

一种基于机器学习的时变面板数据 政策评估方法^①

高华川 白仲林

(天津财经大学统计学院)

研究目标：基于个体间关系时变的面板数据，提出一种估计政策因果效应的反事实机器学习算法——时变 LASSO 方法。**研究方法：**基于真实数据生成过程未知的经典文献中的实际数据和时变的模拟数据，分别将时变 LASSO 方法与两种常系数方法（HCW 方法和常系数 LASSO 方法）进行对比研究；以英国“脱欧公投”对英镑汇率的因果效应分析为例，研究了政策信息提前披露时政策因果效应的动态行为。**研究发现：**相对于两种常系数方法，时变 LASSO 方法能更准确地估计反事实结果，更适用于样本时期数较长、政策干预时间点相对较晚的政策评估；实证研究表明，常系数方法对脱欧公投的影响存在一定的高估。**研究创新：**提出了一种基于机器学习的时变 LASSO 面板数据政策评估方法，以及基于此的两种反事实估计方法。**研究价值：**对于政策信息提前披露的政策，提供了评价政策因果效应动态行为的工具，完善了面板数据政策评估理论方法。

关键词 面板数据 政策评估 机器学习 时变 LASSO 脱欧公投

中图分类号 F224.0 **文献标识码** A

引 言

政策评估也称为项目评估（Program Evaluation），是指应用计量经济学方法与工具，在经济数据的基础上，对社会经济政策进行量化分析，其主要目的是测度某个政策实施后对某个群体、行业或地区的“因果”影响（洪永淼，2015）。因果推断以及政策评估的关键在于，根据未受政策干预的控制组信息来估计处理组如果未接受政策干预的反事实结果，从而政策干预的因果效应被定义为政策干预下的实际结果与反事实结果之差。

对于由处理组和控制组个体组成的面板数据，如果这些个体对政策干预的响应是相似的（Hsiao，2003），或者即使响应不同，只要它们由不受政策干预的某些共同因子所驱动（Gregory 和 Head，1999；Sargent 和 Sims，1977），就可以利用控制组个体所包含的信息估计处理组个体的反事实结果。基于这一思想，国内外学者在面板数据政策评估领域取得了一系列理论成果，提出了多种评估政策因果效应的方法，例如双重差分法（Card 和 Krueger，1994）、合成控制法（Abadie 等，2003，2010）、HCW 方法（Hsiao 等，2012）以及常系数

① 本文得到天津市哲学社会科学研究规划项目“大数据背景下的宏观经济实时预测”（TJ TJ16—001Q）、教育部人文社会科学研究规划基金项目“宏观经济政策因果效应分析方法及其应用研究”（8YJA790005）、国家自然科学基金青年科学基金项目“ACD 模型中的密度估计与模型检验”（11801399）资助。

LASSO方法(Li和Bell, 2017)等。

事实上,每种面板数据政策评估方法均建立于适当的假定条件基础之上,而且弱化假定条件、扩大政策评估方法的适用范围是发现新方法的逻辑。双重差分法对控制组赋予等权重,以构造与处理组具有共同变化趋势的控制组个体。然而,共同趋势假定不可直接检验,且在很多情况下是不成立的,此时政策效应的估计有偏。Abadie等(2003, 2010)放松了双重差分法的共同趋势假定和等权重假定,利用非时变权重加权控制组个体的事后结果估计单个处理组个体的反事实结果,将双重差分方法拓展为合成控制方法。Xu(2017)提出的广义合成控制方法利用不受政策干预的共同因子估计处理组个体的反事实结果,将政策评估方法推广为多处理个体和政策非同期实施的情形。合成控制法要求权重非负且和为1,即处理组个体在控制组的凸包之内。从而,若处理组个体特征大于或小于所有控制组个体特征,将无法找到合适的权重进行合成控制。HCW方法放松了合成控制对于权重的要求,允许控制组个体的权重为负,从而一定能找到合适的权重来得到合成控制。然而,当控制组个体数较多时,HCW应用信息准则来选择控制组个体的方法将会失效。Li和Bell(2017)进一步将LASSO方法(Least Absolute Shrinkage and Selection Operator)应用于控制组个体的选择,改进了HCW依信息准则选择控制组个体的规则,提出常系数LASSO政策评估方法。

因此,从现有文献可见,面板数据政策评估方法的沿革主要涉及三方面的问题,即处理组和控制组的多样性、政策干预的非同期性和控制组个体选择的有效性。由于所有方法均假定控制组个体的权重系数为常数,并未考虑反事实分析中个体间经济关系的时变特征。然而,政策信息的提前披露会影响处理组与控制组个体的相互联系,于是,利用控制组结果变量的非时变权重加权估计处理组个体的反事实结果并不能反映处理组与控制组个体间关系时变的现实。Giraitis(2014)研究认为,若真实数据生成过程(Data Generating Process, 以下简称为DGP)为时变,但错误地假定模型权重系数为常数,会导致参数估计量不具有 consistency。另外,从实证研究的角度讲,当某国将发生制度变革或进行可能产生较大争议的政策调整之前,通常会在某些场合透露一些消息或进行“媒体吹风会”(例如我国的医疗改革、征收房产税等),由此导致的预期变化将使得个体间经济关系发生改变,甚至某些个体会进入或退出控制组。例如,在英国“脱欧”的进程中,英国与其他国家(尤其是欧盟成员国)的经济关系随着脱欧公投的临近以及各种脱欧相关事件的公布而内生地演化。于是,利用控制组国家“事前”的经济特征来估计英国“事后”的经济事实是欠合理的,应当关注英国与控制组国家的经济关系的时变性。由此可见,理论与实证研究均提出了考虑反事实分析中个体间经济关系时变特征的政策评估方法的需求,这正是本文的研究目标。

本文在常系数LASSO方法的基础上,将新近提出的时变LASSO方法(Kapetanios和Zikes, 2018)与反事实分析相结合,提出了一种允许权重系数随时间变化的时变LASSO面板数据政策评估方法。与现有文献相比,本文的主要贡献在于:放松了传统面板数据政策评估方法中的常系数假定,从而更适合于政策信息的提前披露导致预期变化,从而使得个体间经济关系随时间演化的情形。就笔者所掌握的文献来看,在反事实分析中考虑个体间经济关系的时变特征,本文尚属首次。在此基础上又提出了两种反事实估计方法,并将其应用于研究英国“脱欧公投”对英镑汇率的影响。

本文结构安排如下:第一节介绍两种常系数面板数据政策评估方法,包括HCW以及常系数LASSO方法,并在此基础上提出时变LASSO面板数据政策评估方法;第二节依据经典文献中的三组实际数据,探讨了当真实DGP未知时三种方法的比较研究;第三节利用

Monte Carlo 模拟方法, 验证当真实 DGP 确为时变时, 本文提出的两种反事实估计方法的小样本性质; 第四节将时变 LASSO 方法用于分析英国脱欧公投对于英镑汇率的影响; 第五节是本文的结论。

一、面板数据政策评估方法

1. HCW 方法

(1) 基本设定

假定能观测到 $N=J+1$ 个个体共 T 个时期的信息。令 d_{it} 表示是否受到政策干预的虚拟变量, 即 $d_{it}=1$ 表示个体 i 在时期 t 受到政策干预, $d_{it}=0$ 表示个体 i 在时期 t 未受政策干预。令 Y_{it}^1 和 Y_{it}^0 分别表示个体 i 在时期 t 如果受到政策干预和如果未受干预的潜在结果。令 Y_{it} 表示个体 i 在第 t 期实际观测到的结果变量。由于不能同时观测到 Y_{it}^1 和 Y_{it}^0 , 而只能观测到 Y_{it} , 因此

$$Y_{it} = d_{it}Y_{it}^1 + (1 - d_{it})Y_{it}^0$$

HCW 方法适用于处理组仅包含一个个体的情况。不失一般性, 假定仅第一个个体在某一时期 T_0 ($1 < T_0 < T$) 之后不间断地受到政策干预, 其他 J 个个体组成未受干预的控制组, 即

$$d_{it} = \begin{cases} 1, & i = 1, t = T_0 + 1, \dots, T \\ 0, & otherwise \end{cases}$$

因此, 对处理组个体 $i=1$ 而言

$$Y_{1t} = \begin{cases} Y_{1t}^0, & t = 1, \dots, T_0 \\ Y_{1t}^1, & t = T_0 + 1, \dots, T \end{cases}$$

对 J 个控制组个体有

$$Y_{it} = Y_{it}^0, \quad i = 2, 3, \dots, J+1, t = 1, 2, \dots, T$$

政策干预对处理组个体的处理效应为

$$\tau_{1t} = Y_{1t}^1 - Y_{1t}^0, t = T_0 + 1, T_0 + 2, \dots, T$$

在后干预期 $t=T_0+1, \dots, T$, 我们仅能观测到 Y_{1t}^1 而观测不到 Y_{1t}^0 , 因此, 政策评估的关键是估计反事实结果 Y_{1t}^0 , $t=T_0+1, \dots, T$ 。HCW 方法通过为每个控制组个体赋予一个权重, 使得合成控制个体与处理组个体在前干预期 $1, \dots, T_0$ 的变动趋势非常相似, 从而可以用合成控制个体在政策干预之后的结果变量作为处理组个体反事实结果的估计。具体地, HCW 方法假定 Y_{it}^0 服从如下因子模型

$$Y_{it}^0 = \mu_i + b_i' f_t + \epsilon_{it} \quad i = 1, \dots, J+1; \quad t = 1, \dots, T$$

其中, μ_i 表示个体固定效应, f_t 表示 $K \times 1$ 维不可观测的共同因子, b_i 表示个体 i 的因子载荷向量, ϵ_{it} 为均值为 0 的随机误差项。共同因子 f_t 影响所有个体的潜在结果 Y_{it}^0 , 导致处理组个体的潜在结果 Y_{1t}^0 与控制组个体的潜在结果 $Y_{-1,t} = (Y_{2t}^0, \dots, Y_{J+1,t}^0)'$ 存在截面相关性, 这一相关性可用如下回归模型概括

$$Y_{it}^0 = \beta_1 + Y'_{-1,i} \beta_{-1} + \xi_{1,i} \quad (1)$$

其中, β_1 为截距项, 回归系数 $\beta_{-1} = (\beta_2, \dots, \beta_{J+1})'$ 为控制组个体的权重。

在前干预期 $t=1, 2, \dots, T_0$, $Y_{it}^0 = Y_{it}$ 可观测, 利用前干预期的数据估计式 (1) 可得到 $\hat{\beta}_1$ 和 $\hat{\beta}_{-1}$ 。HCW 方法认为, 如果处理组个体没有受到政策干预, 那么在后干预期 $t=T_0+1, T_0+2, \dots, T$, 处理组与控制组还应当保持式 (1) 所表述的关系, 因此, 可用式 (2) 估计处理组个体的反事实结果

$$\hat{Y}_{it}^0 = \hat{\beta}_1 + Y'_{-1,i} \hat{\beta}_{-1}, t = T_0 + 1, \dots, T \quad (2)$$

此时, 政策干预的效应估计为

$$\hat{\tau}_{it} = Y_{it}^1 - \hat{Y}_{it}^0 = Y_{it} - \hat{Y}_{it}^0, t = T_0 + 1, \dots, T$$

政策干预的平均处理效应的估计为

$$\hat{\tau}_1 = \frac{1}{T - T_0} \sum_{t=T_0+1}^T \hat{\tau}_{it}$$

(2) 控制组个体的选择

HCW 方法的关键在于控制组个体的选择。原因在于式 (1) 的样本容量为前干预期时期数 T_0 , 解释变量个数即为控制组个体数 J 。当 T_0 远远大于 J 时, 可以让所有控制组个体进入模型。但是在面板数据政策评估的实际应用中, T_0 往往是有限的, 进入模型的控制组个体越多, 回归模型式 (1) 的样本内拟合优度上升, 而样本外的反事实估计精度会越低。因此, 式 (1) 存在样本内拟合优度与样本外估计精度之间的权衡。

HCW 利用地理上的邻近或经济上的相似性来筛选潜在控制个体, 并提出一种两步法从潜在控制个体中选择进入模型的控制组: 首先, 依次选择 $j=1, 2, \dots, J$ 个控制组个体进入模型, 对于有 j 个控制组个体的模型, 利用拟合优度 R^2 从中选择拟合最好的模型, 记为 $M(j)^*$ 。第一步结束后共得到 J 个模型 $M(1)^*, \dots, M(J)^*$ 。然后, 利用 AIC 等信息准则在 $M(1)^*, \dots, M(J)^*$ 中选择出最优模型 $M(m)^*$ 。

2. 常系数 LASSO 方法

如前所述, HCW 方法的关键在于控制组个体的选择。当潜在控制组个体数较多时, HCW 利用地理上的邻近与经济上的相似性筛选控制组个体的方法存在一定的主观性。另外, 当 $J > T_0$ 时, 即式 (1) 中的解释变量个数大于样本容量, HCW 方法不可行。即使 $J < T_0$, 但控制组个体数量 J 仍较大时, HCW 方法的计算量很大, 例如, 当有 $J=30$ 个控制组个体时, HCW 方法需要估计 $C_{30}^1 + C_{30}^2 + \dots + C_{30}^{30} \approx 10^9$ 个模型。实际上, 式 (1) 中控制组个体的选择问题可归结为高维数据中的变量选择或特征选择问题。另外, 若将估计式 (1) 的前干预期数据看作样本内, 则式 (2) 对反事实的估计也可称为样本外的反事实预测。而特征选择与预测是机器学习的主要研究领域与优势所在, Li 和 Bell (2017) 提出利用机器学习中的常系数 LASSO 方法进行控制组个体选择与反事实估计。

常系数 LASSO 方法由 Tibshirani (1996) 作为一种高维数据降维的统计方法提出, 现在已经成为有监督的机器学习 (Supervised Machine Learning) 中有关特征选择的标准内容。例如, Chernozhukov 等 (2016) 所提出的“双机器学习方法” (Double Machine Learning) 即是利用两次 LASSO 方法筛选同时影响结果变量与处理变量的混杂因素, 并将其作

为控制变量加入回归模型进行因果推断。

(1) 基本设定

LASSO 是回归模型在高维情况下进行变量选择或特征选择的方法, 为便于叙述, 将式 (1) 写为普通回归模型的形式

$$Y_t = X_t' \beta + u_t, t = 1, \dots, T_0 \quad (3)$$

其中, $Y_t = Y_{1t}^0$, $X_t = (1, Y_{-1,t})'$, $\beta = (\beta_1, \beta_{-1})' = (\beta_1, \beta_2, \dots, \beta_{J+1})'$, $u_t = \xi_{1,t}$ 。

众所周知, 当回归模型中解释变量的个数 J 增大时, 样本内拟合优度上升, 但样本外预测性能下降, 这一现象在机器学习领域称为过拟合 (Overfitting)。LASSO 方法通过在 OLS 目标函数中加入惩罚项, 引入样本内估计偏差, 提高样本外反事实预测精度。具体地, LASSO 方法的目标函数为

$$\sum_{t=1}^{T_0} (Y_t - X_t' \beta)^2 + \lambda \sum_{i=2}^{J+1} |\beta_i| \quad (4)$$

其中, 惩罚项 $\sum_{i=2}^{J+1} |\beta_i| = ||\beta_{-1}||_1$ 为 L_1 范数, L_1 惩罚能够使某些系数收缩为 0, 从而起到变量选择 (即控制组个体选择) 的作用。 $\lambda \geq 0$ 称为调节参数 (Turning Parameter), 它控制了对非零系数的惩罚力度, λ 越大, 对非零 β_i 的惩罚越大, 导致越多的 β_i 收缩为 0。两种极端情况是 $\lambda = 0$ 和 $\lambda = \infty$, 前者不对 β_i 施加任何约束, LASSO 退化为 OLS, 而后者将所有 β_i 收缩为 0。

与 HCW 方法相同, 在得到 β 的 LASSO 估计量 $\hat{\beta}$ 后, 利用 $\hat{Y}_t = X_t' \hat{\beta}$, $t = T_0 + 1, \dots, T$ 估计处理组个体的反事实结果。由式 (4) 可知, 为了得到 $\hat{\beta}$, 需要确定调节参数 λ 的值, 机器学习领域常用交叉验证 (Cross Validation) 方法。

(2) 交叉验证

在机器学习领域, 用于估计式 (4) 的数据 ($t = 1, \dots, T_0$) 称为训练数据, 在训练数据上的误差称为训练误差或经验误差, 在新样本 $t = T_0 + 1, \dots, T$ 上的反事实预测误差称为泛化误差。显然, 我们希望得到泛化误差小的模型。而由于新样本上的反事实结果观测不到, 无法得到真实的泛化误差。交叉验证方法通过将训练数据划分出一部分样本作为测试数据, 并用模型在测试数据上的测试误差作为泛化误差的估计。

交叉验证方法先将数据集 $D = \{Y_t, X_t\}_{t=1}^{T_0}$ 划分为 k 个大小相似的互斥子集, 即 $D = D_1 \cup D_2 \cup \dots \cup D_k$, $D_i \cap D_j = \emptyset$, $i \neq j$ 。然后, 每次用 $k-1$ 个子集的并集作为训练集, 余下的子集作为测试集, 这样就可获得 k 组训练集与测试集。利用 k 组训练集进行 k 次模型估计, 并得到测试集上的 k 次测试误差, 最后, 将 k 次测试误差的均值作为泛化误差的估计。这种交叉验证方法又称为 k 折交叉验证 (k -fold Cross Validation)。

面板数据政策评估方法中的 T_0 通常较小, 一般采用留一法 (Leave-One-Out, LOO) 交叉验证, 即令 $k = T_0$ 。具体地, 假定 λ 在一个离散集合 $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ 中取值, 其中 $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ 根据 Friedman 等 (2010) 来设定。对于任一 $\lambda \in \Lambda$ 以及任一 $t = 1, \dots, T_0$, 通过最小化留一法目标函数来估计 β

$$\sum_{s=1, s \neq t}^{T_0} (Y_s - X_s' \beta)^2 + \lambda \sum_{i=2}^{J+1} |\beta_i|$$

使上式最小化的 β 记作 $\hat{\beta}_{-t}(\lambda)$ 。然后, 计算测试集 t 上的测试误差 $e_t(\lambda) = Y_t - X_t' \hat{\beta}_{-t}(\lambda)$

以及 $k=T_0$ 组测试集上的均方误差

$$CV(\lambda) = \frac{1}{T_0} \sum_{t=1}^{T_0} e_t(\lambda)^2 = \frac{1}{T_0} \sum_{t=1}^{T_0} [Y_t - X_t' \hat{\beta}_{-t}(\lambda)]^2$$

选择 $\hat{\lambda} \in \Delta$ 以最小化上式中的 $CV(\lambda)$ 。最后, 将 $\hat{\lambda}$ 代入式 (4) 并用所有样本求出 LASSO 估计量 $\hat{\beta}$ 。

3. 时变 LASSO 方法

HCW 和常系数 LASSO 等政策评估方法均假定控制组个体的权重系数 β 为不变常数。若现实中政策实施前的信息披露使预期产生变化, 从而导致处理组个体与控制组个体之间的结构关系随着时间的推移而发生改变, 则需放松常数假定。因此, 本文将 Kapetanios 和 Zikes (2018) 提出的时变 LASSO (Time-Varying LASSO) 方法与反事实分析相结合, 提出如下基于机器学习的时变 LASSO 面板数据政策评估方法。时变 LASSO 面板数据政策评估方法的“机器学习”特征主要体现在: 借鉴 LASSO 方法将控制组个体的选择看作高维数据的降维问题或特征选择问题, 且将反事实估计看作后干预期的反事实预测。

(1) 基本设定

本文允许常数 LASSO 方法式 (3) 中的权重系数 β 时变, 具体地, 本文设定的模型形式为

$$Y_t = X_t' \beta_t + u_t \quad (5)$$

其中 $\beta_t = (\beta_{1t}, \beta_{2t}, \dots, \beta_{J+1,t})'$ 为时变权重系数。Kapetanios 和 Zikes (2018) 证明, 上式的时变核 LASSO 估计量

$$\hat{\beta}_t = \arg \min_{\beta} \left\{ \sum_{i=1}^{T_0} w_{ij} (Y_i - X_i' \beta)^2 + \lambda \sum_{i=2}^{J+1} |\beta_i| \right\} \quad (6)$$

是 β_t 的一致估计量, 其中, $w_{ij} \equiv \tilde{w}_{ij} / \sum_{j=1}^{T_0} \tilde{w}_{ij}$ 为各期样本的权重, $\tilde{w}_{ij} = K((t-j)/H)$, $K(\cdot)$ 为高斯核函数, $H=T_0^{1/2}$ 为带宽参数。

(2) 留一法交叉验证

假定 λ 在一组离散集合 $\Delta = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ 中取值。首先, 对于任一 $\lambda \in \Delta$ 以及任一 $t=1, \dots, T_0$ 时变 LASSO 的留一法目标函数为

$$\sum_{s=1, s \neq t}^{T_0} w_{tj} (Y_s - X_s' \beta)^2 + \lambda \sum_{i=2}^{J+1} |\beta_i|$$

使上式最小化的 β 记作 $\hat{\beta}_{-t}(\lambda)$ 。然后, 计算验证数据集 t 上的误差 $e_t(\lambda) = Y_t - X_t' \hat{\beta}_{-t}(\lambda)$ 。对每一个 λ 值, 计算所有验证数据集上的均方误差

$$CV(\lambda) = \frac{1}{T_0} \sum_{t=1}^{T_0} e_t(\lambda)^2 = \frac{1}{T_0} \sum_{t=1}^{T_0} [Y_t - X_t' \hat{\beta}_{-t}(\lambda)]^2$$

选择 $\hat{\lambda} \in \Delta$ 以最小化上式中的 $CV(\lambda)$ 。最后将 $\hat{\lambda}$ 代入式 (6) 并用所有样本求出时变 LASSO 估计量 $\hat{\beta}_t, t=1, \dots, T_0$ 。

(3) 反事实估计方法

本文认为权重系数随时间变化的原因有: 有关政策信息的提前披露所导致的预期变化,

会使得处理组与控制组之间的关系提前进行调整；或者所研究政策实施前发生其他具有较长期效应的重大事件，从而导致权重系数随时间变化。与此两种情况相对应，本文在时变LASSO方法的框架下提出了如下两种反事实估计方法。

第一种方法假定政策实施前无其他具有较长期效应的重大事件发生。如果市场对政策的响应是有效率的，政策的效应就得到了充分的反映。因此，政策实施后权重系数不再变化，这时政策干预后处理组与控制组之间保持政策干预前最后一期的关系，即 $\hat{\beta}_\tau = \hat{\beta}_{T_0}$ ， $\tau = T_0 + 1, \dots, T$ ，因此可用 $\hat{Y}_\tau = X'_\tau \hat{\beta}_{T_0}$ 作为反事实结果 Y_τ ， $\tau = T_0 + 1, \dots, T$ 的估计。此时，权重系数中不包含政策效应。

第二种方法假定政策实施前有其他具有较长期效应的重大事件发生。重大事件发生对权重系数的影响可能持续到政策发生之后，此时权重系数的变化是由重大事件引起的，依然不包含政策效应。此时需要递归地估计 $\hat{\beta}_\tau$ ， $\tau = T_0 + 1, \dots, T$ 。首先，利用 $\hat{Y}_{T_0+1} = X'_{T_0+1} \hat{\beta}_{T_0}$ 估计反事实结果 Y_{T_0+1} ；然后，将 $\hat{Y}_{T_0+1}$ 作为 $Y_{T_0+1}^0$ 的真实值，利用数据 $\{Y_t, X_t\}_{t=T_0+1}^T$ 得到 T_0+1 期的参数估计 $\hat{\beta}_{T_0+1}$ ，并估计 T_0+2 期的反事实结果 $\hat{Y}_{T_0+2}$ ；重复以上两步直至得到所有反事实结果的递归估计 $\hat{Y}_\tau$ ， $\tau = T_0 + 1, \dots, T$ 。

4. 面板数据政策评估方法的统一框架

本文引入如下符号，将三种方法置于统一框架进行讨论，并阐述三种方法的区别与联系。令实际数据为如下面板数据

$$Y^{ds} = \begin{pmatrix} Y_{1,1} & Y_{2,1} & Y_{3,1} & \cdots & Y_{N,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{1,T_0} & Y_{2,T_0} & Y_{3,T_0} & \cdots & Y_{N,T_0} \\ Y_{1,T_0+1} & Y_{2,T_0+1} & Y_{3,T_0+1} & \cdots & Y_{N,T_0+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{1,T} & Y_{2,T} & Y_{3,T} & \cdots & Y_{N,T} \end{pmatrix}$$

其中，第 i 列为第 i 个个体，第 1 列为处理组个体，第 2 列至第 N 列为控制组个体，第 t 行为第 t 期观测值。

令 $Y_{pre,tr}^{ds}$ 和 $Y_{post,tr}^{ds}$ 分别表示处理组在政策干预前和政策干预后的观测结果，即

$$Y_{pre,tr}^{ds} = \begin{pmatrix} Y_{1,1} \\ \vdots \\ Y_{1,T_0} \end{pmatrix}, Y_{post,tr}^{ds} = \begin{pmatrix} Y_{1,T_0+1} \\ \vdots \\ Y_{1,T} \end{pmatrix}$$

同样令 $Y_{pre,\omega}^{ds}$ 和 $Y_{post,\omega}^{ds}$ 分别表示控制组在政策干预前和政策干预后的观测结果，即

$$Y_{pre,\omega}^{ds} = \begin{pmatrix} Y_{2,1} & Y_{3,1} & \cdots & Y_{N,1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2,T_0} & Y_{3,T_0} & \cdots & Y_{N,T_0} \end{pmatrix}, Y_{post,\omega}^{ds} = \begin{pmatrix} Y_{2,T_0+1} & Y_{3,T_0+1} & \cdots & Y_{N,T_0+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2,T} & Y_{3,T} & \cdots & Y_{N,T} \end{pmatrix}$$

则观测矩阵 Y^{ds} 可写成如下分块矩阵的形式

$$Y^{ds} = \begin{pmatrix} Y_{pre,tr}^{ds} & Y_{pre,\omega}^{ds} \\ Y_{post,tr}^{ds} & Y_{post,\omega}^{ds} \end{pmatrix}$$

由于仅处理组在后干预期 $t = T_0 + 1, \dots, T$ 期接受处理, 因此,

$$Y^{obs} = \begin{bmatrix} Y_{pre, tr}^{obs} & Y_{pre, co}^{obs} \\ Y_{post, tr}^{obs} & Y_{post, co}^{obs} \end{bmatrix} = \begin{bmatrix} Y_{pre, tr}(0) & Y_{pre, co}(0) \\ Y_{post, tr}(1) & Y_{post, co}(0) \end{bmatrix}$$

政策干预的因果效应取决于矩阵 $Y_{post, tr}(1)$ 和 $Y_{post, tr}(0)$, 前者为观测结果, 后者为不可观测的反事实结果。因此, 面板数据政策评估的关键便是, 如何利用矩阵 $Y_{pre, tr}(0)$, $Y_{pre, co}(0)$ 和 $Y_{post, co}(0)$ 的信息, 尤其是如何对三者与 $Y_{post, tr}(0)$ 的关系进行建模, 来估计反事实结果 $Y_{post, tr}(0)$

$$Y(0) = \begin{bmatrix} Y_{pre, tr}(0) & Y_{pre, co}(0) \\ \vdots & Y_{post, co}(0) \end{bmatrix}$$

HCW 和常系数 LASSO 面板数据政策评估方法是对 $Y_{pre, tr}(0)$ 和 $Y_{pre, co}(0)$ 之间的关系建立常系数模型 (见式 (1) 和式 (3)), 并假定 $Y_{post, tr}(0)$ 和 $Y_{post, co}(0)$ 也保持同一种关系, 然后用 $Y_{post, co}(0)$ 对 $Y_{post, tr}(0)$ 进行估计。

而时变 LASSO 方法考虑了政策信息的提前披露导致处理组与控制组之间结构关系随时间变化的情况, 从而对 $Y_{pre, tr}(0)$ 和 $Y_{pre, co}(0)$ 之间的关系建立时变系数模型 (见式 (5))。若无其他具有较长期效应的其他重大事件发生, 则认为 $Y_{post, tr}(0)$ 和 $Y_{post, co}(0)$ 保持了 $Y_{post, tr}(0)$ 和 $Y_{post, co}(0)$ 最后一期的关系。

二、基于实际数据的比较研究

Doudchenko 和 Imbens (2017)、Athey 等 (2017) 在分别提出一种面板数据政策评估方法时, 利用了经典文献中的实际数据对不同政策评估方法进行比较研究, 原因在于: (1) 经典文献中的实际数据是开放的, 且其数据质量、可信度均得到了学界公认; (2) 更重要的原因在于实际数据的真实 DGP 是未知的, 利用实际数据能够在不对真实 DGP 做任何假定的条件下将不同方法进行对比研究。同样, 经典文献中实际数据的真实 DGP 是否时变是未知的, 本文借鉴 Doudchenko 和 Imbens (2017)、Athey 等 (2017) 的思想, 利用经典文献中的实际数据, 不对真实 DGP 是否时变做出任何假定, 将时变 LASSO 方法与两种常系数方法进行比较研究。

1. 数据集简介

Hsiao 等 (2012) 研究了 1997 年香港回归以及 2003 年 6 月 29 日签署并于 2004 年 1 月 1 日开始实施的《内地与香港关于建立更紧密经贸关系的安排》(Closer Economic Partnership Arrangement, CEPA) 对于香港经济的影响。数据集中包含 $N=25$ 个国家和地区, 其中, 香港是接受政策干预的处理组个体, 而澳大利亚、德国、日本、中国台湾等 $J=24$ 个国家和地区作为控制组。结果变量 Y_{it} 为季度实际 GDP 增长率。数据的起止时间为 1993 年第一季度到 2008 年第一季度, 即 $T=61$ 。

Abadie 等 (2010) 研究了美国加利福尼亚州于 1988 年 11 月通过的香烟控制 99 号法案对加州香烟消费影响。数据集中包含美国 $N=39$ 个州, 其中, 加州是接受政策干预的处理组个体, 而其他 $J=38$ 个州作为控制组。结果变量 Y_{it} 为各州年度人均香烟消费量^①。数据

① Abadie 等 (2010) 采用了合成控制方法, 其数据集中还包括人均 GDP 的对数、州人口中 15 岁至 24 岁人口所占比重、香烟零售价格以及人均啤酒消费量等协变量。本文考虑的 HCW、LASSO 和时变 LASSO 方法均不需要协变量。

的起止时间为 1970 年到 2000 年, 即 $T=31$ 。

Gardeazabal 和 Vega-Bayo (2017) 提供了研究 1991 年爆发的塞拉利昂内战的经济成本的数据集。数据集中包含 $N=41$ 个国家和地区, 其中, 塞拉利昂是受到干预的处理组个体, 而其他 $J=40$ 个国家和地区作为控制组。结果变量 Y_{it} 为年度人均实际 GDP 的对数。数据的起止时间为 1970 年到 2009 年, 即 $T=40$ 。

由以上数据集简介可知, 本文所选择的三组实际数据基本涵盖了政策评估领域常见的数据结构。以上三组实际数据的描述性统计量见表 1。

表 1 实际数据描述性统计量

数据集	N	T	均值	标准差	最小值	最大值
Hsiao 等 (2012)	25	61	0.0376	0.0389	-0.1959	0.2068
Abadie 等 (2010)	39	31	118.8932	32.7674	40.7000	296.2000
Gardeazabal 等 (2017)	41	40	9.1128	1.2267	6.2323	11.6367

2. 模型评价方法

由于 $Y_{post, tr}(0)$ 不可观测, 因此无法对 $Y_{post, tr}(0)$ 的估计 $\hat{Y}_{post, tr}(0)$ 进行评价。因此, 本文借鉴 Doudchenko 和 Imbens (2017)、Athey 等 (2017) 的思想, 首先将处理组从数据中去掉, 此时实际数据均为未接受政策干预的控制组数据

$$Y_{\omega}^{obs} = \begin{pmatrix} Y_{pre, \omega}(0) \\ Y_{post, \omega}(0) \end{pmatrix} = \begin{pmatrix} Y_{2,1}(0) & Y_{3,1}(0) & \cdots & Y_{N,1}(0) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2,T}(0) & Y_{3,T}(0) & \cdots & Y_{N,T}(0) \end{pmatrix}$$

此时, 以上控制组数据已没有任何实际政策背景, 仅用于评价不同政策评估方法的优劣。然后, 将某一控制组个体作为假想的处理组个体, 不失一般性, 假定第一个控制组个体为假想的处理组个体, 并设定一个假想的政策干预时间点 T_0^* , 则实际数据可表示为

$$Y_{\omega}^{obs} = \begin{pmatrix} Y_{2,1} & Y_{3,1} & Y_{4,1} & \cdots & Y_{N,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{2,T_0^*} & Y_{3,T_0^*} & Y_{4,T_0^*} & \cdots & Y_{N,T_0^*} \\ Y_{2,T_0^*+1} & Y_{3,T_0^*+1} & Y_{4,T_0^*+1} & \cdots & Y_{N,T_0^*+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{2,T} & Y_{3,T} & Y_{4,T} & \cdots & Y_{N,T} \end{pmatrix} = \begin{pmatrix} Y_{pre^*, 2}(0) & Y_{pre^*, -2}(0) \\ Y_{post^*, 2}(0) & Y_{post^*, -2}(0) \end{pmatrix}$$

由于假想的处理组个体实际并未接受政策干预, 因此在假想的政策干预时间点 T_0^* 之后的观测结果就是 $Y_{post^*, 2}(0)$, 利用 $Y_{pre^*, 2}(0)$, $Y_{pre^*, -2}(0)$ 和 $Y_{post^*, -2}(0)$ 的信息, 并应用面板数据政策评估方法估计出 $\hat{Y}_{post^*, 2}(0)$ 之后, 可用 $\hat{Y}_{post^*, 2}(0)$ 与 $Y_{post^*, 2}(0)$ 之间的差距来评价相应政策评估方法的效果。

具体地, 本文提出如下评价方法: 由于真实处理组个体的反事实结果无法观测, 从数据集中去掉处理组个体 $j=1$, 依次设定控制组中第 $j=2, 3, \dots, J+1$ 个个体为假定的处理组个体, 并利用其余的 $J-1$ 个控制组个体来合成个体 j 。首先, 设定一个假想的政策干预时间点 T_0^* , 并利用前干预期 $1, \dots, T_0^*$ 的数据分别估计三种模型; 然后, 利用后干预期 $T_0^*+1, \dots, T$ 的数据分别得到三种方法对个体 j 的反事实结果的估计 $\hat{Y}_{jt}$, $t=T_0^*+1, \dots, T$ 。由于假定的

处理组个体 j 实际并未受到政策干预, 因此 $Y_{jt}=Y_{jt}^0$, 可以用 $RMSE_j=\sqrt{\sum_{t=T_0+1}^T (Y_{jt}-\hat{Y}_{jt})^2}$ 衡量每种方法对个体 j 反事实估计的准确性; 最后, 取 $RMSE_j$ 在所有 J 个控制组个体上的中位数^①, 用中位数 $RMSE$ 评价三种方法对反事实结果估计的准确性。

由于每组数据的量纲不同, 为便于比较, 每组数据的结果需选用一种方法作为基准。由于后两组数据的控制组个体数 J 均大于总时期数 T , HCW 方法失效; 因此, 本文以常系数 LASSO 方法的 $RMSE$ 中位数为基准, 即将三种方法的 $RMSE$ 中位数全部除以常系数 LASSO 方法 $RMSE$ 的中位数, 从而常系数 LASSO 方法的 $RMSE$ 中位数比值均为 1; 若另外两种方法的 $RMSE$ 中位数比值小于 1, 表示此方法的反事实估计结果比常系数 LASSO 方法准确, 反之则不如常系数 LASSO 方法准确。另外, 由于前干预期时期数 T_0^* 为估计模型时的样本容量, 考虑到不同 T_0^* 对评估结果的影响, 本文依次考察了 $T_0^*/T=0.5, 0.7, 0.9$ 三种情况。

3. 三种方法的对比与评价

三种方法反事实估计的 $RMSE$ 中位数比值见表 2。从 Hsiao 等 (2012) 数据集的 $RMSE$ 中位数比值来看, 首先, HCW 方法只有在 $T_0^*/T=0.9$ 时对反事实的估计比常系数 LASSO 方法准确, 而在 $T_0^*/T=0.5, 0.7$ 时, 估计性能劣于常系数 LASSO 方法。其次, 时变 LASSO 方法在 $T_0^*/T=0.5, 0.9$ 时的反事实估计性能强于常系数 LASSO 方法, 尤其是在 $T_0^*/T=0.9$ 的情况下, 估计性能较常系数 LASSO 提高了 20% 左右。从时变 LASSO 与 HCW 方法的对比来看, 时变 LASSO 的两种反事实预测方法在三种情况下均优于 HCW 方法。最后, 时变 LASSO 的两种反事实估计方法在此数据集下的估计性能差别不大。因此, 从这一组数据来看, 时变 LASSO 方法总的来说优于 HCW 方法和常系数 LASSO 方法。

从 Abadie 等 (2010) 数据集来看, 时变 LASSO 的第一种反事实估计方法的表现最好, 在 $T_0^*/T=0.5, 0.7, 0.9$ 三种情况下均优于常系数 LASSO 方法, 尤其是当 $T_0^*/T=0.9$ 时, 估计准确性提高了 30% 左右。而第二种反事实估计方法在 $T_0^*/T=0.7, 0.9$ 时估计性能较好。对于 Gardeazabal 和 Vega-Bayo (2017) 数据集, 时变 LASSO 的第一种反事实估计方法在 $T_0^*/T=0.5, 0.9$ 情况下优于常系数 LASSO 方法, 而第二种反事实估计方法在 $T_0^*/T=0.9$ 时估计性能较好。

表 2 三种方法反事实估计的 $RMSE$ 中位数比值

数据集	Hsiao 等 (2012)			Abadie 等 (2010)			Gardeazabal 等 (2017)		
T_0^*/T	0.5	0.7	0.9	0.5	0.7	0.9	0.5	0.7	0.9
LASSO	1	1	1	1	1	1	1	1	1
HCW	1. 2850	1. 1976	0.9613	—	—	—	—	—	—
TV-LASSO1	0.9873	1. 1435	0.8226	0.9482	0.9774	0.7183	0.9272	1. 0120	0.7676
TV-LASSO2	0.9292	1. 1658	0.8198	1. 1139	0.9031	0.7697	1. 0250	1. 1081	0.8861

注: TV-LASSO1 为时变 LASSO 的第一种反事实估计方法, TV-LASSO2 为第二种反事实估计方法。
粗体标示大于 1 的情况, 即劣于基准方法。

综合经典文献中三组实际数据集的情况, 当潜在 DGP 是否时变为未知时, 时变 LASSO 方法依然优于 HCW 与常系数 LASSO 方法, 尤其是当前干预期时期数较大, 即 $T_0^*/T=0.9$

① 由于面板数据政策评估方法中, 控制组个体数 J 通常较小, 且平均值易受异常值影响, 因此, 本文采用 $RMSE$ 中位数, 而没有采用平均值。

时,估计性能较常系数 LASSO 方法提高了 20%~30%。最后,从时变 LASSO 方法的两种反事实估计方法的对比来看,第一种反事实估计方法更为稳健。

值得特别指出的是,数据所来源的三篇经典文献均是采用了常系数方法,然而本文发现利用其提供的数据,时变 LASSO 方法对反事实的估计效果优于常系数方法,原因可能是由于很多情况下政策的实施通常都会提前披露信息,或者政策实施前其他重大事件的发生都可能导致经济个体之间的关系进行提前调整。例如 Hsiao 等 (2012) 所研究的 CEPA 对香港经济的影响问题,虽然 CEPA 于 2004 年 1 月 1 日开始实施,然而 CEPA 在 2003 年 6 月 29 日就已签署 (信息提前披露),从 CEPA 签署到实施这段时间香港与其他国家和地区的经济关系可能会提前进行调整。另外,发生在样本期内 (1993 年第一季度到 2008 年第一季度) 的“香港回归”事件依然可能导致香港与其他国家和地区 (尤其是与中国大陆) 之间的经济关系发生变化。这是常系数方法所没有考虑到的,也说明了考虑个体间经济关系时变特征的必要性及我们所提出的时变 LASSO 方法的广泛适用性。

三、基于模拟数据的比较研究

Kapetanios 和 Zikes (2018) 是作为一种新的机器学习方法提出了时变 LASSO 方法,并证明了时变核 LASSO 估计量是时变参数的一致估计量。虽然其利用由时变 DGP 产生的模拟数据将时变 LASSO 与常系数 LASSO 进行了比较,但其所设定的 DGP 并不满足政策评估方法的应用背景:(1) 其设定协变量之间相互独立。在政策评估领域,协变量即控制组个体的观测结果变量,显然不满足独立条件。HCW 方法更是直接假定存在影响所有处理组与控制组个体的共同因子,此时控制组的观测结果变量必然相关。(2) 其仅进行了样本外一期预测。但在政策评估领域,为了描述政策干预效应的动态特征,并更准确地估计平均因果效应,均是进行多期反事实预测。因此,为了将 Kapetanios 和 Zikes (2018) 的时变 LASSO 方法与反事实分析相结合,必须验证 DGP 确为时变且协变量之间相关的条件下,时变 LASSO 方法的多期预测效果。

首先,由以下因子模型生成控制组数据

$$X_{jt} = b_j f_t + \epsilon_{jt}, j = 1, \dots, J; t = 1, \dots, T$$

其中, $b_j \sim \text{uniform}(-1, 1)$, $f_t = 0.6f_{t-1} + \zeta_t$ 。共同因子 f_t 影响所有控制组个体,从而控制组个体之间存在截面相关,这一设定较 Kapetanios 和 Zikes (2018) 的设定更加符合政策评估方法的实际应用情形。

然后,根据如下时变模型产生处理组数据

$$Y_t = \beta_{0t} + \beta_{1t}X_{1t} + \dots + \beta_{Jt}X_{Jt} + u_t \quad (7)$$

参考 Kapetanios 和 Zikes (2018),令截距项和前 5 个控制组的系数为时变参数过程,其余控制组系数为 0,即

$$\begin{aligned} \beta_{jt} &= t^{-1/2} \tilde{\beta}_{jt}, \tilde{\beta}_{jt} = \tilde{\beta}_{j,t-1} + \eta_{jt}, j = 0, 1, \dots, 5 \\ \beta_{jt} &\equiv 0, j = 6, 7, \dots, J \end{aligned}$$

以上 DGP 中的所有干扰项 ϵ_{jt} , ζ_t , u_t 和 η_{jt} 均为 $i.i.d. N(0, 1)$ 。

由以上 DGP 产生处理组与控制组数据 (Y_t, X_{jt}) 后,设定一个假想的政策干预时间点 T_0^* 。由于处理组数据由式 (7) 产生,并未实际接受处理,从而其在今后干预期 $T_0^* + 1, \dots, T$

的观测结果即是 $Y_t=Y_t^0$, $t=T_0^*+1, \dots, T$ 。得到反事实估计 $\hat{Y}_t$, $t=T_0^*+1, \dots, T$ 后, 可以用 $RMSE=\sqrt{\sum_{t=T_0^*+1}^T (Y_t-\hat{Y}_t)^2}$ 衡量每种方法对反事实估计的准确性。

为考察控制组个体数量 J 、样本量 T 以及前干预预期时期数 T_0^* 对实验结果的影响, 根据政策评估领域常见的数据结构, 设定控制组个体数 $J=20, 50$, 时期数 $T=50, 100$, $T_0^*/T=0.5, 0.7, 0.9$ 。实验重复次数为 1000, 用 1000 次反事实估计 $RMSE$ 的均值代表反事实估计的准确性。另外, 依然选择常系数 LASSO 方法为基准, 实验结果见表 3。

表 3 1000 次试验反事实估计的 RMSE 均值

J	T	T_0^*/T	LASSO	HCW	TV-LASSO1	TV-LASSO2
20	50	0.5	1	1. 2734	0.9848	0.9288
		0.7	1	1. 0783	0.8723	0.8902
		0.9	1	0.9701	0.8701	0.8402
20	100	0.5	1	1. 1279	0.9459	0.9069
		0.7	1	1. 0872	0.8324	0.8007
		0.9	1	0.7976	0.6784	0.6698
50	50	0.5	1	—	0.9938	0.9997
		0.7	1	—	0.9316	0.9821
		0.9	1	—	0.8691	0.8935
50	100	0.5	1	—	0.9312	0.9527
		0.7	1	—	0.9228	0.9509
		0.9	1	—	0.7173	0.7279

注: TV-LASSO1 为时变 LASSO 的第一种反事实估计方法, TV-LASSO2 为第二种反事实估计方法。
粗体标示大于 1 的情况, 即劣于基准方法。

首先, 时变 LASSO 反事实估计方法在所有情况下均比常系数 LASSO 方法和 HCW 方法更加准确, 尤其是当 $T_0^*/T=0.9$ 时, 预测准确性提高了 20%~30%。另外, 常系数 LASSO 和 HCW 各有优势, HCW 方法更适合于 $T_0^*/T=0.9$ 的情形。由此可见, 模拟数据与实际数据所得出的结论基本一致。

其次, 时变 LASSO 方法对反事实的估计准确性随着控制组个数 J 增加而下降, 随着时期数 T 以及 T_0^*/T 的增加而上升。这表明时变 LASSO 方法适用于控制组个体数适中, 时期数较长, 且政策干预时间点相对较晚的情形。

四、实证研究

在 2018 年 6 月 23 日, 英国就是否脱离欧盟进行全民公投, 最终英国民众以 51.9% 的比例选择脱离欧盟。英国脱欧对英国、欧盟乃至全球政治经济格局等方面具有多重影响。其中, 汇率作为一国综合竞争力的体现尤其值得关注。金玲 (2016) 认为, 英国脱欧公投之后, 市场短期避险行为引发的金融市场动荡、政治危机导致的投资者信心下降, 以及英国经济面临的长期不确定性均对英镑汇率产生了不利影响。

汇率是由众多因素决定的, 公投只是可能影响汇率的众多因素之一。然而, Nasir 和 Morgan (2017) 认为, 公投提供了一种自然实验, 在这种情况下可以假定在一定的时期内, 公投这一因素具有压倒性的影响力。

另外，市场预期是汇率变化的重要影响因素，在考虑汇率变动时，市场预期往往比宏观经济“基本面”更为重要（Pilbeam，2001）。需要特别指出的是，早在公投之前便有各种与英国脱欧相关的新闻事件公布。例如，2013 年 1 月 23 日，时任英国首相的卡梅伦首次提及脱欧公投；2015 年 5 月 28 日，英国政府向下议院提交并公布了有关“脱欧公投”的议案。各种脱欧相关信息在这一时期的出现为预期转变创造了空间。因此，公投之前英国与其他国家或地区的经济关系可能随着公投的临近以及各种脱欧相关事件的公布而内生地演化。本文根据这一典型事实假定潜在 DGP 是时变的，采用时变 LASSO 面板数据政策评估方法估计“脱欧公投”对英镑汇率的影响。

1. 数据及数据处理

本文收集了主要币种的美元汇率数据，为尽量消除影响汇率波动的次要因素，本文的结果变量 Y_{it} 为各国或地区货币的美元汇率的季度平均值。另外，由于部分国家或地区的汇率自 1995 年第二季度开始公布，本文数据的起止时间为 1995 年第二季度到 2018 年第四季度，即 $T=95$ 。英国脱欧公投发生于 2016 年 6 月 23 日，因此 $T_0=84$ ，即 $T_0/T\approx 0.88$ 。剔除掉含有缺失值的国家之后，本文最终选取了 $N=22$ 个国家和地区，其中，英国是接受政策干预的处理组个体，日本、加拿大、中国台湾等 $J=21$ 个国家和地区作为控制组。数据的描述性统计量见表 4。

表 4 数据的描述性统计量

国家或地区	均值	标准差	最小值	最大值	国家或地区	均值	标准差	最小值	最大值
印度	49.08	9.95	21.41	72.13	南非	8.32	2.95	3.64	15.79
中国台湾	31.43	2.11	25.59	35.03	墨西哥	11.96	3.38	6.14	20.26
新加坡	1.50	0.18	1.22	1.83	马来西亚	3.53	0.48	2.46	4.44
委内瑞拉	3.35	3.08	0.17	9.98	加拿大	1.26	0.18	0.97	1.59
泰国	35.10	5.39	24.64	46.61	韩国	1106.40	150.37	766.20	1606.00
斯里兰卡	105.01	29.89	49.72	175.77	中国香港	7.77	0.03	7.73	7.85
瑞士	1.20	0.25	0.82	1.75	丹麦	6.29	0.90	4.77	8.58
瑞典	7.77	1.09	5.99	10.58	巴西	2.23	0.79	0.91	3.95
日本	108.09	13.79	77.35	139.84	澳大利亚	1.35	0.26	0.94	1.96
中国大陆	7.44	0.87	6.09	8.36	新西兰	1.57	0.32	1.16	2.44
挪威	7.01	1.10	5.08	9.27					

注：数据来源于 WIND 数据库。

2. 实证分析

图 1 为用时变 LASSO 方法估计的各国或地区时变权重 $\hat{\beta}_{it}$ 。由图 1 可知，除 2003 年第三季度至 2013 年第一季度的权重系数变化较为平缓之外，其他时期，尤其是 1995 年第二季度至 2003 年第二季度各国或地区权重系数波动较大。这表明本文假定潜在 DGP 时变是合理的。值得注意的是，随着公投前各种脱欧相关事件的公布，自 2013 年第二季度开始，一小部分国家或地区的权重系数开始发生变化，这一变化平滑且缓慢，说明本文提出的第一种反事实估计方法用前干预期的最后一期，即 T_0 期权重 $\hat{\beta}_{T_0}$ 来估计反事实是合理的。

时变 LASSO 的两种反事实估计方法均与 T_0 期权重 $\hat{\beta}_{T_0}$ 有关，而两种常系数方法的反事实估计均与用前干预期数据估计的常系数权重 β 有关，因此有必要将时变 LASSO 方法的 T_0

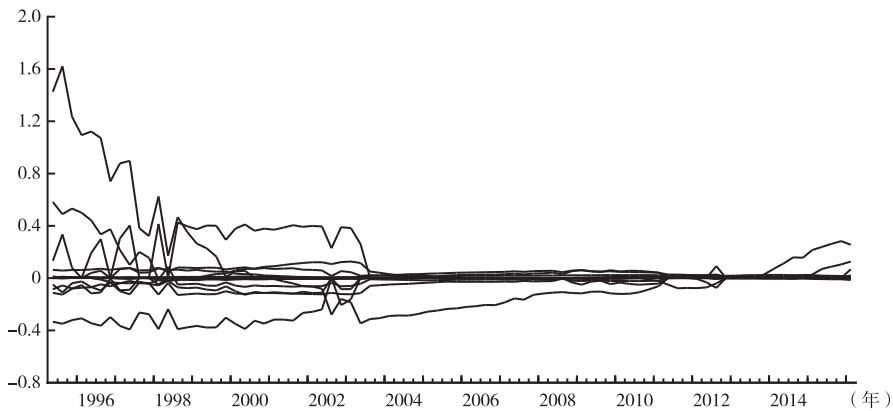


图 1 各国家或地区的时变权重系数

期权重 $\hat{\beta}_{T_0}$ 与常系数方法的常数权重 β 加以比较, 见表 5。由表 5 可知, 两种常系数方法的各国权重符号相同, 大小相似, 而时变 LASSO 方法的权重与两种常系数方法的符号、大小均有较大差异, 这意味着两种常系数方法与时变 LASSO 方法对反事实的估计可能产生很大不同。

表 5 各控制组国家或地区汇率的权重系数

国家或地区	HCW	LASSO	TV-LASSO (T_0)	国家或地区	HCW	LASSO	TV-LASSO (T_0)
印度	—	—	0.0004	南非	—	—	0.0009
中国台湾	—	—	—	墨西哥	0.0126	0.0019	0.0057
新加坡	0.1696	—	0.1088	马来西亚	−0.1187	−0.0721	0.0124
委内瑞拉	0.0137	0.0111	0.1088	加拿大	—	0.0668	—
泰国	—	−0.0001	−0.0152	韩国	0.0001	0.0001	0.0003
斯里兰卡	−0.0008	—	0.0001	中国香港	−0.5282	−0.4827	0.2430
瑞士	−0.1555	−0.0362	0.0925	丹麦	0.0372	0.0077	0.0070
瑞典	—	0.0195	0.0072	巴西	—	—	—
日本	−0.0006	−0.0011	−0.0005	澳大利亚	—	—	—
中国大陆	—	—	—	新西兰	0.1499	0.0783	—
挪威	—	0.0072	—	常数项	4.4732	4.1829	−1.5675

注：时变 LASSO 权重系数为前干预期的最后一期权重。

粗体标示时变 LASSO 方法与两种常系数方法符号相反的权重系数。

真实英镑与用时变 LASSO 方法估计的合成英镑的汇率走势如图 2 (a) 所示。其中, 粗实线为真实英镑汇率走势, 虚线为时变 LASSO 方法估计的合成英镑汇率走势。竖虚线位置为英国脱欧公投发生的时间点。在竖虚线左侧, 真实英镑汇率与合成英镑汇率非常接近, 说明时变 LASSO 方法能很好地拟合公投前英镑汇率的变动趋势。而在竖虚线右侧, 表示时变 LASSO 反事实估计的两条虚线均与实线发生了偏离。第二种时变 LASSO 方法的估计结果认为: 若没有发生脱欧公投, 英镑汇率将基本与公投前持平; 第一种时变 LASSO 方法的估计结果认为: 若没有发生脱欧公投, 英镑汇率将依然保持贬值趋势, 这一结论与 Nasir 和 Morgan (2017) 的观点相符。实线与虚线之间的差值即为脱欧公投对英镑汇率的政策效应,

见图 2 (b)。由图 2 (b) 可知, 两种时变 LASSO 反事实估计方法均认为脱欧公投对英镑汇率的影响仅持续了 1 年, 随后开始反弹。两种方法关于脱欧公投对英镑汇率的平均处理效应估计分别为 0.0064 与 0.0447, 即脱欧公投导致英镑贬值 0.91% 与 6.24%。

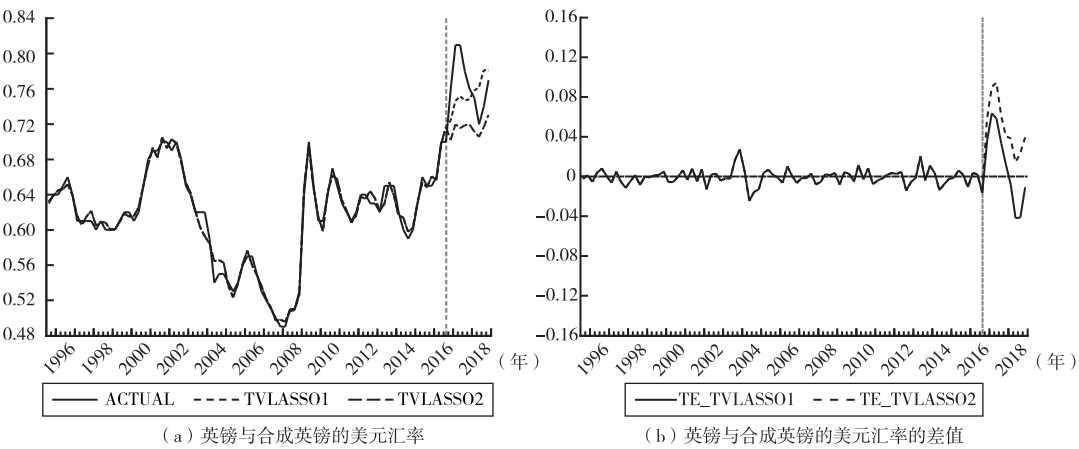


图 2 脱欧公投对英镑汇率的影响 (时变 LASSO 方法)

两种常系数方法的估计结果如图 3 所示。由于表 (2) 中两种常系数方法的权重系数相差不大, 导致图 3 (a) 两条虚线的走势基本相同。在虚线右侧, 实线与两条虚线发生较大偏离, 尤其是两条下降的虚线表明: 若没有发生脱欧公投, 英镑汇率将发生很大幅度的升值, 且影响持续至样本截止时期。但经查阅资料发现, 除脱欧公投外此期间并没有发生可能导致英镑如此大幅升值的事件。因此, 本文认为, 常系数方法有可能会高估脱欧公投对英镑汇率的影响。实线与虚线的差值即为脱欧公投对英镑汇率的影响, 见图 3 (b)。HCW 和常系数 LASSO 方法关于脱欧公投对英镑汇率的平均处理效应估计分别为 0.0928 与 0.0868, 即脱欧公投导致英镑贬值 14.07% 和 12.94%。

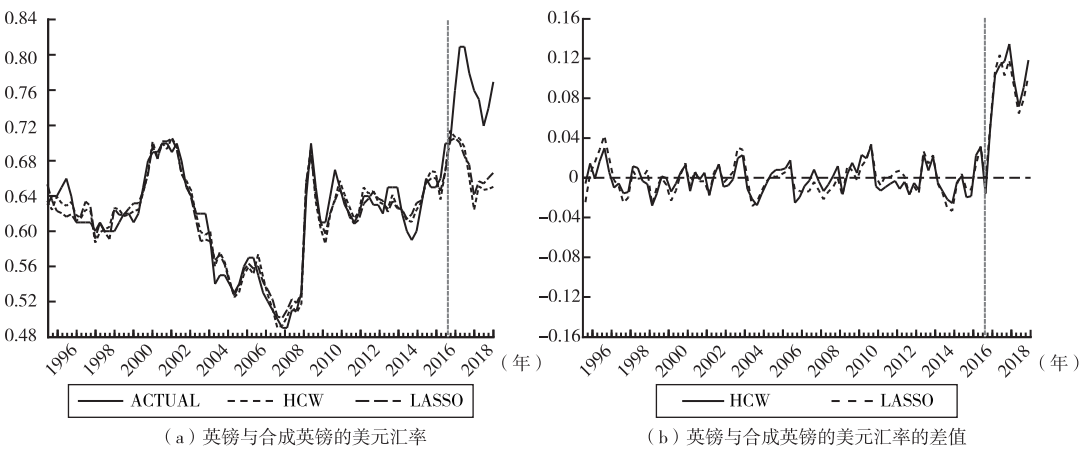


图 3 脱欧公投对英镑汇率的影响 (常系数方法)

3. 安慰剂检验

由表 4 可知, 不同币种的汇率不在一个数量级, 因此, 个体安慰剂检验没有意义。由于

前干预期时期数较长,本文做如下时间安慰剂检验:仅利用前干预期数据,设定一个假想的“公投”时间点,运用三种方法估计政策效应。因为所使用的数据都是真实公投之前的信息,英镑汇率未受公投影响,利用同样的估计方法,如果得到显著的政策效应,则说明,相应的方法存在问题。相反,如果发现没有显著的政策效应,则证明相应的方法可能是有效的。

利用1995年第二季度至2016年第一季度共 $T=84$ 期数据,设定2010年第一季度为假想的“公投”发生时间,即 $T_0=60$ ^①。图4(a)为时变LASSO方法的时间安慰剂检验结果,其中粗实线为真实英镑汇率走势,虚线为时变LASSO方法估计的合成英镑汇率走势。竖虚线位置为假想的“公投”时间点。由图4(a)可知,时变LASSO的两种反事实估计方法均能较好地拟合假想的“公投”之后真实英镑汇率走势。相反,由图4(b)两种常系数方法的反事实估计结果可知,在假想的公投发生之后,两条虚线与实线发生明显的偏离。这表明相对于常系数方法,时变LASSO方法对反事实的估计更加可信。

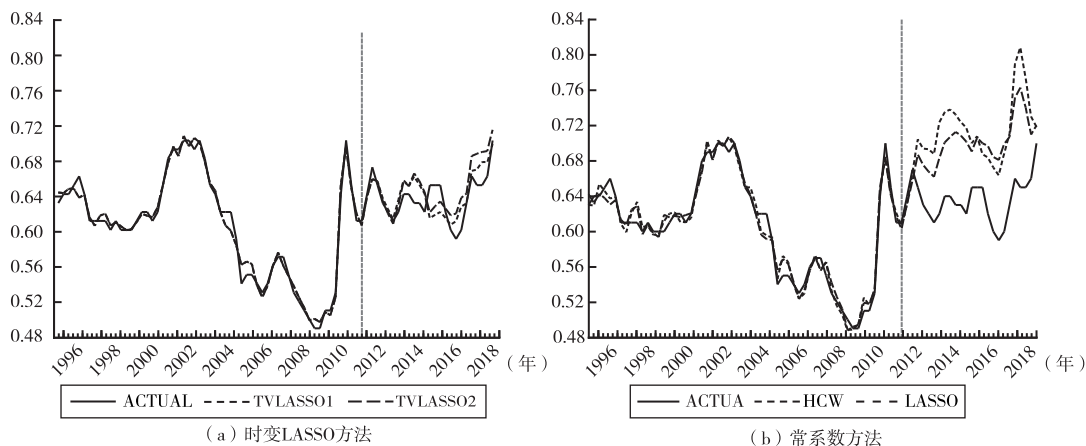


图4 时间安慰剂检验

五、结 论

对于个体间关系时变的面板数据,本文提出了一种估计政策因果效应的反事实机器学习算法——时变LASSO方法,并依据经典文献中的三组实际数据和模拟数据对HCW方法、常系数LASSO方法和时变LASSO方法进行对比研究。研究表明:与两种常系数方法相比,时变LASSO方法对反事实的估计更加准确,尤其是前干预期时期数 T_0 较大时优势更为明显,估计准确性提高了20%~30%。另外,时变LASSO方法适用于样本时期数较长,且政策干预时间点相对较晚的情形。实证研究表明,相对于两种常系数方法,时变LASSO方法的研究结论更加可信,而常系数方法可能高估了脱欧公投对英镑汇率的影响。

本文提出的时变LASSO面板数据政策评估方法是对面板数据政策评估理论的一种扩充和完善。与HCW方法和常系数LASSO方法相比,该方法的优势主要体现在:首先,传统HCW方法与常系数LASSO方法由于假定权重系数为常数,从而只能选择处理组与控制组个体之间关系相对稳定的时期进行研究,例如Hsiao等(2012)的数据截止到2008年第一

① 本文也尝试设定 $T_0=50, 70$, 结果与 $T_0=60$ 相似,由于篇幅所限,不再展示。

季度，避开了 2008 年金融危机时期各国家或地区之间经济关系的可能波动；Abadie 等 (2015) 在挑选潜在控制组时，排除了一些样本期内经历了较大的经济结构变化的国家或地区。而本文提出的时变 LASSO 面板数据政策评估方法允许控制组个体权重随着时间变化而发生改变，尤其适合于政策实施前的信息披露导致预期变化，从而使得个体间经济关系随时间演化的情形。

参 考 文 献

- [1] Abadie A., Gardeazabal J., 2003, *The Economic Costs of Conflict: A Case Study of the Basque Country* [J], *American Economic Review*, 93 (1), 112~132.
- [2] Abadie A., Diamond A., Hainmueller J., 2010, *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program* [J], *Journal of the American Statistical Association*, 105, 493~505.
- [3] Abadie A., Diamond A., Hainmueller J., 2015, *Comparative Politics and the Synthetic Control Method* [J], *American Journal of Political Science*, 59 (2), 495~510.
- [4] Athey S., Bayati M., Doudchenko N., Imbens G., Khosravi K., 2017, *Matrix Completion Methods for Causal Panel Data Models* [R], arXiv preprint arXiv: 1710.10251.
- [5] Card D., Krueger A. B., 1994, *Minimum Wages and Employment: A Case Study of the Fast-food Industry in New Jersey and Pennsylvania* [J], *American Economic Review*, 84 (4), 772~793.
- [6] Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W., Robins J., 2016, *Double Machine Learning for Treatment and Causal Parameters* [R], arXiv preprint arXiv: 1608.00060.
- [7] Doudchenko N., Imbens G. W., 2016, *Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis* [R], NBER Working Paper No. 22791.
- [8] Friedman J., Hastie T., Tibshirani R., 2010, *Regularization Paths for Generalized Linear Models via Coordinate Descent* [J], *Journal of Statistical Software*, 33 (1), 1~22.
- [9] Hsiao C., 2003, *Analysis of Panel Data (2nd edn)* [M], *Econometric Society Monograph* no. 34, New York: Cambridge University Press.
- [10] Hsiao C., Ching H. S., Wan S. K., 2012, *A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China* [J], *Journal of Applied Econometrics*, 27 (5), 705~740.
- [11] Gardeazabal J., Vega-Bayo A., 2017, *An Empirical Comparison between the Synthetic Control Method and Hsiao's et al.'s Panel Data Approach to Program Evaluation* [J], *Journal of Applied Econometrics*, 32 (5), 983~1002.
- [12] Giraitis L., Kapetanios G., Yates T., 2014, *Inference on Stochastic Time-Varying Coefficient Models* [J], *Journal of Econometrics*, 179 (1), 46~65.
- [13] Gregory A., Head A., 1999, *Common and Country-Specific Fluctuations in Productivity, Investment, and the Current Account* [J], *Journal of Monetary Economics*, 44 (3), 423~451.
- [14] Kapetanios G., Zikes F., 2018, *Time-Varying Lasso* [J], *Economics Letters*, 169, 1~6.
- [15] Li K. T., Bell D. R., 2017, *Estimation of Average Treatment Effects with Panel Data: Asymptotic Theory and Implementation* [J], *Journal of Econometrics*, 197 (1), 65~75.
- [16] Nasir M., Morgan J., 2017, *Pre-Brexit: the EU Referendum as an Illustration of the Effects of Uncertainty on the Sterling Exchange Rate* [J], *Journal of Economic Studies*, 45 (5), 910~921.
- [17] Sargent T. J., Sims C. A., 1977, *Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory* [A], in Sims C. et al. (eds.), *New Methods in Business Cycle Research*

[C], Minneapolis: Federal Reserve Bank of Minneapolis.

[18] Tibshirani R., 1996, *Regression Shrinkage and Selection via the Lasso* [J], Journal of the Royal Statistical Society, Series B (methodological), 58 (1), 267~288.

[19] Pilbeam K., 2001, *Economic Fundamentals and Exchange Rate Movements* [J], International Review of Applied Economics, 15 (1), 55~64.

[20] Xu Y., 2017, *Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models* [J], Political Analysis, 25 (1), 57~76.

[21] 洪永森:《提倡定量评估社会经济政策,建设中国特色新型经济学智库》[J],《经济研究》2015年第12期。

[22] 金玲:《英国脱欧:原因、影响及走向》[J],《国际问题研究》2016年第4期。

A Time Varying Panel Data Approach for Program Evaluation Based on Machine Learning

Gao Huachuan Bai Zhonglin

(School of Statistics, Tianjin University of Finance and Economics)

Research Objectives: Based on the panel data of time-varying relationships among individuals, a counterfactual machine learning algorithm, time-varying LASSO method, is proposed to estimate the policy causal effect. **Research Methods:** Using the real data in classical literatures whose data generation process is unknown and time-varying simulated data, the time-varying LASSO method is compared with two constant coefficient methods (HCW method and LASSO method), and by taking analysis of impact of “Brexit” on sterling exchange rate as an example, the dynamic behavior of policy causal effect is studied when policy information is disclosed in advance. **Research Findings:** Compared with the two constant coefficient methods, the time-varying LASSO method can estimate the counterfactuals more accurately, and is more suitable for program evaluation with long sample periods and with relatively late intervention time, which improves the timeliness of policy evaluation. Empirical studies show that the constant coefficient method could overestimate the impact of Brexit. **Research Innovations:** Propose a time-varying LASSO panel data policy evaluation method based on machine learning, and two counterfactual prediction methods. **Research Value:** For the policy of which the information is disclosed in advance, it provides a tool to evaluate the dynamic behavior of causal effect, and improves the theoretical methods of panel data program evaluation.

Key Words: Panel Data; Program Evaluation; Machine Learning; Time Varying LASSO; Brexit

JEL Classification: C10

(责任编辑:韩 君)