

案例：选择预测模型就如给小孩买衣服

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，或法 email 给我——刘宝红，136 5127 1450（微信同）| baohong@hotmail.com。

在选择合适的预测方法上，我们一般是根据一定的历史需求，找到拟合度最高的预测模型。但是，业务总是在变，特别是影响需求的手段越来越多的情况下，我们在选择预测模型时，一定要考虑一定的灵活性。这就如小孩的个子总是在长，我们不能严格地按体裁衣，而是要留有一定富余量一样。当然，灵活度会带来准确度的问题，需要我们来平衡。

案例企业主营家居日用品，线下、线上模式共存，线上业务发展迅速，广受 618、双十一、双十二等电商节日的影响。在天猫旗舰店上，过去一年有销售记录的有 1000 多个产品。除了需求暴增的电商大节外，这些产品平日的需求相对稳定，但并不是说没波动，比如门店时不时办活动，做促销。这么多的产品，除了大型的电商节促销活动，平常还得基于数据来预测需求，计划库存。

我们选择了一个典型产品，摘取其在天猫旗舰店半年（26 周）的净销量¹。半年来，这个产品的整体需求没有明显的趋势、季节性和周期性。但是，这并不意味着没有波动。在大大小小的活动、节假日的影响下，需求的波峰波谷时隐时现，忽高忽低，如图 1。那么，我们该用什么方法来预测这个产品呢？本案例分析了不同的预测方法，阐述在选择预测方法时要注意的地方。

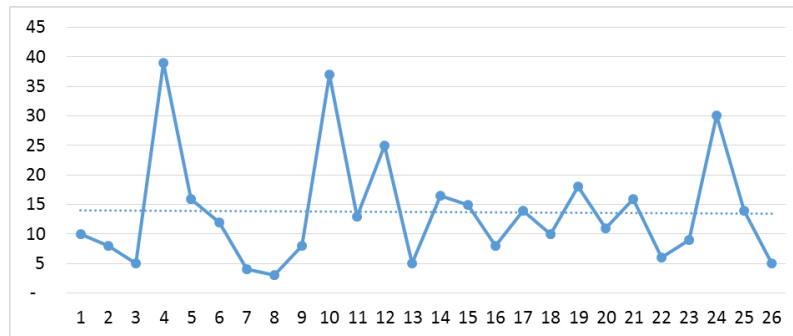


图 1：天猫旗舰店上的典型产品需求

对于需求预测而言，我们有三种最基本，也是应用最广泛的方法，那就是（1）移动平均法，（2）指数平滑法（包括趋势、季节性模型），还有（3）“幼稚预测”，也就是不预测，简单地拿上期的需求当下期的预测。这些方法放到一起，应该能解决日常的绝大多数问题。如果问题超出

¹净销量是销量扣除退货。这里假设退货可以再销。该产品在天猫旗舰店的退货率为 0.4%。

这三种方法，需要更复杂的模型，大多数人没有足够的驾驭能力，倒不如因陋就简，用这三类方法中的一种，再配以职业判断赌一把算了。

对于移动平均法，我们要确定移动时段的长短，也就是说，用过去多少期的数据来预测。这主要取决于两个因素：（1）需求的变动性——需求波动越大，移动时段就越短，移动的期数就越小，这样好尽快响应；（2）随机性——需求的随机性越大，移动时段就越长，移动的期数就越大，这样好尽量消除随机因素。

指数平滑法的考量也类似：波动越大，平滑系数 α 的取值越大，以尽快响应。

别忘了，不预测也是一种选择，那就是单纯地跟着球跑，拿上期的实际值当做下期的预测。实践中，有些预测者有系统的偏见，或者由公司政治驱动，比如销售的预测习惯性地虚高，再比如短缺时每个门店都拔高补货量，这些往往还不如不预测，用“幼稚预测”来代替。“幼稚预测”要尽可能是短期预测。比如门店向分仓、分仓向总仓发出的补货指令，补货周期很短的情况。

对于这个产品，我们分别用上述三种方法来预测，并且调试不同的参数，比较预测的准确度。对于这些预测模型，我们用前 9 周的数据来“预热”，也就是初始化，然后用第 10 周到 26 周的数据来测试预测的准确度，以比较预测模型的优劣。

三类模型的预测数据如图 2，各自的预测准确度汇总在图 3。具体的模型、公式和数据都在 Excel 表格中，可扫描二维码，或者访问链接 scm-blog.com/case.html 来下载。



| 周 | 需求历史 | 幼稚预测 | 移动平均 2 | 移动平均 4 | 移动平均 6 | 移动平均 8 | 指数平滑0.05 | 指数平滑0.10 | 指数平滑0.15 | 指数平滑0.20 |
|----|------|------|--------|--------|--------|--------|----------|----------|----------|----------|
| 1 | 10 | | | | | | | | | |
| 2 | 8 | 10 | | | | | 10 | 10 | 10 | 10 |
| 3 | 5 | 8 | 9 | | | | 10 | 10 | 10 | 10 |
| 4 | 39 | 5 | 7 | | | | 10 | 9 | 9 | 9 |
| 5 | 16 | 39 | 22 | 16 | | | 11 | 12 | 13 | 15 |
| 6 | 12 | 16 | 28 | 17 | | | 11 | 13 | 14 | 15 |
| 7 | 4 | 12 | 14 | 18 | 15 | | 11 | 13 | 14 | 14 |
| 8 | 3 | 4 | 8 | 18 | 14 | | 11 | 12 | 12 | 12 |
| 9 | 8 | 3 | 4 | 9 | 13 | 12 | 11 | 11 | 11 | 10 |
| 10 | 37 | 8 | 6 | 7 | 14 | 12 | 10 | 11 | 10 | 10 |
| 11 | 13 | 37 | 23 | 13 | 13 | 16 | 12 | 13 | 14 | 15 |
| 12 | 25 | 13 | 25 | 15 | 13 | 17 | 12 | 13 | 14 | 15 |
| 13 | 5 | 25 | 19 | 21 | 15 | 15 | 13 | 14 | 16 | 17 |
| 14 | 17 | 5 | 15 | 20 | 15 | 13 | 12 | 13 | 14 | 15 |
| 15 | 15 | 17 | 11 | 15 | 17 | 14 | 12 | 14 | 15 | 15 |
| 16 | 8 | 15 | 16 | 15 | 19 | 15 | 13 | 14 | 15 | 15 |
| 17 | 14 | 8 | 12 | 11 | 14 | 16 | 12 | 13 | 14 | 14 |
| 18 | 10 | 14 | 11 | 13 | 14 | 17 | 12 | 13 | 14 | 14 |
| 19 | 18 | 10 | 12 | 12 | 11 | 13 | 12 | 13 | 13 | 13 |
| 20 | 11 | 18 | 14 | 13 | 14 | 14 | 13 | 14 | 14 | 14 |
| 21 | 16 | 11 | 15 | 13 | 13 | 12 | 12 | 13 | 13 | 13 |
| 22 | 6 | 16 | 14 | 14 | 13 | 14 | 13 | 14 | 14 | 14 |
| 23 | 9 | 6 | 11 | 13 | 13 | 12 | 12 | 13 | 13 | 12 |
| 24 | 30 | 9 | 8 | 11 | 12 | 12 | 12 | 12 | 12 | 12 |
| 25 | 14 | 30 | 20 | 15 | 15 | 14 | 13 | 14 | 15 | 15 |
| 26 | 5 | 14 | 22 | 15 | 14 | 14 | 13 | 14 | 15 | 15 |

图 2：三种不同的预测方法汇总

如图 3，幼稚预测最简单，就是拿上期的需求作为下期的预测。这就如简单的补货机制：用掉多少，就补多少，其实并没有多少“预测”的成分，所以准确度最低，也就不足为奇了。

对于移动平均法，移动时段的长度，也就是用过去多少期的数据，会影响预测模型对变化的响应速度，以及预测的准确度。对于移动期数，我们分别取 2，4，6，8 期，发现期数越大，预测的准确度越高。也就是说，当基于过去 8 期移动平均的时候，预测的准确度最高。

对于指数平滑法，因为该产品的需求历史相对平稳，我们给平滑系数 α 取值 0.05 到 0.2 之间，每相差 0.05，模拟 4 次。我们发现， α 越小，预测的准确度越高，这与整体需求的相对平稳不无关系。

| | 幼稚预测 | 移动平均 2 | 移动平均 4 | 移动平均 6 | 移动平均 8 | 指数平滑0.05 | 指数平滑0.10 | 指数平滑0.15 | 指数平滑0.20 |
|---------|------|--------|--------|--------|--------|----------|----------|----------|----------|
| 平均绝对误差% | 98% | 75% | 65% | 59% | 59% | 51% | 54% | 58% | 60% |
| 平均绝对误差 | 11 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 均方误差 | 190 | 135 | 114 | 87 | 86 | 87 | 87 | 90 | 94 |
| 累计误差 | -3 | 3 | 18 | 15 | 13 | 44 | 27 | 19 | 15 |

图 3：预测准确度的对比分析

与幼稚预测相比，移动平均和指数平滑法的预测结果明显更优，但以期数为 8 的移动平均法模型为最佳（主要是基于均方误差最小，也考虑到尽量最小化累计误差）。

当然，有人或许会说，指数平滑法（ $\alpha=0.05$ ）也看上去不错，其均方误差跟 8 期移动平均法差不多，平均绝对误差百分比甚至比 8 期移动平均要低 8 个百分点。但是，累计误差上，该指数平滑法却显著大于移动平均。在这里，误差是实际值减去预测值，正值表示预测偏低，短缺。对于平滑系数为 0.05 的指数平滑模型，第 10 到 26 期的累计误差是 44，这意味着累计的短缺是 44 个，远高于 8 期移动平均的 13 个。这反映在整体的现货率上，意味着 5 个百分点的差距。我们模拟了第 10 到 26 期的按时交货率²，看得出，8 期移动平均有明显的优势，不但是第 10 到 26 期的综合现货率（图 4），还是每期的具体现货率（图 5），莫不如此。

这是不是说，我们就该选择 8 期移动平均法呢？且慢，让我们再看一下下面的情景。

在第 26 期结束后，让我们再增加 5 期（第 27 到 31 周），这几期的需求波动较大，呈现明显的上升趋势（见图 6）。或许是天猫旗舰店有个小促销（不是大的电商节），或许是有几个团购，或许是某个自媒体大 V 随意点了个赞——我们暂不去了解为什么，但我们知道这种事情会发生，而且经常是供应链不知情的情况下。

² 在模拟的时候，我们假定当期未满足的需求，可以拖到下期；当期的补货量等于当期预测扣除上期库存余额，而且是期初就到，能够满足本期及上期未满足的需求；当期补货量与当期需求的差值就是库存余额。

| 周 | 需求历史 | 移动平均 8 | | | | | | 指数平滑0.05 | | | | | | | |
|-------|------|--------|----|------|------|------|------|----------|----|------|------|------|------|--|-----|
| | | 预测 | 补货 | 累计需求 | 累计补货 | 库存余额 | 现货率 | 预测 | 补货 | 累计需求 | 累计补货 | 库存余额 | 现货率 | | |
| 10 | 37 | 12 | 12 | 37 | 12 | -25 | 32% | 10 | 10 | 37 | 10 | -27 | 28% | | |
| 11 | 13 | 16 | 41 | 50 | 53 | 3 | 100% | 12 | 38 | 50 | 49 | -1 | 91% | | |
| 12 | 25 | 17 | 14 | 75 | 67 | -9 | 66% | 12 | 13 | 75 | 62 | -13 | 48% | | |
| 13 | 5 | 15 | 23 | 80 | 90 | 10 | 100% | 13 | 26 | 80 | 88 | 8 | 100% | | |
| 14 | 17 | 13 | 4 | 97 | 93 | -3 | 81% | 12 | 5 | 97 | 92 | -4 | 74% | | |
| 15 | 15 | 14 | 17 | 112 | 110 | -1 | 93% | 12 | 17 | 112 | 109 | -3 | 83% | | |
| 16 | 8 | 15 | 16 | 120 | 127 | 7 | 100% | 13 | 15 | 120 | 124 | 5 | 100% | | |
| 17 | 14 | 16 | 9 | 134 | 135 | 2 | 100% | 12 | 8 | 134 | 132 | -2 | 88% | | |
| 18 | 10 | 17 | 15 | 144 | 150 | 7 | 100% | 12 | 14 | 144 | 146 | 2 | 100% | | |
| 19 | 18 | 13 | 7 | 162 | 157 | -5 | 74% | 12 | 10 | 162 | 156 | -6 | 68% | | |
| 20 | 11 | 14 | 19 | 173 | 175 | 3 | 100% | 13 | 18 | 173 | 174 | 2 | 100% | | |
| 21 | 16 | 12 | 9 | 189 | 185 | -4 | 76% | 12 | 11 | 189 | 185 | -4 | 78% | | |
| 22 | 6 | 14 | 17 | 195 | 202 | 8 | 100% | 13 | 16 | 195 | 201 | 7 | 100% | | |
| 23 | 9 | 12 | 5 | 204 | 207 | 3 | 100% | 12 | 6 | 204 | 207 | 3 | 100% | | |
| 24 | 30 | 12 | 8 | 234 | 215 | -19 | 38% | 12 | 9 | 234 | 216 | -18 | 40% | | |
| 25 | 14 | 14 | 33 | 248 | 248 | 0 | 100% | 13 | 31 | 248 | 247 | -1 | 93% | | |
| 26 | 5 | 14 | 14 | 253 | 262 | 9 | 100% | 13 | 14 | 253 | 261 | 8 | 100% | | |
| 综合现货率 | | | | | | | 74% | 综合现货率 | | | | | | | 69% |

图 4：两种预测方法的现货率模拟

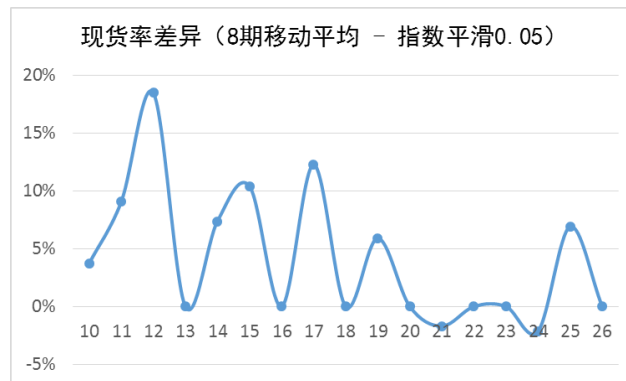


图 5：两种预测方法的现货率差异（8 期平均减去指数平滑法）

让我们针对这 5 期的预测数据进一步比较，发现“最佳”的预测模型大相径庭了：如图 7，在移动平均法领域，原来最好的 8 期移动没法有效应对变动，预测的准确度变成最差；而原来准确度最差的 2 期移动平均，却因为能更快地响应，变成最好的模型。指数平滑法也类似：平滑系数为 0.2 的模型因为响应快，效果反倒明显好过原来平滑系数为 0.05 的模型。

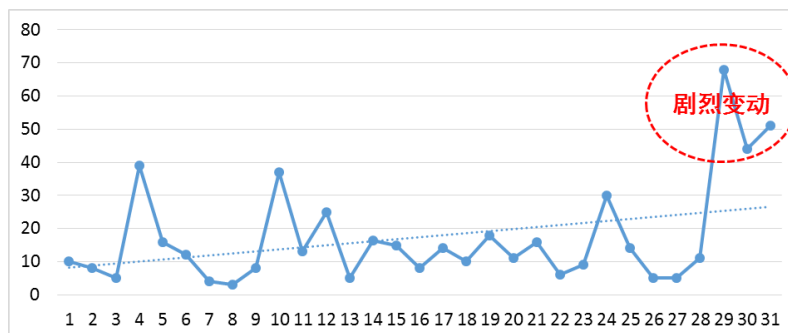


图 6：拐点来临，需求突变

还让人惊奇的是，原来准确度最低的“幼稚预测”，现在的准确度却仅次于 2 期移动平均，在有些准确度指标上，甚至是最好。这并不是说“幼稚预测”本身有多好，也不是推荐大家用“幼稚预测”，而是因为拐点来临，需求模式突变，移动平均和指数平滑法都难以有效应对，反倒不如不预测，“跟着感觉走”，采取简单的用几个，补几个的策略。

第10到第26周：需求相对平稳

| | 幼稚预测 | 移动平均 2 | 移动平均 4 | 移动平均 6 | 移动平均 8 | 指数平滑0.05 | 指数平滑0.10 | 指数平滑0.15 | 指数平滑0.20 |
|---------|------|--------|--------|--------|--------|----------|----------|----------|----------|
| 平均绝对误差% | 98% | 75% | 65% | 59% | 59% | 51% | 54% | 58% | 60% |
| 平均绝对误差 | 11 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 均方误差 | 190 | 135 | 114 | 87 | 86 | 87 | 87 | 90 | 94 |
| 累计误差 | -3 | 3 | 18 | 15 | 13 | 44 | 27 | 19 | 15 |

第27到第31周：需求模式发生拐点变化

| | 幼稚预测 | 移动平均 2 | 移动平均 4 | 移动平均 6 | 移动平均 8 | 指数平滑0.05 | 指数平滑0.10 | 指数平滑0.15 | 指数平滑0.20 |
|---------|------|--------|--------|--------|--------|----------|----------|----------|----------|
| 平均绝对误差% | 41% | 51% | 77% | 71% | 75% | 76% | 76% | 73% | 69% |
| 平均绝对误差 | 19 | 16 | 22 | 23 | 24 | 26 | 24 | 23 | 22 |
| 均方误差 | 782 | 740 | 888 | 869 | 926 | 1041 | 959 | 905 | 862 |
| 累计误差 | 35 | 54 | 85 | 84 | 89 | 116 | 98 | 89 | 84 |

图 7：不同模型的预测准确度

要知道，预测模型都是基于一定的历史数据，而历史不可能完全重复，比如拐点会来临，需求模式会突变，尤其是线上、线下环境复杂，各种社交媒体让需求变化更剧烈的情况下。拐点刚来的时候，我们一般没法及时探知，还不得不依赖原来的预测模型，而这正是真正考验预测模型的时候。这要求预测模型有一定的灵活性。

但是，拐点来临时的灵活性呢，往往会降低平常的预测准确度。在选择合适的预测模型时，我们要权衡利弊，两者兼顾。这也意味着，我们往往不能严格按照模型的拟合度，选择拟合度最高的那一个。这就相当于给小孩子买衣服，你不会严格地按体裁衣，而是留一点富余量一样。前面讲的“过度拟合”，其实就是严格地按体裁衣，预测效果反倒不好。不过留有余量的代价呢，就是现在穿着并不是 100%的合体。当然，你也不会给买一件很大的衣服，让孩子一直穿得松松垮垮，最后终于合体了，衣服却也穿烂了——这些都需要我们的职业判断。

就拿本案例来说，8 期移动平均法看上去是**当前最合身的**：基于 26 期的历史，该方法的均方误差最小。2 期移动平均法是另一个极端：在需求模式改变的时候，2 期移动平均法最能快速响应，但平时会导入太多的波动和“噪音”。这两个极端往往都不是最佳方案。而介于中间的方案呢，比如 4 期移动平均法，往往能更好地兼顾两个极端，因而是更好的解决方案。

我们进一步分析第 27 到 42 周的数据。如图 8，进入后半年后，天气变冷，对该产品的需求显著增加，而需求波动也更大了。需求波动大，客观上要求预测模型更加灵敏。我们再次测试上述各个模型，发现 4 期移动平均法的准确度最高（图 9）。我们也试图优化指数平滑法，发现平滑指数为 0.4 时，该方法最优，但还是没有 4 期移动平均好。这是因为这段时间需求呈现一定的上升趋势，简单指数平滑法没法有效应对（或许霍尔特双参数法会更好，感兴趣的读者可以试试）。

这就是说，在前半年需求较平稳时，8期移动平均最佳；进入后半年，需求波动更大时，4期移动平均最优。理论上，我们可以按照季节的变化来更换预测模型，但往往可操作性较差：很多人没有足够的专业知识来优化、选择预测模型，而是习惯于从头到尾使用一种方法。既然如此，两害权益取其轻，用4期移动平均法来预测所有的时段，也是一种可取的折中方案。

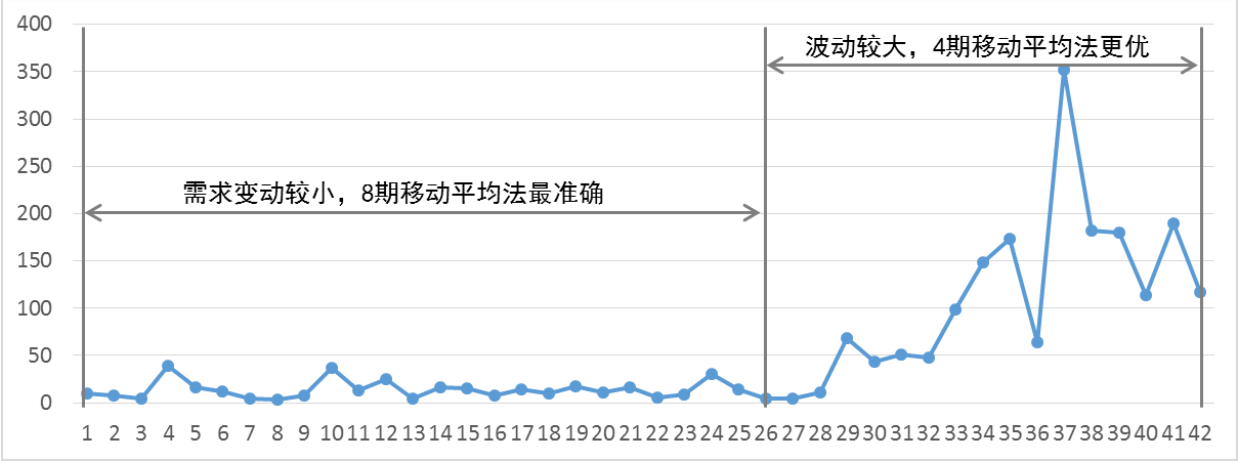


图 8：进入后半年，需求更高，波动也更大

之所以选择4期移动平均，而不是2期移动平均法，还有几个原因：其一，前者的均方误差最小。第37周的需求看上去是个极端值，应该是9月份有什么大型的促销活动（这也反映在第36周的需求偏低上：大型活动前，消费者往往采取观望态度，所以活动前的需求会人为偏低）。如果“清洗”这两周的数据，比如用前4周的平均值来“削峰填谷”，这段时间的需求波动就会更小，让4期移动平均的优势更加明显（相对于2期移动平均法）。其二，2期移动平均法有点太“灵敏”，特别是对于前半年来说，这会给供应链带来过多的波动，以及相应的运营成本。其三，4期平均比较好沟通，也符合很多公司的操作习惯，比如“我们用上月4周的平均作为下月的预测”——这是个次要的原因。

第27到42周：需求模式改变，需求波动更大

| | 幼稚预测 | 移动平均 2 | 移动平均 4 | 移动平均 6 | 移动平均 8 | 指数平滑0.05 | 指数平滑0.10 | 指数平滑0.15 | 指数平滑0.20 |
|---------|------|--------|--------|--------|--------|----------|----------|----------|----------|
| 平均绝对误差% | 51% | 48% | 52% | 52% | 55% | 69% | 56% | 53% | 51% |
| 平均绝对误差 | 63 | 53 | 49 | 54 | 58 | 81 | 65 | 59 | 56 |
| 均方误差 | 9296 | 5862 | 5456 | 6422 | 7222 | 12301 | 9350 | 7773 | 6893 |
| 累计误差 | 112 | 200 | 329 | 515 | 683 | 1286 | 1016 | 818 | 669 |

图 9：需求波动更大时，4期移动平均成为最优模型

到现在为止，我们把该产品从第1周复盘到第42周。一年有52周，那后面的43到52周呢？熟悉电商业务的人知道，这是每年第四季度，“双十一”“双十二”肆虐，需求极端波动，也是需求预测最为挑战的时候。就拿这个产品来说，全年60%以上的营收来自这两个电商节，任何基于需求历史的预测模型都没法有效应对，需要跟市场、销售等前端职能对接，更多借助他们的判断来预测，这里不予赘述。

【小贴士】选择预测模型时，要考虑什么因素？

预测模型的选择是个复杂的过程，需要考虑多方面的因素，再配以职业判断。计划软件往往按照特定的指标判断预测模型的优劣，很难综合考虑多种因素，特别是历史数据没有反映的信息。这就是对计划软件建议的模型，有经验的计划员总是充满戒心的原因了。

在选择合适的预测模型时，我们首先要看**预测准确度**。

预测准确度有多种统计方法，这里介绍常用的**3类**。

第一类是**绝对误差**。这是实际减去预测，取绝对值，除以实际值，就得到百分比。绝对误差指标当然是越小越好，但当实际值太小的时候，百分比可能会非常高而误导，我们的解决方案就是限定最高不超过**100%**。

第二类是**均方误差**。这是实际减去预测，再平方。平方的用意是彰显那些极端值，比如预测太高或太低。要知道，一般的误差我们可以通过安全库存、适当赶工来应对；害死我们的大都是极端案例，要尽量避免。

第三类是**累计误差**。这主要是判断预测方法是否有系统性偏差。一个理想的预测模型应该是有时偏高，有时偏低，但两相抵消，累计误差会较小，甚至接近零。累计误差能够帮助判断总需求与总供应是否匹配：累计误差接近**0**，表明整个补货渠道里总量匹配，这样，如果时间不匹配的话，我们赶工加急，至少也有足够的产品在供应渠道可供加急。

这些准确度指标的缺点是不够直观，跟业务结果有一定的距离，往往并不能帮助我们做出最后的决策。有时候，两种模型的预测准确度非常接近，但“翻译”成业务结果时，差距却很明显，就如图4所示的例子。

这时候，预测模型的选择就要考虑业务结果：**现货率、库存和运营成本**。

现货率和库存比较直观，我们可以通过复盘模拟来评估，上面图4就是这样的例子。

运营成本则否：预测对运营成本的影响如何，我们很难直接量化。但是，我们可以用预测的平滑度来评估：预测不平滑，表现为起伏较大，对供应链运营的影响会大，相应的运营成本也高。而平滑度呢，我们可以标准差来量化：预测的标准差越大，表明平滑效果越差，反之亦然。

比如对于上面的案例，我们计算第**10到26期**预测的标准差，借以判断预测的平滑性。如图10，幼稚预测没有任何平滑效应，预测值的变动最大，表现为标准差最大；移动平均和指数平滑法都有平滑效果，预测值的标准差也明显更小。更具体地看，移动平均法的移动期数越高，指数平滑法的平滑系数越小，平滑效应越明显，都是意料中的事。

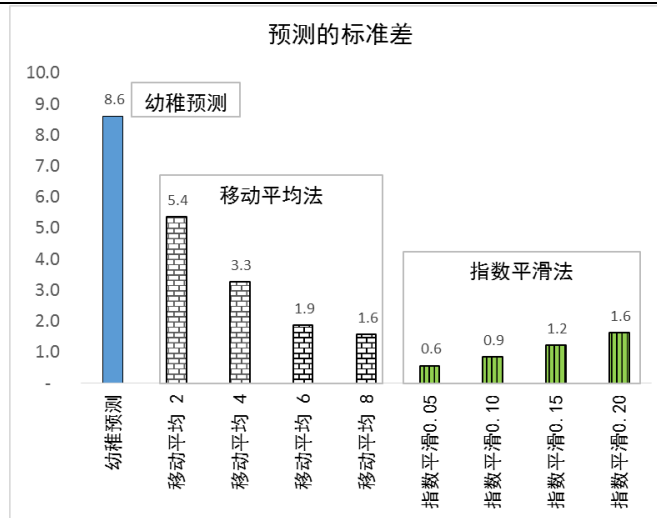


图 10: 预测的平滑度会影响到供应链的运营成本

最后，预测模型的选择还要考虑实施的难易度。人们难以理解的模型，就很难推广；公式复杂，实施难度高的模型也是。就整体预测效果而言，简单的模型往往比复杂的更理想，很大程度上是因为简单的模型更容易落地。这就相当于给你个 100%功能的武器，你能操作 30%；还是 70%功能的武器，你能操作 50%一样。

真正的计划高手，高就高在用最基本的模型，比如移动平均和指数平滑法，解决实际中的大多数问题。这就如真正的武林高手不会花拳绣腿，拖泥带水一样。所以，那些动不动就把大数据、人工智能、机器学习等挂在嘴上的人，八成是在以主题博大，以掩盖基本方法论上的贫瘠。