

案例：进入量产后，预测怎么做？

刘宝红

作者按：这些案例是基于一些学员公司的挑战而开发的，旨在阐述“从数据分析开始，由职业判断结束”的决策流程，不管是需求预测，还是库存计划；不管是新产品，还是老产品。对于大部分案例，我也把原始数据放到网站上，供大家下载、分析，更深入地了解数据模型。这些案例还是半成品，希望大家发表自己的看法，让案例更加完善。您可以直接留言，或发 email 给我——刘宝红，136 5127 1450（微信同）| baohong@hotmail.com。

在《上新结束后，预测怎么做》案例中，我们探讨的是新产品导入。这里要讨论的是，进入量产后，后续的需求该怎么预测？我们这里介绍一种预测方法：指数平滑法。这套方法论在上世纪 50 年代发展成熟，由一些运筹学专家研究出来，在实践中应用很广，主要用在库存预测上¹。

指数平滑法是一种特殊的移动平均法。对不同时段的实际需求，指数平滑法赋予的权重不同：越是近期的实际值，权重越大；越是远期的数据，权重越小，这使得预测模型能够更快反映实际的变化。这些是通过平滑系数 α 来实现： α 越大，预测越灵敏，越能尽快反映实际变化，当然也越受随机因素影响，带给供应链的波动也越大； α 越小，越多的随机性被过滤掉，预测也表现得越平稳，给供应链的运营成本越低，但风险是没法及时反映市场的需求变化。

指数平滑法的公式有两种形式：

$$\text{预测值} = \alpha * \text{上期实际值} + (1 - \alpha) * \text{上期预测值} \quad (\text{公式 1})$$

或者

$$\text{预测值} = \text{上期预测值} + \alpha * (\text{上期实际值} - \text{上期预测值}) \quad (\text{公式 2})$$

这本书是写给实践者看的，从管理的角度来阐述，所以我会尽量避免成串的公式；我会用简单的语言来阐述，让大家从管理的角度理解这些公式的意义。至于具体的公式，大家可以百度更详细的内容。这些公式大都是六七十年前研究出来的，并不算火箭技术²；问题是，这些本来很简单的方法论，却被很多著述解释得异常晦涩难懂——一旦学问们开始卖弄“数学之美”，推导出一串又一串的公式时，对我们实践者而言无异于灾难，导致没法在管理中广泛应用。

¹ Forecasting Methods and Applications, 作者 Spyros Makridakis, Steven C. Wheelwright, Rob J. Hyndman, Wiley 出版, 第三版, 2018 年重印, 第 174 页。

² 说到这些, 我想介绍一下上世纪 50 年代美国的 HMMS 研究团队。这是四位研究者姓氏的第一个字母 (Holt, Modigliani, Muth 和 Simon), 他们都在卡耐基工学院 (后来与梅隆学院合并, 成为卡耐基梅隆大学)。他们的目的是为工业界制定更好的决策机制, 来应对种种库存、生产和计划问题。这些问题在宏观层面导致经济危机, 在微观层面让企业处于不停的危机状态, 不管是赶工加急, 还是产能利用不足, 以及库存积压。Holt 就是我们这里要讲的霍尔特, 他开发了应对平缓需求的简单指数平滑法, 应对趋势的霍尔特双参数线性指数平滑法, 以及应对季节性的霍尔特-温特模型, 都成为工业界最为广泛应用的预测模型。其余的 3 位研究者中, Modigliani 和 Simon 后来获得了诺贝尔经济学奖, 而 Muth 的理性预期模型呢, 又成为卢卡斯获取诺贝尔奖的基石。引自 Charles c. Holt 的文章 Learning How to Plan Production, Inventories, And Work Force, Operations Research, Vol. 50, No. 1, January-February 2002, pp. 96-99.

让我们拿踢足球打个比方，来解释指数平滑法的逻辑。预测就如防守方：防守者是按照自己的预测行事，并随着球的最近落点（实际需求），调整预测，决定下一步该怎么走。下一步的预测介于两个极端：一个极端是“我行我素”，严格按照原来的计划（上个预测），实际上是拿原来的预测作为下一步的预测（平滑系数 α 为 0）；另一个极端是“步步紧逼”，上一次球落到什么地方，就赶到什么地方，实际上是拿最新的实际值作为下一步的预测（平滑系数 α 为 1）。

“我行我素”是以不变应万变，风险是可能没法及时响应变化了的局势；“步步紧逼”看上去更积极，实际上却一直慢一步，永远跟着球跑，永远也没法抢到球。所以，在实际操作中，我们会在这两者之间取值，目标是让防守者尽可能接近球，最大化抢到球的概率。这就是优化平滑系数 α ，以最大化预测的准确度。

【小贴士】预测的灵敏度和准确度

预测中，预测的灵敏度（响应度）和准确度经常相抵触：灵敏度高了，准确度就下降；准确度高了，灵敏度则受影响。对指数平滑法来说，最高的灵敏度就是平滑系数等于 1，那就是跟着球跑，看上去让人觉得很“响应”，却是典型的被动反应，给供应链导入频繁的变动，导致产能利用率低，运营成本高，也注定永远没法超前。如果要超前，就得预判，就得按照一定的战略行事，沿着特定的路径前行，注定不会亦步亦趋地“紧贴”需求，在灵敏度上受限，往往得牺牲短期来获取长期利益，在企业追求无限响应的今天，往往也更不受欢迎。

当然，当需求变动幅度非常大，速度非常快的时候，比如有些快时尚，亦步亦趋地跟着实际跑，可能比人的预测还要准确，那就不如不预测，光靠响应速度来应对。比如在规模还比较小的时候，快鱼基于约束理论（TOC）采取的快反策略就是例子。但是，当企业大了，需求大了，供应链的响应速度就跟不上（比如没有太多的富裕产能，而安全库存也没法对付多时的需求），这时候就是问题，需要通过计划来改善。这又回到预测的本质：要预测，是因为供应链的响应速度不够。

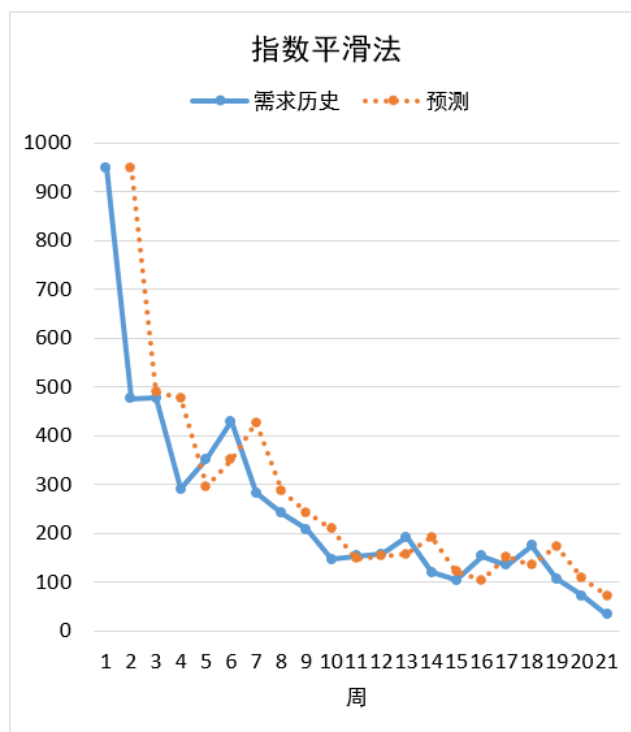
另外，你会注意到，在很多预测模型中，包括这里的指数平滑法，我们都是基于上一次的预测做下一次的预测。这似乎有点违背常理：所有的预测都是错的，为什么还要考虑以前的预测？这里有两个原因。其一，我们现在的落点不是上一步实际发生的，而是上一步预测要发生的（我们是在按预测行事）。所以，下一步的出发点不是上次的实际值，而是上一步的预测。其二，实际发生的有很多偶然性，表现为“噪音”，而以前的预测把部分“噪音”给过滤掉了，所以包含着很多历史经验和智慧，连最简单的平均法、移动平均法都有这种功能。

放到指数平滑法中，这个“过滤器”就是平滑系数。简单地说，平滑系数就是给历史信息“打折”，尽可能多地过滤掉“噪音”。要特别强调的是，这里的“历史信息”更多是综合在上一个预测中，而不是上一个实际值。如果你展开指数平滑法的公式，你会发现，需求历史按照 $(1-\alpha)$ 的等比数列级数综合到预测中（这也是指数平滑法中“指数”的来历）。比如对最新一期的权重是 $(1-\alpha)$ ，二期的是 $(1-\alpha)$ 的平方，三期的是 $(1-\alpha)$ 的三次方，以此类推。因为 $1-\alpha$ 的值介于 0 和 1 之间，所以次数越高，权重就越小，以几何数级衰减。

具体来说，平滑系数 α 的值取决于需求历史的稳定性：需求历史越稳定，需求历史（表现在上次的预测中）的权重越高，最近实际发生的所占权重越低（越是把它当做偶然因素）；反之亦然。当需求历史比较稳定时，选择较小的 α 值，0.05-0.2；需求历史有波动，但长期趋势没有大的变化，可选择稍大的 α 值，0.1-0.4；当波动很大，呈现明显且迅速的上升或下降趋势时，宜选取较大的 α 值，0.6-0.8；当需求历史是上升或者下降序列， α 取较大值，0.6-1³。

在具体应用中，先把需求历史做成散点图，大致判断需求的稳定性和趋势，确定平滑系数的大致范围，然后套用几个不同的 α 值，一般以 0.1 作为阶梯，看哪个的准确度最高。如果是用 Excel 做计划的话，也可以用 Excel 中的插件 Solver，基于特定的优化目标，比如平均绝对误差或者均方差最小，找到最优化的平滑系数。

	周数	需求历史	预测	误差
初始组	1	949		
	2	476	949	-473
	3	478	489	-11
	4	291	478	-187
	5	353	296	57
	6	430	351	79
	7	283	428	-145
	8	242	287	-45
	9	209	243	-34
测试组	10	147	210	-63
	11	154	149	5
	12	158	154	4
	13	193	158	35
	14	120	192	-72
	15	104	122	-18
	16	154	105	49
	17	135	153	-18
	18	175	136	39
	19	107	174	-67
	20	72	109	-37
	21	34	73	-39
预测值	22		35	



测试组	α	0.97
	平均绝对误差	37
	平均绝对误差%	37%
	均方误差	1869

图 1：指数平滑法来预测快消品的需求

如图 1 所示，指数平滑法会用一部分历史数据初始化和优化模型，用另一部分历史数据来验证、测试模型。初始化就如正式比赛前的热身，或者设备投入正式运营前的试车，在试车走合的过程中，我们可能调整有些参数，让设备处于更好的状态（优化）。对于预测模型来说，这种调整更多的是自动调整——好的预测模型往往有一定的“自适应性”。

³ <https://blog.csdn.net/cl1143015961/article/details/41081183>。

就如赛前热身不算正式成绩，我们也按照初始化的结果来判断预测模型的优劣——我们得用专门的数据来判断预测模型的优劣，跟别的模型比较，以选择最佳的模型和参数，然后推测后续的预测值。在预测方面的经典著述中，一般用最初的 9 个数据点作为指数平滑法的初始数据⁴。如果是按照季度汇总，大致就是 2 年的数据；如果按照周来汇总，大致就是 2 个月的数据。

就如指数平滑法的基本公式所示，下期的预测是基于上期的实际值和预测值。在初始化时，预测的初始值可以取第一期的实际值（初始组数据点较多时），或者前几期的平均实际值（初始组数据点较少时）。究竟多少数据点算多，我在有些文献中看到是 15 个以上。实际上，你会很快发现，预测的初始值影响甚微，因为权重只有 $(1-\alpha)$ 的 N 次方，而 $(1-\alpha)$ 介于 0 和 1 之间，次方越高，衰减地越厉害，相应的影响也越轻。就拿下面的案例来说，该初始值对于测试数据的影响系数是 $(1-\alpha)$ 的 8 次方以上，对预测值的影响是 20 次方，权重有多大，算算就知道了。

在这个案例中，我们拿第 1 周的实际值作为第 2 周的预测，然后开始代入公式运算。我们用 1 到 9 周的需求历史供初始化，并优化平滑系数 α 为 0.97；然后用 10 到 21 周的需求历史来测试，检验预测模型的准确度；最后，我们用这个模型，预测第 22 周的需求。具体的模型在 Excel 表格中，可扫描二维码，或者访问链接 scm-blog.com/case.html 来下载。



预测的准确度有多种方法来衡量，比如误差、绝对误差的平均值，以及它们的百分比。文献中常出现的 MAPE⁵，就是绝对误差百分比的平均值。这说起来挺拗口，计算起来其实很简单：实际值减去预测，取绝对值，除以实际值，得到百分比，再把多期的百分比平均就是了。这个值很直观，但也容易误导：当实际值非常小，特别是接近 0 时，这一百分比可能很大。解决方案是设定上限，比如绝对误差的百分比不超过 100%。

这里再介绍一下均方误差，亦即文献中经常提到的 MSE⁶。这是给每期预测和实际值的差值平方，然后再平均。平方的好处是放大极端误差：误差小了我们往往可以应对，比如设置安全库存，或者适当地赶工加急；害死我们的大都是极端误差——预测太低的话，需求很容易击穿安全库存，导致高昂的赶工加急成本；预测太高的话，则容易造成大量的积压，以及由此而来的呆滞库存。这点对很多人来说，相信都有切身体会。给误差平方，就是加倍“惩罚”那些超级误差，凸显那些极端虚高或虚低的预测，这也是我们应该重点避免的。

严格地讲，均方误差并不能直观地体现预测的准确度。比如均方误差是 180 究竟意味着什么，谁也说不清。不过“不怕不识货，就怕货比货”，在比较不同预测方法时，这一参数却非常有用——均方差越小，表明预测模型越准确。所以，均方误差也是优化预测模型的重要考量。比如在这个案例中，我们就是以均方误差最小为目标，建立指数平滑模型，利用初始数据，确定最优的平滑系数 α 为 0.97。

⁴ Forecasting Methods and Applications, 作者 Spyros Makridakis, Steven C. Wheelwright, Rob J. Hyndman, Wiley 出版, 第三版, 2018 年重印。

⁵ 英语的全称是 Mean Absolute Percentage Error, 直译就是绝对误差百分比的平均值。

⁶ 英语的全称是 Mean Square Error, 直译就是误差平方的均值。

值得注意的是，一个目标优化了，并不意味着另一个目标也优化。比如绝对误差百分比的平均值最小时，平均绝对误差、均方差等并不一定最小。

让我们再看一下**累计误差**。累计误差可帮助判断预测模型是否有系统性偏差。理想情况下，一个数据模型会有时候高估，有时候低估，残差会部分互相抵消，累计值会较小，如果不是零的话。在本案例的指数平滑法模型中，大多预测值是高于实际值的，累计误差也体现为过剩，而且持续增加，呈现失控情形（如图 2）。背后的原因呢，是因为该快消品的需求呈现明显的下降趋势，简单的指数平滑法并不是最佳的预测方法——简单指数平滑法最适合于没有趋势、没有周期性的平缓情形。

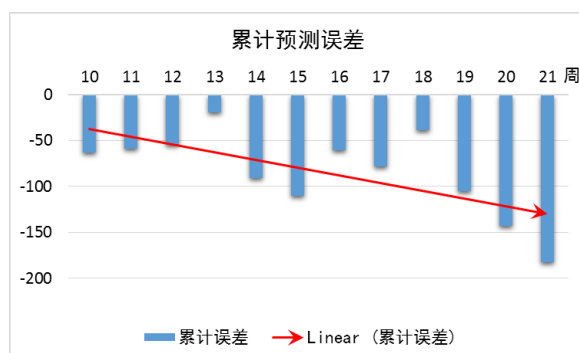


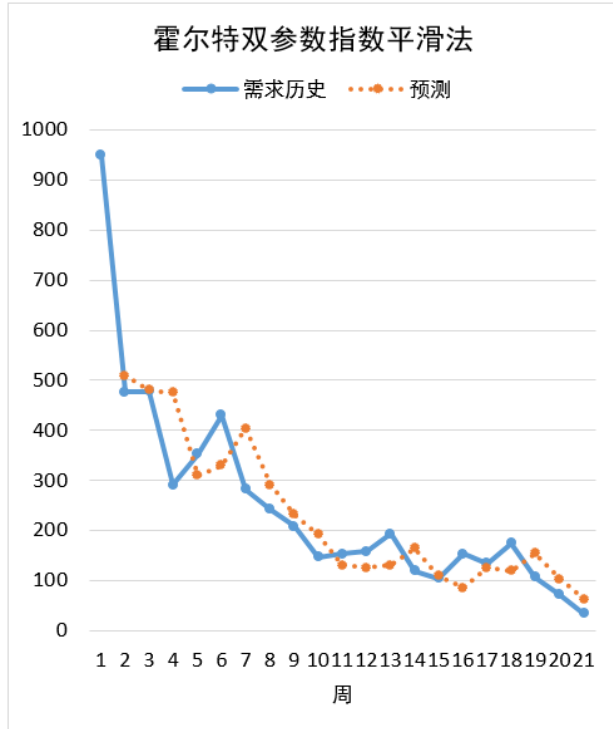
图 2：累计误差反映了系统性偏差

在指数平滑模型中，最优的平滑系数是 0.97，很接近 1，近乎用上期的实际值当做下期的预测值。这相当于跟着球跑，你其实是不需要什么预测模型来告诉你这点。这也表明，在这个案例中，简单指数平滑法没有增加多少价值；我们得选择更加合适的预测方法，这就是下面要讲的霍尔特双参数线性指数平滑法（简称霍尔特法）。

霍尔特法是按照查尔斯·霍尔特命名的。该方法论发表于 1957 年，成为预测上用得最广的模型之一。为了与霍尔特以及霍尔特-温特模型区别，我们也把“指数平滑法”成为“简单指数平滑法”。简单地说，霍尔特法就是在简单指数平滑系数 α 的基础上，增加了一个趋势的平滑系数 β ，所以叫“双参数平滑法”。当 β 等于 0 的时候，霍尔特模型就成了简单指数平滑法。

在上面的例子中，该快消品的需求呈现明显的下滑趋势，简单指数平滑法没法有效应对，就只能“步步紧逼”，“球”跑到哪里，就跟到哪里，这注定是滞后一步，永远没法超前（不会打球的人经常这么干）；霍尔特法增加了趋势的参数，更好地预判“球”的走向，系统增加了抢到“球”的胜算（有经验的球员的做法），这从其提高了的预测准确度上可以得到验证（见下图 3）。

	周数	需求历史	截距	斜率	预测	误差
初始组	1	949	509	0		
	2	476	483	-3	509	-33
	3	478	478	-3	480	-2
	4	291	328	-18	476	-185
	5	353	344	-14	310	43
	6	430	410	-6	330	100
	7	283	307	-16	404	-121
	8	242	252	-20	291	-49
	9	209	214	-22	232	-23
测试组	10	147	156	-25	192	-45
	11	154	149	-23	131	23
	12	158	152	-21	126	32
	13	193	181	-16	131	62
	14	120	129	-19	165	-45
	15	104	105	-20	110	-6
	16	154	140	-14	85	69
	17	135	133	-14	126	9
	18	175	164	-9	120	55
	19	107	117	-13	155	-48
	20	72	78	-16	104	-32
	21	34	40	-18	63	-29
预测值	22				22	
	23				4	



α	0.80
β	0.10
平均绝对误差%	33%
均方误差	1,790

图 3：用霍尔特双参数指数平滑法来预测

在霍尔特双参数平滑法模型中，预测由两部分构成：一部分是水平截距，是在上期截距的基础上，用简单指数平滑法来更新；一部分是斜率，是在上期斜率的基础上平滑调整，也用简单指数平滑法来更新。两者相加，就得到下期的预测。霍尔特法不但持续调整截距，而且持续调整趋势，在横向和纵向两维调整预测，所以能更好地应对趋势的变化。基本的公式如下，更多细节可以百度“霍尔特双参数指数平滑法”，扫描二维码，或者访问链接 scm-blog.com/case.html，下载该案例的 Excel 文件。



$$\text{本期截距} = \alpha * \text{本期需求实际值} + (1 - \alpha) * (\text{上期截距} + \text{上期斜率})$$

$$\text{本期斜率} = \beta * (\text{本期截距} - \text{上期截距}) + (1 - \beta) * \text{上期斜率}$$

$$\text{下期预测} = \text{本期截距} + \text{本期斜率}$$

因为有趋势，霍尔特法可以预测多期的值：未来第 n 期的预测等于本期截距加上 n 倍的斜率。比如在这个案例中，未来第 1 期的预测 $= 40 + 1 * (-18) = 22$ ，未来第 2 期的预测 $= 40 + 2 * (-18) = 4$ 。你已经发现了，第 3 期的预测就成了负数，这显然不合理。这是趋势参数带来的问题，导致霍尔特法有过度预测的倾向，在使用的时候要加以留意。

霍尔特双参数平滑法中，平滑系数 α 和 β 介于 0 和 1 之间。与简单指数平滑法一样，这两个平滑系数越大，预测模型越响应，也就是说最新发生的对下一步的预测影响更大。

在初始化时，可以假定初始截距为第一个实际值，初始趋势为第二个实际值减去第一个实际值；也可以基于前几个实际值，利用线性回归模型来计算截距和斜率。我也看到有些模型把趋势的初始值设为 0⁷。经过一段时间的初始化后，模型会自动纠偏，与实际值更好地拟合。

在这个案例中，鉴于前几个数据相当离散，我们用前 5 期的平均值作为截距的初始值，斜率的初始值设为 0， α 取值 0.8， β 取值 0.1。对于测试组而言，预测的平均绝对误差降低到 33%，均方误差为 1790。在简单指数平滑法模型中，这两个指标分别为 37% 和 1869。显然，对于这个快消品来说，霍尔特比简单指数平滑法更加有效。

就累计误差来说，霍尔特法的累计误差一直在 0 附近徘徊，明显优于简单指数平滑法（如图 4）。这表明对于这个快消品的需求模式，霍尔特法的自我纠偏能力强于简单指数平滑法，而不是像后者一样一路偏差，在需求一路走低的情况下，一直赶不上需求的变化，导致库存风险大增（误差在这里定义为实际需求-预测，负值表明预测高于实际）。

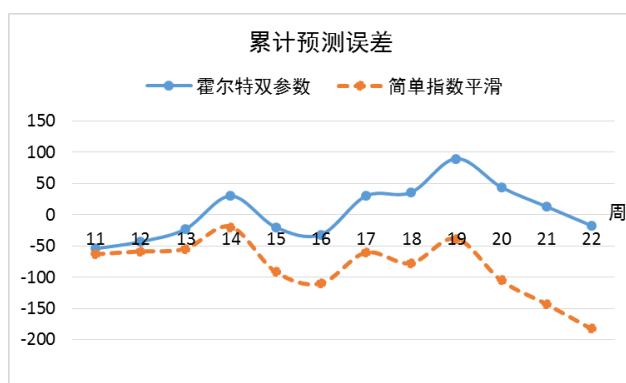


图 4：霍尔特与简单指数平滑法的累计误差对比

对于这个快消品来说，霍尔特法比简单指数平滑法好。但是否该用霍尔特法，我们还得回答一个根本的问题：这种预测方法是否增加价值？这就如你需要一个人帮忙，来了甲乙两个人，甲比乙捣乱地少，这并不是说你会用甲；你需要知道的是，甲是不是比没人帮忙要强。“没人帮忙”就是基准，来判断这人是否增加价值。

在预测上，“没人帮忙”就是“不预测”，也就是亦步亦趋，被动反应，用上期的实际需求作为下期的预测。在英文的预测文献中，有个专用名词叫“幼稚预测”⁸，指的就是这种基准预测。实践中，“幼稚预测”其实挺常用，比如再订货点补货方式用的就是上期用掉多少，下期就补多少（当然还会考虑到经济订货量等因素），是一种地地道道的“幼稚预测”。

⁷ 比如在 Real-Statistics.com 网站，Charles Zaiontz 博士的霍尔特法 Excel 中，斜率的初始值就用 0。这位博士开发了各种各样的 Excel 表格，用来解决数理统计的问题，包括简单指数平滑法和霍尔特法，是学习这些预测模型的很好资料（<http://www.real-statistics.com/time-series-analysis>）。

⁸ 英文为 Naïve Forecast。

就如有的人非但帮不了忙，反而帮倒忙一样，并不是所有的预测方法都增加价值。有时候，“人类一思考，上帝就发笑”，作为还不如不作为，预测还不如不预测。比如在颗粒度很小的地方做预测，或者由销售、用户拍脑袋做预测，预测准确度太低，往往还不如不预测。对这个快消品来说，我们常用的移动平均法也属此列。

因为需求变动相当大，我们就用过去 3 周的平均值作为预测，计算出移动平均法的预测准确度。不难发现，移动平均法在这 4 个准确度指标上，没有一个比“幼稚预测”好（如图 5）。这说明移动平均法不增加价值，还不如不预测，直接用上期的实际需求当做下期的预测。

	周数	需求历史	幼稚预测	移动平均	简单指数平滑	霍尔特双参数
初始化	1	949				
	2	476	949		949	509
	3	478	476		489	480
	4	291	478	634	478	476
	5	353	291	415	296	310
	6	430	353	374	351	330
	7	283	430	358	428	404
	8	242	283	355	287	291
	9	209	242	318	243	232
测试组	10	147	209	245	210	192
	11	154	147	199	149	131
	12	158	154	170	154	126
	13	193	158	153	158	131
	14	120	193	168	192	165
	15	104	120	157	122	110
	16	154	104	139	105	85
	17	135	154	126	153	126
	18	175	135	131	136	120
	19	107	175	155	174	155
	20	72	107	139	109	104
21	34	72	118	73	63	
预测值	22		34	71	35	22
	23					4

平均绝对误差%	36%	53%	37%	33%
平均绝对误差	37	47	37	38
均方误差	1,873	2,879	1,869	1,790
累计误差	(175)	(347)	(180)	48
α			0.97	0.80
β				0.10

图 5：多种预测方法的比较

同理，简单指数平滑法虽然在均方误差上稍优于“幼稚预测”，但别的几个指标上都不如后者，虽然非常接近。这从其平滑参数上也可看出： α 等于 0.97，非常接近 1，而 1 意味着上期的实际值 100% 成为下期的预测，其实就是“幼稚预测”。既然如此，我们也就犯不着花那么多的功夫，来用简单指数平滑法预测这个变动剧烈的快消品；还不如被动反应，把上期的实际需求充当下期的预测得了。

霍尔特法则不。如图 3 所示，在 4 个准确度指标上，霍尔特双参数指数平滑法都比“幼稚预测”好，虽然有些指标的改善看上去微不足道。比如平均绝对误差降低了 3 个百分点，也意味着

预测准确度不过从 64% 上升到 67%。不过不要小看这些微小的改善，对供应链的库存和运营成本来说，它们的影响往往非同小可。

让我们先把预测转化为每周的补货量，看看对供应链的影响。这里假定客户的需求（订单）可以积压，不会因为库存而取消；上次结余的库存也可以用于下期。在这个逻辑下，每周的预测加上积压的需求订单，减去库存，就得到净需求，而净需求就是给生产线或供应商的订单（补货量）。

从图 6 可以看出，霍尔特法下，补货量的波峰、波谷都明显小于“幼稚预测”。这意味着霍尔特法预测的结果波动更小，也意味着波峰时赶工加急更少，波谷时产能浪费更少，降低了供应链的运营成本，提高了产能利用率。要知道，骆驼是被最后一根稻草压死的，而极端的波峰和波谷就是这最后一击，要了生产线、供应商的命。

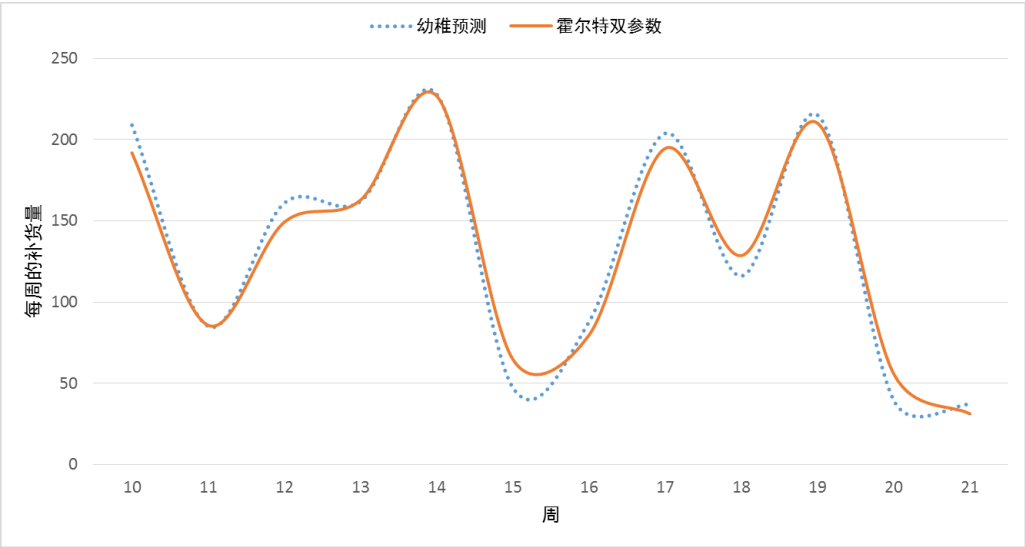


图 6：不同预测方法带来的每周补货量变化

我们进一步计算补货量的变动性，发现霍尔特法比“幼稚预测”的标准差小 8%。对于变动性，供应链的自然应对就是建立安全库存，这意味着霍尔特法下的安全库存要低 8%。我们还模拟每一期末的库存余额（见图 7），发现在第 21 周结束时，霍尔特法的库存为 29 个，比“幼稚预测”的 38 个要低 24%。

当然，计算这些数字，并不是为了得出这样的结论：预测准确度每提高 3 个百分点，安全库存就降低 8%，整体库存下降 24%——这个案例是个例，缺乏那样的通用性。不过全球顶尖的研究机构 Gartner 倒是做过更全面的研究，来评判预测准确度的整体影响（见图 7），相信可以帮助我们更客观地评判霍尔特法来预测这个快消品的优点。

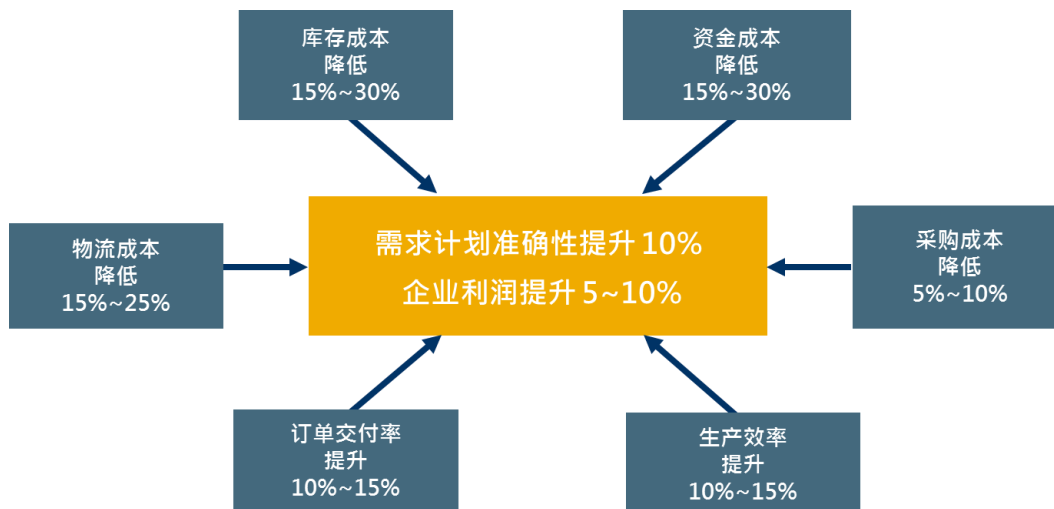


图 7：需求计划对供应链绩效的杠杆作用

出处：Forecast Accuracy and ROI Report 2012, Gartner Group

在累计误差上，霍尔特法看上去也要优于幼稚预测，其累计偏差先是负数，然后纠偏成正值，接下来是上下起伏在 0 附近徘徊。这表明霍尔特法有一定的自我纠偏功能（也叫自适应性，这是趋势平滑系数 β 的功劳）。幼稚预测呢，则完全是随波逐流，没有自我纠偏能力，累计偏差则一直是负数，一路失控，如图 8。

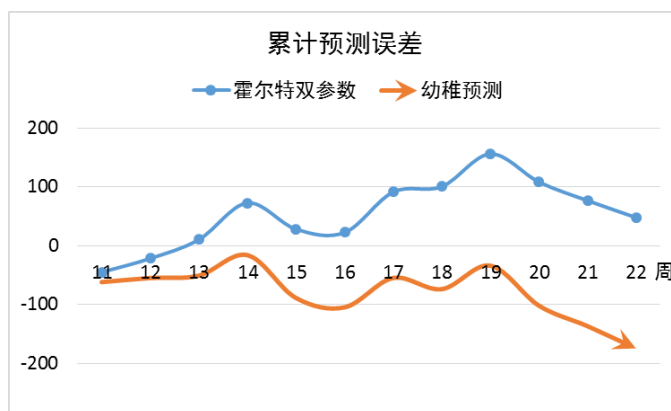


图 8：霍尔特法与幼稚预测的累计误差

看到这里，或许有人会说，霍尔特法这么好，但预测准确度还是只有 67%啊。人们看到准确度，往往联想到的是现货率，或者说库存达成率。对于一个电商来说，67%的现货率当然不行——这会严重影响排名、流量等。这是误解：预测的准确度和现货率相关但是两回事。打个极端的比方，你的需求是 10 个但预测了 100 个，预测准确度当然非常低，但现货率却是 100%——你在牺牲库存来确保现货率。

按照绝对误差来统计预测准确度，其实低估了现货率（库存达成率），因为前者是严格地按照“丁是丁卯是卯”来计算：这期的实际需求是 10 个，不管你预测了 5 个还是 15 个，你的预测

准确度都是 50% ($|实际-预测|/实际*100%$)。但别忘了，如果你预测 15 个的话，你这期的现货率是 100%，还剩下 5 个的库存，可以用来满足后期的需求，提高后期的现货率。

如图 9，我们模拟了第 10 到 21 周的库存余额，来统计每周的现货率。这里假定每周补货，补货量是本周的预测减去上周的库存余额（如果上周的余额为负，表明本周要多补货，以满足上周末满足的需求）。为了简便期间，我们还假定补货是周一就到，可以用来满足本周的需求。看得出，每周的实际现货率都相当高，第 10 到 21 周的综合现货率是 84%，远高于 67% 的预测准确度。这也是为什么你到很多公司去，有些分仓、客户的预测准确度动不动只有百分之二三十，但也没见客户因为现货率太低而跳脚（他们往往是高估需求，或者设立安全库存来提高现货率）。

		霍尔特法：补货、库存余额、现货率						
	周数	需求历史	霍尔特预测	累计需求	补货	累计补货	库存余额	库存现货率
初始化	1	949						
	2	476	509					
	3	478	480					
	4	291	476					
	5	353	310					
	6	430	330					
	7	283	404					
	8	242	291					
	9	209	232					
测试组	10	147	192	147	192	192	45	100%
	11	154	131	301	86	278	-23	85%
	12	158	126	459	149	427	-32	80%
	13	193	131	652	163	590	-62	68%
	14	120	165	772	227	817	45	100%
	15	104	110	876	65	882	6	100%
	16	154	85	1030	80	961	-69	55%
	17	135	126	1165	195	1156	-9	93%
	18	175	120	1340	129	1285	-55	68%
	19	107	155	1447	210	1495	48	100%
	20	72	104	1519	56	1551	32	100%
	21	34	63	1553	31	1582	29	100%
							综合现货率	84%

图 9：模拟现货率（库存达成率）

但严格地讲，预测准确度又是公平的：我们的目标是用多少预测多少，预测多了会积压（库存风险），预测少了会短缺（交付风险），两种风险都不好，都要避免。不过光从预测准确度一个指标来看，是没法得知具体是短缺还是过剩风险。在实际考核中，很多企业把这一指标拆分成两个：按时交货率和库存周转率，用这两个更加直观、跟客户和股东利益直接挂钩的指标来考核。

就这个案例来说，84%的现货率还是不够，库存余额不时出现负值，意味着短缺。怎么办？供应链的第一道防线——需求预测错了，我们就得启动第二道防线——安全库存来弥补。我们知道每期预测的误差，就可以计算其标准差，在这个案例里是 44。假定库存达成率是 95%（意味着需求一到，95%的情况下有现货供应），就得到安全库存 72（假定补货周期是 1 周，详细的公式

见后文的库存计划部分)。看得出，安全库存系统拔高了库存余额的水位，显著降低了断货的风险(图 10)。

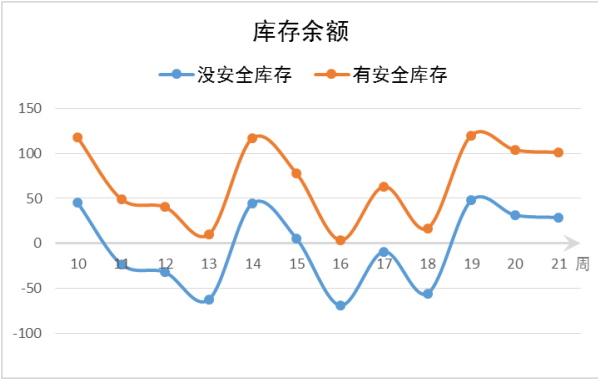


图 10: 设置安全库存，应对预测的准确度不高

当然，有人会问，设置了安全库存，这产品到了下架的时候，岂不是剩下一堆呆滞库存？没错，如果我们不及时调整的话。到了产品生命周期管理末期，要经常评估需求与供应，在合适的时间点切断供应，不再补货，同时把安全库存水位清零，力图消耗掉手头和在途库存。