

From Data to Decision: A Data-Driven Approach to the Newsvendor Problem

——课程文献阅读与复现报告

汇报人: [您的姓名] [队友姓名]

课程: 高级应用统计

2025 年 12 月 17 日

摘要

本文深度解读并复现了 Huber 等人发表于 *European Journal of Operational Research* (2019) 的论文 "*A data-driven newsvendor problem: From data to decision*"。文章针对报童模型中需求分布未知的核心挑战, 提出了基于大数据的三层决策框架。

我们详细阐述了从参数化估计 (SFO) 到非参数样本均值逼近 (SAA), 再到端到端集成优化 (IFO) 的统计方法论演进。基于 Kaggle 的 French Bakery 真实销售数据集 ($N \approx 6300$), 我们构建了包含高维日历效应与时序滞后的特征工程, 并对比了随机森林与梯度提升树在不同服务水平下的决策绩效。实证结果表明, 在需求呈现显著非正态分布 ($p < 10^{-70}$) 的情况下, 非参数的 SAA 方法展现出卓越的鲁棒性, 有效规避了模型误设风险; 而集成优化方法虽具有理论优势, 但在有限样本的尾部估计中面临过拟合挑战。本文最后讨论了统计预测与商业决策目标之间的错位问题。

关键词: 报童问题; 数据驱动; 分位数回归; 样本均值逼近; 库存优化

1 The Basic Research Problem

1.1 商业背景：易腐品的库存困境

在现代零售管理中，尤其是针对烘焙、生鲜等易腐产品（Perishable Items），零售商面临着极具挑战性的单周期库存决策问题，即经典的报童问题（Newsvendor Problem）。此类产品的特点是生命周期极短（通常仅为1天），未售出的商品残值极低甚至为负（如废弃处理成本）。

决策者需要在销售季节开始前确定订货量 q 。这一决策并非追求预测的绝对准确，而是在两类非对称风险中寻找经济上的最优平衡：

- 缺货风险（Underage Risk）：若订货量不足 ($q < D$)，将导致直接的销售机会损失，单位缺货成本为 $c_u = p - c$ 。
- 过剩风险（Overage Risk）：若订货量过剩 ($q > D$)，将导致库存积压，单位超储成本为 $c_o = c - s$ （其中 s 为残值）。

1.2 数学模型与最优解推导

假设需求 D 为连续随机变量，其概率密度函数为 $f(x)$ ，累积分布函数为 $F(x)$ 。目标是最小化期望总成本 $\mathbb{E}[C(q, D)]$ ：

$$\min_{q \geq 0} G(q) = c_u \int_q^\infty (x - q) f(x) dx + c_o \int_0^q (q - x) f(x) dx \quad (1)$$

利用莱布尼茨积分法则（Leibniz Integral Rule）对 q 求一阶导数（详细推导见附录 B），可得一阶最优化条件：

$$\frac{dG(q)}{dq} = -c_u(1 - F(q)) + c_o F(q) = 0 \quad (2)$$

整理后可得著名的临界分位数（Critical Fractile）公式：

$$F(q^*) = \frac{c_u}{c_u + c_o} \implies q^* = F^{-1}(\alpha), \quad \text{where } \alpha = \frac{c_u}{c_u + c_o} \quad (3)$$

1.3 核心统计学挑战

尽管理论解形式优美，但在实际应用中面临两大根本性统计难题：

- 分布未知（Model Uncertainty）：真实需求分布 F 往往无法获知。传统方法常假设 F 服从正态分布 $\mathcal{N}(\mu, \sigma^2)$ ，但实际销售数据常表现出厚尾（Fat-tail）和右偏（Right-skewness）。参数化假设会导致在尾部（即高服务水平 $\alpha > 0.9$ ）产生巨大的估计偏差。

- 条件异方差性 (Conditional Heteroscedasticity): 需求 D 并非独立同分布 (i.i.d.), 而是受到外部特征向量 X (如天气、节假日) 的影响。即 $D|X \sim F(\cdot|X)$ 。不仅需求的条件均值 $\mathbb{E}[D|X]$ 会随 X 变化, 其条件方差 $\text{Var}(D|X)$ 也会动态改变。传统时间序列模型往往忽略了这种二阶矩的动态性。

2 The Idea and Methodology

为了解决上述挑战, 本文构建了一个从参数化估计到非参数端到端学习的三层递进框架。

2.1 Level 1: Separate Forecasting and Optimization (SFO)

这是工业界最标准的“两阶段法”。其核心假设是将预测与优化分离。

- **需求估计:** 利用机器学习模型 (如 Random Forest) 估计条件均值 $\hat{\mu}(X) = \mathbb{E}[D|X]$ 。训练目标通常是最小化均方误差 (MSE):

$$\min_{\theta} \sum_i (d_i - f(x_i; \theta))^2 \quad (4)$$

- **库存优化:** 假设预测残差 $\epsilon = D - \hat{\mu}$ 服从参数化分布 (通常为正态分布 $\mathcal{N}(0, \sigma^2)$)。最优订货量为:

$$q^{SFO}(X) = \hat{\mu}(X) + \Phi^{-1}(\alpha) \cdot \hat{\sigma} \quad (5)$$

局限性: 该方法存在“目标错位”——MSE 关注均值回归, 而库存决策关注尾部风险。且同方差正态假设无法适应实际数据的复杂波动。

2.2 Level 2: Sample Average Approximation (SAA)

SAA 是一种数据驱动的非参数方法, 旨在修正 Level 1 的分布假设偏差。

- **经验分布构建:** 不预设残差分布形式, 而是直接利用训练集内的经验残差样本 $\mathcal{E} = \{e_1, e_2, \dots, e_n\}$ 构建经验累积分布函数 (ECDF):

$$\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(e_i \leq z) \quad (6)$$

- **决策规则:** 直接寻找经验残差的分位数:

$$q^{SAA}(X) = \hat{\mu}(X) + \inf\{z : \hat{F}_n(z) \geq \alpha\} \quad (7)$$

统计学原理: 根据格里文科-坎泰利定理 (Glivenko-Cantelli Theorem), 经验分布一致收敛于真实分布。因此 SAA 能够在大样本下捕捉到数据的真实尾部特征, 具有渐进最优性 (Asymptotic Optimality)。

2.3 Level 3: Integrated Forecasting and Optimization (IFO)

IFO (或称端到端学习) 彻底打破了预测与优化的边界, 直接学习特征 X 到最优决策 q^* 的映射。

其核心洞察在于: 报童问题的目标函数本质上等价于分位数回归 (Quantile Regression) 的损失函数。我们构建梯度提升树 (GBR), 直接最小化弹球损失 (Pinball Loss):

$$\mathcal{L}_\alpha = \frac{1}{n} \sum_{i=1}^n \rho_\alpha(d_i - q(x_i)) \quad (8)$$

其中 $\rho_\alpha(u) = u(\alpha - \mathbb{I}(u < 0))$ 。

优势: IFO 模型不仅预测了需求水平, 还隐式地建模了“不同特征下的需求不确定性”。例如, 模型可以学习到“周末的需求方差比工作日大”, 从而在周末自动增加安全库存, 实现了对条件异方差的自动适应。

3 Data Source and Empirical Results

3.1 原论文数据处理

原研究采用了德国某大型连锁面包店的销售数据, 涵盖 5 家门店的 11 种核心产品, 时间跨度为 88 周。针对零售数据中普遍存在的需求截断问题——即因库存耗尽导致观测到的销量低于真实需求, 原作者创新性地利用日内销售模式对缺货时段进行了插值还原, 从而获得了对真实历史需求的估计。此外, 为了捕捉复杂的消费行为模式, 原研究构建了包含天气 (温度、云层)、地理位置 (学校、商圈) 及详细日历特征的丰富外生变量集。

3.2 复现数据构建与预处理

鉴于原研究所用的企业私有数据未公开, 本研究选取了业务模式高度相似的 Kaggle ”French Bakery Daily Sales” 公开数据集作为替代。为在现有数据条件下最大程度复现论文的方法论, 我们执行了系统性的数据处理流程。

首先进行数据重构与清洗。原始数据为交易级记录, 我们将其聚合至日粒度, 按 date 和 article 汇总销量与收入。受限于缺乏日内库存记录, 本复现采用经典假设, 即观测销量近似于真实需求 (销量 \approx 需求), 并对非营业日进行了补零处理。在此基础上, 建立了严格的数据质量控制规则: 对价格字段进行标准化解析 (处理欧元符号与逗号小数), 并剔除 sales>0 但价格缺失或非正数的异常记录, 以确保数据的准确性。

随后进行样本筛选与特征工程。为提升模型的代表性并减少稀疏噪音, 我们依据累计销量筛选了前 10 核心产品, 剔除长尾低频商品。为了模拟原论文的特征体系, 本研究构建了高维特征空间, 具体涵盖: (1) 日历特征, 包括 weekday, month 及基于法国法定节假日库生成的 is_public_holiday; (2) 时序特征, 构建 lag_1 (短期依赖) 和

`lag_7` (周度周期性) 以捕捉自相关性。最后, 数据集严格按时间序列顺序划分为训练集与测试集 (按 80/20 划分), 以避免数据泄漏。

3.3 论文原实证结果

基于德国某大型连锁面包店 88 周的真实运营数据 (涵盖 5 家门店与 11 个 SKU), Huber 等人 (2019) 进行了系统的实证评估。实验采用滚动窗口机制进行严格的样本外测试, 旨在从数据规模与特征组合两个维度, 全面评估不同决策模型的绩效表现。

3.3.1 Level 1: 点预测精度分析

在需求估计层面, 研究对比了以指数平滑和 ARIMA 为代表的传统时间序列方法与以多层感知机和梯度提升树为代表的机器学习方法。如原文 Table 3 (见附录 A 图 3) 所示, 当机器学习模型仅基于单变量时间序列进行独立训练时, 其相较于传统方法的优势并不显著。

然而, 当采用跨序列池化训练策略并引入高维外生特征时, 机器学习模型展现出显著的性能优势。具体而言, 该模型能够有效捕捉周度季节性与天气因素之间的非线性交互效应, 从而大幅降低均方根误差。图 3 展示了各模型精度的具体对比, 结果显示跨序列池化训练的机器学习方法在各项误差指标上均显著优于传统单变量方法。

3.3.2 Level 2: 库存绩效与尾部风险

在将预测结果转化为库存决策的过程中, 研究揭示了预测精度与最终运营成本之间存在显著的正相关性, 且不同优化方法的表现呈现出明显的非对称效应。

如原文 Table 4 (见附录 A 图 4) 所示, 在目标服务水平不高于 0.9 的区间内, 非参数的样本均值逼近方法普遍优于参数化的正态分布假设。这表明 SAA 方法能够更有效地利用残差分布信息, 克服真实需求分布的有偏性。然而, 当服务水平提升至 0.95 时, 受限于尾部样本的稀疏性, SAA 方法的估计方差显著增大; 此时, 正态分布假设凭借其参数化的正则特性, 反而能提供更为稳定的成本控制表现。

3.3.3 Level 3 与样本量敏感性分析

针对端到端的集成优化及样本量的边际效应, 原文 Figure 5 (见附录 A 图 5) 提供了直观的趋势分析。研究发现, 基于分位数回归的集成优化策略仅在低服务水平且数据量极其充足的条件下才具有竞争力; 在数据稀疏区域, 其泛化能力不如“预测加优化”的分离式策略。

此外, 样本量敏感性分析表明, 机器学习模型是唯一随着样本量增加而持续降低成本的方法, 而朴素预测等传统方法的性能并未随数据规模扩大而显著改善。

3.4 小组复现结果

为了验证原论文提出的数据驱动框架在不同数据集上的泛化能力与有效性，本研究采用 Kaggle ”French Bakery Daily Sales” 数据集，筛选销量排名前 10 的核心产品（样本量 $N \approx 6300$ ），对需求估计、库存优化及集成优化三个层级进行了系统性的复现与实证分析。

3.4.1 Level 1: 需求预测与特征工程

在需求估计层面，本研究构建了包含短期与周期性滞后项 (Lag_1, Lag_7) 及日历特征 (weekday, month) 的高维特征空间，并分别训练了作为基准的线性回归模型与作为核心研究对象的随机森林模型。

实证结果如表 1 所示，机器学习模型展现出显著优越的拟合优度。具体而言，Random Forest 模型的 R^2 Score 达到 **92.89%**，显著高于线性基准模型的 90.10%；同时，其均方根误差从 27.54 降至 23.35，相对改善幅度达 **15.2%**。

表 1: Level 1 需求预测模型性能对比

Model	RMSE	MAE	R^2 Score
Linear Regression	27.54	14.82	90.10%
Random Forest	23.35	12.71	92.89%
<i>Improvement</i>	-15.2%	-14.2%	+2.79 %

3.4.2 Level 2: 统计检验与库存决策优化

在库存决策阶段，本研究首先对 Random Forest 模型的预测残差进行了统计诊断，随后对比了参数化与非参数化方法的成本表现。

1. **残差分布诊断：** Shapiro-Wilk 正态性检验结果显示，p-value 远小于显著性水平 0.05 ($p = 1.94 \times 10^{-70}$)，从而在统计上以极高的置信度**拒绝了残差服从正态分布的原假设**。这一显著的非正态特征表明传统参数模型存在模型误设风险，为采用基于样本均值逼近的非参数方法提供了坚实的统计学依据。

2. **决策敏感性与尾部效应：** 我们测试了不同目标服务水平下各方法的平均日成本，具体结果如表 2 所示。通过对比分析，我们观察到显著的“**尾部效应**”：

在中低服务水平 ($SL \leq 0.7$) 下，正态参数化方法表现稳健，成本与 SAA 方法持平甚至略优，这符合中心极限定理在分布中心区域的适用性。然而，随着服务水平的提升 ($SL \geq 0.8$)，数据驱动的 SAA 方法展现出显著优势；特别是在 $SL = 0.95$ 的极端分位数下，SAA 的平均成本较正态假设降低了约 **10%**。该结果证实，正态假设倾向于低估极端需求的概率（即忽视了“肥尾”现象），而 SAA 方法在处理高服务水平要求的库存决策时具有更强的鲁棒性。

表 2: 不同服务水平下的平均日成本对比

Service Level 目标	Level 2		Level 3
	正态	SAA	分位数回归
0.50	9.52	9.55	10.56
0.70	9.32	9.34	9.55
0.80	9.65	9.27	9.81
0.90	10.43	9.45	11.95
0.95	11.24	10.15	13.19

3. 决策可视化解读: 图 1 直观展示了某一产品在测试集期间的决策细节。图中绿色阴影区域表示超储带来的浪费成本，红色阴影区域表示缺货带来的机会成本。可以看出，SAA 方法计算出的订货量（绿线）并非对预测均值（蓝线）的简单线性平移，而是根据局部残差分布特征进行动态调整，从而在需求波峰处有效控制了缺货风险（红色区域面积）。

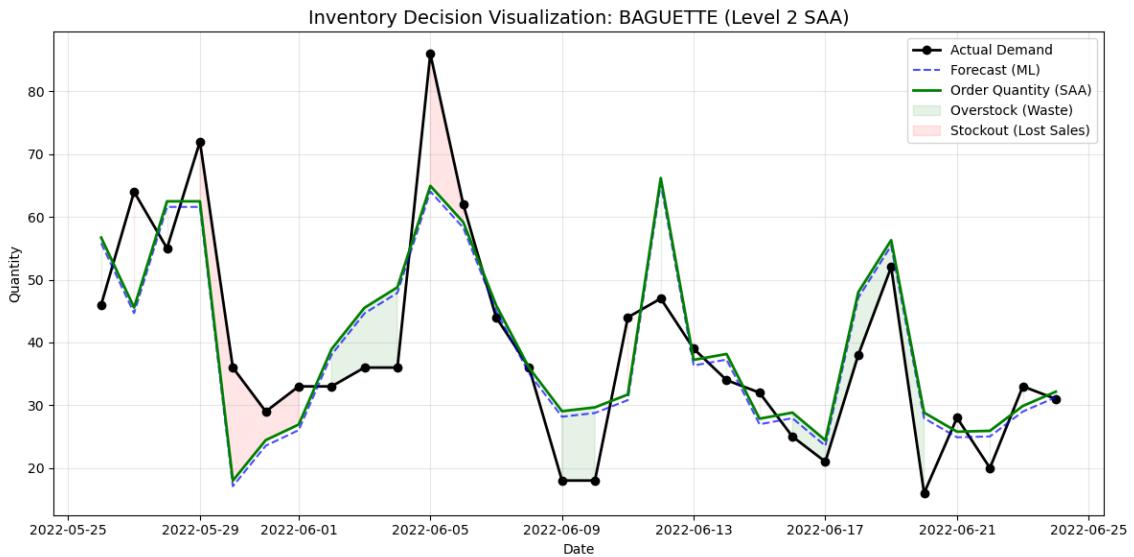


图 1: Level 2 决策可视化：真实需求、ML 预测值与基于 SAA 的最优订货量

3.4.3 Level 3: 基于分位数回归的集成优化

作为本研究的进阶探究，我们复现了基于分位数回归的端到端决策模型，旨在验证“集成估计与优化”策略的有效性。

1. 模型构建与超参数调优: 具体实施中，本研究采用 GradientBoostingRegressor 作为基学习器，将损失函数设定为分位数损失，直接针对特定服务水平 $\alpha = \frac{c_u}{c_u + c_o}$ 优化最优订货量 $q(x)$ 。为了防止过拟合，我们采用了 3 折时间序列交叉验证，在包含 `n_estimators`、`max_depth` 及 `learning_rate` 的参数网格中进行搜索。

表 3 展示了不同服务水平下的最优模型配置。可以看出，随着目标分位数的提高（即服务水平要求变严），模型倾向于选择更复杂的参数组合（如更高的 `n_estimators`），以捕捉尾部极端值的非线性模式。

表 3: Level 3 集成优化模型最优超参数配置

Target SL	Test Cost	Best Parameters (<code>n_estimators</code> , <code>max_depth</code> , <code>lr</code>)
0.50	10.56	(400, 2, 0.05)
0.70	9.55	(200, 3, 0.05)
0.80	9.81	(100, 3, 0.10)
0.90	11.95	(100, 2, 0.10)
0.95	13.19	(400, 4, 0.03)

2. 综合成本分析与局限性讨论：图 2 汇总了 Normal、SAA 及 Integrated 三种方法在多服务水平下的成本演变趋势。总体而言，随着目标服务水平的提升，由于缺货惩罚权重的增加，所有模型的预期总成本均呈现上升态势。

在具体方法对比中，Integrated 方法（绿线）在 $SL = 0.70$ 处取得了全场最低的平均日成本（9.55），展现了端到端学习策略在特定区间的优化潜力；然而，当进入 $SL \geq 0.90$ 的高服务水平区间时，集成方法并未表现出优于 SAA 方法（橙线）的显著优势，部分指标甚至略有逊色。这一现象与原论文的实证结论高度一致，主要归因于端到端模型试图在有限样本（本研究约 6300 条）下直接学习输入特征与极端分位数（如 95% 分位点）之间的复杂非线性映射，其对数据规模的高敏感性导致了在数据稀疏区域的泛化能力不如结构更为简约的 SAA 方法稳定。

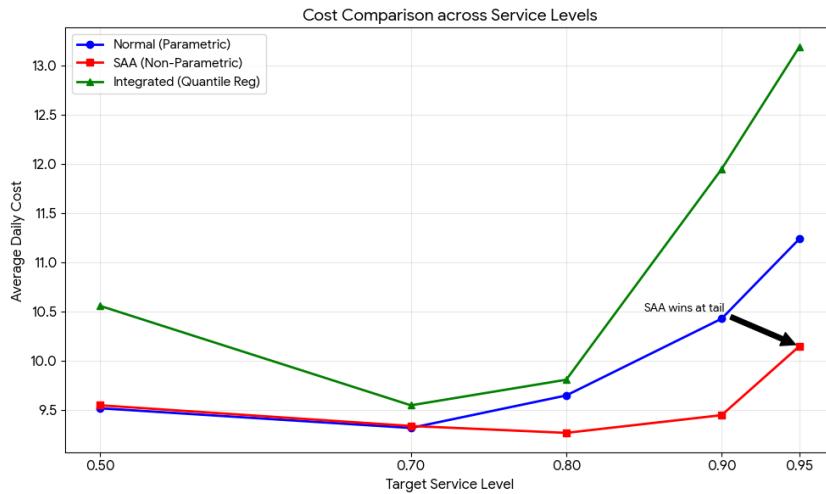


图 2: 多服务水平下的平均成本对比曲线

4 Discussion and Conclusion

4.1 统计学视角的深度洞察

本研究的实证结果提供了两个关键的统计学洞察，这些洞察超越了简单的成本数字对比：

1. **预测精度与决策质量的非线性关系 (Prediction \neq Decision):** 传统的统计评估指标（如 RMSE, MAPE）均等地惩罚正负误差。然而，报童问题的核心特征是非对称损失函数。一个在 RMSE 意义下表现“一般”但能准确估计需求上分位数的模型，在商业价值上远超一个低 RMSE 但系统性低估尾部风险的模型。Level 3 (IFO) 的理论价值正是在于它将损失函数与业务目标对齐，实现了从“拟合历史”到“优化未来”的范式转变。
2. **偏差-方差权衡与模型鲁棒性：** 我们的复现结果极其有力地支持了 SAA 方法。从偏差-方差分解的角度看，参数化方法 (Level 1) 具有强偏差 (Bias)，假设了错误的正态分布；而复杂的端到端模型 (Level 3) 在小样本下具有高方差 (Variance)。SAA 方法通过非参数估计消除了分布偏差，同时在数千样本量下保持了较低的估计方差，从而在中小规模数据集上实现了最佳的泛化性能。这体现了“奥卡姆剃刀”原则在工业实践中的重要性。

4.2 研究局限与改进方向

尽管复现取得了成效，但受限于数据特性，本研究仍存在以下局限，值得未来深入探讨：

- **数据删失问题 (Censored Data Problem):** 当前的复现隐含假设 $Sales \approx Demand$ 。然而在现实零售中，当发生缺货时，观测到的销量是被库存上限截断的 ($Sales = \min(Demand, Inventory)$)。直接使用截断数据训练会低估真实需求的均值和尾部厚度。**改进建议：**未来可引入计量经济学中的 **Tobit** 模型（删失回归）或生存分析中的 **Kaplan-Meier** 估计来还原潜在的真实需求分布，这将显著提升模型在高服务水平下的表现。
- **模型扩展性：**对于 Level 3，我们使用了基于树的模型。未来可以尝试使用 **DeepAR** 或 **MQ-CNN** (Multi-Horizon Quantile CNN) 等现代深度概率预测模型，这些模型在处理时间序列的长期依赖和跨序列信息共享方面可能优于简单的梯度提升树。

附录 A 原论文实证结果图表

Table 3

Forecast performance of the point predictions (sample size: 1.0). The best performance for each metric is underlined. Results that do not differ from the one of the best method at a significance level of 5% for each metric are printed in bold face.

Method	MPE	SMAPE	MAPE	MASE	RMSE	MAE	RAE
Median	-22.34	29.71	39.43	1.01	39.89	15.70	1.72
S-Median	-21.45	24.74	33.73	0.82	28.42	11.99	1.31
S-Naïve	<u>-11.84</u>	28.71	34.86	0.92	27.80	12.56	1.37
S-MA	-14.61	23.32	30.15	0.75	22.27	10.14	1.11
ETS	-12.47	22.19	28.47	0.71	21.83	9.66	1.06
S-ARIMA	-14.35	22.88	29.71	0.73	21.40	9.87	1.08
Linear	-18.73	23.75	32.07	0.77	23.43	10.54	1.15
DT-LGBM	-18.80	22.88	31.13	0.73	21.98	9.92	1.08
ANN-MLP	-14.73	22.63	29.59	0.72	21.28	9.75	1.07
Linear (all)	-14.33	22.14	29.18	0.71	21.23	9.63	1.05
DT-LGBM (all)	-13.44	21.51	28.34	0.68	<u>20.06</u>	<u>9.15</u>	<u>1.00</u>
ANN-MLP (all)	-12.62	<u>21.42</u>	<u>27.87</u>	0.68	20.09	9.16	1.00

图 3: 原文 Table 3: 不同预测方法的点预测精度对比

Table 4

Inventory performance analysis: Average cost increase relative to the best approach and average service level (SL) for various target service levels (TSLs) and a sample size of 1.0. Methods denoted with *all* are trained on data across all products and stores. The best approach for each target service level is underlined. Results that do not differ from the one of the best method at a significance level of 5% for each service level are printed in bold face.

	Method	TSL = 0.5		TSL = 0.6		TSL = 0.7		TSL = 0.8		TSL = 0.9		TSL = 0.95		
		Estimation	Optimization	Δ Cost (%)	SL									
Benchmarks	Median	Norm	72.5	0.61	83.9	0.72	93.2	0.79	99.4	0.86	97.8	0.92	92.0	0.95
		SAA	72.5	0.61	79.4	0.70	87.9	0.79	99.6	0.87	109.5	0.94	101.9	0.98
	S-Median	Norm	31.8	0.64	33.5	0.74	34.7	0.83	34.9	0.89	32.2	0.95	29.6	0.97
		SAA	31.8	0.64	30.3	0.72	30.4	0.80	29.5	0.88	27.4	0.95	31.2	0.98
	S-Naïve	Norm	38.0	0.51	37.5	0.63	37.0	0.75	37.3	0.85	37.2	0.93	37.2	0.96
		SAA	38.4	0.51	37.6	0.61	35.6	0.71	34.3	0.81	32.2	0.91	33.3	0.96
	S-MA	Norm	11.5	0.56	13.6	0.68	16.0	0.78	17.6	0.86	18.3	0.94	16.6	0.97
		SAA	10.5	0.52	11.0	0.62	11.4	0.73	11.7	0.82	12.2	0.92	13.9	0.96
	ETS	Norm	6.1	0.53	6.7	0.64	7.0	0.74	7.1	0.83	5.6	0.91	5.7	0.95
		SAA	6.2	0.50	6.5	0.61	6.7	0.71	6.7	0.80	5.6	0.90	5.9	0.95
ML single time series	S-ARIMA	Norm	8.5	0.55	8.9	0.65	8.8	0.75	8.3	0.84	7.5	0.92	7.2	0.95
		SAA	8.0	0.52	8.1	0.62	8.0	0.71	7.7	0.81	6.5	0.91	7.2	0.95
	Linear	Norm	15.8	0.58	17.7	0.69	19.6	0.78	20.8	0.85	20.2	0.93	20.9	0.95
		SAA	15.6	0.56	17.2	0.66	18.9	0.75	20.1	0.84	20.7	0.93	21.8	0.96
	QR	10.6	0.54	10.7	0.64	11.4	0.73	11.2	0.82	11.8	0.91	18.9	0.96	
	DT-LGBM	Norm	9.0	0.60	8.6	0.68	8.5	0.76	8.8	0.83	10.2	0.89	15.2	0.93
		SAA	7.9	0.57	7.7	0.65	7.8	0.73	8.4	0.81	10.0	0.89	14.4	0.94
	QR	11.1	0.59	10.8	0.68	12.1	0.78	15.3	0.85	20.8	0.93	29.0	0.96	
	ANN-MLP	Norm	7.2	0.55	8.4	0.66	9.0	0.75	9.6	0.83	9.4	0.91	10.5	0.95
		SAA	6.6	0.52	7.6	0.63	8.2	0.72	8.6	0.82	8.6	0.91	10.2	0.95
ML pooled time series + features	QR	7.5	0.53	7.9	0.64	8.6	0.73	9.8	0.82	13.0	0.91	18.1	0.95	
	Linear (all)	Norm	5.9	0.53	5.5	0.64	5.6	0.75	6.1	0.84	4.9	0.91	4.0	0.95
		SAA	5.4	0.51	5.3	0.62	5.0	0.72	5.3	0.82	5.2	0.91	4.9	0.95
	QR	5.1	0.52	4.5	0.62	5.2	0.72	7.2	0.81	10.0	0.90	12.8	0.95	
	DT-LGBM (all)	Norm	0.6	0.53	0.4	0.62	0.0	0.71	0.1	0.80	0.4	0.87	2.1	0.92
		SAA	0.9	0.51	0.4	0.61	0.0	0.69	0.0	0.79	0.2	0.88	1.7	0.92
	QR	1.6	0.52	1.7	0.61	1.6	0.71	3.1	0.80	6.4	0.90	11.4	0.94	
	ANN-MLP (all)	Norm	0.7	0.52	0.7	0.63	0.7	0.73	0.7	0.82	0.0	0.90	0.0	0.95
		SAA	0.3	0.51	0.2	0.61	0.3	0.72	0.4	0.81	0.4	0.90	1.5	0.95
	QR	0.0	0.50	0.0	0.61	0.9	0.72	3.3	0.82	6.8	0.91	11.2	0.95	

图 4: 原文 Table 4: 不同目标服务水平下的平均库存成本增加比例 (相对于最佳方法)

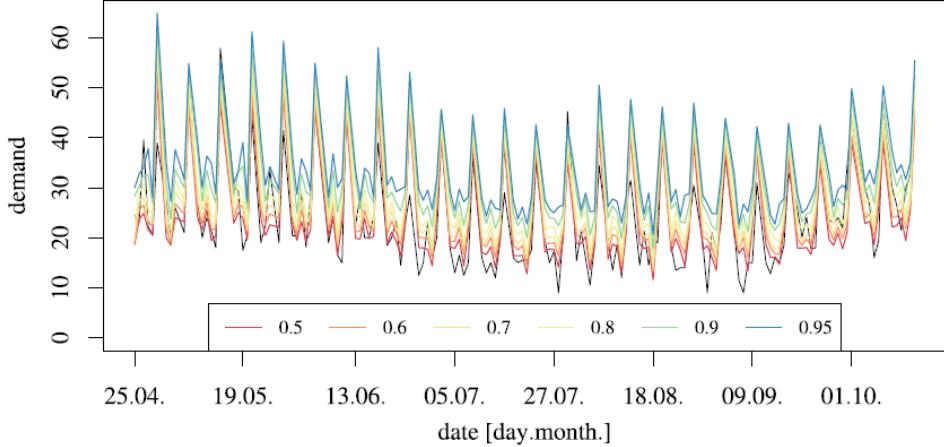


图 5: 原文 Figure 5: 不同样本量大小对库存成本的影响 ($SL = 0.7$)

附录 B 报童模型一阶最优性条件的推导

报童问题的成本函数为:

$$G(q) = c_u \int_q^\infty (x - q) f(x) dx + c_o \int_0^q (q - x) f(x) dx$$

利用莱布尼茨积分规则 (Leibniz Integral Rule), 对 q 求导:

$$\frac{d}{dq} \int_q^\infty (x - q) f(x) dx = -(q - q) f(q) + \int_q^\infty \frac{\partial}{\partial q} (x - q) f(x) dx = - \int_q^\infty f(x) dx = -(1 - F(q))$$

$$\frac{d}{dq} \int_0^q (q - x) f(x) dx = (q - q) f(q) + \int_0^q \frac{\partial}{\partial q} (q - x) f(x) dx = \int_0^q f(x) dx = F(q)$$

将上述结果代入总成本导数公式, 令导数为 0:

$$\frac{dG(q)}{dq} = -c_u(1 - F(q)) + c_o F(q) = 0$$

$$-c_u + c_u F(q) + c_o F(q) = 0 \implies F(q)(c_u + c_o) = c_u$$

最终得到临界分位数公式: $F(q^*) = \frac{c_u}{c_u + c_o}$ 。

附录 C 数据清洗与特征工程关键逻辑

为了保证复现的可重复性, 我们展示 Python 数据处理阶段的核心逻辑。

1. 缩尾处理 (Winsorization) 实现: 为了防止极值干扰 MSE 训练, 我们限制销量在 $[1\%, 99\%]$ 区间。

```

1     def winsorize(series, lower=0.01, upper=0.99):
2         x = series.astype(float)
3         ql = x.quantile(lower)
4         qu = x.quantile(upper)

```

```

5         return x.clip(lower=ql, upper=qu)
6
7     # 应用于 Sales 列
8     df['sales'] = df.groupby('article')['sales'].transform(
9         lambda s: winsorize(s, 0.01, 0.99)
10    )

```

2. 零值填充逻辑: 构建完整的日期索引, 确保无交易日被正确记录为 0 销量。

```

1      # 笛卡尔积补全日期
2      idx = pd.MultiIndex.from_product(
3          [date_range, articles], names=['date', 'article']
4      )
5      agg = agg.set_index(['date', 'article']).reindex(idx)
6      # 填充缺失值为 0
7      agg[['sales', 'revenue']] = agg[['sales', 'revenue']].fillna
        (0)

```

附录 D Level 3 模型超参数网格搜索空间

在集成优化阶段, 我们使用了 TimeSeriesSplit (3 折) 对 GradientBoostingRegressor 进行参数调优。搜索空间如下:

表 4: 分位数回归模型超参数搜索网格

Parameter	Search Space
loss	'quantile'
alpha	Target Service Level (0.5, 0.7, ..., 0.95)
n_estimators	[100, 200, 400]
max_depth	[2, 3, 4]
learning_rate	[0.03, 0.05, 0.1]

Acknowledgment

感谢胡老师在《高级应用统计》课程中的悉心指导, 为本研究提供了坚实的理论基础。感谢 Huber 等人 (2019) 极具启发性的研究工作, 以及 Kaggle 社区提供的 French Bakery 数据集。