

# Exploiting Virtual Elasticity of Production Systems for Respecting OTD-Part 1: Post-Optimality Conditions for Ergodic Order Arrivals in Fixed Capacity Regimes

Bruno G. Rüttimann<sup>1</sup>, Martin T. Stöckli<sup>2</sup>

<sup>1</sup>ETH Zürich IWF, Zurich, Switzerland

<sup>2</sup>Inspire AG, Zurich, Switzerland

Email: bruno.ruettimann@inspire.ethz.ch, stoeckli@inspire.ethz.ch

**How to cite this paper:** Rüttimann, B.G. and Stöckli, M.T. (2020) Exploiting Virtual Elasticity of Production Systems for Respecting OTD-Part 1: Post-Optimality Conditions for Ergodic Order Arrivals in Fixed Capacity Regimes. *American Journal of Operations Research*, 10, 321-342.

<https://doi.org/10.4236/ajor.2020.106018>

**Received:** September 3, 2020

**Accepted:** November 27, 2020

**Published:** November 30, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Respecting the on-time-delivery (OTD) for manufacturing orders is mandatory. This depends, however, on the probability distribution of incoming order rate. The case of non-equal distribution, such as aggregated arrivals, may compromise the observance of on-time supplies for some orders. The purpose of this paper is to evaluate the conditions of post-optimality for stochastic order rate governed production systems in order to observe OTD. Instead of a heuristic or a simulative exploration, a Cartesian-based approach is applied to developing the necessary and sufficient mathematical condition to solve the problem statement. The research result demonstrates that increasing speed of throughput reveals a latent capacity, which allows arrival orders above capacity limits to be backlog-buffered and rescheduled for OTD, exploiting the virtual manufacturing elasticity inherent to all production systems to increase OTD reliability of non JIT-based production systems.

## Keywords

On-Time-Delivery, Production System, Lean Manufacturing, Industry 4.0, Arrival Rate, Markovian Arrival Distribution, Production Backlog, Manufacturing Elasticity, Production Capacity, Bottle Neck, Break-Even Point, Optimal Production Volume, Ergodic Processes

---

## 1. Introduction

According to the production-specific *SPQR axiom*, which addresses the observance of business requirements of speed, punctuality, quality, return [1], the ob-

servance of on-time-deliveries (OTD) for customer orders is compulsory to have a viable and sustainable business model. Punctuality and quality are at the centre of customer satisfaction. Many operational excellence projects are targeted to increase OTD, hence the importance of this topic. Therefore, directing the research attention to the requirements and conditions to observe OTD generally for all types of production systems by applying a structured Cartesian view of the topic is essential [2].

Just in time (JIT) production systems are mainly taktrate ( $TR$ )-governed manufacturing systems with  $SD[TR] = 0$  where  $SD$  denoting the standard deviation and  $TR$  representing the pace of ordering. The necessary and sufficient conditions for a  $TR$ -governed manufacturing system to observe OTD is given by the Equation system 1 [1]

$$\begin{cases} \forall i : \inf \{ER_i\} \geq TR \\ Z : PLT_z \leq EDT \end{cases} \quad (1)$$

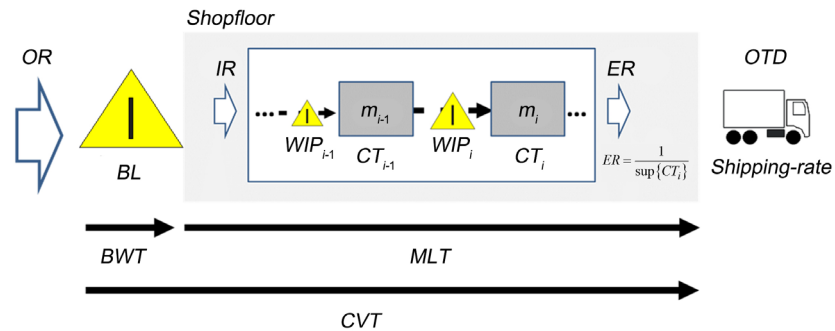
The first condition for OTD requires that the capacity is sufficient—capacity given by the exit rate ( $ER$ ) which is determined by the smallest  $ER_i$  of each process step. The ER has to be large enough to process the arriving orders expressed by the taktrate ( $TR$ ). The second condition to be observed requires that the process lead-time ( $PLT$ ) is shorter than the customer’s expected delivery time ( $EDT$ ) confirmed at the moment of the ordering. Having a  $TR$ -governed production system with a deterministic call-off characteristic of production orders within a defined timespan, we can assume an even spreading of incoming orders within this timespan. With spreading we intend here the time-dimension of distribution.

For non-deterministic order-based JIT production systems, or more generally, for batch& queue (B&Q) production systems based on a push manufacturing principle—which is a manufacturing system modus operandi usually applied in the case of erratic order entry—Equation (1) has to be modified. Indeed, the order arrival rate is stochastic and we should not talk any more about a  $TR$ , but about an order rate ( $OR$ ) with the characteristic of  $SD[OR] > 0$ . To observe  $OTD$  in this case, the more general Equation system 2 is valid

$$\begin{cases} \text{for : } SD[OR] > 0 \\ \forall i : \inf \{ER_i\} > E[OR] \\ Z : BWT_z + PLT_z \leq EDT \end{cases} \quad (2)$$

The variable  $OR$  will sometimes exceed  $ER$ , and therefore according to the *Theorem of WIP* build-up a work in progress (WIP) stock at the process entry point  $Z$  which we will name backlog ( $BL$ ). Usually the production order input rate ( $IR$ ) to the shopfloor is reduced to match the  $ER$ , *i.e.*  $IR = ER$ . The production system to be modelled is schematically shown in **Figure 1**.

The customer visible time ( $CVT$ ) will therefore comprise the  $PLT$  plus the WIP-generated backlog waiting time ( $BWT$ ) which has to be shorter than  $EDT$ ; for simplicity reason, we are not considering the shipping time, which is not an



**Figure 1.** Production system with variable  $OR$  forming a  $BL$  if  $OR$  exceeds  $ER$ .

issue of manufacturing. Equation (2) is directly derived from the following important theorem, the *Theorem of General Production Requirements or OTD Theorem*, which sets the necessary conditions for OTD [1]:

Theorem of General Production Requirements (or OTD Theorem):

The necessary and sufficient conditions to supply a customer with OTD, *i.e.* with the right quantity at the right time, is that first the capacity requirement and second the lead time requirement have to be satisfied simultaneously, independent of the applied transfer principle, *i.e.* SPF or B&Q. The capacity requirement is given by the Corollary of Weak WIP Stationarity and the lead time requirement necessitates that  $MLT$  plus  $BWT$  is shorter than  $EDT$ .

As mentioned in the OTD theorem, it would be more appropriate to talk about  $MLT$  (Manufacturing Lead Time) instead of  $PLT$ , which indicates the timespan one single piece needs to transit the process. The  $MLT$  comprises the lead time to produce the entire batch  $B(n)$  with the transfer principle of  $n$  pieces, *i.e.* when the first piece enters the process and the last piece of the batch leaves the process. In Western production systems, with its dominant batch-culture and B&Q production systems making the lion's part, the difference between  $MLT$  and  $PLT$  can be considerably big. The application of the  $MLT$  metric will be mandatory in multi-product manufacturing systems where order-batches of different products are produced and have to be delivered again on time as whole batches to the customer. In this introductory paper of conceptualization regarding post-optimality for OTD we will refrain from considering batches and will therefore view each order as a single product to be delivered.

These are the general, theorem-based requirements to observe OTD. However, such as Equation (1) for the  $TR$ , also Equation (2) for the  $OR$  applies the discrete Poissonian view of average occurrences per timespan represented by  $\lambda$ . If  $\lambda$  is an average, the same average can be the result of different distributions of the phenomenon to be modeled. The two extreme distributions are on the one hand the ideal situation of the even spreading over time of occurrences with constant inter-arrival time ( $TR$ -like) and on the other hand the bulk occurrence, *i.e.* the verification of concentrated arrivals. In this context we rather prefer the word spreading instead of distribution, indicating the focus on the random sequence

of arrival orders (or the sequence of changing patterns which would lead to a non-ergodic process) in the time-dimension  $t$  of the distribution  $\tau$  of the stochastic variable  $x(t)$ . However, we use the appropriate word distribution to characterize the sequence-independent view of arrival orders represented by the aggregated shape of the inter-arrival time  $\tau = x^{-1}(t)$  of the random variable. If the process is ergodic, it is not necessary to make this difference because the process is time-invariant. It seems to be clear that in the case of an ideal situation of even arrival sequence of orders, Equation (1) can be observed theoretically depending in reality only on the characteristics of underlying manufacturing systems to transform efficiently “order arrivals” into “product deliveries”. However, if the arrival spreading of orders is aggregated (bulk arrivals), the instantly needed capacity is not enough to process all the orders at the same instant forcing them to be queued and compromising eventually the observance of OTD. This is true even though the average condition of Equation (2) may be observed. Indeed, whereas usually the constant  $ER$  for a mono-product manufacturing system always presents an even deterministic spreading over time, this might not be the case for the  $OR$ . In the case of non-deterministic arrival rate the question now is, in which post-optimal range of a non-evenly time-distributed  $OR$ , Equation system 2 will be observed.

The treated topic is complex and until presently it has not yet been investigated although queuing theory application is wide-spread. In fact, the findings have far-reaching implications in production economics and manufacturing theory as well as optimization under uncertainty. Therefore, this paper does only have introductory character. The complex character of the topic requires a swing back to basic theory. This entails rather a didactic connotation of the presentation and might be seen as an unnecessary step in scientific papers. However, in this context for sure it increases understanding of the whole topic allowing to reaching a larger readership.

In the following, we will use for the solution finding a new production mathematics’ reasoning based on a Cartesian scientific approach [1] [2], which is a theory-cognitive foundation allowing to *understand the behavior* of a production system instead of *explore the behavior* via solution-oriented discrete dynamic simulation techniques. This Cartesian approach is a didactically sensible extension to today’s still heuristic-cognitive based production-theory teaching.

## 2. A Little Bit More Theory

Let us have a closer look at two important production topics, on the one hand the production capacity and on the other hand the manufacturing elasticity. We prefer to call it manufacturing elasticity and not production elasticity in order not to confuse it with the economic output elasticity. In this paper, we are explicitly not interested in the topic of production flexibility, *i.e.* the capability to transform different type of products resulting in a high mix variability. Not because this topic is not of importance, on contrary, however, because it adds a further complexity to the problem statement, which is a variable  $ER$  instead of a

deterministic  $ER$  with a mix-specific distribution. Indeed, the concurrent optimization of two target functions, such as elasticity and flexibility, is a problem of Pareto efficiency [2] [3], multivariate-type of problem optimization, which would only increase the complexity of this paper.

Production capacity defines the amount of products, or generally, the volume of produced items, which a certain production infrastructure can transform from raw materials or components into intermediate or finished goods. The expressions throughput or exit rate ( $ER$ ) define the concept of capacity figuratively. The concept of production capacity is directly linked to the concept of bottleneck, modelled in Equation (3)

$$\begin{cases} CT_b = \sup\{CT_i\} \\ ER = ER_b = \frac{1}{CT_b} \end{cases} \quad (3)$$

which is a mathematical representation directly derived from the following *Theorem of Throughput* [1]. This theorem unequivocally defines the concept of bottleneck

Theorem of Throughput (or Bottleneck Theorem):

Given is a sequence of production steps with each process step having a deterministic, but different cycle time  $CT$ . The maximum throughput, *i.e.* the maximum exit rate  $ER$ , of a process is given by the slowest process step, *i.e.* the process step with the longest cycle time  $CT$ ; this process step is called bottleneck.

The  $ER$  is a performance concept, defined as number of pieces per time unit and represents the productivity of the process constituting the nominal specific capacity. The absolute production volume  $V$  is given by the timespan  $t$  the  $ER$  is applied, *i.e.*  $V = ER \cdot t$ . The production capacity is therefore also a concept of the applied work shift regime.

Manufacturing elasticity, however, deals with the cost of a variable production regime subject to load changes for the quantity to be produced. Usually, production capacity is defined for an optimal volume range, production minimum given by the long-term minimal economic sustainable production level  $V_{BEP}$ , *i.e.* by the break-even point (BEP). In this paper we are not primarily interested in the BEP, but rather in the concept of insensitivity to cost change for various production level regimes. The BEP is a function of fixed cost  $K_{fix}$  of investment depreciation as well as fixed operating cost (Equation (4))

$$V_{BEP} := \frac{K_{fix}}{p - k_v(V)} \quad (4)$$

where  $p$  is the price and  $k_v(V)$  represents the variable unitary specific cost. Therefore  $m(V) = p - k_v(V)$  is the margin contribution per unit sold. Please note that the variable cost show in reality a variable unitary characteristic with cubical exponent in function of increasing production volume  $V$  due to inefficiencies at

low volume and increased expenses for additional shifts. Moreover, the production volume where the variable cost  $k_v$  are minimum is given by

$$V_{\min\{k_v\}} := \{V | \min\{k_v(V)\}\} \rightarrow \left\{ V_{\min\{k_v\}} \left| \frac{d}{dV} k_v(V) = 0 \right. \right\}$$

This leads to a non-linear model with an optimal finite operational production level  $V_{opt}$  above the BEP and not in the infinite like a linear model would suggest. The optimal production level is the associated production volume where the marginal price decrease equals the marginal variable cost increase

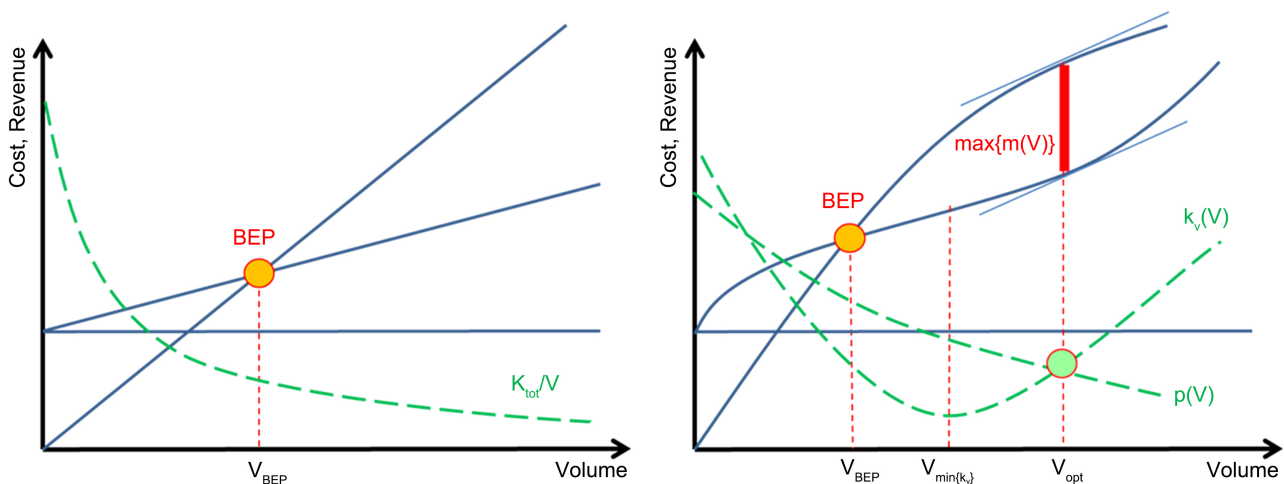
$$V_{opt} := \left\{ V \left| \frac{d}{dV} p(V) = \frac{d}{dV} k_v(V) \right. \right\}$$

or

$$V_{opt} := \left\{ V | \max\{(p(V) - k_v(V)) \cdot V\} \right\} \rightarrow \left\{ V_{opt} \left| \frac{d}{dV} (m(V) \cdot V) = 0 \right. \right\}$$

In addition, it is evident from the equation that the optimal production regime  $V_{opt}$  is not at the volume level where  $k_v(V)$  is minimal, but where the margin  $m(V)$  is maximal, *i.e.* marginal cost equal marginal price (Figure 2, right side). Please note that in reality also the price  $p$  might be a function of volume  $V$  resulting in  $p(V)$ . However, the average specific cost are based on the full costing model divided by the production volume (Figure 2, left side).

The “art” of maximizing manufacturing elasticity is therefore not reduced to the problem of minimizing investment in order to lower the BEP, but to have an extended range of volume with reduced  $k_v(V)$ . How industry 4.0-type production system, allowing batch size one (one-offs), will solve this contradiction, is left out in this paper. However, this is very important. Because to define correctly elasticity of production we have to distinguish two different elasticities: we can call it improperly the *average elasticity* based on a full costing model and the classical is the *marginal elasticity* based only on variable cost. Please note that



**Figure 2.** Didactic break-even model (left side) vs. realistic BEP model (right side) showing the basic differences adding also the average specific full cost (left side) as well as marginal unitary variable cost and price (qualitative representation).

the rarely precisely defined word *elasticity* in production theory of manufacturing, if ever defined, is intended of having low sensitivity of cost change to volume change  $\varepsilon = \min\{\partial k/\partial V\}$ . However, nota bene, this corresponds in economics to inelasticity. In production theory elasticity has a different meaning. Production systems with high elasticity allow the usage of the equipment over a large production volume range without significant cost changes. Elasticity is linked to the kind of invested equipment and work shift model. Nevertheless, automation generally comprises high investment increasing BEP and therefore lowering, *i.e.* worsening the *average elasticity*, presenting a high *average elasticity* only in the high volume range (Figure 2, left side); this confirms the fact that high investment are not apt for low volume production. However, high investment do not affect *marginal elasticity* depending only from variable cost or, if ever, due to lower specific variable cost, even increases marginal elasticity. It is very important to keep in mind this behaviour for future investment in automation. If an appropriate investment will allow a high marginal elasticity it will depend on the specific investment, which reduces the employment of blue-collar workforce, mainly responsible for the non-linear shape of the variable cost curve. Therefore, high investment in automation allows to approach the linear didactic break-even model. Solving the enigma of how to obtain a high elasticity corresponds therefore to the quest of solving “the squaring of the circle”. This leads directly to enounce the:

***Theorem of Optimal Production Range (or Limited Elasticity Theorem)***

*Given is a production system characterized by fixed cost  $K_{fix}$  and non-constant specific variable cost  $k_v(V)$  in function of production volume. The marginal elasticity is the insensitivity of variable cost in presence of changes in production volume. Due to the variability of the specific variable cost, the optimal manufacturing elasticity is confined within a limited range around minimal cost volume  $V_{\min\{k_v\}}$  and with economic meaningful extension towards the volume around the optimal production level  $V_{opt}$ .*

Therefore, this theorem reveals that the exploitation of optimal manufacturing elasticity is not only a function of cost, but also of revenue. It is directly linked to the *Postulate of Incompatibility (or Flexibility-Elasticity Contradiction Postulate)* enounced in [2] stating the impossibility to have at the same time a high flexible and a high elastic production system under realistic conditions (*i.e.* without near-infinite investment in production machinery which eliminates the necessity of blue-collar workforce). In any case, this would result in a low average elasticity and a high marginal elasticity. In the following, we will assume  $ER \approx E[OR]$  where *ER* stands at the base of the optimal operational level in order to comply with the just enounced theorem.

### 3. The Paradoxon of the Mean

We are used to apply statistical parameters, especially such as the mean value, everywhere we have to describe synthetically the characteristic of a physical or

social phenomenon. The mean is the average or better the expected value of a stochastic data set of a random variable  $X$ , represented with  $E[X]$ . Moreover, many think that the mean is a good representation of this random variable  $X$  and don't further care about the dispersion of the data around the mean described as the variance  $Var[X]$  or the standard deviation  $SD[X]$ . However, the "mean" does not exist physically, but is basically an artificial value—a "fiction" if you want—derived from a dataset of variable  $X$  with a mathematical description (the first statistical moment). In fact, a random variable is usually represented by its real distribution. Generally, statistical parameters are not sufficient to define a distribution unequivocally [4].

Let us suppose a random process  $OR = X(\tau, t)$  with an average arrival rate of call-offs (representing potentially immediate executable manufacturing orders)  $E[OR]$  of  $\lambda = 12$  orders/day during a certain timeframe  $t$  and that the supplier has agreed with the customers to respect a commonly accepted *EDT*. These 12 orders per day might already be an average scale down from 60 weekly orders. We are used to transform weeks in days, days in hours, or even hours in minutes for high performance lean-takt lines or paced manufacturing cells, to bring it to shopfloor cycle time (*CT*) dimensions shown in Equation (5) which looks logic and apparently not questionable.

$$\lambda = 12/\text{day} = 1.5/\text{h} \quad (5)$$

with 8 working hours/day

The mean is an interesting parameter because it is a so-called non-biased estimator of a sample statistics for the population with a reduced SE of the mean given by  $\sigma_{E[X]} = \frac{\sigma_X}{\sqrt{n}}$ . Nevertheless, Equation (5) makes implicitly an important simplification of an even distribution of the order arrivals.

If we are bringing the average consideration of 12 orders/day to a proportional value of 1.5 order/h, we imply an even uniform spreading of arrivals over the time span of one day, which could be represented again by a TR. Then this TR can be approximated to a "deterministic" ergodic process. With an OR of 1.5/h, most probably, it is potentially possible to fulfill the OTD theorem of Equation 2 for all orders. It will, however, most probably not be possible to fulfill it with the mean logic of considering  $E[X] = 12/\text{day}$ , because it depends on the underlying order arrival distribution  $X$ . Indeed, for a potential bulk arrival of all 12 orders within one hour, only for some of the first orders of the bulk OTD may be respected, but not for all orders. This leads directly to the apparent paradoxical contradiction

$$\lambda = 12/\text{day} \neq 1.5/\text{h}$$

Therefore, we can generalize and let evolve the attention catching title of this section "Paradoxon of the mean" into the more appropriate statement "paradoxon of proportions". While an infinitesimal approach applied to the distribution would not rise any concern, the so often applied discrete view, using for simplification an average approach based on the mean, however, may become a



concern—in certain cases.

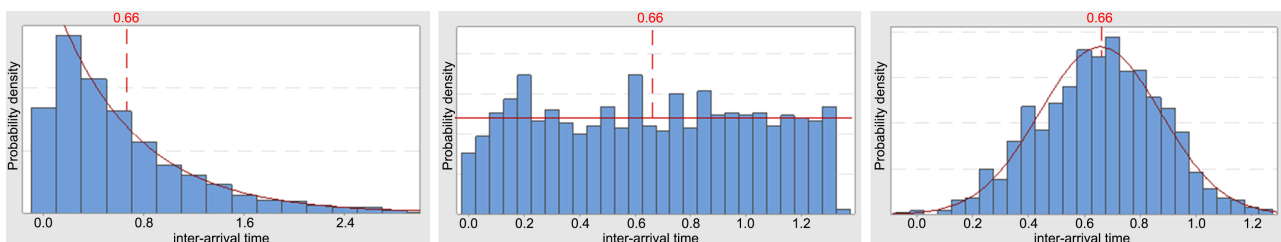
Indeed, we have to analyze the arrival process by taking into account its real distribution. Therefore, we will look at the OTD performance of a defined production system under different arrival random variables. For this purpose, we define the following queuing system taking alternative arrival processes into consideration. We consider a  $M$  = Markovian (*i.e.* exponential),  $U$  = uniform (even),  $N$  = normal distribution, and  $D$  = deterministic according to Dirac's  $\delta$ -function, all distributions with the same mean inter-arrival time  $E[\tau(OR)] = 1/\lambda$  where  $\tau(OR)$  denotes the inter-arrival time of the order rate distribution. In our example  $1/\lambda = 0.66$  h. Furthermore, let us take a deterministic  $D$  characterized server process, with only a single server, which in our case is a machine or workstation (no parallelization of processing capacity) and unlimited buffer capacity  $b$ , queuing system formalized as

$$\begin{cases} (X_i / D / 1 / \infty) \\ X_i := M, U, N, D \end{cases}$$

If the machine  $s$  is not a single stage server but a process, we get the case of a manufacturing process. Let us compare  $E_X[\tau(OR)] = 1/\lambda$  of the four arrival rate distributions  $X$  to the deterministic  $CT$  of the workstation or the  $CT_b$  of the bottleneck of the process, which determines the  $ER$  of the process, supposing to be  $CT_b = 1/\lambda$ . The distributions  $M, U, N$  are shown in **Figure 3**.

The probability of symmetric distributions  $P_X$  (the third statistical moment is  $\mu_3 = 0$ ) for shorter or longer arrival intervals than the cycle time of the bottleneck is intrinsically given by the definition of the mean (and median) and equals 0.5. The random variable  $X$  stands for the four specific distributions and  $P_X$  is the cumulative distribution function (*i.e.* it is the integral of the probability density function  $p_X$  of the arrival intervals).

Nota bene, the  $OR$  cannot really be influenced and hardly a company decides to refuse an incoming order. Therefore, to maintain an efficient and lean manufacturing process, the input rate ( $IR$ ) of the order release to the shopfloor will be adapted down to the existing  $ER$  capacity, with the consequence that orders exceeding instant capacity will become backlog ( $BL$ ) and therefore experience a  $BWT$ . This has not to be a concern—it is just a natural consequence of the capacity limit of the production line. However, OTD might be compromised. Nevertheless, this  $BL$  could then be rescheduled according to customer-specific



**Figure 3.** Arrival probability density functions (exponential, uniform, normal) generated randomly by 1000 data points with the same mean  $E_X[\tau] = 1/\lambda$  and  $SD_X[\tau] > 0$  where  $\tau$  is the inter-arrival time of  $OR$ .

urgencies with different scheduling principles, for instance LIFO instead of FIFO or a preferential “VIP” scheduling might be applied to observe OTD.

Now, transforming the Poissonian rate-view of the first condition of Equation 2 into the Markovian time-view becomes

$$\inf \{ER_i\} := \sup \{CT_i\} = CT_b < \frac{1}{E[OR]} = \frac{1}{\lambda}$$

We can therefore assume that for inter arrival times  $\tau = 1/x$  longer than  $CT_b$  (i.e.  $1/x > 1/\lambda = CT_b$ ) the first condition is observed, being possible to process immediately arriving orders if no WIP exists. On the other side, for the contrary with short arrivals (i.e.  $1/x < 1/\lambda = CT_b$ ), the capacity (i.e.  $ER$ ) will not be enough to process the orders immediately. Indeed, according to the *Theorem of WIP* [1], the orders have to queue forming a WIP, which we call BL at the begin of the process, with a certain BWT.

For the exponential distribution the probability  $P_M(\tau)$  of arrival times up to  $CT_b$  and above  $CT_b$  is given by the Equation system 6

$$\begin{cases} P_M(0 \leq \tau \leq CT_b) = 1 - e^{-\lambda \cdot CT_b} \\ P_M(CT_b < \tau \leq \infty) = e^{-\lambda \cdot CT_b} \end{cases} \tag{6}$$

The Markov distribution is very interesting, because it is characterized by only one parameter  $\lambda$ . This reality-matching probability density function is often used to model queuing systems, covering partially also the bulk arrival characteristics. It has the advantage of an easy to calculate integral of the cumulative distribution function, not necessitating statistical tables, which makes it the preferred stochastic function to model queuing systems. Without showing the calculus, assuming  $CT_b = E[\tau] = 1/\lambda$ , where  $\tau = 1/x$  then

$$\begin{cases} P_M(0 \leq \tau \leq CT_b) \approx 0.63 \\ P_M(CT_b < \tau \leq \infty) \approx 0.37 \end{cases} \tag{7}$$

meaning the probability of short inter-arrivals is higher than the long inter-arrivals above the mean, asymmetry which is typical of skewed probability density functions. Again, without showing the calculus, the median results to be

$$\tilde{E}[\tau] = \frac{1}{\lambda} \ln 2 < E[\tau] = \frac{1}{\lambda}$$

what demonstrates that the median is below the mean being

$$\begin{cases} P_M(0 \leq \tau \leq \tilde{E}[\tau]) = 0.5 \\ P_M(\tilde{E}[\tau] < \tau \leq \infty) = 0.5 \end{cases}$$

Remember the median corresponds to the 50% percentile. This is an indication that the mean arrivals should not be applied for the first condition of Equation (2); the median of arrivals would be more appropriate, because the number of arrivals is relevant. If the order arrivals are not single orders but each order has an own batch-size of its own, the situation changes further. This situation has to be taken into consideration in the case of a multi-product manufacturing

system. In this case here, we can neglect this further complication which would require an adaptation of the *OR* and *ER*.

For the uniform distribution, the probability  $P_U(\tau)$  of arrival times up to  $CT_b$  and above (assuming  $\tau_{\max} = 2CT_b$ ) is shown in Equation (8)

$$\begin{cases} P_U(0 \leq \tau \leq CT_b) = 1 - \frac{1}{\tau_{\max}} \cdot (\tau_{\max} - CT_b) = 0.5 \\ P_U(CT_b < \tau \leq \tau_{\max}) = \frac{1}{\tau_{\max}} \cdot (\tau_{\max} - CT_b) = 0.5 \end{cases} \tag{8}$$

The uniform distribution is simple, but doesn't reflect well the reality of arrival rates. Nevertheless, through its simplicity, it is very easy to analyze the post-optimal conditions to observe OTD.

The normal cumulative distribution function (we do not apply the better suited lognormal distribution for simplicity reason, although being conscious of non-existing negative arrival times) is shown in Equation (9). If we assume  $CT_b = 1/\lambda$  we obtain

$$\begin{cases} P_N(-\infty \leq \tau \leq CT_b) = 1 - \Phi(z) = 0.5 \\ P_N(CT_b < \tau \leq +\infty) = \Phi(z) = 0.5 \end{cases} \tag{9}$$

where  $\Phi(z)$ , the standardized integral of the probability density function  $p_N(\tau)$ , is a non-elementary function difficult to calculate

$$\Phi(z) := \left\{ \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}\tau^2} d\tau \mid z = \frac{\tau - \mu}{\sigma} \right\}$$

The symmetric normal distribution is made-up of two independent parameters, the mean and the variance. We will conceive the variance as small as necessary in order not to get negative arrivals. This parametric analysis of variance leads at the limit to Dirac's  $\delta$ -function

$$\lim_{\sigma \rightarrow 0} N(\mu, \sigma^2) = \delta(\tau)$$

which we model as deterministic arrivals in the following with  $SD_X[\tau] = 0$ . For deterministic arrivals following Dirac's  $\delta$ -function, the probability density function is summarized in Equation (10)

$$\begin{cases} p_D(0 \leq \tau < CT_b) = 0 \\ p_D(CT_b) = \infty \\ p_D(CT_b < \tau \leq +\infty) = 0 \end{cases} \tag{10}$$

with  $P_D(CT_b = \tau < \infty) = 1$ . Being  $X$  in this case a deterministic arrival process, we can substitute the concept of *OR* with that of  $TR = \lambda$  for which the Equation (1) is valid building-up no backlog with no *BWT*. Assuming no saturation effects of the process, the production process will therefore be able to process all arriving orders on-time. This is the ideal case for a deterministic *TR*-based *JIT*-production system. As this case is deterministic, no post-optimality analysis is necessary, because the *TR* has to be forcedly below the nominal *ER* to have a sustainable solution. However, we will analyze the post-optimality conditions for

a non-ergodic type process within a restricted time-period for which the deterministic  $OR$  is above the  $ER$ .

#### 4. Developing Post-Optimality Conditions to Respect OTD

To analyze the post-optimality conditions to respect OTD we have to distinguish different cases. The simplest case is an arrival order process of a single product and stage-changing deterministic OR (“deterministically non-ergodic” process). A more general case, always based on a mono-product manufacturing line, assumes the OR as a stochastic variable of ergodic process characteristic. For these two cases, the *Second Corollary to the Theorem of Throughput (Corollary of Bottleneck Time-invariance)* applies which states that the ER for a mono-product manufacturing line is time-invariant [1]. The most generalized case (not treated here) considers a multi-product line with stable product-mix or, increasing the complexity of the problem, with a variable product-mix of OR. In these multi-product cases the before mentioned *Corollary of Bottleneck Time-invariance* does not apply any more. We will concentrate in this paper on the first two cases of a mono-product manufacturing line and analyze the multi-product OR regime in a following paper.

##### 4.1. The Case of a Non-Ergodic Process with a Time-Bound Deterministic $OR$ above $ER$

In our case, the post-optimality conditions are determined by the deviation of the  $OR$  from the nominal capacity of the production system, which, according to the *Theorem of Throughput*, is given for a mono-product manufacturing line by the deterministic cycle time  $CT_b$  at the bottleneck. According to the OTD theorem and Equation system 2, the solution space is also limited by  $EDT$ . Whereas  $OR < ER$  is not critical,  $OR > ER$  might become a concern. The post-optimality question is: what is the  $\sup\{OR\}$  and the maximum timespan  $\sup\{\Delta t\}$  of a temporarily higher order-entry regime while still observing the second condition of Equation system 2. During the  $\Delta t$  the  $OR$  behaves like a  $TR$ . We are expressly writing  $\sup\{OR\}$  and not using the diction  $\max\{OR\}$ , because we are not intending to maximize a target function, *i.e.* solving an optimization problem, but we are interested in the largest admissible value of the random-deterministic variable  $OR$ .

In this case of non-ergodicity we will assume for  $OR(t)$  and especially for the  $\sup\{OR\}$  deterministic values over the timespan  $\Delta t$ . To develop the post-optimality conditions for observing OTD we have to take Equation system 2 and solve it in view to determine the solution space for the  $OR$ . Please note that apart from the SPQR axiom, if the *Theorem of General Production Requirements* would not have been enounced, the problem of post-optimality of order entry cannot be defined, hence the fundamental importance of this OTD theorem for production systems. Notice that the compliance of OTD refers directly to the punctuality axiom.

For developing the post-optimality conditions, we can rewrite the second condition of Equation (2) as

$$BWT \leq EDT - PLT = \Delta T$$

where  $\Delta T$  allows to let aside the real values of  $EDT$  and  $PLT$ . In any case it has to be  $BWT < \Delta T$ . The backlog waiting time ( $BWT$ ) can be modelled by setting  $OR = x$  and  $ER = \lambda$

$$BWT \leq \frac{BL}{ER} = BL \cdot CT_b = BL \cdot \frac{1}{\lambda} = (x - \lambda) \Delta t \cdot \frac{1}{\lambda}$$

and where the backlog ( $BL$ ) has been computed as

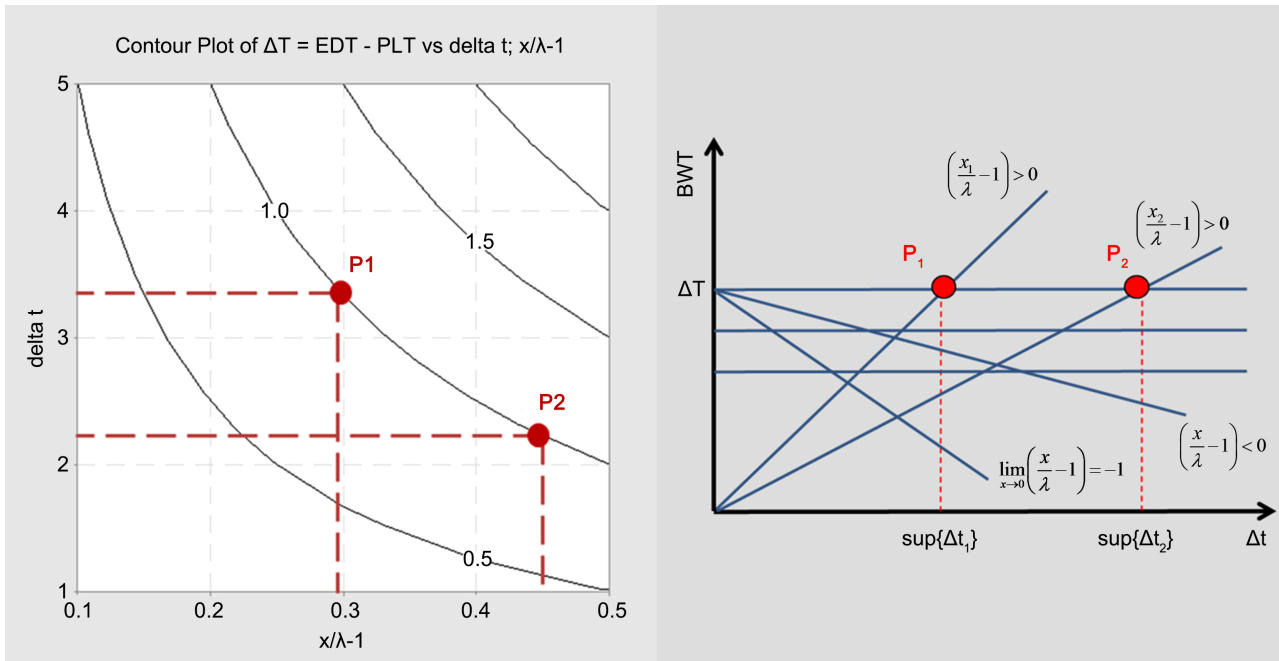
$$BL = \int_{t_1}^{t_2} (OR - ER) dt = (x - \lambda) \cdot t \Big|_{t_1}^{t_2} = (x - \lambda) \cdot \Delta t$$

and where  $\Delta t$  is the timespan of duration the higher order regime ( $x > \lambda$ ) is active. The resulting post-optimality condition to determine the admissible range of data for  $x$  and  $\Delta t$  is shown in Equation (11)

$$\left( \frac{x(t)}{\lambda} - 1 \right) \cdot \Delta t \leq f(\Delta T) \quad (11)$$

Equation (11) is a parametric relation of implicit type, such as  $f(x, y, k) = 0$ , with the variables  $x(t)$  and  $\Delta t$  in parametric function of  $\Delta T$ , *i.e.* the difference between the customer's expected delivery time and the process lead time. **Therefore, given  $EDT$  and being  $\Delta T = EDT - PLT$ , the faster the throughput speed  $PLT$  is, the production systems can process a higher order rate during a certain timespan while respecting OTD. This intuitive, but now finally derived and explicitly formulated conclusion is very important, because it allows for non-TR governed production systems to play with the backlog. That means high order entry can be managed by buffering temporarily instant-capacity exceeding orders as  $BL$  and reschedule them accordingly to observe OTD.** These findings are compliant to the *Lemma of Flexible Scheduling Principle* [1] which refers to how to manage the shopfloor order release of a forming backlog. This possibility avoids expensive investment in overcapacities with their costly consequences, which are a big issue in today's supply chains of fragmented value-add economy [5] [6]. The graphical interdependence of variables resulting from Equation (11) is shown in **Figure 4** for parametric iso  $\Delta T$  lines.

From **Figure 4** is evident that different combinations of higher order entry during a certain timespan are possible satisfying the iso  $\Delta T$  lines. It reveals that it is possible to have in the shown case a 30% higher order entry than the capacity during 3.3 time units (P1) or 45% higher OR during 2.2 time units (P2) and still observe OTD. If  $\left( \frac{x}{\lambda} - 1 \right) > 0$  then BWT will increase, else decrease (**Figure 4**, right side). The virtual elasticity can be exploited until the BWT has reached  $\Delta T$ , which represents the maximum intrinsic virtual capacity reserve of fast production lines. When this virtual production capacity has been exploited, backlog has to diminish with an  $OR$  below  $ER$  to be exploited again. The maximum reduction



**Figure 4.** Graphic representation of post-optimality (in function of  $x, \Delta t$ ) regarding OTD for higher order rate  $x$  than equilibrium  $\lambda$  modelled with Equation (11). The graphic shows exemplarily two equivalent post-optimality situations P1 and P2 for a given isoline  $\Delta T = 1$  (of any given time-unit) from different possible  $\Delta T = EDT - PLT$  isolines (left side). Right side: Alternative representation of the same concept adding *BL*-reducing *OR* conditions (qualitative representation).

rate of *BL* is when  $\lim_{x \rightarrow 0} \left( \frac{x}{\lambda} - 1 \right) = -1$  representing the case of no-order entry.

These time-bound discrete values of  $x$  refer to deterministic order entry, an *OR* situation which corresponds seldom to reality. Such a changing deterministic process behavior with  $E_D[\tau_1, \Delta t] \neq E_D[\tau_2, \Delta t]$  and  $SD_D[\tau, \Delta t] = 0$  within defined timespans  $\Delta t$  can be assimilated to a non-ergodic process.

#### 4.2. The Case of a Stochastic *OR* with Stationary Ergodic Process Characteristic

If the order rate  $x$  is not a deterministic value, as assumed before, but more realistically a random variable  $X$ , the probability to experience a higher order rate  $x$  than  $\lambda$  depends on the probability density function and is (written in the Markovian time-dimension of inter-arrival time  $\tau$ )

$$P_x \left( \frac{1}{x} < \tau < \frac{1}{\lambda} \right) = \int_{1/x}^{1/\lambda} p_x(\lambda, \tau) d\tau$$

and depends not only on the increased *OR*, but also on the type of probability density function  $p_x(\tau)$ . Please note that here  $\tau$  stands for the inter-arrival time. Do not confuse it with the timespan  $\Delta t$  during which a higher *OR* is present. Let us take as an example an *OR* higher up to 30% respect to  $\lambda = 1.5/\text{h}$  resulting in extremis  $x = 1.95$  orders/h. This corresponds for the Markovian time view to a shorter inter-arrival time of  $1/x = 0.51$  h compared to the  $CT_b = 1/\lambda = 0.66$  h resulting into a 23% shorter inter-arrival time. **Figure 5** shows the probability for

the three different probability density functions for shorter inter-arrival times up to 0.51 h, as deviations from the mean  $1/\lambda = 0.66$  h (red area).

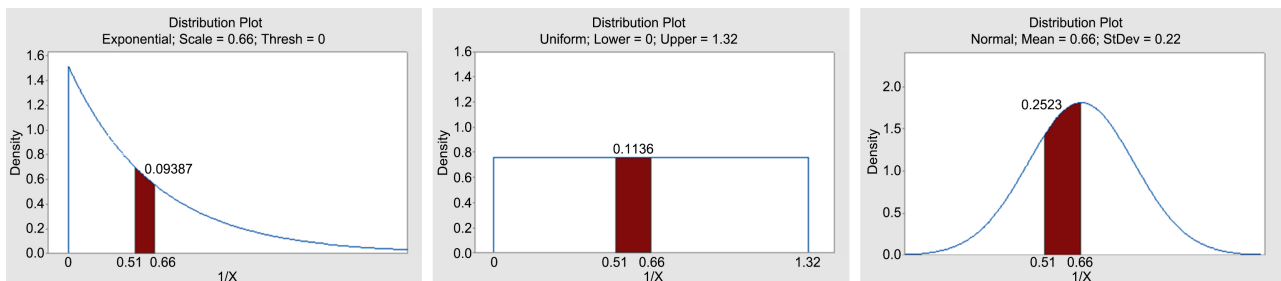
The computation reveals for a Markovian exponential process a probability of 9.4%, for an even distribution a probability of 11.4%, and for a normal distribution a probability of 25.2%. The big difference between 9.4% and 25.2% shows that the assumed underlying distribution of arrival rate  $OR(\tau)$  is of utmost importance. Therefore, for non-deterministic arrival rates, *i.e.* for a random variable  $X$ , Equation (11) has to be rewritten resulting in Equation (12)

$$\int_{t_1}^{t_2} \left( \frac{x(\tau, t)}{\lambda} - 1 \right) \cdot dt \leq f(\Delta T) \quad (12)$$

Equation (12) models the generic condition for post-optimality of a stochastic process to observe OTD. If the stochastic variable  $x(\tau, t)$  becomes time-invariant  $x(\tau)$ , *i.e.*  $E[X] = \text{const}$  and  $SD[X] = \text{const}$ , then the process is ergodic. The computational solution of this integral equation for different  $OR$  distribution  $x(\tau)$  is left to the interested reader.

In Equation (11) we have assumed a deterministic  $OR$  temporarily higher than the nominal capacity calculating the timespan during which a higher  $OR$  regime might be “digested” by the manufacturing system. However, this is not the typical case of a random variable. Indeed, how a random variable is defined, also lower  $OR$  than the mean  $E[X]$ , *i.e.* the assumed capacity given here by  $\lambda$ , have a certain probability to occur reducing the BL-WIP. Equation (12) has a general validity covering also this real situation of a random variable. **Equation (12) gives the general post-optimality condition for stochastic OR in presence of a fixed capacity characterized by the ER in parametric function of faster PLT than the requested EDT. This difference  $\Delta T$  between PLT and EDT can be considered to be the built-in virtual elastic capacity of a production system.**

We can therefore enounce the following *Corollary of Post-Optimality* (or *Corollary of Virtual Elasticity*). This new corollary belongs to the mathematical category of *problem existence* and in this case, not giving the solution, it is of non-constructive type for which Equation (12) is a non-constructive proof for Equation (2) (OTD theorem):



**Figure 5.** Probabilities of occurrences with inter-arrival time  $\tau = 1/x$  in the interval of  $[0.51; 0.66]$  hours for exponential, uniform, and normal distribution of  $\tau$ .

***First Corollary to the Theorem of General Production Requirements (Corollary of Post-Optimality or Virtual Elasticity)***

*Given is a production system with a deterministic exit rate  $ER$  and a stable process lead-time  $PLT$ . The market is characterized by a stochastic order arrival rate  $OR$  each single order with expected delivery time  $EDT$ . The post-optimality range to comply to on-time-deliveries  $OTD$  depends on the difference  $\Delta T$  between  $EDT$  and  $PLT$  as well as from the distribution of  $OR$ . This virtually increased capacity is called virtual elasticity of a production system.*

This corollary seems to be in contradiction to the  $OTD$  theorem itself because according to the corollary it can temporarily also result  $OR > ER$  while still respecting  $OTD$ . This has not to be a contradiction because the  $OTD$  theorem is generally valid and over a long time-span whereas this corollary is valid for a limited time-span (post-optimality condition) according to Equation (12) (indeed, by increasing virtually production capacity) and constitute generally only a weak solution for  $OTD$ . However, we can therefore take a practical recommendation, synthesized in the following

***Lemma to the Corollary of Post-Optimality (Lemma of Increasing Speed)***

*Based on  $\Delta T = EDT - PLT$ , accelerating  $PLT$  of the production process increases the manufacturing-inherent reserve  $\Delta T$  of production capacity, i.e. the virtual elasticity, in order to observe  $OTD$  for stochastic  $OR$ .*

Therefore, it is always recommendable to speed-up processes and to reduce  $PLT$ . This is in perfect concordance with the *Lemma to the Theorem of Generalized Lead Time (Lemma of SPF Desirability)* [1], which invites trying to install always a fast single piece flow (SPF) in order to have a shorter  $PLT$ .

The new coined term *virtual elasticity* is opposed to costly conceived *physical elasticity* of production equipment, which is depending on the BEP. Elasticity, i.e. the non-sensitivity of production cost to load-changes, will become very important for so-called Industry 4.0 type production systems allowing batch size one [7] [8]. Increasing speed, i.e. reducing  $PLT$  and at the same time shorten cash-to-cash cycle, is therefore a very important topic on every production manager's agenda and a topic of core attention in the domain of JIT Lean production systems. In order to respect  $OTD$ , the BL can be rescheduled accordingly to optimize the scheduling sequence of orders to respect  $OTD$ . Why is Equation (12) a very important result especially in view of Industry 4.0 type of manufacturing systems? These types of production systems show, or aim to allow, a high variance of  $OR$  as well as a high mix variability, stating literally, for instance, to have the possibility "to put a Porsche seat into a VW" [7] [8]. Such generalized  $OR$  processes are highly non-ergodic. However, the aim to be able to produce a one-off, or more generally a variable non-defined mix, entails a dynamic bottleneck, i.e. a changing  $ER = \lambda$ , leading to the undesired effect of unevenness, which Toyota called Mura, leading finally to Muda (waste). However, the Toyota JIT lean production system bases on ergodic-type of processes. This undesired consequence of changing mix and dynamically changing bottleneck leads to another mathematical problem—an optimization problem to be further investi-



gated. Indeed, instead of applying discrete brute-force simulation methods, Equation (12) allows to frame a functional understanding of the “physics” to respects OTD for general conditions. Please note that the capacity is only virtually increased. Indeed, the nominal specific capacity is still given by  $CT_b$  *i.e.* the cycle time at the bottleneck.

Furthermore, symmetric probability density functions, such as uniform ( $U$ ) or normal ( $N$ ) distributions, show the same cumulative distribution probability on the left and on the right side of the mean or median, levelling out in the average higher and lower  $OR$  than the  $ER$ . This is not the case for a skewed probability density such as the exponential Markovian ( $M$ ) density function, *i.e.*

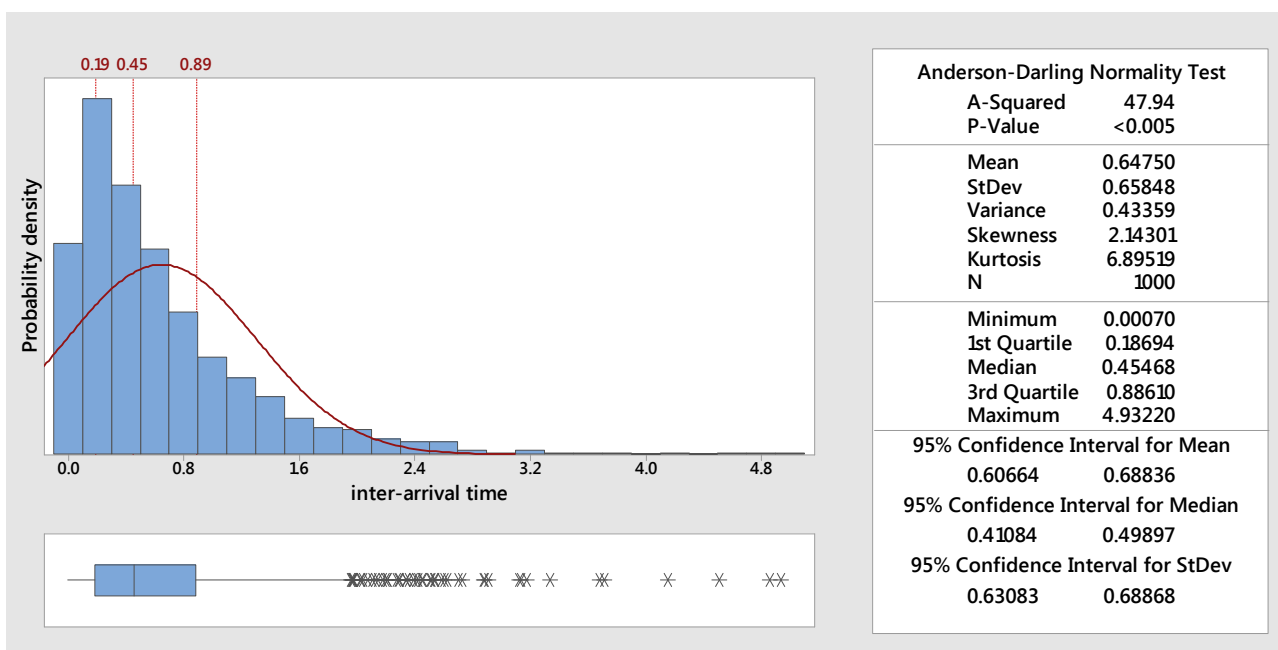
$$\int_0^{1/\lambda} p_M(\tau) d\tau \neq \int_{1/\lambda}^{\infty} p_M(\tau) d\tau \tag{13}$$

where  $E[\tau] = 1/\lambda$ , and referring to the case before with  $\lambda = 1.5$

$$P_M(0 < \tau < 0.66) > P_M(0.66 < \tau < \infty) \text{ i.e. } 0.63 > 0.37 \text{ (QED)}$$

which confirms the statement of Equation (7). Due to the asymmetric probability of the probability function with regard to the mean (Equation (13)), this would imply an increasing backlog. Excluding seasonality effects and economic growth, this suggests that not the mean, but the median has to be taken as reference for capacity considerations. **Figure 6** shows the Markovian probability density function of inter-arrival times with its descriptive statistics.

Now, the asymmetric extension of inter-arrival time of probability  $P_M(0.19 < x < 0.45)$  with higher density and  $P_M(0.45 < x < 0.89)$  with lower density have the same probabilities to occur which is 0.25 corresponding to the definition of quartile.



**Figure 6.** Randomly generated Markovian distributed inter-arrival times with  $\lambda = 1.5$  showing 1<sup>st</sup>, 2<sup>nd</sup> (median), and 3<sup>rd</sup> quartile.

These quartile/median figures are quite different from the figures computed with Equation (13) referring to the mean. These differences do not exist in symmetric distributions.

Conclusion: If symmetric distributions show random arrivals, they represent stochastically a strong solution for OTD depending only on Equation (2a). If they show an enhanced but non-random behavior above the mean over a certain timespan, they can only have a weak solution for OTD. Asymmetric distributions always represent only weak solutions. A strong solution means that OTD can probabilistically always be fulfilled, weak solution means that it can be fulfilled under certain circumstances given by the *Corollary of Post-Optimality* just enounced before. We can therefore enounce the

***Second Corollary to the Theorem of General Production Requirements (Corollary of Strong and Weak OTD Solutions)***

*Given is an order rate with generally random arrival character within a fixed-capacity regime. Necessary and sufficient condition for a strong OTD solution is the ergodic characteristic of a stochastic order process paired with asymmetric distribution of order rates of the random OR variable and if they comply with the requirements of the OTD theorem. However, non-ergodic processes with arbitrary arrivals, or ergodic processes with asymmetric random distributions of order rate, constitute only a weak solution for OTD observance that is depending on the specific circumstances. The circumstances are the distribution of arrivals, the sequence of arrivals, and timespan of non-randomness, as well as dependency on the virtual elasticity of the production system.*

Please note that this just enounced corollary constitutes a constructive proof to Equation (12), whereas the *Corollary of Post-Optimality* belongs to the mathematical category of *existence problem* with non-constructive proof. The *Corollary of Strong and Weak OTD Solutions* states clearly that a Poissonian-distributed order rate process in this case might only constitute a weak solution to respect OTD.

### 4.3. Practical Application of Theoretic Findings

An increased order entry is beneficial for every enterprise. With the utilization of the virtual elasticity and despite limited capacity, in certain cases, it is possible to observe the mandatory OTD in order to maintain a high related customer satisfaction referred to punctuality of deliveries. However, in extreme cases it might be necessary to increase capacity by additional shifts. In this case, additional aspects have to be taken into consideration to benefit from increased order intake:

Two aspects have to be satisfied

- Maximizing marginal manufacturing elasticity
- Maximizing margin contribution  $m(V)$

These two conditions translate into (see **Figure 2** and **Figure 7**)

$$\frac{d}{dV}k_v(V) = 0$$

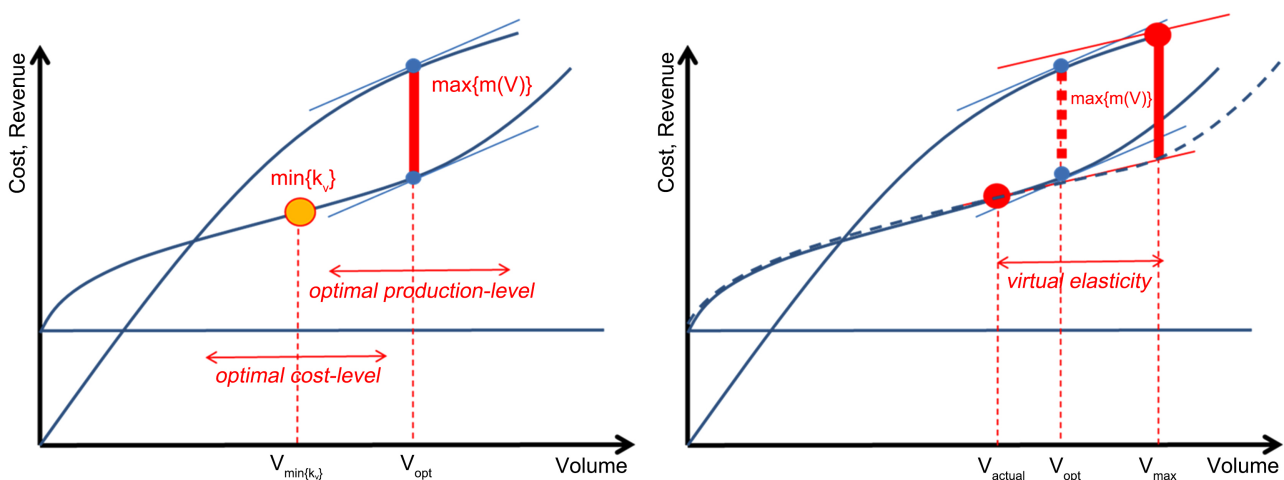
and

$$\frac{d}{dV}m(V) = 0$$

The concept of optimal cost-level and optimal production-level are schematically shown in **Figure 7** on the left side. The optimal cost-level is subordinate to the optimal production-level to maximize profit, and generally, the optimal production-level is above the optimal cost-level.

However, the virtual elasticity allows to process increased order entry while still satisfying the OTD customer requirements. Interesting is that the variable production cost are not those of higher production volume, because the temporarily increased volume are produced at the same cost of the actual (not speed-up produced) volume  $V_{act}$  within the  $\Delta T$  timespan, if staying in the same shift regime. Given that the maximum production level depends also on the price curve, the maximum volume  $V_{max}$  can go theoretically beyond the optimal volume  $V_{opt}$  to maximize margin contribution if the  $\Delta T$  timespan is large enough (right graph of **Figure 7**). Please note that this statement is valid for a certain timespan  $\Delta t$  and not for the entire production. This reasoning is based on the assumption, that the further order intake in the future is unknown. Indeed, in a real situation we have a case after case order-intake evaluation and not the micro-economic aggregated ex-post overview according to **Figure 7**. In other words, the price reality of the arriving orders is not of ordered decreasing character but with varying prices of random orders according to the exact chronological order entry sequence. However, with these order prices in random sequence the revenue curve can be approximated with an average linear specific price curve.

From an economic point of view, the additional order entry is limited to the condition of marginal sales equaling marginal virtual elasticity variable cost, condition given in Equation (14)



**Figure 7.** Graphical representation of optimal cost and optimal production level (left side) as well as the representation of virtual elasticity (right side) enabling to increase maximum margin contribution theoretically even beyond the level of classic optimal volume for respecting OTD (idealized and much exaggerated conceptual representation).

$$V_{\max} := \left\{ V \mid \frac{d}{dV} p(\Delta V_{\text{OTD}}) = \frac{d}{dV} k_v(V_{\text{elast}}) \right\} > V_{\text{opt}} \quad (14)$$

Although Equation (14) might be of limited applicability in high performance production systems given by reduced  $\Delta T$  it is for sure of theoretic importance. Nevertheless, we can come back to the *Theorem of Optimal Production Range* and therefore enounce the following lemma:

***Lemma to the Theorem of Optimal Production Range (Lemma of Hidden Potential Exploitation)***

*Based on the realistic BEP model determining an optimal production volume to maximize margin contribution with an existent equipment, exploiting the property of virtual elasticity it is possible to extend maximization of margin contribution theoretically even beyond the common optimal production volume. This is possible due to the property of virtual elasticity of production systems and the resulting virtually increased capacity by speeding-up processes (Lemma of Increased Speed) in order to observe OTD also for increased order entry.*

The finding summarized in Equation (14) and paraphrased with the *Lemma of Hidden Potential Exploitation* is an important revelation in the world of economic manufacturing and adds new insights of manufacturing-related properties of every production systems. This revelation is as well as important for practical reasons as for theoretical understanding of the value of speed. It helps to face the challenge of increased order entry with variable arrival rate above the nominal capacity allowing to satisfy OTD without physical capacity extension, at least for a limited timespan. Although this finding is very interesting, it has to be admitted that the virtual elasticity cannot replace consistent physical nominal extension of capacity and are limited within the scope of  $\Delta T$  to observe OTD. It is ideally suited to absorb order peaks, but not structurally increased order levels.

It may emerge the question of the practical applicability, or better practical implementation, of these findings. Two related topics have to be investigated: firstly, finding manufacturing solutions to speed-up processes to increase virtual capacity, and secondly, finding scheduling algorithms allowing most of the backlogged orders to satisfy OTD. The first topic can be easily resolved by TPS-derived Lean JIT techniques, hence the importance to transform B&Q Systems into Lean JIT production systems [2]. Traditional concepts applied to speed-up processes are e.g.:

- Shorten CT at the bottleneck (increase capacity without investing)
- Eliminate non value-add steps in the process such as searching (eliminate Muda)
- Reduce batch-size (in order to reduce WIP).

To generate significantly the virtual elasticity, however, introducing a SPF is mandatory, see the *Main Theorem of Production Time (or SPF Dominance Theorem)* and its corollary and lemma [1]. In lean JIT manufacturing systems, however, the virtual elasticity is usually already exploited.

## 5. Future Research Opportunities

Lean JIT-based production systems show a superior performance versus traditional B&Q production systems and stand at the base to reduce *PLT* and respecting OTD [2]. However, production systems are usually not *TR*-based but stochastic *OR*-governed due to variable order entry rate reflecting natural market demand fluctuation. This topic has just been treated summarily in this paper for a single product with deterministic *ER*. However, Industry 4.0 production systems aim at enabling an infinite mix variability including to attain batch size one with variable *PLT* according to change of mix [7] [8]. This high mix variability induces a variable bottleneck with variable *ER* of the production system. The questions, however, are what the constructive proof of post-optimality conditions in such a production regime will be and, especially, what type of scheduling algorithms are recommendable for non-ergodic stochastic variable mix regime with consequent variable *ER* and variable *PLT*. This quest might be extended beyond the usual application of brute force simulation methods, which can be of NP-complexity. The question is which new laws and principles can be enounced to optimize *ER* and *PLT* to meet OTD in a highly variable production context. In the case of additional complexity given by flexible non-deterministic product-mix order entry, and thus a variable *ER*, an intelligent scheduling algorithm must also take into consideration the set-up time of different equipment as well as transfer time in a graph-modeled equipment layout [2]. The Lean SMED Technique (Single Minute Exchange of Die) might get, in this case, an additional core aspect to be focused on.

## 6. Summary

Originally, Lean JIT production systems have been conceived for stable *TR*-based production regimes for the manufacturing of a deterministic product mix with adequate, but inelastic capacities. These production systems are largely representative for the automotive industry. However, in many industries, order entry is a random variable. This makes it difficult, but not impossible to implement JIT-like Lean systems. In addition, economy shows seasonal patterns requiring a variable capacity, not mentioning the variability originated by pipeline-filling effects of fragmented global supply chains during economic cycles. This complex variability of order entry may lead to costly investment in order to observe OTD resulting in production overcapacities. This does not have to be the case, at least not for limited variability and limited timespans, if the virtual elasticity of a production system is discovered, implemented, and exploited. The virtual elasticity is the result of the inherent latent production capacity resulting from the difference between *EDT* and *PLT*. This insight is an important result. Even though the findings of the paper may seem trivial, it has to be stated that this is not self-evident. Indeed, instead of investing in costly production capacities covering the upper range of order variability that is leading to overcapacities, the latent capacity intrinsic to every production systems given by increased speed of SPF and

JIT production systems can be exploited especially for arrival distributions constituting a weak solution for OTD. It goes without saying that reducing *PLT* is not intended simply to operate the equipment at higher speeds, but by applying appropriate manufacturing and transfer principles in order to implement a high performance production system.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Rüttimann, B.G. (2017) Lean Compendium—Introduction to Modern Manufacturing Theory. Springer. <https://doi.org/10.1007/978-3-319-58601-4>
- [2] Rüttimann, B.G. and Stöckli, M.T. (2020) From Batch & Queue to Industry 4.0—Type Manufacturing Systems: A Taxonomy of Alternative Production Models. *JSSM*, **13**, 299-316. <https://doi.org/10.4236/jssm.2020.132019>
- [3] Rüttimann, B.G. and Stöckli, M.T. (2016) Lean and Industry 4.0—Twins, Partners, or Contenders? A Due Clarification Regarding the Supposed Clash of Two Production Systems. *JSSM*, **9**, 485-500. <https://doi.org/10.4236/jssm.2016.96051>
- [4] Matejka, J. and Fitzmaurice, G. (2017) White Paper, Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. Autodesk Research, Toronto. <https://doi.org/10.1145/3025453.3025912>
- [5] Rüttimann (2001) The Pipeline-Filling Effect. *Aluminium International Journal*, Giesel Verlag.
- [6] Fischer and Rüttimann (2009) The Curse of Globalization—Must We Expect Crisis in the Aluminium Industry that Are More Abrupt in the Future? *Aluminium International Journal*, Giesel Verlag.
- [7] (2013) Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0—Abschlussbericht des Arbeitskreises Industrie 4.0. <https://www.plattform-i40.de/PI40/Navigation/DE/Home/home.html>
- [8] Zukunftsbild Industrie 4.0. <https://www.plattform-i40.de/PI40/Navigation/DE/Home/home.html>