Movement and the State

Lianzhou Tang* Qiong Zhang†

August 17, 2026

Abstract

This paper examines how a short-lived social movement, widely seen as a failure, shaped state formation over subsequent decades. We study the May Thirtieth Movement of 1925 in China, a nationwide anti-imperialist mobilization led primarily by the newly founded Chinese Communist Party (CCP) and swiftly repressed. Exploiting spatial variation in exposure to the movement, we establish five main findings: (1) the movement sharply accelerated the local organizational development of the CCP and (2) generated sizable spatial spillovers through party-member migration and the diffusion of organizational templates. (3) The rapid growth of the CCP was associated with local ideological change, as reflected in increased publication of communist-related books and stronger working-class connotations in newborn names, with little evidence of enhanced military capacity or fewer attacks by rival parties. (4) During the full-scale civil war, the CCP captured May Thirtieth areas earlier than comparable non-exposed areas. (5) After the founding of the People's Republic of China (PRC), exposed areas exhibited more radical behavior during subsequent far-left political campaigns, adjusted more slowly to rightward policy shifts, and, by the twenty-first century, displayed greater dissatisfaction with government and stronger anti-foreign attitudes.

JEL Codes: *D72; D74; H11; N45; P16.*

Keywords: *Movement; state; ideology; ideological legacy.*

*School of Applied Economics and †School of Public Administration, Renmin University of China. Email: cy.tang@ruc.edu.cn. We thank David Yang for helpful discussions. Scheduled presentations include the 12th Monash–Paris–Warwick–Zurich–CEPR Text-As-Data Workshop and the Quantitative History Workshop at the University of Hong Kong. We gratefully acknowledge financial support from the International Growth Centre (Grant 7933416-01) and the General Program of the National Natural Science Foundation of China (NSFC; Grant No. 72473147). The authors declare no conflicts of interest.

"A Single Spark Can Start a Prairie Fire"

— *Mao Zedong, 1930 letter, on the mobilizational energy of the May Thirtieth Movement*

I. Introduction

Where do the forces capable of creating new states come from? Social scientists have long sought to answer this question,[1] with a particularly large literature emphasizing the role of war.[2] Yet even when forces conducive to state formation are present, including state-making wars, it remains unclear what circumstances make organized collective action valuable to participants and allow it to succeed. We approach this question through social movements, which frequently accompany episodes of social revolution (Tilly, 1978, 1984, 1993).[3] During mass mobilization, the grievances of relatively disadvantaged groups are transformed into collective claims against incumbent rulers whose interests conflict with theirs. We know surprisingly little about whether movements contribute to state formation or merely reflect it.

This paper studies the May Thirtieth Movement of 1925 in China,[4] a large-scale social movement that has received limited attention even among historians (Ku, 1979). The movement was an anti-imperialist mobilization led primarily by the newly founded Chinese Communist Party (CCP), and did not explicitly aim to overthrow the existing regime. Relative to many consequential social movements, it was extraordinarily brief, lasting only about two months before being met by countermeasures and repression from both foreign actors and domestic authorities. Its direct geographic reach was also limited, as official accounts indicate that only a subset of counties participated. Yet this short-lived episode coincided with a striking political transformation. Around 1925, the CCP evolved from what has been described as a nearly stillborn weak party (van de Ven, 1992) into a rapidly expanding organization.[5] Whether, and through what mechanisms, this brief movement contributed to the CCP's eventual founding of the People's Republic of China (PRC) in 1949 is an unresolved empirical question.

Our analysis is made possible by extensive data collection. By compiling all available local, municipal, and provincial gazetteers, together with organizational histories, we cross-validate the first CCP member, the first party organization, and the subsequent

[1] Existing explanations include the emergence of new ruling elites enabled by taxable surpluses (Olson, 1993; Sánchez de la Sierra, 2020), crises (Skocpol, 1979), transformations of the international order and the recognition of sovereignty (Jackson and Rosberg, 1982; Jackson, 1990; Meyer et al., 1997; Krasner, 1999; Philpott, 2010), and trade-offs shaped by economic growth (Alesina and Spolaore, 1997).

[2] See, for example, Tilly (1975), Parker (1988), Downing (1992), Tilly (1992), and Holsti (1996).

[3] Prominent examples include the Russian October Revolution of 1917, which brought the Bolsheviks to power and paved the way for the Soviet state; the Iranian Revolution of 1979, which toppled the Pahlavi monarchy and established the Islamic Republic; and the Haitian Revolution of 1791–1804, which ended French colonial rule and created an independent state.

[4] The movement is named after the events of May 30, 1925.

[5] See Appendix Figure A5 for changes in CCP membership.

establishment and dissolution of county party committees in each county.[6] By manually coding historical materials on the May Thirtieth Movement, we locate the movement's spatial coverage across China. We further compile a large body of available historical sources to examine what role the May Thirtieth Movement played and how it did so.

We first show that party organization expanded slowly before 1925. Although the CCP had already made substantial efforts to extend member activity, organizational development lagged far behind. County coverage of party organizations was only one quarter of county coverage of member activity. Fewer than 20% of counties established a party organization in the year of the first local CCP member's arrival or in the following year. After 1925, the same-year-or-following-year share exceeded 50%.

The analysis then examines the role of the May Thirtieth Movement in the CCP's rapid expansion. Our analysis exploits the substantial spatial variation in exposure to the movement. To address spatial selection, we control for a rich set of industrial, agricultural, and transportation endowments that may have shaped May Thirtieth exposure, together with province-by-year fixed effects, so that treated and untreated counties are compared within narrowly defined cells (Heckman et al., 1998b). To address potential selection concerns, we conduct two additional specifications. (1) We use the synthetic difference-in-differences estimator to closely match pre-treatment trends between treated and control counties, thereby strengthening the credibility of the post-treatment comparison (Arkhangelsky et al., 2021). We also implement the changes-in-changes estimator, which allows for substantial distributional differences between treated and control counties (Athey and Imbens, 2006). (2) We construct the boundary between May Thirtieth and non-May Thirtieth areas and apply a spatial regression discontinuity design. Across the empirical strategies, the results show a highly consistent pattern: relative to CCP member activity, party organizations and county party committees increased much more sharply in May Thirtieth areas. The pattern suggests that the movement directly promoted the intensive margin of local CCP development. Using biographical dictionaries and martyr records, we also find that May Thirtieth areas experienced increased CCP entry across social groups, with the strongest response among students. In addition, counties that already had traces of CCP member activity before 1925 developed party organizations more rapidly after the movement.

We next show how the May Thirtieth Movement radiated beyond its direct geographic coverage. Its effects remained visible in both first-order and second-order counties adjacent to May Thirtieth areas (with effects on party organizations extending farther). Adjacent-county effects do not appear to reflect the geographic spread of the movement itself.[7] We document two channels of diffusion. First, counties adjacent to May Thirtieth

[6] County party committees are themselves party organizations, but they are the highest-ranking CCP committees at the county level. In the analysis that follows, we therefore distinguish them from other local party organizations.

[7] Using chronicle entries from local gazetteers, we show that protest and demonstration activities surged in 1925 only in May Thirtieth areas, rather than in other places or other years.

areas received more returning native CCP members from May Thirtieth areas, especially students and workers, consistent with the social composition of participants in the movement. Second, party organizations in adjacent counties became more similar to those in May Thirtieth areas. The diffusion patterns help explain how a movement that directly covered only a limited share of Chinese territory could shape CCP development across the country.

The analysis then examines why a massive but short-lived movement produced persistent effects. We assess three explanations. (1) Local ideological change. We provide two complementary pieces of evidence. First, using about 420,000 books, we identify books that promoted communism, socialism, or broader leftist revolutionary ideas (hereafter, red books). We find that, in 1925–1927, after the May Thirtieth Movement began, the publication of red books increased sharply in May Thirtieth areas. This initial increase was not driven by the CCP itself, since fewer than 20% of red books were written by authors who were already party members at the time of publication. Nor was it driven by a general expansion in book production, since science and engineering books did not increase. As relations between the Kuomintang (KMT) and the CCP deteriorated sharply, red-book growth remained weak during 1928–1945. Red-book publication surged again only after the KMT and the CCP entered full-scale war, with faster growth than before. Second, using martyrs' names and birth years, together with available worker and business-elite rosters as training data, we find a sustained increase in names with stronger working-class connotations. This increase persisted even during periods in which the CCP was subject to extensive KMT purges. (2) Strengthened coercive capacity. County-level data on CCP garrisons show no increase in CCP garrison presence in May Thirtieth areas, which rules out the possibility that the CCP achieved organizational expansion through coercive capacity. By contrast, KMT troops increased, possibly to suppress CCP activity in May Thirtieth areas. (3) Differences in the intensity of attacks. We find no clear increase in martyr deaths in May Thirtieth areas after the 1927 purge,[8] and no systematic difference between May Thirtieth and non-May Thirtieth areas in the ease of establishing or revoking county committees. The absence of differential purging is consistent with the highly top-down nature of the KMT purge of the CCP after 1927. However, after county committees were revoked, martyr deaths did not decline significantly in May Thirtieth areas. The post-revocation pattern suggests stronger organizational resilience: revolutionary activity could continue even after formal county committees were dismantled.

We next investigate the role of May Thirtieth areas during the KMT-CCP civil war after 1945. We find that these areas were systematically captured earlier by the CCP. This pattern admits two explanations: the CCP may have prioritized attacks on May Thirtieth areas, reflecting path choice, and May Thirtieth areas may have been easier

[8]After controlling for the presence of county committees, May Thirtieth areas exhibit no unusually large increase in martyr deaths. This finding indicates that the purge did not specifically target May Thirtieth areas, but instead targeted counties with party committees.

for the CCP to capture. These channels are complementary, since May Thirtieth areas had substantially more developed CCP organizations by 1945. We provide suggestive evidence for both. First, the routes through which the CCP connected two CCP-held areas contained more May Thirtieth counties than the corresponding shortest-distance paths. Second, martyrs in May Thirtieth areas were more likely to come from local native places. This suggests that their capture involved greater local mobilization and internal support, in addition to external military force.

Finally, we examine the further influence of the May Thirtieth Movement on the new state after the founding of the PRC. We find that May Thirtieth areas retained a visible left-wing ideological imprint. This imprint appears in two patterns. (1) May Thirtieth areas acted more aggressively during the Great Leap Forward and suffered more severely during the Great Famine. (2) During subsequent bottom-up rightward reforms, their response was slower. We also document that, by the twenty-first century, May Thirtieth areas expressed greater dissatisfaction with government, stronger demand for democracy, and stronger anti-foreign attitudes. Given the many confounding factors that accumulated over such a long horizon, we cannot determine whether later dissatisfaction reflected unmet expectations formed through earlier ideological diffusion or the severe harm these areas later experienced during radical political campaigns.

Related Literature. Our first contribution is to the literature on the political consequences of social movements. Compared with the large literature on the emergence and diffusion of movements, especially protests,[9] the political consequences of movements remain less fully explored. Existing theoretical and empirical work examines how protests shape the short-run political preferences of voters and politicians,[10] as well as subsequent policy choices (Gillion, 2012; Madestam et al., 2013; Aidt and Franck, 2015; Battaglini, 2017; Enos et al., 2019; Wasow, 2020; Casanueva, 2021; Gethin and Pons, 2024). Part of our analysis is consistent with prior evidence that protest activity can affect short-run political demands, while our main contribution goes further by studying how a short-lived movement shaped, from below, the extensive margin of long-run organized politics that later underpinned a new state. Notably, unlike much of the existing literature focusing on voters in democratic settings, this paper helps illuminate the dynamic consequences of movements under autocratic rule.[11] Another difference from existing research is that the stated demands of the movement do not directly map onto the outcomes we study, allowing us to examine the unintended consequences of social movements (Giugni, 1998; Amenta and Polletta, 2019). Owing to the paper's distinctive setting (sample period and spatial scope), we also speak to three additional questions: (1) the mutual shaping of social movements and war, a link often assumed in studies of state formation (Skocpol,

[9] See Cantoni et al. (2024) for an excellent review.

[10] One exception that examines long-run effects is Hungerman and Moorthy (2023), who study how the 1970 Earth Day events shaped children's environmental attitudes.

[11] The question is especially relevant today, when more than 70% of the world's population lives under autocratic rule (Nord et al., 2025).

1979; Tarrow, 2015); (2) how the consequences generated by protest activity persisted and diffused geographically, whereas existing work focuses more on how protest activity itself spreads across space (Madestam et al., 2013; Bursztyn et al., 2021); and (3) the long-run social legacies of social movements (McAdam, 1989; Andrews, 2004).

Second, this paper contributes to the literature on state formation. Existing work has developed rich insights into the origins of early states (Scott, 2017; Mayoral and Olsson, 2019; Mayshar et al., 2022; Allen et al., 2023; Fernández-Villaverde et al., 2023). Allen et al. (2023) study the origins of early governments and show that citizens voluntarily delegated rights to rulers in exchange for public goods (Locke, 1966; Hobbes and Klenner, 2005), rather than being ruled by coercive power formed by a small elite class (Engels, 1934). Our findings on modern state formation echo this logic. We show that a large-scale social movement helped form a new state by reshaping political identity, a force especially important in modern state building (Akerlof and Kranton, 2000; Alesina et al., 2000). Identity, in turn, entered citizens' utility calculations when they decided whether to delegate rights to a prospective government, eventually contributing to state formation. Beyond the evidence we provide, direct evidence on the origins of modern states remains scarce.[12]

A large literature studies how war shapes states and their capacity (Hintze, 1975; Tilly, 1985; Parker, 1988; Downing, 1992; Ertman, 1997; Gennaioli and Rainer, 2007; Besley and Persson, 2009; Dincecco and Prado, 2012). The literature emphasizes how war promotes administration, consolidates the monopoly of violence, and generates regular extractive capacity. Our question is why some wars instead lead states toward decline (Kaspersen and Strandsbjerg, 2017). War requires broad social cooperation (Centeno, 2002; Tarrow, 2015; Hutchinson, 2017), and citizen support depends on the trade-off between expected welfare gains and current costs. By showing how a large-scale social movement shaped citizen preferences, this paper illuminates the interaction among movements, war, and state formation.

Finally, this paper contributes to our understanding of the rise of the CCP. How a party that was tiny at birth and lacked orthodox legitimacy seized power in less than three decades and founded the PRC is itself a long-standing question for historians. One strand of the literature emphasizes the CCP's ability to mobilize popular emotions, build ideological attachments, and construct political legitimacy as a key source of its rise (Chesneaux, 1968; Smith, 2002; Zarrow, 2006). Using modern econometric tools, we formalize this view and provide evidence beyond institutional explanations for why the CCP began to organize rapidly only a few years after its founding (van de Ven, 1992). The closest work in economics is Chen and Kung (2026), which studies how the full-scale Second Sino-Japanese War beginning in 1937 promoted CCP development and finds greater CCP presence in Japanese-occupied areas. Our findings are complementary. We focus on

[12] Mastrorocco and Teso (2023) study how regulatory capacity shaped the geographic presence of the U.S. federal government in a principal-agent framework.

the CCP's earlier, near-from-scratch rise, a process that Chen and Kung (2026) also highlights as important but leaves open. We also show that omitting controls for Japanese occupation slightly attenuates most of our estimates, because May Thirtieth exposure, despite its broad geographic distribution, was negatively correlated with Japanese military activity, which was concentrated mainly in northern China for much of the period.

II. Historical Background

II.A. China's Political Landscape: 1911–1949

In 1911–1912, the Qing state rapidly lost authority and collapsed. The sudden dissolution of the old imperial order left Chinese society institutionally unprepared, ushering in decades of social unrest (Young, 1980). In 1912–1913 alone, China had more than 300 political parties (Phillips, 1996), all competing for power in a near vacuum. Drawing on its revolutionary legitimacy[13] and absorbing many former allies and smaller parties, the KMT quickly became the largest party in China. In the 1913 elections, it won 269 of 596 seats in the House of Representatives and 123 of 274 seats in the Senate.

Yet the KMT's early electoral strength did not translate into broad social recognition. From 1915 onward, foreign encroachments on Chinese sovereignty[14] and the perceived weakness of Republican diplomacy generated widespread disappointment with the existing political order and governing strategy. The 1917 Russian Revolution then offered an apparently successful model of rule, prompting some Chinese intellectuals to look to Marxism as a solution to China's predicament (Chow, 1960). Russia was also willing to foster an ideological proxy in China. In 1921, the CCP was formally founded.

At the time of the CCP's founding, the KMT was still led by Sun Yat-sen, and the CCP remained a target of active KMT cooperation. Because the CCP was backed by the Soviet Union, cooperation with it promised organizational assistance and a stronger basis for resisting Western encroachment. This alliance, however, did not last. After Sun's death in 1925, Chiang Kai-shek gradually came to dominate the KMT and adopted an increasingly hostile stance toward the CCP. First, the KMT right wing relied heavily on chambers of commerce and other local elites, many of whom were tied to foreign capital, which encouraged a more ambiguous posture toward the Western powers. By contrast, the CCP sought to mobilize students, workers, and peasants, creating a sharp conflict in the two parties' social bases. Second, the CCP's frequent propaganda work and rapid organizational expansion made many within the KMT increasingly uneasy. These tensions culminated in the systematic purges and massacres of CCP members in 1927–1928. For the following decades, the two parties remained broadly antagonistic. After

[13]The KMT's predecessor, the Tongmenghui, had long been committed to overthrowing the Qing government.

[14]In 1915, Japan, taking advantage of the European powers' preoccupation with World War I, issued the Twenty-One Demands to the Chinese government, seeking expanded privileges in Shandong, South Manchuria, eastern Inner Mongolia, mining, railways, and other areas.

the Second Sino-Japanese War ended in 1945, the KMT and the CCP entered a full-scale civil war, which ended in CCP victory. In 1949, CCP chairman Mao Zedong proclaimed the founding of the PRC.

II.B. The May Thirtieth Movement

Origins of the Movement. In the early twentieth century, large amounts of foreign capital entered China, especially Shanghai, where numerous concessions provided space for foreign economic and political activity. Chinese workers in foreign-owned enterprises were often poorly treated. In Japanese-owned cotton mills, for example, workers commonly worked more than twelve hours a day and were subject to arbitrary beatings by Japanese supervisors (Smith, 2002). These conditions generated widespread labor tension and hostility toward foreigners.

The immediate trigger of the May Thirtieth Incident was the attempt by Japanese cotton mills to further reduce male employment and replace men with women and girls, whom employers viewed as easier to control. This provoked broad discontent among Shanghai workers and led to protests. In May 1925, the worker representative Gu Zhenghong was shot dead by Japanese personnel during a protest, sharply escalating the situation and leading to citywide demonstrations in Shanghai at the end of the month. The demonstrations were initially confined to Shanghai, and their demands centered on punishing the perpetrator and limiting the power of foreign-owned firms. The movement reached an interim climax in Shanghai on May 30, from which later generations derived the name "May Thirtieth Movement"; at the time, it was more commonly known as the "Shanghai Incident". CCP leaders had already recognized the importance of binding mass ideology to the party and increasing the party's public visibility (Saich, 2020). In the early stage of the movement, the CCP quickly seized the opportunity to lead the movement and expand its influence. The movement was primarily led by Qu Qiubai, a later CCP leader.[15] Left-wing figures within the KMT did not oppose the movement at first and even supported it (Zarrow, 2006). Improving the lives of ordinary people and securing China's substantive independence were central to the KMT's early political program, helping the party claim political legitimacy in a period of disorder. Strikes could also strengthen the Chinese government's bargaining position in subsequent negotiations, including over tariff policy (Orchard, 1930). Compared with the KMT, however, the CCP was far more adept at channeling public anger and therefore assumed an almost leading role from the outset (Zarrow, 2006; Perry, 2007).

Diffusion and Repression of the Movement. The CCP made strenuous efforts to shape the spatial diffusion of the movement, deliberately turning a local disturbance into a nationwide political campaign (Institute of History, Shanghai Academy of Social Sciences, 2005). The movement spread rapidly through the postal system, which had

[15] Qu twice served as the de facto top leader of the CCP (August 1927–July 1928; September 1930–January 1931). In 1935, he was arrested and executed by the KMT.

become extensive after the fall of the Qing (Harris, 2012), as well as through newspapers and other media. CCP members also traveled widely, gave public speeches, and called on students to join the movement.[16] As the movement spread, its slogans broadened from "punish the perpetrators" to anti-imperialist resistance.[17] By tapping into long-standing grievances over imperialist aggression, this broader framing helped mobilize participation on a much larger scale (Zarrow, 2006). Consistent with this nationwide wave, strike intensity increased by a factor of 5–8 (Appendix Figure A1). It also greatly raised the CCP's visibility and gave the party a vehicle for spreading communist and left-revolutionary ideas, including class struggle and worker-peasant solidarity (van de Ven, 1992; Smith, 2002). Mao also recalled that after the May Thirtieth Movement, peasants' political consciousness became increasingly radical (Womack, 2018).

Official CCP sources estimate that 600–700 counties participated in the May Thirtieth Movement.[18] Why, then, did a movement that appealed to national solidarity fail to achieve even wider coverage? A key factor was obstruction by the KMT over ideological conflict. From the outset, the KMT right wing repeatedly spread alarm, emphasized the negative consequences of protests, and portrayed the movement as one of the CCP's "tricks" (Perry, 1993). As the movement expanded, the CCP's leading role and broad popular support made the benefits of continued KMT support increasingly disproportionate to its costs, further angering the KMT (Perry, 2007). Anti-CCP propaganda began to appear in July (Chen, 2015). Beyond the threat to political control, the movement exceeded the limits that the local elites on whom the KMT relied were willing to tolerate, since they were themselves targets of protest in a broader sense. Some records show that right-wing unions secretly hired local gangs to carry out violence against protest organizations (Perry, 2007). In addition, when foreign troops (particularly British and Japanese forces, which had the deepest ties to China) carried out extensive armed repression and retaliated through large-scale power cutoffs, the KMT feared that the relatively stable order that had emerged after the fall of the Qing would be disrupted (Smith, 2002). The public did not accept the KMT's obstructive role. For example, people mocked KMT members who refused to join the movement and discouraged others from doing so as engaging in "destructive, shameful behavior" (Perry, 1993).

Discussions of the Movement. Supporters of the movement argued that its first defining feature was that it marked China's first mass-mobilizational political movement (Qu, 1926), and that it signaled the beginning of the National Revolution (the rise of the CCP) (Qu, 1926). In 1930, when Mao Zedong sought to encourage party members demoralized by KMT attacks, he argued that the source of the CCP's development had not been extinguished: "sparks" such as the May Thirtieth Movement could still "start

[16] Figure 1(a) plots the coverage of CCP member activity and shows a sharp increase in 1925.

[17] Figure 2 presents a word cloud based on 109 declarations and petitions collected from Institute of History, Shanghai Academy of Social Sciences (1981, 1986, 2005).

[18] See `https://cpc.people.com.cn/BIG5/33837/2533866.html?`.

a prairie fire" (Mao, 1944). It was also the May Thirtieth Movement that confirmed the CCP's later political line of broadly uniting the masses (Deng, 1930). Critics, including former CCP members, figures close to the CCP, and other cultural elites, raised a wider range of objections. Some dismissed the movement as no more than an expression of the Chinese tendency to "strike high-sounding tones" (Ding, 1925), or as another adventurist action by the CCP (Chen, 1925). Others argued that the protest should have remained confined to Shanghai and focused on the incident itself (Liang, 1925), criticized the CCP's ineffective leadership during the movement (Zhang, 1998), or worried that it would alarm the bourgeoisie and ultimately backfire against the CCP itself (Chen, 2009). Yet supporters and critics largely agreed that, in immediate terms, the May Thirtieth Movement failed: it was quickly suppressed, its anti-imperialist demands were largely unmet, and it produced few visible benefits or direct consequences (Deng, 1930).

II.C. CCP Party Organizations

Before 1925, CCP members had been active across China in an effort to build the party's organizational presence, but with limited success (van de Ven, 1992). Establishing a local party organization required overcoming at least five obstacles. First, it required a sufficient number of party members. A party branch, for example, needed at least three members. Second, the formation of a party branch required its members to move from individual ideological commitment to strict organizational obedience, a defining feature of the CCP (van de Ven, 1992; Ip, 1994). Third, in a politically fragmented China dominated by the KMT, the creation of a party organization required broad local acceptance, since it entailed potential political risks. Unless local residents believed that such an organization would bring future benefits exceeding those risks (Johnson, 1962), its establishment was difficult; it also posed a threat to local elites (Perry, 1980). Fourth, creating a party organization increased the security risks faced by its members in subsequent party work (Stranahan, 1998). Fifth, a local party organization also required recognition from higher-level party authorities (van de Ven, 1992).

County-level CCP organizations commonly took several forms. First, small groups or branches constituted the most basic party units, although their names varied across periods. Second, party fractions or party groups were CCP organizations embedded in non-party associations, such as peasant associations and labor unions. Third, county work committees, work teams, and similar bodies were usually dispatched organs created by higher-level party organizations. Fourth, the county party committee was the highest-ranking party organization at the county level and coordinated overall local party activity.

III. Data

III.A. CCP Development

One central focus of this paper is the development of party organizations across counties between 1921 and 1950. We integrate a wide range of data sources to ensure that each record can be checked against at least two independent sources. Our baseline source consists of 3,647 local gazetteers, with some counties publishing multiple editions. Each county typically describes the origins of local party activity in its chronology and CCP history sections. Because these gazetteers were published after the CCP had come to power, recording such information did not entail political risk; it was also a way for counties to proclaim their revolutionary tradition.

First, each county usually reports both the first CCP member who appeared within the county and the first CCP member native to the county. If the first CCP member who appeared in the county was not a local native, we record the member's name,[19] arrival time, place of origin, sending organization, and identity. If the member was native to the county, the record often contains richer biographical information. We therefore record the location where the member joined the party, the organization through which the member joined, the position held at the time of joining, the date of joining, and the date of return to the home county. At a minimum, this procedure gives us the timing of the first CCP member's appearance in each county and that member's migration path. Second, we record the founding date of the first party organization reported in the county, which was often a branch, small group, party fraction, work team, or dispatched organ. Third, we record the founding date of the county party committee. Information on county party committees is richer, allowing us to trace subsequent episodes of reestablishment and reconstruction. Using a county gazetteer as an example, Figure 3 illustrates the search procedure.

Our validation sources include: (1) 503 provincial or municipal gazetteers, especially their party-government volumes; (2) 774 organizational histories at the provincial, municipal, or county level, which were often published separately from local gazetteers; and (3) Wikipedia or Baidu Baike entries, mainly for information on party members. The first CCP members appearing in a county often later became martyrs or senior CCP figures, making their biographical information relatively easy to obtain and cross-check. Whenever information conflicts across two sources, or within a single source, we drop the entire observation.

For administrative matching, because most local gazetteers were published around 1990, we use 1990 administrative boundaries as the matching benchmark. To facilitate matching with other datasets, we rely on several sources for administrative boundaries and

[19] Some individuals used aliases, but we treat a record as valid only when a specific name is reported. The same rule applies below.

changes across different periods: (1) the 1911 and 1990 administrative division projects from CHGIS V6 (CHGIS, 2016); (2) administrative division changes during 1912–1949 (Zheng, 2000); and (3) post-1949 administrative division histories released by the CCP.

Under our strict screening strategy and after matching to administrative units, 1,747 counties contain information on the first CCP member, 2,330 counties contain information on the first party organization, and 2,283 counties contain information on the county party committee. Appendix Figure A2 reports the spatial coverage of the CCP data, and Figure 1 reports annual development patterns.

III.B. May Thirtieth Movement

Our data on the May Thirtieth Movement mainly come from *Historical Materials on the May Thirtieth Movement* (Institute of History, Shanghai Academy of Social Sciences, 1981, 1986, 2005), especially the third volume (Institute of History, Shanghai Academy of Social Sciences, 2005), which provides the most detailed account of movement activity across China. The three-volume collection was compiled over several decades, contains 2.12 million Chinese characters, and brings together archival materials, diplomatic documents, firsthand accounts, oral histories, and meeting records. It is the most comprehensive source on the May Thirtieth Movement available to us.

Based on the descriptions in the collection, we define a locality as participating in the May Thirtieth Movement if it experienced the Three Strikes (workers' strikes, student class boycotts, and merchants' market closures), demonstrations, public assemblies, or public expressions of support. Because our analysis is conducted mainly at the county level, the historical materials allow us to identify 384 counties with May Thirtieth activity, which we refer to as the strict definition. Many records, however, are not reported at the county level. For example, some entries only describe large-scale mobilization in Nanjing, without specifying which counties under the city participated. We therefore construct a broader definition for such cases, under which 877 counties are coded as participating in the movement. Figure 5 reports the spatial coverage of May Thirtieth exposure.

The comparison between the two coding rules also provides a natural robustness check. In most specifications, if the estimates are statistically significant, we expect the absolute value of the coefficient under the strict definition to be larger than that under the broader definition.

To assess geographic precision, we use local gazetteers, a data source independent of *Historical Materials on the May Thirtieth Movement*, and collect the frequency of social-movement-related terms from chronicle entries.[20] Figure 6 reports the results. We find three patterns. (1) In May Thirtieth areas, the frequency of these terms in chronicle entries rose sharply in 1925. (2) Outside 1925, later years show no comparable effect, which helps rule out the influence of other possible social movements. (3) At the monthly

[20]The terms are *march, demonstration, assembly, petition, merchant strike, labor strike, student strike, and protest.*

level, movement-related terms began to appear in May, consistent with the outbreak of the movement in Shanghai at the end of May, and almost disappeared by July, consistent with subsequent repression by multiple authorities.

III.C. Additional Data

Books. We collect 427,519 books published in or before 1949 from the Republican-Era Book Database. The database contains each book's title, author, editor, translator, publisher, publication place and year, keywords, abstract, and classification code. After geocoding publication places, 330,402 books can be matched to city-level administrative units and 10,594 to county-level administrative units. We therefore conduct this part of the analysis mainly at the city level. Using information on book classification, abstracts, keywords, and publishers, we construct a "red-book" measure. Under the strict definition, 17,296 books are related to socialism or communism. Under the broader definition, which also includes left-wing literature and themes related to red revolution, the number rises to 23,205.[21]

Martyrs. We extract information on 340,883 martyrs from the martyr sections of local gazetteers, including native place, place of death, event of death, CCP membership status, and date of party entry. Appendix Figure A3 gives examples of the book-publication and martyr source records.

Republican-era Biographies. We extract native place, party affiliation, and detailed career histories for 15,995 individuals from the *Minguo biographical dictionary* (Xu, 2007).

Timing of CCP Capture. Xie et al. (1986) provides the date on which the CCP captured each county during the full-scale KMT–CCP civil war.

Northern Expedition. We identify counties traversed by the Northern Expedition beginning in 1926 from volumes 12 to 21 of *Revolutionary Documents* (Party History Materials Compilation Committee, Central Executive Committee of the Kuomintang, 1956–1959).

Japanese Occupation. We can readily obtain the timing of Japanese military entry into and withdrawal from each county from the regime, war, and chronology sections of local gazetteers.

Agricultural and Industrial Output. We collect annual grain output and industrial output from county gazetteers. Grain output is standardized to 10,000 jin, and industrial output is standardized to constant 1980 prices, which are the units most commonly used in county gazetteers.

Control Variables. We collect the following control variables: **population census** (county-level population data come from the 1928 Republican population census (Statistics Department, Ministry of Interior, Republic of China, 1931), matched to 1,058

[21] See Appendix Section A for details.

counties); **firm directory** (Du (1991) provides information on large firms established in or before 1927, although most firms can only be matched to the city level); **middle schools** (Department of General Education, Ministry of Education (1935) provides a directory of Republican-era middle schools in operation in 1934); **Christian institutions** (Mayfield et al. (2024) provides data on Christian institutions established in or before 1950); **railways** (we manually vectorize railway maps provided by Wu (1999)); **river networks** (river network data come from Natural Earth); **agricultural endowments** (Galor and Özak (2016) provides fine-grained raster data on post-1500 agricultural endowments, measured in calories); **population and pasture land** (data come from the 1920 raster data released by HYDE (Klein Goldewijk et al., 2017); although their spatial precision is limited, they provide broad coverage); and **precipitation** (data come from Jones and Harris (2013); annual precipitation data cover 1921–1949; a county-year is classified as anomalous rainfall when precipitation exceeds the county-specific mean by more than one standard deviation).

III.D. Determinants of May Thirtieth Movement Coverage

As Table 1 shows, counties affected by the May Thirtieth Movement had several distinctive characteristics. First, they were more likely to host large firms, consistent with the movement's initial focus on workers' rights. Second, they had greater agricultural and pastoral endowments, consistent with the movement's broader appeal to the rights and welfare of ordinary people during its diffusion. Third, they were closer to railways, rivers, and coastlines, suggesting that transport accessibility was crucial for the spread of a movement that remained active for only about two months. Fourth, they were more likely to have schools, consistent with the movement's reliance on student speeches and class strikes. Fifth, they were more likely to have Christian institutions, suggesting that globally circulating liberal ideas may have played a role. Sixth, counties closer to provincial capitals or with denser populations were more likely to experience the movement, possibly reflecting the strategic selection of places where popular emotions could be more effectively mobilized.

In the subsequent empirical analysis, the variables reported in Table 1 serve as potential controls. Following Belloni et al. (2014), we use post-double-selection to select covariates from a battery of baseline variables. This approach reduces overfitting and provides a principled procedure for covariate selection. Except in exercises designed to assess the validity of the results, we exclude two variables based on the 1928 population census, because they cover only a limited number of counties; for example, the employment rate is available for only 562 counties. We also exclude variables related to Christian institutions and middle schools, because both have missing establishment years to varying degrees, making it difficult to construct pre-treatment measures.

IV. Effects of the May Thirtieth Movement

IV.A. Stylized Facts

Before 1925, China already had widespread traces of CCP member activity: 20% of counties had evidence of such activity (Appendix Figure A4 and Figure 1(b)). Yet the number of party organizations was strikingly disproportionate. Only 5% of counties had a party organization, and only 15% established one in the same year that the first CCP member arrived (Appendix Table A1, Panel A, column (2)). The arrival of a CCP member did not predict the establishment of a local party organization even in the following year (Appendix Table A1, Panel A, column (4)).

After 1925, this pattern changed sharply. When a CCP member arrived in a county, the probability that a party organization was established in the same year exceeded 30% (Appendix Table A1, Panel B, column (2)), and the probability of establishment in the following year exceeded 22% (Appendix Table A1, Panel B, column (4)). In other words, after 1925, the arrival of a CCP member was associated with a greater than 50% probability that the county would establish a party organization within two years. Another striking pattern is that 7.6% of counties established a new party organization in 1925 alone, exceeding the entire pre-1925 stock of counties with party organizations. In 1926, nearly another 16% of counties established new party organizations. The number of county party committees also surged after 1925. Counties with a first CCP member likewise peaked in 1925, but this increase largely continued the pre-1925 trend and quickly declined after reaching its peak. Consistent with these patterns, CCP membership increased tenfold in 1925 (Appendix Figure A5).

Appendix Table A2 sharpens this result by restricting the sample to counties that already had CCP member activity before 1925 but had no party organization or county party committee before 1925. Under the strict May Thirtieth definition, these counties experienced significantly faster organizational conversion: with PDS controls and province fixed effects, the annual first-establishment probability increased by 13.9% for party organizations and 7.9% for county party committees (Panel B, column (3)).

IV.B. The May Thirtieth Movement and CCP Development

IV.B.1. Identification Strategy

We estimate the effects of the May Thirtieth Movement on CCP development using the following event-study specification:

$$Y_{ct}^{CCP} = \sum_{\substack{\tau=1921 \\ \tau\neq 1924}}^{1949} \beta_\tau \left(MT_c \times \mathbf{1}\{t=\tau\}\right) + \mu_c + \lambda_{p(c)t} + \mathbf{X}_c'\boldsymbol{\theta}_t + \varepsilon_{ct}. \tag{1}$$

Here, Y_{ct}^{CCP} denotes a CCP development outcome in county c and year t. In the baseline regressions, we define $Y_{ct}^{CCP} = \mathbf{1}$[the first year in which county c had a CCP member, a party organization, or a county party committee $\leq t$], the cumulative analogue of a variable measuring whether party development had ever appeared in county c by year t.

We discuss the potential biases under this specification later.

MT_c indicates exposure to the May Thirtieth Movement. We use two definitions. The strict definition includes only counties for which we can clearly verify that the movement occurred within county c. The broader definition includes all counties under the strict definition, as well as counties located in cities where May Thirtieth activity was reported, even when the source does not specify which subordinate counties participated. Under the strict definition, we exclude observations that appear only under the broader definition but cannot be verified at the county level. In the regressions, we report estimates under both definitions and examine whether the absolute value of the strict-definition coefficient is systematically larger than that of the broader-definition coefficient, which is itself informative about measurement precision.

The omitted comparison year is 1924, so each β_τ measures the difference between exposed and unexposed counties in year τ relative to that baseline year. In estimating the fixed-effects specification, we include county fixed effects, μ_c, to absorb time-invariant county characteristics. Baseline controls are interacted with year fixed effects, $\mathbf{X}_c'\boldsymbol{\theta}_t$, to allow their effects to vary over time. We also include province-by-year fixed effects, $\lambda_{p(c)t}$, so that identification comes from comparisons within the same province and year. This is a stringent specification. As shown below, the May Thirtieth Movement generated spatial spillovers, so absorbing province-year variation likely yields a conservative estimate.

IV.B.2. Threats to Identification Validity

Cumulative Outcome Variable. This paper primarily focuses on whether a county *ever* experienced CCP development. Even so, we provide three pieces of evidence showing that, at least apart from county committees, once a county had CCP members or party organizations, this status was difficult to reverse: (1) In our data collection, we rarely encountered accounts documenting a transition from the presence of a party organization to its complete absence. (2) More importantly, once the CCP took root in a locality, it almost never effectively withdrew from that place. Localities capable of establishing CCP organizations were themselves places where the CCP could more easily take root and expand (Schoenhals, 1999). Even after KMT purges, larger existing party organizations quickly split into smaller but still operational units (Stranahan, 1998) and were rapidly rebuilt (Sullivan, 1985). (3) In the ten years after the arrival of the first party member, the establishment of the first party organization, and the subsequent establishment of the county committee, the number of martyrs increased persistently and substantially (Figure A6).[22] Even when county committees were revoked or rebuilt, the decline in the number of martyrs after revocation and the increase after rebuilding were much smaller than the initial increase at establishment (Appendix Figure A7(c) and Appendix Figure A7(d)). This suggests that, at a minimum, the coverage of CCP activity and party

[22]During the sample period, among martyrs whose causes of death can be verified, 73% died in internal wars or incidents.

organizations was difficult to eradicate completely. For county committees, although their higher organizational level made them more vulnerable to purges, we find that the patterns of revocation and rebuilding were highly similar between May Thirtieth and non-May Thirtieth areas (Table A3). In the baseline regressions, we therefore do not make additional adjustments to the county-committee outcome, so that all coefficients remain comparable. The rich dynamics of CCP county committee establishment, revocation, and rebuilding provide the variation we exploit in the subsequent analysis.

Biases from the Cumulative Outcome Variable. Although absorbing outcome variables are common in applied work (Cantoni, 2012; Jácome, 2020; Clay et al., 2021; Chambru et al., 2024), they may generate bias in DID settings. The reason is straightforward: (1) DID compares changes in mean outcomes, and (2) even in the absence of any intervention, the mean outcome in both the treated and control groups mechanically converges to one. This implies that if group A had a higher pre-intervention share of counties with outcome equal to one than group B, group A would mechanically exhibit less subsequent growth, generating a negative estimated effect even without any treatment. In our context, as Figure 7 shows, the treated group had significantly higher coverage of party organizations. This mechanical force therefore biases our estimates downward. Although such attenuation is conservative for our main interpretation, we further address this concern in the subsequent analysis. Motivated by Deaner and Ku (2024), we construct a simple counterfactual based on pre-May Thirtieth trends. We also provide an alternative estimation using a hazard model.

Stable Unit Treatment Value Assumption (SUTVA). The evaluation of a natural experiment requires that control units remain unaffected by the treatment. In our setting, this means that counties without May Thirtieth activity should not be influenced by the movement. This assumption may be easily violated. For example, once counties exposed to the May Thirtieth Movement developed organized institutions, they may have exported revolutionary activity to surrounding areas. If so, especially given our inclusion of province-by-year fixed effects, the baseline estimates would likely be substantially attenuated. Below, we address this concern by excluding counties adjacent to May Thirtieth counties. At the same time, the spillover effects of the May Thirtieth Movement are themselves an important question, which we discuss later.

Comparability of Exposed and Unexposed Counties. The occurrence of the May Thirtieth Movement was not fully random. However, the evaluation of a natural experiment does not require random assignment of treatment and control units (Heckman et al., 1998a); it is primarily concerned with whether the two groups would have followed similar trends in the absence of treatment. We implement a relative-magnitude restriction to assess how sensitive the baseline results are to violations of this assumption. We then use two complementary strategies to strengthen the argument. First, still within a difference-in-differences framework, we apply the synthetic difference-in-differences estimator (Arkhangelsky et al., 2021), which reweights units and matches pre-May Thirtieth

trends, improving pre-treatment comparability between treated and control counties and thereby increasing the credibility of post-treatment estimates. As an additional specification, we use the changes-in-changes estimator (Athey and Imbens, 2006), which nonparametrically recovers the full counterfactual outcome distribution and estimates quantile treatment effects by comparing this distribution with the treated group's actual second-period distribution. This method imposes no functional-form assumption on the outcome distribution, allows treated and control groups to differ arbitrarily in the distribution of unobservables as long as these distributions are time-invariant within each group, and accommodates systematic heterogeneity in treatment effects. We also separately control for interactions between year fixed effects and indicators for whether a county had CCP member activity or a party organization in 1924.

We additionally implement a spatial regression discontinuity design. Based on the contiguous geographic distribution of May Thirtieth counties, the design partitions counties nationwide into several contiguous May Thirtieth regions and non-May Thirtieth areas.[23] The analysis then compares CCP development in adjacent counties on opposite sides of the boundary, which differed in their exposure to the May Thirtieth Movement. We focus on 1927, when CCP development had reached an interim peak, while county party committees remained rare in 1925. Our outcomes are the changes in each CCP development indicator between 1924 and 1927.

Potentially Confounding Major Events. (1) Northern Expedition. The Northern Expedition, launched in 1926 under broad KMT–CCP cooperation, is an important event that may affect the validity of our results. Existing scholarship emphasizes the positive role of this wide-ranging military campaign in the expansion of the CCP (van de Ven, 1992). We compile detailed information on the geographic coverage of the Northern Expedition from *Revolutionary Documents* (Party History Materials Compilation Committee, Central Executive Committee of the Kuomintang, 1956–1959) to rule out this competing explanation. (2) May Fourth Movement. The May Fourth Movement was a large-scale student movement that occurred in 1919, before the founding of the CCP, and centered on demands for Chinese sovereignty. We proxy for exposure to the movement using the frequency of protest-related terms in local gazetteers in 1919 and a list, compiled by the Institute of Modern History, CASS (2013), of cities (rather than counties) where the May Fourth Movement explicitly occurred. (3) Second Sino-Japanese War. Although full-scale war between China and Japan broke out in 1937, Japan had already occupied Manchuria beginning in 1931. We collect the timing of Japan's first entry into and final withdrawal from each county from the foreign-relations, war, and related sections of local gazetteers.

[23] See Appendix Figure A25 for details on how these boundaries are defined. The general strategy of constructing a contiguous region around dispersed event locations is motivated by König et al. (2017); our implementation uses a concave hull. Subsequent studies have applied this approach to regression discontinuity designs, such as Asmus and Franck (2022).

IV.B.3. Effects of the May Thirtieth Movement on CCP Development

Figure 7 reports the direct effects of the May Thirtieth Movement on whether a county had CCP member activity, a party organization, and a county party committee. For simplicity, we mainly report the coefficients based on the strict definition of May Thirtieth exposure; the coefficients under the broader definition are systematically smaller. Our preferred estimates control for province-by-year fixed effects, county fixed effects, and covariates. We also report, in parentheses, estimates with only county and year fixed effects, since the preferred specification is stringent in some respects.

As Figure 7 shows, over 1925–1949, May Thirtieth counties were not significantly more likely to have CCP member activity under the estimating strategy in equation (1). By contrast, they were 8.5% more likely to have a party organization (25.7%) and 22.1% more likely to have a county party committee (34.3%). Turning to the peak annual effects, relative to 1924, May Thirtieth counties were 7.9% more likely to have CCP member activity in 1926 (6.2%), although this estimate does not fully rule out the influence of pre-trends. The probability of having a party organization increased by an additional 17.1% in May Thirtieth counties in 1928 (18.9%), while the probability of having a county party committee increased by an additional 28.2% in 1930 (29.6%). Overall, we find that the May Thirtieth Movement had no clear effect on the geographic scope of CCP member activity, or that any such effect was substantially weaker than its effect on the CCP's organizational consolidation. In addition, counties with pre-1925 CCP member activity completed CCP organizational consolidation more rapidly after the May Thirtieth Movement (Appendix Table A2).

We reestimate the baseline regression using a hazard model (Figure 8), which allows us to trace the time-varying dynamics of CCP development. The estimates show that, in most years, May Thirtieth areas exhibited higher annual rates of CCP development.

Using information on the establishment of Republican-era political parties and associations (Wang and Yang, 1989), we find no similar increase in May Thirtieth areas (Appendix Table A4). Instead, the estimated coefficient turns negative in 1926, possibly because these organizations' competitive position vis-a-vis the CCP declined.

Robustness checks for party-development effects. (i) DID bias under cumulative outcome variables. Because all CCP development indicators in May Thirtieth areas were no lower than those in non-May Thirtieth areas before the movement, the mechanical convergence of the treated and control means toward one can only lead us to underestimate the effect of the May Thirtieth Movement. Even so, following the setup in Deaner and Ku (2024), we construct a simple counterfactual based on pre-movement changes in the outcome variables (Appendix Figure A27).

(ii) SUTVA. Appendix Figure A28 excludes counties adjacent to May Thirtieth counties. The estimated coefficients increase substantially, suggesting that the May Thirtieth Movement may have generated positive spillovers to neighboring counties. The coefficient

for CCP member activity also becomes significant, indicating that the movement likely increased the geographic coverage of CCP activity, although this effect did not clearly exceed that in adjacent counties in the baseline specification.

(iii) Comparability between treated and control counties. We first report sensitivity analyses of the parallel-trends assumption (Appendix Figure A29). We find that only the result for CCP member coverage becomes statistically insignificant at the 5% level when the permitted violation of parallel trends is no larger than 50% of the pre-May Thirtieth violation. Second, applying the synthetic difference-in-differences estimator, which relies less heavily on the parallel-trends assumption, produces larger coefficients than the baseline regression (Appendix Figure A30): the estimated effects are 12% for CCP member coverage, 26% for party organization coverage, and 34% for county party committee coverage. Applying the changes-in-changes estimator (Athey and Imbens, 2006), which further accommodates incomparability between treated and control counties, yields a consistent pattern of estimates (Appendix Figure A31). We also separately control for interactions between year fixed effects and indicators for whether a county had CCP member activity or a party organization in 1924.

We further construct a boundary around May Thirtieth areas and implement a spatial regression discontinuity design, comparing CCP development in geographically adjacent May Thirtieth and non-May Thirtieth counties. Figure 9 reports the results. Party organizations and county party committees exhibit clear discontinuous increases at the boundary, while the presence of CCP members also rises to some extent. Appendix Figure A26 examines the sensitivity of these results to alternative boundary definitions.

(iv) Major historical events. For the Northern Expedition, we compile the locations and timing of battles during 1926–1928. In addition, using the shortest path between temporally adjacent battles, we infer the counties likely traversed by Northern Expedition forces. We combine these inferred transit counties with battle counties to construct a broader measure of Northern Expedition exposure. The spatial overlap between the Northern Expedition and the May Thirtieth Movement is limited: the correlation is 0.17 under the strict definitions of both events and 0.25 under the broader definitions. Appendix Figure A33 reports estimates controlling for Northern Expedition exposure.

In addition, Appendix Figure A34 reports estimates of the effect of the Northern Expedition in areas without May Thirtieth exposure. When we control only for county and year fixed effects, most outcomes are economically large and statistically significant, consistent with historians' qualitative accounts of the Northern Expedition's role in the expansion of CCP organizations. Once comparisons are restricted to within-province variation, however, all estimates shrink substantially and become markedly smaller than the estimated effects of the May Thirtieth Movement.

For the role of Japanese occupation, using the timing of Japanese entry into and withdrawal from each county, we do not simply add a dummy for whether a county was under Japanese occupation. Instead, we saturate the specification with interactions

between province-by-year fixed effects and indicators for the county's Japanese-entry and Japanese-withdrawal cohorts. This allows for rich dynamic effects, including the possibility that counties occupied by Japan may have already been on different trajectories before occupation. The results are unaffected (Appendix Figure A35). In some outcomes, the estimated coefficients after 1931 shift slightly upward, possibly because Japan began its gradual advance into Manchuria and North China in 1931, where Japanese military activity was negatively correlated with May Thirtieth exposure.

For the May Fourth Movement, we construct two datasets. First, following Figure 6, we measure the frequency of protest-related terms in each county in 1919. Although the May Fourth Movement was also large in scale, its geographic overlap with May Thirtieth exposure was limited (Appendix Figure A36(a)): the overlap between counties with positive protest-keyword frequencies in 1919 and 1925 was only 36%. Second, we compile from Institute of Modern History, Chinese Academy of Social Sciences, Modern History Materials Editorial Office (2013) a list of 60 cities where May Fourth activities were explicitly documented (Appendix Figure A36(b)). To rule out confounding from the May Fourth Movement, we conduct three tests: (1) excluding May Fourth counties or cities identified by the two measures; (2) interacting each May Fourth measure with year fixed effects; and (3) using May Fourth cities as placebo locations for the May Thirtieth Movement. Controlling for May Fourth exposure leaves our baseline results essentially unchanged, while the placebo test yields null effects (Appendix Figures A36(c)–A36(h)).

(v) Placebo. One concern is that the observed CCP development may simply reflect post-1925 growth in counties surrounding large cities such as Shanghai. We select seven major cities, including Shanghai, where the May Thirtieth Movement attracted substantial attention, and conduct a set of distance-based placebo exercises. In Appendix Figure A37, panel (a), for each city, we calculate the distance from each county to that city and assign the nearest counties to the placebo treatment group, with the number of treated counties determined by the share of strict May Thirtieth counties among all counties. In panel (b), around each city, the nearest counties are assigned to the placebo treatment group in proportion to the actual 1924 shares of different CCP development outcomes in May Thirtieth counties. In panel (c), we repeat the exercise in panel (a) after excluding the original May Thirtieth counties. In panel (d), based on the province-specific incidence of May Thirtieth exposure, counties closer to these cities are assigned within each province to the placebo treatment group. Across all specifications, no estimates are comparable in magnitude to the baseline coefficients. In addition, we drop each province one at a time to ensure that the results are not driven by any single province (Appendix Figure A38), and randomly reassign May Thirtieth exposure across counties as another placebo exercise (Appendix Table A5).

Intensive-margin effects on party development. Using information on 2,097 individuals in the *Minguo biographical dictionary* who had ever been CCP members, we find that the probability of joining the CCP increased significantly in May Thirtieth

counties in 1925 (Appendix Table A6). In column (3) of Appendix Table A6, for example, CCP entry in 1925 increased by 22%.

Appendix Table A7 further classifies individuals by their occupation in 1925, or by their pre-entry occupation if they joined the CCP in 1925. The increase in CCP entry was most pronounced among students. Peasants, soldiers, and cultural workers also experienced significant increases. These patterns confirm that the May Thirtieth Movement mobilized groups beyond workers. By contrast, the probability of CCP entry among workers increased but was not statistically significant. The probability of CCP entry among merchants declined, although the estimate is not statistically significant, consistent with the movement's growing conflict with merchant interests.

One concern is that the *Minguo biographical dictionary* may have limited external validity. We therefore also examine party-entry years in the martyr rosters. Because each martyr's location and occupation in 1925 cannot be tracked, exposure is assigned based on native place. The results show a large increase in the probability of CCP entry among these more ordinary individuals during 1925–1927 (Appendix Table A8).

IV.B.4. Spillover Effects

A natural question is how the May Thirtieth Movement, which directly covered only several hundred counties, can explain the CCP's rapid expansion after 1925. We focus here on the movement's spillover effects. First, we define counties adjacent to May Thirtieth counties and second-order adjacent counties, namely counties adjacent to those adjacent counties. Including May Thirtieth counties themselves, these areas account for 39% of all counties; under the broader definition of May Thirtieth exposure, they account for 75%. Using counties beyond second-order adjacency as the control group, Table 2 reports CCP expansion after the May Thirtieth Movement in May Thirtieth counties, first-order adjacent counties, and second-order adjacent counties. We find clear spillover effects that decline with distance. Even when comparisons are restricted within provinces, second-order adjacent counties experienced a 6.1% increase in party organization coverage and a 5.0% increase in county party committee coverage relative to counties beyond second-order adjacency (Table 2, Panel A, columns (5) and (6)). The hazard-model estimates show a similar pattern (Table 2, Panel B). In addition, we extend the analysis to fourth-order neighboring counties and find that, for some CCP development outcomes and specifications, the movement's influence did not stop at the second order but remained detectable through the fourth order (Figure A39).

We then examine two channels of diffusion, focusing on counties adjacent to May Thirtieth counties. The first channel is the migration of CCP members. Appendix Figures A8a and A8c examine whether the arriving CCP member came from a May Thirtieth area. Continuing to compare adjacent and non-adjacent counties, we find that after the May Thirtieth Movement, adjacent counties were more likely to receive CCP members from May Thirtieth areas, helping fill local gaps in CCP activity. Appendix Figures A8b

and A8d replace the outcome with whether the arriving CCP member did not come from a May Thirtieth area, to rule out the possibility that the results merely reflect the extensive-margin expansion of CCP member activity.

We find that CCP member migration relied primarily on returning natives rather than purely outside entrants (Appendix Figure A9), suggesting that May Thirtieth areas became more receptive to the CCP. In terms of coefficient magnitudes, these returning CCP members were especially likely to come from student and educator backgrounds and from worker backgrounds, both of which were central constituencies of the movement (Appendix Figure A10).

The second channel is imitation in party-organization building. We classify party organizations into common types, including small groups or branches, work teams, party members, and centrally dispatched organs. For each adjacent county, we estimate outcomes indicating whether the county established a party organization of the same type as that in a neighboring May Thirtieth area (Appendix Figures A11a and A11c). These outcomes rose significantly during 1925–1927. By contrast, outcomes indicating that a county established a party organization of a different type declined, ruling out the possibility that the result is driven only by a general difference in party-organization formation.

V. How the May Thirtieth Movement Worked

This section builds on social-contract and coercive accounts of state formation (Service, 1978; Locke, 1988). In the absence of coercion, local residents join a party, or allow an emerging regime to operate locally, only when they expect the welfare gains it offers to exceed those promised by other parties or by the status quo. Otherwise, they have little reason to incur the associated political and other costs. This logic became especially relevant after 1927, when large-scale violence against the CCP made tolerating its local presence increasingly unsafe.

V.A. Ideology

Evidence from books. Directly measuring changes in local residents' beliefs or welfare expectations is difficult. We therefore use whether books published in May Thirtieth areas carried communist, socialist, or broader red ideological content as a proxy. Appendix Figure A12 shows that, beyond books classified under Marxism-Leninism and Mao Zedong Thought or politics, red propaganda was also widely present in art and literature (for example, poems praising communism and biographies of communist leaders) and in economics (for example, works arguing for the superiority of Marxist political economy). Appendix Figure A13 reports a word cloud based on the available text of red books. The most frequent terms are "politics," "Mao Zedong," "Leninism," "Marxism," and "Mao Zedong Thought."

Because locating book publication places at the county level is difficult, we geocode 17,296 communist and socialist books, and 23,205 books including broader red propaganda, to the city level.[24] To define May Thirtieth exposure, a city enters the treatment group if at least one county in that city experienced the movement.[25] Table 3 and Appendix Table A11 report the results. We highlight three findings. First, with the spread of the May Thirtieth Movement, red books grew rapidly under both definitions, and the number of published books reached a local peak in the second year after the movement. By contrast, book categories in Appendix Figure A12 with few or no red books show no comparable increase (Appendix Table A13).[26] Second, this increase disappeared during the intensive KMT purges against the CCP in 1927–1928. Third, after 1945, when the civil war between the CCP and the KMT resumed, the number of red books surged far beyond the levels observed in the first few years after the May Thirtieth Movement.

Appendix Figure A14 reports the growth of major book categories across different periods. The seemingly most orthodox category, Marxism-Leninism and Mao Zedong Thought, did not lead in every period. By contrast, literature grew substantially in 1929–1945, a period when other types of red books showed little increase.

Evidence from Names. Our second, more suggestive, piece of evidence comes from martyrs' names. Following Fouka and Serlin (2026), who argue that parents' naming choices may reflect class-related perceptions, we ask whether names became more associated with the working class rather than with the elite and commercial strata on which the KMT relied after the May Thirtieth Movement. We focus on martyrs' given names. For each name, its share among workers measures the extent to which it carries working-class connotations. We refer to this measure as the working-class score.[27] A central challenge is the absence of a suitable name-training dataset: China conducted no comprehensive population census before 1953, much less one that disclosed individual names. We therefore construct an alternative training set from the NLC Republican-era Books Database. The resulting source pool contains 176 randomly collected worker

[24]Illustrative examples include Friedrich Engels's *Socialism: Utopian and Scientific* (1938), a strict red book that explains the theoretical foundations and core principles of scientific socialism, and Kang Youwei's *The Book of Great Unity* (1935), a broad red book because it replaces private ownership and production for private profit with common ownership and collective production. Its vision of a classless economic order, in which "all under heaven is held in common," closely resembles utopian socialism, although Kang developed it outside Marxist theory. Both titles were reissued in multiple editions.

[25]In Appendix Table A12, two additional tests are reported. (1) We retain as treated only cities mentioned in *Historical Materials on the May Thirtieth Movement* and exclude cities in which only a few counties participated in the movement. (2) We combine the strict and broad definitions of May Thirtieth exposure and use the share of participating counties within each city as a measure of treatment intensity.

[26]A few categories with too few books to obtain valid estimates are excluded.

[27]Under the algorithm described below, illustrative worker-associated names include 高壽山 (Gao Shoushan), with a score of 97.5, and 李福春 (Li Fuchun), with a score of 98.6. In 高壽山, the combined given name 壽山 alludes to the traditional wish for a life as enduring as the Southern Mountain. In 李福春, 福 means good fortune and 春 means spring. Illustrative business-elite-associated names include 車文軒 (Che Wenxuan), with a score of 3.5, and 陳思明 (Chen Siming), with a score of 2.2. In 車文軒, 文軒 suggests literary refinement; in 陳思明, 思 means reflection and 明 means clarity.

rosters, mainly lists of railway workers and union members, and 27 books containing entrepreneurs' names, mainly business directories that list managers and chamber-of-commerce members.[28] To ensure statistical power, the baseline exercise uses 100 worker rosters and 5 entrepreneur-name books. Together, these sources contain 41,628 worker names and 21,751 business-elite names. Matching these scores to martyrs with recorded birth years yields 33,851 matched martyrs. To assess sensitivity to sampling, we conduct 200 resampling exercises in which 15 worker rosters and 2 entrepreneur-name books are added to or removed from the baseline sample.

We then construct a county-year panel by averaging this score across martyrs by county and birth year, and apply the inverse hyperbolic sine transformation to deal with zero values. As shown in Appendix Table A15, although some specifications are not statistically significant, most coefficients exceed 0.1. This implies that the use of names with stronger working-class connotations increased by more than 10% after 1925. Unlike books, whose publication declined sharply several years after the movement under KMT repression, the naming evidence remains economically meaningful in most periods, consistent with the idea that naming choices were less subject to censorship.

V.B. Coercive Capacity

A competing explanation is that, after broad mobilization in May Thirtieth areas, the CCP may have developed stronger military capacity, which in turn secured its organizational expansion or increased party membership through coercion. To examine this possibility, we collect information on the presence of KMT and CCP garrisons from the military and garrison sections of local gazetteers and organizational histories. We also code the timing of subsequent withdrawals, returns, and new arrivals, thereby constructing a county-level panel of garrison presence. Figure 10 yields two findings. First, CCP garrisons did not increase visibly in May Thirtieth areas. Second, KMT garrisons increased sharply after 1927. As shown in Figure 10(c), garrison coverage rose by about 10 percentage points in 1927, consistent with intensified repression of the rapidly expanding CCP after 1927.

V.C. Weaker Purges

One major reason Chiang Kai-shek decided to launch the purge was the class struggle and radical worker-peasant mobilization frequently initiated by the CCP (Tsui, 2018). These demands fundamentally conflicted with the interests of local elites and other groups on which the KMT relied. A possible explanation for faster CCP development in May Thirtieth areas is therefore that the KMT found it harder to carry out purges in these places because of their stronger potential mass base.

We first examine the party-cleansing campaign launched by the KMT in 1927. As

[28] Appendix Table A14 lists examples of both types of source books. The table also shows that entrepreneur-name books contain more individual names per book, on average, than worker rosters.

Appendix Figure A15 shows, among counties with a county committee in 1927, martyr deaths increased by nearly 50% in 1927–1928. We use the establishment and revocation of county committees as a revealed measure of exposure to purges. County committees coordinated party organizations within a region and directly received assignments from higher-level authorities, making them much harder to operate clandestinely than ordinary party organizations (Stranahan, 1998). Panel A of Appendix Table A3 shows that, although more than half of county committees existing in 1927 were revoked within the next two years, the probability of revocation did not differ significantly between May Thirtieth and non-May Thirtieth areas. Panels B and C show that, in most relative years, existing county committees had no clear difference in the probability of revocation over the next decade between the two types of areas. Revoked county committees also had no clear difference in the probability of reconstruction over the next decade. Appendix Figure A16 further shows that, after controlling for the presence of a county committee, martyr deaths around 1927 did not differ significantly between May Thirtieth and non-May Thirtieth areas. Taken together, the evidence does not support the view that May Thirtieth areas faced weaker purges, consistent with the fact that the KMT purge of the CCP was a centrally led and systematic campaign rather than one that depended mainly on local implementers (Fitzgerald, 1996).

As an additional discussion, we examine whether CCP organizations in May Thirtieth areas were more resilient. This discussion builds on the observation that, after county committees were purged and revoked, they often continued to operate through underground organizations or fragmented party branches. Appendix Figure A17 supports this interpretation. After county committees were revoked, martyr deaths changed little in May Thirtieth areas, while they declined sharply in non-May Thirtieth areas.

VI. The May Thirtieth Movement's Further Role in State Formation

VI.A. May Thirtieth Areas during the Civil War

In this section, we provide suggestive evidence on how the CCP development generated by the May Thirtieth Movement affected the full-scale civil war between the KMT and the CCP after 1945. Our first piece of evidence is that May Thirtieth areas were more likely to be captured earlier by the CCP. The analysis uses a dataset on county-level CCP capture dates (Xie et al., 1986). Our baseline analysis starts on December 31, 1945, the end of the year in which the Second Sino-Japanese War ended.[29] At that point, the KMT and the CCP had inherited dispersed local control from Japanese forces, and there were 18 CCP-held zones across the country.[30]

[29] We later show sensitivity to alternative baseline dates.

[30] A CCP-held zone is defined as a connected set of counties captured by the CCP that is not adjacent to another CCP-held county.

We then use the following algorithm to examine whether May Thirtieth exposure predicts which county the CCP captured next. The first step identifies the CCP-held zones. For each CCP-held zone, we identify its adjacent counties, denoted by n, and the number of adjacent May Thirtieth counties, denoted by m. Daily capture dates identify the next county captured adjacent to each CCP-held zone.[31] After a capture event, a new CCP-held zone formed or an existing one expanded, and the algorithm repeats the procedure. Appendix Figure A18 illustrates the procedure. For each event around each CCP-held zone, if May Thirtieth counties were not special, the probability that the next capture occurred in a May Thirtieth county should equal m/n. Equivalently, in the following expression,

$$\Pr(\text{May Thirtieth county selected}) = \frac{\theta m}{\theta m + (n - m)}. \tag{2}$$

the null hypothesis is $\theta = 1$. The parameter θ captures how much more likely a May Thirtieth county was to be selected relative to a non-May Thirtieth county. After weighting by the number of capture events, we find that θ is 1.61 for strictly defined May Thirtieth areas and 1.41 for broadly defined May Thirtieth areas. Thus, relative to non-May Thirtieth counties, these counties were 61% and 41% more likely to be captured next, respectively. Appendix Figure A19 reports sensitivity to alternative baseline dates.

Although this evidence shows that the May Thirtieth Movement affected the course of the civil war in the long run, it combines two channels. First, the CCP may have preferred to attack May Thirtieth counties earlier, reflecting path choice. Second, the CCP may have attacked different counties, but May Thirtieth counties were easier to capture. We provide additional evidence on these channels. Ideally, one would observe precise troop movements, which would allow a clean separation of the two mechanisms. Such movements, however, cannot be measured with sufficiently fine data.

To examine the path-choice hypothesis, we compute the actual paths through which initially disconnected CCP-held zones became connected. We then compare them with the minimum-distance paths that would have connected the same zones. Because CCP-held zones could expand during the connection process, the minimum path could change before two zones were connected. We therefore define an iterative minimum-distance path (Appendix Figure A20). Compared with this benchmark, actual paths were longer, but the share of strictly defined May Thirtieth counties doubled, from 13.3% to 28.6%, and the share of broadly defined May Thirtieth counties increased by 50%, from 33.3% to 50.0%. Under the assumption that other determinants of military routes were orthogonal to May Thirtieth exposure, this pattern suggests that the CCP was more likely to use May Thirtieth areas as marching routes.

We provide two pieces of evidence for why May Thirtieth areas shaped the geography of the war. First, as shown earlier, May Thirtieth areas had higher CCP organizational

[31] If the CCP captured more than one county adjacent to a CCP-held zone on the same day, we apply the same procedure to each event.

coverage, including in 1945 (Appendix Table A16, Panel A), and county-committee coverage was strongly negatively associated with the timing of CCP capture (Appendix Table A16, Panel B). We further show that, although May Thirtieth areas already had broad county-committee coverage by 1945 ($> 50\%$), they still experienced faster growth during the war (Appendix Figure A21). Second, May Thirtieth areas displayed stronger ideological intensity during the civil war, as reflected in book publications (Table 3 and Appendix Table A11). We further support this interpretation in Table 4, which shows that, among martyrs in May Thirtieth areas between 1945 and 1949, a larger share came from local native places.[32] This suggests that May Thirtieth areas helped generate coercive capacity through local mobilization, rather than relying entirely on external military attack.

VI.B. May Thirtieth Areas after PRC Founding

Before the founding of the PRC, May Thirtieth areas were more deeply exposed to red ideology and experienced faster and more extensive CCP organizational development. We therefore ask to what extent these ideological legacies continued to shape political behavior and beliefs after 1949.

First, we examine Mao-era far-left campaigns. The Great Leap Forward, which began nationwide in 1958, was a mass political and economic campaign led by Mao Zedong, centered on the pursuit of steel output to "surpass Britain and catch up with America." Its implementation relied on highly unscientific methods, including melting farm tools for steel and diverting labor away from grain production. The campaign generated long-lasting negative economic consequences. We find that, relative to other areas, May Thirtieth areas experienced a more than 50% larger surge in steel output in the peak year (Figure 11(c)) and a 20% decline in grain output in the peak year (Appendix Figure A23). During the subsequent famine in 1959–1961, births fell by 20%. Overall, May Thirtieth areas behaved more extremely in later far-left political campaigns during the Mao era.

Second, we examine rightward reforms after the Mao era through the timing of the Household Responsibility System (HRS), a major economic reform that began to spread after 1978, two years after Mao's death. The HRS emerged from clandestine household contracting by peasants trapped in hunger and poverty under the collective-era people's commune system, a practice that remained officially prohibited until its eventual legalization. We define two dates: when the first production team in a county adopted the HRS, and when more than 50% of production teams in a county had done so. We find that May Thirtieth areas began to adopt this policy later, by about 0.2 years (Table 5). As a placebo test, we use the county-government-promoted, relatively lenient one-child policy of the period and find no similar difference. Taken together, as China's political line shifted rightward after the Mao era, May Thirtieth areas lagged behind in policy

[32]For comparability, counties whose martyr records report only martyrs from the county itself are excluded.

adjustment.

For political beliefs in the twenty-first century, we use two representative social surveys: one conducted in 2018, covering 135 counties, and another conducted in 2013 and 2015, covering 153 counties. Together, they include 22,204 respondents. We find no systematic differences between May Thirtieth and other areas in demographic characteristics, income, or health (Appendix Figure A24). We document three patterns (Figure 12): (1) respondents in May Thirtieth areas are more likely to believe that the current government performs poorly and that individuals should not submit to the government; (2) they express stronger support for democracy; and (3) they display stronger anti-foreign attitudes. The last pattern may be related to the anti-imperialist sentiment mobilized during the May Thirtieth Movement. For their broader dissatisfaction with government, the intervening century contains too many confounding forces for us to distinguish whether it reflects rebellious sentiments rooted in the May Thirtieth period or distrust of government generated by long exposure to more radical political campaigns after the founding of the PRC.

VII. Concluding Remarks

Social movements remain powerful forces, even today (Cantoni et al., 2024). Set in China during this period of profound turbulence, we study a large-scale movement and show that it helped a weak party rise rapidly and ultimately shaped the process through which a new state was formed.

History provides a rich experimental setting that allows us to offer a coherent account of the movement and its consequences. We quantitatively characterize the movement's role in ideological change and highlight a path to state formation distinct from violence. This suggests that a central claim in theories of the origins of government, namely the social-contract view, remains useful for understanding the formation of modern states. In this framework, parties foster broad social cooperation by changing citizens' expectations about welfare under alternative rulers. We also study a less-examined form of diffusion in social movements: the diffusion of the movement's consequences, which offers a new perspective on how protests and related movements spread. The long time span of our setting further allows us to examine what happened after the movement helped shape a new state. Although the results are mostly negative, they raise a broader question: social legacies that help a revolutionary party complete state formation may later become the basis for implementing radical policies, allowing state capacity to serve high-risk ideological goals.

Our findings are suggestive, and at least three questions lie beyond the scope of this paper. (1) The role of modern technology in social movements. Mastrorocco and Teso (2023) find that railways reduced monitoring costs and alleviated principal-agent problems, thereby contributing to the formation of modern bureaucratic states. Our evidence

suggests that railways were also associated with the diffusion of the movement, and that this diffusion ultimately undermined the old regime. This makes it important to study how different information technologies and physical-transport technologies differentially empower protesters and repressors. (2) The evolution of movement organizations. During the May Thirtieth Movement, Shanghai, for example, established a headquarters organization for the movement. Yet this organizational experiment was ultimately unsuccessful: by August, it had been crushed by local gangs hired to attack it. This points to the value of future work tracing how organizations created during movements transform into party organizations. (3) How protests become movements. In collecting historical materials, we found that local declarations were often highly inflammatory and inspiring. This makes the power of language in protest an urgent question for future research.

References

Aidt, Toke S. and Raphaël Franck. Democratization under the threat of revolution: Evidence from the great reform act of 1832. *Econometrica*, 83(2):505–547, 2015.

Akerlof, George A. and Rachel E. Kranton. Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753, 2000.

Alesina, Alberto and Enrico Spolaore. On the number and size of nations. *The Quarterly Journal of Economics*, 112(4):1027–1056, 1997.

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. Economic integration and political disintegration. *American Economic Review*, 90(5):1276–1296, 2000.

Allen, Robert C., Mattia C. Bertazzini, and Leander Heldring. The economic origins of government. *American Economic Review*, 113(10):2507–2545, 2023.

Amenta, Edwin and Francesca Polletta. The cultural impacts of social movements. *Annual Review of Sociology*, 45:279–299, 2019. doi: 10.1146/annurev-soc-073018-022342.

Andrews, Kenneth T. *Freedom Is a Constant Struggle: The Mississippi Civil Rights Movement and Its Legacy*. University of Chicago Press, Chicago, 2004.

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference-in-differences. *American economic review*, 111(12):4088–4118, 2021.

Asmus, Gerda and Raphaël Franck. State capacity, national economic policies and local development: The russian state in the southern urals. CEPR Discussion Paper DP17103, Centre for Economic Policy Research, 2022.

Athey, Susan and Guido W Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, 2006.

Battaglini, Marco. Public protests and policy making. *The Quarterly Journal of Economics*, 132(1):485–549, 2017.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650, 2014.

Besley, Timothy and Torsten Persson. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4):1218–1244, 2009. doi: 10.1257/aer.99.4.1218.

Bursztyn, Leonardo, Davide Cantoni, David Y. Yang, Noam Yuchtman, and Y. Jane

Zhang. Persistent political engagement: Social interactions and the dynamics of protest movements. *American Economic Review: Insights*, 3(2):233–250, 2021.

Cantoni, Davide. Adopting a new religion: the case of protestantism in 16th century germany. *The Economic Journal*, 122(560):502–531, 2012.

Cantoni, Davide, Andrew Kao, David Y. Yang, and Noam Yuchtman. Protests. *Annual Review of Economics*, 16(1):519–543, 2024.

Casanueva, Annalí. Can chants in the street change politics' tune? evidence from the 15m movement in spain. *Evidence From the 15M Movement in Spain (April 22, 2021)*, 2021.

Centeno, Miguel Angel. *Blood and Debt: War and the Nation-State in Latin America*. Penn State Press, 2002.

Chambru, Cédric, Emeric Henry, and Benjamin Marx. The dynamic consequences of state building: Evidence from the french revolution. *American Economic Review*, 114 (11):3578–3622, 2024.

Chen, Da. Minguo shiwu nian guonei gongren bagong de fenxi [an analysis of domestic workers' strikes in 1926], 1926.

Chen, Duxiu. Gei dai jitao de yi feng xin [a letter to dai jitao]. *Xiangdao [The Guide]*, (129–130):1187; 1196–1197, September 1925. Written August 30, 1925; published in issues dated September 11 and September 18, 1925.

Chen, Duxiu. Gao quan dang tongzhi shu [letter to all party comrades]. In Ren, Jianshu, editor, *Chen Duxiu zhuzuo xuanbian: di si juan, 1926–1931 [Selected Writings of Chen Duxiu, Vol. 4, 1926–1931]*. Shanghai renmin chubanshe, Shanghai, 2009. ISBN 9787208081413. Written December 10, 1929; the text begins on p. 415 in this edition.

Chen, Duxiu. Appeal to all comrades of the chinese communist party (10 december 1929). In *Prophets Unarmed*, pages 397–416. Brill, 2015.

Chen, Ting and James Kai-Sing Kung. The rise of the chinese communist party. *Journal of the European Economic Association*, page jvag029, 2026.

Chesneaux, Jean. Egalitarian and utopian traditions in the east. *Diogenes*, 16(62):76–102, 1968.

CHGIS. CHGIS V6: China Historical Geographic Information System. China Historical GIS, Version 6, 2016. Published December 2016.

Chow, Tse-tsung. *The May fourth movement: Intellectual revolution in modern China*. Harvard University Press, 1960.

Clay, Karen, Akshaya Jha, Joshua A Lewis, and Edson R Severnini. Impacts of the clean air act on the power sector from 1938-1994: Anticipation and adaptation. Technical report, National Bureau of Economic Research, 2021.

Deaner, Ben and Hyejin Ku. Causal duration analysis with diff-in-diff. Technical report, Technical report, Working Paper, 2024.

Deng, Zhongxia. *Zhongguo zhigong yundong jianshi (1919–1926) [A Brief History of the Chinese Labor Movement, 1919–1926]*. 1930. Later republished by Renmin chubanshe.

Department of General Education, Ministry of Education. *Quanguo zhongdeng xuexiao xiaoming dizhi yilanbiao, Zhonghua Minguo ershisan niandu [National Directory of Secondary-School Names and Addresses, Republic of China Year 23]*. Commercial Press, Shanghai, 1935. First edition, December 1935. The directory is based on secondary-school survey reports for Republic of China Year 23, i.e. 1934.

Dincecco, Mark and Mauricio Prado. Warfare, fiscal capacity, and performance. *Journal of Economic Growth*, 17(3):171–203, 2012. doi: 10.1007/s10887-012-9079-4.

Ding, Wenjiang. Gaodiao yu zeren [high-sounding talk and responsibility]. *Chenbao*

[Morning Post], page 3, June 1925.

Downing, Brian M. *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. Princeton University Press, Princeton, NJ, 1992.

Du, Xuncheng. *Minzu zibenzhuyi yu jiu Zhongguo zhengfu, 1840–1937 [National Capitalism and the Old Chinese Government, 1840–1937]*. Shanghai Academy of Social Sciences Press, Shanghai, 1991.

Edelsbrunner, Herbert, David Kirkpatrick, and Raimund Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.

Engels, Friedrich. *Herr Eugen Dühring's Revolution in Science*. Martin Lawrence, London, 1934.

Enos, Ryan D., Aaron R. Kaufman, and Melissa L. Sands. Can violent protest change local policy support? evidence from the aftermath of the 1992 los angeles riot. *American Political Science Review*, 113(4):1012–1028, 2019.

Ertman, Thomas. *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. Cambridge University Press, Cambridge, 1997.

Fernández-Villaverde, Jesús, Mark Koyama, Youhong Lin, and Tuan-Hwee Sng. The fractured-land hypothesis. *The Quarterly Journal of Economics*, 138(2):1173–1231, 2023.

Fitzgerald, John. *Awakening China: Politics, Culture, and Class in the Nationalist Revolution*. Stanford University Press, Stanford, CA, 1996.

Fouka, Vasiliki and Theo Serlin. Migration and the making of the english middle class. Technical report, National Bureau of Economic Research, 2026.

Galor, Oded and Ömer Özak. The agricultural origins of time preference. *American Economic Review*, 106(10):3064–3103, 2016.

Gennaioli, Nicola and Ilia Rainer. The modern impact of precolonial centralization in africa. *Journal of Economic Growth*, 12(3):185–234, 2007. doi: 10.1007/s10887-007-9017-z.

Gethin, Amory and Vincent Pons. Social movements and public opinion in the united states. Technical report, National Bureau of Economic Research, 2024.

Gillion, Daniel Q. The influence of protest activity on congressional behavior: The scope of minority protests in the district. *Journal of Politics*, 74(4):950–962, 2012.

Giugni, Marco G. Was it worth the effort? the outcomes and consequences of social movements. *Annual Review of Sociology*, 24:371–393, 1998. doi: 10.1146/annurev.soc.24.1.371.

Harris, Lane Jeremy. *The Post office and state formation in modern China, 1896-1949*. University of Illinois at Urbana-Champaign, 2012.

Heckman, James J, Hidehiko Ichimura, Jeffrey A Smith, and Petra E Todd. Characterizing selection bias using experimental data, 1998a.

Heckman, James J, Hidehiko Ichimura, and Petra Todd. Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2):261–294, 1998b.

Hintze, Otto. *The Historical Essays of Otto Hintze*. Oxford University Press, New York, 1975.

Hobbes, Thomas and Hermann Klenner. *Leviathan*. Felix Meiner Verlag, 2005.

Holsti, Kalevi Jaakko. *The State, War, and the State of War*. Number 51. Cambridge University Press, Cambridge, 1996.

Hungerman, Daniel and Vivek Moorthy. Every day is earth day: Evidence on the long-term impact of environmental activism. *American Economic Journal: Applied Eco-*

nomics, 15(1):230–258, 2023.

Hutchinson, John. *Nationalism and War*. Oxford University Press, 2017.

Institute of History, Shanghai Academy of Social Sciences, editor. *Historical Materials on the May Thirtieth Movement, Volume 1*. Shanghai renmin chubanshe, Shanghai, November 1981.

Institute of History, Shanghai Academy of Social Sciences, editor. *Historical Materials on the May Thirtieth Movement, Volume 2*. Shanghai renmin chubanshe, Shanghai, August 1986.

Institute of History, Shanghai Academy of Social Sciences, editor. *Historical Materials on the May Thirtieth Movement, Volume 3*. Shanghai renmin chubanshe, Shanghai, December 2005.

Institute of Modern History, Chinese Academy of Social Sciences, Modern History Materials Editorial Office, editor. *Wusi aiguo yundong [The May Fourth Patriotic Movement]*. Modern History Materials Monograph Series. Intellectual Property Publishing House, Beijing, 2013. ISBN 978-7-5130-1697-1. 2 volumes.

Ip, Hung-Yok. The origins of chinese communism: A new interpretation. *Modern China*, 20(1):34–63, 1994.

Jackson, Robert H. *Quasi-States: Sovereignty, International Relations and the Third World*, volume 12. Cambridge University Press, Cambridge, 1990.

Jackson, Robert H. and Carl G. Rosberg. Why africa's weak states persist: The empirical and the juridical in statehood. *World Politics*, 35(1):1–24, 1982.

Jácome, Elisa. Mental health and criminal involvement: Evidence from losing medicaid eligibility. *Job Market Paper, Princeton University*, 2020.

Johnson, Chalmers A. *Peasant nationalism and communist power: the emergence of revolutionary China, 1937-1945*. Stanford University Press, 1962.

Jones, P. D. and I. Harris. Climatic research unit (cru) time series (ts) version 3.21 of high resolution gridded data of month-by-month variation of climate (jan. 1901–dec. 2012). *University of East Anglia and NCAS British Atmospheric Data Centre*, 2013.

Kaspersen, Lars Bo and Jeppe Strandsbjerg. *Does War Make States? Investigations of Charles Tilly's Historical Sociology*. Cambridge University Press, 2017.

Klein Goldewijk, Kees, Arthur Beusen, Jonathan Doelman, and Elke Stehfest. Anthropogenic land use estimates for the holocene–hyde 3.2. *Earth System Science Data*, 9 (2):927–953, 2017.

König, Michael D., Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. Networks in conflict: Theory and evidence from the great war of africa. *Econometrica*, 85(4): 1093–1132, 2017.

Krasner, Stephen D. *Sovereignty: Organized Hypocrisy*. Princeton University Press, Princeton, NJ, 1999.

Ku, Hung-Ting. Urban mass movement: the may thirtieth movement in shanghai. *Modern Asian Studies*, 13(2):197–216, 1979.

Liang, Qichao. Women gai zenmeyang yingfu shanghai cansha shijian [how should we respond to the shanghai massacre?]. *Chenbao [Morning Post]*, page 6, June 1925.

Locke, John. *The Second Treatise of Government: An Essay Concerning the True Original, Extent and End of Civil Government, and A Letter Concerning Toleration*. Library of Alexandria, 1966.

Locke, John. *Two Treatises of Government*. Cambridge Texts in the History of Political Thought. Cambridge University Press, Cambridge, 1988. Edited by Peter Laslett.

Ma, Yuping and Yuchong Huang, editors. *Zhongguo zuotian yu jintian: 1840–1987*

nian guoqing shouce [China Yesterday and Today: A Handbook of National Conditions, 1840–1987]. Jiefangjun chubanshe, Beijing, 1989. ISBN 7506510006.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. Do political protests matter? evidence from the tea party movement. *The Quarterly Journal of Economics*, 128(4):1633–1685, 2013.

Mao, Zedong. Shiju guliang he hongjun xingdong wenti [assessment of the current situation and the question of red army action]. In *Kangzhan yiqian xuanji [Selected Works before the War of Resistance]*. Zhonggong Zhongyang Beifangju, 1944. Written January 5, 1930; unabridged original version later retitled Xingxing zhi huo, keyi liaoyuan [A Single Spark Can Start a Prairie Fire].

Mastrorocco, Nicola and Edoardo Teso. State capacity as an organizational problem. *Working Paper*, 2023.

Mayfield, Alex, Margaret Frei, Daryl Ireland, and Eugenio Menegon. The china historical christian database: A dataset quantifying christianity in china from 1550 to 1950. *Data*, 9(6):76, 2024.

Mayoral, Laura and Ola Olsson. Pharaoh's cage: Environmental circumscription and appropriability in early state development. *Available at SSRN 3496225*, 2019.

Mayshar, Joram, Omer Moav, and Luigi Pascali. The origin of the state: Land productivity or appropriability? *Journal of Political Economy*, 130(4):1091–1144, 2022. doi: 10.1086/718372.

McAdam, Doug. The biographical consequences of activism. *American Sociological Review*, 54(5):744–760, 1989.

Meyer, John W., John Boli, George M. Thomas, and Francisco O. Ramirez. World society and the nation-state. *American Journal of Sociology*, 103(1):144–181, 1997.

Nord, Marina, David Altman, Fabio Angiolillo, Tiago Fernandes, Ana Good God, and Staffan I. Lindberg. 25 years of autocratization-democracy trumped? democracy report 2025. *Democracy Report*, 2025.

Olson, Mancur. Dictatorship, democracy, and development. *American Political Science Review*, 87(3):567–576, 1993.

Orchard, Dorothy J. China's use of the boycott as a political weapon. *The Annals of the American Academy of Political and Social Science*, 152(1):252–261, 1930.

Parker, Geoffrey. *The Military Revolution: Military Innovation and the Rise of the West, 1500–1800*. Cambridge University Press, Cambridge and New York, 1988.

Party History Materials Compilation Committee, Central Executive Committee of the Kuomintang, editor. *Geming wenxian: Guomin gemingjun chushi beifa shiliao, di 12–21 ji [Revolutionary Documents: Historical Materials on the National Revolutionary Army's Northern Expedition, Volumes 12–21]*. Party History Materials Compilation Committee, Central Executive Committee of the Kuomintang, Taipei, 1956–1959.

Perry, Elizabeth. *Rebels and Revolutionaries in North China*. Stanford University Press, Stanford, CA, 1980.

Perry, Elizabeth J. *Shanghai on strike: The politics of Chinese labor*. Stanford University Press, 1993.

Perry, Elizabeth J. Popular protest in shanghai, 1919–1927: Social networks, collective identities, and political parties. In Dillon, Nara and Jean C. Oi, editors, *At the Crossroads of Empires: Middlemen, Social Networks, and State-Building in Republican Shanghai*, pages 87–109. Stanford University Press, Stanford, CA, 2007.

Phillips, Richard T. *China since 1911*. Bloomsbury Publishing, 1996.

Philpott, Daniel. *Revolutions in Sovereignty: How Ideas Shaped Modern International*

Relations. Princeton University Press, Princeton, NJ, 2010.

Qu, Qiubai. Wusa zhong zichan jieji yu wuchan jieji zhi huzheng lingxiuquan [the contest for leadership between the bourgeoisie and the proletariat in the may thirtieth movement]. *Xin qingnian [New Youth]*, (3), March 1926. Section 2 of Guomin huiyi yu Wusa yundong [The National Assembly and the May Thirtieth Movement]; written January 22, 1926.

Rambachan, Ashesh and Jonathan Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.

Saich, Tony. *Finding Allies and Making Revolution: The Early Years of the Chinese Communist Party*. Brill, Leiden, 2020.

Sánchez de la Sierra, Raúl. On the origins of the state: Stationary bandits and taxation in eastern congo. *Journal of Political Economy*, 128(1):32–74, 2020.

Schoenhals, Michael. Political movements, change and stability: The chinese communist party in power. *The China Quarterly*, 159:595–605, 1999.

Scott, James C. *Against the Grain: A Deep History of the Earliest States*. Yale University Press, 2017.

Service, Elman R. Classical and modern theories of the origins of government. In Cohen, Ronald and Elman R. Service, editors, *Origins of the State: The Anthropology of Political Evolution*, pages 21–34. Institute for the Study of Human Issues, Philadelphia, 1978.

Skocpol, Theda. *States and Social Revolutions: A Comparative Analysis of France, Russia and China*. Cambridge University Press, Cambridge, 1979.

Smith, Stephen Anthony. *Like cattle and horses: nationalism and labor in Shanghai, 1895-1927*. Duke University Press, 2002.

Statistics Department, Ministry of Interior, Republic of China, editor. *Minguo shiqi nian ge sheng shi hukou diaocha tongji baogao [Statistical Report on the 1928 Household Registration Survey of Provinces and Municipalities]*. General Affairs Department, Ministry of Interior, Nanjing, 1931.

Stranahan, Patricia. *Underground: The Shanghai Communist Party and the Politics of Survival, 1927–1937*. Bloomsbury Publishing PLC, 1998.

Sullivan, Lawrence R. Reconstruction and rectification of the communist party in the shanghai underground: 1931–34. *The China Quarterly*, 101:78–97, 1985.

Tarrow, Sidney. *War, States, and Contention: A Comparative Historical Study*. Cornell University Press, Ithaca, NY, 2015.

Tilly, Charles. Reflections on the history of european state-making. In Tilly, Charles, editor, *The Formation of National States in Western Europe*, pages 3–83. Princeton University Press, Princeton, NJ, 1975.

Tilly, Charles. *From Mobilization to Revolution*. Addison-Wesley, Reading, MA, 1978.

Tilly, Charles. Social movements and national politics. In Bright, Charles and Susan Friend Harding, editors, *Statemaking and Social Movements: Essays in History and Theory*, pages 297–317. University of Michigan Press, Ann Arbor, 1984.

Tilly, Charles. War making and state making as organized crime. In Evans, Peter B., Dietrich Rueschemeyer, and Theda Skocpol, editors, *Bringing the State Back In*, pages 169–191. Cambridge University Press, Cambridge, 1985.

Tilly, Charles. *Coercion, Capital, and European States, AD 990–1992*. Blackwell, Oxford, 1992.

Tilly, Charles. *European Revolutions, 1492–1992*. Blackwell, Oxford, 1993.

Tsui, Brian. *China's Conservative Revolution: The Quest for a New Order, 1927–1949*.

Cambridge University Press, Cambridge, 2018.

van de Ven, Hans J. *From Friend to Comrade: The Founding of the Chinese Communist Party, 1920–1927*. University of California Press, Berkeley, 1992.

Wang, Jin and Jianghua Yang, editors. *Zhongguo Dangpai Shetuan Cidian [Dictionary of Chinese Political Parties and Associations]*. Zhonggong dangshi ziliao chubanshe, Beijing, 1989.

Wasow, Omar. Agenda seeding: How 1960s black protests moved elites, public opinion and voting. *American Political Science Review*, 114(3):638–659, 2020.

Womack, Brantly. Mao, the party, and the national revolution: 1923–1927. In *The Foundations of Mao Zedong's Political Thought, 1917–1935*. University of Hawai'i Press, 2018.

Wu, Yuexing, editor. *Zhongguo xiandai shi dituji: 1919–1949 [Atlas of Modern Chinese History, 1919–1949]*. Zhongguo ditu chubanshe, Beijing, 1999.

Xie, Dayuan, Jun Ma, and Yanming Lu. *Xianxue huanlai de fengbei: Quanguo ge shi, xian (qi) lici jiefang shijian ziliao huibian [A Monument Bought with Blood: A Compilation of CCP Capture Dates for Chinese Cities, Counties, and Banners]*. Shaanxi shifan daxue chubanshe, Xi'an, di 1 ban edition, 1986.

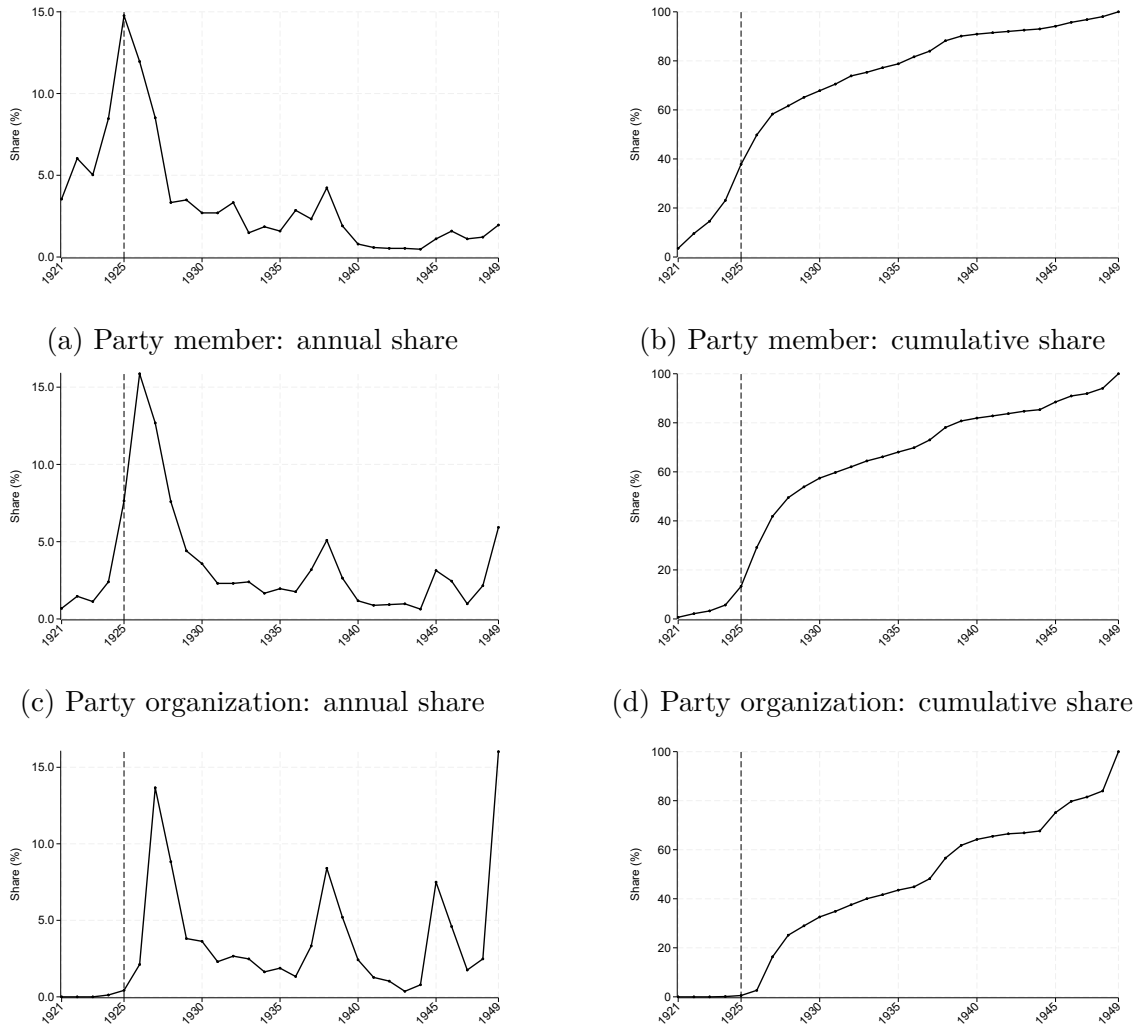
Xu, Youchun, editor. *Minguo renwu dacidian: zengding ben [Biographical Dictionary of Republican China: Revised Edition]*. Hebei People's Publishing House, Shijiazhuang, revised edition edition, 2007.

Young, Ernest P. Politics in the aftermath of revolution: the era of yuan shih-k'ai, 1912-1916. *The Cambridge History of China*, 12:1912–1949, 1980.

Zarrow, Peter. *China in war and revolution, 1895-1949*. Routledge, 2006.

Zhang, Guotao. Zai "wusa" yundong zhong [in the may thirtieth movement]. In *Wo de huiyi [My Memoirs]*, volume 2 of *Xiandai xijian shiliao shuxi*, pages 11–19. Dongfang chubanshe, Beijing, 1998. ISBN 7506009854. Di jiu pian, Gongyun yu fandi, di san zhang.

Zheng, Baoheng. *Minguo shiqi zhengqu yange [Evolution of Administrative Divisions during the Republican Period]*. Hubei Education Press, Wuhan, 2000. Covers the Republican period, 1912–1949.



(a) Party member: annual share

(b) Party member: cumulative share

(c) Party organization: annual share

(d) Party organization: cumulative share

(e) County committee: annual share

(f) County committee: cumulative share

Figure 1. Party Development Over Time: Annual and Cumulative Shares

Note: This figure reports annual and cumulative county shares for first CCP member activity, first party organization, and first county party committee. The left panels plot the first year in which the corresponding status appears, and the right panels plot whether the status had appeared by each year. For visual clarity, each outcome sample keeps counties that experienced the corresponding type of development before 1949. Back to Sections II.C., III.A., and IV.A..



(a) Chinese terms with readable English approximation



(b) English approximation

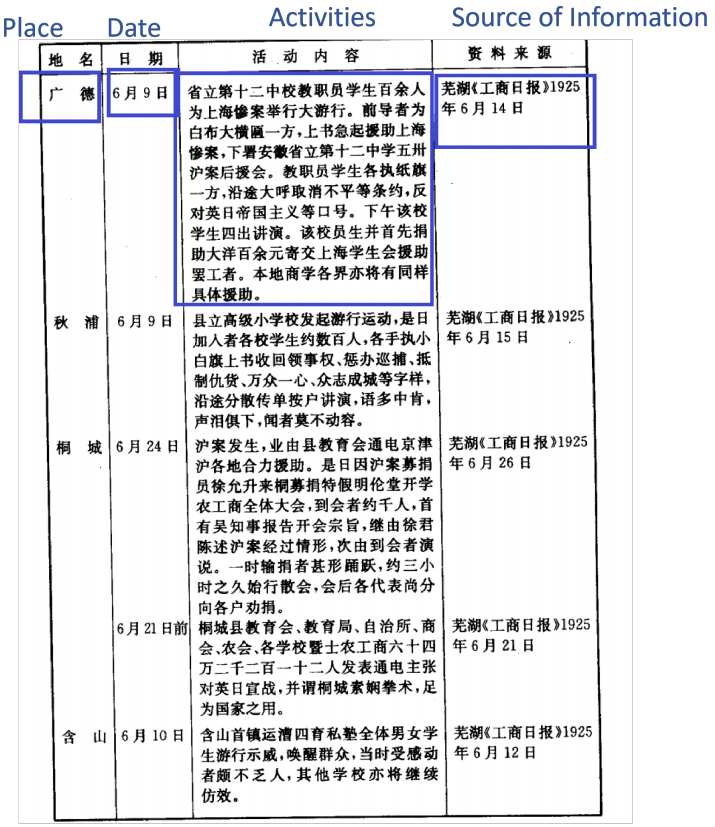
Figure 2. Top 100 May Thirtieth Declaration Terms: Word Clouds

Note: Using 109 declarations and petitions collected from Institute of History, Shanghai Academy of Social Sciences (1981, 1986, 2005), this figure plots word clouds for the original Chinese terms and their English approximations. Back to Section II.B..



Figure 3. County-Gazetteer Search Example for CCP Development Histories

Note: This figure documents the source workflow for CCP development data. The screenshot shows gazetteer sections used to code the first CCP member, the first party organization, and county committee establishment, together with Wikipedia and Baidu Baike cross-checks for the first arriving CCP member where available. Back to Section III.A..



地 名	日 期	活 动 内 容	资 料 来 源
广 德	6月9日	省立第十二中校教职员学生百余人为上海惨案举行大游行。前导者为白布大横匾一方，上书急起援助上海惨案，下署安徽省立第十二中学五卅沪案后援会。教职员学生各执纸旗一方，沿途大呼取消不平等条约，反对英日帝国主义等口号。下午该校学生四出讲演。该校员生并首先捐助大洋百余元寄交上海学生会援助罢工者。本地商学各界亦将有同样具体援助。	芜湖《工商日报》1925年6月14日
秋 浦	6月9日	县立高级小学校发起游行运动，是日加入者各校学生约数百人，各手执小白旗上书收回领事权、惩办巡捕、抵制仇货、万众一心、众志成城等字样，沿途分散传单按户讲演，语多中肯，声泪俱下，闻者莫不动容。	芜湖《工商日报》1925年6月15日
桐 城	6月24日	沪案发生，业由县教育会通电京津沪各地合力援助。是日因沪案募捐员徐允升来桐募捐特假明伦堂开学农工商全体大会，到会者约千人，首有吴知事报告开会宗旨，继由徐君陈述沪案经过情形，次由到会者演说。一时输捐者甚形踊跃，约三小时之久始行散会，会后各代表尚分向各户劝捐。	芜湖《工商日报》1925年6月26日
	6月21日前	桐城县教育会、教育局、自治所、商会、农会、各学校暨士农工商六十四万二千二百一十二人发表通电主张对英日宣战，并谓桐城素娴拳术，足为国家之用。	芜湖《工商日报》1925年6月21日
含 山	6月10日	含山首镇运漕四育私塾全体男女学生游行示威，唤醒群众，当时受感动者颇不乏人，其他学校亦将继续仿效。	芜湖《工商日报》1925年6月12日

Activity-content translations.

Guangde. More than 100 faculty members and students of Anhui Provincial No. 12 Middle School held a large march in response to the Shanghai Massacre. At the front was a large white-cloth banner reading, "Rise at once to aid the victims of the Shanghai Massacre," signed by the May Thirtieth Shanghai Incident Support Association of Anhui Provincial No. 12 Middle School. Faculty members and students each carried a paper flag and shouted slogans along the route, including "abolish the unequal treaties" and "oppose British and Japanese imperialism." In the afternoon, students from the school went out in different directions to give public speeches. Faculty members and students also took the lead in donating more than 100 silver dollars, which were sent to the Shanghai Student Union to support the strikers. Local commercial and educational circles were also expected to provide similar concrete assistance.

Qiupu. The county higher primary school initiated a march. On that day, several hundred students from various schools joined. Each carried a small white flag bearing slogans such as "restore consular jurisdiction," "punish the police," "boycott enemy goods," "unite as one," and "unity is strength." Along the route, they distributed leaflets and gave household-by-household speeches. Their words were pointed and often delivered with tears, moving nearly all who heard them.

Tongcheng. After the Shanghai Incident occurred, the county education association had already sent telegrams to Beijing, Tianjin, Shanghai, and other places calling for joint assistance. On that day, because Xu Yunsheng, a fundraiser for the Shanghai Incident, came to Tongcheng to solicit donations, a general assembly of students, peasants, workers, merchants, and other groups was convened at Minglun Hall. About 1,000 people attended. County magistrate Wu first reported the purpose of the meeting, after which Xu described the course of the Shanghai Incident. Attendees then delivered speeches. Donations were made with great enthusiasm, and the meeting lasted about three hours before dispersing. After the meeting, representatives went separately to households to encourage further donations. The Tongcheng County Education Association, Education Bureau, Self-Government Office, Chamber of Commerce, Farmers' Association, various schools, and 642,212 people from the gentry, peasantry, industry, and commerce issued a public telegram advocating war against Britain and Japan. The telegram also stated that Tongcheng had long been skilled in martial arts and could serve the nation.

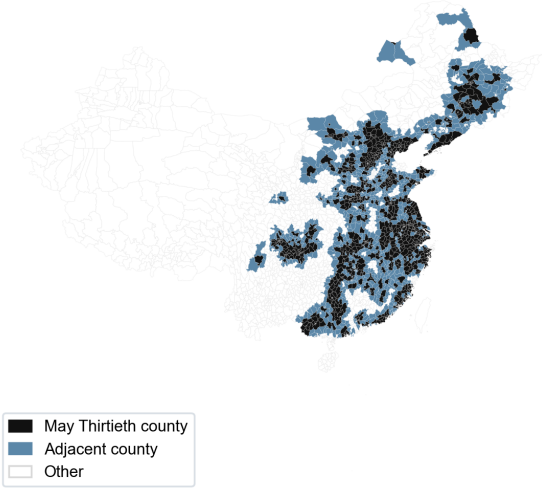
Hanshan. In Yuncao, the leading town of Hanshan, all male and female students of the Siyu private school held a demonstration march to awaken the masses. Many people were moved at the time, and other schools were expected to follow suit.

Figure 4. May Thirtieth Movement Historical Source Information Example

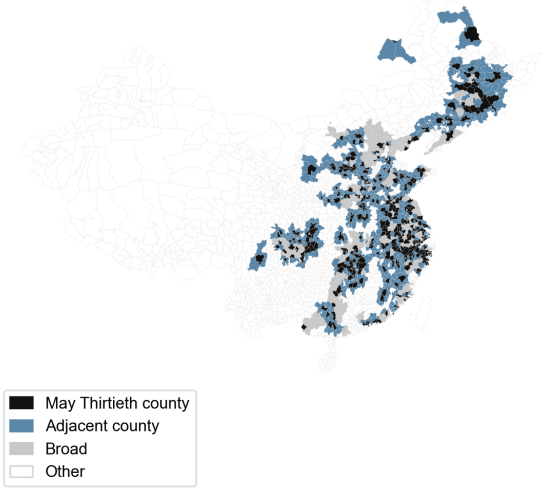
Note: The screenshot shows how May Thirtieth county entries are coded from the third volume of *Historical Materials on the May Thirtieth Movement* (Institute of History, Shanghai Academy of Social Sciences, 2005). The left panel lists county names, dates, and activity content; the activity descriptions identify whether each entry refers to the movement itself, such as protests, marches, or demonstrations. The right panel translates the displayed activity records. Back to Section III.B..



(a) Existing May Thirtieth county map



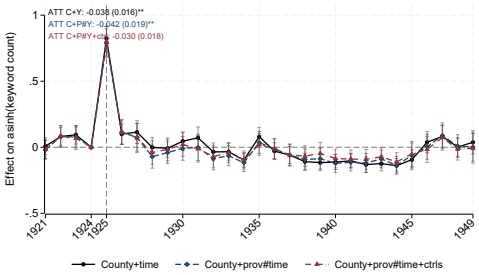
(b) Adjacent counties: broad May Thirtieth



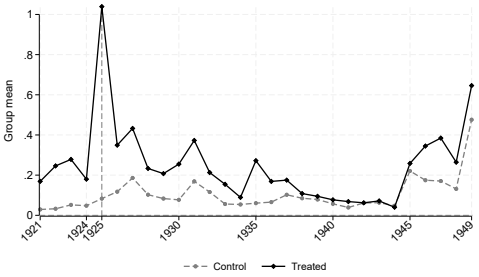
(c) Adjacent counties: strict May Thirtieth

Figure 5. May Thirtieth Movement Counties and Adjacent Counties

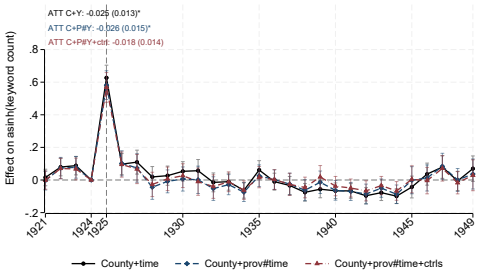
Note: This figure reports the spatial coverage of May Thirtieth exposure extracted for the paper. Panel (a) maps strict May Thirtieth counties, with broad-only counties shown in gray. Panel (b) maps broad May Thirtieth counties and their adjacent counties. Panel (c) maps strict May Thirtieth counties and their adjacent counties; analyses using the strict definition exclude broad-only counties from the comparison group. Back to Section III.B..



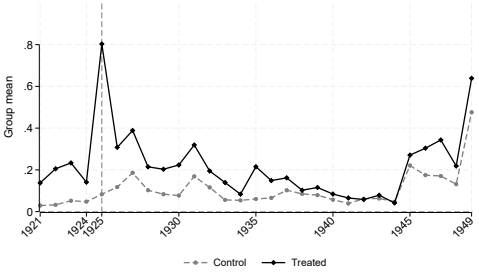
(a) Event study: annual strict



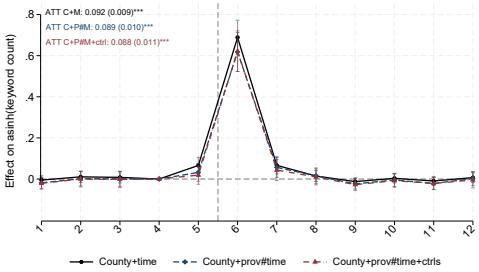
(b) Group means: annual strict



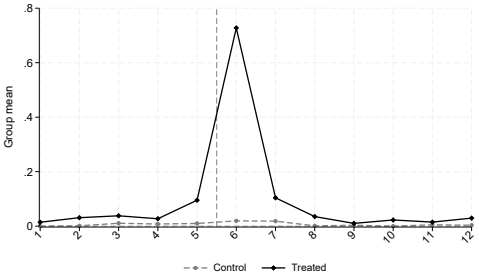
(c) Event study: annual broad



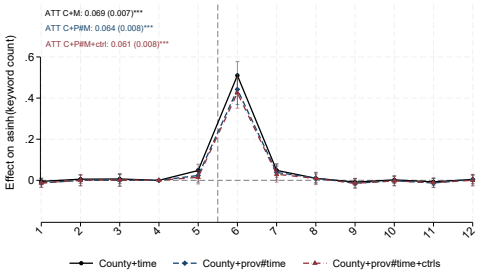
(d) Group means: annual broad



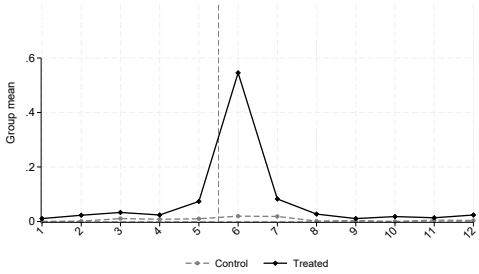
(e) Event study: monthly strict



(f) Group means: monthly strict



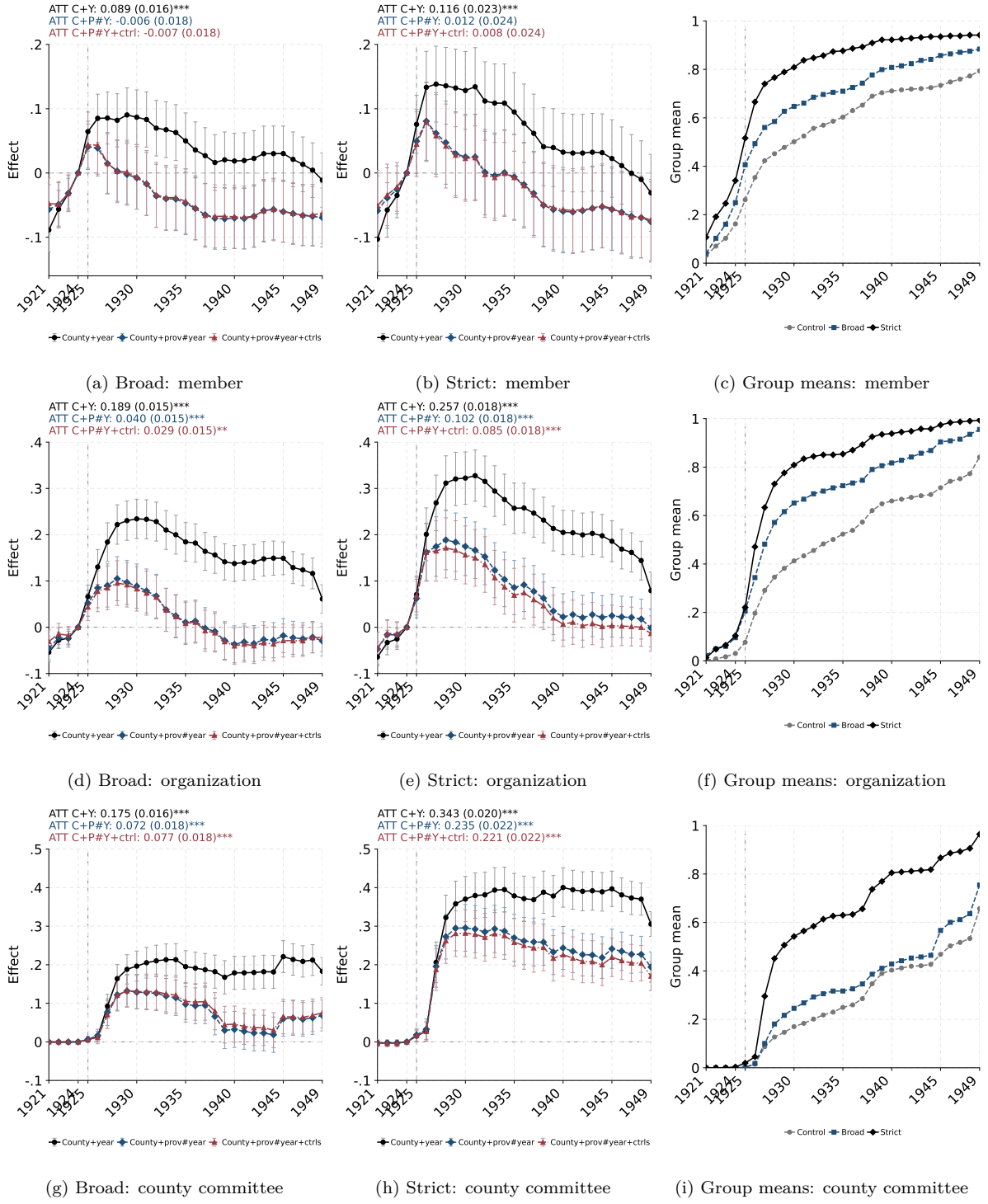
(g) Event study: monthly broad



(h) Group means: monthly broad

Figure 6. Effect of the May Thirtieth Movement on Protest Activity

Note: From the chronicle sections of county gazetteers, we collect counts of terms such as class boycott, labor strike, market closure, protest, march, and demonstration by year, and by month for 1925. The left panels report event-study estimates and the right panels report descriptive statistics by treatment status. The event-study panels use the inverse hyperbolic sine transformation. Panels (a) and (c) are annual regressions, while panels (e) and (g) are within-1925 monthly regressions. Back to Sections III.B. and IV.B..



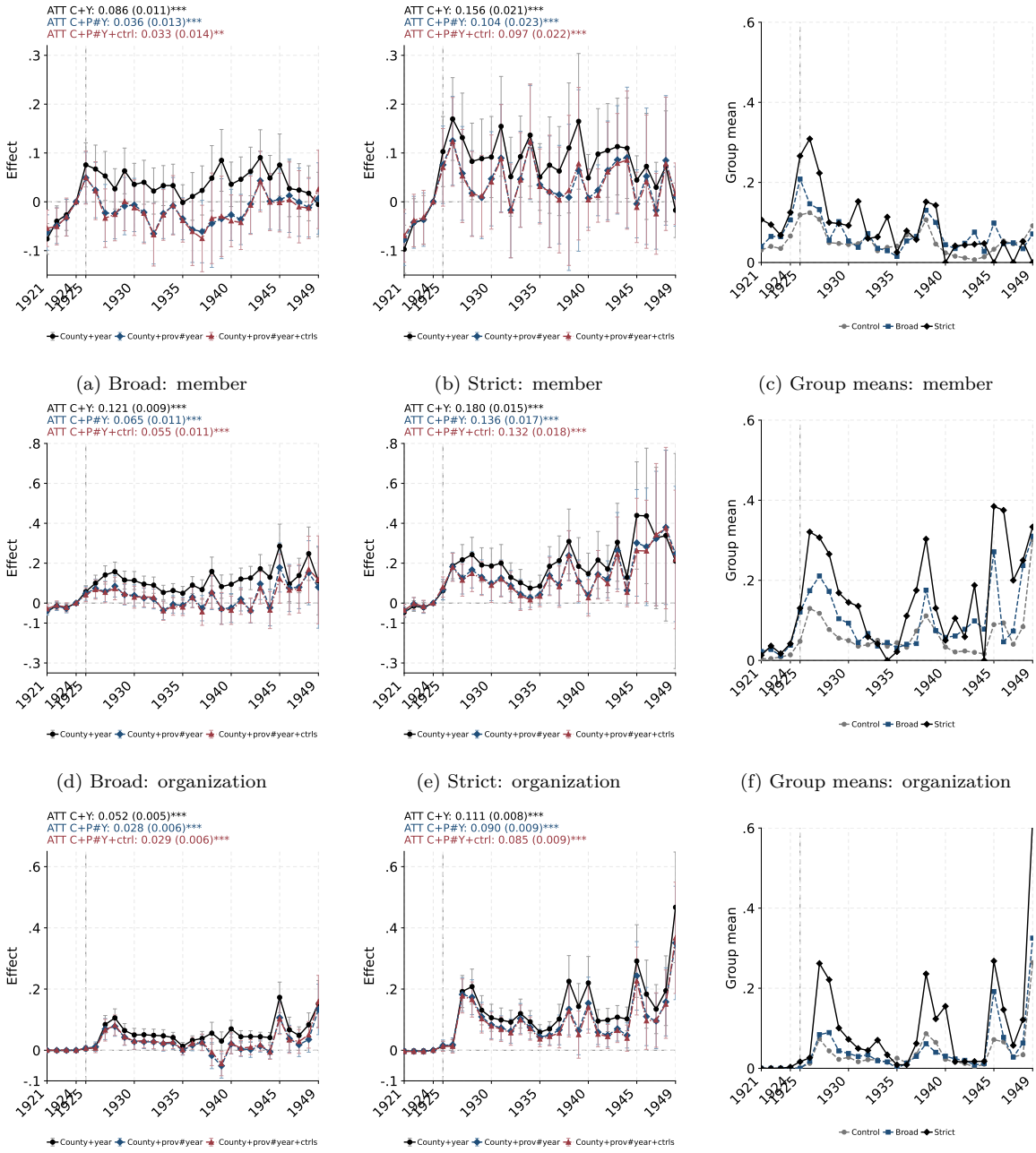
(a) Broad: member (b) Strict: member (c) Group means: member

(d) Broad: organization (e) Strict: organization (f) Group means: organization

(g) Broad: county committee (h) Strict: county committee (i) Group means: county committee

Figure 7. May Thirtieth Movement and CCP Development

Note: This figure is estimated within the framework of equation (1). Here, C+Y denotes county and year fixed effects; C+P#Y denotes county and province-year fixed effects; and C+P#Y+ctrl denotes county and province-year fixed effects plus controls selected by the control-selection equation. Panels report broad and strict May Thirtieth specifications and group means for CCP member activity, party organizations, and county committees. Standard errors are clustered by county; intervals are 95% confidence intervals. The same notation applies below. Back to Section IV.B..



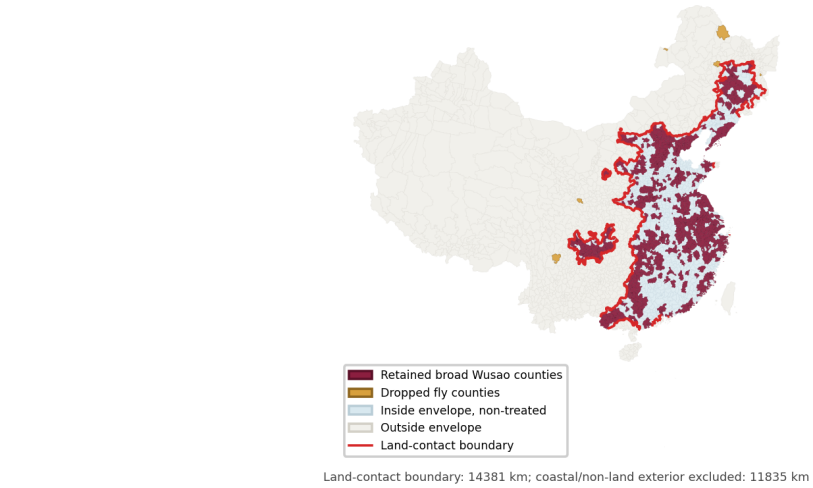
(a) Broad: member (b) Strict: member (c) Group means: member

(d) Broad: organization (e) Strict: organization (f) Group means: organization

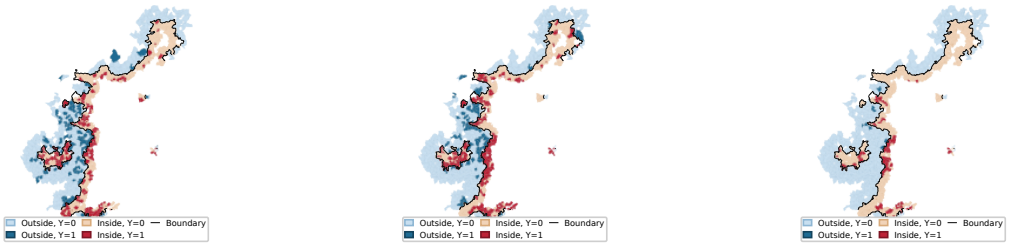
(g) Broad: county committee (h) Strict: county committee (i) Group means: county committee

Figure 8. Hazard model

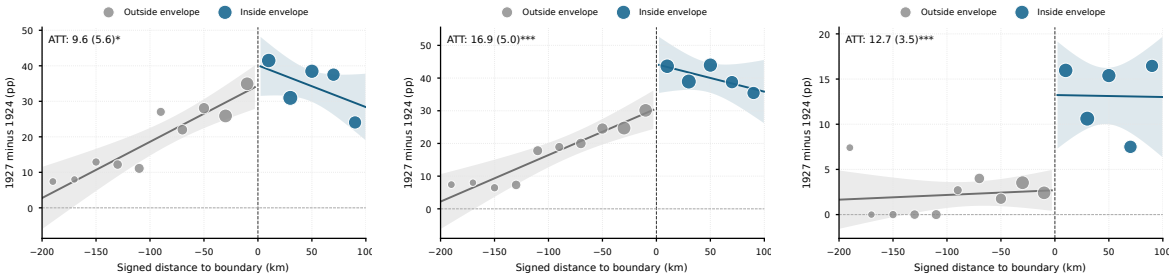
Note: This figure uses a hazard specification: after an outcome first turns to one, later observations for that outcome are removed from the regression. This specification makes year-by-year changes easier to read for absorbing CCP development outcomes. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



(a) Boundary sample



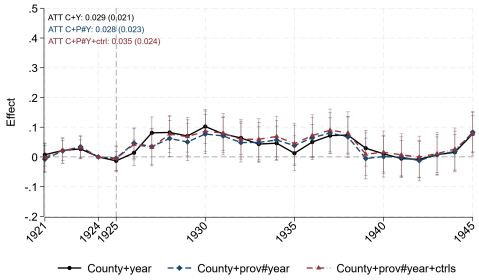
(b) Contrast: member (c) Contrast: organization (d) Contrast: committee



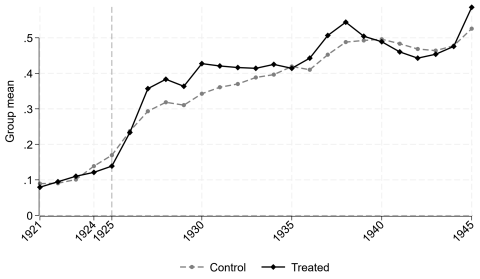
(e) RD: member (f) RD: organization (g) RD: committee

Figure 9. Spatial Boundary RD around May Thirtieth Areas

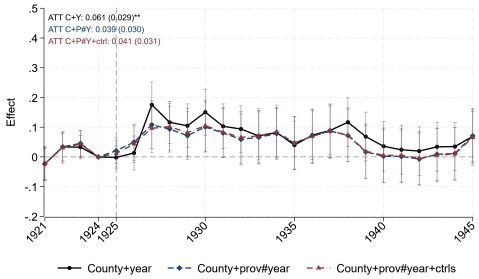
Note: This figure presents the results of our spatial regression discontinuity analysis. Panel (a) shows the boundary separating May Thirtieth and non-May Thirtieth areas, extracted according to the concave-hull procedure described in Figure A25. The area within the boundary constitutes the contiguous May Thirtieth region. We use the broad definition of May Thirtieth participation to connect May Thirtieth areas as extensively as possible. Red denotes May Thirtieth counties, while blue denotes non-May Thirtieth counties located within the boundary. Panels (b)–(d) map CCP development on both sides of the boundary, using bandwidths of 100 km inside and 200 km outside the boundary. Dark blue denotes counties outside the boundary whose CCP indicator increased from 0 in 1924 to 1 in 1927 (light blue denotes no change), while dark red denotes counties inside the boundary whose CCP indicator increased from 0 to 1 over the same period (light red denotes no change). Panels (e)–(g) present the spatial regression discontinuity estimates. The intervals denote 95% confidence intervals. Back to Section IV.B..



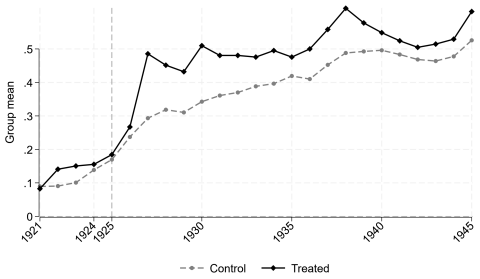
(a) Event study: Broad KMT



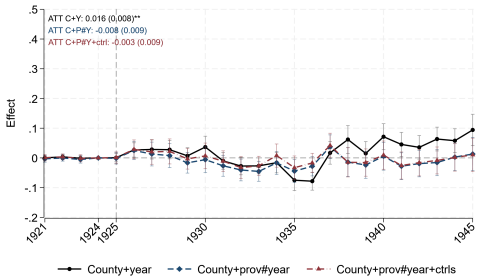
(b) Group means: Broad KMT



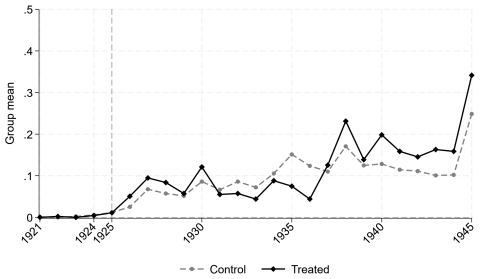
(c) Event study: Strict KMT



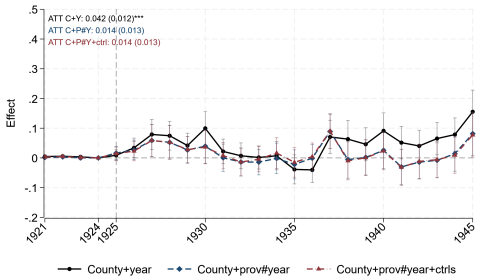
(d) Group means: Strict KMT



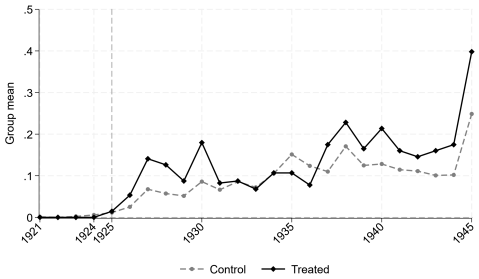
(e) Event study: Broad CCP



(f) Group means: Broad CCP



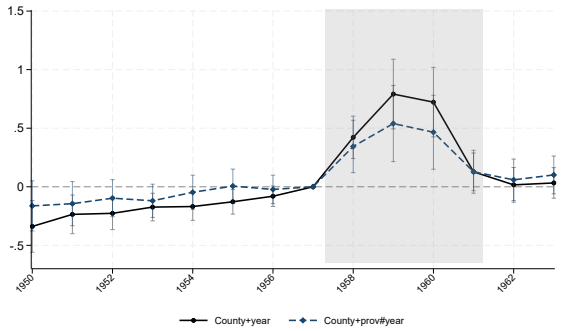
(g) Event study: Strict CCP



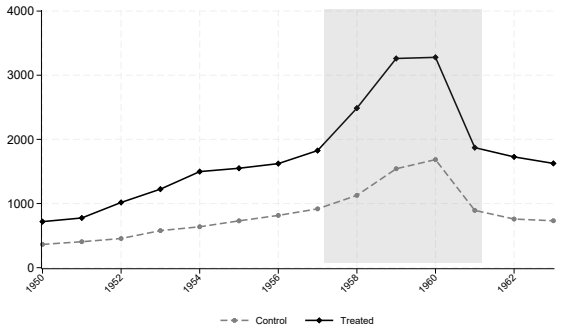
(h) Group means: Strict CCP

Figure 10. May Thirtieth Movement and Garrison Presence

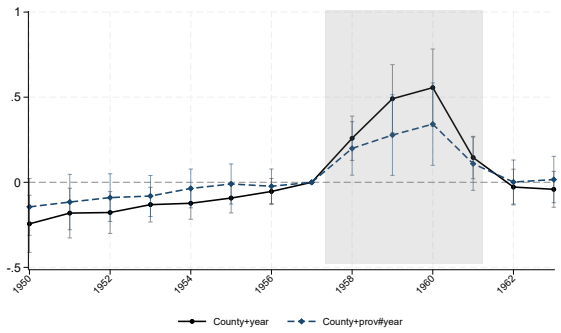
Note: The dependent variables indicate whether KMT or CCP garrisons were present, based on information collected from the military or garrison sections of county gazetteers. We exclude counties with no garrison record between 1921 and 1945. We do not additionally collect garrison data after 1945, because the full-scale civil war makes these records difficult to structure systematically. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section V.B..



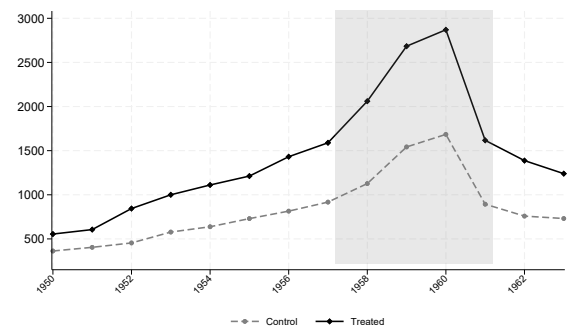
(a) Event study: strict May Thirtieth



(b) Group means: strict May Thirtieth



(c) Event study: broad May Thirtieth



(d) Group means: broad May Thirtieth

Figure 11. Industrial Great Leap in May Thirtieth Areas

Note: This figure examines whether May Thirtieth areas experienced a stronger leap in industrial output during 1958–1961. Industrial output data come from county gazetteers and are measured in units of ten million yuan. Panels (b) and (d) report descriptive group means for May Thirtieth and non-May Thirtieth areas. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section VI.B..

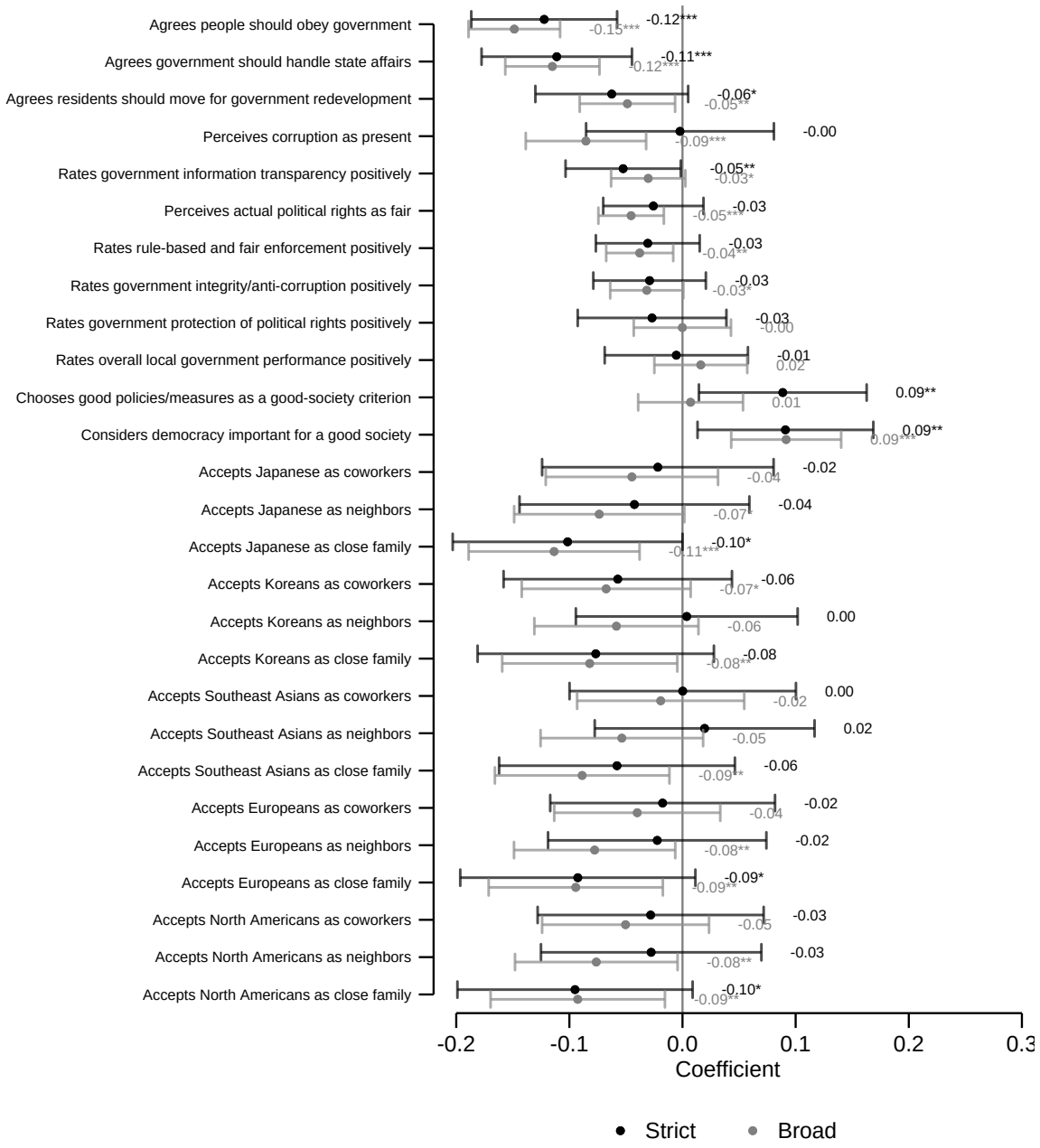


Figure 12. May Thirtieth Movement and Ideology in the Twenty-First Century

Note: The data come from two large social surveys: one conducted in 2018 covering 135 counties, and another conducted in 2013 and 2015 covering 153 counties. Together, they include 22,204 respondents. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section VI.B..

Table 1. County Covariates and May Thirtieth Coverage

	Strict May Thirtieth	Broad May Thirtieth
Panel A. Control Variables for Baseline		
Pre-1921 enterprise	0.14***	0.28***
	(0.02)	(0.02)
Agricultural endowment (s.d.)	0.08***	0.14***
	(0.01)	(0.01)
HYDE 1920 population density (s.d.)	0.03***	0.10***
	(0.01)	(0.01)
HYDE 1920 pasture share (s.d.)	0.03***	0.05***
	(0.01)	(0.01)
Closer to railway (s.d.)	0.06***	0.10***
	(0.00)	(0.01)
Closer to river (s.d.)	0.01**	0.01*
	(0.01)	(0.01)
Closer to coastline (s.d.)	0.07***	0.13***
	(0.01)	(0.01)
Closer to 1911 provincial capital (s.d.)	0.05***	0.10***
	(0.01)	(0.01)
Precipitation anomaly in prior 3 years	0.07***	0.10***
	(0.02)	(0.02)
Panel B. Control Variables for Robustness		
Males per 100 females, Minguo 17 (s.d.)	0.00	-0.01
	(0.02)	(0.02)
Employed population share, Minguo 17 (s.d.)	-0.04**	-0.05***
	(0.02)	(0.02)
Secondary school in county, 1934	0.20***	0.30***
	(0.02)	(0.02)
Christian school in county	0.18***	0.22***
	(0.06)	(0.05)
Christian hospital in county	0.24***	0.23***
	(0.05)	(0.04)
Any CHCD institution in county	0.13***	0.10***
	(0.02)	(0.02)

Note: The dependent variable equals one if the county is a May Thirtieth area. Each coefficient is estimated in a separate linear probability model. Continuous covariates enter as standardized variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Section III.D..

Table 2. Spatial Gradient Around May Thirtieth Counties

	Broad May Thirtieth			Strict May Thirtieth		
	Member (1)	Org. (2)	Committee (3)	Member (4)	Org. (5)	Committee (6)
Panel A. Outcome status						
May Thirtieth counties × Post-1925	0.082***	0.178***	0.216***	0.034	0.184***	0.300***
	(0.031)	(0.027)	(0.029)	(0.031)	(0.025)	(0.028)
First-ring counties × Post-1925	0.114***	0.183***	0.168***	0.052*	0.142***	0.114***
	(0.029)	(0.025)	(0.026)	(0.027)	(0.022)	(0.025)
Second-ring counties × Post-1925	0.043	0.104***	0.102***	-0.018	0.061***	0.050**
	(0.029)	(0.026)	(0.025)	(0.026)	(0.022)	(0.023)
Panel B. Annual first-appearance probability						
May Thirtieth counties × Post-1925	0.092***	0.127***	0.064***	0.138***	0.194***	0.107***
	(0.018)	(0.016)	(0.008)	(0.024)	(0.020)	(0.010)
First-ring counties × Post-1925	0.082***	0.093***	0.043***	0.069***	0.090***	0.033***
	(0.016)	(0.013)	(0.007)	(0.016)	(0.013)	(0.007)
Second-ring counties × Post-1925	0.025*	0.039***	0.023***	0.004	0.034***	0.012*
	(0.014)	(0.012)	(0.006)	(0.013)	(0.011)	(0.006)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Province × year FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS controls × year	Yes	Yes	Yes	Yes	Yes	Yes
Rain anomaly	Yes	Yes	Yes	Yes	Yes	Yes
Observations, Panel A	66,468	66,468	66,468	53,476	53,476	53,476
Counties, Panel A	2,292	2,292	2,292	1,844	1,844	1,844
Observations, Panel B	28,833	32,515	47,217	23,890	27,383	37,984
Counties, Panel B	2,292	2,292	2,292	1,844	1,844	1,844

Note: The omitted comparison group is always counties outside the second ring. May Thirtieth, first-ring, and second-ring indicators are made mutually exclusive before constructing Post-1925 interactions. Panel A uses the original county-year status outcomes. Panel B uses a hazard model. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.B..

Table 3. May Thirtieth Movement and Broad-Definition Red Books

	asinh(Count)			log(1+Count)		
	(1)	(2)	(3)	(4)	(5)	(6)
1922-1924	0.000	-0.013	-0.011	0.001	-0.010	-0.008
	(0.010)	(0.015)	(0.014)	(0.008)	(0.012)	(0.011)
1925	0.059***	0.061*	0.063*	0.048***	0.052*	0.054*
	(0.020)	(0.033)	(0.032)	(0.016)	(0.029)	(0.028)
1926	0.092***	0.138**	0.126**	0.076***	0.116**	0.106**
	(0.031)	(0.062)	(0.058)	(0.026)	(0.054)	(0.050)
1927	0.081**	0.137**	0.122**	0.068**	0.115**	0.103**
	(0.040)	(0.062)	(0.058)	(0.032)	(0.052)	(0.048)
1928	-0.000	-0.046	-0.045	0.003	-0.035	-0.034
	(0.031)	(0.037)	(0.037)	(0.025)	(0.029)	(0.029)
1929-1936	0.056**	0.011	0.006	0.046**	0.009	0.006
	(0.025)	(0.030)	(0.029)	(0.020)	(0.023)	(0.023)
1937-1945	0.165**	0.106	0.088	0.132**	0.087	0.071
	(0.071)	(0.082)	(0.086)	(0.060)	(0.068)	(0.072)
1946	0.546***	0.224	0.190	0.448***	0.182	0.152
	(0.131)	(0.169)	(0.165)	(0.111)	(0.143)	(0.139)
1947	0.540***	0.134	0.162	0.449***	0.109	0.135
	(0.140)	(0.167)	(0.161)	(0.119)	(0.142)	(0.137)
1948	0.700***	0.289	0.341*	0.579***	0.234	0.278*
	(0.151)	(0.183)	(0.179)	(0.128)	(0.155)	(0.152)
1949	0.884***	0.539***	0.507***	0.743***	0.453***	0.423***
	(0.163)	(0.188)	(0.184)	(0.140)	(0.161)	(0.158)
ATT 1925-1949	0.193***	0.111*	0.101*	0.159***	0.091*	0.082*
	(0.047)	(0.057)	(0.057)	(0.039)	(0.048)	(0.047)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	189	187	187	189	187	187
Control cities	117	115	115	117	115	115

Note: This table uses a city-level panel. A city is coded as treated if at least one county within the city participated in the May Thirtieth Movement. Broad-definition red books include books on communism, socialism, and other left-wing revolutionary ideas; the strict definition excludes the broader left-wing revolutionary category. Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Sections V.A. and VI.A..

Table 4. May Thirtieth Movement and Local-Native Share among Martyrs

	Broad May Thirtieth		Strict May Thirtieth	
	(1)	(2)	(3)	(4)
May Thirtieth county	0.115**	0.097*	0.168***	0.119*
	(0.047)	(0.052)	(0.056)	(0.068)
Treated mean	0.725	0.725	0.778	0.778
Control mean	0.610	0.610	0.610	0.610
Observations	233	233	192	192
Treated counties	75	75	34	34
Control counties	158	158	158	158
Province FE	No	Yes	No	Yes

Note: We retain counties with records of martyrs who died in the county but were not natives of that county, and then compute the share of 1945–1949 martyrs whose native places were local. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section VI.A..

Table 5. Rightward Reforms in May Thirtieth Areas

	HRS First-Team Year		HRS Majority-Team Year		One-Child Policy Year	
	Strict	Broad	Strict	Broad	Strict	Broad
	(1)	(2)	(3)	(4)	(5)	(6)
May Thirtieth county	0.223**	0.201***	0.253***	0.169***	-0.111	-0.089
	(0.108)	(0.073)	(0.089)	(0.060)	(0.149)	(0.102)
Observations	1,588	1,993	1,574	1,975	1,586	1,985
Treated counties	242	634	237	625	263	647
Control counties	1,346	1,359	1,337	1,350	1,323	1,338
Control mean	1980.18	1980.18	1981.78	1981.78	1979.96	1979.95
Province FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is the policy implementation year. All policy-timing measures come from local gazetteers. Columns (1)–(2) use the year when the first production team in the county implemented the Household Responsibility System (HRS). Columns (3)–(4) use the year when at least half of production teams in the county had implemented HRS. Columns (5)–(6) use the county-government-promoted one-child policy (OCP) year, coded as the earliest local-gazetteer year in which the county (i) explicitly advocated one birth, controlled second births, and prohibited third births; (ii) penalized third or higher-order births; or (iii) rewarded one-child families. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Back to Section VI.B..

Appendix A. Definitions of Red-Book Indicators

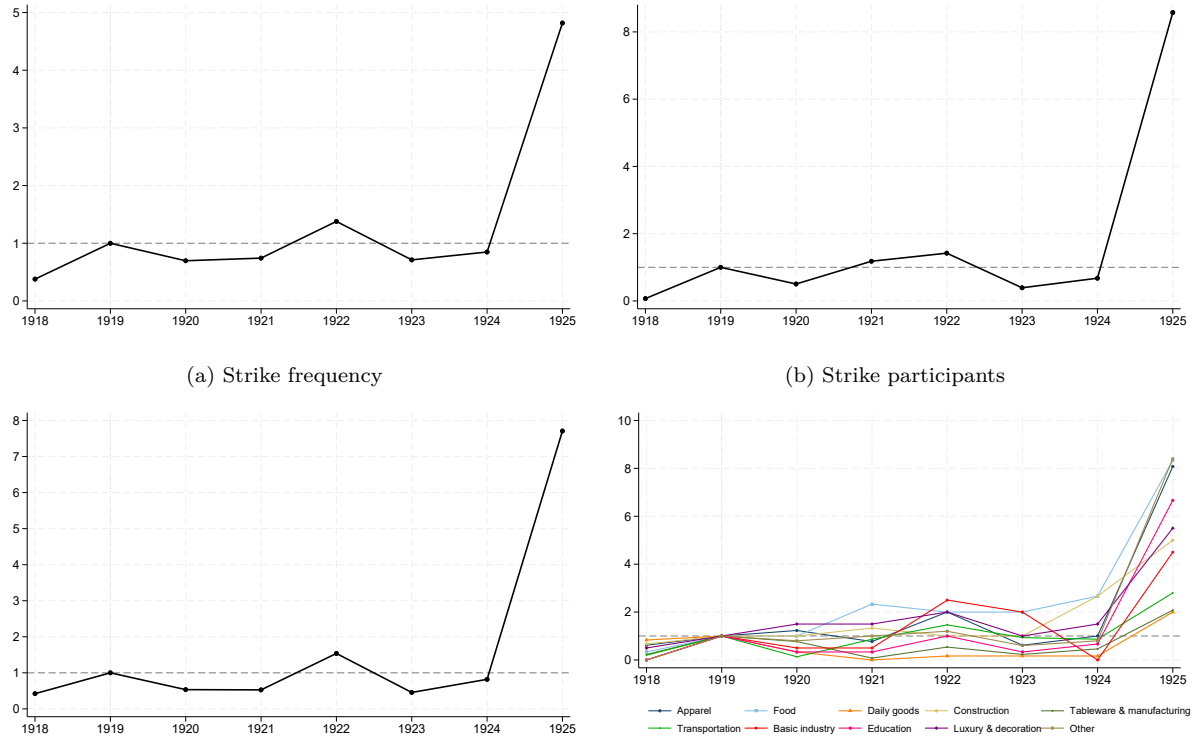
We compile about 450,000 books from Republican-period bibliographic records of the National Library of China (NLC), of which roughly 340,000 can be matched to city- or county-level places of publication. The available fields include title, author, editor, translator, publisher, place and year of publication, keywords, abstract, classification number, and classification node. Our primary basis for ideological coding is the abstract, but only 43 percent of books contain one. For many books, the title, keywords, and classification information are insufficient to determine ideological content, and the absence of an abstract prevents reliable classification. Thus, in the main analysis, we focus on books whose ideological categories can be clearly identified. For placebo outcomes, we use categories in which, among books with available information, only a very small share, or none, are red books, such as science and engineering books.

Communist/socialist indicator. This indicator equals one only when a book directly concerns communism, socialism, Marxism-Leninism, the Soviet socialist system, communist organizations, or movements and political ideas clearly associated with these traditions. Positive cases include records that explicitly mention the Communist Party, the CCP, communist organizations, the Comintern, the Third International, communism, socialism, scientific socialism, utopian socialism, Marxism, Leninism, Marxism-Leninism, or Bolshevism.

Broad red-ideology indicator. Relative to the previous indicator, this measure includes not only communism and socialism, but also broader leftist revolutionary content. It equals one when the metadata indicates that the book disseminates, explains, records, studies, or mobilizes such content. Examples include abstracts that advocate a political ideology centered on social equality, social justice, or resistance to class oppression, as well as books that promote the economic, legal, or political systems of the Soviet Union or Soviet Russia. A mere mention of the Soviet Union, the Soviets, Russia, or Moscow is insufficient, because such books may simply describe Soviet or Russian geography or the environment.

Anti-communist materials hostile to communism are coded as zero under both indicators. Materials criticizing CCP leaders are also coded as zero. General introductions to ideologies across countries are excluded unless the abstract explicitly emphasizes the superiority of CCP or Soviet Russian ideology relative to other ideologies.

Appendix B. Additional Figures



(a) Strike frequency

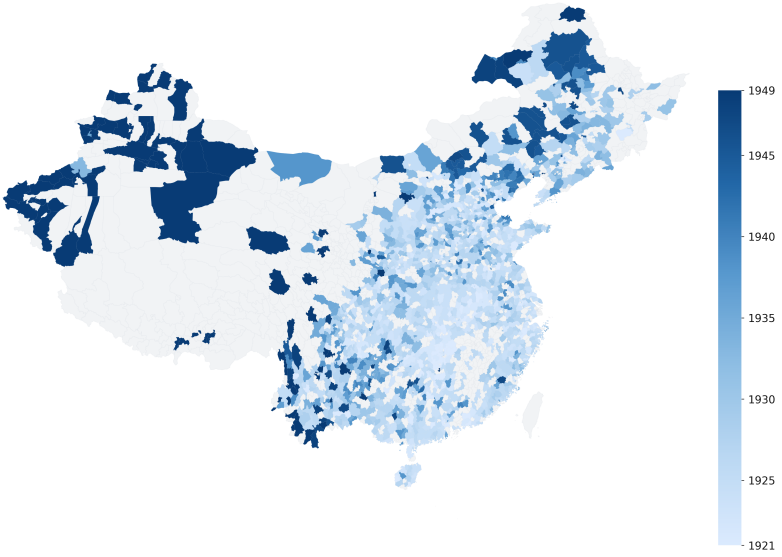
(b) Strike participants

(c) Strike days

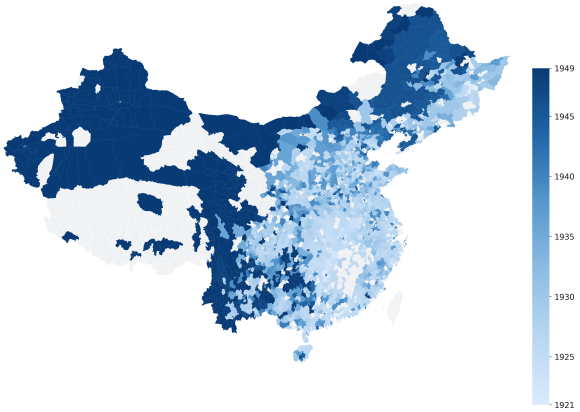
(d) Industry-specific strike activity

Figure A1. Strike Activity Normalized to 1919, 1918–1925

Note: Each panel normalizes the corresponding strike measure to its 1919 level. Panels report strike frequency, strike participants, strike days, and industry-specific strike activity. Data source: Chen (1926). Back to Section II.B..



(a) First party member year



(b) First party organization year



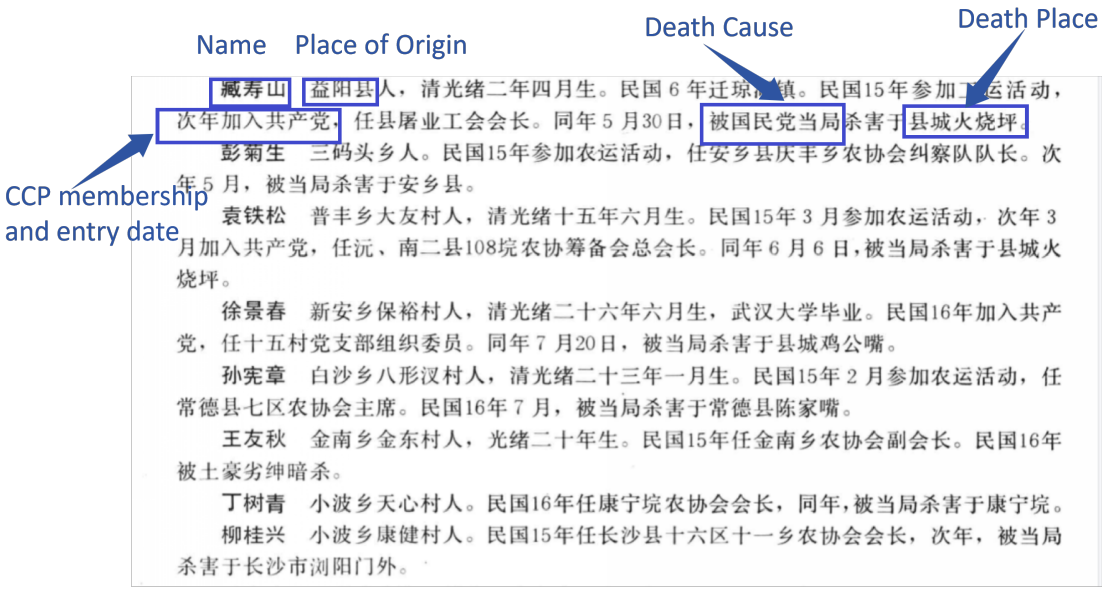
(c) First county party committee year

Figure A2. Geographic Distribution of CCP Development

Note: These maps report the spatial coverage of the CCP development data compiled for the paper. Gray areas are counties with missing data for the corresponding outcome. Counties observed in 1949 or later are assigned the same color to make the maps easier to read. Back to Section III.A..



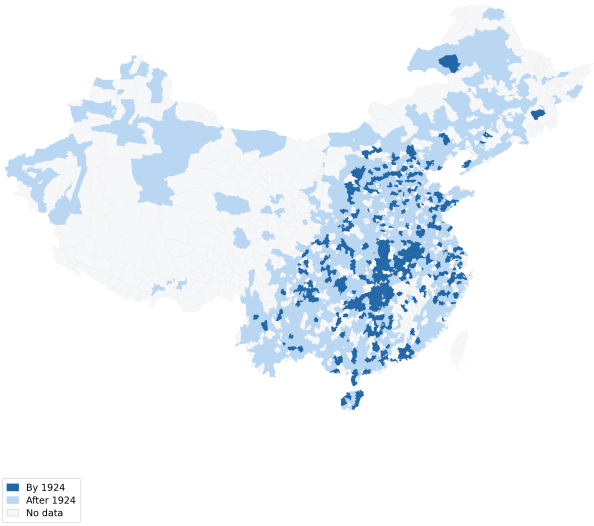
(a) Book-publication records



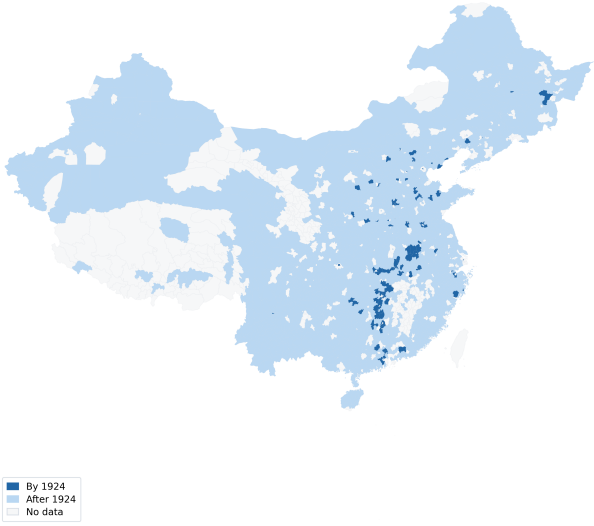
(b) Martyr records

Figure A3. Examples of Book-Publication and Martyr Records

Note: Panel (a) reports a typical book record from the NLC Republican-era Books Database. Panel (b) reports an example of martyr data from a county gazetteer. Back to Section III.C..



(a) Counties with party member by 1924



(b) Counties with party organization by 1924

Figure A4. Party Presence by 1924

Note: This figure maps counties with CCP member activity and party organizations before the May Thirtieth Movement. The maps use the CCP development data compiled from local gazetteers. Back to Section IV.A..

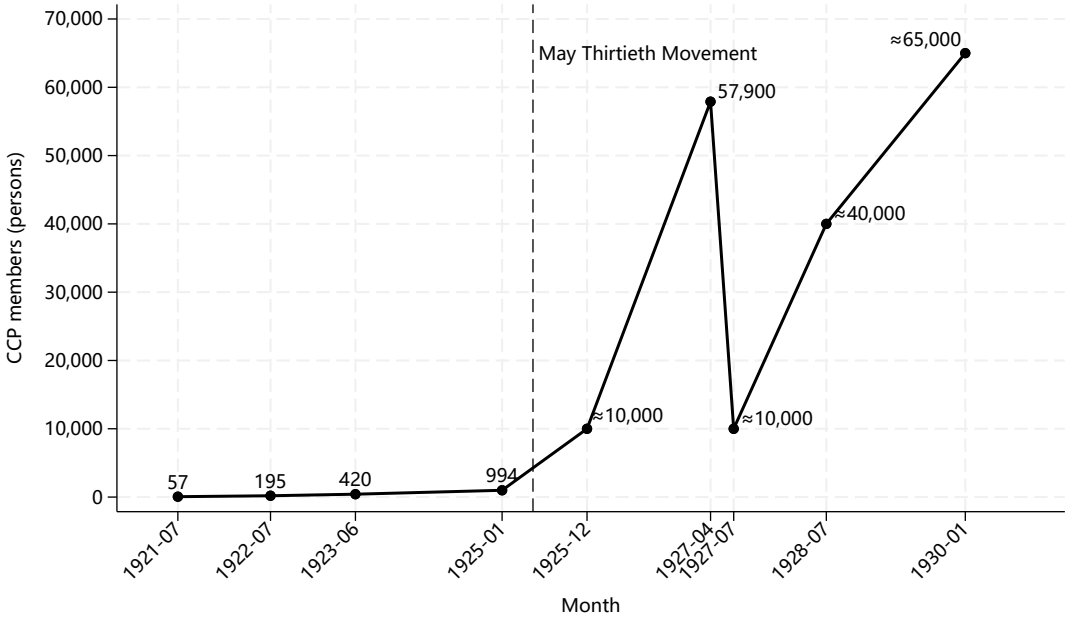
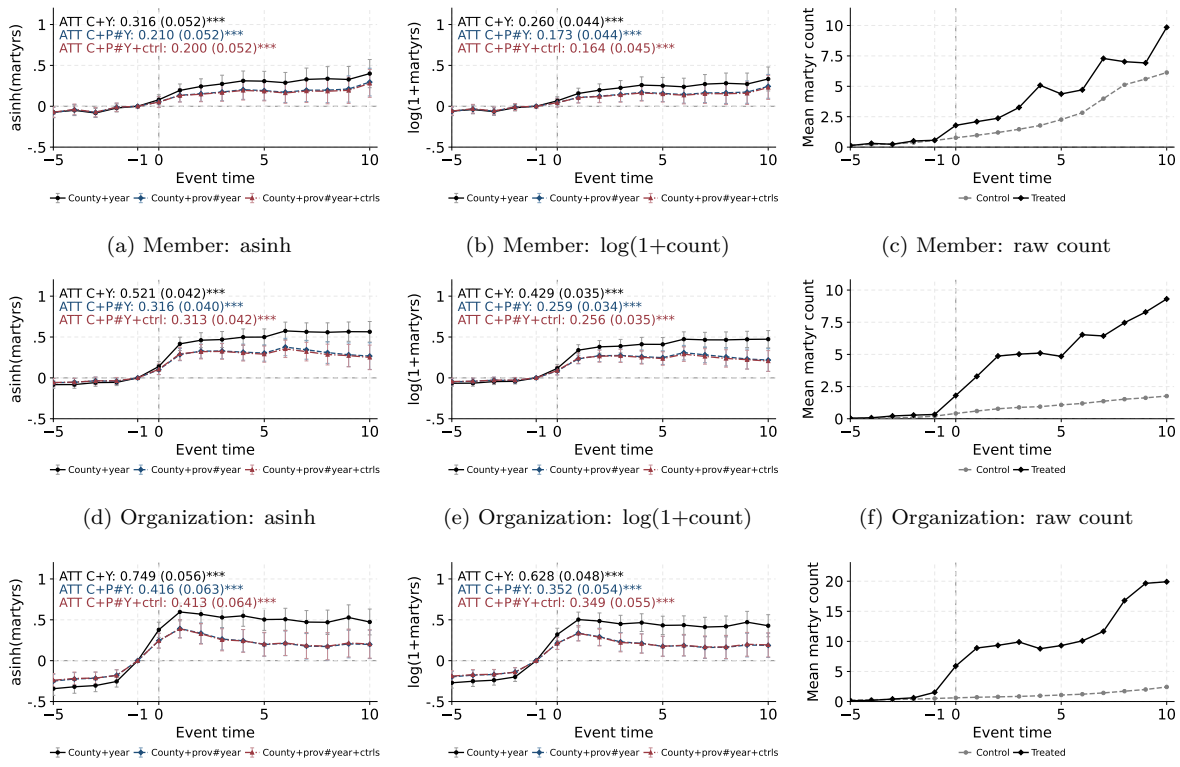


Figure A5. CCP Membership Growth, 1921–1930

Note: Data source: Ma and Huang (1989). There is no settled consensus on CCP membership in 1921, but most historical estimates place it between 50 and 60 members. Back to Section IV.A..



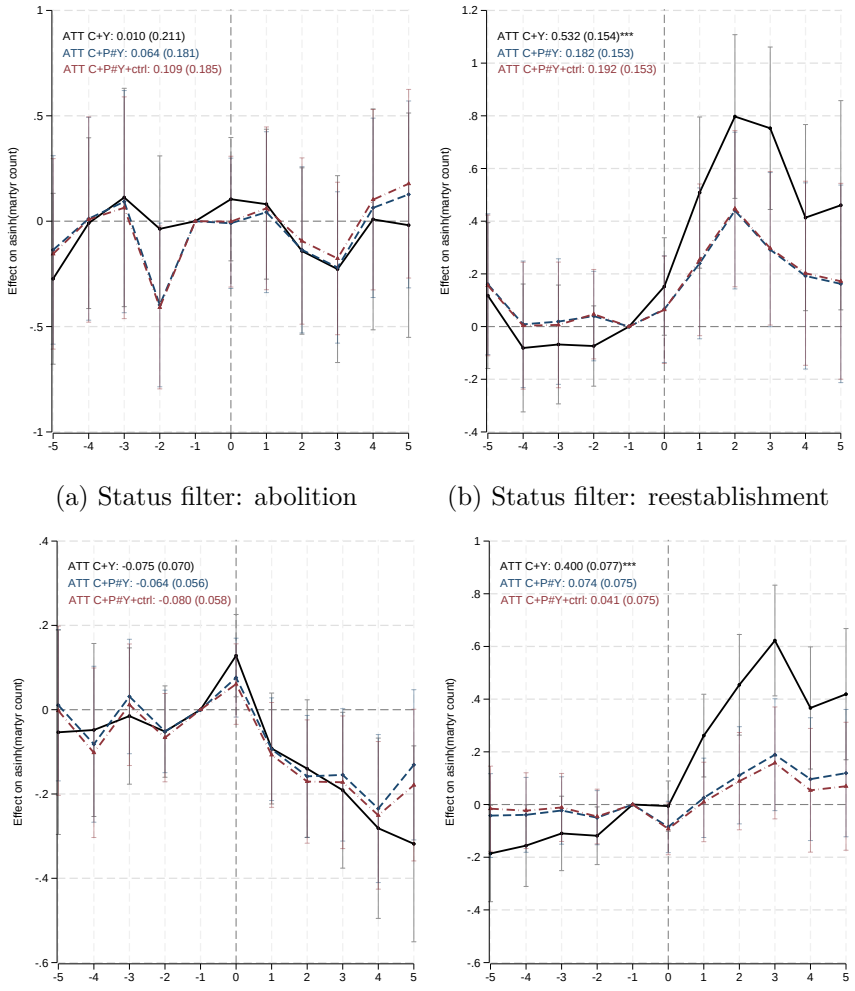
(a) Member: asinh (b) Member: log(1+count) (c) Member: raw count

(d) Organization: asinh (e) Organization: log(1+count) (f) Organization: raw count

(g) County committee: asinh (h) County committee: log(1+count) (i) County committee: raw count

Figure A6. Martyr Counts Around CCP Development

Note: This figure reports changes in martyr counts from five years before to ten years after each type of CCP development event. The left column uses the inverse hyperbolic sine transformation of county martyr counts, the middle column uses log(1+count), and the right column reports descriptive group means. All regressions use stacked difference-in-differences to avoid negative-weight problems. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



(a) Status filter: abolition

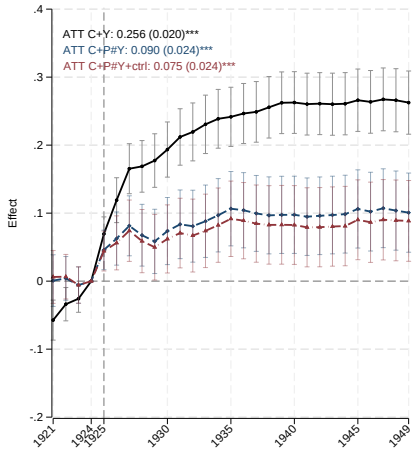
(b) Status filter: reestablishment

(c) Observation-level filter: abolition

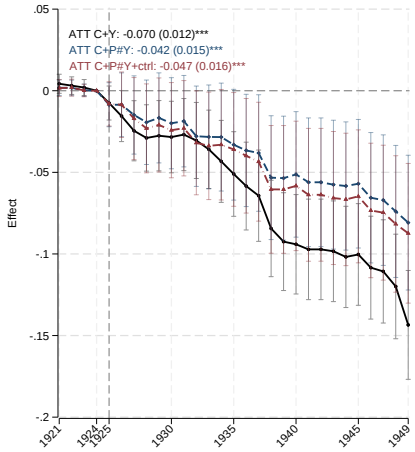
(d) Observation-level filter: reestablishment

Figure A7. Martyr Counts Around County-Committee Reestablishment and Abolition

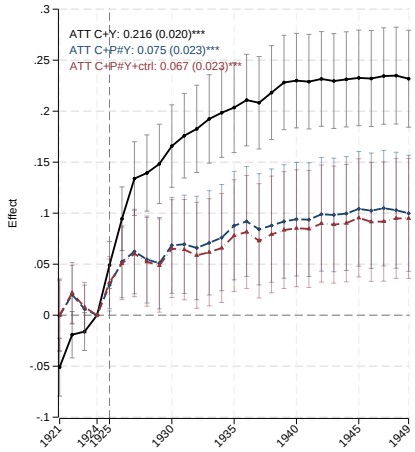
Note: This figure reports changes in martyr counts after county-committee abolition or reestablishment. Panels (a)–(b) use a county-level subdataset. For example, in panel (a), treated counties had a county committee throughout the five years before abolition and no county committee throughout the five years after abolition; control counties had a county committee throughout the same window around the treated county's abolition year. Panels (c)–(d) use an observation-level subdataset. For example, in panel (c), treated observations are kept only while the county had a committee before abolition and only while the county remained without a committee after abolition; observations before first establishment or after reestablishment are excluded. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



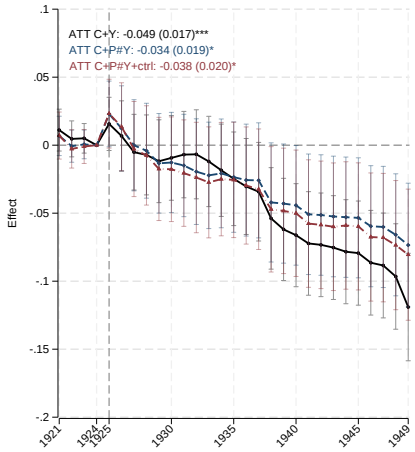
(a) Broad: from May Thirtieth Area



(b) Broad: from Non-May Thirtieth Area



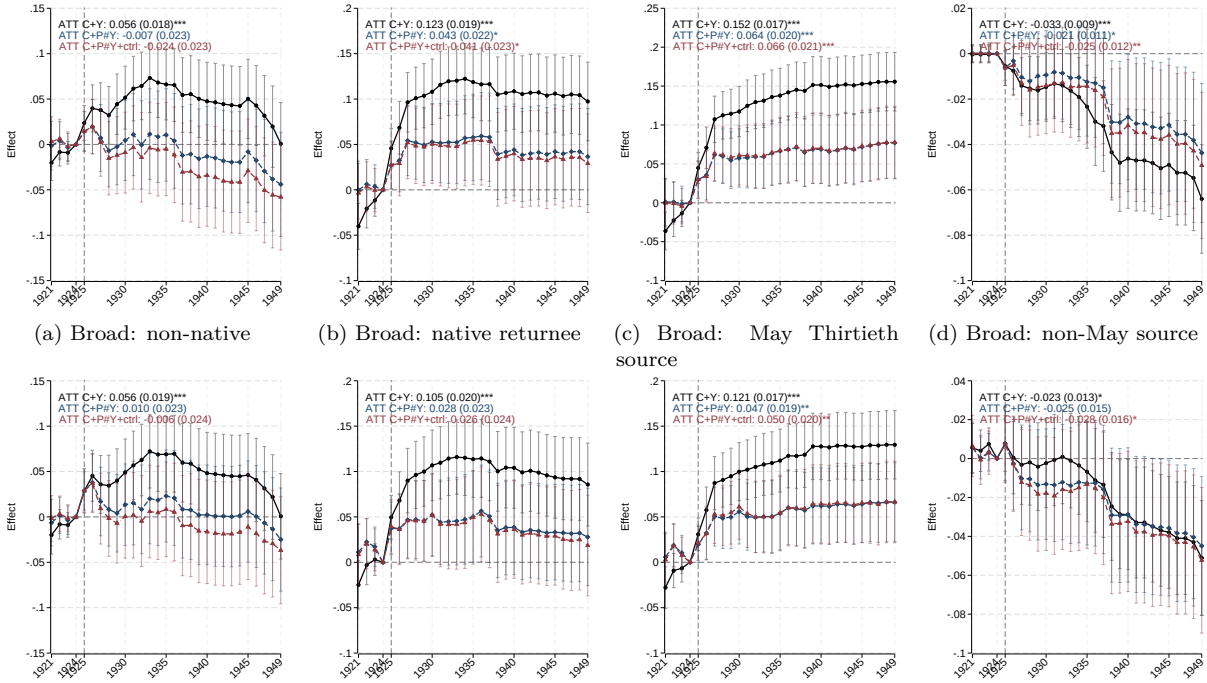
(c) Strict: from May Thirtieth Area



(d) Strict: from Non-May Thirtieth Area

Figure A8. May Thirtieth Movement and Party-Member Outflow to Adjacent Counties

Note: This figure examines whether counties adjacent to May Thirtieth areas received more CCP members from May Thirtieth areas. In panels (a) and (c), the outcome is a cumulative indicator for whether the county's first CCP member came from a broad May Thirtieth area. In panels (b) and (d), the outcome is a cumulative indicator for whether the county's first CCP member did not come from a May Thirtieth area. Treatment is first-order adjacency under the corresponding May Thirtieth definition, excluding all counties directly exposed to the May Thirtieth Movement. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

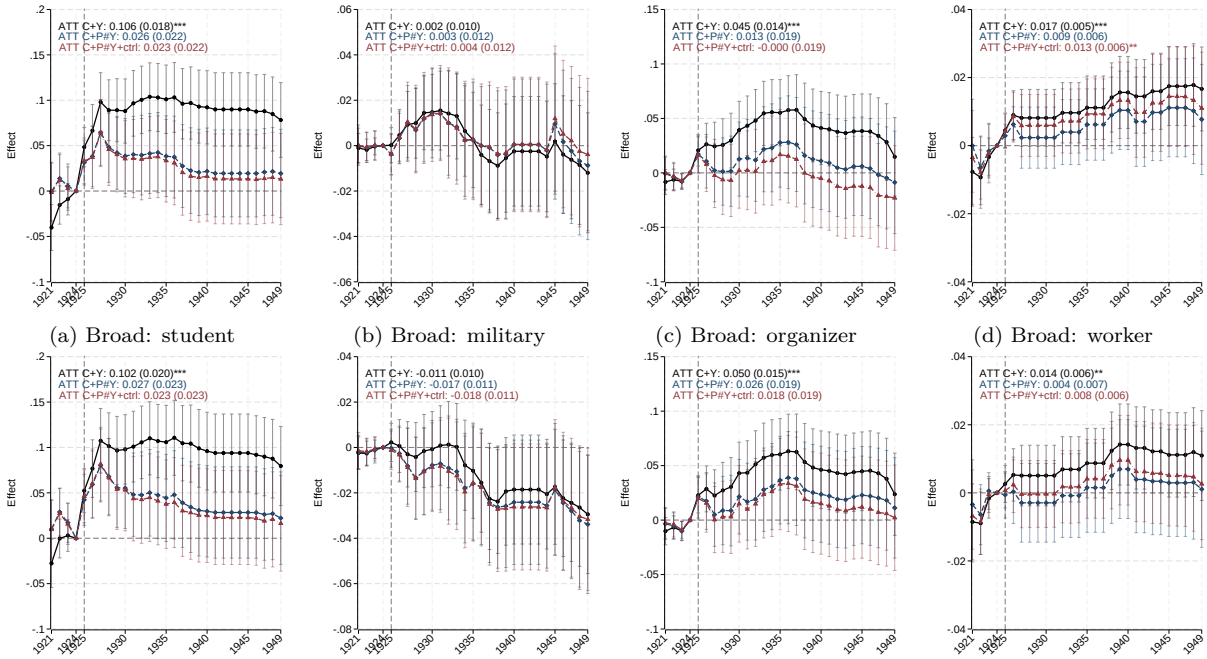


(a) Broad: non-native (b) Broad: native returnee (c) Broad: May Thirtieth source (d) Broad: non-May source

(e) Strict: non-native (f) Strict: native returnee (g) Strict: May Thirtieth source (h) Strict: non-May source

Figure A9. Adjacent Counties and First-Member Source Type

Note: This figure examines the source type of the first CCP member arriving in first-order counties adjacent to May Thirtieth areas. The first row uses the broad May Thirtieth definition, and the second row uses the strict definition. Columns report cumulative indicators for whether the county's first CCP member was a non-native entrant, a native returnee, a native returnee whose source place was a May Thirtieth area, and a native returnee whose source place was not a May Thirtieth area, respectively. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

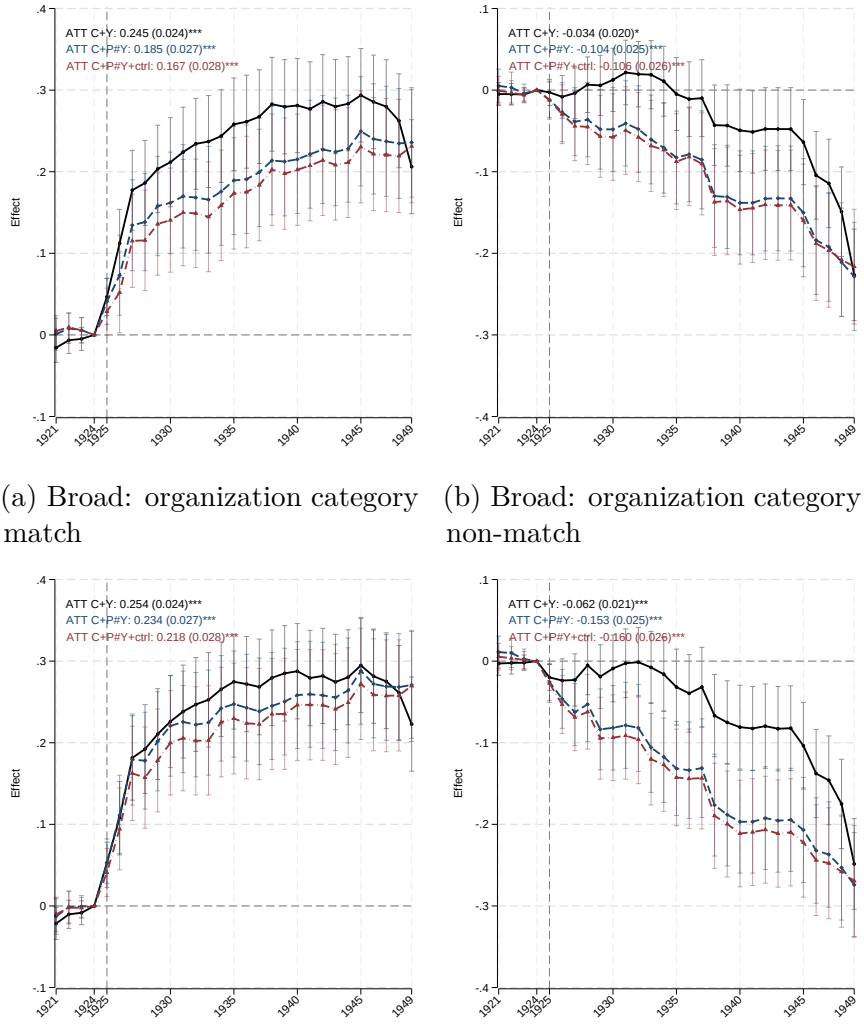


(a) Broad: student (b) Broad: military (c) Broad: organizer (d) Broad: worker

(e) Strict: student (f) Strict: military (g) Strict: organizer (h) Strict: worker

Figure A10. Adjacent Counties and First-Member Identity

Note: This figure examines the occupation or identity of the first CCP member arriving in first-order counties adjacent to May Thirtieth areas. The first row uses the broad May Thirtieth definition, and the second row uses the strict definition. Columns report cumulative indicators for education-related, military-related, party-organizer, and worker backgrounds, respectively. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



(a) Broad: organization category match

(b) Broad: organization category non-match

(c) Strict: organization category match

(d) Strict: organization category non-match

Figure A11. May Thirtieth Movement and Party-Organization Similarity in Adjacent Counties

Note: This figure examines whether party organizations established in first-order counties adjacent to May Thirtieth areas were more similar to nearby May Thirtieth organizations. Similarity to a neighboring May Thirtieth organization equals one if the county established a party organization whose type matches that in an adjacent May Thirtieth county; if no directly adjacent May Thirtieth county has the relevant organization, the comparison expands to second-order adjacent May Thirtieth counties. Similarity to the nearest county organization equals one if the organization type matches that of the geographically nearest county with an existing party organization. Common categories include party branches, party small groups, dispatched organs, and party-group organizations. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

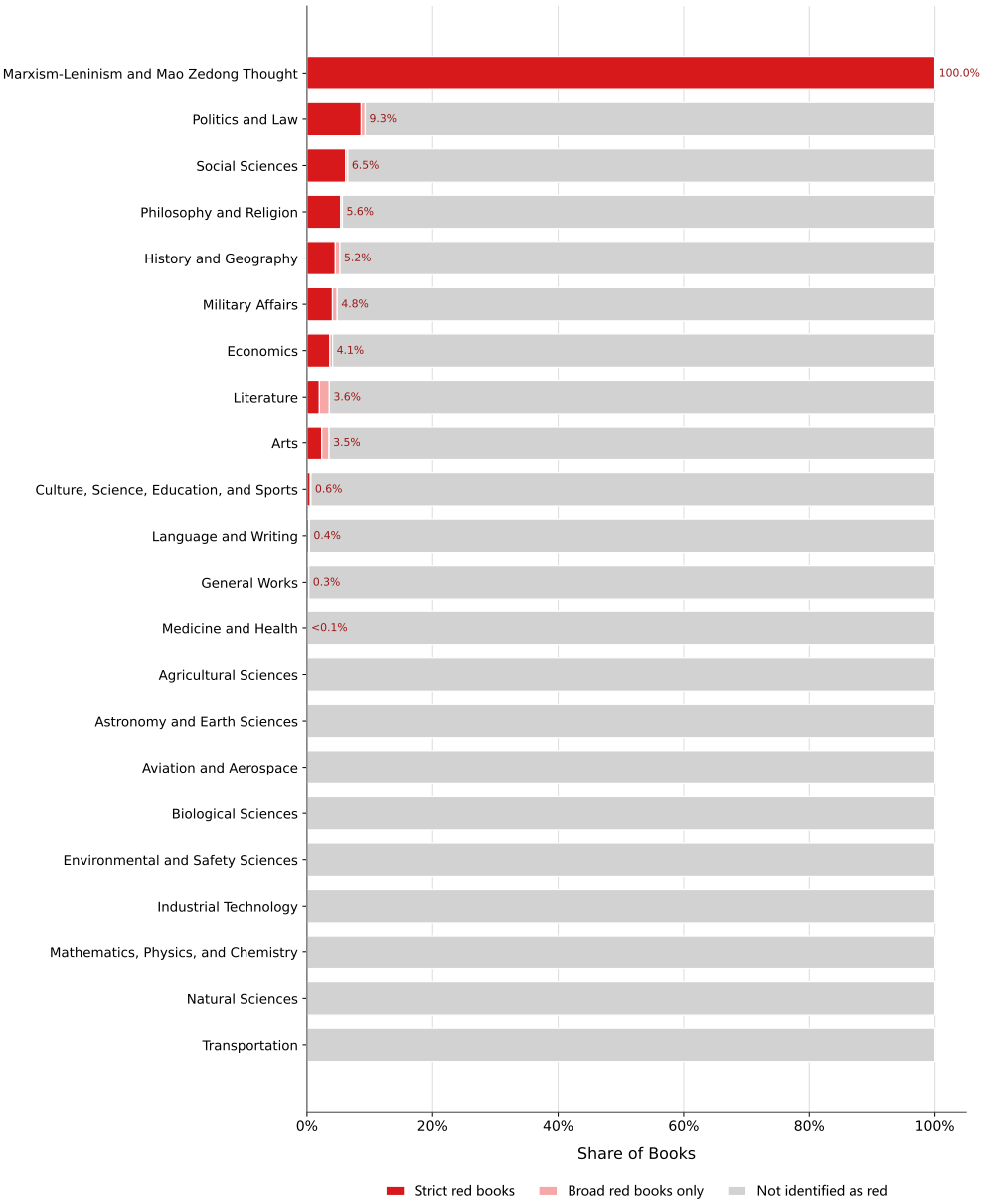


Figure A12. Red-Book Category Shares by First Category

Note: Using the NLC Republican-era Books Database categories, this figure reports, within each first-category class, the share of books classified as strict-definition red books and the additional share classified only under the broad red-book definition. Back to Section V.A..

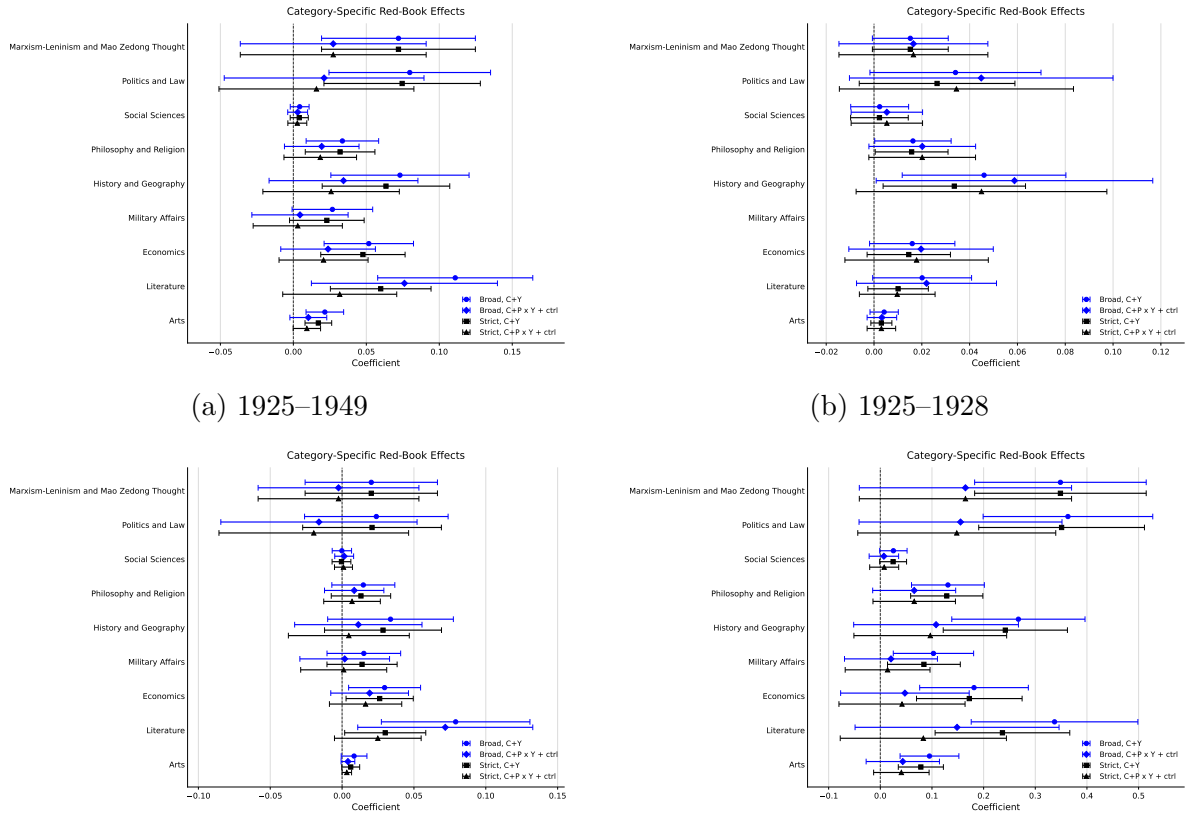


(a) Chinese terms with readable English approximation

(b) English approximation

Figure A13. Top 100 Strict-Definition Red-Book Terms: Word Clouds

Note: This figure plots word clouds using the 100 most frequent terms extracted from the titles and abstracts of strict-definition red books. Back to Section V.A..



(a) 1925–1949

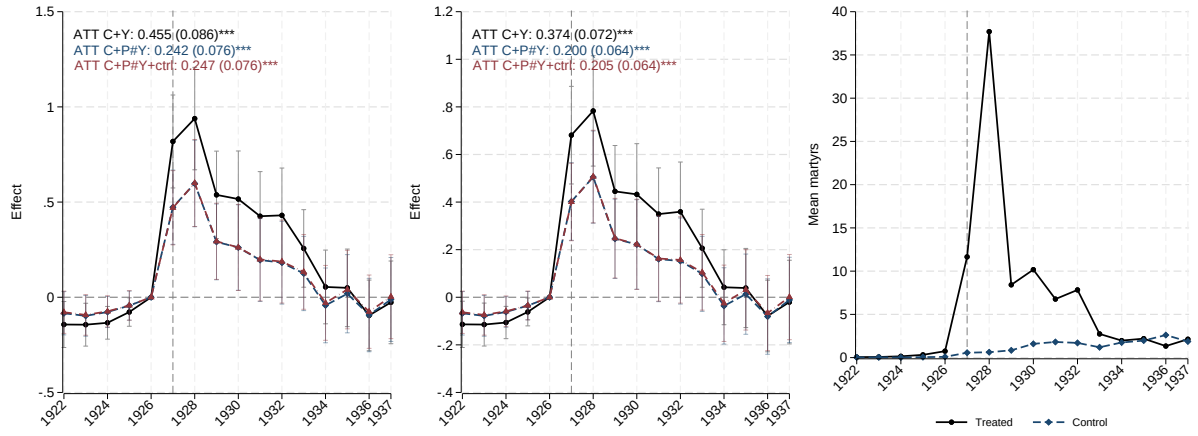
(b) 1925–1928

(c) 1929–1945

(d) 1946–1949

Figure A14. Red-Book Category Coefficients by Period

Note: This figure reports four sets of coefficients: strict red books and broad-definition red books, each estimated with city and year fixed effects and with city and province-by-year fixed effects. Bars show 95% confidence intervals, with standard errors clustered by city. Back to Section V.A..



(a) asinh count (b) log(1+count) (c) Group means

Figure A15. County Committees Active in 1927 and Martyr Counts Around 1927

Note: This figure reports martyr counts around 1927 in counties with active county committees. Panel (a) reports estimates using the inverse hyperbolic sine transformation, panel (b) reports estimates using log(1+count), and panel (c) reports group means. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section V.C..

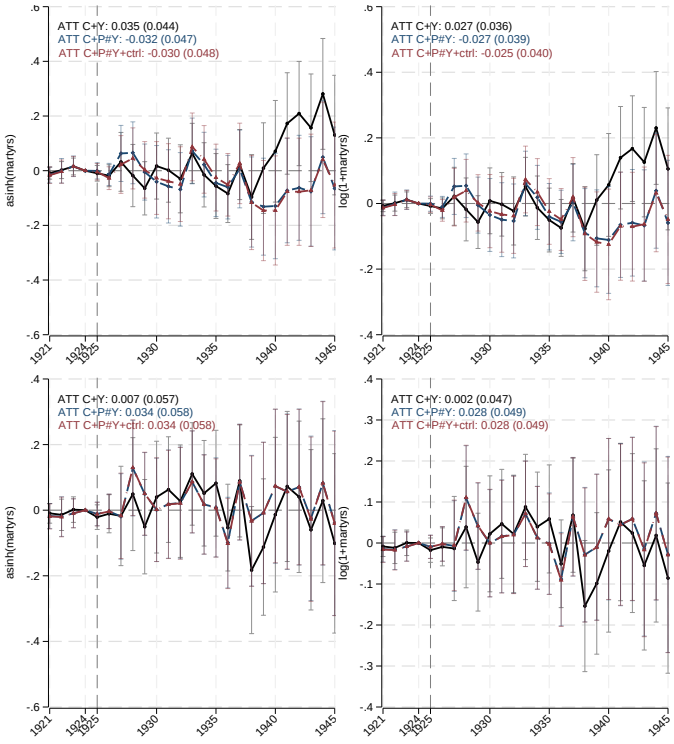
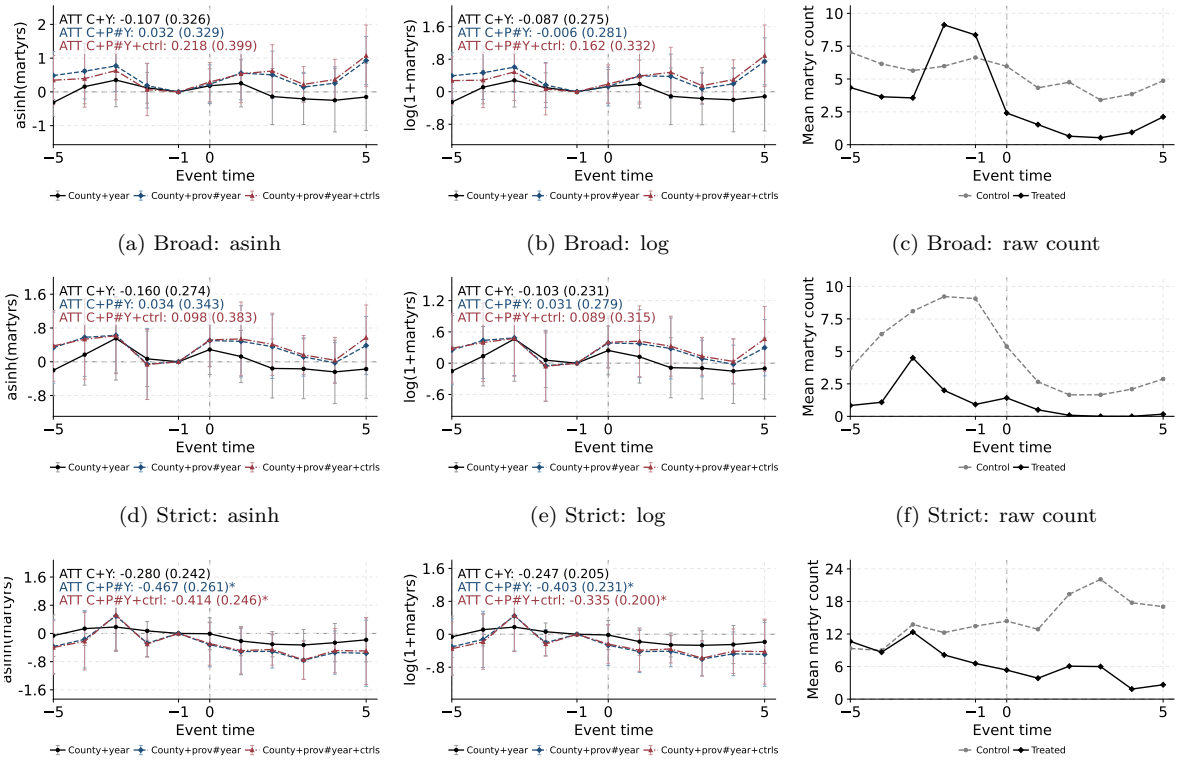


Figure A16. Purge Intensity by May Thirtieth Status After Controlling for County-Committee Status

Note: This figure controls for whether the county had a county committee, separating the martyr increase associated with May Thirtieth exposure from the increase associated with more developed party organizations. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section V.C..



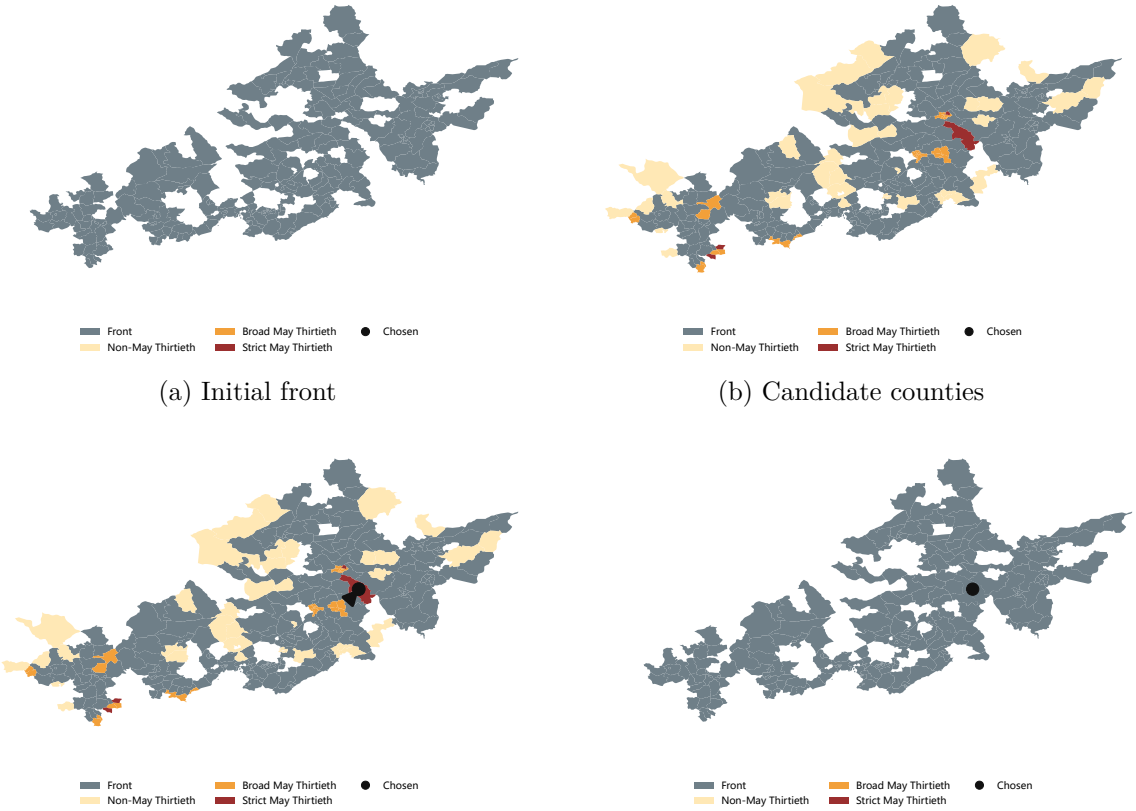
(a) Broad: asinh (b) Broad: log (c) Broad: raw count

(d) Strict: asinh (e) Strict: log (f) Strict: raw count

(g) Non-May Thirtieth: asinh (h) Non-May Thirtieth: log (i) Non-May Thirtieth: raw count

Figure A17. Martyr Counts Around County-Committee Abolition: By May Thirtieth Status

Note: This figure reports changes in martyr counts around county-committee abolition for broad May Thirtieth counties, strict May Thirtieth counties, and non-May Thirtieth counties. The first two columns plot stacked difference-in-differences event-study estimates using the inverse hyperbolic sine and log(1+count) transformations. The right column plots untransformed annual mean martyr counts for stacked treated and control counties. Event time is defined relative to county-committee abolition, the window is -5 to 5, and event time -1 is the omitted baseline. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section V.C..



(a) Initial front

(b) Candidate counties

(c) Chosen county

(d) Updated front

Figure A18. Choice of CCP Capture Timing

Note: This figure illustrates how we use existing CCP-held zones to identify the next capture event and determine whether the newly captured area was covered by the May Thirtieth Movement. Panel (a) shows the initial CCP-held front. In panel (b), the non-black areas are counties adjacent to that front that the CCP had not yet captured. In panel (c), the black point marks the county captured on the next date. Panel (d) adds the newly captured county to the connected CCP-held zone, after which the procedure repeats. Back to Section VI.A..

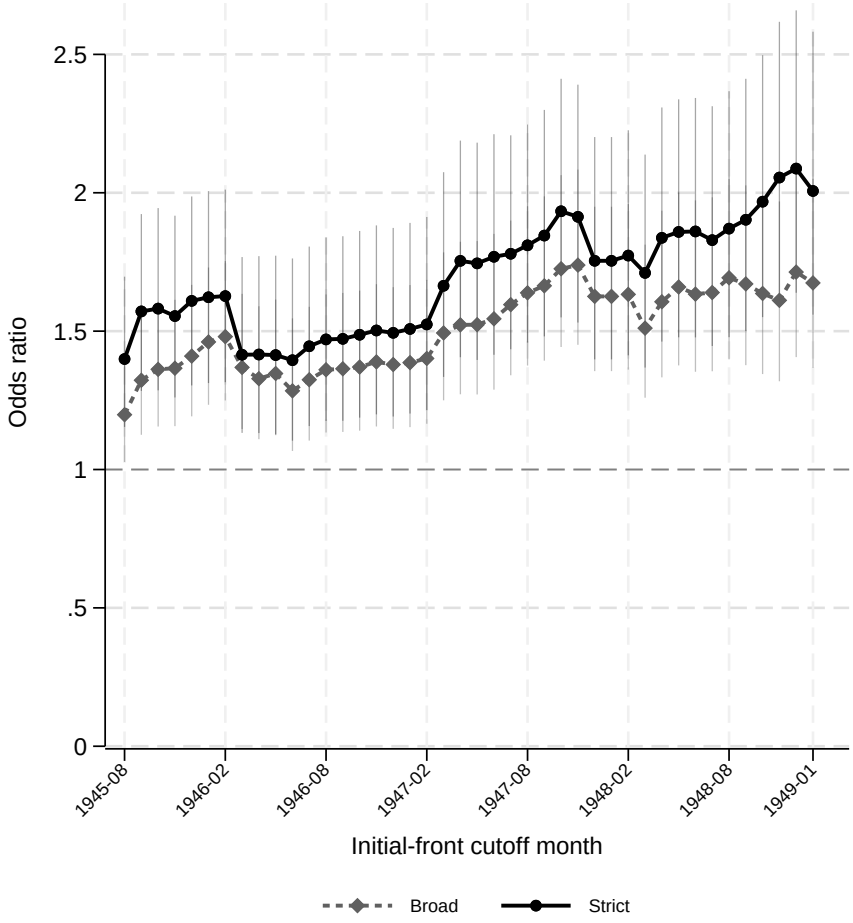


Figure A19. Seed-Cutoff Sensitivity

Note: For each alternative seed cutoff, this figure estimates the relative odds that the next county captured by the CCP was a May Thirtieth county. At each frontier step j, let n_j denote the number of candidate counties adjacent to the current CCP-held front, m_j the number of May Thirtieth counties among those candidates, k_j the number of counties the CCP actually captured in the next step, and y_j the number of May Thirtieth counties among the realized choices. We parameterize the relative selection odds by $\eta = \log(\theta)$, where $\theta = 1$ corresponds to equal selection odds for May Thirtieth and non-May Thirtieth counties. Conditional on n_j, m_j, and k_j, the realized number of May Thirtieth choices follows a noncentral hypergeometric likelihood, $P(Y_j = y) \propto \binom{m_j}{y}\binom{n_j - m_j}{k_j - y}\exp(\eta y)$. This likelihood compares May Thirtieth and non-May Thirtieth candidates within the same local frontier choice set, holding fixed the size and composition of each realized choice opportunity. The estimate $\hat{\eta}$ solves the conditional likelihood score equation, $\sum_j y_j = \sum_j E_\eta(Y_j)$, where $E_\eta(Y_j)$ is computed under the noncentral hypergeometric distribution for frontier step j. The reported odds ratio is $\hat{\theta} = \exp(\hat{\eta})$. The standard error is computed from the inverse Fisher information on the log-odds scale, $I(\hat{\eta}) = \sum_j Var_{\hat{\eta}}(Y_j)$, so that $SE(\hat{\eta}) = 1/\sqrt{I(\hat{\eta})}$. Back to Section VI.A..

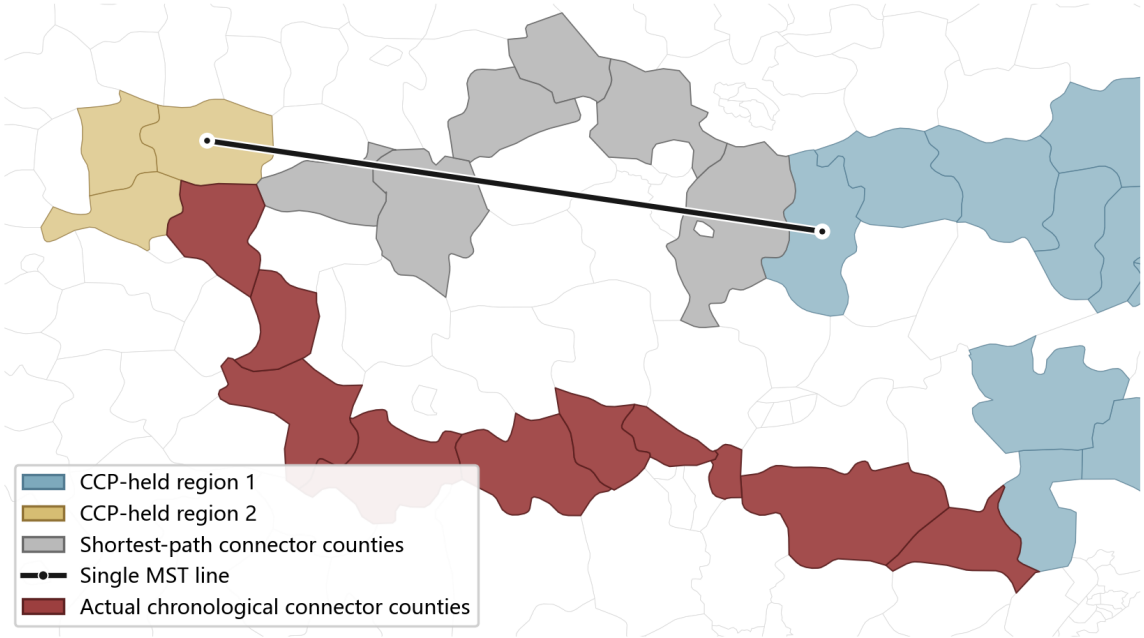
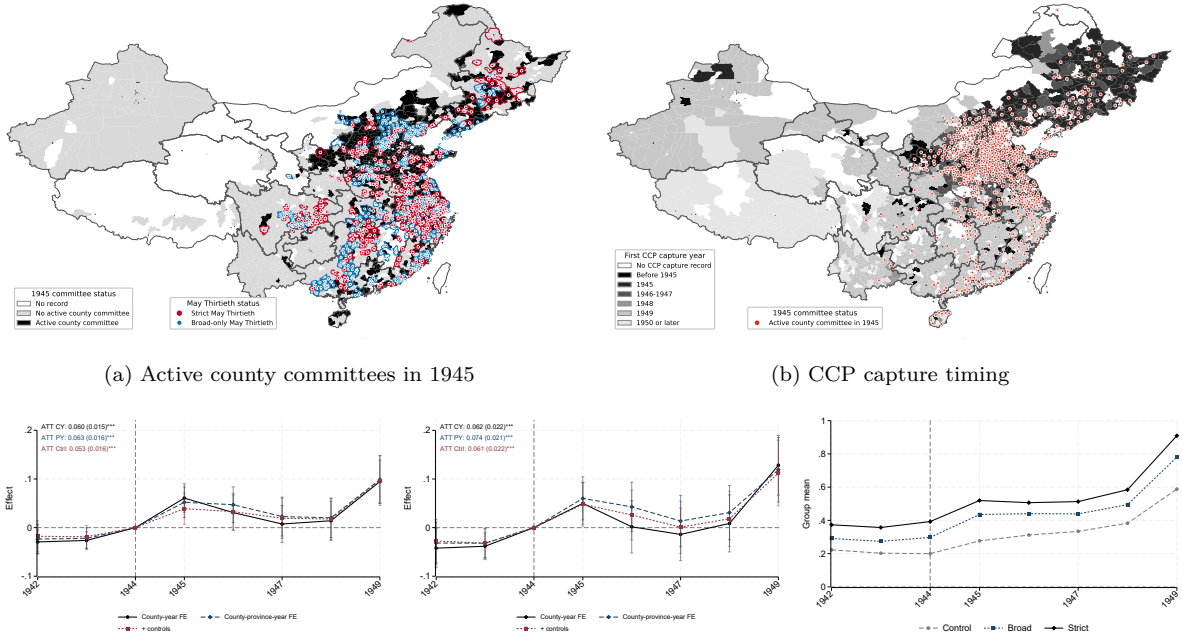


Figure A20. Actual Path and MST Link Example

Note: This figure reports three examples. The black line is the shortest distance between two CCP-held areas computed using the minimum-spanning-tree (MST) algorithm. The red area is the path through which the two CCP-held areas first became linked. The gray area shows the MST before the two CCP-held areas became linked in practice, allowing for boundary shifts before the actual connection. Back to Section VI.A..



(a) Active county committees in 1945

(b) CCP capture timing

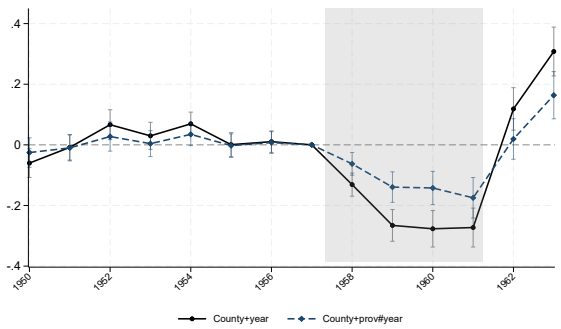
(c) Event study: broad May Thirtieth

(d) Event study: strict May Thirtieth

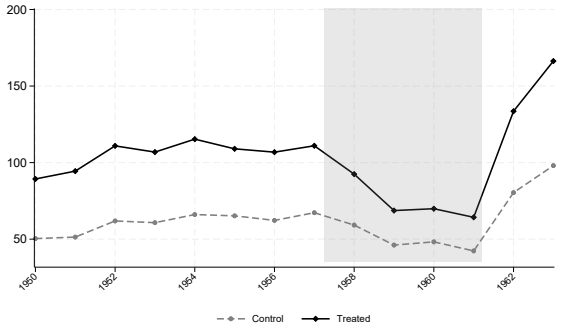
(e) Group means

Figure A21. May Thirtieth Movement and County Committee Establishment: Civil War Period

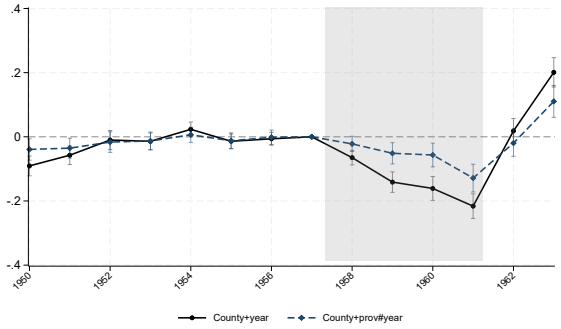
Note: Panels (a) and (b) plot the distribution of county committees in 1945 and the distribution of CCP capture timing, respectively. The remaining three panels examine county-committee growth in May Thirtieth areas during the civil war. Back to Section VI.A..



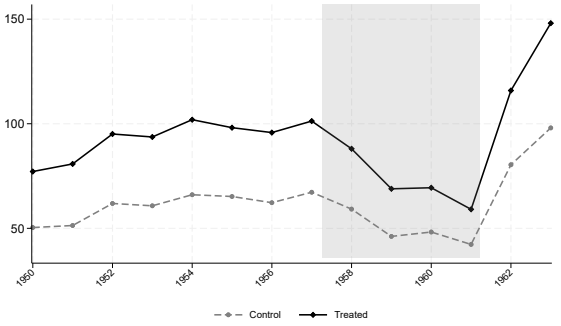
(a) Event study: strict May Thirtieth



(b) Group means: strict May Thirtieth



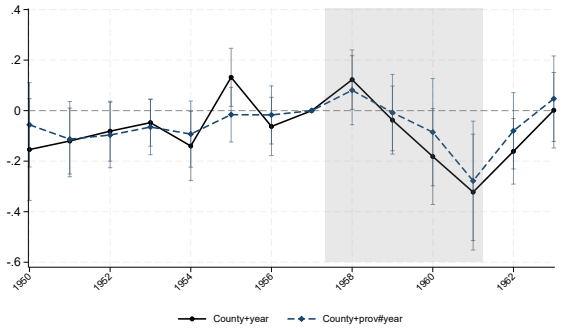
(c) Event study: broad May Thirtieth



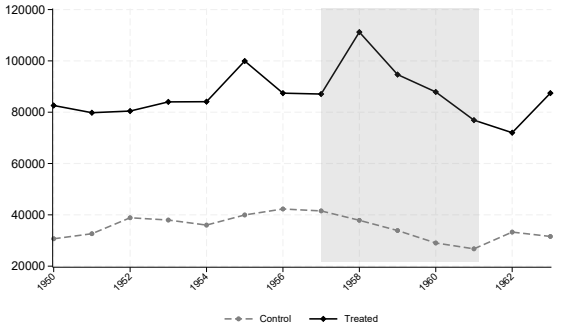
(d) Group means: broad May Thirtieth

Figure A22. May Thirtieth Areas and Births During the Great Famine

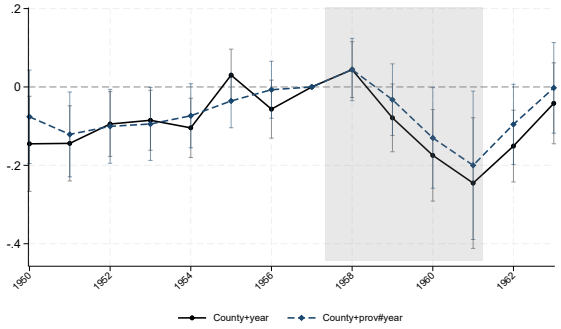
Note: This figure examines the Great Famine of 1959–1961, which occurred almost contemporaneously with the Great Leap Forward. We measure the shock using changes in birth counts. The event-study coefficients in the left panels are reported as ratios, while the descriptive group-mean panels on the right are measured in persons. The data come from the 1990 Chinese population census, from which we construct a county-level panel by birth cohort. Back to Section VI.B..



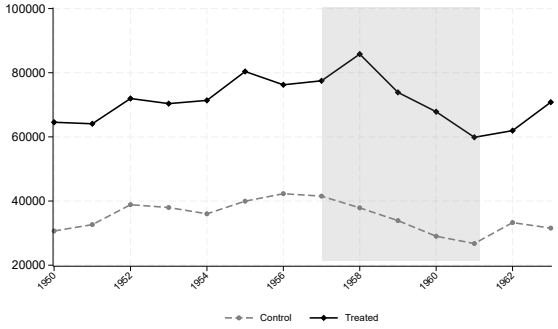
(a) Event study: strict May Thirtieth



(b) Group means: strict May Thirtieth



(c) Event study: broad May Thirtieth



(d) Group means: broad May Thirtieth

Figure A23. May Thirtieth Movement and Grain Output, 1950–1963

Note: This figure examines whether May Thirtieth areas experienced a sharper decline in agricultural output during 1958–1961. Grain-output data come from county gazetteers and are measured in tons. Panels (b) and (d) report descriptive group means for May Thirtieth and non-May Thirtieth areas. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section VI.B..

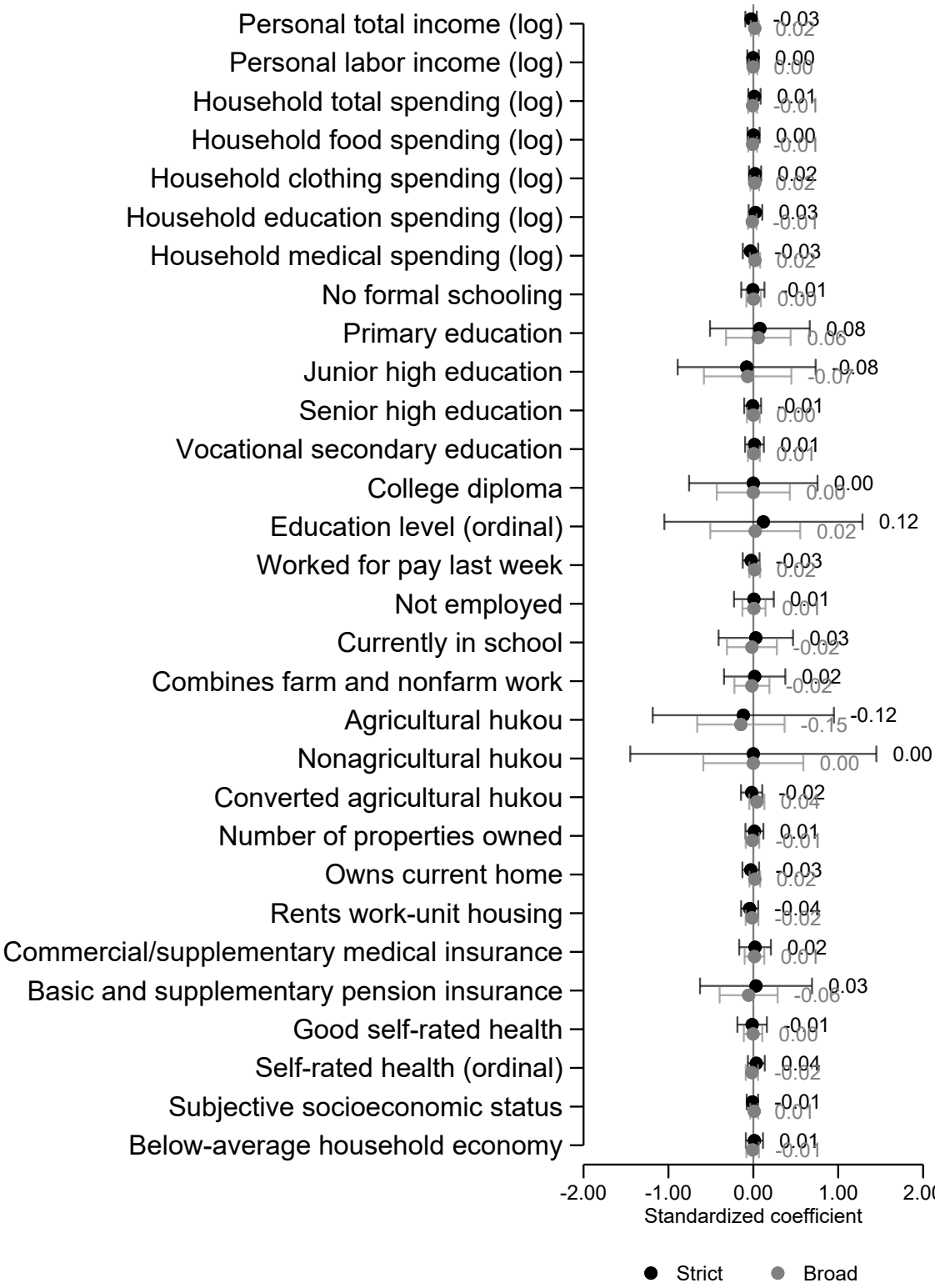


Figure A24. May Thirtieth Areas and Objective Status in the Twenty-First Century

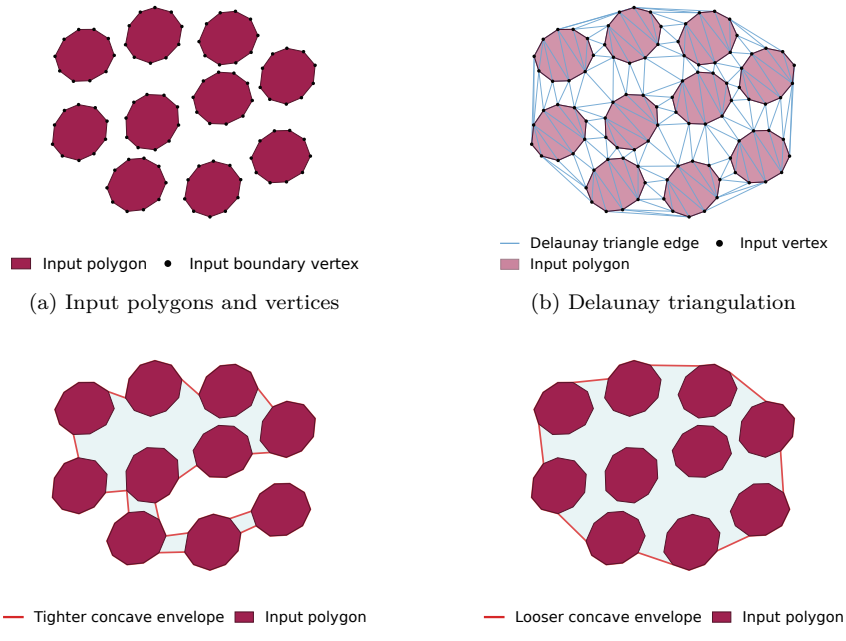
Note: The data come from two large social surveys: one conducted in 2018 covering 135 counties, and another conducted in 2013 and 2015 covering 153 counties. Together, they include 22,204 respondents. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section VI.B..

Figure A26 illustrates how we determine the boundaries of May Thirtieth areas using a concave-hull computational-geometry approach (Edelsbrunner et al., 1983). Panel (a) plots the polygon vertices of different counties (in GIS, county boundaries are likewise represented by line segments connecting a large number of vertices). In Panel (b), we construct the Delaunay triangulation over all vertices, such that the circumcircle of each triangle contains no other input point. If all triangles are retained, the outer edges of the outermost triangles form a region that provides the most expansive boundary. In practice, such a boundary would encompass nearly all counties. We therefore first exclude singleton counties located more than 100 km from any other contiguous cluster. We then use a ratio parameter r to peel triangles inward from the outer edge. Specifically, the total edge length L of the retained triangles is required to remain above L^*, where

$$L^* = L_{\min} + r(L_{\max} - L_{\min}).$$

A larger r implies that fewer triangles are removed. $L_{\max}$ corresponds to retaining all triangles, whereas $L_{\min}$ yields a tight region containing only May Thirtieth counties, making the resulting May Thirtieth area extremely narrow. Panels (c) and (d) illustrate the boundaries generated by a relatively small and a relatively large ratio, respectively, with the outer envelope defining the boundary of the May Thirtieth area.

After determining the boundary for a given ratio, we classify a county as belonging to the region if its centroid falls within the constructed boundary, and then delineate each contiguous May Thirtieth region using the actual boundaries of its outermost counties. We then remove boundary segments with no counties on their exterior side, which occurs mainly along coastlines. Our baseline specification sets $r = 0.08$, a heuristic choice. In Appendix Figure A26, we also consider ratios of 0.07, 0.09, and 0.10. With a smaller ratio, the boundary becomes sharply indented in many places and collapses into a narrow contour. This brings many May Thirtieth counties into the boundary comparison and undermines the purpose of comparing the outermost May Thirtieth counties. With a larger ratio, as shown in Appendix Figure A26, Panel (i), the interior side of the boundary contains few May Thirtieth counties, and the resulting boundary is effectively supported by only a few peripheral May Thirtieth counties. The regressions include 975 counties when $r = 0.07$ and 857 counties when $r = 0.10$.



(a) Input polygons and vertices

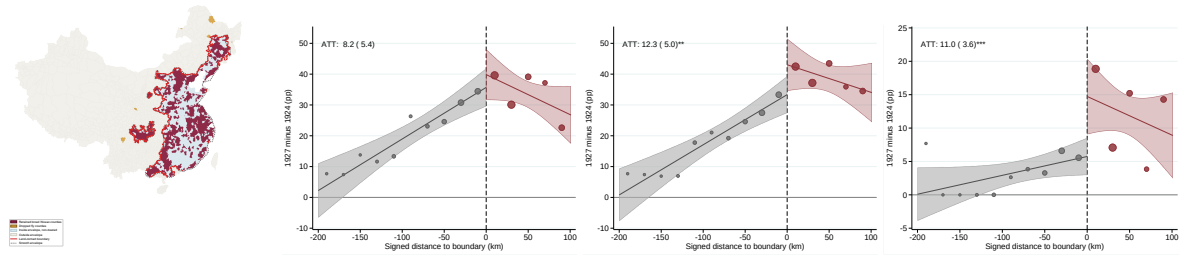
(b) Delaunay triangulation

(c) More inward concave-hull boundary

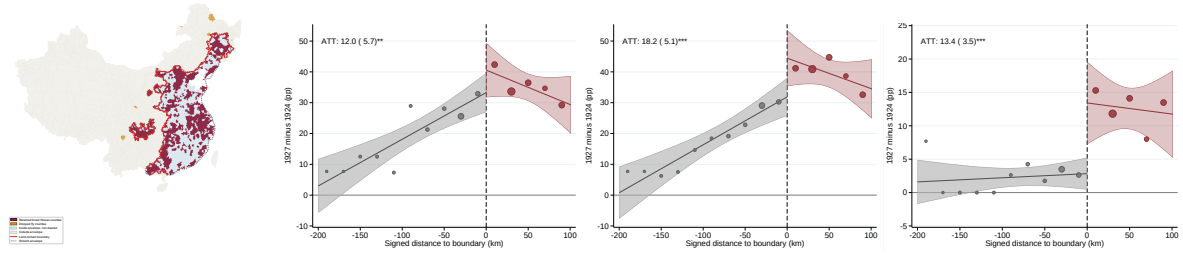
(d) Looser concave-hull boundary

Figure A25. Illustration of Concave-Hull Boundary Construction

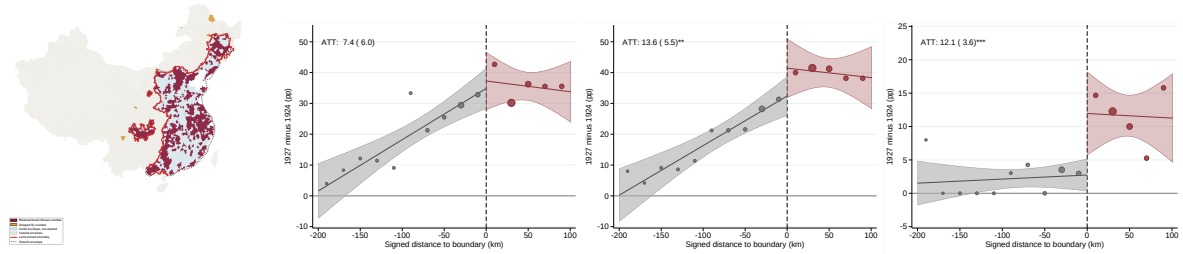
Note: Back to Section IV.B.



(a) Ratio 0.07: boundary (b) Ratio 0.07: member (c) Ratio 0.07: organization (d) Ratio 0.07: committee



(e) Ratio 0.09: boundary (f) Ratio 0.09: member (g) Ratio 0.09: organization (h) Ratio 0.09: committee



(i) Ratio 0.10: boundary (j) Ratio 0.10: member (k) Ratio 0.10: organization (l) Ratio 0.10: committee

Figure A26. Spatial Boundary RD Sensitivity to the Concave-Hull Boundary Parameter

Note: This figure examines the sensitivity of the spatial regression discontinuity design to alternative values of the concave-hull boundary parameter. Each row reports a different value (0.07 in the top row, 0.09 in the middle row, and 0.10 in the bottom row). The first column maps the corresponding boundary sample, and the remaining columns present RD plots for CCP member activity, party organizations, and county party committees. The bandwidth is 200 km outside the boundary and 100 km inside the boundary. The intervals denote 95% confidence intervals. Back to Section IV.B..

Appendix C. Robustness Checks

This appendix reports robustness checks discussed in Section IV.B.3..

In constructing the counterfactual paths in Appendix Figure A27, county i belongs to group $g \in \{0, 1\}$, where $G_i = 1$ denotes the May Thirtieth treated group and $G_i = 0$ denotes the control group. The outcome Y_{it} is an absorbing indicator. The event time T_i is the first year in which the relevant event occurs: the first appearance of CCP members, the establishment of a party organization, or the first establishment of a county party committee. Let t denote calendar year. Then

$$T_i = \min\{t : Y_{it} = 1\}. \tag{A1}$$

The survival probability $S_{g,t}$ is the share of counties in group g that have not yet experienced the event by year t:

$$S_{g,t} = \Pr(T_i > t \mid G_i = g). \tag{A2}$$

The cumulative incidence $F_{g,t}$ is the corresponding event-experienced share,

$$F_{g,t} = 1 - S_{g,t}. \tag{A3}$$

The observed treated and control cumulative means plotted in the figure are $F_{1,t}$ and $F_{0,t}$.

The cumulative-DID counterfactual shifts the control group's cumulative-incidence path by the average pre-period treated-control gap. We estimate this gap over 1922–1924:

$$\hat{c}^{\text{cum}} = \frac{1}{3} \sum_{t=1922}^{1924} \left(F_{1,t} - F_{0,t}\right). \tag{A4}$$

For years after the May Thirtieth shock, the treated group's cumulative-incidence counterfactual is

$$\widehat{F}_{1,t}^{0,\text{cum}} = F_{0,t} + \hat{c}^{\text{cum}}. \tag{A5}$$

This construction treats the pre-period level difference in cumulative incidence as the treated-control gap that would have prevailed after 1925 in the absence of the May Thirtieth shock.

The hazard-DID counterfactual first maps cumulative incidence into a time-average hazard and then constructs the no-May Thirtieth path in hazard space. For $t > 1921$, the time-average hazard is

$$H_{g,t} = \frac{1}{t - 1921} \log\left(\frac{1 - F_{g,1921}}{1 - F_{g,t}}\right). \tag{A6}$$

Equivalently,

$$H_{g,t} = -\frac{\log S_{g,t} - \log S_{g,1921}}{t - 1921}. \tag{A7}$$

The hazard $H_{g,t}$ measures the average speed at which counties still at risk in 1921 are absorbed between 1921 and t. We estimate the pre-period hazard gap over 1922–1924:

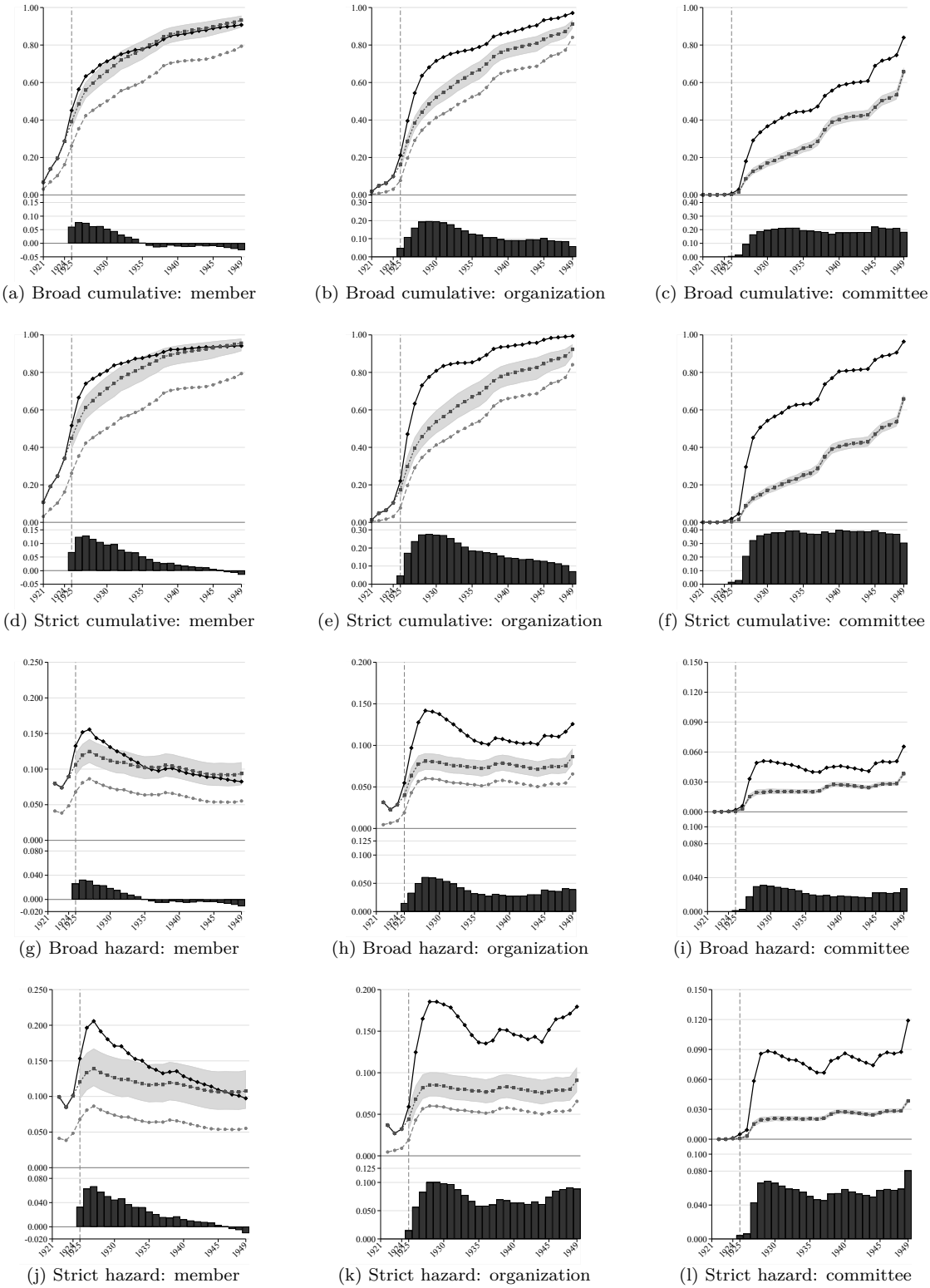
$$\hat{c}^{H} = \frac{1}{3} \sum_{t=1922}^{1924} \left(H_{1,t} - H_{0,t}\right). \tag{A8}$$

For years after the May Thirtieth shock, the treated group's hazard counterfactual is

$$\widehat{H}_{1,t}^{0} = H_{0,t} + \hat{c}^{H}. \tag{A9}$$

Finally, we invert the hazard counterfactual back into cumulative incidence:

$$\widehat{F}_{1,t}^{0,H} = 1 - (1 - F_{1,1921}) \exp\left[-(t - 1921)\widehat{H}_{1,t}^{0}\right]. \tag{A10}$$



(a) Broad cumulative: member (b) Broad cumulative: organization (c) Broad cumulative: committee

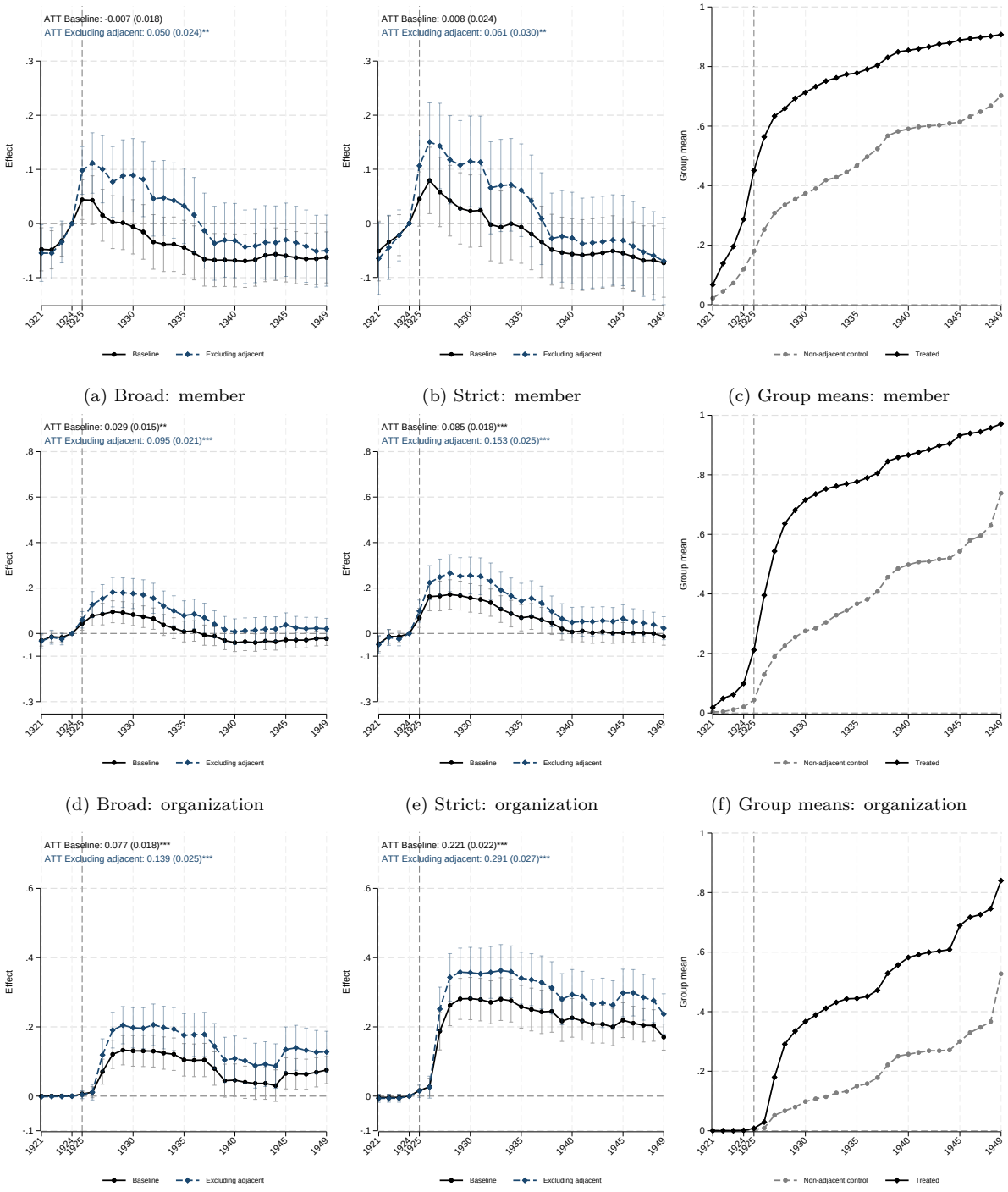
(d) Strict cumulative: member (e) Strict cumulative: organization (f) Strict cumulative: committee

(g) Broad hazard: member (h) Broad hazard: organization (i) Broad hazard: committee

(j) Strict hazard: member (k) Strict hazard: organization (l) Strict hazard: committee

Figure A27. Counterfactual and Hazard-Model Checks for Cumulative Outcomes

Note: Following the logic of Deaner and Ku (2024), this figure plots counterfactual paths for cumulative outcomes and the corresponding hazard-adjusted counterfactuals. Rows report broad cumulative outcomes, strict cumulative outcomes, broad hazard specifications, and strict hazard specifications. Columns report member activity, party organizations, and county committees. The bars show the difference between the treated group and its counterfactual. Back to Section IV.B..



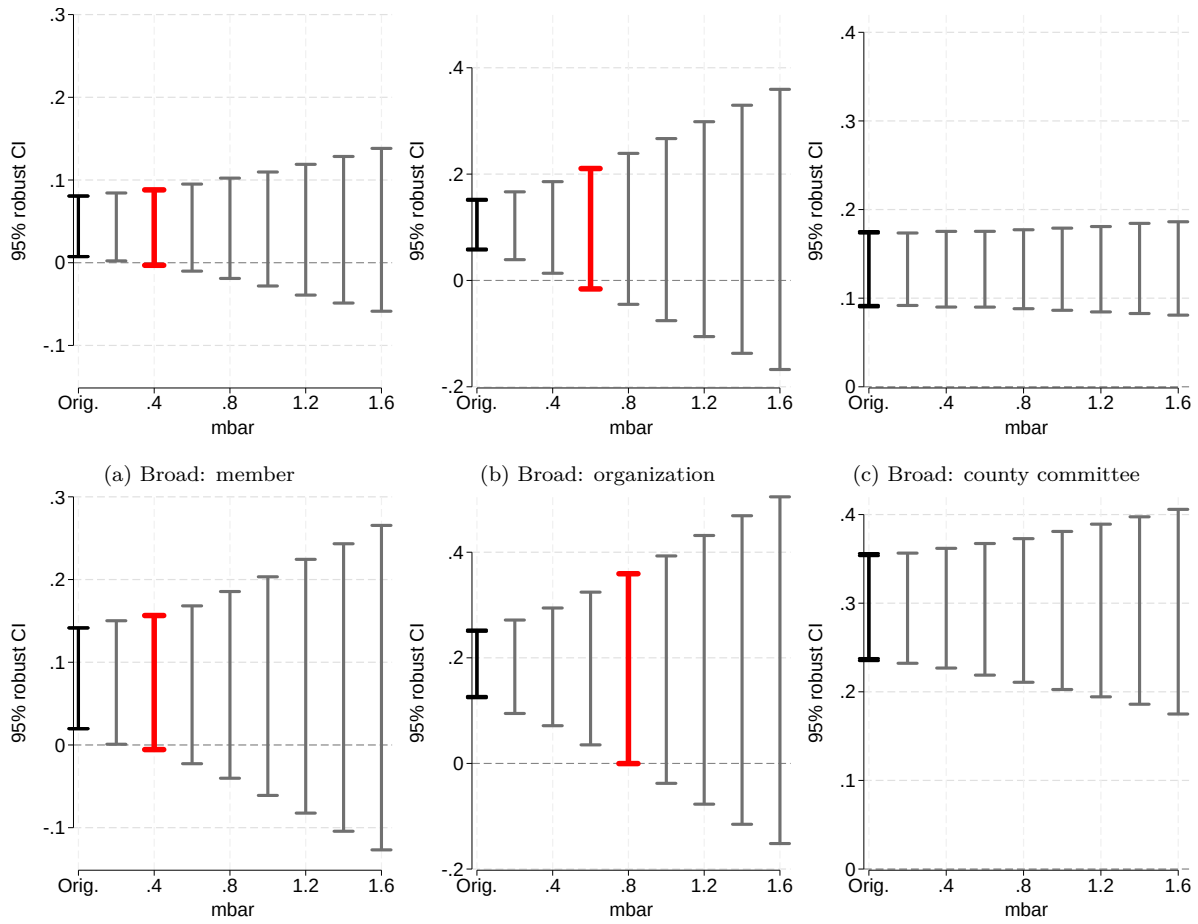
(a) Broad: member (b) Strict: member (c) Group means: member

(d) Broad: organization (e) Strict: organization (f) Group means: organization

(g) Broad: county committee (h) Strict: county committee (i) Group means: county committee

Figure A28. May Thirtieth Movement and CCP Development: Excluding Adjacent Counties

Note: This figure follows Figure 7. In the first two columns, black curves plot the baseline estimates and blue curves rerun the same specification after excluding first-order counties adjacent to May Thirtieth counties. The right column reports group means for May Thirtieth counties under the broad definition and for non-adjacent control counties. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

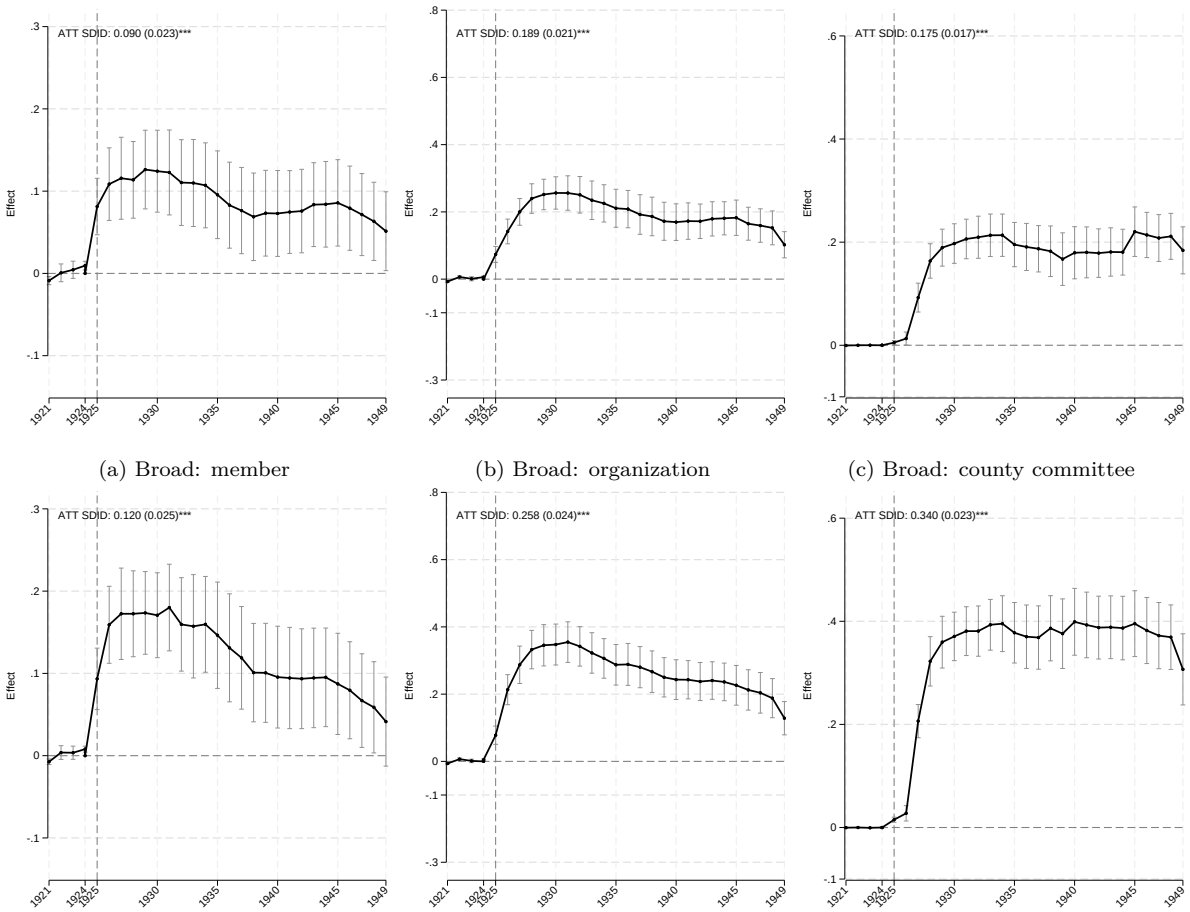


(a) Broad: member (b) Broad: organization (c) Broad: county committee

(d) Strict: member (e) Strict: organization (f) Strict: county committee

Figure A29. Sensitivity to Violations of Parallel Trends

Note: This figure reports the relative-magnitude sensitivity analysis recommended by Rambachan and Roth (2023). For each outcome, we evaluate the post-treatment period with the largest estimated effect: panel (a) 1925, panel (b) 1928, panel (c) 1929, panel (d) 1926, panel (e) 1928, and panel (f) 1930. Black intervals report the original estimates, gray intervals report estimates as the allowed violation of parallel trends is gradually relaxed, and the red interval marks the first confidence interval that crosses zero. Intervals are 95% confidence intervals. Back to Section IV.B..

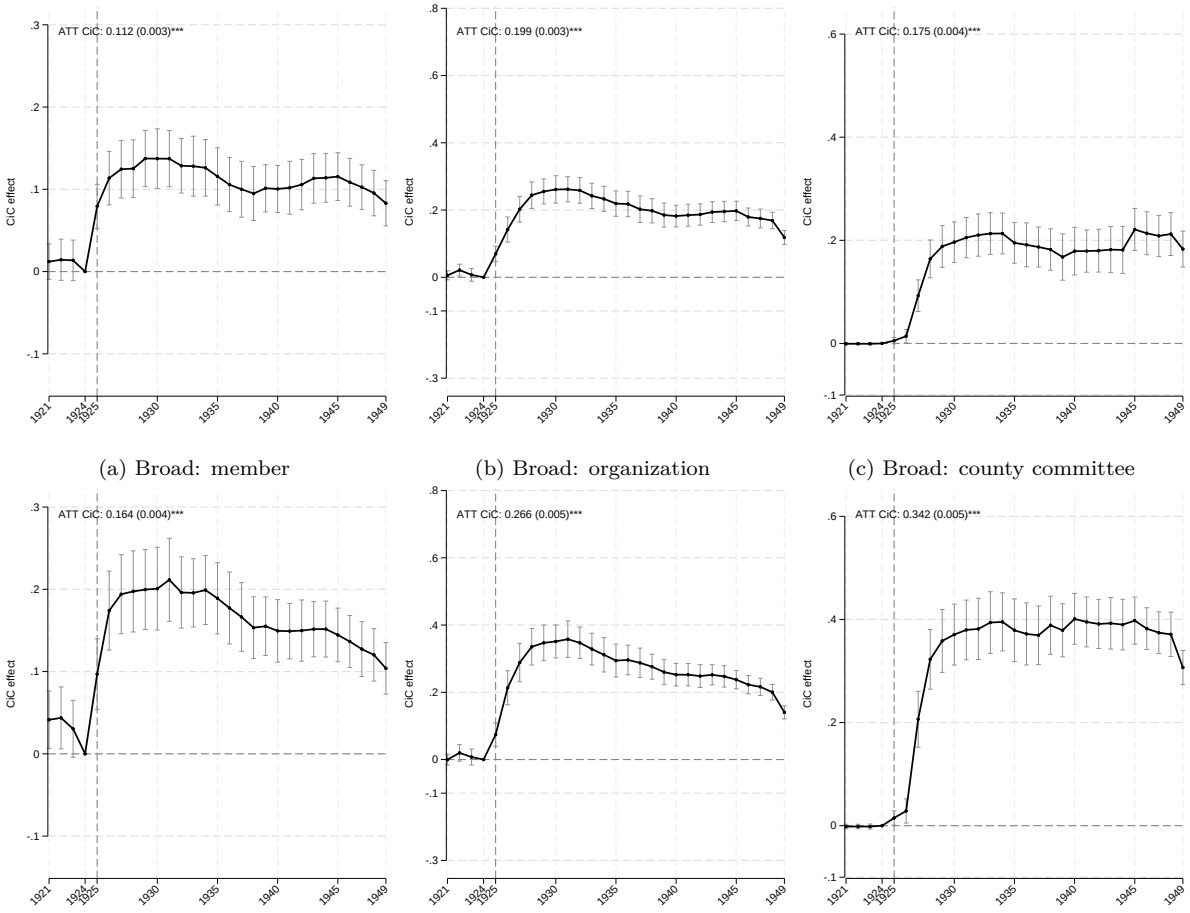


(a) Broad: member (b) Broad: organization (c) Broad: county committee

(d) Strict: member (e) Strict: organization (f) Strict: county committee

Figure A30. Synthetic Difference-in-Differences (Arkhangelsky et al., 2021)

Note: This figure reestimates the baseline CCP development outcomes using the synthetic difference-in-differences estimator. Panels report broad and strict May Thirtieth specifications for CCP member activity, party organizations, and county committees. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

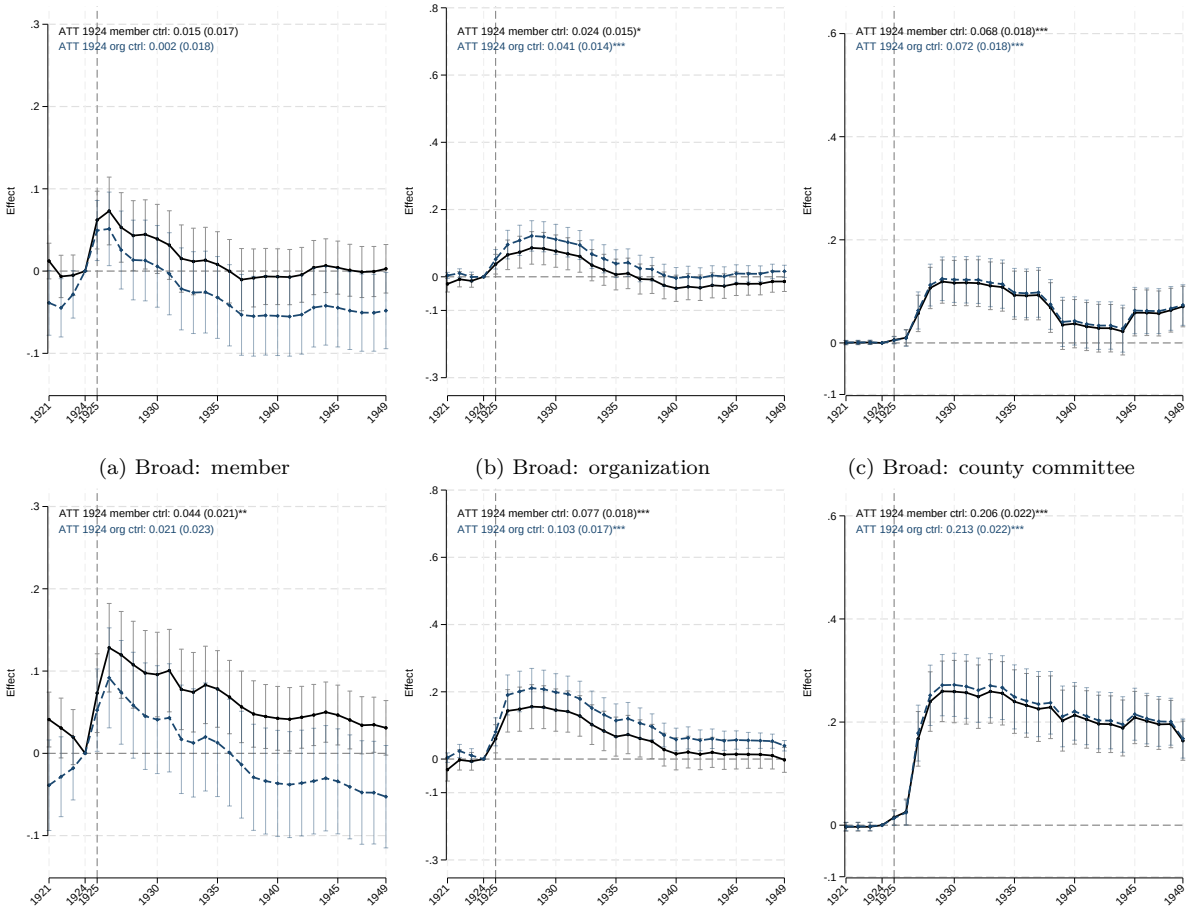


(a) Broad: member (b) Broad: organization (c) Broad: county committee

(d) Strict: member (e) Strict: organization (f) Strict: county committee

Figure A31. Changes-in-Changes (Athey and Imbens, 2006)

Note: This figure reestimates the baseline CCP development outcomes using the changes-in-changes estimator. Panels report broad and strict May Thirtieth specifications for CCP member activity, party organizations, and county committees. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

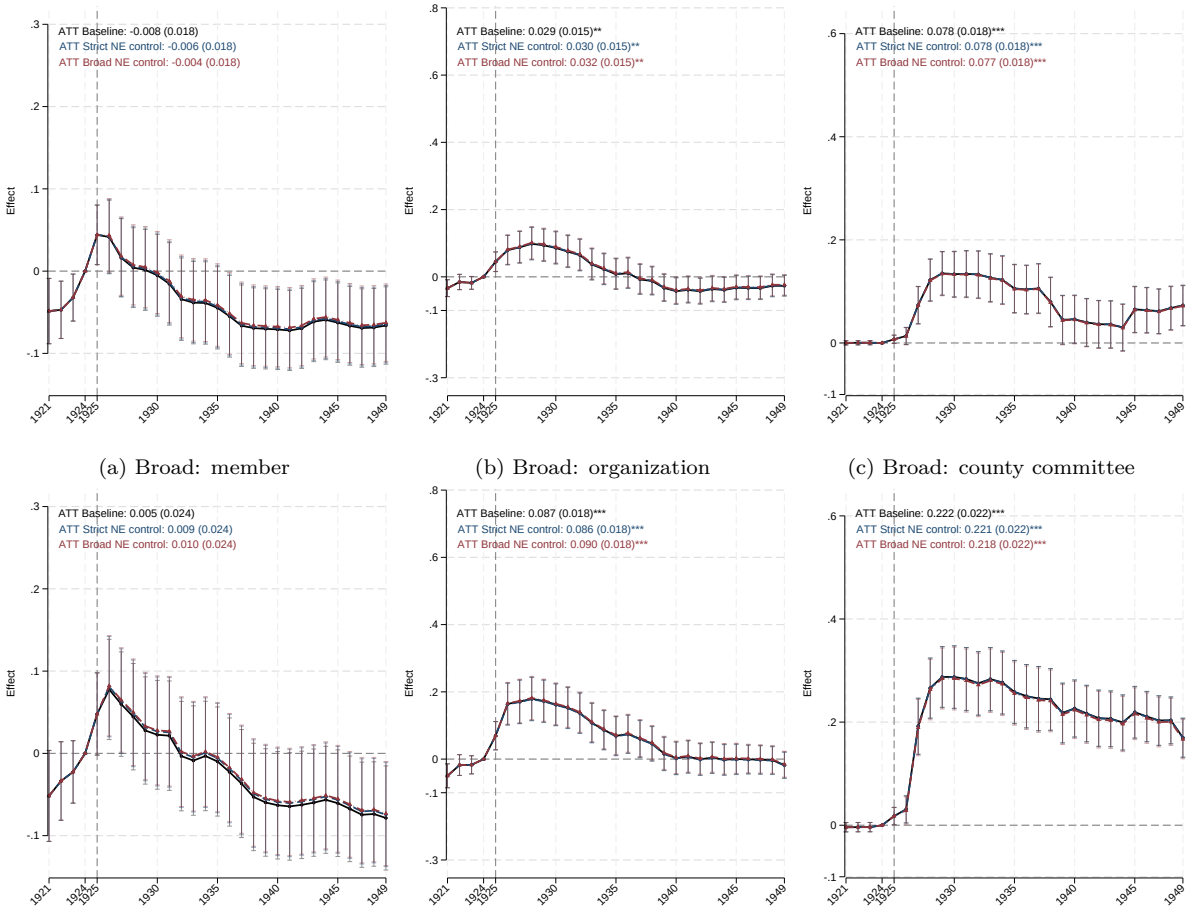


(a) Broad: member (b) Broad: organization (c) Broad: county committee

(d) Strict: member (e) Strict: organization (f) Strict: county committee

Figure A32. Controlling for CCP Member Presence in 1924

Note: This figure adds controls for interactions between year fixed effects and indicators for whether a county had CCP member activity or a party organization in 1924. Baseline estimates are included for comparison. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

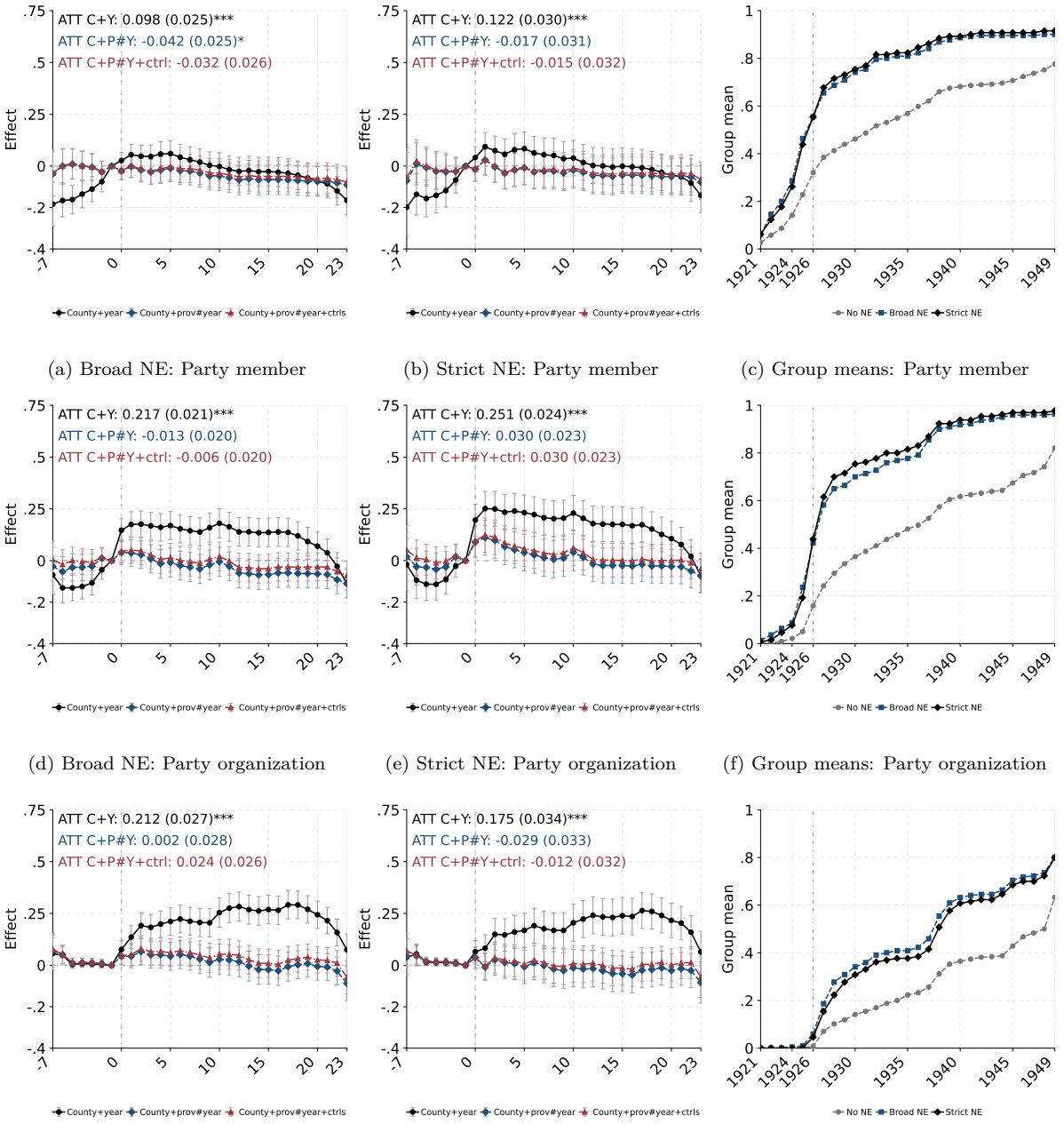


(a) Broad: member (b) Broad: organization (c) Broad: county committee

(d) Strict: member (e) Strict: organization (f) Strict: county committee

Figure A33. Controlling for Northern Expedition Exposure

Note: This figure reports how estimates change after controlling for Northern Expedition exposure. Battle-location data come from volumes 12–21 of *Revolutionary Documents* (Party History Materials Compilation Committee, Central Executive Committee of the Kuomintang, 1956–1959). We use two exposure definitions for 1926–1928: recorded battle locations, and a broader measure that also includes counties along the minimum-distance paths linking temporally adjacent battles. Both definitions are coded by the year in which the corresponding Northern Expedition exposure occurred. Baseline estimates are included for comparison. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



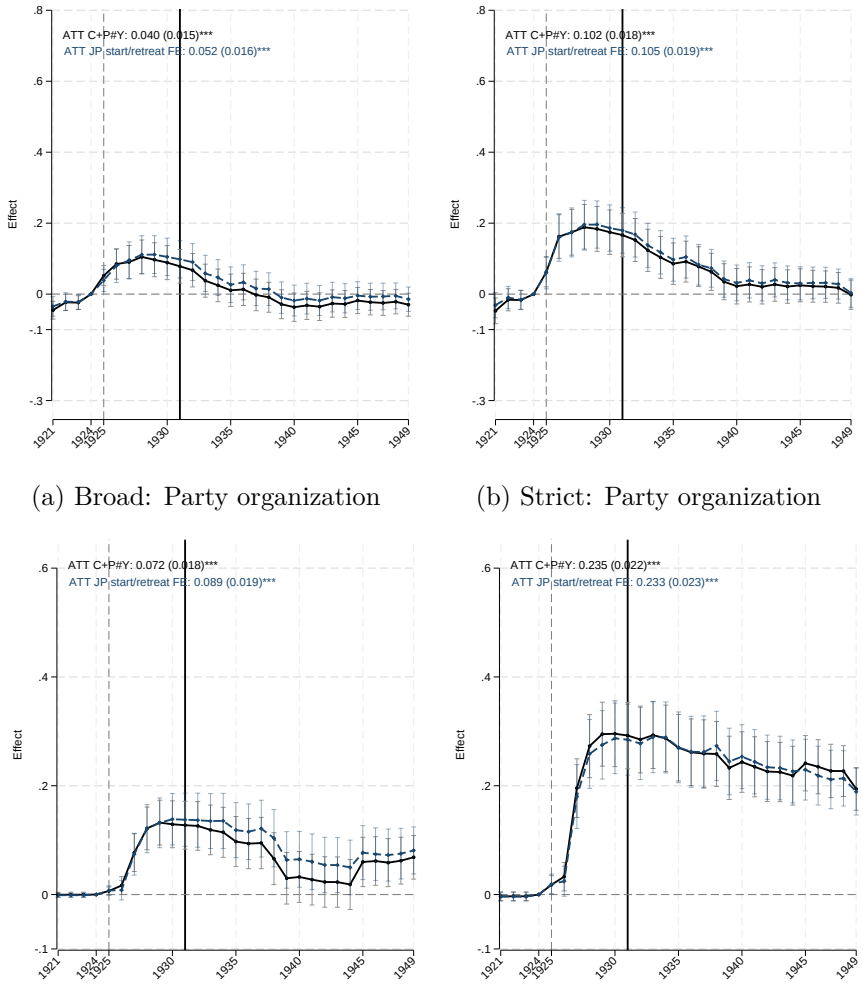
(a) Broad NE: Party member (b) Strict NE: Party member (c) Group means: Party member

(d) Broad NE: Party organization (e) Strict NE: Party organization (f) Group means: Party organization

(g) Broad NE: County party committee (h) Strict NE: County party committee (i) Group means: County party committee

Figure A34. Northern Expedition and CCP Development

Note: This figure reports the effect of Northern Expedition exposure on CCP development after excluding counties exposed to the broad May Thirtieth definition. Because the Northern Expedition occurred over 1926–1928, the horizontal axis is relative event time. We use two exposure definitions: recorded battle locations, and a broader measure that also includes counties along the minimum-distance paths linking temporally adjacent battles. Both definitions are coded by the year in which the corresponding Northern Expedition exposure occurred. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

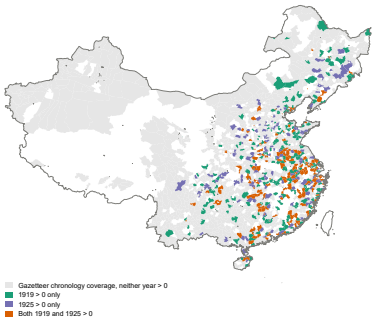


(a) Broad: Party organization (b) Strict: Party organization

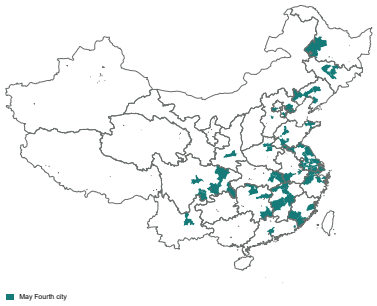
(c) Broad: County party committee (d) Strict: County party committee

Figure A35. Controlling for Japanese Occupation

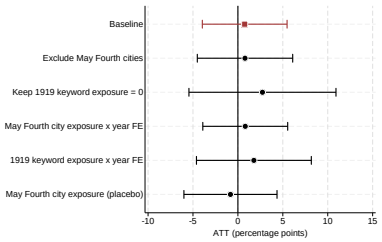
Note: This figure examines whether Japanese occupation confounds the baseline results. We control for the first year in which each county was occupied by Japanese forces and the last year in which Japanese forces withdrew. These Japanese-entry and Japanese-withdrawal cohorts are interacted with year fixed effects, so estimates are identified within cohorts defined by Second Sino-Japanese War timing. Baseline estimates are included for comparison. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



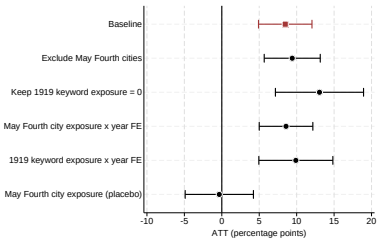
(a) County-level protest keywords, 1919 and 1925



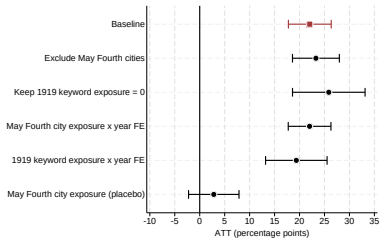
(b) City-level May Fourth exposure



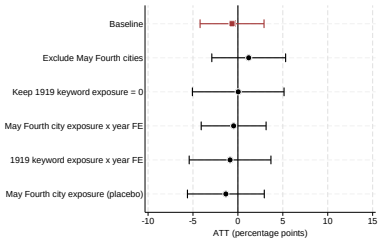
(c) Strict: Party member



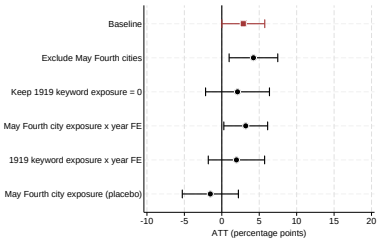
(d) Strict: Party organization



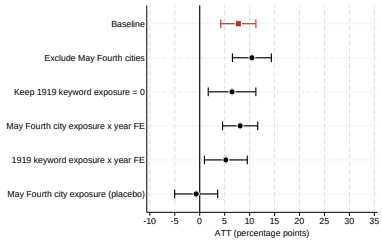
(e) Strict: County party committee



(f) Broad: Party member



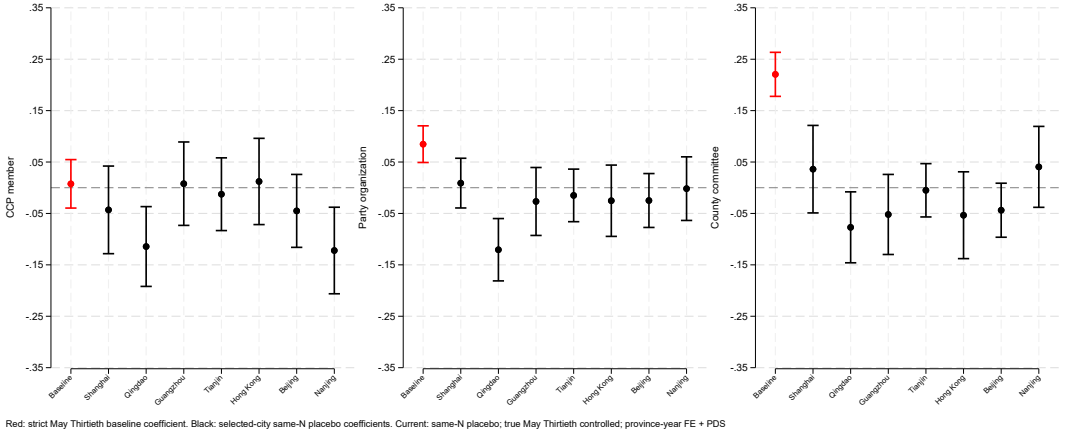
(g) Broad: Party organization



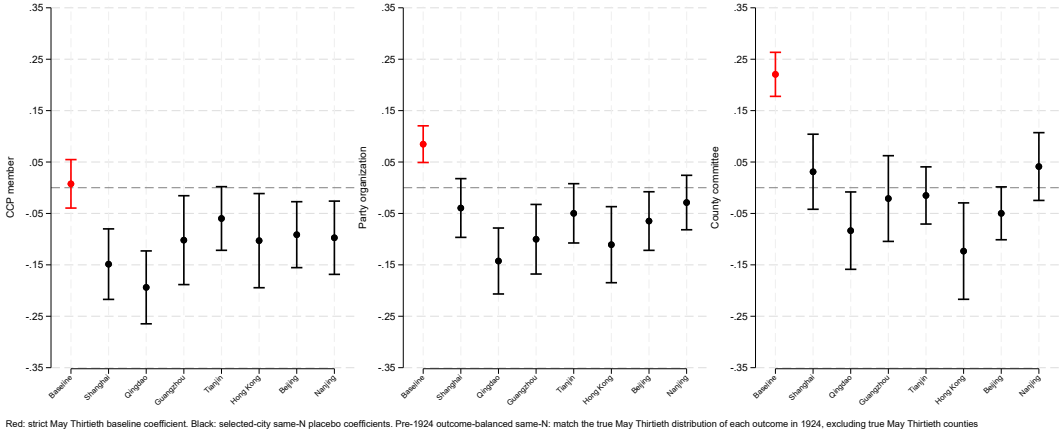
(h) Broad: County party committee

Figure A36. May Fourth Proxy and Party-Development Robustness

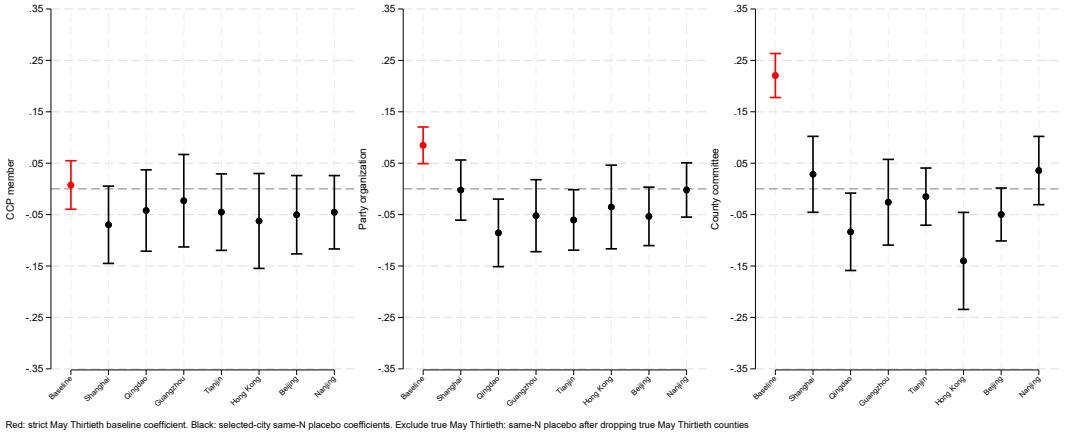
Note: This figure examines whether May Fourth Movement exposure confounds the baseline results. We measure May Fourth exposure using protest-related terms in the 1919 chronicle sections of local gazetteers and a unified city-level indicator. Panel (a) maps county-level overlap in nonzero protest-keyword exposure in 1919 and 1925. Panel (b) maps city-level exposure to the May Fourth Movement, compiled from the Institute of Modern History, CASS (2013). Panels (c)–(h) compare the baseline estimates with five alternative specifications: (1) excluding the identified May Fourth cities; (2) excluding counties with positive protest-keyword frequency in 1919; (3) controlling for interactions between 1919 protest-keyword frequency and year fixed effects; (4) controlling for interactions between the May Fourth city indicator and year fixed effects; and (5) conducting a placebo test that uses May Fourth cities as the treatment group in the baseline specification. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



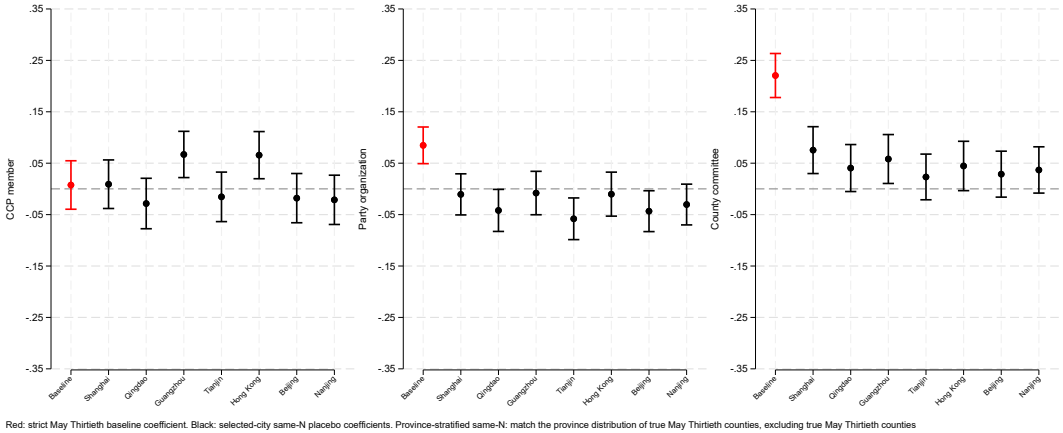
(a) Current specification



(b) Balanced pre-1924 placebo



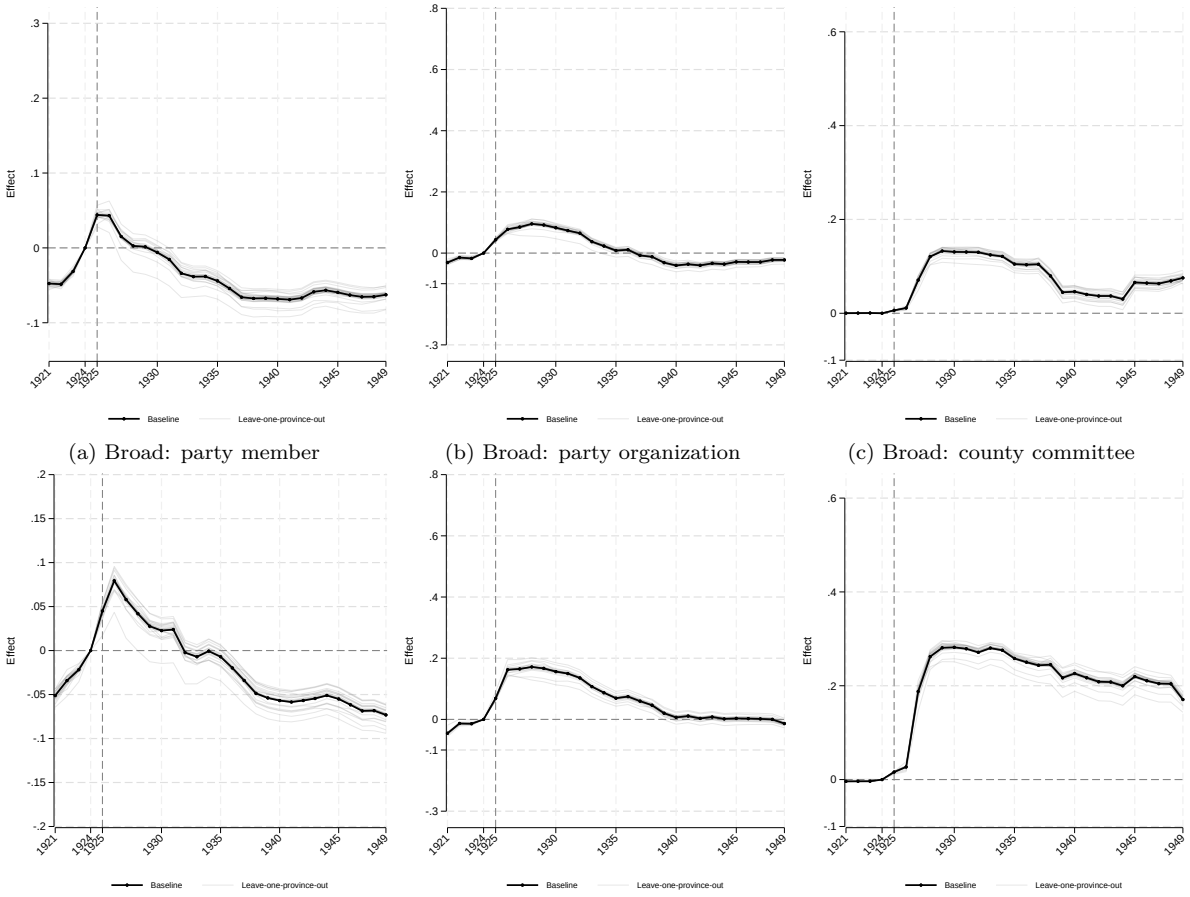
(c) Excluding true May Thirtieth counties



(d) Province-stratified placebo

Figure A37. Placebo Exercises for Spillover Robustness

Note: Each panel plots the actual strict May Thirtieth baseline coefficient in red and same-N selected-city placebo coefficients in black for CCP members, party organizations, and county committees. Panel (a) assigns placebo May Thirtieth counties using counties nearest to each selected city. Panel (b) draws placebo cities after balancing on the pre-1924 outcome distribution, assigning placebo counties near cities in proportions matching pre-1924 May Thirtieth-area member, organization, and county-committee presence. Panel (c) draws placebo exposure after excluding true May Thirtieth counties. Panel (d) draws placebo cities within provinces, assigning counties nearest to selected cities within each province according to the province-specific incidence of May Thirtieth exposure. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

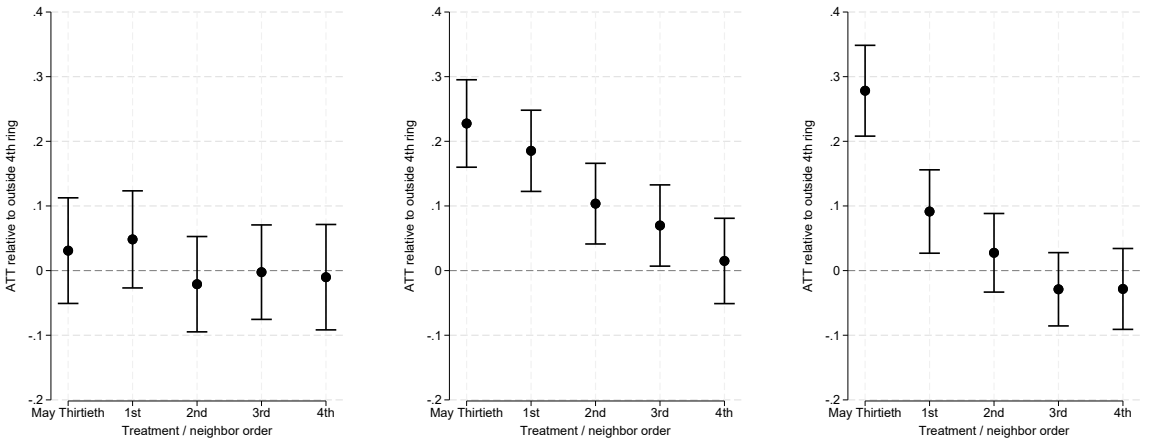


(a) Broad: party member (b) Broad: party organization (c) Broad: county committee

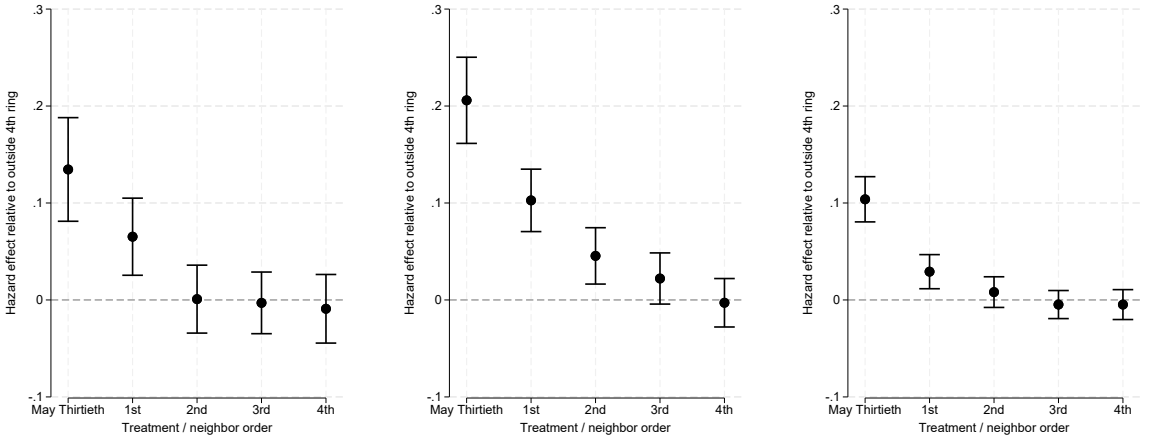
(d) Strict: party member (e) Strict: party organization (f) Strict: county committee

Figure A38. Leave-One-Province-Out Estimates

Note: Gray curves report event-study estimates after dropping one province at a time. Black curves report the baseline event-study estimates. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..



(a) Status: Party member (b) Status: Party organization (c) Status: County party committee



(d) Hazard: Party member (e) Hazard: Party organization (f) Hazard: County party committee

Figure A39. Spatial Spillover Gradient by Neighbor Ring

Note: Each panel reports estimates for May Thirtieth counties and first- through fourth-order neighboring counties, using counties outside the fourth ring as the comparison group. Status panels use cumulative outcome specifications; hazard panels use first-appearance risk-set specifications. Standard errors are clustered by county; intervals are 95% confidence intervals. Back to Section IV.B..

Appendix D. Additional Tables

Table A1. First Party Members and Subsequent Party Organization Formation

	Organization in t		Organization in $t+1$	
	(1)	(2)	(3)	(4)
Panel A. Before 1925				
Party organization	0.081^{***}	0.151^{***}	0.043^{***}	0.007
	(0.013)	(0.017)	(0.012)	(0.017)
Panel B. 1925 and Later				
Party organization	0.330^{***}	0.304^{***}	0.216^{***}	0.225^{***}
	(0.014)	(0.014)	(0.012)	(0.012)
Panel C. Difference				
later - before	0.249^{***}	0.153^{***}	0.173^{***}	0.219^{***}
	(0.019)	(0.024)	(0.017)	(0.023)
First-member-year control	No	Yes	No	Yes
Observations	1669	1669	1515	1515

Note: Columns (1)–(2) use an indicator for whether the county's first party organization appears in the same calendar year as its first party member. Columns (3)–(4) use an indicator for whether the first party organization appears exactly one year after the first party member; these columns drop counties whose first party member appears in 1924, because their $t+1$ outcome falls in 1925. Panels A and B report the organization probability for counties whose first party member appears before 1925 and in 1925 or later, respectively; Panel C reports the difference between the two panels. Columns (2) and (4) control for the first-member year. Standard errors in parentheses are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.A..

Table A2. Organizational Effects of the May Thirtieth Movement in Counties with a CCP Foundation

	(1)	(2)	(3)
Panel A. Broad			
Party organization	0.040**	0.018	0.029
	(0.019)	(0.021)	(0.025)
County committee	0.021***	0.019**	0.021**
	(0.007)	(0.008)	(0.010)
Observations	2,043 / 4,732	2,043 / 4,732	2,043 / 4,732
Panel B. Strict			
Party organization	0.115***	0.093***	0.139***
	(0.032)	(0.031)	(0.031)
County committee	0.073***	0.067***	0.079***
	(0.013)	(0.013)	(0.014)
Observations	1,542 / 3,491	1,542 / 3,491	1,542 / 3,491
Year FE	Yes	Yes	Yes
Province FE	No	No	Yes
PDS controls	No	Yes	Yes
Rain anomaly	No	Yes	Yes

Note: The sample keeps counties that had at least one party member before 1925 but had no party organization and no county party committee before 1925. The dependent variables are annual establishment indicators in the post-1925 period, and the sample is restricted to 1925 and later. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.A..

Table A3. Committee Survival Dynamics in May Thirtieth Areas

	Descriptive statistics			Group differences			
	Strict May Thirtieth (1)	Broad May Thirtieth (2)	Non-May Thirtieth (3)	Strict–Non (4)	Broad–Non (5)	Strict–Non (6)	Broad–Non (7)
Panel A. Abolition proportions among counties with an active county committee in 1927							
1927	0.122	0.156	0.165	-0.043	-0.010	-0.076	-0.031
				(0.047)	(0.045)	(0.053)	(0.052)
1928	0.356	0.356	0.353	0.002	0.002	-0.013	0.009
				(0.066)	(0.059)	(0.071)	(0.064)
1929	0.089	0.096	0.083	0.006	0.014	0.000	0.005
				(0.039)	(0.035)	(0.041)	(0.037)
Province FE				No	No	Yes	Yes
Observations	90	135	133	223	268	223	268
Panel B. Probability of abolition in year $t+k$ after establishment or reestablishment							
t	0.125	0.122	0.084	0.041***	0.038***	0.020	0.023**
				(0.013)	(0.010)	(0.014)	(0.011)
t+1	0.206	0.202	0.177	0.029*	0.025*	0.016	0.015
				(0.017)	(0.013)	(0.019)	(0.015)
t+2	0.118	0.110	0.131	-0.013	-0.020*	-0.034**	-0.035***
				(0.015)	(0.012)	(0.015)	(0.013)
t+3	0.090	0.087	0.082	0.008	0.005	-0.007	0.000
				(0.013)	(0.011)	(0.015)	(0.012)
t+4	0.061	0.050	0.053	0.008	-0.003	0.009	-0.003
				(0.011)	(0.009)	(0.013)	(0.010)
t+5	0.028	0.033	0.027	0.001	0.006	0.005	0.006
				(0.008)	(0.007)	(0.010)	(0.008)
t+6	0.018	0.019	0.025	-0.007	-0.006	-0.001	-0.003
				(0.007)	(0.006)	(0.007)	(0.006)
t+7	0.008	0.009	0.018	-0.011*	-0.010*	-0.009	-0.010*
				(0.006)	(0.005)	(0.006)	(0.006)
t+8	0.015	0.013	0.007	0.008	0.006	0.015*	0.011*
				(0.006)	(0.005)	(0.008)	(0.006)
t+9	0.008	0.008	0.008	0.000	0.000	0.003	0.002
				(0.005)	(0.004)	(0.006)	(0.005)
t+10	0.004	0.006	0.006	-0.002	-0.000	0.000	-0.000
				(0.004)	(0.004)	(0.005)	(0.004)
Province FE				No	No	Yes	Yes
Panel C. Probability of reestablishment in year $t+k$ after abolition							
t	0.351	0.353	0.313	0.038	0.040	0.034	0.056*
				(0.035)	(0.029)	(0.041)	(0.034)
t+1	0.083	0.100	0.140	-0.057**	-0.040*	-0.028	-0.013
				(0.022)	(0.020)	(0.027)	(0.023)
t+2	0.058	0.070	0.081	-0.022	-0.010	0.014	0.003
				(0.019)	(0.017)	(0.020)	(0.018)
t+3	0.074	0.082	0.080	-0.005	0.002	0.013	0.027
				(0.020)	(0.018)	(0.026)	(0.022)
t+4	0.060	0.057	0.066	-0.006	-0.008	0.001	-0.021
				(0.018)	(0.016)	(0.022)	(0.018)
t+5	0.042	0.043	0.061	-0.018	-0.018	-0.008	-0.014
				(0.017)	(0.015)	(0.020)	(0.017)
t+6	0.024	0.033	0.037	-0.013	-0.003	-0.010	-0.000
				(0.013)	(0.013)	(0.014)	(0.014)
t+7	0.089	0.068	0.053	0.036*	0.015	0.003	-0.006
				(0.021)	(0.017)	(0.024)	(0.018)
t+8	0.057	0.072	0.042	0.015	0.029*	-0.011	0.001
				(0.018)	(0.016)	(0.020)	(0.017)
t+9	0.051	0.036	0.028	0.023	0.008	0.020	0.008
				(0.017)	(0.013)	(0.017)	(0.014)
t+10	0.038	0.049	0.050	-0.012	-0.001	-0.023	-0.001
				(0.018)	(0.017)	(0.022)	(0.021)
Province FE				No	No	Yes	Yes

Note: Panel A reports the post-1927 abolition dynamics for counties with active county committees in 1927. Panel B reports the probability of abolition in each year after establishment or reestablishment. Panel C reports the probability of reestablishment in each year after abolition. Columns (1)–(3) report raw proportions. Columns (4)–(7) report LPM differences; columns (6)–(7) control for province FE. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Sections IV.B. and V.C..

Table A4. Minguo Association/Party Establishments

	asinh(Count)			log(1+Count)		
	(1)	(2)	(3)	(4)	(5)	(6)
1922-1924	0.009	0.003	0.003	0.007	0.003	0.002
	(0.014)	(0.019)	(0.018)	(0.011)	(0.015)	(0.014)
1925	0.013	0.015	0.018	0.010	0.012	0.014
	(0.015)	(0.024)	(0.023)	(0.012)	(0.019)	(0.017)
1926	-0.031**	-0.041*	-0.041*	-0.025**	-0.033**	-0.033**
	(0.015)	(0.021)	(0.021)	(0.012)	(0.017)	(0.017)
1927	0.016	-0.013	-0.018	0.013	-0.011	-0.016
	(0.025)	(0.024)	(0.023)	(0.021)	(0.019)	(0.019)
1928	0.005	-0.039	-0.042	0.003	-0.033	-0.035
	(0.027)	(0.030)	(0.028)	(0.022)	(0.024)	(0.023)
1929-1936	-0.005	-0.021	-0.027	-0.005	-0.019	-0.023
	(0.030)	(0.028)	(0.025)	(0.024)	(0.023)	(0.022)
1937-1945	-0.029	-0.033	-0.043	-0.024	-0.027	-0.035
	(0.031)	(0.040)	(0.039)	(0.025)	(0.033)	(0.032)
1946	0.016	-0.012	-0.021	0.015	-0.012	-0.019
	(0.047)	(0.040)	(0.039)	(0.039)	(0.032)	(0.031)
1947	0.020	-0.007	-0.008	0.017	-0.008	-0.009
	(0.044)	(0.038)	(0.039)	(0.037)	(0.031)	(0.032)
1948	-0.018	-0.049	-0.049*	-0.015	-0.040	-0.041
	(0.023)	(0.030)	(0.029)	(0.019)	(0.025)	(0.025)
1949	-0.030**	-0.029	-0.026	-0.025**	-0.026	-0.023
	(0.014)	(0.022)	(0.021)	(0.012)	(0.019)	(0.018)
ATT 1925-1949	-0.020	-0.028	-0.033	-0.016	-0.024	-0.028
	(0.021)	(0.027)	(0.027)	(0.018)	(0.023)	(0.022)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	189	187	187	189	187	187
Control cities	117	115	115	117	115	115

Note: This table is estimated using city-level panel regressions. We obtain year-by-year information on the establishment of political parties and associations from Wang and Yang (1989). Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.B..

Table A5. Fake May Thirtieth Randomization Check

	Party organization (1)	County committee (2)
Panel A. Outcome status		
Actual coefficient	0.189	0.175
Placebo mean	-0.000	0.001
Placebo s.d.	0.015	0.015
Placebo p95	0.027	0.026
Actual percentile	1.000	1.000
Panel B. Annual first-appearance probability		
Actual coefficient	0.160	0.066
Placebo mean	0.011	0.007
Placebo s.d.	0.010	0.005
Placebo p95	0.027	0.015
Actual percentile	1.000	1.000
Repetitions	200	200
Observations, Panel A	66,468	66,468
Counties, Panel A	2,292	2,292
Observations, Panel B	32,515	47,217
Counties, Panel B	2,292	2,292

Note: Using the share of strict May Thirtieth counties in the original data, this table randomly reassigns treatment status and reports the distribution of placebo coefficients. Back to Section IV.B..

Table A6. Elite Individuals' CCP Join Year

	asinh(Count)			log(1+Count)		
	(1)	(2)	(3)	(4)	(5)	(6)
1922-1924	0.055	0.028	0.034	0.042	0.021	0.026
	(0.037)	(0.044)	(0.046)	(0.028)	(0.034)	(0.035)
1925	0.259***	0.229***	0.219***	0.200***	0.176***	0.167***
	(0.055)	(0.072)	(0.080)	(0.043)	(0.055)	(0.061)
1926	0.160***	0.120	0.096	0.124***	0.093	0.074
	(0.054)	(0.076)	(0.083)	(0.042)	(0.059)	(0.064)
1927	0.162***	0.095	0.089	0.126***	0.074	0.070
	(0.056)	(0.069)	(0.073)	(0.043)	(0.053)	(0.056)
1928	0.060	0.044	0.033	0.046	0.034	0.026
	(0.043)	(0.064)	(0.066)	(0.033)	(0.050)	(0.051)
1929-1936	0.065*	0.003	-0.001	0.051*	0.002	-0.001
	(0.039)	(0.044)	(0.045)	(0.030)	(0.034)	(0.035)
1937-1945	0.004	-0.057	-0.048	0.003	-0.044	-0.037
	(0.038)	(0.045)	(0.045)	(0.029)	(0.035)	(0.034)
1946	-0.039	-0.082*	-0.081	-0.030	-0.064*	-0.063
	(0.045)	(0.050)	(0.051)	(0.035)	(0.038)	(0.040)
1947	-0.012	-0.065	-0.053	-0.009	-0.050	-0.041
	(0.041)	(0.046)	(0.046)	(0.032)	(0.035)	(0.036)
1948	-0.008	-0.045	-0.036	-0.006	-0.035	-0.027
	(0.044)	(0.055)	(0.055)	(0.034)	(0.043)	(0.042)
1949	0.012	-0.052	-0.041	0.009	-0.040	-0.031
	(0.042)	(0.047)	(0.048)	(0.033)	(0.036)	(0.037)
ATT 1925-1949	0.005	-0.031	-0.034	0.004	-0.024	-0.026
	(0.021)	(0.024)	(0.025)	(0.016)	(0.019)	(0.019)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	189	187	187	189	187	187
Control cities	117	115	115	117	115	115

Note: This table examines the timing of CCP entry among notable individuals. The data come from the *Minguo Biographical Dictionary* (Xu, 2007). Regressions use city-level panel specifications. The omitted baseline period is 1921–1924; period coefficients are treatment-period interactions relative to 1921–1924. Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.B..

Table A7. Elite Individuals' CCP Join Year by Pre-Join Occupation Category

	Students	Peasants	Workers	Soldiers	Gov/party	Cultural workers	Commerce
Panel A. asinh(Count)							
1925	0.112**	0.032	0.028	0.011	-0.002	0.051*	-0.013
	(0.052)	(0.021)	(0.022)	(0.035)	(0.008)	(0.026)	(0.013)
1926	0.010	0.044**	0.015	0.046**	-0.010	-0.005	-0.000
	(0.062)	(0.019)	(0.021)	(0.023)	(0.019)	(0.036)	(0.001)
1927	0.057	-0.005	0.018	0.038**	-0.003	0.010	-0.000
	(0.042)	(0.034)	(0.016)	(0.018)	(0.019)	(0.019)	(0.002)
1928	-0.006	0.034	0.003	-0.019	-0.002	-0.006	-0.001
	(0.035)	(0.035)	(0.010)	(0.018)	(0.009)	(0.019)	(0.001)
Panel B. log(1+Count)							
1925	0.087**	0.025	0.023	0.009	-0.002	0.040*	-0.010
	(0.040)	(0.017)	(0.017)	(0.028)	(0.006)	(0.021)	(0.010)
1926	0.007	0.034**	0.012	0.036**	-0.008	-0.004	-0.000
	(0.048)	(0.014)	(0.016)	(0.018)	(0.015)	(0.028)	(0.001)
1927	0.043	-0.004	0.014	0.030**	-0.003	0.008	-0.000
	(0.033)	(0.026)	(0.012)	(0.014)	(0.015)	(0.015)	(0.002)
1928	-0.005	0.026	0.002	-0.015	-0.002	-0.005	-0.001
	(0.027)	(0.027)	(0.008)	(0.014)	(0.007)	(0.015)	(0.001)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
PDS controls × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rain anomaly	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,416	2,416	2,416	2,416	2,416	2,416	2,416
Cities	302	302	302	302	302	302	302
Treated cities	187	187	187	187	187	187	187
Control cities	115	115	115	115	115	115	115

Note: This table examines the timing of CCP entry among notable individuals by occupation. The data come from the *Minguo Biographical Dictionary* (Xu, 2007). Regressions use city-level panel specifications. The omitted baseline period is 1921–1924; period coefficients are treatment-period interactions relative to 1921–1924. We use this pooled pre-1925 baseline because too few individuals in the disaggregated occupation sample joined the CCP in 1921 to support precise single-year estimates. Occupation categories are recoded as pre-entry social and economic occupations. Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.B..

Table A8. Martyr Party-Join Year

	asinh(Count)			log(1+Count)		
	(1)	(2)	(3)	(4)	(5)	(6)
1922-1924	0.039*	0.032	0.032	0.030*	0.025	0.025
	(0.020)	(0.022)	(0.022)	(0.016)	(0.017)	(0.017)
1925	0.121***	0.099**	0.099**	0.094***	0.077**	0.077**
	(0.041)	(0.043)	(0.043)	(0.032)	(0.033)	(0.034)
1926	0.153***	0.118**	0.121**	0.120***	0.092**	0.094**
	(0.048)	(0.049)	(0.049)	(0.037)	(0.038)	(0.038)
1927	0.162***	0.120**	0.123***	0.126***	0.093**	0.095***
	(0.047)	(0.047)	(0.047)	(0.036)	(0.037)	(0.037)
1928	0.070*	0.013	0.013	0.055*	0.010	0.010
	(0.036)	(0.035)	(0.034)	(0.029)	(0.027)	(0.027)
1929-1936	0.006	-0.012	-0.013	0.004	-0.010	-0.011
	(0.016)	(0.019)	(0.019)	(0.012)	(0.015)	(0.015)
1937-1945	0.044	0.021	0.018	0.035	0.017	0.015
	(0.038)	(0.039)	(0.039)	(0.031)	(0.031)	(0.031)
1946	0.029	0.013	0.012	0.025	0.012	0.011
	(0.052)	(0.054)	(0.054)	(0.042)	(0.044)	(0.044)
1947	0.037	0.046	0.045	0.031	0.038	0.037
	(0.050)	(0.053)	(0.054)	(0.040)	(0.043)	(0.043)
1948	-0.003	-0.000	0.001	-0.001	0.001	0.002
	(0.046)	(0.050)	(0.051)	(0.037)	(0.040)	(0.041)
1949	0.016	0.027	0.027	0.012	0.020	0.020
	(0.041)	(0.047)	(0.047)	(0.032)	(0.037)	(0.037)
ATT 1925-1949	0.012	-0.003	-0.004	0.010	-0.002	-0.003
	(0.026)	(0.026)	(0.026)	(0.021)	(0.020)	(0.020)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	23,200	23,171	23,171	23,200	23,171	23,171
Counties	800	799	799	800	799	799
Treated counties	172	172	172	172	172	172
Control counties	628	627	627	628	627	627

Note: This table uses martyr records for which the specific timing of party entry can be distinguished and aggregates them at the county level. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section IV.B..

Table A11. May Thirtieth Movement and Strict-Definition Red Books

	asinh(Count)			log(1+Count)		
	(1)	(2)	(3)	(4)	(5)	(6)
1922-1924	-0.003	-0.012	-0.010	-0.002	-0.009	-0.007
	(0.010)	(0.016)	(0.014)	(0.008)	(0.013)	(0.011)
1925	0.042**	0.049	0.050	0.034**	0.042	0.042
	(0.018)	(0.033)	(0.032)	(0.015)	(0.029)	(0.028)
1926	0.083***	0.124**	0.111**	0.068***	0.103**	0.093**
	(0.029)	(0.058)	(0.054)	(0.024)	(0.050)	(0.047)
1927	0.054	0.112*	0.095*	0.046	0.094*	0.080*
	(0.036)	(0.058)	(0.055)	(0.029)	(0.048)	(0.045)
1928	-0.010	-0.051	-0.050	-0.005	-0.040	-0.039
	(0.030)	(0.038)	(0.037)	(0.024)	(0.030)	(0.029)
1929-1936	0.035	-0.010	-0.013	0.030	-0.007	-0.009
	(0.023)	(0.033)	(0.031)	(0.018)	(0.026)	(0.024)
1937-1945	0.114*	0.051	0.038	0.090*	0.042	0.029
	(0.063)	(0.072)	(0.075)	(0.052)	(0.059)	(0.062)
1946	0.447***	0.175	0.145	0.366***	0.141	0.116
	(0.122)	(0.158)	(0.154)	(0.102)	(0.133)	(0.129)
1947	0.500***	0.136	0.161	0.413***	0.111	0.133
	(0.130)	(0.157)	(0.152)	(0.110)	(0.133)	(0.129)
1948	0.633***	0.244	0.290*	0.521***	0.196	0.235
	(0.144)	(0.175)	(0.172)	(0.122)	(0.148)	(0.146)
1949	0.835***	0.505***	0.463**	0.700***	0.423***	0.385**
	(0.158)	(0.181)	(0.179)	(0.135)	(0.154)	(0.153)
ATT 1925-1949	0.158***	0.076	0.068	0.129***	0.062	0.055
	(0.042)	(0.053)	(0.052)	(0.035)	(0.043)	(0.043)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	189	187	187	189	187	187
Control cities	117	115	115	117	115	115

Note: This table uses a city-level panel. A city is coded as treated if at least one county within the city participated in the May Thirtieth Movement. Strict-definition red books include books on communism and socialism. Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Sections V.A. and VI.A..

Table A12. May Thirtieth Exposure and Red Books: Alternative Specifications

Panel A. Broad red books: city-level broad May Thirtieth exposure

	asinh(Count)			log(1+Count)		
1922-1924	-0.014 (0.024)	-0.028 (0.041)	-0.021 (0.039)	-0.010 (0.018)	-0.020 (0.031)	-0.015 (0.030)
1925	0.088** (0.039)	0.102 (0.064)	0.110* (0.065)	0.073** (0.032)	0.086 (0.055)	0.092* (0.055)
1926	0.139** (0.064)	0.199* (0.108)	0.187* (0.102)	0.116** (0.055)	0.172* (0.095)	0.162* (0.090)
1927	0.182** (0.079)	0.269** (0.116)	0.220** (0.103)	0.152** (0.065)	0.227** (0.100)	0.188** (0.090)
1928	-0.000 (0.063)	-0.077 (0.061)	-0.095 (0.065)	0.002 (0.052)	-0.059 (0.047)	-0.073 (0.050)
1929-1936	0.066 (0.051)	-0.005 (0.053)	-0.019 (0.051)	0.054 (0.040)	-0.004 (0.041)	-0.014 (0.040)
1937-1945	0.287** (0.117)	0.377** (0.152)	0.318** (0.143)	0.234** (0.098)	0.315** (0.129)	0.265** (0.120)
1946	0.751*** (0.205)	0.746*** (0.265)	0.657** (0.258)	0.621*** (0.174)	0.622*** (0.226)	0.548** (0.220)
1947	0.754*** (0.211)	0.683** (0.263)	0.630** (0.255)	0.623*** (0.179)	0.568** (0.226)	0.524** (0.218)
1948	0.922*** (0.238)	0.954*** (0.295)	0.929*** (0.297)	0.772*** (0.203)	0.807*** (0.253)	0.788*** (0.255)
1949	1.415*** (0.266)	1.372*** (0.301)	1.252*** (0.300)	1.212*** (0.231)	1.175*** (0.260)	1.073*** (0.258)
ATT 1925-1949	0.305*** (0.077)	0.325*** (0.099)	0.280*** (0.092)	0.252*** (0.064)	0.271*** (0.082)	0.234*** (0.076)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls×year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	5,452	5,365	5,365	5,452	5,365	5,365
Cities	188	185	185	188	185	185
Treated cities	72	72	72	72	72	72
Control cities	116	113	113	116	113	113

Panel B. Strict red books: city-level broad May Thirtieth exposure

	asinh(Count)			log(1+Count)		
1922-1924	-0.021 (0.023)	-0.025 (0.042)	-0.020 (0.040)	-0.016 (0.018)	-0.018 (0.032)	-0.014 (0.031)
1925	0.045 (0.035)	0.076 (0.062)	0.084 (0.063)	0.038 (0.029)	0.066 (0.054)	0.072 (0.054)
1926	0.121** (0.059)	0.181* (0.100)	0.171* (0.095)	0.100** (0.050)	0.156* (0.088)	0.147* (0.084)
1927	0.133* (0.068)	0.235** (0.107)	0.189** (0.095)	0.111* (0.057)	0.198** (0.092)	0.162** (0.082)
1928	-0.009 (0.064)	-0.082 (0.062)	-0.100 (0.066)	-0.004 (0.052)	-0.062 (0.048)	-0.077 (0.051)
1929-1936	0.026 (0.048)	-0.043 (0.061)	-0.053 (0.059)	0.022 (0.037)	-0.032 (0.047)	-0.040 (0.045)
1937-1945	0.201** (0.100)	0.260** (0.131)	0.221* (0.122)	0.162* (0.083)	0.217** (0.109)	0.183* (0.101)
1946	0.631*** (0.191)	0.639** (0.250)	0.573** (0.244)	0.521*** (0.161)	0.531** (0.212)	0.478** (0.207)
1947	0.681*** (0.201)	0.614** (0.253)	0.568** (0.244)	0.563*** (0.169)	0.512** (0.215)	0.474** (0.207)
1948	0.852*** (0.228)	0.903*** (0.283)	0.877*** (0.287)	0.710*** (0.194)	0.759*** (0.242)	0.740*** (0.245)
1949	1.360*** (0.256)	1.318*** (0.289)	1.197*** (0.287)	1.159*** (0.221)	1.122*** (0.248)	1.019*** (0.246)
ATT 1925-1949	0.249*** (0.071)	0.254*** (0.090)	0.220*** (0.084)	0.205*** (0.058)	0.212*** (0.074)	0.184*** (0.069)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls×year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	5,452	5,365	5,365	5,452	5,365	5,365
Cities	188	185	185	188	185	185
Treated cities	72	72	72	72	72	72
Control cities	116	113	113	116	113	113

Panel C. Broad red books: broad+strict county-share intensity

	asinh(Count)			log(1+Count)		
1922-1924	-0.013 (0.021)	-0.023 (0.032)	-0.017 (0.029)	-0.010 (0.016)	-0.017 (0.025)	-0.013 (0.023)
1925	0.110*** (0.042)	0.125** (0.058)	0.113** (0.050)	0.089*** (0.034)	0.102** (0.048)	0.091** (0.041)
1926	0.142** (0.065)	0.168* (0.089)	0.122 (0.076)	0.118** (0.054)	0.143* (0.076)	0.105 (0.065)
1927	0.190*** (0.072)	0.265*** (0.094)	0.199** (0.078)	0.156*** (0.059)	0.218*** (0.080)	0.164** (0.065)
1928	0.008 (0.056)	-0.053 (0.045)	-0.066 (0.043)	0.009 (0.046)	-0.040 (0.035)	-0.051 (0.033)
1929-1936	0.069 (0.046)	0.019 (0.044)	0.002 (0.040)	0.056 (0.037)	0.015 (0.034)	0.002 (0.031)
1937-1945	0.265** (0.105)	0.320** (0.135)	0.245* (0.127)	0.217** (0.088)	0.268** (0.113)	0.206* (0.107)
1946	0.545*** (0.190)	0.482** (0.228)	0.407* (0.225)	0.451*** (0.161)	0.403** (0.194)	0.338* (0.192)
1947	0.507** (0.198)	0.476** (0.214)	0.459** (0.214)	0.415** (0.168)	0.394** (0.183)	0.380** (0.182)
1948	0.626*** (0.222)	0.602** (0.258)	0.593** (0.258)	0.524*** (0.189)	0.511** (0.222)	0.504** (0.223)
1949	1.297*** (0.250)	1.130*** (0.277)	1.029*** (0.270)	1.111*** (0.217)	0.968*** (0.240)	0.879*** (0.233)
ATT 1925-1949	0.264*** (0.070)	0.266*** (0.084)	0.216*** (0.078)	0.218*** (0.058)	0.222*** (0.070)	0.180*** (0.065)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls×year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	190	189	189	190	189	189
Control cities	116	113	113	116	113	113

Panel D. Strict red books: broad+strict county-share intensity

	asinh(Count)			log(1+Count)		
1922-1924	-0.019 (0.021)	-0.024 (0.033)	-0.019 (0.030)	-0.015 (0.016)	-0.017 (0.025)	-0.014 (0.023)
1925	0.072* (0.040)	0.098* (0.056)	0.083* (0.047)	0.058* (0.032)	0.080* (0.046)	0.069* (0.039)
1926	0.121** (0.059)	0.148* (0.081)	0.106 (0.069)	0.100** (0.049)	0.125* (0.069)	0.091 (0.059)
1927	0.140** (0.061)	0.220** (0.085)	0.163** (0.070)	0.116** (0.051)	0.183** (0.072)	0.136** (0.059)
1928	-0.004 (0.056)	-0.049 (0.046)	-0.062 (0.043)	-0.001 (0.046)	-0.037 (0.036)	-0.048 (0.033)
1929-1936	0.028 (0.043)	-0.020 (0.047)	-0.033 (0.042)	0.024 (0.034)	-0.014 (0.036)	-0.025 (0.033)
1937-1945	0.185** (0.089)	0.228** (0.113)	0.174 (0.106)	0.150** (0.073)	0.189** (0.094)	0.144 (0.088)
1946	0.426** (0.177)	0.396* (0.214)	0.346 (0.213)	0.353** (0.148)	0.332* (0.182)	0.287 (0.180)
1947	0.460** (0.187)	0.421** (0.204)	0.409** (0.204)	0.379** (0.157)	0.352** (0.173)	0.340** (0.172)
1948	0.585*** (0.212)	0.573** (0.249)	0.566** (0.250)	0.486*** (0.180)	0.484** (0.212)	0.479** (0.214)
1949	1.261*** (0.241)	1.104*** (0.266)	0.996*** (0.258)	1.075*** (0.208)	0.939*** (0.229)	0.844*** (0.222)
ATT 1925-1949	0.213*** (0.064)	0.210*** (0.075)	0.171** (0.069)	0.175*** (0.052)	0.175*** (0.062)	0.142** (0.057)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province×Year FE	No	Yes	Yes	No	Yes	Yes
PDS controls×year	No	No	Yes	No	No	Yes
Rain anomaly	No	No	Yes	No	No	Yes
Observations	8,874	8,758	8,758	8,874	8,758	8,758
Cities	306	302	302	306	302	302
Treated cities	190	189	189	190	189	189
Control cities	116	113	113	116	113	113

Note: We conduct two additional tests. (1) We retain as treated only cities mentioned in *Historical Materials on the May Thirtieth Movement* and exclude cities in which only a few counties participated in the movement. (2) We combine the strict and broad definitions of May Thirtieth exposure and use the share of participating counties within each city as a measure of treatment intensity. Cells report coefficients with clustered standard errors in parentheses. Standard errors are clustered by city. * p<0.10, ** p<0.05, *** p<0.01. Back to Section V.A..

Table A13. May Thirtieth Movement and Natural Science Books

Period	CLC science category					
	Natural Sci. General	Math./ Phys./Chem.	Astron./ Earth Sci.	Biological Sci.	Agricultural Sci.	Trans- portation
Panel A. Inverse hyperbolic sine						
1922–1924	-0.001	-0.002	0.004	0.015	-0.012	0.007
	(0.004)	(0.007)	(0.005)	(0.014)	(0.010)	(0.004)
1925	-0.004	0.002	0.021	0.015	0.012	0.008
	(0.008)	(0.002)	(0.017)	(0.014)	(0.015)	(0.008)
1926	-0.001	0.000	0.020	0.028	0.003	0.010
	(0.002)	(0.003)	(0.017)	(0.021)	(0.013)	(0.009)
1927	0.000	0.002	0.020	0.023	-0.017	0.004
	(0.010)	(0.003)	(0.015)	(0.022)	(0.017)	(0.009)
1928	0.001	-0.004	0.078	0.057	0.046	0.013
	(0.003)	(0.003)	(0.050)	(0.044)	(0.031)	(0.008)
Panel B. Log plus one						
1922–1924	-0.001	-0.001	0.003	0.012	-0.009	0.006
	(0.003)	(0.005)	(0.004)	(0.011)	(0.008)	(0.004)
1925	-0.003	0.002	0.016	0.012	0.008	0.007
	(0.006)	(0.002)	(0.013)	(0.011)	(0.012)	(0.006)
1926	-0.001	0.000	0.015	0.021	0.002	0.008
	(0.001)	(0.003)	(0.014)	(0.017)	(0.010)	(0.007)
1927	0.000	0.002	0.016	0.019	-0.014	0.003
	(0.008)	(0.002)	(0.012)	(0.019)	(0.013)	(0.007)
1928	0.000	-0.003	0.063	0.045	0.036	0.010
	(0.002)	(0.003)	(0.040)	(0.034)	(0.024)	(0.007)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Province × year FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS controls × year	Yes	Yes	Yes	Yes	Yes	Yes
Rain anomaly	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,758	8,758	8,758	8,758	8,758	8,758
Cities	302	302	302	302	302	302
Treated cities	187	187	187	187	187	187
Control cities	115	115	115	115	115	115

Note: This table examines whether May Thirtieth areas experienced clear post-movement growth in other book categories that contain few or no red books. Natural Sci. General denotes Natural Sciences: General; Math./Phys./Chem. denotes Mathematics, Physics, and Chemistry; Astron./Earth Sci. denotes Astronomy and Earth Sciences; Biological Sci. denotes Biological Sciences; Agricultural Sci. denotes Agricultural Sciences; Transportation denotes Transportation. Standard errors are clustered by city. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section V.A..

Table A14. Examples of Worker and Entrepreneur Rosters

Workers	Entrepreneurs
Membership Directory of the Shanghai Shenbao Employees' Federation	Shanghai Directory of All Trades
Membership Correspondence Directory of the Shanghai Postal Workers' Union	Directory of Chinese Commercial Firms in Shanghai (Chinese-English)
Special Issue of the First Membership Delegates' Congress of the National Federation of Salt Industry Unions of the Republic of China	Shanghai Commercial Directory
Register of Chinese Seamen (No. 3)	Directory of Shanghai Domestic-Goods Manufacturers
Membership Directory of the Taiwan Provincial Railway Union	Directory of Directors and Supervisors of Shanghai Trade Associations
Membership Directory of the Ping-Han Railway Union	Membership Roster of the Shanghai Chamber of Commerce
Membership Directory of the Zhejiang Postal Workers' Union	Industrial and Commercial Firms Directory of Shanghai and the Nation
Membership Directory of the Yue-Han District Railway Union (Dashatou Branch)	Directory of Shanghai Import and Export Trading Firms (1948-1949, Chinese-English)
Membership Directory of the Zhuzhou Machine Works Union, Central Railway Workshop	Directory of Chinese Commercial Firms (1947)
List of Factories in Lanzhou	Membership Roster of the Chinese Manufacturers' Association of Hong Kong

Note: This table lists examples of the worker rosters and entrepreneur-name sources used to construct the name-training data. Worker sources mainly record union members and railway or factory workers, while entrepreneur sources come from commercial directories and chamber-of-commerce rosters. Back to Section V.A..

Table A15. May Thirtieth Areas and Worker Name Score

	Strict May Thirtieth			Broad May Thirtieth		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Current baseline sample						
1925–1928	0.162***	0.141**	0.150**	0.142***	0.152***	0.166***
	(0.061)	(0.068)	(0.071)	(0.043)	(0.049)	(0.053)
1929–1936	0.124*	0.108	0.096	0.078	0.085	0.087
	(0.064)	(0.080)	(0.081)	(0.047)	(0.060)	(0.063)
1937–1945	0.149*	0.161	0.149	0.108*	0.135*	0.138
	(0.084)	(0.106)	(0.117)	(0.064)	(0.081)	(0.087)
1946–1949	0.189**	0.213**	0.209*	0.155**	0.144*	0.150
	(0.096)	(0.103)	(0.119)	(0.073)	(0.083)	(0.096)
Panel B. Mean over 200 random book perturbations						
1925–1928	0.158	0.136	0.145	0.139	0.148	0.161
1929–1936	0.122	0.106	0.094	0.076	0.084	0.085
1937–1945	0.146	0.154	0.143	0.105	0.129	0.132
1946–1949	0.186	0.208	0.201	0.155	0.147	0.149
Panel A observations	5,182	5,084	5,084	6,652	6,562	6,562
Panel B mean observations	5,189	5,091	5,091	6,662	6,572	6,572
Panel A mean dep. var.	4.566	4.565	4.565	4.566	4.566	4.566
Panel B mean dep. var.	4.559	4.558	4.558	4.558	4.558	4.558
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
Province-year FE	No	Yes	Yes	No	Yes	Yes
PDS controls × year	No	No	Yes	No	No	Yes

Note: The dependent variable is the inverse hyperbolic sine of the worker-peasant name score. Treatment-year interactions for birth years 1921–1924 are omitted, so reported rows are relative to the pooled 1921–1924 baseline. Panel A reproduces the current candidate. Panel B averages 200 random perturbations that randomly add or remove worker rosters or entrepreneur-name rosters. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section V.A..

Table A16. May Thirtieth Counties, County Committees, and CCP Capture Timing

	Broad May Thirtieth		Strict May Thirtieth	
	(1)	(2)	(3)	(4)
Panel A. May Thirtieth counties and active county committees in 1945				
May Thirtieth county	0.158***	0.055**	0.243***	0.157***
	(0.047)	(0.027)	(0.056)	(0.028)
Outcome mean	0.329	0.329	0.317	0.317
Observations	2,292	2,292	1,844	1,844
Panel B. Active county committees in 1945 and CCP capture timing				
Active county committee in 1945	-2.439***	-1.294***	-2.738***	-1.474***
	(0.430)	(0.496)	(0.427)	(0.557)
Outcome mean	1946.987	1946.987	1947.022	1947.022
Observations	1,936	1,936	1,566	1,566
Province FE	No	Yes	No	Yes

Note: This table examines two correlations: the relationship between May Thirtieth areas and the presence of county committees in the 1945 cross section, and the relationship between county-committee presence and CCP capture timing. Standard errors are clustered by county. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Back to Section VI.A..