Appendix

We develop a model for a single stage, which represents a work center, e.g., a machining center or unit process. We term this as the tactical planning model (TPM). The intent is to develop a simple model for understanding the impact of the planned lead time and getting some guidance on how it depends on the characteristics of the demand process and the level of capacity.

The single stage transforms some input into an output. We assume that the stage operates as follows:

- At the start of each time period, the work to be processed by the stage arrives to the stage and joins a queue. We denote the arrival in time period *t* by $A_t$, where the units are in terms of the workload on the stage. For instance, the units of the arrivals could be hours of processing time.
- We denote the production in period *t* by $P_t$, where the units are in terms of the workload, e.g., hours of processing time.
- We assume a linear control rule: $P_t = Q_t / n$ where $P_t$ is the amount of work processed in time period t, $Q_t$ is the queue at start of period t, and *n* is the planned lead time, $n \geq 1$.

In Figure A1 we depict the single-stage system and notation for the model: each period the system demand $A_t$ enters a queue or work-in-process, denoted by $Q_t$; each period the stage processes some portion of the work-in-process, with the output denoted by $P_t$.

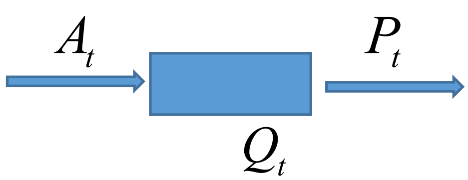


Figure A1: Single-Stage System

## Interpretation:

The amount of work processed at each time period is a fixed portion of the queue in front of that work station. When the queue grows, the work center works harder; when the queue is short, the work rate is lower. The underlying assumption is that the system has the flexibility to adjust its production rate, e.g. by shifting workers, working overtime, outsourcing, etc.

We can show using Little's Law that the average time for a unit of work to complete processing is *n* time periods. Furthermore we find experimentally that the variability in the completion time is very small, and a very high percentage of the work is completed in $\lfloor n \rfloor or \lceil n \rceil$ time periods[1]. Thus we interpret *n* as

[1] *n* need not be an integer, but the work completion time is measured as an integer number of time periods. So the statement is in terms of *n* being rounded down or up to the nearest integer, which is denoted with the notation $\lfloor \ \rfloor or \lceil \ \rceil$.

the ***Planned Lead Time*** (PLT) at the work station. That is, we plan on work to require *n* time periods to complete processing at the stage.

## Model Analysis:

For the analysis of this model we need state the linear control rule (1) and an inventory balance equation (2) for each time period *t*:

$$P_t = Q_t \big/ n \tag{1}$$

$$Q_t = Q_{t-1} + A_t - P_{t-1}. \tag{2}$$

We first divide the second equation (2) by *n*, and then substitute (1) into the balance equation to get a smoothing equation:

$$\begin{aligned} &\frac{Q_t}{n} = \frac{Q_{t-1}}{n} + \frac{A_t}{n} - \frac{P_{t-1}}{n} \\ &\Rightarrow P_t = (1-\alpha)P_{t-1} + \alpha A_t \\ &where\ \alpha = 1 \big/ n. \end{aligned} \tag{3}$$

Thus, we can express the linear control rule (1) as a simple smoothing equation, whereby the production in the current period is the weighted average of the arrivals in the period and the production level in the prior period. We see immediately the impact of the planned lead time *n*: a longer planned lead time results in a smaller smoothing parameter α; and a smaller α means that there is more smoothing of the demand series, and hence, the resulting production series is smoother.

We can do repeated substitution into (3) to express the current production in terms of all prior demand:

$$\begin{aligned} P_t &= \alpha A_t + (1-\alpha)\alpha A_{t-1} + (1-\alpha)^2 \alpha A_{t-2} + \ldots \\ &= \sum_{j=0}^{\infty} (1-\alpha)^j \alpha A_{t-j} \end{aligned} \tag{4}$$

For (4) we assume that we have an infinite history of arrivals. If we were to assume (more realistically) that we have a finite arrival history, say, back to time period 0, then the above expressions would be:

$$P_t = \sum_{j=0}^{t-1} (1-\alpha)^j \alpha A_{t-j} + (1-\alpha)^t A_0 \tag{5}$$

where we fix $P_0 = A_0$. There is very little difference between (4) and (5) for $t > 10$ and typical values of α (α < 0.5). It is easier to work with equation (4), and thus, we will continue to make the (unrealistic) assumption of an infinite history so as to simplify the presentation.

We assume that the demand arrival $A_t$ in every period *t* is an independent and identically distributed random variable (i.i.d.) with mean $E\left[A_t\right]=\mu$ and variance $Var\left[A_t\right]=\sigma^2$. From (4) we see that $P_t$ is also a random variable and we can use (4) to obtain its moments:

$$E\left[P_t\right]=\sum_{j=0}^{\infty}(1-\alpha)^j \alpha E\left[A_{t-j}\right]=\sum_{j=0}^{\infty}(1-\alpha)^j \alpha\mu=\mu \tag{6}$$

where we use the fact that the geometric weights sum to one. We find the variance from:

$$\begin{aligned}Var\left[P_t\right]&=\sum_{j=0}^{\infty}(1-\alpha)^{2j}\alpha^2\times Var\left[A_{t-j}\right]\\&=\sum_{j=0}^{\infty}(1-\alpha)^{2j}\alpha^2\sigma^2=\frac{\alpha^2\sigma^2}{1-(1-\alpha)^2}=\frac{\alpha\sigma^2}{2-\alpha}\\&=\frac{\sigma^2}{2n-1}.\end{aligned} \tag{7}$$

Thus we have the standard deviation for production being:

$$SD\left[P_t\right]=\sigma\sqrt{\frac{1}{2n-1}}. \tag{8}$$

From the linear rule (1) and from (6) and (7) , we find that:

$$E\left[Q_t\right]={E\left[P_t\right]}\big/{\alpha}=n\mu. \tag{9}$$

From (8) and (9) we can see the basic behavior of the model. As we increase the planned lead time *n*, we reduce the standard deviation of the production random variable; that is, we smooth the production. But a longer planned lead time *n* results in a longer queue of work at the stage, namely a larger work-in-process.

## Setting the planned lead time

To set the planned lead time, on the one hand, we want to make it as short as possible so as to minimize the amount of WIP. On the other hand, we want it to be long so as to smooth the production requirements.

To determine the level of smoothing, we assume the stage has a maximum reasonable capacity given by $\mu+\chi$; this represents the amount of output that the stage can produce per time period, under normal circumstances. We will set the planned lead time so that the production given by the linear control policy is within the maximum reasonable capacity most of the time; to make this operational, we assume a service level β, where β is the percent of time that production is within the maximum reasonable capacity. Thus we want to set the planned lead time so that:

$$\Pr\left[P_t \le \mu + \chi\right] = \beta. \tag{10}$$

Now we make an additional assumption: we assume that the system demand $A_t$ is normally distributed, with mean and standard deviation given by $\mu, \sigma$. Then $P_t$ is also normally distributed with mean and standard deviation given by (6), (8). We can now transform $P_t$ into a standard normal random variable and re-write the left-hand-side of (10) as:

$$\Pr\left[P_t \le \mu + \chi\right] = \Pr\left[\frac{P_t - \mu}{\sigma / \sqrt{2n-1}} \le \frac{\chi}{\sigma / \sqrt{2n-1}}\right] = \Phi\left(\frac{\chi}{\sigma / \sqrt{2n-1}}\right). \tag{11}$$

Where $\Phi(\ )$ is the cumulative distribution function for the standard normal variable. Now we need to set the expression in (11) to the service level β:

$$\begin{aligned}
&\Phi\left(\frac{\chi}{\sigma / \sqrt{2n-1}}\right) = \beta \\
&\Rightarrow \frac{\chi}{\sigma / \sqrt{2n-1}} = \Phi^{-1}(\beta) = z \\
&\Rightarrow n = \frac{(z\sigma)^2 + \chi^2}{2\chi^2}.
\end{aligned} \tag{12}$$

where we have introduced the service factor $z = \Phi^{-1}(\beta)$; for instance if the service level $\beta = 0.95$, then z = 1.64.

The above equation (12) relates the planned lead time to the variability of the demand process, the service level, and the headroom.

We note from (12) that it suggests n < 1 if $z\sigma < \chi$; this is inconsistent with the model setup as we assume $n \ge 1$. Hence, we understand equation (12) to apply as long as $z\sigma \ge \chi$; when $z\sigma < \chi$, then the planned lead time is one time period.

For additional reading we cite the following papers from our research that develop the *Tactical Planning Model* for multi-stage systems and develop various extensions and applications.


Graves, Stephen C. "A tactical planning model for a job shop." *Operations Research* 34.4 (1986): 522-533.

Graves, Stephen C. "Safety stocks in manufacturing systems." *Journal of Manufacturing and Operations Management,* 1988, Vol. 1, No. 1, pp. 67-101. (Sloan WP, 1987).

Teo, Chee-Chong, Rohit Bhatnagar, and Stephen C. Graves. "Setting planned lead times for a make-to-order production system with master schedule smoothing." *IIE Transactions* 43.6 (2011): 399-414.

Teo, Chee-Chong, Rohit Bhatnagar, and Stephen C. Graves. "An application of master schedule smoothing and planned lead time control." *Production and Operations Management* 21.2 (2012): 211-223.

Chhaochhria, Pallav, and Stephen C. Graves. "A forecast-driven tactical planning model for a serial manufacturing system." *International Journal of Production Research* 51.23-24 (2013): 6860-6879.

Yuan, Rong, and Stephen C. Graves. "Setting optimal production lot sizes and planned lead times in a job shop." *International Journal of Production Research* 54.20 (2016): 6105-6120.