

2022.6.11

进一步实验

（固定效应）**分位数回归组合**：为进一步验证之前实验的结论，在11/10模型（0.01与0.025分位数去掉lgb）的模型池上针对固定效应正则参数 λ 进行讨论，预测的实验结果如下：

	0.01	0.025	0.165	0.25	0.5	0.75	0.835	0.975	0.99
1	0.014451	0.038885	0.221515	0.316337	0.521207	0.551110	0.484800	0.192568	0.076493
10	0.014290	0.035515	0.215269	0.309268	0.503979	0.503253	0.423758	0.134453	0.069830
50	0.014288	0.035551	0.215033	0.306400	0.488956	0.489211	0.420438	0.134042	0.069532
10000	0.014288	0.035531	0.215041	0.306432	0.488936	0.489433	0.420405	0.134032	0.069436
10000-无截距	0.014288	0.035531	0.214937	0.306395	0.488947	0.489182	0.420276	0.134838	0.070271
加权平均	0.0155	0.0376	0.221	0.3145	0.5035	0.5003	0.4269	0.1342	0.0696
简单平均	0.0171	0.0389	0.2208	0.3147	0.5046	0.501	0.427	0.1337	0.0694
相对加权平均提升	7.82%	5.50%	2.70%	2.57%	2.89%	2.17%	1.52%	0.12%	0.24%
相对简单平均提升	16.45%	8.66%	2.61%	2.63%	3.10%	2.31%	1.54%	-0.25%	-0.05%

由上述结果知：

- 回归组合比平均/加权有一定提升；
- 截距对于回归组合是必需的；（如果实际求得截距较大，则截距必需；否则有截距的组合未必胜过无截距组合，但是前者显然更稳健）
- 在固定效应模型与Pool模型的选择中，更倾向于Pool模型。——问题：为什么固定效应不需要？可能原因：过度控制

升级固定效应控制等级：将M5数据有的四个类别（dept、cat、store、state）作为固定效应，结果对比如下：结果显示，只要对固定效应施加一定的惩罚项，不同固定效应模型组合结果以Pinball loss计的差距并不大（在四位小数之后才有差距，差距比例在千分之一量级）。所以，对于升级后的固定效应，比sku级别的效应要有用一些，但是有无固定效应的区别依然不大。

λ	0.01	0.025	0.165	0.25	0.5	0.75	0.835	0.975	0.99
0	0.014413	0.036246	0.221798	0.310055	0.495514	0.496054	0.434980	0.148771	0.098958
1	0.014288	0.035531	0.215191	0.306399	0.488959	0.489445	0.420459	0.134016	0.069345
10	0.014288	0.035531	0.215056	0.306399	0.489018	0.489381	0.420443	0.134012	0.069427
50	0.014288	0.036373	0.215092	0.306399	0.488958	0.489671	0.420392	0.134013	0.069453
10000	0.014288	0.035531	0.215113	0.306404	0.489008	0.489277	0.420428	0.134042	0.069427

目前工作总结与进一步想法

本学期截止目前，围绕着间断数据的分位数组合，做了如下工作：

- 完善了组合模型池，包括Bootstrap、分位数回归、基于假设分布、机器学习方法等模型。
- 尝试了基于（固定效应）分位数回归的组合方式。
- 初步尝试了基于LSTM进行基于特征的（时变权重）预测组合。

工作评价

- **组合模型池**：组合工作的基础，包含了使用信息与方法具有差异性的模型。不过可能有几个问题：
(1) 关于Bootstrap, WSS与VZ已然不是较新的方法，是否要进行一些更新？(2) 机器学习方法在模型池中仅是较好而非绝对优于其它方法，鉴于M5的比赛结果，机器学习模型在池里是否有说服力？
- **(固定效应) 分位数回归组合**：有理论支持回归组合的优势，并且实验结果好于简单平均/倒数损失加权平均/（以及目前实验设置中的元学习组合）；相对过往文献，试图针对**数据集加入固定效应**，结合Koenker (2004) 的分位数固定效应模型进行研究，但在此框架下固定效应没有显著改进预测组合。另外一个问题是，回归组合是一个相当老的方法了。（能否解释引入固定效应效果不佳的原因？）
- **基于特征（元学习）的时变权重预测组合**：在上次汇报的基础上，又尝试了使用真实数据的滞后特征（而非在测试集上使用基于基准预测分布的抽样），结果没有明显区别，权重仍然会收敛而非时变，不再展示。不对其进行进一步实验，因其设计有问题：
 - 如果想研究时变权重，那么关注的时间段不能过短，否则没有研究意义；对于一个长度4-5年的日度数据，28天参考/测试可能不够长。
 - 更重要的是，时变权重的意义在于面对不同预测环境（往往是外生冲击所致）时选择合适模型，28天涉及的事件可能不够完善丰富。对于销售数据来说，重大节假日/重大活动与普通日期可能会导致这些区别。（对比 Negro et al.(2016) 的动态预测池研究，其训练数据是1964-1991季度GDP与通胀数据，测试数据是1992-2010，并且重点考察了2008金融危机与其它时间的差别）
 - 如果做时变权重的研究，基准模型的运算花费不宜太高（因需要时间长）。
 - 即使研究时变并试图引入特征，直接上LSTM是否过急？因为其是一个黑箱；并且根据Negro et al.(2016)的讨论，直接优化时变参数可能倾向于角解，不能完全体现组合的意义。（事实上，他们的工作是基于一个参数化的时变过程）

进一步想法

应该进一步围绕**间断需求**的特点展开：目前的研究没有体现出间断需求的特性，目前看到/想到如下扩展点：

- **整数输出**：在现有文献中，有一个术语 **coherent forecasting** (Freeland & McCabe, 2004) 专门描述：计数过程的预测计算结果也应该是计数数值（实际意思是预测与原始数据支撑相同）。目前读了Freeland & McCabe (2004, IJF) 以及Homburg et al. (2020, Journal of Forecasting) 的文献，发现其coherent forecasting 都是围绕着整数自回归 (INAR) 模型，即：

$$X_t = \alpha \circ X_{t-1} + \epsilon_t$$

X_0 是均值 λ_0 的泊松分布； ϵ_t 是独立、均值为 λ 的泊松分布； $\alpha \circ X_{t-1}$ 代表二项分布 $B(X_{t-1}, \alpha)$ ，基于 X_{t-1} 的条件分布。这或许是一个基准模型的设定，但不是组合方法。需要进一步考虑；目的是得到符合实际取值的尽可能精确预测。

- **概率预测及评价方式**：
 - 目前的结果报告中，优化目标是Pinball loss，而方法对比的评价指标也是 Pinball loss，评价有失稳健，结果可能过拟合；
 - Kolassa (2016) 的 *Evaluating predictive count data distributions in retail sales forecasting* 一文 (IJF) 对大多数常见点预测指标提出了批评，并给出了衡量间断预测（特别是概率预测）的一些指标推荐（实际上还是calibration与sharpness），之后对一些简单的概率预测进行了对比。其建议是预测间断预测的分布（甚至不只是几个分位数）。
 - 虽然每次汇报都只展示9个分位数的对比指标，事实上每个方法都生成了0.01-0.99的分位数预测。其目的是，试图用99个分位数来拼出整个分布列（借助离散整数性质，以及对两端的信念）。使用分位数组合而非密度组合最主要的考虑是其可以方便的结合不同预测方法/模型。而在回归框架下，进行组合得到99个组合分位数结果也不是太复杂的计算。

- 结合库存改进评价/预测中就考虑库存：结合安全库存，右侧分位数的重要性（这一点也与评价方式有关）。

Thanos E. Goltsos , Aris A. Syntetos , Christoph H. Glock , George Ioannou四人在EJOR上的综述 ***Inventory – forecasting: Mind the gap*** (2021) 强调了把预测与库存集成起来的重要性与期望。（当然，结合的深度，以及相关物流知识的理解，是一个问题）