

概率预测组合文献

Combining Density Forecasts (Stephen G. Hall, James Mitchell, 2007) :

- **组合方式：**密度预测的组合形式为一线性组合

$$p_t(y_t) = \sum_{i=1}^N w_i g_{it}(y_t)$$

为得到权重，优化真实密度与估计密度的“KLIC距离”，如下：

$$\text{KLIC}_t = E[\ln f_t(y_t) - \ln p_t(y_t)]$$

$$\overline{\text{KLIC}} = \frac{1}{T} \sum_{t=1}^T [\ln f_t(y_t) - \ln p_t(y_t)]$$

真实分布与待优化参数无关，故可直接优化估计密度的对数似然函数（对数分数）；KLIC距离的性质可保证估计结果在样本内不会更差。

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{1}{T} \sum_{t=1}^T \ln p_t(y_t)$$

- **方法说明：**该方法可以和一种拟合优度检验——Berkowitz (2001) LR test 联系起来，其检验统计量与KLIC距离仅差一个常系数；但是该文章未证明其最优性。不过提供了一种检验估计密度的方法。

$$\text{LR} = 2 \sum_{t=1}^T [\ln q_t(z_t^*) - \ln \phi(z_t^*)]$$

- **组合实证*：**1994-2004英国通胀数据，个体预测为 *National Institute Economic Review*、the Bank of England 的预测以及一个自行设定的简单时间序列预测（高斯分布）。使用最优权重，确实在样本内不会更差，但样本外不一定能优于最优个体预测（取决于组合的质量）。

Optimal prediction pools (John Geweke, Gianni Amisano, 2011) :

- **组合方式：**线性组合，使用最大化对数分数优化权重（与上一篇类似）

$$\sum_{i=1}^n w_i p(\mathbf{y}_t; \mathbf{Y}_{t-1}^o, A_i); \quad \sum_{i=1}^n w_i = 1; \quad w_i \geq 0 \\ (i = 1, \dots, n).$$

$$\sum_{t=1}^T \log \left[\sum_{i=1}^n w_i p(\mathbf{y}_t^o; \mathbf{Y}_{t-1}^o, A_i) \right]$$

- **方法说明**: 证明了两模型、多模型组合优化函数的凸性，讨论了模型池中某些模型占优时权重的性质（会自行排除不好的模型），说明了在足够数据的基础上，组合预测会优于单独模型。
- **组合实证**: S&P500 数据1250交易日的日对数收益率数据。个体模型为：独立同分布高斯模型、GARCH、E-GARCH、t-GARCH、stochastic volatility model (SV)、hierarchical Markov normal mixture model (HMMM)。两模型组合中组合的结果基本都优于最优模型，甚至组合中可以不包括最优模型；多模型组合效果更好，但是结果（对数分数）对权重的变化并不敏感；作者怀疑更精细方法（bagging、boosting等）进一步优化的意义（有待研究）。

Combining probability forecasts(Michael P . Clements, David I. Harvey, 2011):

- **组合方式**: 三种方式：**线性组合** (LiOP) 、**乘积组合** (LoOP) 、Kamstra and Kennedy(1998) 的**组合方式 (KK)**

$$\text{LiOP } C_t(f_{1t}, f_{2t}; \beta) = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t}$$

$$\text{LoOP } C_t(f_{1t}, f_{2t}; \beta) = \alpha f_{1t}^{\beta_1} f_{2t}^{\beta_2}$$

$$\begin{aligned} \text{KK } C_t(f_{1t}, f_{2t}; \beta) \\ = \frac{\exp \left[\alpha + \beta_1 \ln \left(\frac{f_{1t}}{1-f_{1t}} \right) + \beta_2 \ln \left(\frac{f_{2t}}{1-f_{2t}} \right) \right]}{1 + \exp \left[\alpha + \beta_1 \ln \left(\frac{f_{1t}}{1-f_{1t}} \right) + \beta_2 \ln \left(\frac{f_{2t}}{1-f_{2t}} \right) \right]} \end{aligned}$$

- **组合实证**: 大致是说在数据有限时LiOP较好，数据大时KK较好

（涉及了一个基于设定DGP的模拟实验和基于美国经济衰退概率的实证。（关于DGP的相关讨论有待进一步理解）

M5的概率预测

- **Robust recurrent network model for intermittent time-series forecasting** (Yunho Jeon, Siheyeon Seong, 2021): 概率预测未必和点预测对立；（M5点预测第三名）尽管是点预测竞赛，但仍然使用 **DeepAR** 优化 Tweedie 分布的似然函数（将连接函数一侧作为网络输出）；更进一步，在预测过程中使用来自拟合分布的多个样本进行滚动预测以生成最终预测序列。
- **M5 Competition Uncertainty: Overdispersion,distributional forecasting, GAMLSS and beyond** (Florian Ziel, 2021) (arXiv):

提出了一个基于**广义加性模型**的概率预测框架 (generalized additive models location shape scale, GAMLSS)，针对于大多数模型没有考虑到M5数据的过度分散等问题。直接对分布进行建模，而非对分位数建模。基本模型如下：

$$g_m(\theta_m) = \sum_{n=1}^{N_m} f_{m,n}(\mathbf{X}_t)$$

实证： M5数据（做了一些量的缩减），尝试7种分布 (Poisson, geometric, negative binomial, Waring, generalized Poisson, double Poisson and zero-inflated Poisson)

使用Lasso进行正则化

考虑了一个简单的模型组合（选择）：基于HQC标准为每个预测选择上述7分布中最优的分布

结果：基本都优于Arima的Benchmark（M5简单方法最优）；经HQC选择后，改进超过10%，而M5最优方法改进为12%。主要优化在靠中间的分位数（与M5结果相似）