Continuous time Restless Bandit and dynamic scheduling for make-to-stock production

F. Dusonchet* and M.-O.Hongler
EPFL
Institut de Production et Robotique (IPR)
Laboratoire de production Microtechnique (LPM)

Abstract— **We study the Whittle's relaxed version of the continuous time, discrete and continuous states space Restless Bandit problem under the discounted cost criterion. Explicit expressions for the priority indices, which generalize the Gittins indices, are derived. This formalism is then used in the context of flexible make-to-stock production manufacturing to construct dynamic scheduling rules. These analytical results are finally compared with the numerically derived optimal policy, obtained for a server delivering two types of items. It is observed that the Whittle's relaxed version of the Restless Bandit model yields nearly optimal dynamic scheduling rules.**

Keywords— **Flexible Make-to-Stock Production, Random Production Flows, Hedging Stocks, Multi-Armed Bandit Problem, Restless Bandit Problem, Priority Indices.**

I. Introduction

Consider the following classical problem of a multiclass, make-to-stock production system subject to breakdowns and repairs:

A machine is able to produce N different types of items, finished items are stored into N different respective finished good inventories (FGI). Write

$$\vec{X}(t) = (X_1(t), \ldots, X_k(t), \ldots, X_N(t)) \in \Omega^N \subset \mathbb{R}^N \quad (1)$$

to describe the net inventory process characterizing the population levels in the FGI's at time t. The negative values of $X_k(t)$ describe the presence of backorder (i.e. demands that cannot be immediately satisfied). Let us assume that we can schedule the production facility by using a "bang-bang" type control variable

$$\vec{u}(t) = \Big\{ (u_1(t), \ldots, u_N(t)) \ ; \ \ u_k(t) \in \{0,1\} \Big\}$$

i.e. for the machine in its operating state $u_k(t) = 1$, means that the production of the items of type k is engaged and $u_k(t) = 0$ means that no item of type k is produced. Note that $u_k(t) = 1$ does not imply that the machine is actually producing an item of type k, it can indeed be failed.

Both, the instantaneous controlled production rate $\vec{P}(t)$ and demand rate $\vec{D}(t)$ are supposed to be independent, stationary Markovian random vector-processes defined on a probability space $(\tilde{\Omega}, \mathcal{F}, \mathbb{P})$. We shall respectively write:

$$\vec{P}(t) = \big(P_1(t), \ldots, P_k(t), \ldots, P_N(t)\big) \quad (2)$$

*In part supported by the "Fonds National Suisse pour la Recherche Scientifique"

and

$$\vec{D}(t) = \big(D_1(t), \ldots, D_k(t), \ldots, D_N(t)\big). \quad (3)$$

The $P_k(t)$ are independent random processes which model the controlled production rate for the items of type k and the $D_k(t)$ are independent random processes which model the demand for the items of type k. Moreover, the $P_k(t)$ processes satisfy the constraints:

$$0 \leq P_k(t) \leq \mathcal{C}_k(t), \ t \geq 0, \ k = 1, \ldots N,$$

where $\mathcal{C}_k(t) \in \{0, c_k\}$ denote the stochastic total production capacity processes of the failure prone machine. Here we assume the $\mathcal{C}_k(t)$ to be alternating Markov renewal processes (i.e. two states Markov processes). For an item of type k the value $\mathcal{C}_k(t) = c_k$ is its maximal instantaneous production and $\mathcal{C}_k(t) = 0$ represents the failure state of the machine.

We impose the following conditions on the controlled production process $\vec{P}(t)$:

a) The production is non-preemptive (i.e. the machine must finish the item it is currently producing before starting a new one).

b) The production facility has a limited capacity. Accordingly it can engage, at each time, the production of only a single item (i.e. at each time $t \in \mathbb{R}^+$, at most one of the N controls $u_k(t)$ equals to one).

In order to schedule the production, we introduce a scheduling policy π, which is a mapping:

$$\begin{array}{rccc} \pi \ : & \Omega^N & \to & \{0,1\}^N \\ & \vec{X}(t) & \mapsto & \vec{u}(t). \end{array}$$

The function $\vec{u}(t)$ defines which type of item the machine is engaged on at time t.

Definition (Admissible policy): A policy $\pi(t)$ is admissible if it depends only on the past and on the present state of the random net inventory process $\vec{X}(t)$.

We shall only consider in the sequel admissible policies for which $\vec{P}(t)$ and $\vec{u}(t)$ fulfill both conditions a) and b) above.

The set of admissible policies will be denoted by $\mathcal{U}$.

Given a policy $\pi \in \mathcal{U}$, let us denote the cumulated production up to time t by

$$\vec{\mathcal{P}}_\pi(t) = \int_0^t \vec{P}(t)\, dt$$

and the cumulated demand up to time t by

$$\vec{\mathcal{D}}(t) = \int_0^t \vec{D}(t)\, dt.$$

In terms of $\vec{\mathcal{P}}_\pi(t)$ and $\vec{\mathcal{D}}(t)$, the net inventory process can therefore be expressed as:

$$\vec{X}(t) = \vec{\mathcal{P}}_\pi(t) - \vec{\mathcal{D}}(t), \tag{4}$$

with the initial conditions

$$\vec{X}(t_0) = \vec{x}_0 \in \Omega^N \text{ and } \vec{u}(t_0) = \vec{u}_0.$$

Let us further introduce the instantaneous running cost $h(\vec{X}(t), \vec{u}(t))$ associated with the global state of the net inventory $\vec{X}(t)$ and the production engagement $\vec{u}(t)$:

$$h \ : \ \Omega^N \times \{0,1\}^N \ \rightarrow \ \mathbb{R}^+$$

$$(\vec{X}(t), \vec{u}(t)) \, h(\vec{X}(t), \vec{u}(t)) = (h_1^\theta(x_1), \ldots, h_N^\theta(x_N)),$$

where $\theta = a$ when the production of the k-type items is engaged in the state x_k (i.e. $u_k(t) = 1$) and $\theta = p$ when the production is switched off (i.e. $u_k(t) = 0$).

Assuming an infinite time horizon, we define, for an initial condition $(\vec{x}_0, \vec{u}_0)$, the total production cost $J^\pi(\vec{x}_0, \vec{u}_0)$ under policy π by:

$$J^\pi(\vec{x}_0, \vec{u}_0) := E^\pi_{(\vec{x}_0, \vec{u}_0)} \int_0^\infty e^{-\delta t} h(\vec{X}(t), \vec{u}(t))\, dt, \tag{5}$$

with $E^\pi_{\vec{x}_0, \vec{u}_0}$ denoting the expectation operator, conditional to the initial condition $(\vec{x}_0, \vec{u}_0)$. The term $e^{-\delta t} > 0$ is a discounting factor. Clearly, the integral in Eq.(5) does not converge for arbitrary cost functions $h(\vec{x}, \vec{u})$. In the following, we assume that $h(\vec{x}, \vec{u})$ satisfies the required regularities conditions, to insure that Eq.(5) exists $\forall \pi \in \mathcal{U}$ and all initial conditions.

The optimal control problem is to determine the optimal policy π^* minimizing the total production cost given in Eq.(5). The value of the production cost under the optimal policy π^* will be denoted by:

$$J^*(\vec{x}_0, \vec{u}_0) := J^{\pi^*}(\vec{x}_0, \vec{u}_0) = \inf_{\pi \in \mathcal{U}} J^\pi(\vec{x}_0, \vec{u}_0).$$

In the following, we shall focus our attention to problems for which the **set-up costs and/or time delays**, often needed to switch the production from one type to another, **are negligible**. Hence, the initial condition $\vec{u}_0$ is not needed to characterize the dynamic of the processes $X_k(t)$. We can indeed initially change the production engagement, without incurring setup penalties. The explicit mention of the initial state $\vec{u}_0$ will hence be omitted in the sequel.

In absence of setup penalties, we will see that the formalism given by the so called "Restless Bandit" problem (RBP) is relevant to model the scheduling of flexible production systems (the RBP formalism was first discussed by P. Whittle [21]). Finding the optimal scheduling rule for a flexible production then essentially amounts to find the optimal policy for the RBP. In section II, we will observe, that the absence of setup penalties enables to approximately decouple the original N-armed RBP into N single-item make-to-stock production processes [21]:

$$X_k(t) = \mathcal{P}_k(t) - \mathcal{D}_k(t), \ k = 1, \ldots, N, \tag{6}$$

for which the scheduling policy is much easier to determine. In fact, the decoupling into single production processes enables us to construct indices $\nu_k(x_k)$, similar to those proposed by P. Whittle in [21], from which a so called "Priority index policy" can be derived.

Definition (Priority index policy) : A priority index policy is a scheduling rule, based on the existence of indices $\nu_k(x_k)$, depending only on the inventory state $X_k(t) = x_k$ of the items of type k. In terms of the $\nu_k(x_k)$, the priority index policy commands: "*At each decision time, engage the project exhibiting the smallest index value* $\nu_k(x_k)$".

Remark: For the scheduling production problem, an index $\nu_{N+1}(x)$ will be be introduced to take into account the fact that the machine can be idle. In the following, we will call this index "*the idle index*". When this idle index is smaller than all the other indices, the machine does not produce any item.

For the general k-items scheduling problems, similar decoupling approximation methods form the core of several recent contributions such as [14], [17], [18], where priority index policies are shown to provide suboptimal, but efficient, scheduling rules. In particular, in [17], the authors use the RBP to describe an heuristic scheduling rule for a multiclass, make-to-stock $M/M/1$ queuing system working with lost sale and under the time average cost criterion (i.e. $\delta = 0$).

Remember that priority index policies have been shown to give optimal rules for the class of Static Bandit problem (SBP, also called Classical Bandit problems), characterized by the fact that when not engaged, the projects remain "frozen" (i.e do not evolve in time) and are costless (i.e. $h_k^0(x) \equiv 0$). A detailed account for the SBP can be found in [2]. The "frozen" assumption is however not fulfilled for the multi-class production systems we are dealing with. Indeed, even unserved, the demands continue to accrue and hence the global state of the system is in permanent evolution (i.e. the net inventory of an item not currently pro-

duced, evolves with time). In this case, the scheduling of a flexible production naturally belongs to the class of the RBP. Contrary to the static case, it is known that for the RBP, any priority index policy yields only suboptimal, yet often efficient, scheduling policies. Several basic difficulties inherent to the RBP stimulate an ongoing research activity ([11] and [12]), in which basic questions such as:

- How to calculate the priority index, for specific production and demand stochastic processes ?

- How far from the optimal solution is a priority index policy when RBP are considered ?

are addressed. Due to their simplicity, the scheduling rules based on priority index policies are very appealing for applications at the shop-floor level. It is therefore important to study explicitly some tractable situations. This is the approach adopted in the present paper, where priority indices are calculated to schedule allocation problems of the RBP-type.

Our paper is organized as follows: in section II, we present the general theory of the continuous time, continuous states space, Restless Bandit together with the Whittle's relaxation principle. In section III, we perform explicit calculations of the priority indices for several underlying random dynamics, including the diffusion processes in section III-A and the Markov chain processes in section III-B. We derive new explicit expressions for the priority indices which complete former results obtained in [7] and [5]. In section IV, the results of section III are applied to the specific make-to-stock production context. Finally, section V is devoted to numerical illustrations.

II. The Whittle relaxation for the continuous time, continuous states, discounted "Restless Bandit" problem

First, we briefly present the basic formulation of the continuous time, continuous state space version of the multi-armed *Restless Bandit problem* (RBP), along the lines pioneered by P. Whittle [21].

Consider a collection of N projects (i.e. N dynamical systems):

$$X_k(t) \in \mathbb{R},\ \ k = 1, 2, ..., N.$$

At each instant $t \in \mathbb{R}^+$ **exactly $\mathbf{M < N}$ projects must be engaged** (i.e. must be in their active phase). If at time t, the project k is in state $X_k(t) = x_k$ and is engaged, then an *active running cost* $h_k^a(x_k)$ is incurred and the project evolves following an active transition probability. We suppose in the following that the $X_k(t)$ are stationary Markovian stochastic processes and we use the notation:

$$P_k\Big(X_k(t+dt) \ \Big|\ X_k(t), a\Big)$$

to describe the transition probability, where a indicates that the active action is selected. The running cost is discounted over time by a factor $e^{-\delta t}$. This means that the present value of one unit of tax, equals $e^{-\delta t}$ when received t units of time in the future.

The other $N - M$ projects remain disengaged (i.e. remain in their passive phases). They generate *passive running costs* $h_j^p(x_j)$ and evolve according to stationary Markovian transition probabilities:

$$P_j\Big(X_j(t+dt) \ \Big|\ X_j(t), p\Big),$$

where p indicates that the passive action is taken. The passive costs are discounted by the same factor $e^{-\delta t}$.

Projects are to be selected for operation according to an *scheduling policy* $\pi \in \mathcal{U}$ ($\mathcal{U}$ is the set of admissible policies). We write

$$\vec{X}(t) = (X_1(t), \ldots, X_N(t))$$

for the state of the system at time t. As noted in section I, the absence of switching penalties implies that the initial condition $\vec{u}_0 \in \{a, p\}^N$ is not necessary to characterize the evolution of $\vec{X}(t)$. Then the general problem, for a given initial condition

$$\vec{X}(0) = \vec{x}_0 = (x_1^0, \ldots, x_N^0),$$

is to compute, on a infinite time horizon, the optimal scheduling policy π^* which minimizes the total expected discounted cost

$$J^*(\vec{x}_0) = \min_{\pi \in \mathcal{U}} E^{\pi}_{\vec{x}_0}\left[\int_0^{\infty} \sum_{k=1}^{N} h_k^{\Pi_k(s)}(X_k(s)) e^{-\delta s} ds\right]. \quad (7)$$

The index $\Pi_k(t) \in \{a, p\}$ denotes the operating state (i.e. active, passive) of the project k at time t when the policy π is adopted. The operator $E^{\pi}_{\vec{x}_0}$ denotes the expectation, under policy π, conditional to $\vec{x}_0$. Moreover, the optimization of Eq.(7) has to be performed under the constraint that **exactly M projects are engaged at each time t.**

Let us define the action function:

$$I_k(t) = \begin{cases} 1 & \text{if the project } k \text{ is active at time t,} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\bar{I}_k(t) = 1 - I_k(t).$$

In terms of $I_k(t)$ and $\bar{I}_k(t)$, we can rewrite the RBP as:

$$\text{RBP} \equiv \begin{cases} J^*(\vec{x}_0) = \\ \quad \min_{\pi \in \mathcal{U}} E^{\pi}_{\vec{x}_0}\left[\int_0^{\infty} \sum_{k=1}^{N} h_k^a(X_k(t)) I_k(t) e^{-\delta t} dt + \right. \\ \qquad \left. \int_0^{\infty} \sum_{k=1}^{N} h_k^p(X_k(t)) \bar{I}_k(t) e^{-\delta t} dt\right] \\ \\ \text{subject to the constraint} \\ \quad \sum_{k=1}^{N} I_k(t) = M, \quad \forall t \geq 0. \end{cases}$$

The complexity of the RBP as exposed above has been shown in [13] to be PSPACE-hard. To solve this type of problems, one therefore relies, in general, on approximations. P. Whittle proposed in [21] an approximation scheme know as the *Whittle's relaxation* problem (WR), which consists in relaxing the requirement that **exactly M projects** must be active at each time t, to the weaker requirement that M **projects must be active on average**. Accordingly, the WR reads as:

$$WR \equiv \begin{cases} J^W(\vec{x}_0) = \\ \min_{\pi\in\mathcal{U}} E^{\pi}_{\vec{x}_0}\left[\int_0^{\infty}\sum_{k=1}^{N} h^a_k(X_k(t))I_k(t)e^{-\delta t}dt + \right. \\ \qquad\qquad \left. \int_0^{\infty}\sum_{k=1}^{N} h^p_k(X_k(t))\bar{I}_k(t)e^{-\delta t}dt\right] \\ \\ \text{subject to the constraint} \\ E^{\pi}_{\vec{x}_0}\left[\int_0^{\infty}\sum_{i=1}^{N} I_i(t)e^{-\delta t}dt\right] = \frac{M}{\delta}. \end{cases} \tag{8}$$

From the definition of the action function $I_k(t)$, we have:

$$E^{\pi}_{\vec{x}_0}\left[\int_0^{\infty}\sum_{k=1}^{N}\bar{I}_k(t)e^{-\delta t}dt\right] = \frac{N-M}{\delta}.$$

Along the lines pioneered by P. Whittle, we use the Lagrangian multiplier formalism to solve the problem (8). Accordingly, the Lagrange function $J^W(\vec{x}_0,\gamma)$ associated with Eq.(8) reads as:

$$J^W(\vec{x}_0,\gamma) = \sum_{k=1}^{N} J^k(x^0_k,\gamma) - (N-M)\frac{\gamma}{\delta}, \tag{9}$$

with

$$J^k(x^0_k,\gamma) = \min_{\pi\in\mathcal{U}E^{\pi}_{x^0_k}}\Big[\int_0^{\infty} h^a_k(X_k(t))I_k(t)e^{-\delta t}dt + \int_0^{\infty}(h^p_k(X_k(t)) + \gamma)\bar{I}_k(t)e^{-\delta t}dt\Big]. \tag{10}$$

Clearly the problem given in Eq.(8) is now decoupled into N single-project subproblems $J^k(x_k,\gamma)$ of the type given in Eq.(10). Following [12], we interpret the multiplier γ as playing the economic role of a constant *tax incurred when not producing*. Each single problem of the type arising in Eq.(10) is known as a γ-**penalty problem**.

The single armed RBP that results from the decoupling belongs to the class of stationary Markovian decision problems [15]. It has been proved that, for these problems, the optimal policy is stationary [15] (i.e. it only depend on the state x_k of project k). We can therefore define the set $\mathcal{O}_k(\gamma) \subseteq \mathbb{R}$ containing all the states $x_k \in \mathbb{R}$ for which it is optimal to take the active action in the γ-penalty problem (10). In terms of the set $\mathcal{O}_k(\gamma)$, the following structural property is essential:

Definition 2 (indexability): We say that problem (10) is indexable if the set $\mathcal{O}_k(\gamma)$ increases monotonically from the empty set to the full state space as the tax γ increases from $-\infty$ to $+\infty$.

Note that when the problem defined by Eq.(10) is indexable, it follows from Definition 2, that indices $\nu_k(x_k)$ exist, for each $x_k \in \mathbb{R}$. The values $\nu_k(x_k)$ correspond to the smallest values of γ for which $x_k \in \mathcal{O}_k(\gamma)$. The indices $\nu_k(x_k)$ can be used to characterize the optimal solution of the γ-penalty problems for each value of $\gamma \in \mathbb{R}$ as follows:

Take the active action in all states x_k for which $\gamma > \nu_k(x_k)$, and the passive action otherwise.

Hence, given the tax γ and being in the state x_k, $\nu_k(x_k)$ is the unique breakpoint where both the active and the passive actions are optimal. In other words, the index $\nu_k(x_k)$ is the smallest value of γ that would make the active action suboptimal in the state x_k.

From now on, we shall focus on indexable RBP's. Under this assumption, we shall propose a derivation of the indices $\nu_k(x_k)$, $k = 1,\ldots,N$, for the decoupled single-armed RBP and we will construct an associated heuristic scheduling rule.

A dynamic programming argument (see [22]), implies that the optimal cost function $J^k(x^0_k,\gamma)$ fulfills the property:

$$\inf_{\theta\in\{a,p\}}\Big[h^{\theta}_k(X_k(t_0)) - \delta J^k(x^0_k,\gamma) + \frac{\partial J^k(x^0_k,\gamma)}{\partial t} + L(\theta,t_0)J^k(x^0_k,\gamma)\Big] = 0, \tag{11}$$

where $L(\theta,t_0)$ is the infinitesimal generator of the controlled process (see [22] chapter 4, p.177).

For notational ease, we define $J^k_{\theta}(x^0_k,\gamma)$ to be the solution of:

$$\Big[h^{\theta}_k(X_k(t_0)) - \delta J^k(x^0_k,\gamma) + \frac{\partial J^k(x^0_k,\gamma)}{\partial t} + L(\theta,t_0)J^k(x^0_k,\gamma)\Big] = 0,$$

i.e. $J^k_{\theta}(x^0_k,\gamma)$ stand for the minimum discounted cost, when it is decided to take action θ for the project k, at time t_0.

As we assumed the indexability of the γ-penalized problem, the index value $\nu_k(x^0_k)$, for the project k in the state $X_k(0) = x^0_k$, will be the minimal value of γ such that:

$$J^k_a(x^0_k,\gamma) = J^k_p(x^0_k,\gamma) \;\Leftrightarrow\; J^k_a(x^0_k,\nu_k(x^0_k)) = J^k_p(x^0_k,\nu_k(x^0_k)). \tag{12}$$

By deriving the index value for each initial condition $x_k \in \mathbb{R}$, we get the index function of project k:

$$\begin{array}{rccc} \nu_k : & \mathbb{R} & \to & \mathbb{R} \\ & x_k & \mapsto & \nu_k(x_k) \end{array}$$

As a result, we set the **generalized heuristic scheduling rule** proposed by Whittle for the multi-armed Restless Bandit problem, as:

Definition (Whittle heuristic): Suppose that each project $k = 1, \ldots, N$, of an RBP is indexable, then the Whittle heuristic commands: *engage at each time t the M projects exhibiting the M smallest index values $\nu_k(x_k)$* where $x_k = X_k(t)$.

Note that when γ is fixed, the optimal discounted cost $J^k(x_k, \gamma)$ is equal to $J_a^k(x_k, \gamma)$ for each $x_k \in \mathcal{O}_k(\gamma)$ and is equal to $J_p^k(x_k, \gamma)$ elsewhere. In order to entirely define the optimal discounted cost $J^k(x_k, \gamma)$ in terms of $J_a^k(x_k, \gamma)$ and $J_p^k(x_k, \gamma)$, it remains to fit $J_a^k(x_k, \gamma)$ and $J_p^k(x_k, \gamma)$ on the active/passive boundary (i.e. the boundary of $\mathcal{O}_k(\gamma)$). In the following, we shall use a "smooth-fit" principle:

Definition (Smooth-fit principle): Suppose that x_k is on the active/passive boundary. Then the smooth-fit of $J_a^k(x_k, \gamma)$ and $J_p^k(x_k, \gamma)$ reads as:

$$\begin{aligned} & J_a^k(x_k, \gamma) = J_p^k(x_k, \gamma), \\ & \tfrac{d}{dx} J_a^k(x, \gamma)\Big|_{x=x_k} = \tfrac{d}{dx} J_p^k(x, \gamma)\Big|_{x=x_k}, \\ & \tfrac{d^2}{dx^2} J_a^k(x, \gamma)\Big|_{x=x_k} = \tfrac{d^2}{dx^2} J_p^k(x, \gamma)\Big|_{x=x_k}. \end{aligned} \tag{13}$$

Remark: When the evolution of the $X_k(t)$ are given by diffusion processes, the smooth-fit principle yields the optimum (see section 3.8 of [16]). This is not necessarily so for cases where non-diffusive processes occur. We will, in this paper, systematically use the smooth-fit principle to derive the indices and we will verify a posteriori, (appendix D), that this principle yields optimal results for the markov chains dynamics considered in section IV-A.

III. Explicitly solved examples

The explicit calculations of the priority index for an arbitrary underlying stochastic process, is generally an elaborate exercise. Let us now explicitly calculate this index for several simple types of dynamics. Some of the expressions derived in this section will later be used in the production engineering context.

For general processes $X_k(t)$ and arbitrary cost functions $h_k^\theta(x)$, $\theta \in \{a, p\}$, the indexability of the γ-penalty problem (10) is not guaranteed. Accordingly, we shall proceed in a first step by assuming that the processes $X_k(t)$ and the cost functions $h_k^\theta(x)$ possess the required properties to ensure the indexability. In a second step, we will verify that indexability indeed holds for the particular choices made.

As an introductive illustration, we develop in appendix A the derivation of the index for a deterministic RBP.

A. One-dimensional RBP with diffusion dynamics.

As only single-armed Bandits are considered in this section, we will omit the item index k. Consider now the situation where the project $X(t)$ is a diffusion process solving the stochastic differential equation:

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t), \tag{14}$$

with $dW(t)$ a White Gaussian noise process. The drift term $\mu(X(t))$ and the variance $\sigma(X(t))$ obey to:

$$\mu(X(t)) = \begin{cases} \mu_a > 0 & \text{if the active action is chosen,} \\ \mu_p < 0 & \text{if the passive action is chosen,} \end{cases}$$

respectively:

$$\sigma(X(t)) = \begin{cases} \sigma_a & \text{if the active action is chosen,} \\ \sigma_p & \text{if the passive action is chosen.} \end{cases}$$

Using the Ito formula, Eq.(11) can be written in the form (see for example [6]):

$$\begin{cases} \frac{1}{2}\sigma_a^2 \frac{d^2}{dx^2} J_a(x, \gamma) + \mu_a \frac{d}{dx} J_a(x, \gamma) - \delta J_a(x, \gamma) + h_a(x) = 0, \\ \frac{1}{2}\sigma_p^2 \frac{d^2}{dx^2} J_p(x, \gamma) + \mu_p \frac{d}{dx} J_p(x, \gamma) - \delta J_p(x, \gamma) + \\ \qquad\qquad h_p(x) + \gamma = 0. \end{cases} \tag{15}$$

Due to the linearity of Eq.(15), its general solution is:

$$J_a(x, \gamma) = C_a^+ e^{-w_a^+ x} + C_a^- e^{w_a^- x} + S_a(x, \gamma)$$

and

$$J_p(x, \gamma) = C_p^+ e^{-w_p^+ x} + C_p^- e^{w_p^- x} + S_p(x, \gamma),$$

with the notations:

$$w_\theta^+ = \frac{\sqrt{\mu_\theta^2 + 2\delta\sigma_\theta^2} + \mu_\theta}{\sigma_\theta^2} > 0,$$

$$w_\theta^- = \frac{\sqrt{\mu_\theta^2 + 2\delta\sigma_\theta^2} - \mu_\theta}{\sigma_\theta^2} > 0$$

and $S_\theta(x)$ are the particular solutions of Eq.(15) corresponding to engage [respectively disengage] the project X_k forever [6]. We obtain:

$$\begin{aligned} & S_\theta(x, \gamma) = E_{(x,\gamma)} \int_0^\infty e^{-\delta t} h_\theta(x(t))\, dt = \\ & = \tfrac{2}{\sigma_\theta^2 (w_\theta^+ + w_\theta^-)} \Bigg[e^{-w_\theta^+ x} \int_{-\infty}^{x} h_\theta(y)\, e^{w_\theta^+ y} dy + \\ & \qquad\qquad e^{w_\theta^- x} \int_x^\infty (h_\theta(y) + \gamma I_{\theta=p})\, e^{-w_\theta^- y} dy \Bigg], \end{aligned}$$

with

$$I_{\theta=p} = \begin{cases} 1 & \text{if } \theta = \text{p}, \\ 0 & \text{if } \theta = \text{a} \end{cases} \tag{16}$$

and the C_θ^+ and C_θ^- are integration constants.

From the fact that $\mu_a > 0$, $\mu_p \leq 0$, we can check that $w_a^- > 0$ and that $w_\theta^+ < 0$. Moreover, if $x_0 = +\infty$ the optimal discounted cost is reached by remaining passive forever. Similarly, when $x_0 = -\infty$ the optimal discounted cost is reached by remaining active forever. This implies:

$$\lim_{x\to\infty} J_p(x,\gamma) = S_p(x,\gamma) \Rightarrow C_p^+ = 0$$

and

$$\lim_{x\to-\infty} J_a(x,\gamma) = S_a(x,\gamma) \Rightarrow C_a^- = 0.$$

Using the smooth-fit principle described in Eq.(13), we can write, after straightforward but lengthy algebra, the index $\nu(x)$ in the compact form:

$$\nu(x) = \frac{-2}{w_a^+ w_p^- \sigma_a^2 \sigma_p^2}\Big[\mathcal{I}_2^a \sigma_p^2 \frac{(w_a^- - w_p^-)}{w_a^-} + \mathcal{I}_1^p \sigma_a^2 \frac{(w_a^+ - w_p^+)}{w_p^+} + h_p(x)\sigma_a^2 - h_a(x)\sigma_p^2\Big]. \quad (17)$$

where:

$$\mathcal{I}_1^\theta = \int_0^\infty h_\theta(x - \frac{y}{w_\theta^+})\, e^{-y} dy$$

and

$$\mathcal{I}_2^\theta = \int_0^\infty h_\theta(x + \frac{y}{w_\theta^-})\, e^{-y} dy.$$

Remark: When $h_p(x) = 0$, $\mu_p = 0$ and for the limit $\sigma_p \to 0$, the RBP converges to the static Bandit problem for which the passive project remains "frozen" and does not incur cost. In this limit, the index given by Eq.(17) converges directly to the result derived by I. Karatzas [6], for a static single-armed diffusive Bandit, which reads as:

$$\nu_{\text{class}}(x) = \frac{1}{\delta}\int_0^\infty h_a(x + \frac{y}{w_a^-}) e^{-y}\, dy.$$

B. One-dimensional RBP with continuous time Markov chains

Let us now consider the case where the processes $X_k(t)$ is a birth and death process (i.e. a continuous-time and discrete state Markov chain for which $X_k(t) \in \mathbb{Z}^N$). Assume that the holding time between the transitions from the state x to $x+1$ is exponentially distributed with parameter μ_a when the active action is chosen and μ_p for the passive action. Conversely, for the transitions from the state x to $x-1$, the parameter is λ_a for the active action and λ_p for the passive action. We impose that $\mu_a > \lambda_a$ and that $\mu_p < \lambda_p$. In other words, the time average of the process $X_k(t)$ increases when active and decreases when passive. The associated running costs rates are $h_\theta(x)$, $\theta \in \{a, p\}$.

Lemma 1: Under the above assumptions and following the optimal policy π^*, Eq.(11) takes the form:

$$\begin{cases} \delta J_a(x,\gamma) = h_a(x) + \lambda_a J_a(x-1,\gamma) + \mu_a J_a(x+1,\gamma) + \\ \qquad\qquad -(\lambda_a + \mu_a) J_a(x,\gamma) \\ \qquad \text{if } \pi^*(X(0) = x) = a. \\ \hline \delta J_p(x,\gamma) = h_p(x) + \lambda_p J_p(x-1,\gamma) + \mu_p J_p(x+1,\gamma) + \\ \qquad\qquad (\lambda_p + \mu_p) J_p(x,\gamma) + \gamma \\ \qquad \text{if } \pi^*(X(0) = x) = p. \end{cases} \quad (18)$$

Proof Assume that $\pi^*(X(0) = x) = a$, then the first order time expansion of Eq.(11) reads as:

$$J_a(x,\gamma) = \xi h_a(x) + (1 - \delta\xi)\Big[$$

$$\xi\lambda_a J_a(x-1,\gamma) + \xi\mu_a J_a(x+1,\gamma) + (1 - \xi\lambda_a - \xi\mu_a) J_a(x,\gamma)\Big].$$

Neglecting the terms of order $\mathcal{O}(\xi^2)$ in the above expansion yields the required result. A similar expression can also be directly derived when $\pi^*(X(0) = x) = p$.

□

Due to the linearity of Eq.(18), we have:

$$J_a(x,\gamma) = C_{a^+}(w_a^+)^x + C_{a^-}(w_a^-)^x + S_a(x,\gamma),$$
$$J_p(x,\gamma) = C_{p^+}(w_p^+)^x + C_{p^-}(w_p^-)^x + S_p(x,\gamma),$$

with C_{θ^+} and C_{θ^-}, $\theta \in \{a, p\}$, being integration constants,

$$w_\theta^+ = \frac{(\delta + \lambda_\theta + \mu_\theta) + \sqrt{(\delta + \lambda_\theta + \mu_\theta)^2 - 4\lambda_\theta\mu_\theta}}{2\mu_\theta}$$

$$w_\theta^- = \frac{(\delta + \lambda_\theta + \mu_\theta) - \sqrt{(\delta + \lambda_\theta + \mu_\theta)^2 - 4\lambda_\theta\mu_\theta}}{2\mu_\theta}$$

and $S_a(x,\gamma)$, $S_p(x,\gamma)$ being the particular solutions which correspond to remain active [respectively passive] forever. We derive $S_a(x,\gamma)$ in the appendix B and obtain:

$$S_a(x,\gamma) = \frac{1}{\mu_a(w_a^+ - w_a^-)}\Big\{ h_a(x) + (w_a^-)^x \sum_{k=-\infty}^{x-1} h_a(k)(w_a^-)^{-k} + (w_a^+)^x \sum_{k=x+1}^{\infty} h_a(k)(w_a^+)^{-k} \Big\}.$$

A calculation along the same lines yields:

$$S_p(x,\gamma) = \frac{1}{\mu_p(w_p^+ - w_p^-)}\Big\{ (h_p(x) + \gamma) + (w_p^-)^x \sum_{k=-\infty}^{x-1} (h_p(k) + \gamma)(w_p^-)^{-k} + (w_p^+)^x \sum_{k=x+1}^{\infty} (h_p(k) + \gamma)(w_p^+)^{-k} \Big\}.$$

For consistency, it is required that the total cost incurred, when $x \to -\infty$, equals the cost incurred when engaging the server forever. Respectively the total cost incurred, when

$x \to +\infty$ equals the cost incurred when letting the server be idle forever. Using the property that

$$0 \leq w_\theta^- \leq 1 \leq w_\theta^+,$$

which is straightforward to establish, these asymptotic behaviours imply:

$$\lim_{x\to\infty} J_p(x,\gamma) = S_p(x,\gamma) \Rightarrow C_{p^+} = 0$$

and

$$\lim_{x\to-\infty} J_a(x,\gamma) = S_a(x,\gamma) \Rightarrow C_{a^-} = 0.$$

Again, the index $\nu(x)$ is derived by fit both functions $J_a(x,\gamma)$ and $J_p(x,\gamma)$. We use here a discrete version of the smooth-fit principle, described in (13) by writing:

$$\tfrac{d}{dx} J_\theta(x_0,\gamma) := J_\theta(x_0,\gamma) - J_\theta(x_0-1,\gamma)$$

$$\tfrac{d^2}{dx^2} J_\theta(x_0,\gamma) := J_\theta(x_0+1,\gamma) - 2J_\theta(x_0,\gamma) + J_\theta(x_0-1,\gamma).$$

Explicit expressions for $\nu(x)$ will be given below for a specific form of the cost function $h_\theta(x)$.

IV. Dynamic scheduling of multiclass make-to-stock production.

Let us now apply the general framework of section II to the production context. The make-to-stock problem can be naturally formulated as a multi-armed RBP with $N+1$ projects as follows:

- Identify the positions of the N net inventories $X_k(t)$, $k = 1, 2, \ldots N$, with the first N projects of the RBP.

- Take the active and the passive cost functions identical for each item (i.e. $h_k^a(x) = h_k^p(x) := h_k(x)$).

- Add an extra project $X_{N+1}(t)$, called the *idling* project, with "frozen" dynamics given by: $X_{N+1}(t) \equiv x_c,\ t \in \mathbb{R}^+$ Engaging the idling project represents the decision to be idle.

- Impose, the idling project to incur no cost (i.e. $h_{N+1}(x) \equiv 0$).

- Assume the cost vector function to have a piecewise linear form, namely:

$$h_k(x) = A_k x^+ + B_k x^-, \quad k = 1, 2, \ldots N, \tag{19}$$

where $x^+ = \max(x, 0)$; $x^- = \max(-x, 0)$ and $A_k,\ B_k \geq 0$.

By construction, the index of the idling project is $\nu_{N+1}(x_c) \equiv 0$. Therefore, following the Whittle heuristic, the machine is left idle (i.e. does not produce any item) when all indices are positive (i.e. $\nu_k(x_k) > 0$, $k = 1, \ldots, N$). Moreover, when all indices are strictly increasing, the positions $d_k^* \in \mathbb{R}$ such that $\nu_k(d_k^*) = 0$ correspond to the hedging stocks for the type k items, $k = 1, \ldots, N$.

A. Markovian queue dynamics

Assume that the net inventory process Eq.(4) is described by a continuous-time discrete state Markov chain with parameters λ_k and μ_k. The resulting make-to-stock problem is identical to the one studied in [4]. Note also that the contribution [17] discusses a similar problem, but the optimization is done under the average cost criterion.

Following [17], we apply the standard uniformization argument given in [9] and the dynamic programming equation, Eq.(11) becomes:

$$J^*(\vec{X}) = \tfrac{1}{\Lambda+\delta}\left[h(\vec{X}) + \sum_{k=1}^{N} \lambda_k J^*(\vec{X} - e_k) + \mu J^*(\vec{X}) + \min\left\{0, \min_k\left(\mu_k \Delta_k J^*(\vec{X})\right)\right\}\right], \tag{20}$$

with $\Delta_k J^*(\vec{X}) = J^*(\vec{X} + e_k) - J^*(\vec{X})$ and e_k is the unit vector with the k-th component equal to unity,

$$\mu = \max_k\{\mu_k\} \quad \text{and} \quad \Lambda = \mu + \sum_{i=1}^{N} \lambda_i.$$

As stated in [17], the form of Eq.(20) suggests that the optimal policy can be described with switching curves and hedging stocks. We shall see that the priority index policies lead to a similar structure and hence the Whittle relaxation method is suitable to discuss the production problems.

Note first that the policy resulting from the Whittle relaxation is intrinsically preemptive, contrary to the make-to-stock problem (i.e. the machine must complete its currently engaged item before starting a new one). Hence, strictly speaking, a direct use of the multi-armed RBP is not possible. To overcome this difficulty and to nevertheless use the RBP in the production context, we will suitably renormalize the service time to approximately transform our original problem into a preemptive one. The renormalization process is done by imposing:

$$\mu_k \mapsto \tilde{\mu}_k = \rho_k \mu_k + (\rho - \rho_k)\frac{1}{\frac{1}{\mu_k} + T} + (1-\rho)\mu_k, \tag{21}$$

$$\rho_k = \frac{\lambda_k}{\mu_k}, \quad \rho = \sum_{k=1}^{N} \rho_k.$$

In writing Eq.(21), we have used the fact that when the production priority (derived from the Whittle heuristic) is to engage the type k production, three alternatives may occur, namely:

a) The server is already engaged on the type k products and the average service time is $\frac{1}{\mu_k}$.

b) The server is engaged on a production type $j \neq k$ and the average service time is $\frac{1}{\mu_k} + T_j$ where T_j is the average time needed to finish the production of the type j item.

We denote by T the average of the T_j.

c) The server is idle and, as we consider only problems without switching time, the average service time is $\frac{1}{\mu_k}$.

For infinite time horizon, $\tilde{\mu}_k$ is therefore a weighted average taking into account the relative contribution of a), b) and c). The respective weights are determined as follows:

i) The average sojourn time in situation a) is proportional to the partial traffic $\rho_k = \frac{\lambda_k}{\mu_k}$.

ii) The average sojourn time in situation b) is proportional to the global traffic, minus the partial traffic of the k-type items: $(\rho - \rho_k)$.

iii) The average sojourn time in situation c) is proportional to the percentage of idle time: $(1 - \rho)$.

Note that the value of T is bounded:

$$0 \leq T \leq \max_{j \in \{1,\dots,N\}} \left(\frac{1}{\mu_j}\right) = \frac{1}{\mu}$$

From now on and for simplicity, we shall keep the notation μ_k for $\tilde{\mu}_k$.

As we have seen, the application of the Whittle relaxation enables us to focus on a single item problem, say item k. Then the γ-penalty given by problem (10) reads here as:

$$J^k(x,\gamma) = \frac{1}{\lambda_k+\mu_k+\delta}\left[h_k(x) + \lambda_k J^k(x-1) + \mu_k J^k(x) + \min\{\gamma, \mu_k \Delta J^k(x)\}\right]. \quad (22)$$

From now on, we can again suppress the index k as the calculation involve only a single item. To make headway, we assume that the indexability property holds (we will verify, a posteriori, that this is indeed the case). Following the derivation of section III-B, Eq.(11) reads as:

$$\begin{cases} \delta J_a(x,\gamma) = h(x) + \lambda J_a(x-1,\gamma) + \mu J_a(x+1,\gamma) - (\lambda+\mu) J_a(x,\gamma) \\ \\ \delta J_p(x,\gamma) = h(x) + \lambda J_p(x-1,\gamma) - \lambda J_p(x,\gamma) + \gamma \end{cases} \quad (23)$$

which is a special case of the system given by Eq.(18) with $\mu_p = 0$.

From section III-B, we have:

$$\begin{aligned} J_a(x,\gamma) &= C_a^+ (w_+)^x + C_a^- (w_-)^x + S_a(x,\gamma) \\ J_p(x,\gamma) &= C_p (w_0)^x + S_p(x,\gamma), \end{aligned} \quad (24)$$

with C_a^+, C_a^-, C_p being integration constants,

$$\begin{aligned} w_+ &= \frac{(\delta+\lambda+\mu)+\sqrt{(\delta+\lambda+\mu)^2-4\lambda\mu}}{2\mu}, \\ w_- &= \frac{(\delta+\lambda+\mu)-\sqrt{(\delta+\lambda+\mu)^2-4\lambda\mu}}{2\mu}, \\ w_0 &= \frac{\lambda}{\lambda+\delta} \end{aligned} \quad (25)$$

and $S_a(x,\gamma)$, $S_p(x,\gamma)$ are the relevant particular solutions. Explicit calculations are given in Appendix B, where we find:

$$S_a(x,\gamma) = \frac{1}{\mu(w_+ - w_-)}\left\{ h(x) + (w_-)^x \sum_{k=-\infty}^{x-1} h(k)(w_-)^{-k} + (w_+)^x \sum_{k=x+1}^{\infty} h(k)(w_+)^{-k} \right\},$$

$$S_p(x,\gamma) = (w_0)^{x+1} \sum_{k=-\infty}^{x} \frac{(h(k)+\gamma)(w_0)^{-k}}{\lambda}.$$

Using the fact that $0 \leq w_- \leq w_0 \leq 1 \leq w_+$ and the asymptotic behaviour of the solutions, we have:

$$\lim_{x \to -\infty} J_a(x,\gamma) = S_a(x,\gamma) \Rightarrow C_a^- = 0.$$

As before, the index $\nu(x)$ is calculated with the smooth-fit principle expressed in Eq.(13). For the piecewise linear cost function $h(x)$ given by Eq.(19), the corresponding explicit expressions, derived in Appendix C, read:

$$\nu(x) = \begin{cases} \frac{2A\mu - \mu(A+B)(w_+ + w_-)(w_-)^x + \mu(A+B)(w_+ - w_-)(w_-)^x}{2\delta(\delta+1)}, \\ \qquad \text{if } x \geq 0 \\ -\frac{B\mu}{\delta(\delta+1)} \\ \qquad \text{if } x < 0. \end{cases} \quad (26)$$

Remarks:

i) Observe in Eq.(26) that, with the choice of $h(x)$ given by Eq.(19), the indexability property does indeed hold as $\nu(x)$ is monotonically increasing.

ii) When $x < 0$, the Restless priority index directly reduces to the well known $B\mu$ policy, which is optimal in this case [4].

iii) As it is emphasized in [17], the index Eq.(26) does not exist in the limit $\delta \to 0$ (i.e. $\nu(x) = -\infty$). In the contrary, for $\delta > 0$, the scheduling policy does not need to serve a fixed time-average number of classes. Accordingly, for the relaxed version of the RBP, priority indices exist when $\delta > 0$.

iv) Note that the random time look ahead policy (RTLA) derived in [4] is also a priority index rules. There is however no direct correspondence between the RTLA and the RBP indices as, by the remark *ii*), the underlying optimization problem are structurally different for $\delta = 0$ and $\delta > 0$.

v) The asymptotic behaviour:

$$\nu(x) = \begin{cases} \frac{A\mu}{\delta(\delta+1)} & \text{when } x \to +\infty, \\ \\ \frac{-B\mu}{\delta(\delta+1)} & \text{when } x < 0, \end{cases} \quad (27)$$

exhibits the structure of the well known $B\mu/A\mu$ policy (see section V-A below for a more detailed discussion devoted to this policy).

vi) When the indexability property is fulfilled, the make-to-stock single item production problem is optimally solved by the Whittle relaxation. This can be proved as follows: Consider the two-armed RBP formulation of the single item production problem and remember that the idling project does not incur cost. This two-armed RBP is equivalent to the γ-penalized problem (22) with $\gamma = 0$. The optimal policy for this problem is to engage the production in all states $x \in \mathbb{Z}$ for which $\nu(x) < \gamma = 0$. As the idle index is $\nu_{N+1}(x) \equiv 0$, the optimal policy for the $\gamma = 0$-penalized problem is therefore equivalent to engage the project with the smallest priority index. Hence the Whittle relaxation indeed solves optimally the single item production problem.

vii) For a multi-item production machine, the priority index policy solves the scheduling production problem only sub-optimally.

viii) The index $\nu(x)$ given by Eq.(26) is monotonically increasing. Hence the Whittle heuristic for the single item problem coincides with a hedging stock policy (i.e. *produce only when the stock level is below the hedging stock*), with a hedging level d^* defined as:

$$\nu(d^*) = 0. \tag{28}$$

The hedging stock policy is known to be optimal for a single-item production (this is consistent with the remarks i) and vi) above). Solving Eq.(28) with Eq.(26), we find:

$$d^* = \left\lceil \ln\left(\frac{A}{A+B}\right)\frac{1}{\ln(w_-)} - 1 \right\rceil. \tag{29}$$

ix) Despite to the fact that the smooth-fit principle used to derive Eq.(28) was not proven to yield the optimal solution, it here does so. Indeed, the optimal hedging level d^* can be derived by using the approach as explained in [3]. This alternative calculation, which is performed in Appendix D yields:

$$d^* = \left\lfloor \ln\left(\frac{A}{A+B}\right)\frac{1}{\ln(w_-)} \right\rfloor. \tag{30}$$

Clearly both hedging stocks Eqs.(29) and (30) are identical.

Asymptotic regimes

• In the limit $\delta \to 0$, we have that $w_- \to \rho := \frac{\lambda}{\mu}$ and therefore:

$$\lim_{\delta\to 0} d^* = \left\lfloor \frac{1}{\ln(\rho)} \ln\left(\frac{A}{A+B}\right) \right\rfloor.$$

This is the optimal hedging point for the single-item problem under the average cost criterion, derived in [17].

• When $\delta \to 0$ and $\rho \to 1$ (heavy traffic limit), the optimal hedging stock tends to infinity (i.e. $d^* \to \infty$). Note that, such a behaviour does also follow from Eq.(3.7) of [8], when $c \to 0$ and $\gamma \to 0$.

B. Diffusive dynamics

Let us finally consider the case where we model the demand respectively the production by diffusive processes following stochastic differential equations:

$$dP_k(u_k(t),t) = u_k(t)\left[\mathcal{U}_k dt + \sigma_{k,P} dW_{k,P}(t)\right],\ k = 1,2,\dots N \tag{31}$$

respectively

$$dD_k(t) = \mathcal{V}_k dt + \sigma_{k,D} dW_{k,D}(t),\ k = 1,2,\dots N, \tag{32}$$

where $dW_{k,P}(t)$ and $dW_{k,D}(t)$ are independent White Gaussian Noise processes (WGN), $\mathcal{U}_k$ and $\mathcal{V}_k$ are the drifts and $\sigma_{k,P}$ and $\sigma_{k,D}$ are the variances of the diffusion processes.

Using Eq.(31) and Eq.(32), it is straightforward to write the time evolution of the net inventory $X_k(t)$ given by Eq.(4) in the form:

$$dX_k(t) = (\mathcal{U}_k u_k(t) - \mathcal{V}_k)\, dt + \sigma_k(u_k(t)) dW_k(t), \tag{33}$$

where $dW_k(t)$ are standard independent WGN's for $k = 1,2\dots N$, and the controlled variances $\sigma_k(u_k(t))$ reads:

$$\sigma_k^2(u(t)) = (\sigma_{k,D})^2 + (u(t)\sigma_{k,P})^2,\ k = 1,2\dots N. \tag{34}$$

Assume again that the running cost is piecewise linear, as given in Eq.(19). Then for the Whittle relaxation problem, the value of the index $\nu_k(x_k)$ follows from Eq.(17), provided that the following relations hold:

$$\mu_{a,k} = (\mathcal{U}_k - \mathcal{V}_k) > 0,$$

$$\mu_{p,k} = -\mathcal{V}_k < 0,$$

$$\sigma_{a,k}^2 = \sigma_k^2(1) = \sigma_{k,D}^2 + \sigma_{k,P}^2,$$

$$\sigma_{p,k}^2 = \sigma_k^2(0) = \sigma_{k,D}^2.$$

To further simplify the analysis, we will assume that only the demand process $\vec{D}(t)$ fluctuates. Hence, we take $\sigma_{a,k} = \sigma_{p,k}$. In this case, we can establish:

Lemma 2: With the above assumptions, the problem is indexable for every positive convex function $h_k(x)$ satisfying:

$$\int_0^\infty e^{-\delta t} h_k(\vec{X}(t)) dt < \infty.$$

Proof: With the above assumptions, the index Eq.(17) simplifies and reads as (we drop the index k):

$$\nu(x) = \mathcal{K} \int_0^\infty e^{-y} \left[h\left(x + \frac{y}{w_a^-}\right) - h\left(x - \frac{y}{w_p^+}\right)\right] dy,$$

with

$$\mathcal{K} = \frac{-2}{w_a^+ w_p^- \sigma^2}\left(1 - \frac{2\delta}{\sigma^2 w_a^- w_p^+}\right) > 0.$$

Hence:

$$\frac{d}{dx}\nu(x) = \mathcal{K}\int_0^\infty e^{-y}\left[\frac{d}{dx}h\left(x + \frac{y}{w_a^-}\right) - \frac{d}{dx}h\left(x - \frac{y}{w_p^+}\right)\right]dy.$$

This last expression is positive as all terms in the integral are positive. Then the index $\nu_k(x)$ is monotonically increasing and hence the problem is indexable.

□

Lemma 3: When $B > A$, the hedging stock level d^* is given by

$$d^* = \max\left\{0, \frac{\sigma^2}{\mu_p + \sqrt{\mu_p^2 + 2\delta\sigma^2}} \ln\left[\frac{(A+B)}{A}\left\{\frac{\mu_a - \mu_p - \sqrt{\mu_a^2 + 2\delta\sigma^2} + \sqrt{\mu_p^2 + 2\delta\sigma^2}}{2(\mu_a - \mu_p)}\right\}\right]\right\}. \quad (35)$$

Proof: For the cost function Eq.(19), the index can be explicitly written as:

if x is positive:

$$\nu(x) = \begin{cases} \frac{1}{2\delta^2}\Big(2A(\mu_a - \mu_p) + \\ \quad e^{-\frac{\left(\mu_p + \sqrt{\mu_p^2 + 2\delta\sigma^2}\right)x}{\sigma^2}}\Big[-(A+B)(\mu_a - \mu_p) + \\ (A+B)\sqrt{\mu_a^2 + 2\delta\sigma^2} - (A+B)\sqrt{\mu_p^2 + 2\delta\sigma^2}\Big]\Big) \end{cases} \quad (36)$$

and if x is strictly negative:

$$\nu(x) = \begin{cases} \frac{1}{2\delta^2}\Big(-2B(\mu_a - \mu_p) + \\ \quad e^{\frac{\left(\sqrt{\mu_a^2 + 2\delta\sigma^2} - \mu_a\right)x}{\sigma^2}}\Big[(A+B)(\mu_a - \mu_p) + \\ (A+B)\sqrt{\mu_a^2 + 2\delta\sigma^2} - (A+B)\sqrt{\mu_p^2 + 2\delta\sigma^2}\Big]\Big) \end{cases} \quad (37)$$

Now when $B > A$, solving $\nu(x) = 0$, with $\nu(x)$ given by Eq.(36), gives the required form.

□

Remarks:

i) When $B = 0$ and $A > 0$, the logarithm of the expression for d^* given in Eq.(35) is smaller than unity and we consistently conclude that the optimal hedging is located at 0 (i.e."just-in-time" production-rule).

ii) The expression form of the hedging point d^* given by Eq.(35) exhibits the same structural form as the one derived by E.V. Krichagina, [8]. Nevertheless, both expressions are not directly comparable. Indeed, the control rules considered in [8] are of "regulator" types while the class of controls considered here are of "bang bang" types. Accordingly, the local time process on the hedging level are different. This obviously implies different values of d^*.

iii) As in section IV-A, if $B \neq 0$ and for a vanishing discounting factor $\delta \to 0$, the index $\nu(x)$ does not exist (i.e. tends to $-\infty$).

iv) The hedging stock given by Eq.(36), in the limit $\delta \to 0$, is not directly comparable with the result given by L.M. Wein (1992). Indeed, L. Wein considers the fluid approximation of the throughput delivered by a failure prone machine. In this limit, the dynamics converges to a diffusion process with a reflecting boundary on the hedging point. Although we do also have a diffusive dynamics, the behaviour on the hedging level is not purely reflecting. This is due to the fact that the drift of our process is dynamically controlled. We hence deal with a "bang-bang" regulated diffusive process and the resulting local time process on the barrier differs from a standard reflecting boundary. Note however, that in the limit $\rho \to 1$ and for $\delta = 0$ both hedging levels tend to infinity (see [19] page 731).

v) Observe from Eqs.(36) and (37) that we have the asymptotic behaviours:

$$\nu(x) = \begin{cases} \frac{A}{\delta^2}(\mu_a - \mu_p) = \frac{A}{\delta^2}\mathcal{U} & \text{when } x \to \infty \\ \frac{-B}{\delta^2}(\mu_a - \mu_p) = \frac{-B}{\delta^2}\mathcal{U} & \text{when } x \to -\infty, \end{cases} \quad (38)$$

which again corresponds to a $B\mathcal{U}/A\mathcal{U}$ type scheduling policy.

V. Numerical experiments

In this section the Restless Bandit heuristic will be compared with the optimal policy calculated by A.Y. Ha [4] for the two items problem. In a second group of experiments, we study the case where the machine can produce more than two different items. The discussion will be based on a comparison between the RBP and other classical heuristic policies which we shall briefly recall. In our simulations, the production rates are all equal (i.e. $\mu_k = \mu,\ k = 1, \ldots, N$), the discounting factor is $\delta = 0.01$, and the net inventory is empty initially.

A. Review of some priority rules for make-to-stock productions

The three heuristics that we have studied in our numerical experimentations are:

1) The **static $\mathbf{h_k\mu/b_k\mu}$ rule** (where b_k is the cost rate for backorder type k products, h_k the storage cost rate for type k products and μ_k is the production rate of type k product) which is as follows:

a) **If demands are backordered:** *Produce the item with the largest $b_k\mu_k$ among all products for which there exists backordered demands.*

b) **If no demands are backordered:** *Produce the item with the smallest $h_k\mu_k$ among all products for witch the inventory levels are under their hedging stock d_k^*.*
This heuristic is fully myopic as it directly minimizes the instantaneous cost $h(x)$.

2) The **switching rule** which is obtained by modifying the static $h_k\mu/b_k\mu$ rule as follows:
a) **If demands are backordered:** *Produce the item with the largest $b_k\mu_k$ among all products for which there exists backordered demands (similar to the static $\mathbf{h_k\mu/b_k\mu}$ rule).*
b) **If no demand is backordered:** *Produce the item with the largest $b_k\mu_k(1-x_k/d_k^*)$ value if positive or let the server be idle.*
The quantity $(1-x_k/d_k^*)$ can be interpreted as the proportion of unfilled stock. The larger it is, the more probable a product will be backordered. This heuristic implies the existence of a linear switching curve in the positive quadrant of the state space that ends at the hedging point

$$\vec{d}^* = (d_1^*, \ldots, d_N^*).$$

3) The priority index policy based on the RBP indices given by Eq.(26). For this heuristic, we apply the renormalization procedure of μ given by Eq.(21), with $T = \frac{1}{\mu}$. Unlike the first two policies, this third one is not myopic, as the priority indices fully take into account the infinite time horizon.

B. Numerical results

In our experimentation, we have measured the total average discounted cost over 4000 simulation runs. The horizon H is chosen to be large enough to guarantee that the results become invariant on H (the presence of the discounting factor makes this possible). For the $h\mu/b\mu$ policy and for the switching rule, we have chosen the hedging stocks which minimize the total discounted cost. To find them for a machine able to produce N types of items, we have used of a N-dimensional search around the optimal hedging stock, known for a single item problem.

Experiment 1):

The number of item types is: $N = 2$
The demand rates are: $\lambda_1 = 0.4$; $\lambda_2 = 0.5$
The production rate is: $\mu = 1$
The costs are: $B_1 = 30$; $B_2 = 40$; $A_1 = 1$; $A_2 = 1$

For $\delta = 0.01$, the optimal policy has been derived numerically in [4]. It is given by a switching curve and a hedging point $\vec{d}^*$ as follows: When $x < 0$, The optimal policy commands to engage the item having the largest $B_k\mu_k$. When $x \geq 0$, the switching curve is almost equal to the strait line $y = x + 1$ and ends at the hedging levels $\vec{d}^* = (9, 11)$ (i.e. the hedging stock is $d_1^* = 9$ and $d_2^* = 11$).

	Hedging	Cost	Percentage more
Ha	$(9, 11)$	7401	optimal
Restless	(5,8)	7437	0.5% $\geq$ optimal
Switching	$(4, 7)$	7639	3.2% $\geq$ optimal
$h\mu/b\mu$	$(1, 8)$	7838	5.9% $\geq$ optimal

Experiment 2):

The number of item types is: $N = 3$
The demand rate are: $\lambda_k = 0.3$; $k \in \{1, \ldots, 3\}$
The production rate is $\mu = 1$
The cost are $B_1 = 80$; $B_2 = 90$; $B_3 = 100$; $A_1 = 3$; $A_2 = 2$; $A_3 = 3$

	Hedging	Cost	Percentage more
Restless	(4, 4 ,4)	19706	Best
Switching	$(4, 5, 4)$	20763	5.3% $\geq$ Restless
$h\mu/b\mu$	$(0, 1, 8)$	24672	25.2% $\geq$ Restless

Experiment 3):

The number of item types is: $N = 4$
The demand rate are: $1/\lambda_1 = 4$; $1/\lambda_2 = 4.1$; $1/\lambda_3 = 4.2$; $1/\lambda_4 = 4.3$
The production rate is $\mu = 1$
The cost are $B_1 = 40$; $B_2 = 30$; $B_3 = 20$; $B_4 = 10$; $A_k = 1$, $k \in \{1, \ldots, 4\}$

	Hedging	Cost	Percentage more
Restless	(2,3,3,3)	7745	Best
Switching	$(2, 3, 3, 4)$	8366	10.4% $\geq$ Restless
$h\mu/b\mu$	$(1, 1, 3, 3)$	8983	18.6% $\geq$ Restless

Remark: As remarked in [4], we do also find that, under the $h\mu/b\mu$ policy, the optimal discounted cost is reached for a strongly uneven distribution of the hedging stock levels.

VI. Conclusions

Using the relaxed version of the "Restless Bandit" problem, we are able to calculate explicitly the generalized Gittins' priority indices for several underlying stochastic process governing the arms dynamic. These explicit expressions are then used in the context of production manufacturing to discuss the dynamic scheduling of jobs in a flexible shop floor. A direct comparison with the optimal policy, known for the two products case, shows that Restless priority indices yield a scheduling policy which is very closed to the optimal one. In particular if backorder exists, the priority index rule reduces to the scheduling policy which produce the item with the largest $b\mu$ over all backordered items. This scheduling policy is known to be optimal in this case. In all simulations experiments performed, we observe that the Restless priority index rule is always better or at least equivalent than previously studied scheduling rules. In any case, the Restless priority index policy per-

forms much better than any purely myopic allocation rules.

Acknowledgments

We are grateful to the referees for stimulating comments which enable to strongly improve the presentation.

Appendix

A. One-dimensional deterministic project

We derive here the index for the discounted version of the deterministic problem presented in section 5 of [21]. Consider a continuous time project $X(t) \in \mathbb{R}$ satisfying

$$\frac{d}{dt}X(t) = f_{\theta(X(t))}(X(t)), \tag{39}$$

where $f_{\theta(X(t))}(X(t))$ equals $f_a(X(t))$, when the project is in state $X(t)$ and is in its active phase, and equals to $f_p(X(t))$ when the project is in the state $X(t)$ and is in its passive phase. Let us also introduce the two instantaneous reward rates $h_\theta(x)$, $\theta \in \{a, p\}$.

To derive the index $\nu(x)$ of this problem, we first solve the corresponding deterministic γ-penalty problem $J^k(x_k, \gamma)$:

Lemma 4: Following the optimal policy π^* for the γ-penalty problem, Eq.(11) read as:

$$\begin{cases} f_a(x)\frac{d}{dx}J_a(x,\gamma) - \delta J_a(x,\gamma) + h_a(x) = 0 \\ \qquad \text{if } \pi^*(X(0)=x) = a, \\ \hline f_p(x)\frac{d}{dx}J_p(x,\gamma) - \delta J_p(x,\gamma) + h_p(x) + \gamma = 0. \\ \qquad \text{if } \pi^*(X(0)=x) = p. \end{cases} \tag{40}$$

Proof Assume that $\pi^*(X(0)=x) = a$, then the first order time expansion of Eq.(11) reads as:

$$J_a(x,\gamma) = \xi h_a(x) + (1-\delta\xi)(J_a(x,\gamma) + \xi\frac{d}{dt}X(0)\frac{d}{dx}J_a(x,\gamma)),$$

After neglecting the terms of order $\mathcal{O}(\xi^2)$ in the above expansion, one gets the required result. A similar expression can be directly derived when $\pi^*(X(0)=x) = p$.

□

We solve Eq.(40) for $\frac{d}{dx}J_\theta(x,\gamma)$, $\theta = a, p$ and obtain:

$$\frac{d}{dx}J_\theta(x,\gamma) = \begin{cases} \frac{\delta J_a(x,\gamma) - h_a(x)}{f_a(x)} & \text{in the active phase,} \\ \frac{\delta J_p(x,\gamma) - h_p(x) - \gamma}{f_p(x)} & \text{in the passive phase.} \end{cases}$$

Using the smooth-fit principle given in Eq.(13) and solving for γ we obtain:

$$\nu(x) = h_a(x) - h_p(x) + \frac{(f_p(x) - f_a(x))(f_p(x)h_a'(x) - f_a(x)h_p'(x))}{f_p(x)(f_a'(x) - \delta) - f_a(x)(f_p'(x) - \delta)}. \tag{41}$$

Observe that in the limit $\delta \to 0$, Eq.(41) consistently reduces to the result given in the Proposition 8 of [21].

B. Optimal cost functions

Here we derive the optimal cost functions $J_a(x,\gamma)$, $J_p(x,\gamma)$ for a single class make-to-stock server which dynamics is given by a Markov chain in continuous time.

We saw in section IV that the optimal cost functions for the γ-penalized problem obey to the Eq.(23):

$$\begin{cases} \delta J_a(x,\gamma) = h(x) + \lambda J_a(x-1,\gamma) + \mu J_a(x+1,\gamma) - \\ \qquad\qquad (\lambda+\mu)J_a(x,\gamma) \\ \delta J_p(x,\gamma) = h(x) + \lambda J_p(x-1,\gamma) - \lambda J_p(x,\gamma) + \gamma. \end{cases}$$

This system is linear and the general solutions of the homogenous system are

$$\begin{aligned} J_a(x,\gamma) &= C_{a^+}(w_+)^x + C_{a^-}(w_-)^x, \\ J_p(x,\gamma) &= C_p(w_0)^x, \end{aligned}$$

where C_{a^+}, C_{a^-}, C_p are integration constants and

$$\begin{aligned} w_+ &= \tfrac{(\delta+\lambda+\mu)+\sqrt{(\delta+\lambda+\mu)^2-4\lambda\mu}}{2\mu}, \\ w_- &= \tfrac{(\delta+\lambda+\mu)-\sqrt{(\delta+\lambda+\mu)^2-4\lambda\mu}}{2\mu}, \\ w_0 &= \tfrac{\lambda}{\lambda+\delta}. \end{aligned} \tag{42}$$

The particular solutions correspond to engage or to let the server be idle forever. Being active forever, we get:

$$\begin{aligned} &S_a(x,\gamma) = E_x\left[\int_0^\infty e^{-\delta s}h(X(s))ds\right] = \\ &\int_0^\infty e^{-\delta s}\sum_{k=-\infty}^{\infty} h(k)P\{X(s)=k|X(0)=x\}ds = \\ &\int_0^\infty e^{-\delta s}\sum_{k=-\infty}^{\infty} h(k)P\{X(s)=k-x|X(0)=0\}ds. \end{aligned} \tag{43}$$

Let us now calculate now the transition probability density $P\{X(s)=k-x|X(0)=0\}$. Consider a space homogeneous markov chain process $X(t)$ with parameter λ and μ. Define $\rho = \frac{\lambda}{\mu}$ and $P_n(t) = P\{X(t)=n|X(0)=0\}$. We know that $P_n(t)$ follows the equation (see for example [15]):

$$\frac{d}{dt}P_n(t) = -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t). \tag{44}$$

Define:

$$P_n(t) = Q_n(t)\rho^{-\frac{n}{2}}e^{-(\lambda+\mu-2\sqrt{\lambda\mu})t}$$

in terms of which the Eq.(44) becomes:

$$\frac{d}{dt}Q_n(t) = \sqrt{\lambda\mu}(Q_{n+1}(t) + Q_{n-1}(t) - 2Q_n(t)).$$

The solution of this last equation reads (see for example [1]):

$$Q_n(t) = e^{-2\sqrt{\lambda\mu}t}\mathbb{I}_{|n|}(2\sqrt{\lambda\mu}t)$$

with $\mathbb{I}_n(x)$ being the modified Bessel function:

$$e^{\frac{1}{2}x(t+1/s)} = \sum_{k=-\infty}^{\infty} t^k \mathbb{I}_n(x).$$

Hence, we obtain:

$$P_n(t) = \rho^{-\frac{n}{2}} e^{-(\lambda+\mu)t}\mathbb{I}_{|n|}(2\sqrt{\lambda\mu}t)$$

Using the Laplace transform of $P_n(t)$, which directly occurs in Eq.(43), we end with:

$$S_a(x,\gamma) = \sum_{k=-\infty}^{\infty} \frac{h(k)\delta^{-\frac{k-x}{2}}}{\sqrt{(\delta+\lambda+\mu)^2-4\lambda\mu}} \cdot \left[\frac{(\rho+\lambda+\mu)-\sqrt{(\rho+\lambda+\mu)^2-4\lambda\mu}}{2\sqrt{\lambda\mu}}\right]^{|k-x|}$$

From Eq.(42) and the fact that $w_+w_- = \frac{\lambda}{\mu} = \rho$, we can show that $S_a(x,\gamma)$ takes the form:

$$S_a(x,\gamma) = \frac{1}{\mu(w_+-w_-)}\Big\{h(x) + (w_-)^x \sum_{k=-\infty}^{x-1} h(k)(w_-)^{-k} + (w_+)^x \sum_{k=x+1}^{\infty} h(k)(w_+)^{-k}\Big\}. \quad (45)$$

Along the same lines, when the server is idle forever, we obtain:

$$S_p(x,\gamma) = \int_0^\infty e^{-\delta s}\sum_{k=-\infty}^{\infty}(h(k)+\gamma)(P\{X(s)=k-x|X(0)=0\}ds.$$

In this case $P\{X(s)=k-x|X(0)=0\}$ is a Poisson process and we end with:

$$S_p(x) = (w_0)^{x+1}\sum_{k=-\infty}^{x}\frac{(h(k)+\gamma)(w_0)^{-k}}{\lambda}. \quad (46)$$

C. Index obtained for a piecewise linear running cost

Here we calculate the index for the single class make-to-stock server problem described in section IV-A when the dynamics is a continuous time Markov chain and the cost rate function $h(x)$ is piecewise linear:

$$h(x) = \begin{cases} +Ax \;;\; A>0 & \text{if } x \geq 0 \\ -Bx \;;\; B>0 & \text{if } x < 0. \end{cases}$$

We have shown in section IV-A that the cost functions $J_a(x)$ and $J_p(x)$ are:

$$J_a(x,\gamma) = C_a(w_+)^x + S_a(x,\gamma)$$
$$J_p(x,\gamma) = C_p(w_0)^x + S_p(x,\gamma),$$

where $S_a(x)$ and $S_p(x)$ are given by Eq.(45) and Eq.(46) respectively.

Using the form of $h(x)$, we derive the closed form of $J_a(x,\gamma)$ and $J_p(x,\gamma)$. For the region $x \geq 0$, the summation formula for geometric series implies:

$$J_a(x,\gamma) = \frac{1}{\delta^2\mu(w_+-w_-)}\Big\{2^{-x-1}\Big[(A+B)\Big((\lambda-\mu)^2+\delta(\lambda+\mu)\Big)(2w_-)^x + \mu(w_+-w_-)\Big(2^{x+1}A(\delta x-\lambda+\mu)+A(\lambda-\mu)(2w_-)^x + B(\lambda-\mu)(2w_-)^x + 2C_{a^+}\delta^2(2w_+)^x\Big)\Big]\Big\},$$

and

$$J_p(x,\gamma) = \frac{(\delta+\lambda)(A+\delta-A\lambda+\delta\gamma)+(w_0)^x\Big((\delta+\lambda)(C_p\delta^2+(A+B)\lambda)+\delta^2\gamma\Big)}{\delta^2(\delta+\lambda)}.$$

Similarly, for the region $x < 0$ region, we obtain:

$$J_a(x,\gamma) = \frac{1}{\delta^2\mu(w_+-w_-)}\Big\{2^{-x-1}\Big[(A+B)\Big((\lambda-\mu)^2+\delta(\lambda+\mu)\Big)(2w_+)^x + \mu(w_+-w_-)\Big(-2^{x+1}B(\delta x-\lambda+\mu)+\big(2C_{a^+}\delta^2 - (A+B)(\lambda-\mu)\big)(2w_+)^x\Big)\Big]\Big\},$$

and

$$J_p(x,\gamma) = \frac{B(-\delta x+\lambda)+\delta(C_p\delta(W_0)^x+\gamma)}{\delta^2}.$$

Using the smooth-fit principle given in Eq.(13) and the definition Eq.(25), we can derive the index $\nu(x)$ in the form given by Eq.(26).

D. Position of the hedging stock

Here we calculate the optimal position of the hedging level for a single class make-to-stock server with Markov dynamics $X(t)$.

In section IV-A, we saw that the optimal policy for a single item make-to-stock problem is a hedging stock policy. Under such a policy, the stochastic process:

$$Y(t) = d^* - X(t)$$

is isomorphic to a $M/M/1$ queue [17]. Let us denote by $J^{d^*}(x)$, the cost incurred under policy π. Optimality implies that the discrete derivative with respect to d^* vanishes, namely:

$$J^{d^*}(x) - J^{d^*+1}(x) = 0. \quad (47)$$

To solve Eq.(47) let us first recall that:

$$J^{d^*}(x) = E_x\int_0^\infty e^{-\delta t}h(X(t))dt.$$

Let τ be an exponentially distributed random variable with mean $1/\delta$ independent of $X(t)$ and $h(x)$. Then

$$J^{d^*}(x) = \frac{1}{\delta} E_x E_\tau \left[h(X(\tau)) \right]. \quad (48)$$

Permutating the expectation operators in Eq.(48), we have:

$$\delta J^{d^*}(x) = E_\tau \left[\sum_{i=-\infty}^{d^*} h(i) P(X(\tau) = i \,|\, X(0) = x) \right]. \quad (49)$$

Letting $Y(\tau) = d^* - X(\tau)$ we obtain:

$$\delta J^{d^*}(x) = E_\tau \Bigg[A \sum_{y=0}^{d^*-1} (d^* - y) P(Y(\tau) = y \,|\, Y(0) = 0) + B \sum_{y=d^*+1}^{\infty} (y - d^*) P(Y(\tau) = y \,|\, Y(0) = 0) \Bigg].$$

From Eq.(47) we obtain:

$$0 = \delta J^{d^*+1}(x) - \delta J^{d^*}(x) = E_\tau \left[(A+B) \left(1 - \sum_{y=d^*+1}^{\infty} P(Y(\tau) = y \,|\, Y(0) = 0) \right) - B \right].$$

Calculating the expectation with respect to τ, we find that:

$$0 = \delta J^{d^*+1}(x) - \delta J^{d^*}(x) =$$

$$A - (A+B)\delta \int_0^\infty e^{-\delta t} \sum_{y=d^*+1}^{\infty} P\left(Y(t) = y \,|\, Y(0) = d^* - x\right) dt.$$

It is known that when $x = d^*$, the Laplace transform $f(\delta, y)$ of the transient probability density of the $M/M/1$ queue $P\left(Y(t) = y \,|\, Y(0) = d^* - x\right)$ reads simply as (see [10] for example):

$$f(\delta, y) = \frac{(1 - \omega_-)\omega_-^y}{\delta},$$

where:

$$\omega_- = \frac{(\delta + \lambda + \mu) - \sqrt{(\delta + \lambda + \mu)^2 - 4\lambda\mu}}{2\mu}.$$

So we obtain:

$$\delta \int_0^\infty e^{-\delta t} \sum_{y=d^*+1}^{\infty} P\left(Y(t) = y \,|\, Y(0) = d^* - x\right) dt = \omega_-^{d^*+1},$$

and we end with:

$$0 = A - (A+B)\omega_-^{d^*+1} \quad \Rightarrow$$

$$d^* = \left\lfloor \frac{1}{\ln(\omega_-)} \ln\left(\frac{A}{A+B} \right) \right\rfloor.$$

References

[1] W. Feller *"An introduction to probability theory and its applications"*. Vol II, 2nd. J. Wiley, (1970)

[2] J. C. Gittins. *"Multi-Armed Bandits Allocation Indices"*. J. Wiley, (1989).

[3] P. Glasserman "Hedging-Point Production Control with Multiple Failure Modes". Trans. Automat. Cont. **40** (4), (1995), 707-712.

[4] A. Y. Ha "Optimal dynamic scheduling policy for a make-to-stock production system". Oper. Res. **45**, (1997), 42-53 .

[5] M.-O. Hongler and F. Dusonchet "Optimal stopping and Gittins' indices for piecewise deterministic evolution process". J. of discrete Events Systems **11** (3), (2001), 235-248.

[6] I. Karatzas "Gittins indices in the dynamic allocation problem for diffusion processes". Ann. Appl. Probab. **15**, (1987), 1527-1556.

[7] H. Kaspi and A. Mandelbaum " Lévy Bandit: Multi-armed Bandit driven by Lévy processes". The Ann. of Appl. Probab. **5**, (1995), 541-565.

[8] E. V. Krichagina, Sheldon X. C. Lou, Suresh P. Sethi and Michael I. Taksar "Production control in a failure-prone manufacturing system: Diffusion approximation and asymptotic optimality". The Ann. of Appl. Probab. **3**, (1993), 421-453 .

[9] S. A. Lippman "Applying a New Device in the Optimization of Exponential Queuing System". In Oper. Res. **23**, (1975), 687-710.

[10] J. Medhi "Stochastic processes". J. Wiley, (1994).

[11] J. Nino-Mora "Restless Bandit, partial conservation law and indexability". Adv. Appl. Probab. **33** (1), (2001), 76-98.

[12] J. Nino-Mora "On certain greedoid polyhedra, partially indexable scheduling problems, and extended Restless Bandit allocation indices". Submitted to *mathematical programming*

[13] C. H. Papadimitriou and J. N. Tsitsiklis "The complexity of optimal queueing network control". Math. Oper. Res **24**, (1999), 293-305.

[14] A. Pena and P. Zipkin "Dynamic scheduling rules for a multiproduct make-to-stock queue". Oper. Res. **15**, (1997), 919-930.

[15] M. L. Puterman "Markov Decision Processes. Discrete stochastic dynamic programing". Wiley-Interscience Publication (1994).

[16] A.N. Shiryayev "Optimal Stopping Rules". Springer, verlag (1978). Applications of mathematic 8.

[17] M.H Veatch and L. M. Wein "Scheduling a make-to-stock queue: index policies and hedging points". Oper. Res. **44**, (1996), 634-647.

[18] F. de Véricourt, F. Karaesman and Y. Dallery. "Dynamic Scheduling in a Make-to stock system: A partial characterization of optimal policies". Oper. Res. **48** (5), (2000), 811-819.

[19] L. M. Wein "Dynamic scheduling of a multi-class make-to-stock queue". Oper. Res. **40**, (1992), 724-735.

[20] P. Whittle. "Optimization over Time. Dynamic Programming and Stochastic Control". J. Wiley, New-York, (1982) Volume I.

[21] P. Whittle "Restless Bandit: Activity in a changing world". A Celebration of Applied Probability, J. Gani (ed.), J. Appl. Probab **25A**, (1988), 287-298.

[22] P. Whittle "Optimal control : Basics and Beyond". John Wiley and Sons. Collection: Wiley-Interscience series in systems and optimization (1996).