案例分析：世界杯是否影响卡塔尔新冠疫情？

基于增广合成对照方法

史长浩

中国人民大学统计学院

2022 年 12 月 19 日

RENMIN UNIVERSITY OF CHINA

案例背景

- 卡塔尔世界杯于 2022 年 11 月 20 日开幕，吸引了超 120 万球迷到场看球，为疫情以来首届不限观众人数的大型体育赛事。而一年前，同样在疫情期间举办的日本夏季奥运会、北京冬奥会则是“全封闭”的大型赛事。卡塔尔世界杯几乎全面放开，恢复了 2019 年之前的赛事形态。
- 卡塔尔当局自 11 月 1 日起取消了几乎所有与新冠疫情相关的旅行限制措施，包括抵达前的强制核酸检测，不再提供疫苗证明、注册相关健康码的信息，以及在比赛时不需要佩戴口罩等。而球迷们主要来自德、意、法、日、巴西等疫情仍然高位运行的国家，他们的到来，是否会让卡塔尔这个仅有 250.5 万人的小国，变成新的疫情暴发地？
- 本文尝试基于增广合成对照方法研究这一问题。

数据来源

- 本案例所用数据来自公益项目 Our World in Data [Source: ourworldindata.org].
- 从中截取了 2022 年 10 月 2 日到 12 月 13 日共 73 天的新增病例数据；因为卡塔尔地处连接欧、亚、非三洲的中东地区，故只选择地处这三洲的其他国家作为对照，剔除一些无效记录之后，得到 127 个国家作为对照组。
- 卡塔尔世界杯于 11 月 20 日开幕，故认为自开幕日起，卡塔尔的新增确诊人数受到赛事干预。本文关心的问题即：若不举行世界杯，11 月 20 日之后卡塔尔的新增病例数是否与当前观测有显著区别？

数据描述

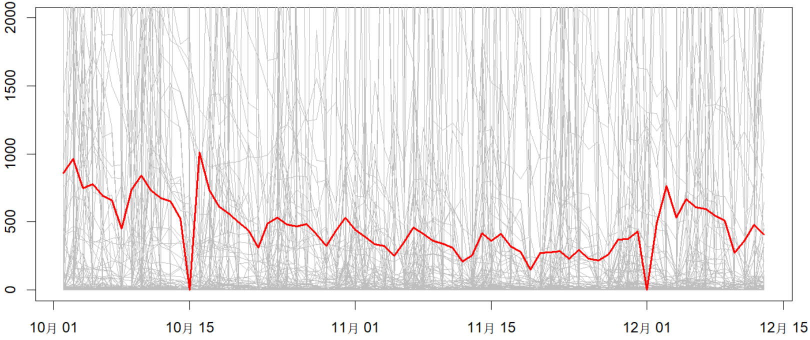


图 1: 128 个国家的日增病例数据可视化。横轴为日期，纵轴为确诊人数（个）。红色实线对应卡塔尔。

数据描述

表 1: 10 月 2 日至 12 月 13 日 128 个国家新增新冠确诊病例数。

	10.2	$\cdots$	11.19	11.20	$\cdots$
Qatar	$Y_{1,1}(0)$	$\cdots$	$Y_{1,T_0}(0)$	$Y_{1,T_0+1}(1)$	$\cdots$
German	$Y_{2,1}(0)$	$\cdots$	$Y_{2,T_0}(0)$	$Y_{2,T_0+1}(0)$	$\cdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\cdots$
Japan	$Y_{N,1}(0)$	$\cdots$	$Y_{N,T_0}(0)$	$Y_{N,T_0+1}(0)$	$\cdots$

- 表中 $Y_{i,j}(0)$ 表示未受世界杯干预的单元，而 $Y_{i,j}(1)$ 表示受到世界杯干预的单元。为了评估世界杯对卡塔尔国内疫情的影响，本质上需要对任意 $t > T_0$ 估计 $Y_{1,t}(1) - Y_{1,t}(0)$，核心问题在于估计缺失值 $Y_{1,t}(0)$，可以基于其他国家的病例数“构造”一个对照组 $\hat{Y}_{1,t}(0) = \sum_{n=2}^{N} \gamma_n Y_{n,t}(0)$，这即是合成对照法的思想。

合成对照方法

合成对照方法归结为求解如下约束优化问题 [Abadie, Diamond and Hainmueller, 2015]：

$$
\begin{aligned}
&\min_{\gamma} \left\| \boldsymbol{V}_{\boldsymbol{x}}^{1/2} \left(\boldsymbol{X}_{1\cdot} - \boldsymbol{X}_{0\cdot}' \gamma \right) \right\|_2^2 + \zeta \sum_{W_i=0} f\left(\gamma_i\right), \\
&\text{subject to} \quad \sum_{W_i=0} \gamma_i = 1, \\
&\qquad \gamma_i \geq 0 \quad i : W_i = 0.
\end{aligned} \tag{1}
$$

其中 $\boldsymbol{V}_{\boldsymbol{x}} \in \mathbb{R}^{T_0 \times T_0}$ 是对称重要性矩阵（symmetric importance matrix），为了符号简单计，本文取其为单位阵；$f\left(\gamma_i\right)$ 是对权重的分散惩罚。

增广合成对照方法

设 $\hat{m}_{iT}$ 是 $Y_{iT}(0)$ 的估计量，则定义 $Y_{1T}(0)$ 增广合成对照方法（ASCM）估计量为 [Eli Ben-Michael, Avi Feller and Jesse Rothstein, 2021]：

$$\begin{aligned}\hat{Y}_{1T}^{\text{aug}}(0) &= \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} Y_{iT} + \left(\hat{m}_{1T} - \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} \hat{m}_{iT} \right) \\ &= \hat{m}_{1T} + \sum_{W_i=0} \hat{\gamma}_i^{\text{scm}} \left(Y_{iT} - \hat{m}_{iT} \right),\end{aligned} \tag{2}$$

其中，$\hat{\gamma}_i^{\text{scm}}$ 是前文 SCM 方法得到的权重，标准的 SCM 估计是该估计的一个特例，只需令 $\hat{m}_{iT}$ 为常数。原则上 SCM 只在处理前结果拟合较好时才可使用，而 ASCM 通过权重的修正拓宽了 SCM 的使用场景。

Ridge ASCM

- 若 $\hat{m}(\boldsymbol{X}_i)$ 通过岭回归确定，即 $\hat{m}(\boldsymbol{X}_i) = \hat{\eta}_0^{\text{ridge}} + \boldsymbol{X}_i'\hat{\eta}^{\text{ridge}}$，其中

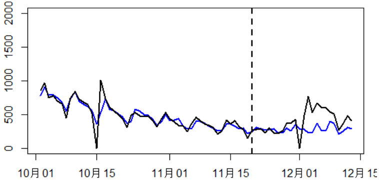
$$\left\{\hat{\eta}_0^{\text{ridge}}, \hat{\eta}^{\text{ridge}}\right\} = \arg\min_{\eta_0,\eta} \frac{1}{2}\sum_{W_i=0}\left(Y_i - (\eta_0 + X_i'\eta)\right)^2 + \lambda^{\text{ridge}}\|\eta\|_2^2. \tag{3}$$

- 则得到岭 ASCM 估计：

$$\hat{Y}_{1T}^{\text{aug}}(0) = \sum_{W_i=0}\hat{\gamma}_i^{\text{scm}} Y_{iT} + \left(\boldsymbol{X}_1 - \sum_{W_i=0}\hat{\gamma}_i^{\text{scm}}\boldsymbol{X}_{i\cdot}\right)\cdot\hat{\eta}^{\text{ridge}}. \tag{4}$$

- Ridge ASCM 是案例分析中主要使用的方法。

结果分析



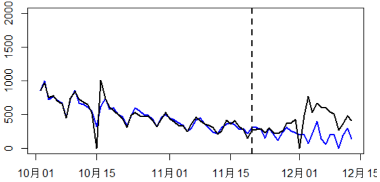


图 2: 卡塔尔确诊病例随时间变化曲线。黑色为实际观测，蓝色为合成对照（左：SCM；右：Ridge ASCM）。

- Ridge ASCM 相较于 SCM 实现了对处理前结果更好的拟合：$\mathsf{MSE}^{\mathsf{ascm}} = 6352.01$，$\mathsf{MSE}^{\mathsf{scm}} = 9421.58$。
- Ridge ASCM 相较于 SCM 的处理组中的平均处理效应 $\mathsf{ATT} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}(Y_{1,t}(1) - Y_{1,t}(0))$ 更大：$\mathsf{ATT}^{\mathsf{ascm}} = 186.67$，$\mathsf{ATT}^{\mathsf{scm}} = 111.08$。

结果分析

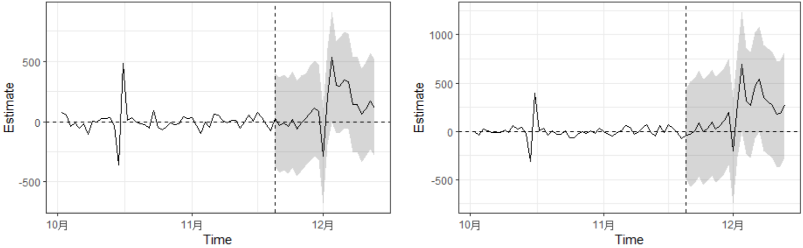


图 3: 卡塔尔确诊病例实际观测与合成对照之差随时间变化曲线。（左：SCM；右：Ridge ASCM）。

- 在 sharp 零假设（$H_0: Y_{1,t}(0) = Y_{1,t}(1), \forall t > T_0$）下，可以作 permutation test，检验的 p 值分别为 0.013 (SCM) 和 0.023 (ASCM)，即两种方法都表明处理组的处理效应是显著的，世界杯确实使得卡塔尔的疫情更加严重。

结果分析

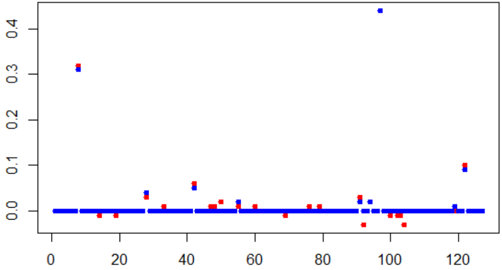


图 4: Ridge ASCM 与 SCM 合成对照组的权重对比。横轴为国家编号，纵轴为权重大小。蓝色对应 SCM，红色对应 ASCM。

- SCM 的权重更加稀疏，127 个国家之中，只有 9 个国家权重不为 0，且严格非负；Ridge ASCM 允许出现负权，且有 23 个国家具有非零的权。

讨论

- 合成对照方法还可通过引入额外的协变量信息来提高拟合的效果，但在本案例中并无显著提升。
- 关于世界杯是否真的对卡特尔当地的疫情防控产生不利影响，由于存在诸多不可观测的混杂，该问题极其复杂，不同的学者有不同的观点，本文仅从数据层面给出一种尝试性地回答，并不全面。

Thanks!