STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree project in economics
Spring 2020

Close to Home: The Effect of Proximity to Violent Protests on Hong Kong's 2019 Electoral Outcomes

Matilda Fors (24246) and Andrea Schmidt (24181)

Abstract: This paper aims to investigate the effect of violent pro-democracy protests on electoral support for pro-democracy candidates in the case of Hong Kong's 2019 District Council election. More specifically, we estimate the size of an effect resting on three theoretically and contextually informed microfoundations: violence in protests, geographical proximity of protests to voters' homes, and temporal proximity of protests to election day. We argue that these three microfoundations are all incorporated in our designed proxy measure of MTR (metro) closures, allowing us to test their aggregate effect on constituency electoral outcomes. The analysis is pursued through an ordinary least squares method of multiple linear regression as well as a probit regression. Results show that the effect is statistically significant and robust across several model specifications. We conclude that geographical proximity to violent pro-democracy protests, within a specified time frame before election day, indeed has a positive marginal effect on support for pro-democracy candidates in the case of Hong Kong 2019. However, this effect should be viewed in relation to what is, in all likelihood, a dramatic shift in baseline support for pro-democracy candidates.

Keywords: democracy, election, Hong Kong 2019 pro-democracy protests, proximity, violent protests

JEL: D72, D74, D91

Supervisor:	Anders Olofsgård
Date submitted:	May 13, 2020
Date examined:	May 25, 2020
Discussant:	Lovisa Qvarner
Examiner:	Johanna Wallenius

Acknowledgements

Firstly, we would like to express our sincerest gratitude to our supervisor Anders Olofsgård, Associate Professor, for his invaluable support and insights. In addition, we would like to thank Professor Örjan Sjöberg for guidance, and Assistant Professor Yangguang (Sunny) Huang at the Hong Kong University of Science and Technology for support in the early research stage. Finally, we are grateful to our classmates Johan Allen and Johan Callermo, as well as everyone else who has provided feedback during this writing process.

Table of Contents

1. Introduction

Political elections are typically thought of as the outcome of rational processes in which voters gauge candidates' commitment to topics dear to their hearts, casting their vote according to a clearly specified set of political beliefs. However, just as behavioral economics has cast a light on how economic actors often behave irrationally in relation to the textbook principles of *homo economicus*, there should also be room for considering psychological, social and cognitive aspects of political life. Of course, the risk of this endeavor is, by undermining the legitimacy of the mechanism through which officials are elected, to undermine the foundations of democracy itself. On the other hand, political campaigners have known since the dawn of democracy that voters can be coaxed, cajoled, intimidated or empowered through various forms of political messaging. Analyzing how voters respond to cognitive and psychological stimuli should therefore be of interest to anyone who wants to understand how democratic processes might be supported and solidified.

In this paper, we examine the relationship between violent political protests and electoral outcomes by studying Hong Kong's District Council election, which was held on November 24, 2019. The election took place in the midst of a period of political turmoil and violent pro-democracy protests. As a matter of fact, it followed immediately upon a seventeen-day period of particular violence and disruption to everyday life, brought on by the death of an undergraduate student, which happened as clashes between protesters and police went on nearby. This should make the Hong Kong 2019 District Council election a particularly interesting natural experiment in order to study voter behavior as a response to violent protests. As noted by Enos, Kaufman and Sands (2019, p. 1027) "[t]he implicit threat of violence underlies the relationship between governments and citizens in many places." The question, however, is how violence changes this relationship.

1.1. Research Purpose

The Hong Kong 2019 case is unique. The pro-democracy movement initially enjoyed broad support – early rallies drew as many as one million people, more than 10 % of the population, to peacefully march the streets (Robles, 2019) – and, in spite of high levels of violence and long-drawn disruptions to Hongkongers' daily lives, translated into a landslide victory at the polls of the District Council election in November 2019. The current literature provides competing theories that suggest that violent protests both increase and decrease support for the protest movements; a macro snapshot of Hong Kong seems to indicate the former. However, in this study we wish to delve deeper into the microfoundations of these results.

The aspect of violence in protests plays a critical role for electoral outcomes since violence has been argued to increase the salience and support for protests (Gillion and Soule, 2018, p. 1660) as well as signal credible discontent to both other protesters and the regime, which may thereby affect even the support from non-participating voters (Desai, Olofsgård and Yousef, 2020). On the other hand, violent protests have also been suggested to reduce the public's identification with the movement, thereby also reducing the public's support (Muñoz and Anduiza, 2019), or even increase support for politically opposing groups (Simpson, Willer and Feinberg, 2018). Furthermore, it has been suggested by scholars such as Aidt and Franck (2015) and Enos, Kaufman and Sands (2019) that geographical proximity to violent protests may play a crucial role in electoral outcomes, by increasing support for protesters' causes. Theories from behavioral science (Tversky et al., 1973) also suggest that temporal proximity to an event can increase its salience, thereby reminding voters of important topics and influencing their decision making process.

Building on these theories about violence in protests as well as proximity – both geographical and temporal – the purpose of this paper is to test whether violent pro-democracy protests in people's geographical vicinity during the final days before election day had any explanatory power in the case of Hong Kong and the 2019 District Council election. More specifically, we examine whether violent pro-democracy protests in geographical proximity to one's home and temporal proximity to election day affects people's preferences to vote for pro-democracy candidates. The aim of this exercise is not only to learn more about the specific case of Hong Kong in 2019, but also to contribute to a broader understanding of the role of violent protests in electoral change from a behavioral perspective.

In our analysis, we employ closures of MTR (metro) stations as a proxy for where violent pro-democracy protests took place within a specified time frame before the election, and test whether proximity to MTR closures has any effect on the proportion of votes for pro-democracy candidates in the 2019 District Council election. This is done by pursuing both an OLS method of multiple linear regression and a probit regression, which allow us to test the hypothesis across several model specifications. The analysis is limited to the seventeen-day period leading up to election day, using data from 409 of Hong Kong's 452 constituencies. Results indicate that, on constituency level, geographical proximity to violent pro-democracy protests in the days before election day indeed has a marginal effect on the voter decision in the form of increased support for pro-democracy candidates. The results are statistically significant and robust across several model specifications.

Our paper is organized as follows. We begin by providing a review of previous literature concerning the relationship between violent political protests and changes in political support, particularly about the impact of proximity to violent protests on electoral outcomes. In Section 3, we move on to provide a background on the 2019 pro-democracy protests of Hong Kong and the 2019 District Council election. Section 4 formulates our research focus, contribution to the current state of knowledge and hypothesis. Section 5 describes our data collection methods and Section 6 introduces our constructed variables. In Section 7, we provide an empirical framework which aims to test the relationship between violent pro-democracy protests and the preference to vote for pro-democracy candidates in the case of Hong Kong 2019. Section 8 presents our empirical results and sensitivity analyses. Section 9 discusses these results and their implications. Lastly, we conclude the thesis in Section 10 with some final remarks as well as suggestions for future research.

2. Literature Review

Movements of violent protests and social unrest are important vehicles for translating demands for change into political reform. However, there is no consensus on the exact mechanism through which this translation takes place, especially considering that there is typically no institutional procedure that takes the views of protestors into account (Gillion and Soule, 2018). Even though some work has been done on the link between protests and government policy, the link between protests and elections is largely uncharted territory (ibid). The main empirical difficulty lies in estimating the direction of causality, as well as finding appropriate data (Enos, Kaufman and Sands, 2019).

In this section, we will first go over the main models explaining how political protests relate to changes in political support, with special focus on the competing theories about the impact of violent protests. Thereafter, we will give a brief overview of the previous literature regarding proximity to violent protests in relation to voter decision making, both from a geographical and temporal perspective.

2.1. Protests and Elections

2.1.1. Protests and Increased Support

Do protests affect electoral returns in the political direction expressed by protesters, and if so, what is the mechanism through which this happens? Scholars such as Lee (2002, cited in Gillion and Soule, 2018, p. 1650) have argued for how protesters are “mobilizing public opinion” by spreading information to their peers in a bottom-up manner, which stands in contrast to the top-down information flows from political elites to voters. In Lee’s view, the peer-to-peer dynamic of political information has a higher influence on voters’ decisions; leading to the conclusion that protests are effective in mobilizing voters to support the electoral outcomes sought by the protest movement. Similarly, Gillion and Soule (2018) present the voter effect hypothesis; namely, that protests expressing liberal (conservative) opinions tend to increase the share of the two-party vote for the Democratic (Republican) candidate in U.S. elections. More specifically, Gillion and Soule (2018) think of protests as informative cues that voters may use to re-evaluate candidates and reassess particular topics. The scope of the informative cue can be measured in terms of salience. In Gillion and Soule’s definition, some factors that make a protest more salient include high turnout rates (>100 people), long duration (>1 day) of the protest, property damage, arrests, police presence, injuries, the use of weapons and deaths. Furthermore, the informative cues of protests do not only serve the electorate, but politicians as well. Gillion and Soule introduce the vulnerability hypothesis; namely that protests expressing liberal (conservative) positions will lead to an increase in emerging Democratic (Republican) candidates. They test their hypotheses on election results for the U.S. House of Representatives from 1960 to 1990 and based on protest data from the Dynamics of Collective Action (DCA) database, finding that protests indeed have an effect on electoral outcomes. Firstly, the findings confirm the voter effect hypothesis and argue that protests can educate, inform and remind the electorate of important questions, thus influencing electoral outcomes. Secondly, the findings confirm the vulnerability effect and support the view that protests serve as important cues for emerging political candidates for when to enter the race. Gillion and Soule (2018, p. 1660) conclude by saying that "[protests] are signals of political information that do not go unnoticed by the local electorate."

2.1.2. Protests As Signals

Similar to the view of protests as informative signals, Desai, Olofsgård and Yousef (2020) propose that the use of violence or non-violence in protests is best viewed as a strategic choice. Violent protests, they argue, come at a higher risk for the protesting individual and thus send stronger signals of credible discontent to both one's peers and the regime. Furthermore, if there is a common and idiosyncratic source of discontent among voters, violent protests may – under the assumption that politically active groups are better informed – cause others to update their beliefs about the political situation. The more costly the action, i.e., the more violent, the stronger is the update, which will affect voting. As such, violent protests may increase the support for the movement even among non-participating voters. In seeing violence as a strategic signals, one can think of the informative cues (cf. Gillion and Soule, 2018) of protests as amplified, thus eliciting even bigger responses in the electorate.

2.1.3. When Protests Backfire

Just as violence can increase the salience of, and thereby the support for, protests it can also lead to the opposite. Simpson, Willer and Feinberg (2018) develop a theory for what happens when protesters resort to violence. Violence, they argue, causes protesters to be perceived as less reasonable, thereby reducing the public's identification with them and consequently also reducing public support. As a matter of fact, Simpson, Willer and Feinberg (2018) suggest that violent protests may increase support for politically opposing groups. They derive their results from a large (n = 800) survey experiment in which four groups of participants were asked to rate their perception of white nationalists and anti-racist counter-protesters, respectively, after having read four different manipulated articles regarding the use of violence during a clash between the two protesting groups. Interestingly, the authors find that violent anti-racist counter-protestors led to stronger support for white nationalists. However, violence from white nationalists did not translate into the same support for anti-racists. This is interpreted as resulting from prior expectations on the violent tendencies of the two groups; in short, that violent behavior from historically or stereotypically non-violent groups would be more salient and thus create stronger counter reactions among the public.

Muñoz and Anduiza (2019) find similar effects following a sudden riot in Barcelona: witnessing violence on the streets reduced the public's support for the movement. Though the overall support decreased, results were heterogeneous across partisan groups. Among those classified as core supporters of the protest movement, support dropped by 6.8 percentage points. Among those classified as weak supporters, opposers and nonaligned, support dropped by 15.5, 13 and 15.5 percentage points, respectively. This indicates that although violence in protests may reduce the public's support, the size of this effect largely depends upon priors such as political beliefs.

In conclusion, the research of both Muñoz and Anduiza (2019) and Simpson, Willer and Feinberg (2018) suggests that though protesters receive less support from the public after they engage in violent activities, political beliefs and expectations play key roles in determining the size of this reduction.

2.2. Introducing Proximity

Evidently, the literature suggests that violent protests can both increase and decrease political support for protest movements. In order to narrow the analysis, one interesting point of departure is that of proximity; namely, the hypothesis that the closer a protest is to a voter or political decision maker, the bigger influence it will have on his or her decisions. Here we will focus on two kinds of proximity: geographical proximity and temporal proximity.

2.2.1. Geographical Proximity

Aidt and Franck (2015) have studied the relationship between the Swing riots, an uprising of landless agricultural workers in rural Britain, and the 1832 Great Reform Act which paved the way toward universal suffrage in Britain. The riots were highly violent, often including the destruction of agricultural machinery. In their study, Aidt and Franck (2015) note that historical evidence suggests that the Swing riots were neither an organized political movement with a clear aim of parliamentary reform nor that the rioters were involved in political associations. Rather, the riots seem to have been caused by rural socioeconomic deprivation, particularly by the poor harvest in 1828-1829. Therefore, Aidt and Franck make the assumption that the Swing riots arose spontaneously in certain areas rather than were organized by supporters of parliamentary reform. Building on this assumption, they show that the geographical proximity to Swing riots had an impact on the proportion of Whigs (reform-friendly politicians) elected to the House of Commons in the 1831 election, in turn paving the way for the Great Reform Act. Aidt and Franck interpret this result, which holds up robustly after falsification tests, as evidence of how patrons and voters came to support the Whigs after having witnessed first-hand the destruction and violence in their immediate surroundings. This interpretation, in turn, builds on the "threat of revolution" theory, which entails that the political elite would vouch for democratic policy changes out of fear of being overthrown by violent revolutionaries.

Similarly, Enos, Kaufman and Sands (2019) have studied support for improvements in urban minorities' living conditions, such as public school spending, after the 1992 Los Angeles riots. The riots took place in the wake of four white police officers' assault of African-American Rodney King in Los Angeles. The event, which was videotaped, sparked outrage in the African-American community and led to one month of protests, vandalism, more than 11,000 arrests and an estimated $1 billion of material damages (CNN, 2013, cited in Enos, Kaufman and Sands, 2019, p. 1013). Just one week after the last troops of the National Guard, which were sent in to stifle the protests, withdrew, the 1992 primary elections were held. Using geocoded voter data, Enos, Kaufman and Sands find that support for higher public school spending (used as a proxy for support for improvements in urban minorities' living conditions) increased among voters who lived in close proximity to where riots had taken place. By comparing election results from Los Angeles County to election results from other parts of California, such as San Diego and San Francisco, Enos, Kaufman and Sands are able to isolate the effect of geographical proximity on policy support. Even though other parts of California, and even the nation, were indirectly exposed to the riots via media coverage, this did not translate into a change in policy support that is comparable to the change seen in Los Angeles County. In conclusion, the study finds that witnessing large and violent riots in one's immediate vicinity has a positive effect on changes in policy support.

Drawing on the work of Aidt and Franck (2015) and Enos, Kaufman and Sands (2019), it is possible to think of policy support as a function of geographical proximity to violent protests.

2.2.2. Temporal Proximity and the Availability Heuristic

The concept of the availability heuristic is widely used to study decision making, especially in behavioral economics and behavioral finance. At its heart, the availability heuristic is a mental shortcut used to assess frequency of a class or probability of an event, based on how easily such cases can be brought to mind. As a consequence, information that is more retrievable will have a disproportionately large effect on decision making, leading to systematic and predictable biases. In particular, emotionally salient experiences or recent occurrences tend to be more retrievable than those that are mundane, everyday and happened a long time ago (Tversky et al., 1973). For example, if one has just heard about a traffic accident, one might be more careful when driving, even though the objective probability of a traffic accident has not changed. Furthermore, witnessing the traffic accident first-hand rather than reading about it in the newspaper increases the individual's subjective probability assessment regarding the occurrence of traffic accidents.

When individuals cast their vote on election day, they are essentially making a decision on what future they would like to see, based on, among many other things, the probability they ascribe to certain scenarios. This should put the availability heuristic in focus. For example, Krupnikov (2014) sets out to answer the question whether negativity in political campaigning affects voter turnout, and instead finds that the political content seems to be less important than the timing of the message; that is, negativity in political campaigning can both mobilize and demobilize voters depending on the timing of the exposure to the message. This offers support for a boundedly rational view of political life, where voters may cast their votes on far less rational grounds than their pronounced support for certain policies or candidates.

3. Background

Relying on the previous literature about violent political protests and electoral outcomes, we now present a background of the 2019 pro-democracy protests of Hong Kong and the following District Council election, which serves as this study's empirical focus. To begin with, we discuss the underlying causes of the protests and provide a brief overview of the key events. Secondly, we explain the form and function of Hong Kong's District Council elections as well as discuss the results of the 2019 election. Finally, we present Hong Kong's MTR (metro) system and how it relates to the protest movement.

3.1. The 2019 Hong Kong Protests

Hong Kong, which is a special administrative region (SAR) of China, has been governed under the principle of "one country, two systems" since its handover from Britain to China in 1997. It is set to remain so until 2047, when the region is planned to be integrated into the mainland. In practice, this means that Hong Kong has an economic and administrative system separate from mainland China, as well as judicial independence and freedom of press (Yeung, 2019). However, Hong Kong's autonomy in relation to mainland China remains a heated political question. In 2014, the central government issued a decision regarding proposed reforms to the Hong Kong electoral system – namely to implement pre-screening of the candidates for the Chief Executive of Hong Kong – which triggered a series of protests known as the Umbrella Movement (Kaiman, 2014). In February 2019, the Hong Kong government proposed a bill that would enable extradition of Hong Kong citizens to mainland China (Reuters, 2019). In much the same way as in 2014, the extradition bill tapped into growing feelings among Hongkongers that its independence was being infringed upon ahead of time. For example, many worried that the new law could be used to extradite journalists and political dissidents to the mainland, where chances of a fair trial were perceived as slim (BBC, 2019). As a consequence, large protests again erupted in June 2019, forming the starting point of the 2019 pro-democracy protests of Hong Kong.

On June 9th, 2019, approximately one million people took to the streets to march peacefully in protest against the extradition bill (Robles, 2019). Initially, protesters only demanded the withdrawal of the bill, but following an escalation of the police's tactics against protesters, the objective shifted to demands for an independent investigation of the police's use of force in stifling the protests, which was one of the protesters' "five demands."[1] Starting out as peaceful, the protests soon took a violent turn with storming of the Legislative Council (LegCo) on July 1st, a violent mob attack on protesters in rural Yuen Long MTR station, general strikes throughout August and numerous violent clashes between protesters and police, where tear gas, rubber bullets and live ammunition were used by the police and protesters threw petrol bombs and bricks (Robles, 2019). The bill was formally withdrawn by Hong Kong's chief executive Carrie Lam on September 4th, 2019, though many critics claim it was "too little, too late" (Reuters, 2019). Violence ramped up in October and November following a ban on wearing masks or facial covering to public gatherings, which were used extensively by protesters in order to protect their identity. On November 8, an undergraduate student from the Hong Kong University of Science and Technology (HKUST) died due to an injury sustained in a fall from a

[1] The five demands of the pro-democracy protest movement are: (1) the full withdrawal of the extradition bill, (2) an independent investigation into alleged police brutality, (3) that the protesters should not be classified as "rioters", which is a crime punishable by up to 10 years in prison according to Hong Kong law, (4) amnesty for arrested protesters, and (5) universal suffrage for Hong Kong's two major governing bodies, the Chief Executive and the Legislative Council (Wong, 2019).

parking garage as clashes between protesters and police went on nearby (Lum, 2019). This event – the first confirmed death directly related to clashes between police and protesters – sparked a new wave of rage in Hong Kong (Leung, 2019). University students occupied campus areas in violent standoffs that lasted for days, and protesters set out to paralyze the city by blocking vital infrastructure such as tunnels and highways. On November 11, just thirteen days before the 2019 District Council election, protesters called upon the public to join them in a five-day city-wide strike as a response to the death of the HKUST student, which included purposeful disruptions to MTR and road traffic. On the same day, a man was doused with a flammable liquid and set on fire during a dispute with a protester in Ma On Shan, and another young protester was shot with live ammunition in the abdomen by a police officer (South China Morning Post, 2019a). All in all, the seventeen days between the death of the HKUST student and the election were marked by extreme violence, even compared to the previous months of protests.

3.2. Hong Kong's District Council Election

The District Council election is Hong Kong's closest equivalent to universal suffrage. Hong Kong is divided into eighteen districts, each with their own district council. The districts are then subdivided into several constituencies. All in all, Hong Kong has 452 constituencies. There can be multiple candidates competing in the same constituency, but the one who receives the most votes will be the one elected to represent their constituency. For example, the Wan Chai district has thirteen constituencies and the Wan Chai district council thereby has thirteen seats. Apart from the 452 seats for all of the constituencies, the rural districts of Tsuen Wan, Tuen Mun, Yuen Long, North, Tai Po, Sai Kung, Sha Tin, Kwai Tsing and Islands also have seats reserved for ex-officio Rural Committee Chairmen in the New Territories. Representatives holding these 27 seats are not elected by the public (District Council Election, 2019). Altogether, the number of seats in the district councils sum up to 479.

The function of the district councils is mainly to serve as a link between the Hong Kong government and the local population. District councilors play a small but not negligible role in the election of Hong Kong's highest political leader, the Chief Executive, as they hold 120 seats in the 1,200-member committee that elects the Chief Executive (Lahiri and Hui, 2019). Still, the primary function of the district councils is to advise the government on matters such as public facilities and services, community well-being, the local environment and the use of public funds in the community. They are also responsible for promoting community and cultural activities as well as making environmental improvements in their respective districts. For that reason, District Council elections have typically been focused on local matters such as traffic improvements and management of public spaces rather than regional politics (District Council Election, 2019).

However, due to the intensity of the protests, many analysts consider the 2019 District Council elections a *de facto* referendum on the pro-democracy movement. Ivan Choy Chi-keung, senior lecturer in politics at Chinese University (cited in Magramo, 2019) said the 2019 election would "have a more symbolic meaning rather than an actual change in governance." Nonetheless, voter turnout was historic; on election day, 71.2 % of 4.1 million registered voters cast their vote, in comparison to the 47 % turnout in the 2015 election and the previous record of 58 % in the 2016 Legislative Council election (South China Morning Post, 2019b). Pro-democracy candidates won a landslide victory, taking home – together with independent candidates who expressed democratic sympathies – 392 seats, compared to the 2015 result of 116 seats. Conversely, the pro-Beijing camp suffered a defeating loss as it went from 292 to 60 seats (South China Morning Post, 2019c).

Table 1: Hong Kong's District Council Election Results

	Seats won 2015	**Seats won 2019**
Pro-Beijing camp	292 (67.75 %)	60 (13.27 %)
Pro-democracy camp	116 (26.91 %)	392 (86.73 %)
Independent candidates	23 (5.34 %)	*
Total	431 (100 %)	452 (100 %)
Voter turnout rate	47.0 %	71.2 %

Table 1: Distribution of seats won in the District Council 2015 and 2019, percentage in parentheses. On the last row, voter turnout is reported in percent. *Source:* South China Morning Post, 2019c.

* In this categorization, conducted by South China Morning Post (2019c), pro-democracy and independent candidates are viewed as belonging to one and the same category in the 2019 results. However, we will in this study follow a more granular approach, described in detail in Section 5.1, in our categorization of the candidates and throughout the analysis. Lastly, in a concluding sensitivity analysis we will extend the definition of pro-democracy candidates – for a selected group of constituencies – to also include independents.

3.3. The Mass Transit Railway (MTR) System

The Mass Transit Railway (MTR) is the operator of public transport in Hong Kong. More specifically, MTR operates heavy rail, light rail and buses all across Hong Kong.[2] The heavy rail is the metro system of Hong Kong, consisting of eleven lines with a total of 93 stations as of 2019.

Throughout Hong Kong's months of protests in 2019, MTR corporation and stations were targeted by pro-democracy protesters. Stations were regularly subject to vandalism such as destruction of property, fires or intentional floodings; already in October, MTR corporation stated that 1,200 ticket barriers, 800 ticket machines, 900 CCTV cameras, 50 escalators and 40 lifts had been damaged by protesters (Low, 2019). The reason for this outrage was that protesters perceived MTR as doing Beijing's, by way of the HKSAR government's, bidding by suspending its services early in the evenings, thereby imposing a *de facto* curfew and curbing demonstrations (Yau, 2019). In addition, protesters perceived MTR to support police violence by allowing riot police to enter stations and trains in order to conduct mass arrests, for example as happened

[2] A note to the reader: throughout this paper, "MTR stations" will be used to refer to metro stations only, thus excluding bus lines and local light rail lines from the analysis.

when special forces stormed a stationary train at Prince Edward station on August 31 and arrested 63 people (Siu, Lum and Low, 2019). The event quickly gained the public's attention as police were accused of using excessive force, giving MTR the epithet "Murderer Transit Railway" in protester circles. The government, however, accused MTR corporation of being too lenient in the early days of the protests by not shutting down service in protest-stricken areas, thereby letting violent protesters escape arrest (Low, 2019).

4. Research Design

This section presents a closer description of the relationship we aim to analyze. First, we develop a theoretically and contextually informed framework for how closures of MTR stations relate to the 2019 District Council election outcomes. We then move on to argue for a crucial assumption of our analysis, namely that the geographical variation in the occurrence of protests is independent of constituency characteristics, in the Hong Kong 2019 case. At last, we present our research question and following baseline hypothesis.

4.1. The Link Between MTR Closures and Election Results

Based on previous research, we believe that there are three microfoundations that all increase the salience of protests among voters and thus, in part, can explain how protests translate into electoral outcomes. The first microfoundation is the element of violence in protests, which functions as an informative cue that leads the electorate to update their beliefs about important topics (Gillion and Soule, 2018). Violence also signals credible discontent (Desai, Olofsgård and Yousef, 2020), which has a similar effect. The second microfoundation is voters' geographical proximity to protests, which also creates stronger updates in political support, compared to protests that do not take place in the immediate vicinity (Aidt and Franck, 2015; Enos, Kaufman and Sands, 2019). The third and final microfoundation is protests taking place in temporal proximity to election day. Temporal proximity to protests make them more salient, causing them, via the availability heuristic (Tversky et al., 1973) to have a disproportionately large impact on decision making such as the voting decision.

We suggest that our measure of MTR closures incorporates these three microfoundations, making it possible to test their aggregate impact on electoral outcomes. Firstly, as we have argued in Section 3.3, there is a clear link between violence in protests and MTR closures in the Hong Kong 2019 case. Causality seems to run both ways. On the one hand, MTR stations closed down as a precautionary measure if protests could be anticipated in an area, causing more outrage among protesters. On the other hand, violent protesters targeted MTR stations, forcing them to shut down and in many cases remain shut for repairs for some time thereafter. Secondly, as each MTR station has a specific physical location, MTR closures allow us to map out the geographical location of violent protests. Thirdly, as MTR stations closed down more or less simultaneously as when protests took place, this proxy makes us able to capture the time frame of violent protests. In this paper we only wish to study violent protests during the seventeen particularly violent days leading up to election day, and because MTR closures capture the timing of violent protests, it allows us to pursue this limitation by only including MTR closures occuring during these days.

To sum up, we believe that georeferenced MTR closures for this seventeen-day period is a proxy that allows us to quantify the violent protests experienced by each of Hong Kong's districts and their respective constituencies right before election day.[3] As such, we consider MTR closures to be the most appropriate measure for capturing where and when violent protests took place, compared to, for instance, the number of participating protesters or compiled media coverage of the intensity of the protests. As always, a word of caution is due when analyzing the results because all proxies, of course, pose a risk of disregarding real observations.

4.2. The Conditional Independence Assumption

Importantly, our analysis is based on the assumption that protests in a particular constituency occurred independently of the characteristics (e.g., demography) of the people living in that constituency. In other words, we see the geographical variation in occurrence of protests as an exogenous variable rather than an endogenous characteristic of a constituency. As such, we argue that, on constituency level, geographical variation in protests and voting behavior are conditionally independent of constituency characteristics. This assumption is critical because that is what allows us to make inferences about the link between geographical proximity to protests and the preference for voting for pro-democracy candidates at constituency level. Without this assumption, i.e., if people were instead systematically protesting in their own neighborhoods, our estimate would be biased and thus not valid. The conditional independence assumption is supported by three main arguments, all based on historical evidence.

Firstly, this assumption is strengthened by the protesters' choices of protest locations. Protests were typically centered around hotspots, such as universities or the car park where the HKUST student fell, rather than in residential areas. Protest sites appear to have been chosen rather arbitrarily, especially since protesters began employing the "be water" tactic, inspired by the philosophy of Hong Kong martial arts star Bruce Lee. In the context of the protests, "be water" meant that protesters would gather quickly, often in multiple locations all across the city, and disperse just as quickly. "[The protesters] are really fluid, and sometimes they get together very quickly and they disperse very quickly, and it really looks like water is flowing and flowing through different parts of the city," said Masato Kajimoto, an assistant professor of Journalism and Media Studies Center at the University of Hong Kong (cited in Satoshi, 2019). Lam, Ng and Xinqi (2019) describe the protesters as moving in "unexpected waves" and, having learned the lessons from the 2014 Umbrella Movement, "traded the strategy of prolonged mass sit-ins for spontaneous road blockades and circling of buildings – a 'formless' protest in Lee's words – to sustain their momentum and secure the continued goodwill of the public." The "be water" tactic offers support for the view that protesters traveled to protest sites, often selected randomly and spontaneously, rather than protested in their own neighborhoods.

---

[3] A note to the reader: throughout this paper, "MTR closures" and "violent pro-democracy protests" will sometimes be used in alteration to refer to the same thing, as we, in the context of this paper, see the former as a proxy for the latter.

Secondly, the assumption is supported by the protesters' organizing tactics. As a matter of fact, downloads of encrypted messaging app Telegram surged in Hong Kong during the months of protests, increasing by a factor of ten between July 2018 and July 2019 (Banjo, 2019). Thousands of Hong Kong protesters used Telegram to anonymously coordinate activities and spread information about upcoming rallies and flash mobs, as the encrypted app implied less risk of police infiltration (Schectman, 2019). This, too, corroborates the view that protesters gathered at selected hotspots rather than protested outside their front doors.

Thirdly, it is further strengthened by the geographical location of the lunchtime protests that were held throughout November in various districts including Central, Tai Koo, Causeway Bay, Wong Chuk Hang and Kwun Tong, with hundreds of participants at each spot and major roads blocked (Lau, Low and Cheung, 2019). Particularly in Central, the city's financial district, office workers together with other protesters persistently marched the streets during lunch hours, chanting slogans and singing songs, despite being confronted by police officers and tear gas (RTHK, 2019). Since the average amount of time people spend commuting with public transit in Hong Kong on a weekday is 73 minutes (Moovit, 2017), one can assume that most people did not live in the same area as where their workplace was located. Consequently, as the lunch demonstrations were held outside people's offices, that was usually not the same constituency as these people were voting in.

Altogether, our analysis rests on the assumption that, on constituency level, there is no variable that simultaneously affects the likelihood of pro-democracy protests and pro-democratic votes in a particular constituency. This is not to deny the fact that there is always a risk of factors that are correlated with both of these events, thereby causing omitted variable bias. To take this into account and further solidify the important assumption of conditional independence, we will still control for a carefully selected set of characteristics of the constituencies in our regressions, as presented in Section 6.3.

4.3. Contribution and Research Question

It is true that pro-democracy candidates succeeded overall in the 2019 District Council election, winning approximately 87 % of the seats (South China Morning Post, 2019c). However, we have theoretical as well as empirical grounds for believing that there are at least three microfoundations of this result that can make it more nuanced: violent protests, geographical proximity of voters to protests, as well as temporal proximity of protests to election day. As such, we build on Gillion and Soule's (2018) view of violent protests as strong informative cues, Aidt and Franck's (2015) work on the electoral consequences of geographical proximity to protests, as well as the theories of temporal proximity, via the availability heuristic (Tversky et al., 1973). As we argue that MTR closures incorporate these three microfoundations, using this proxy enables us to test their aggregated impact. We also wish to contrast the overall election result to the findings of Muñoz and Anduiza (2019) as well as Simpson, Willer and Feinberg (2018), whose research suggests that the Hong Kong 2019 District Council election results are an empirical anomaly and that violent pro-democracy protests should actually reduce support for pro-democracy candidates.

Accordingly, we wish to contribute to the current state of knowledge in three separate ways. Firstly, as the previous literature does not provide any consensus about the impact of violent protests on electoral outcomes, we wish to give further insights by analyzing a new empirical case on the basis of current theories. Secondly, as the main previous studies have primarily focused on countries outside East Asia, we believe that shifting the focus to Hong Kong would broaden the view and contribute with new perspectives on the contexts in which violent protests can be linked to electoral outcomes. Thirdly, we want to contribute to the field by employing a so far unexplored proxy for capturing violent protests: MTR (metro) closures. As protesters today both organize and protest in entirely different ways compared to the Swing riots during the 19th century, which the study of Aidt and Franck (2015) focuses on, we believe that we could contribute to the current research by pursuing an innovative analysis of contemporary protests.

By studying the Hong Kong 2019 case, we wish to test whether geographical proximity to violent pro-democracy protests in the days before election day had an impact on election results. More specifically, this paper intends to answer the following research question:

Does geographical proximity to violent pro-democracy protests in the days before election day affect the electoral success of pro-democracy candidates?

Relying on the conclusions suggested by Gillion and Soule (2018), Aidt and Franck (2015) and Tversky et al. (1973), we hypothesize that people living in constituencies that were strongly affected by violent pro-democracy protests in the seventeen days before election day were more likely to vote for pro-democracy candidates in the 2019 District Council election. This leaves us with the following baseline hypothesis:

The larger the number of violent pro-democracy protests a constituency has experienced in the seventeen days before election day, the greater the proportion of votes for pro-democracy candidates in that constituency.

We test this hypothesis by employing MTR closures as a proxy for violent pro-democracy protests, and use data on MTR closures all across Hong Kong during the seventeen days before election day together with data on the 2019 District Council election results. As such, we have defined the geographical distance to a particular radius, and the temporal aspect to a particular period, aiming to test our hypothesis holding these conditions fixed.

5. Method

This section presents how the data used for this study was collected. In order to answer the research question, the following three data sets were needed:

1) One data set containing the election results, on the level of constituencies, from the District Council election on November 24, 2019.

2) One data set containing the number of closed MTR stations within a specified radius from each constituency during the period from November 8 to November 24, 2019.
3) One data set of demographic, economic and spatial control variables on the level of constituencies or districts, to avoid omitted variable bias.

As the desired data was not available from any external database, we created the data sets manually. Each one in turn, we now present the data collection methods of election results, then MTR closures and, lastly, control variables.

5.1. Data Collection of District Council Election Results

The data about the number of votes received by each candidate as well as the candidate affiliation was provided by the Government of the HKSAR District Council Election Results (2019a, 2019b). The candidate affiliation was stated by the candidate on the Candidate Introduction Form. Although there were several alternative formulations for candidate affiliation stated on the forms, we argue that all of these could be sorted into three distinct categories: "pro-democracy", "pro-Beijing" or "independent", thereby suiting the aim of this study's analysis. We applied the following decision rules when categorizing the candidates for our created data set of election results:

1) If a candidate belonged to a party, their candidate affiliation was determined by Wikipedia's (2019) "List of political parties in Hong Kong" into one of the three categories. Wikipedia's list was used as it provides the most compiled information regarding the political affiliation of Hong Kong's many political parties.
    a) As 6 of the parties were so small that they were only represented by a total of 1-3 candidates each and, accordingly, not included in Wikipedia's list, these parties were categorized using the same rule as for the independent candidates below.
2) If a candidate claimed to be "independent", "N/A", "nil", "not affiliated", "沒有" or "無", this was categorized as being independent. Also, if the candidate did not write anything at all, or the information was too vague, this was categorized as being independent. As such, this category mostly covers candidates who are mainly concerned with local issues such as traffic, local infrastructure and construction, and are not involved in the regional politics of the Hong Kong–mainland China conflict.
    a) If a candidate had not claimed any political affiliation *and* received less than 5 % of the votes in their constituency, this was categorized as independent without making further inquiries.
3) If a candidate explicitly claimed to be "independent pro-democracy 獨立民主派", "democrat", "democracy camp", "pro-democracy camp", "democratic progressive camp" or simply "pro-democracy 民主派" this was categorized as pro-democracy. This category does also cover those who formally claimed to be just "independent" but used symbols of democracy; pro-democracy slogans such as speaking about the "five demands" of the protest movement ("五大訴求 缺一不可"), "come on Hongkongers!"

(香港人, 加油!") or "Free Hong Kong, revolution of our times!" ("光復香港 時代革命!"); stated demands for increased democracy and an investigation of police violence on their introduction sheets.[4]

4) If a candidate claimed to be "independent pro-establishment 建制派", this was categorized as independent pro-Beijing. However, none of these cases were found.

5.2. Data Collection of MTR Stations' Locations and Closures

The data about the location of MTR stations was provided by the Government of the HKSAR Constituency Boundary Maps (2019). By using a radius tool available in the interactive Constituency Boundary Maps, we mapped out the constituencies located within 300, 500 and 1,000 meters radius from each MTR station. An example of MTR stations georeferenced to constituencies is illustrated in Appendix B Figure B1. As such, one MTR station could be matched to several constituencies. The main argument behind this is that people in multiple constituencies may be equally geographically close to one and the same MTR station, even if that station is located only inside the boundaries of one of these constituencies; therefore the violent protests proxied by the MTR closure should be assigned to all constituencies within that radius. Matching MTR stations to constituencies in this way also provides a larger sample than matching stations to constituencies one-to-one, as the total number of stations is limited to 93. However, some stations and lines were excluded from the data sets as these were seen as misleading exceptions, leaving us with a total of 91 MTR stations. The list of excluded MTR stations as well as the final list of MTR stations included in our data sets are presented in Appendix A Table A1-A2.

The data about MTR closures during the selected period was provided by MTR's official Twitter account @mtrupdate (MTR Twitter Update, 2019). Via the Twitter account, MTR releases updates about closed stations, service delays and other traffic information several times per day. By collecting and compiling these updates, we created a data set of the number of closed stations per day between November 8 and November 24, 2019.

A MTR station was marked as closed if this particular station was announced as closed or if a line intersecting with this MTR station was labeled as "suspended", which means that the entire line was closed. The reasoning behind this rule is that both of these events, being consequences of ongoing protests, are assumed to remarkably affect people living in the vicinity. Naturally, stations could have been closed for other reasons such as maintenance or repairs. However, given the turbulent overall situation in Hong Kong within the specified time frame, we perceive it as reasonable to assume that all closures were caused by violent protests.

As the question of interest is to proxy the number of violent protests per day, it was not taken into account for how long each station was closed – also implying that there were no minimum time limit in order to be marked as closed – or if it had been closed several times

[4] However, we struggled with questions about replicability and subjectivity as we classified the independent candidates. Even though we established different criteria for what should separate an "independent" candidate from an "independent pro-democratic" candidate, there will always be a measure of subjective judgment. In the end, though, we decided that this more granular approach was better than simply bundling all independents together, as there obviously existed sub-categories of the independent candidates.

during the same day. Accordingly, each station received for each day either the value of 1 (if closed during some point of time) or 0 (if open all day).

5.2.1. Key Considerations of the MTR Data

Importantly, we do not exclude constituencies without any MTR stations within the chosen radius from the analyses. As violent protesters often targeted MTR stations actively as part of the protests, we believe that violent protests far away from MTR stations only happened on occasion, and that these protests were likely both smaller and non-violent. This implies that constituencies without MTR stations indeed contribute to our analysis, as these were simply not strongly affected by violent protests. Furthermore, excluding constituencies without MTR stations from the data set would imply a further risk of selection bias, as well as reduce our sample. As such, we do not differentiate between constituencies with zero MTR closures because of absence of MTR stations and those with zero MTR closures because of absence of violent protests, and we do not exclude any of these from our data sets.

In much the same way, we make no difference between constituencies located in rural or urban areas of Hong Kong. Since people in rural areas might be more dependent on the MTR system to be able to travel to work, and thus more negatively affected by a MTR closure, one may believe that there is a need to control for the rural or urban location of a constituency. However, as MTR closures should be interpreted as a proxy for capturing where violent protests took place within the specified time frame, rather than a measure of how affected people were by MTR closures, we assume each MTR closure to represent an equally violent protest regardless of where in Hong Kong the MTR station is located; accordingly, we do not construct any such control variables.

Lastly, we have chosen to account for MTR closures in absolute values rather than as a proportion of available stations. Obviously, it is more inconvenient if 80 % of the nearby MTR stations are closed compared to 10 %, implying that it might be a reason to account for closures as a proportion of the total number of available MTR stations in a constituency. However, as our study does not seek to investigate how much one's everyday life was affected by closed MTR stations, but rather aims to estimate the number of violent protests in each constituency during our chosen period, using absolute values is the most appropriate for our study.

5.3. Data Collection of Control Variables

The data about demographic, economic and spatial characteristics of each constituency was provided by the HKSAR Census and Statistics Department (2016) from the 2016 Population By-Census (16BC). We collected data on age, educational attainment, proportion of students and median monthly income on constituency level. We collected data on the proportion of mainlanders on district level. We collected data on the geographical distance to the border to mainland China on regional level.

6. Data

This section presents our constructed variables made from the three data sets. Firstly, we have created two dependent variables that we employ for our different types of regressions; one dependent variable denoting the support for pro-democracy candidates, and another dependent variable coded to describe if the victorious candidate was pro-democracy or not. Secondly, our key independent variables denote the number of local MTR closures across various specifications. Thirdly, we have created control variables of the constituencies' demographic, economic and spatial characteristics. Descriptive statistics of the baseline data sets are presented in Table 2.

6.1. Electoral Success of Pro-democracy Candidates

We construct two types of measures of the electoral success of pro-democracy candidates in the District Council election 2019 in Hong Kong, which are our dependent variables. As the question of interest is to estimate the proportion of votes for pro-democracy candidates in each constituency, we do not differentiate between votes for pro-Beijing or independent candidates, but rather treat these as constituting the "non-pro-democratic choice." In other words, we have simplified the voter choice into a binary one, where people can vote either pro-democratic or not pro-democratic.

Firstly, in our linear model we use the dependent variable *PDSupport,* which is the proportion of votes for pro-democracy candidates in each constituency. Secondly, in our probit model we use the dependent variable *PDWinner,* a dummy variable which takes the value 1 if the victorious candidate in a constituency was pro-democracy, and 0 otherwise. Although the binary variable could have been used also for the linear model, we opt for the continuous as this one will give more variation in data, thereby enabling us to obtain the most precise results possible for the linear regression.

An important feature of these dependent variables is that they entail sorting out, from Hong Kong's 452 constituencies, those constituencies that did not have at least one pro-democracy candidate. This procedure reduces the number of constituencies in our analysis from 452 to 409.[5]

6.2. Local MTR Closures

Following the approach of Aidt and Franck (2015), we use the georeferenced MTR closures to construct two types of measures of violent protests. These are our independent variables of interest, throughout this paper referred to as our "key independent variables." To begin with, the variables *MTR300, MTR500* and *MTR1000* measure protests on the intensive margin. As such,

[5] Among the 43 constituencies without at least one pro-democracy candidate, 31 constituencies have candidates that are both pro-Beijing and independent. Results from these 31 constituencies will be a part of Sensitivity Analysis III. The remaining 12 constituencies have only independent candidates and thus the results cannot be analyzed in a meaningful way.

they measure the number of MTR stations that at some point of time were closed during the seventeen-day period, in such geographical proximity to the constituency that a circle with a 300, 500 or 1,000 m radius and with its center at the MTR station covers a part of the constituency. The size of the radiuses were chosen with 500 m as a point of departure, because the average distance between two MTR stations in Hong Kong is about 1,000 m (Government of the HKSAR Constituency Boundary Maps, 2019), making it reasonable to assume that most people have their closest MTR stations within 500 m from their home, and will hence be affected by violent protests occuring there. We then followed Aidt and Franck (2015) and created both a smaller and larger radius than this. Our aim with this strategy is to later find the most significant radius of these three, and utilize the other ones as robustness checks. As shown in the descriptive statistics presented in Table 2, the number of MTR closures ranges from 0-23 for *MTR300,* 0-30 for *MTR500*, and 0-57 for *MTR1000*.

Secondly, the dummy variable *MTRClosuresTreatment* measures protests on the extensive margin, and is used in Sensitivity Analysis I. It takes the value of 1 if there were strictly more than zero MTR closures and 0 otherwise, employing the 500 m radius. In our econometric models, *MTRClosuresTreatment* is coded as an indicator variable where *MTRClosuresTreatment 0* serves as the baseline.

6.3. Control Variables of the Constituencies

Although we argue that the geographical variation in the occurrence of protests is independent from the constituency's characteristics, there is always a possibility of variables that are correlated both with our dependent and independent variable, thereby leading to omitted variable bias. For this reason, we create a set of independent variables aimed to be used as controls for the constituencies' demographic, economic and spatial characteristics. In this section, we motivate the choice of each control variable on the basis of theoretical grounds, grounds context-specific to Hong Kong as well as our own plausibility assessments.

Demographic controls

To begin with, students and university-educated young people played central roles in the Hong Kong 2019 pro-democracy protests. In fact, field surveys show that the support for the protests varied sharply by age, educational level and class (Sum, 2019). Moreover, many students lived on campus because this is usually less expensive than living off-campus.[6] As such, many students that participated in university protests actually lived on university campuses, implying that they both protested and voted in the same constituency. For this reason, we constructed the variable *Students* as the proportion of university students in each constituency. To account for age, we coded the variables *Age2029* and *Age65* which are the proportion of residents in each constituency aged 20-29 years as well as 65 years or older. To account for educational attainment, we created the variable *Education* which measures the proportion of residents in each constituency who have completed post-secondary education.

[6] Hong Kong, together with Singapore and Paris, is one of the world's cities with the highest costs of accommodation, according to The Economist Intelligence Unit (2019).

Further, one may hypothesize that the proportion of the constituency's population who are originally from mainland China would influence election results. Though unclear in which direction this effect may run – mainlanders could either be supporters of the policies of the Beijing central government, or, having left the mainland for political reasons, firmly oppose them – we believe that controlling for the proportion of mainlanders is necessary. We therefore coded the variable *Mainlanders,* which is the proportion of mainland Chinese residents in each district who have resided in Hong Kong for less than 7 years. This data was only available on the district level; we have made the assumption that district data should be a fairly good proxy for constituency data, thereby providing each constituency of a district with the same district's average proportion of mainland Chinese residents.

Economic controls

Again referring to the above mentioned field surveys, socioeconomic status appears to have influenced people's support for the 2019 protests, and most likely then also their voting decision. This is further strengthened by theories from the field of political economy, suggesting that an individual's political attitudes are often influenced by their socioeconomic status, as noted by for example Brown-Iannuzzi, Lundberg and McKee (2017). In much the same way, Wolfinger and Rosenstone (1980) suggests voter turnout to be correlated with education and income levels, which again suggests that socioeconomic variables should be controlled when analyzing election data. To take into account socioeconomic status, we have thus, in addition to the variable *Education,* created the variable *Income* as the median monthly income from main employment in Hong Kong dollars (HKD) in each constituency.

Spatial controls

It is evident that many of the historically pro-Beijing strongholds have been located close to the border to mainland China. Indeed, the majority of the constituencies that in the District Council election 2019 continued to vote in large numbers for pro-Beijing candidates, were those bordering mainland China (South China Morning Post, 2019c). For this reason, we constructed the ordinal variable *ChinaDist,* ranging from 1 to 3 depending on the distance to the mainland China border and coded according to the constituency's location in one of Hong Kong's three regions. The index is equal to 1 if the constituency is located on Hong Kong Island (the furthest from the border), 2 if located in Kowloon and 3 if located in New Territories (closest to the border). In our econometric models, *ChinaDist* is coded as a series of indicator variables where *ChinaDist 1* serves as the baseline.

6.4. Methodological Limitations

The data set and the collection of it is, however, subject to several limitations that must be taken into account when considering the robustness and wider applicability of our results. To begin with, the 16BC data is unfortunately only released every fifth year, so the latest data available is from 2016. Therefore, the data for the control variables will be from 2016 while the data for the dependent and independent variables will be from 2019. Even though using 2016 data is not optimal, this is the best alternative as such detailed data has not been released for 2019 yet. Also, this is an acceptable proxy assuming that Hong Kong's demography has not changed radically during the short period from 2016-2019.[7]

Another possible shortcoming of our data is the self-selection bias inherent in election data. As the voter turnout rate in the 2019 District Council election was 71.2 %, the election results only capture the opinions of those with at least a minimal political interest, and the results are thus not fully representative of the electorate. It is also somewhat problematic to apply control variables, collected from the entire population, to the smaller sample of voters in the District Council election. Hence, an alternative strategy would have been to base our study on an opinion poll of a randomly selected sample. On the other hand, an opinion poll creates a risk of respondents not answering truthfully, implying that it is not as accurate as the actual election results. Above all, as the 2019 pro-democracy protests of Hong Kong happened so recently – being still ongoing at the time of writing – there were simply no such opinion polls at hand, which made us conclude that the 2019 District Council election data was, after all, the best available proxy for public opinion, as well as an acceptable representation of the total population.

Moreover, our election data set reveals the problem that some constituencies have more candidates to vote for than others, and some constituencies not even having candidates from all three categories to vote for. One reason for this variation might be that candidates are endogenous to constituency preferences. For example, pro-democracy candidates could be absent in a constituency because they know that the chance of victory is slim, implying that the voters in this constituency do not have strong pro-democratic beliefs. This will lead our estimates to be biased and we do not know the direction of the bias as there could also be numerous other reasons for why, for instance, pro-democracy candidates are missing. To take into account the problem of some constituencies not having any pro-democracy candidates to vote for, we pursue a sensitivity analysis where we classify independent candidates, in those constituencies where pro-democracy candidates are absent, as pro-democratic (Sensitivity Analysis III). In doing this, we wish to examine whether the absence of pro-democracy candidates in a constituency has any effect on the support for independent candidates in that constituency.

Regarding control variables, it would have been desirable to control for former political beliefs. That is because constituencies' prior political beliefs might be correlated both with electoral outcomes and the occurrence of protests, thereby posing a risk of bias for our estimates. The direction of the potential bias is unknown, as it could either be the case that protests were more widespread in areas with strong pro-democratic beliefs, or in areas with strong pro-Beijing beliefs. To give an example of political controls, Aidt and Franck (2015)

[7] As there have been changes regarding some of the constituencies' names or geographical boundaries between 2016-2019, we simply corrected for these changes in our data sets. The changes made are shown in Appendix C Table C1-C2. In 2017, the Hong Kong Electoral Affair Commission (EAC) proposed to create 21 new elected seats across 10 District Councils and, accordingly, the total number of elected seats for the elections increased from 431 (2016) to 452 (2019) (Legislative Council of Hong Kong, 2017). As such, new constituencies have been created from one or several existing constituencies in order to give particular districts more elected seats. A list of the newly created constituencies is available in Appendix C Table C3. To adjust for this, we have used district averages of the 16BC data as proxies for the newly created constituencies as we believe these should be fairly appropriate.

wanted to quantify the general support for the Whigs prior to the riots in 1830-31, and therefore they controlled for the "Whig share" in 1826. However, conducting these types of controls for Hong Kong poses empirical difficulties. The available data for former political beliefs on constituency level is the 2015 District Council election results, where nearly all of the constituencies voted pro-Beijing. Although the results appear unambiguous, it is inappropriate to assume that a majority of Hong Kong's population did not support democracy at the time. An explanation for the electoral outcome could rather be that the 2015 elections did not carry the same political undertones as the 2019 one, but rather represented a more typical District Council election focusing on local questions instead of regional politics. Consequently, we could not eliminate the potential bias of former political beliefs by using the 2015 District Council election results as a control variable, because employing that data set would lead to even bigger problems in terms of misleading comparisons. In sum, we choose to pursue our analyses without controlling for former political beliefs as this is our best alternative available. Nonetheless, we are aware that this could be a source of potential bias in our empirical strategy.

In addition to political beliefs, there is always a possibility of missing important explanatory variables and thereby obtaining invalid results. However, we are confident that we have included an extensive set of control variables that account for many relevant biases.

6.5. Descriptive Statistics

We report descriptive statistics for all variables of our baseline data sets in Table 2. The data consists of percentage, binary, numerical and ordinal variables. For variable list and descriptions, we refer to Appendix D Table D1. As mentioned in Section 6.1, constituencies that did not have at least one pro-democracy candidate were sorted out, thereby reducing our sample size of the baseline data set from 452 observations to 409. While this sample represents 90.45 % of the real number of constituencies, we believe that the bias of reducing the sample is negligible. However, dropping 43 observations poses a risk of selection bias in our analyses, which we must consider when examining the robustness of our results.

Looking at our dependent continuous variable *PDSupport,* the data suggests a mean of 0.56 accompanied with a low standard deviation (0.09), implying that the share of votes for pro-democracy candidates is rather high on average. Regarding our dependent dummy variable *PDWinner,* we find a mean of 0.85 but a relatively higher standard deviation (0.36); altogether, showing that the majority of the constituencies had a pro-democracy winner. Furthermore, our binary key independent variable *MTRClosuresTreatment* displays a mean of 0.56, thereby showing that slightly more than half of the constituencies in our sample experienced strictly more than zero MTR closures during the investigated period. The means of our continuous key independent variables are, however, located in the lower ranges of the intervals; this suggests that although most constituencies experienced MTR closures, the average amount of MTR closures was still low.

Table 2: Descriptive Statistics of the Baseline Data Sets

	Mean	Std. Dev.	Min	Max	Obs.
Dependent variables					
PDSupport	.5614379	.0872317	.0111	.8836	409
PDWinner	.8508557	.3566669	0	1	409
0 Not pro-demodratic					*61*
1 Pro-democratic					*348*
Key independent variables					
MTR300	2.413203	4.044837	0	23	409
MTR500	4.05868	5.332244	0	30	409
MTR1000	8.381418	8.303746	0	57	409
MTRClosuresTreatment	.5672372	.4960653	0	1	409
0 No Closures					*177*
1 Closures					*232*
Control variables					
Age2029	.1276663	.0324437	.0515458	.2690399	409
Age65	.1671123	.0452495	.066702	.3685957	409
Education	.2723872	.1047367	.115864	.552929	409
Mainlanders	.0231673	.0099097	.0105983	.0492386	409
Students	.14676	.0307391	.0557664	.2678379	409
Income	17996.79	7954.445	11020	75000	409
ChinaDist	2.359413	.7672306	1	3	409
1 Hong Kong					*73*
2 Kowloon					*116*
3 New Territories					*220*

Table 2: Descriptive statistics for all dependent and independent variables of the baseline data sets. For the variable *Mainlanders* with district-level variation, we attribute the district average to each constituency within that district. For the newly created constituencies presented in Appendix C Table C3, we attribute the district average. *Sources, dependent variables:* The Government of the HKSAR District Council Election Results (2019a, 2019b). *Sources, key independent variables:* The Government of the HKSAR Constituency Boundary Maps (2019); MTR Twitter Update (2019). *Source, control variables:* The HKSAR Census and Statistics Department (2016).

7. Empirical Framework

Our empirical strategy aims to test whether geographical proximity to pro-democracy protests during the seventeen particularly violent days leading up to the election affected the electoral success of pro-democracy candidates in the District Council Election 2019. First, we conduct a t-test for the difference in means for *PDSupport* between constituencies who have experienced zero MTR closures and constituencies who have experienced strictly more than zero closures during the period. We then continue to test the nature of this relationship by first applying an ordinary least squares (OLS) method of multiple linear regression. Second, we follow the strategy of Aidt and Franck (2015) and add an alternative model by estimating a binary choice model with a probit estimator, where we test whether the victorious candidate in each constituency was pro-democratic or not.

The dependent variables of the tests are either *PDSupport* or *PDWinner* – depending on the aim of the test – and are treated as continuous or binary, respectively. The key independent variables of the tests are either any of the continuous MTR variables or the binary variable *MTRClosuresTreatment,* again depending on the aim of the test. The independent variables *Students, Age2029, Age65, Mainlanders, Income* and *Education* are treated as continuous. The independent ordinal variable *ChinaDist* is treated as a series of indicator variables.

7.1. Test for the Difference in Unconditional Means

To test if there is a relationship between geographical proximity to violent pro-democracy protests and the preference for voting for pro-democracy candidates on constituency level, a test for the difference in unconditional means was performed as a point of departure. Accordingly, we differentiate between constituencies that have experienced MTR closures and those that have not, and test if these groups were associated with statistically different mean outcomes in votes for pro-democracy candidates through an independent sample t-test.

Given the different intensity of our treatment variable, i.e., that constituencies have experienced different numbers of MTR closures, we have in this test chosen to simplify the data by using a binary model, where constituencies could either have been treated or not. This binary specification is created by dividing the constituencies into two groups based on a cut-off: those with strictly more than zero MTR closures during the seventeen-day period in such geographical proximity to the constituency that a circle with a 500 meter radius around the station covers a part of the constituency (treatment group) and those with no MTR closures within that radius (control group). The key independent variable *MTRClosuresTreatment* is coded as an indicator variable, taking the value 1 if there were more than zero MTR closures, and 0 otherwise. Because our dependent variable *PDSupport* denotes the proportion of votes for pro-democracy candidates*,* constituencies without at least one pro-democracy candidate are excluded from the analysis, making the number of observations n = 409. As there were 177 constituencies that experienced zero MTR closures during the period, and 232 that experienced strictly more than zero, this cut-off implies that we have 43.3 % of the constituencies below the cut-off, and 56.7 % above.

In order to test that two groups are not significantly different, we produced a balance table, available in Appendix E Table E3, showing the means and variances within the treatment and control groups, respectively, on all control variables as well as t-statistics on the hypothesis that the means of the two groups are the same. The table reveals that the treatment and control groups are fairly equal in terms of the distribution of the control variables, with significant differences in means of the control variables present in only three out of seven variables (*Mainlanders*, *Age2029* and *ChinaDist*). Consequently, this indicates that the two groups are on average equal which serves the aim of our empirical strategy, but also that there is indeed a need to control for these variables in our regressions.

Our test for difference in unconditional means was performed by employing the following hypothesis testing for *PDSupport*.

$$H_0 : \mu_T - \mu_C = 0 \qquad \text{(Hypothesis 1)}$$
$$H_1 : \mu_T - \mu_C \neq 0$$

In other words, we test if geographical proximity to violent pro-democracy protests in the final days before election day has any effect on the preference for voting for pro-democracy candidates on a constituency level. If we can reject the null hypothesis, there are grounds for believing that there is a significant difference in the preference for voting for pro-democracy candidates between constituencies being geographically proximate to violent pro-democracy protests, and those that were not.

7.2. OLS Method of Multiple Linear Regression

In order to advance our understanding of the variables' relationship, we test if the preference for voting for pro-democracy candidates is proportional to the number of local violent pro-democracy protests on constituency level, by executing an ordinary least squares (OLS) method of multiple linear regression. Apart from the argument of extending our analysis, the linear regression is needed because of the statistically significant differences in means in three out of seven possible control variables, as presented in Appendix E Table E3, which makes omitted variable bias a concern. In this baseline linear regression model, we employ the key independent variable *MTR500* and, again, the dependent variable *PDSupport*.

We estimate the following multiple linear regression model

$$(PD\ Support)_i = \beta_0 + \beta_1(MTR\ 500\ _i) + X_i\beta_2 + \varepsilon_i \qquad (1)$$

where the dependent variable is the proportion of votes for pro-democracy candidates in constituency i in the District Council election 2019, *MTR500* is the baseline measure of local MTR closures, X is a vector of constituency level observables, and ε is an error term. By using the vector of observables, we initially include controls for age, education, mainlanders living in Hong Kong, students, income and geographical distance to China.

We follow a stepwise forward selection method of regressors, originating from an algorithm first proposed by Efroymson (1960). Even though all regressors, including the control variables, are theoretically of interest, we wish to limit the risk of multicollinearity and also take the relatively low degrees of freedom into account. Furthermore, the fact that both the research question and the data sets have been generated by us, there is no strong theoretical background for exactly which control variables should be included. Therefore we employ a selection method to confirm that neither too many nor too few regressors are included in the final model. In the forward stepwise regression method, each regressor is evaluated according to the absolute t-value of its beta coefficient. The overall significance level of the selection is 15 %. We begin with a null model containing no regressors, and successively add regressors according to the mentioned criterion. As such, we will begin by running multiple simple linear regressions with *PDSupport* as the dependent variable, in order to find the regressor with the highest absolute t-value. Once this is added, we run the regression again, but this time with two regressors. A detailed description of the procedure is included in Appendix E Table E1. In the end, the following regressors are included in the final OLS model:

$$PDSupport_i = \beta_0 + \beta_1 MTR500_i + \beta_2 ChinaDist_i + \beta_3 Mainlanders_i + \beta_4 Income_i + \beta_5 Education_i + \varepsilon_i \quad (2)$$

As demonstrated in our forward stepwise regression, *MTR500* turns out to be the most significant out of our three continuous MTR variables. Consequently, *MTR500* will be employed in all baseline models and the other radiuses (300 m and 1,000 m) will be investigated in a later sensitivity analysis as robustness checks.

It is important to note that the forward stepwise regression revealed that none of the variables *Students, Age2029* and *Age65* were statistically significant at a 15 % level. This further strengthens our assumption that the geographical occurrence of protests were not related to the characteristics of the constituencies, and we can now leave out the possibility that the proportion of students or young people simultaneously affects the likelihood of pro-democracy protests and pro-democratic votes in a particular constituency. Consequently, this reinforces our argument that there is an observable link between experiencing pro-democracy protests and voting for pro-democracy candidates.

Further, we test for heteroscedasticity (Table 3) in our linear regression model using the Breusch-Pagan/Cook-Weisberg test, developed by Breusch and Pagan (1979) and Cook and Weisberg (1982). By evaluating the null hypothesis of constant variance of the error term against the alternative hypothesis of heteroscedasticity, the test reveals that there is heteroscedasticity of the error term for all continuous key independent variables employed in the linear regression model. We also perform the test for the binary key independent variable *MTRClosuresTreatment,* used for Sensitivity Analysis I, and find the same p-value as for the continuous ones.

Table 3: Breusch-Pagan / Cook-Weisberg Test for Heteroscedasticity

	MTR300	**MTR500**	**MTR1000**
chi 2(1)	20.20	30.07	20.74
Prob > chi	0.0000	0.0000	0.0000

Table 3: Breusch-Pagan / Cook-Weisberg test for heteroscedasticity of the error term reported for the key independent variables *MTR300, MTR500* and *MTR1000.* H_0 : Constant variance. Variables: fitted values of PDSupport.

Because we can reject the null of constant variance of the error term at satisfactory levels of significance, we have evidence of heteroscedasticity and thus a violation of the assumption that our data is identically and independently distributed. Since we are looking at geographical data, it is reasonable to assume that the heteroscedasticity is a result of spatial autocorrelation. Theoretically, this could be resolved by clustering the standard errors on district level; however, with only 18 districts, i.e., 18 clusters, we run the risk of small sample bias of the standard errors. For that reason, White heteroscedasticity-robust standard errors will be used in the analysis.

In addition, we will test for multicollinearity among all variables by employing a Variance Inflation Factor (VIF). A common rule of thumb is that if the VIF-value is 10 or higher, multicollinearity may be a concern (Williams, 2015). As the mean VIF-value of our variables is 2.03 and the highest VIF-value is 3.58, as shown in Appendix E Table E2, we see no sign of multicollinearity in our data.

Finally, our primary aim with our linear regression model is to test whether the estimated coefficient for our explanatory variable *MTR500* differs significantly from 0. We evaluate our hypothesis using the following test

$$H_0 : \beta_1 = 0 \qquad \text{(Hypothesis 2)}$$
$$H_1 : \beta_1 \neq 0$$

Once more, this hypothesis is tested by computing the coefficient's t-statistic. By rejecting the null hypothesis, we would be able to conclude that, on constituency level, geographical proximity to violent pro-democracy protests in the final days before election day has explanatory power in the success of pro-democracy candidates.

7.3. Probit Model

Lastly, to complement our previous models we move on to test if the number of local violent pro-democracy protests in a particular constituency affects the probability of a victorious pro-democracy candidate in that constituency, by using a probit model with a binary dependent variable. As such, the values of interest are probabilities and the marginal changes in these, given changes in our key independent variable *MTR500*. In this discrete choice model, the dependent variable *PDWinner* takes on the value of 1 if a seat in a constituency was won by a pro-democracy candidate, and 0 otherwise. Accordingly, the probit model is used to estimate the probability that an observation belongs to one of two categories; in this case, the probability that the winning candidate in a constituency was pro-democracy. One important feature of the probit model is that it allows for different rates of change at high and low values of the independent variable, unlike the linear model that only describes linear change. Another argument for choosing the probit model for this test is that when true probabilities are extreme, the linear model may lead to predictions outside of the relevant range, i.e., predictions that are greater than 1 or less than 0.

The probit model takes the following form

$$P(Y = 1 | x) = \Phi(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) \tag{3}$$

where $\Phi(z)$ is the cumulative distribution function (CDF) of the standard normal distribution yielding values strictly between 0 and 1.

The probit model is derived from the following underlying variable model, where y* is an unobserved variable

$$y^* = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon \tag{4}$$

This will cause the observed variable y to take the value 0 or 1 following

$$y = (0 \textit{ if } y^* \leq 0; 1 \textit{ if } y^* > 0) \tag{5}$$

Thereby, the unobserved variable y* determines the value of y from Equation 4.

The error term in a probit model is assumed to be independent from X and follow a standard normal distribution (Wooldridge, 2013).

To estimate our probit model we use the maximum likelihood estimator (MLE) framework, which automatically accounts for heteroscedasticity in the variance of the dependent variable. The MLE will maximize the log-likelihood function, so that it calculates the betas maximizing the product of the log-likelihoods for all observations in our data (Wooldridge, 2013).

7.3.1 Application of the Probit Model

We employ the following probit model, using the binary dependent variable *PDWinner* and the key independent variable *MTR500*:

$$Pr(PDWinner = 1) = \beta_0 + \beta_1(MTR500_i) + X_i\beta_2 + \varepsilon_i \tag{6}$$

where β_0 is a constant term, β_1 and β_2 are parameters estimated using MLE technique, X is a vector of observables and ε is an error term. The vector includes the same control variables as used in the multiple linear regression (selected through the forward stepwise regression method) of Section 7.1, which makes us able to compare the effect of violent pro-democracy protests *ceteris paribus*. As in the multiple linear regression, White heteroscedasticity-robust standard errors will be used.

The main question of interest in the probit model is to test whether our key independent variable *MTR500* has a significant effect on *PDWinner*, as the other independent variables are solely used as control variables. This is evaluated through the following hypothesis test

$$H_0 : \beta_1 = 0 \qquad \text{(Hypothesis 3)}$$
$$H_1 : \beta_1 \neq 0$$

Accordingly, we test the null hypothesis that the coefficient of *MTR500* is equal to 0. If we could reject the null hypothesis, we would be able to conclude that, on constituency level, geographical proximity to violent pro-democracy protests in the final days before election day has explanatory power in predicting the probability of a victorious pro-democracy candidate.

Importantly, the coefficients from the probit regression output cannot be interpreted directly as in a linear probability model. Rather, we estimate the marginal effects on the predicted response probability of the dependent variable given changing values in the corresponding independent variables. As such, the marginal effect of *MTR500* is estimated to find the constant effect violent pro-democracy protests have on the probability of voting for pro-democracy candidates.

8. Empirical Results

This section presents the empirical results obtained from applying the models to our data sets. Following the order of the previous section, we first provide the results from the test for unconditional difference in means, secondly the linear regression and, third, the probit regression. The last part of this section presents the results of our three sensitivity analyses.

8.1. Difference in Unconditional Means

Firstly, the test performed for difference in unconditional means indicates that there indeed is a significant difference in support for pro-democracy candidates between constituencies that have experienced strictly more than zero MTR closures during the investigated period, and those that have not. As presented in Table 4, the treatment group showed a higher mean support for pro-democracy candidates (56.83 %) than the control group (55.24 %), which implies that the direction of the difference is in line with our baseline hypothesis. Since the p-value of the test is 0.0671, Hypothesis 1 can be rejected on a 10 % significance level. Although the difference in means is considered to be relatively small and the significance level is considered to be weak, we believe that this relationship is worth investigating further.

Table 4: Two-sample T-test with Equal Variances

	Mean	Std. Err.	Obs.
Control group	.5523989	.0074104	177
Treatment group	.5683341	.0050532	232
Combined	.5614379	.0043133	409
diff	-.0159352	.0086806	

diff = mean (0) – mean(1)

H_0 : diff = 0

$H_1 : diff \neq 0$

t = -1.8257

Pr(|T| > |t|) = 0.0671

Table 4: Results from two-sample t-test on treatment and control group. Full test results are available in Appendix E Table E4.

8.2. Linear Regression Results

The results from the baseline linear regression are presented in Table 5. To begin with, the model's overall F-test, for which we obtain a p-value of 0.0007, indicates that the linear model is superior to the intercept-only model in explaining the variation in the data. More specifically, the R^2 statistic informs us that the selected regressors can explain 6.31 % of the variation in the dependent variable *PDSupport*.

The linear regression results, summarized in Table 5, show a positive and significant relationship between our key independent variable, *MTR500,* and *PDSupport.* As shown in Table 5, a one-unit change in the values of *MTR500* implies an increase of 0.149 percentage points on the outcome of *PDSupport.* Since the coefficient of a linear regression can be interpreted as their direct effect on the dependent variable, we interpret these results as if a one-unit change in the number of closed MTR stations within the previously defined 500 m radius from the constituency increases the proportion of votes for pro-democracy candidates in that constituency by 0.149 percentage points. As the coefficient's p-value is 0.029, we can reject Hypothesis 2 and conclude that, on a 5 % significance level, our data supports that proximity to violent pro-democracy in the final days before election day protests has explanatory power in predicting the success of pro-democracy candidates, on the level of constituencies.

Moreover, the results show significant relationships, at varying levels of significance, for the independent variables *Education*, *Income* and *ChinaDist 3,* implying that these too have explanatory power in predicting the outcome of our dependent variable; however, it should be noted that these are control variables used to avoid omitted variable bias, and not our variables of interest.

Table 5: Linear Regression Output

	Support for pro-democracy candidates (%)
MTR500	0.00149** (0.000682)
Education	0.186** (0.0813403)
Mainlanders	-0.881 (0.6446766)
Income	-0.00000296*** (0.00000101)
ChinaDist 2	0.0140 (0.0185764)
ChinaDist 3	0.0340** (0.0132446)
Constant	0.556*** (0.0197432)
Goodness of Fit	
Number of observations	409
F(6, 402)	3.97
Prob > F	0.0007
R-squared	0.0631
Root MSE	.08506

Table 5: Regression output, linear baseline model, *PDSupport* as the dependent variable, *MTR500* as the key independent variable. Coefficients of all independent variables, White heteroscedasticity-robust standard errors in parentheses. Full regression output is available in Appendix E Table E5. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

Furthermore, as the independent variables are denoted in different units, the economic significance of each independent variable was computed by taking the coefficient times the standard deviation of each independent variable (Table 6). Economic significance allows us to gauge the relative importance of the independent variables in their effect on the dependent variable. Even though the economic significance of a one-unit change in *Education* or *Income* is comparatively larger, this is to be expected as these socioeconomic variables are traditionally correlated with voting behavior. Consequently, the economic significance of a one-unit change in *MTR500* is small but not negligible, compared to the other coefficients.

Table 6: Economic Significance

	Support for pro-democracy candidates (%)
MTR500	0.0079557
Education	0.0195246
Mainlanders	-0.0087260
Income	-0.0235452
ChinaDist 2	0.0107280
ChinaDist 3	0.0260789

Table 6: Economic significance, linear baseline model. *PDSupport* as the dependent variable, *MTR500* as the key independent variable. Economic significance is calculated as coefficient times standard deviation of each independent variable. The standard deviation of *ChinaDist* was used for both of its indicator variables.

8.2.1. Linear Regression Postestimation

After concluding that a one-unit change in the value of *MTR500* implies an approximate increase of 0.149 percentage points in support for pro-democracy candidates, we conduct three predictions by varying the value of *MTR500*. In prediction (1), *MTR500* takes on its minimum value of 0. In prediction (2), *MTR500* takes on its sample mean of 4.05868. Finally, in prediction (3) it takes on its maximum value of 30. For all predictions, the additional independent variables are held fixed at their sample means. Results are presented in Table 7. We conclude that the difference in predicted support for pro-democracy candidates between those constituencies that have experienced zero MTR closures and the constituency with the maximum number of closures, *ceteris paribus,* is approximately 4.48 percentage points.

Table 7: Adjusted Predictions of Support for Pro-Democracy Candidates

	MTR500 **value**	**Predicted value of *PDSupport***
Prediction 1	0	.5553821
Prediction 2	4.05868	.5614379
Prediction 3	30	.6001436

Table 7: Predicted values of the key independent variable *PDSupport,* varying *MTR500* and keeping all other independent variables fixed at their sample means. Full results are available in Appendix E Table E6.

To put this result in relation to the election data, we construct the continuous variable *ResultsDiff* which measures the difference in percentage points between percent of votes for the winning candidate and the runner-up, in the 409 investigated constituencies. Descriptive statistics of *ResultsDiff* are reported below (Table 8 and 9).

Table 8: Descriptive Statistics of *ResultsDiff*

	Mean	Std. Dev.	Min	Max	Obs.
ResultsDiff	.1627942	.1039941	.0004505	.5625926	409

Table 8: Descriptive statistics of *ResultsDiff*.

Table 9: Five-number Summary of *ResultsDiff*

	Min	1st Quartile	Median	3rd Quartile	Max
ResultsDiff	.0004505	.0779743	.1513779	.2282391	.5625926

Table 9: Five-number summary of *ResultsDiff*.

When studying the distribution of the created variable *ResultsDiff*, it is apparent that while some candidates won big, others succeeded only by a slim margin. Near the top of the distribution, the vote gap between the winning candidate and the runner-up is close to 40 percentage points. Here, any marginal changes in support that can be attributed to *MTR500* (i.e., approximately 0.149 percentage points per additional MTR closure) will not have had an effect on the electoral outcome. However, near the lower end of the distribution the vote gap is much smaller, revealing that marginal changes in support actually have influenced the electoral outcome. In the first quartile, electoral differences between the winning candidate and the runner-up lie in the range of 0.45 to 7.8 percentage points; indicating that marginal differences (in increments of approximately one percentage point) in support for candidates had a direct impact on the electoral outcome. To specify further, a total of 55 constituencies (13.45 % of the sample) display values for *ResultsDiff* that are equal to or less than 4.48 percentage points, which is the maximum predicted marginal effect of *MTR500*. These results suggest that the marginal differences in support for pro-democracy candidates captured by our proxy variable may hypothetically have had an impact on the electoral outcome in 13.45 % of the studied constituencies.

8.3. Probit Regression Results

The results from the baseline probit regression are presented in Table 10. To begin with, the Wald chi-square statistic implies that all independent variables, tested simultaneously, have an effect separated from zero on the dependent variable *PDWinner*, on a 1 % significance level. Although not our main area of interest, this test of overall significance indicates that at least one of the independent variables of our probit model has explanatory power for the outcome of *PDWinner*.

Furthermore, the explanatory value of a probit regression can be estimated by McFadden's pseudo R^2, where a higher pseudo R^2 indicates higher explanatory power of the model (Wooldridge, 2013). In our model, McFadden's pseudo R^2 is 8.46 % which indicates a reasonably good fit of the model, since values between 0.2 and 0.4 are generally considered a very good fit of the model (Louviere, Hensher and Swait, 2000).

Thirdly, the coefficients and Z-values of the probit regression are also presented in Table 10. These results suggest that the key independent variable *MTR500* has a positive and statistically significant effect (p-value of 0.032) on *PDWinner*. Consequently, we can reject the null of Hypothesis 3 on a 5 % significance level. Similarly to the linear regression results, the probit regression results also show that the control variables *Education* and *Income* have an effect on *PDWinner* at the 1 % significance level.

Table 10: Probit Regression Output

Victorious pro-democracy candidate		
	Coefficient	**Z-value**
MTR500	.0453439**	2.15
Education	4.145259***	2.65
Mainlanders	.8827067	0.09
Income	-.0000536***	-2.87
ChinaDist 2	-.4622304	-1.47
ChinaDist 3	.1141359	0.46
Constant	.8313466**	2.50
Goodness of Fit		
Number of obs.	409	
Wald chi2 (6)	25.75	
Prob > chi2	0.0002	
Pseudo R2	0.0846	

Table 10: Regression output, probit baseline model. *PDWinner* as the dependent variable, *MTR500* as the key independent variable. Coefficient and Z-values of all independent variables. Full regression output is available in Appendix E Table E7. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

We then move on to estimate the marginal effects from the probit baseline model, presented in Table 11, in order to find the constant effect of *MTR500* on the predicted probability of *PDWinner*. As such, the marginal effects allow us to estimate the change in probability that the victorious candidate of a constituency was pro-democratic, based on a one-unit change in the number of closed MTR stations within the previously specified 500 m radius. In addition, similarly to the economic significance calculation for the coefficients of the OLS regression, the marginal effect calculation makes the variables comparable although they were initially denoted in different units.

Data indicates that the marginal effect of a one-unit change in *MTR500* implies a 0.97 % increase in the predicted probability of a pro-democracy winner in that constituency. As we obtain a p-value of 0.03, it can be concluded that this effect is statistically significant on a 5 % significance level. Our interpretation of these results is that a one-unit change in the number of violent pro-democracy protests in a constituency's geographical proximity in the final days before election day increases the probability of a victorious pro-democracy candidate in that constituency by approximately one percent.

Table 11: Marginal Effects

Victorious pro-democracy candidate		
	$\delta y/\delta x$	**Z-value**
MTR500	.0096555**	2.17
Education	.882692 ***	2.71
Mainlanders	.1879637	0.09
Income	-.0000114***	-2.94
ChinaDist 2	-.1146518	-1.54
ChinaDist 3	.0212793	0.44

Table 11: Marginal effects and Z-values from the probit baseline model. Full results are available in Appendix E Table E8. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

8.3.1. Probit Regression Postestimation

To demonstrate the predicted effect of our probit estimates, we conduct three predictions. All control variables are set to their sample means. In prediction (4), *MTR500* takes on the value of 0. In prediction (5), *MTR500* takes on its sample mean, which is 4.05868. Finally, in prediction (6) *MTR500* takes on its maximum value of 30. Results are reported in Table 12. We conclude that, with all control variables set to their sample means, the difference in predicted probability of electing a pro-democracy candidate between those constituencies that have experienced zero MTR closures and the constituency with the maximum number of closures, is approximately 16.2 %.

Table 12: Adjusted Predictions of Victorious Pro-Democracy Candidate

	MTR500 **value**	**Predicted value of *PDWinner***
Prediction 4	0	.8278901
Prediction 5	4.05868	.8707402
Prediction 6	30	.9894496

Table 12: Predicted probability of the key independent variable *PDWinner*. Full results are available in Appendix E Table E9.

8.4. Sensitivity Analyses

Finally, we have pursued three sensitivity analyses in order to test the robustness of our linear and probit baseline models. In the first sensitivity analysis, we divide the constituencies into the treatment and control group used in Section 7.1 and pursue our regressions on this binary model. By doing this, we are able to investigate whether the preference for voting for pro-democracy candidates depends on whether one has experienced violent pro-democracy protests or not, instead of our baseline hypothesis that the preference depends on the number of violent pro-democracy protests experienced. In the second sensitivity analysis, we vary the radius of our key independent variable and pursue our regressions using these adjusted variables. More specifically, we utilize two other radiuses (300 and 1,000 m) to test the robustness of our baseline radius of 500 m. The last sensitivity analysis entails classifying independent candidates as pro-democracy, in those constituencies where the only available candidates are either independent or pro-Beijing. This procedure allows us to test whether the pro-democracy support caused by violent pro-democracy protests also manifested itself in constituencies where there were no explicitly pro-democratic candidates to vote for.

8.4.1. Sensitivity Analysis I: Examining A Binary Relationship

Our first sensitivity analysis aims to test the robustness of our study, by investigating whether our baseline hypothesis holds that more MTR closures lead to a larger proportion of votes for pro-democracy candidates, or if the tendency to vote pro-democratic rather is binary, i.e., it depends on whether one has experienced MTR closures or not. In other words, we test if the effect is not linear and proportional to the number of closures, but rather a step function, where people's perceptions change whenever there is at least one closed MTR station. We pursue this test using the treatment and control group defined in Section 7.1, and adopt the dummy variable *MTRClosuresTreatment* as our key independent variable. First, we present the results from the linear regression, and second, from the probit regression.

When pursuing the sensitivity analysis through a linear regression, we obtain the same p-value of 0.0007 as in the baseline linear model. The R^2 statistic is 6.43 %, which indicates that this model can explain the variation in our dependent variable somewhat better than the baseline model displaying an R^2 statistic of 6.31 %; however, this difference is considered to be minor. As presented in Table 13, a one-unit change in the values of the key independent variable *MTRClosuresTreatment* implies an increase of 1.78 percentage points on the outcome of our dependent variable *PDSupport,* compared to the increase of 0.149 percentage points of the baseline linear regression. We consider the approximately ten times larger coefficient of *MTRClosuresTreatment* to be logical, as the key independent variable now only can take on the value of 0 or 1, implying that a one-unit change must be larger than when using the continuous variable *MTR500*; most importantly, the relationship between the independent and dependent variable remains positive. Using the binary variable *MTRClosuresTreatment* seems to have made the effect on the dependent variable less significant, as the coefficient's p-value is now 0.060, which is higher than the earlier p-value of 0.029, suggesting that our hypothesis of the baseline linear regression is the most appropriate for explaining the actual relationship. As such, this indicates that our baseline specification, i.e., the relationship where the support for democratic candidates are affected linearly by the number of MTR closures, seem to fit our data somewhat better than assuming a binary relationship. However, changing our key independent variable seems to have only minor effects on the overall effect, and we thus conclude that the results from our baseline linear regression remain robust.

Table 13: Linear Regression Output, Sensitivity Analysis I

	Support for pro-democracy candidates (%)
MTRClosuresTreatment 1	0.0178* (0.0094563)
Education	0.192** (0.0825032)
Mainlanders	-0.936 (0.6254766)
Income	-0.00000304*** (0.00000104)
ChinaDist 2	0.0106 (0.0195633)
ChinaDist 3	0.0314** (0.0136688)
Constant	.555479*** (0.0195131)
	Goodness of Fit
Number of observations	409
F(6, 402)	3.97
Prob > F	0.0007
R-squared	0.0643
Root MSE	.08501

Table 13: Regression output, linear model for Sensitivity Analysis I. *PDSupport* as the dependent variable, *MTRClosuresTreatment* as the key independent variable. Coefficients of all independent variables, White heteroscedasticity-robust standard errors in parentheses. Full regression output is available in Appendix G Table G1. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

We also examine the binary relationship using our probit model. Firstly, the probit model's overall F-test (Table 14) confirms that at least one of the coefficients in our probit regression is significantly different from zero. Moving on in the analysis, we find that our key independent variable *MTRClosuresTreatment* exhibits a positive and statistically significant effect on the dependent variable, with a p-value of 0.002. We are thus able to reject the null of Hypothesis 3 at 1 % significance levels, implying that the binary MTR variable indeed has a significant effect on support for pro-democracy candidates.

Moreover, the marginal change in probability of electing a pro-democracy candidate, as a constituency moves from experiencing zero MTR closures to experiencing strictly more than zero, is 11.7 % (Table 15). Results are once again significant on the 1 % level. In other words, a constituency that has experienced strictly more than zero MTR closures is 11.7 % more likely to elect a pro-democracy candidate than a constituency which has not experienced any MTR closures. Compared to the baseline probit model developed in Section 8.3., where the marginal effect of one additional MTR closure on the probability of a victorious pro-democracy candidate was estimated at 0.97 %, the binary model exhibits stronger marginal effects. This is to be expected as a one-unit change in the continuous variable *MTR500* would logically elicit a smaller marginal change in the dependent variable, than a one-unit change in the binary variable *MTRClosuresTreatment*. Above all, the effect of MTR closures on pro-democratic support remains significant and positive, indicating that our results are indeed robust. However, the significant marginal effect of the baseline probit model indeed reveals that increasing the number of closed MTR stations seem to add up in terms of the vote outcome; from this reasoning, we draw the conclusion that the continuous key independent variable seems to fit the data better than the binary one.

Table 14: Probit Regression Output, Sensitivity Analysis I

Victorious pro-democracy candidate		
	Coefficient	**Z-value**
MTRClosuresTreatment 1	.5371378***	3.13
Education	4.376511***	2.81
Mainlanders	-1.686126	-0.16
Income	-.0000573***	-3.07
ChinaDist 2	-.5693764*	-1.75
ChinaDist 3	.0255763	0.10
Constant	.8576563**	2.58
Goodness of Fit		
Number of obs.	409	
Wald chi2(6)	28.40	
Prob > chi2	0.0001	
Pseudo R2	0.0927	

Table 14: Regression output, probit model for Sensitivity Analysis I. *PDWinner* as the dependent variable, *MTRClosuresTreatment* as the key independent variable. Coefficients and Z-values of all independent variables. Full regression output is available in Appendix G Table G2. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

Table 15: Marginal Effects, Sensitivity Analysis I

Victorious pro-democracy candidate		
	$\delta y/\delta x$	**Z-value**
MTRClosuresTreatment 1	.1169332***	3.08
Education	.9233769***	2.86
Mainlanders	-.355747	-0.16
Income	-.0000121***	-3.13
ChinaDist 2	-.1360115*	-1.85
ChinaDist3	.004530	0.10

Table 15: Marginal effects and Z-values from the probit model for Sensitivity Analysis I. Full results are available in Appendix G Table G3. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

8.4.2. Sensitivity Analysis II: Varying the Geographical Radius

Our second sensitivity analysis aims to further test the robustness of the observed relationship between our dependent and key independent variables, by varying the geographical radius used in our constructed MTR variable. As such, we are able to investigate whether the positive effect running from MTR closures to pro-democracy support is statistically significant across different radiuses, or if the relationship seems to weaken at a particular point. This sensitivity analysis is executed by employing, in turn, the variables *MTR300* and *MTR1000* as the key independent variable in the linear as well as the probit baseline model.

To begin with, the results obtained from employing the adjusted variables to the linear baseline model are positive but not statistically significant, since the p-values for *MTR300* and *MTR1000* are 0.130 and 0.122, respectively. Full linear regression outputs for Sensitivity Analysis II are available in Appendix G Table G4 and G7. On the other hand, in the probit baseline model, both variables show a positive and statistically significant marginal effect on a 5 % significance level. Full probit regression outputs for Sensitivity Analysis II are available in Appendix G Table G5 and G8. Our interpretation of these results is that our linear baseline model is not sufficiently robust when varying the radius from 500 to 300 or 1000 m, however, our probit baseline model is.

Importantly, we believe that two additional comments should be made when discussing the interpretation of the linear regression results. First, when we go from using *MTR500* to *MTR300,* the number of MTR stations assigned to each constituency naturally decreases as the radius shrinks. As such, the sample size remains the same, but the variation in our key independent variable decreases remarkably as each constituency is now connected to fewer MTR closures. The relatively small variation in our data may make it difficult to establish a significant relationship between the dependent and independent variable, which could explain the lack of statistical significance. Second, changing the key independent variable to *MTR1000* implies that we are doubling our initial radius. As Hong Kong is relatively small in size, setting the radius to

1,000 m makes the geographical boundaries around each MTR station overlap with substantially more constituencies than before. This means that constituencies are now – to a larger extent than before – connected to the same MTR closures as one another, which causes the variation of MTR closures between constituencies to, once again, decrease remarkably. This could be seen as an underlying reason for the statistically insignificant relationship.

Altogether, we believe that this result is of importance as it shows that the observed relationship actually disappears when employing our two selected radiuses. This underlines that the effect from MTR closures is highly dependent on the exact geographical radius; in other words, that geographical proximity plays a crucial role for the relationship to hold. For this reason, we also believe that a relevant area for future research would be to further test the robustness of the baseline model by employing even more radiuses at smaller intervals, hopefully making it possible to note the particular radius at which the statistically significant relationship disappears.

8.4.3. Sensitivity Analysis III: Modifying the Candidate Classification

This sensitivity analysis aims to investigate whether support for independent candidates increased in constituencies where pro-democracy candidates were absent, thereby testing the robustness of our baseline results by modifying the underlying data set. To test this, we run the same linear and probit regressions as in our baseline models. The only difference is that we have now added 31 new observations to our baseline sample, constituting those constituencies in which the only available candidates were pro-Beijing and independents, and classified these independent candidates as pro-democracy. This increases the size of our sample from 409 to 440 observations. Descriptive statistics of this slightly modified data set are reported in Appendix G Table G10. The rationale behind this modification is that if there indeed is an observable link between MTR closures and support for democracy, voters in constituencies without at least one pro-democracy candidate would instead support independents. Importantly, we thus believe that the voting decision in constituencies with available pro-democracy candidates will remain unchanged, and therefore we have not changed the categorization of the candidates belonging to the baseline sample of 409 observations.

The results from Sensitivity Analysis III do not remarkably differ from our baseline results. Linear as well as probit regression outputs and the probit model's marginal effects are included in Appendix G Table G11-G13. While the linear regression's coefficient for *MTR500* (0.00141) is somewhat lower compared to the linear baseline model (0.00149), the relationship between *MTR500* and *PDSupport* is still positive and significant which strengthens the robustness of our baseline results. However, pursuing our linear regression on the modified data set resulted in a larger R^2 statistic (7.83 %) compared to the baseline model (6.31 %), which suggests that the modified data set indeed makes the model slightly better in explaining the variation in the data. For the probit model, McFadden's pseudo R^2 is nearly the same in the model using the modified data set (8.43 %) as in the probit baseline model (8.46 %), indicating that none of the models is remarkably superior in its fit. The small difference in marginal effects between the two models (0.98 % for modified model; 0.97 % for probit baseline model) is also negligible.

Altogether, as the modification made in Sensitivity Analysis III did not give rise to any major changes in the relationship between MTR closures and support for democracy, we have grounds to believe that voters in constituencies without at least one pro-democracy candidate preferred to support the independent candidate rather than the pro-Beijing candidate, if being geographically proximate violent pro-democracy protests. Consequently, this sensitivity analysis strengthens our baseline hypothesis that constituencies in geographical proximity to violent pro-democracy protests in the final days before election day will vote for the candidate that is not pro-Beijing; that is, a pro-democracy candidate if such a candidate is available, and otherwise an independent candidate.

9. Discussion

This section provides a qualitative discussion of this study's results. We begin by discussing the interpretations and implications of our findings, namely the observed relationship between MTR closures and the support for pro-democracy candidates in the Hong Kong 2019 District Council election. Thereafter, we turn to a more general analysis of how our results respond to the underlying theories, which leads up to answering this paper's research question.

9.1. Results Discussion

Our results conclusively indicate that there is an effect from geographical and temporal proximity to protests on the voter decision. More specifically, geographical proximity to violent pro-democracy protests in the days before election day has a statistically significant positive effect on the support for pro-democracy candidates, on constituency level. Results are robust when modifying the categorization rule of what makes a pro-democracy candidate. Furthermore, results are robust across several model specifications (linear, non-linear, and using both a binary and a heterogeneous treatment). Nonetheless, the models that best seem to fit the data is the one utilizing a continuous (*MTR500*), rather than binary (*MTRClosuresTreatment*), proxy for violent pro-democracy protests. This suggests that support for pro-democracy candidates increases in proportion to the absolute number of protests, rather than follows a step function where one protest is enough to garner the public's support. Importantly, results were not robust to our variations in the radius determining the geographical proximity, suggesting that exact geographical proximity indeed plays a crucial role for the observed relationship between protests and electoral outcomes to hold. We also suggest that all results are interpreted with caution as the exclusion of 43 observations could possibly cause selection bias in our data.

Another common feature across model specifications is the relatively small size of the effect from proximity to violent protests on support for pro-democracy candidates. When comparing the size of the impact through economic significance (linear regression model) and marginal effects (probit regression model), MTR closures have a consistently smaller effect on pro-democracy support than the other significant variables measuring education and income levels. Two things about this are worth noting: firstly, that this finding is entirely in line with our expectations, given the theoretical background on socioeconomic variables' impact on voting behavior. Secondly, in light of our assumption of conditional independence between the location of protests and the political attitudes of the residents of that area, this finding should be viewed as the isolated effect of geographical and temporal proximity to violent protests on support for pro-democracy candidates. In other words, we see this effect as a behavioral response emanating solely from the psychological stimulus of experiencing violent protests close to one's home, near election day. Because of the intricate nature of this effect, we do not expect it to be substantial in size.

More specifically, the size of the effect is such that a one-unit increase in the number of MTR closures implies an increase of 0.149 percentage points on support for pro-democratic candidates, according to our linear baseline model. Predictions from this model along with an analysis of the distribution of the difference in votes for the winning candidate and the runner-up indicate that the isolated effect from MTR closures could hypothetically have impacted the electoral outcome in 13.45 % of the studied constituencies. Similar conclusions can be drawn from the probit baseline model, in which the marginal effect of a one-unit increase in the number of MTR closures implies that the probability of a victorious pro-democracy candidate increased by approximately one percent. Predictions from the probit baseline model also confirm that the probability of electing a pro-democracy candidate clearly increases when going from zero to the maximum amount of MTR closures in the vicinity, in this case by 16.2 %. To sum up, we have evidence of a small but statistically and economically significant effect on pro-democracy support, stemming solely from geographical and temporal proximity to violent pro-democracy protests.

9.2. General Discussion

This paper's research question can now be answered in the affirmative. More specifically, we have shown that for our data, proximity (geographical and temporal) to violent pro-democracy protests translates into a small but statistically and economically significant increase in support for pro-democracy candidates. Turning back to our theoretical framework, our results are in line with the voter effect hypothesis formulated by Gillion and Soule (2018). We see this result as resting on three theoretically sound microfoundations: first, that violence functions as an informative cue and a signal to the electorate (Gillion and Soule, 2018; Desai, Olofsgård and Yousef, 2020), thereby amplifying the salient effect of protests and influencing the voter decision. Second, we turn to scholars such as Aidt and Franck (2015) and Enos, Kaufman and Sands (2019) whose work on geographical proximity to violent protests helps us to view the observed patterns of pro-democratic support as a function of geographical distance to protests. The final microfoundation can be traced to Tversky's et al. (1973) theories regarding the availability heuristic and how recent and emotionally salient experiences will have a systematic impact on decisions. In sum, our results indicate that the aggregated impact of these three microfoundations, i.e., the informative cue of violence, and geographical as well as temporal proximity, can explain voting in the case of Hong Kong 2019.

Opposingly, our results are contrary to empirical findings such as those from Simpson, Willer and Feinberg (2018) as well as Muñoz and Anduiza (2019), who both claim that violent protests typically decrease the public's support for the protesting groups. This gap, we claim, should not be seen as an empirical anomaly but rather highlight the importance of context in establishing external validity. When it comes to drawing conclusions about how our findings may contribute to the current state of knowledge about protests in relation to electoral outcomes, Hong Kong's complex political and economic history must be kept in mind. We are humble to the fact that our results are contingent on many factors specific to the Hong Kong 2019 context, and do not claim our results to necessarily be applicable beyond that. As scholars before us have pointed to, violent protests must be interpreted contextually, with special attention paid to prior beliefs and opinions in the electorate. As such, though our data indicates that proximity to violent pro-democracy protests could have been a determinant of the 2019 electoral outcomes in a portion of Hong Kong's constituencies, we recognize that many other factors were at play. For example, the Umbrella movement of 2014 might have laid the basis for the dramatically increased support for democracy, and the widespread dissatisfaction about the growing influence from Beijing on Hong Kong's politics and economy might have contributed to the decreased support for the pro-Beijing camp. Furthermore, the heated political climate of the summer and fall of 2019 had likely created a whole new baseline for pro-democratic support, on which geographical proximity to protests likely had marginal effects. Similar to conclusions drawn by Aidt and Franck (2015), who claim that the Great Reform Act of 1832 would likely have passed even without the Swing riots, but at a different point in time, it is reasonable to assume that demands for democracy had presumably been building up in Hong Kong, onto which the extradition bill and subsequent protests were catalysts for a major shift in electoral support. A tentative conclusion that is worthy of further investigation is thus that violence in protests only translates into increased political support as long as it is preceded by a shift in baseline support.

10. Concluding Remarks

In this paper, we have investigated whether geographical proximity to violent pro-democracy protests, in temporal proximity to election day, had an impact on the success of pro-democracy candidates in Hong Kong's 2019 District Council election. We have created the data sets used for the analysis manually, and these consist of election results on constituency level, protests data proxied by Hong Kong's MTR (metro) closures as well as demographic, economic and spatial characteristics of the analyzed constituencies. The study was done by performing both an OLS method of multiple linear regression and a probit regression on the collected data sets, consisting of a sample of 409 of Hong Kong's 452 constituencies, thereby aiming to answer the following research question:

Does geographical proximity to violent pro-democracy protests in the days before election day affect the electoral success of pro-democracy candidates?

The results suggest that geographical proximity to violent protests, in temporal proximity to the election day, indeed has a small, but statistically significant, effect on the voter decision in the form of increased support for pro-democracy candidates. Nonetheless, this effect must be viewed in relation to what we see as a dramatic shift in baseline pro-democracy support in Hong Kong, caused by months of political turbulence.

Our study aims to contribute to the current state of knowledge in three separate ways: provide further insights to the rather ambiguous research field of protests and elections, broaden the view by studying a country in East Asia, and suggest using innovative proxies for measuring protests in today's world. All in all, we propose that protests' effect on elections must be studied contextually, with particular attention paid to the context-specific informative cues that increase their salience and act as carriers of the public's support.

Finally, we suggest three main areas for future research. The first one consists of methodological modifications such as finding and utilizing reasonable political control variables, or changing the pro-democracy candidate classification to include all candidates not explicitly pro-Beijing, as done by South China Morning Post (2019c). Controlling for former political beliefs would also allow for making use of panel data, thus making other econometric approaches possible. Another methodological opportunity would be to investigate the underlying reasons for the variation of available candidates between constituencies to see if this affects electoral results. For example, investigations into variation of available candidates could be used to reject or confirm Gillion and Soule's (2018) vulnerability hypothesis, brought up in Section 2.1.1. Our second main suggestion for future research is to evaluate the external validity of our results by running similar tests on larger samples from other contexts. Finally, we encourage longitudinal studies that are able to separate our three microfoundations and isolate their respective effects, in order to determine how they relate to one another. Longitudinal studies could also aim to specify the general preconditions for protests to translate into electoral support, again utilizing panel data to account for political and socioeconomic shifts over time.

Altogether, we hope that this paper can inspire more research into how supposedly rational behavior such as political voting may be influenced by contextual factors, heuristics and psychological stimuli – a rich and ample field of study that, in a fast-paced world, is perhaps more relevant than ever.

References

Aidt, T. S., and Franck, R., 2015. Democratization Under the Threat of Revolution: Evidence from the Great Reform Act of 1832. *Econometrica,* 83(2), pp. 505-547.

Banjo, S., 2019. Hong Kong Protests Drive Surge in Telegram Chat App. *Bloomberg,* [online] August 15. Available at: <https://www.bloomberg.com/news/articles/2019-08-15/hong-kong-protests-drive-surge-in-popular-telegram-chat-app> [Accessed March 24, 2020]

Brown-Iannuzzi, J. L., Lundberg, K. B. and McKee, S., 2017. The Politics of Socioeconomic Status: How Socioeconomic Status May Influence Political Attitudes and Engagement. *Current Opinion in Psychology,* 18, pp. 11-14.

BBC, 2019. The Hong Kong Protests Explained in 100 and 500 Words. *BBC,* [online] November 28. Available at: <https://www.bbc.com/news/world-asia-china-49317695> [Accessed March 23, 2020]

Breusch, T.S. and Pagan, A. R., 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica,* 47(5), pp. 1287-1294.

Cook, D. R. and Weisberg, S., 1982. *Diagnostics for Heteroscedasticity in Regression*. St. Paul: University of Minnesota, School of Statistics.

Desai, R. M., Olofsgård, A. and Yousef, T. M., 2020. Anti-Regime Politics and the Middle Class: Evidence from the Arab World. *SITE working paper.*

District Council Election, 2019. *District Councils Brief.* [online] Available at: <https://www.elections.gov.hk/dc2019/eng/dcbriefs.html> [Accessed March 23, 2020]

The Economist Intelligence Unit, 2019. *Worldwide Cost of Living 2019: Which Global Cities Have the Highest Cost of Living?* [pdf] London: The Economist Intelligence Unit. Available at: <https://www.df.cl/noticias/site/artic/20190319/asocfile/20190319171329/reporte.pdf> [Accessed April 11, 2020]

Efroymson, M. A., 1960. Multiple Regression Analysis. In Ralston, A., and Wilf, H.S., eds., 1960. *Mathematical Methods for Digital Computers.* New York: Wiley. pp. 91-103

Enos, R., Kaufman, A. R., and Sands, M. L., 2019. Can Violent Protests Change Local Policy Support? Evidence from the Aftermath of the 1992 Los Angeles Riot. *American Political Science Review,* 113(4), pp. 1012-1028.

Gillion, D., and Soule, S., 2018. The Impact of Protest on Elections in the United States. *Social Science Quarterly,* 99(5), pp. 1649-1664.

Government of the HKSAR (Hong Kong Special Administrative Region) Constituency Boundary Maps, 2019. *2019 District Council Ordinary Election Constituency Boundaries* [online] Available at: <https://www.eac.hk/en/distco/2019dc_elect_map.htm?fbclid=IwAR2JLNr1wPoZB7x66fTeESED6-KTNg2v6XVCoG-R1mwNdx0lLhBrwMH4gm4> [Accessed February 21, 2020]

Government of the HKSAR (Hong Kong Special Administrative Region) District Council Election Results, 2019a. *Election Results at Districts* [online] Available at: <https://www.elections.gov.hk/dc2019/eng/results.html> [Accessed March 1, 2020]

Government of the HKSAR (Hong Kong Special Administrative Region) District Council Election Results, 2019b. *Introduction to Candidates* [online] Available at: <https://www.elections.gov.hk/dc2019/eng/intro_to_can.html> [Accessed March 18, 2020]

HKSAR (Hong Kong Special Administrative Region) Census and Statistics Department, 2016. *2016 Population By-Census (16BC)* [online] Available at: <https://www.bycensus2016.gov.hk/en/bc-index.html> [Accessed March 15, 2020]

Kaiman, J., 2014. Hong Kong's Umbrella Revolution – the Guardian Briefing. *The Guardian* [online] September 30. Available at: <https://www.theguardian.com/world/2014/sep/30/-sp-hong-kong-umbrella-revolution-pro-democracy-protests> [Accessed April 18, 2020]

Krupnikov, Y., 2014. How Negativity Can Increase and Decrease Voter Turnout: The Effect of Timing. *Political Communication,* 31(3), pp. 446-466.

Lahiri, T. and Hui, M., 2019. Hong Kong Will Have Its Only Truly Democratic Election This Weekend. *Quartz* [online] November 22. Available at: <https://qz.com/1753318/what-you-need-to-know-about-hong-kongs-district-council-elections/> [Accessed May 11, 2020]

Lam, J., Ng, N., and Xinqi, S., 2019. Be Water, My Friend: Hong Kong Protesters Take Bruce Lee's Wise Saying to Heart and Go with the Flow. *South China Morning Post,* [online] June

22. Available at:
<https://www.scmp.com/news/hong-kong/politics/article/3015627/be-water-my-friend-protesters-take-bruce-lees-wise-saying> [Accessed April 16, 2020]

Lau, C., Low, Z., and Cheung, T., 2019. Hong Kong Protests: Lunchtime Demonstrations Extend across City with Hundreds of People Blocking Roads. *South China Morning Post,* [online] November 15. Available at:
<https://www.scmp.com/news/hong-kong/society/article/3037877/hong-kong-protests-lunchtime-demonstrations-extend-across> [Accessed April 17, 2020]

Legislative Council of Hong Kong, 2017. Panel on Constitutional Affairs: Review of the Number of Elected Seats for the Sixth-Term District Councils, (LC Paper No. CB(2)1848/16-17(01)). Hong Kong: Legislative Council.

Leung, H., 2019. Hong Kong Student Who Fell From Height During a Protest Dies. *TIME,* [online] November 8. Available at:
<https://time.com/5721979/hong-kong-student-brain-injury-death/> [Accessed March 23, 2020]

Louviere, J. J., Hensher, D. A. and Swait, J. D., 2000. *Stated Choice Methods: Analysis and Applications.* Cambridge: Cambridge University Press.

Low, Z., 2019. MTR Corporation Rumbles Back to Life amid Extensive Repairs – But Offers No Date for Full Restoration of Hong Kong Rail Services. *South China Morning Post,* [online] October 11. Available at:
<https://www.scmp.com/news/hong-kong/transport/article/3032609/mtr-corporation-rumbles-back-life-amid-extensive-repairs> [Accessed March 24, 2020]

Lum, A., 2019. Hong Kong Student Chow Tsz-lok Dies After Suffering Severe Brain Injury in Car Park Fall. *South China Morning Post,* [online] November 8. Available at:
<https://www.scmp.com/news/hong-kong/politics/article/3036833/hong-kong-student-who-suffered-severe-brain-injury-after> [Accessed March 22, 2020]

Magramo, K., 2019. Why Hong Kong District Council Elections Are Important and How Ongoing Protests May Affect Vote. *South China Morning Post* [online] November 23. Available at:
<https://www.scmp.com/news/hong-kong/politics/article/3039015/why-hong-kong-district-council-elections-are-important-and> [Accessed March 23, 2020]

Moovit, 2017. *Global Public Transit Index: Facts and Usage Statistics about Public Transit in Hong Kong, China* [online]. Available at:

<https://moovitapp.com/insights/en/Moovit_Insights_Public_Transit_Index_China_Hong_Kong-2741> [Accessed April 4, 2020]

MTR Twitter Update, 2019. *MTR Update* (@mtrupdate) [Twitter]. Tweets between November 8, 2019 – November 24, 2019. Available at: <https://twitter.com/mtrupdate> [Accessed February 23, 2020]

MTR System Map, 2020. *MTR Corporation* [online] Available at: <http://www.mtr.com.hk/archive/en/services/routemap.pdf> [Accessed February 23, 2020]

Muñoz, J., and Anduiza, E., 2019. 'If a Fight Starts, Watch the Crowd': The Effect of Violence on Popular Support for Social Movements. *Journal of Peace Research*, 56(4), pp. 485-498.

Reuters, 2019. Timeline: Key Dates in Hong Kong's Anti-government Protests. *Reuters,* [online] November 11. Available at: <https://www.reuters.com/article/us-hongkong-protests-timeline/timeline-key-dates-in-hong-kongs-anti-government-protests-idUSKBN1XL0N3> [Accessed March 23, 2020]

Robles, P., 2019. Key Events from Hong Kong's Anti-government Protests. *South China Morning Post,* [online] December 9. Available at: <https://multimedia.scmp.com/infographics/news/hong-kong/article/3032146/hong-kong-protests/index.html?src=sheld> [Accessed March 23, 2020]

RTHK, 2019. Lunch Time Protesters Appear at Kowloon Bay. *Radio Television Hong Kong*, [online] November 26. Available at: <https://news.rthk.hk/rthk/en/component/k2/1494549-20191126> [Accessed April 17, 2020]

Satoshi, Y., 2019. "Be Water": Hong Kong Protesters Learn from Bruce Lee. *NHK World–Japan*, [online] November 13. Available at: <https://www3.nhk.or.jp/nhkworld/en/news/backstories/745/> [Accessed April 16, 2020]

Schectman, J., 2019. Exclusive: Messaging App Telegram Moves to Protect Identity of Hong Kong Protesters. *Reuters,* [online] August 31. Available at: <https://www.reuters.com/article/us-hongkong-telegram-exclusive/exclusive-messaging-app-telegram-moves-to-protect-identity-of-hong-kong-protesters-idUSKCN1VK2NI> [Accessed April 17, 2020]

South China Morning Post, 2019a. As it Happened: 16 Hours of Violence in Hong Kong Continue with Mong Kok Unrest, While Carrie Lam Says 'Wishful Thinking' to Believe

Government Will Yield to Political Demands. *South China Morning Post,* [online] November 11. Available at: <https://www.scmp.com/news/hong-kong/politics/article/3037135/shots-fired-hong-kong-anti-government-protesters-cause> [Accessed March 23, 2020]

South China Morning Post, 2019b. Hong Kong's District Councillors Traditionally Handle Municipal Matters. But This Election Was a Global Event as Almost 3 Million Residents Clamoured to Be Heard. *South China Morning Post,* [online] November 25. Available at: <https://www.scmp.com/news/hong-kong/politics/article/3039142/hong-kongs-district-council-elections-traditionally-handle> [Accessed March 23, 2020]

South China Morning Post, 2019c. Hong Kong District Council Elections: Winners and Losers. *South China Morning Post,* [online] November 27. Available at: <https://multimedia.scmp.com/infographics/news/hong-kong/article/3039600/hong-kong-district-council-elections-2019/index.html?src=social> [Accessed March 23, 2020]

Simpson, B., Willer, R., and Feinberg, M., 2018. Does Violent Protest Backfire? Testing a Theory of Public Reactions to Activist Violence. *Socius: Sociological Research for a Dynamic World*, 4, pp. 1-14.

Siu, P., Lum, A., and Low, Z., 2019. Chaos on Hong Kong's MTR Network as Police Chase Protesters into Station and Arrest 63. *South China Morning Post,* [online] September 1. Available at: <https://www.scmp.com/news/hong-kong/law-and-crime/article/3025241/chaos-hong-kongs-mtr-network-police-chase-protesters> [Accessed March 24, 2020]

Sum, L., 2019. Young, Educated and Middle Class: First Field Study of Hong Kong Protesters Reveals Demographic Trends. *South China Morning Post,* [online] August 12. Available at: <https://www.scmp.com/news/hong-kong/politics/article/3022345/young-educated-and-middle-class-first-field-study-hong-kong> [Accessed April 17, 2020]

Tversky, A., et al., 1973. *Judgment under Uncertainty: Heuristics and Biases*. Springfield: U.S. Department of Commerce.

Wikipedia, 2019. *List of Political Parties in Hong Kong* [online] Available at: <https://en.wikipedia.org/wiki/List_of_political_parties_in_Hong_Kong> Accessed March 15, 2020.

Williams, R., 2015. *Multicollinearity.* [pdf]. University of Notre Dame. Available at: <https://www3.nd.edu/~rwilliam/stats2/l11.pdf> [Accessed April 18, 2020]

Wolfinger, R. E. and Rosenstone, S. J., 1980. *Who votes?* New Haven: Yale University Press.

Wooldridge, J. M., 2013. *Introductory Econometrics: A Modern Approach.* 5th ed. Mason: South-Western Cengage Learning.

Wong, T. K., 2019. Hong Kong Protests: What Are the 'Five Demands'? What Do Protesters Want? *South China Morning Post,* [online] August 20. Available at: <https://www.scmp.com/yp/discover/news/hong-kong/article/3065950/hong-kong-protests-what-are-five-demands-what-do> [Accessed May 12, 2020]

Yau, C., 2019. Are the MTR's Early Closures Necessary for Repairs or a 'De Facto Curfew'? Either Way, Hong Kong Commuters Are in for More Delays. *South China Morning Post,* [online] October 19. Available at: <https://www.scmp.com/news/hong-kong/politics/article/3033692/are-mtrs-early-closures-necessary-repairs-or-are-they-de> [Accessed March 24, 2020]

Yeung, J., 2019. From an Extradition Bill to a Political Crisis: A Guide to the Hong Kong Protests. *CNN,* [online] December 20. Available at: <https://edition.cnn.com/2019/11/15/asia/hong-kong-protests-explainer-intl-hnk-scli/index.html> [Accessed March 23, 2020]

Appendices

Appendix A: MTR Data

Table A1: Excluded Stations and Lines from the Data Sets

Excluded station or line	Reason for exclusion
All buses and light rail lines	These operate very locally and often in rural areas. They were also not targeted by protesters in the same way as regular MTR stations.
“Airport Express” line	Not an ordinary MTR line that people use in their daily life. (However, “Tsing Yi” and “Kowloon”, stations along the Airport Express line, intersect with other lines as well, so these were not excluded).
“AsiaWorld-Expo” station	Train station on the Airport Express line which could not be reached by other MTR lines.
“Airport” station	Train station on the Airport Express line which could not be reached by other MTR lines.
“Hong Kong West Kowloon” station	Train station on the HSR (high-speed railway) line, going directly from Hong Kong to Shenzhen, China.
“Hin King” station	A MTR station which opened on February 14th, 2020, so it did not exist during the investigated period.

Table A1: List of excluded stations and lines from the data sets, and reasons for exclusion. *Source:* MTR System Map (2020).

Table A2: List of MTR Stations Included in the Data Sets

Name of MTR stations			
Admiralty	Kam Sheung Road	North Point	Tai Wai
Austin	Kennedy Town	Ocean Park	Tai Wo
Causeway Bay	Kowloon	Olympic	Tai Wo Hau
Central	Kowloon Bay	Po Lam	Tin Hau
Chai Wan	Kowloon Tong	Prince Edward	Tin Shui Wai
Che Kung Temple	Kwai Fong	Quarry Bay	Tiu Keng Leng
Cheung Sha Wan	Kwai Hing	Racecourse	Tseung Kwan O
Choi Hung	Kwun Tong	Sai Wan Ho	Tsim Sha Tsui
City One	Lai Chi Kok	Sai Ying Pun	Tsing Yi
Diamond Hill	Lai King	Sha Tin	Tsuen Wan
Disneyland Resort	Lam Tin	Sha Tin Wai	Tsuen Wan West
East Tsim Sha Tsui	Lei Tung	Sham Shui Po	Tuen Mun
Fanling	Lo Wu	Shau Kei Wan	Tung Chung
Fo Tan	LOHAS Park	Shek Kip Mei	University
Fortress Hill	Lok Fu	Shek Mun	Wan Chai
Hang Hau	Lok Ma Chau	Sheung Shui	Whampoa
Heng Fa Chuen	Long Ping	Sheung Wan	Wong Chuk Hang
Heng On	Ma On Shan	Siu Hong	Wong Tai Sin
HKU	Mei Foo	South Horizons	Wu Kai Sha
Ho Man Tin	Mong Kok	Sunny Bay	Yau Ma Tei
Hong Kong	Mong Kok East	Tai Koo	Yau Tong
Hung Hom	Nam Cheong	Tai Po Market	Yuen Long
Jordan	Ngau Tau Kok	Tai Shui Hang	

Table A2: List of MTR stations included in the data sets (total = 91). *Source:* MTR System Map (2020).

Appendix B: Matching MTR Stations to Constituencies

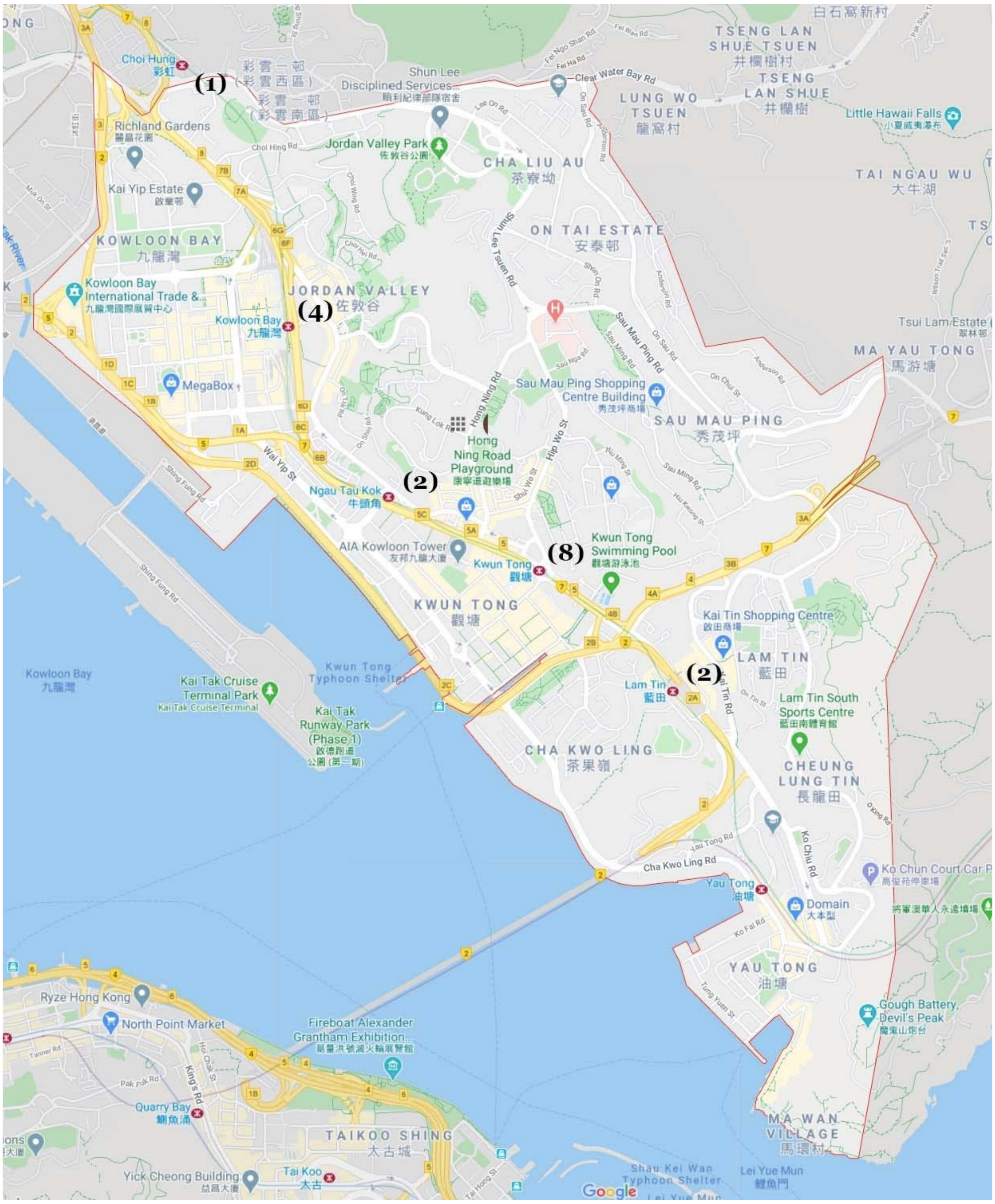

Figure B1: Georeferencing MTR closures in Kwun Tong District. Within the geographical boundaries of Kwun Tong District, the following MTR stations were closed at some point of time during the seventeen-day period (the number indicates how many days it has been closed in total): Choi Hung (1), Kowloon Bay (4), Kwun Tong (8), Lam Tin (2), and Ngau Tak Kok (2). Each MTR symbol represents one of these MTR stations. The MTR station Yau Tong is also located inside Kwun Tong, however, this station was never closed and thereby not marked on this map. The black numbers are an alteration to the original Google Map. *Sources:* Google Maps (2020) and MTR's official Twitter account @mtrupdate (MTR Twitter Update, 2019).

Appendix C: Constituency Changes Between 2016-2019

Table C1: Changes of Constituency Names

Constituency Code	Constituency Name in 2016	Constituency Name in 2019
J06	Sheung Choi	Choi Tak
J24	Chui Cheung	Yau Chui
G14	Kai Tak South	Kai Tak Central & South
S09	Tai Pak Tin	Tai Pak Tin West
T02	Yat Tung Estate South	Mun Yat
R10	Jat Min	Jat Chuen

Table C1: Changes of constituency names between 2016-2019. *Source:* Government of the HKSAR District Council Election Results, 2019a; HKSAR Census and Statistics Department, 2016.

Table C2: Changes of Constituency Geographical Boundaries

Constituency in 2016	Constituency in 2019
F18 Un Chau and So Uk	F19 Un Chau F20 So Uk
T09 Cheung Chau South T10 Cheung Chau North	T10 Cheung Chau

Table C2: Changes of constituency geographical boundaries between 2016-2019. *Source:* ibid.

Table C3: Constituencies Created For The 2019 District Council Election

Constituency Code in 2019	Constituency Name in 2019
E02	Kowloon Station
F12	Pik Wui
G13	Kai Tak East
J11	Kwung Tong On Tai
J14	On Tat
J22	Chun Cheung
K03	Tsuen Wan South
L12	So Kwun Wat
L28	Yan Tin
M09	Yuen Long Tung Tau
M10	Shap Pat Heung North
M14	Hung Fuk
M17	Shing Yan
Q10	Hoi Chun
Q20	Wai Yan
R09	Shui Chuen O
R25	Hoi Nam
R38	Di Yee
S02	Kwai Luen
S10	Tai Pak Tin East
T05	Tung Chung Central

Table C3: Newly created constituencies for the 2019 District Council election. *Source:* Legislative Council of Hong Kong, 2017; Government of HKSAR Constituency Boundary Maps, 2019.

Appendix D: Variables

Table D1: Variable List

Type of Variable	Name	Description	Unit
Dependent variables	*PDSupport*	Proportion of votes for pro-democracy candidates in each constituency.	%
	PDWinner	Dummy variable which takes the value 1 if the victorious candidate in a constituency was pro-democratic, and zero otherwise.	Binary
Key independent variables	*MTR300*	Number of closed MTR stations, during the 17-day period between November 8 and November 24, 2019, in such geographical proximity to the constituency that a circle with a 300 m radius and with its center at the MTR station covers a part of the constituency.	Numerical
	MTR500	Number of closed MTR stations, during the 17-day period between November 8 and November 24, 2019, in such geographical proximity to the constituency that a circle with a 500 m radius and with its center at the MTR station covers a part of the constituency.	Numerical
	MTR1000	Number of closed MTR stations, during the 17-day period between November 8 and November 24, 2019, in such geographical proximity to the constituency that a circle with a 1,000 m radius and with its center at the MTR station covers a part of the constituency.	Numerical
	MTR Closures Treatment	Dummy variable which takes the value 1 if a constituency has strictly more than zero closed MTR stations during the 17-day period between November 8 and November 24, 2019, in such geographical proximity to the constituency that a circle with a 500 meter radius around the station covers a part of the constituency.	Binary
Demographic control variables	*Age2029*	Proportion of residents in each constituency age 20–29 years (excluding foreign domestic helpers).	%
	Age65	Proportion of residents in each constituency age 65 years or older (excluding foreign domestic helpers).	%
	Education	Proportion of residents in each constituency who have completed post-secondary education.	%
	Mainlanders	Proportion of mainland Chinese residents in each constituency, who have resided in Hong Kong for less than 7 years (excluding foreign domestic helpers).	%
	Students	Proportion of university students in each constituency.	%
Economic control variable	*Income*	Median monthly income from main employment in each constituency (excluding foreign domestic helpers) in Hong Kong Dollars (HKD).	Numerical
Spatial control variable	*ChinaDist*	Distance to the mainland China border, coded according to the constituency's location in one of Hong Kong's three regions. Hong Kong Island = 1, Kowloon = 2, New Territories = 3.	Ordinal

Table D1: Variable list. *Sources, dependent variables:* The Government of the HKSAR District Council Election Results (2019a, 2019b). *Sources, key independent variables:* The Government of the HKSAR Constituency Boundary Maps (2019); MTR Twitter Update (2019). *Source, control variables:* The HKSAR Census and Statistics Department (2016).

Appendix E: Estimation Outputs from Baseline Analyses

Table E1: Forward Stepwise Regression for *PDSupport*

	Education	Income	Students	Age2029	Age65	ChinaDist	Mainlanders	MTR300	MTR500	MTR1000
Round 1	-0.41 (0.685)	-1.71 (0.088)	-0.57 (0.571)	0.42 (0.675)	-0.93 (0.353)	**3.23** **(0.001)**	-1.70 (0.089)	1.94 (0.053)	2.23 (0.026)	1.42 (0.156)
Round 2	0.43 (0.665)	-1.18 (0.240)	-0.84 (0.401)	0.28 (0.781)	-0.44 (0.657)		-1.70 (0.091)	1.69 (0.092)	**2.16** **(0.031)**	1.34 (0.181)
Round 3	0.18 (0.856)	-1.31 (0.191)	-0.87 (0.383)	0.29 (0.773)	-0.44 (0.659)		**-1.91** **(0.056)**			
Round 4	-0.08 (0.939)	**-1.56** **(0.120)**	-0.90 (0.366)	0.41 (0.684)	-0.45 (0.650)					
Round 5	**2.33** **(0.020)**		-0.32 (0.752)	-0.17 (0.866)	-1.21 (0.227)					
Round 6			-0.26 (0.798)	-0.08 (0.932)	-0.48 (0.634)					

Table E1: Forward stepwise regression. *PDSupport* as the dependent variable. T-statistics for all independent variables, p-values in parentheses. Selected regressor marked in **bold.** All regressions are conducted using White heteroscedasticity-robust standard errors. Regressors are selected according to the criterion of the highest absolute t-value and a p-value of <15 %. Round 3 note: as *MTR500* is now included, the variables *MTR300* and *MTR1000* will not be further analyzed in the regression since they often include the same data points and would therefore bring multicollinearity to the model. Round 6 note: None of the three variables are statistically significant at the 15 % level. We therefore terminate the search and conclude that the regression model that will be used is:

$$PDSupport_i = \beta_0 + \beta_1 ChinaDist_i + \beta_2 MTR500_i + \beta_3 Mainlanders_i + \beta_4 Income_i + \beta_5 Education_i + \varepsilon_i$$

Table E2: Variance Inflation Factor

Variable	VIF	1/VIF
MTR500	1.06	0.943270
Education	3.58	0.278951
Income	3.36	0.297796
ChinaDist	1.11	0.902984
Mainlanders	1.04	0.964338
Mean VIF	**2.03**	

Table E2: Variance Inflation Factor for all independent variables. Test for multicollinearity.

Table E3: Balance table for Treatment and Control Group

	Treatment Group (MTRClosuresTreatment = 1)	Control Group (MTRClosuresTreatment = 0)	T-statistic for difference in means, $H_0 : \mu_T - \mu_C = 0$ $H_1 : \mu_T - \mu_C \neq 0$
Education	.2789292 (.006946)	.2638123 (.0077417)	-1.4482
Income	18292.93 (537.8564)	17608.63 (574.0281)	-0.8617
Mainlanders	.025333 (.000746)	.0203286 (.000498)	-5.2204***
Students	.1468755 (.0019593)	.1466086 (.0024023)	-0.0869
Age2029	.124629 (.0020678)	.1316474 (.0025042)	2.1775**
Age65	.1670717 (.0029406)	.1671656 (.0034554)	0.0208
ChinaDist	2.439655 (.0438728)	2.254237 (.0654778)	-2.4362**

Table E3: Balance table for treatment and control group. Mean for all control variables, standard errors in parentheses. Significance levels denoted ***=p<0.01, **=p<0.05, *=p<0.10.

Table E4: Test for Difference in Unconditional Means

	Mean	Std. Err.	Std. Dev.	95 % Conf. Interval		Obs.
Control group	.5523989	.0074104	.0985886	.5377742	.5670235	177
Treatment group	.5683341	.0050532	.0769679	.5583778	.5782903	232
Combined	.5614379	.0043133	.0872317	.5529588	.569917	409
diff	-.0159352	.0086806		-.0329995	.0011291	

diff = mean (0) – mean(1)
H_0 : diff = 0

t = -1.8257
degrees of freedom = 407

H_1 : *diff* < 0	H_1 : *diff* ≠ 0	H_1 : *diff* > 0
Pr(T < t) = 0.0336	Pr(|T| > |t|) = 0.0671	Pr(T > t) = 0.9664

Table E4: Test for difference in unconditional means between treatment (*MTRClosuresTreatment* = 1) and control (*MTRClosuresTreatment* = 0) group. Two-sample t-test with equal variances.

Table E5: Linear Regression

	Coefficient	Robust Std. Err.	t	P>|t|	[95 % Conf. Interval]	
ChinaDist 2	.0139828	.0185764	0.75	0.452	-.0225362	.0505018
ChinaDist 3	.0339909	.0132446	2.57	0.011	.0079536	.0600283
MTR500	.001492	.000682	2.19	0.029	.0001514	.0028327
Mainlanders	-.8805491	.6446766	-1.37	0.173	-2.147908	.3868095
Income	-2.96e-06	1.01e-06	-2.93	0.004	-4.94e-06	-9.74e-07
Education	.1864158	.0813403	2.29	0.022	.0265102	.3463213
Constant	.5559528	.0197432	28.16	0.000	.5171401	.5947655

Table E5: Regression output, linear baseline model. *PDSupport* as the dependent variable, *MTR500* as the key independent variable.

Table E6: Adjusted Predictions of Linear Regression

		Margin	Delta method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Prediction (1) MTR500 = 0 (min)	*Constant*	.5553821	.0060442	91.89	0.000	.5434999	.5672644
Prediction (2) MTR500 = 4.05868 (mean)	*Constant*	.5614379	.0042061	133.48	0.000	.5531692	.5697066
Prediction (3) MTR500 = 30 (max)	*Constant*	.6001436	.0160995	37.28	0.000	.5684938	.6317934

Table E6: Adjusted predictions of linear regression. *PDSupport* as the dependent variable, *MTR500* as the key independent variable. Model VCE: Robust. Number of obs = 409. *MTR500* is set respectively to 0, to its mean of 4.05868, and to 30. The remaining independent variables are set to their mean values (as presented in Table 2).

Table E7: Probit Regression

	Coef.	Robust Std. Err.	z	P>|z|	[95 % Conf. Interval]	
ChinaDist 2	-.4622304	.3133945	-1.47	0.140	-1.076472	.1520117
ChinaDist 3	.1141359	.2506213	0.46	0.649	-.3770729	.6053447
MTR500	.0453439	.0210836	2.15	0.032	.0040209	.0866669
Mainlanders	.8827067	10.23239	0.09	0.931	-19.17241	20.93782
Income	-.0000536	.0000187	-2.87	0.004	-.0000903	-.000017
Education	4.145259	1.561322	2.65	0.008	1.085124	7.205393
Constant	.8313466	.332741	2.50	0.012	.1791862	1.483507

Table E7: Regression output, probit baseline model. *PDWinner* as the dependent variable, *MTR500* as the key independent variable.

Table E8: Average Marginal Effects

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Education	.882692	.3257611	2.71	0.007	.244212	1.521172
Income	-.0000114	3.88e-06	-2.94	0.003	-.000019	-3.81e-06
Mainlanders	.1879637	2.178474	0.09	0.931	-4.081767	4.457695
ChinaDist 2	-.1146518	.0746602	-1.54	0.125	-.2609832	.0316796
ChinaDist 3	.0212793	.0486221	0.44	0.662	-.0740182	.1165769
MTR500	.0096555	.0044536	2.17	0.030	.0009266	.0183845

Table E8: Average marginal effects from the probit baseline model. Note: dy/dx for factor levels is the discrete change from the base level. Model VCE: Robust. Number of obs = 409.

Table E9: Adjusted Predictions of Probit Regression

		Margin	Delta method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Prediction (4) MTR500 = 0 (min)	*Constant*	.8278901	.0266736	31.04	0.000	.7756109	.8801693
Prediction (5) MTR500 = 4.05868 (mean)	*Constant*	.8707402	.0178857	48.68	0.000	.8356848	.9057955
Prediction (6) MTR500 = 30 (max)	*Constant*	.9894496	.0160256	61.74	0.000	.95804	1.020859

Table E9: Adjusted predictions of probit regression. *PDWinner* as the dependent variable, *MTR500* as the key independent variable. Model VCE: Robust. Number of obs = 409. *MTR500* is set respectively to 0, to its mean of 4.05868, and to 30. The remaining independent variables are set to their mean values (as presented in Table 2).

Appendix G: Estimation Outputs from Sensitivity Analyses

Table G1: Linear Regression, Sensitivity Analysis I

	Coefficient	Robust Std. Err.	t	P>|t|	[95 % Conf. Interval]	
ChinaDist 2	.010631	.0195633	0.54	0.587	-.0278282	.0490901
ChinaDist 3	.0313915	.0136688	2.30	0.022	.0045204	.0582627
MTRClosuresTreatment 1	.0178013	.0094563	1.88	0.060	-.0007887	.0363913
Mainlanders	-.9359544	.6254766	-1.50	0.135	-2.165568	.2936592
Income	-3.04e-06	1.04e-06	-2.92	0.004	-5.09e-06	-9.94e-07
Education	.1922703	.0825032	2.33	0.020	.0300786	.354462
Constant	.555479	.0195131	28.47	0.000	.5171186	.5938394

Table G1: Regression output, linear model for Sensitivity Analysis I. *PDSupport* as the dependent variable, *MTRClosuresTreatment* as the key independent variable.

Table G2: Probit Regression, Sensitivity Analysis I

	Coef.	Robust Std. Err.	z	P>|z|	[95 % Conf. Interval]	
ChinaDist 2	-.5693764	.324697	-1.75	0.080	-1.205771	.067018
ChinaDist 3	.0255763	.2517661	0.10	0.919	-.4678762	.5190289
MTRClosuresTreatment 1	.5371378	.1716157	3.13	0.002	.2007772	.8734984
Mainlanders	-1.686126	10.4817	-0.16	0.872	-22.22987	18.85762
Income	-.0000573	.0000187	-3.07	0.002	-.000094	-.0000207
Education	4.376511	1.556814	2.81	0.005	1.325212	7.42781
Constant	.8576563	.3317917	2.58	0.010	.2073565	1.507956

Table G2: Regression output, probit model for Sensitivity Analysis I. *PDWinner* as the dependent variable, *MTRClosuresTreatment* as the key independent variable.

Table G3: Average Marginal Effects, Sensitivity Analysis I

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Education	.9233769	.3232915	2.86	0.004	.2897373	1.557017
Income	-.0000121	3.87e-06	-3.13	0.002	-.0000197	-4.52e-06
Mainlanders	-.355747	2.212331	-0.16	0.872	-4.691836	3.980342
ChinaDist 2	-.1360115	.0736276	-1.85	0.065	-.2803191	.008296
ChinaDist 3	.0045301	.0449984	0.10	0.920	-.0836652	.0927254
MTRClosuresTreatment 1	.1169332	.0379193	3.08	0.002	.0426127	.1912537

Table G3: Average marginal effects from the probit model for Sensitivity Analysis I. Note: dy/dx for factor levels is the discrete change from the base level. Model VCE: Robust. Number of obs = 409.

Table G4: Linear Regression, Sensitivity Analysis II (MTR300)

	Coefficient	Robust Std. Err.	t	P>|t|	[95 % Conf. Interval]	
ChinaDist 2	.0146215	.018807	0.78	0.437	-.022351	.0515939
ChinaDist 3	.0344984	.0134064	2.57	0.010	.0081431	.0608538
MTR300	.0011793	.0007782	1.52	0.130	-.0003506	.0027092
Mainlanders	-.82444	.644383	-1.28	0.201	-2.091221	.4423414
Income	-3.01e-06	1.00e-06	-3.00	0.003	-4.97e-06	-1.04e-06
Education	.1964182	.0815664	2.41	0.016	.0360683	.3567681
Constant	.5555869	.0197802	28.09	0.000	.5167014	.5944725

Table G4: Regression output, linear model for Sensitivity Analysis II. *PDSupport* as the dependent variable, *MTR300* as the key independent variable.

Table G5: Probit Regression, Sensitivity Analysis II (MTR300)

	Coef.	Robust Std. Err.	z	P>|z|	[95 % Conf. Interval]	
ChinaDist 2	-.4849508	.3186771	-1.52	0.128	-1.109547	.1396449
ChinaDist 3	.077433	.2532011	0.31	0.760	-.418832	.5736979
MTR300	.0680359	.0283948	2.40	0.017	.0123832	.1236886
Mainlanders	1.573831	10.20668	0.15	0.877	-18.43088	21.57855
Income	-.0000533	.0000187	-2.86	0.004	-.0000899	-.0000168
Education	4.144282	1.562818	2.65	0.008	1.081214	7.20735
Constant	.862543	.3377058	2.55	0.011	.2006519	1.524434

Table G5: Regression output, probit model for Sensitivity Analysis II. *PDWinner* as the dependent variable, *MTR300* as the key independent variable.

Table G6: Average Marginal Effects, Sensitivity Analysis II (MTR300)

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Education	.8815182	.3256089	2.71	0.007	.2433365	1.5197
Income	-.0000113	3.88e-06	-2.93	0.003	-.0000189	-3.75e-06
Mainlanders	.3347651	2.170135	0.15	0.877	-3.918622	4.588152
ChinaDist 2	-.1185041	.0744806	-1.59	0.112	-.2644835	.0274753
ChinaDist 3	.0143003	.0480477	0.30	0.766	-.0798714	.108472
MTR300	.0144717	.0060047	2.41	0.016	.0027026	.0262408

Table G6: Average marginal effect from the probit model for Sensitivity Analysis II. Note: dy/dx for factor levels is the discrete change from the base level. Model VCE: Robust. Number of obs = 409.

Table G7: Linear Regression, Sensitivity Analysis II (MTR1000)

	Coefficient	Robust Std. Err.	t	P>|t|	[95 % Conf. Interval]	
ChinaDist 2	.011368	.018829	0.60	0.546	-.0256476	.0483836
ChinaDist 3	.0332273	.0130238	2.55	0.011	.007624	.0588307
MTR1000	.0008124	.0005249	1.55	0.122	-.0002195	.0018444
Mainlanders	-.8710966	.6556781	-1.33	0.185	-2.160083	.4178896
Income	-2.96e-06	1.00e-06	-2.95	0.003	-4.92e-06	-9.89e-07
Education	.1857509	.0778678	2.39	0.018	.0326718	.3388299
Constant	.556314	.0197209	28.21	0.000	.5175449	.595083

Table G7: Regression output, linear model for Sensitivity Analysis II. *PDSupport* as the dependent variable, *MTR1000* as the key independent variable.

Table G8: Probit Regression, Sensitivity Analysis II (MTR1000)

	Coef.	Robust Std. Err.	z	P>|z|	[95 % Conf. Interval]	
ChinaDist 2	-.5417774	.3139529	-1.73	0.084	-1.157114	.073559
ChinaDist 3	.0853864	.2534668	0.34	0.736	-.4113994	.5821722
MTR1000	.0282525	.0121818	2.32	0.020	.0043766	.0521284
Mainlanders	.9734922	10.53074	0.09	0.926	-19.66639	21.61337
Income	-.0000518	.0000185	-2.80	0.005	-.0000881	-.0000156
Education	3.834093	1.542614	2.49	0.013	.8106242	6.857562
Constant	.8599165	.3346188	2.57	0.010	.2040757	1.515757

Table G8: Regression output, probit model for Sensitivity Analysis II. *PDWinner* as the dependent variable, *MTR1000* as the key independent variable.

Table G9: Average Marginal Effects, Sensitivity Analysis II (MTR1000)

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Education	.8171838	.3217286	2.54	0.011	.1866073	1.44776
Income	-.0000111	3.86e-06	-2.86	0.004	-.0000186	-3.48e-06
Mainlanders	.2074864	2.244841	0.09	0.926	-4.192322	4.607294
ChinaDist 2	-.1350295	.074713	-1.81	0.071	-.2814644	.0114053
ChinaDist 3	.0155503	.0475324	0.33	0.744	-.0776115	.108712
MTR1000	.0060216	.0025762	2.34	0.019	.0009724	.0110708

Table G9: Average marginal effect from the probit model for Sensitivity Analysis II. Note: dy/dx for factor levels is the discrete change from the base level. Model VCE: Robust. Number of obs = 409.

Table G10: Descriptive Statistics of the Data Set for Sensitivity Analysis III

	Mean	Std. Dev.	Min	Max	Obs.
Dependent variables					
PDSupport	.5607927	.0869406	.0111	.8836	440
PDWinner	.8477273	.359694	0	1	440
0 Not pro-demodratic					*67*
1 Pro-democratic					*373*
Key independent variables					
MTR300	2.325	3.966409	0	23	440
MTR500	3.895455	5.225919	0	30	440
MTR1000	8.172727	8.213117	0	57	440
MTR Closures Treatment	.5545455	.4975816	0	1	440
0 No Closures					*196*
1 Closures					*244*
Control variables					
Age2029	.1277696	.0327852	.0515458	.2690399	440
Age65	.1677731	.0460144	.066702	.3685957	440
Education	.2687397	.1048113	.0861442	.552929	440
Mainlanders	.0234409	.0100748	.0105983	.0492386	440
Students	.1471068	.0308602	.0557664	.2797719	440
Income	17908.8	8101.3	11000	75000	440
ChinaDist	2.35	.7637875	1	3	440
1 Hong Kong					*78*
2 Kowloon					*130*
3 New Territories					*232*

Table G10: Descriptive statistics for all dependent and independent variables of the data set for Sensitivity Analysis III. For the variable *Mainlanders* with district-level variation, we attribute the district average to each constituency within that district. For the newly created constituencies presented in Appendix C Table C3, we attribute the district average. *Sources, dependent variables:* The Government of the HKSAR District Council Election Results (2019a, 2019b). *Sources, key independent variables:* The Government of the HKSAR Constituency Boundary Maps (2019); MTR Twitter Update (2019). *Source, control variables:* The HKSAR Census and Statistics Department (2016).

Table G11: Linear Regression, Sensitivity Analysis III

	Coefficient	Robust Std. Err.	t	P>|t|	[95 % Conf. Interval]	
ChinaDist 2	.0130398	.0172866	0.75	0.451	-.0209362	.0470158
ChinaDist 3	.0337306	.0124963	2.70	0.007	.0091697	.0582914
MTR500	.0014126	.0006649	2.12	0.034	.0001059	.0027194
Mainlanders	-.8429728	.5926552	-1.42	0.156	-2.007811	.3218659
Income	-3.61e-06	9.51e-07	-3.79	0.000	-5.48e-06	-1.74e-06
Education	.2253089	.075498	2.98	0.003	.0769207	.373697
Constant	.5574634	.0187874	29.67	0.000	.5205375	.5943893

Table G11: Regression output, linear model for Sensitivity Analysis III. *PDSupport* as the dependent variable, *MTR500* as the key independent variable.

Table G12: Probit Regression, Sensitivity Analysis III

	Coef.	Robust Std. Err.	z	P>|z|	[95 % Conf. Interval]	
ChinaDist 2	-.3173598	.2959279	-1.07	0.284	-.8973679	.2626483
ChinaDist 3	.1625832	.237965	0.68	0.494	-.3038196	.628986
MTR500	.0454777	.0211112	2.15	0.031	.0041005	.086855
Mainlanders	-2.472107	9.459303	-0.26	0.794	-21.012	16.06779
Income	-.0000556	.0000168	-3.30	0.001	-.0000886	-.0000226
Education	4.219553	1.429602	2.95	0.003	1.417584	7.021522
Constant	.8637494	.3155623	2.74	0.006	.2452587	1.48224

Table G12: Regression output, probit model for Sensitivity Analysis III. *PDWinner* as the dependent variable, *MTR500* as the key independent variable.

Table G13: Average Marginal Effects, Sensitivity Analysis III

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
Education	.9107706	.3014402	3.02	0.003	.3199586	1.501583
Income	-.000012	3.52e-06	-3.41	0.001	-.0000189	-5.10e-06
Mainlanders	-.5335927	2.04227	-0.26	0.794	-4.536368	3.469182
ChinaDist 2	-.0789112	.0713899	-1.11	0.269	-.2188329	.0610105
ChinaDist 3	.032078	.0495775	0.65	0.518	-.065092	.1292481
MTR500	.0098162	.0045182	2.17	0.030	.0009606	.0186717

Table G13: Average marginal effect from the probit model for Sensitivity Analysis III. Note: dy/dx for factor levels is the discrete change from the base level. Model VCE: Robust. Number of obs = 440.