



夏春：确诊预测模型与投资 组合模型有何相似之处？



文/意见领袖专栏作者 夏春



近期，媒体持续报道香港各大学医学院对未来确诊人数的预测，引起了广泛的关注。3月14日港大医学院院长根据模型估计，在2021年12月31日开启的第五波Omicron变种病毒冲击下，香港实际感染人数为358万。同日，中大医学院一位教授估计第五波最终约有500万人感染。

到3月22日，香港政府公布第五波疫情确诊病例106万人，港大医学院院长更新估计，认为感染人数已超过440万，又预测第六波疫情将在6月初前两周出现，持续约两个月，额外感染220万人。

作为对比，大家应该注意到内地医学专家很少发布类似的预测。当然这并不代表内地没有这样的预测，差距可能来自媒体的报道偏好，也可能香港政府期待更多市民接种疫苗，鼓励医学专家发布他们的预测。

确诊人数估计引发的争议

专家预测的数据高于政府数据倒是可以理解，毕竟在目前医疗资源挤兑下，许多自我抗原检测为阳性的个人并未在政府公共卫生系统上报。

值得一提的是，医学院在公布预测的点估计数据之外，还包含了 95% 置信区间的估计，对应 358 万人点估计的是 230 万至 460 万人的区间估计。但媒体通常只报道点估计，可能是担心读者不理解置信区间的含义。

大家自然注意到不仅区间估计的下限要明显高于已确诊病例，而且区间估计的宽度居然高达 230 万（460 万-230 万），怎么看都会觉得这样的预测不够精确。

而且，从 3 月 14 日到 22 日才过去 8 天，点预测的变化就高达 82 万，相当于政府确诊数据的 80%，怎么看都会产生一种预测不太靠谱的印象。

于是，有趣的一幕发生了，港大统计系系主任忍不住撰文批评港大医学院的估计，指模型估计过于夸张而且误差大，可能造成市民不必要的恐慌，增加医疗系统的负担，结果事与愿违。

根据医学院的区间估计数据，系主任甚至反向推导出完全随机分布的有效样本只有大约 40 人，如此小的样本量，可信度自然不高。

同时，根据目前香港病毒感染最高危的楼宇（每栋病例均超过 40 人）检测数据来看，平均初步阳性比率约为 11.6%，远远低于港大医学院模型推算的 48% 感染率（也就是用 358 万点估计除以香港总人口 741 万）。

医学院的区间估计避免了“过度自信”

在心理学和行为金融学里有一个重要的概念叫做“过度自信”，指的就是大多数人在做预测时，给出的区间估计过于狭窄。这本质上来源于对预测精度的过度自信，用统计术语表达就是，估计的“标准差”太小。

举例来说，各大券商在年底发布来年上证指数的估计，3500-4200 的区间预测要比 3650-3950 区间的缺乏信心。各种研究都显示，“过度自信”是投资的主要杀手之一，频繁买卖就是最典型的体现。

因此，在我看来，医学院的区间估计这么宽，恰恰是医学院对预测模型不太有信心的一种侧面体现，从理性角度来看，其实不算是坏事。

为什么疫情预测模型都很难精确？

据我所知，港大医学院并未公布预测模型。但如果回溯过去，可以看到 2020 年疫情爆发以来全球各国医学专家基于不同模型给出的预测结果，都不是太靠谱。例如港大医学院在 2020 年就曾经预测香港会有四分之一的人口感染 Covid-19（传染性低于 Omicron 变种）。

归根到底，无论如何建模，疫情预测都离不开一个关键参数 R_0 ，也就是基本繁殖数，代表一个确诊病人可以把病毒平均传给多少人。如果 R_0

等于 3，那么一个病人传播 10 轮后，累计确诊就会达到 29524 人。

但是，指数传播有一个特点就是结果对参数变化非常敏感。如果 R_0 分别等于 2.9 和 3.1，那么传播 10 轮对应的累计确诊就分别是 22142 和 39030 人。初始参数高出约 7%，结果就会多出 76%。而且随着传播轮数的增加，结果的差异性会放得更大。

在这个完全不考虑防疫措施的最简单模型下，指数传播很快就会超过一个地区的总人数，而所谓“全民免疫”的思想就是建立在 R_0 会随着已感染人数的增加和未感染人数的减少而逐步减少到小于 1。

因此，改进后的模型应该考虑到在单（多）次疫情冲击下，累计确诊随时间变化的分布应该接近一个单波浪 S（多波浪）的形状。医学专家也做了这样的尝试，把 R_0 之外的疫情影响因素也考虑进来，比如防疫措施的力度，方式，个人行为变量，社会经济条件包括医疗，住宿，交通等。

但是，变量越多，模型越复杂，结果的“蝴蝶效应”也就更加明显，失之毫厘，差之千里。仅就 Omicron 病毒的 R_0 估值来看，虽然一些研究认为在 3 左右，但是在 2 和 16 之间的估计值也同样存在。这样，即使基于一个地区的经验构造出有效模型，换一个地区就可能完全失灵。

也因此，一些严肃的专家认为建立这样的模型价值不高，而且常常会引起适得其反的效果。例如预测误差太大的研究可能导致市民对“专家”、“精英”的反感，不利于科学知识的传播。

“蝴蝶效应”同样存在于证券投资组合模型

确诊人数预测模型和证券投资组合模型是风马牛不相及的，但他们之间却有一个至关重要但鲜为人知的相似点，那就是对参数的高度敏感性。我相信即使是金融从业人员，也很少知道投资组合理论会和“蝴蝶效应”联系在一起。

最简单的投资组合模型会针对可选择证券的期望收益率，以及他们之间的反差-协方差矩阵，来构造一个在给定风险下，投资者预期收益率最大，或者在给定收益率下，风险最小的最优组合权重。从数学角度来说，求解这样一组权重非常简单。

大家可能会认为权重的敏感性来源于未来的期望收益率难以精准衡量，的确收益率变化大，很不稳定，但其实是相对稳定的证券反差-协方差矩阵的估值对模型结果的影响更大。

归根结底在于，求解过程中要对多个证券的反差-协方差矩阵求一个逆矩阵。这个过程虽然不涉及到任何类似 R0 指数递增带来的结果变化，但矩

阵的细微变化就会让结果大不相同，甚至导致完全相反的结果。

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_39722

