

清 华 大 学

综 合 论 文 训 练

题目：在制品库存与产品库存的关系
研究

培养单位：工业工程系

学 科：工业工程

姓 名：徐驰

指导教师：吴甦教授

2014 年 6 月 27 日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名：_____ 导师签名：_____ 日 期：_____

中文摘要

随着制造业的发展，企业越来越重视库存的优化。对于某些从相同的半成品分化出来的产品，企业希望能够使其库存停留在未分化的在制品阶段，待收到订单之后再按要求加工，以此来减少该系列产品的总库存。本文主要研究这种库存改进策略的效果，分析需求分布、需求相关性、服务水平、生产能力等参数对改进效果的影响，为企业提供决策支持。

关键词：在制品库存 安全库存 风险分担 延迟差异化

ABSTRACT

As manufacturing industry develops, enterprises are paying more attention to optimizing their inventory. There are some products that come from the same kind of semi-product. Enterprise would like to keep these products at in-process inventory until specific orders arrive, so as to reduce the total inventory. This paper studies the sequence of the strategy, shows how the parameters such as the distribution of demand, correlation of demand, service level, production capacity would affect the improvement of the strategy. These conclusions could be enterprises' decision support.

Key words: In-process inventory Safety stock Risk-pooling Postponement

目 录

第 1 章 引言	1
1.1 课题背景	1
1.2 问题描述	1
1.3 研究目标	2
1.4 意义和价值	2
第 2 章 文献综述	3
2.1 安全库存和安全时间	3
2.2 多阶段生产和在途库存	5
2.3 风险分担	7
2.4 延迟差异化	8
第 3 章 研究思路	10
3.1 证明优化潜力存在性	10
3.2 改进失效现象的研究	10
3.3 需求相关性的影响	11
3.4 延迟订单的影响	11
第 4 章 优化潜力存在性	13
4.1 安全库存的制定标准	13
4.2 两种成品需求独立正态分布	16
4.3 两种以上的成品需求独立正态分布	18
4.4 一些参数对改进效果的影响	19
第 5 章 改进失效现象	22
5.1 对称稳定分布的参数与改进失效的关系	22
5.2 右偏斜的需求分布一定存在改进失效的可能	24
5.3 任何服务水平都存在改进失效的可能	26
5.4 泊松分布下发生改进失效的区间图示	29

5.5 第二类服务水平与改进失效	32
5.6 避免改进失效	34
第 6 章 需求相关性对改进效果的影响	35
6.1 需求服从二维正态分布的情况	35
6.2 需求服从多维正态分布的情况	39
6.3 需求相关性影响总结	40
6.4 需求相关性实例	41
第 7 章 延迟订单对改进效果的影响	44
7.1 考虑延迟订单成本的任意更新过程	44
7.2 考虑延迟订单成本的泊松过程	47
第 8 章 总结与展望	51
8.1 总结	51
8.2 展望	51
插图索引	53
表格索引	54
公式索引	55
参考文献	59
致 谢	62
声 明	63
附录 A 外文文献翻译	64
A.1 简介	64
A.1.1 传统生产订单系统及它面临的问题	64
A.1.2 拉式系统的目标	65
A.1.3 本研究的目标	65
A.2 拉式系统概要	65
A.2.1 拉式系统方法论	65
A.2.2 看板	66

A.2.3 操作过程	66
A.2.3.1 过程内看板	66
A.2.3.2 过程间看板	67
A.3 系统模型	68
A.3.1 标记方式	68
A.3.2 信息和原材料流动示意图	68
A.3.3 系统的基本公式	68
A.4 生产和库存的波动分析	70
A.4.1 单位订货量相对较小的情况	70
A.4.1.1 生产波动	70
A.4.1.2 库存波动	71
A.4.2 单位订货量相对较大的情况	72
A.4.2.1 标记方式	72
A.4.2.2 仿真模型	73
A.4.2.3 输入数据和参数值	75
A.4.2.4 实验结果	75
A.4.3 与推式系统的比较	75
A.5 总结	77
参考文献	78
附录 B 汽车销售数据	79

第 1 章 引言

1.1 课题背景

随着 MIS、MRP 等技术在汽车制造业的应用，汽车行业对零部件供应的及时性和服务水平要求越来越高。

G 公司是全球最大的汽车制造商之一。M 公司作为汽车零部件供应商，为 G 公司某些型号的汽车生产保险杠。

保险杠的生产流程需要经过注塑、喷涂、装配三个阶段，下面对每个阶段的作用和特点作简要介绍。

注塑 注塑是保险杠生产的第一步，即通过注塑机器和模具，制造出特定形状的保险杠。注塑不同型号的保险杠时需要换模，换模时间相当长。为了提高生产效率，需要在注塑阶段设置合适的生产批量，尽量做到少换模。公司当前拥有注塑机器 3 台，生产的保险杠型号共十余种，并且要求每台机器每天换模不超过 2 次。这种策略下，每天能投入生产的保险杠型号是有限的，因此有必要保持一定量的安全库存。

喷涂 喷涂的作用是对注塑成型的保险杠上色。每种型号的保险杠有几种不同的颜色可选。喷涂工序是在连续运转的环形传送带上完成的，工人在某个位置将未喷涂的保险杠注塑件挂到传送带上的支架上，传送一段距离后进入喷涂机器，完成喷涂后在某个位置取下保险杠，然后空支架又运转到上货处。喷涂机器更换喷涂颜料需要一定的时间和人力，因此也要尽量设置合理的喷涂批量，减少颜料更换次数。

装配 装配是指给保险杠安装必要的配件。经过前两道工序，保险杠的主体已经基本完成。在打包发往 G 公司之前，还需要在保险杠上安装一些配件。这道工序耗时不多，且相当稳定，一般能保证在规定时间内完成装配并发货。

1.2 问题描述

如前所述，受限于注塑机器的换模次数，M 公司无法在一天内生产每一种型号的保险杠塑件，因此需要保有足量的安全库存，以应对可能发生的需求波动。不仅如此，由于每种型号的保险杠有不同颜色之分，进一步增大了需求的

不确定性。如果每种型号每种颜色的保险杠都保有足量安全库存，无疑使得安全库存量变得更大。

公司希望能在不影响服务水平、不降低生产效率的前提下，通过采用合理的生产和库存策略，减少安全库存总量。

1.3 研究目标

通过数学模型分析将成品安全库存合并前置到在制品库存是否有助于减少安全库存，并进一步分析此改进方案的适用范围和优化效果。最后根据前面的分析，给出一些定性或定量的结论，帮助企业对改进方案可能产生的效果建立初步的认识，以决定是否值得继续推进方案。

1.4 意义和价值

对于一般企业而言，在不降低服务水平和生产效率的基础上，通过库存合并前移减少安全库存，能够减少生产线的库存成本，节省场地和设备，解放更多的流动资金，对生产线的生产柔性有一定的提升。

但是，企业作出改进的同时也会付出一些必要的成本。因此，企业希望能在改进尚未开始的时候，就对改进方案可能带来的利益和后果有一定的了解。如果在改进之前发现改进不能起到正向的效果，或者改进的收益太小，企业就能避免无效的投入；如果在改进之前发现改进存在巨大的潜在收益，企业就可以投入更多的精力，充分挖掘改进的效果。

第 2 章 文献综述

为了更深入地研究上述问题，我们首先要明确安全库存、在途库存等概念的准确定义和计算方法，然后了解多阶段生产的特点以及用于研究多阶段生产的现有数学模型。

2.1 安全库存和安全时间

Fetter 和 Dalleck (1961)^[1] 认为安全库存是为了满足一定服务水平而保留的库存。作者根据日销售量和补货周期等因素，提出了计算一定服务水平下的安全库存的方法。

Lambrecht 等人 (1984)^[2] 在研究物料需求计划系统 (MRP) 时指出，安全库存是为了对抗需求预测、生产批量、生产运输提前期等因素中存在的不确定性。作者根据供应链中的级库存模型，将多阶段生产计划建模成动态规划问题和马尔科夫决策问题，改进了 Clark 和 Scarf (1962)^[3] 的启发式算法，并指出了启发式算法的上下界和获得最优解条件。

Baker 等人 (1986)^[4] 研究了多阶段生产线中，提高零部件通用性对安全库存的影响。作者结合风险分担思想在供应链等领域中的应用，用数学模型证明，在给定最低服务水平的前提下，引入通用件能够降低两阶段、两产品生产线的安全库存，并总结了引入通用件产生的影响：所有零件的安全库存总数降低；通用件的安全库存低于它所替代的两种零件安全库存之和；没有被通用件替代的其他零件安全库存增加。

Eppen 和 Martin (1988)^[5] 总结了各种情况下安全库存的计算方式。当需求和提前期分布已知时，令 W 代表提前期内需求， X 代表单位时间需求， Y 代表提前期长度， R 代表补货点。则有以下公式。

$$\mu_W = \mu_X \mu_Y \quad (2-1)$$

$$\sigma_W^2 = \sigma_X^2 \mu_Y + \sigma_Y^2 \mu_X^2 \quad (2-2)$$

$$R = \mu_W + Z_\delta \sigma_W \quad (2-3)$$

当需求和提前期分布未知时，需要视情况建立合适的模型估算它们的分布。

Molinder (1997)^[6] 认为生产中的不确定性主要来自需求、制造、运输等环节, 并进一步指出制造环节的不确定性是由产能负载、机器故障、排队效应、返工等多种因素引起的。作者认为这些不确定性对多阶段生产的影响尤其显著, 因为某一工序的延迟可能导致物料清单 (BOM) 上所有更高级别的物料都受到影响。作者通过模拟退火优化了多种参数情形下的订货批量、安全库存和提前期, 并比较了每种情形下用安全库存或安全时间作为对抗风险准则的效果。实验结果表明:

1. 需求不确定性高、提前期不确定性低时, 宜采用安全库存作为优化准则; 需求和提前期不确定性都高时, 宜采用安全时间作为准则。
2. 产品结构 (或 BOM) 基本不影响采用哪种准则的抉择。
3. 当缺货成本与库存成本之比较低时, 安全库存准则比其他准则都好。
4. 当上述比值较高时, 准则的抉择往往取决于需求不确定性。需求不确定性低, 宜采用安全库存; 反则宜采用安全时间。

Chopra 等人 (2004)^[7] 认为降低供应商的供货提前期和降低供货提前期的不确定性是两种降低安全库存的有效方法。一般认为, 在正态模型下, 降低供货提前期的不确定性比降低供货提前期更为有效, 尤其是在供货提前期不确定性较大的情况下。本文作者以 Eppen 和 Martin (1988)^[5] 的工作为基础, 证明了该结论不是完全正确的, 而是存在一个不小于 50% 的服务水平临界点, 当服务水平低于这个临界点时, 降低提前期的不确定性反而会使安全库存增大。若模型是正态的, 则这个临界点为 50%; 若不是正态的, 临界点一般在 50% 到 70%, 正好是一般企业的服务水平范围。因此企业如果希望降低安全库存, 必须谨慎检查自己的服务水平和提前期分布模型, 盲目降低提前期的不确定性反而可能增大安全库存。文章的最后, 作者还提出了通过累积分布函数 (CDF) 估算临界点的方法。

Fang 等人 (2013)^[8] 考虑需求不确定性、提前期、提前期不确定性的边际价值, 建立了能同时处理需求和供应两边不确定性的模型, 并且不依赖于最优库存策略。相反, 非最优的库存策略下, 从这几个方面获得的边际价值更大。因此, 即使面临糟糕的库存管理策略, 管理人员也能够根据此模型很快找出最有效率的改进方向。

2.2 多阶段生产和在途库存

Gershwin (1994)^[9] 研究了如图 2.1 所示的单产品多阶段生产线。其中 M_i 代表机器， B_i 代表缓存。

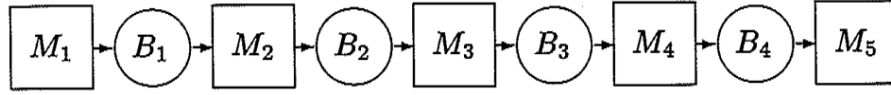


图 2.1 单产品多阶段生产

多阶段生产中的重要制约因素是机器故障。机器故障有两种模式：基于时间的故障模式和基于运作的故障模式。两种模式都是随机的。基于时间的故障模式下，机器故障的发生取决于使用的时间长短，比如材料的强度失效；基于运作的故障模式下，机器故障的发生取决于使用次数或产量，比如刀具失效。

机器即使没有出故障，也不一定运转，因为还受到上下游库存的限制。如果上游库存不足，导致下游机器停工，我们称下游机器为缺料状态；如果下游库存已满，导致上游机器停工，我们称上游机器为阻塞状态。

缓存区的库存被称为在途库存。在多阶段生产中，在途库存往往是必不可少的，因为它能减少机器故障带来的损失。当一台机器出现故障时，上游的机器可以继续运转，生产的产品进入在途库存；下游机器也能依靠故障机器的在途库存提供原料继续运转。但在途库存也会带来一些问题，Gershwin 在书中总结了在途库存的主要缺点：

1. 在途库存是消耗资金制造出来的，占用了资金却不产生任何效益。
2. 根据 Little's Law^[10]，在途库存对提前期有影响。在途库存越多，提前期越长。提前期过长可能导致顾客不满，同时也使得生产中的故障更难以被观察到。在观察到故障之前，故障机器可能已经制造了很多残次品。
3. 零部件在库存期间可能遭受破损、偷窃等损失。库存量越大、时间越长，遭受的损失越大。
4. 储存用的场地和设备会消耗空间和资金。

Gershwin 还用马尔科夫模型分析了多阶段生产线在缓存容量上限 $N = 0$ 和 $N = \infty$ 的极端情形下的表现，得出了两个重要的结论：

1. 在 $N = 0$ 的情况下，多阶段生产线的运作时间与总时间之比总是随机器数目增加而减小，并且收敛于 0。也就是说，如果没有在途库存，当产品需要的生产阶段很多时，生产线的效率将趋近于 0。

2. 在 $N = \infty$ 的情况下，多阶段生产线的产能总是取决于最慢的一台机器。也就是说，总体产能只有在瓶颈阶段可能得到优化。

Gershwin 指出，用马尔科夫模型研究多阶段生产还存在一些难点。最突出的难点就是状态空间过大。每个机器有 2 种状态，每个缓存有 $N + 1$ 种状态，马尔科夫模型的状态空间就是它们的乘积，即

$$M = 2^k \prod_{i=0}^{k-1} (N_i + 1)$$

随着阶段数目的增加，状态空间 M 很快就变得非常庞大。

Goyal (1978)^[11] 考虑需求分布均匀、不允许缺货的多阶段生产，建立了以减少总成本为优化准则的数学模型，用于分析最优的生产批量以及各阶段之间最优的运输批量。

Kimura 和 Terada (1981)^[12] 分析了多阶段拉动式生产的需求和库存波动是如何逐级传递的，证明了拉动式生产系统比推动式生产系统更能适应需求的波动，并能够防止下游的生产和库存波动向上游逐级扩大。Kimura 和 Terada 所定义的拉动式生产系统需要满足以下运行方式：建立标准的再订货点和订货批量大小；随时了解库存水平和延期订单；连续检查库存水平，一旦某些零部件数量低于再订货点，就立即向上游订货。拉动式生产系统中使用看板来维持运作，流程如图 2.2 所示。

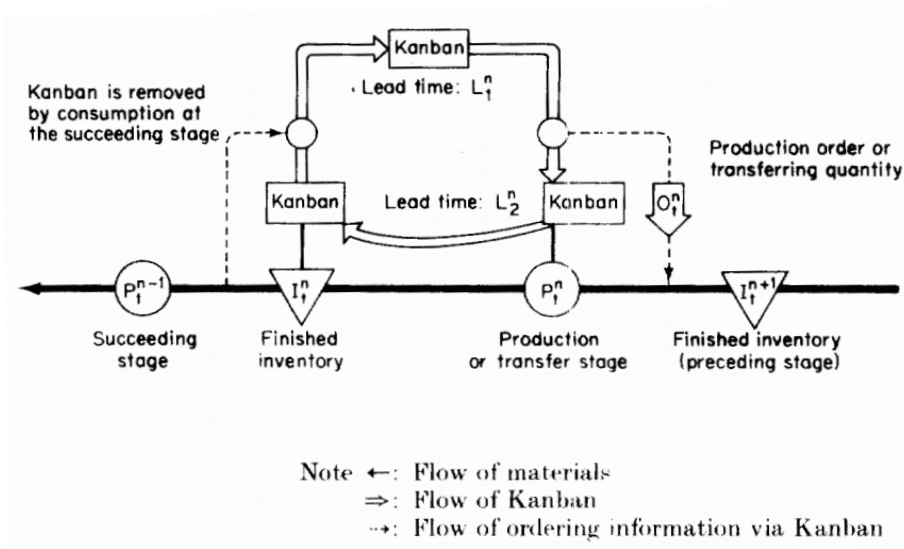


图 2.2 拉动式生产系统流程图

研究中使用的推动式生产系统是 Tabe 等人 (1980)^[13] 建立的模型, 如图 2.3 所示。

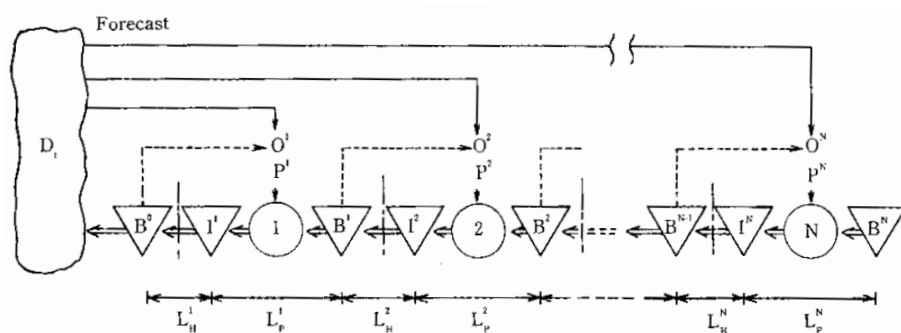


图 2.3 推动式生产系统流程图

通过数学模型分析和仿真模型实验, Kimura 和 Terada 得到如下结论:

1. 拉动式生产系统中, 单位订货量的大小非常重要。在单位订货量与产量水平相比很小的情况下, 拉动式生产中的下游生产波动不会在向上游传递的过程中放大。
2. 推动式生产系统中, 生产和库存波动的放大率取决于预测误差。以波动放大率为准则时, 推动式系统和拉动式系统之间的抉择取决于预测误差的大小。
3. 拉动式生产系统中另一个影响波动放大率的参数是从看板离开容器到该阶段完成生产的时间差。这个时间差越长, 波动放大率越高。

Conway 等人 (1988)^[14] 反对一味追求降低在途库存的做法, 全面总结了在途库存在多阶段生产中的意义和作用, 并分析各种情况下如何真正改善多阶段生产的绩效。该研究中有以下一些重要的结论: 靠前的几个阶段对整条生产线的产能影响最大, 长生产线的绩效比短生产线差得不多; 添加缓存对绩效的提升较大, 但增加缓存容量的效果不显著; 在生产线靠中间的位置添加缓存, 效果比靠两端的位置好得多; 在未平衡的生产线中添加 buffer 效果不显著。

2.3 风险分担

Daskin 等人 (2002)^[15] 研究了 LMRP (location model with risk pooling) 问题, 在考虑选址造成的风险分担效应下通过数学规划进行选址。Snyder 等人 (2007)^[16] 在 LMRP 的基础上考虑参数的随机性, 利用拉格朗日松弛求解

SLMRP (stochastic LMRP) 问题。Ozsen 等人 (2008)^[17] 在 LMRP 基础上, 考虑了增大仓库容量所需的成本, 建模求解 CLMRP (capacited LMRP) 问题。Shen 等人 (2003)^[18] 考虑将一些零售商转化成分销商, 使得整个系统的成本降低, 并建模。

Eppen 和 Schrage (1981)^[19] 研究了 supplier-depot-warehouse 系统。这套系统中, 补给站从供应商进货, 然后分配给仓库, 使得仓库之间的存货量达到一定的比值。文中分类讨论了无固定成本和有固定成本情况下的不同订货策略。Josson 和 Silver (1987)^[20] 考虑两阶段的库存系统, 中央仓库每隔固定时间从外部供货商进货, 然后按照“期望总缺货量最少”的原则分配到支线仓库的情况。文中研究了如何在每个周期内部增加一次完全的重新分配, 以减少总缺货量。Jackson (1988)^[21] 研究了有仓库向多个零售商供货的情况下如何调整一些参数达到最优库存。考虑了需求服从正态分布和泊松分布的情况。使用的方法主要是仿真实验。文中引入了风险分担期的概念, 仓库有货时可向各零售商补货, 此时处于风险分担期; 仓库无货时各零售商只能靠自己的库存来满足订单, 此时不属于风险分担期。Schwarz (1989)^[22] 研究了两种不同的风险分担方式, 找到了它们绩效优劣的临界点, 并指出了管道库存成本 (pipeline inventory cost) 的在这个系统中的重要性。

Clark 和 Scarf (2004)^[23] 考虑了订货时多阶段提前期的情况, 讨论了提前期随很多因素变化时的最优订货策略。Benjaafar 等人 (2005)^[24] 利用生灭过程等模型, 探讨了生产库存系统中多种因素、多种假设对风险分担效果的影响。Yang 和 Schrage (2009)^[25] 研究了部分产品可替代的情况下, 风险分担可能达到的利益或造成的后果。

2.4 延迟差异化

Zinn 和 Bowersox (1988)^[26] 总结归纳了 5 种延迟差异化的方式及它们各自的优势和劣势, 以及 5 种不同的成本模型, 随后在一些参数下对这些模型做了仿真。Lee 和 Tang (1997)^[27] 对多种不同的延迟差异化方式和通用件设计方式进行了研究, 用数学公式表达出了它们的目标函数。

Su 等人 (2005)^[28] 详细研究了时间延迟差异化 (time postponement) 和结构延迟差异化 (form postponement) 之前的区别, 并做了理论分析和仿真实验。本质上是运筹学中多排与单排问题的应用和扩展。Cheng 等人 (2010)^[29] 对两

阶段产品在均匀分布、正态分布、泊松分布等各种需求分布下的延迟差异化，并做了一系列的仿真实验。

Zeng 等人^[30] 的书中考虑了延迟差异化导致的提前期变化，并用数值实验证明了部分提前比全部提前的改进效果更好。**Krishnamoorthy** 和 **Islam** (2004)^[31] 用生灭过程来研究泊松到达的需求分布下的延迟差异化。文中假设未满足的顾客需求会在一个订单池中储存，直到有富余的货物时（指库存大于某个临界值时）再发货。订单池的容量是有限的，超出容量的订单将会被舍弃。**Jayaraman** 等人 (2012)^[32] 在上述假设的基础上，考虑了库存货物自然损耗的情况，通过建模的方式求解稳态下的成本。

第3章 研究思路

3.1 证明优化潜力存在性

将成品库存合并前移到在途库存，会造成安全库存的变动。影响这个变动的因素是多方面的：风险分担（risk pooling）使安全库存减少；但在发货之前必须先把在途库存制成成品，相当于增加了提前期，使安全库存增大。

为了证明生产线确实存在优化潜力，我们从极端情况入手，先假设喷涂工序不消耗时间，即忽略库存前移对提前期的影响。如果这种情况下安全库存减少了，那么至少可以认为：新的库存点与成品阶段之间的工序耗时不多的情况下，库存合并前移能减少安全库存。如此即证明了生产线确实存在可优化的潜力。

研究这个问题时，可以首先从最简单的情况入手：假设某注塑件在喷涂阶段分化成两种不同颜色的保险杠成品，并且两种保险杠成品的需求服从独立同分布的正态分布。我们需要证明的是，在这种情况下，如果不保留两种保险杠成品的库存，而是保留喷涂前的注塑件在制品库存，是否能在保持原有服务水平的同时，使得安全库存总量降低。

在此基础上，可以把结论推广到两种不同颜色的保险杠需求独立，但服从不同的正态分布的情况。最后可以推广到多种不同颜色的保险杠需求独立，且服从不同的正态分布的情况。这样，我们就能够证明，当保险杠需求服从独立的正态分布时，将在制品库存保留在分化前的阶段，能够使安全库存总量降低。

最后，可以通过一些数值实验，探讨系统中的各参数对改进效果的影响。

3.2 改进失效现象的研究

通过前期的一些数值实验，我们发现一些值得注意的事情：在某些情况下，改进方案不仅不能使总库存减小，反而会使总库存变大。

这个现象的成因是什么？具体表现是什么？影响它的因素有哪些？这些问题都是值得探索研究的。

我们需要从不同的需求分布、不同的服务水平入手，讨论各种参数对这种现象的影响。只有我们找到引发这种现象的参数规律之后，才能总结出我们最

希望得到的结论：如何判断企业当前环境下是否会发生改进失效，以及如何避免改进失效。

3.3 需求相关性的影响

将不同颜色相同型号的成品库存合并前移，成品之间的相关性必然会影响到方案的效果。因此需要分类讨论成品需求相互独立、需求正相关、需求负相关等情况。

独立 成品需求相互独立，意味着一种颜色的保险杠需求量对相同型号另一种颜色的保险杠需求量没有影响。这是最简单的假设，但有时候可能不符合实际情况。

正相关 某种颜色的保险杠需求变化，可能是由于该车型的市场需求变化。此时该车型的各种颜色需求可能同时变化。也就是说，相同型号不同颜色的保险杠需求可能倾向于同时增大或减小。这是生产实际中可能发生的情况。

负相关 如果某种车型的市场需求相对稳定，则意味着，一种颜色的需求增长会抑制另一种颜色的需求，表现为相同型号不同颜色的保险杠需求负相关。这也是生产实际中可能发生的情况。

不同的需求相关性会影响库存改进方案的效果。显然，成品需求负相关的情况下，此改进方案的效果将会更加突出。分析成品需求完全正相关和完全负相关的极端情况，能帮助我们明确改进方案的绩效范围，找到改进效果的上下界，从而更好地作出决策。

研究的最后，需要对企业生产线的实际数据进行统计分析，找出不同颜色相同型号的成品需求之间的实际相关性，并针对该结果做具体分析。

3.4 延迟订单的影响

在前面的思路中，我们始终默认每一期的需求对后一期没有影响，也就是说未满足的订单将会被直接舍弃。实际上，作为供应商，如果不能按时交货，有时需要将订单记下，尽快补货，并且承受一定的惩罚。这种订单称为延迟订单。在这一部分中，我们将会讨论有延迟订单的情况下，改进的具体效果。

由于前期的需求和库存会对后期造成一定的影响，此处需要引入马尔科夫过程来进行建模。特别的，如果需求和生产遵循泊松过程的话，我们可以利用

一些已有的模型来进行研究，如生灭过程。以一些现有的数学结论作为基础，就能尝试一些定量的研究，如计算改进效果的上限等。

第 4 章 优化潜力存在性

在本章中，我们将忽略相关性和提前期等一切影响因素，单纯从需求的概率分布变化来探究在制品库存前置策略对安全库存的优化作用。

4.1 安全库存的制定标准

安全库存的计算方式是总库存减去需求的期望值。由于本文讨论的所有情况都不涉及需求期望值的变动，因此在比较策略实施前后安全库存的变化时，总库存的减少量与安全库存的减少量始终是等价的。本文后续讨论中不再刻意强调安全库存与总库存的区别，以减少繁琐的数学表达，使数学公式的推导更简洁。

在讨论安全库存时，我们常用的标准有惩罚成本和服务水平。

惩罚成本是企业缺货时需要付出的成本。合同罚款就是最常见的惩罚成本之一。在汽车配件的生产过程中，配件生产商会和需求方签订合同，当生产商不能按时供货时，就需要承受一笔罚款。惩罚成本能够比较直观地反应缺货带来的损失，体现安全库存的作用。

然而，惩罚成本在使用过程中有一个难以解决的问题：如何准确确定惩罚成本。惩罚成本不仅包括实际的资金损失，还应该考虑其他隐性损失。以汽车配件生产商为例，合同规定的罚款只是表面上的惩罚成本，实际的损失还应该包括企业的声誉损失。汽车行业非常重视精益生产，缺货造成的声誉损失可能使企业丢失更多的潜在订单，甚至失去合作伙伴。这些损失比表面上的罚款更严重，但很难对其进行量化。

因此，如果仅仅使用合同的罚款数额作为惩罚成本，那么分析过程使用的惩罚成本是小于实际损失的。这种情况下，企业做出的决策，会更倾向于承受损失。这是对企业有害的。

本文中，我们将主要使用服务水平作为安全库存的制定标准。这样做的原因不仅包括以上讨论的惩罚成本的缺陷，还有给出更为科学的理由。在接下来的论述中，我们将会证明，如果使用惩罚成本作为安全库存的制定标准，与使用服务水平相比，实质上是等价的。

若只考虑惩罚成本，不考虑库存成本，则企业需要设定一个自己能接受的惩罚成本期望的最大值。我们首先分析惩罚成本的期望。设企业的库存为 K ，惩罚成本为 p ，需求服从分布 $P(X = x) = f(x)$ ，其累积分布为 $P(X < x) = F(x)$ 。则惩罚成本为

$$C(x, K) = \begin{cases} 0, & x \leq K \\ p(x - K), & x > K \end{cases} \quad (4-1)$$

如果期望存在的话，其对应的期望是

$$\begin{aligned} E[C(x, K)] &= \int_{-\infty}^{+\infty} C(x, K) f(x) dx \\ &= \int_K^{+\infty} p(x - K) f(x) dx \\ &= p(x - K) F(x) \Big|_K^{+\infty} - \int_K^{+\infty} p F(x) dx \\ &= \lim_{M \rightarrow +\infty} \left[p(M - K) - \int_K^M p F(x) dx \right] \\ &= \lim_{M \rightarrow +\infty} \left[\int_K^M p dx - \int_K^M p F(x) dx \right] \\ &= p \int_K^{+\infty} (1 - F(x)) dx \end{aligned} \quad (4-2)$$

另一方面，库存为 K 时，对应的服务水平是

$$\eta = \int_{-\infty}^K f(x) dx \quad (4-3)$$

由公式4-2和4-3可知，惩罚成本期望是随 K 单调递减的，服务水平是随 K 单调递增的。二者可以一一对应，即每个惩罚成本期望值都对应某个特定的服务水平。企业制定自己的惩罚成本期望值，和制定自己的服务水平，是等价的。

使用服务水平来计算安全库存时，不需要考虑库存成本。使用惩罚成本时，可以把库存成本也加入到模型中。接下来我们将证明，加入库存成本之后，惩罚成本与服务水平也是一一对应的。以下推导过程中使用离散的库存量和需求量。

设企业的库存为 K ，需求为 D ，惩罚成本为 p ，库存成本为 h 。则总成本为

$$C(K) = p \max\{D - K, 0\} + h \left(\frac{1}{2} \min\{D, K\} + \max\{K - D, 0\} \right) \quad (4-4)$$

企业的目标是寻找最优的库存 K ，使得库存成本和惩罚成本之和的期望最小。我们对公式4-4中的几个部分分别求期望。

$$\begin{aligned}
 E(\min\{D, K\}) &= \sum_{i=1}^{\infty} \min\{i, K\} P(D = i) \\
 &= \sum_{i=1}^{\infty} \sum_{j=1}^{\min\{i, K\}} P(D = i) \\
 &= \sum_{j=1}^K \sum_{i=j}^{\infty} P(D = i) \\
 &= \sum_{j=1}^K P(D \geq j)
 \end{aligned} \tag{4-5}$$

$$\begin{aligned}
 E(\max\{D - K, 0\}) &= -\min\{K - D, 0\} \\
 &= -(\min\{K, D\} - D) \\
 &= D - \sum_{j=1}^K P(D \geq j)
 \end{aligned} \tag{4-6}$$

$$\begin{aligned}
 E(\max\{K - D, 0\}) &= \sum_{i=1}^{\infty} \max\{K - i, 0\} P(D = i) \\
 &= \sum_{i=1}^{\infty} \sum_{j=1}^{\max\{K-i, 0\}} P(D = i) \\
 &= \sum_{j=1}^{K-1} \sum_{i=1}^j P(D = i) \\
 &= \sum_{j=1}^{K-1} P(D \leq j)
 \end{aligned} \tag{4-7}$$

将公式4-5、4-6和4-7代入公式4-4可得

$$E[C(K)] = pD + \left(\frac{h}{2} - p\right) \sum_{j=1}^K P(D \geq j) + h \sum_{j=1}^{K-1} P(D \leq j) \tag{4-8}$$

考虑库存增加时，总成本的增量

$$\begin{aligned}
 E[C(K+1)] - E[C(K)] &= \left(\frac{h}{2} - p\right) P(D \geq K+1) + h P(D \leq K) \\
 &= \frac{h}{2} - p + \left(\frac{h}{2} + p\right) P(D \leq K)
 \end{aligned} \tag{4-9}$$

公式4-9是随 K 单调递增的，因此，当 $E[C(K+1)] - E[C(K)]$ 取得 0 值时， $E[C(K)]$

取得最小值。此时有

$$P(D \leq K) = \frac{2p-h}{2p+h} \quad (4-10)$$

根据服务水平的定义，公式4-10中的 $P(D \leq K)$ 即是企业的服务水平 η 。本文讨论的改进方案仅移动在制品库存在生产线上的位置，不影响库存费用和需求分布。因此，给出任意的惩罚成本，都有一个唯一的服务水平与之对应。

通过以上的讨论，我们已经证明了，以惩罚成本和库存成本来制定安全库存的问题，可以转化为以服务水平来制定安全库存的问题。为了使数学推导更加简洁，本文余下的部分将主要使用服务水平作为安全库存的制定标准。企业可以根据实际情况选取合适的制定方式，且不影响本文中结论的适用性。

4.2 两种成品需求独立正态分布

仍然以汽车保险杠生产为例，假设某种汽车保险杠只有两种颜色，且两种颜色的需求服从独立的正态分布。第一种颜色的保险杠，需求 D_1 服从正态分布 $N_1(\mu_1, \sigma_1^2)$ ；第二种颜色的保险杠，需求 D_2 服从正态分布 $N_2(\mu_2, \sigma_2^2)$ 。企业设定的服务水平为 η 。设 z_η 是标准正态分布的 η 分位数，即 z_η 满足

$$\int_{-\infty}^{z_\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \eta$$

则此服务水平下，两种成品的库存分别为

$$\xi_1 = \mu_1 + z_\eta \sigma_1 \quad (4-11)$$

$$\xi_2 = \mu_2 + z_\eta \sigma_2 \quad (4-12)$$

现在假设我们按照改进策略，取消两种颜色的成品库存，改为保留未喷涂的在制品库存。忽略供货提前期等变化的影响。企业的服务水平仍然为 η 。为了满足该服务水平，所需的在制品库存为 ξ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为 $D = D_1 + D_2$ 。

设 D_1 、 D_2 的联合分布为 $f(x_1, x_2)$ 。因为 D_1 、 D_2 是独立的，所以 $f(x_1, x_2)$ 的概率分布为

$$\begin{aligned} f(x_1, x_2) &= f(x_1) \cdot f(x_2) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \end{aligned}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]} \quad (4-13)$$

设 D 、 D_2 的联合分布为 $g(y_1, y_2)$ 。由 $y_1 = x_1 + x_2, y_2 = x_2$ 得 $x_1 = y_1 - y_2, x_2 = y_2$ 。因此，雅可比矩阵为

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

由此可得 $g(y_1, y_2)$ 的概率分布为

$$\begin{aligned} g(y_1, y_2) &= f(x_1, x_2) \cdot |J| \\ &= f(y_1 - y_2, y_2) \cdot \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]} \end{aligned} \quad (4-14)$$

$g(y_1, y_2)$ 的边缘分布 $h(y_1)$ 即是在制品需求 D 的概率分布。对 y_2 积分可得

$$h(y_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]} dy_2 \quad (4-15)$$

令

$$A = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

则公式4-15可变形为

$$\begin{aligned} h(y_1) &= A \int_{-\infty}^{+\infty} \frac{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right] + \frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}} dy_2 \\ &= A \int_{-\infty}^{+\infty} \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2}y_2 - \left(\frac{\sigma_2}{\sigma_1\sqrt{\sigma_1^2 + \sigma_2^2}}(y_1 - \mu_1) + \frac{\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\mu_2\right)\right]^2} dy_2 \end{aligned} \quad (4-16)$$

再令

$$t = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} y_2 - \left(\frac{\sigma_2}{\sigma_1\sqrt{\sigma_1^2 + \sigma_2^2}}(y_1 - \mu_1) + \frac{\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\mu_2 \right)$$

则公式4-16可变形为

$$h(y_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (4-17)$$

公式4-17的右半部分是标准正态分布的累积概率函数，总累积概率应该为 1，即

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1$$

因此得到 $h(y_1) = A$ ，即

$$h(y_1) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}} \quad (4-18)$$

由公式4-18可知，在制品的需求 D 服从正态分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 。为了满足服务水平 η ，所需的在制品库存为

$$\xi = \mu_1 + \mu_2 + z_\eta \sqrt{\sigma_1^2 + \sigma_2^2} \quad (4-19)$$

至此，我们可以利用公式4-11、4-12和4-19将改进前后的库存总量进行比较

$$\begin{aligned} \xi_1 + \xi_2 - \xi &= \mu_1 + z_\eta \sigma_1 + \mu_2 + z_\eta \sigma_2 - \left(\mu_1 + \mu_2 + z_\eta \sqrt{\sigma_1^2 + \sigma_2^2} \right) \\ &= z_\eta \left(\sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2} \right) \\ &> z_\eta \left(\sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} \right) \\ &= z_\eta \left(\sigma_1 + \sigma_2 - \sqrt{(\sigma_1 + \sigma_2)^2} \right) \\ &= 0 \end{aligned} \quad (4-20)$$

由公式4-20可知 $\xi_1 + \xi_2 > \xi$ ，即改进后的在制品库存严格小于改进前的成品库存之和。这个结果说明，当两种不同颜色的成品需求服从独立正态分布时，通过把成品库存合并到在制品库存，能够降低企业的库存总量。

4.3 两种以上的成品需求独立正态分布

现在我们将上述结果进行推广。假设某种汽车保险杠有 N 种颜色 ($N \geq 2$)，并且服从两两相互独立的正态分布。第 i 种颜色的保险杠，需求 D_i 服从正态分

布 $N_i(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, N$ 。企业的服务水平仍为 η ，其对应的标准正态分布 η 分位数仍为 z_η 。

为了满足此服务水平，各成品需要保留的库存为

$$\xi_i = \mu_i + z_\eta \sigma_i, \quad i = 1, 2, \dots, N \quad (4-21)$$

现在假设我们按照改进策略，取消所有颜色的成品库存，改为保留未喷涂的在制品库存。为了满足服务水平 η ，所需的在制品库存为 ξ 。此时在制品库存的需求为 $D = \sum_{i=1}^N D_i$ 。前面已证任意两个独立的正态分布之和仍然服从正态分布，由于所有的 D_i 都是相互独立的，因此它们可以连续加和，其结果仍然服从正态分布。所以我们有 $D \sim N(\sum_{i=1}^N \mu_i, \sum_{i=1}^N \sigma_i^2)$ 。进而可以得到在制品库存为

$$\xi = \sum_{i=1}^N \mu_i + z_\eta \sqrt{\sum_{i=1}^N \sigma_i^2} \quad (4-22)$$

通过公式4-21和4-22比较改进前后的库存变化

$$\begin{aligned} \sum_{i=1}^N \xi_i - \xi &= \sum_{i=1}^N (\mu_i + z_\eta \sigma_i) - \left(\sum_{i=1}^N \mu_i + z_\eta \sqrt{\sum_{i=1}^N \sigma_i^2} \right) \\ &= z_\eta \left(\sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=1}^N \sigma_i^2} \right) \\ &> z_\eta \left(\sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sigma_i \sigma_j} \right) \\ &= z_\eta \left(\sum_{i=1}^N \sigma_i - \sqrt{\left(\sum_{i=1}^N \sigma_i \right)^2} \right) \\ &= 0 \end{aligned} \quad (4-23)$$

由公式4-23可知 $\sum_{i=1}^N \xi_i > \xi$ ，即改进后的在制品库存严格小于改进前的成品库存。至此，我们已证明，当成品库存服从两两相互独立的正态分布时，通过保留在制品库存能够降低企业的库存总量。

4.4 一些参数对改进效果的影响

前面已经证明了保留在制品库存的改进能够达到降低安全库存的目的，本节将讨论影响改进效果的参数，找出最适宜进行改进的环境。

到目前为止，我们探讨的过程中出现过的参数主要有需求的均值、方差和企业的服务水平。其中，需求的均值决定的是库存中固定的部分而非波动的部分，因此需求均值对于改进效果是没有影响的。我们需要考虑的是需求的方差和企业的服务水平对改进效果的影响。

为了使问题更简单，我们假设只有两种成品参与改进，它们的需求相互独立，且都服从相同的正态分布 $N(\mu, \sigma^2)$ 。由公式4-18可知，改进后的需求服从正态分布 $N(2\mu, 2\sigma^2)$ 。企业的服务水平为 η ，则改进前的安全库存为 $2 \cdot z_\eta \cdot \sigma$ ，改进后的安全库存为 $z_\eta \cdot \sqrt{2}\sigma$ 。每一个标准差 σ 和每一个服务水平 η ，都对应着一个改进前后的库存差值 $\Delta = 2z_\eta\sigma - \sqrt{2}z_\eta\sigma$ 。利用计算机程序求出 $\sigma \in [0, 50]$ 、 $\eta \in (0, 1)$ 的所有对应 Δ 值，并作图如图4.1所示。

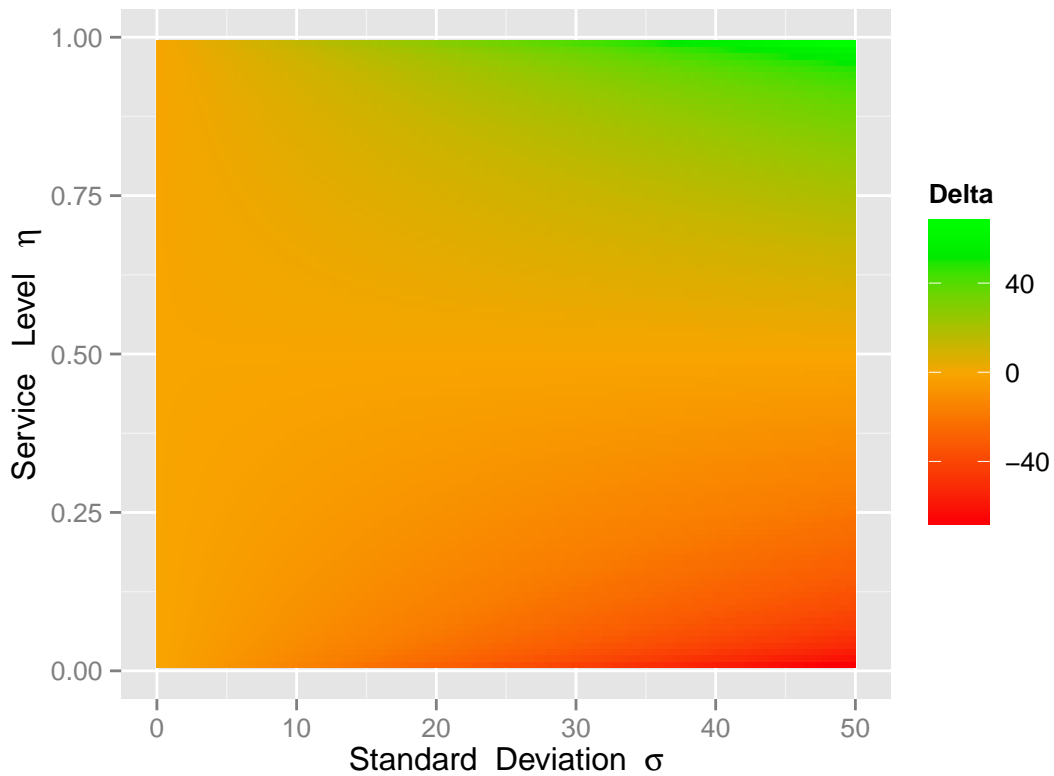


图 4.1 需求波动和服务水平对改进效果的影响

图4.1中每个点的位置代表一种 (σ, η) 组合，每个点的颜色代表该处的 Δ 值。颜色越绿，代表改进效果越好；颜色越红，代表改进效果越差。

观察图中颜色的分布区域和渐变趋势，可以得到以下结论：

1. 绿色区域集中在右上角，并向右上方渐变，说明需求波动大、企业服务水平高的环境下，改进效果是最好的，改进策略值得考虑；
2. 红色区域集中在右下角，并向右下方渐变，说明需求波动大、企业服务水平低的环境下，改进效果是最差的，改进之前需要慎重；
3. 需求波动相同的情况下，企业服务水平越高，改进的效果越好；
4. 企业服务水平相同的情况下，需求的波动越强，改进的效果越好。

除此之外还能发现，在 $\eta = 0.5$ 处出现了一条“对称轴”，并且该轴两侧的改进效果有正负之别。下一章中，我们将详细讨论这个问题。

第 5 章 改进失效现象

在第4章的证明过程中，我们通过添加项的方式对不等式进行放缩，从而证明了 $\sum_{i=1}^N \xi_i > \xi$ 是严格成立的。但是这一推导过程中，我们默认了 $z_\eta > 0$ 。由 z_η 的定义可知

$$z_\eta > 0 \iff \eta > 0.5$$

也就是说，只有当企业设定的服务水平高于 50% 时，才能够通过保留在制品库存降低企业库存总量。

当 $\eta = 0.5$ 时，有 $z_\eta = 0$ ，参考公式4-23的推导过程，可知此时 $\sum_{i=1}^N \xi_i - \xi = 0$ ，即改进前后企业的总库存没有发生变化。

当 $\eta < 0.5$ 时，有 $z_\eta < 0$ ，参考公式4-23的推导过程，可以发现此时的不等式放缩方向是相反的，应该得到 $\sum_{i=1}^N \xi_i - \xi < 0$ ，即改进后反而使得企业的总库存增加。

由此可见，当需求服从正态分布时，在某些服务水平下，将成品库存合并前移到在制品库存不仅没有起到改进效果，反而会使总库存增大。我们将这种情况称为改进失效。

本章将结合常用的需求分布函数，详细讨论改进失效现象与需求分布、服务水平等因素的关系。

5.1 对称稳定分布的参数与改进失效的关系

前面的部分主要考虑了需求为正态分布时，改进策略对库存总量的影响。之所以研究正态分布，是因为正态分布确实广泛存在于现实生活中。实际生产中的需求分布可能是各种各样的，甚至可能写不出解析式，但只要是方差有限的独立同分布，根据中心极限定理，在进行叠加的时候总会趋近于正态分布。而正态分布自身叠加仍是正态分布。因此，正态分布就像一个吸引域，从各种分布出发最终都落入正态分布。正态分布下的库存改进效果，也能在很大程度上反应实际的改进效果。

第4章中已经提到，安全库存等于总库存与需求期望值之差。回顾公式4-21和4-22可知，改进前和改进后的安全库存分别为 $z_\eta \sum_{i=1}^n \sigma_i$ 和 $z_\eta \sqrt{\sum_{i=1}^n \sigma_i^2}$ 。为了使考虑的问题更简单，我们假设所有的需求是独立同分布的 $N(\mu, \sigma^2)$ ，则改进前和改进后的安全库存分别为 $nz_\eta \sigma$ 和 $\sqrt{n}z_\eta \sigma$ 。

当 $\eta < 0.5$ 的时候，安全库存是负值，这是实际生产中一般不会出现的情况。当 $\eta > 0.5$ 时，可以看到， n 种成品合并之后的安全库存是单个成品安全库存的 \sqrt{n} 倍，而不合并的情况下则是单个成品的 n 倍。此时改进是一定有效果的。

事实上，像正态分布这种“合并后安全库存是单个成品安全库存的 $n^{1/\alpha}$ 倍”的性质，还出现在其他的对称稳定分布中。Feller (2008)^[33] 提到，对称稳定分布有 $f_{n,\alpha}(x) = f_\alpha(n^{1/\alpha}x)$ 的性质。也就是说， n 个独立同分布的对称稳定分布的随机变量 X 相加，和的概率分布与随机变量 $n^{1/\alpha}X$ 是相同的。而正态分布正是对称稳定分布取 $\alpha = 2$ 的一个特例。

稳定分布的概率密度函数为 $f(x; \alpha, \beta, c, \mu)$ ，其中参数 $\alpha \in (0, 2]$ ，影响曲线的陡峭程度； $\beta \in [-1, 1]$ ，影响曲线的偏斜程度； $c \in (0, +\infty)$ ，影响曲线的水平尺度； $\mu \in (-\infty, +\infty)$ ，影响曲线的水平中心。当 $\beta = 0$ 时，即为对称稳定分布。图 5.1 展示了 $\beta = 0, c = 1, \mu = 0$ 的对称稳定分布在不同的 α 取值下的图形，其中 $\alpha = 2$ 时即为正态分布。

广义中心极限定理指出， n 个独立同分布的随机变量（方差可以无限）相加，最终会收敛到一个稳定分布。因此，和正态分布一样，稳定分布也广泛存在于实际生活中。Newman (2005)^[34] 的一系列实证分析证实了这一点。

假设单个成品需求服从对称稳定分布 $f(x; \alpha, 0, c, \mu)$ ，则它的安全库存服从对称稳定分布 $f(x; \alpha, 0, c, 0)$ 。设此需求分布下，满足服务水平 η 需要的安全库存为 s ， n 个成品需要的总安全库存是 ns 。根据 (Feller) 所述的性质可以推知，改进后需要的安全库存为 $n^{1/\alpha}s$ 。 $\eta > 0.5$ 时 $s > 0$ ，故改进是否有效就取决于 α 的取值。将改进前后的库存作比较：

$$\frac{ns}{n^{1/\alpha}s} = n^{1-1/\alpha}$$

其中 α 的取值范围是 $(0, 2]$ 。

当 $\alpha \in (1, 2]$ 时， $n^{1-1/\alpha} > 1$ ，改进前的库存大于改进后的库存，改进有效。

当 $\alpha \in (0, 1]$ 时， $n^{1-1/\alpha} \leq 1$ ，改进前的库存不大于改进后的库存，改进无

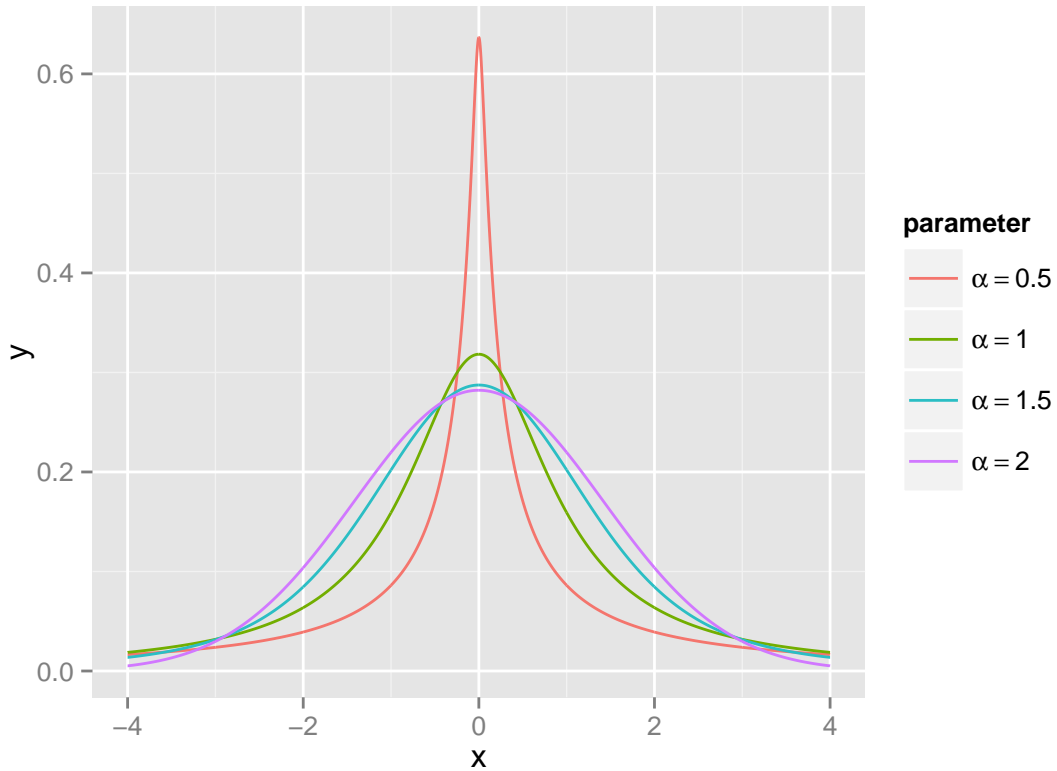


图 5.1 不同参数 α 下的对称稳定分布

效。

5.2 右偏斜的需求分布一定存在改进失效的可能

对称分布存在一个普遍的缺点——需求可能出现负值，因此我们在实际生产中也常常使用一些其他的分布函数来描述需求的分布。现实的需求不可能为负值，其概率密度函数一般不会有左侧的长尾。因此，需求分布一般都是右偏斜的。Agrawal 和 Smith (1996)^[35] 使用一些实际的需求数据，拟合一些常用的需求分布，包括泊松分布、指数分布、负二项分布等。它们都是右偏斜的分布。

假设有 n 种颜色的成品，其需求 $D_i (i = 1, 2, \dots, n)$ 独立同分布且该分布是右偏斜的，累积分布函数都为 F ，均值为 μ ，方差为 σ^2 。将成品库存合并前移到在制品库存后，在制品的总需求为 $D_n = \sum_{i=1}^n D_i$ ，其累积分布函数为 F_n ，均值为 $n\mu$ ，方差为 $n\sigma^2$ 。

设每种颜色的成品库存量为 s ，则对应的服务水平为 $\eta = F(s)$ 。由 F_n 的定

义知

$$F_n(ns) = P(D_n < ns) = P\left(\frac{\sum_{i=1}^n D_i}{n} < s\right) \quad (5-1)$$

根据中心极限定理，当 $n \rightarrow \infty$ 时， $\frac{1}{n} \sum_{i=1}^n D_i$ 的分布趋近于正态分布 $N(\mu, \sigma^2/n)$ 。设正态分布 $N(\mu, \sigma^2/n)$ 的累积分布函数为 G 。由中心极限定理得

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n D_i}{n} < s\right) = \lim_{n \rightarrow \infty} \Phi\left(\frac{s - \mu}{\sigma/\sqrt{n}}\right) = \lim_{n \rightarrow \infty} G(s) \quad (5-2)$$

其中 Φ 为标准正态分布的累积分布函数。由公式5-1和5-2可知

$$\lim_{n \rightarrow \infty} [F_n(ns) - G(s)] = 0 \quad (5-3)$$

接下来我们将证明，对任何的右偏斜需求分布，一定存在一个区间，当服务水平在此区间内时，就存在改进失效的可能。已知需求 D_i 的分布是右偏的，因此有

$$F(\mu) > 0.5 = \Phi(0) = G(\mu) \quad (5-4)$$

由 F 和 G 的连续性，至少存在一个 μ 的领域 $[\mu, s_0)$ ，使得这个区间内的所有库存值 s 都满足

$$F(s) > G(s), \quad \forall s \in [\mu, s_0)$$

令 $\delta = F(s) - G(s) > 0$ 。在公式5-3中，根据极限的定义，存在一个正数 N ，使得对任意的 $n > N$ ，都有

$$|F_n(ns) - G(s)| < \delta = F(s) - G(s) \quad (5-5)$$

公式5-5显示，对于任意的 $s \in [\mu, s_0)$ ，都存在一个 N ，使得 $n > N$ 时恒有 $F_n(ns) < F(s)$ 。

我们知道 $F(s)$ 代表需求量小于 s 的概率，即库存量 s 对应的服务水平。反过来，某个服务水平 η 也对应着一个库存量，我们将这个对应关系定义为函数 F^{-1} 。函数 F^{-1} 的表达式为

$$F^{-1}(\eta) = \inf \{s | F(s) \geq \eta\}$$

其中 \inf 表示集合的下确界。同理可定义 F_n^{-1} 。

若企业需要满足的服务水平为 η ，则改进前和改进后的库存分别为 $F^{-1}(\eta)$ 和 $F_n^{-1}(\eta)$ 。若改进后的在制品库存大于改进前的各颜色成品库存之和，就说明改进失效了。也就是说，改进失效的具体表现是 $F_n^{-1}(\eta) > nF^{-1}(\eta)$ 。下面我们将证明，当 $F^{-1}(\eta) \in [\mu, s_0)$ 时，存在改进失效的可能。

现在假设企业的设定的服务水平为 η ，满足 $\eta \in [F(\mu), F(s_0))$ ，则改进前各颜色成品的库存为 $s = F^{-1}(\eta) \in [\mu, s_0)$ 。前面已经证明，对于任意的 $s \in [\mu, s_0)$ ，都存在一个 N ，使得 $n > N$ 时恒有

$$F_n(ns) < F(s) = \eta \quad (5-6)$$

根据反函数的性质，由于累积分布函数 F 和 F_n 是单调递增的，所以 F^{-1} 和 F_n^{-1} 也是单调递增的。公式5-6显示 $F_n(ns) < \eta$ ，结合 F_n^{-1} 的单调性可知

$$F_n^{-1}(\eta) > ns \quad (5-7)$$

把 $s = F^{-1}(\eta)$ 代入公式5-7有

$$F_n^{-1}(\eta) > nF^{-1}(\eta)$$

前面已经提到，上式即为改进失效的具体表现。

通过本节分析，我们证明了，当各颜色成品需求服从独立同分布的任意一种右偏斜分布时，存在一个大于 $F^{-1}(\mu)$ 的服务水平区间，企业的服务水平位于这个区间时，存在改进失效的可能。当服务水平小于 $F^{-1}(\mu)$ 时，安全库存是负值，无需过多讨论。

5.3 任何服务水平都存在改进失效的可能

在上一节的讨论中，我们已经证明，如果需求服从独立同分布的右偏斜分布，则存在一个大于 $F^{-1}(\mu)$ 的服务水平区间，使得这个区间内的服务水平下可能发生改进失效的风险。那么，这个区间是否存在上限？当服务水平远高于 $F^{-1}(\mu)$ 的时候，是否存在某些服务水平，使得在这些服务水平下一定不会发生改进失效？我们以最简单的伯努利分布为例，来研究这个问题。

如果单个产品的需求服从伯努利分布：需求量为 d_1 的概率为 α ，需求量为 d_2 的概率为 $1 - \alpha$ 。为了保证需求是右偏的，我们设定 $d_1 < d_2$ ，且 $\alpha > 0.5$ 。

假设有两种颜色的成品，需求相互独立，且都服从上述的伯努利分布。企业的服务水平为 η ，则 η 和 α 的相对大小会对库存产生影响：当 $\eta > \alpha$ 时，每种成品需要保留的库存为 d_2 ，总库存为 $2d_2$ ；当 $\eta \leq \alpha$ 时，每种成品需要保留的库存为 d_1 ，总库存为 $2d_1$ 。因此，改进前的服务水平与库存的关系如表5.1所示。

表 5.1 改进前服务水平与库存的关系

服务水平	单个成品库存	总库存
$\eta > \alpha$	d_2	$2d_2$
$\eta \leq \alpha$	d_1	$2d_1$

现将这两种成品的库存合并保留到在制品库存。首先计算这两种成品需求的联合分布

表 5.2 两种成品需求的联合分布

第一种成品需求	第二种成品需求	概率
d_1	d_1	α^2
d_1	d_2	$\alpha(1 - \alpha)$
d_2	d_1	$\alpha(1 - \alpha)$
d_2	d_2	$(1 - \alpha)^2$

由表5.2可以得到合并后在制品的需求分布，如表5.3所示。

表 5.3 在制品需求分布

需求量	概率	累积概率
$2d_1$	α^2	α^2
$d_1 + d_2$	$2\alpha(1 - \alpha)$	$2\alpha - \alpha^2$
$2d_2$	$(1 - \alpha)^2$	1

企业的服务水平仍为 η ，根据表5.3在制品的需求累积概率分布与 η 的相对大小，可以得出需要保留的在制品库存：当 $\eta > 2\alpha - \alpha^2$ 时，在制品需要保留的库存为 $2d_2$ ；当 $\alpha^2 < \eta \leq 2\alpha - \alpha^2$ 时，在制品需要保留的库存为 $d_1 + d_2$ ；当 $\eta \leq \alpha^2$ 时，在制品需要保留的库存为 $2d_1$ 。因此，改进后的服务水平与库存的关系如表5.4所示。

表 5.4 改进后服务水平与库存的关系

服务水平	在制品库存
$\eta > 2\alpha - \alpha^2$	$2d_2$
$\alpha^2 < \eta \leq 2\alpha - \alpha^2$	$d_1 + d_2$
$\eta \leq \alpha^2$	$2d_1$

对比表5.1和表5.4可知：当 $\eta \leq \alpha^2$ 时，改进前后的库存都为 $2d_1$ ；当 $\alpha^2 < \eta \leq \alpha$ 时，改进后的库存 $2d_2$ 大于改进前的库存 $2d_1$ ； $\alpha < \eta \leq 2\alpha - \alpha^2$ 时，改进后的库存 $d_1 + d_2$ 小于改进前的库存 $2d_2$ ；当 $\eta > 2\alpha - \alpha^2$ 时，改进前后的库存都为 $2d_2$ 。汇总如表5.5所示。

表 5.5 不同服务水平下改进前后库存对比

服务水平	改进前库存量	改进后库存量
$\eta \leq \alpha^2$	$2d_1$	$2d_1$
$\alpha^2 < \eta \leq \alpha$	$2d_1$	$2d_2$
$\alpha < \eta \leq 2\alpha - \alpha^2$	$2d_2$	$d_1 + d_2$
$\eta > 2\alpha - \alpha^2$	$2d_2$	$2d_2$

从表5.5中可以看出，当 $\alpha^2 < \eta \leq \alpha$ 时，一定会发生改进失效。显然，对任意的服务水平 η ，一定存在某些 α ，使得 $\alpha^2 < \eta \leq \alpha$ 。因此，任意的服务水平下都能够找到某些右偏斜的伯努利分布，使得改进失效。

上述结论很容易推广到三种以上成品合并库存的情况。假设有 n 种颜色的成品，需求服从独立同分布的伯努利分布，需求量为 d_1 的概率为 α ，需求量为 d_2 的概率为 $1 - \alpha$ ， $d_1 < d_2$ ， $\alpha > 0.5$ 。易知，当服务水平 $\eta \leq \alpha$ 时，需要保留的单个成品库存量为 d_1 ，总库存量为 nd_1 。将它们的库存合并到在制品库存后，在制品的需求服从二项分布，即

$$P(kd_1 + (n - k)d_2) = C_n^k \alpha^k (1 - \alpha)^{n-k}$$

取 $k = n$ 可知 $P(nd_1) = \alpha^n$ ，因此当服务水平 $\eta > \alpha^n$ 时，需要保留的在制品库存量是大于 nd_1 的。当服务水平满足 $\alpha^n < \eta \leq \alpha^n$ 时，一定发生改进失效。

通过以上讨论我们证明了，将 n 种颜色的成品库存合并为在制品库存时，对于任意的服务水平 η ，都可能在某些右偏的需求分布下发生改进失效。因此，

企业在做出决策之前不能盲目乐观，必须根据自身的服务水平和实际的需求分布，谨慎地进行模拟和推算，防止发生改进失效。

5.4 泊松分布下发生改进失效的区间图示

在前两节的讨论中，我们从理论上证明了改进失效现象的存在性。但是，对改进失效现象发生的具体范围和如何避免，仍然没有清晰直观的认识。本节中，我们将以实际生产中常用的泊松分布为例，展示改进失效的实际影响和避免改进失效的方法。

为了精确描述库存总量在改进前后的变化，我们首先需要知道改进前后需求分布的变化。设需求 D 服从泊松分布 $Po(\lambda)$ ，则其概率为

$$P(D=x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x \in \mathbb{N}$$

设企业的服务水平为 η ，满足此服务水平所需的库存为 ξ 。则库存 ξ 的表达式为

$$\xi = \inf \left\{ \xi \in \mathbb{N} \left| \sum_{x=0}^{\xi} \frac{e^{-\lambda} \lambda^x}{x!} \geq \eta \right. \right\} \quad (5-8)$$

其中符号 \inf 表示集合的下确界。

设两种颜色的成品需求服从相互独立的泊松分布 $Po(\lambda_1)$ 、 $Po(\lambda_2)$ 。企业的服务水平为 η 。则两种成品需要保持的库存 ξ_1 、 ξ_2 分别为

$$\xi_1 = \inf \left\{ \xi_1 \in \mathbb{N} \left| \sum_{x=0}^{\xi_1} \frac{e^{-\lambda_1} \lambda_1^x}{x!} \geq \eta \right. \right\} \quad (5-9)$$

$$\xi_2 = \inf \left\{ \xi_2 \in \mathbb{N} \left| \sum_{x=0}^{\xi_2} \frac{e^{-\lambda_2} \lambda_2^x}{x!} \geq \eta \right. \right\} \quad (5-10)$$

现在假设取消两种颜色的成品库存，改为保留共同的在制品库存。企业的服务水平仍然为 η 。为了满足该服务水平，所需的在制品库存为 ξ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为 $D = D_1 + D_2$ 。

设 D_1 、 D_2 的联合分布为 $f(x_1, x_2)$ 。因为 D_1 、 D_2 是独立的，所以 $f(x_1, x_2)$ 的概率分布为

$$f(x_1, x_2) = f(x_1)f(x_2)$$

$$\begin{aligned}
&= \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2} \lambda_2^{x_2}}{x_2!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!}
\end{aligned} \tag{5-11}$$

设 D, D_2 的联合分布为 $g(y_1, y_2)$ 。由 $y_1 = x_1 + x_2, y_2 = x_2$ 得 $x_1 = y_1 - y_2, x_2 = y_2$ 。与正态分布下的证明过程类似，我们求出雅可比矩阵

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

由此可得 $g(y_1, y_2)$ 的概率分布为

$$\begin{aligned}
g(y_1, y_2) &= f(y_1 - y_2, y_2) \cdot |J| \\
&= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!}
\end{aligned} \tag{5-12}$$

然后求边缘分布 $h(y_1)$ 。与前面不同的是，泊松分布的自变量 $y_1 - y_2 \geq 0$ ，因此有 $y_2 \leq y_1$ 。

$$\begin{aligned}
h(y_1) &= \sum_{y_2=0}^{y_1} g(y_1, y_2) = \sum_{y_2=0}^{y_1} \frac{e^{-(\lambda_1+\lambda_2)} \lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^{y_1}}{y_1!} \sum_{y_2=0}^{y_1} \frac{y_1!}{y_2! (y_1 - y_2)!} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{y_1-y_2} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{y_2}
\end{aligned} \tag{5-13}$$

公式5-13的右半部分是二项分布 $B(y_1, \frac{\lambda_2}{\lambda_1+\lambda_2})$ 的全部累积概率，其和应该为1，即

$$\sum_{y_2=0}^{y_1} \frac{y_1!}{y_2! (y_1 - y_2)!} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{y_1-y_2} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{y_2} = 1$$

因此得到边缘分布为

$$h(y_1) = \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^{y_1}}{y_1!} \tag{5-14}$$

由公式5-14可知，在制品的总需求 D 服从泊松分布 $Po(\lambda_1 + \lambda_2)$ 。所以应保留的在制品库存为

$$\xi = \inf \left\{ \xi \in \mathbb{N} \left| \sum_{x=0}^{\xi} \frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1 + \lambda_2)^x}{x!} \geq \eta \right. \right\} \tag{5-15}$$

根据公式5-9、5-10和5-15，如果改进前的需求分布服从 $Po(\lambda_1)$ 和 $Po(\lambda_2)$ ，则改进后的需求服从 $Po(\lambda_1 + \lambda_2)$ 。据此，我们就可以算出各种情况下改进前后的库存具体数值，从而进行比较。

为了方便作图，我们假设两种成品的需求是独立同分布的，都服从泊松分布 $Po(\lambda)$ ，则改进后的在制品需求服从泊松分布 $Po(2\lambda)$ 。企业的服务水平为 η 。 $Po_{\lambda}^{-1}(\eta)$ 表示改进前的每种成品的库存， $Po_{2\lambda}^{-1}(\eta)$ 表示改进后的在制品总库存。每一个参数 λ 和每一个服务水平 η ，都对应着一个改进前后的库存差值 $\Delta = 2Po_{\lambda}^{-1}(\eta) - Po_{2\lambda}^{-1}(\eta)$ 。利用计算机程序求出 $\lambda \in (0, 100)$ 、 $\eta \in (0.5, 1)$ 的所有对应 Δ 值，并作图如图5.2所示。

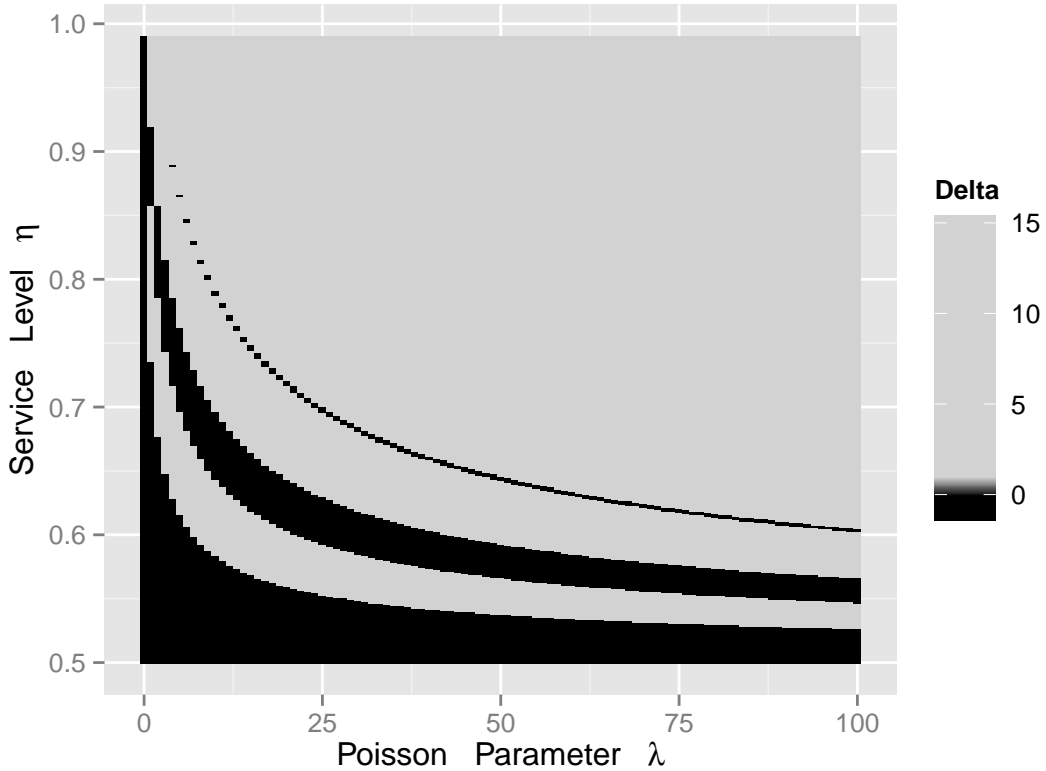


图 5.2 泊松分布下的改进失效区间图示

图5.2中每个点的位置代表一种 (λ, η) 组合，每个点的颜色代表该处的 Δ 值。如果 $\Delta \leq 0$ ，则该点为黑色；如果 $\Delta > 0$ ，则该点为浅灰色。因此，图中的黑色区域代表发生改进失效的区域，灰色区域代表改进有效果的区域。

由于泊松分布的离散性，图5.2中的黑色区域边缘存在锯齿，甚至是断断续

续的。从这幅图形中，我们能够得到很多有效的信息。

首先，它验证了前两节证明的一些论断。作为右偏斜分布的泊松分布，在服务水平大于 0.5 的某个区间内，确实出现了改进失效的一片区域；另外，当 $\lambda \rightarrow 0$ 时，黑色区域的延伸趋势也验证了上一节的论断，即不存在“不会发生改进失效的服务水平”。

其次，黑色区域在横轴方向的变化趋势告诉我们，需求量越大，发生改进失效的服务水平区间就越小，也就是说越不容易发生改进失效。当 λ 增大到一定程度时，泊松分布会趋近于正态分布。我们已经了解到，正态分布在 $\eta > 0.5$ 的时候是不会出现改进失效的。因此，随着 λ 的增长，黑色区域会越来越小，最终小到可以忽略的程度。结合上一节中伯努利分布的讨论可以看出，对于右偏斜的需求分布，偏斜程度越大，越可能发生改进失效。

最后，黑色区域在纵轴方向的变化趋势显示，服务水平越高，发生改进失效的区域越少。也就是说，这样的改进方法对于服务水平高的企业是有利的。虽然理论上不存在“不会发生改进失效的服务水平”，但是如果已经知道了需求的参数 λ 的一个确定的下限，就能够指出一个对应的服务水平的安全线，使得服务水平高于这个安全线时不可能发生改进失效。

5.5 第二类服务水平与改进失效

到目前为止，我们在提到服务水平时，都是指第一类服务水平，即得到满足的订单数占总订单数的比例。在实际生产中，也有一些企业采用第二类服务水平作为库存的制定标准。第二类服务水平是指得到满足的需求量占总需求量的比例。如果库存不足以满足需求，就从需求中减去现有的全部库存，剩余的量才作为未满足的需求量。下面我们将证明，如果改进前后都以相同的第二类服务水平作为制定库存的标准，则不可能发生改进失效。

我们知道，发生改进失效的判定标准是“服务水平不变的情况下，改进后的总库存大于改进前的总库存”。事实上，因为服务水平是随库存增加而单调增加的，因此上述判定标准等价于“总库存不变的情况下，改进后的服务水平低于改进前的服务水平”。所以，如果保持改进前后库存总量不变，只需证明改进后的第二类服务水平不可能低于改进前的第二类服务水平即可。

设两种成品的需求 D_1 、 D_2 相互独立，概率密度函数分别为 $f_1(x_1)$ 和 $f_2(x_2)$ 。企业的服务水平（第二类服务水平）为 η 。该服务水平下，两种成品的库存分

别为 ξ_1 和 ξ_2 。根据第二类服务水平的定义，得到满足的需求量与总需求量之比的期望值为第二类服务水平。两种成品得到满足的需求量分别为 $\min(x_1, \xi_1)$ 和 $\min(x_2, \xi_2)$ ，因此有

$$\eta = \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1)}{x_1} f(x_1) dx_1 = \int_{-\infty}^{+\infty} \frac{\min(x_2, \xi_2)}{x_2} f(x_2) dx_2$$

考虑两种成品需求的联合分布 $f(x_1, x_2)$ 。由于需求是相互独立的，因此有 $f(x_1, x_2) = f(x_1)f(x_2)$ 。将两种成品看做一个整体，则得到满足的需求量为 $\min(x_1, \xi_1) + \min(x_2, \xi_2)$ ，总需求量为 $x_1 + x_2$ 。因为它们各自的服务水平都是 η ，所以综合服务水平也是 η 。因此有

$$\eta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1) + \min(x_2, \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \quad (5-16)$$

现将两种成品库存合并到在制品库存，保持总量不变，即合并后的库存为 $\xi = \xi_1 + \xi_2$ 。合并后的服务水平为 η' 。合并后的总需求量为 $x_1 + x_2$ ，其中得到满足的需求量为 $\min(x_1 + x_2, \xi_1 + \xi_2)$ ，因此合并后的服务水平为

$$\eta' = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1 + x_2, \xi_1 + \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \quad (5-17)$$

为了比较 η 和 η' 的大小，我们先证明 $\min(x_1, \xi_1) + \min(x_2, \xi_2) \leq \min(x_1 + x_2, \xi_1 + \xi_2)$ 。设 $m_1 = \min(x_1, \xi_1)$ ， $m_2 = \min(x_2, \xi_2)$ ，则有 $m_1 \leq x_1$ ， $m_1 \leq \xi_1$ ， $m_2 \leq x_2$ ， $m_2 \leq \xi_2$ 。因此 $x_1 + x_2 \geq m_1 + m_2$ ， $\xi_1 + \xi_2 \geq m_1 + m_2$ 。由此可证

$$\begin{aligned} \min(x_1 + x_2, \xi_1 + \xi_2) &\geq \min(m_1 + m_2, m_1 + m_2) \\ &= m_1 + m_2 \\ &= \min(x_1, \xi_1) + \min(x_2, \xi_2) \end{aligned} \quad (5-18)$$

将不等式5-18代入公式5-16和5-17中，就能比较 η 和 η' 的大小，具体过程如下：

$$\begin{aligned} \eta' &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1 + x_2, \xi_1 + \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \\ &\geq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\min(x_1, \xi_1) + \min(x_2, \xi_2)}{x_1 + x_2} f(x_1)f(x_2) dx_1 dx_2 \\ &= \eta \end{aligned} \quad (5-19)$$

由公式5-19可知，总库存保持不变的情况下，改进后的第二类服务水平不低于改进前的第二类服务水平。因此，保持第二类服务水平不变的情况下，改进后的库存不会大于改进前的库存。

至此，我们证明了本小节开头提出的观点，即“如果改进前后都以相同的第二类服务水平作为制定库存的标准，则不可能发生改进失效”。因此，对于使用第二类服务水平作为库存制定标准的企业来说，这样的改进是值得考虑的。

5.6 避免改进失效

通过本章的探讨，我们初步了解了改进失效现象的影响因素和具体表现。作为对本章的一个总结，我们在此简单归纳各种参数对改进失效的影响，作为企业改进前需要考虑的因素：

1. 如果需求服从正态分布，只需保证企业的服务水平大于 0.5，即可避免改进失效；
2. 如果需求服从对称稳定分布，需要保证决定分布陡峭程度的参数 α 的值在 (1,2] 区间内，才能避免改进失效；
3. 如果需求服从右偏斜的分布，需要关注需求分布的偏斜程度和企业的服务水平。分布的偏斜程度越大，越可能发生改进失效；企业的服务水平越低，越可能发生改进失效。

总而言之，企业做出改进决策之前，首先需要考虑的就是需求分布的特点，其次需要检查自身的服务水平在这个需求分布下有无改进失效的可能。

第 6 章 需求相关性对改进效果的影响

在之前的讨论中，我们总是假设成品的需求分布是相互独立的。事实上，现实中的两种相似成品的需求往往存在一些相关性。本章将讨论成品需求之间的相关性对改进方案的效果的影响。

6.1 需求服从二维正态分布的情况

为了构造相关系数为 ρ 的两个正态分布的需求变量，我们假设两种颜色的成品需求 (D_1, D_2) 服从二维正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，则它们的联合分布为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]} \quad (6-1)$$

对 x_2 积分，可以得到 D_1 的边缘分布

$$f_1(x_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]} dx_2 \quad (6-2)$$

令

$$A = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}$$

则公式6-2可变形为

$$f_1(x_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} e^{-\frac{1}{2} \left[\frac{\sigma_1(x_2-\mu_2)-\rho\sigma_2(x_1-\mu_1)}{\sqrt{1-\rho^2}\sigma_1\sigma_2} \right]^2} dx_2 \quad (6-3)$$

再令

$$t = \frac{\sigma_1(x_2-\mu_2)-\rho\sigma_2(x_1-\mu_1)}{\sqrt{1-\rho^2}\sigma_1\sigma_2}$$

则公式6-3可变形为

$$f_1(x_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (6-4)$$

公式6-4的有半部分是标准正态分布的累积概率，总累积概率应为 1，即

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1$$

因此得到 $f_1(x_1) = A$ ，即

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \quad (6-5)$$

同理有

$$f_2(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \quad (6-6)$$

公式6-5和6-6表明，当需求 D_1 、 D_2 服从二维正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 时，它们分别服从一维正态分布 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 。下面证明，正态分布的两个变量 D_1 、 D_2 之间的相关系数为 ρ 。

由相关系数的定义可知

$$\begin{aligned} \text{Corr}(D_1, D_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \int_{-\infty}^{+\infty} \frac{x_1 - \mu_1}{\sigma_1} f(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (6-7)$$

与公式6-3类似，令

$$t = \frac{\sigma_2(x_1 - \mu_1) - \rho\sigma_1(x_2 - \mu_2)}{\sqrt{1 - \rho^2}\sigma_1\sigma_2}$$

则公式6-7变形为

$$\text{Corr}(D_1, D_2) = \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \int_{-\infty}^{+\infty} \frac{e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}}}{2\pi\sigma_2} \left(t \sqrt{1 - \rho^2} + \frac{\rho(x_2 - \mu_2)}{\sigma_2} \right) e^{-\frac{t^2}{2}} dt dx_2 \quad (6-8)$$

易知

$$\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt = 0$$

且

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

因此公式6-8可得

$$\begin{aligned}\text{Corr}(D_1, D_2) &= \int_{-\infty}^{+\infty} \frac{x_2 - \mu_2}{\sigma_2} \left[\frac{\rho(x_2 - \mu_2)}{\sqrt{2\pi}\sigma_2^2} \right] dx_2 \\ &= \frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 e^{-\frac{1}{2} \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2} d \left(\frac{x_2 - \mu_2}{\sigma_2} \right)\end{aligned}\quad (6-9)$$

令

$$w = \frac{x_2 - \mu_2}{\sigma_2}$$

则公式6-9可变形为

$$\begin{aligned}\text{Corr}(D_1, D_2) &= \frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w^2 e^{-\frac{w^2}{2}} dw \\ &= -\frac{\rho}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w d \left(e^{-\frac{w^2}{2}} \right) \\ &= -\frac{\rho}{\sqrt{2\pi}} \left[w e^{-\frac{w^2}{2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{w^2}{2}} dw \right] \\ &= -\frac{\rho}{\sqrt{2\pi}} (0 - \sqrt{2\pi}) \\ &= \rho\end{aligned}\quad (6-10)$$

至此，我们已构造出两个相关性为 ρ 的需求变量 D_1 和 D_2 ，它们各自服从正态分布 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ ，并且知道它们的联合分布服从二维正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。

设企业的服务水平为 η ，对应的标准正态分布分位数为 z_η 。则两种颜色的成品需要保留的库存 ξ_1 、 ξ_2 分别为

$$\xi_1 = \mu_1 + z_\eta \sigma_1 \quad (6-11)$$

$$\xi_2 = \mu_2 + z_\eta \sigma_2 \quad (6-12)$$

现在假设我们按照改进策略，取消两种颜色的成品库存，改为保留未喷涂的在制品库存。忽略供货提前期等变化的影响。企业的服务水平仍然为 η 。为了满足该服务水平，所需的在制品库存为 ξ 。此时的在制品库存需要同时应对两种颜色的成品需求，因此，对未喷涂的在制品需求为 $D = D_1 + D_2$ 。设 D 、 D_2 的联合分布为 $g(y_1, y_2)$ 。由 $y_1 = x_1 + x_2, y_2 = x_2$ 得 $x_1 = y_1 - y_2, x_2 = y_2$ 。因此，雅

可比矩阵为

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

由此可得 $g(y_1, y_2)$ 的概率分布为

$$\begin{aligned} g(y_1, y_2) &= f(x_1, x_2) \cdot |J| \\ &= f(y_1 - y_2, y_2) \cdot \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(y_1-y_2-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} \right]} \end{aligned} \quad (6-13)$$

$g(y_1, y_2)$ 的边缘分布 $h(y_1)$ 即是在制品需求 D 的概率分布。对 y_2 积分可得

$$h(y_1) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(y_1-y_2-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(y_1-y_2-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} \right]} dy_2 \quad (6-14)$$

令

$$A = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)}}$$

则公式6-14可变形为

$$h(y_1) = A \int_{-\infty}^{+\infty} \frac{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \left[y_2 - \frac{(y_1 - \mu_1)\sigma_2^2 + \mu_2\sigma_1^2 + \rho\sigma_1\sigma_2 y_1}{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right]^2} dy_2 \quad (6-15)$$

再令

$$t = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2(1-\rho^2)}} \left[y_2 - \frac{(y_1 - \mu_1)\sigma_2^2 + \mu_2\sigma_1^2 + \rho\sigma_1\sigma_2 y_1}{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right]$$

则公式6-15可变形为

$$h(y_1) = A \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (6-16)$$

公式6-16的右半部分是标准正态分布的累积概率函数，总累积概率应该为 1，因

此得到 $h(y_1) = A$, 即

$$h(y_1) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)}} e^{-\frac{[y_1 - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)}} \quad (6-17)$$

由以上推导可知, 在制品需求 D 服从正态分布 $N(\mu, \sigma^2)$, 其中 $\mu = \mu_1 + \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$ 。因此, 需要保留的在制品库存 ξ 为

$$\xi = \mu + z_\alpha \sigma = \mu_1 + \mu_2 + z_\alpha \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \quad (6-18)$$

根据公式6-11、6-12和6-18, 可以比较改进前后的库存变化。

$$\begin{aligned} \xi_1 + \xi_2 - \xi &= \mu_1 + z_\alpha \sigma_1 + \mu_2 + z_\alpha \sigma_2 - \left(\mu_1 + \mu_2 + z_\alpha \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right) \\ &= z_\alpha \left(\sigma_1 + \sigma_2 - \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \right) \end{aligned} \quad (6-19)$$

我们知道 $\sigma_1 > 0$, $\sigma_2 > 0$ 以及 $|\rho| \leq 1$, 因此有

$$\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} \leq \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} = \sigma_1 + \sigma_2 \quad (6-20)$$

将公式6-20代入公式6-19, 得

$$\xi_1 + \xi_2 - \xi \geq z_\alpha [\sigma_1 + \sigma_2 - (\sigma_1 + \sigma_2)] = 0 \quad (6-21)$$

公式6-21的结果表明, 对于需求有相关性、服从二维正态分布的两种成品, 如果取消成品库存, 改为保留共同的在制品库存, 在相同的服务水平下, 改进后的在制品库存不会比改进前的两种成品库存之和更大。

6.2 需求服从多维正态分布的情况

现在我们将上述结果进行推广。假设某成品有 N 种颜色 ($N \geq 2$), 它们的需求服从多维正态分布, 且两两之间的相关系数为 ρ_{ij} 。则每种颜色的库存为

$$\xi_i = \mu_i + z_\alpha \sigma_i, \quad i = 1, 2, \dots, N \quad (6-22)$$

假设取消全部成品库存, 改为保留共同的在制品库存, 则在制品库存 ξ 应为

$$\xi = \sum_{i=1}^N \mu_i + z_\alpha \sqrt{\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \quad (6-23)$$

根据公式6-22和6-23作比较

$$\begin{aligned}
\sum_{i=1}^N \xi_i - \xi &= \sum_{i=1}^N (\mu_i + z_\alpha \sigma_i) - \left(\sum_{i=1}^N \mu_i + z_\alpha \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \right) \\
&= z_\alpha \left(\sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j} \right) \\
&\geq z_\alpha \left(\sum_{i=1}^N \sigma_i - \sqrt{\sum_{i=0}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sigma_i \sigma_j} \right) \\
&= 0
\end{aligned} \tag{6-24}$$

公式6-24表明，对于两种以上颜色、需求具有相关性的成品，改进后的在制品库存不会比改进前的各颜色成品库存之和更大。

6.3 需求相关性影响总结

公式6-23中， ρ_{ij} 的取值范围是 $[-1, 1]$ 。容易发现，在其他参数不变的情况下，当所有的 ρ_{ij} 都同时满足 $\rho_{ij} = -1$ 时，在制品库存 ξ 取得最小值；当所有的 ρ_{ij} 都同时满足 $\rho_{ij} = 1$ 时，在制品库存 ξ 取得最大值。

$\rho_{ij} = -1$ 时，各种颜色的成品需求都是完全负相关的。一种颜色的成品需求增加，就会抑制其他颜色的需求。这种情况主要发生在需求总量稳定、需求偏好不确定的条件下。在制品库存 ξ 取得最小值，说明此时采取保留在制品库存而不是成品库存，优化的收益最大。这与我们生活经验是相符的。比如全校毕业生征订毕业衫，总需求量几乎是确定的，每个人一般只订一件，且每个人偏好的颜色不同。此时，提前给每种颜色的毕业衫印制足量的库存显然是不妥的，一般会采用先收集订单再印制的方式。

$\rho_{ij} = 1$ 时，各种颜色成品需求都是完全正相关的。一种颜色的成品需求增加，会导致其他颜色的需求同步增加。在制品库存 ξ 取得最大值，说明此时采取保留在制品库存的策略，收益最小。事实上， $\rho_{ij} = 1$ 时，公式6-24中的等号是成立的，即 $\sum_{i=1}^N \xi_i - \xi = 0$ 。换句话说，此时采取保留在制品库存的改进措施毫无效果。

其他情况下，不论各成品的需求是正相关还是负相关，只要不是完全正相关（ $\rho_{ij} = 1$ ），都能从保留在制品库存的策略中获得收益。获得收益的大小取决

于公式6-24中计算的改进前后的库存之差，该式体现了相关系数 ρ 的正负与改进效果的关系。 $\rho < 0$ 时，需求负相关。与需求相互独立的情况相比，此时的改进收益更大； $\rho > 0$ 时，需求正相关。与需求相互独立的情况相比，此时的改进收益更小。

因此，需求的相关性是企业在做出改进决策之前需要考虑的重要因素。如果需求呈现负相关，企业可能更倾向于做出改进的决策；反之，则需要谨慎考虑改进效果是否值得为之付出努力。

6.4 需求相关性实例

本文是以汽车行业为课题背景的，研究相同型号不同颜色的保险杠能否合并成品库存为在制品库存。本节将研究一些汽车行业的实例，并探讨改进效果。由于汽车的销量数据中很少包括颜色信息，我们略为折衷，研究使用相同部件的不同车型的销量相关性。本节的数据来源为新浪汽车^{[36][37]}，详细数据在附录中，见表格B.1。

骊威 (LIVINA) 和骐达 (TIIDA) 是东风日产的两个汽车品牌，它们有很多相同型号的配件，如保险杠、汽油格等。如果使用类似本文的改进方式，将它们的配件库存进行合并，能够降低总库存量吗？我们首先观察两种汽车在过去两年中的销售数据，如图6.1所示。



图 6.1 骊威、骐达月销售量

图6.1显示，两种车型的销量可能存在较强的正相关。我们通过散点图和线性拟合来进一步观察，如图6.2所示。

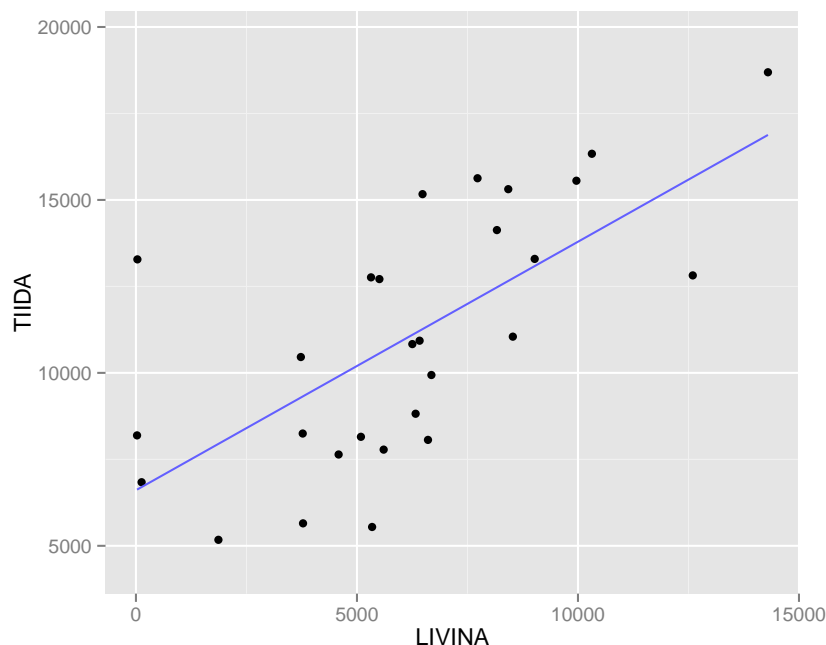


图 6.2 骊威、骐达销量散点图

图6.2中也显示了两种汽车销量的正相关性。经过计算，骊威和骐达两者销量的相关系数为 $\rho = 0.67$ 。

再选择另一组数据进行比较。英朗 (EXCELLE) 和英朗 XT (EXCELLEXT) 是上海通用别克两个汽车品牌，他们也有很多相同型号的配件，如主控板、火花塞等。这两个品牌过去两年的销售数据如图6.3所示。

图6.3中的两条曲线变化趋势基本相同，显示了比较强的正相关。相应的散点图和线性拟合如图6.4所示。

经过计算，英朗和英朗 XT 两者销量的相关系数为 $\rho = 0.82$ 。这两个品牌销量的相关性比骊威、骐达的相关性更强。同时，这两组数据说明，汽车行业中一些定位比较接近，有较多共同配件的汽车，它们的销量很可能是正相关的。

在本章的讨论中，我们已经知道，需求之间的正相关性会使得改进的效果减弱，改进前后的库存差别不明显。因此，对于汽车行业来说，是否要通过合并部件库存来降低总库存，是需要谨慎考虑，仔细论证的一件事情。

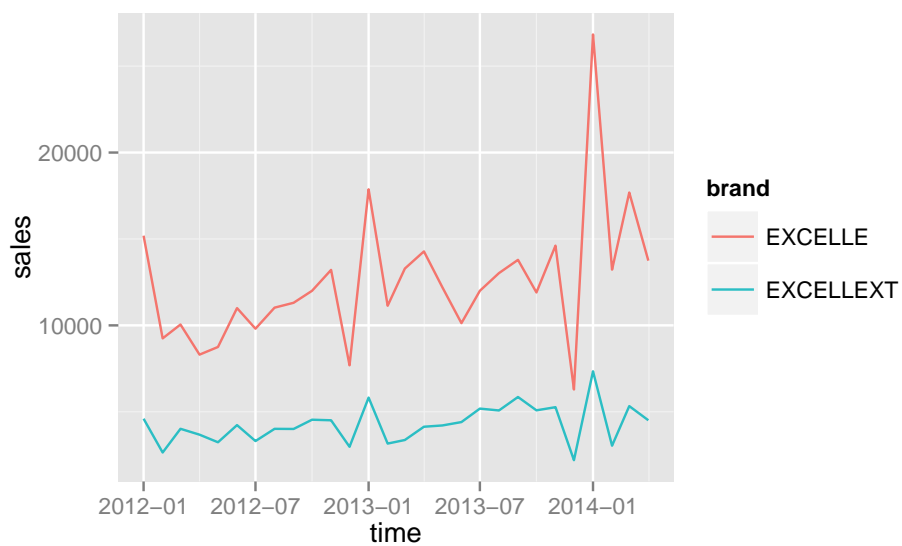


图 6.3 英朗、英朗 XT 月销售量

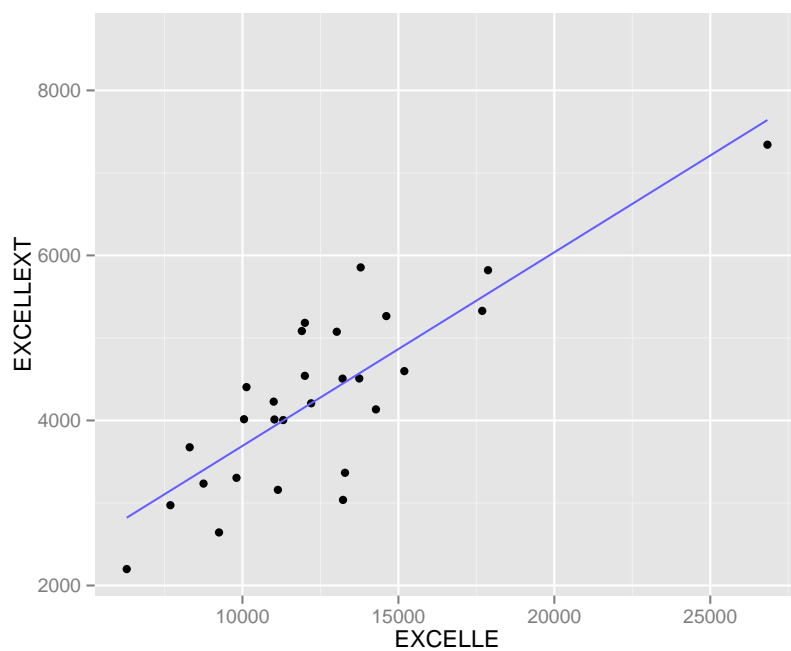


图 6.4 英朗、英朗 XT 销量散点图

第 7 章 延迟订单对改进效果的影响

在之前的讨论中，我们始终默认每一期的需求对后一期没有影响，也就是说未满足的订单将会被直接舍弃。实际上，作为供应商，如果不能按时交货，有时需要将订单记下，尽快补货，并且承受一定的惩罚。这种订单称为延迟订单。本章中，我们将讨论延迟订单对改进效果的影响。与前文不同的是，由于引入惩罚成本，本章中评价改进效果的指标不再是服务水平，而是考虑降低总成本。

7.1 考虑延迟订单成本的任意更新过程

以汽车保险杠的生产线为例，为了使讨论简化，我们假设注塑环节需要的时间远大于喷涂环节，从而忽略喷涂环节的生产时间。也就是说，改进前后相比，单件产品的生产时间是不变的。

假设有 N 种颜色的成品，每种产品的需求都是独立到来的。每当需求到来时，如果该产品有库存，则立即满足该需求，且注塑机产生一个生产订单；如果没有库存，则产生一个延迟订单，并且在注塑机处产生一个生产订单。所有的生产订单采用先到先服务的策略。定义变量如下：

s_i : 每种颜色初始库存；

h : 库存成本率（成品与在制品相同）；

b : 延迟订单成本率（成品与在制品相同）；

$D_i(t)$: 直到 t 时刻， i 产品的总需求；

$M_i(t)$: 直到 t 时刻， i 产品的总产量；

$I_i(t)$: t 时刻的 i 产品库存， $I_i(t) = s_i + M_i(t) - \min\{s_i + M_i(t), D_i(t)\}$ ；

$B_i(t)$: t 时刻的 i 产品延迟订单， $B_i(t) = D_i(t) - \min\{s_i + M_i(t), D_i(t)\}$ ；

$\phi_{i,N}(t)$: t 时刻 i 产品的成本率， $\phi_{i,N}(t) = hI_i(t) + bB_i(t)$ ；

$\phi_N(t)$: t 时刻所有产品的成本率， $\phi_N(t) = \sum_{i=1}^N \phi_{i,N}(t)$ ；

$\Phi_N(t)$: 从初始时刻到 t 时刻的总成本， $\Phi_N(t) = \int_0^t \phi_N(u) du$ ；

s : 改进后的在制品初始库存；

$D(t)$: 直到 t 时刻，改进后的在制品总需求；

$M(t)$: 直到 t 时刻，改进后的在制品总产量；

$I(t)$: t 时刻的在制品库存, $I(t) = s + M(t) - \min\{s + M(t), D(t)\}$;

$B(t)$: t 时刻的在制品延迟订单, $B(t) = D(t) - \min\{s + M(t), D(t)\}$;

$\phi(t)$: t 时刻在制品的成本率, $\phi(t) = hI(t) + bB(t)$;

$\Phi(t)$: 改进后, 从初始时刻到 t 时刻的总成本, $\Phi(t) = \int_0^t \phi(u) du$ 。

接下来我们将证明, 如果单纯把几种成品的库存合并为在制品库存, 而不改变初始的库存总量, 则改进后的总成本是低于改进前的总成本的。首先分析改进前后各变量之间的关系, 显然有

$$s = \sum_{i=1}^N s_i, \quad D(t) = \sum_{i=1}^N D_i(t), \quad M(t) = \sum_{i=1}^N M_i(t) \quad (7-1)$$

根据以上关系可以推导出 $\min\{s + M(t), D(t)\} \geq \sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\}$, 过程如下:

$$\begin{aligned} \min\{s + M(t), D(t)\} &= \min\left\{\sum_{i=1}^N s_i + \sum_{i=1}^N M_i(t), \sum_{i=1}^N D_i(t)\right\} \\ &= \min\left\{\sum_{i=1}^N [s_i + M_i(t)], \sum_{i=1}^N D_i(t)\right\} \\ &\geq \min\left\{\sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\}, \sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\}\right\} \\ &= \sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\} \end{aligned} \quad (7-2)$$

将不等式7-2代入 $I(t)$ 的定义中, 就能比较改进前后同一时刻的库存大小:

$$\begin{aligned} I(t) &= s + M(t) - \min\{s + M(t), D(t)\} \\ &\leq \sum_{i=1}^N s_i + \sum_{i=1}^N M_i(t) - \sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\} \\ &= \sum_{i=1}^N [s_i + M_i(t) - \min\{s_i + M_i(t), D_i(t)\}] \\ &= \sum_{i=1}^N I_i(t) \end{aligned} \quad (7-3)$$

同理可比较改进前后同一时刻延迟订单数的大小:

$$B(t) = D(t) - \min\{s + M(t), D(t)\}$$

$$\begin{aligned}
&\leq \sum_{i=1}^N D_i(t) - \sum_{i=1}^N \min\{s_i + M_i(t), D_i(t)\} \\
&= \sum_{i=1}^N [D_i(t) - \min\{s_i + M_i(t), D_i(t)\}] \\
&= \sum_{i=1}^N B_i(t)
\end{aligned} \tag{7-4}$$

再将不等式7-3和7-4代入 $\phi(t)$ 的定义中，得到

$$\phi(t) = hI(t) + bB(t) \leq h \sum_{i=1}^N I_i(t) + b \sum_{i=1}^N B_i(t) = \phi_N(t) \tag{7-5}$$

公式7-5说明，改进后每时每刻的成本率都低于改进前同一时刻的成本率，自然地，改进后的总成本就低于改进前的总成本，即

$$\Phi(t) = \int_0^t \phi(u) du \leq \int_0^t \phi_N(u) du = \Phi_N(t) \tag{7-6}$$

至此，已经证明了：如果单纯把几种成品的库存合并为在制品库存，而不改变初始的库存总量，则改进后的总成本低于改进前的总成本。

在这个系统中，需要企业决定的参数是初始库存。当需求分布和生产速率确定时，企业会选择一个最优的初始库存，使得成本最小。此处的成本是指长期运行的成本，所以这个最优的初始库存是由需求和生产在极限状态下的稳定分布决定的。Wolff (1989)^[38] 证明：设 $Q_i(t) = D_i(t) - M_i(t)$ ，且所有的到达都是独立同分布的更新过程，到达速率严格小于生产速率，则 $Q_i(t)$ 是存在极限分布的。设此极限分布为 Q_i 。那么极限状态下的总成本率为

$$\psi_N(s_1, s_2, \dots, s_N) = \sum_{i=1}^N \left\{ b[E(Q_i) - E(\min\{s_i, Q_i\})] + h[s_i - E(\min\{s_i, Q_i\})] \right\} \tag{7-7}$$

公式7-7是初始库存 (s_1, s_2, \dots, s_N) 的函数。对于每一个 $b[E(Q_i) - E(\min\{s_i, Q_i\})] + h[s_i - E(\min\{s_i, Q_i\})]$ ，存在一个 s_i 使它取得最小值，我们将这个最优的 s_i 记为 s_i^* 。显然，使得总成本率 $\psi_N(s_1, s_2, \dots, s_N)$ 最小的一组 (s_1, s_2, \dots, s_N) 是 $(s_1^*, s_2^*, \dots, s_N^*)$ 。

同理，对于改进后的系统，设 $Q(t) = D(t) - M(t)$ ， $Q(t)$ 的极限分布为 Q 。则改进后的极限状态下的成本率为

$$\psi(s) = b[E(Q) - E(\min\{s, Q\})] + h[s - E(\min\{s, Q\})] \tag{7-8}$$

同样存在一个 s 使 $\psi(s)$ 取得最小值，将这个最优的 s 记为 s^* 。

显然，一般情况下， s^* 和 $\sum_{i=1}^N s_i^*$ 是不相等的。也就是说，如果企业在改进前后都按照最优的库存策略来制定初始库存，那么改进前后的总库存不一定相等。这种情况下，改进后的总成本是否仍然低于改进前的总成本呢？利用之前的结论，我们仍然能够得到肯定的回答。

将我们要证明的结论转化为数学公式，目标就是证明 $\psi(s^*) \leq \psi_N(s_1^*, s_2^*, \dots, s_N^*)$ 。首先利用公式7-6可以得到

$$\psi\left(\sum_{i=1}^N s_i^*\right) \leq \psi_N(s_1^*, s_2^*, \dots, s_N^*) \quad (7-9)$$

然后利用 s^* 的定义，由于 s^* 是使 $\psi(s)$ 取得最小值的初始库存，所以 $\psi(s^*)$ 小于等于任何 $\psi(s)$ 值，其中包括 $\psi(\sum_{i=1}^N s_i^*)$ 。

$$\psi(s^*) \leq \psi\left(\sum_{i=1}^N s_i^*\right) \quad (7-10)$$

由不等式7-9和7-10可得

$$\psi(s^*) \leq \psi_N(s_1^*, s_2^*, \dots, s_N^*)$$

因此，即使在分别采取最优初始库存的情况下，改进后的成本率仍然比改进前要低。

7.2 考虑延迟订单成本的泊松过程

上一节的定性结论适用于需求为任意更新过程的情况。如果要对改进前后的库存变化进行更量化的分析，需要知道需求的具体分布。在生产线的研究中，我们常常使用泊松过程作为需求和生产的分布。因此，现在我们假设需求和生产都是泊松过程。定义：

λ_i ： i 产品的到达速率；

λ ： 所有产品到达的总速率， $\lambda = \sum_{i=1}^N \lambda_i$ ；

p_i ： i 产品的需求占比， $p_i = \frac{\lambda_i}{\lambda}$ ；

μ ： 生产速率；

ρ ： 需求产能比， $\rho = \frac{\lambda}{\mu}$ ，且 $\rho < 1$ ；

D 、 M 、 Q 等变量定义与上一节相同。

根据 Buzacott 和 Shanthikumar (1993)^[39], $Q_i(t)$ 的极限分布是几何分布, 记为 Q_i , 其概率分布为

$$P(Q_i = n) = (1 - r_i)r_i^n$$

其中 r_i 的定义如下

$$r_i = \frac{p_i \rho}{1 - \rho + p_i \rho}$$

由此可以计算出 $I_i(t)$ 和 $B_i(t)$ 各自的极限分布期望。分别记为 I_i 和 B_i , 则它们的期望为

$$E(I_i) = s_i - \frac{r_i(1 - r_i^{s_i})}{1 - r_i}$$

$$E(B_i) = \frac{r_i^{s_i+1}}{1 - r_i}$$

然后得到极限成本率

$$\psi_N(s_1, s_2, \dots, s_N) = E \left(\sum_{i=1}^N [hI_i + bB_i] \right) \quad (7-11)$$

公式7-11中, 每一个 $hI_i + bB_i$ 实际是是关于自变量 s_i 的函数, 记为 $C(s_i)$, 则可以采用 $C(s_i + 1) - C(s_i)$ 的方式逐渐增大 s_i , 找到最优的 s_i^* 。根据 Veatch 和 Wein (1996)^[40], 这个最优值为

$$s_i^* = \left\lfloor \frac{\ln \gamma}{\ln r_i} \right\rfloor \quad (7-12)$$

其中 $\gamma = \frac{h}{h+b}$ 。

公式7-12中为了符合实际需求而加入了取整符号。为了方便计算和推导, 我们不对 s_i^* 向下取整, 而是直接代入 $\psi_N(s_1, s_2, \dots, s_N)$, 得到的是

$$\psi_N(s_1^*, s_2^*, \dots, s_N^*) = h \sum_{i=1}^N \frac{\ln \gamma}{\ln \frac{p_i \rho}{1 - \rho + p_i \rho}} \quad (7-13)$$

同样的方法可以推出改进后的极限成本率为

$$\psi(s^*) = \frac{h \ln \gamma}{\ln \rho} \quad (7-14)$$

由公式7-13和7-14可以比较改进前后的库存差, 将这个差值记为 $\Delta\psi$, 则 $\Delta\psi = \psi_N(s_1^*, s_2^*, \dots, s_N^*) - \psi(s^*)$ 。

现在我们有了解算改进前后差值的公式，就可以进行一些数值实验。取 $h = 1$, $b = 5$, $p_i = \frac{1}{N}$, 计算 $\Delta\psi$ 随产能比 ρ 的变化。数值实验的结果如图7.1所示。

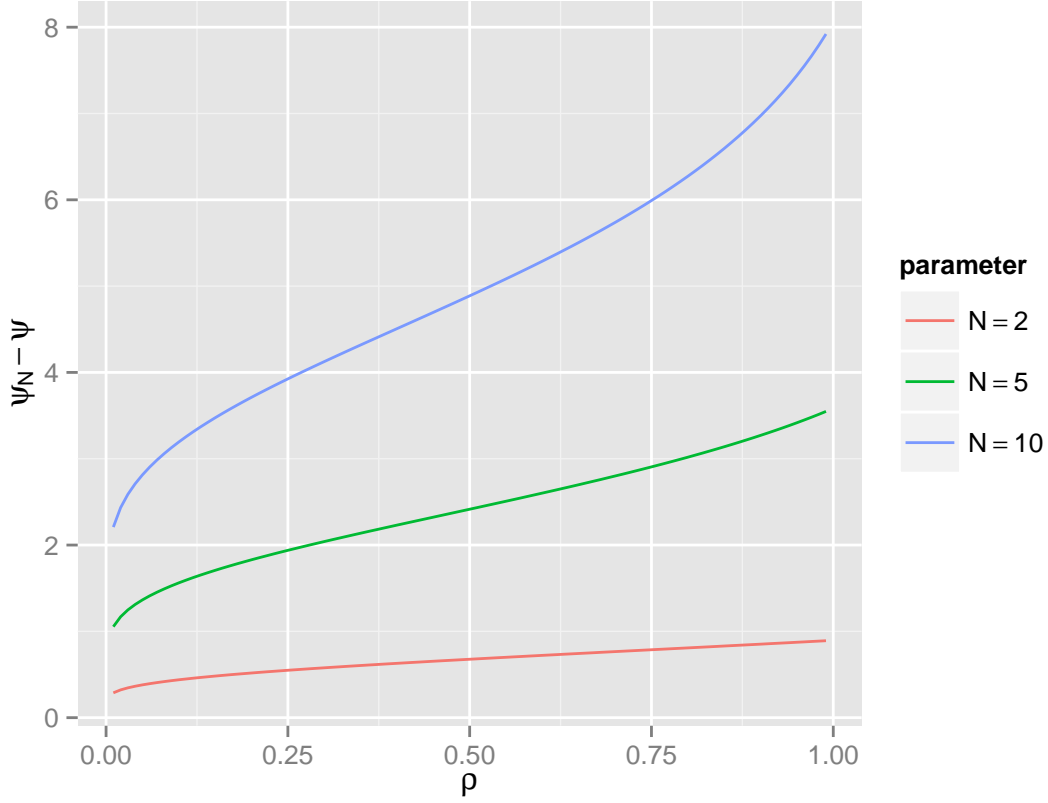


图 7.1 改进前后极限成本率之差随 ρ 的变化

毫无疑问，改进后的成本是小于改进前的成本的，这一点在前一节中已经得到定性的证明。除此之外，图7.1的三条曲线都显示， ρ 越接近 1，改进前后的库存之差越大。 ρ 代表的是需求产能比， $\rho = \frac{\lambda}{\mu}$ 。也就是说，需求量越接近产能，改进的效果越显著。这是与日常经验相符的，因为产能严重过剩的生产线一般不需要通过优化来挖掘潜力。

我们可以利用上述特性来估计改进效果的上限。求出 $\rho \rightarrow 1$ 时， $\Delta\psi$ 的极限，就可以作为改进效果上限的一个估计。设所有需求的到达速率相同，都是 $p_i = \frac{1}{N}$ 。则

$$\lim_{\rho \rightarrow 1} \Delta\psi = h \ln \gamma \lim_{\rho \rightarrow 1} \sum_{i=1}^N \left[\frac{1}{\ln \frac{p_i \rho}{1 - \rho + p_i \rho}} - \frac{p_i}{\ln \rho} \right] \quad (7-15)$$

对公式7-15使用洛必达法则，可以得到

$$\lim_{\rho \rightarrow 1} \Delta \psi = \frac{1}{2} h(N-1) \ln \frac{h+b}{h} \quad (7-16)$$

只要将具体的 h 、 b 、 N 的值代入公式7-16，就可以求出 $\Delta \psi$ 在 $\rho \rightarrow 1$ 时的极限值。该极限值可以作为对改进效果的上限的一种估计。企业可以通过这个简单的公式先估算可能获得的收益，再决定是否把其他因素考虑进来作更深的研究。

第 8 章 总结与展望

8.1 总结

本文主要从四个不同的方面探究了“合并同型号成品库存到在制品库存”这一改进方案可能产生的效果。本文在探究过程中采用的主要方式是数学建模,这使得本文的结论具有一定的局限性——基于很多理想化的假设条件;但同时也赋予了这些结论更强的普适性——没有局限于某个具体的企业生产线,而是给出普遍意义的结论。

在研究改进方案是否能够起到改进效果时,本文从数学性质良好的正态分布入手,证明了改进方案的优化潜力。还通过数值实验,研究了系统中一些参数对改进效果的影响。结果表明,需求波动越大、对服务水平要求越高的企业,在改进中获得的收益越大。

在研究改进失效现象时,本文既分析了数学性质良好的对称稳定分布,又分析了实际经常使用的右偏斜分布。同时也关注了服务水平的影响,探讨了两种不同的服务水平定义带来的不同结果。最后通过数值实验,找出参数影响改进失效现象的规律,提出了避免改进失效的方法,并且验证了之前证明的一些结论。

在研究需求相关性对改进效果的影响时,本文关注了很多文献都不曾注意的细节,把证明的基础建立在严密的构造上。讨论了极端状态下的改进效果,找到了需求相关性对改进效果的影响规律。得到结论后,又通过一些实际的数据为例,展示了企业应该如何利用这些结论来做决定。

研究延迟订单对改进效果的影响时,本文采用了运筹学的一些方法和结论,得到了定性和定量的结果。同时,还给出了估算改进效果上限的方法,为企业的决策提供一个参考。

8.2 展望

本文对需求分布、服务水平、生产时间等参数都有一定的研究,但尚未进展到对提前期变化的研究。企业将成品的库存合并前移到在制品,那么从在制品到成品所需的生产时间势必对改进效果产生相当重要的影响。再加上提前期

的波动，整个系统在时间方面可能受到很大的影响。这一部分的研究也是比较有难度的。对时间问题的讨论可能需要继续使用马尔科夫过程，必要的时候还可能用到杰克逊网络。

除此之外，本文也没有涉及库存策略的影响。实际生产中对库存的控制往往是通过各种既定的策略来完成的，比如 (s,S) 策略，推动式生产或拉动式生产等等。当库存遵循这些策略的时候，改进方案还能起到效果吗？为了适应改进方案，库存策略中的一些参数应该取何值？

以上这些问题都是具有较强实际意义的，同时，它们也具有比较高的难度。若时间和精力允许，我将在未来的研究中尝试解决它们。

插图索引

图 2.1	单产品多阶段生产	5
图 2.2	拉动式生产系统流程图	6
图 2.3	推动式生产系统流程图	7
图 4.1	需求波动和服务水平对改进效果的影响	20
图 5.1	不同参数 α 下的对称稳定分布	24
图 5.2	泊松分布下的改进失效区间图示	31
图 6.1	骊威、骐达月销售量	41
图 6.2	骊威、骐达销量散点图	42
图 6.3	英朗、英朗 XT 月销售量	43
图 6.4	英朗、英朗 XT 销量散点图	43
图 7.1	改进前后极限成本率之差随 ρ 的变化	49
图 A.1	看板设计	66
图 A.2	过程内看板的循环	67
图 A.3	过程间看板的循环	67
图 A.4	拉式系统流程图	69
图 A.5	仿真模型概念图	74
图 A.6	单位订货量 M^n 对生产波动的影响	75
图 A.7	单位订货量 M^n 对库存波动的影响	76
图 A.8	生产波动放大率	76
图 A.9	库存波动放大率	77

表格索引

表 5.1	改进前服务水平与库存的关系	27
表 5.2	两种成品需求的联合分布	27
表 5.3	在制品需求分布	27
表 5.4	改进后服务水平与库存的关系	28
表 5.5	不同服务水平下改进前后库存对比	28
表 B.1	汽车销售数据	79

公式索引

公式 2-1	3
公式 2-2	3
公式 2-3	3
公式 4-1	14
公式 4-2	14
公式 4-3	14
公式 4-4	14
公式 4-5	15
公式 4-6	15
公式 4-7	15
公式 4-8	15
公式 4-9	15
公式 4-10	16
公式 4-11	16
公式 4-12	16
公式 4-13	17
公式 4-14	17
公式 4-15	17
公式 4-16	17
公式 4-17	18
公式 4-18	18
公式 4-19	18
公式 4-20	18
公式 4-21	19
公式 4-22	19
公式 4-23	19
公式 5-1	25
公式 5-2	25

公式 5-3	25
公式 5-4	25
公式 5-5	25
公式 5-6	26
公式 5-7	26
公式 5-8	29
公式 5-9	29
公式 5-10.....	29
公式 5-11.....	30
公式 5-12.....	30
公式 5-13.....	30
公式 5-14.....	30
公式 5-15.....	30
公式 5-16.....	33
公式 5-17.....	33
公式 5-18.....	33
公式 5-19.....	33
公式 6-1	35
公式 6-2	35
公式 6-3	35
公式 6-4	35
公式 6-5	36
公式 6-6	36
公式 6-7	36
公式 6-8	36
公式 6-9	37
公式 6-10.....	37
公式 6-11.....	37
公式 6-12.....	37
公式 6-13.....	38
公式 6-14.....	38

公式 6-15	38
公式 6-16	38
公式 6-17	39
公式 6-18	39
公式 6-19	39
公式 6-20	39
公式 6-21	39
公式 6-22	39
公式 6-23	39
公式 6-24	40
公式 7-1	45
公式 7-2	45
公式 7-3	45
公式 7-4	46
公式 7-5	46
公式 7-6	46
公式 7-7	46
公式 7-8	46
公式 7-9	47
公式 7-10	47
公式 7-11	48
公式 7-12	48
公式 7-13	48
公式 7-14	48
公式 7-15	49
公式 7-16	50
公式 A-1	69
公式 A-2	70
公式 A-3	70
公式 A-4	70
公式 A-5	70

公式 A-6	70
公式 A-7	70
公式 A-8	71
公式 A-9	71
公式 A-10.....	71
公式 A-11.....	71
公式 A-12.....	72
公式 A-13.....	72
公式 A-14.....	72
公式 A-15.....	73
公式 A-16.....	73
公式 A-17.....	73
公式 A-18.....	73

参考文献

- [1] Fetter R B, Dalleck W C. Decision models for inventory management. RD Irwin, 1961
- [2] Lambrecht M R, Muckstadt J A, Luyten R. Protective stocks in multi-stage production systems. THE INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH, 1984, 22(6):1001–1025
- [3] Clark A J, Scarf H. Approximate solutions to a simple multi-echelon inventory problem. Studies in Applied Probability and Management Science, 1962. 88–110
- [4] Baker K R, Magazine M J, Nuttle H L. The effect of commonality on safety stock in a simple inventory model. Management Science, 1986, 32(8):982–988
- [5] Eppen G D, Martin R K. Determining safety stock in the presence of stochastic lead time and demand. Management Science, 1988, 34(11):1380–1390
- [6] Molinder A. Joint optimization of lot-sizes, safety stocks and safety lead times in an MRP system. International Journal of Production Research, 1997, 35(4):983–994
- [7] Chopra S, Reinhardt G, Dada M. The Effect of Lead Time Uncertainty on Safety Stocks. Decision Sciences, 2004, 35(1):1–24
- [8] Fang X, Zhang C, Robb D J, et al. Decision support for lead time and demand variability reduction. Omega, 2013, 41(2):390–396
- [9] Gershwin S B. Manufacturing systems engineering. PTR Prentice Hall (Englewood Cliffs, NJ), 1994
- [10] Little J D. A Proof for the Queuing Formula: $L = \lambda W$. Operations research, 1961, 9(3):383–387
- [11] Goyal S K. Economic batch quantity in a multi-stage production system. THE INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH, 1978, 16(4):267–273
- [12] Kimura O, Terada H. Design and analysis of Pull System, a method of multi-stage production control. The International Journal Of Production Research, 1981, 19(3):241–253
- [13] Tabe T, Muramatsu R, Tanaka Y. Analysis of production ordering quantities and inventory variations in a multi-stage production ordering system. International Journal of Production Research, 1980, 18(2):245
- [14] Conway R, Maxwell W, McClain J O, et al. The role of work-in-process inventory in serial production lines. Operations Research, 1988, 36(2):229–241
- [15] Daskin M S, Coullard C R, Shen Z J M. An inventory-location model: Formulation, solution algorithm and computational results. Annals of Operations Research, 2002, 110(1-4):83–106
- [16] Snyder L V, Daskin M S, Teo C P. The stochastic location model with risk pooling. European Journal of Operational Research, 2007, 179(3):1221–1238

- [17] Ozsen L, Coullard C R, Daskin M S. Capacitated warehouse location model with risk pooling. *Naval Research Logistics (NRL)*, 2008, 55(4):295–312
- [18] Shen Z J M, Coullard C, Daskin M S. A joint location-inventory model. *Transportation Science*, 2003, 37(1):40–55
- [19] Eppen G, Schrage L. Centralized ordering policies in a multi-warehouse system with lead times and random demand. *Multi-level production/inventory control systems: Theory and practice*, 1981, 16:51–67
- [20] Jönsson H, Silver E A. Analysis of a Two-Echelon Inventory Control System with Complete Redistribution. *Management Science*, 1987, 33(2):215–227
- [21] Jackson P L. Stock Allocation in a Two-Echelon Distribution System or "what to Do Until Your Ship Comes In". *Management Science*, 1988, 34(7):880–895
- [22] Schwarz L B. A Model for Assessing the Value of Warehouse Risk-Pooling: Risk-Pooling Over Outside-Supplier Leadtimes. *Management Science*, 1989, 35(7):828–842
- [23] Clark A J, Scarf H. Optimal Policies for a Multi-Echelon Inventory Problem. *Management Science*, 2004, 50(12_supplement):1782–1790
- [24] Benjaafar S, Cooper W L, Kim J S. On the Benefits of Pooling in Production-Inventory Systems. *Management Science*, 2005, 51(4):548–565
- [25] Yang H, Schrage L. Conditions that cause risk pooling to increase inventory. *European Journal of Operational Research*, 2009, 192(3):837–851
- [26] Zinn W, Bowersox D J. Planning physical distribution with the principle of postponement. *PLANNING*, 1988, 9(2):111
- [27] Lee H L, Tang C S. Modelling the Costs and Benefits of Delayed Product Differentiation. *Management Science*, 1997, 43(1):40–53
- [28] Su J C P, Chang Y L, Ferguson M. Evaluation of postponement structures to accommodate mass customization. *Journal of Operations Management*, 2005, 23(3–4):305–318
- [29] Cheng T E, Li J, Wan C J, et al. Simulation of a Two-End-Product Postponement System. *Proceedings of Postponement Strategies in Supply Chain Management*. Springer, 2010: 109–123
- [30] Zeng D, An W, Luo X. Research on Hybrid Postponement Strategies of Supply Chain. *Proceedings of Logistics*. American Society of Civil Engineers, : 3262–3269
- [31] Krishnamoorthy A, Islam M E. (s, S) inventory system with postponed demands. *Stochastic analysis and applications*, 2004, 22(3):827–842
- [32] Jayaraman R, Sivakumar B, Arivarignan G. A perishable inventory system with postponed demands and multiple server vacations. *Modelling and Simulation in Engineering*, 2012, 2012:8
- [33] Feller W. An introduction to probability theory and its applications, volume 2. John Wiley & Sons, 2008

- [34] Newman M. Power laws, Pareto distributions and Zip's law. *Contemporary Physics*, 2005, 46(5):323–351
- [35] Agrawal N, Smith S A. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics (NRL)*, 1996, 43(6):839–861
- [36] 新浪汽车：东风日产销售数据. http://data.auto.sina.com.cn/xlsjk/brand.php?start_date=2012-01&end_date=2014-4&bid=94
- [37] 新浪汽车：上海通用别克销售数据. http://data.auto.sina.com.cn/xlsjk/brand.php?start_date=2012-01&end_date=2014-4&bid=58
- [38] Wolff R W. *Stochastic modeling and the theory of queues*, volume 14. Prentice hall Englewood Cliffs, NJ, 1989
- [39] Buzacott J A, Shanthikumar J G. *Stochastic models of manufacturing systems*, volume 4. Prentice Hall Englewood Cliffs, NJ, 1993
- [40] Veatch M H, Wein L M. Scheduling a Make-To-Stock Queue: Index Policies and Hedging Points. *Operations Research*, 1996, 44(4):634–647

致 谢

衷心感谢导师吴甦老师对我的帮助和指导。吴老师在选题和整体研究思路方面给了我很大的启发，为我的研究搭起了基本的框架。在完成这篇论文的过程中，吴老师始终尊重和鼓励我，为我答疑解惑而又不干涉我的想法。

感谢开题和中期答辩的评委皋琴老师、朱万山老师、曹晖老师，你们提出了很多宝贵的意见和建议，不仅修正了论文中可能出现的一些错误，也启发了很多新的思路和方法。

感谢吴振卿学长和陈昌国学长，你们在论文写作过程中给予了我很多实质性的帮助和建议。

感谢这篇论文所涉及到的各位学者，你们的研究使我受益良多，既是坚固的基石，又是灵感的源泉。

感谢父母、朋友和同学对我的关心，你们使我的生活更加精彩，使我能够以更饱满的热情投入到论文的写作中。

感谢 L^AT_EX 模板 ThuThesis，为我省去了调整论文格式的繁琐过程，使我能够专注于论文内容的写作。

声 明

本人郑重声明：所提交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

附录 A 外文文献翻译

标题：一种多阶段生产的控制方法——拉式生产的设计和分析

原标题：Design and analysis of Pull System, a method of multi-stage production control

作者：OSAMU KIMURA, HIROSUKE TERADA

摘要：我们把多阶段生产过程的控制系统分为两种，分别称它们为推式系统和拉式系统。前者是一种传统的方法——每个阶段库存的零部件是根据它到最终阶段的总流程时间来预测的。生产和库存控制是基于这些预测值的。后者是本文提出的一种方法——每个阶段保持一定总量的库存，其补给是根据后续生产过程的消耗速率而定的。我们先制定拉式系统的定义，然后从批量大小、提前期等系统参数出发，给出整个生产过程中的生产、库存波动的仿真模型。

A.1 简介

A.1.1 传统生产订单系统及它面临的问题

一般来说，由于制造需要的时间比允许的供货延迟时间长，多阶段制造过程需要预先生产一些产品。我们能把这些过程的生产控制系统归为以下两类。

1. 推式系统：这类系统考虑每个阶段到最终阶段的流程时间，预测每阶段库存零部件或在制材料的需求。基于这些预测，它们通过调整最终成品和零部件的库存来控制整个多阶段生产。我们把这一类称为推动型生产订单系统，或者简称推式系统。
2. 拉式系统：每个阶段保有一定量的库存。下游工序只根据自己消耗零部件的速率和时间，向上游工序订货并且从其储存区提取零部件。我们把这一类称为拉动型生产订单系统，或者简称拉式系统。

大多数传统生产控制系统属于前一种。系统越大，以下这些固有的问题就越突出。

- (a) 当需求发生剧变或者生产出现障碍时，几乎不可能恢复每个阶段的生产计划。因此，这样的问题很可能导致过量库存甚至呆滞库存。

- (b) 实际生产控制中，员工不可能仔细检查所有与生产速率和库存水平相关的情况。因此，生产计划必须包含过量的安全库存。
 - (c) 无法通过生产批量和定时生产来进行优化，因为每一个细节都最优的生产计划太复杂了，没法计算。
- 拉式系统作为解决这些问题的方法被设计出来。我们可以按照下游工序的消耗速率，通过简单可靠的补货规则就能实现优化。

A.1.2 拉式系统的目标

在包含外部供应商的多阶段生产过程中：

1. 阻止需求或产量波动从下游工序扩大到上游工序。
2. 使在制品库存波动最小化，以简化库存控制。
3. 通过去中心化提升车间控制水平：使车间主管或工头不仅作为生产控制者，同时也作为库存控制者。

A.1.3 本研究的目标

本研究的目标是证明拉式系统确实满足以上列出的这些目标。我们尤其从订货单位、送货提前期等系统参数特点出发，考虑下游工序的需求或产量波动对上游工序产量和库存的影响。

A.2 拉式系统概要

A.2.1 拉式系统方法论

正如前面所定义的，拉式系统是这样的系统：

1. 保持每阶段的库存在特定水平上。
2. 下游工序为了补充原材料，从上游工序订取已消耗的材料。

为了实践这样的系统，应该遵循以下程序：

- (a) 建立标准的再订货点和批量大小。
- (b) 随时了解库存水平和延期订单。
- (c) 连续检查库存水平，一旦某些零部件数量少于再订货点，就重新订货。

拉式系统必须满足上面提到的所有要求。然而，在复杂的多阶段生产过程中设计一个满足所有条件的系统是相当困难的，因为要满足条件（2）所需的成本和绩效是冲突的。

在丰田公司，我们用看板——一种标记——来解决这个问题。

A.2.2 看板

看板携带了以下信息。

1. 零部件名称、编号。
2. 数量通常是根据容器容量来标定的。再订货点和订货数量是容器容量的整数倍。
3. 上游工序：制造车间、装配线或者储存点。
4. 下游工序：同上。

其他信息如打包方式、看板数量等也被标示出来作为参考。典型的看板如图A.1所示。

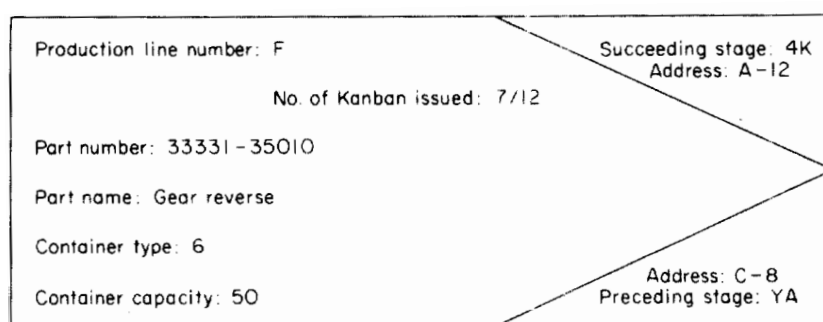


图 A.1 看板设计

A.2.3 操作过程

A.2.3.1 过程内看板

某阶段的零部件被放在一个容器里。一个看板被贴在或挂在容器上，然后把容器放在看板指定的储存地点（见图A.2 ①）。

当下游工序提取零部件或原材料时，工人把看板取下来，放在看板盒里（见图A.2 ②）。盒子里的看板被定期取出来挂在计划板的钩子上。计划板上各种看板的顺序向工人们展示了生产过程中各项任务的订单派送情况（见图A.2 ③）。

工人按照预先设计的速率，根据计划板上各种看板的顺序生产各种物品。看板自身也会随着某批产品投入生产而移动到生产过程中（见图A.2 ④）。

步骤①到④反复运行，生产过程也就持续有效地进行。

我们应该牢记，在步骤②中，如果下游工序始终不提货，那么看板就不会进入看板盒或者挂上计划板。因此，这个车间就始终不会生产这种物品。

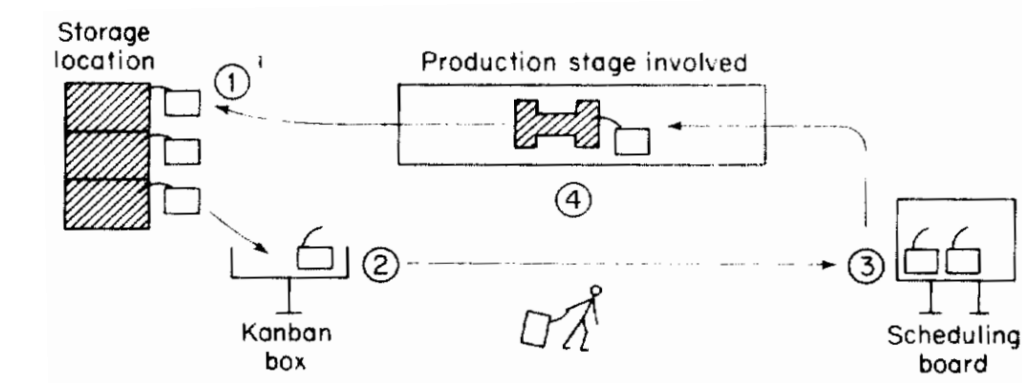


图 A.2 过程内看板的循环

A.2.3.2 过程间看板

过程间看板操作跟过程内看板是一样的，只要把运输视为类似于制造的操作就行了（图A.3）。

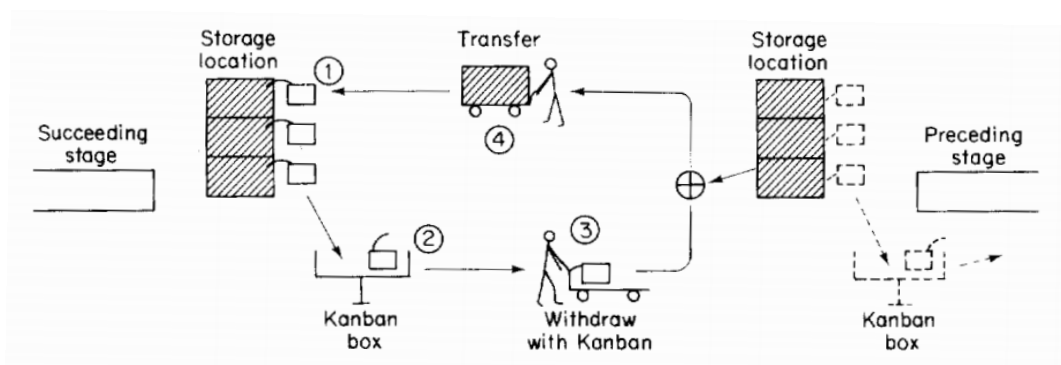


图 A.3 过程间看板的循环

我们同样应该牢记，提货数量等于看板表示的数量，盒子里没有看板就不会提取任何东西。

图 3 中的虚线表示过程内看板，它在上游车间中流动。当原材料或零部件从储存点被提走，容器上的过程内看板被替换为过程间看板。换下来的过程内看板会被放到过程内看板盒中。

正如我们在上面阐述的那样，下游生产过程的生产速率（数量）通过过程内看板和过程间看板传递给了上游工序。这样一个多阶段生产过程中的所有看板维持着所有阶段完成生产。

A.3 系统模型

我们按照如下说明，对拉式系统建模，以阐明系统特点。这里把系统考虑为只有一种产品的简单线性多阶段生产系统。

A.3.1 标记方式

t 时段。

O_t^n 第 n 道工序在 t 时段的订货量。

P_t^n 第 n 道工序在 t 时段的产量。

I_t^n 第 n 道工序在 t 时段末的库存。

M^n 订货单位（一个看板订多少）。

Z_t^n I 除以 M 的余数。

C^n 第 n 道工序的生产能力。

L_1^n 看板从容器中拿出来，的时间到开始生产的时间差。

L_2^n 生产开始和结束的时间差。

D_t t 时段的需求。

$D_{t,t+L}$ 在 t 时段末对 $(t+L)$ 时段的需求预测。

A.3.2 信息和原材料流动示意图

拉式系统中信息和原材料的流动可以用图A.4来表示。

A.3.3 系统的基本公式

在拉式系统中，当容器里的物品开始被使用时，看板就从容器上取下来。因此，在每一个库存站都有这样一个容器，它里面的物品已经被使用了一部分，

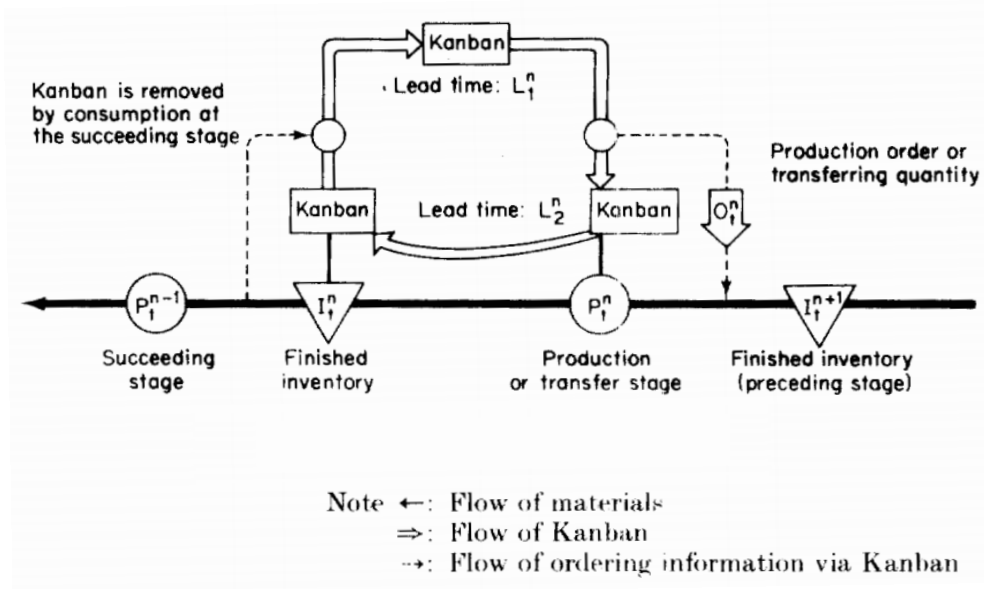


图 A.4 拉式系统流程图

看板已经被取下来了。其他的容器里物品都是满的，看板也贴在上面。令 Z_t^n 为已经部分使用的容器中剩余物品数，记作

$$Z_t^n = \text{Mod}(I_t^n, M^n)$$

其中符号 $\text{Mod}(A, B)$ 表示 A/B 的余数。第 n 个库存点已经移走的看板数由下游阶段的产量 P_t^{n-1} 决定。但是 P_t^{n-1} 的一部分可以被看板已移走的容器里的原材料抵消掉，这部分最多对应 Z_{t-1}^n 。

因此，可以用

$$\left. \begin{aligned} X &= P_t^{n-1} - Z_{t-1}^n & (P_t^{n-1} \geq Z_{t-1}^n) \\ \text{or} \\ X &= 0 & (P_t^{n-1} < Z_{t-1}^n) \end{aligned} \right\} \quad (\text{A-1})$$

表示这个时段内移走的看板数所对应的原材料数。然后可以把第 t 时段移走的看板数表示为

$$\left[\frac{X}{M^n} \right]_+ = \left[\frac{\max(0, P_t^{n-1} - Z_{t-1}^n)}{M^n} \right]_+$$

符号 $[A]_+$ 表示把 A 向上取整的高斯函数。经过 L_1^n 时间之后，被移走的看板会

被添加到第 n 生产阶段的订单里。所以生产订单 O_t^n 为

$$O_t^n = \left[\frac{\max(0, P_{t-L_1^n}^{n-1} - Z_{t-L_1^n-1}^n)}{M^n} \right]_+ M^n + O_{t-1}^n - P_{t-1}^n \quad (\text{A-2})$$

其中 $O_{t-1}^n - P_{t-1}^n$ 是前一时段末留下的未满足订单。生产量为

$$P_t^n = \min(O_t^n, C^n, I_{t-1}^{n+1} + P_{t-L_2^n}^{n+1}) \quad (\text{A-3})$$

其中 $I_{t-1}^{n+1} + P_{t-L_2^n}^{n+1}$ 表示上游阶段的库存限制。

第 n 阶段的库存水平可以表示为

$$I_t^n = I_{t-1}^n + P_{t-L_2^n}^n - P_t^{n-1} \textcircled{1} \quad (\text{A-4})$$

以上公式 (A-2)-(A-4) 是拉式系统的基本公式。

A.4 生产和库存的波动分析

A.4.1 单位订货量相对较小的情况

A.4.1.1 生产波动

当单位订货量 M^n 与产量 P_t^n 相比相对较小时，我们可以令

$$\left[\frac{\max(0, P_{t-L_1^n}^{n-1} - Z_{t-L_1^n-1}^n)}{M^n} \right]_+ M^n \doteq P_{t-L_1^n}^{n-1}$$

然后根据 (A-2) 有

$$O_t^n \doteq P_{t-L_1^n}^{n-1} + O_{t-1}^n - P_{t-1}^n \quad (\text{A-5})$$

当生产能力 C^n 和上游阶段库存 I_t^{n+1} 没有约束时，(A-3) 可以写作

$$P_t^n = O_t^n \quad (\text{A-6})$$

由 (A-5) 和 (A-6) 可得

$$\begin{aligned} P_t^n &= P_{t-L_1^n}^{n-1} \\ &= P_{t-(L_1^n+L_1^{n-1})}^{n-2} \end{aligned} \quad (\text{A-7})$$

① 译者注：此处原文为 $I_t^n = I_t^{n-1} + P_{t-L_2^n}^n - P_t^{n-1}$ ，疑为原作者笔误。

$$\begin{aligned} & \vdots \\ & = P_{t-(L_1^n+L_1^{n-1}+\dots+L_1^2)}^1 \end{aligned} \quad (\text{A-8})$$

正如拉式系统中得出的这个结果，当单位订货量与产量相比相对较小时，下游阶段的生产波动始终按照原始情况传递给上游阶段。传递过程中的时滞等于流程中各阶段从容器上移走看板和开始生产之间的时差之和。

特别地，如果最终阶段 $\{P^1\}$ 的生产是独立的，则 $\{P_1^n\}$ 的方差为

$$V(P^n) = V(P^{n-1}) = \dots = V(P^1) \quad (\text{A-9})$$

定义放大率：

$$\text{Amp}(P^n) = \frac{V(P^n)}{V(P_1)}$$

根据 (A-9)

$$\text{Amp}(P^n) = \text{Amp}(P^{n+1}) = \dots = 1 \quad (\text{A-10})$$

A.4.1.2 库存波动

根据 (A-7)

$$P_{t-L_2^n}^n = P_{t-(L_1^n+L_2^n)}^{n-1}$$

因此根据 (A-4)

$$I_t^n = I_{t-1}^n + P_{t-(L_1^n+L_2^n)}^{n-1} - P_t^{n-1} \textcircled{1}$$

解这个方程可得

$$I_t^n = A - \sum_{R=t-(L_1^n+L_2^n)+1}^t P_R^{n-1} \quad (\text{A-11})$$

其中

$$A = I_0^n + \sum_{R \geq 0} P_R^{n-1}$$

表示初始状态。

① 译者注：此处原文为 $I_t^n = I_{t-1}^n + P_{t-(L_1^n+L_2^n)}^{n-1} - P_t^{n-1}$ ，应为原作者笔误。

当 $\{P_t^1\}$ 相互独立时, 由 (A-1) 可得

$$V(I^n) = (L_1^n + L_2^n)V(P^{n-1}) \quad (\text{A-12})$$

将 (A-9) 代入 (A-12) 可得

$$V(I^n) = (L_1^n + L_2^n)V(P^1) \quad (\text{A-13})$$

令

$$\text{Amp}(I^n) = \frac{V(I^n)}{V(P^1)}$$

然后由 (A-13) 可得

$$\begin{aligned} \text{Amp}(I^n) &= \frac{V(I^n)}{V(P^1)} \\ \text{Amp}(I^n) &= L_1^n + L_2^n \end{aligned} \quad (\text{A-14})$$

$L_1^n + L_2^n$ 表示从容器上移走看板直到该阶段完成生产的时间差。

因此, 在拉式系统中, 当 $\{P_1\}$ 的波动独立时, 每阶段的库存波动相比于最终阶段 $\{P_1\}$ 来说是放大了。放大的程度随着看板从容器上取走的时间和该阶段完成生产的时间之差增大而增大。但是, 放大的程度不会随某阶段在流水线上的位置越靠上而越大。

到目前为止, 我们假设的是没有生产能力和库存水平的约束。如果这些约束存在, $\{P_t^n\}$ 的上限会受到 C^n 或者 $I_t^{n+1} + P_{t-L_2^{n+1}}^{n+1}$ 的限制 (见公式A-3), 因此在这些情况下波动会比无约束情况下更小。但是, 伴随着波动变小, 未满足订单数和生产延迟都会变大。无论如何, 这些管理问题不属于我们要研究的目标。

A.4.2 单位订货量相对较大的情况

在第 n 阶段单位订货量 M^n 与产量 P_{t-1}^n 相比较大的情况下, 公式 (A-2) 中的订货量^① O_t^n 不能按照公式 (A-5) 那样近似, 所以很难作出理论分析。因此, 我们尝试用以下仿真模型进行分析。

A.4.2.1 标记方式

O_t^n 在 t 时段, 第 $(n-1)$ 阶段向第 n 阶段订货的数量。

① 译者注: 此处原文为 “production quantity”, 疑应为 “order quantity”。

- R_t^n 在 t 时段, 第 n 阶段向第 $(n-1)$ 阶段运货的数量。
- B_t^n t 时段末, 从第 $(n+1)$ 阶段的库存运到第 n 阶段, 正等待处理的数量。
- $L_{P_1}^n$ 与前面定义的 L_1^n 意义相同。
- $L_{P_2}^n$ 与前面定义的 L_2^n 意义相同。
- $L_{H_1}^n$ 从容器上取走看板直到订单开始发货的时长。
- $L_{H_2}^n$ 从订单开始发货到操作完成的时长。
- S^0 最终阶段的安全库存。

除了以上定义的标记, 其余标记与之前定义的相同。

A.4.2.2 仿真模型

1. 一般模型; 如图A.5所示。
这是包含 n 个生产阶段和 n 个运输阶段的单产品多阶段生产过程。
2. 最终产品的需求量 D_t ; 服从均值为 \bar{D} , 方差为 σ_D^2 的正态分布。均值和方差都是已知的。
3. 生产和运输计时; 生产和运输在每个时段初开始。
4. 系统中的公式
最终阶段的库存量

$$B_t^0 = B_{t-1}^0 + R_{t-L_{H_2}^1}^1 - D_t \quad (\text{A-15})$$

产品的运送

$$R_t^1 = I_{t-1}^1 \quad (\text{A-16})$$

运送前的库存

$$I_t^1 = P_{t-L_{P_2}^1}^1 \quad (\text{A-17})$$

第一阶段的订货量。

(我们只对最终阶段采用推式系统。)

$$O_t^1 = \hat{D}_{t-1,t+L} + \hat{D}_{t-1,t+L-1} - P_{t-1}^1 - B_{t-1}^0 + S^0 \quad (\text{A-18})$$

其中 $L = L_{P_2}^1 + L_{H_2}^1$ 。

第二阶段的生产、运输、库存公式可以用上面提到的方法得到。

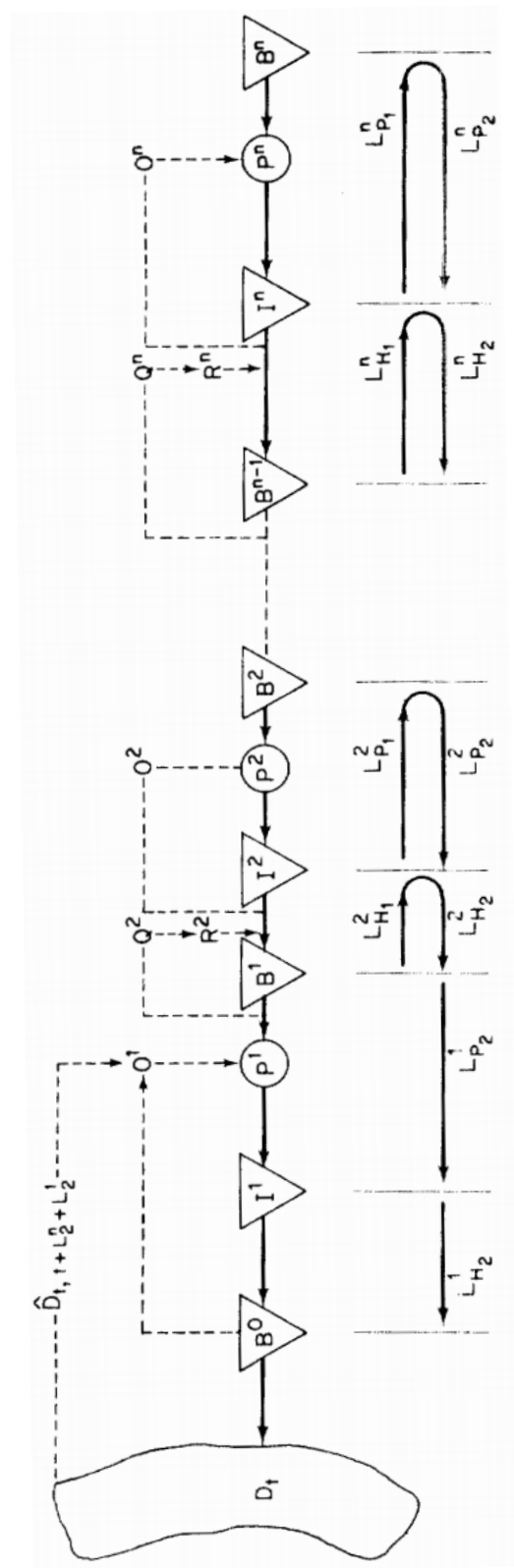


图 A.5 仿真模型概念图

A.4.2.3 输入数据和参数值

- (1) t $1, 2, 3, \dots, 100$
- (2) D_t $\bar{D} = 100 \frac{\sigma_D}{\bar{D}} = 0.1$
- (3) \hat{D} \bar{D}
- (4) n 5
- (5) I_0^n, C^n very large numbers
- (6) $L_{P_1}^n = L_{P_2}^n = L_{H_1}^n = L_{H_2}^n = 0$ ($n = 1, 2, \dots, 5$)

A.4.2.4 实验结果

在上述条件下，改变单位订货量 M^n ，结果如图A.6和图A.7所示。

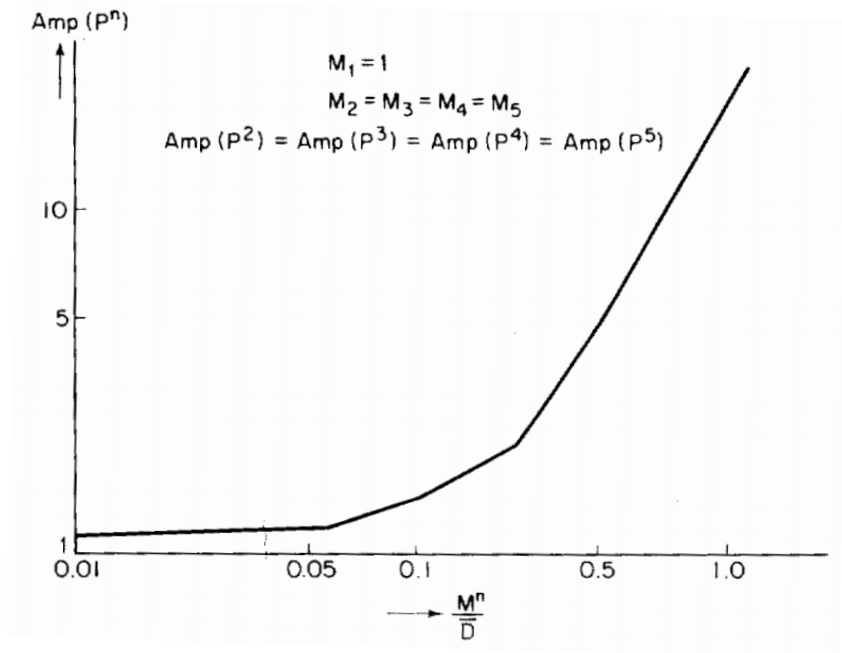


图 A.6 单位订货量 M^n 对生产波动的影响

在拉式系统中，随着单位订货量相对于产量水平增大，生产和库存的波动会增大，但是波动不会在向上游传递的过程中扩大。

A.4.3 与推式系统的比较

为了比较推式系统和拉式系统内下游阶段生产波动对上游阶段的影响，推式系统（由早稻田大学的 Y.Tanaka 和 T.Tabe 建立）和本研究中的拉式系统的实验结果用图A.8和图A.9展示。

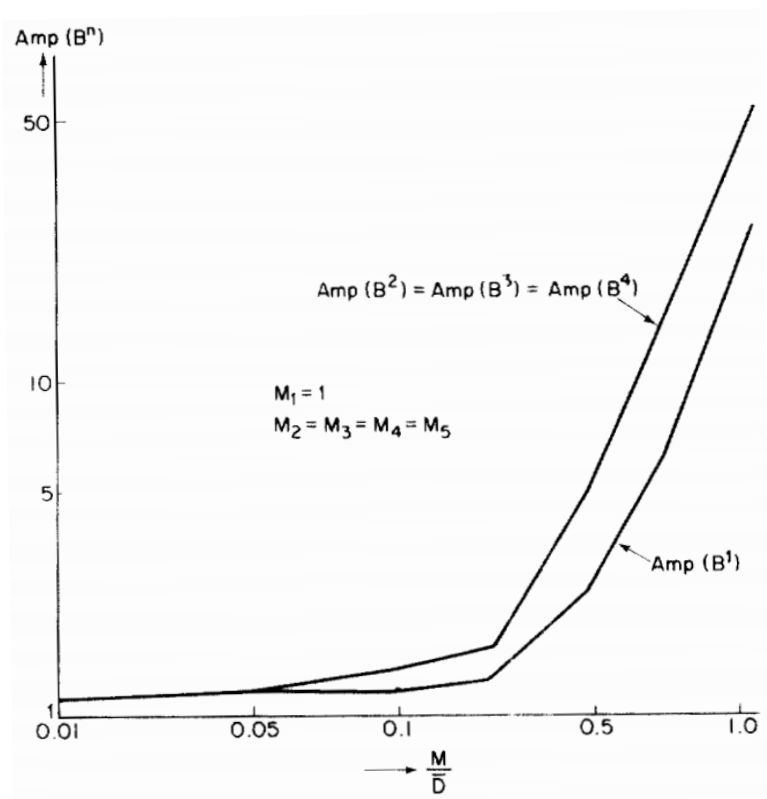


图 A.7 单位订货量 M^n 对库存波动的影响

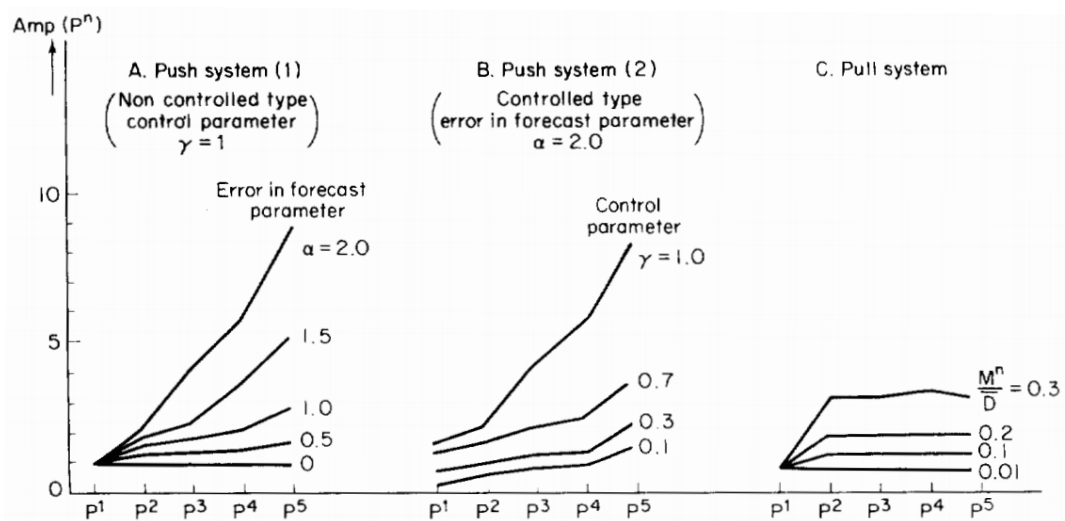


图 A.8 生产波动放大率

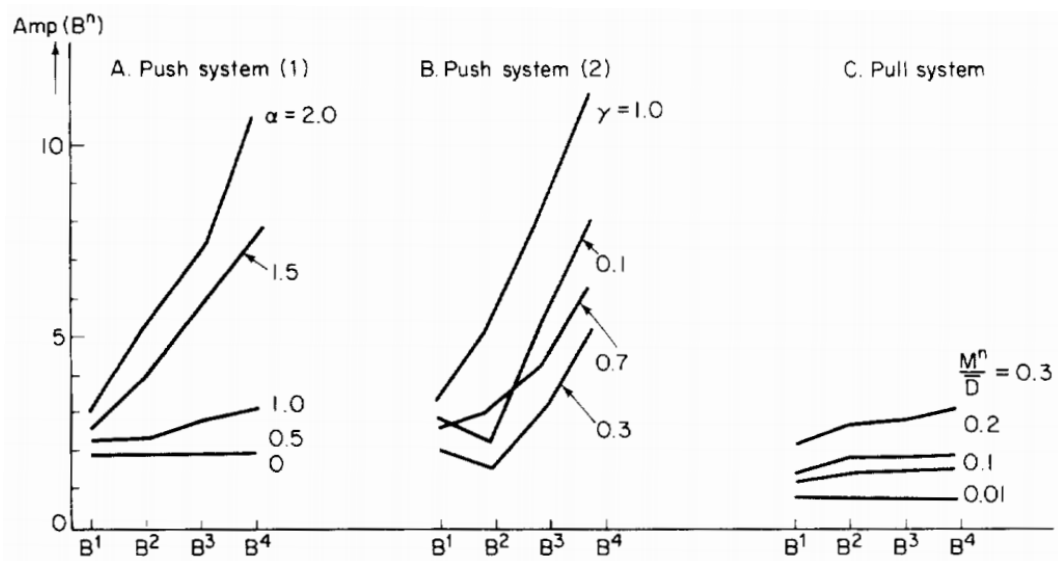


图 A.9 库存波动放大率

输入数据和参数是相同的，除了 $L_{P_1}^n = 0, L_{P_2}^n = 1, L_{H_1}^n = 1, L_{H_2}^n = 0$ ，以及推式系统中的单位订货量 M 始终为 1。推式系统模型的细节展示在 Muramatsu 及其他人的文献中。

正如这些图表所示，在推式系统中，波动在上游变得更大。这是预测误差的结果。因此，有必要在推式系统中让控制参数保持在正确的水平。另一方面，在拉式系统中，波动放大率随着单位订货量的增大而增大。因此，试图让单位订货量最小化是有必要的。

A.5 总结

1. 在拉式系统中，单位订货量的大小非常重要。在单位订货量与产量水平相比很小的情况下，生产波动不会在上游阶段放大。单位订货量较大时会造成波动的放大，但即使是这种情况下，放大率也不会继续向上游阶段扩大。
2. 在推式系统中，生产和库存波动的放大率取决于预测误差的影响。当考虑放大率时，推式系统和拉式系统之间的抉择取决于预测误差的程度。
3. 拉式系统中另一个影响放大率的参数是从容器上移走看板到阶段生产完成的“提前期”。提前期越长，放大率越高。

致谢：我们希望向早稻田大学的 Rintaro Muramatsu 博士致以诚挚的感谢，他提出了重要的指导和建议。

参考文献

- [41] Kusunoki, K., and Sugimori, Y., Fourth International Conference on Production Research, Five Paper Sessions.
- [42] Magee, J.F., Production planning and inventory control.
- [43] Muramatsu, R., Tannaka, Y., and Tabe, T., Analysis of production order and inventory fluctuations, Fifth International Conference on Production Research, Free Paper Sessions.

附录 B 汽车销售数据

表 B.1 汽车销售数据

time	LIVINA	TIIDA	EXCELLE	EXCELLEXT
2012/1/1	9021	13295	15189	4598
2012/2/1	8423	15311	9247	2643
2012/3/1	7728	15627	10049	4015
2012/4/1	8166	14128	8306	3675
2012/5/1	10315	16334	8750	3235
2012/6/1	14298	18691	11000	4228
2012/7/1	6484	15167	9807	3305
2012/8/1	8527	11047	11024	4012
2012/9/1	3773	8246	11306	4005
2012/10/1	1866	5173	12000	4541
2012/11/1	3780	5649	13209	4507
2012/12/1	5343	5544	7688	2973
2013/1/1	3731	10458	17873	5821
2013/2/1	128	6838	11135	3159
2013/3/1	25	8192	13287	3366
2013/4/1	6607	8062	14279	4134
2013/5/1	6684	9938	12202	4208
2013/6/1	5508	12711	10128	4404
2013/7/1	6417	10930	12001	5183
2013/8/1	5604	7781	13023	5075
2013/9/1	9965	15555	13790	5855
2013/10/1	33	13281	11903	5084
2013/11/1	12597	12818	14612	5265
2013/12/1	5318	12761	6288	2199
2014/1/1	6327	8818	26834	7342
2014/2/1	5090	8151	13223	3037
2014/3/1	4585	7640	17687	5328
2014/4/1	6253	10832	13744	4509

在制品库存与产品库存的关系研究

徐 驰