模型平均化

Model Averaging

仓库: https://gitee.com/arlionn/MA

连玉君 (中山大学)

arlionn@163.com

模型不确定性

- 影响经济增长的因素：Captial，Labor，Law，Culture，... ...
- 假设有 K 个潜在的因素，则可能存在 2^K 个 线性模型：

$$y = c + X_\ell \beta_\ell + u, \quad \ell = 1, \ldots, 2^K$$

- 例如，$2^6 = 64, \quad 2^{10} = 1024$
- 如何确定最佳模型?
  - p-hacking: $*** > ** > *$
  - R^2，RSS，AIC，BIC，... ...
  - 变量筛选：Lasso，Lasso-IV，DML (Double Machine Learning)
  - Model Averaging (模型平均化)

谋杀率与死刑 Moral-Benito (2015, PDF)

Table 1. The Deterrent Effect of Capital Punishment.

	Dependent variable is the murder rate		
	(1)	(2)	(3)
Execution rate	0.19	−0.44	−11.49
	(0.02)	(−0.18)	(−2.72***)
Controls included	None	KLS (2003)	DS (2006)
N	51	51	51
R^2	0.01	0.68	0.85
Net lives saved per execution	−1.47	0.06	26.93
	(−0.07)	(0.01)	(2.62***)

- KLS (2003) 控制包括监狱死亡率、每起暴力犯罪的囚犯、每 100,000 名居民的囚犯、人均收入、保险失业率、黑人和城市人口的百分比、年龄分布变量（0-24 岁和 25-44 岁的分数 -olds) 和婴儿死亡率。
- DS (2006) 控制包括人均收入、失业率、警察就业、少数族裔百分比和年龄分布变量（15-19 岁和 20-24 岁的部分）。

Table 2. The Deterrent Effect of Capital Punishment Via Model Averaging.

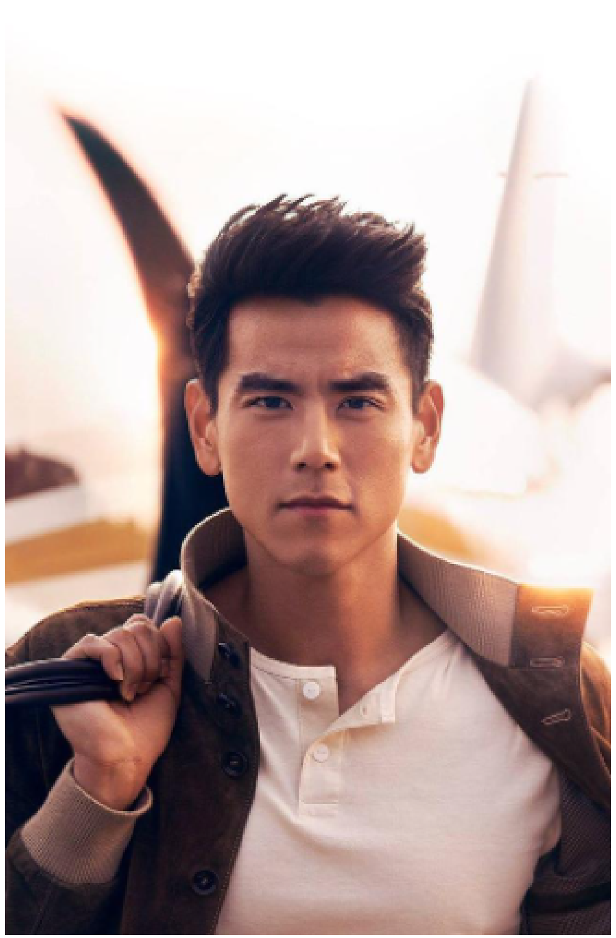
	BMA-g (1)	WALS (2)	BMA-bic (3)	FMA-aic (4)	MMA (5)	JMA (6)
Execution rate	−2.477	−1.786	−2.652	−2.701	−2.712	−2.717
	(2.357)	(2.264)	(2.343)	(2.331)	(2.327)	(2.303)
Infant mortality rate	0.000	0.003	0.000	0.001	0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Prison deaths rate	0.002	0.035	0.007	0.020	0.018	0.019
	(0.058)	(0.211)	(0.088)	(0.084)	(0.076)	(0.088)
Prisoners per 100,000 residents	0.007	0.005	0.007	0.007	0.007	0.007
	(0.005)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
Prisoners per violent crime	−5.716	−4.072	−5.676	−5.783	−5.783	−5.946
	(5.270)	(3.106)	(4.894)	(4.275)	(4.398)	(4.280)
Fraction black	9.604	6.898	9.415	9.171	9.404	9.540
	(4.189)	(3.582)	(4.097)	(3.957)	(3.777)	(3.647)
Fraction urban	0.067	1.959	0.176	0.517	0.583	0.648
	(1.201)	(2.769)	(1.633)	(1.621)	(1.585)	(1.754)
Fraction 25–44 year-olds	2.488	23.296	5.401	11.209	12.553	12.061
	(10.382)	(20.588)	(14.417)	(17.490)	(18.232)	(17.818)
Fraction 0–24 year-olds	0.700	5.849	1.304	2.746	2.981	2.925
	(4.541)	(23.022)	(7.364)	(8.585)	(8.039)	(8.757)

> Source: Moral-Benito, E., **2015**, Model averaging in economics: An overview, **Journal of Economic Surveys**, 29 (1): 46-75. -Link-, -PDF-

MA：基本思路

- 备选模型： $\overline{\mathcal{M}} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$
- 对应的估计值： $\{\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_M\}$
- 权重： $w = \{w_1, \ldots, w_M\}$
- 平均化估计量 (Averaging estimator)

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{w}) = \sum_{m=1}^{M} w_m \widehat{\boldsymbol{\theta}}_m$$



w_m^* 的选择与目的有关

- 模型不确定性
- 变量筛选
- 预测

前提：

- 备选模型不存在系统性偏误

```
sysuse "nlsw88.dta", clear
eststo m1: reg wage union hours 2.race tenure
eststo m2: reg wage union hours if e(sample)
esttab m2 m1, nogap s(rss rmse aic bic r2 N) drop(_cons)
--------------------------------------------
wage    N=1867       (1)              (2)
--------------------------------------------
union              1.378***         1.193***
                  (6.26)           (5.58)
hours             0.0539***        0.0422***
                  (5.66)           (4.56)
2.race                            -1.356***
                                  (-6.55)
tenure                             0.179***
                                  (10.88)
--------------------------------------------
rss              31161.0          28745.0
rmse               4.089            3.929
aic              10559.6          10412.9
bic              10576.2          10440.6
r2                0.0392            0.114
--------------------------------------------
(t), * p<0.05, ** p<0.01, *** p<0.001
```

线性回归回顾

- $y_i = x_i'\beta + e_i, \quad i = 1, \ldots, n, \qquad x_i$ 和 β 均为 $k \times 1$
- $\mathbb{E}\left(e_i \mid x_i\right) = 0, \quad \sigma^2 = \mathbb{E}\left(e_i^2\right)$
- 参数个数为：$\left(\beta \text{ and } \sigma^2\right)$ is $K = k + 1$
- $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{y}\right)$ 是 LS 的系数估计值
- $\widehat{e_i} = y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}$ 为 LS 的残差
- $\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^n \hat{e}_i^2$ 为方差估计量

AIC 和 BIC： MLE 估计

Akaike's (1974) AIC 和 Schwarz's (1978) BIC 都用于评估模型拟合程度 ([R] estat ic)：

$$\mathrm{AIC} = -2 \ln L + 2 \cdot k$$

$$\mathrm{BIC} = -2 \ln L + \ln(n) \cdot k$$

其中，k 是模型中的参数个数， $\ln(L)$ 是对数似然函数。

- $f(y, \theta)$ is a parametric density with a $K \times 1$ parameter $\boldsymbol{\theta}$
  - e.g. $y_i \sim N(x_i' \beta, \sigma^2)$， $\boldsymbol{\theta} = (\beta, \sigma)$
  - $$f(y_i; \beta, \sigma^2) = \phi(x_i' \beta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right)$$
- 似然函数 (likelihood) 定义为：$L(\boldsymbol{\theta}) = f(\boldsymbol{y}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i, \boldsymbol{\theta})$
- The MLE $\widehat{\boldsymbol{\theta}}$ 极大化 $\ln L(\boldsymbol{\theta})$

AIC 和 BIC：OLS 估计

当模型的干扰项服从独立正态分布时，AIC 和 AIC 可表示为：

$$\mathrm{AIC} = n\ln(RSS/n) + 2k$$
$$\mathrm{BIC} = n\ln(RSS/n) + k\ln(n)$$

其中，n 为观察值个数，RSS 为残差平方和。

也可以用 $\mathrm{MSE} = \hat{\sigma}^2$ 替代上式中的 RSS/n，因为 $\hat{\sigma}^2 = RSS/n$。

Mallow's c_p

Source: Gilmour (1996), Hansen (2007, PDF)

另一种常用的评价准则是 Mallow's (1973) c_p，定义为：

$$C_p=\frac{RSS_k}{\widehat{\sigma}^2}-(n-2k).$$

- $RSS_k=\sum_{i=1}^n\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^n\hat{e}_i^2$，为包含 k 个变量的模型的残差平方和
- $\widehat{\sigma}^2=RSS_{FULL}/n$，基于完整模型 (full model) 计算而得。

> **Remark 1**
> Atilgan (1996) provides a relationship between AIC and Mallows's C_p, shows that under some conditions AIC selection behaves like minimum mean squared error selection, and notes that AIC and C_p are somewhat equivalent criteria.

AIC, BIC 和 C_p 的关系

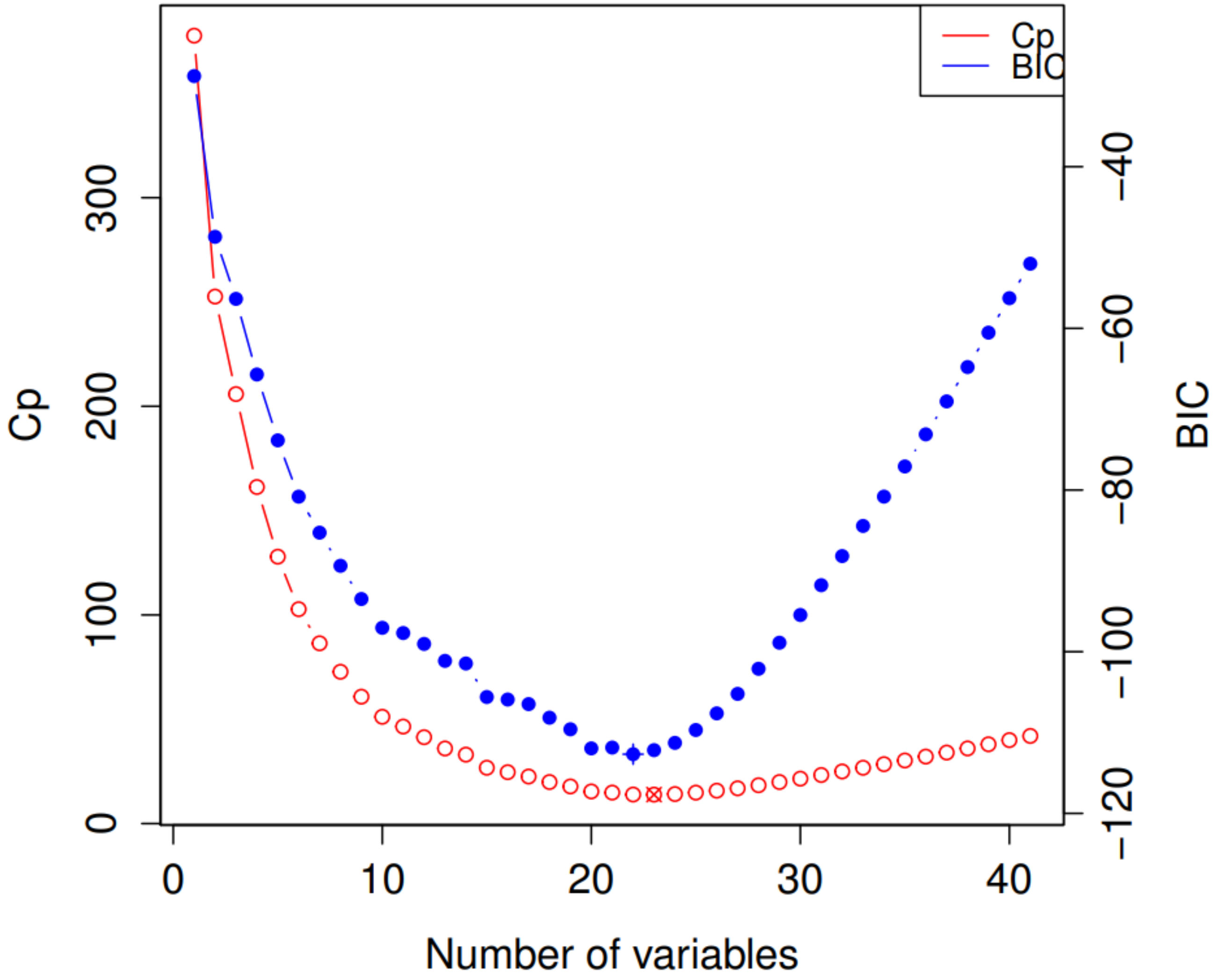
- 在具有高斯误差的线性模型中：
    - (1) MLE 和 OLS 等价；
    - (2) C_p 和 AIC 也等价
- BIC：倾向于选择相对 **精简** 的模型
- AIC：倾向于选择相对 **饱满** 的模型
    -

更一般化的理解

AIC 和 BIC 都试图在「拟合程度」和「模型复杂度」之间进行权衡，通用表达式为：

$$IC = -2\ln(\text{likelihood}) + c(n, K)$$

- 拟合程度：用 (负数项) 来衡量，值越小越好 (拟合的好)
- 模型复杂度：用 (正数项) 来衡量，差别在于惩罚力度不同。
  - AIC：$c = 2K$
  - BIC：$c = K\log(n)$
- 对于一笔相同的数据，IC 值越小的模型越佳。



I just ran two trillion regressions

Christoph Hanck
Universität Duisburg-Essen

Hanck 2016

模型筛选 v.s. 模型平均化

- 模型集：$\overline{\mathcal{M}} = \{\mathcal{M}_1, \ldots, \mathcal{M}_M\}$
- 模型筛选：从集合中挑选一个（`help stepwise`，[R] stepwise）
- 这是最好的做法吗？
- 类比：投资组合选择
- 假设我们有一组资产
  - 我们应该把所有银子都用来买房子或买股票吗？
  - 还是说应该分散投资？
  - 我们知道：多元化可减少差异和风险
  - 因为：虽然高度相关，但资产并非完全相关
  - 最关键的是：我们永远不知道「**True Model**」

MA：基本思路 (review)

- 备选模型：$\overline{\mathcal{M}}=\{\mathcal{M}_1,\ldots,\mathcal{M}_M\} \quad \Rightarrow \quad \{\widehat{\boldsymbol{\theta}}_1,\ldots,\widehat{\boldsymbol{\theta}}_M\}$
- 权重：$w=\{w_1,\ldots,w_M\}$
- 平均化估计量 (Averaging estimator)

$$\hat{\theta}_{MA}=\sum_{m=1}^{M} w_m\hat{\theta}_m$$

- 方差：

$$\operatorname{var}\left(\hat{\theta}_{MA}\right)=\sum_{m=1}^{M} w_m^2\left[\operatorname{var}\left(\hat{\theta}_m \mid \mathcal{M}_m\right)+\left(\hat{\theta}_m-\hat{\theta}_{MA}\right)^2\right]$$

如何选择权重 w_m？

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{w})=\sum_{m=1}^{M} w_m \widehat{\boldsymbol{\theta}}_m$$

- 传统频数视角：**FMA**, Jackknife-MA (**JMA**), **LS-MA**，**MMA**
  - 基于 RRS，AIC/BIC, Mallow's C_p 等指标确定权重
  - 多以大样本理论为基础
- 贝叶斯视角：**BMA**
  - 以 Bayes 法则为基础
  - 以先验概率和先验分布为基础，根据数据信息得到后验分布 $Pr(\mathcal{M}_i \mid y)$
- **WLMA** (Weighted Average Least Squre)
  - 区分 **Focus Variables** 和 **Auxiliary Variables**：$y=x_1\beta_1+x_2\beta_2+u$
  - 以 FWL 定理为基础的 BMA + FMA

模型选择概率

- 假设我们有两个模型 $\mathcal{M}_1$ 和 $\mathcal{M}_2$，边际概率为 $p_1(\boldsymbol{y})$ 和 $p_2(\boldsymbol{y})$
- 贝叶斯定理表明：

$$
\begin{aligned}
\operatorname{Pr}\left(\mathcal{M}_1 \mid \boldsymbol{y}\right) &= \frac{p_1(\boldsymbol{y})}{p_1(\boldsymbol{y})+p_2(\boldsymbol{y})} \\
\operatorname{Pr}\left(\mathcal{M}_2 \mid \boldsymbol{y}\right) &= \frac{p_2(\boldsymbol{y})}{p_1(\boldsymbol{y})+p_2(\boldsymbol{y})} .
\end{aligned}
$$

- 贝叶斯选择概率较大的模型：
  - 如果 $p_1(\boldsymbol{y}) > p_2(\boldsymbol{y})$，选择 $\mathcal{M}_1$，反之，则选择 $\mathcal{M}_2$
- 极大化 $p(\boldsymbol{y})$ 与极小化 $-2\log p(\boldsymbol{y}) \simeq \mathrm{BIC}$ 是等价的
- 因此，基于 BIC 准则选择模型近似于贝叶斯选择，我们以 BIC 为基础来确定权重

Smooth BIC (BIC weighting) - 近似 BMA

> Burham and Anderson (2002).

根据贝叶斯定理：

$$-2\log p(\boldsymbol{y}) \simeq \mathrm{BIC} \quad \Longrightarrow \quad p(\boldsymbol{y}) \simeq \exp(-\mathrm{BIC}/2)$$

这意味着，给定数据集，一个模型为真的概率 $p(\boldsymbol{y})$ 与 $\exp(-\mathrm{BIC}/2)$ 成正比。

因此，我们可以设定如下权重：

$$w_m = \frac{p_m(\boldsymbol{y})}{\sum_{j=1}^{M} p_j(\boldsymbol{y})} = \frac{\exp\left(-\mathrm{BIC}_m/2\right)}{\sum_{j=1}^{M} \exp\left(-\mathrm{BIC}_j/2\right)}$$

Computational SBIC

- 从计算的角度来看，使用 BIC 的差值更好一些
- 在有些模型中，BIC_m 的值很大，致使 $\exp\left(-\mathrm{BIC}_m/2\right)$ 会超出计算能力
- 令 $\mathrm{BIC}^* = \min_m \mathrm{BIC}_m$ 表示所有模型中的 BIC 的最小值
- 设 $\Delta\mathrm{BIC}_m = \mathrm{BIC}_m - \mathrm{BIC}^* \geq 0$，则

$$w_m = \frac{\exp\left(-\mathrm{BIC}_m/2\right)}{\sum_{j=1}^{M}\exp\left(-\mathrm{BIC}_j/2\right)} = \frac{\exp\left(-\Delta\mathrm{BIC}_m/2\right)}{\sum_{j=1}^{M}\exp\left(-\Delta\mathrm{BIC}_j/2\right)}$$

- 显然，二者是等价的，但采用第二个式子的计算负担会小很多。

更简洁的表述：

- 令 $\Delta_k = BIC_k - BIC_{\min}$ 或 $\Delta_k = AIC_k - AIC_{\min}$
- 则权重 w_k 表示为：

$$w_k = \frac{\exp\left(0.5\Delta_k\right)}{\sum_{j=1}^{K} \exp\left(0.5\Delta_j\right)}$$

Smoothed AIC (SAIC)

Buckland et al. (1997), Burnham and Anderson (1998) 认为也可以基于 AIC 构建权重：

$$w_m=\frac{\exp\left(-\mathrm{AIC}_m/2\right)}{\sum_{j=1}^M \exp\left(-\mathrm{AIC}_j/2\right)}=\frac{\exp\left(-\Delta\mathrm{AIC}_m/2\right)}{\sum_{j=1}^M \exp\left(-\Delta\mathrm{AIC}_j/2\right)}$$

显然，

- AIC 或 BIC 越小的模型被赋予的权重越大

Bootstrap MA：基于 Bootstrap 确定权重

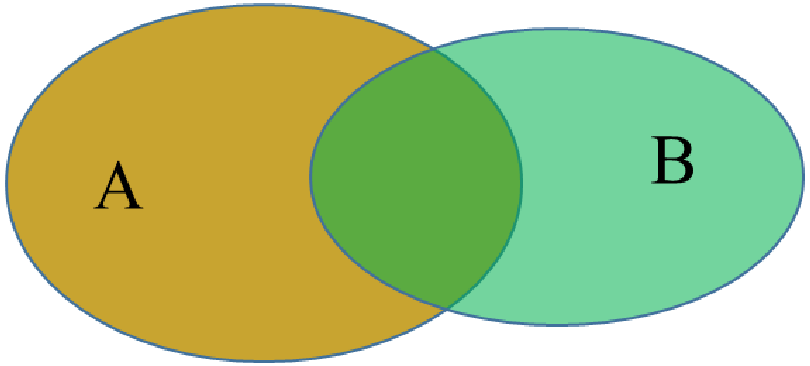
> Source: link

- BMA 和 FMA 的问题：模型冗余
  - `reg y c.(x1 x2)#(z1-z5) c.(x1 x2)##c.(x1 x2)`
  - 很多解释变量高度相关，导致有诸多相似的模型，会挤占其它有用模型的权重
- 解决办法：
  - A. Bootstrap 再抽样 → BS 经验样本 $Samp_1^{bs}$
  - B. 基于 AIC, BIC 等准则，筛选出基于 $Samp_1^{bs}$ 样本的最优模型，记为 $\mathcal{M}_1^{bs}$
  - C. 重复上述两步 $B = 1000$ 次，得到 $\{\mathcal{M}_b^{bs}\}, \ b = 1, 2, \cdots B$
  - D. 计算权重：

$$w_j = \frac{\#(\mathcal{M}_b^{bs} = \mathcal{M}_j)}{B}$$

Bayesian Model Averaging (BMA)

- 预先确定一组模型 (变量组合)
- 为每个模型设定一个先验概率
- 假设被解释变量的分布特征
- 估计后验概率
- 后验概率越大的模型获得权重越高



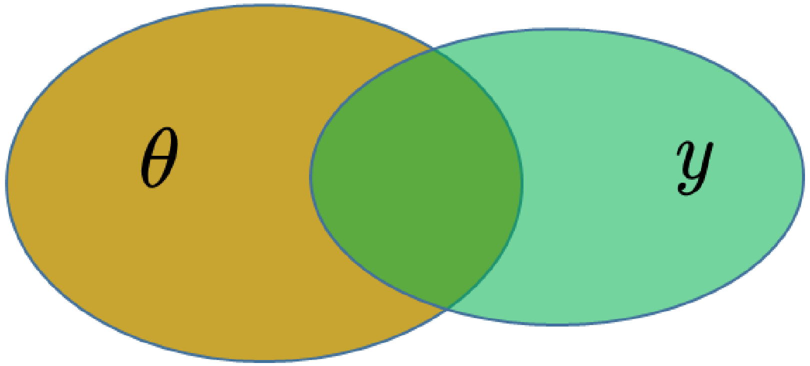
Bayes 定理

$$p(A \cap B) = p(A) \cdot p(B \mid A)$$
$$p(A \cap B) = p(B) \cdot p(A \mid B)$$

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B \mid A)}{p(B)}$$

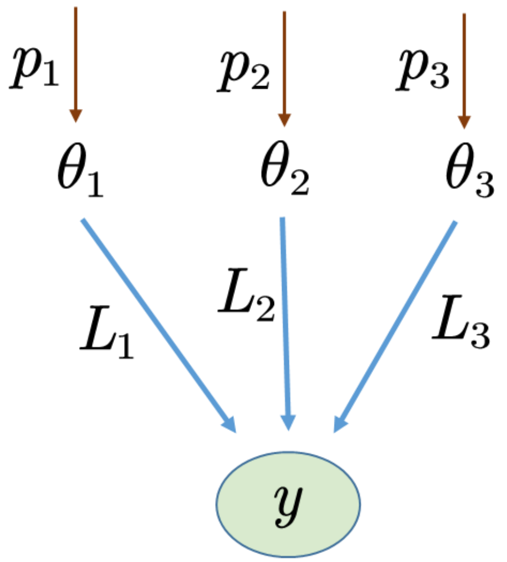
后验分布 (posterior)

$$p(\theta \mid y) = \frac{p(\theta) \cdot p(y \mid \theta)}{p(y)}$$



其中,

- $p(\theta)$: 先验分布 (prior), eg. $p(\theta_j) = 1/J$
- $p(y \mid \theta)$: y 的概率模型 (likelihood),
  - eg. $y_i \sim N(\mu, \sigma^2)$, 或 $y_i \sim U(\lambda)$
- $p(y)$: marginal probability of y
  - 若 θ 连续: $p(y) = \int p(y \mid \theta) \cdot p(\theta) \cdot d\theta$
  - 若 θ 离散: $p(y) = \sum p(\theta) \cdot p(y \mid \theta)$



$$L_j = p\,(y|\theta_j),\ j = 1, 2, 3$$

全概率公式:

$$\begin{aligned} p(y) &= p_1 L_1 + p_2 L_2 + p_3 L_3 \\ &= p(\theta_1)p(y|\theta_1) + p(\theta_2)p(y|\theta_2) + p(\theta_3)p(y|\theta_3) \end{aligned}$$

逆概率：贝叶斯法则

$$p(\theta_j \mid y) = \frac{p(\theta_j) \cdot p(y \mid \theta_j)}{\sum_{j=1}^{3} p(\theta_j) \cdot p(y \mid \theta_j)}$$

$$p(\theta_1 \mid y) = \underbrace{\frac{L_1}{p_1 L_1 + p_2 L_2 + p_3 L_3}}_{BayesFactor} \cdot p_1$$

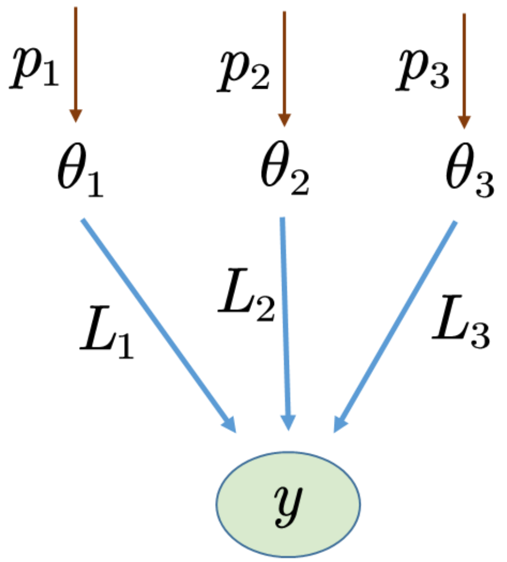
如何理解后验分布？

$$p(\theta_1 \mid y) = \underbrace{\frac{L_1}{p_1 L_1 + p_2 L_2 + p_3 L_3}}_{BayesFactor} \cdot p_1$$

> 相亲：有三个相亲对象，但对方信息一无所知

- 假设 1：相中概率 1/3 (先验概率)
- 假设 2：理想中对象 - 暖男、有担当 (先验分布)
- 新信息：见到对方 → 重新审视 → 后验概率
- e.g. $L_1 = 2,\ L_2 = 5,\ L_3 = 1$
    - $p(\theta_1 \mid y) = 2/8,\ p(\theta_2 \mid y) = 5/8,\ p(\theta_3 \mid y) = 1/8$
    - $BF_1 = \frac{2}{(1/3)\times 2+(1/3)\times 5+(1/3)\times 1} = 6/8 < 1$
    - $BF_2 = 15/8 > 1$; $\quad BF_2 = 3/8 < 1$

例子：求解 y 的均值



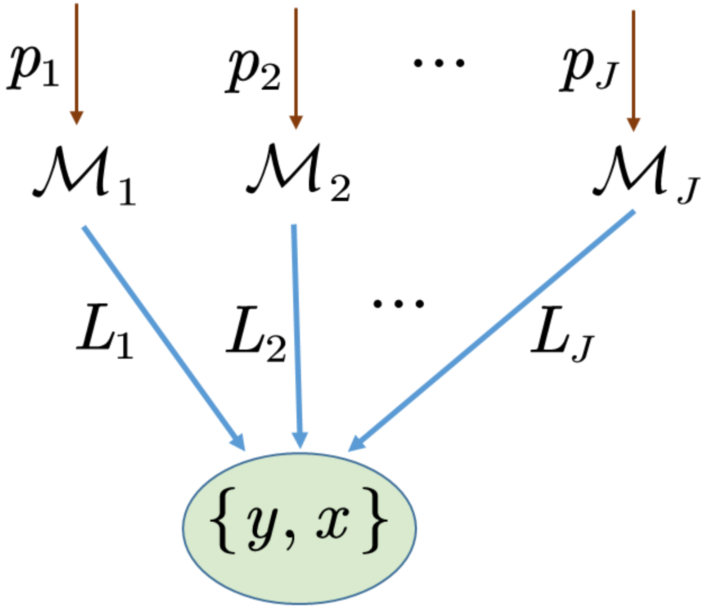
$$L_j = p(y|\theta_j),\ j=1,2,3$$

- Data： $y = \{1.1,\ 3.5,\ 1.2,\ 3.4,\ 2.6\}$ $(n=5)$
- 参数空间： $\theta_1 = 1,\ \theta_2 = 2,\ \theta_3 = 5$
  - 假设 1 - 先验概率： $p(\theta_j) = 1/3$
  - 假设 2 - 分布函数： $y_i \sim N(\mu, 1)$
    - Note: $(y_i - \mu)/1 \sim N(0,1)$
- $\widehat{L}_1 = p(y|\theta_1) = \prod_{i=1}^{5}(1/\sqrt{2\pi})\exp(-0.5(y_i-\theta_1)^2)$
- $\widehat{L}_2 = p(y|\theta_2), \widehat{L}_3 = p(y|\theta_3)$

$$\hat{p}(\theta_1 \mid y) = \frac{\widehat{L}_1}{(1/3)\widehat{L}_1 + (1/3)\widehat{L}_2 + (1/3)\widehat{L}_3} \cdot \frac{1}{3}$$

- Notes：
  - θ 可以在一个很大的范围内连续取值

BMA: 多个模型的平均化

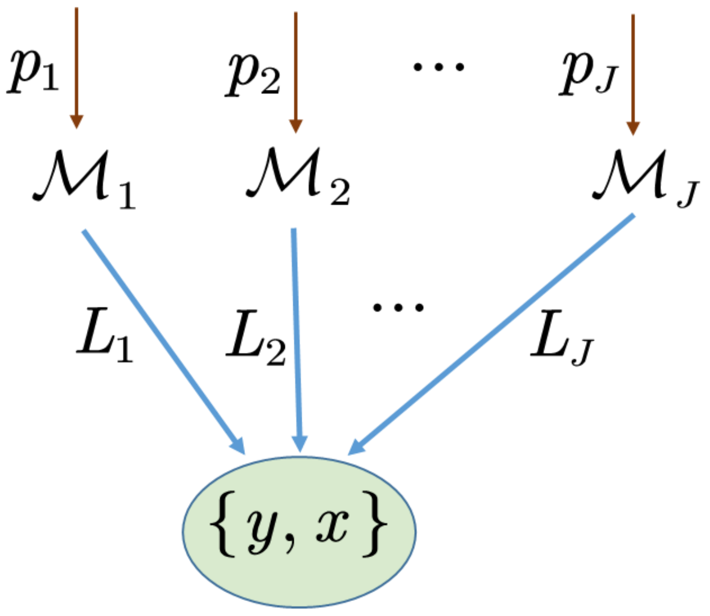


$$p_j = p(\mathcal{M}_j), j = 1, 2, \cdots, J$$

$$L_j = p(y \mid x, \mathcal{M}_j)$$

$$p(\mathcal{M}_j \mid y) = \frac{p(\mathcal{M}_j) \cdot p(y \mid \mathcal{M}_j)}{\sum_{j=1}^{J} p(\mathcal{M}_j) \cdot p(y \mid \mathcal{M}_j)} = \underbrace{\frac{L_1}{p_1 L_1 + p_2 L_2 + \cdots p_J L_J}}_{\text{BayesFactor}} \cdot p_j$$

- 需要预先设定：
  - $p(\mathcal{M}_j)$: 模型的先验概率, 如 $p(\mathcal{M}_j) = 1/J$
  - $p(y \mid \mathcal{M}_j)$: 参数的先验分布，如正态线性模型、Logit 模型等



$$p_j = p(\mathcal{M}_j), j = 1, 2, \cdots, J$$

$$L_j = p(y \mid x, \mathcal{M}_j)$$

BMA: 多个模型的平均化

$$p(\mathcal{M}_j \mid y) = \frac{L_1}{p_1 L_1 + p_2 L_2 + \cdots p_J L_J} \cdot p_j$$

- 需要预先设定：
  - 设 $p(\mathcal{M}_j) = 1/J$，即，等概率

$$p(\mathcal{M}_j \mid y) = \frac{L_1}{L_1 + L_2 + \cdots L_J}$$

BMA 估计量:

$$\mathrm{E}(\theta \mid y)=\sum_{m=1}^{M} p(m \mid y) \cdot \mathrm{E}(\theta \mid y, m)$$

> 对比:

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{w})=\sum_{m=1}^{M} w_m \widehat{\boldsymbol{\theta}}_m$$

MA 有何用?

- 参数估计 - 克服模型不确定性
  - 敏感性分析
  - 稳健性检验
- 变量筛选 - 变量相对重要性比较 (PIP-BMA; $\widehat{w}_i$ —Bootstrap MA)
  - 理论对比和验证
- 预测 / 补漏 →
  - 因果推断 $\hat{\tau} = E[(y_i - y_i^N)|T]$
  - 类似于 Lasso (CV - λ)

几个例子

MA：Stata 实操 (1)

```
sysuse "nlsw88.dta", clear

*-OLS
gen black = 2.race
qui reg wage union hours black tenure
keep if e(sample)
eststo OLS1: reg wage union hours
eststo OLS2: reg wage union hours black tenure
```

MA：Stata 实操 (2)

```
*-Model averaging
global cx "hours black tenure" // test variables

bma  wage union, auxiliary($cx)
est store bma

wals wage union, auxiliary($cx)
est store wals

miinc wage $cx, all(union) reg(regress) ic(aic)
est store miinc
```

MA：Stata 实操 (3)

```
local m "OLS1 OLS2 bma wals miinc"
esttab `m', mtitle(`m') nogap compress
```

```
----------------------------------------------------------------------------
                 (1)          (2)          (3)          (4)          (5)
                OLS1         OLS2          bma         wals        miinc
----------------------------------------------------------------------------
union          1.378***     1.193***     1.193***     1.220***     1.193***
              (6.26)       (5.58)       (5.58)       (5.71)       (5.58)
hours         0.0539***    0.0422***    0.0421***    0.0419***    0.0422***
              (5.66)       (4.56)       (4.49)       (4.53)       (4.55)
black                     -1.356***    -1.355***    -1.281***    -1.356***
                          (-6.55)      (-6.55)      (-6.19)      (-6.55)
tenure                     0.179***     0.179***     0.160***     0.179***
                          (10.88)      (10.88)       (9.76)      (10.88)
_cons          5.221***     4.891***     4.894***     5.000***     4.891***
             (14.07)      (13.54)      (13.40)      (13.85)      (13.53)
----------------------------------------------------------------------------
N               1867         1867         1867         1867         1867
----------------------------------------------------------------------------
t statistics in parentheses, * p<0.05, ** p<0.01, *** p<0.001
```

bma 结果演示

```
sysuse "nlsw88.dta", clear
tab industry, gen(dum)
global cx "hours black tenure dum*"

bma  wage union, auxiliary($cx)
```

```
Model space: 16384 models

BMA estimates                                    Number of obs =     1854
                                                 k1            =        2
                                                 k2            =       14
--------------------------------------------------------------------------
        wage |      Coef.   Std. Err.        t      pip   [1-Std. Err. Bands]
-------------+------------------------------------------------------------
       _cons |   6.357651   .5171292     12.29     1.00     5.840522  6.874781
       union |   .8269925   .2140942      3.86     1.00     .6128983  1.041087
-------------+------------------------------------------------------------
       hours |   .0333596   .0110856      3.01     0.96      .022274  .0444452
       black |  -1.232202   .2032808     -6.06     1.00    -1.435482 -1.028921
      tenure |   .1513812   .0163189      9.28     1.00     .1350623  .1677001
        dum1 |  -.3363967    .915286     -0.37     0.15    -1.251683  .5788893
        dum2 |  -.0105626   .4179666     -0.03     0.02    -.4285292   .407404
        dum3 |    .030671   .2399515      0.13     0.04    -.2092805  .2706225
        dum4 |  -1.303369   .4657277     -2.80     0.94    -1.769097 -.8376415
        dum5 |   1.944851   .5683615      3.42     0.99     1.376489  2.513212
        dum6 |  -2.347272   .4099492     -5.73     1.00    -2.757221 -1.937323
        dum7 |   .1341793   .3762751      0.36     0.14    -.2420959  .5104544
        dum8 |  -.1527826   .4327174     -0.35     0.14       -.5855  .2799348
        dum9 |  -3.180391   .5961395     -5.33     1.00    -3.776531 -2.584252
       dum10 |  -.0216761   .2216672     -0.10     0.03    -.2433433  .1999911
       dum11 |  -.8434916    .388299     -2.17     0.89    -1.231791 -.4551926
--------------------------------------------------------------------------
```

参考：

- Luca, Giuseppe De, Jan R. Magnus, 2011, Bayesian Model Averaging and Weighted-Average Least Squares: Equivariance, Stability, and Numerical Issues, **Stata Journal**, 11(4): 518–544. -PDF-， -cited-

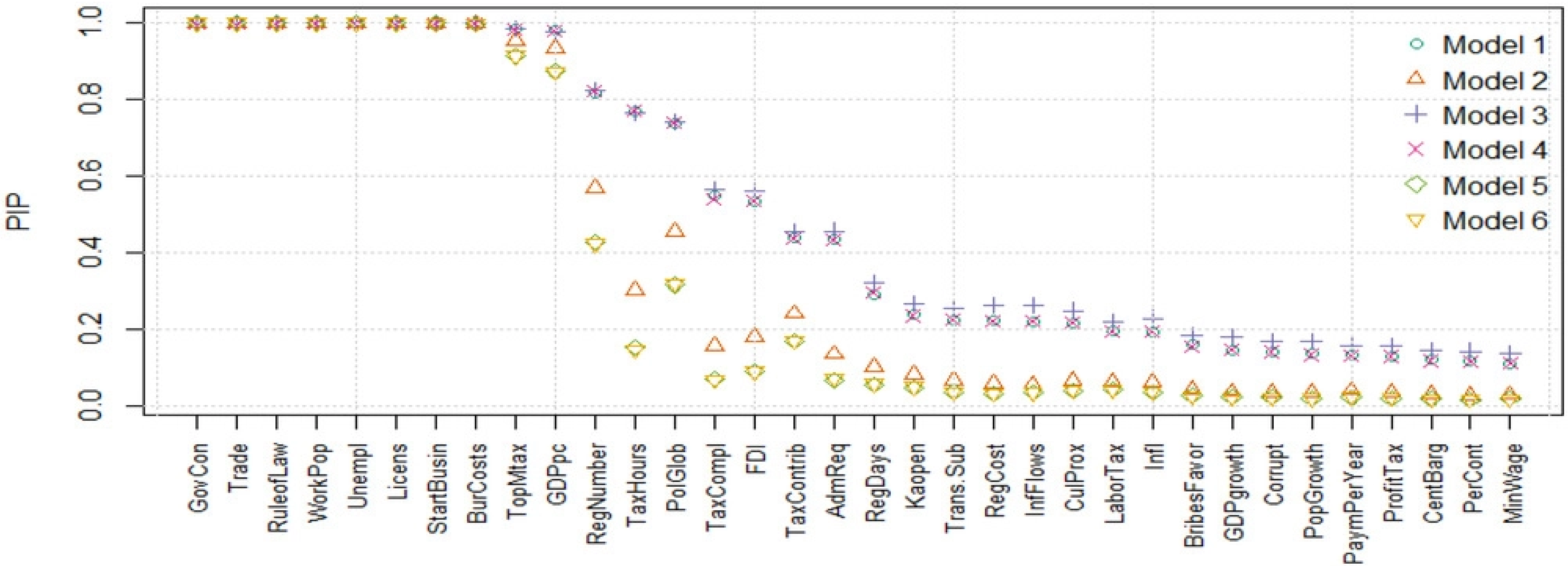
BMA 应用：地下经济的决定因素

Zhanabekov, S., **2022**, Robust determinants of the shadow economy, **Bulletin of Economic Research**: 1-36. -Link-, -PDF-

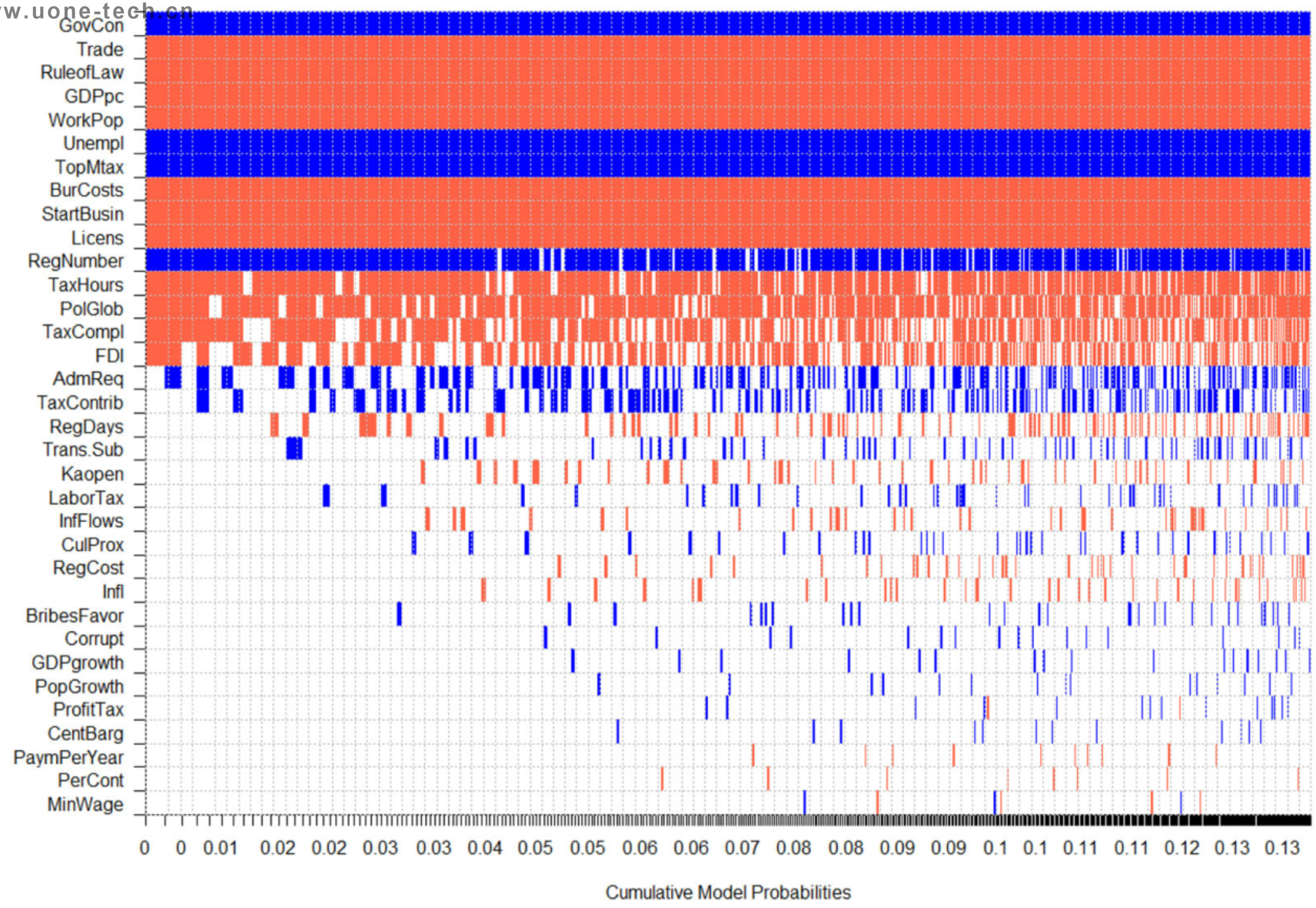
TABLE 4 Baseline model results

		Shadow1		
		PIP	**Post mean**	**Post SD**
1	GovCon	1.00000	0.51614***	0.04432
2	Trade	1.00000	−0.05673***	0.00649
3	RuleofLaw	1.00000	−5.54624***	0.60854
4	WorkPop	1.00000	−0.58846***	0.08395
5	Unempl	1.00000	0.20968***	0.03472
6	Licens	1.00000	−0.36236***	0.05658
7	StartBusin	0.99986	−0.72017***	0.15475
8	BurCosts	0.99934	−0.15682***	0.03605
9	TopMtax	0.98539	0.23852***	0.07357
10	GDPpc	0.98091	−0.00026**	0.00008
11	RegNumber	0.81982	0.18666	0.11618
12	TaxHours	0.76745	−0.00116	0.00085
13	PolGlob	0.73802	−0.03785	0.02864
14	TaxCompl	0.54494	−0.13438	0.15028
15	FDI	0.53152	−0.01792	0.02048

> Source: Zhanabekov (2022, PDF)



Source: Zhanabekov (2022, PDF)



Source: Zhanabekov (2022, PDF)

MA：参考资料

参考资料：参考书

- Hansen, B. E. Econometrics[M]. Princeton University Press, 2021. -Bruce E. Hansen 主页-， -PDF-, Chapter 28
- Fletcher, D. 2018, Model averaging, Springer Berlin Heidelberg, Berlin, Heidelberg. -Link-, -PDF1-, -PDF-
- Bayesian Model Averaging Home Page

参考资料：综述

- Moral-Benito, E. 2015. "Model averaging in economics: An overview". **Journal of Economic Surveys**, 29 (1): 46-75. Link, Link, PDF
- Steel, M. F. J., **2020**, Model averaging and its use in economics, **Journal of Economic Literature**, 58 (3): 644-719. -Link-, -PDF1-, -PDF2-, -cited-
- Nonejad, N., **2021**, An overview of dynamic model averaging techniques in time-series econometrics, **Journal of Economic Surveys**, 35 (2): 566-614. -Link-, -PDF-

- Luca, Giuseppe De, Jan R. Magnus, 2011, Bayesian Model Averaging and Weighted-Average Least Squares: Equivariance, Stability, and Numerical Issues, **Stata Journal**, 11(4): 518–544. -PDF- , -cited-
- Magnus, J. R., G. De Luca, **2016**, Weighted-average least squares (wals): A survey, **Journal of Economic Surveys**, 30 (1): 117-148. -Link-, -PDF-
- Hoeting, J. A., D. Madigan, A. E. Raftery, C. T. Volinsky, **1999**, Bayesian model averaging: A tutorial, **Statistical Science**, 14 (4): 382-417. -Link-, -PDF-
  - Hoeting, 2022, Methodology for Bayesian Model Averaging: An Update. PDF, 提供了一些软件实操方面的链接。
- Fragoso, Tiago M.; Bertoli, Wesley; Louzada, Francisco (2017). Bayesian Model Averaging: A Systematic Review and Conceptual Classification. International Statistical Review, -PDF-

- Schomaker, M., 2019, Model Averaging and Model Selection after Multiple Imputation using the R-package MAMI, -PDF-. 讲得很全面.
- BMA： Bayesian model averaging, Raftery et al., 2005, Fragoso et al., 2018.
- **Codes**
  - General Model Averaging Programs (Gauss, Matlab, R). Codes.
  - General Model Averaging Programs (Stata). Codes.
- Hansen, B. E., **2007**, Least squares model averaging, **Econometrica**, 75 (4): 1175-1189. -Link-, -PDF-, Replication
- Hansen, B. E., **2008**, Least-squares forecast averaging, **Journal of Econometrics**, 146 (2): 342-350. -Link-, -PDF-, Replication
- Kaplan - BMA, -Link-
- 连玉君 - Model-Averaging笔记

MA 应用：经济学 (1)

- Magnus, J. R., O. Powell, P. Prufer, **2010**, A comparison of two model averaging techniques with an application to growth empirics, **Journal of Econometrics**, 154 (2): 139-153. -Link-, -PDF- `经济增长`
- Camarero, M., S. Moliner, C. Tamarit, **2021**, Japan's FDI drivers in a time of financial uncertainty. New evidence based on bayesian model averaging, **Japan and the World Economy**, 57: 101058. -Link-, -PDF- `FDI`
- Butler, A. W., M. O. Keefe, R. Kieschnick, **2014**, Robust determinants of ipo underpricing and their implications for IPO research, **Journal of Corporate Finance**, 27: 367-383. -Link-, -PDF1- IPO 定价

MA 应用：经济学 (2)

- Enrique Moral-Benito, "Growth Empirics in Panel Data under Model Uncertainty and Weak Exogeneity", **Journal of Applied Econometrics**, 2016, 31 (3): 584-602. -Link-, -PDF-, -Replication-, -readme-
- Aller, C., L. Ductor, D. Grechyna, **2021**, Robust determinants of co2 emissions, **Energy Economics**, 96: 105154. -Link-, -PDF-

MA 应用：金融/会计

- Breuer, M., H. H. Schutt, **2021**, Accounting for uncertainty: An application of bayesian methods to accruals models, **Review of Accounting Studies**: 1-43. -Link-, -PDF-
- Zhanabekov, S., **2022**, Robust determinants of the shadow economy, **Bulletin of Economic Research**: 1-36. -Link-, -PDF-

MA 应用：中文

- 柯忠义, **2017**, 创业板上市公司经济绩效及影响因素——基于贝叶斯模型平均法 (bma) 的实证研究, **数量经济技术经济研究**, 34 (1): 146-160.
- 陈浪南,王升泉,陈文博.中国城乡居民收入差距决定因素的实证研究——基于贝叶斯模型平均(BMA)法[J]. **经济学报**, 2016, 3(03): 61-78.
- 陈伟, 牛霖琳. 基于贝叶斯模型平均方法的中国通货膨胀的建模及预测. **金融研究**, 2013(11): 15-27.

MA：新进展

- Zhu, R., X. Y. Zhang, A. T. K. Wan, G. H. Zou, **2021**, Kernel averaging estimators, **Journal of Business & Economic Statistics**: 1-13. -Link-, -PDF-
- Sun, Y. Y., Y. M. Hong, T. H. Lee, S. Y. Wang,X. Y. Zhang, **2021**, Time-varying model averaging?, **Journal of Econometrics**, 222 (2): 974-992. -Link-, -PDF-
- Nonejad, N., **2021**, An overview of dynamic model averaging techniques in time-series econometrics, **Journal of Economic Surveys**, 35 (2): 566-614. -Link-, -PDF-

MA：新进展

- Lu, X., K. Miao, L. J. Su, **2021**, Determination of different types of fixed effects in three-dimensional panels*, **Econometric Reviews**, 40 (9): 867-904. -Link-, -PDF-
- Feng, Y., Q. F. Liu, Q. S. Yao, G. Q. Zhao, **2021**, Model averaging for nonlinear regression models, **Journal of Business & Economic Statistics**: 1-14. -Link-, -PDF-