

2020.12.2

参数模型概率预测尝试

来自于 *Forecasting the intermittent demand for slow-moving inventories: A modelling approach* (Snyder et al., 2012) 的方法，其尝试需求服从以下分布：

- 泊松分布
- 负二项分布
- hurdle shifted Poisson process (原始的泊松分布向正值平移1，然后概率为 1-p 取 0, p 取正值)

具体如下表所示：

Table 1
Count distributions used in the empirical study.

Distribution	Mass function	Parameter restrictions	Mean (μ)
Poisson	$\frac{\lambda^y}{y!} \exp(-\lambda)$	$\lambda > 0$	λ
Negative binomial	$\frac{\Gamma(a+y)}{\Gamma(a)y!} \left(\frac{b}{1+b}\right)^a \left(\frac{1}{1+b}\right)^y$	$a > 0, b > 0$	$\frac{a}{b}$
Hurdle shifted Poisson	$\begin{cases} q & y = 0 \\ p\lambda^{y-1} \exp(-\lambda)/(y-1)! & y = 1, 2, \dots \end{cases}$	$p \geq 0, q > 0, \lambda > 0, p + q = 1$	$p(\lambda + 1)$

同时，考虑均值随时间变化的状况，设定以下三种均值变化模式（以及shifted poisson的0值概率p的变化模式）：

Table 2
Recurrence relationships for the mean.

Relationship	Recurrence relationship	Restrictions
Static	$\mu_t = \mu_{t-1}$	
Damped dynamic	$\mu_t = (1 - \phi - \alpha)\mu + \phi\mu_{t-1} + \alpha y_{t-1}$	$\mu > 0, \phi > 0, \alpha > 0$ $\phi + \alpha < 1$
Undamped dynamic	$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}$	$\delta > 0, \alpha > 0$ $\delta + \alpha = 1$

(感觉Hurdle Shifted Poisson与TSB有一些联系；或许还可以对需求间隔指数平滑并设定参数分布（如几何分布））

在预测过程中，使用MLE获得相关参数。Poisson与Hurdle Shifted Poisson在static均值下参数可以解析求得，其它模型需用数值计算。

预测时，由于需要滞后值，可通过模拟的方式（1000次）获得多条时间序列以查看分位数。同时，预测期分布的参数（基于时变的均值）个人通过对模拟取平均的方法获得。向前预测28期。

已实现模型回顾

自11.4以后，近一个月实现了以下概率预测模型：

- 基于历史数据：依托WSS与VZ的历史数据模拟。
- 基于分位数估计：基于GEFcom2014第一名的方法，先用GAM提取均值成分，再结合计数数据分位数回归的方法 (Machado & Silva, 2005)对成分进行线性分位数回归。分位数点预测
- 基于分布模型：Snyder et al. (2012) 的方法（如上）。

这三个方法使用的信息有所不同：一、三使用的是历史销量数据及其滞后，二使用了外生变量，但并未使用任何销量的滞后项。虽然法二的安排有简化的目的（引入滞后变量需要大量模拟，失去了分位数回归直接估计分布的优势），但是这形成了模型在使用信息上的异质性。

这三个方法均是**local模型**，即只使用本条时间序列的信息进行预测。

之后想法

原计划是打算实现以下两种模型：（1）GAM对分布参数的估计，使用外生变量（2）DeepAR等M5使用的机器学习、深度学习方法（同时引入滞后与外生变量，且做Global模型），但近期可用时间有限！

近期还是转向对probabilistic combination 的文献研究，主要考虑：

- 之前对BMA及BPS(Bayesian predictive synthesis)不够理解，需要再看看
- 基于模型的预测分布及分位数预测结果怎么组合（分布组合、分位数组合等）
- 可否**Global化**
- 离散问题
- 相关性（问题：分布预测结果的相关性度量是计算不同基准模型预测分布的相关系数吗，感觉同一个数据预测的模型应该是很像的，相关性大概率不低。。。）
- 相似性：比如KL散度，频域上的熵等。。。

quantile regression averaging：单个模型点预测作regressor

降维、加feature..... ensemble?

比赛选手特征工程等重视.....