = \tau \min(Q(t), CAP) \tag{22}$$

where $F(t, t+\tau)$ is the demand from time t to $t+\tau$, and $P(t, t+\tau)$ is the production in that interval. Let $f()$ be the ergodic density function for $Q(t)$ which approximates Q_t. Then, for $x \leq CAP$,

$$f(x|x \leq CAP) = K \exp[(-(x-\mu)^2)/\sigma^2] \tag{23}$$

where "exp" denotes the exponential operator, and K is a normalizing constant. When $x > CAP$,

$$f(x|x > CAP) = K \exp[((CAP-\mu)(CAP+\mu-2x))/\sigma^2] \tag{24}$$

These equations are from Graves (1988). There is an issue as to the specification of σ^2, which represents the variance of the accumulated demand over one time period. Graves (1988) reports that in the uncapacitated case (as $CAP \to \infty$), the steady-state variance of $Q(t)$ is given by $\sigma^2/2$, which is half the variance accumulated over one period. This happens because of the nature of the approximation process, which takes the limit of the queue size as the size of the time unit goes to zero, thereby causing a smoother transition from period to period. The result is that the computation of the variance of production as shown below in (25) does not match simulation results if $\sigma^2 = \text{tr}(W\Sigma W^T)$, but matches simulation very well if $\sigma^2 = 2\text{tr}(W\Sigma W^T)$, to account for the continuous approximation mentioned above.

From (23) and (24), K can be numerically evaluated. Computation of $f(x)$ allows us to compute the moments of actual production, denoted by $P(t)$, as follows:

$$E[P(t)] = \mu$$
$$E[P(t)^2] = \int_0^{CAP} x^2 f(x)dx + CAP^2 \int_{CAP}^{\infty} f(x)dx \tag{25}$$
$$Var[P(t)] = E[P(t)^2] - (E[P(t)])^2$$

23

I have developed some simple C code to numerically evaluate this function, as well as a simulation with which to compare it. For 100,000 time periods the simulation required 47 seconds and gave a relative error of 0.029 for the uncapacitated production (compared to the analytical DRP results). For evaluation at 6000 points, the integral code required less than one second and gave a relative error of 0.049 for the capacitated production (compared to simulation results of 1,000,000 time periods).

Autocorrelation in the DRP/MMFE Demand Process

It is worthwhile to be able to compare in some manner the characterization of demand by the MMFE model with standard time-series processes. Heath and Jackson (1994) provide a qualitative comparison between the two. Here, we note that it is very easy to compute the autocorrelation function of the MMFE process.

The autocorrelation function for a stationary stochastic process $\{X_n\}$ will be denoted $\rho_X(\tau)$; it is the correlation coefficient of X_t and $X_{t+\tau}$ for all t (because of stationarity). The argument τ is called the lag, and the value of $\rho_X(\tau)$ is the lag-τ autocorrelation. Thus,

$$\rho_X(\tau) = \frac{E[X_0 X_\tau] - \mu_X^2}{\sigma_X^2} \qquad \tau = 0, 1, \ldots$$

where $E[X_i] = \mu_x$ and $Var[X_i] = \sigma_X^2$ for all i. This function and autocorrelation in general, as well as the impact of autocorrelation in the arrival process and service times on performance measures in a single-server queueing system are described in Livny, Melamed, and Tsiolis (1993).

For the MMFE process, where the demand in period t is $f_t(t)$, we have

$$\rho_{f_t}(\tau) = \frac{E[f_0(0) f_\tau(\tau)] - \mu^2}{\sigma^2} \qquad \tau = 0, 1, \ldots \tag{26}$$

where μ is the long-run average demand and σ^2 is the variance of demand, given by $\sigma^2 = tr(\Sigma) = \sum_{i=0}^{H} \Sigma_{i,i}$. To compute $E[f_0(0) f_\tau(\tau)]$:

First, the demands are given by:

$$f_0(0) = \mu + \sum_{i=0}^{H} \Delta f_{-i}(i)$$

$$f_\tau(\tau) = \mu + \sum_{j=0}^{H} \Delta f_{\tau-j}(j)$$

and so,

24

$$f_0(0)f_\tau(\tau) = \mu^2 + \mu\sum_{i=0}^{H}\Delta f_{-i}(i) + \mu\sum_{j=0}^{H}\Delta f_{\tau-j}(j) + \sum_{i=0}^{H}\sum_{j=0}^{H}\Delta f_{-i}(i)\Delta f_{\tau-j}(j)$$

Taking expected values, and since $E[\Delta f_t(t+i)] = 0$ for all t and i, we get:

$$E[f_0(0)f_\tau(\tau)] = \mu^2 + \sum_{i=0}^{H}\sum_{j=0}^{H}E[\Delta f_{-i}(i)\Delta f_{\tau-j}(j)] \qquad (27)$$

Now, within the summation in (27), $\Delta f_{-i}(i)$ and $\Delta f_{\tau-j}(j)$ may have non-zero correlation only if $-i = \tau - j$, because forecast revisions made in different periods are independent. Looking at the covariance matrix of forecast revisions, we see that

$$\Sigma_{ij} = E[(\Delta f_t(t+i) - E[\Delta f_t(t+i)])(\Delta f_t(t+j) - E[\Delta f_t(t+j)])]$$

but since the expectation of a forecast revision is zero,

$$\Sigma_{ij} = E[\Delta f_t(t+i)\Delta f_t(t+j)] \qquad (28)$$

Furthermore, if $-i = \tau - j$ then we have

$$E[\Delta f_{-i}(i)\Delta f_{\tau-j}(j)] = E[\Delta f_{-i}(i)\Delta f_{-i}(\tau+i)]$$
$$= \begin{cases} \Sigma_{i,\tau+i} & \text{if } \tau+i \leq H \\ 0 & \text{if } \tau+i > H \end{cases} \qquad (29)$$

and so, from (27) and (29), the end result is

$$E[f_0(0)f_\tau(\tau)] = \mu^2 + \sum_{i=0}^{H-\tau}\Sigma_{i,\tau+i} \qquad (30)$$

and

$$\boxed{\rho_{f_t}(\tau) = \left(\sum_{i=0}^{H-\tau}\Sigma_{i,\tau+i}\right) \Big/ \left(\sum_{i=0}^{H}\Sigma_{i,i}\right)} \qquad (31)$$

It is therefore easy to compute the lag-τ autocorrelation; it is simply the sum of the elements on the diagonal that starts in the first row and the τth column and ends in the τth row, divided by the sum of the elements on the main diagonal.

Example. For the "circular" covariance matrix, where $\Sigma_{ij} = p^{|i-j|}C$ for some C and p with $0 \leq p < 1$, it is easily seen that $\rho_{f_t}(\tau) = p^\tau\left(\dfrac{H-\tau+1}{H+1}\right)$.

# ■ Appendix

This section summarizes some other efforts and options explored during the research.

Lot-Sizing and the Transition Equations

The body of the paper describes a method for handling lot-sizing that uses a periodic review <nQ,R> inventory submodel to capture the effects of lot-sizing. Another approach is to modify the DRP transition equations themselves to account for the lot-sizing effects. Some means of doing this are laid out here, although no analysis is completed.

In Graves, Kletter, and Hetzel (1994), the transition equations for planned production and production starts define the flow of information in the DRP model. This is shown in Figure A. This figure also shows the planned production start revisions and planned production starts, indicated by ΔG and G, respectively. These have been largely ignored in this paper, but are an important part of the information flow process.
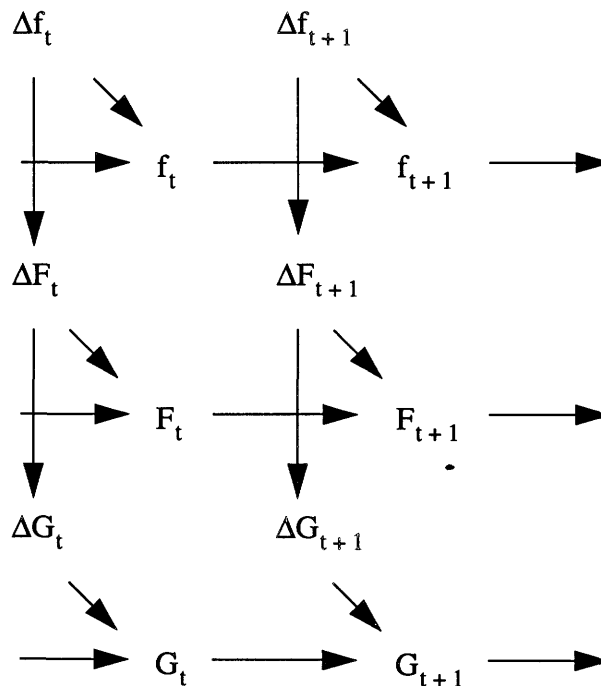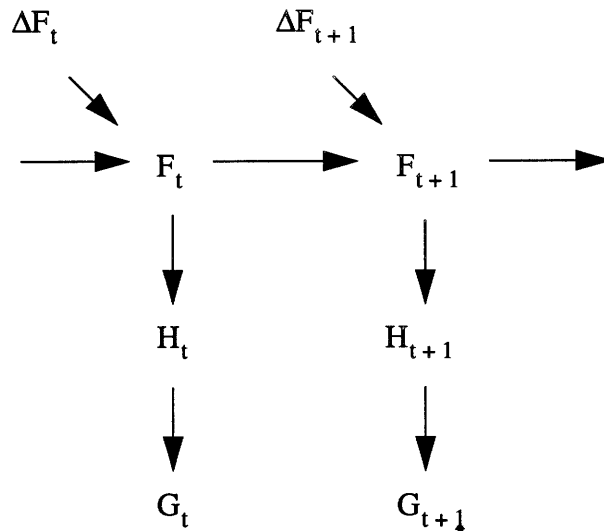

Figure A. Information flow in original DRP model

We have identified at least three ways to modify this information flow to incorporate the effect of lot-sizing on production. Each of thse methods involves adding a step in the above diagram, inbetween the planned production outputs and planned production starts. We call this step the "actual production outputs," and it represents the planned production output taking into account lot-sizing and capacity effects, and any other effects which cannot be modeled by the weight matrix explic-

itly. Each method introduces a different way of thinking about the transition equations, and the flow of dependent variables in the process.

Direct, no-feedback method. This method, as shown in Figure B., shows that actual production outputs are generated directly from the planned production output vector in some manner (such as rounding or aggregating). The actual production outputs vector is given in the diagrams by H_t.

$$\Delta F_t \qquad \qquad \Delta F_{t+1}$$

$$\searrow \qquad \qquad \searrow$$

$$\longrightarrow F_t \longrightarrow F_{t+1} \longrightarrow$$

$$\downarrow \qquad \qquad \downarrow$$

$$H_t \qquad \qquad H_{t+1}$$

$$\downarrow \qquad \qquad \downarrow$$

$$G_t \qquad \qquad G_{t+1}$$

**Figure B. Information flow in direct, no-feedback method**

There is no feedback of information from actual outputs to planned outputs. This has important implications for inventory control if actual outputs will differ from planned outputs, because the system cannot account for inventory balance equations. This method would probably be the most amenable to analysis.

Direct, feedback method. This method improves on the previous by introducing feedback of production information so that the system can account for inventory and production imbalances between the desired production and the actual production outputs. The flow is shown in Figure C. This method requires that the transition equations for planned production outputs in the DRP model be changed, and thus the analysis is more complicated.

Indirect method. This method is similar to the DRP information flows in that we add an actual production outputs revision vector from which the actual production outputs are derived. This is the basis of the inventory submodel used in the body of the paper. The flow is shown in Figure D.
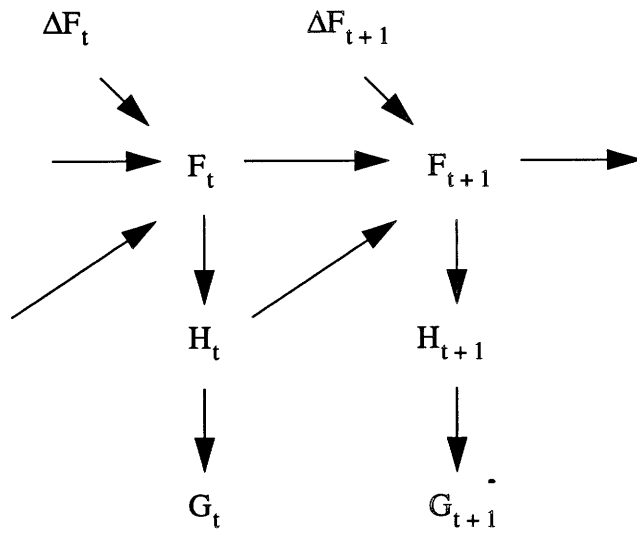
Figure C. Information flow in direct, feedback method
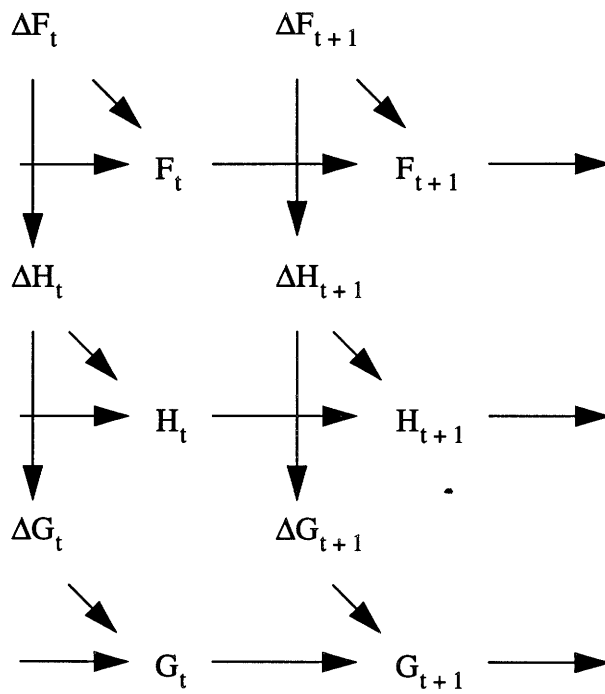
Figure D. Information flow in indirect method

Capacity With Dependent Production

Our prior attempt at analyzing capacity required that we turn the planned production outputs vector into just the planned output for the current period in every period. Thus, H periods of forecast information is lost every period. (I.e., instead of using the entire vector F_t we had to use only the first element, $F_t(t)$.) This was required to approximate the production requirements as a diffusion process. We were interested in determining if the production requirements could be modeled as a multidimensional diffusion process, which would then capture the planned production vector and its various correlations, which had been lost in the previous method.

The defining characteristic of the diffusion process would be the infinitesmal mean and variance. In Graves (1988) these are calculated for the extension to the Tactical Planning Model with capacity for two cases: (1) when desired production is less than or equal to capacity, (2) when desired production is greater than capacity. These cases correspond to the effects of the equation $P_t = \min(\alpha Q_t, CAP)$, where production equals either the desired production or the capacity of the system. We would need to develop similar sets of equations.

We examine only the first case $P_t = \alpha Q_t$, where production is uncapacitated. To analysis this, we start with the transition equations. We define $F_i(t)$ as the plan at time t for production in time t+i, $\Delta f_i(t, t + \tau)$ as the change in forecast demand from time t to time $t + \tau$, and $P_i(t, t + \tau)$ as the amount of production that occurs between times t and $t + \tau$. Following Graves (1988), the transition equations are:

$$F_i(t + \tau) = F_t(t) + P_{i+1}(t, t + \tau) + \sum_k w_{ik} \Delta f_k(t, t + \tau) - P_i(t, t + \tau) \quad i = 0, ..., H - 1$$

$$F_H(t + \tau) = F_H(t) + \mu\tau + \sum_k w_{Hk} \Delta f_k(t, t + \tau) - P_H(t, t + \tau) \tag{A1}$$

$$P_i(t, t + \tau) = \tau F_i(t) \qquad i = 0, ..., H$$

Substituting to remove the production values and simplifying:

$$F_i(t + \tau) = (1 - \tau) F_t(t) + \tau F_{i+1}(t) + \sum_k w_{ik} \Delta f_k(t, t + \tau) \qquad i = 0, ..., H - 1$$

$$F_H(t + \tau) = (1 - \tau) F_H(t) + \mu\tau + \sum_k w_{Hk} \Delta f_k(t, t + \tau) \tag{A2}$$

We may then take $\tau \to 0$ to get local growth rates for mean and variance.

Mean local growth rate. We want

29

$$\mu_i(x) = \frac{d}{dt}E[F_i(t)|F(t) = x] \qquad \forall i = 0...H$$

$$= \lim_{\tau \to 0}\frac{1}{\tau}(E[F_i(t+\tau)|F(t) = x] - E[F_i(t)|F(t) = x]) \tag{A3}$$

The left part of the numerator is derived from (A2):

$$E[F_i(t+\tau)|F(t) = x] = (1-\tau)x_i + \tau x_{i+1} \qquad i = 0, ..., H-1$$

$$E[F_H(t+\tau)|F(t) = x] = (1-\tau)x_H + \tau\mu$$

because $E[\Delta f(t, t+\tau)] = 0$. The right hand side of numerator is:

$$E[F_i(t)|F(t) = x] = x_i \qquad i = 0, ..., H$$

and so local mean growth rate is, for i=0,...,H-1:

$$\frac{1}{\tau}(E[F_i(t+\tau)|F(t) = x] - E[F_i(t)|F(t) = x]) = \frac{1}{\tau}((1-\tau)x_i + \tau x_{i+1} - x_i)$$

$$= \frac{1}{\tau}(\tau x_{i+1} - \tau x_i) \tag{A4}$$

$$= x_{i+1} - x_i$$

and for i=H:

$$\frac{1}{\tau}(E[F_H(t+\tau)|F(t) = x] - E[F_H(t)|F(t) = x]) = \frac{1}{\tau}((1-\tau)x_H + \tau\mu - x_H)$$

$$= \frac{1}{\tau}(\tau\mu - \tau x_H) \tag{A5}$$

$$= \mu - x_H$$

and as $\tau \to 0$, we get:

$$\mu_i(x) = x_{i+1} - x_i \qquad i = 1, ..., H-1$$

$$\mu_H(x) = \mu - x_H \tag{A6}$$

Variance local growth rate. Assumptions about forecast revision process were:

$$E[\Delta f(t)] = 0$$

$$Var[\Delta f(t)] = \Sigma t \tag{A7}$$

So, the variance of the local growth rate is given by:

$$\frac{d}{dt}Cov[F(t)|F(t) = x]$$

$$= \lim_{\tau \to 0}\frac{1}{\tau}(Cov[F(t+\tau)|F(t) = x] - Cov[F(t)|F(t) = x]) \tag{A8}$$

Here is the left hand term, broken down element by element:

$$\text{Cov}\left[F_i(t+\tau), F_j(t+\tau) \mid F(t) = x\right]$$
$$= E\left[F_i(t+\tau)F_j(t+\tau) \mid F(t) = x\right] - E\left[F_i(t+\tau) \mid F(t) = x\right]E\left[F_j(t+\tau) \mid F(t) = x\right]$$

$$= E\left[\left((1-\tau)x_i + \tau x_{i+1} + \sum_k w_{ik}\Delta f_k(t, t+\tau)\right)\left((1-\tau)x_j + \tau x_{j+1} + \sum_k w_{jk}\Delta f_k(t, t+\tau)\right)\right]$$
$$- E\left[(1-\tau)x_i + \tau x_{i+1} + \sum_k w_{ik}\Delta f_k(t, t+\tau)\right]E\left[(1-\tau)x_j + \tau x_{j+1} + \sum_k w_{jk}\Delta f_k(t, t+\tau)\right]$$

If we multiply out the left expectation and take the expectation inside on the right, we have:

$$= E\left[\left((1-\tau)x_i + \tau x_{i+1}\right)\left((1-\tau)x_j + \tau x_{j+1}\right) + \left((1-\tau)x_i + \tau x_{i+1}\right)\sum_k w_{jk}\Delta f_k(t, t+\tau)\right.$$
$$\left. + \left((1-\tau)x_j + \tau x_{j+1}\right)\sum_k w_{ik}\Delta f_k(t, t+\tau) + \sum_k w_{ik}\Delta f_k(t, t+\tau)\sum_k w_{jk}\Delta f_k(t, t+\tau)\right]$$
$$- \left((1-\tau)x_i + \tau x_{i+1}\right)\left((1-\tau)x_j + \tau x_{j+1}\right)$$

The forecast revisions in the right-hand term's factors drop out because $E[\Delta f(t)] = 0$. Now if we take the remaining expectation into the sums, we have:

$$= \left((1-\tau)x_i + \tau x_{i+1}\right)\left((1-\tau)x_j + \tau x_{j+1}\right) + E\left[\sum_k w_{ik}\Delta f_k(t, t+\tau)\sum_k w_{jk}\Delta f_k(t, t+\tau)\right]$$
$$- \left((1-\tau)x_i + \tau x_{i+1}\right)\left((1-\tau)x_j + \tau x_{j+1}\right)$$

the first and third terms cancel, and the middle term is the (i,j)th element of the matrix $\tau W\Sigma W^T$.

Now, for the right hand term of the Covariance equation, by-element:

$$\text{Cov}\left[F_i(t)F_j(t) \mid F(t) = x\right] = E\left[F_i(t)F_j(t) \mid F(t) = x\right] - E\left[F_i(t) \mid F(t) = x\right]E\left[F_j(t) \mid F(t) = x\right]$$
$$= x_i x_j - x_i x_j = 0$$

and so the local growth rate is:

$$\frac{d}{dt}\text{Cov}(F_i(t), F_j(t)) = \frac{1}{\tau}\left(\tau\{W\Sigma W^T\}_{ij} - 0\right)$$
$$= \{W\Sigma W^T\}_{ij} \tag{A9}$$

or, in matrix form,

$$\frac{d}{dt}\text{Cov}(F(t)) = W\Sigma W^T \tag{A10}$$

Equations for the capacitated form need to be derived, and from them a diffusion characterized.

Bibliography

Graves, S.C., 1986, "A Tactical Planning Model for a Job Shop," *Operations Research,* 34:4, 522–533.

Graves, S.C., 1988, "Extensions to a Tactical Planning Model for a Job Shop," Sloan School Working Paper #20-9688, Massachusetts Institute of Technology.

Graves, S.C., D.B. Kletter, and W.B. Hetzel, 1994, "A Dynamic Model for Requirements Planning with Application to Supply Chain Optimization," Sloan School Working Paper #3669-94, Massachusetts Institute of Technology.

Hausman, W.H., 1969, "Sequential Decision Problems: A Model to Exploit Existing Forecasters," *Management Science,* 16:2, B-93–B-111.

Heath, D.C., and P.L. Jackson, 1994, "Modeling the Evolution of Demand Forecasts With Application to Safety Stock Analysis in Production/Distribution Systems," *IIE Transactions,* 26:3, 17–30.

Jackson, P.L., D.C. Heath, and J.A. Muckstadt, 1993, "Modeling the Impact of Assembly Demand Forecast Accuracy on Customer Service, Inventory, and Expediting," *1993 NSF Conference Proceedings,* 1075–1089.

Whitt, W., 1983, "The Queueing Network Analyzer," *The Bell Systems Technical Journal,* 62:9, 2779–2815.