The Differential Impact of the Hong Kong National Security Law on Political Sensitivity Bias in Local Opinion Polls

Dianyi Yang [1] and Leike Huang [2]

[1]Department of Government, London School of Economics and Political Science
[2]Oxford School of Global and Area Studies, University of Oxford

November 5, 2023

Abstract

Political sensitivity bias refers to the deviation of observed poll results from the genuine public opinion due to perceived government pressure on respondents. This bias can mislead analysts and politicians into making inaccurate assessments of political climates. Hong Kong, having experienced a rapid reduction in political freedoms under the Hong Kong National Security Law (HKNSL), provides a unique case for studying political sensitivity bias where the impacts of other transitory factors can be isolated. We show alternative evidence of such bias in the city by using public opinion poll data and the newly invented Synthetic Difference-in-Differences (SDID) estimator to investigate the differential impact of the law on local opinion polls. Our results indicate a significant positive differential treatment effect on polls regarding attitudes towards Taiwan and Tibet independence compared with polls of lower political sensitivity. This implies that the Hong Kong National Security Law has contributed to the increased political sensitivity bias in these polls, thereby proving the existence of such bias.

Keywords— Hong Kong National Security Law, political sensitivity bias, public opinion, preference falsification

1 Introduction and Definition

Following the de-escalation of the 2019-20 Hong Kong protests, China passed the Hong Kong National Security Law (NSL) at the end of June 2020.[1] The law criminalises four types of activities that threaten the security and the stability of the Special Administrative Region or the People's Republic of China[2]. This was immediately followed by the disbandment of Demosisto[3] (Ma 2020) and the arrests of hundreds of protesters (Kuo and Wintour 2020), and pro-independence activists (Ho 2020). Thenceforth, it has been widely observed that various political freedoms in Hong Kong have diminished under the law (Angeli 2021). This study is motivated by the possibility that the impact of the NSL on freedom of speech could be reflected in local opinion polls. Specifically, this paper explores whether the NSL leads to differential increases in political sensitivity bias in opinion polls that are more related to the clauses of the NSL.

This paper uses Kobayashi and Chan's (2022) definition of political sensitivity bias as the difference between observed poll results and the latent genuine public opinion "arising when respondents selectively refuse to respond or falsify their answers to sensitive questions for fear of punishment by the state apparatus, such as police and national security agents". This concept is especially pertinent to political science because biased public opinion results misguide political scientists and even politicians. The most notable example raised by Kuran (1997) was the sudden fall of Communism in Eastern Europe to the surprise of scholars, intelligence organisations and even the East European leaders themselves, who overestimated support for the communist regimes.

The potential samples for research on political sensitivity bias are nevertheless restricted. Democratic states where political scientists enjoy copious opinion poll data and feel safe conducting experiments, by definition, do not punish their citizens for their opinions (Whitehead 2002, pp.10-11). However, this does not preclude them from studies on social desirability bias, which is present across regime types (Kalinin 2016). In contrast, open data may exist but can be manipulated in less democratic states (Carlitz and McLellan 2021). Moreover, conducting list experiments in a robust dictatorship can encounter ethical concerns. Researchers may rely on existing international or national survey data, but the questions and designs of the survey are out of researchers' control –this will be discussed in detail with examples in the literature review section. More importantly, as Kobayashi and Chan (2022) argue, falsified preference can be indistinguishable from genuine regime support in robust autocracies.

Under these constraints, the case of Hong Kong is valuable for this field of study for three reasons. First, Hong Kong is one of the very few regions which have undergone a rapid transition in the data age with respect to increasing government control over speech and expression (Kobayashi and P. Chan 2022). Before the National Security Law, people in Hong Kong generally enjoyed the freedom to express their personal views on politically sensitive topics. This is evident as Hong Kong scored 4 out of 4 in this field in Freedom House's 2019 report[4], which dropped to 3 in the 2020 report and then to 2 in the 2021 report[5]. The rapid transition helps to isolate the effect on political sensitivity bias from social desirability bias and genuine public opinion. The proximity of this transition allows for the application of newer data science techniques in research. Secondly, open opinion poll data have been copious before and under the NSL. For example, the Hong Kong Public Opinion Research Institute (HKPORI 2023) has been conducting randomised opinion surveys on various topics ranging from the popularity of the government to the opinion on the independence of Taiwan, since 1993 [6]. This provides grounds for observational studies such as this paper. Thirdly, academic freedom in Hong Kong could be considered unimpeded at the time of writing. Kobayashi, a scholar based in Hong Kong, has completed two politically sensitive studies on the impacts of the NSL involving original surveys and experiments and managed to publish results against Beijing. Both articles will be examined further in the literature review section . This implies that future field studies in Hong Kong are still possible.

This paper contributes to the debate by providing an alternative way to reveal political sensitivity bias through open public poll data and showing that an event, such as the enactment of the NSL, can have a heterogeneous impact on the political sensitivity bias in different opinion polls. The rest of this paper is divided as follows: an introduction to the background, a review of the literature and theory, an empirical strategy section that introduces the data and methodology, the analysis of results, the discussion of the limitations of this study, and the conclusion.

[1] A more detailed introduction to the background is in Section 2.

[2] More details in Section 2.2

[3] A movement-based pro-democracy youth political party in Hong Kong that "aims to achieve democratic self-determination in Hong Kong", headed by Agnes Chow Ting, Nathan Law Kwun-chung and Joshua Wong Chi-fung (Demosistō 2023).

[4] The 2019 report summarises the situation in 2018.

[5] First report under the NSL.

[6] More discussion about the data source is in Section 4.1.

2 Background

2.1 Prelude –The Anti-Extradition Law Amendment Bill Movement

In 2019, widespread protests broke out in Hong Kong over the introduction of the Fugitive Offenders Amendment Bill on Extradition by the Hong Kong government (BBC 2019b). This is known as the Anti-Extradition Law Amendment Bill Movement or the 2019-2020 Hong Kong Protests. Opponents of the bill feared it would expose people in Hong Kong to unfair trials and give Beijing greater influence over Hong Kong as it could be abused to target Anti-Beijing activists and journalists. The protesters demanded that the event cease to be characterised as a riot, amnesty for arrested protesters, an independent inquiry into the alleged police brutality, complete universal suffrage, and the withdrawal of the bill (BBC 2019b) – these are later known as the "Five Demands". In response to this outcry, the Carrie Lam administration withdrew the extradition bill on the 4th of September 2019 but declined the other four demands (Kuo 2019)[7]. Consequently, the protests escalated and culminated in the sieges of the universities (Yu et al. 2019; HKFP 2019), and the landslide victory of the pro-democracy camp in the Hong Kong District Council elections in November 2019 (Lam, Sum, and Ng 2019). While protests in Hong Kong were largely silenced by the breakout of the COVID-19 pandemic in early 2020[8], Beijing was on the move to regain control over the defiant city by drafting a new national security law[9].

2.2 The Hong Kong National Security Law (HKNSL)

As introduced in Section 2.1, the Standing Committee of the National People's Congress of China (NPCSC) drafted and eventually passed the Hong Kong National Security Law[10] on 30 June 2020. The new law criminalises any act of *secession, subversion, terrorism and collusion* with foreign or external forces; the law also empowers Beijing with law interpretation and trials.

Although the turmoil related to the 2019-2020 protests facilitated the formulation of a National Security Law for Hong Kong by Beijing, the legal roots of this legislation date back to before the handover of the city from the United Kingdom to China in 1997: Article 23 of the Hong Kong Basic Law[11], created before the handover, stipulates that Hong Kong

> "shall enact laws on its own to prohibit any act of treason, secession, sedition, subversion against the Central People's Government, or theft of state secrets, to prohibit foreign political organisations or bodies from conducting political activities in the Region, and to prohibit political organisations or bodies of the Region from establishing ties with foreign political organizations or bodies." (Basic Law 2021, p.77)

Despite Article 23, Hong Kong had never successfully passed a national security law before 2020[12]. In 2002-2003, there was an attempt to pass a national security bill by the HKSAR government and Legislature, which similarly sparked a mass demonstration on 1 July joined by about half-million people[13] (Wong 2004, p.67). As Beijing understood the impossibility of passing a national security law by the HKSAR itself due to unpopularity and the need for such a law regarding the 2019-2020 protests, Beijing decided to invoke Article 18 of the Hong Kong Basic Law, which empowers the NPCSC to add national laws to (and delete laws from) Annex III of the Hong Kong Basic Law, which shall be applied in HKSAR[14] (Basic Law 2021, p.63).

3 Theory

3.1 Literature Review

The term *political sensitivity bias* was first clearly defined by Kobayashi and Chan (2022) as the discrepancy between observed poll results and the latent genuine public opinion due to preference falsification and non-

[7]The administration later exacerbated the tension by introducing the Prohibition on Face Covering Regulation (Cap. 241K), which annoyed the masked protesters (BBC 2019a).

[8]This is evident as the Hong Kong people did not protest when 15 activists were arrested by the Hong Kong police in April in fear of infection with the deadly virus (Davidson 2020).

[9]China's annually assembled legislature, the National People's Congress, authorised its standing committee to formulate a new national security law for Hong Kong on 29 May (Kuo 2020)

[10]Full name: Law of the People's Republic of China on Safeguarding National Security in the Hong Kong Special Administrative Region.

[11]The organic law for the Hong Kong Special Administrative Region, akin to a constitution for a sovereign state.

[12]Under the British rule before 1997, there were national security provisions in Hong Kong's Official Secrets Ordinance, Crimes Ordinance and Societies Ordinance. However, this concerned the security of the British rule rather than the Chinese rule.

[13]Interestingly, Wong (2004) suspected that the HKSAR government similarly took advantage of the SARS pandemic as an opportunity for the national security bill

[14]Consultation with the NPCSC Committee for the Basic Law of the Hong Kong Special Administrative Region and the government of the Region. Article 18 even authorises the Central People's Government (State Council) to issue an order applying the relevant national laws in the region in the events of war and *turmoil* in the region beyond the control of the HKSAR government, which endangers national unity or security.

response. Similar concepts have been proposed by earlier scholars, for example, by Blair et al. (2020) as "sensitivity bias in authoritarian regimes" , and by Tang (2016) as "political desirability bias" . However, only Kobayashi and Chan (2022) distinguish political sensitivity bias from social desirability bias (where the pressure is not from the government), and from autocracy-driven preference falsification, which does not consider non-response in surveys.

Preference falsification is a concept to which political sensitivity bias is closely related. Tim Kuran (1987) introduces this concept in his open-voting public choice model as the act of misrepresenting one's preference under perceived public pressures, as the benefits outweigh the costs. However, this concept does not take into account the bias that arises from non-response in surveys, which constitutes *avoidance* but not falsification. In this research, we do not distinguish between these two sources of bias and are instead interested in the combined effect of them.

Nevertheless, the political sensitivity bias that this paper explores should not be confused with the *social desirability bias*, which arises when the pressure to falsify preference or selectively avoid surveys comes from other members of the society rather than the government (Krumpal 2013). The latter is not the focus of this study. Despite the distinction, the social desirability bias and political sensitivity bias can theoretically co-exist if people feel genuine support for the government around them (Buckley et al. 2022). The impacts of the two are difficult to separate in empirical research.

Empirical studies that attempt to quantitatively measure political sensitivity bias or preference falsification in authoritarian contexts are rare due to the difficulties in measurement. Nonetheless, researchers have developed creative techniques to tackle this problem.

Numerous researchers have used the list experiment technique to directly measure levels of social desirability and political sensitivity bias in public opinion polls across different regimes (Chapkovski and Schaub 2022; Frye et al. 2017; Kalinin 2016; Koehler, Grewal, and Albrecht 2022; Robinson and Tannenberg 2019; Tang 2016). This is done by asking the control and treatment groups two different lists of questions, respectively, with the treatment survey having one additional item –the sensitive one. Unlike a regular survey, respondents are only asked to report how many items in the list pertain to them, rather than responses to each item. The researcher then estimates the genuine public opinion by comparing the results from two groups.

The strength of this approach is its simplicity and clarity –the difference in means simply reveals the genuine public opinion. The downsides are that a large sample size is required, the results are subject to the ceiling and floor effects, and the sample is not always representative due to the nature of web-based surveys (Blair and Imai 2012). In most of the studies mentioned above, young and better-educated people are over-represented. In addition, there are ethical concerns about list experiments –those who give anti-government responses in the direct-question group might be subject to state retaliation. This is even more problematic if the researchers outsource the survey to institutions *within* the autocratic country, which are more prone to state pressure to expose the respondents. In contrast, this study relies on publicly available secondary data and does not increase the risks of survey respondents.

Public surveys, when conducted in rare coincidences with political events, could be used in quantitative research on preference falsification and political sensitivity bias. Jiang and Yang (2016) used a nationwide survey in 2006 which coincided with the political purge of Chen Liangyu, the Chinese Communist Party's secretary of Shanghai. They employed a difference-in-difference (DID) approach to estimate the effects of the purge on Shanghai respondents' expressed and actual support for the government[15]. They found evidence for increased preference falsification among Shanghai residents. However, this method relies on rare opportunities of coincidences of public surveys and political events, which are difficult to imitate for other researchers. Interestingly, Jiang and Yang (2016) also tested the differential sensitivities of the explicit measures using a survey experiment technique developed by Bradburn (1979), whereby the researchers asked two groups of respondents their perceived sensitivities to a list of questions while asking only the treatment group to "give answers that *most people would give*", and the control group to give *their real answer* (Jiang and Yang 2016). The differences between the answers could indicate the differential sensitivities of the survey question.

Two studies conducted by Kobayashi nonetheless use different empirical strategies and focus on the context of Hong Kong National Security Law. Kobayashi et al. (2021) fielded comparable conjoint experiments and found the support for the "Five Demands" [16] was largely stable despite the HKNSL. In Kobayashi et al. (2021), the realised demand for the retraction of the extradition bill is replaced by the demand for the resignation of the Chief Executive –Carrie Lam, which was another popular demand at the time. The researchers conducted two conjoint experiments[17] before and after the HKNSL came into effect. They found no significant changes in the pattern of aggregate preferences between the two experiments despite the repression of the NSL. A more recent study by Kobayashi and Chan (2022) attempts to measure precisely the political sensitivity bias in Hong Kong under NSL with online panel survey data collected before and after the enactment of the NSL. They reveal that pro-democracy citizens, who are

[15] Levels of actual support are proxied with items on happiness, judicial independence, and beneficiary group.

[16] The five demands have been explained in Section 2.1.

[17] Conjoint experiments identify people's preferences by breaking down a "bundle" into components and testing the popularities of different combinations of components. In this research, the common components in the two experiments are the Five Demands, fiscal policy, rates for residential properties, with the National Security Law component listed only in the second experiment.

potentially subject to the NSL, are less likely to respond to political polls, more likely to falsify sensitive past behaviour and engage in preference falsification under the NSL than when the NSL was not in force.

Overall, the previous literature reveals that the HKNSL has induced political sensitivity bias and preference falsification in the polls (Kobayashi and P. Chan 2022). Nevertheless, the differential treatment effects of events on political sensitivity bias across political topics remain a vacuum in the literature. The differential treatment effect approach employed by this paper also bypasses the experimental strategies (Chapkovski and Schaub 2022; Robinson and Tannenberg 2019; Kobayashi, Song, and P. Chan 2021; Bradburn 1979), which are common in the fields of preference falsification and political sensitivity bias but can be practically and ethically intractable.

3.2 Hypotheses

This paper explores whether polls that are more closely related to the NSL receive a differential impact from the NSL on political sensitivity bias compared to those less related to the NSL. Although the differential change is likely to be positive (i.e. those more linked to the NSL see larger increases in their bias), we will use two-tail tests for the most conservative results. Therefore, we are testing the following pair of hypotheses:

Hypothesis 0 (H_0): *The impact of the NSL on the political sensitivity bias does not differ by public opinion polls in Hong Kong ($\Delta\tau = 0$).*

Hypothesis 1 (H_1): *The impact of the NSL on the political sensitivity bias differs by public opinion polls in Hong Kong ($\Delta\tau \neq 0$).*

Here $\Delta\tau$ denotes the differential treatment effect, which is the difference between the treatment effect on the treated group and that on the control group[18]:

$$\Delta\tau = \tau_{treated} - \tau_{control} \tag{1}$$

4 Empirical Strategy

4.1 Choice of data

4.1.1 Population of Interest

The population of interest here is Cantonese-speaking Hong Kong residents aged 18 or above. This largely isolates the public opinion of local residents from that of tourists, students, and those temporarily working in Hong Kong who come from mainland China or the rest of the world (whose views tend to differ from that of the local population). This aligns with the target population of my data source (HKPORI 2021).

4.1.2 Data Source

All survey data used in this paper are sourced from the Hong Kong Public Opinion Research Institute (HKPORI). The HKPORI succeeded the Hong Kong University Public Opinion Programme (HKUPOP), which existed for 28 years between 1991 and 2019 (HKPORI 2023).

The Hong Kong Public Opinion Research Institute has been conducting up-to-date randomised opinion surveys on various politically sensitive topics, including the attitudes toward Taiwan and Tibet independence since 1993 (HKPORI 2023). The survey results have been rim-weighted according to the gender, age, educational attainment, and economic activity status distributions of the Hong Kong population (HKPORI 2021).

For a typical HKPORI poll, as in Figure 1, there are time series data for the survey start dates, survey end dates, successful cases, subsamples, response rates, positive responses, negative responses, and the net value[19].

For some polls, there is more than one positive/negative response and includes a response option of don't know/hard to say (DK/HS). Raw datasets for some polls are also available.

A full list of HKPORI polls and related information are available here.

4.1.3 Choice of Polls

As enumerated in Supplementary Material, there are around 160 polls from the HKPORI. However, some polls are excluded for their discontinuities, being inaccessible (Public Sentiment Index), being too recent, being collected at long intervals (e.g. annually), no recent updates[20], being covered by other polls, or

[18] more explanation on the differential treatment effect is explained in the SDID explanation and implications sections.

[19] Net value calculated as the proportion of positive responses less that of the negative responses.

[20] Polls are excluded if the last collection was before October 2020, more is explained in the Time span section.

你贊唔贊成台灣獨立？（按次計算）
Do you agree to Taiwan becoming independent? (Per Poll)

顯示/隱藏 Show/Hide | 複製 Copy | 下載CSV Download CSV | 下載Excel Download Excel | Search:

調查開始日期 Survey Start Date	調查結束日期 Survey End Date	成功樣本 Successful cases	次樣本人數 Subsample	回應率 Response rate	贊成 Agree	唔贊成 Disagree	唔知/難講 DK/HS	總數 Total	淨值 Net value	備注 Remark
All	All	All	All	All	All	All	All	All	All	All
2022-10-10	2022-10-17	505	505	62.3%	20.9%	62.2%	17.0%	100.0%	-41.3%	

Figure 1: Poll on Taiwan independence in Oct 2022 (Source: HKPORI, 2023)

being irrelevant. Irrelevance refers to when a "pro-Beijing" response cannot be identified in the poll (e.g. corporate responsibility) or when the polls are not for Hong Kong (e.g. Macau Surveys).

96 polls passed this initial screening. Their raw data have been downloaded, cleaned and the information about these polls is listed in Supplementary Material.

4.1.4 Time Span

The theoretical time span for the main analysis is between 1 January 2010 and 1 May 2023 (the time of data access). In practice, the effective end time of the main analysis is earlier than May of 2023. This is because the last poll of interest can be earlier than May of 2023, sometimes as early as October of 2022– in that case, the time span of the panel data would be chosen based on the poll of interest. However, a poll is excluded from the first screening, as explained in Section 4.1.3, if the last collection of that poll was before October of 2023.

The starting time of the model, 2010, is based on the results of trend tests (in Section 5.1) against two pivotal events in Hong Kong –the 2007-2008 financial crisis and the following economic recession, and the 2014 Hong Kong protests.

The 2007-2008 Global Financial Crisis (GFC) dealt a heavy blow to Hong Kong. Lee et al. (2010) found a significant increase in the risk of mental depression in the Hong Kong population associated with the economic contraction triggered by the global financial crisis. Therefore, it is reasonable to suspect that the psychological impact of the crisis may affect people's behaviour at polls –specifically, this concerns the ability of the model to predict post-crisis poll behaviour with pre-2008 results. According to the pre-treatment trend test results in Section 5.1, there is no predictability. Hence, data before 2010 are removed to exclude the impacts of the GFC.

Similarly, the Hong Kong protests and electoral reform in 2014 could have also influenced people's behaviour at polls. The Hong Kong protests were sparked off by the proposed reforms to the Hong Kong electoral system, which would introduce political screening against the dissidents of Beijing (J. Chan 2014). Despite the ongoing protests, the NPCSC passed an even more conservative version of electoral reform than the system used in the 2012 election (J. Chan 2014). The protests ended in December 2014 when the last protesters were cleared by the police with little resistance (BBC 2014).

Hence, another series of pre-treatment trend tests are conducted in Section 5.1 to check if the model using poll data between 2010 and 2014 can plausibly predict poll results from 2015 onwards. The test results indicate that the model using pre-2015 data can accurately predict post-2015 trends. Therefore, pre-2015 data are included in the pre-treatment period.

In short, the time span for the main analysis is set to be as long as possible while the validity of the model can be plausibly assumed. The starting time of 2010 is the earliest time from when the model can consistently predict future poll trends, and the end dates are constrained by the latest polls available.

4.2 Classification of Polls by Political Sensitivity and the Choice of "Treated Units"

The 96 polls that have passed the screening are categorised by levels of political sensitivity, which can be "extremely high", "high", "medium", and "low".

Polls of "extremely high" political sensitivity are defined as polls to which giving a negative response in public constitutes an *ad verbum* violation of the NSL in Hong Kong. These are the polls of interest or the "treated units' for the main analysis. According to this criterion, polls on the independence of Taiwan/Tibet and those on the June Fourth (Tiananmen) incident are categorised as "extremely high" political sensitivity. The former two violate Part 1 of the NSL concerning secession; the latter violates Part 2 of the NSL concerning subversion, as Beijing regards the 1989 protests as attempts to subvert the rule of the Communist Party (Nathan 2001, p.330; People's Daily 1986).

Polls of "high" political sensitivity are defined as polls to which giving a negative response in public in Hong Kong does not violate the NSL *ad verbum* but could nevertheless result in an arrest and, sometimes, a conviction, as the rule of law and judicial autonomy are now in question. One possibility is the re-invocation of the obsolete anti-sedition provisions in the colonial-era Crimes Ordinance, which were not used after 1967 until the NSL, with references to the British authorities construed as references to the HKSAR government (Chau 2022). For example, the poll for the popularity of Xi Jinping is classified as "high political sensitivity" since a negative response in public can be interpreted as exciting disaffection against the central government.

Polls of medium political sensitivity are defined as polls to which giving a negative response in public is under political pressure but at low risks of persecution. An example of this is doubting the impartiality of Hong Kong courts. Polls of low political sensitivity are polls to which giving a negative response in public is generally considered politically acceptable, such as expressing dissatisfaction with the financial budget.

4.3 Treatment and (No) Covariates

The treatment in the main analysis is the differential impact of the NSL. Unlike other panel data research with binary treatment status where the control group are assumed to be under no treatment effect, this paper assumes all polls are affected by the NSL (Section 4.5.5), but to different degrees. This thesis explores the degree to which the polls of "extremely high" political sensitivity are affected by the NSL more than those of medium and low political sensitivity.

However, although the NSL has a clear starting time of legal effect ——30 June 2020, its political and behavioural impacts could have started as early as 29 May, when China's legislature authorised its standing committee to formulate a new national security law for Hong Kong. However, in this study, one cannot distinguish between the two as no polls by HKPORI were conducted between the two dates. For convenience, 30 June 2020 is chosen as the treatment date in the algorithm.

Moreover, no covariates (control variables) are included in the analysis. This is due to the special characteristics of the panel data used by this paper –unlike other panel data where units are distinguished geographically, our panel data distinguish units by poll, and all polls are applied to the same geographical space (Hong Kong) and population. This precludes this paper from finding exogeneous covariates that are both time-variant and unit-variant. Factors that change over time are implicitly included in the time-fixed effects, and their heterogeneous impacts on polls are absorbed by the algorithm when building synthetic counterfactuals. This is explored in Section 4.6 in detail.

4.4 Outcome Variable: Standardised Public Opinion Results Using the Probit Model

For the empirical tests, this paper uses the z-scores of the opinion poll results, rather than the poll results *per se*, as a standardised measure for the outcome variable (Y_{it}). This will be referred to as the *standardised public opinion results* hereafter. The formula is as follows:

$$Y_{it} = probit(\frac{A_{it}}{A_{it} + B_{it}}) = \Phi^{-1}(\frac{A_{it}}{A_{it} + B_{it}}) \quad (2)$$

Where A_{it} denotes the share of respondents who give a pro-Beijing response in the poll i in time period t. B_{it} denotes the share of respondents who give a non-pro-Beijing response in the poll i in time period t. Therefore, $\frac{A_{it}}{A_{it}+B_{it}}$ can be understood as the net approval rate of Beijing that excludes non-participants and other responses such as "don't know" and "hard to say".

The following poll in Table 1 is used as an example for illustration. Since the independence of Taiwan is opposed by Beijing, "Agree" responses would therefore be regarded as anti-Beijing, and "Disagree" responses would be regarded as pro-Beijing.

Response	Agree	Disagree	DK/HS
Percentage (%)	20.9	62.2	17.0

Table 1: Selected information for the poll on Taiwan independence in Oct 2022 (Source: HKPORI, 2023)

Hence, $A_{it} = 62.2\%$, $B_{it} = 20.9\%$, $\frac{A_{it}}{A_{it}+B_{it}} = \frac{62.2\%}{62.2\%+20.9\%} = 74.85\%$ (2 d.p.), and $Y_{it} = probit(74.85\%) = 0.67$.

The justification for using the z-score or the probit of the pro-Beijing vote share, rather than the vote share per se, is that this corrects for the distortion of results when the vote share is extreme (i.e. close to 0% or 100%). This is because the survey results used in this research are inherently binary rather than ordinal[21] –people cannot express the magnitudes of their satisfaction/discontent. Those who become more content but already gave a positive response/less content but already gave a negative response are not affecting the poll results. Consequently, the vote shares are more difficult to change at extreme values. Using the vote shares per se, therefore, underestimates the change in public opinion at extreme values. Using the probit/z-score results alleviate this problem. Further explanation is in Appendix A.

[21] There are few exceptions, but they are treated as binary anyway.

4.5 Assumptions

4.5.1 Independent Unobserved Transitory Shocks Across Units and Time

The first assumption is that the unobserved transitory shocks are "independent across units and in time" (Ferman and Pinto 2017, p.4). This means that no unit substantially undermines the predictability of the model, and no events other than the treatment dramatically impact the trend.

This assumption can be understood twofold –in the pre-treatment and the post-treatment periods. This independence in the pre-treatment period is relaxed under the SDID, our main estimator. This is because the SDID estimator assigns more even weights to control units compared to the SC estimator and is therefore more robust against unobserved transitory shocks. This is explained in Section 4.6. This first fold of the assumption is tested in Section 5.1. However, for the post-treatment period, we can only assume this is the case.

4.5.2 Existence of Weights

The second assumption is the existence of weights - there is a "stable linear combination of the control units that absorbs all time-correlated shocks" of the treated units (Ferman and Pinto 2017, 2017, p.4). This means that we can produce a counterfactual of interest based on a linear algorithm and past data of the sample units. Again, this assumption is more relaxed under SDID than under the SC estimator. This is because the robustness of the SDID estimator is enhanced by the time weights. More is explained in Section 4.6. This assumption is also tested in Section 5.1.

4.5.3 Non-requirement of Equidistance

The third assumption is that the SDID estimator does not require that the sample periods are equidistant in time. This is of concern as some polls were not conducted at regular intervals (e.g. 3 to 6 months). The SC estimator, the foundation of the SDID estimator, does not require that the sample periods are equidistant in time (Abadie, Diamond, and Hainmueller 2015). We assume this is also true for the SDID estimator although it is not mentioned by Arkhangelsky et al. (2021).

4.5.4 Unfalsified HKPORI Results

The fourth assumption is that the poll results by HKPORI are not falsified, and there is no under-reporting.

One might be concerned about political pressure implied by the irregular intervals. There is evidence that the HKPORI once postponed the publication of poll results in 2022 upon Xi's visit (P. Lee 2022). However, we argue this implies the authenticity of results –there is no incentive to postpone fabricated "happy" results. Also, if they were under political pressure, why not postpone the *survey* instead of the publication?

4.5.5 Treatment Effect Across All Polls

The fifth assumption is that all polls have been affected by the NSL in terms of increased political sensitivity bias ($\tau_{control} > 0$), for it allows for the most conservative interpretation. This is also in line with Kobayashi and Chan's (2022) findings, although they asked different questions in their surveys.

4.5.6 The NSL does not affect the latent genuine public opinion

The sixth assumption is that the latent genuine public opinion is not affected by the NSL, so any impacts of the law on the observed public opinion can be considered impacts on the political sensitivity bias.

4.6 The Synthetic Difference-in-Differences Model

The Synthetic Difference-in-Differences (SDID) approach was developed by Arkhangelsky et al. (2021), which combines the desirable features of the difference-in-differences (DID) and synthetic control (SC) methods for estimating causal effects with panel data. This method is claimed by Arkhangelsky et al. (2021) to be applicable and perform at least as well, if not better, in all settings where the DID and SC estimators are commonly used and can be applied in some scenarios where the conditions of the two conventional estimators are not met[22]. Moreover, the algorithm provided by Arkhangelsky et al. (2021) allows researchers to run three estimators simultaneously for comparison. Therefore, results using the DID and SC estimators are also reported as a robustness check to complement the SDID results.

The synthetic difference-in-differences requires a strongly balanced panel with N units (in this case, opinion polls) and T time periods (poll dates). Time periods T are assigned based on the time of the

[22] SDID relaxes the parallel trends assumptions on which traditional diff-in-diff methods are based and can be applied to large numbers of treatment units, or when the outcomes of the treated units are extreme to the donor pool, which are not applicable to the traditional SC method. SDID can also be applied with staggered treatment timing (Porreca 2022). One caveat from Arkhangelsky et al. (2021) is that systematic heterogeneity in outcomes by either units or time periods may worsen the precision of SDID relative to the DID estimator.

"treated" poll, the periods of the control units are aligned with the treated unit if their dates are different. More explanation is in Appendix B. Y_{it} denotes the outcome (standardised public opinion results) for unit i in period t. $W_{it} \in 0,1$ denotes the exposure to the binary treatment. N_{co} denotes the number of control units in the donor pool; hence $N - N_{co}$ is the number of treated units, which are exposed to the treatment after time T_{pre}.

At this point, the comparison of the three estimators can be summarised as follows by their regression problems, where $\Delta\tau$ denotes the *differential* treatment effect in this paper[23]:

$$(\widehat{\Delta\tau}^{sdid},\ \widehat{\mu},\ \widehat{\alpha},\ \widehat{\beta}) = \arg\min_{\tau,\mu,\alpha,\beta}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\mu-\alpha_i-\beta_t-W_{it}\tau)^2\,\widehat{w}_i^{sdid}\widehat{\lambda}_t^{sdid}\right\}.^{24} \tag{3}$$

$$(\widehat{\Delta\tau}^{did},\ \widehat{\mu},\ \widehat{\alpha},\ \widehat{\beta}) = \arg\min_{\tau,\mu,\alpha,\beta}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\mu-\alpha_i-\beta_t-W_{it}\tau)^2\right\}. \tag{4}$$

$$(\widehat{\Delta\tau}^{sc},\ \widehat{\mu},\ \widehat{\beta}) = \arg\min_{\tau,\mu,\beta}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\mu-\beta_t-W_{it}\tau)^2\,\widehat{w}_i^{sc}\right\}. \tag{5}$$

Note that in SDID and SC, weights ($\hat{w}_i^{sdid}$ and $\hat{w}_i^{sc}$) are assigned to control units in the donor pool, which does not apply to DID. The SDID and SC methods alike find weights $\hat{w}_i^{sdid}$ and $\hat{w}_i^{sc}$ to "align pre-exposure trends of the outcome variable of the unexposed units with those of the exposed units" (Arkhangelsky, Athey, et al. 2021). However, SDID allows the actual trend of the treated units to be merely parallel to the synthetic trend (i.e. leaving a gap in between) while the SC seeks to minimise the distance between the two trends[25].

Both the SDID and DID estimators account for unit fixed effects α_i, which the SC estimator does not. The unit fixed effects account for the gaps in the parallel trends between units rather than leaving the gaps in the error terms, which the SC estimator implies.

The SDID estimator, however, is distinct for its time weights λ_t^{sdid}. The SDID estimator calculates time weights for pre-treatment periods to minimise the *difference in adjusted outcome differences among control units*[26].

The coexistence of the unit and time weights provides the SDID estimator with a "double robustness property" (Arkhangelsky, Athey, et al. 2021), whereby the estimate is resistant to bias associated with the systematic component of the error matrix (p.4096), as long as either of the balancing approaches is effective (p.4102). Moreover, including the two-way fixed effects and intercept terms makes the SDID estimator invariant to additive shocks to units and times, which is shared by the DID but not the SC estimator.

Specific to this paper, we could demonstrate the superiority of the SDID estimator with the following example in Figure 2 [27]:

Figure 2 shows that the parallel assumption, vital for the difference-in-differences estimator, is not perfectly met in the pre-treatment trends. The SDID estimator, which adjusts time and unit weights, produces a more parallel synthetic pre-treatment trend compared with the DID estimator.

Moreover, the control unit contribution plot (second row of Figure 2) shows that after adjusting for time weights, the SDID estimator has more similar trends among the control unit compared to the DID and SC estimators, as the dots are closer to the horizontal line. Also, the distribution of control unit weights is more even in the SDID estimator than in the SC estimator, as indicated by the dot sizes. In fact, the SC estimator uses only 9 out of 43 control units in the donor pool[30], whereas SDID uses 27. This means that SDID estimates is more resistant to stochastic shocks to those individual control units assigned non-zero weights than the SC estimator.

In short, the SDID estimator is more suited for this study than DID or SC estimators for its double robustness, better adjusted parallel trends, and resistance to stochastic shocks to individual control units.

---

[23]Here we replaced τ in Arkhangelsky et al. (2021), which denotes the average treatment effect, with $\Delta\tau$ since we assume the control units are also affected by the treatment, as explained in Sections 3.2 and 4.5.5.

[24]In all three regression problems, an intercept μ is present alongside fixed effects. To avoid multicollinearity, one must drop the intercept and a pair of fixed effects as "reference" . Alternatively, one could add conditions to the fixed effects: e.g. $\sum_i \alpha_i = \sum_t \beta_t$ =0 for DiD, as suggested by Arkhangelsky (2021) for the SDID, this would be $\sum_i \alpha_i \hat{w}_i^{sdid} = \sum_t \beta_t \hat{\lambda}_t^{sdid}$ =0.

[25]Technically, in the algorithm for time weights in the SDID estimator, Arkhangelsky et al. (2021) also add a regularisation parameter ζ to prevent overfitting. For technical details, see (4) and (5) in Arkhangelsky et al. (2021).

[26]For more information on the algorithm for time weights, see (6) in Arkhangelsky et al. (2021).

[27]Real poll data from HKPORI are used. The "treated unit" here is the rejection of Taiwan independence.

[29]The more centred the dots are around the line, the better: This indicates better parallel trends among control units for SDID and DID, and similar post-treatment trends for SC. Note that a point above the line indicates a *less positive* difference in trend compared with the treatment group. Further explanation is in Appendix C.

[30] "Killing the donor pool" is a common phenomenon in SC estimation as the SC forces the sum of weights to be 1 without allowing a parallel trend to exist. This means that control units which have parallelly higher/lower trends are marginalised by SC but included by SDID.

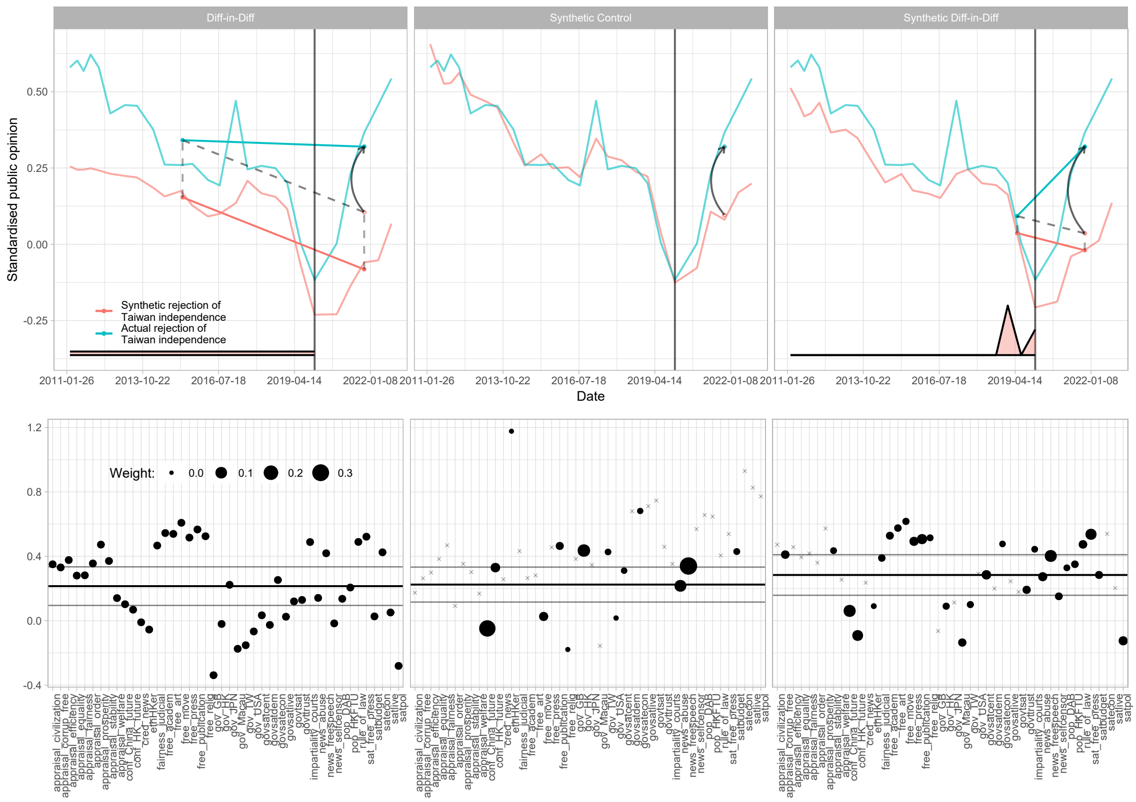

Figure 2: A comparison between difference-in-differences, synthetic control, and synthetic differences-in-differences estimates for the differential effect of the HKNSL on the standardised demo public opinion. The first row shows trends in standardised public opinion over time for the demo poll and the weighted average of relevant control polls; time weights are at the bottom of the graphs; arrows and parallelograms are drawn based on (time) weighted averages. The estimated differential effect is indicated by an arrow. The second row shows the poll-by-poll adjusted outcome difference $\hat{\delta}_{tr} - \hat{\delta}_i$[29], with the unit weights $\hat{\omega}_i$ indicated by dot size and the weighted average of these differences —the estimated differential effect —indicated by a black horizontal line. Control units with zero weight are denoted by an ×-symbol. CIs are shown by grey horizontal lines.

4.7 Choice of Standard Errors –the Driscoll Kraay (DK) Covariance Matrix

For the main analysis of this paper, the standard errors are calculated with the variance-covariance matrix proposed by Driscoll and Kraay (1998). This method is consistent in the presence of heteroscedastic, cross-sectionally and serially correlated errors with many time periods[31]. Serially correlated errors are common among panel data, to which our multi-survey data belong. The cross-sectionally dependence is crucial for the panel data used in this study since some polls are more interlinked, and their errors could be more cross-sectionally correlated to some than to others. For example, the support for Taiwan's independence is linked more closely to the support for Taiwan re-joining the UN than the appraisal of the public order. However, we also include the covariance matrix proposed by Newey and West (1987) as a robustness check to the results, but not the clustered standard errors. More justification for the use of the DK standard errors is in Appendix D.

[31] A panel data with $N = 33$, $T = 25$ could qualify for a "large T", according to Professor Jeff Wooldridge, an econometrics expert on panel data at UCSD, in a forum post. We have a larger T.

5 Results

5.1 Pre-Treatment Trend Tests

This section tests the ability of the SDID estimator to accurately predict pre-treatment trends for three treated groups of interest[32], as identified in Section 4.2, against the two pivotal events identified in Section 4.1.4, and assumptions of independent unobserved transitory shocks and existence of weights in Sections 4.5.1 and 4.5.2.

First, we tested the predictability of the models against the 2007-2008 Global Financial crisis (GFC). The results of these trend tests are summarised in Figure 3. The placebo "treatment" date is set to be 15 Sep 2008, the Bankruptcy of the Lehman Brothers ——the symbolic event for the GFC. The starting time is set to 2000, to avoid the impacts of the Asian Financial Crisis in the late 1990s, and to maximise the number of control polls (many polls were not available before the handover of Hong Kong in 1997). The end time is set to 30 June 2020, the treatment date of the main analysis.

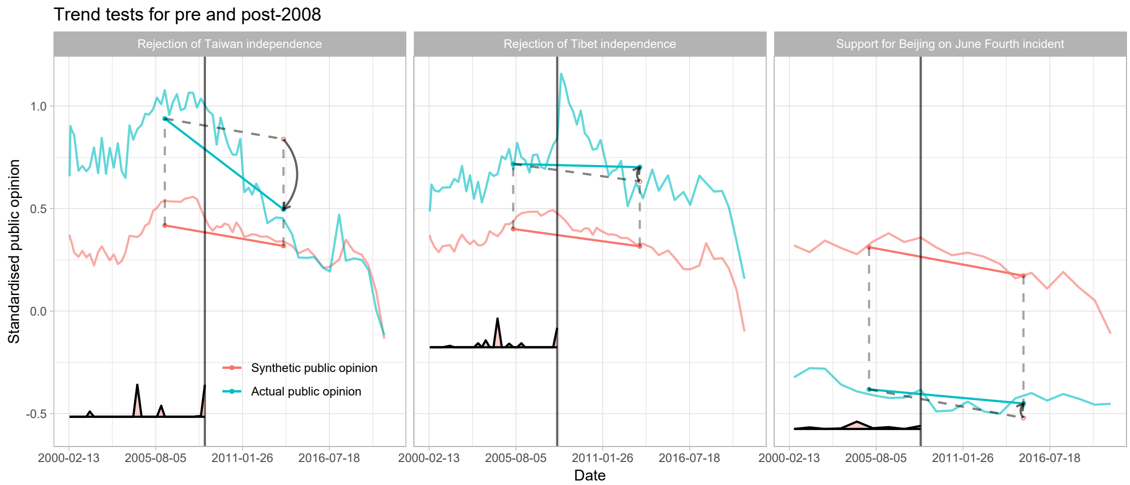

Figure 3: Trends of actual and synthetic public opinion based on pre-2008 information for polls on Taiwan independence, Tibet independence and the June Fourth (Tiananmen) incident. Synthetic trends are calculated using the weighted average of relevant control polls; time weights are at the bottom of the graphs; arrows and parallelograms are drawn based on (time) weighted averages.

Apparently, none of the three sets of treated units passed the 2008 tests. None of the post-2008 trends is accurately predicted by the models trained on pre-2008 data. For the poll on Taiwan independence, there is a significant decrease[33] in the actual trend compared to the synthetic trend. For the poll on Tibet independence, there is a spike in the actual trend right after the placebo treatment date, which is not predicted by the synthetic trend. For the poll on the June Fourth (Tiananmen) incident, the fall in the actual public opinion during the 2019-20 Hong Kong protests is not predicted by the synthetic trend.

In short, the pre-treatment period for the main analysis cannot include dates before 2008 as the correlation between the polls are different after 2008 from before 2008.

We then tested the predictability of the models against the 2014 Hong Kong electoral reform protests, as summarised in Figure 4. The starting time for the 2015 placebo tests is set to 2010 to avoid the impacts of the GFC as previously explored. The SDID models can plausibly predict the post-2015 trends based on pre-2015 data for the polls on Taiwan independence and Tibet independence, but not for the poll on the June Fourth (Tiananmen) incident. The actual trends for the former are almost exactly parallel to the synthetic trends. This is evident as the parallelograms in the corresponding plots show almost 0 placebo differential treatment effects (the blue line segments are almost overlapped with the dashed line segments). This means that for these polls, one can more confidently assume the synthetic trends are plausible counterfactuals of the actual trends in the main analysis (i.e. the synthetic trends represent a world where the NSL has homogeneous impacts across polls). However, the model fails to predict the post-2015 trend based on pre-2015 data for the poll on the June Fourth (Tiananmen) incident. This failure means that we cannot include the June Fourth (Tiananmen) incident in our main analysis.

Alternatively, one could read the regression results in Appendix E for the same interpretation.

In summary, all three groups fail the trend tests for 2008, indicating that the correlation between polls has changed since 2008; the polls for the independence of Taiwan/Tibet have passed the trend tests for 2015, hence they will be analysed in the main analysis. The poll for the June Fourth (Tiananmen) incident fails both tests and is excluded from the main analysis.

[32]The first two "groups" are single treated units (`indep_Taiwan` and `indep_Tibet`), whereas the last group on the June Fourth (Tiananmen) incident consists of three treated polls (`64_student`, `64_CHN_gov` and `64_reverse`).

[33]Placebo differential treatment effect=-0.342, se=0.0605 (DK), significant at 0.001.

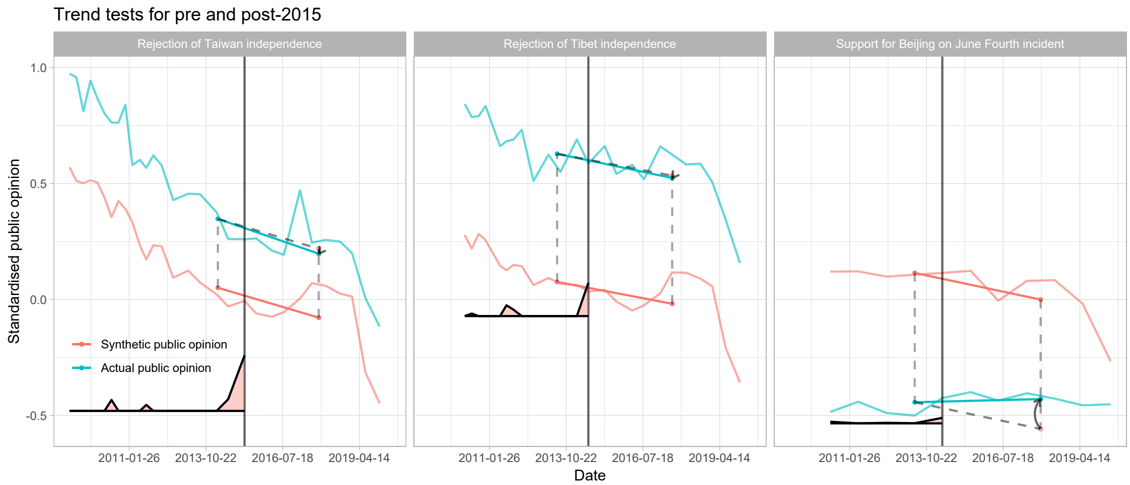

Figure 4: Trends of actual and synthetic public opinion trained on pre-2015 information for polls on Taiwan independence, Tibet independence and the June Fourth (Tiananmen) incident. Synthetic trends are calculated using the weighted average of relevant control polls; time weights are at the bottom of the graphs; arrows and parallelograms are drawn based on (time) weighted averages.

5.2 Main Result Analysis and Hypothesis Testing

Since only two treated groups have passed the pre-treatment trend tests in Section 5.1, and we use the polls of low and medium sensitivity as the control group, we divide and reformulate the alternative hypothesis H_1 into the following and test them individually:

Hypothesis 1a (H_{1a}): *The NSL has a differential impact on the political sensitivity bias in the poll on the independence of Taiwan, from the polls of low and medium sensitivity in Hong Kong* ($\Delta\tau_{taiwan} \neq 0$).

Hypothesis 1b (H_{2a}): *The NSL has a differential impact on the political sensitivity bias in the poll on the independence of Tibet, from the polls of low and medium sensitivity in Hong Kong* ($\Delta\tau_{tibet} \neq 0$).

As shown in the following subsections, both sub-hypotheses are supported by statistical evidence, and the differential impact is larger in the poll on Taiwan independence than in the poll on Tibet independence.

5.2.1 Treated Poll of Interest: Rejection of the Independence of Taiwan

We assessed the differential impact of the NSL on the poll for the rejection of Taiwan independence contemporaneously with three estimators. The results are visualised in Figure 5.

All three estimators indicate a statistically significant positive differential treatment effect (DTE) of the NSL on the rejection of Taiwan independence in polls, (see Table 2). Both the SC and SDID estimators show similar synthetic trends and sizes of the DTE, which implies robustness of our results.

	SDID	DID	SC
Differential Treatment Effect	0.284***	0.215***	0.224***
	(0.064)	(0.061)	(0.055)
Std.Errors	Driscoll-Kraay (L=1)	Driscoll-Kraay (L=2)	Driscoll-Kraay (L=2)
FE: Poll	Yes	Yes	No
FE: Time	Yes	Yes	Yes
Effective N	210	1170	390
R2	0.813	0.869	0.486
R2 Adj.(within)	0.240	0.006	0.055
RMSE	0.17	0.15	0.48

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 2: Poll: Rejection of Taiwan independence.

Table 2 suggests that all three estimated coefficients are statistically significant at the 0.01 level. The significance is robust against other standard errors (see Section 5.3.2). Since the pre-treatment tests have

[35]The more centred the dots are around the line, the better: This indicates better parallel trends among control units for SDID and DID, and similar post-treatment trends for SC. Note that a point above the line indicates a *less positive* difference in trend compared with the treatment group. Further explanation is in Appendix C.

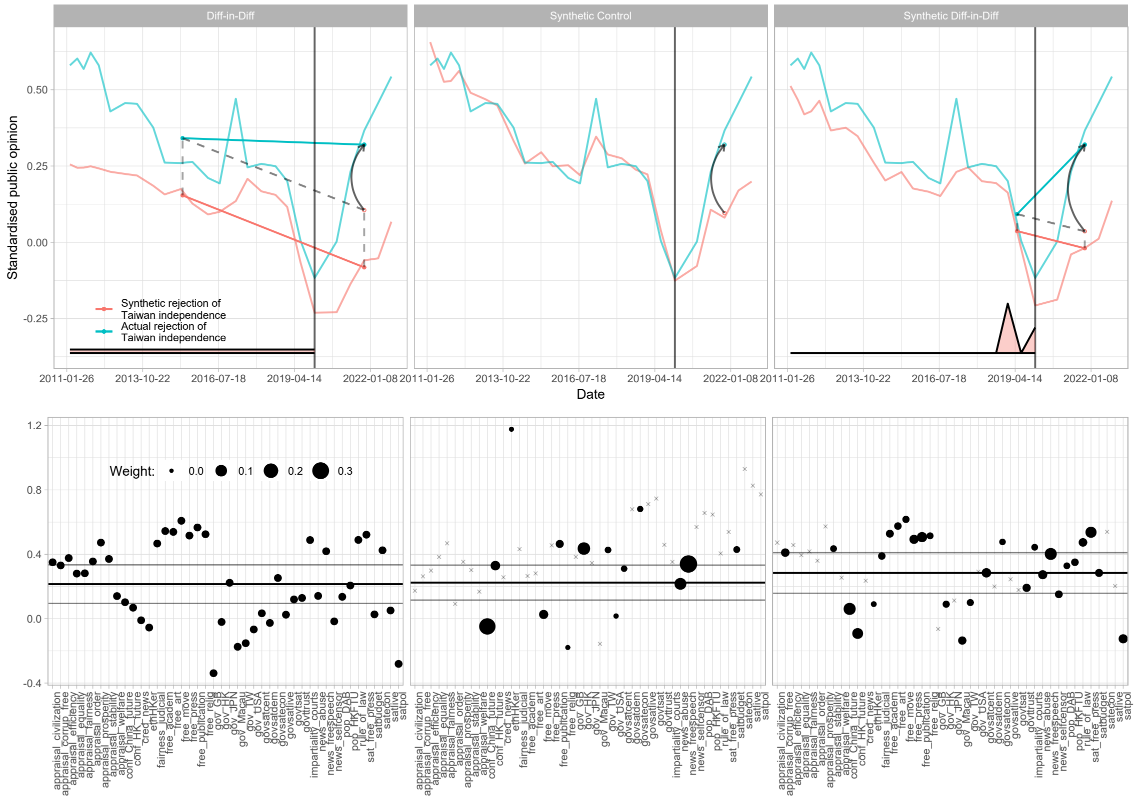

Figure 5: A comparison between difference-in-differences, synthetic control, and synthetic differences-in-differences estimates for the differential effect of the HKNSL on the standardised poll results for the rejection of Taiwan independence by the Hong Kong public. The first row shows trends in standardised public opinion over time on the rejection of Taiwan independence and the weighted average of relevant control polls; time weights are at the bottom of the graphs; arrows and parallelograms are drawn based on (time) weighted averages. The estimated differential effect is indicated by an arrow. The second row shows the poll-by-poll adjusted outcome difference $\hat{\delta}_{tr} - \hat{\delta}_i$[35], with the unit weights $\hat{\omega}_i$ indicated by dot size and the weighted average of these differences —the estimated differential effect —indicated by a black horizontal line. Control units with zero weight are denoted by an ×-symbol. CIs are shown by grey horizontal lines.

shown that the synthetic trend can accurately reflect the change in genuine public opinion, and we assume that the poll results are not falsified, the DTE can be attributed to a change in the political sensitivity bias.

This means that there is statistical evidence for:

Hypothesis 1a (H_{1a}): *The NSL has a differential impact on the political sensitivity bias in the poll on the independence of Taiwan, from the polls of low and medium sensitivity in Hong Kong ($\Delta\tau_{taiwan} \neq 0$).*

In terms of the magnitude, we take the SDID result, which indicates a DTE of 0.284. This means that the extra political sensitivity of the poll on Taiwan independence compared to the polls of medium and low political sensitivity has led to an extra increase in the political sensitivity bias of 0.284 in the standardised public opinion (z-score) due to the NSL. Given that the SDID estimation suggests a DTE-led increase in post-treatment average from 0.0361 to 0.3200, this means the proportion of pro-Beijing responses has increased from 51.4% to 62.5%, or an 11.1 percentage point unstandardised DTE.

One may wonder how the DTE varies over time. Therefore, we provide an event study style plot for the SDID results with 95% and 90% confidence intervals in Figure 6. Figure 6 implies that the DTE peaked at just below 0.4 around 22 months after the NSL took legal effect.

5.2.2 Treated Poll of Interest: Rejection of the Independence of Tibet

Similarly, the differential impact of the NSL on the poll for the rejection of Tibet independence is also assessed contemporaneously with three estimators, as shown in Figure 7.

Again, all three estimators indicate a positive DTE on the rejection of Tibet independence in polls, the statistical significance of which is examined in Table 3. Both the DID and SDID estimators show similar trends and sizes of the DTE, with the SC estimator showing a higher DTE.

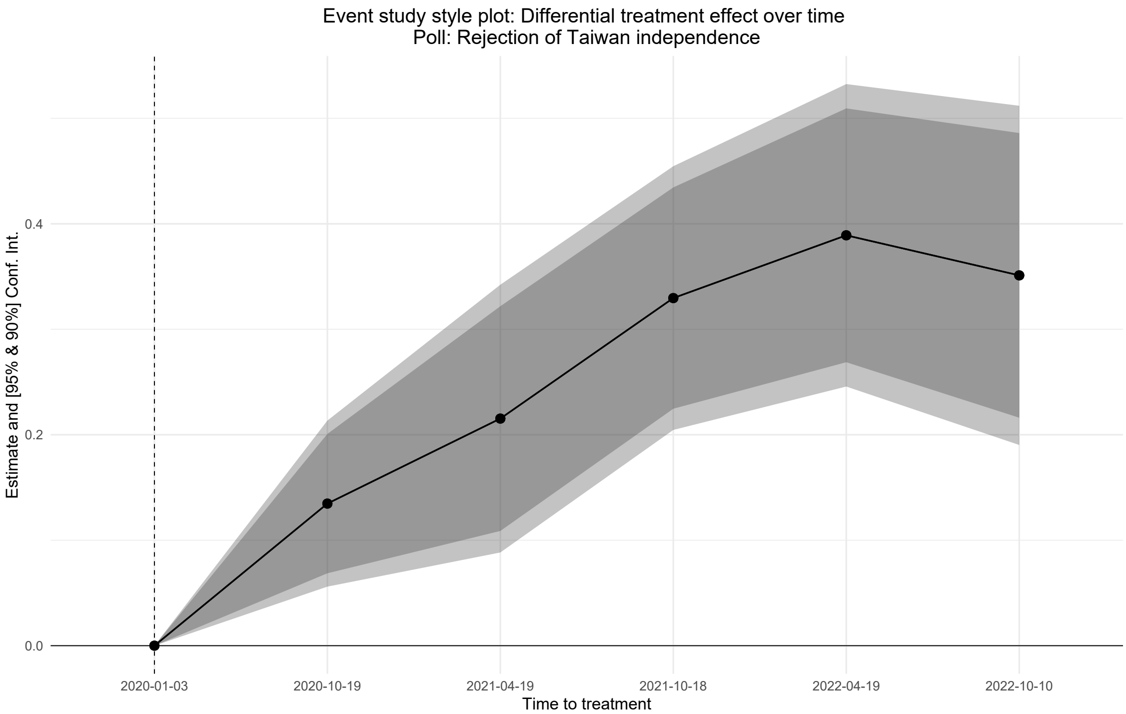

Figure 6: Event study style plot for standardised differential treatment effect over time for the poll on Taiwan independence.

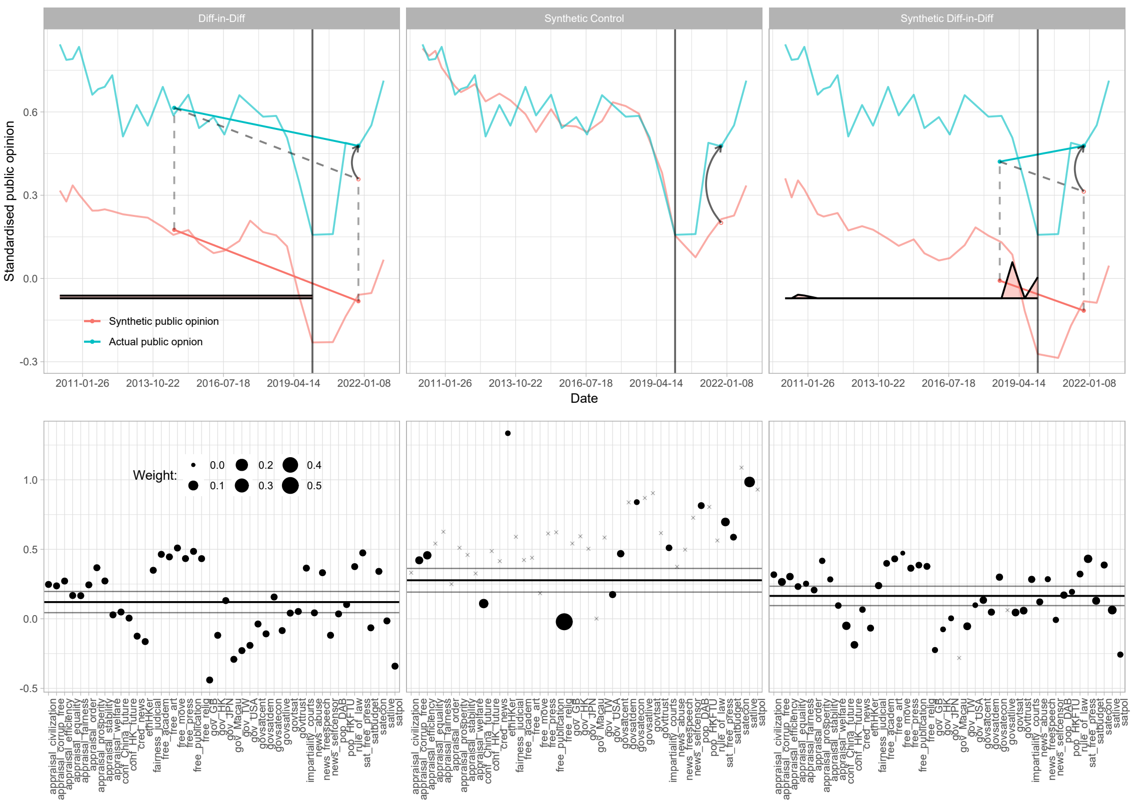

Figure 7: A comparison between difference-in-differences, synthetic control, and synthetic differences-in-differences estimates for the differential effect of the HKNSL on the standardised poll results for the public rejection of Tibet independence in Hong Kong.

Table 3 suggests that the coefficients from all three estimators are statistically significant at the 0.01 level. The significance is robust against other standard errors (see Section 5.3.2). As explained before, DTE can be attributed to a change in the political sensitivity bias.

This means that there is statistical evidence for:

	SDID	DID	SC
Differential Treatment Effect	0.165***	0.120***	0.277***
	(0.036)	(0.039)	(0.043)
Std.Errors	Driscoll-Kraay (L=1)	Driscoll-Kraay (L=2)	Driscoll-Kraay (L=2)
FE: Poll	Yes	Yes	No
FE: Time	Yes	Yes	Yes
Effective N	387	1305	406
R2	0.912	0.873	0.197
R2 Adj.(within)	0.109	0.001	0.021
RMSE	0.17	0.15	0.61

* p < 0.1, ** p < 0.05, *** p < 0.01

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 3: Poll: Rejection of Tibet independence.

Hypothesis 1b (H_{1b}): *The NSL has a differential impact on the political sensitivity bias in the poll on the independence of Tibet, from the polls of low and medium sensitivity in Hong Kong ($\Delta\tau_{tibet} \neq 0$).*

As for the magnitude, although the SC estimator shows a higher DTE, we conservatively take the result of the SDID, which is 0.165. This means that the extra political sensitivity of the poll on Tibet independence compared to the polls of medium and low political sensitivity has led to an extra increase in the political sensitivity bias of 0.165 in the standardised public opinion (z-score) due to the NSL. Given that the SDID estimation suggests a DTE-led increase in post-treatment average from 0.3131 to 0.4779, this means the proportion of pro-Beijing responses has increased from 62.3% to 68.4%, or a 6.1 percentage point DTE.

The event study style plot for the SDID results for the poll on Tibet independence with 95% and 90% confidence intervals in Figure 8. Figure 8 implies that the DTE was initially insignificant, then first peaked at 0.23 around 9 months after the NSL took legal effect, fell to 0.13 and then gradually went back up to 0.24 in October 2022, when the NSL had been in force for around two years. One should not be too puzzled about the fall between April 2021 and October 2021 as the 95% confidence interval suggests that the fall is not definitive.

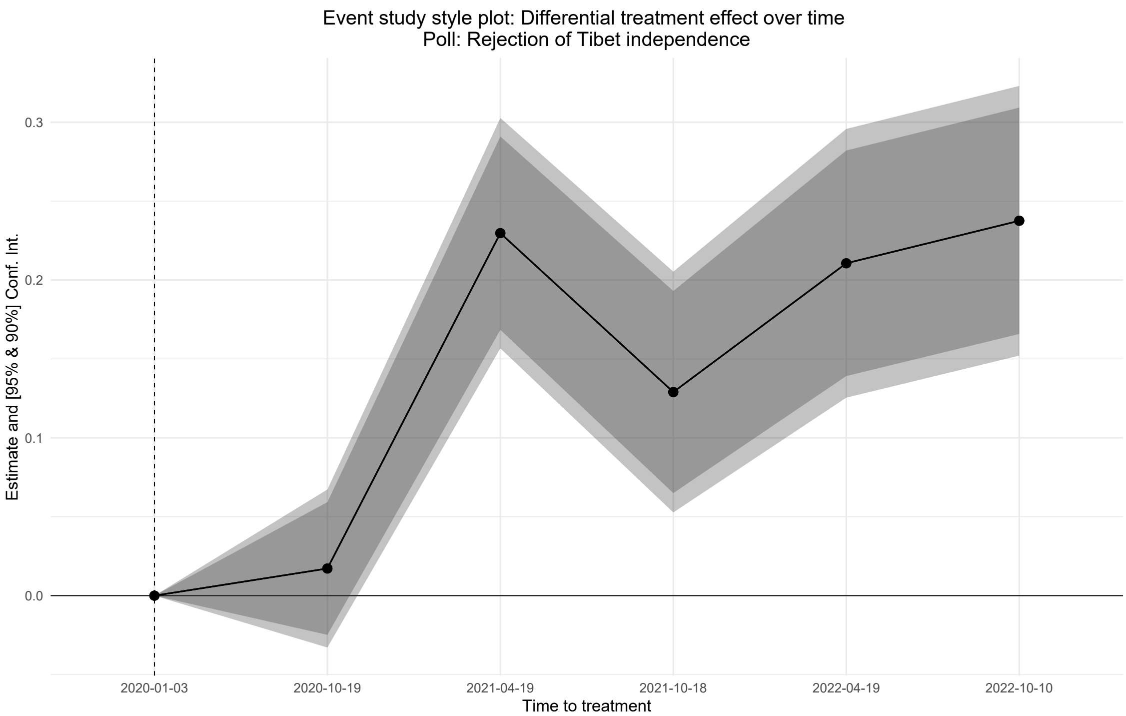

Figure 8: Event study style plot for standardised differential treatment effect over time for the poll on Tibet independence.

5.3 Robustness Checks

This study employs two sets of robustness checks to the main analysis.

5.3.1 Different Estimators

As introduced before, the main SDID estimator is complemented by the conventional DID and SC estimators. The results for these alternative estimators have already been reported in Section 5.2. Albeit varying in magnitude, the results are positive and statistically significant across estimators.

5.3.2 Different Standard Errors

Section 4.7 explains why the Driscoll-Kraay standard errors are the most suitable in this case, as it corrects for cross-sectionally dependent errors. However, we nevertheless use the Newey-West (NW) standard errors, as a robustness check. The latter also accounts for serially correlated and heteroskedastic errors, but not cross-sectionally correlated errors, which are likely present in our data. Alternative results using the NW standard errors are summarised in Tables 4 and 5.

	SDID	DID	SC
Differential Treatment Effect	0.284***	0.215***	0.224***
	(0.049)	(0.063)	(0.079)
Std.Errors	Newey-West (L=1)	Newey-West (L=2)	Newey-West (L=2)
FE: Poll	Yes	Yes	No
FE: Time	Yes	Yes	Yes
Effective N	210	1170	390
R2	0.813	0.869	0.486
R2 Adj.(within)	0.240	0.006	0.055
RMSE	0.17	0.15	0.48

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 4: Poll: Rejection of Taiwan independence.

	SDID	DID	SC
Differential Treatment Effect	0.165***	0.120**	0.277**
	(0.029)	(0.044)	(0.130)
Std.Errors	Newey-West (L=1)	Newey-West (L=2)	Newey-West (L=2)
FE: Poll	Yes	Yes	No
FE: Time	Yes	Yes	Yes
Effective N	387	1305	406
R2	0.912	0.873	0.197
R2 Adj.(within)	0.109	0.001	0.021
RMSE	0.17	0.15	0.61

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 5: Poll: Rejection of Tibet independence.

Table 4 indicates that the NW method yields a smaller standard error for SDID compared to the DK method, meaning that the main result for SDID is more conservative. However, the NW standard errors for DID and SC are larger than those using the DK method. Despite this, all results are significant at the 0.05 level, which suggests our results are very robust.

Similarly, Table 5 also indicates that the NW method yields a smaller standard error for SDID compared to the DK method, suggesting the main result for SDID is more conservative. However, the NW standard errors for DID and SC are also larger than those using the DK method. Nevertheless, all results are significant at the 0.05 level, indicating the conceivable robustness of our results.

5.4 Implications of the Results

The results of this study only demonstrate that the *changes* in the political sensitivity bias in the polls more related to the NSL (independence of Taiwan/Tibet) have been higher than that in polls that are less related to the NSL (polls of medium/low political sensitivity). That is to say,

$$\Delta\tau = \tau_{treated} - \tau_{control} > 0 \tag{1}$$

The differential impact is likely smaller than the treatment effect on the treated ($\Delta\tau < \tau_{treated}$) for the assumption that the control group has also received a positive treatment effect ($\tau_{control} > 0$)[36].

However, the differential treatment effect (DTE) confirms the *existence* of political sensitivity bias in Hong Kong under the NSL, at least for the treated polls in this paper. This is in line with the previous literature. One can infer this conclusion logically from our results:

Let η_i denotes the political sensitivity bias present in poll i, then

$$\begin{aligned}
\tau_i &= \eta^{post} - \eta^{pre} \\
\Delta\tau &= \tau_{treated} - \tau_{control} \\
&= (\eta^{post}_{treated} - \eta^{pre}_{treated}) - (\eta^{post}_{control} - \eta^{pre}_{control}) > 0
\end{aligned}$$

Since the political sensitivity bias cannot be negative ($\eta^{pre}_{treated} \geq 0$)[37] and the impact of the NSL on the control group cannot be negative ($\tau_{control} = \eta^{post}_{control} - \eta^{pre}_{control} \geq 0$), then the post-treatment levels of political sensitivity bias in the treated polls must be positive ($\eta^{post}_{treated} > 0$).

To make further inferences, one needs further information on the pre-existing levels of bias prior to the NSL in the treated polls ($\eta^{pre}_{treated}$) and the control polls ($\eta^{pre}_{control}$), as well as the post-treatment levels of bias in the control polls ($\eta^{post}_{control}$). This is beyond the scope of this paper.

6 Limitations

6.1 Bias from Time Alignment

One source of bias in our results may be the alignment of time periods when forming the panel data (Appendix B). This is because the results of the closest polls in time would not necessarily equal the hypothetical results if the polls were conducted at the same time as the treated poll. The differences between the two generate measurement errors, which is a source of endogeneity, thereby biasing our estimates. However, this bias should not significantly affect our results as the treated polls selected have been conducted at relatively longer intervals (3 to 6 months) compared to the control polls. This limitation is difficult to remedy as no other Hong Kong sources provide comparable poll data to the HKPORI in terms of time span and variety.

6.2 Different Classifications of Polls

Another source of bias could be the classification of polls, which is inherently subjective, although qualitative criteria are used in Section 4.2. The problem arises as the high-sensitivity cohort is excluded from the control unit pool while the medium and low cohorts are included. If some high-sensitivity polls are wrongly included as control units, a downward bias could arise; wrongly excluded medium/low-sensitivity polls would cause an upward bias. However, we assign those on edge between medium and high sensitivity into the medium cohort so that the bias is downward and the results would be conservative. Therefore, it is likely that we have underestimated the DTE, not overestimating it. Future research can improve the classification by conducting list experiments to detect the current levels of political sensitivity bias in the polls and classify the polls accordingly.

6.3 Inability to Measure the Political Sensitivity Bias and the Treatment Effects

As discussed in Section 5.4, our results do not provide a direct measure of the political sensitivity bias or the treatment effects *per se*. This is because the pre-existing levels of bias before the NSL and the current levels of bias in the control group are unknown. Unfortunately, the past levels of bias are now impossible to measure quantitatively. Nevertheless, one can rely on their qualitative knowledge and relevant quantitative literature, for example, Kobayashi and Chan' s (2022), to assess the bias and effects qualitatively.

7 Conclusion

The coincidence of the abundant poll data in Hong Kong and the enactment of the Hong Kong National Security Law allows us to assess the *differential* treatment effect of the law on the political sensitivity bias in different polls. We have found that polls that are more relevant to the NSL, namely those on the independence of Taiwan/Tibet, have seen larger increases in their political sensitivity bias compared to those of lower political sensitivity. Specifically, the differential impact has been larger on the poll for

[36]See Assumption 5

[37]A negative η would indicate an anti-government bias, which is against the theory.

Taiwan independence than that for Tibet independence. The differential treatment effects on the two polls peaked at 22 and 9 months, respectively, after the law took legal effect.

The differential treatment effects can be considered evidence of political sensitivity bias in Hong Kong under the NSL, in line with the previous literature. This means that public opinion poll results have been unreliable under the NSL, even though they were collected scientifically and not falsified. However, this paper contributes to the literature by arguing that not only does the political sensitivity bias vary by polls and over time, but also the *differences* in bias between the polls vary over time. This implies that researchers and policymakers should be cautious when evaluating the dynamics of political sensitivity bias –low bias in some polls does not imply an equally low bias in others; the fact that some polls are unaffected by an event does not imply others are unaffected.

Future research can distinguish between the two sources of political sensitivity bias –preference falsification and non-response. This requires more detailed knowledge of the composition of respondents (and non-respondents) in each poll –which is currently unavailable from HKPORI open data. Furthermore, another interesting area to be explored is the impact of the institutions emphper se that conduct the survey. The detected bias could have been different if a government-sponsored institute, a domestic survey company, or a foreign-based institute were to conduct the survey or experiment. Such studies would shed light on the validity of outsourced survey results typical in the literature.

Acknowledgements We thank Dr Melissa Sands for her advice on the causal inference research design. DY expresses special gratitude to Dr Melissa Sands, the supervisor of his undergraduate dissertation upon which this paper is based. She provided pivotal guidance on the structure of this paper and encouraged us to explore novel approaches even when no precedents had been set in the field. Furthermore, we express our appreciation to Yifan Wang for his expertise and feedback on the Synthetic Difference-in-Differences estimator. We extend our gratitude to other people in the LSE community who have exchanged thoughts on this study with us and whose names can not be enumerated in this short section.

Data Availability Statement The replication files for this paper are available at https://doi.org/10.7910/DVN/2CGPGO.

Supplementary Material The supplementary material for this article can be accessed using this link.

Funding Statement This research is not funded.

Competing Interests The authors declear none.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (Feb. 2015). "Comparative Politics and the Synthetic Control Method". In: *American Journal of Political Science* 59 (2), pp. 495–510. ISSN: 00925853. DOI: 10.1111/ajps.12116.

Angeli, Datt (Sept. 2021). *The Impact of the National Security Law on Media and Internet Freedom in Hong Kong | Freedom House*. URL: https://freedomhouse.org/article/impact-national-security-law-media-and-internet-freedom-hong-kong (visited on 05/23/2023).

Arkhangelsky, Dmitry, Susan Athey, et al. (Dec. 2021). "Synthetic Difference-in-Differences". In: *American Economic Review* 111 (12), pp. 4088–4118. ISSN: 0002-8282. DOI: 10.1257/aer.20190159.

Arkhangelsky, Dmitry, Guido W. Imbens, et al. (July 2021). "Double-Robust Two-Way-Fixed-Effects Regression For Panel Data". In.

Basic Law (2021). *The Basic Law of the Hong Kong Special Administrative Region of the People's Republic of China*. URL: https://www.basiclaw.gov.hk/filemanager/content/en/files/basiclawtext/basiclaw_full_text.pdf (visited on 05/23/2023).

BBC (Dec. 2014). *Police clear final Hong Kong protest site at Causeway Bay - BBC News*. URL: https://www.bbc.co.uk/news/world-asia-30474687 (visited on 05/23/2023).

— (Oct. 2019a). *Hong Kong: Anger as face masks banned after months of protests - BBC News*. URL: https://www.bbc.co.uk/news/world-asia-china-49931598 (visited on 05/23/2023).

— (Nov. 2019b). *The Hong Kong protests explained in 100 and 500 words - BBC News*. URL: https://www.bbc.co.uk/news/world-asia-china-49317695 (visited on 05/23/2023).

Blair, Graeme, Alexander Coppock, and Margaret Moor (Nov. 2020). "When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments". In: *American Political Science Review* 114 (4), pp. 1297–1315. ISSN: 0003-0554. DOI: 10.1017/S0003055420000374.

Blair, Graeme and Kosuke Imai (2012). "Statistical Analysis of List Experiments". In: *Political Analysis* 20.1, pp. 47–77. DOI: 10.1093/pan/mpr048.

Bradburn, Norman M (1979). *Improving interview method and questionnaire design*. 1st ed. Jossey-Bass. ISBN: 087589402X.

Buckley, Noah et al. (2022). "Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes". URL: www.v-dem.net..

Carlitz, Ruth D. and Rachael McLellan (Mar. 2021). "Open Data from Authoritarian Regimes: New Opportunities, New Challenges". In: *Perspectives on Politics* 19 (1), pp. 160–170. ISSN: 1537-5927. DOI: 10.1017/S1537592720001346.

Chan, Johannes (Nov. 2014). "Hong Kong' s Umbrella Movement". In: *The Round Table* 103 (6), pp. 571–580. ISSN: 0035-8533. DOI: 10.1080/00358533.2014.985465.

Chapkovski, Philipp and Max Schaub (Apr. 2022). "Solid support or secret dissent? A list experiment on preference falsification during the Russian war against Ukraine". In: *Research & Politics* 9 (2), p. 205316802211083. ISSN: 2053-1680. DOI: 10.1177/20531680221108328.

Chau, Candice (July 2022). *Explainer: Hong Kong's sedition law - a colonial relic revived after half a century - Hong Kong Free Press HKFP*. URL: https://hongkongfp.com/2022/07/30/explainer-hong-kongs-sedition-law-a-colonial-relic-revived-after-half-a-century/ (visited on 05/23/2023).

Davidson, Helen (Apr. 2020). *Hong Kong using Covid-19 crisis as 'golden opportunity' for crackdown, says arrested leader | Hong Kong | The Guardian*. URL: https://www.theguardian.com/world/2020/apr/20/hong-kong-using-covid-19-crisis-as-golden-opportunity-for-crackdown-says-arrested-leader (visited on 05/23/2023).

Demosistō (2023). *About Us - 香港众志 Demosistō*. URL: https://web.archive.org/web/20180514213620/https:/www.demosisto.hk/about?lang=en (visited on 05/23/2023).

Driscoll, John C. and Aart C. Kraay (Nov. 1998). "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data". In: *Review of Economics and Statistics* 80 (4), pp. 549–560. ISSN: 0034-6535. DOI: 10.1162/003465398557825.

Ferman, Bruno and Cristine Pinto (2017). "Revisiting the Synthetic Control Estimator".

Freedom House (2019). *Hong Kong: Freedom in the World 2019 Country Report | Freedom House*. URL: https://freedomhouse.org/country/hong-kong/freedom-world/2019 (visited on 05/23/2023).

Frye, Timothy et al. (Jan. 2017). "Is Putin' s popularity real?" In: *Post-Soviet Affairs* 33 (1), pp. 1–15. ISSN: 1060-586X. DOI: 10.1080/1060586X.2016.1144334.

HKFP (Nov. 2019). *'A heartbreaking day': PolyU student leaders call for public help as hundreds trapped in campus - Hong Kong Free Press HKFP*. URL: https://hongkongfp.com/2019/11/18/heartbreaking-day-polyu-student-leaders-call-public-hundreds-trapped-campus/ (visited on 05/23/2023).

HKPORI (Jan. 2021). *POP releases five core social indicators (2021-01-19) - HKPORI - 香港民意研究所*. URL: https://www.pori.hk/press-release-en/2021-01-19-pm.html?lang=en (visited on 05/23/2023).

— (2023). *POP Poll - HKPORI - 香港民意研究所*. URL: https://www.pori.hk/pop-poll.html?lang=en (visited on 05/23/2023).

Ho, Kelly (July 2020). *Hong Kong teen activist Tony Chung bailed; told to remove social media posts within 72 hrs following security law arrest - Hong Kong Free Press HKFP*. URL: https://hongkongfp.com/2020/07/31/hong-kong-teen-activist-tony-chung-bailed-told-to-remove-social-media-posts-within-72-hrs-following-security-law-arrest/ (visited on 05/23/2023).

Jiang, Junyan and Dali L. Yang (Apr. 2016). "Lying or Believing? Measuring Preference Falsification From a Political Purge in China". In: *Comparative Political Studies* 49 (5), pp. 600–634. ISSN: 0010-4140. DOI: 10.1177/0010414015626450.

Kalinin, Kirill (Apr. 2016). "The social desirability bias in autocrat's electoral ratings: evidence from the 2012 Russian presidential elections". In: *Journal of Elections, Public Opinion and Parties* 26 (2), pp. 191–211. ISSN: 1745-7289. DOI: 10.1080/17457289.2016.1150284.

Kobayashi, Tetsuro and Polly Chan (Dec. 2022). "Political sensitivity bias in autocratizing Hong Kong". In: *International Journal of Public Opinion Research* 34 (4). ISSN: 1471-6909. DOI: 10.1093/ijpor/edac028.

Kobayashi, Tetsuro, Jaehyun Song, and Polly Chan (Dec. 2021). "Does repression undermine opposition demands? The case of the Hong Kong National Security Law". In: *Japanese Journal of Political Science* 22 (4), pp. 268–286. ISSN: 1468-1099. DOI: 10.1017/S1468109921000256.

Koehler, Kevin, Sharan Grewal, and Holger Albrecht (Aug. 2022). "Who fakes support for the military? Experimental evidence from Tunisia". In: *Democratization* 29 (6), pp. 1055–1076. ISSN: 1351-0347. DOI: 10.1080/13510347.2022.2032666.

Krumpal, Ivar (June 2013). "Determinants of social desirability bias in sensitive surveys: a literature review". In: *Quality & Quantity* 47 (4), pp. 2025–2047. ISSN: 0033-5177. DOI: 10.1007/s11135-011-9640-9.

Kuo, Lily (Apr. 2019). *Hong Kong's leader withdraws extradition bill that ignited mass protests | Hong Kong | The Guardian*. URL: https://www.theguardian.com/world/2019/sep/04/hong-kong-lam-to-withdraw-extradition-bill-say-reports (visited on 05/23/2023).

— (May 2020). *China threatens 'countermeasures' against UK over Hong Kong crisis | Hong Kong | The Guardian*. URL: https://www.theguardian.com/world/2020/may/29/hong-kong-crisis-china-support-police-warns-us-interfere-trump (visited on 05/23/2023).

Kuo, Lily and Patrick Wintour (July 2020). *Hong Kong: China threatens retaliation against UK for offer to Hongkongers | Hong Kong | The Guardian*. URL: https://www.theguardian.com/world/2020/jul/02/china-could-prevent-hongkongers-moving-to-uk-says-dominic-raab (visited on 05/23/2023).

Kuran, Timur (1987). "Chameleon Voters and Public Choice". In: *Public Choice* 53 (1), pp. 53–78. ISSN: 00485829, 15737101. URL: http://www.jstor.org/stable/30024732.

— (1997). *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press. ISBN: 9780674707573.

Lam, Jeffie, Lok-kei Sum, and Kang-chung Ng (Nov. 2019). *Hong Kong elections: pro-democracy camp wins 17 out of 18 districts while city leader says she will reflect on the result | South China Morning Post*. URL: https://www.scmp.com/news/hong-kong/politics/article/3039151/hong-kong-elections-tsunami-disaffection-washes-over-city (visited on 05/23/2023).

Lee, Peter (July 2022). *China's policies on Hong Kong receive record high 'positive appraisal' in Handover survey - Hong Kong Free Press HKFP*. URL: https://hongkongfp.com/2022/07/06/chinas-policies-on-hong-kong-receive-record-high-positive-appraisal-in-handover-survey/ (visited on 05/23/2023).

Lee, Sing et al. (Oct. 2010). "Evidence for the 2008 economic crisis exacerbating depression in Hong Kong". In: *Journal of Affective Disorders* 126 (1-2), pp. 125–133. ISSN: 01650327. DOI: 10.1016/j.jad.2010.03.007.

Ma, Joanne (June 2020). *Hong Kong pro-democracy group Demosisto disbands - YP | South China Morning Post*. URL: https://www.scmp.com/yp/discover/news/hong-kong/article/3091186/hong-kong-pro-democracy-group-demosisto-disbands (visited on 05/23/2023).

Nathan, Andrew J. (2001). "The Tiananmen Papers". In: *Foreign Affairs* 80 (1), p. 2. ISSN: 00157120. DOI: 10.2307/20050041.

Newey, Whitney K. and Kenneth D. West (May 1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55 (3), p. 703. ISSN: 00129682. DOI: 10.2307/1913610.

People's Daily (Apr. 1986). 必须旗帜鲜明地反对动乱 *[Must be clearly opposed to unrest]*. Chinese. URL: http://cpc.people.com.cn/GB/64184/64186/66681/4494004.html (visited on 04/29/2017).

Porreca, Zachary (2022). "Synthetic Difference-In-Differences Estimation With Staggered Treatment Timing". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.4015931.

Robinson, Darrel and Marcus Tannenberg (July 2019). "Self-censorship of regime support in authoritarian states: Evidence from list experiments in China". In: *Research & Politics* 6 (3), p. 205316801985644. ISSN: 2053-1680. DOI: 10.1177/2053168019856449.

Tang, Wenfang (Apr. 2016). "Political Trust". In: *Populist Authoritarianism*. Oxford University Press, pp. 134–151. DOI: 10.1093/acprof:oso/9780190205782.003.0008.

Whitehead, Laurence. (2002). *Democratization theory and experience*. Oxford University Press. ISBN: 0-19-160032-6.

Wong, Yiu-chung (2004). *"One country, two systems" in crisis : Hong Kong's transformation since the handover*. Lexington Books. ISBN: 0739104926.

Yu, Elaine et al. (Nov. 2019). *In Hong Kong, Daring Escapes From a Violent Siege at a University - The New York Times*. URL: https://www.nytimes.com/2019/11/18/world/asia/hong-kong-protests-university.html (visited on 05/23/2023).

Zeileis, Achim, Susanne Köll, and Nathaniel Graham (2020). "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R". In: *Journal of Statistical Software* 95 (1). ISSN: 1548-7660. DOI: 10.18637/jss.v095.i01.

A An Illustration of the Probit Model

The mechanism through which the use of z-scores can avoid the distortion of public opinion changes at extreme values can be demonstrated by the following simple probit model:

Assume the public opinion X_{it} on a given poll at a given time is one-dimensional and can be approximated by a normal distribution with mean μ and standard deviation σ. Assume there is a threshold γ for the minimum level of support required to give a pro-Beijing response in a survey. This can be visualised as Figure 9:

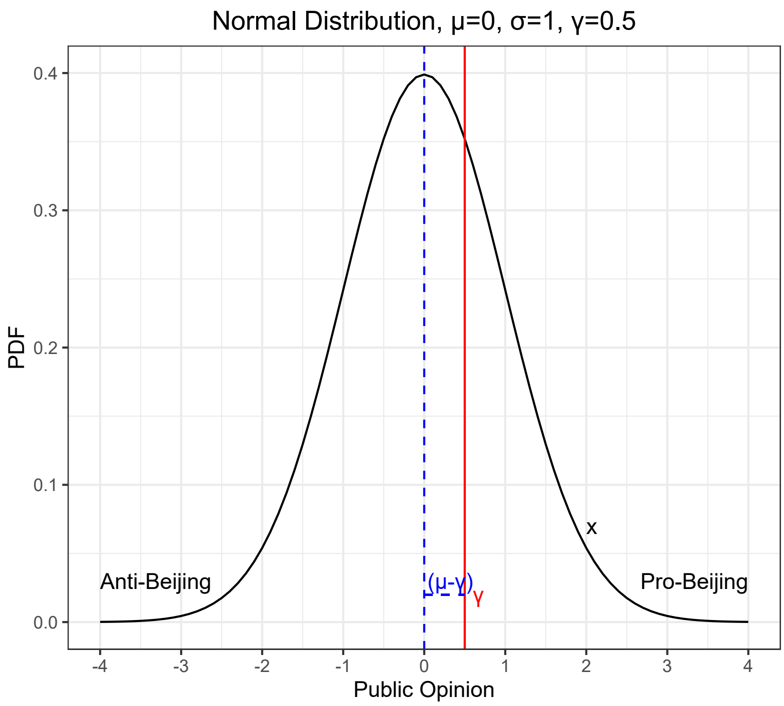

Figure 9: Illustration of a normally distributied public opinion with a threshold

The threshold line (γ) cuts the distribution in two halves –the right half represents those who will give a pro-Beijing response in the survey, and the left half represents the anti-Beijing respondents –the exact sizes of them are reflected in poll results. However, we argue that there is a better indicator of public opinion than the poll results per se: the difference between the threshold and the public opinion mean ($\mu - \gamma$). The superiority of the latter arises when the poll result is far from 50% (i.e. close to 0% or 100%) - in such situations, the changes in the poll result would be insensitive to changes in the public opinion distribution. This can be shown with the following example:

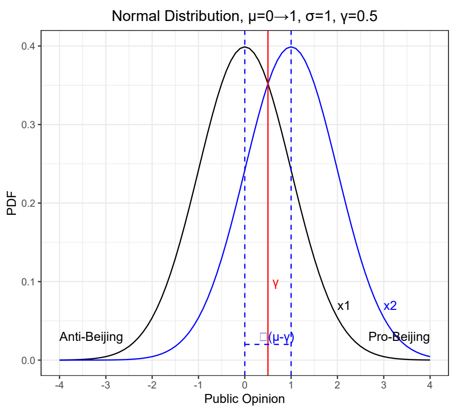

Figure 10: A shift in the distribution when μ is close to γ.

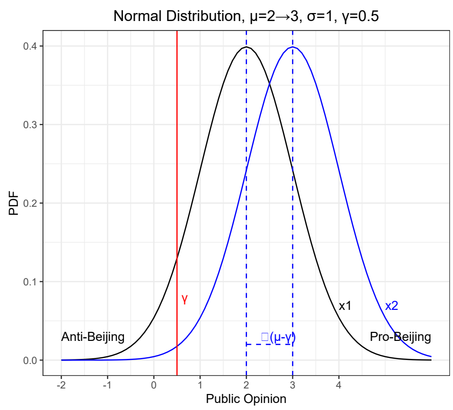

Figure 11: A shift in the distribution when μ is far from γ.

Suppose an event (e.g. the Hong Kong National Security Law) shifts the expressed public opinion distribution in favour of Beijing by one unit (assuming homogeneous effect of the event). Consequently, the difference between the threshold and the public opinion mean ($\mu - \gamma$) and the threshold increases from -0.5 to 0.5 (1 unit increase). When μ is close to γ, or the poll result is relatively close to 50% (as in Figure

10), the change in the poll result would be obvious; in this case, the share of the pro-Beijing responses increases from 31% to 69% (a 38 percentage point increase). However, when μ is far from γ, or the poll result is relatively far from 50% (as in Figure 11), the same change to the distribution $(\mu - \gamma)$ would only yield a small change in the poll results. In this case, $\Delta(\mu - \gamma)$ is still 1, but the corresponding share of the pro-Beijing responses increases by only 6.4 percentage points (from 93% to 99.4%).

So far, we have shown how $(\mu - \gamma)$ is a more reliable measurement for the outcome variable Y_{it} than the poll result per se $(\frac{A_{it}}{A_{it}+B_{it}})$ when the event has a homogeneous effect across the public opinion distribution (change in μ). Nevertheless, a heterogeneous impact (change in σ) could also be captured if a small modification is made. Now we use the standardised difference, instead of the absolute distance, between the threshold and the public opinion mean $(\frac{\mu-\gamma}{\sigma})$ to measure Y_{it}, which accounts for changes in the standard deviation, and this would be equivalent to the z-score formula we introduced in Equation (2):

$$Y_{it} = \frac{\mu - \gamma}{\sigma} = probit(\frac{A_{it}}{A_{it} + B_{it}}) = \Phi^{-1}(\frac{A_{it}}{A_{it} + B_{it}}) \tag{2}$$

The ability of this z-score measure to capture heterogeneous impact of an event on the public opinion distribution can be shown with the following examples with fixed distribution means but different standard deviations.

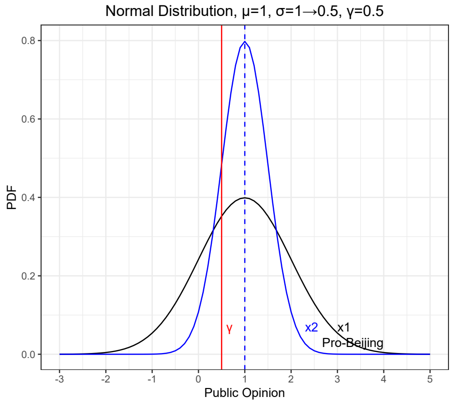

Figure 12: A change in   when μ is close to γ.

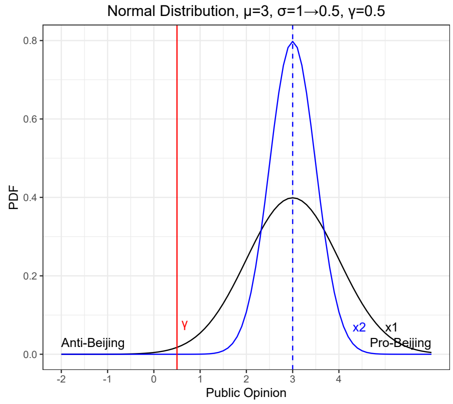

Figure 13: A change in   when μ is far from γ.

When μ is close to γ (poll result is close to 50%), the standardised difference (z-score) is fairly insensitive to a change in the standard deviation –in Figure 12 Y_{it} increases from 0.5 to 1. Whereas when μ is far from γ (poll result is far from 50%), as in Figure 13, the standardised difference (z-score) would be sensitive to such changes –in this case, it increases from 2.5 to 5. This means that the standardised method/z-score gives more weight to changes in the distribution away from the threshold –this aligns with our intention well: we want to highlight the concentration and dispersion of extreme political stances, which are rarer, and downplay that of the moderate stances.

In short, we propose the z-scores of the poll results using the probit function as a standardised measure of public opinion that can be summarised in the following formula:

$$Y_{it} = \frac{\mu - \gamma}{\sigma} = probit(\frac{A_{it}}{A_{it} + B_{it}}) = \Phi^{-1}(\frac{A_{it}}{A_{it} + B_{it}}) \tag{2}$$

This measure can correct for extreme values in the poll results, which can be distortionary in their crude forms. A change in Y_{it} could be simply interpreted as a homogeneous effect of an event on the public opinion distribution (change in μ). This could also be interpreted as a heterogeneous effect (changes in μ and σ), in which case more knowledge (perhaps qualitative) beyond the poll results is required for interpretation.

B Alignment of Time Periods

Since the poll surveys were not always conducted at the same time, we need to align them by their starting dates to form panel data. In the alignment process, for each control unit poll, we extract the data with the closest dates to those of the treated unit and dropped other data. Some data are duplicated for them being the closest to two or more dates of the treated unit.

Here is a visual demonstration of the alignment process:

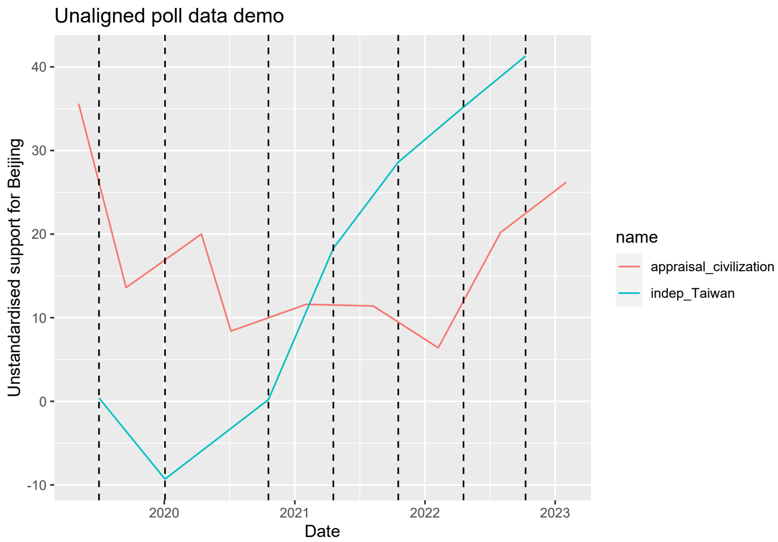

Figure 14: Selected unaligned poll data for demonstration.

Figure 14 shows the unstandardised poll results for Taiwan independence (`indep_Taiwan`) and the appraisal of the level of civilisation (`appraisal_civilization`) in Hong Kong. The 7 vertical dashed lines indicate the dates of the treated unit –the poll on Taiwan independence. We then identify the corresponding 7 time points in the poll data for `appraisal_civilization` that are the closest to the corresponding seven time points in `appraisal_civilization`, as drop other irrelevant data, as in Figure 15.

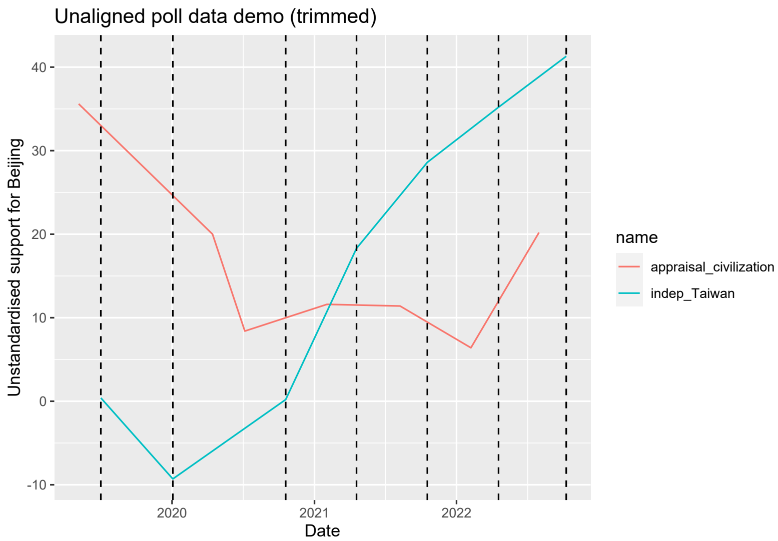

Figure 15: Selected trimmed poll data for demonstration.

After that, the remaining `appraisal_civilization` data can be aligned with the corresponding `indep_Taiwan` data, as in Figure 16.

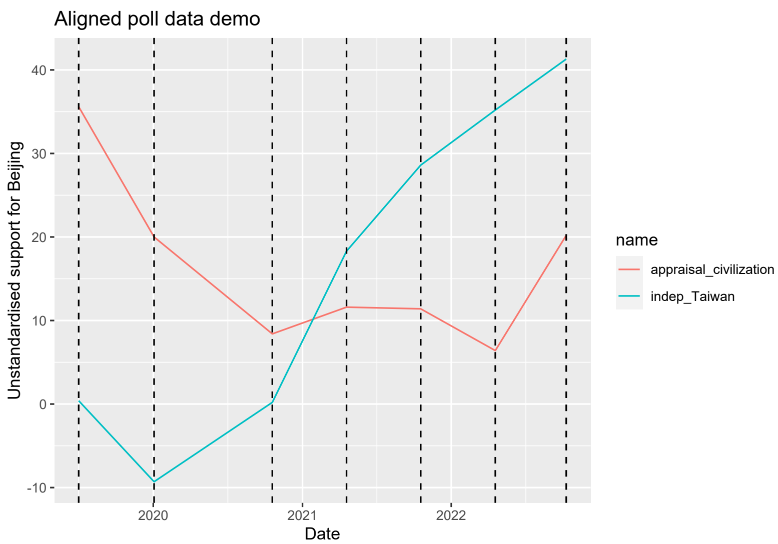

Figure 16: Selected aligned poll data for demonstration.

C Adjusted Outcome Differences in the Control Unit Plots

The poll-by-poll adjusted outcome difference is calculated by $\hat{\delta}_{tr} - \hat{\delta}_i$, where:

$$\hat{\delta}_{tr} = \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^{N} \hat{\delta}_i \tag{6}$$

$$\hat{\delta}_i^{sc} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} \tag{7}$$

$$\hat{\delta}_i^{did} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} - \frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} Y_{it} \tag{8}$$

$$\hat{\delta}_i^{sdid} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} - \frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} \hat{\lambda}_t^{sdid} Y_{it} \tag{9}$$

D More on the Choice of Standard Errors

Arkhangelsky et al. (2021) have provided three choices of standard errors - the *placebo, jackknife and bootstrap*. However, the latter two methods do not apply to estimations with only one treated unit (p.28). The placebo compromise given by Akhangelsky et al. (2021), however, typically yields way larger standard errors than expected to the point that even the sample result given by the authors are "statistically insignificant". The authors admit this in the introduction to the R package that "We have only one treated unit (California), so the only implemented method to estimate the standard error is the 'placebo' method described in Section 5 of Arkhangelsky et al. Because this is not trustworthy, it is not used by default: vcov instead returns NA. Here it is probably too large".

More on how the alternative *fixest* package that we use calculates the standard errors can be found using this link.

As for the Newey-West standard errors, heteroscedastic and serially correlated errors are corrected (Newey and West 1987), but not cross-sectionally correlated errors. Since cross-sectionally correlated errors are likely present in my data (some polls belong to the same "cohort" as shown in Supplementary Material), the NW method is a less desirable substitute for the DK method.

Another common method used for panel data is clustered standard errors at the unit level, which guards against within-cluster correlation of *independent clusters* (Zeileis, Köll, and Graham 2020). Nevertheless, the independence among units (polls) is in question ——a problem shared by the NW covariance matrix. Moreover, clustered standard errors do not directly account for serially correlated errors, which are likely in poll data. However, those two shortcomings alone do not preclude the clustered standard errors from being a valid robustness check –after all, the NW results suffer similar weaknesses. What precludes the clustered standard errors is that the SC regression does *not include unit (poll) fixed effects*, and results based on clustering by unit (poll) are thus unfair in this case. we nevertheless include the results with the clustered standard errors in Tables 6 and 7.

It is evident in those tables that the clustered standard errors show equivalent results to the NW and DK standard errors for SDID and DID estimators. However, those for SC are inconsistent with the NW and DK results due to missing poll fixed effects.

	SDID	SDID	DID	DID	SC	SC
Differential Treatment Effect	0.284***	0.284***	0.215***	0.215***	0.224***	0.224**
	(0.049)	(0.058)	(0.063)	(0.038)	(0.079)	(0.096)
Std.Errors	Newey-West (L=1)	Clustered by poll	Newey-West (L=2)	Clustered by poll	Newey-West (L=2)	Clustered by poll
FE: Poll	Yes	Yes	Yes	Yes	No	No
FE: Time	Yes	Yes	Yes	Yes	Yes	Yes
Effective N	210	210	1170	1170	390	390
R2	0.813	0.813	0.869	0.869	0.486	0.486
R2 Adj.(within)	0.240	0.240	0.006	0.006	0.055	0.055
RMSE	0.17	0.17	0.15	0.15	0.48	0.48

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 6: Poll: Rejection of Taiwan independence.

	SDID	SDID	DID	DID	SC	SC
Differential Treatment Effect	0.165***	0.165***	0.120**	0.120***	0.277**	0.277
	(0.029)	(0.032)	(0.044)	(0.038)	(0.130)	(0.193)
Std.Errors	Newey-West (L=1)	Clustered by poll	Newey-West (L=2)	Clustered by poll	Newey-West (L=2)	Clustered by poll
FE: Poll	Yes	Yes	Yes	Yes	No	No
FE: Time	Yes	Yes	Yes	Yes	Yes	Yes
Effective N	387	387	1305	1305	406	406
R2	0.912	0.912	0.873	0.873	0.197	0.197
R2 Adj.(within)	0.109	0.109	0.001	0.001	0.021	0.021
RMSE	0.17	0.17	0.15	0.15	0.61	0.61

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 7: Poll: Rejection of Tibet independence.

E Regression Result Table for Pre-Treatment Trend Tests.

	Taiwan.indep	Tibet.indep	June Fourth Incident
2008			
Placebo DTE	-0.342***	0.069*	0.070
	(0.060)	(0.037)	(0.060)
Std.Errors	Driscoll-Kraay (L=2)	Driscoll-Kraay (L=2)	Driscoll-Kraay (L=2)
Effective N	1440	1444	666
R2	0.907	0.914	0.971
R2 Adj.(within)	0.315	0.023	0.029
RMSE	0.17	0.16	0.17
2015			
Placebo DTE	-0.021	-0.010	0.129*
	(0.042)	(0.021)	(0.058)
Std.Errors	Driscoll-Kraay (L=1)	Driscoll-Kraay (L=1)	Driscoll-Kraay (L=1)
Effective N	252	600	400
R2	0.939	0.958	0.978
R2 Adj.(within)	-0.001	-0.001	0.136
RMSE	0.14	0.12	0.14
FE: Poll	Yes	Yes	Yes
FE: Time	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Effective observations for SC and SDID are smaller due to zero time and unit weights.

Table 8: Regression Result Table for Pre-Treatment Trend Tests.