2021.10.21

概率预测组合文献（二）

主要关注内容：组合权重的估计——**由固定权重转向时变（动态）权重**

Combining Forecast Densities from VARs with Uncertain Instabilities (Jore et al. , 2010)

- **组合方式**：

  依然是线性组合，但是权重初步考虑了时间的影响：

$$p_\tau(y_{\tau,h}) = \sum_{i=1}^{N} w_{i,\tau,h}\ g(y_{\tau,h} \mid I_{i,\tau}), \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}$$

  这里权重被称作“递归对数分数权重”（recursive logarithmic score weights），定义为前第10期到最近观测期的对数分数占比：

$$w_{i,\tau,h} = \frac{\exp\left[\sum_{\underline{\tau}-10}^{\tau-1-h} \ln g(y_{\tau,h} \mid I_{i,\tau})\right]}{\sum_{i=1}^{N} \exp\left[\sum_{\underline{\tau}-10}^{\tau-1-h} \ln g(y_{\tau,h} \mid I_{i,\tau})\right]}, \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}$$

- **组合实证**：

  数据与方法：估计美国产出、通胀、短期利率的模型，数据为1965Q4-2005Q4的季度数据，个体模型为AR及VAR类（8种），数据划分与Clark 和 McCracken (2008) 的研究保持一致以保证可比性。组合策略还考虑了等权组合的方法。实验结果主要是考察与PIT（概率积分变换）相关的正态性检验，进以评估拟合优度。

  实证结论：

  1. 点预测与区间预测的不同：基准的AR方法及CM（2008）推荐的点预测方法几乎没有证据支持有利于组合预测（权重偏小甚至为0）；相关的统计检验也支持上述结论。
  2. 在概率校准（Calibration）检验方面，等权（EW）与本文提出的递归权重（RW）大多数时候表现相当，但在1985-2005时间段EW稍差；其它方法在此方面不如这两个组合方法。

此外，在*Nowcasting GDP in Real Time: A Density Combination Approach* (Aastveit et al., 2014) 一文中，也利用了相同的权重设置，而个体方法改为Bridge、Dynamic Factor Models、MF-VAR三种方法，对美国1990Q2到2010Q3的GDP数据进行预测，在对数分数和校准检验两方面进一步验证了RW权重相对与个体方法的优势。

Confronting Model Misspecification in Macroeconomics (Waggoner, Zha, 2012)

- **组合方式**：

  使用马尔可夫状态切换机制决定两个模型权重如何变化。两个个体模型分别是DSGE（动态随机一般均衡，参数化较紧密）和BVAR（贝叶斯向量自回归，参数化较松散）。其组合预测如下所示：

$$p\left(y_t \mid Y_{t-1}^o, \Theta, Q, w, s_t\right)=\sum_{i=1}^n w_{i, s_t} p\left(y_t \mid Y_{t-1}^o, \Theta_i, M_i\right)$$

其中 Y 指观测数据，Θ 指个体模型参数，Q 指状态转移矩阵，s_t 指不可观测的状态（服从马尔科夫链）。

这里作者选择将参数、转移矩阵与权重同时进行估计，通过最大化后验以及Gibbs、MCMC等方法来估计参数与权重。（最终的估计　　结果是：估计出两个状态，每个状态有一个固定的权重分配，模型组合在这两个状态中选一个。）

$$\log p\left(Y_T^o \mid \Theta, Q, w\right)=\sum_{t=1}^T \log p\left(y_t^o \mid Y_{t-1}^o, \Theta, Q, w\right)$$

$$p\left(\Theta, Q, w \mid Y_T^o\right) \propto p\left(Y_T^o \mid \Theta, Q, w\right) p(\Theta, Q, w)$$

- **组合实证**：

  数据：宏观经济数据（1960-2010）

  方法：个体方法、等权组合、马尔可夫切换的对比，参考指标为对数MDD（边际数据密度）、LPS（对数预测分数）、LCPL（对数集中预测似然）。

  结论：等权组合已经带来了较大提升，而马尔可夫切换在此基础上进一步提升（适应不同的冲击）。文章还用大量篇幅讨论了如何基于可解释的DSGE来讨论如何从组合结果中分析外生冲击的影响（可解释性研究）。

Generalised density forecast combinations (Kapetanios et al., 2015)

- **组合方式**：

  提出了“广义意见池”（Generalised opinion pool）的概念，每个方法的权重时变。

$$p_t(y)=\sum_{i=1}^N w_{it}(y)\, q_{it}(y)$$

such that

$$\int p_t(y)\, dy=1,$$

  为了能解出权重，文中将权重表示成基函数回归的形式，以解决最小化损失问题。

$$\Phi_{q_i}^T=\left\{w_{T\eta i}(.): w_{T\eta i}(.)=\tilde{\nu}_{i0}+\sum_{s=1}^{p_T} \tilde{\nu}_{is} \eta_s\left(y, \boldsymbol{\theta}_s\right),\right.$$
$$\left.\tilde{\nu}_{is} \geq 0, \eta_s\left(y, \boldsymbol{\theta}_s\right) \geq 0\right\}, \quad i=1, \ldots, N,$$

  权重的时变性基于基回归系数 v 的标准化实现：

$$\tilde{\nu}_{it0} = \tilde{\nu}_t(\nu_{i0}; \boldsymbol{\nu}_i^{(-0)}) = \tilde{\nu}_t(\nu_{i0}) = \frac{\nu_{i0}}{\sum_{i=1}^{N}\left(\nu_{i0} + \sum_{s=1}^{p_T} \nu_{is}\kappa_{its}\right)}$$

$$\tilde{\nu}_{its} = \tilde{\nu}_t(\nu_{is}; \boldsymbol{\nu}_i^{(-s)}) = \tilde{\nu}_t(\nu_{is}) = \frac{\nu_{is}}{\sum_{i=1}^{N}\left(\nu_{i0} + \sum_{s=1}^{p_T} \nu_{is}\kappa_{its}\right)},$$

基的个数作者建议通过交叉验证选取；基的形式作者认为可以有多种设置，文中用的是分段回归的基函数（分段示性函数）。而分段的边界如果不事先指定，通过网格搜索解决。损失函数仍是对数分数（再加负号）。

（分段线性回归可能带来不光滑的性质，文中也提出了一种改进措施来缓解这一问题，以得到更光滑的密度）

- **方法说明**:

  权重的一致性：证明了在一定假设下，权重基回归的参数 υ 和 θ 满足渐进正态性，依概率收敛到真值；

  回归系数 υ 的估计也可证明具有相合性、渐进正态性。

- **组合实证**:

  该文章进行了模拟实验与实证实验。

  蒙特卡洛模拟：做了四个实验以考察**对错误指定的模型的改进**：1.真实密度非正态，用基于正态的广义线性组合去捕捉非正态分布；2.真实密度正态，用两个错误指定的正态组合分析组合改进能力；3.考虑门限自回归模型，评估不同自回归参数线性正态模型组合逼近非线性过程的能力；4.真实密度随时间演变，其根据一个带随机波动的不可观测成分趋势周期模型而变化（现实背景是美国的通货膨胀），考察两个没有随机波动的模型组合预测能力。将广义池与线性池做比较。

  实证分析：使用 S&P500 在1972.1.3-2013.9.9每日对数回报率数据，对标 John Geweke 与 Gianni Amisano (2011) 的结果。个体模型为 GARCH, T-GARCH, E-GARCH, SV四种，尝试2模型、3模型、4模型组合。将广义池与线性池做比较。

  结论：模拟与实证的结果证明对数情况下广义池的优越性；此外的发现是，**交叉验证可以避免过拟合**（倾向于少分段）；**广义池更适用于大样本**（因其需要参数更多，故需要更长时间的数据）；在实证中作者的建议是**对复杂的组合使用简单的权重估计形式**（少分段）；以及**个体模型数不是越多越好**；同时其还发现线性池**在样本外可以但通常不会胜过所有个体模型的预测结果。**

Dynamic prediction pools: An investigation of financial frictions and forecasting performance (Del Negro et al., 2016)

- **组合方式**:

  动态池，依然是线性组合，下式是联合概率：

$$p(y_{1:T}|\lambda_{1:T}, \mathcal{P}) = \prod_{t=1}^{T} \{\lambda_t p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1-\lambda_t)p(y_t|y_{1:t-1}, \mathcal{M}_2)\}, \quad \lambda_t \in [0, 1].$$

  与“广义池”不同的是，其设置了权重的一个先验随机过程：

$$\begin{aligned} x_t &= (1-\rho)\mu + \rho x_{t-1} + \sqrt{1-\rho^2}\sigma\varepsilon_t, \\ &\qquad \varepsilon_t \sim iid\, N(0, 1),\ x_0 \sim N(\mu, \sigma^2), \\ \lambda_t &= \Phi(x_t), \end{aligned}$$

而权重的在 $t-1$ 、 t 时刻的后验如下所示，而密度序列使用Bootstrap来近似；最后对后验概率求期望得到权重的估计。文章对向前 h 步的权重估计也进行了讨论。

$$p(\lambda_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P}) = \int p(\lambda_t|\theta, \lambda_{t-1}, \mathcal{P})p(\lambda_{t-1}|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})d\lambda_{t-1}$$

$$p(\lambda_t|\theta, \mathcal{I}_t^{\mathcal{P}}, \mathcal{P}) = \frac{[\lambda_t p(y_t|y_{1:t-1}, \mathcal{M}_1) + (1-\lambda_t)p(y_t|y_{1:t-1}, \mathcal{M}_2)]p(\lambda_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})}{p(y_t|\theta, \mathcal{I}_{t-1}^{\mathcal{P}}, \mathcal{P})}$$

- **组合实证**：

  数据：1992-2011产出增长、通胀等宏观经济数据

  方法：个体方法是两个不同的DSGE（动态随机一般均衡模型），与等权组合、BMA（贝叶斯平均）、最优（静态）池、动态模型平均等组合方法对比。

  结论：动态池在**面对冲击时权重变化惯性小**，易于对模型预测结果的相对变化作出反应；在以对数得分表示的结果中，动态池略优于等权组合，大优于其它组合（把握不好权重变化规律，不如一视同仁）。文章也说明了本方法在政策分析的应用。

Dynamic Bayesian predictive synthesis in time series forecasting (McAlinn and West, 2019)

- **组合方式**：本文提出了一个BPS（贝叶斯预测合成）框架，其对当前时刻的概率预测可以表示为下式：

$$p(y_t|\boldsymbol{\Phi}_t, y_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(y_t|\boldsymbol{\Phi}_t, \mathcal{H}_t) = \int \alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t) \prod_{j=1:J} h_{tj}(x_{tj})dx_{tj}$$

  这里 h 指个体预测模型（文中是代理模型，agent），α 是 y 在给定 x 的条件下 y 先验的条件概率密度函数（合成函数，模型的主要目标），Φ 是其中的时变参数，h 是当前预测的密度。。进一步，使用动态线性模型（DLM）来作为预测合成的方法，式中主要量可以如下表示：

$$\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t) = N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t) \quad \text{with} \quad \boldsymbol{F}_t = (1, \boldsymbol{x}_t')' \quad \text{and} \quad \boldsymbol{\theta}_t = (\theta_{t0}, \theta_{t1}, \ldots, \theta_{tJ})'$$

$$\begin{aligned} y_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t), \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\boldsymbol{W}_t) \end{aligned}$$

$$p(\boldsymbol{x}_t|\boldsymbol{\Phi}_t, y_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(\boldsymbol{x}_t|\mathcal{H}_t) = \prod_{j=1:J} h_{tj}(x_{tj})$$

  生成预测时通过MCMC依次采样 Φ, θ, y 来实现。考虑到短期模型和长期模型有各自的适用性，文中尝试了提前1步和4步的预测。

- **组合实证**：

  数据：美国宏观经济数据，1961年第一季度至2014年第四季度美国经济的年通货膨胀率(P)、短期名义利率(R)和失业率(U)

  方法：个体方法为四个DLM代理模型（变量不同），将BPS与BMA、等权池、对数池比较，使用指标为其它方法与BPS密度比的对数

  结论：BPS能够快速适应个体模型预测的变化而修改权重，预测的标准差较低

随便想想

所看论文主线是如何进行概率预测，大多数是如何安排线性组合的权重，从定值转向时变权重。所用方法也由直接优化对数分数转向贝叶斯方法（对数分数仍是一常用评价指标）。

Dynamic Bayesian predictive synthesis in time series forecasting (McAlinn and West, 2019) 一文在结论中提出，本文的方法可以用在包括金融(例如股票、指数和债券)、业务(例如产品需求和收益)、气象和风险(例如地震和环境风险)等领域。**对于多变量综合、非正态预测和离散数据，以及缺失或不完整/部分预测，都需要方法论上的扩展。**据此，有下列想法：

- 所见论文的实证都基于宏观经济或金融数据，（一般基于正态假设的个体模型进行组合），那对于不适于用正态假设的数据（如intermittent demand），是否对于概率组合预测需要有较大方法上的变化；
- 这些论文的组合预测都是基于一条时间序列，那么如果基于数据集，利用数据集的信息对概率组合预测是否有更好的效果，即对于数据集的概率预测，是每条单独训练权重，还是需要有一个共通的模式生成各序列的权重；global combination ensemble
- 多数都是线性权重组合的概率预测，那么非线性的组合方式是否还要进一步思考；
- M5的不确定性预测似乎没有考虑方法组合的问题，那么，假设组合前几名的预测结果（不同设置的网络），或者将网络与统计方法组合，能否获得严格更好的效果，或者利用较简单的统计方法的组合以尽力接近网络的结果?
- 效用?