

網路輿情聲量對高端疫苗施打量的影響

Internet opinion and sentiment on the willingness of getting the
MVC covid-19 vaccine

政大經研所碩二 楊士逸



目錄

1. 引言
2. 摘要
3. 文獻回顧
4. 資料處理
5. 挑選介入效果
6. 實證結果
7. 安慰劑檢定
8. 結論
9. 研究限制與未來方向



引言

- 2021 年 5 月中，台灣爆發了第一波新冠疫情，經歷了長達約兩個多月的三級警戒
- 因為這一波疫情，民眾對新冠疫苗的需求提高，網路上對於疫苗的討論度也不斷地提升，例如：疫苗的效果、疫苗施打的陰謀論、打疫苗後的副作用與不良反應等等
- 每天都會看到鋪天蓋地的新冠疫苗相關資訊，因此我們對於網路輿情的正面、負面或中立聲量是否直接影響民眾對疫苗的施打意願產生興趣，目前也未看過研究此議題的文獻



摘要

- 本文利用合成控制法 (synthetic control method)，以 AZ、BNT、Moderna 三種新冠疫苗模擬高端疫苗施打量，研究網路輿情負面聲量與高端疫苗施打量之間的因果關係
- 以台灣衛生福利部疾病管制署 (CDC) 提供的疫苗施打期數為基準，在施打區間上做些許的調整，並搭配 OpView 社群口碑資料庫的聲量資料
- 探討 2021 年 9 月 24 日至 10 月 2 日之間，社群媒體上的負面聲量介入效果 (treatment effect)，是否會對之後的高端疫苗施打量有顯著的影響



文獻回顧

- Abadie et al. (2010) 探討加州菸草管制政策 (Proposition 99) 對其菸草銷售量的影響
- 挑選一些變數，使用合成控制法，以美國的 38 個州模擬加州 1970 到 2000 年的菸草銷售量
- 產生一個模擬加州當作對照組，討論若未受到介入效果，加州的菸草銷售會如何變化

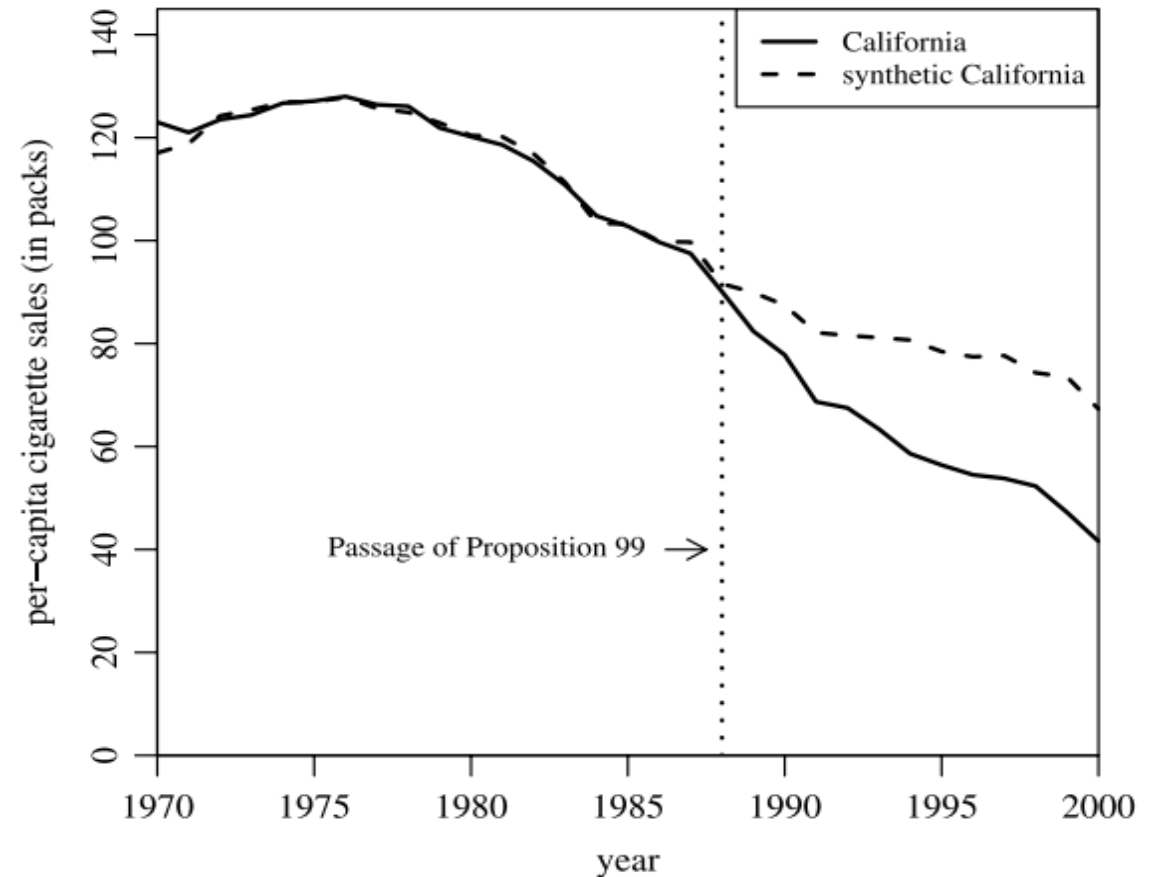


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.



文獻回顧

- 以下為 Abadie 挑選的變數，使加州和「模擬加州」在介入前足夠相似
- 控制變數放平均數，outcome variable 放落後期

Table 1. Cigarette sales predictor means

Variables	California		Average of 38 control states
	Real	Synthetic	
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81



資料處理：OpView - 聲量資料

- OpView 資料庫撈取資料的方法為設定標題關鍵字，得到的資料包括：內容、主文回文數、點閱按讚數、發佈時間、聲量強度、情緒標記等
- 資料來源包括社群網站、新聞媒體、討論區，例如：PPT、FB、Dcard
- 情緒標記是由 OpView 公司的演算法判斷此篇主文或回文為正面、負面還是中立的貼文
- 高端關鍵字：EUA、三期、股價、登記、我OK你先打、副作用、死亡、免疫橋接、第一劑、第二劑、不良反應、出國、支持國產疫苗、保護力、解盲、球員兼裁判、反應(正、負、中立的數量平均)



資料處理：新冠疫苗資料

- 台灣新冠疫苗第一、二劑是採預約制，且預約後有一段明定的施打期間，並非隨到隨打
- 因此我們根據衛福部公布的施打日期來做期間的劃分，探討 t 期間在社群媒體發生的「所有事件」，對 $t + 1$ 期疫苗施打量的影響
- 高端疫苗在 8 月 23 日才開始施打，因此我們把 CDC 的 06 期當作本論文第 1 期
- 而 CDC 明訂的施打日期不連續，但這段時間仍有疫苗的施打量被登記，推測可能為國外補登和殘劑的施打



資料處理：新冠疫苗資料

- 2、4、7、10期為自行創造的期數，使施打期間變成連續的資料

	CDC 明定的施打期數	施打日期	天數
第 1 期	06 期 (含加開)	8/23 - 8/29	7
第 2 期	-	8/30 - 9/2	4
第 3 期	07 期 (含加開)	9/3 - 9/10	8
第 4 期	-	9/11 - 9/14	4
第 5 期	08 期	9/15 - 9/23	9
第 6 期	09 期	9/24 - 10/2	9
第 7 期	-	10/3 - 10/5	3
第 8 期	10 期	10/6 - 10/14	9
第 9 期	11 期	10/15 - 10/21	7
第 10 期	-	10/22 - 10/28	7
第 11 期	12 期	10/29 - 11/3	6
第 12 期	13 期	11/4 - 11/12	9



資料處理

總結我們會使用的資料：

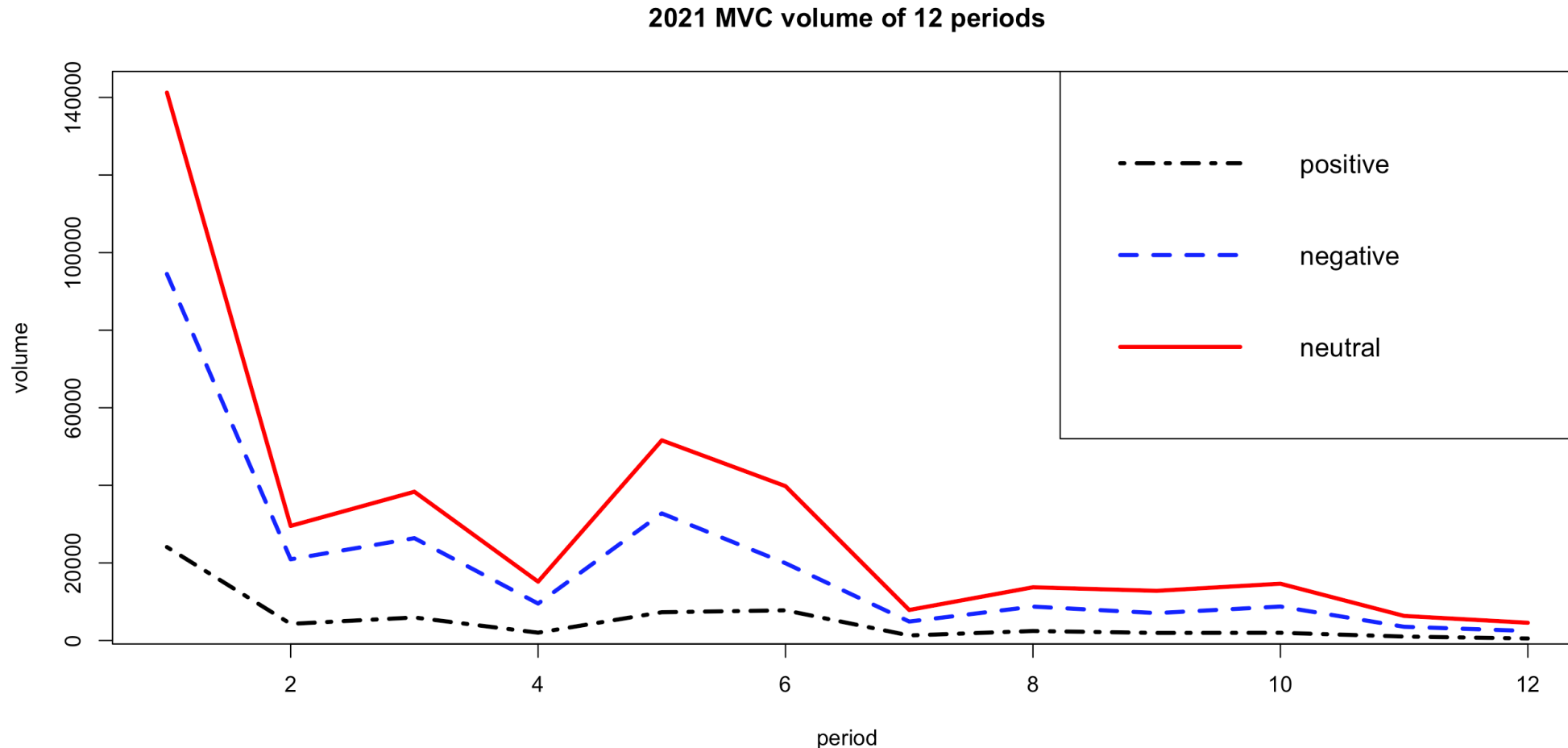
- AZ 疫苗：12 期的 OpView 正、負、中立聲量貼文數，12 期的施打量
- BNT 疫苗：12 期的 OpView 正、負、中立聲量貼文數，12 期的施打量
- Moderna 疫苗：12 期的 OpView 正、負、中立聲量貼文數，12 期的施打量
- MVC 疫苗：12 期的 OpView 正、負、中立聲量貼文數，12 期的施打量

以上資料期間為 8/23 – 11/9



挑選介入效果

- 下圖為高端疫苗的正面、負面、中立聲量總貼文數折線圖，可以看出相較於中立和負面聲量，正聲量幾乎沒有波動，代表負面聲量的效果會主宰正面聲量，因此我們的介入效果打算從負面聲量下手。



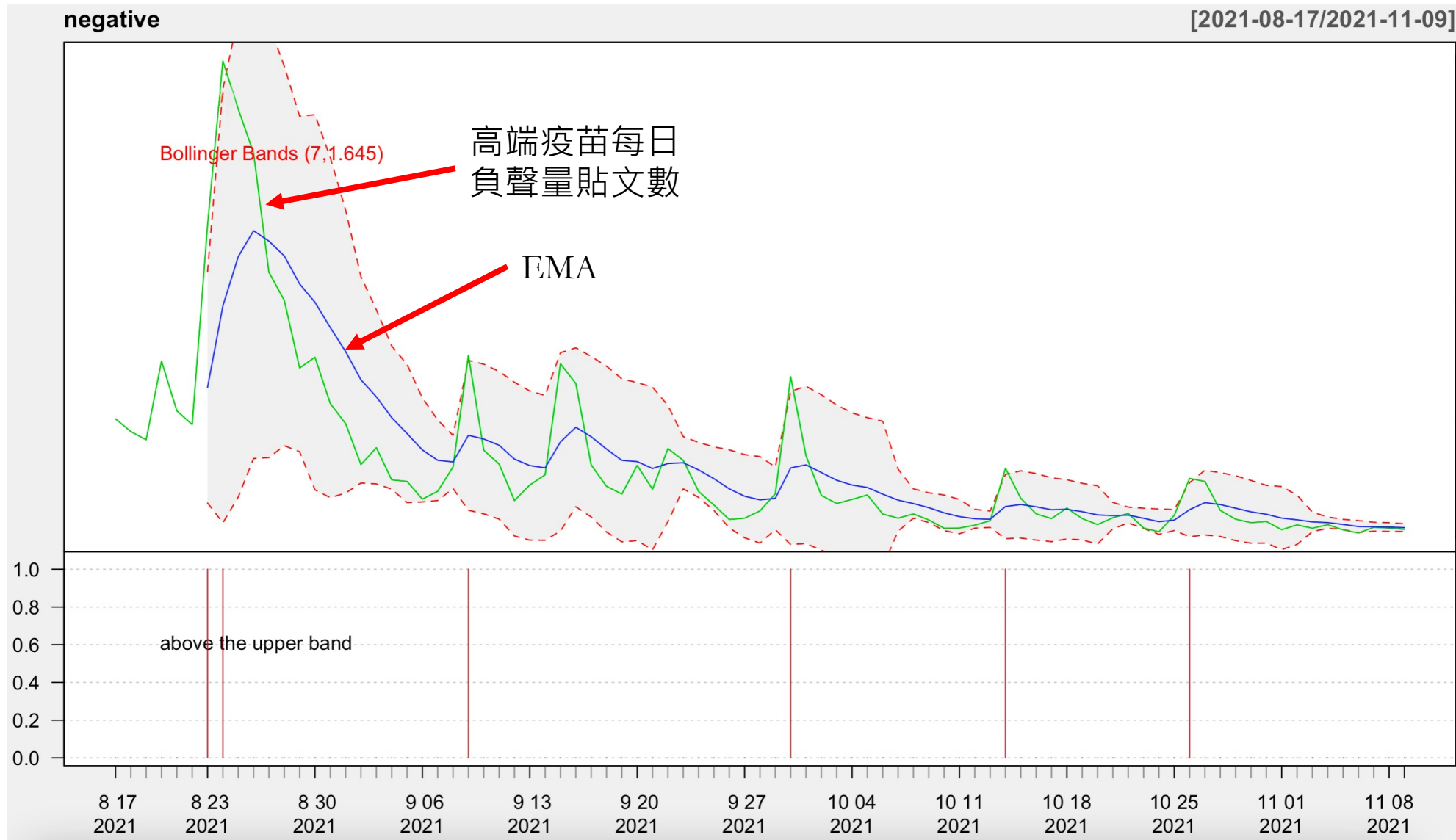
挑選介入效果

- 第 5、6 期負聲量貼文數為高點，特別是在第 6 期發生了幾個嚴重的事件：美國、英國、日本相繼把高端疫苗列為入境不認證疫苗
- 我們認為第 6 期是負面介入效果的好選擇，打算使用布林通道 (Bollinger Bands) 來驗證資料產生的結果與經濟直覺挑選的重大事件 (event) 之時間點是否一致
- 我們畫出高端疫苗每日負面聲量貼文數的布林通道，以通過布林通道上緣為基準，若通過則代表這一期的負聲量數「影響夠大」
- 不使用 SMA (simple moving average)，使用 EMA (exponential moving average)，差別在於越近的期數要給越大的權重，較符合聲量的性質
- 7 日均線，標準差設定為 1.645，代表常態分配的 90 %



挑選介入效果

- 在第 6 期 9/30 時突破上通道，與先前推測相符



挑選介入效果

- 各期上突破總數，1、3、6、8、10 期皆有突破上通道

	施打日期	上突破總數
第 1 期	8/23 - 8/29	2
第 2 期	8/30 - 9/2	0
第 3 期	9/3 - 9/10	1
第 4 期	9/11 - 9/14	0
第 5 期	9/15 - 9/23	0
第 6 期	9/24 - 10/2	1
第 7 期	10/3 - 10/5	0
第 8 期	10/6 - 10/14	1
第 9 期	10/15 - 10/21	0
第 10 期	10/22 - 10/28	1
第 11 期	10/29 - 11/3	0
第 12 期	11/4 - 11/12	0



挑選介入效果

- 在挑選介入期間時有幾點要特別注意，若介入前 (pre-treatment) 的期間不夠長，容易模擬得太差，介入後 (post-treatment) 期間太短，沒辦法看政策效果的持續性
- 1、3、8、10 期雖然也有突破上通道，但基於上述的理由，第 3 和 8 期不能挑選為介入期間。而第 1 期高端疫苗剛開始施打，沒有介入前的資料，也無法當成介入效果發生的期間
- 第 8 期雖然有突破上通道，但在這期已經開始廣泛施打各種疫苗，高端疫苗的介入效果可能會污染 (contaminate) 到對照組疫苗 (donor pool)，若受到污染，介入效果會不夠純粹 (pure)



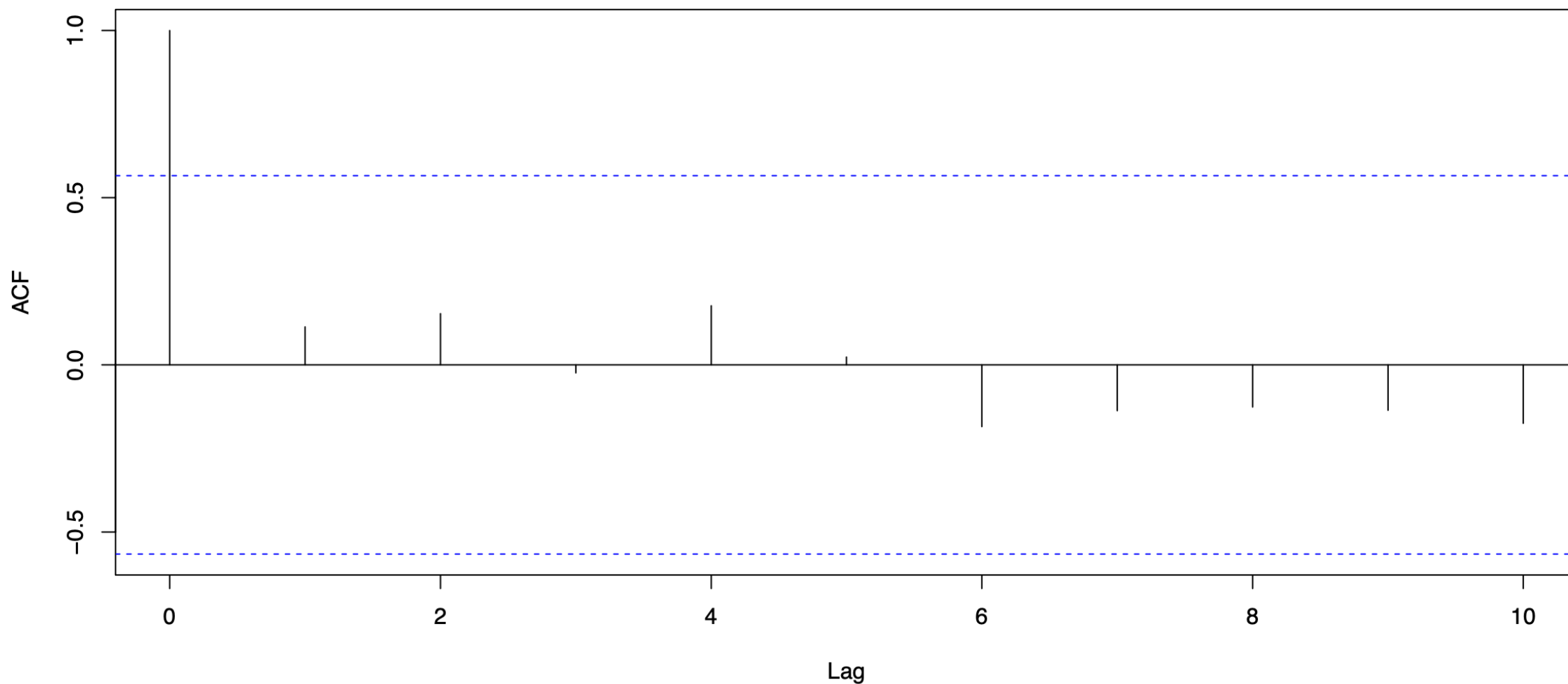
資料處理

- 接著要確認一件事，究竟 $t + 1$ 期高端疫苗施打量的改變，是只受到 t 期社群媒體事件的影響，還是也會受到 $t - 1$ 、 $t - 2$... 期的影響，若會受到更早期的事件影響，要釐清因果關係就會更複雜
- 簡單來說，我們需要找到負聲量資料跨期之間的相關，使用 Time Series 的 ACF (Auto Correlation Function)、PACF (Partial Auto Correlation Function) 驗證
- 由 ACF、PACF 知，負聲量跨期間線性相關性在統計上不顯著，故 t 期的負聲量不會有 $t - 1$ 期負聲量的遞延效果



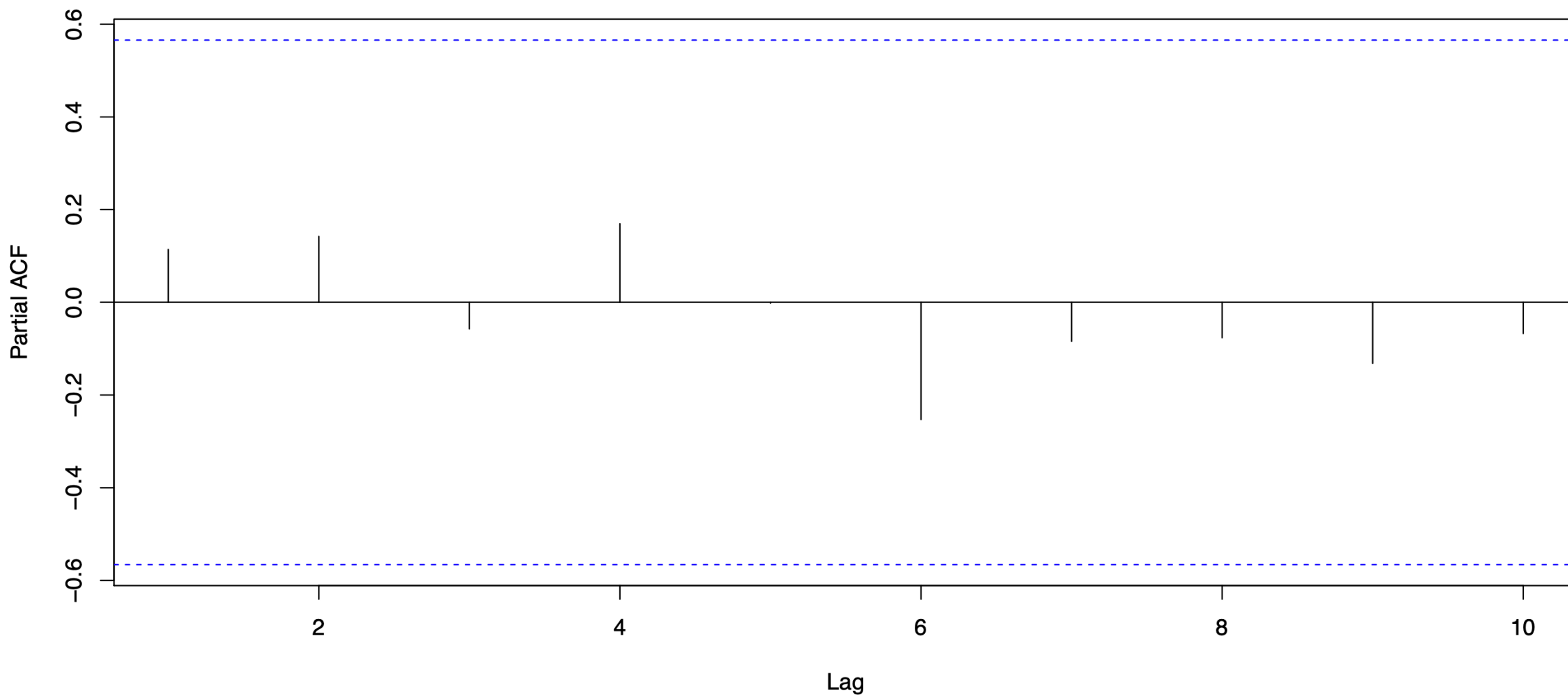
資料處理

- ACF：本期與落後 1~10 期之間的相關性，皆不超過 95% C.I.



資料處理

- PACF：本期與落後 1~10 期之間的「偏」相關性，皆不超過 95% C.I.



資料處理

給定一個重要的假設：

- 我們假設社群媒體影響的效果的大部分為「即時」的， t 期的負聲量會導致民眾在 t 期的意願馬上產生變化
- 而因為衛福部疫苗施打政策的關係，決策的改變大部分會在 $t + 1$ 期反應出來
- 但不是所有接收到影響的人都剛好在下一期施打，所以部分的變化在 $t + 1$ 期後才會有反應，但這樣只會影響到介入效果的大小，不會影響我們的結果



實證結果 - 變數

- 第 1 期高端疫苗才開始施打，第 1 期以前沒有真實施打案例的聲量

	高端	模擬高端	對照組平均
2~6 期平均負聲量貼文數	21903.6	17194.6	16150.93
2~6 期平均正聲量貼文數	5458.4	3434.75	3896.20
2~6 期平均中立聲量貼文數	34888.8	29383.2	30859.93
第 2 期高端疫苗施打量	2467.0	33190.25	30826.67
第 3 期高端疫苗施打量	57034.0	26301.2	420812.00
第 4 期高端疫苗施打量	8393.0	14713.64	58491.33
第 5 期高端疫苗施打量	3286.0	81712.48	463128.67
第 6 期高端疫苗施打量	517897.0	514887.13	734009.00



實證結果 – 權重

3 個疫苗的權重 w 為

- AZ 疫苗 3 月 22 日就開始施打，與高端施打期間差距很大，權重為 0 %
- BNT 疫苗 9 月 22 日 (第五期) 才開始施打，先前為健保卡補登資料，故佔比也相對較小

疫苗	AZ	BNT	Moderna
權重	0 %	3.5 %	96.5 %

8 個變數的權重 v 為

- 負聲量的權重，與介入效果相互呼應
- 5 期的施打量落後期權重呈遞增狀態，代表施打量越靠近介入效果影響越大

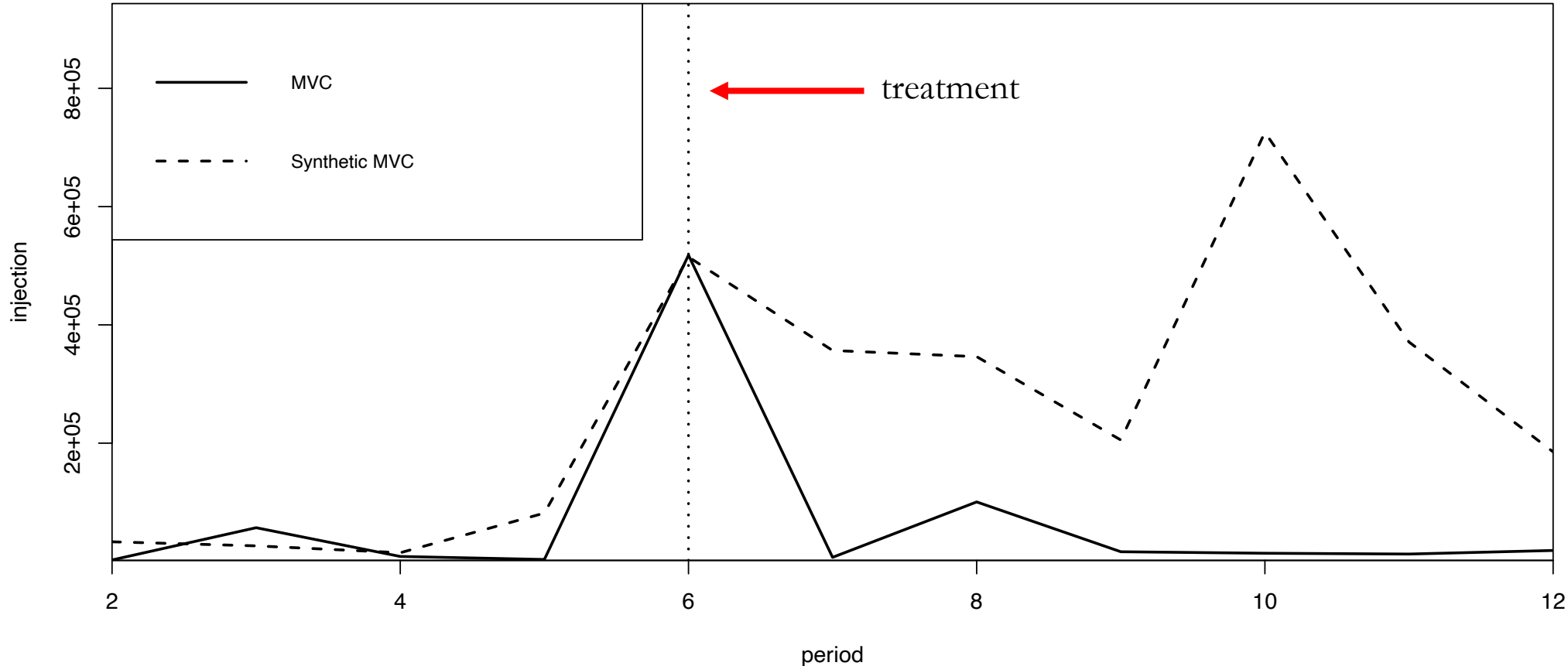
疫苗	negative volume	positive volume	neutral volume	injection period 2	injection period 3	injection period 4	injection period 5	injection period 6
權重	17 %	0.3 %	0.1 %	0.5 %	4.1 %	17.4 %	15 %	45.5 %



實證結果 – 模擬

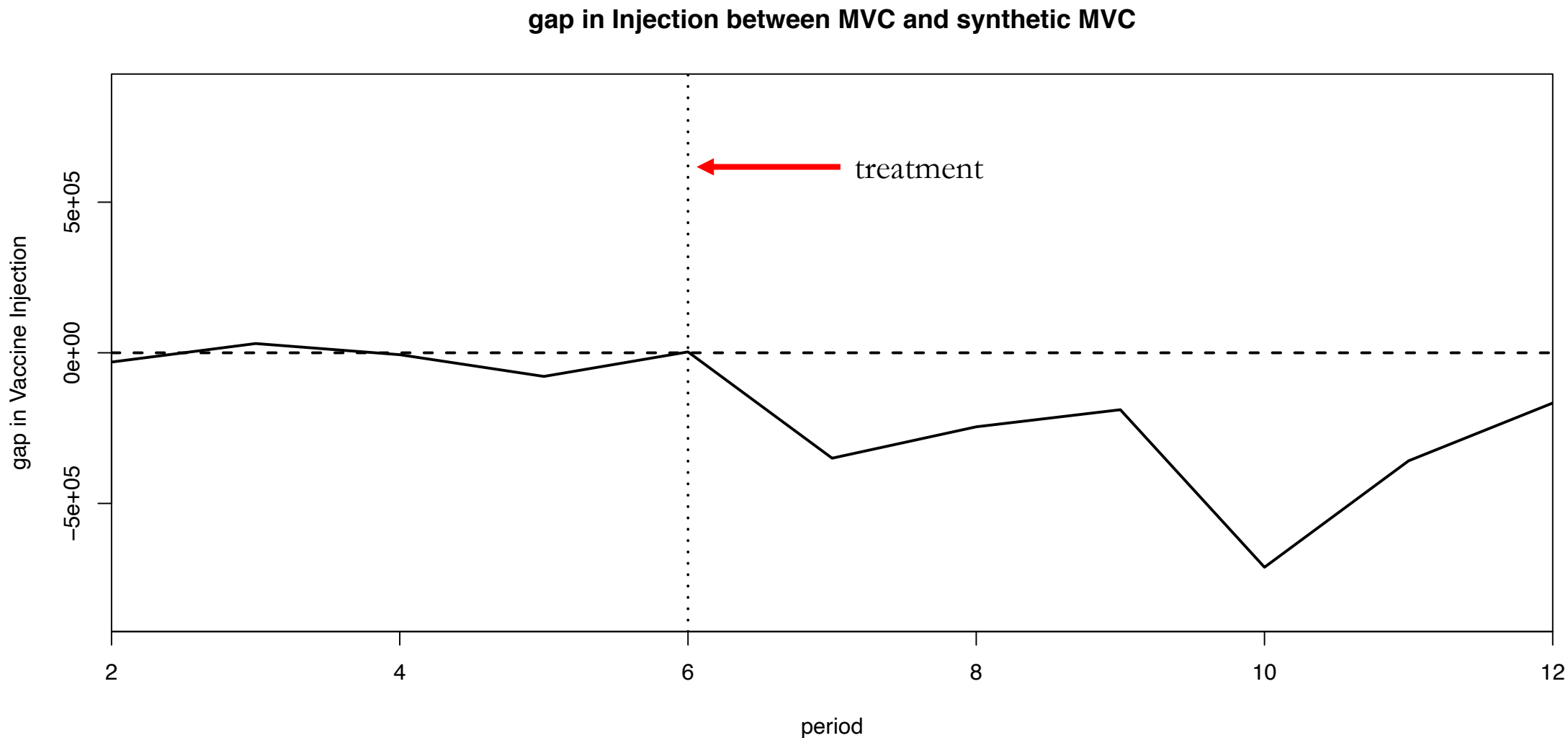
- 極小化預測均方誤差 (MSPE) 的期間挑選 2 – 6 期，第 6 期發生負面介入效果後，真實的高端疫苗施打量下降了 98 %，模擬的施打量下降了 29 %，相比之下多下降了約 70 %，代表若無介入效果，真實施打量不會下降這麼多

MVC Injection vs. synthetic MVC Injection



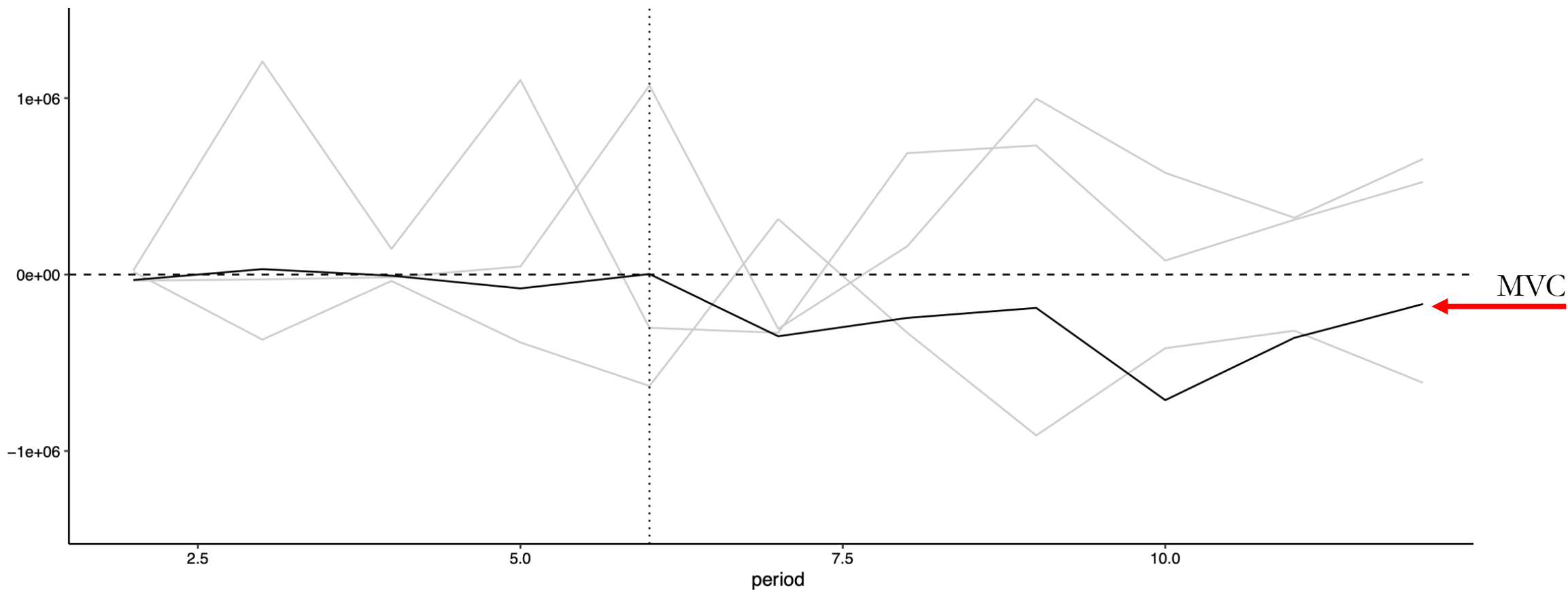
實證結果

- 介入前的預測均方誤差幾乎趨近於 0，代表對照單位模擬的結果已經非常優異了



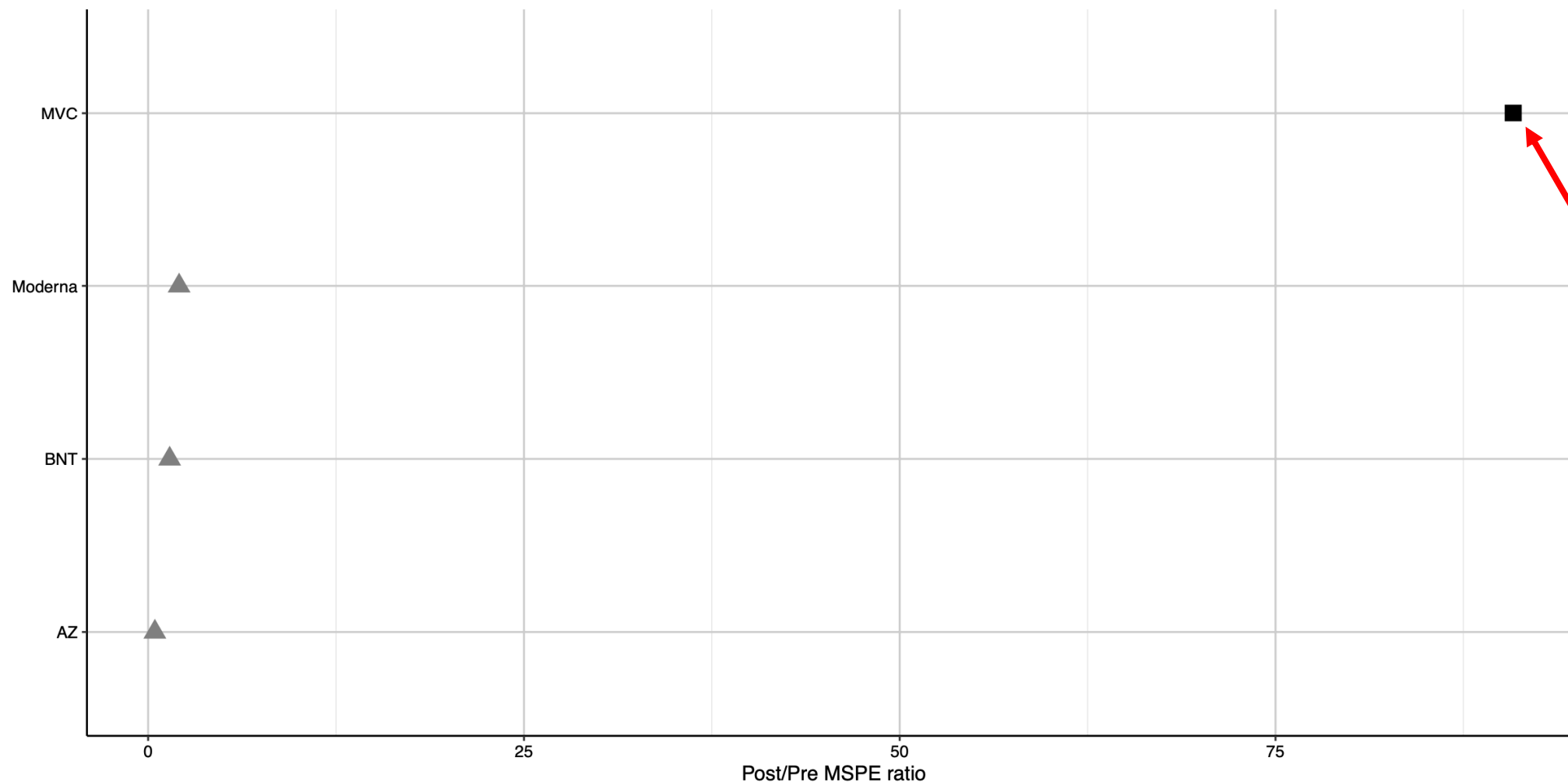
安慰劑檢定 - 1

- 三條灰色線為把對照組中的三種疫苗分別轉換為實驗組，跑合成控制法得到的預測均方誤差
- 在介入前預測均方誤差夠小的線，只有高端疫苗為實驗組所產生的黑色線



安慰劑檢定 - 2

- 高端疫苗的 $ratio\ of\ MSPE = \frac{post\ treatment\ MSPE}{pre\ treatment\ MSPE}$ 在最右邊，遠遠大於另外三個疫苗



結論

1. 本文利用合成控制法，證實了社群媒體的負面聲量與高端疫苗施打量下降之間，具有因果關係
2. 使用安慰劑檢定證明了此結果具有統計顯著性
3. 意味著若網路上的負面事件影響力夠大，且通過統計方法驗證後，皆可能會對社會產生實質的負面影響，因此消息是否準確就顯得極為重要
4. 以疫苗廠牌分群使用合成控制法，並釐清出別人未研究過的因果關係是本文最大的貢獻



研究限制與未來方向

1. 總期數只有 12 期，期間過短，若拉長區間或許可以使預測更精準
2. 對照單位 (control unit) 只有三個，使用 SCM 可能會發生差擬合 (poor matching) 問題，也會使安慰劑檢定產生 p-value 統計顯著性有疑慮
3. 未來的研究方向可以嘗試其他國家的新冠疫苗，來提高對照單位的數量，但要注意控制變數與施打期間的問題
4. 若要再提升本文合成控制法的擬合程度，可以嘗試改良過後的合成控制法

