2021.11.24

方法指标

Calibration的检验

对预测区间可靠性的检验，有无条件覆盖（Kupiec test）与独立条件覆盖（ Christoffersen test）两种。（PIT由于分布函数不连续，不好表示出分布函数逆函数，可能不允许做；同时这个操作只能观察个案）

Kupiec test：认为一定比例（如90%）预测区间应该覆盖一定比例（90%）的真实数据，不考虑序列相关。统计真值击中区间的个数，之后构造统计量，服从自由度为1卡方分布（c 是名义比例，π 是实际比例，n_0 与 n_1 是0、1个数）：

$$I_t = \begin{cases} 1 & \text{if } P_t \in [\widehat{L}_t, \widehat{U}_t] \rightarrow \text{'hit'}, \\ 0 & \text{if } P_t \notin [\widehat{L}_t, \widehat{U}_t] \rightarrow \text{' miss' (or 'violation')}. \end{cases}$$

$$LR_{UC} = -2\log\left\{\frac{(1-c)^{n_0}c^{n_1}}{(1-\pi)^{n_0}\pi^{n_1}}\right\}$$

Christoffersen test：考虑上一个区间是否覆盖真值对下一个时间的影响，也是服从自由度为1卡方分布。

$$LR_{Ind} = -2\log\left\{\frac{(1-\pi_2)^{n_{00}+n_{10}}\pi_2^{n_{01}+n_{11}}}{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}\right\}$$

Sharpness的评价

Pinball loss：$$Pinball(\widehat{Q}_{P_t}(q), P_t, q) = \begin{cases} (1-q)(\widehat{Q}_{P_t}(q) - P_t), & \text{for } P_t < \widehat{Q}_{P_t}(q), \\ q(P_t - \widehat{Q}_{P_t}(q)), & \text{for } P_t \geq \widehat{Q}_{P_t}(q), \end{cases}$$

在M5中使用了SPL作为评价指标，类似MASE的方式，考虑了数据规模对预测的影响。

$$\mathbf{SPL}(\mathbf{u}) = \frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t - Q_t(u))u\mathbf{1}\{Q_t(u) \leq Y_t\} + (Q_t(u) - Y_t)(1-u)\mathbf{1}\{Q_t(u) > Y_t\}}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|}$$

CRPS（Continuous Ranked Probability Score）：

$$CRPS(\widehat{F}_{P_t}, P_t) = \int_{-\infty}^{\infty}(\widehat{F}_{P_t}(x) - \mathbf{1}_{\{P_t \leq x\}})^2 dx,$$

DRPS是CRPS离散求和版本。这是对分布整体的评价。

评价结果

预测方法现有如下三种：

WSS：历史销量与是否有销量的概率模拟，重复1000次求分位数

VZ：历史销量与正销量间隔时间的模拟，重复1000次求分位数

quantGAM_count: 之前提到的GAM与quantile回归结合，应用在计数intermittent demand的结果；但是由于协变量不太好（取值种类少，矩阵共线性强），有些估计很离谱（销量估计100000+。。。），故加一个修正：估计分位数大于历史最大值，设为最大值+1。

评价的区间为99%双侧区间、95%双侧区间、67%双侧区间、50%双侧区间。涉及的分位数为0.995与0.005、0.975与0.025、0.835与0.165、0.75与0.25，以及0.5分位数。对最后28期预测的结果进行评价。

Kupiec test检验67、50预测区间拟合不好：如果预测区间包括0，则数据覆盖面更广，67/50区间可能覆盖了实际更多的数据，这是数据本身特点导致的。

考虑到数据的序列相关性，使用Christoffersen test检验，99、95、50区间不拒绝率都较高，67区间不拒绝率有所升高，但仍不佳。

2022/3/31更新：以上结果应该做单侧区间检验而非双侧，因为 intermittent demand 本身就是偏态分布。

Kupiec test不拒绝率	**99PI**	**95PI**	**67PI**	**50PI**
WSS	98.78%	99.50%	29.17%	23.59%
VZ	97.34%	99.47%	23.50%	17.74%
quantGAM_count	87.39%	96.93%	30.63%	22.09%

Christoffersen test不拒绝率	**99PI**	**95PI**	**67PI**	**50PI**
WSS	99.96%	99.83%	47.00%	95.12%
VZ	99.95%	99.81%	53.88%	95.78%
quantGAM_count	99.68%	99.50%	53.15%	95.59%

VZ在每个分位数上表现都最好，quantGAM表现略差。中间的pinball loss偏高，两侧较低。左侧比右侧loss要低。（左侧0多）

Pinball loss	**0.995**	**0.975**	**0.835**	**0.75**	**0.5**	**0.25**	**0.165**	**0.025**	**0.005**
WSS	0.0438	0.1497	0.4692	0.5527	0.5559	0.3189	0.2271	0.0357	0.0072
VZ	0.0445	0.1420	0.4512	0.5281	0.5306	0.2817	0.2266	0.0356	0.0072
quantGAM_count	0.0733	0.1668	0.4728	0.5486	0.5439	0.2910	0.2312	0.0364	0.0075

就分布总体的评价而言，VZ最好，quantGAM次之。

	DPRS
WSS	0.9497
VZ	0.8866
quantGAM_count	0.9384