BOTTLENECKS IN SERIAL PRODUCTION LINES: A SYSTEM-THEORETIC APPROACH

C.-T. KUO, J.-T. LIM* and S. M. MEERKOV

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122

In this work, a new definition of production systems bottlenecks is formulated and analyzed. Specifically, a machine is defined as the bottleneck if the sensitivity of the system's performance index to this machine's production rate in isolation is the largest. Although appealing from the systems point of view, this definition suffers a deficiency due to the fact that the sensitivities involved cannot be either measured on-line or efficiently calculated off-line. To avoid this, the paper develops a method based on indirect but real-time data. From this point of view, the main result of the work is as follows: The bottleneck machine in a serial production line can be identified by analyzing relationships between the so-called manufacturing blockage and manufacturing starvation of each machine. This leads to a simple rule for bottleneck identification. The rule requires neither the calculation of the production rate sensitivities nor the production rate itself. When the probabilities of manufacturing blockages and starvations are not available from on-line measurements, the paper presents their analytical estimates which, under certain conditions, can be used for bottleneck identification. Finally, a case study at an automotive component plant is described.

1. INTRODUCTION AND PROBLEM FORMULATION

Serial production lines are manufacturing systems consisting of machines arranged in the consecutive order and buffers separating each two adjacent machines. For the purposes of this study, we formalize a serial production line by the following assumptions [1]:

(i) The system consists of M machines arranged serially, and $M - 1$ buffers separating each consecutive pair of machines.

(ii) The machines have identical cycle time T_c. The time axis is slotted with the slot duration T_c. Machines begin operating at the beginning of each time slot.

(iii) Each buffer is characterized by its capacity, $N_i < \infty$, $1 \leq i \leq M - 1$.

(iv) Machine i is starved during a time slot if buffer $i - 1$ is empty at the beginning of the time slot. Machine 1 is never starved for parts.

(v) Machine i is blocked during a time slot if buffer i has N_i parts at the beginning of

---

*On Sabbatical leave from the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Taejon, Korea.

the time slot and machine $i + 1$ fails to take a part during the time slot. Machine M is never blocked by ready goods buffer.

(vi) Machine i, being neither blocked nor starved during a time slot, produces a part with probability p_i and fails to do so with probability $1 - p_i$. Parameter p_i is referred to as the **production rate** of machine i **in isolation**.

For this system, the Production Rate $(\widetilde{PR})$, defined as the average number of parts produced by the last machine, m_M, per cycle of time, T_c, is often considered as a performance index of interest. According to the model (i)–(vi), $\widetilde{PR}$ is a function of all system's parameters, i.e.,

$$\widetilde{PR} = \widetilde{PR}(p_1, \ldots, p_M, N_1, \ldots, N_{M-1}).$$

Given system (i)–(vi), an important practical problem is the identification of a machine that impedes the system performance (i.e., $\widetilde{PR}$) in the strongest manner. Such a machine is typically referred to as the **bottleneck (BN)**, and many believe that a key to a successful production line operation is a continuous process of identification and elimination of bottlenecks. Often, this is referred to as the Process of Continuous Improvement or **Kaizan**, using the Toyota production system terminology. In fact, a large body of work, called the Theory of Constraints [2], is devoted to this subject.

In spite of its importance, the notion of a bottleneck does not seem to have a unanimous and unambiguous interpretation (see [3] for a review of the subject). Specifically, the following definitions are often found in the literature:

(a). A machine is the bottleneck if its isolation production rate is the smallest among all the machines in the system. In terms of the model (i)–(vi), this implies that m_i is the bottleneck if $p_i < p_j$, $\forall j \neq i$.

(b). A machine is the bottleneck if the Work-in-Process inventory (WIP) in the buffer immediately preceding this machine is the largest. In terms of the model (i)–(vi), m_i is the bottleneck if $E[h_{i-1}] > E[h_j]$, $\forall j \neq i - 1$, where h_i is the occupancy of b_i and $E[\cdot]$ denotes the expectation.

In this work, we define the bottleneck as follows:

DEFINITION 1.1.: A machine is the **bottleneck** if the sensitivity of the system's performance index to its production rate in isolation is the largest, as compared to all other machines. In terms of the model (i)–(vi), m_i is the bottleneck if

$$\frac{\partial \widetilde{PR}(p_1, \ldots, p_M, N_1, \ldots, N_{M-1})}{\partial p_i} > \frac{\partial \widetilde{PR}(p_1, \ldots, p_M, N_1, \ldots, N_{M-1})}{\partial p_j}, \quad \forall j \neq i. \quad (1.1)$$

Although in some cases, all three definitions, (a), (b) and (1.1), result in the same machine as the bottleneck, they are not always identical. Indeed, Figure 1 gives an
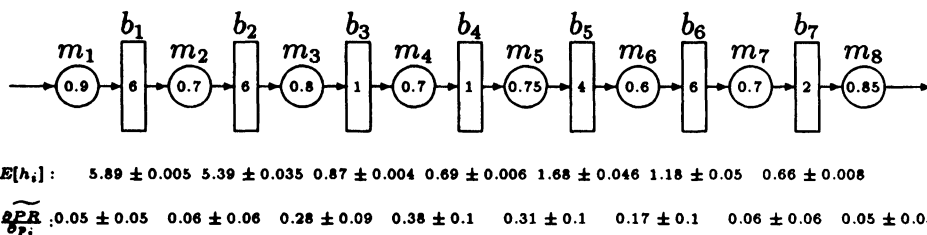
$E[h_i]$:   5.89 ± 0.005 5.39 ± 0.035 0.87 ± 0.004 0.69 ± 0.006 1.68 ± 0.046 1.18 ± 0.05   0.66 ± 0.008

$\frac{\widetilde{\partial PR}}{\partial p_i}$:0.05 ± 0.05   0.06 ± 0.06   0.28 ± 0.09   0.38 ± 0.1   0.31 ± 0.1   0.17 ± 0.1   0.06 ± 0.06   0.05 ± 0.05

Figure 1   Example of bottleneck machines for different definitions

example where the machine with the largest sensitivity is neither the one with the smallest p_i nor with the largest $E[h_{i-1}]$. In Figure 1 and all the subsequent figures, the numbers in the circles and rectangles denote the machines' and the buffers' parameters, p_i and N_i, respectively. The two rows of numbers below the buffers and the machines show the

values of $E[h_i]$ and $\partial \widetilde{PR}/\partial p_i$ along with 95% confidence intervals. Thus, according to definition (a), the bottleneck is machine m_6, according to definition (b)—machine m_2, and according to (1.1)—machine m_4. Consequently, the largest increase in the system's production rate is obtained if machine m_4, rather than m_6 and m_2, is improved.

Remark 1.1: In Figure 1 and in all the subsequent figures, the simulations are based on GPSS/H software package [4]. The confidence intervals for all statistical estimates have been evaluated as follows [5]: Let w be the length of a simulation run in which the random variable Y is measured. Divide the data measured, i.e., Y_1, Y_2,..., Y_w into n batches of

length $c = \dfrac{w}{n}$. Hence, $Y_1,...,Y_c$ form batch 1, $Y_{c+1},...,Y_{2c}$ – batch 2, etc. Let $\tilde{Y}_j(c), j = 1,...,n$, be the j-th batch mean and $\tilde{\tilde{Y}}(n\ c) = \sum_{j=1}^{n} \tilde{Y}_j(c)/n = \sum_{k=1}^{w} Y_k/w$ the grand sample mean. Then, an approximate $100(1 - \beta)\%$ confidence interval is

$$\tilde{\tilde{Y}}(n, c) \pm t_{n-1,1-\beta/2} \sqrt{\frac{\sigma^2_{\tilde{Y}(c)}(n)}{n}}, \qquad (1.2)$$

where $t_{n-1,1-\beta/2}$ is the upper $1 - \beta/2$ critical point for the t distribution with $n - 1$ degrees of freedom and

$$\sigma^2_{\tilde{Y}(c)}(n) = \frac{\sum_{j=1}^{n} [\tilde{Y}_j(c) - \tilde{\tilde{Y}}(n, c)]^2}{n - 1}. \qquad (1.3)$$

Definition 1.1 was introduced in [1] where, in addition, it was shown that if a serial production line is unimprovable with respect to workforce re-distribution (i.e., $\widetilde{PR}$ cannot be increased by re-assigning work among the operations), then the machine with the smallest p_i is indeed the bottleneck; otherwise any machine, including the one with the largest p_i, can be the bottleneck. However, no characterization of bottlenecks in "improvable" systems has been given. Such a characterization is the topic of this paper.

Each of the Definitions, (a), (b), and (1.1), have their advantages and disadvantages. Definitions (a) and (b) are advantageous because they are formulated in terms of on-line data which can be measured during real-time system operation. However, being local in nature, they may not identify the worst machine from the total system point of view. Indeed, an increase of p_2 and p_6 in the system shown in Figure 1 leads to a smaller improvement of $\widetilde{PR}$ than the increase in p_4.

On the other hand, Definition 1.1 takes into account not only the local but also the global properties of system (i)–(vi): the sensitivity of $\widetilde{PR}$ to p_i depends not only on p_i but also on all other p_j's, $j = 1,..., i - 1, i + 1,..., M$, and all N_j's, $j = 1,..., M - 1$, simultaneously. Thus, Definition 1.1 identifies the bottleneck from the total system point of view. However, it is not formulated in terms of on-line data, and the quantities involved in (1.1) cannot be measured directly during real-time system operation. Moreover, since even the mere calculation of $\widetilde{PR}$ is a formidable task and cannot be accomplished in a closed from (see [6], [7]), the calculation of its derivatives may be a more difficult or even a prohibitive problem. Therefore, to make Definition 1.1 practical, it has to be re-formulated in terms of data available on the factory floor. **The goal of this paper is to derive an indicator that permits one to identify the bottleneck machine, defined by Definition 1.1, using real-time data.**

From this point of view, the main result of this work is as follows: **Inequality (1.1) can be reformulated in terms of inequalities between the probabilities of the so-called manufacturing blockages and manufacturing starvations of each machine in the system. Since the frequency of blockages and starvations can be measured on-line, this leads to a rule for bottleneck identification which uses only real-time data, without calculating either $\partial\widetilde{PR}/\partial p_j$, or $\widetilde{PR}$, and even without the knowledge of p_i's and N_i's.**

When the probabilities of manufacturing blockages and starvations are not available from real-time measurements, the papers gives their analytical estimates which, under certain conditions, can be used for bottleneck identification as well. The calculation of these estimates does require the knowledge of p_i's and N_i's.

The remainder of this paper is structured as follows: In Section 2, a simple case of a two machines—one buffer system is considered. The general case is studied in Sections 3–5. In section 6, the notion of potency of a material handling system is introduced. A practical application is described in Section 7. Finally, the conclusions are formulated in Section 8. All proofs are collected in Appendices A and B. The term bottleneck is used throughout in the sense of Definition 1.1.

2. TWO MACHINES—ONE BUFFER SYSTEM

Introduce the notions of manufacturing blockage and starvation:

DEFINITION 2.1: Machine m_i is said to be **blocked in the manufacturing sense** during a time slot if it is up during this time slot, b_i is full at the beginning of this time slot, and m_{i+1} fails to take a part from b_i at the beginning of this time slot.

DEFINITION 2.2: Machine m_i is said to be **starved in the manufacturing sense** during a time slot if it is up during this time slot and b_{i-1} is empty at the beginning of this time slot.

Remark 2.1: The blockage and starvation introduced by assumptions (iv) and (v) of Section 1 are sometimes referred to as communication blockage and starvation; they differ from the above by deleting the requirement of the machine being up.

Let $\widetilde{mb}_i$ and $\widetilde{ms}_i$ denote the probabilities of manufacturing blockage and starvation, respectively, i.e.,

$$\widetilde{mb}_i = \text{Prob } (\{m_i \text{ is up during a time slot}\}$$
$$\cap \{b_i \text{ is full at the beginning of this slot}\}$$
$$\cap \{m_{i+1} \text{ fails to take a part from } b_i \text{ at the beginning of this slot}\}),$$

$$\widetilde{ms}_i = \text{Prob } (\{m_i \text{ is up during a time slot}\}$$
$$\cap \{b_{i-1} \text{ is empty at the beginning of this slot}\}).$$

LEMMA 2.1: *For the serial production line (i)–(vi) with $M = 2$,*

$$\widetilde{mb}_1 = p_1 Q(p_2, p_1, N_1),$$

$$\widetilde{ms}_2 = p_2 Q(p_1, p_2, N_1),\tag{2.1}$$

where

$$Q(x, y, N) = \begin{cases} \dfrac{(1 - x)(1 - \alpha)}{1 - \dfrac{x}{y}\alpha^N}, & x \neq y, \\[3ex] \dfrac{1 - x}{N + 1 - x}, & x = y, \end{cases}$$

$$\alpha = \frac{x(1 - y)}{y(1 - x)}.\tag{2.2}$$

Proof: See Appendix A.

THEOREM 2.1: *For the serial production line (i)–(vi) with $M = 2$, the inequality*

$$\frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_1} > \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_2} \qquad (or \; \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_1} < \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_2}).$$

takes place if and only if

$$\widetilde{mb}_1 < \widetilde{ms}_2 \qquad (or \; \widetilde{mb}_1 > \widetilde{ms}_2, \; respectively). \qquad (2.3)$$

Proof: See Appendix A.

This Theorem relates inequality (1.1) with the inequality between blockages and starvations (2.3). Specifically, it states that the bottleneck can be identified without calculating $\partial \widetilde{PR}/\partial p_i$ but just observing blockages and starvations of the machines: if the former is larger than the latter, the bottleneck is downstream of the buffer, otherwise the bottleneck is upstream.

Obviously, the case of $M = 2$ is quite simple and might be of a little practical importance. Moreover,

COROLLARY 2.1: *For $M = 2$, the bottlenecks identified by definitions (a) and (1.1) are always the same.*

Proof: See Appendix A.

However, as is it is shown below, the insight provided by $M = 2$ can be used as a guide for analysis of the general case.

3. GENERAL CASE

Similar to $\widetilde{PR}$, the probabilities $\widetilde{mb}_i$ and $\widetilde{ms}_i$ cannot be calculated exactly for $M > 2$. Therefore, a direct generalization of Theorem 2.1 for $M > 2$ is impossible. Nevertheless, there are two avenues available for the extension of the result of Section 2 to the general case: In the first one, guided by the insight provided by the two machine—one buffer case, we postulate a rule for bottleneck identification, which uses machine blockages and starvation in a manner similar to Theorem 2.1, and show, using discrete event simulations, that in most cases it indeed results in a correct bottleneck identification. This is carried out in this Section. In the second avenue, we develop an analytical approach to the rule postulated. Here, based on the techniques developed in [1], we derive estimates of $\widetilde{mb}_i$ and $\widetilde{ms}_i$, denoted as mb_i and ms_i, and show that, under certain conditions, the bottleneck identified using the measured data, $\widetilde{mb}_i$ and $\widetilde{ms}_i$, and the calculated data, mb_i and ms_i, do coincide. This is carried out in Section 4.

Consider the system (i)–(vi) with $M > 2$. Assume that the following assumption holds:

HYPOTHESIS H.1: *There exists a single machine, m_i, such that*

$$\widetilde{mb}_{j-1} > \widetilde{ms}_j, j = 2, \dots, i,$$

$$\widetilde{mb}_j < \widetilde{ms}_{j+1}, j = i, \dots, M - 1.$$

Remark 3.1: As it will become clear below, this hypothesis is introduced to ensure that the

system has a unique bottleneck. The case of multiple bottlenecks is considered in Section 5.

RULE 3.1: *Assume that Hypothesis H.1 holds. Then if*

$$\widetilde{mb}_j > \widetilde{ms}_{j+1}, \quad j = 1, \ldots, M - 1, \tag{3.1}$$

the bottleneck is downstream of machine m_j. If

$$\widetilde{ms}_j > \widetilde{mb}_{j-1}, \quad j = 2, \ldots, M, \tag{3.2}$$

the bottleneck is upstream of machine m_j.

Numerical Justification: As it has been pointed out above, Rule 3.1 was justified by discrete event simulations of serial production lines defined by assumptions (i)–(vi). Two typical examples are shown in Figure 2. The three rows of numbers below the machines show the values of $\widetilde{ms}_i$, $\widetilde{mb}_i$, and $\partial\widetilde{PR}/\partial p_i$, respectively, along with 95% confidence intervals. In each run of the discrete event model, zero initial conditions for all buffers have been assumed and a 10,000 time slots warm up period has been carried out. As it is illustrated in Figure 3, after this period, the transients have largely decayed, and the next 40,000 slots of the stationary regime have been used to statistically evaluate the quantities shown in Figure 2 and the variance (1.3) with $n = 40$ and $c = 1000$. The sensitivities $\partial\widetilde{PR}/\partial p_i$ have been estimated, using finite differences $\Delta\widetilde{PR}/\Delta p_i$, $\forall i$, with the step $\Delta p_i = 0.01$. The arrows in Figure 2 are obtained by placing a line within the sign of inequality between $\widetilde{mb}_j$ and $\widetilde{ms}_{j+1}$. Then, according to Rule 3.1, **the bottleneck is the machine that has the arrows pointed towards it from both sides (or one side, for the first and the last machine).** Thus, the bottlenecks in Figure 2 (a) and (b) are machines m_1 and m_4, respectively. This conclusion is supported by the values of $\partial\widetilde{PR}/\partial p_i$ as well. Note that, in view of the arrows mentioned above, Hypothesis 3.1 indeed implies that the system has a unique bottleneck.

With a very few exception, in all systems considered, the bottlenecks identified using Rule 3.1 and $\partial\widetilde{PR}/\partial p_i$ were the same. An example in which the two conclusions differ is shown in Figure 4. Here Rule 3.1 resulted in m_2, whereas $\partial\widetilde{PR}/\partial p_i$ in m_1, as the bottleneck. As it can be seen from the last row in Figure 4, the difference between $\partial\widetilde{PR}/\partial p_2$ and $\partial\widetilde{PR}/\partial p_1$ is quite small. This discrepancy could be attributed either to the fact that Δp_i is not small enough to result in a precise value of $\partial\widetilde{PR}/\partial p_i$ (smaller Δp_i results in numerical instability) or to a relative inaccuracy of Rule 3.1. Based on the above, we conclude
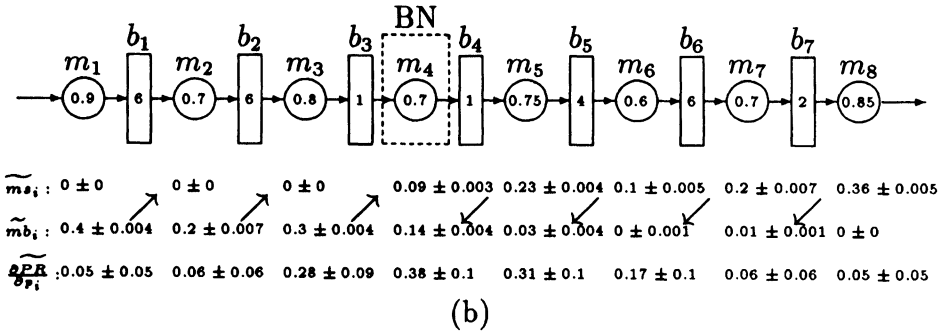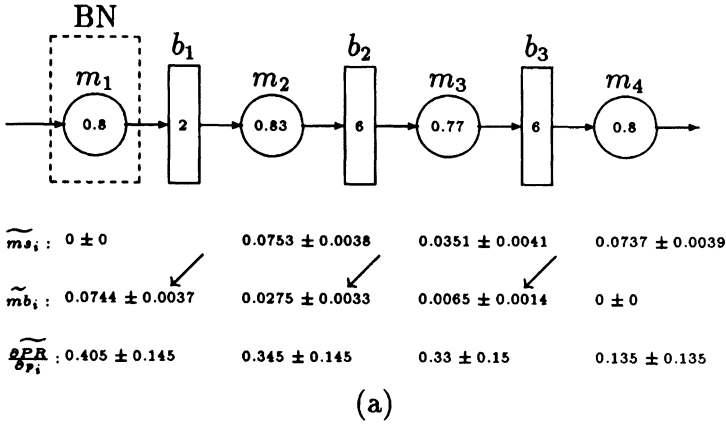
BN

m_1 (0.8) b_1 2 m_2 (0.83) b_2 6 m_3 (0.77) b_3 6 m_4 (0.8)

$\widetilde{ms_i}$: 0 ± 0   0.0753 ± 0.0038   0.0351 ± 0.0041   0.0737 ± 0.0039

$\widetilde{mb_i}$: 0.0744 ± 0.0037   0.0275 ± 0.0033   0.0065 ± 0.0014   0 ± 0

$\widetilde{\frac{\partial PR}{\partial p_i}}$: 0.405 ± 0.145   0.345 ± 0.145   0.33 ± 0.15   0.135 ± 0.135

(a)

BN

m_1 (0.9) b_1 6 m_2 (0.7) b_2 6 m_3 (0.8) b_3 1 m_4 (0.7) b_4 1 m_5 (0.75) b_5 4 m_6 (0.6) b_6 6 m_7 (0.7) b_7 2 m_8 (0.85)

$\widetilde{ms_i}$: 0 ± 0   0 ± 0   0 ± 0   0.09 ± 0.003 0.23 ± 0.004 0.1 ± 0.005 0.2 ± 0.007 0.36 ± 0.005

$\widetilde{mb_i}$: 0.4 ± 0.004 0.2 ± 0.007 0.3 ± 0.004 0.14 ± 0.004 0.03 ± 0.004 0 ± 0.001 0.01 ± 0.001 0 ± 0

$\widetilde{\frac{\partial PR}{\partial p_i}}$: 0.05 ± 0.05 0.06 ± 0.06 0.28 ± 0.09 0.38 ± 0.1 0.31 ± 0.1 0.17 ± 0.1 0.06 ± 0.06 0.05 ± 0.05

(b)

Figure 2   Illustration of Rule 3.1

that Rule 3.1 identifies either the machine with the largest $\widetilde{\partial PR/\partial p_i}$ or a machine with $\widetilde{\partial PR/\partial p_i}$ being close to the largest one. ∎

4. ANALYTICAL JUSTIFICATION

In this Section, we provide an analytical justification of Rule 3.1. Introduce

$\tilde{p}_i^f$ = Prob({m_i produces a part during a time slot} | {it is not blocked at the beginning of this time slot}),

$\tilde{p}_i^b$ = Prob({m_i produces a part during a time slot} | {it is not starved at the beginning of this time slot}).

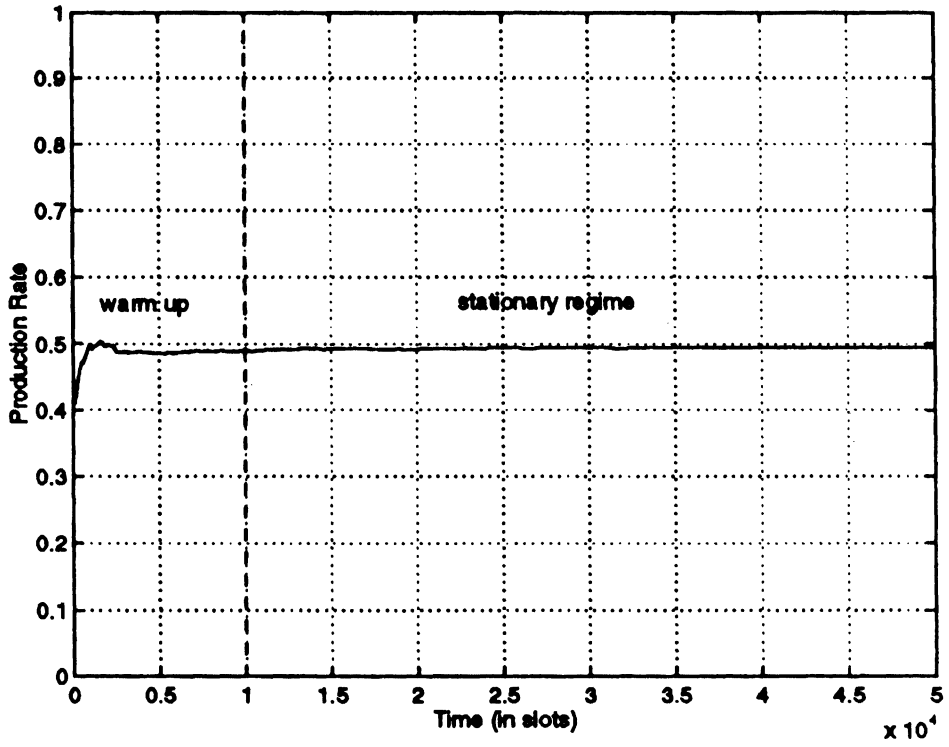Since the last machine cannot be blocked and the first machine starved,
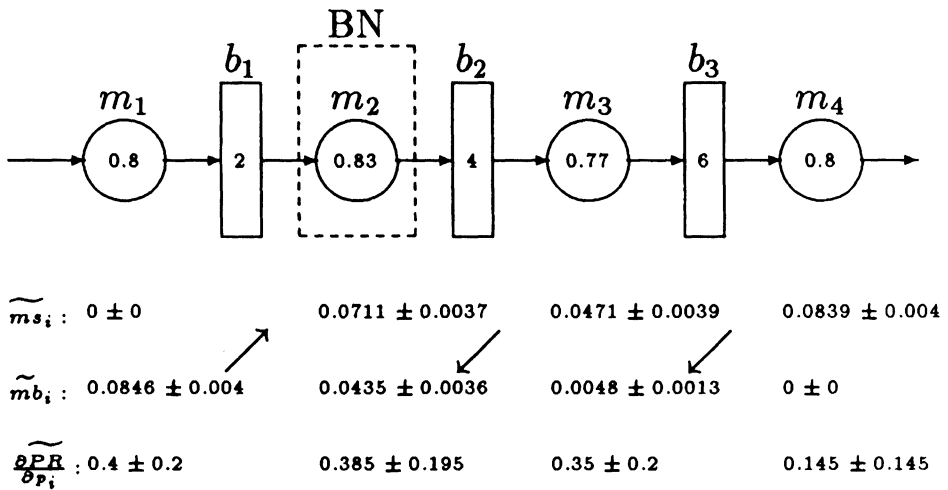
Figure 3 Warm up period and stationary regime

Figure 4 Counter example to Rule 3.1

$$\widetilde{PR} = \tilde{p}_M^f = \tilde{p}_1^b.$$

In addition, by Definitions 2.1 and 2.2,

$$\widetilde{mb}_i = p_i \sum_{j=i+1}^{M} (\prod_{r=i+1}^{j-1} p_r)(1 - p_j) X_{i, \ldots j-1}(N_{i, \ldots}, N_{j-1}), i = 1, \ldots, M - 1,$$

$$\widetilde{ms}_i = p_i X_{i-1}(0), i = 2, \ldots, M. \tag{4.1}$$

Unfortunately, $\tilde{p}_i^f$ and $\tilde{p}_i^b$ are unknown and cannot be calculated in the closed form. However, as it has been shown in [1], they can be evaluated using the following recursive procedure:

$$p_i^b(s + 1) = p_i[1 - Q(p_{i+1}^b(s + 1), p_i^f(s), N_i)], 1 \le i \le M - 1,$$

$$p_i^f(s + 1) = p_i[1 - Q(p_{i-1}^f(s + 1), p_i^b(s + 1), N_{i-1})], 2 \le i \le M,$$

$$p_1^f(s) = p_1, \ p_M^b(s) = p_M, \ s = 0, 1, 2, \ldots, \tag{4.2}$$

with initial conditions

$$p_i^f(0) = p_i, \ i = 1, \ldots, M,$$

where $Q(x, y, N)$ is defined in (2.2).

LEMMA 4.1: *Recursive procedure (4.2) is convergent and, therefore, the following limits exist*:

$$p_i^f = \lim_{s \to \infty} p_i^f(s), \quad p_i^b = \lim_{s \to \infty} p_i^b(s), \quad i = 1, \ldots, M. \tag{4.3}$$

Moreover,

$$|\tilde{p}_i^f - p_i^f| \sim \mathbb{O}(\delta), i = 1, \ldots, M - 1,$$

$$|\tilde{p}_i^b - p_i^b| \sim \mathbb{O}(\delta), i = 2, \ldots, M.$$

and

$$PR =: p_M^f = \widetilde{PR} + \mathbb{O}(\delta), \tag{4.4}$$

where δ is defined by (2.5) in [1] as the magnitude of the difference between the joint and the product of appropriate marginal probabilities of buffers occupancy; according to Numerical Fact 2.1 of [1], δ ≪ 1.

Proof: See Lemma A.10 and Theorem 2.1 in [1].

Given p_i^f and p_i^b evaluated according to (4.2), (4.3), introduce the **calculated** probabilities of manufacturing blocking and manufacturing starvation:

$$mb_i = p_i Q(p_{i+1}^b, p_i^f, N_i), \ i = 1, \dots, M - 1,$$

$$ms_i = p_i Q(p_{i-1}^f, p_i^b, N_{i-1}), \ i = 2, \dots, M. \tag{4.5}$$

LEMMA 4.2: *The calculated and the measured probabilities of manufacturing blockages and starvations are related as follows:*

$$\widetilde{|mb_i} - mb_i| \sim \mathcal{O}(\delta), \quad i = 1, \dots, M - 1,$$

$$\widetilde{|ms_i} - ms_i| \sim \mathcal{O}(\delta), \quad i = 2, \dots, M. \tag{4.6}$$

Proof: See Appendix B.

Since, as stated above, the calculated probabilities of blockages and starvations approximate the measured ones, the analytical justification of Rule 3.1 is carried out in terms of mb_i and ms_i. For this justification we need the following

HYPOTHESIS H.2: *For a production line (i)–(vi), if*

$$mb_{j-1} > ms_j, \quad j = 2, \dots, M,$$

then

$$X_{j-1}(0) \approx Q(p_{j-1}^f, p_j^b, N_{j-1}) =: \epsilon_{j1} \ll 1. \tag{4.7}$$

If

$$mb_j < ms_{j+1}, j = 1, \dots, M - 1,$$

then

$$(1 - p_{j+1}^b)X_j(N_j) \approx Q(p_{j+1}^b, p_j^f, N_j) =: \epsilon_{j2} \ll 1. \tag{4.8}$$

Here $X_{j-1}(0)$ and $X_j(N_j)$ are the probabilities that buffer $j - 1$ is empty and buffer j is full, respectively; function $Q(x, y, N)$ is defined in (2.2), and p_j^b and p_j^f are the steady states of the recursive procedure (4.2).

Remark 4.1: Expression (4.7) implies that if the upstream machine is blocked more often than the downstream machine is starved, the probability that the buffer, separating these

two machines, is empty is small. Similarly, (4.8) implies that if the upstream machine is blocked less frequently than the downstream machine is starved, the probability that the buffer is full is small.

Conditions under which Hypothesis H.2 takes place can be characterized formally in a number of situations. These include situations where the buffers are large enough or when p_j^f and p_j^b are sufficiently different from each other. The former is carried out in Lemma 4.3 and the latter is illustrated in Figures 5 and 6 (a).

LEMMA 4.3: *Under Hypothesis H.1, for any* $0 < \epsilon_0 \ll 1$ *there exists* N^* *such that if* $N_j > N^*, j = 1, \ldots, M - 1,$

$$\epsilon =: \max_{j,s}(\epsilon_{j1}, \epsilon_{s2}) < \epsilon_0,$$

where ϵ_{j1} *and* ϵ_{s2} *are defined in (4.7) and (4.8), respectively.*

Proof: See Appendix B.

It should not be necessarily concluded that N^* has to be large in order for ϵ to be small. Indeed, Figure 5 shows a system with $N^* = 2$ but ϵ quite small. On the other hand, the same system with machine m_3 being different ($p_3 = 0.75$ instead of 0.6) needs larger buffers to have ϵ small (see Figures 6 (a) and (b)). In Figures 5 and 6, the first four rows of numbers under the machines show the measured (during the process of discrete event
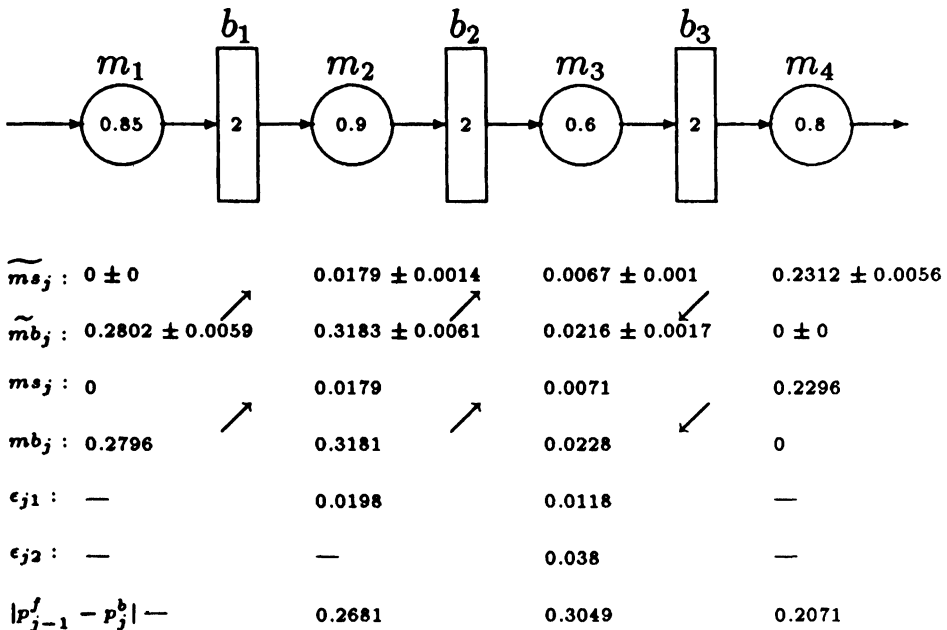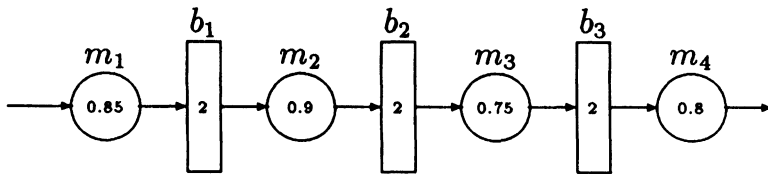

	m_1		m_2		m_3		m_4
$\widetilde{ms}_j$:	0 ± 0		0.0179 ± 0.0014		0.0067 ± 0.001		0.2312 ± 0.0056
$\widetilde{mb}_j$:	0.2802 ± 0.0059		0.3183 ± 0.0061		0.0216 ± 0.0017		0 ± 0
ms_j :	0		0.0179		0.0071		0.2296
mb_j :	0.2796		0.3181		0.0228		0
ϵ_{j1} :	—		0.0198		0.0118		—
ϵ_{j2} :	—		—		0.038		—
$|p_{j-1}^f - p_j^b|$	—		0.2681		0.3049		0.2071

Figure 5 Example of a system with a small N^* resulting in a small ϵ ($N^* = 2, \epsilon = 0.038$)

b_1 b_2 b_3

m_1 m_2 m_3 m_4

0.85 2 0.9 2 0.75 2 0.8

$\widetilde{ms}_j$: 0 ± 0 0.0338 ± 0.0023 0.0213 ± 0.0017 0.1202 ± 0.0037

$\widetilde{mb}_j$: 0.1692 ± 0.0048 0.1918 ± 0.0051 0.0498 ± 0.0024 0 ± 0

ms_j : 0 0.032 0.0202 0.1202

mb_j : 0.1702 0.1951 0.0514 0

ϵ_{j1} : — 0.0356 0.027 —

ϵ_{j2} : — — 0.0685 —

$|p_{j-1}^f - p_j^b|$: — 0.1451 0.1694 0.0702

(a)

b_1 b_2 b_3

m_1 m_2 m_3 m_4

0.85 4 0.9 4 0.75 4 0.8
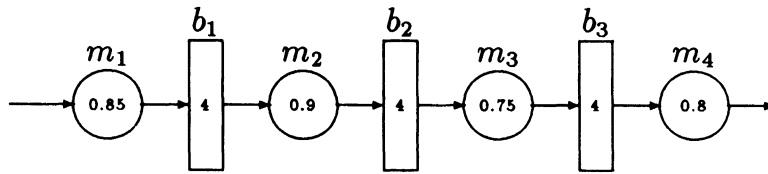
$\widetilde{ms}_j$: 0 ± 0 0.0082 ± 0.0017 0.0022 ± 0.0008 0.0726 ± 0.0036

$\widetilde{mb}_j$: 0.1216 ± 0.0048 0.1646 ± 0.005 0.0192 ± 0.0023 0 ± 0

ms_j : 0 0.0072 0.0015 0.0722

mb_j : 0.1222 0.1663 0.0207 0

ϵ_{j1} : — 0.008 0.002 —

ϵ_{j2} : — — 0.0277 —

$|p_{j-1}^f - p_j^b|$: — 0.1163 0.1635 0.0515

(b)

Figure 6 Illustration of the effect of $|p_{j-1}^f - p_j^b|$ on N^* ((a) $N^* = 2$, $\epsilon = 0.0685$, (b) $N^* = 4$, $\epsilon = 0.0277$)

simulations) and the calculated probabilities of manufacturing blockages and starvations whereas the last three rows give ϵ_{j1}, ϵ_{j2} and $|p^f_{j-1} - p^b_j|$, respectively.

Figures 5 and 6 (a) indicate that larger values of $|p^f_{j-1} - p^b_j|$ result in small values of ϵ. This illustrates another situation when Hypothesis H.2 takes place.

The main result of this Section is in the following:

THEOREM 4.1: *Under Hypothesis H.1 and H.2, the following two statements hold:*

(a). If

$$mb_j > ms_{j+1},$$

then there exists i > j such that

$$\frac{\partial PR}{\partial p_j} < \frac{\partial PR}{\partial p_i}.$$

(b). If

$$ms_j > mb_{j-1},$$

then there exists i < j such that

$$\frac{\partial PR}{\partial p_j} < \frac{\partial PR}{\partial p_i}.$$

Proof: See Appendix B.

Theorem 4.1 justifies Rule 3.1: Indeed, it says that under the conditions stated, if $mb_j > ms_{j+1}$, the bottleneck is downstream of machine m_j; if $ms_j > mb_{j-1}$, the bottleneck is upstream of machine m_j. In addition, this Theorem offers a possibility for bottleneck analysis in systems where the measured data, $\widetilde{mb}_i$ and $\widetilde{ms}_i$, is not available, e.g., in the design stage of a serial production line or in situations where production lines are not instrumented to measure $\widetilde{mb}_i$ and $\widetilde{ms}_i$. In such situations, given p_i's and N_i's, the values of $\widetilde{mb}_i$ and $\widetilde{ms}_i$ can be evaluated using (4.5), and Rule 3.1 (i.e., Theorem 4.1) could be applied, resulting in a correct bottleneck identification, if ϵ is sufficiently small. A case study, based on the calculated data mb_i and ms_i, is described in Section 7.

5. MULTIPLE BOTTLENECKS

Assume that Hypothesis H.1 does not hold, i.e., there are at least two machines that have the arrows pointed towards them from both side (or one side for the boundary machines). Examples of such systems are shown in Figures 7 and 8. Which one, then, is the "primary"
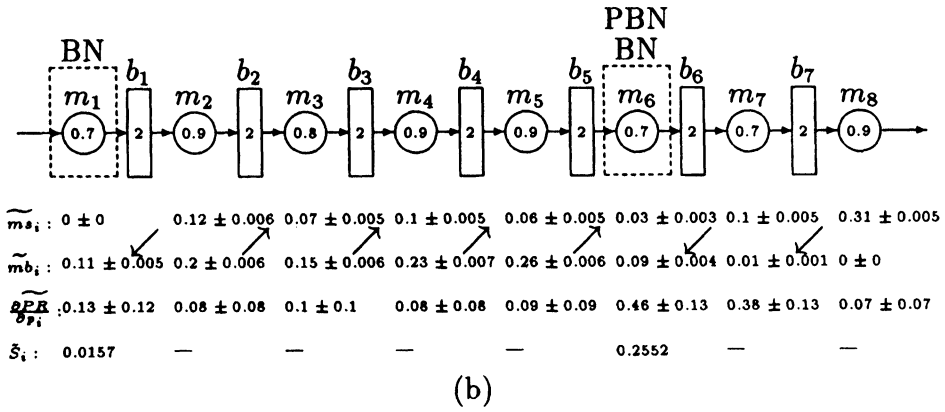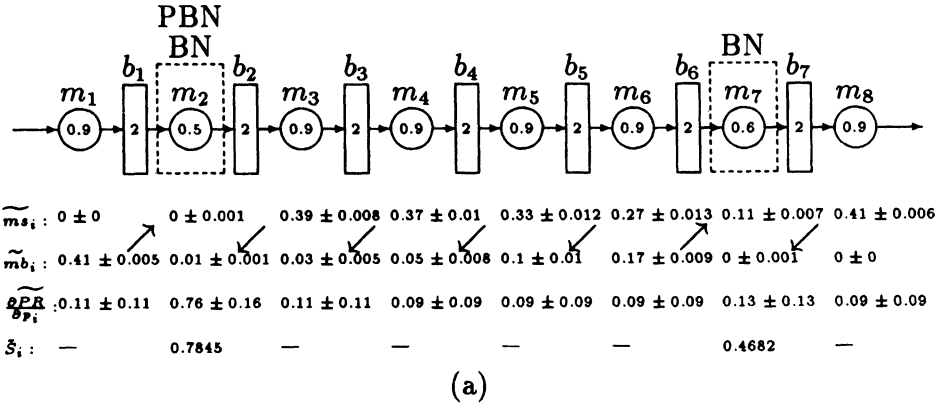
$\widetilde{ms}_i$: 0 ± 0 0 ± 0.001 0.39 ± 0.008 0.37 ± 0.01 0.33 ± 0.012 0.27 ± 0.013 0.11 ± 0.007 0.41 ± 0.006

$\widetilde{mb}_i$: 0.41 ± 0.005 0.01 ± 0.001 0.03 ± 0.005 0.05 ± 0.008 0.1 ± 0.01 0.17 ± 0.009 0 ± 0.001 0 ± 0

$\widetilde{\frac{\partial PR}{\partial p_i}}$: 0.11 ± 0.11 0.76 ± 0.16 0.11 ± 0.11 0.09 ± 0.09 0.09 ± 0.09 0.09 ± 0.09 0.13 ± 0.13 0.09 ± 0.09

$\tilde{S}_i$:   — 0.7845 —   —   —   — 0.4682 —

(a)

$\widetilde{ms}_i$: 0 ± 0 0.12 ± 0.006 0.07 ± 0.005 0.1 ± 0.005 0.06 ± 0.005 0.03 ± 0.003 0.1 ± 0.005 0.31 ± 0.005

$\widetilde{mb}_i$: 0.11 ± 0.005 0.2 ± 0.006 0.15 ± 0.006 0.23 ± 0.007 0.26 ± 0.006 0.09 ± 0.004 0.01 ± 0.001 0 ± 0

$\widetilde{\frac{\partial PR}{\partial p_i}}$: 0.13 ± 0.12 0.08 ± 0.08 0.1 ± 0.1 0.08 ± 0.08 0.09 ± 0.09 0.46 ± 0.13 0.38 ± 0.13 0.07 ± 0.07

$\tilde{S}_i$: 0.0157 —   —   —   — 0.2552 —   —

(b)

Figure 7 Illustration of Rule 5.1

$\widetilde{ms}_i$: 0 ± 0 0.01 ± 0.001 0.14 ± 0.007 0.09 ± 0.007 0.04 ± 0.003 0.19 ± 0.007 0.09 ± 0.005 0.3 ± 0.005

$\widetilde{mb}_i$: 0.29 ± 0.004 0.09 ± 0.006 0.17 ± 0.008 0.22 ± 0.007 0.06 ± 0.004 0.13 ± 0.006 0.01 ± 0.001 0 ± 0

$\widetilde{\frac{\partial PR}{\partial p_i}}$: 0.06 ± 0.06 0.27 ± 0.1 0.1 ± 0.1 0.08 ± 0.08 0.37 ± 0.1 0.1 ± 0.09 0.27 ± 0.1 0.06 ± 0.06

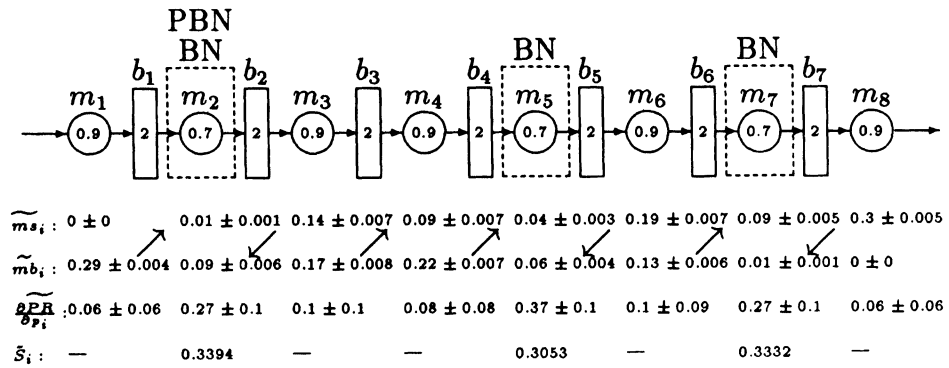$\tilde{S}_i$:   — 0.3394 —   — 0.3053 — 0.3332 —

Figure 8 Counter example to Rule 5.1

bottleneck? To answer this question, we introduce a heuristic notion of bottlenecks Severity. This notion can be defined in a number of way. The simplest (but, possibly, not the most accurate) one is as follows:

The **severity** of the i-th bottleneck is defined as

$$S_i = (mb_{i-1} + ms_{i+1}) - (mb_i + ms_i), i = 2, \dots, M - 1, \tag{5.1}$$

using the calculated data, or

$$\tilde{S}_i = (\widetilde{mb}_{i-1} + \widetilde{ms}_{i+1}) - (\widetilde{mb}_i + \widetilde{ms}_i), i = 2, \dots, M - 1, \tag{5.2}$$

using the measured data. For the first and the last machine as the bottleneck severity becomes, respectively,

$$S_i = ms_2 - mb_1 \text{ or } \tilde{S}_1 = \widetilde{ms}_2 - \widetilde{mb}_1,$$

$$S_M = mb_{M-1} - ms_M \text{ or } \tilde{S}_M = \widetilde{mb}_{M-1} - \widetilde{ms}_M.$$

RULE 5.1: Assume $m_{i_1}, \dots, m_{i_k}$, $k < M$, are the bottlenecks identified according to the Rule 3.1 (with Hypothesis H.1 being omitted). Let m_{i_j}, $j \in (1, \dots, k)$, be the bottleneck with the largest sensitivity. Then m_{i_j} is the primary bottleneck (PBN) in the sense that $\partial PR/\partial p_{i_j} > \partial PR/\partial p_j$ (or $\widetilde{\partial PR}/\partial p_{i_j} > \widetilde{\partial PR}/\partial p_i$), $i \neq i_j$.

Numerical Justification: This Rule also has been analyzed through discrete event simulations. Using the numerical procedure described in the Numerical Justification of Rule 3.1, a large number of systems (i)–(vi) with multiple bottlenecks has been investigated.

Two typical examples are shown in Figure 7, where the last row of numbers is the bottleneck severity. Although in most cases, Rule 5.1 resulted in correct identification of the primary bottleneck, a number of counterexamples has also been discovered. One of them is shown in Figure 8. Moreover, counterexamples to Rule 5.1 seem to be more numerous than those to Rule 3.1. Nevertheless, since in the majority of cases analyzed Rule 5.1 identified the machine with the largest effect on the production rate, we conclude that it can be used as a tool for the primary bottleneck identification.

At this point in time, we do not have an analytical justification of this Rule.   ∎

6. POTENCY OF THE MATERIAL HANDLING SYSTEMS

Production lines consist of machines and buffers or, more generally, material handling system (MHS). In the framework of model (i)–(vi), the potency of machines is defined by $\min_i p_i$. The potency of the MHS, however, does not seem to have a quantitative
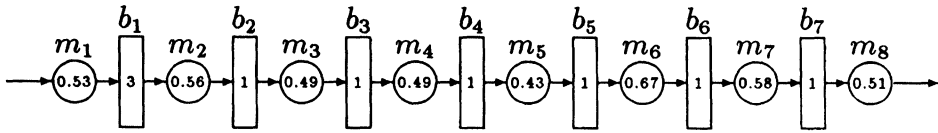
Figure 9 Automotive component production line

characterization. Indeed, buffers capacities, $N_1, \ldots, N_{M-1}$, do not, by themselves, define how efficient MHS is. Based on the bottleneck analysis presented above, such a characterization is introduced below:

DEFINITION 6.1: MHS is **weakly potent** if the machine with the smallest isolation production rate is the bottleneck; otherwise, MHS is **not potent**. MHS is **potent** if it is weakly potent and, in addition, production rate of the system is close to that of the slowest machine in isolation. MHS is **strongly potent** if it is potent and this production is achieved using the smallest total buffer capacity $\sum_{i=1}^{M-1} N_i$.

Verification of the weak potency, non-potency, and potency can be carried out using Rules 3.1 and 5.1. A tool for verification of the strong potency will be addressed elsewhere.

Below we apply the notion of the MHS potency, along with Rule 3.1 and 5.1, to analysis and improvement of a practical production line.

7. APPLICATION AT AN AUTOMOTIVE COMPONENT PLANT

The production system considered in this case study is shown in Figure 9. Parameters of the machines and the buffers identified are shown in the Figure, where the numbers in the circles and rectangles represent, as before, the p_i's and N_i's, respectively (see [8] for details). The system was not instrumented to measure the frequency of blockages and starvations and, therefore, the bottleneck analysis that follows has been carried out in term of the calculated data, mb_i and ms_i.

Base on the parameters shown in Figure 9 and using the recursive procedure (4.2), the production rate of the line has been evaluated to be 0.2068 parts/cycle which is less than 50% of the production rate of the slowest machine (m_5). Thus, the system incurs substantial losses, due to a relative inefficacy on the material handling system. The goal of the continuous improvement project was to recover some of these losses by modifying the MHS.

Table I shows the values of mb_i's and ms_i's calculated according to (4.5). From these

Table I Starvation and blockage of the system without the MHS improvement

i	1	2	3	4	5	6	7	8
ms_i	0	0.0042	0.0815	0.1494	0.1737	0.3978	0.3207	0.2982
mb_i	0.3252	0.3477	0.2365	0.1912	0.0862	0.161	0.1175	0

Bottleneck: m_5; Production rate = 0.2068 parts/cycle

Table II Starvation and blockage of the system with $N_4 = N_5 = 5$

i	1	2	3	4	5	6	7	8
ms_i	0	0.01	0.1083	0.2173	0.1604	0.2511	0.2159	0.2367
mb_i	0.2634	0.2825	0.1392	0.0057	0.0048	0.2408	0.1525	0

Bottlenecks: m_3 ($S_3 = 0.2523$), m_7 ($S_7 = 0.1091$);
Production rate = 0.2683 parts/cycle

data and Rule 3.1, it follows that m_5 is the bottleneck, i.e., the MHS is weakly potent. To protect the bottleneck, increase b_4 and b_5 to N_4 to $N_5 = 5$. The resulting system has the production rate of 0.2683 parts/cycle and the new bottlenecks, as it follows from the data shown in Table II, are m_3 and m_7 with the severities $S_3 = 0.2523$ and $S_7 = 0.1091$. This renders the MHS not potent. Protecting the primary bottleneck, we increase b_2 and b_3 to $N_2 = N_3 = 5$. Resulting system has $PR = 0.3249$ parts/cycle with the bottleneck at m_8 (Table III). Again, the potency is not achieved. Increasing now b_7 to $N_7 = 5$, we, finally, obtain $PR = 0.3882$ parts/cycle and the primary bottleneck at m_5 (Table IV). Since this production rate is within 10% of the slowest machine in isolation and this machine is the primary bottleneck, we conclude that the MHS arrived at is a potent one.

A version of the system designed above has been implemented on the factory floor. A new, robotic MHS has been installed to accommodate the buffering recommendations developed. At present, the line exhibits a satisfactory performance. More details can be found in [8].

8. CONCLUSIONS

Using the largest sensitivity of the system performance index with respect to the isolation production rate of each machine as the definition of a bottleneck, it is shown in this work that the location of the bottleneck in a serial production line can be determined by analyzing frequencies of blockages and starvations of each machine. Since these

Table III Starvation and blockage of the system with $N_2 = N_3 = N_4 = N_5 = 5$

i	1	2	3	4	5	6	7	8
ms_i	0	0.0192	0.0025	0.0065	0.0057	0.0238	0.1032	0.1801
mb_i	0.2071	0.2194	0.1584	0.1597	0.1027	0.3331	0.1847	0

Bottleneck: m_8;
Production rate = 0.3249 parts/cycle

Table IV Starvation and blockage of the system with $N_2 = N_3 = N_4 = N_5 = N_7 = 5$

i	1	2	3	4	5	6	7	8
ms_i	0	0.0369	0.0102	0.0225	0.023	0.1156	0.181	0.1168
mb_i	0.1438	0.1402	0.0884	0.0821	0.0219	0.2008	0.0157	0

Bottlenecks: m_5 ($S_5 = 0.1528$), m_7 ($S_7 = 0.1209$);
Production rate = 0.3882 parts/cycle

frequencies could be either measured (real-time operation environment) or calculated (production system design environment or for systems not instrumented for on-line measurements), this offers a simple and systematic tool for bottleneck identification. For both, single and multiple bottlenecks situations; this rule can be summarized as follows:

Bottleneck Identification Rule: *If the frequency of manufacturing blockage of machine m_i is larger than the frequency of manufacturing starvation of machine m_{i+1} (either measured or calculated), the bottleneck is downstream of machine m_i. If the frequency of the manufacturing starvation of machine m_i is larger than the frequency of the manufacturing blockage of m_{i-1}, the bottleneck is upstream of machine m_i. If, according to this rule, there exist multiple bottlenecks, the primary one is the bottleneck with the largest Severity.*

Although this rule is to a certain extend justified in this work, both numerically (Section 3 and 5) and analytically (Sections 4), additional research is needed to more rigorously quantify its properties, especially in the case of multiple bottlenecks.

APPENDIX A. PROOFS FOR SECTION 2

Proof of Lemma 2.1: The dynamics of buffer occupancy in system (i)–(vi) with $M = 2$ can be described by an irreducible, ergodic Markov chain with states $0, 1,\ldots, N_1$ and transition matrix:

$$P = \begin{bmatrix} 1 - p_1 & (1 - p_1)p_2 & & & \\ p_1 & p_1p_2 + (1 - p_1)(1 - p_2) & \cdot & & \\ & p_1(1 - p_2) & \cdot & \cdot & \\ & & \cdot & \cdot & \\ & & \cdot & p_1p_2 + (1 - p_1)(1 - p_2) & (1 - p_1)p_2 \\ & & & p_1(1 - p_2) & 1 - p_2 + p_1p_2 \end{bmatrix}$$

As it has been shown in [1], the invariant measure of this Markov chain has the following form:

$$\text{Prob}(h = j) =: X(j) = \frac{\alpha^j}{\alpha + \cdots + \alpha^{N_1} + 1 - p_2}, \quad 1 \le j \le N_1,$$

where h is the buffer occupancy at $t = \infty$ and $\alpha = p_1(1 - p_2) / p_2(1 - p_1)$. Therefore, after some algebra,

$$X(0) = \begin{cases} \dfrac{(1 - p_1)(1 - \alpha)}{1 - \dfrac{p_1}{p_2} \alpha^{N_1}}, & p_1 \neq p_2, \\[4mm] \dfrac{1 - p_1}{N_1 + 1 - p_1}, & p_1 = p_2, \end{cases}$$

$$X(N_1) = \begin{cases} \dfrac{1 - \dfrac{1}{\alpha}}{1 - \dfrac{p_2}{p_1}(\dfrac{1}{\alpha})^{N_1}}, & p_1 \neq p_2, \\[4mm] \dfrac{1}{N_1 + 1 - p_1}, & p_1 = p_2. \end{cases}$$

Thus, the probabilities of manufacturing blockage and manufacturing starvation are:

$$\widetilde{mb}_1 = \text{Prob}\ (\{m_1 \text{ is up during a time slot}\}$$

$$\cap\ \{b_1 \text{ is full at the beginning of this slot}\}$$

$$\cap\ \{m_2 \text{ fails to take a part from } b_1 \text{ at the beginning of this slot}\})$$

$$= p_1\, X\,(N_1)(1 - p_2)$$

$$= \begin{cases} p_1\dfrac{(1 - p_2)(1 - \dfrac{1}{\alpha})}{1 - \dfrac{p_2}{p_1}(\dfrac{1}{\alpha})^{N_1}}, & p_1 \neq p_2 \\[4mm] p_1\dfrac{1 - p_1}{N_1 + 1 - p_1}, & p_1 = p_2 \end{cases}$$

$$= p_1 Q(p_2, p_1, N_1),$$

$$\widetilde{ms}_2 = \text{Prob}\ (\{m_2 \text{ is up during a time slot}\}$$

$$\cap\ \{b_1 \text{ is empty at the beginning of this slot}\})$$

$$= p_2 X(0)$$

$$= \begin{cases} p_2\dfrac{(1 - p_1)(1 - \alpha)}{1 - \dfrac{p_1}{p_2}\alpha^{N_1}}, & p_1 \neq p_2 \\[4mm] p_2\dfrac{1 - p_1}{N_1 + 1 - p_1}, & p_1 = p_2 \end{cases}$$

$$= p_2 Q(p_1, p_2, N_1),$$

where $Q(x, y, N)$ is defined in (2.2).

To prove Theorem 2.1, we need the following three lemmas: ∎

LEMMA A.1: *Let $f_i(p_1, p_2, N_1)$ be the sensitivity of $\widetilde{PR}$ with respect to p_i, i.e., $f_i(p_1, p_2, N_1)$ $= \partial\widetilde{PR}(p_1, p_2, N_1)/\partial p_i$, $i = 1, 2$. Then functions $f_1(p_1, p_2, N_1)$ and $f_2(p_1, p_2, N_1)$ are monotonically decreasing and increasing in p_1, respectively.*

Proof: Recall [1] that

$$\widetilde{PR}(p_1, p_2, N_1) = p_2[1 - Q(p_1, p_2, N_1)] = p_1[1 - Q(p_2, p_1, N_1)] \tag{A.1}$$

Hence,

$$f_1(p_1, p_2, N_1) = \frac{\partial\widetilde{PR}(p_1, p_2, N_1)}{\partial p_1} = -p_2 \frac{\partial Q(p_1, p_2, N_1)}{\partial p_1}. \tag{A.2}$$

From equation (2.2), function $Q(p_1, p_2, N_1)$ can be re-written as follows:

$$Q(p_1, p_2, N_1) = \frac{1 - p_2}{\alpha + \cdots + \alpha^{N_1} + 1 - p_2},$$

where α is given in (2.2). Then

$$\frac{\partial Q(p_1, p_2, N_1)}{\partial p_1} = \frac{-(1 + 2\alpha + \cdots + N_1\alpha^{N_1-1})\dfrac{(1 - p_2)^2}{p_2(1 - p_1)^2}}{[\alpha + \cdots + \alpha^{N_1} + (1 - p_2)]^2}.$$

After some algebra, this simplifies to

$$\frac{\partial Q(p_1, p_2, N_1)}{\partial p_1} = \begin{cases} \dfrac{-1 + \alpha^{N_1} + N_1(1 - \alpha)\alpha^{N_1}}{p_2\left(1 - \dfrac{p_1}{p_2}\alpha^{N_1}\right)^2}, & p_1 \neq p_2, \\[4mm] \dfrac{-N_1(N_1 + 1)}{2p_1(N_1 + 1 - p_1)^2}, & p_1 = p_2. \end{cases} \tag{A.3}$$

Therefore, from (A.2) and (A.3),

$$f_1(p_1, p_2, N_1) = \begin{cases} \dfrac{(1 - \alpha^{N_1}) - N_1(1 - \alpha)\alpha^{N_1}}{\left(1 - \dfrac{p_1}{p_2}\alpha^{N_1}\right)^2}, & p_1 \neq p_2, \\[4mm] \dfrac{N_1^2 + N_1}{2(N_1 + 1 - p_1)^2}, & p_1 = p_2 \end{cases} \tag{A.4}$$

Analogously, it can be shown that

$$f_2(p_1, p_2, N_1) = \begin{cases} \dfrac{[1 - (\frac{1}{\alpha})^{N_1}] - N_1 (1 - \frac{1}{\alpha}) (\frac{1}{\alpha})^{N_1}}{(1 - \frac{p_2}{p_1} (\frac{1}{\alpha})^{N_1})^2}, & p_1 \neq p_2, \\[1.5em] \dfrac{N_1^2 + N_1}{2(N_1 + 1 - p_1)^2}, & p_1 = p_2 \end{cases} \qquad (A.5)$$

Consider now the two cases, $p_1 \neq p_2$ and $p_1 = p_2$, separately. For the case $p_1 \neq p_2$, we write:

$f_1(p_1, p_2, N_1)$

$$= \frac{(1 - \alpha)(1 + \alpha + \alpha^2 + \ldots + \alpha^{N_1 - 1} - N_1\alpha^{N_1})}{(1 - \frac{p_1}{p_2}\alpha^{N_1})^2}$$

$$= \frac{1}{\dfrac{(1 - \frac{p_1}{p_2}\alpha^{N_1})}{(1 - \alpha)} \dfrac{(1 - \frac{p_1}{p_2}\alpha^{N_1})}{(1 + \alpha + \alpha^2 + \ldots + \alpha^{N_1 - 1} - N_1 \alpha^{N_1})}}$$

$$= \frac{1}{\dfrac{[1 - \alpha^{N_1} + (1 - \frac{p_1}{p_2})\alpha^{N_1}]}{(1 - \alpha)} \dfrac{[1 - \alpha^{N_1} + (1 - \frac{p_1}{p_2})\alpha^{N_1}]}{(1 - \alpha^{N_1}) + (\alpha - \alpha^{N_1}) + \ldots + (\alpha^{N_1 - 1} - \alpha^{N_1})}}$$

$$= \frac{1}{\left(1 + \alpha + \ldots + \alpha^{N_1 - 1} + \dfrac{\frac{(p_2 - p_1)}{p_2}}{\frac{p_2(1 - p_1) - p_1(1 - p_2)}{p_2(1 - p_1)}}\alpha^{N_1}\right) \dfrac{[1 - \alpha^{N_1} + (1 - \frac{p_1}{p_2})\alpha^{N_1}]}{(1 - \alpha)(1 + \alpha + \ldots + \alpha^{N_1 - 1}) + \ldots + (1 - \alpha)\alpha^{N_1 - 1}}}$$

$$= \frac{1 + 2\alpha + 3\alpha^2 + \ldots + N_1\alpha^{N_1 - 1}}{[1 + \alpha + \alpha^2 + \ldots + \alpha^{N_1 - 2} + (1 + \frac{p_1(1 - p_2)}{p_2})\alpha^{N_1 - 1}]^2}$$

$$= \frac{1 + 2\alpha + 3\alpha^2 + \ldots + N_1\alpha^{N_1 - 1}}{1 + 2\alpha + \ldots + (N_1 - 1)\alpha^{N_1 - 2} + (N_1 + 2A)\alpha^{N_1 - 1} + \ldots + (2 + 2A)\alpha^{2N_1 - 3} + (1 + A)^2\alpha^{2N_1 - 2}}$$

$$= \frac{k(\alpha)}{k(\alpha) + l(\alpha, A)}, \qquad (A.6)$$

where

$$A = \frac{p_1(1 - p_2)}{p_2},$$

$$k(\alpha) = 1 + 2\alpha + 3\alpha^2 + \ldots + N_1\alpha^{N_1 - 1},$$

$$l(\alpha, A) = \begin{cases} 2A\alpha^{N_1-1} + \ldots + (2 + 2A)\alpha^{2N_1-3} + (1 + A)^2\alpha^{2N_1-2}, & N_1 \geq 2, \\ 2A + A^2, & N_1 = 1. \end{cases}$$

For an arbitrary $p_1 \in (0, 1)$, choose $\bar{p}_1$ so that $\bar{p}_1 > p_1$. Then $\bar{\alpha} = \dfrac{\bar{p}_1(1 - p_2)}{p_2(1 - \bar{p}_1)} > \alpha$ and

$\bar{A} = \dfrac{\bar{p}_1(1 - p_2)}{p_2} > A$. Therefore, for $N_1 = 1$,

$$f_1(\bar{p}_1, p_2, N_1) - f_1(p_1, p_2, N_1) = \frac{1}{1 + \bar{A} + \bar{A}^2} - \frac{1}{1 + A + A^2} < 0.$$

For $N_1 \geq 2$,

$$f_1(\bar{p}_1, p_2, N_1) - f_1(p_1, p_2, N_1)$$

$$= \frac{k(\bar{\alpha})}{k(\bar{\alpha}) + l(\bar{\alpha}, \bar{A})} - \frac{k(\alpha)}{k(\alpha) + l(\alpha, A)}$$

$$= \frac{k(\bar{\alpha})l(\alpha, A) - k(\alpha)l(\bar{\alpha}, \bar{A})}{[k(\bar{\alpha}) + l(\bar{\alpha}, \bar{A})][k(\alpha) + l(\alpha, A)]}. \tag{A.6}$$

The numerator of (A.7) can be re-written as

$$k(\bar{\alpha})l(\alpha, A) - k(\alpha)l(\bar{\alpha}, \bar{A})$$

$$= (1 + 2\bar{\alpha} + \ldots + N_1\bar{\alpha}^{N_1-1}) [2A\alpha^{N_1-1} + \ldots + (1 + A)^2 \alpha_1^{2N_1-2}]$$

$$- (1 + 2\alpha + \ldots + N_1\alpha^{N_1-1}) [2\bar{A}\bar{\alpha}^{N_1-1} + \ldots + (1 + \bar{A})^2 \bar{\alpha}^{2N_1-2}]$$

$$= (\alpha^{N_1-1} + 2\bar{\alpha}\alpha^{N_1-1} + \ldots + N_1\bar{\alpha}^{N_1-1} \alpha^{N_1-1}) [2A + \ldots + (1 + A)^2 \alpha^{N_1-1}]$$

$$- (\bar{\alpha}^{N_1-1} + 2\bar{\alpha}^{N_1-1} \alpha + \ldots + N_1\bar{\alpha}^{N_1-1} \alpha^{N_1-1}) [2\bar{A} + \ldots + (1 + \bar{A})^2 \bar{\alpha}^{N_1-1}]. \tag{A.7}$$

Since $\bar{\alpha} > \alpha$ and $\bar{A} > A$,

$$(\alpha^{N_1-1} + 2\bar{\alpha}\alpha^{N_1-1} + \ldots + N_1\bar{\alpha}^{N_1-1} \alpha^{N_1-1}) < (\bar{\alpha}^{N_1-1} + 2\bar{\alpha}^{N_1-1} \alpha + \ldots + N_1\bar{\alpha}^{N_1-1} \alpha^{N_1-1})$$

and

$$[2A + \ldots + (1 + A)^2 \alpha_1^{N_1-1}] < [2\bar{A} + \ldots + (1 + \bar{A})^2\bar{\alpha}^{N_1-1}].$$

Hence,

$$k(\overline{\alpha})l(\alpha, A) - k(\alpha)l(\overline{\alpha}, \overline{A}) < 0,$$

and, therefore,

$$f_1(\overline{p}_1, p_2, N_1) - f_1(p_1, p_2, N_1) < 0,$$

i.e., $f_1(p_1, p_2, N_1)$ is monotonically decreasing in p_1 if $p_1 \neq p_2$.

For the case $p_1 = p_2$, expression (A.6) holds as well, since in this case $\alpha = 1$. Hence, $f_1(p_1, p_2, N_1)$ is again monotonically decreasing in p_1.

The monotonicity property of $f_2(p_1, p_2, N_1)$ with respect to p_1 is proved analogously.

∎

LEMMA A.2: *Functions* $\widetilde{mb}_1$ *and* $\widetilde{ms}_2$ *defined by (2.1) are monotonically increasing and decreasing in* p_1, *respectively.*

Proof: By Lemma 2.1, we have

$$\widetilde{mb}_1 = p_1 Q(p_2, p_1, N_1).$$

Since $Q(p_2, p_1, N_1)$ is monotonically increasing in p_1 (see [1]), function $\widetilde{mb}_1$ is also monotonically increasing in p_1. Similar argument is used to prove that $\widetilde{ms}_2$ is monotonically decreasing in p_1.

∎

LEMMA A.3: *For the case* $p_1 = p_2$,

$$\widetilde{mb}_1 = \widetilde{ms}_2 \; and \, f_1(p_2, p_2, N_1) = f_2(p_2, p_2, N_1).$$

Proof: Follows directly from (2.2), (A.4) and (A.5).

∎

Proof of Theorem 2.1: Denote $\widetilde{mb}_1 = \widetilde{mb}_1(p_1, p_2, N_1)$ and $\widetilde{ms}_2 = \widetilde{ms}_2(p_1, p_2, N_1)$. Suppose $\widetilde{mb}_1 < \widetilde{ms}_2$. Then we show, by contradiction, that $p_1 < p_2$. Indeed, assume $p_1 \geq p_2$. Then, by Lemmas A.2 and A.3,

$$\widetilde{mb}_1(p_1, p_2, N_1) \geq \widetilde{mb}_1(p_2, p_2, N_1) = \widetilde{ms}_2(p_2, p_2, N_1) \geq \widetilde{ms}_2(p_1, p_2, N_1),$$

which contradicts the assumption. Therefore, $p_1 < p_2$.

Since $p_1 < p_2$, using Lemmas A.1 and A.3, we have:

$$f_1(p_1, p_2, N_1) > f_1(p_2, p_2, N_1) = f_2(p_2, p_2, N_1) > f_2(p_1, p_2, N_1),$$

i.e., $\partial \widetilde{PR}(p_1, p_2, N_1)/\partial p_1 > \partial \widetilde{PR}(p_1, p_2, N_1)/\partial p_2$. This proves the "if" part. The "only if" part is proved by assuming $\partial \widetilde{PR}(p_1, p_2, N_1)/\partial p_1 > \partial \widetilde{PR}(p_1, p_2, N_1)/\partial p_2$ and using arguments analogous to the above.

∎

Proof of Corollary 2.1: From (A.1) and Lemma 2.1, we have:

$$\widetilde{PR}(p_1, p_2, N_1) = p_2 - \widetilde{ms}_2 = p_1 - \widetilde{mb}_1.$$

Therefore, the inequality $p_1 < p_2$ implies and is implied by the inequality $\widetilde{mb}_1 < \widetilde{ms}_2$. Similarly, $p_1 > p_2$ implies and is implied by $\widetilde{mb}_1 > \widetilde{ms}_2$. Thus, by Theorem 2.1,

$$\frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_1} > \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_2} \qquad (\text{or } \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_1} < \frac{\partial \widetilde{PR}(p_1, p_2, N_1)}{\partial p_2})$$

takes place if and only if

$$p_1 < p_2, \qquad (\text{or } p_1 > p_2, \text{ respectively})$$

i.e., the bottlenecks identified by definitions (a) and (1.1) are the same.

∎

APPENDIX B. PROOFS FOR SECTION 4

Proof of Lemma 4.2: From equation (4.1), we have:

$$\widetilde{mb}_i = p_i \sum_{j=i+1}^{M} \left(\prod_{r=i+1}^{j-1} p_r \right) (1 - p_j) X_{i,\ldots,j-1} (N_i, \ldots, N_{j-1}), i = 1, \ldots, M - 1.$$

By Lemma A.6 of [1], $\widetilde{mb}_i$ can be expressed in terms of $\tilde{p}_i^b$ as

$$\widetilde{mb}_i = p_i - \tilde{p}_i^b + \mathbb{O}(\delta), i = 1, \ldots, M - 1.$$

Using Lemma 4.1,

$$\widetilde{mb}_i = p_i - p_i^b + \mathcal{O}(\delta), \ i = 1, \dots, M - 1.$$

Thus, from the steady state equation of recursive procedure (4.2),

$$
\begin{aligned}
\widetilde{mb}_i &= p_i - p_i[1 - Q(p_{i+1}^b, p_i^f, N_i)] + \mathcal{O}(\delta) \\
&= p_i Q(p_{i+1}^b, p_i^f, N_i) + \mathcal{O}(\delta) \\
&= mb_i + \mathcal{O}(\delta), \ i = 1, \dots, M - 1.
\end{aligned}
$$

Therefore,

$$|\widetilde{mb}_i - mb_i| \sim \mathcal{O}(\delta), \ i = 1, \dots, M - 1.$$

Analogously, from equation (4.1), $\widetilde{ms}_i$ can be written as

$$\widetilde{ms}_i = p_i X_{i-1}(0), \ i = 2, \dots, M.$$

By Lemma A.6 of [1] and Lemma 4.1, this can be approximated as

$$
\begin{aligned}
\widetilde{ms}_i &= p_i - \tilde{p}_i^f + \mathcal{O}(\delta) \\
&= p_i - p_i^f + \mathcal{O}(\delta), \ i = 2, \dots, M.
\end{aligned}
$$

Using the steady state equations of recursive procedure (4.2), we have:

$$
\begin{aligned}
\widetilde{ms}_i &= p_i - p_i[1 - Q(p_{i-1}^f, p_i^b, N_{i-1})] + \mathcal{O}(\delta) \\
&= p_i Q(p_{i-1}^f, p_i^b, N_{i-1}) + \mathcal{O}(\delta) \\
&= ms_i + \mathcal{O}(\delta), \ i = 2, \dots, M.
\end{aligned}
$$

Hence, we conclude that

$$|\widetilde{ms}_i - ms_i| \sim \mathcal{O}(\delta), \ i = 2, \dots, M.$$

Proof of Lemma 4.3: Under Hypothesis H.1,   ∎

$$mb_{j-1} > ms_j, \ \forall j \leq i.$$

Under this condition, we show, by contradiction, that there exists $N_1^* < \infty$ such that $p_{j-1}^f > p_j^b$, for all $N_{j-1} > N_1^*$, $\forall j \le i$. Assume that, for all $N_1^* < \infty$, there exists $N_{j-1} > N_1^*$, $\forall j \le i$, such that $p_{j-1}^f \le p_j^b$. From (2.2) and (4.5),

$$\lim_{N_{j-1}\to\infty} mb_{j-1} = \lim_{N_{j-1}\to\infty} p_{j-1} Q(p_j^b, p_{j-1}^f, N_{j-1}).$$

$$= \begin{cases} \lim_{N_{j-1}\to\infty} p_{j-1} \dfrac{(1-p_j^b)(1-\dfrac{1}{\alpha})}{1-\dfrac{p_j^b}{p_{j-1}^f}(\dfrac{1}{\alpha})^{N_{j-1}}}, & p_j^b > p_{j-1}^f \\[4ex] \lim_{N_{j-1}\to\infty} p_{j-1} \dfrac{1-p_j^b}{N_{j-1}+1-p_j^b}, & p_j^b = p_{j-1}^f \end{cases}$$

$$0,$$

$$\lim_{N_{j-1}\to\infty} ms_j = \lim_{N_{j-1}\to\infty} p_j Q(p_{j-1}^f, p_j^b, N_{j-1})$$

$$= \begin{cases} \lim_{N_{j-1}\to\infty} p_j \dfrac{(1-p_{j-1}^f)(1-\alpha)}{1-\dfrac{p_{j-1}^f}{p_j^b}\alpha^{N_{j-1}}}, & p_j^b > p_{j-1}^f \\[4ex] \lim_{N_{j-1}\to\infty} p_j \dfrac{1-p_j^b}{N_{j-1}+1-p_j^b}, & p_j^b = p_{j-1}^f \end{cases}$$

$$= \begin{cases} p_j \dfrac{p_j^b - p_{j-1}^f}{p_j^b} > 0, & p_j^b > p_{j-1}^f, \\[3ex] 0, & p_j^b = p_{j-1}^f, \end{cases}$$

where

$$\alpha = \frac{p_{j-1}^f(1-p_j^b)}{p_j^b(1-p_{j-1}^f)} < 1.$$

Therefore, there exists $N_{j-1} > N_1^*$, $\forall j \le i$, such that

$$mb_{j-1} \le ms_j,$$

which contradicts the assumption.

Choose N_1^* large enough so that $p_{j-1}^f > p_j^b$, for all $N_{j-1} > N_1^*$, $\forall j \le i$. Then,

$$\lim_{N_{j-1} \to \infty} Q(p_{j-1}^f, p_j^b, N_{j-1}) = \frac{(1 - p_{j-1}^f)(1 - \alpha)}{1 - \dfrac{p_{j-1}^f}{p_j^b} \alpha^{N_{j-1}}} = 0,$$

where

$$\alpha = \frac{p_{j-1}^f(1 - p_j^b)}{p_j^b(1 - p_{j-1}^f)} > 1.$$

Thus, for any $0 < \epsilon_{10} \ll 1$, there exists $N_1^{**} > N_1^*$ such that, for all $N_{j-1} > N_1^{**}$, $\forall j \leq i$,

$$Q(p_{j-1}^f, p_j^b, N_{j-1}) = \epsilon_{j1} < \epsilon_{10}.$$

Similarly, for $j \geq i$, we can conclude that, for any $0 < \epsilon_{20} \ll 1$, there exists $N_2^{**} > N_2^*$ such that, for all $N_j > N_2^{**}$,

$$Q(p_{j+1}^b, p_j^f, N_j) = \epsilon_{j2} < \epsilon_{20}.$$

Therefore, for given ϵ_0, choose $N^* = \max(N_1^{**}, N_2^{**})$ so that

$$\epsilon = \max_{j,s}(\epsilon_{j1}, \epsilon_{s2}) < \epsilon_0.$$

$\blacksquare$

The proof of Theorem 4.1 requires the following four lemmas:

LEMMA B.1: *Let $g_1(x, y, N)$ be the sensitivity of $Q(x, y, N)$ with respect to x and $g_2(x, y, N)$ be the sensitivity of $Q(x, y, N)$ with respect to y, i.e., $g_1(x, y, N) = \partial Q(x, y, N)/\partial x$ and $g_2(x, y, N) = \partial Q(x, y, N)/\partial y$. Then*

$$g_1(x, y, N) = \begin{cases} \dfrac{-1 + \alpha^N + N \dfrac{y - x}{y(1 - x)} \alpha^N}{y(1 - \dfrac{x}{y} \alpha^N)^2}, & x \neq y, \\[20pt] \dfrac{-N(N + 1)}{2x(N + 1 - x)^2}, & x = y, \end{cases} \tag{B.1}$$

$$g_2(x, y, N) = \begin{cases} \dfrac{x - [x + N \dfrac{x(y - x)}{y(1 - y)}]\alpha^N}{y^2(1 - \dfrac{x}{y} \alpha^N)^2}, & x \neq y, \\[20pt] \dfrac{N(N + 1 - 2x)}{2x(N + 1 - x)^2}, & x = y, \end{cases} \tag{B.2}$$

where α is defined in (2.2).

Proof: Follows directly from (A.3).

LEMMA B.2: *Introduce*

$$k_{1j} = \frac{\partial p_{j-1}^f}{\partial p_j}, j = 2, \ldots, M, \ k_{2j} = \frac{\partial p_j^b}{\partial p_j}, j = 2, \ldots, M,$$ (B.3)

$$k_{3j} = \frac{\partial p_{j+1}^b}{\partial p_j}, j = 1, \ldots, M - 1, \ k_{4j} = \frac{\partial p_j^f}{\partial p_j}, j = 1, \ldots, M - 1,$$

where $p_j^f, p_j^b, j = 1, \ldots, M$, are the steady states of recursive procedure (4.2). Then

$$k_{31} = V_2, \ k_{41} = 1, \ k_{12} = 0,$$

$$k_{22} = \frac{V_{23}}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] - \frac{V_{23} p_2 g_2(p_3^b, p_2^f, N_2) + V_{13} p_2 g_1(p_3^b, p_2^f, N_2)}{B_2} [1 - Q(p_1, p_2^b, N_1)],$$

$$k_{32} = -\frac{V_{13} p_2 g_2(p_1, p_2^b, N_1)}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] + \frac{V_{13}}{B_2} [1 - Q(p_1, p_2^b, N_1)],$$

$$k_{42} = -\frac{V_{23} p_2 g_2(p_1, p_2^b, N_1)}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] + \frac{V_{23}}{B_2} [1 - Q(p_1, p_2^b, N_1)],$$

$$k_{1j} = \frac{U_{1,j-1} V_{2,j+1}}{B_j} [1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- \frac{U_{1,j-1} V_{2,j+1} p_j g_2(p_{j+1}^b, p_j^f, N_j) + U_{1,j-1} V_{1,j+1} p_j g_1(p_{j+1}^b, p_j^f, N_j)}{B_j} [1 - Q(p_{j-1}^f, p_j^b, N_{j-1})],$$

$$j = 3, \ldots, M - 2,$$

$$k_{2j} = \frac{U_{2,j-1} V_{2,j+1}}{B_j} [1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- \frac{U_{2,j-1} V_{2,j+1} p_j g_2(p_{j+1}^b, p_j^f, N_j) + U_{2,j-1} V_{1,j+1} p_j g_1(p_{j+1}^b, p_j^f, N_j)}{B_j} [1 - Q(p_{j-1}^f, p_j^b, N_{j-1})],$$

$$j = 3, \ldots, M - 2,$$ (B.4)

$$k_{3j} = -\frac{U_{2,j-1}V_{1,j+1}p_j g_2(p_{j-1}^f, p_j^b, N_{j-1}) + U_{1,j-1}V_{1,j+1}p_j g_1(p_{j-1}^f, p_j^b, N_{j-1})}{B_j}[1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$+ \frac{U_{2,j-1}V_{1,j+1}}{B_j}[1 - Q(p_{j-1}^f, p_j^b, N_{j-1})],$$

$$j = 3, \dots, M-2,$$

$$k_{4j} = -\frac{U_{2,j-1}V_{2,j+1}p_j g_2(p_{j-1}^f, p_j^b, N_{j-1}) + U_{1,j-1}V_{2,j+1}p_j g_1(p_{j-1}^f, p_j^b, N_{j-1})}{B_j}[1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$+ \frac{U_{2,j-1}V_{2,j+1}}{B_j}[1 - Q(p_{j-1}^f, p_j^b, N_{j-1})],$$

$$j = 3, \dots, M-2,$$

$$k_{1,M-1} = \frac{U_{1,M-2}}{B_{M-1}}[1 - Q(p_M, p_{M-1}^f, N_{M-1})]$$

$$- \frac{U_{1,M-2}p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1})}{B_{M-1}}[1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})],$$

$$k_{2,M-1} = \frac{U_{2,M-2}}{B_{M-1}}[1 - Q(p_M, p_{M-1}^f, N_{M-1})]$$

$$- \frac{U_{2,M-2}p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1})}{B_{M-1}}[1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})],$$

$$k_{3,M-1} = 0,$$

$$k_{4,M-1} = -\frac{U_{2,M-2}p_{M-1}g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2}) + U_{1,M-2}p_{M-1}g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{B_{M-1}}$$

$$[1 - Q(p_M, p_{M-1}^f, N_{M-1})]$$

$$+ \frac{U_{2,M-2}}{B_{M-1}}[1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})],$$

$$k_{1M} = U_{M-1}, \quad k_{2M} = 1,$$

where $g_1(x, y, N)$ *and* $g_2(x, y, N)$ *are given in (B.1) and (B.2),*

$$U_1 = 0,$$

$$U_{1j} = [U_{j-1}p_j g_1(p_{j-1}^f, p_j^b, N_{j-1}) + p_j g_2(p_{j-1}^f, p_j^b, N_{j-1})][p_j g_1(p_{j+1}^b, p_j^f, N_j)], j = 2, \dots, M-1,$$

$$U_{2j} = 1 - [U_{j-1}p_jg_1(p_{j-1}^f, p_j^b, N_{j-1}) + p_jg_2(p_{j-1}^f, p_j^b, N_{j-1})][p_jg_2(p_{j+1}^b, p_j^f, N_j)], j = 2, \ldots, M - 1,$$

$$U_j = \frac{U_{1j}}{U_{2j}}, j = 2, \ldots, M - 1,$$

$$V_M = 0,$$

$$V_{1j} = [V_{j+1}p_jg_1(p_{j+1}^b, p_j^f, N_j) + p_jg_2(p_{j+1}^b, p_j^f, N_j)][p_jg_1(p_{j-1}^f, p_j^b, N_{j-1})], j = 2, \ldots, M - 1,$$

$$V_{2j} = 1 - [V_{j+1}p_jg_1(p_{j+1}^b, p_j^f, N_j) + p_jg_2(p_{j+1}^b, p_j^f, N_j)][p_jg_2(p_{j-1}^f, p_j^b, N_{j-1})], j = 2, \ldots, M - 1,$$

$$V_j = \frac{V_{1j}}{V_{2j}}, j = 2, \ldots, M - 1, \tag{B.5}$$

and

$$B_2 = V_{23} - V_{13}p_2^2g_2(p_1, p_2^b, N_1)g_1(p_3^b, p_2^f, N_2) - V_{23}p_2^2g_2(p_1, p_2^b, N_1)g_2(p_3^b, p_2^f, N_2),$$

$$B_j = U_{2,j-1}[V_{2,j+1} - V_{1,j+1}p_j^2g_2(p_{j-1}^f, p_j^b, N_{j-1})g_1(p_{j+1}^b, p_j^f, N_j)$$

$$- V_{2,j+1}p_j^2g_2(p_{j-1}^f, p_j^b, N_{j-1})g_2(p_{j+1}^b, p_j^f, N_j)]$$

$$+ U_{1,j-1}[-V_{1,j+1}p_j^2g_1(p_{j-1}^f, p_j^b, N_{j-1})g_1(p_{j+1}^b, p_j^f, N_j)$$

$$- V_{2,j+1}p_j^2g_1(p_{j-1}^f, p_j^b, N_{j-1})g_2(p_{j+1}^b, p_j^f, N_j)], j = 3, \ldots, M - 2,$$

$$B_{M-1} = U_{2,M-2} - U_{1,M-2}p_{M-1}^2g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2})g_2(p_M, p_{M-1}^f, N_{M-1})$$

$$- U_{2,M-2}p_{M-1}^2g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2})g_2(p_M, p_{M-1}^f, N_{M-1}). \tag{B.6}$$

Proof: The expression for k_{31} is obtained using the induction: Introduce $\partial p_s^b/\partial p_1$, $s = 2, \ldots,$ $M - 1$. For $s = M - 1$, we have

$$\frac{\partial p_{M-1}^b}{\partial p_1} = \frac{\partial p_{M-1}[1 - Q(p_M^b, p_{M-1}^f, N_{M-1})]}{\partial p_1}$$

$$= -p_{M-1}\frac{\partial Q(p_M^b, p_{M-1}^f, N_{M-1})}{\partial p_M^b}\frac{\partial p_M^b}{\partial p_1} - p_{M-1}\frac{\partial Q(p_M^b, p_{M-1}^f, N_{M-1})}{\partial p_{M-1}^f}\frac{\partial p_{M-1}^f}{\partial p_1}$$

$$= -p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1})\frac{\partial p_{M-1}^f}{\partial p_1}$$

$$= -p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1})\frac{\partial p_{M-1}[1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})]}{\partial p_1}$$

$$= -p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1})$$

$$[-p_{M-1}\frac{\partial Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{\partial p_{M-2}^f}\frac{\partial p_{M-2}^f}{\partial p_1} - p_{M-1}\frac{\partial Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{\partial p_{M-1}^b}\frac{\partial p_{M-1}^b}{\partial p_1}]$$

$$= p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2}) \frac{\partial p_{M-2}^f}{\partial p_1}$$

$$+ p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2}) \frac{\partial p_{M-1}^b}{\partial p_1},$$

where $g_1(x, \dot{y}, N)$ and $g_2(x, y, N)$ are given by (B.1) and (B.2), respectively. Thus,

$$\frac{\partial p_{M-1}^b}{\partial p_1} = \frac{p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{1 - p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2})} \frac{\partial p_{M-2}^f}{\partial p_1}$$

$$= \frac{V_{1,M-1}}{V_{2,M-1}} \frac{\partial p_{M-2}^f}{\partial p_1}$$

$$= V_{M-1} \frac{\partial p_{M-2}^f}{\partial p_1}, \tag{B.7}$$

where $V_{1,M-1}$, $V_{2,M-1}$ and V_{M-1} are defined in (B.5). Assume that for $2 < s < M - 1$,

$$\frac{\partial p_s^b}{\partial p_1} = V_s \frac{\partial p_{s-1}^f}{\partial p_1}, \tag{B.8}$$

where V_s, $2 < s < M - 1$, are given by (B.5). Then,

$$\frac{\partial p_{s-1}^b}{\partial p_1} = \frac{\partial p_{s-1}[1 - Q(p_s^b, p_{s-1}^f, N_{s-1})]}{\partial p_1}$$

$$= -p_{s-1} \frac{\partial Q(p_s^b, p_{s-1}^f, N_{s-1})}{\partial p_s^b} \frac{\partial p_s^b}{\partial p_1} - p_{s-1} \frac{\partial Q(p_s^b, p_{s-1}^f, N_{s-1})}{\partial p_{s-1}^f} \frac{\partial p_{s-1}^f}{\partial p_1}$$

$$= -p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) \frac{\partial p_s^b}{\partial p_1} - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1}) \frac{\partial p_{s-1}^f}{\partial p_1}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})] \frac{\partial p_{s-1}^f}{\partial p_1}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1})$$

$$\quad - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})] \frac{\partial p_{s-1}[1 - Q(p_{s-2}^f, p_{s-1}^b, N_{s-2})]}{\partial p_1}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})]$$

$$\left[-p_{s-1} \frac{\partial Q(p^f_{s-2}, p^b_{s-1}, N_{s-2})}{\partial p^f_{s-2}} \frac{\partial p^f_{s-2}}{\partial p_1} - p_{s-1} \frac{\partial Q(p^f_{s-2}, p^b_{s-1}, N_{s-2})}{\partial p^b_{s-1}} \frac{\partial p^b_{s-1}}{\partial p_1} \right]$$

$$= [V_s p_{s-1} g_1(p^b_s, p^f_{s-1}, N_{s-1}) + p_{s-1} g_2(p^b_s, p^f_{s-1}, N_{s-1})][p_{s-1} g_1(p^f_{s-2}, p^b_{s-1}, N_{s-2})] \frac{\partial p^f_{s-2}}{\partial p_1}$$

$$+ [V_s p_{s-1} g_1(p^b_s, p^f_{s-1}, N_{s-1})$$

$$+ p_{s-1} g_2(p^b_s, p^f_{s-1}, N_{s-1})][p_{s-1} g_2(p^f_{s-2}, p^b_{s-1}, N_{s-2})] \frac{\partial p^b_{s-1}}{\partial p_1}.$$

Hence, using notations (B.5),

$$\frac{\partial p^b_{s-1}}{\partial p_1} = \frac{[V_s p_{s-1} g_1(p^b_s, p^f_{s-1}, N_{s-1}) + p_{s-1} g_2(p^b_s, p^f_{s-1}, N_{s-1})][p_{s-1} g_1(p^f_{s-2}, p^b_{s-1}, N_{s-2})]}{1 - [V_s p_{s-1} g_1(p^b_s, p^f_{s-1}, N_{s-1}) + p_{s-1} g_2(p^b_s, p^f_{s-1}, N_{s-1})][p_{s-1} g_2(p^f_{s-2}, p^b_{s-1}, N_{s-2})]} \frac{\partial p^f_{s-2}}{\partial p_1}$$

$$= \frac{V_{1,s-1}}{V_{2,s-1}} \frac{\partial p^f_{s-2}}{\partial p_1}$$

$$= V_{s-1} \frac{\partial p^f_{s-2}}{\partial p_1}. \tag{B.9}$$

Thus, the inductive hypothesis is established, and therefore,

$$\frac{\partial p^b_s}{\partial p_1} = V_s \frac{\partial p^f_{s-1}}{\partial p_1}, \ 1 < s < M. \tag{B.10}$$

In particular, for $s = 2$,

$$k_{31} = \frac{\partial p^b_2}{\partial p_1} = V_2 \frac{\partial p^f_1}{\partial p_1} = V_2, \tag{B.11}$$

since $p^f_1 = p_1$. For the same reason,

$$k_{41} = \frac{\partial p^f_1}{\partial p_1} = 1, \tag{B.12}$$

$$k_{12} = \frac{\partial p^f_1}{\partial p_2} = 0. \tag{B.13}$$

The expressions for k_{1M} and k_{2M}, given in (B.4), are derived in a similar manner.
The expressions for k_{22}, k_{32} and k_{42} are derived as follows:

$$k_{22} = \frac{\partial p_2^b}{\partial p_2} = \frac{\partial p_2[1 - Q(p_3^b, p_2^f, N_2)]}{\partial p_2}$$

$$= 1 - Q(p_3^b, p_2^f, N_2) - p_2 \frac{\partial Q(p_3^b, p_2^f, N_2)}{\partial p_3^b} \frac{\partial p_3^b}{\partial p_2} - p_2 \frac{\partial Q(p_3^b, p_2^f, N_2)}{\partial p_2^f} \frac{\partial p_2^f}{\partial p_2}$$

$$= 1 - Q(p_3^b, p_2^f, N_2) - p_2 g_1(p_3^b, p_2^f, N_2)k_{32} - p_2 g_2(p_3^b, p_2^f, N_2)k_{42},$$

(B.14)

$$k_{42} = \frac{\partial p_2^f}{\partial p_2} = \frac{\partial p_2[1 - Q(p_1^f, p_2^b, N_1)]}{\partial p_2}$$

$$= 1 - Q(p_1^f, p_2^b, N_1) - p_2 \frac{\partial Q(p_1^f, p_2^b, N_1)}{\partial p_1^f} \frac{\partial p_1^f}{\partial p_2} - p_2 \frac{\partial Q(p_1^f, p_2^b, N_1)}{\partial p_2^b} \frac{\partial p_2^b}{\partial p_2}$$

$$= 1 - Q(p_1, p_2^b, N_1) - p_2 g_2(p_1, p_2^b, N_1)k_{22}.$$

(B.15)

To find the expression for k_{32} in terms of k_{42}, we use again the induction. Introduce $\partial p_s^b/\partial p_2$, $s = 3,\ldots, M - 1$. For $s = M - 1$, we write

$$\frac{\partial p_{M-1}^b}{\partial p_2} = \frac{\partial p_{M-1}[1 - Q(p_M^b, p_{M-1}^f, N_{M-1})]}{\partial p_2}$$

$$= -p_{M-1} \frac{\partial Q(p_M^b, p_{M-1}^f, N_{M-1})}{\partial p_M^b} \frac{\partial p_M^b}{\partial p_2} - p_{M-1} \frac{\partial Q(p_M^b, p_{M-1}^f, N_{M-1})}{\partial p_{M-1}^f} \frac{\partial p_{M-1}^f}{\partial p_2}$$

$$= -p_{M-1} g_2(p_M, p_{M-1}^f, N_{M-1}) \frac{\partial p_{M-1}^f}{\partial p_2}$$

$$= -p_{M-1} g_2(p_M, p_{M-1}^f, N_{M-1}) \frac{\partial p_{M-1}[1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})]}{\partial p_2}$$

$$= -p_{M-1} g_2(p_M, p_{M-1}^f, N_{M-1})$$

$$[-p_{M-1} \frac{\partial Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{\partial p_{M-2}^f} \frac{\partial p_{M-2}^f}{\partial p_2} - p_{M-1} \frac{\partial Q(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{\partial p_{M-1}^b} \frac{\partial p_{M-1}^b}{\partial p_2}]$$

$$= p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1})g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2}) \frac{\partial p_{M-2}^f}{\partial p_2}$$

$$+ p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1})g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2}) \frac{\partial p_{M-1}^b}{\partial p_2}.$$

Hence,

$$\frac{\partial p_{M-1}^{b}}{\partial p_2} = \frac{p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2})}{1 - p_{M-1}^2 g_2(p_M, p_{M-1}^f, N_{M-1}) g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2})} \frac{\partial p_{M-2}^f}{\partial p_2}$$

$$= \frac{V_{1,M-1}}{V_{2,M-1}} \frac{\partial p_{M-2}^f}{\partial p_2}$$

$$= V_{M-1} \frac{\partial p_{M-2}^f}{\partial p_2}, \tag{B.16}$$

where $V_{1,M-1}$, $V_{2,M-1}$ and V_{M-1} are defined in (B.5). Assume that for $3 < s < M - 1$,

$$\frac{\partial p_s^b}{\partial p_2} = V_s \frac{\partial p_{s-1}^f}{\partial p_2}, \tag{B.17}$$

where V_s, $3 < s < M - 1$, can be found in (B.5). Then,

$$\frac{\partial p_{s-1}^b}{\partial p_2} = \frac{\partial p_{s-1}[1 - Q(p_s^b, p_{s-1}^f, N_{s-1})]}{\partial p_2}$$

$$= -p_{s-1} \frac{\partial Q(p_s^b, p_{s-1}^f, N_{s-1})}{\partial p_s^b} \frac{\partial p_s^b}{\partial p_2} - p_{s-1} \frac{\partial Q(p_s^b, p_{s-1}^f, N_{s-1})}{\partial p_{s-1}^f} \frac{\partial p_{s-1}^f}{\partial p_2}$$

$$= -p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) \frac{\partial p_s^b}{\partial p_2} - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1}) \frac{\partial p_{s-1}^f}{\partial p_2}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})] \frac{\partial p_{s-1}^f}{\partial p_2}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})] \frac{\partial p_{s-1}[1 - Q(p_{s-2}^f, p_{s-1}^b, N_{s-2})]}{\partial p_2}$$

$$= [-V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) - p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})]$$

$$[-p_{s-1} \frac{\partial Q(p_{s-2}^f, p_{s-1}^b, N_{s-2})}{\partial p_{s-2}^f} \frac{\partial p_{s-2}^f}{\partial p_2} - p_{s-1} \frac{\partial Q(p_{s-2}^f, p_{s-1}^b, N_{s-2})}{\partial p_{s-1}^b} \frac{\partial p_{s-1}^b}{\partial p_2}]$$

$$= [V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) + p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})][p_{s-1} g_1(p_{s-2}^f, p_{s-1}^b, N_{s-2})] \frac{\partial p_{s-2}^f}{\partial p_2}$$

$$+ [V_s p_{s-1} g_1(p_s^b, p_{s-1}^f, N_{s-1}) + p_{s-1} g_2(p_s^b, p_{s-1}^f, N_{s-1})][p_{s-1} g_2(p_{s-2}^f, p_{s-1}^b, N_{s-2})] \frac{\partial p_{s-1}^b}{\partial p_2}.$$

Hence, taking into account (B.5),

$$\frac{\partial p_{s-1}^b}{\partial p_2} = \frac{[V_s p_{s-1} g_1 (p_s^b, p_{s-1}^f, N_{s-1}) + p_{s-1} g_2 (p_s^b, p_{s-1}^f, N_{s-1})][p_{s-1} g_1 (p_{s-2}^f, p_{s-1}^b, N_{s-2})]}{1 - [V_s p_{s-1} g_1 (p_s^b, p_{s-1}^f, N_{s-1}) + p_{s-1} g_2 (p_s^b, p_{s-1}^f, N_{s-1})][p_{s-1} g_2 (p_{s-2}^f, p_{s-1}^b, N_{s-2})]} \frac{\partial p_{s-2}^f}{\partial p_2}$$

$$= \frac{V_{1,s-1}}{V_{2,s-1}} \frac{\partial p_{s-2}^f}{\partial p_2}$$

$$= V_{s-1} \frac{\partial p_{s-2}^f}{\partial p_2}. \tag{B.18}$$

Therefore, the inductive hypothesis is established, and we have

$$\frac{\partial p_s^b}{\partial p_2} = \frac{V_{1s}}{V_{2s}} \frac{\partial p_{s-1}^f}{\partial p_2} = V_s \frac{\partial p_{s-1}^f}{\partial p_2}, \ 2 < s < M. \tag{B.19}$$

The case of $s = 3$ yields

$$k_{32} = \frac{\partial p_3^b}{\partial p_2} = \frac{V_{13}}{V_{23}} \frac{\partial p_2^f}{\partial p_2} = \frac{V_{13}}{V_{23}} k_{42}. \tag{B.20}$$

Equations (B.14), (B.15) and (B.20) can be re-written in the matrix form as follows:

$$\begin{bmatrix} 1 & p_2 g_2 (p_3^b, p_2^f, N_2) & p_2 g_1 (p_3^b, p_2^f, N_2) \\ p_2 g_2 (p_1, p_2^b, N_1) & 1 & 0 \\ 0 & -V_{13} & V_{23} \end{bmatrix} \begin{bmatrix} k_{22} \\ k_{42} \\ k_{32} \end{bmatrix} = \begin{bmatrix} 1 - Q(p_3^b, p_2^f, N_2) \\ 1 - Q(p_1, p_2^b, N_1) \\ 0 \end{bmatrix}. \tag{B.21}$$

Solving (B.21) for k_{22}, k_{32} and k_{42}, we get

$$k_{22} = \frac{V_{23}}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] - \frac{V_{23} p_2 g_2 (p_3^b, p_2^f, N_2) + V_{13} p_2 g_1 (p_3^b, p_2^f, N_2)}{B_2}$$

$$[1 - Q(p_1, p_2^b, N_1)],$$

$$k_{32} = -\frac{V_{13} p_2 g_2 (p_1, p_2^b, N_1)}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] + \frac{V_{13}}{B_2} [1 - Q(p_1, p_2^b, N_1)],$$

$$k_{42} = -\frac{V_{23} p_2 g_2 (p_1, p_2^b, N_1)}{B_2} [1 - Q(p_3^b, p_2^f, N_2)] + \frac{V_{23}}{B_2} [1 - Q(p_1, p_2^b, N_1)],$$

where

$$B_2 = \det \begin{bmatrix} 1 & p_2 g_2(p_3^b, p_2^f, N_2) & p_2 g_1(p_3^b, p_2^f, N_2) \\ p_2 g_2(p_1, p_2^b, N_1) & 1 & 0 \\ 0 & -V_{13} & V_{23} \end{bmatrix}$$

$$= V_{23} - V_{13} p_2^2 g_2(p_1, p_2^b, N_1) g_1(p_3^b, p_2^f, N_2) - V_{23} p_2^2 g_2(p_1, p_2^b, N_1) g_2(p_3^b, p_2^f, N_2).$$

For k_{1j}, k_{2j}, k_{3j}, k_{4j}, $j = 3,\ldots, M - 2$, and for $k_{1,M-1}$, $k_{2,M-1}$, $k_{4,M-1}$, using the same procedure, we obtain, respectively,

$$\begin{bmatrix} U_{2,j-1} & -U_{1,j-1} & 0 & 0 \\ 0 & 1 & p_j g_2(p_{j+1}^b, p_j^f, N_j) & p_j g_1(p_{j+1}^b, p_j^f, N_j) \\ p_j g_1(p_{j-1}^f, p_j^b, N_{j-1}) & p_j g_2(p_{j-1}^f, p_j^b, N_{j-1}) & 1 & 0 \\ 0 & 0 & -V_{1,j+1} & V_{2,j+1} \end{bmatrix} \begin{bmatrix} k_{1j} \\ k_{2j} \\ k_{4j} \\ k_{3j} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1 - Q(p_{j+1}^b, p_j^f, N_j) \\ 1 - Q(p_{j-1}^f, p_j^b, N_{j-1}) \\ 0 \end{bmatrix}, \tag{B.22}$$

$$\begin{bmatrix} U_{2,M-2} & -U_{1,M-2} & 0 \\ 0 & 1 & p_{M-1} g_2(p_M, p_{M-1}^f, N_{M-1}) \\ p_{M-1} g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2}) & p_{M-1} g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2}) & 1 \end{bmatrix} \begin{bmatrix} k_{1,M-1} \\ k_{2,M-1} \\ k_{4,M-1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1 - Q(p_M, p_{M-1}^f, N_{M-1}) \\ 1 - Q(p_{M-2}^f, p_{M-1}^b, N_{M-2}) \end{bmatrix}. \tag{B.23}$$

Solving (B.22) and (B.23), we obtain k_{1j}, k_{2j}, k_{3j}, k_{4j}, $j = 3,\ldots, M - 2$, and $k_{1,M-1}$, $k_{2,M-1}$, $k_{4,M-1}$, as shown in (B.4) with B_j, $j = 3,\ldots, M - 1$, given as follows:

$$B_i = \det \begin{bmatrix} U_{2,j-1} & -U_{1,j-1} & 0 & 0 \\ 0 & 1 & p_j g_2(p_{j+1}^b, p_j^f, N_j) & p_j g_1(p_{j+1}^b, p_j^f, N_j) \\ p_j g_1(p_{j-1}^f, p_j^b, N_{j-1}) & p_j g_2(p_{j-1}^f, p_j^b, N_{j-1}) & 1 & 0 \\ 0 & 0 & -V_{1,j+1} & V_{2,j+1} \end{bmatrix}$$

$$= U_{2,j-1}[V_{2,j+1} - V_{1,j+1} p_j^2 g_2(p_{j-1}^f, p_j^b, N_{j-1}) g_1(p_{j+1}^b, p_j^f, N_j)$$

$$- V_{2,j+1} p_j^2 g_2(p_{j-1}^f, p_j^b, N_{j-1}) g_2(p_{j+1}^b, p_j^f, N_j)]$$

$$+ U_{1,j-1}[-V_{1,j+1} p_j^2 g_1(p_{j-1}^f, p_j^b, N_{j-1}) g_1(p_{j+1}^b, p_j^f, N_j)$$

$$- V_{2,j+1} p_j^2 g_1(p_{j-1}^f, p_j^b, N_{j-1}) g_2(p_{j+1}^b, p_j^f, N_j)], j = 3, \ldots, M - 2,$$

$$B_{M-1} = \det \begin{bmatrix} U_{2,M-2} & -U_{1,M-2} & 0 \\ 0 & 1 & p_{M-1}g_2(p_M, p_{M-1}^f, N_{M-1}) \\ p_{M-1}g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2}) & p_{M-1}g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2}) & 1 \end{bmatrix}$$

$$= U_{2,M-2} - U_{1,M-2}p_{M-1}^2 g_1(p_{M-2}^f, p_{M-1}^b, N_{M-2})g_2(p_M, p_{M-1}^f, N_{M-1})$$

$$- U_{2,M-2}p_{M-1}^2 g_2(p_{M-2}^f, p_{M-1}^b, N_{M-2})g_2(p_M, p_{M-1}^f, N_{M-1}).$$

Finally, since $p_M^b = p_M$, $k_{3,M-1}$ is evaluated as follows:

$$k_{3,M-1} = \frac{\partial p_M^b}{\partial p_{M-1}} = \frac{\partial p_M}{\partial p_{M-1}} = 0. \tag{B.24}$$

∎

LEMMA B.3: *For a serial production line (i)–(vi),*

$$\frac{\partial PR}{\partial p_1} = 1 - Q(p_2^b, p_1, N_1) - p_1[k_{31}g_1(p_2^b, p_1, N_1) + k_{41}g_2(p_2^b, p_1, N_1)],$$

$$\frac{\partial PR}{\partial p_j} = [1 - Q(p_{j-1}^f, p_j^b, N_{j-1})][1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- p_j[1 - Q(p_{j+1}^b, p_j^f, N_j)][k_{1j}g_1(p_{j-1}^f, p_j^b, N_{j-1}) + k_{2j}g_2(p_{j-1}^f, p_j^b, N_{j-1})]$$

$$- p_j[1 - Q(p_{j-1}^f, p_j^b, N_{j-1})][k_{3j}g_1(p_{j+1}^b, p_j^f, N_j) + k_{4j}g_2(p_{j+1}^b, p_j^f, N_j)],$$

$$j = 2, \ldots, M - 1, \tag{B.25}$$

$$\frac{\partial PR}{\partial p_M} = 1 - Q(p_{M-1}^f, p_M, N_{M-1}) - p_M[k_{1M}g_1(p_{M-1}^f, p_M, N_{M-1}) + k_{2M}g_2(p_{M-1}^f, p_M, N_{M-1})],$$

where $g_1(x, y, N)$, $g_2(x, y, N)$, $k_{1j}, k_{2j}, j = 2, \ldots, M$, and $k_{3j}, k_{4j}, j = 1, \ldots, M - 1$, are given by (B.1), (B.2) and (B.4), respectively.

Proof: Since $PR = (p_j^f p_j^b)/p_j$, $\forall j = 1, \ldots, M$, and $p_1^f = p_1, p_M^b = p_M$, we obtain

$$\frac{\partial PR}{\partial p_1} = 1 - Q(p_2^b, p_1^f, N_1) - p_1 \frac{\partial Q(p_2^b, p_1^f, N_1)}{\partial p_1}$$

$$= 1 - Q(p_2^b, p_1, N_1) - p_1 \left[\frac{\partial Q(p_2^b, p_1^f, N_1)}{\partial p_2^b} \frac{\partial p_2^b}{\partial p_1} + \frac{\partial Q(p_2^b, p_1^f, N_1)}{\partial p_1^f} \frac{\partial p_1^f}{\partial p_1} \right],$$

$$\frac{\partial PR}{\partial p_j} = [1 - Q(p_{j-1}^f, p_j^b, N_{j-1})][1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- p_j[1 - Q(p_{j+1}^b, p_j^f, N_j)] \frac{\partial Q(p_{j-1}^f, p_j^b, N_{j-1})}{\partial p_j}$$

$$- p_j[1 - Q(p_{j-1}^f, p_j^b, N_{j-1})] \frac{\partial Q(p_{j+1}^b, p_j^f, N_j)}{\partial p_j}$$

$$= [1 - Q(p_{j-1}^f, p_j^b, N_{j-1})][1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- p_j[1 - Q(p_{j+1}^b, p_j^f, N_j)]\left[\frac{\partial Q(p_{j-1}^f, p_j^b, N_{j-1})}{\partial p_{j-1}^f} \frac{\partial p_{j-1}^f}{\partial p_j} + \frac{\partial Q(p_{j-1}^f, p_j^b, N_{j-1})}{\partial p_j^b} \frac{\partial p_j^b}{\partial p_j} \right]$$

$$- p_j[1 - Q(p_{j-1}^f, p_j^b, N_{j-1})]\left[\frac{\partial Q(p_{j+1}^b, p_j^f, N_j)}{\partial p_{j+1}^b} \frac{\partial p_{j+1}^b}{\partial p_j} + \frac{\partial Q(p_{j+1}^b, p_j^f, N_j)}{\partial p_j^f} \frac{\partial p_j^f}{\partial p_j} \right],$$

$$j = 2, \ldots, M - 1, \tag{B.26}$$

$$\frac{\partial PR}{\partial p_M} = 1 - Q(p_{M-1}^f, p_M^b, N_{M-1}) - p_M \frac{\partial Q(p_{M-1}^f, p_M^b, N_{M-1})}{\partial p_M}$$

$$= 1 - Q(p_{M-1}^f, p_M, N_{M-1}) - p_M\left[\frac{\partial Q(p_{M-1}^f, p_M^b, N_{M-1})}{\partial p_{M-1}^f} \frac{\partial p_{M-1}^f}{\partial p_M} + \frac{\partial Q(p_{M-1}^f, p_M^b, N_{M-1})}{\partial p_M^b} \frac{\partial p_M^b}{\partial p_M} \right].$$

Using the expressions for $\partial Q(x, y, N)/\partial x$ and $\partial Q(x, y, N)/\partial y$ from Lemma B.1 and the expressions for $\partial p_{j-1}^f/\partial p_j$, $\partial p_j^b/\partial p_j$, $j = 2, \ldots, M$, and $\partial p_{j+1}^b/\partial p_j$, $\partial p_j^f/\partial p_j$, $j = 1, \ldots, M - 1$, from Lemma B.2, (B.26) is reduced to (B.1), (B.2), (B.4) and (B.25). ∎

LEMMA B.4: *Under Hypothesis H.2, the following hold:*

(a) if $mb_{j-1} > ms_j$, then

$$k_{1j}g_1(p_{j-1}^f, p_j^b, N_{j-1}) + k_{2j}g_2(p_{j-1}^f, p_j^b, N_{j-1}) \sim \mathbb{O}(\epsilon), j = 2, \ldots, M; \tag{B.27}$$

(b) if $mb_j < ms_{j+1}$, then

$$k_{3j}g_1(p_{j+1}^b, p_j^f, N_j) + k_{4j}g_2(p_{j+1}^b, p_j^f, N_j) \sim \mathbb{O}(\epsilon), j = 1, \ldots, M - 1, \tag{B.28}$$

where $\epsilon = \max_{j,s}(\epsilon_{j1}, \epsilon_{s2})$ and $g_1(x, y, N)$, $g_2(x, y, N)$, k_{1j}, k_{2j}, $j = 2, \ldots, M$, and k_{3j}, k_{4j}, $j = 1, \ldots, M - 1$, are given by (B.1), (B.2) and (B.4), respectively.

Proof: Statement (a) is proved as follows: From Hypothesis H.2,

$$\epsilon_{j1} = Q(p_{j-1}^f, p_j^b, N_{j-1}) = \frac{(1 - p_{j-1}^f)(1 - \alpha)}{1 - \frac{p_{j-1}^f}{p_j^b}\alpha^{N_{j-1}}}$$

$$= \frac{\frac{p_j^b - p_{j-1}^f}{p_j^b}}{1 - \frac{p_{j-1}^f}{p_j^b}\alpha^{N_{j-1}}} << 1, j = 2, \ldots, M,$$

where

$$\alpha = \frac{p_{j-1}^f(1 - p_j^b)}{p_j^b(1 - p_{j-1}^f)}.$$

Thus,

$$1 - \frac{p_{j-1}^f}{p_j^b}\alpha^{N_{j-1}} = \frac{p_j^b - p_{j-1}^f}{\epsilon_{j1}p_j^b}, j = 2, \ldots, M. \tag{B.29}$$

Using Lemma B.1 and (B.29), we write:

$$g_1(p_{j-1}^f, p_j^b, N_{j-1}) = \frac{\epsilon_{j1}^2 p_j^b \left\{-1 + \left[\frac{p_j^b}{p_{j-1}^f} + \frac{N_{j-1}(p_j^b - p_{j-1}^f)}{p_{j-1}^f(1 - p_{j-1}^f)}\right](1 - \frac{p_j^b - p_{j-1}^f}{\epsilon_{j1}p_j^b})\right\}}{(p_j^b - p_{j-1}^f)^2} \sim \mathcal{O}(\epsilon),$$
$$\tag{B.30}$$

$$g_2(p_{j-1}^f, p_j^b, N_{j-1}) = \frac{\epsilon_{j1}^2 \left\{p_{j-1}^f - \left[p_{j-1}^f + \frac{N_{j-1}(p_j^b - p_{j-1}^f)}{p_j^b(1 - p_j^b)}\right](1 - \frac{p_j^b - p_{j-1}^f}{\epsilon_{j1}p_j^b})\frac{p_j^b}{p_{j-1}^f}\right\}}{(p_j^b - p_{j-1}^f)^2} \sim \mathcal{O}(\epsilon).$$

By Lemma B.2 and (B.30), the k_{ij}s, involved in (B.27), are calculated as follows: For $j = 2$,

$$k_{12} = 0, \tag{B.31}$$

$$k_{22} = \frac{V_{23}}{V_{23} - V_{13}p_2^2 g_1(p_3^b, p_2^f, N_2)\mathcal{O}(\epsilon) - V_{23}p_2^2 g_2(p_3^b, p_2^f, N_2)\mathcal{O}(\epsilon)}[1 - Q(p_3^b, p_2^f, N_2)]$$

$$- \frac{V_{23}p_2 g_2(p_3^b, p_2^f, N_2) + V_{13}p_2 g_1(p_3^b, p_2^f, N_2)}{V_{23} - V_{13}p_2^2 g_1(p_3^b, p_2^f, N_2)\mathcal{O}(\epsilon) - V_{23}p_2^2 g_2(p_3^b, p_2^f, N_2)\mathcal{O}(\epsilon)}[1 - \epsilon_{j1}].$$

Neglecting the $\mathcal{O}(\epsilon)$ terms, we have

$$k_{22} \approx 1 - Q(p_3^b, p_2^f, N_2) - p_2 g_2 \, (p_3^b, p_2^f, N_2) - V_3 p_2 g_1 \, (p_3^b, p_2^f, N_2). \tag{B.32}$$

For $3 \leq j \leq M - 2$, we first calculate

$$\begin{aligned} B_j &= U_{2,j-1}[V_{2,j+1} - V_{1,j+1} p_j^2 \mathbb{O}(\epsilon) g_1(p_{j+1}^b, p_j^f, N_j) \\ &\quad - V_{2,j+1} p_j^2 \mathbb{O}(\epsilon) g_2 \, (p_{j+1}^b, p_j^f, N_j)] \\ &\quad + U_{1,j-1}[-V_{1,j+1} p_j^2 \mathbb{O}(\epsilon) g_1 \, (p_{j+1}^b, p_j^f, N_j) \\ &\quad - V_{2,j+1} p_j^2 \mathbb{O}(\epsilon) g_2 \, (p_{j+1}^b, p_j^f, N_j)] \\ &\approx U_{2,j-1} V_{2,j+1}. \end{aligned} \tag{B:33}$$

Then, using (B.33),

$$k_{1j} \approx U_{j-1}[1 - Q(p_{j+1}^b, p_j^f, N_j)] \\ - U_{j-1} p_j g_2 \, (p_{j+1}^b, p_j^f, N_j) - U_{j-1} V_{j+1} p_j g_1 \, (p_{j+1}^b, p_j^f, N_j), \tag{B.34}$$

$$k_{2j} \approx 1 - Q(p_{j+1}^b, p_j^f, N_j) - p_j g_2 \, (p_{j+1}^b, p_j^f, N_j) - V_{j+1} p_j g_1 \, (p_{j+1}^b, p_j^f, N_j). \tag{B.35}$$

Similarly, for $j = M - 1$,

$$k_{1,M-1} \approx U_{M-2}[1 - Q(p_M, p_{M-1}^f, N_{M-1})] - U_{M-2} p_{M-1} g_2 \, (p_M, p_{M-1}^f, N_{M-1}), \tag{B.36}$$

$$k_{2,M-1} \approx 1 - Q(p_M, p_{M-1}^f, N_{M-1}) - p_{M-1} g_2 \, (p_M, p_{M-1}^f, N_{M-1}). \tag{B.37}$$

For $j = M$,

$$k_{1M} = U_{M-1}, \tag{B.38}$$

$$k_{2M} = 1. \tag{B.39}$$

Therefore, substituting (B.31), (B.32), and (B.34)–(B.39) into the left hand side of (B.27) and using (B.30), we conclude that

$$k_{1j} g_1(p_{j-1}^f, p_j^b, N_{j-1}) + k_{2j} g_2(p_{j-1}^f, p_j^b, N_{j-1}) \sim \mathbb{O}(\epsilon).$$

Statement (b) is proved analogously. ∎

Proof of Theorem 4.1: Suppose first that

$$mb_j > ms_{j+1}. \tag{B.40}$$

Then, under Hypothesis H.1, the following two cases are possible:

(α) $mb_{j+1} > ms_{j+2},..., mb_{i-1} > ms_i, mb_i < ms_{i+1},..., mb_{M-1} < ms_M, j < i < M$, i.e., according to Rule 3.1, the bottleneck is in the interior of the line,

(β) $mb_{j+1} > ms_{j+2},..., mb_{M-1} > ms_M$, i.e., according to Rule 3.1, the bottleneck is the last machine.

Consider first case (α). By Hypothesis H.2 and Lemmas B.3 and B.4,

$$\frac{\partial PR}{\partial p_i} = [1 - Q(p_{i-1}^f, p_i^b, N_{i-1})][1 - Q(p_{i+1}^b, p_i^f, N_i)]$$

$$- p_i[1 - Q(p_{i+1}^b, p_i^f, N_i)][k_{1i}g_1(p_{i-1}^f, p_i^b, N_{i-1}) + k_{2i}g_2(p_{i-1}^f, p_i^b, N_{i-1})]$$

$$- p_i[1 - Q(p_{i-1}^f, p_i^b, N_{i-1})][k_{3i}g_1(p_{i+1}^b, p_i^f, N_i) + k_{4i}g_2(p_{i+1}^b, p_i^f, N_i)],$$

$$= (1 - \epsilon_{i1})(1 - \epsilon_{i2})$$

$$- p_i(1 - \epsilon_{i2})[k_{1i}g_1(p_{i-1}^f, p_i^b, N_{i-1}) + k_{2i}g_2(p_{i-1}^f, p_i^b, N_{i-1})]$$

$$- p_i(1 - \epsilon_{i1})[k_{3i}g_1(p_{i+1}^b, p_i^f, N_i) + k_{4i}g_2(p_{i+1}^b, p_i^f, N_i)],$$

$$= 1 - \alpha_i, \qquad (B.41)$$

where

$$\alpha_i = \epsilon_{i1} + (1 - \epsilon_{i1})\epsilon_{i2}$$

$$+ p_i(1 - \epsilon_{i2})[k_{1i}g_1(p_{i-1}^f, p_i^b, N_{i-1}) + k_{2i}g_2(p_{i-1}^f, p_i^b, N_{i-1})]$$

$$+ p_i(1 - \epsilon_{i1})[k_{3i}g_1(p_{i+1}^b, p_i^f, N_i) + k_{4i}g_2(p_{i+1}^b, p_i^f, N_i)] \sim \mathbb{O}(\epsilon).$$

Now we calculate $\partial PR/\partial p_j, j < i$. This calculation is different for $j = 1$ and $j > 1$. For $j = 1$, using Lemma B.3, we write

$$\frac{\partial PR}{\partial p_1} = 1 - Q(p_2^b, p_1, N_1) - p_1[k_{31}g_1(p_2^b, p_1, N_1) + k_{41}g_2(p_2^b, p_1, N_1)]$$

$$= 1 - \alpha_1, \qquad (B.42)$$

where

$$\alpha_1 = Q(p_2^b, p_1, N_1) + p_1[k_{31}g_1(p_2^b, p_1, N_1) + k_{41}g_2(p_2^b, p_1, N_1)] \sim \mathbb{O}(1).$$

For $j > 1$, by Hypothesis H.2 and Lemmas B.3 and B.4, we obtain

$$\frac{\partial PR}{\partial p_j} = (1 - \epsilon_{j1})[1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$- p_j [1 - Q(p_{j+1}^b, p_j^f, N_j)][k_{1j}g_1 (p_{j-1}^f, p_j^b, N_{j-1}) + k_{2j}g_2 (p_{j-1}^f, p_j^b, N_{j-1})]$$

$$- p_j (1 - \epsilon_{j,1})[k_{3j}g_1 (p_{j+1}^b, p_j^f, N_j) + k_{4j}g_2 (p_{j+1}^b, p_j^f, N_j)],$$

$$= 1 - \alpha_j, \tag{B.43}$$

where

$$\alpha_j = Q(p_{j+1}^b, p_j^f, N_j) + \epsilon_{j1}[1 - Q(p_{j+1}^b, p_j^f, N_j)]$$

$$+ p_j [1 - Q(p_{j+1}^b, p_j^f, N_j)][k_{1j}g_1 (p_{j-1}^f, p_j^b, N_{j-1}) + k_{2j}g_2 (p_{j-1}^f, p_j^b, N_{j-1})]$$

$$+ p_j (1 - \epsilon_{j1})[k_{3j}g_1 (p_{j+1}^b, p_j^f, N_j) + k_{4j}g_2 (p_{j+1}^b, p_j^f, N_j)]$$

$$\sim \mathcal{O}(1).$$

Thus, for all $1 \leq j < i$,

$$\frac{\partial PR}{\partial p_j} < \frac{\partial PR}{\partial p_i},$$

i.e., under the assumption (B.40), the effect on the *PR* of the machine, upstream of machine i, is less than that of machine i.

Consider now case (β). Using Hypothesis H.2 and Lemmas B.3 and B.4, $\partial PR/\partial p_M$ is estimated as follows:

$$\frac{\partial PR}{\partial p_M} = 1 - \epsilon_{M1} - p_M [k_{1M}g_1 (p_{M-1}^f, p_M, N_{M-1}) + k_{2M}g_2 (p_{M-1}^f, p_M, N_{M-1})] = 1 - \alpha_M,$$
$$\tag{B.44}$$

where

$$\alpha_M = \epsilon_{M1} + p_M [k_{1M}g_1 (p_{M-1}^f, p_M, N_{M-1}) + k_{2M}g_2 (p_{M-1}^f, p_M, N_{M-1})] \sim \mathcal{O}(\epsilon).$$

The expressions for $\partial PR/\partial p_j$, $M > j \geq 1$, remain the same as in case (α), i.e., (B.42) and (B.43). Therefore, again,

$$\frac{\partial PR}{\partial p_j} < \frac{\partial PR}{\partial p_M}, \quad M > j \geq 1,$$

i.e., under the assumption (B.40), the effect on the *PR* of the machine, upstream of machine M, is less than that of machine M. Thus, the Rule 3.1 is proved for the case (3.1).

The case of $ms_j > mb_{j-1}$ is proved similarly. Therefore, the Rule 3.1 is proved for the case (3.2).   ■

References

1. D. Jacobs and S.M. Meerkov, "Mathematical Theory of Improvability for Production Systems", *Mathematical Problems in Engineering*, vol. 1, No. 2, pp. 95–137, 1995.
2. E. Goldratt and J. Cox, *The Goal*, North Rivers Press, NY, 1986.
3. S.R. Lawrence and A.H. Buss, "Economic Analysis of Production Bottlenecks," *Mathematical Problems in Engineering*, vol. 1, No. 4, 1995.
4. J.O. Henriksen and R.C. Crain, *GPSS/H Reference Manual*, Wolverine Software Corp., VA, 1989.
5. A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, 1991.
6. S.B. Gershwin, *Manufacturing Systems Engineering*, Prentice Hall, 1994.
7. J.A. Buzacott and J.G. Shanthikumar, *Stochastic Models of Manufacturing Systems*, Prentice Hall, 1993.
8. C.-T. Kuo, J.-T. Lim and S.M. Meerkov, "Improvability Analysis of a Machining Transfer Line: An Application," *Proceeding of the 33rd IEEE CDC*, Orlando, FL, Dec. 1994; also *Control Group Report* No. CGR-94-08, The University of Michigan, Ann Arbor, MI, April 1994.