

附录

正文未报告部分

A. 测算框架的详细说明

A. 1. 单个变量解释效力的测算

给定推测函数 \hat{f} , 我们希望衡量某个具体的解释变量 x_{ki} 的样本变动, 最终带来了多少样本内推测值 \hat{y}_i 的变动, 从而得到该变量对推测值内样本内变动的贡献份额。这里的难点, 在于如何将 x_{ki} 样本变动的同时, 其他变量 $x_{ji}, j \neq k$ 样本变动带来的推测值 \hat{y}_i 变动, 纳入考虑。在通常的“加法”思维下, 我们希望固定住其他变量, 只让 x_{ki} 变化并观察推测值 $\hat{f}(\dots, x_{ki}, \dots)$ 出现了多少变化。但这样一来, 我们所衡量的 x_{ki} 对推测值变动的贡献, 本质上都是“条件”评估——测算结果取决于其他变量被固定住的取值水平; 与此同时, 给定其他变量样本取值的多样性, 并没有任何一组特定的取值具有明显的“合理性”。这样一来, 上述条件评估本身又变成一个高维度的对象, 从而失去了对单一变量解释力贡献测度的简明性。与此相反, 我们可以采取一个“减法”思维, 直接获得一个“无条件”测度。具体而言, 我们定义推测函数 \hat{f} 关于变量 x_k 的平均可解释预测变动 (average explained prediction variation) 如下:

$$AEPV(\hat{f}; k) = \frac{1}{N} \sum_{i=1}^N \left(\hat{f}(x_i) - \frac{1}{N} \sum_{j=1}^N \hat{f}(x_{1i}, \dots, x_{k-1,i}, x_{kj}, x_{k+1,i}, \dots, x_{Ki}) \right)^2. \quad (A1)$$

上式中的第 i 个求和项解释如下: 首先, 给定样本点 x_i 及其对应的推测值 $\hat{f}(x_i) = \hat{f}(x_{1i}, \dots, x_{ki}, \dots, x_{Ki})$, 固定 x_k 之外的变量样本取值不变, 但让 x_k 取遍样本取值 $\{x_{k1}, \dots, x_{kN}\}$ 并计算对应推测值的平均, 则这一平均推测值反映了除 x_k 之外所有其余变量对该样本点处推测值的贡献, 同时仅保留 x_k 自身在整个样本取值范围内预测贡献的平均说明; 其次, $\hat{f}(x_i)$ 与该预测均值的离差, 刻画了该样本点处, x_k 取对应样本值 x_{ki} 时, 推测值的改进程度, 从而反映了变量 x_k 在该样本点对于推测值的贡献大小; 最后, 将所有样本点处的离差取平方和再取平均, 即得到变量 x_k 在所有样本点处, 对样本内推测值变动的总体平均解释度。下面第 A.4 小节, 我们会进一步用线性推测函数的例子, 说明上述公式的意义。

我们可以进一步定义平均预测变动 (average prediction variation) 如下:

$$APV(\hat{f}) = \frac{1}{N} \sum_{i=1}^N \left(\hat{f}(x_i) - \frac{1}{N} \sum_{j=1}^N \hat{f}(x_j) \right)^2, \quad (A2)$$

换言之, $APV(\hat{f})$ 就是样本内推测值的方差。在此基础上, 我们定义推测函数 \hat{f} 关于变量 x_k 的可解释预测变动占比 (fraction of EPV):

$$FEPV(\hat{f}; k) = \frac{AEPV(\hat{f}; k)}{APV(\hat{f})}. \quad (A3)$$

这一占比变量, 简洁的反映了 x_k 对推测函数 \hat{f} 样本内预测的贡献比例。

A. 2. 两个变量联合解释效力的测算

有了单变量情形作为铺垫, 我们下面讨论两个变量的情形。首先, 沿用 $AEPV(\hat{f}; \cdot)$ 的记号, 我们定义两个变量 x_k, x_m 的联合可解释预测变动如下:

$$AEPV(\hat{f}; k, m) = \frac{1}{N} \sum_{i=1}^N \left(\hat{f}(x_i) - \frac{1}{N} \sum_{j=1}^N \hat{f}(\dots, x_{kj}, \dots, x_{mj}, \dots) \right)^2, \quad (A4)$$

上式求和项中 $\hat{f}(\dots, x_{kj}, \dots, x_{mj}, \dots)$ 的省略号表示除 x_k, x_m 外所有其他变量均为样本点 i 处的取值。这里定义的 $AEPV(\hat{f}; k, m)$ 包括了 x_k, x_m 两个变量各自对推测值的独立贡献, 以及两个

变量对推测值的交互贡献。为测量这后一部分预测贡献, 我们可以进一步定义交互可解释预测变动 (interactive EPV):

$$IEPV(\hat{f}; k, m) = AEPV(\hat{f}; k, m) - AEPV(\hat{f}; k) - AEPV(\hat{f}; m). \quad (A5)$$

同样, 我们将在第 A.4 节中用线性推测函数为例, 进一步说明 $AEPV(\hat{f}; k, m)$ 及 $IEPV(\hat{f}; k, m)$ 的涵义。与单变量情形类似, 可以计算上述两个特征量与 $AEPV(\hat{f})$ 的比例, 用来简洁测度两个变量的联合解释贡献占比与交互解释贡献占比:

$$FEPV(\hat{f}; k, m) = \frac{AEPV(\hat{f}; k, m)}{APV(\hat{f})}, \quad FIEPV(\hat{f}; k) = \frac{IEPV(\hat{f}; k, m)}{APV(\hat{f})}. \quad (A6)$$

A. 3. 推测函数非线性性的度量

正如前文所述, 机器学习方法较传统计量方法——特别是以线性回归为基础的线性预测方法——的一个突出优势是可以灵活的捕捉、处理数据样本中存在的非线性性。但如何刻画机器学习方法所生成推测函数的非线性性, 却并没有一套现成的技术。不过, 我们下面马上会说明, 在上述变量解释力测算框架下, 我们可以很自然的度量推测函数的非线性性。

参照 A.1 小节, 我们可以将 $\Delta\hat{y}_{[k]i} \equiv \hat{f}(x_i) - \frac{1}{N} \sum_j \hat{f}(\dots, x_{kj}, \dots)$ 视作变量样本值 x_{ki} 对推测值 \hat{y}_i 的预测贡献。若以 $\Delta\hat{y}_{[k]i}$ 为分析对象, 并将其视作仅与 x_k 有关, 那么我们可以进一步将 x_k 对目标变量的预测效果——即 $\Delta\hat{y}_{[k]i}$ 体现的部分——分解为线性趋势项与非线性项两个部分。为此, 我们只需要用 $\Delta\hat{y}_{[k]i}$ 对 x_k 进行如下的辅助性线性回归:

$$\Delta\hat{y}_{[k]i} = \alpha + \beta_k x_{ki} + \xi_i, \quad i = 1, \dots, N, \quad (A7)$$

进而可以将回归所得的线性部分 $\alpha + \beta_k x_{ki}$ 解释为 x_k 预测贡献中的线性趋势项, 残差项 ξ_{ki} 则可视为相应的非线性部分。同时, 该回归的 R^2 自然可以解释为 x_k 解释能力中线性趋势部分的比例, 而 $1 - R^2$ 则表示非线性变动部分的比例。

类似的, 我们还可以衡量多个变量的联合非线性性。以两个变量为例说明, 我们仅需定义 $\Delta\hat{y}_{[km]i} \equiv \hat{f}(x_i) - \frac{1}{N} \sum_j \hat{f}(\dots, x_{kj}, \dots, x_{mj}, \dots)$, 并通过 2-元辅助回归

$$\Delta\hat{y}_{[km]i} = \alpha + \beta_k x_{ki} + \beta_m x_{mi} + \xi_i, \quad i = 1, \dots, N$$

进行分析即可。

A. 4. 测算框架的直观说明: 三个示例

我们首先考察线性推测函数的例子。假设推测函数由如下形式:

$$\hat{f}(x_i) = \beta^\top x_i = \beta_1 x_{1i} + \dots + \beta_K x_{Ki}.$$

直接计算可知, $\Delta\hat{y}_{[k]i} = \hat{f}(x_i) - \frac{1}{N} \sum_j \hat{f}(\dots, x_{kj}, \dots) = \beta_k x_{ki} - \beta_k \bar{x}_k$, 其中 \bar{x}_k 表示样本均值。因此, (1)式定义的 $AEPV(\hat{f}; k) = \frac{\beta_k^2}{N} \sum_i (x_{ki} - \bar{x}_k)^2$, 即线性回归中变量 x_k 的可解释样本方差 (explained sample variance)。类似的, 两个变量 x_k, x_m 联合时, $AEPV(\hat{f}; k, m) = \frac{1}{N} \sum_i (\beta_k (x_{ki} - \bar{x}_k) + \beta_m (x_{mi} - \bar{x}_m))^2$; 此时存在两个变量交互作用对预测的贡献, 相应的 $IEPV(\hat{f}; k, m) = \frac{2\beta_k \beta_m}{N} \sum_i (x_{ki} - \bar{x}_k)(x_{mi} - \bar{x}_m)$ 。此外, 由于 $\Delta\hat{y}_{[k]i}$ 此时是线性依赖于 x_{ki} , 故非线性性分解所用的辅助回归 (7) 可实现完美拟合, $R^2 = 1$; 同时 $\Delta\hat{y}_{[k]i} = \beta_k x_{ki} - \beta_k \bar{x}_k$ 也说明辅助回归方程 (7) 中包括常数项的必要性。

接下来, 我们考虑一般的可加性推测函数:

$$\hat{f}(x_i) = g_1(x_{1i}) + \dots + g_K(x_{Ki}),$$

上式中各个 $g_k(\cdot)$ 均假设为非线性函数。直接计算可知, $\Delta\hat{y}_{[k]i} = g_k(x_{ki}) - \bar{g}_k$, 其中 \bar{g}_k 表示 $g_k(x_{ki})$ 的样本均值。此时, 单变量或双变量下 $AEPV$ 和 $IEPV$ 的表达式与线性推测函数时类似, 不再赘述。但此时 $\Delta\hat{y}_{[k]i}$ 不再是线性的依赖于 x_{ki} , 因此辅助回归 (7) 将通过 $R^2 < 1$ 捕捉到 $g_k(\cdot)$ 反映的非线性特征。

最后, 我们考虑一个含有交叉项的 (拟) 线性预测模型:

$$\hat{f}(x_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \lambda x_{1i} x_{2i} + \beta_3 x_{3i} + \dots + \beta_K x_{Ki}.$$

此时, 对应的

$$\begin{aligned}
\Delta \hat{y}_{[1]i} &= (x_{1i} - \bar{x}_1)(\beta_1 + \lambda x_{2i}), \\
\Delta \hat{y}_{[2]i} &= (x_{2i} - \bar{x}_2)(\beta_2 + \lambda x_{1i}), \\
\Delta \hat{y}_{[12]i} &= \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \lambda(x_{1i}x_{2i} - \overline{x_1x_2}) \\
&= (x_{1i} - \bar{x}_1)(\beta_1 + \lambda x_{2i}) + (x_{2i} - \bar{x}_2)(\beta_2 + \lambda x_{1i}) \\
&\quad + \lambda(\bar{x}_1x_{2i} + x_{1i}\bar{x}_2 - x_{1i}x_{2i} - \overline{x_1x_2}),
\end{aligned}$$

其中 $\overline{x_1x_2}$ 表示 $x_{1i}x_{2i}$ 的样本均值。由此可算得

$$\begin{aligned}
IEPV(\hat{f}; 1, 2) &= AEPV(\hat{f}; 1, 2) - AEPV(\hat{f}; 1) - AEPV(\hat{f}; 2) \\
&= \frac{\lambda^2}{N} \sum_i (\bar{x}_1x_{2i} + x_{1i}\bar{x}_2 - x_{1i}x_{2i} - \overline{x_1x_2})^2 \\
&\quad + \frac{2\lambda}{N} \sum_i (x_{1i} - \bar{x}_1)(\beta_1 + \lambda x_{2i})(\bar{x}_1x_{2i} + x_{1i}\bar{x}_2 - x_{1i}x_{2i} - \overline{x_1x_2}) \\
&\quad + \frac{2\lambda}{N} \sum_i (x_{2i} - \bar{x}_2)(\beta_2 + \lambda x_{1i})(\bar{x}_1x_{2i} + x_{1i}\bar{x}_2 - x_{1i}x_{2i} - \overline{x_1x_2}) \\
&\quad + \frac{2}{N} \sum_i (x_{1i} - \bar{x}_1)(\beta_1 + \lambda x_{2i})(x_{2i} - \bar{x}_2)(\beta_2 + \lambda x_{1i}).
\end{aligned}$$

B. 完整变量列表、计算程序及信息

表 B1: 完整样本变量

变量代码	释义	分组	变量代码	释义	分组
GR6096	1960-96 年人均 GDP 增速	Na	LIFE060	1960 年预期寿命	2
ABSLATIT	纬度绝对值	4	LT100CR	通航水域土地占比	7
AIRDIST	距大城市的空中距离	7	MALFAL66	60 年代疟疾虚拟变量	4
<u>AVELF</u>	民族语言分化	Na	<u>MINING</u>	采矿业的 GDP 比例	11
BRIT	英国殖民地虚拟变量	10	<u>MUSLIM00</u>	穆斯林比例	6
<u>BUDDHA</u>	佛教徒比例	6	NEWSTATE	独立时间	10
CATH00	天主教徒比例	6	OIL	石油生产国虚拟变量	11
CIV72	公民自由	8	OPENDEC1	1965-74 开放度	3
COLONY	殖民地虚拟变量	10	ORTH00	东正教比例	6
CONFUC	儒教比例	6	OTHFRAC	外语人口比例	na
DENS60	人口密度 1960	7	P60	1960 小学教育	1
DENS65C	60 年代沿海人口密度	7	PI6090	1960-90 通胀率	3
DENS65I	60 年代内陆人口密度	7	SQPI6090	1960-90 通胀率平方	3
DPOP6090	1960-90 人口增长率	1	PRIGHTS	政治权利	8
EAST	东亚虚拟变量	5	POP1560	15 岁以下人口比例	2
ECORG	资本主义	8	POP60	1960 年人口	7
ENGFRAC	英语人口	na	POP6560	65 岁以上人口比例	2
EUROPE	欧洲虚拟变量	5	PRIEXP70	1970 年初级产品出口	11
FERTLDC1	60 年代的生育率	2	PROT00	新教徒比例	6
GDE1	国防开支	3	RERD	实际汇率扭曲	3
GDPCH60L	60 年人均 GDP (log)	1	REVCoup	革命和政变	9
GEEREC1	60 年代公共教育支出占比	1	SAFRICA	非洲虚拟变量	5
GGCFD3	公共投资份额	3	SCOUT	经济外向性	3
GOVNOM1	60 年代政府名义支出占比	3	SIZE60	经济规模	7
GOVSH61	60 年代政府实际支出占比	3	SOCIALIST	社会主义虚拟变量	8
<u>GVR61</u>	60 年代政府消费占比	3	<u>SPAIN</u>	西班牙殖民地虚拟变量	10
H60	1960 高等教育水平	1	TOT1DEC1	60 年代的贸易增长	12
HERF00	信教程度	6	TOTIND	贸易条件排名	12
HINDU00	印度教徒比例	6	TROPICAR	热带地区面积占比	4
IPRICE1	投资品价格	3	TROPPOP	热带地区人口占比	4
<u>LAAM</u>	拉丁美洲虚拟变量	5	WARTIME	1960-90 战争花费比例	9
LANDAREA	土地面积	7	WARTORN	1960-90 战争虚拟变量	9
LANDLOCK	内陆国家虚拟变量	7	<u>YRSOPEN</u>	1950-94 年开放年数	3
LHCPC	油气储量 1993	11	ZTROPICS	热带气候虚拟变量	4

注: 数据来源为 Sala-i-Martin et al. (2004); 部分变量无分组, 记为 na; 加粗变量为 Sala-i-Martin et al. 识别出的前十显著变量, 加下划线变量表示非前十但显著的变量。

表 B2: 计算程序基本信息

算法	MATLAB 命令	随机性	Bootstrap 数量	计算时间
LASSO	lasso	否		约 30 秒
回归树	fitrtree	否		约 30 秒
套袋法	TreeBagger	是	10,000	约 1800 秒
随机森林	TreeBagger	是	10,000	约 1000 秒
神经网络	feedforwardnet	是	10,000	约 1500 秒
支持向量回归	fitrsvm	否		约 30 秒

弹性网络	lasso	否		约 30 秒
最小二乘提升法	fitrensemble	否		约 30 秒
M5P-BG	m5pparamsensemble	是	10,000	约 1800 秒
M5P-RF	m5pparamsensemble	是	10,000	约 1500 秒

注: 计算时间为使用 i7 8700K 芯片配合 32G DDR4 内存所得出的时间; 程序经过并行优化, 调用 6 个逻辑核心; 因为硬件条件差异该项指标会有较大变化

上表中所汇报的时间成本仅指计算一次推测函数值及变量排序所用的时间。测算推测函数交互效应及非线性时, 因为需要对所有变量及其两两组合, 在所有样本点上反复进行推测函数的赋值, 而每次赋值都需要重新运行一遍整个算法, 因此累计计算成本很高。当 Bootstrap 抽样次数取 10,000 时, 全套分析程序在 6 核并行设定下, 运行一遍约需 4 周。

C. 推测函数关键特征度量

C. 1. 单一变量预测解释力度结果

表 C1: 常规方法单一变量解释力度排序

LASSO		RT		BG		RF		NN	
P60	0.04	MALFAL66	0.15	MALFAL66	0.02	MALFAL66	0.01	EAST	0.02
EAST	0.02	BUDDHA	0.13	BUDDHA	0.02	BUDDHA	0.01	SAFRICA	0.01
IPRICE1	0.02	ABSLATIT	0.07	EAST	0.01	LIFE060	0.01	P60	0.01
CONFUC	0.02	LANDAREA	0.04	LIFE060	0.01	EAST	0.01	YRSOPEN	0.01
TROPICAR	0.01	GDE1	0.03	P60	0.01	P60	0.01	BRIT	0.01
BUDDHA	0.01	IPRICE1	0.02	ABSLATIT	0.01	YRSOPEN	0.00	MALFAL66	0.01
YRSOPEN	0.01	AIRDIST	0.01	YRSOPEN	0.00	ABSLATIT	0.00	COLONY	0.00
MINING	0.01	SIZE60	0.01	CONFUC	0.00	CONFUC	0.00	CIV72	0.00
GDPCH60L	0.01	GVR61	0.01	DENS60	0.00	H60	0.00	LIFE060	0.00
RERD	0.01	POP6560	0.01	IPRICE1	0.00	DENS60	0.00	PROT00	0.00
SPAIN	0.00	RERD	0.01	AIRDIST	0.00	DENS65C	0.00	TROPICAR	0.00
DENS65C	0.00	GDPCH60L	0.00	TROPPOP	0.00	TROPPOP	0.00	OTHFRAC	0.00
MALFAL66	0.00	DENS60	0.00	RERD	0.00	RERD	0.00	SPAIN	0.00
GVR61	0.00	CATH00	0.00	LANDAREA	0.00	IPRICE1	0.00	IPRICE1	0.00
MUSLIM00	0.00	AVELF	0.00	DENS65C	0.00	TROPICAR	0.00	GDPCH60L	0.00
OTHFRAC	0.00	BRIT	0.00	TOTIND	0.00	SAFRICA	0.00	AVELF	0.00
REVCoup	0.00	CIV72	0.00	POP6560	0.00	TOTIND	0.00	RERD	0.00
GGCFD3	0.00	COLONY	0.00	H60	0.00	FERTLDC1	0.00	TROPPOP	0.00
AVELF	0.00	CONFUC	0.00	TROPICAR	0.00	AIRDIST	0.00	CATH00	0.00
SAFRICA	0.00	DENS65C	0.00	GDPCH60L	0.00	ZTROPICS	0.00	LAAM	0.00
GOVNOM1	0.00	DENS65I	0.00	DENS65I	0.00	GDPCH60L	0.00	MINING	0.00
HINDU00	0.00	DPOP6090	0.00	ZTROPICS	0.00	DENS65I	0.00	NEWSTATE	0.00
CIV72	0.00	EAST	0.00	OPENDEC1	0.00	LANDAREA	0.00	BUDDHA	0.00
OPENDEC1	0.00	ECORG	0.00	LHCPC	0.00	AVELF	0.00	PRIEXP70	0.00
TROPPOP	0.00	ENGFRAC	0.00	FERTLDC1	0.00	PRIGHTS	0.00	CONFUC	0.00
PRIEXP70	0.00	EUROPE	0.00	HERF00	0.00	POP6560	0.00	SOCIALIST	0.00
POP60	0.00	FERTLDC1	0.00	GDE1	0.00	CATH00	0.00	DENS60	0.00
SCOUT	0.00	GEEREC1	0.00	SIZE60	0.00	GDE1	0.00	WARTORN	0.00
PROT00	0.00	GGCFD3	0.00	PRIGHTS	0.00	PRIEXP70	0.00	POP1560	0.00
GEEREC1	0.00	GOVNOM1	0.00	SPAIN	0.00	SPAIN	0.00	PRIGHTS	0.00
ABSLATIT	0.00	GOVSH61	0.00	SAFRICA	0.00	LHCPC	0.00	REVCoup	0.00
AIRDIST	0.00	H60	0.00	GGCFD3	0.00	LAAM	0.00	GGCFD3	0.00
BRIT	0.00	HERF00	0.00	PROT00	0.00	HERF00	0.00	GOVNOM1	0.00
CATH00	0.00	HINDU00	0.00	CATH00	0.00	GEEREC1	0.00	SCOUT	0.00
COLONY	0.00	LAAM	0.00	GEEREC1	0.00	PROT00	0.00	LT100CR	0.00
DENS60	0.00	LANDLOCK	0.00	LT100CR	0.00	OPENDEC1	0.00	GVR61	0.00
DENS65I	0.00	LHCPC	0.00	GOVSH61	0.00	SIZE60	0.00	LHCPC	0.00
DPOP6090	0.00	LIFE060	0.00	REVCoup	0.00	GOVSH61	0.00	GOVSH61	0.00
ECORG	0.00	LT100CR	0.00	AVELF	0.00	GVR61	0.00	LANDLOCK	0.00
ENGFRAC	0.00	MINING	0.00	PRIEXP70	0.00	LT100CR	0.00	DENS65C	0.00
EUROPE	0.00	MUSLIM00	0.00	CIV72	0.00	GGCFD3	0.00	OPENDEC1	0.00
FERTLDC1	0.00	NEWSTATE	0.00	LAAM	0.00	REVCoup	0.00	ECORG	0.00
GDE1	0.00	OIL	0.00	DPOP6090	0.00	DPOP6090	0.00	OIL	0.00
GOVSH61	0.00	OPENDEC1	0.00	GOVNOM1	0.00	CIV72	0.00	DPOP6090	0.00

H60	0.00	ORTH00	0.00	POP1560	0.00	POP1560	0.00	SIZE60	0.00
HERF00	0.00	OTHFRAC	0.00	PI6090	0.00	GOVNOM1	0.00	ZTROPICS	0.00
LAAM	0.00	P60	0.00	ENGFRAC	0.00	MUSLIM00	0.00	TOT1DEC1	0.00
LANDAREA	0.00	PI6090	0.00	GVR61	0.00	POP60	0.00	FERTLDC1	0.00
LANDLOCK	0.00	SQPI6090	0.00	MUSLIM00	0.00	PI6090	0.00	ENGFRAC	0.00
LHCPC	0.00	PRIGHTS	0.00	POP60	0.00	SQPI6090	0.00	TOTIND	0.00
LIFE060	0.00	POP1560	0.00	TOT1DEC1	0.00	TOT1DEC1	0.00	H60	0.00
LT100CR	0.00	POP60	0.00	MINING	0.00	MINING	0.00	GDE1	0.00
NEWSTATE	0.00	PRIEXP70	0.00	OTHFRAC	0.00	COLONY	0.00	POP60	0.00
OIL	0.00	PROT00	0.00	ECORG	0.00	ENGFRAC	0.00	GEEREC1	0.00
ORTH00	0.00	REVCoup	0.00	WARTIME	0.00	OTHFRAC	0.00	PI6090	0.00
PI6090	0.00	SAFRICA	0.00	NEWSTATE	0.00	HINDU00	0.00	AIRDIST	0.00
SQPI6090	0.00	SCOUT	0.00	HINDU00	0.00	WARTIME	0.00	EUROPE	0.00
PRIGHTS	0.00	SOCIALIST	0.00	COLONY	0.00	ECORG	0.00	LANDAREA	0.00
POP1560	0.00	SPAIN	0.00	SCOUT	0.00	NEWSTATE	0.00	HERF00	0.00
POP6560	0.00	TOT1DEC1	0.00	WARTORN	0.00	WARTORN	0.00	MUSLIM00	0.00
SIZE60	0.00	TOTIND	0.00	BRIT	0.00	SCOUT	0.00	POP6560	0.00
SOCIALIST	0.00	TROPICAR	0.00	LANDLOCK	0.00	BRIT	0.00	WARTIME	0.00
TOT1DEC1	0.00	TROPPOP	0.00	EUROPE	0.00	LANDLOCK	0.00	ABSLATIT	0.00
TOTIND	0.00	WARTIME	0.00	ORTH00	0.00	EUROPE	0.00	HINDU00	0.00
WARTIME	0.00	WARTORN	0.00	SOCIALIST	0.00	SOCIALIST	0.00	SQPI6090	0.00
WARTORN	0.00	YRSOPEN	0.00	OIL	0.00	ORTH00	0.00	DENS65I	0.00
ZTROPICS	0.00	ZTROPICS	0.00	SQPI6090	0.00	OIL	0.00	ORTH00	0.00

注: 每种方法对应变量按照解释力度从大到小排序; 每个变量右侧数字为百分比解释力度, 单位为百分比

C. 2. 单一变量预测非线性性测度结果

表 C2: 常规方法单一变量预测非线性性排序

LASSO		RT		BG		RF		NN	
IPRICE1	0.00	BUDDHA	24.68	EAST	4.53	EAST	4.10	CONFUC	0.16
P60	0.00	AIRDIST	71.61	SPAIN	6.71	SPAIN	7.14	GDPCH60L	0.50
GDPCH60L	0.00	MALFAL66	72.65	ENGFRAC	9.93	ENGFRAC	10.78	H60	1.02
YRSOPEN	0.00	GDPCH60L	84.08	SAFRICA	10.44	SAFRICA	11.17	EAST	1.10
RERD	0.00	EAST	84.75	TROPPOP	19.76	LAAM	15.28	DENS65C	1.13
MINING	0.00	SIZE60	85.36	LAAM	19.78	TROPPOP	16.99	POP60	1.19
GVR61	0.00	CONFUC	85.47	BUDDHA	23.48	YRSOPEN	21.87	SAFRICA	1.22
DENS65C	0.00	DENS65C	88.79	YRSOPEN	23.77	BUDDHA	24.74	P60	1.23
TROPICAR	0.00	ENGFRAC	90.68	TROPICAR	25.32	RERD	24.99	PROT00	1.52
EAST	0.00	SCOUT	93.78	COLONY	25.72	P60	25.06	LIFE060	1.61
CONFUC	0.00	GDE1	93.88	RERD	27.45	TROPICAR	25.64	RERD	1.67
BUDDHA	0.00	DENS60	94.40	CONFUC	28.86	MALFAL66	26.89	ORTH00	1.69
AVELF	0.00	CATH00	94.57	MALFAL66	29.15	CONFUC	28.07	POP1560	1.70
GGCFD3	0.00	PROT00	95.00	P60	29.42	COLONY	28.36	YRSOPEN	1.89
OTHFRAC	0.00	DENS65I	95.60	SOCIALIST	30.11	PRIEXP70	32.22	CIV72	1.91
MUSLIM00	0.00	GVR61	95.97	EUROPE	39.96	NEWSTATE	33.03	BUDDHA	2.06
MALFAL66	0.00	RERD	95.97	NEWSTATE	43.40	SOCIALIST	34.10	SIZE60	2.44
SPAIN	0.00	EUROPE	96.22	ABSLATIT	44.12	AVELF	37.38	TROPICAR	2.66
REVCoup	0.00	TOT1DEC1	96.32	CIV72	44.46	REVCoup	42.09	IPRICE1	2.84
GOVNOM1	0.00	COLONY	96.87	REVCoup	45.70	GVR61	42.36	AVELF	2.85
OPENDEC1	0.00	MINING	97.06	LIFE060	47.02	CIV72	42.70	COLONY	3.00
CIV72	0.00	LAAM	97.26	AIRDIST	47.19	ABSLATIT	42.92	SPAIN	3.54

PRIEXP70	0.00	ZTROPICS	97.27	AVELF	48.58	PROT00	44.06	REVC0UP	3.60
POP60	0.00	OPENDEC1	97.54	PRIEXP70	49.39	LIFE060	46.98	OPENDEC1	4.15
HINDU00	0.00	H60	97.63	ZTROPICS	50.25	ZTROPICS	49.20	GOVNOM1	4.38
SAFRICA	0.00	POP6560	97.64	HINDU00	50.31	AIRDIST	49.71	MALFAL66	4.55
TROPPOP	0.00	SPAIN	97.67	ECORG	52.97	ECORG	52.12	PRIEXP70	5.26
SCOUT	0.00	OIL	97.90	PROT00	53.67	DENS65C	58.32	TOTIND	6.05
PROT00	0.00	LHCPC	97.90	PRIGHTS	57.90	DPOP6090	58.88	CATH00	6.06
GEEREC1	0.00	TOTIND	98.08	DPOP6090	57.99	PRIGHTS	60.39	GOVSH61	6.58
ZTROPICS	95.45	HINDU00	98.20	MUSLIM00	65.77	GOVSH61	62.23	DPOP6090	6.61
BRIT	96.77	PRIGHTS	98.20	SCOUT	66.85	IPRICE1	62.84	LAAM	6.88
SIZE60	97.55	PRIEXP70	98.35	GDPCH60L	68.56	OTHFRAC	63.51	GVR61	7.46
GOVSH61	97.60	HERF00	98.50	GDE1	68.93	MUSLIM00	64.73	TROPPOP	7.70
WARTORN	97.92	POP1560	98.51	IPRICE1	69.48	GDE1	65.59	OTHFRAC	8.49
OIL	97.93	IPRICE1	98.52	GOVSH61	71.02	TOT1DEC1	67.83	MINING	8.51
PI6090	98.09	NEWSTATE	98.58	GVR61	71.27	OPENDEC1	67.94	FERTLDC1	10.91
DPOP6090	98.46	POP60	98.71	DENS65C	71.31	FERTLDC1	69.45	ENGFRAC	12.07
SOCIALIST	98.56	TROPICAR	98.80	SIZE60	71.72	HINDU00	70.45	DENS65I	12.75
ECORG	98.89	PI6090	98.90	DENS60	73.41	TOTIND	72.66	TOT1DEC1	13.13
DENS65I	99.03	ORTH00	98.97	OPENDEC1	74.21	GGCFD3	72.88	DENS60	13.46
SQPI6090	99.16	GEEREC1	99.09	FERTLDC1	75.33	OIL	75.97	BRIT	13.70
DENS60	99.28	GGCFD3	99.40	GGCFD3	79.82	GDPCH60L	79.69	GGCFD3	15.29
LHCPC	99.30	LT100CR	99.40	TOT1DEC1	80.09	DENS60	79.72	ECORG	16.27
POP6560	99.31	AVELF	99.40	TOTIND	80.26	H60	84.49	NEWSTATE	19.05
ABSLATIT	99.34	WARTORN	99.47	OTHFRAC	82.18	ORTH00	85.24	PI6090	19.80
ORTH00	99.39	TROPPOP	99.47	MINING	83.82	EUROPE	88.07	LHCPC	20.29
PRIGHTS	99.49	YRSOPEN	99.55	H60	86.64	SQPI6090	88.39	LT100CR	20.46
POP1560	99.53	CIV72	99.58	WARTIME	88.78	PI6090	88.54	EUROPE	24.36
COLONY	99.55	DPOP6090	99.58	PI6090	89.06	GEEREC1	89.76	SOCIALIST	24.90
LIFE060	99.62	LANDAREA	99.61	POP6560	91.08	SCOUT	90.90	GEEREC1	28.34
GDE1	99.71	WARTIME	99.64	BRIT	91.68	DENS65I	91.59	PRIGHTS	29.72
EUROPE	99.73	LIFE060	99.65	WARTORN	92.03	LT100CR	91.92	WARTORN	33.79
HERF00	99.75	GOVSH61	99.72	DENS65I	92.22	HERF00	93.18	ZTROPICS	35.35
TOT1DEC1	99.81	GOVNOM1	99.75	ORTH00	94.48	CATH00	94.30	AIRDIST	36.71
ENGFRAC	99.81	LANDLOCK	99.82	LANDLOCK	94.98	LANDAREA	94.83	WARTIME	38.37
H60	99.86	SQPI6090	99.83	LANDAREA	96.96	BRIT	95.28	SQPI6090	39.28
NEWSTATE	99.87	ECORG	99.92	HERF00	97.01	WARTORN	95.44	GDE1	39.35
LAAM	99.93	ABSLATIT	99.93	GEEREC1	97.57	MINING	95.44	LANDAREA	43.41
WARTIME	99.95	OTHFRAC	99.94	POP1560	98.31	POP60	96.58	ABSLATIT	50.92
LANDLOCK	99.98	P60	99.95	POP60	99.10	POP1560	99.13	HINDU00	50.93
FERTLDC1	99.98	SAFRICA	99.96	CATH00	99.46	POP6560	99.17	LANDLOCK	53.22
AIRDIST	99.98	SOCIALIST	99.99	LT100CR	99.56	SIZE60	99.56	SCOUT	65.56
LT100CR	99.98	BRIT	99.99	OIL	99.93	WARTIME	99.60	OIL	86.66
LANDAREA	99.99	REVC0UP	100.00	LHCPC	99.95	LHCPC	99.93	MUSLIM00	88.85
TOTIND	99.99	MUSLIM00	100.00	SQPI6090	99.97	GOVNOM1	99.98	POP6560	89.61
CATH00	100.00	FERTLDC1	100.00	GOVNOM1	99.98	LANDLOCK	100.00	HERF00	98.74

注: 每种方法对应变量按照非线性性从小到大排序; 变量右侧数字表示其非线性性程度, 单位为百分比, 依照正文 (7) 式的辅助回归进行测算

C. 3. 前十变量预测非线性性图示

本节图片均按照 A.3 所示绘制 $\Delta y_{[k]i}$ 关于 x_{ki} 的图形。

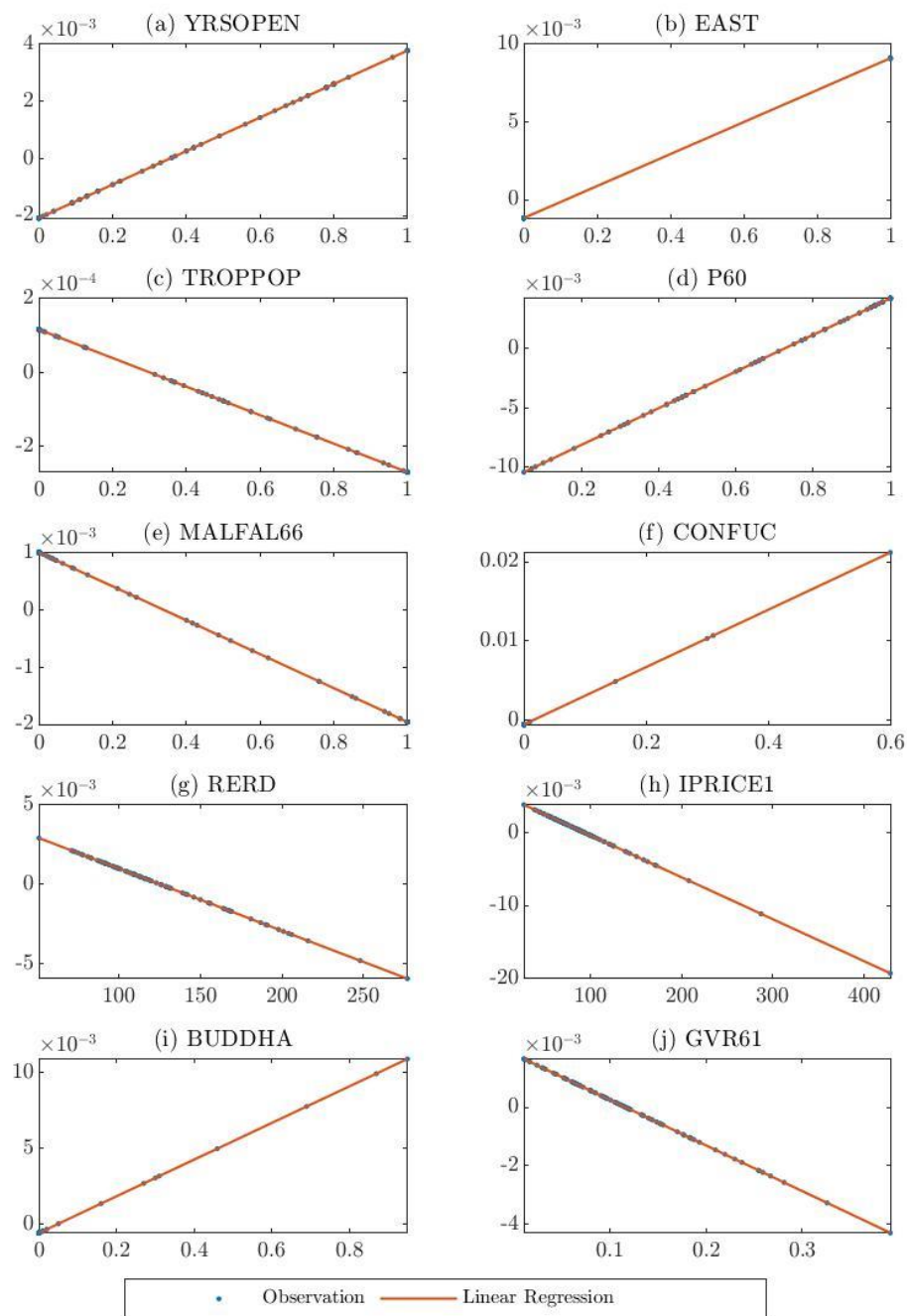


图 C1: LASSO 方法下变量预测非线性性及趋势特征

注: 图中蓝色圆点代表该样本点处的变量 k 对推测值的贡献 $\Delta \hat{y}_{[k]}$

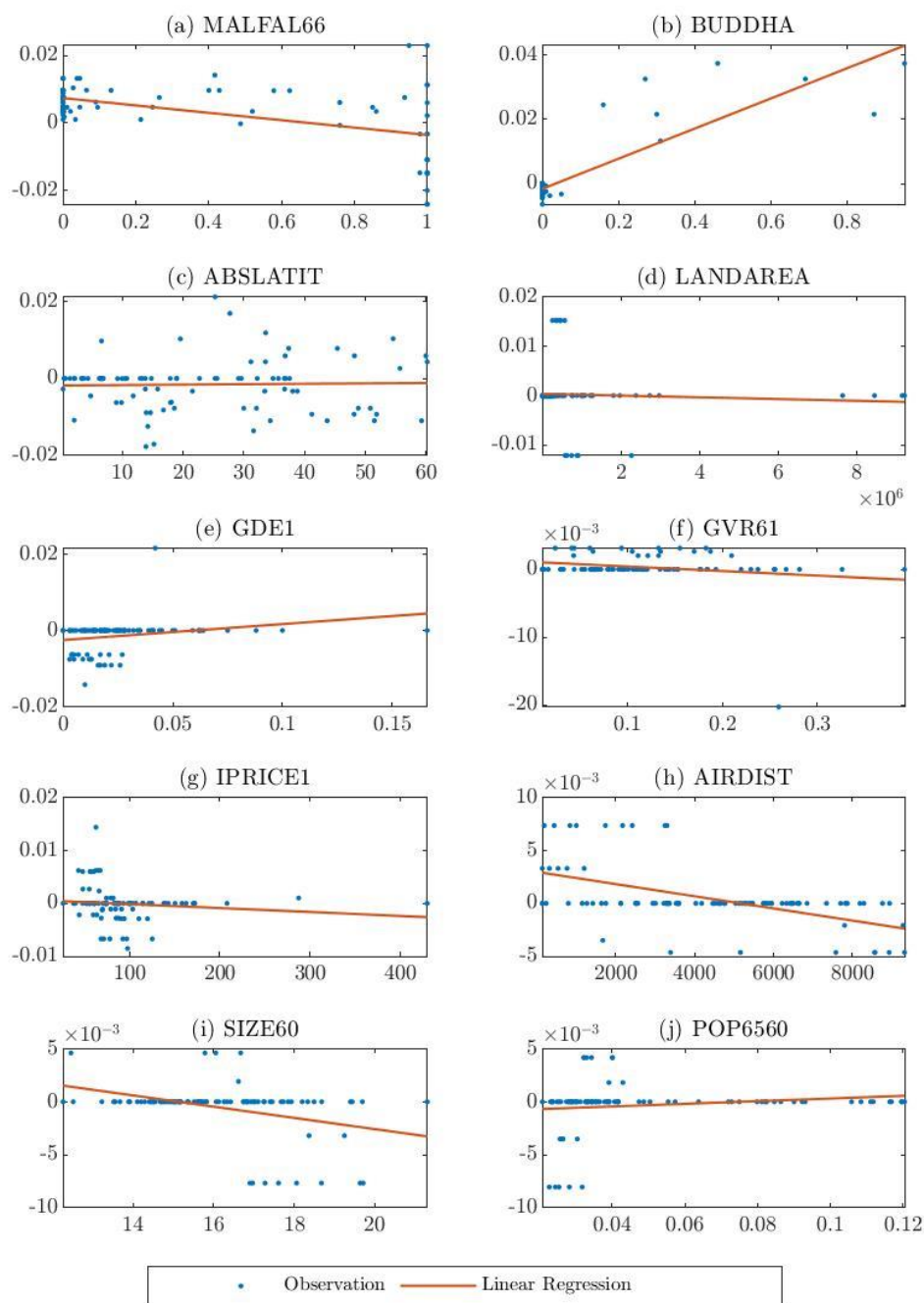


图 C2: 回归树方法下变量预测非线性性及趋势特征

注: 图中蓝色圆点代表该样本点处的变量 k 对推测值的贡献 $\Delta \hat{y}_{[k]}$

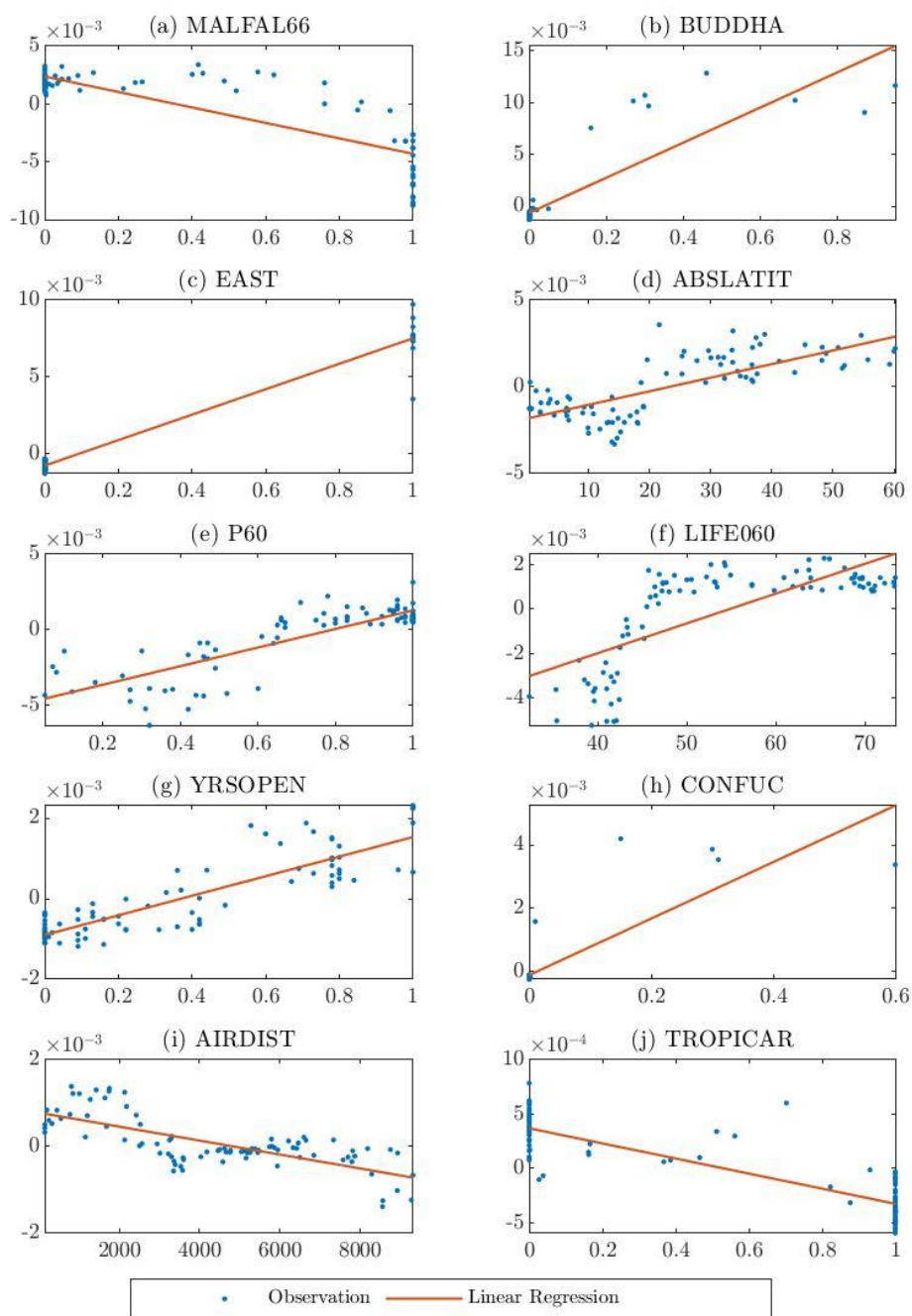


图 C3: 套袋法方法下变量预测非线性性及趋势特征

注: 图中蓝色圆点代表该样本点处的变量 k 对推测值的贡献 $\Delta \hat{y}_{[k]}$

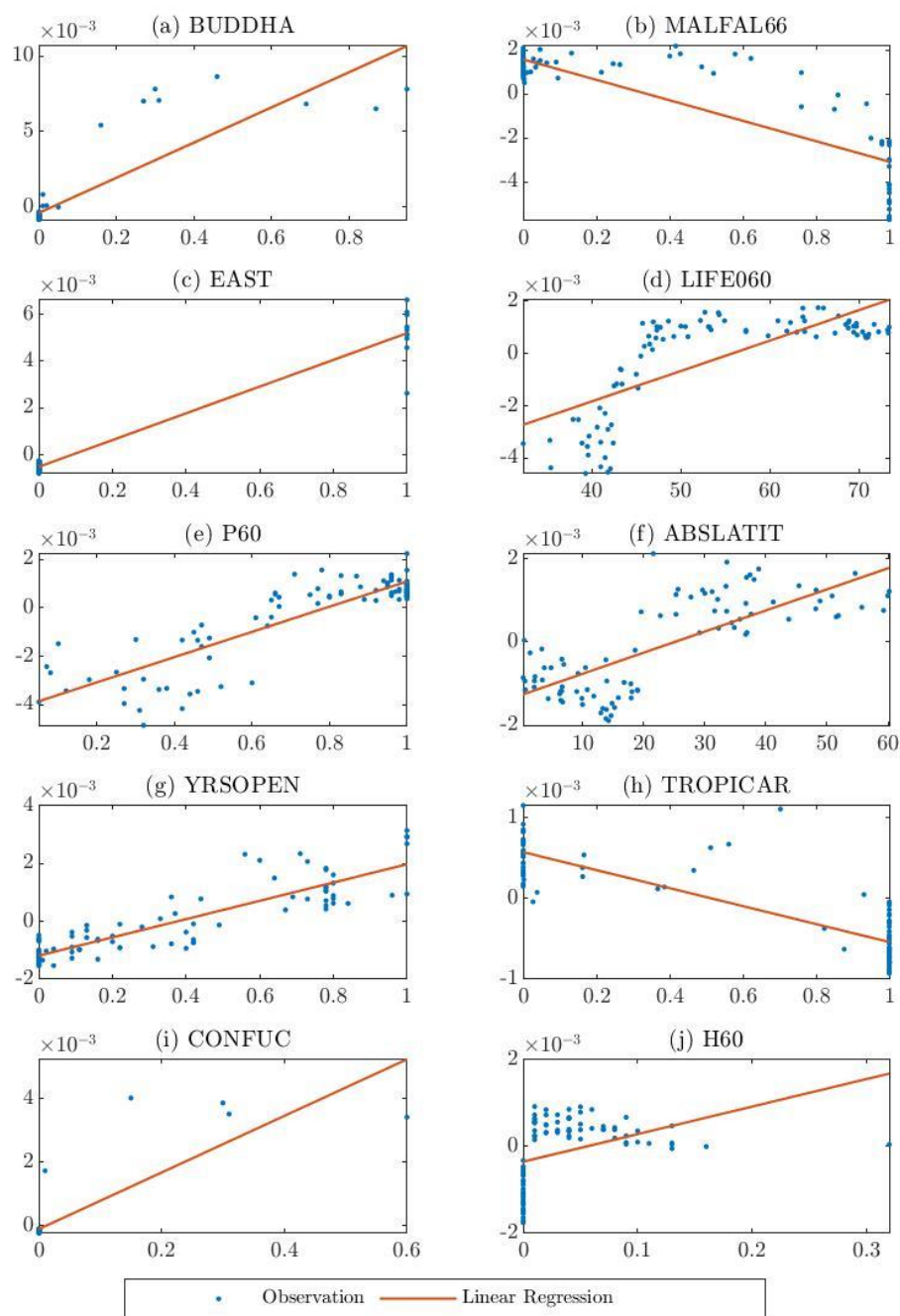


图 C4: 随机森林方法下变量预测非线性性及趋势特征

注: 图中蓝色圆点代表该样本点处的变量 k 对推测值的贡献 $\Delta\hat{y}_{[k]}$

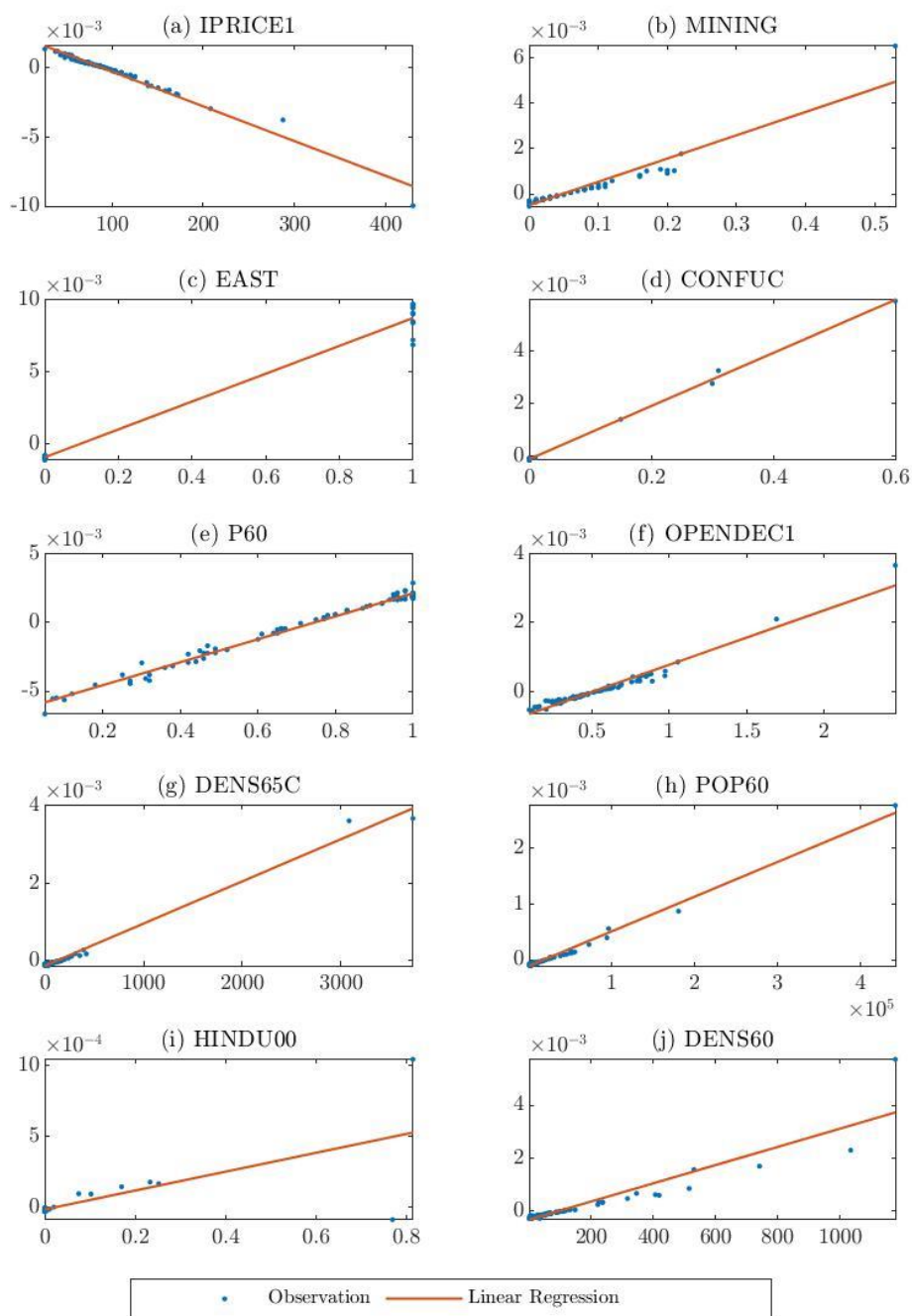


图 C5: 神经网络方法下变量预测非线性性及趋势特征

注: 图中蓝色圆点代表该样本点处的变量 k 对推测值的贡献 $\Delta \hat{y}_{[k]}$

C. 4. 前十变量交互解释力度

热力图矩阵中, 每个方块的颜色代表两两变量组合交互推测解释力 (FIEPV) (见上文(A6)式定义), 颜色越深表示解释力越大; 方块中的数字代表解释力数值, 单位为百分比。

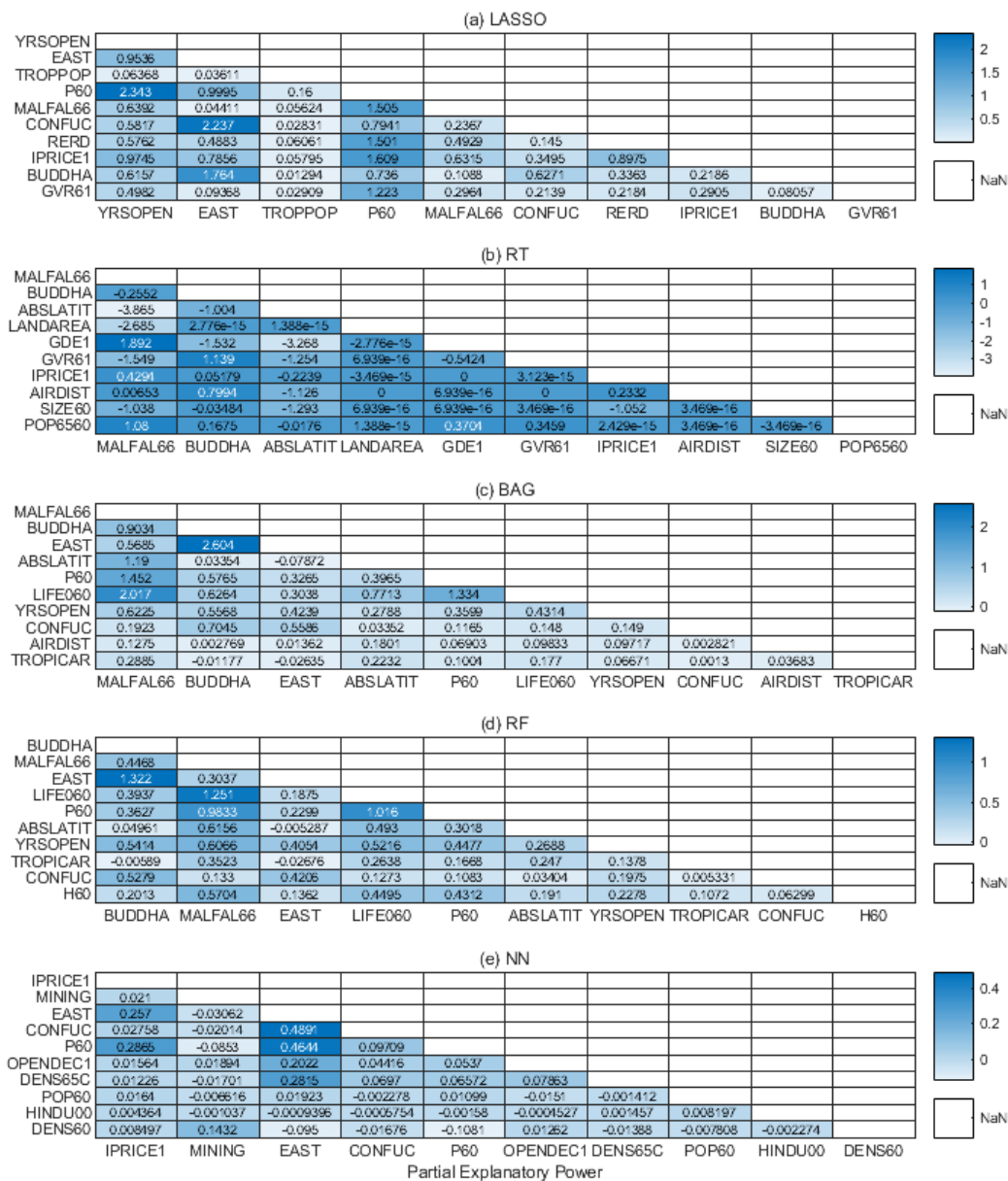


图 C6: 变量交互作用解释力

注: 该附录是期刊所发表论文的组成部分, 同样视为作者公开发表的内容。如研究中使用该附录中的内容, 请务必在研究成果上注明引文和下载附件出处。

引用示例:

参考文献引用范例:

[1] 朱军. 技术吸收、政府推动与中国全要素生产率提升[J]. 中国工业经济. 2017, (1): 5-24.

如果研究中使用了未在《中国工业经济》纸质版刊发、但在杂志网站上正式公开发表的数字内容（包括数据、程序、附录文件），请务必在研究成果正文中注明：

数据（及程序等附件）来自朱军（2017），参见在《中国工业经济》网站（<http://www.ciejjournal.org>）附件下载。